

边界镜像对称延拓双正交小波变换矩阵的构造

杨爱萍 侯正信 王成优

(天津大学电子信息工程学院, 天津 300072)

摘要 计算小波变换的 Mallat 算法需要进行逐级分解和重构, 对于有限长信号的小波变换来说, 为了保证其完全重构, 有必要对其进行边界延拓。基于边界周期延拓的小波变换算法极易实现, 也常见于文献, 而边界对称延拓较周期延拓则更适合用于信号和图像的处理, 但基于边界对称延拓的小波变换矩阵实现方法却很少出现在文献中。为了用矩阵-向量乘积实现信号的小波变换, 给出了一种在信号镜像对称延拓方式下, 任意深度小波变换矩阵的构造方法, 并证明了该延拓方式下实现 Mallat 算法的完全重构条件。作为实例, 绘出了 Bior3.3 小波的分解和重构矩阵的基向量及波形图。将构造的变换矩阵用于基于小波的图像处理中, 不仅可以避免逐级迭代, 大大简化运算量, 而且边界效应也明显减少。

关键词 小波变换 镜像对称延拓 双正交小波 Mallat 算法

中图分类号: TN911.73 TP391.41 文献标识码: A 文章编号: 1006-8961(2008)02-0198-06

Construction of Biorthogonal Wavelet Transform Matrices with Mirror-symmetric Boundary-extension

YANG Aiping HOU Zhengxin WANG Chengyou

(School of Electronic Information Engineering, Tianjin University, Tianjin 300072)

Abstract Iterative decomposition and reconstruction are needed in Mallat algorithm. In order to realize perfect reconstruction, finite-length signals must be extended to some extent before they can be transformed. The algorithm based on periodic boundary-extension always can be seen in the literature. Symmetric boundary-extension has better performance than periodic method in image processing, whereas the matrix transform method based on symmetric boundary-extension is seldom mentioned in the literature. A method of constructing decomposition and reconstruction matrices with arbitrary wavelet transform depth in mirror-symmetric boundary-extension is proposed for wavelet transform in matrix-vector multiplication, and the condition for perfect reconstruction of Mallat algorithm is proved. As an example, the base vectors and base graphs of Bior3.3 wavelet were given. The application of wavelet transform matrices in the wavelet-based image processing can avoid iterative operation, simplify the calculation and meanwhile reduce the edge effect evidently.

Keywords wavelet transform, mirror-symmetric extension, biorthogonal wavelet, Mallat algorithm

1 引言

为了利用 Mallat 算法对有限长信号进行小波变换, 需要对信号边界进行延拓。无论周期延拓还是对称延拓, 由于实现小波变换都需要逐级迭代, 计算量大, 因此, 文献 [1] 构造了周期延拓下任意分解深度的

小波变换矩阵, 使得小波变换可以通过矩阵-向量乘积实现。虽然周期延拓容易实现 Mallat 算法, 但会产生明显的边界效应^[2]。而对称延拓则可减少 Gibbs 效应^[3], 更适合用于信号、图像的处理。但基于边界对称延拓的小波变换矩阵算法很少出现在文献中。文献 [4] 构造了信号整点对称延拓下的双正交小波变换矩阵, 用于 FRIT (finite ridgelet transform)^[5] 图像

基金项目: 国家自然科学基金项目 (60302018); 高等学校博士学科点专项科研基金资助课题 (20060056051)

收稿日期: 2006-07-21; 改回日期: 2006-10-10

第一作者简介: 杨爱萍 (1977~), 女。现为天津大学电子信息工程学院讲师、博士生。主要研究方向为数字图像处理、小波分析等。

E-mail: yangaiping@tju.edu.cn

去噪, 比周期延拓取得了更好的去噪效果。本文在此基础上给出了信号镜像对称延拓下构造双正交小波变换矩阵的方法, 并证明了在该对称延拓方式下实现 Mallat 算法的完全重构条件, 同时完善了小波变换矩阵变换算法, 构造的小波变换矩阵可直接用于基于小波的图像分块处理中, 如 FR II 去噪^[1,5]、图像压缩编码等, 并可使得边界效应明显减少。

2 双正交小波变换的 Mallat 算法

由 Mallat 分解、重构算法知, 小波变换可以通过分解、重构滤波器系数逐级迭代实现, 其中, 双正交小波变换算法最具一般性, 对于某一尺度 k 由 $\{c_{k+1,n}\}$ 求 $\{c_{k,n}\}$ 和 $\{d_{k,n}\}$ 的算法称为分解算法, 其中 $\{c_{k+1,n}\}$ 为原始数字信号在尺度 $k+1$ 下的离散逼近, $\{c_{k,n}\}$ 和 $\{d_{k,n}\}$ 为经过一级分解后在尺度 k 下的离散逼近信号和离散细节信号。对于双正交小波, 分解关系式可写为

$$\begin{cases} c_{k,l} = \sum_n \tilde{p}_{n-2l} c_{k+1,n} \\ d_{k,l} = \sum_n \tilde{q}_{n-2l} c_{k+1,n} \end{cases} \quad (1)$$

重构关系式可写为

$$c_{k+1,l} = \sum_{n \in \mathbf{Z}} p_{l-2n} c_{k,n} + q_{l-2n} d_{k,n} \quad (2)$$

其中, $\{\tilde{p}_n\}$ 和 $\{\tilde{q}_n\}$ 为分解滤波器系数, $\{p_n\}$ 和 $\{q_n\}$ 为重构滤波器系数, 对于双正交小波而言, 且有下列关系成立^[6]:

$$\begin{cases} \tilde{q}_n = (-1)^{1-n} p_{1-n}, & (p_n = (-1)^n \tilde{q}_{1-n}) \\ \tilde{p}_n = (-1)^n q_{1-n}, & (q_n = (-1)^{1-n} \tilde{p}_{1-n}) \end{cases} \quad (3)$$

3 有限长信号小波变换的矩阵表示

对于长度为 $L = 2^{k+1}$ 的有限长信号 $C_{k+1} = [c_{k+1,0}, c_{k+1,1}, \dots, c_{k+1,L-1}]$ 的小波分解与重构过程可用矩阵乘法来实现。小波分解过程为

$$F_k = \begin{bmatrix} C_k \\ \vdots \\ D_k \end{bmatrix} = DE_k \cdot C_{k+1} \quad (4)$$

其中, $C_k = [c_{k,0}, c_{k,1}, \dots, c_{k,L/2-1}]^T$ 为分解后的逼近信号, $D_k = [d_{k,0}, d_{k,1}, \dots, d_{k,L/2-1}]^T$ 为分解后的细节信号, $DE_k = \begin{bmatrix} \tilde{P}_k \\ \tilde{Q}_k \end{bmatrix}$ 称为分解矩阵, 其中 \tilde{P}_k, \tilde{Q}_k 均为

$2^k \times 2^{k+1}$ 大小的矩阵。小波重构过程可写为

$$C_{k+1} = RE_k \cdot F_k = \begin{bmatrix} P_k \\ \vdots \\ Q_k \end{bmatrix} \cdot F_k \quad (5)$$

称 $RE_k = \begin{bmatrix} P_k \\ \vdots \\ Q_k \end{bmatrix}$ 为重构矩阵, P_k, Q_k 均为 $2^{k+1} \times 2^k$ 大小

的矩阵。对于双正交小波变换, C_{k+1} 的对称延拓方式可以按照式 (6) 对边界数据进行对称延拓^[7] (整点对称)

$$\begin{cases} C_{j+1,n} = C_{j+1,-n} & n < 0 \\ C_{j+1,n} = C_{j+1,2n-2} & n > L-1 \end{cases} \quad (6)$$

本文先对 C_{k+1} 的边界数据按照式 (7) 进行镜像对称延拓^[7] (半点对称), 然后对小波变换矩阵进行构造, 并证明了镜像对称延拓方式下实现 Mallat 算法的完全重构条件。

$$\begin{cases} C_{j+1,n} = C_{j+1,-n-1} & n < 0 \\ C_{j+1,n} = C_{j+1,2n-1} & n > L-1 \end{cases} \quad (7)$$

4 信号镜像对称延拓 Mallat 算法的完全重构条件及小波变换矩阵的构造

4.1 镜像对称延拓 Mallat 算法的完全重构条件

镜像对称延拓后的数据是以原数据长度的 2 倍为周期, 为保证信号变换前后数据量不变, 一般采用的滤波器特点及小波变换系数的延拓方式需要满足一定的关系^[8], 可以证明如下结论:

定理 1 采用 Mallat 算法, 对长度为 L 的数据 $C_{k+1} = [c_{k+1,0}, c_{k+1,1}, \dots, c_{k+1,L-1}]$ 进行小波变换。若分解低通、重构低通滤波器系数 $\{\tilde{p}_n\}, \{p_n\}$ 长度为偶数, 并满足 $\tilde{p}_n = \tilde{p}_{1-n}, p_n = p_{1-n}$, 则被分解信号可按照式 (7) 进行镜像对称延拓。重构时若低频数据 $\{c_{k,l}\}$ 按照式 (8) 延拓, 高频数据 $\{d_{k,l}\}$ 按照式 (9) 延拓, 则 Mallat 算法可完全重构。

$$\begin{cases} c_{k-l} = c_{k+l} & l > 0 \text{ 左端} \\ c_{k,(L/2)-1+l} = c_{k,(L/2)-l} & l > 0 \text{ 右端} \end{cases} \quad (8)$$

$$\begin{cases} d_{k-l} = -d_{k+l} & l > 0 \text{ 左端} \\ d_{k,(L/2)-1+l} = -d_{k,(L/2)-l} & l > 0 \text{ 右端} \end{cases} \quad (9)$$

证明 只要证明原始信号按照式 (7) 镜像延拓后分解得到的低频数据的对称性满足式 (8), 高频数据的对称性满足式 (9) 即可。

由 $\tilde{q}_n = (-1)^{1-n} p_{1-n}$, 可得

$$\begin{aligned} \tilde{q}_{1-n} &= (-1)^{1-(1-n)} p_{1-(1-n)} = (-1)^n p_n \\ &= (-1)^n p_{1-n} = -\tilde{q}_n \end{aligned}$$

即有 $\tilde{q}_{1-n} = -\tilde{q}_n$ 。Mallat 算法分解式为

$$\begin{cases} c_{k,l} = \sum_n c_{k+1,n} \tilde{p}_{n-2l} = \sum_n c_{k+1,n+2l} \tilde{p}_n \\ d_{k,l} = \sum_n c_{k+1,n} \tilde{q}_{n-2l} = \sum_n c_{k+1,n+2l} \tilde{q}_n \end{cases}$$

因此, 低频分解系数的左端为

$$\begin{aligned} c_{k-l} &= \sum_n c_{k+1,n-2l} \tilde{p}_n \\ &= \sum_n c_{k+1,-n+2l-1} \tilde{p}_n = \sum_n c_{k+1,-n+2l-1} \tilde{p}_{1-n} \\ &= \sum_n c_{k+1,n+2l-1} \tilde{p}_{1+n} = \sum_n c_{k+1,n+2l-2} \tilde{p}_n = c_{k,l-1} \end{aligned}$$

低频分解系数的右端为

$$\begin{aligned} c_{k,(L/2)-1+l} &= \sum_n c_{k+1,n+2l[(L/2)-1+l]} \tilde{p}_n \\ &= \sum_n c_{k+1,n+L-2+2l} \tilde{p}_n = \sum_n c_{k+1,L-n+1-2l} \tilde{p}_n \\ &= \sum_n c_{k+1,L-(n-1)-2l} \tilde{p}_n = \sum_n c_{k+1,L+n-2l} \tilde{p}_{1-n} \\ &= \sum_n c_{k+1,L+n-2l} \tilde{p}_n = c_{k,(L/2)-l} \end{aligned}$$

相应地, 高频分解系数的左端为

$$\begin{aligned} d_{k-l} &= \sum_n d_{k+1,n-2l} \tilde{q}_n \\ &= \sum_n d_{k+1,-n+2l-1} \tilde{q}_n = \sum_n c_{k+1,-n+2l-1} (-\tilde{q}_{1-n}) \\ &= -\sum_n d_{k+1,-n+2l-1} \tilde{q}_{1-n} = \sum_n d_{k+1,n+2l-2} \tilde{q}_n \\ &= -d_{k,l-1} \end{aligned}$$

高频分解系数的右端为

$$\begin{aligned} d_{k,(L/2)-1+l} &= \sum_n d_{k+1,n+L-2l-2} \tilde{q}_n \\ &= \sum_n d_{k+1,L-(n-1)-2l} \tilde{q}_n = -\sum_n d_{k+1,L-(n-1)-2l} \tilde{q}_{1-n} \\ &= -\sum_n d_{k+1,L+n-2l} \tilde{q}_n = -\sum_n d_{k,(L/2)-l} \end{aligned}$$

至此, 问题得证。

例子 下面给出一个例子说明上述定理, 设长度为 8 的原始信号

$$S = [160 \ 157 \ 155 \ 173 \ 96 \ 109 \ 126 \ 102]$$

对其做边界镜像对称延拓的 Bior3_3 小波变换时, 首先做 S 的镜像对称延拓, 得到的

$$S_{\text{exten}} = [102 \ 126 \ 109 \ 96 \ 173 \ 155 \ 157 \ 160 \ 160 \ 157 \ 155 \ 173 \ 96 \ 109 \ 126 \ 102]$$

以长度 16 为周期。其单尺度 Bior3_3 小波变换后

的低频系数为

$$L_{\text{Coef}} = [168 \ 622 \ 9 \ 123 \ 655 \ 3 \ 252 \ 481 \ 3 \ 217 \ 501 \ 6 \ 217 \ 501 \ 6 \ 252 \ 481 \ 3 \ 123 \ 655 \ 3 \ 168 \ 622 \ 9]$$

高频系数为

$$H_{\text{Coef}} = [11 \ 490 \ 5 \ -15 \ 202 \ 8 \ -20 \ 329 \ 3 \ 0 \ 707 \ 0 \ 1 \ -0 \ 707 \ 1 \ 20 \ 329 \ 3 \ 15 \ 202 \ 8 \ -11 \ 490 \ 5]$$

可见, 其低频系数具有偶对称性, 高频系数具有奇对称性。若对低、高频系数只取一半, 则可得小波变换的结果为

$$\begin{aligned} W_{\text{Coef}} &= [L_{\text{Coef}} \ H_{\text{Coef}}] \\ &= [217 \ 501 \ 6 \ 252 \ 481 \ 3 \ 123 \ 655 \ 3 \ 68 \ 622 \ 9 \ -0 \ 707 \ 1 \ 20 \ 329 \ 3 \ 15 \ 202 \ 8 \ -11 \ 490 \ 5] \end{aligned}$$

在重构时, 低频部分做镜像对称延拓, 高频部分做镜像反对称延拓, 这样既能使得信号小波变换前后保持数据量不变, 又能保证完全重构原信号。

4.2 镜像对称延拓小波变换矩阵的构造

下面以 Bior3_3 双正交小波为例来说明该类小波变换矩阵的构造。低通、高通分解滤波器系数分别为 $\tilde{p}_n = \{\tilde{p}_{-3}, \tilde{p}_{-2}, \tilde{p}_{-1}, \tilde{p}_0, \tilde{p}_1, \tilde{p}_2, \tilde{p}_3, \tilde{p}_4\}$, $\tilde{q}_n = \{\tilde{q}_{-1}, \tilde{q}_0, \tilde{q}_1, \tilde{q}_2\}$ 。

低通、高通重建滤波器系数分别为

$$\begin{aligned} p_n &= \{p_{-1}, p_0, p_1, p_2\} \\ q_n &= \{q_{-3}, q_{-2}, q_{-1}, q_0, q_1, q_2, q_3, q_4\} \end{aligned}$$

并满足 $\tilde{p}_n = \tilde{p}_{1-n}$, $p_n = p_{1-n}$ 。

小波分解矩阵组成形式为 $DE_k = \begin{bmatrix} \tilde{P}_k \\ \tilde{Q}_k \end{bmatrix}$, \tilde{P}_k 、 \tilde{Q}_k

均为 $2^k \times 2^{k+1}$ 大小的矩阵, 分别由系数 $\{\tilde{p}_k\}$ 、 $\{\tilde{q}_k\}$ 组成。根据 Mallat 算法和信号镜像对称延拓, 分解矩阵 DE_k 的构造方法如下:

先设置一个 $2^k \times 2^{k+1}$ 大小的待赋值空矩阵 (图 1 中整个阴影部分), 同时让滤波器系数 $\{\tilde{p}_k\}$ 以

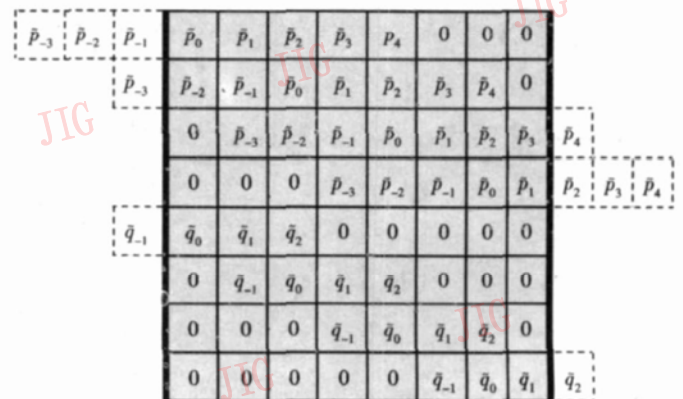


图 1 构造矩阵 DE_2

Fig 1 Construction of DE_2

下标递增的方式排成一行, 放置在空矩阵的第 1 行 (不够则补零), 其中令 \tilde{p}_0 放置在 (0 0) 处, 其左方是下标为负的滤波器系数, 右方是下标为正的滤波器系数, 部分系数可能出现在设置的空矩阵外面 (虚线框); 图 1 下面各行分别由上一行向右移两位得到 (共移位 $2^k - 1$ 次), 图 1 的上半部分给出了 $k=2$ 时的情况; 然后, 以矩阵左边框和右边框为对称轴 (图中的加粗加黑竖线), 将所有放置在 $2^k \times 2^{k+1}$

$$\begin{array}{cccccccc}
 \tilde{p}_{-1} + \tilde{p}_0 & \tilde{p}_{-2} + \tilde{p}_1 & \tilde{p}_{-3} + \tilde{p}_2 & \tilde{p}_3 & \tilde{p}_4 & 0 & 0 & 0 \\
 \tilde{p}_{-3} + \tilde{p}_{-2} & \tilde{p}_{-1} & \tilde{p}_0 & \tilde{p}_1 & \tilde{p}_2 & \tilde{p}_3 & \tilde{p}_4 & 0 \\
 0 & \tilde{p}_{-3} & \tilde{p}_{-2} & \tilde{p}_{-1} & \tilde{p}_0 & \tilde{p}_1 & \tilde{p}_2 & \tilde{p}_3 + \tilde{p}_4 \\
 0 & 0 & 0 & \tilde{p}_{-3} & \tilde{p}_{-2} & \tilde{p}_{-1} + \tilde{p}_4 & \tilde{p}_0 + \tilde{p}_3 & \tilde{p}_1 + \tilde{p}_2 \\
 \tilde{q}_{-1} + \tilde{q}_0 & \tilde{q}_1 & \tilde{q}_2 & 0 & 0 & 0 & 0 & 0 \\
 0 & \tilde{q}_{-1} & \tilde{q}_0 & \tilde{q}_1 & \tilde{q}_2 & 0 & 0 & 0 \\
 0 & 0 & 0 & \tilde{q}_{-1} & \tilde{q}_0 & \tilde{q}_1 & \tilde{q}_2 & 0 \\
 0 & 0 & 0 & 0 & 0 & \tilde{q}_{-1} & \tilde{q}_0 & \tilde{q}_1 + \tilde{q}_2
 \end{array}$$

可用同样方法构造出 4×4 ($k=1$) 大小的小波分解矩阵 (参见图 2)

$$DE_1 = \begin{bmatrix} \tilde{p}_{-1} + \tilde{p}_0 & \tilde{p}_{-2} + \tilde{p}_1 & \tilde{p}_{-3} + \tilde{p}_2 & \tilde{p}_3 + \tilde{p}_4 \\ \tilde{p}_{-3} + \tilde{p}_{-2} & \tilde{p}_{-1} + \tilde{p}_4 & \tilde{p}_0 + \tilde{p}_3 & \tilde{p}_1 + \tilde{p}_2 \\ \tilde{q}_{-1} + \tilde{q}_0 & \tilde{q}_1 & \tilde{q}_2 & 0 \\ 0 & \tilde{q}_{-1} & \tilde{q}_0 & \tilde{q}_1 + \tilde{q}_2 \end{bmatrix}$$

同样, 可构造出 2×2 大小的小波分解矩阵 (参见图 3)

$$DE_0 = \begin{bmatrix} \tilde{p}_{-1} + \tilde{p}_0 + \tilde{p}_3 + \tilde{p}_4 & \tilde{p}_{-3} + \tilde{p}_{-2} + \tilde{p}_2 + \tilde{p}_3 \\ \tilde{a}_{-1} + \tilde{a}_0 & \tilde{a}_1 + \tilde{a}_2 \end{bmatrix}$$

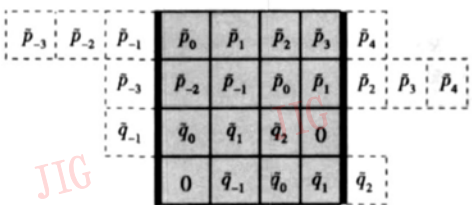


图 2 构造小波分解矩阵 DE_1
Fig 2 Construction of DE_1

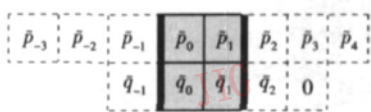


图 3 构造小波分解矩阵 DE_0
Fig 3 Construction of DE_0

这里需要注意的是, 在滤波器系数相对较长时, 可能会出现系数折叠叠加多次的情况 (如图 3

的矩阵外面的系数折叠, 叠加到矩阵里面, 即得到 $2^k \times 2^{k+1}$ 大小的矩阵 \tilde{P}_k 。例如第 1 行中, \tilde{p}_{-1} 叠加到 \tilde{p}_0 上, \tilde{p}_{-2} 叠加到 \tilde{p}_1 上; 第 4 行中, \tilde{p}_2 叠加到 \tilde{p}_1 上, \tilde{p}_3 叠加到 \tilde{p}_0 上, \tilde{p}_4 叠加到 \tilde{p}_{-1} 上等。用同样的方法在矩阵 \tilde{P}_k 的下方得到 $2^k \times 2^{k+1}$ 大小的矩阵 \tilde{Q}_k ; 最后, 由 $\begin{bmatrix} \tilde{P}_k \\ \tilde{Q}_k \end{bmatrix}$ 组成大小为 $2^{k+1} \times 2^{k+1}$ 的分解矩阵 DE_k 。

当 $k=2$ 时, 构造出 8×8 大小的小波分解矩阵 DE_2 为

所示), 在构造 2×2 的小波分解矩阵时, 系数 \tilde{p}_{-3} 先以矩阵左边框为对称轴第 1 次叠加到 \tilde{p}_2 上, 然后再以矩阵右边框为对称轴叠加到 \tilde{p}_1 上。

根据以上方法就可构造出任意阶的分解矩阵 DE_k 。对于任意长度为 $L = 2^{s+1}$ 的有限长信号 $C_{k+1} = [c_{k+1,0}, c_{k+1,1}, \dots, c_{k+1,L-1}]$ 可以利用矩阵-向量乘积的方法来实现该信号的单尺度双正交小波分解:

$$\begin{bmatrix} C_k \\ \vdots \\ D_k \end{bmatrix} = DE_k \cdot C_{k+1}$$

对于长度为 $N = 2^{s+1}$ 的数据可

以进行 m ($1 \leq m \leq s+1$) 级小波分解, 这里称 $s+1$ 为最大分解深度, 有了各级小波分解矩阵 $DE_0, DE_1, DE_2, \dots, DE_s$, 就可得到任意深度的小波分解矩阵。设小波分解深度为 m ($1 \leq m \leq s+1$), 则相应的分解矩阵 DE 可通过如下公式得到:

$$DE = \begin{bmatrix} DE_{s-m+1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{\frac{(2^m-1)}{2^{m-1}}} \end{bmatrix} \cdot \begin{bmatrix} DE_{s-m+2} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{\frac{(2^m-2)}{2^{m-2}}} \end{bmatrix} \cdots \begin{bmatrix} DE_{s-2} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{\frac{(2^2-1)}{2}} \end{bmatrix} \cdot \begin{bmatrix} DE_{s-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_2 \end{bmatrix} \cdot [DE_s] \quad (10)$$

对长度 $N = 8$ 的信号, 在进行 3 级双正交小波变换时, 分解深度为 3 的小波分解矩阵可以通过各级小波分解矩阵 DE_2, DE_1, DE_0 及式 (10) 得到:

$$DE = \begin{bmatrix} DE_0 & 0 & 0 & 0 \\ 0 & \mathbf{I}_{2 \times 2} & 0 & 0 \\ 0 & 0 & \mathbf{I}_{2 \times 2} & 0 \\ 0 & 0 & 0 & \mathbf{I}_{2 \times 2} \end{bmatrix} \cdot \begin{bmatrix} DE_1 & 0 \\ 0 & \mathbf{I}_{4 \times 4} \end{bmatrix} \cdot [DE_2] = [W_0^T \ W_1^T \ W_2^T \ W_3^T \ W_4^T \ W_5^T \ W_6^T \ W_7^T]^T$$

$$= \begin{bmatrix} 0.3536 & 0.3536 & 0.3536 & 0.3536 & 0.3536 & 0.3536 & 0.3536 & 0.3536 \\ 0.2472 & 0.2044 & 0.3425 & 0.2679 & -0.2679 & -0.3425 & 0.2044 & -0.2472 \\ 0.3672 & 0.3750 & -0.5937 & -0.6250 & 0.2813 & 0.2813 & -0.0625 & 0.0234 \\ 0.0234 & 0.0625 & -0.2813 & -0.2813 & 0.6250 & 0.5937 & -0.3750 & -0.3672 \\ 0.3536 & -0.5303 & 0.1768 & 0 & 0 & 0 & 0 & 0 \\ 0 & -0.1768 & 0.5303 & -0.5303 & 0.1768 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0.1768 & 0.5303 & -0.5303 & 0.1768 & 0 \\ 0 & 0 & 0 & 0 & 0 & -0.1768 & 0.5303 & -0.3536 \end{bmatrix}$$

分解矩阵的每一行称为分解基向量 $W_i (i=0, 1, \dots, N-1)$, 当 $N=8$ 时, 小波变换的分解基向量图形如图 4 所示, 那么对数据向量 X 进行最大深度的边界镜像对称延拓方式的 Bior3.3 小波分解后的系数向量 Y 可用以下的矩阵乘法实现:

$$Y = DE \cdot X \quad (11)$$

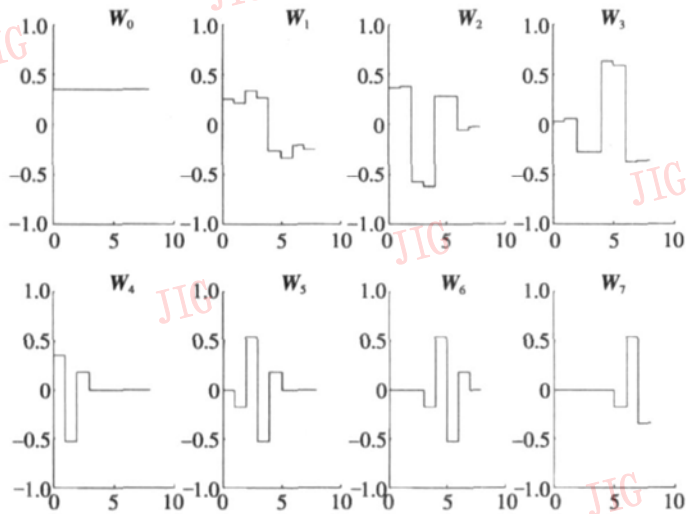


图 4 Bior3.3 最大深度分解基向量 ($N=8$)

Fig 4 Decomposition base vectors of Bior3.3 wavelet in maximum depth ($N=8$)

$$RE = [V_0 \ V_1 \ V_2 \ V_3 \ V_4 \ V_5 \ V_6 \ V_7] = \begin{bmatrix} 0.3536 & 0.7071 & 0.8125 & -0.1875 & 1.1490 & -0.2652 & 0.0000 & -0.0000 \\ 0.3536 & 0.6187 & 0.3984 & -0.1016 & -1.1932 & 0.1547 & -0.0663 & 0.0000 \\ 0.3536 & 0.4419 & -0.4297 & 0.0703 & -0.2210 & 0.9944 & -0.1989 & 0.0000 \\ 0.3536 & 0.1768 & -0.6719 & 0.3281 & 0.1989 & -0.9944 & 0.1547 & -0.0663 \\ 0.3536 & -0.1768 & -0.3281 & 0.6719 & 0.0663 & -0.1547 & 0.9944 & -0.1989 \\ 0.3536 & -0.4419 & -0.0703 & 0.4297 & -0.0000 & 0.1989 & -0.9944 & 0.2210 \\ 0.3536 & -0.6187 & 0.1016 & -0.3984 & -0.0000 & 0.0663 & -0.1547 & 1.1932 \\ 0.3536 & -0.7071 & 0.8125 & -0.1875 & 0.0000 & -0.0000 & 0.2652 & -1.1490 \end{bmatrix} \quad (14)$$

由此易知, 二者完全可以相互交换, 既可以用 DE_k 作为分解矩阵, 同时让 RE_k 作为重构矩阵; 也

Y 的每个分量 $y_i (i=0, 1, \dots, 7)$ 是 X 在基向量 W_i 上的投影: $y_i = W_i^T \cdot X$.

4.3 分解、重构矩阵的互易性

由定理 1 易知, 不但存在各级分解矩阵的逆矩阵

$$RE_k = DE_k^{-1} \quad (k=0, 1, \dots, s) \quad (12)$$

作为其完全重构矩阵, 显然也存在任意分解深度的小波分解矩阵的逆矩阵 $RE = DE^{-1}$ 作为其完全重构矩阵。设各级重构矩阵为 $RE_0, RE_1, RE_2, \dots, RE_s$, 若分解时深度为 $m (1 \leq m \leq s+1)$, 则重构矩阵为

$$RE = RE_s \begin{bmatrix} RE_{s-1} & 0 \\ 0 & \mathbf{I}_2 \end{bmatrix} \cdot \begin{bmatrix} RE_{s-2} & 0 \\ 0 & \mathbf{I}_{\frac{(2-1)}{2^2}} \end{bmatrix} \cdots \begin{bmatrix} RE_{s-m+1} & 0 \\ 0 & \mathbf{I}_{\frac{(2m-1-1)}{2^{m-1}}} \end{bmatrix} \quad (13)$$

重构矩阵 RE 的每一列向量称为重构基向量 $V_i (i=0, 1, \dots, N-1)$ 。对于 Bior3.3 小波来说, 当 $N=8$ 时, 边界镜像对称延拓的最大分解深度小波重构矩阵如下:

可以用 RE_k 作为分解矩阵, 同时让 DE_k 作为重构矩阵。同理, 也可以用 RE 作为任意深度的分解矩阵,

用 DE 作为相应的重构矩阵。其基向量图形如图 5 所示, 而小波重构可用如下矩阵乘法实现:

$$X = RE \cdot Y \quad (15)$$

由式 (11) 式 (15) 可见, 有限长度数据对称延拓双正交小波变换无需逐级计算, 而直接利用矩阵乘法即可实现。

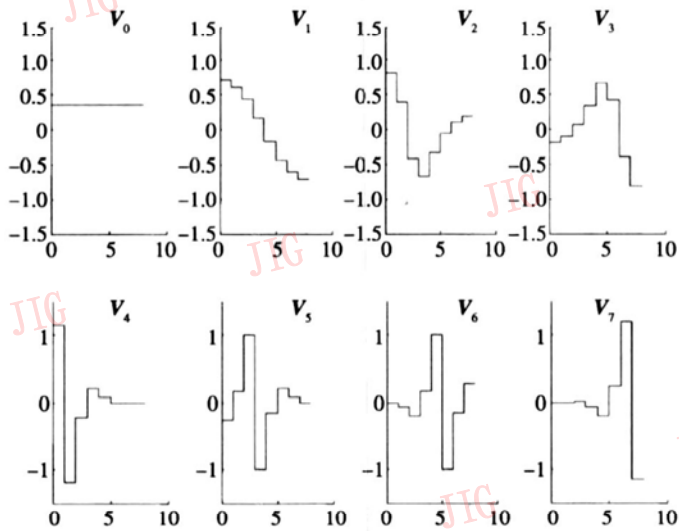


图 5 Biort3_3 最大深度重构基向量 ($N = 8$)

Fig 5 Reconstruction base vectors of Biort3_3 wavelet in maximal depth ($N = 8$)

5 结 论

本文给出了一种信号镜像对称延拓方式下双正交小波任意深度小波变换矩阵的构造方法, 并证明了该延拓方式下 Mallat 算法的完全重构条件。同时举例构造了 Biort3_3 小波的三层分解、重构矩阵, 并给出了基向量及其基图形。这样任意深度的小波变换都不需要逐级迭代而直接用矩阵、向量乘积即可实现, 其是对文献 [1]、[4] 方法的补充和完善。实践证明, 基于边界对称延拓的双正交小波变换矩阵用于信号、图像处理中, 不仅可有效减少边界效应, 而且实现简单, 使得运算量大大降低。

参考文献 (References)

- 1 Guo Xu-jing, Hou Zheng-xin. An improved FRIT denoising method based on wavelet base vector [J]. Journal of Electronic Measurement and Instrument, 2005, 19(6): 49~54 [郭旭静, 侯正信. 基于小波基向量的改进 FRIT 去噪方法 [J]. 电子测量与仪器学报, 2005, 19(6): 49~54]
- 2 Smith M. J.T., Eddins S.L. Analysis/synthesis techniques for subband image coding [J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1990, 38(8): 1446~1456
- 3 Strela V., Heller P.N., Strang G., et al. The application of multiwavelet filter banks to image processing [J]. IEEE Transactions on Image Processing, 1999, 8(4): 548~563
- 4 Hou Zheng-xin, Yang Ai-ping. FRIT denoising using wavelet transform matrices with symmetric boundary-extension [J]. Journal of Optoelectronics • Laser, 2007, 18(5): 603~607. [侯正信, 杨爱萍. 对称延拓小波变换矩阵用于 FRIT 去噪 [J]. 光电子 • 激光, 2007, 18(5): 603~607.]
- 5 Do M.N., Vetterli M. The finite ridgelet transform for image representation [J]. IEEE Transactions on Image Processing, 2003, 12(1): 16~28
- 6 Mallat S. A Wavelet Tour of Signal Processing [M]. Beijing: China Machine Press, 2003 [Mallat Stephane. 信号处理的小波导引 (英文版) [M]. 北京: 机械工业出版社, 2003.]
- 7 Strang G., Nguyen T.Q. Wavelets and Filter Banks [M]. Cambridge, Boston, MA, USA: Wellesley Press, 1996
- 8 Banberger R.H., Eddins S.L., Nuri V. Generalized symmetric extension for size-limited multirate filter banks [J]. IEEE Transactions on Image Processing, 1994, 3(1): 82~87.

基于 Hash 函数敏感性的医学图像精确认证

钟晓燕 冯前进 陈武凡 江贵平

(南方医科大学生物医学工程学院, 广州 510515)

关键词 医学图像认证 易碎水印 整数小波变换 小波树 Hash 函数

中图法分类号: TP309.2 文献标识码: A 文章编号: 1006-8961(2008)02-0204-05

Hard Authentication for Medical Image Based on Sensitivity of Hash Function

ZHONG Xiao-yan, FENG Qian-jin, CHEN Wu-fan, JIANG Gui-ping

(Department of Biomedical Engineering, South Medical University, Guangzhou 510515)

Abstract A fragile watermark is designed using integral wavelet transform combined with hash function in order to verify the integrity of medical images. Even 1 bit of change in the picture can be sensed. Furthermore the region distorted can be easily oriented in this algorithm without an original image. Compared with the traditional digital watermark based on wavelet transform, the integer wavelets not only simplify the calculation but also improve the quality of watermarked image. And the hash function constructed with MD5 algorithm increases the security and ability of tamper localization of watermark. The results of experiment indicates that watermark scheme is highly sensitive to the distortion of the image. Since it requires a key during both the insertion and the extraction procedures, it has high security and the whole process is simple and accurate.

Keywords medical image authentication, fragile watermark, integral wavelet transform, quadtree, Hash function

1 引言

随着图像处理和计算机技术的不断进步和发展, 数字成像技术在医学中的应用日益广泛, 如它在医学超声成像技术、X-CT、MR 及核医学成像中的应用等。另一方面以计算机网络为基础的图像存储和传输系统 PACS (picture archive and communication system) 及其应用也在不断地发展。医学图像的数字化使得对医学数字图像的恶意篡改成为可能, 由此将造成较多的医疗纠纷, 因此在这个信息数字的时代如何对

数字媒体内容的真实性和完整性实施有效保护已成为一个严峻的现实问题。对于大多数应用而言, 图像内容的微小改变可能都是可以接受的, 但是在医疗诊断时, 原有图像是否发生变化对于诊断结果就是非常重要的。因此在对医学图像进行真实性验证时, 应当采用精确认证, 即使用易碎水印。易碎水印算法大都由空间域 LSB 水印算法演变而来。Walton 首次提出用易碎数字水印的方法实现图像认证^[1], Yeung 利用一个密钥 K 为每个基本色产生一个查询表 LUT (look up tables) 通过修改每个像素值使之恰等于对应的水印信息, 从而完成水印信息的嵌入, 从

基金项目: 国家自然科学基金重点项目 (30730036); 国家重点基础研究发展计划 973 项目 (2003CB716104)

收稿日期: 2007-02-07; 改回日期: 2007-07-02

第一作者简介: 钟晓燕 (1977~), 女。现为南方医科大学医学信息研究所硕士研究生。主要研究方向为医学图像认证和编码等。

Email: dawnswallow212@163.com

通讯作者: 江贵平, Email: gzjiang@263.net

而可以把篡改定位精确到一个像素^[2]。Wong 提出基于易碎水印的分块认证算法^[3], 不久 Wong 又将密码学中的公开密钥密码系统引入到水印信息加密中, 从而提高了原系统的安全性^[4]。必须指出的是空间域方法的优点是能够嵌入较多的水印, 但难以定位频率篡改, 因为标记加在某些特定像素上, 常常不能检测对整幅图像的频率改动。

本文在 Wong 的基础上提出了一种基于提升格式的整数小波变换和 Hash 函数敏感性相结合的易碎水印方法。即首先应用整数提升小波变换, 然后构造出小波树结构, 利用小波变换系数的树结构关系和各层小波系数的特点在每棵小波树上嵌入一位水印信息。同时结合 MD5 算法构造的 Hash 函数和经典的密码学理论来设计水印算法, 以实现医学图像的精确认证。由于这种易碎数字水印是将加密算法与图像处理技术相结合, 因此算法可完全公开, 使用者只要持有密钥, 就可以保证安全。实验结果表明, 本文提出的易碎水印算法不仅具有文献 [3] 所述的较好篡改敏感性, 而且具有比文献 [3] 较好的峰值信噪比 PSNR (peak signal to noise ratio)。

2 基于整数小波变换的易碎水印算法

2.1 水印的嵌入

设 W 是二值水印图像, 大小为 $p \times p$, $W(u, v) \in \mathbf{Z}(u \in [0, 1] \wedge \mathbf{Z}, v \in [0, 1] \wedge \mathbf{Z})$, 其中 $W(u, v)$ 表示水印图像像素在 (u, v) 处的灰度值。为了提高嵌入水印图像的质量, 采取在每颗小波树上嵌入一位水印信息

和 3 个分块子图, 如图 1 所示。升阶基于提升框架的整数小波变换可以保证图像在进行小波变换和重构过程中不存在能量损失, 适用于医学图像这种感觉质量要求较高的数字水印领域。同时由于 DWT 具有多分辨率特性, 使得算法可以检测到各个分辨率下篡改的特性, 从而定位篡改攻击的能力强。

(2) 小波树结构 小波变换 (DWT) 后给出了图像在不同分辨率上的表示, 每个 DWT 系数反映了图像在局部空域和局部频域上的内容。图像的 DWT 系数具有一个 4 分树结构, 即某一分辨率上的图像区域在更为精细的下一分辨率上被分成 4 块较

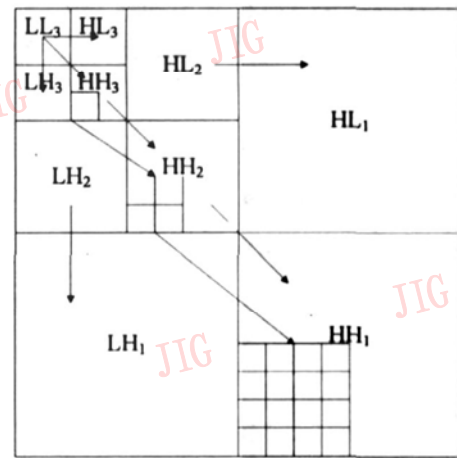
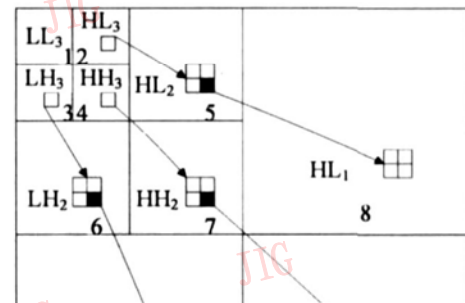


图 1 小波多分辨率分解

Fig 1 Wavelet multiresolution decomposition

小的区域来表示 (如图 2 所示)。第 1 层小波变换系数的检测分辨率 2×2 pixel 块, 第 2 层小波变换系数的检测分辨率 4×4 pixel 块, 依次类推, L 层小波系数的检测分辨率为 $2L \times 2L$ pixels 块。为了获得较高的定位精度和较好的图像质量, 将水印嵌入到第 1 层的细节分量上。采用 3 级小波分解, 因此篡改区域的定位精度为 8×8 pixels 块。



(3) 密钥的生成 为了保证隐藏信息的安全, 构造出了基于 MD5 算法的 Hash 函数, 生成嵌入点密钥 E_k 。

所谓 Hash 函数, 要求满足:

已知 Hash 函数的输出, 要求它的输入是困难的, 即已知 $c = \text{Hash}(m)$, 求 m 是困难的。

已知 m , 计算 $\text{Hash}(m)$ 是容易的。

已知 $c_1 = \text{Hash}(m)$, 构造 m_2 使得 $\text{Hash}(m_2) = c_1$ 是困难的。

$c = \text{Hash}(m)$, c 的每一比特都与 m 的每一比特相关, 并有高度的敏感性。即每改变 m 的一比特, 都将对 c 产生影响。

基于 MD5 算法的 Hash 函数是把输入任意长的消息, 以 512 位输入数据块为处理单位, 且每一分组又被划分为 16 个 32 位子分组, 经过了一系列的处理后, 算法的输出由 4 个 32 位分组组成, 将这 4 个 32 位分组合级联后将生成 128 bits 长的摘要, 发生相同摘要值的碰撞的几率为 $(1/2)^{128}$ 。从而保证当 MD5 输入 (如图 3 所示的图像内容、图像 ID、密钥 K) 发生任何微小变化时, 则生成的嵌入点密钥必然发生改变, 也就不能正确提取出水印。并且可以利用该函数的初值敏感性进行篡改定位, 定位是通过检测 Hash 值

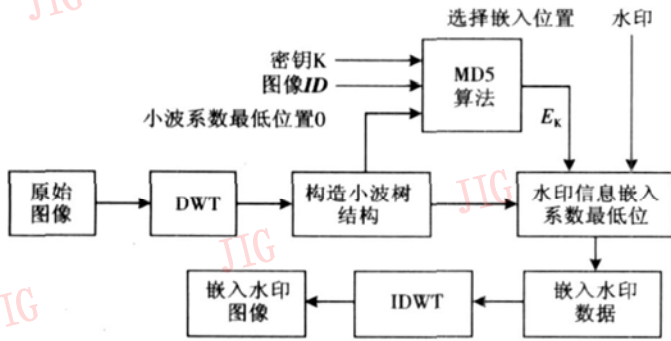


图 3 水印嵌入框图

Fig 3 The block diagram of watermark embedding system

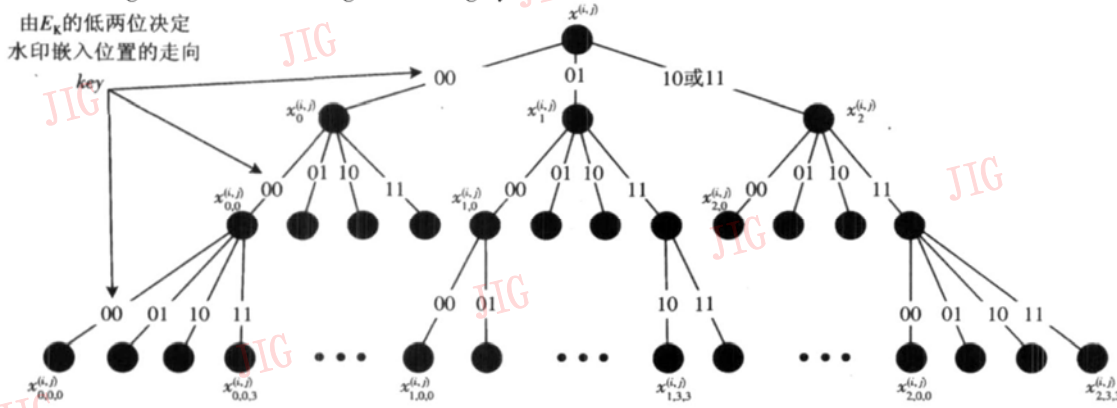


图 4 水印嵌入位置选择

Fig 4 Choose embedded position of watermarking

(5) 嵌入运算 为了提高水印敏感性, 在嵌入脆弱水印的过程中, 是采用嵌入小波系数最低位的方法来嵌入水印, 即对每棵树上的相应位置的小波系数嵌入一位的水印信息。嵌入规则如下: $W(i)$ 为第 i 位的水印信息, 取 0 或 1, 如果值为 0 则将相应节点最低有效位置 0, 反之, 置 1。重复至所有的小波树都已经嵌入水印信息。

(6) 对嵌入了水印的小波系数进行小波逆变换, 得

所确定的嵌入位置提取出的信息来进行, Hash 函数的高度敏感性使得篡改检测非常直观, 如果检测出来的水印图像是噪声图像, 则说明图像已经被篡改。

(4) 定位嵌入点

取根节点为 $LL_3(1, 1)$ 的这棵小波树 $tree(LL_3(1, 1))$; 将属于 $tree(LL_3(1, 1))$ 上的所有节点值 LSB 置 0 得到 $tree'(LL_3(1, 1))$ 。

以 $tree'(LL_3(1, 1))$ 和图像 ID、密钥一起作为 MD5 的输入, 得到 $H(tree'(LL_3(1, 1)))$, 记为 E_k , 以它作为水印嵌入的密钥, 每次循环左移两位, 取低两位确定嵌入位置的走向, 最终定位于相应的叶节点处, 嵌入水印信息。

如图 4 所示, 如果用 key 表示取出的低两位信息。当处于根节点时:

$$\begin{cases} key = 00 & \text{则定位于 } x_0^{(i,j)} \\ key = 01 & \text{则定位于 } x_1^{(i,j)} \\ key = 10, 11 & \text{则定位于 } x_2^{(i,j)} \end{cases} \quad (1)$$

当处于叶节点时:

$$\begin{cases} key = 00 & \text{则定位于 } x_{r_0}^{(i,j)} \\ key = 01 & \text{则定位于 } x_{r_1}^{(i,j)} \\ key = 10 & \text{则定位于 } x_{r_2}^{(i,j)} \\ key = 11 & \text{则定位于 } x_{r_3}^{(i,j)} \end{cases} \quad (2)$$

到嵌入水印图像。嵌入算法的基本框图如图 3 所示。

2.2 水印的提取

水印的提取是嵌入的逆过程。检测水印时, 首先对水印图像进行 3 级小波分解, 按小波树形式组织小波系数, 按照嵌入的过程寻找到嵌入点提取出水印信息。

$$\begin{cases} \tilde{W}(i) = 0 & \tilde{W}_F(x_b, y_i) \bmod 2 = 0 \\ \tilde{W}(i) = 1 & \tilde{W}_F(x_b, y_i) \bmod 2 = 1 \end{cases} \quad (3)$$

式中, $\tilde{W}_F(x_b, y_i)$ 为嵌入水印后图像在点 (x_b, y_i) 处的小波变换系数, mod 表示取余, $\tilde{W}(i)$ 为提取出的第 i 位的水印信息。可见水印抽取并不需要原始图像, 从而实现了所谓的不需要原始图像的盲检 (blind detect) 算法。

3 实验结果与分析

(1) 水印不可见性实验

以一幅 512×512 的 256 级灰度图像 Lena 为例做模拟实验, 算法采用 Daubechies5/3 滤波器 (可

逆) 和基于 2 维离散小波变换 (DWT), 嵌入一幅 64×64 的二值图像。对图像的质量采用峰值信噪比 (PSNR) 来进行评价。

$$PSNR = 10 \lg \frac{M^2 \times \max_{m,n}^2(x(m,n) - x'(m,n))}{\sum_{m,n} (x(m,n) - x'(m,n))^2} \quad (4)$$

图 5 演示了标准图 Lena 和医学 CT 图像嵌入和提取水印的实验。其中嵌入水印后的 Lena 图像和医学 CT 图像的 PSNR 分别为 65.2dB 和 65.5dB, 可见这种易碎水印不仅满足不可见性的要求, 而且具有较高的峰值信噪比, 适用于医学图像认证。



图 5 水印不可见性实验

Fig 5 Invisibility experiments of watermarking

(2) 敏感性实验

通过图 6 的实验来验证水印对篡改的敏感性。图 6 (c) ~ (f) 分别为图 6 (a) 中水印图像随机的选择位置和幅度其一、二、五、十个像素点篡改以后的图像中相应提取出来的水印图像。由此可见, 这种易碎水印对篡改极度敏感, 可以检测到图像 1bit 的篡改并且可直观地将篡改反映出来。这种敏感性主要来源于 Hash 函数对输入的强相关性, 即输出的每

一比特都与输入的每一比特相关, 以及对输入的高度敏感性, 即每改变输入一比特, 都将对输出值产生明显影响, 实验中常以雪崩性描述该性质: 输入改变任一比特, 输出的每一比特都以 1/2 的概率改变。

(3) 篡改定位实验

按照与水印嵌入相同的方法在相应的位置提取出水印, 与原始水印进行比较, 可进行篡改的检测与定位。为了使得认证算法具有一定的定位能力, 往往需要将宿主图像分割为独立的子块。显然图像子块尺寸越小, 定位的精度越高, 但计算量却越大, 所以需在定位精度和计算量之间进行折衷。这里选择 3 级小波分解, 由小波树的结构分析相当于定位 8×8 的块。图 7 是易碎水印对篡改区域定位的实验结果。嵌入水印后的图 7 (a), 经过篡改变为图 7 (b)。从图 7 (c) 可以看出该方法能有效地检测到图像被篡改了, 图 7 (f) 中白色的图像域代表篡改区域。

(4) 对比实验

为了验证本文算法的有效性, 将本文算法与文献 [3]、[8] 做了比较, 如表 1 所示。用篡改评估函数 TAF (tamper assessment function) 来对比 3 种算法对篡改的敏感性。

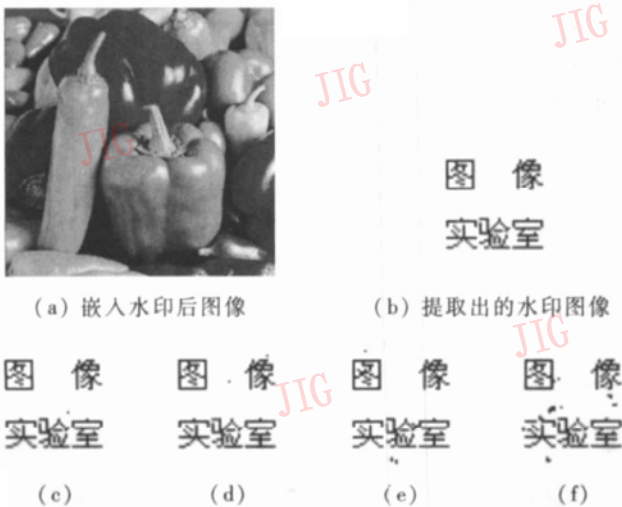


图 6 水印对图像篡改的敏感性实验

Fig. 6 Sensitivity experiments of watermarking towards tampered image

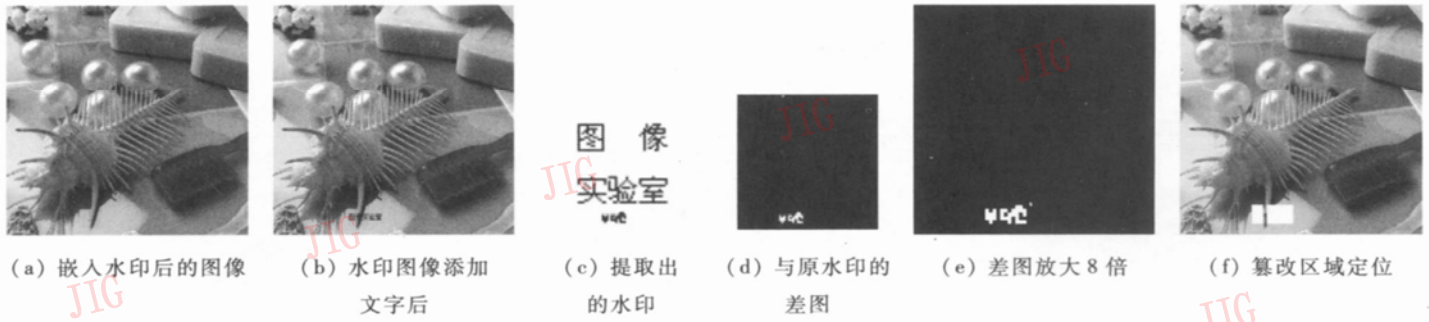


图 7 篡改区域定位试验

Fig. 7 Orientation experiment of tampered region

表 1 与其他算法的敏感性比较

Tab 1 Sensitivity comparison with other algorithms

算法	嵌入水印后 PSNR (dB)	各种攻击方式 (测试图像为 peppers)					
		椒盐噪声	高通滤波	锐化	旋转	压缩	剪切
本文算法	65.4	0.501	0.632	0.502	0.613	0.584	0.013
文献 [3] 中算法	51.1	0.353	0.500	0.498	0.498	0.499	0.006
文献 [8] 中算法	65.3	0.486	0.629	0.501	0.598	0.553	0.011

$$TAF = TAF(\omega, \bar{\omega}) = \frac{1}{N_l} \sum_{i=0}^{N_l} \omega_i \oplus \bar{\omega}_i \quad (5)$$

式中, ω 为原始水印序列, $\bar{\omega}$ 为提取水印序列, \oplus 代表异或操作, N_l 为水印的长度。

对水印图像进行加扰处理, 包括噪声、滤波、锐化、旋转、压缩、剪切等, 表 1 给出了不同干扰, 水印图像的篡改评估值。由比较可知, 在各种干扰下本文算法的篡改评估值较文献 [3]、[8] 均有不同程度的提高, 即表明本文提出的易碎水印算法的篡改敏感性要优于文献 [3]、[8], 同时本文算法的 PSNR 也高于文献 [3]、[8] 的算法, 从而可以进一步地提高嵌入水印图像的质量。

4 结 论

本文提出了一种基于整数小波变换、结合小波树结构和 Hash 函数的易碎水印算法。整数小波变换能够在图像分解和重构过程中, 使得图像损失为零, 可提高水印图像质量。在小波树结构上嵌入水印信息, 可在时频域上定位篡改的区域。而 MD5 算法构造的 Hash 函数的运用提高了水印安全性和水印检测的能力。实验结果表明, 本文算法对篡改具有高度敏感性, 且整个认证过程还需密钥完成检测, 安全性很高, 认证过程计算简单且准确性很高, 是一种医学图像鉴定的有效方法。

参考文献 (References)

- Walton S. Information authentication for a slippery new age [J]. Doctor Dobbs Journal 1995, 20(4): 18~26
- Yeung M, Mintzer F. Invisible watermarking for image verification [J]. Journal of Electronic Imaging 1998, 7(3): 578~591
- Wong P W, Memon N. Secret and public key image watermarking schemes for image authentication and ownership verification [J]. IEEE Transactions on Image Processing 2001, 10(10): 593~601
- Wong P W. A public key watermark for image verification and authentication [A]. In Proceedings of the IEEE International Conference on Image Processing [C], Chicago, USA, 1998: 455~459
- Lin E T, Delp E J. A review of fragile image watermarks [A]. In Proceedings of the Multimedia and Security Workshop (ACM Multimedia'99) Multimedia Contents [C], Orlando, FL, USA, 1999: 25~29
- Coatrieux G, Maître H, Sankur B, et al. Relevance of watermarking in medical imaging [A]. In IEEE-embs Information Technology Applications in Biomedicine [C], Arlington, USA, 2000: 250~255
- Wu Jin-hai, Lin Fu-zong. Image authentication based on digital watermarking [J]. Chinese Journal of Computers 2004, 27(9): 1153~1161. [吴金海, 林福宗. 基于数字水印的图像认证技术 [J]. 计算机学报, 2004, 27(9): 1153~1161.]
- Feng Qian-jin, Chen Ling-jian, Yang Feng. A fragile watermarking scheme for medical images based on integral wavelet transform [J]. Journal of Image and Graphics 2006, 11(5): 736~741. [冯前进, 陈凌剑, 杨丰. 基于整数小波变换的医学图像易碎水印方法 [J]. 中国图象图形学报, 2006, 11(5): 736~741.]

一种新的边缘保留各向异性扩散方法

郭圣文

(华南理工大学生物科学与工程学院生物医学工程系, 广州 510640)

摘要 由于超声图像中的斑点严重影响了图像质量, 也增加了临床诊断与治疗的困难, 因此对于噪声图像, 往往要先用高斯卷积对图像进行一定的平滑, 再通过求微分来检测边缘。为了更好地滤除超声图像中的斑点, 通过构造基于高斯卷积的结构张量, 并将其引入到各向异性扩散方法中, 实验结果表明, 这种新的各向异性扩散方法不仅能有效地抑制斑点噪声, 而且能检测并保留图像边缘与细节特征。

关键词 斑点 各向异性扩散 结构张量

中图分类号: TP391.41 **文献标识码:** A **文章编号:** 1006-8961(2008)02-0209-05

A Novel Approach on Edge Preserving Anisotropic Diffusion

GUO Shengwen

(Department of Biomedical Engineering, College of Bioscience and Bioengineering, South China University of Technology, Guangzhou 510006)

Abstract The Ultrasound images are usually degraded by speckle which brings much difficulties in diagnosis and treatment. In noisy images, the Gaussian convolution is often applied to smooth image and then gray differential is calculated to detect image edges. A structure tensor based on Gaussian convolution is devised and introduced to an isotropic diffusion in order to reduce speckle. Experiment shows that the novel anisotropic diffusion based on structure tensor can reducing speckle, detect and preserve edges and useful details effectively.

Keywords speckle, an isotropic diffusion(ANDI), structure tensor

1 引言

超声图像中存在一种特殊的斑点噪声, 由于它严重影响了图像分辨率与质量, 从而给临床诊断、分析与图像分割带来了较大困难。滤波是去噪常用的方法, 它的基本思想是设计一种基于噪声统计特性的滤波器, 如 Loupas 等人提出了自适应加权中值滤波^[1], Dutt 等人采用的自适应斑点抑制滤波^[2]等, 这些方法虽然对斑点有一定的抑制作用, 但由于它们与传统的滤波方法一样, 存在阈值选择困难、对滤波窗口的大小与形状很敏感、无方向性等问题, 因此它们在保留边缘与重要结构特征方面的能力有限。

各向异性扩散 (anisotropic diffusion, ANDI) 方法^[3]最早由 Perona 与 Malik 提出, 随后引起了人们

广泛的关注, 并取得了相当大的进展, 如多尺度分析^[4-5]、鲁棒性分析^[6]等。尤其是 Weickert 对各向异性扩散的适定性进行了研究, 得出了其适定性的条件^[7], 并通过模拟斑点图像的处理, 验证了各向异性扩散方法在抑制图像中斑点噪声方面具有良好的性能。You 等人提出了一种改进的四阶各向异性扩散方法^[8], 可以有效地消除“块效应”。

各向异性扩散与传统滤波方法的不同之处在于它的各向异性, 即在不同局部区域、不同方向上, 滤波与平滑的强度各异, 它与图像的局部结构 (灰度梯度) 相关, 在灰度变化较大的区域或方向上, 由于灰度梯度较大, 扩散作用较小, 从而应尽量保留这些局部细节特征; 若灰度变化较小或只有孤立噪声点的区域, 则采用较强的平滑处理, 以达到抑制噪声, 保留边缘与细节的目的。近年来, 各向异性扩散方

基金项目: 广东省自然科学基金项目 (05300233)

收稿日期: 2006-04-26 改回日期: 2006-09-26

第一作者简介: 郭圣文 (1971~), 男, 博士, 讲师。主要研究方向为图像处理、分析与可视化技术。Email: shwguo@scut.edu.cn

法在图像多尺度非线性分析、边缘检测、抑噪与边缘保留、分割等方面得到广泛应用。

在边缘检测、图像分割与分析过程中,人们往往希望能保留或增强某些重要的边缘与细节特征,但各向异性扩散方法,随着扩散过程的继续,边缘与细节会变得越来越模糊,而且对于存在大量特殊斑点的超声图像,这种现象变得更为明显。在对噪声图像进行边缘检测时,人们常用高斯卷积方法(即高斯滤波)来对噪声进行一定的平滑后,再提取边缘,由于减少了噪声的干扰,因此检测结果会更准确。为此,可以考虑先构造一个基于高斯卷积的结构张量,然后将其应用于各向异性扩散过程,以便在抑制斑点噪声的同时,还能保留边缘与重要细节。

2 边缘保留各向异性扩散方法

2.1 Gaussian 卷积微分

图像局部区域中像素灰度值的变化,为边缘检测、结构分析与理解提供了重要的信息,这种灰度的变化一般可从图像灰度的微分与梯度得到。然而,由于微分过程是非适定的,即原始图像的初始值发生微小的扰动,会导致微分过程产生很大的偏移,而且它对噪声极为敏感,因此,可以先通过一定的平滑后再求微分的方法来检测图像边缘与结构,即引入 Gaussian 卷积。

考虑图像与 Gaussian 核卷积的微分

$\partial_x \partial_y (K_\rho * I) = K_\rho * \partial_x \partial_y (I) = \partial_x \partial_y (K_\rho) * I$ (1)
 式中, I 表示图像, K_ρ 是尺度为 ρ 的高斯函数(卷积核), m, n 为阶数。上式说明,图像与 Gaussian 核卷积的微分等于微分图像与 Gaussian 核的卷积,也等同于图像与 Gaussian 核微分的卷积。

Gaussian 卷积常用于图像抑噪与去模糊,也可与其他非线性方法相结合,用以检测边缘、屋脊、交叉点等特征。应用 2 维图像 Gaussian 卷积在不同方向与阶数的微分,可以用于检测与分析图像结构。

式(1)中的尺度 ρ 即 Gaussian 函数的标准差,它的选择取决于噪声与检测特征的大小。其选取原则是大于噪声的尺度,小于特征的尺度,由于它忽略了比 ρ 小的部分,因此可利用 Gaussian 卷积来平滑噪声和检测一定大小的边缘与特征。实际应用时,一般可根据先验知识,如有关噪声及图像特征的知识及噪声及图像的局部统计特性(如均值、方差等)来确定。

2.2 各向异性扩散

各向异性扩散方程为

$$\frac{\partial I(x, y, t)}{\partial t} = \text{div}(\nabla I) \quad (2)$$

初始条件为 $I(x, y, 0)$, 表示初始图像, (x, y) 为空间位置, t 是时间参数, ∇I 为图像灰度梯度, div 表示散度。

Perona 与 Malik 对其进行了改进^[3]:

$$\frac{\partial I(x, y, t)}{\partial t} = \text{div}(g(|\nabla I|) \nabla I) \quad (3)$$

式中, $|\nabla I|$ 为梯度大小, $g(\cdot)$ 为“边缘停止”函数, 又称作扩散系数。当 $x \rightarrow \infty$ 时, $g(\cdot) \rightarrow 0$ 。 $g(\cdot)$ 对各向异性扩散的抑噪与边缘保留能力有非常重要的影响。 $g(\cdot)$ 的选取原则为在梯度较小或只有孤立噪声点的平坦区域, $g(\cdot)$ 的值较大, 平滑作用较强; 而梯度较大的突变区域, 则往往是边缘与细节, $g(\cdot)$ 值较小, 从而应保留这些局部特征。另外, 在信噪比较低时, 各向异性扩散方法也能取得较好的效果。

文献 [3] 给出了以下两个函数:

$$g(|\nabla I|) = e^{-\lambda (|\nabla I|)^2} \quad (4)$$

$$g(|\nabla I|) = \frac{1}{1 + (|\nabla I|/\kappa)^2} \quad (5)$$

它们均为单调递减函数, 当自变量增加到一定的值后, 其值急剧减小。

各向异扩散方法的离散化计算过程是, 首先分别求邻域中心像素的上、下、左、右 4 个方向的灰度差分, 以得到梯度, 再计算扩散系数, 最后求散度即可得到最终结果。

2.3 边缘保留各向异性扩散

结构张量常用于图像的局部方向估计与结构分析, 它在图像处理与计算机视觉领域应用广泛, 如纹理分析、光流场估计与非线性扩散滤波等^[9~11]。但是, 如果在噪声环境中, 利用基于梯度的结构张量来分析图像局部结构与方向, 则噪声对梯度的大小与方向均会产生较大影响, 从而使边缘及局部结构方向发生偏移。

为此, 本文先对图像进行高斯平滑, 然后构造一个高斯卷积微分的结构张量, 以减弱噪声对图像局部方向检测与结构估计的影响。

不妨令

$$D = R^T \begin{pmatrix} c_1 & 0 \\ 0 & c_2 \end{pmatrix} R \quad (6)$$

式中,

$$R = \frac{1}{\sqrt{(L_x^\rho)^2 + (L_y^\rho)^2}} \begin{vmatrix} L_x^\rho & -L_y^\rho \\ L_y^\rho & L_x^\rho \end{vmatrix} \quad (7)$$

它是在尺度 ρ 下与图像梯度方向一致的矩阵。

$$L_x^\rho = \frac{\partial(K_\rho^* I)}{\partial x} \quad (8)$$

$$L_y^\rho = \frac{\partial(K_\rho^* I)}{\partial y} \quad (9)$$

式中, L_x^ρ 与 L_y^ρ 分别表示图像的高斯卷积在 x 方向和 y 方向的微分。

则

$$D = \frac{1}{(L_x^\rho)^2 + (L_y^\rho)^2} \begin{vmatrix} c_1(L_x^\rho)^2 + c_2(L_y^\rho)^2 & (c_2 - c_1)L_x^\rho L_y^\rho \\ (c_2 - c_1)L_x^\rho L_y^\rho & c_1(L_x^\rho)^2 + c_2(L_y^\rho)^2 \end{vmatrix} \quad (10)$$

其中 c_1, c_2 为 0~1 之间的常数。

结构张量 D 实质上是在尺度 ρ 上对图像局部结构的估计, 因此也称为结构张量。若将它引入到各向异性扩散过程, 则称之为边缘保留各向异性扩散方法 (edge preserving an isotropic diffusion, EPANDI)。由于尺度 ρ 与图像的局部结构特征相关, 因此可取其方差或均值的函数作为尺度 ρ 。当没有图像局部结构

特征的任何先验知识时, ρ 可通过多次实验确定。

3 实验与比较

3.1 仿真实验

为验证边缘保留各向异性扩散方法的性能, 选取 Lina 图像 (128 × 128) 生成一系列不同信噪比的仿真斑点图像进行了处理实验, 部分图像如图 1 所示。

由于式 (10) 中的尺度 ρ 对算法的结果有极为重要的影响, 且它与噪声与图像特征相关, 因此可令

$$\rho = r \cdot \sigma_N / \sigma_S \quad (11)$$

式中, r 为常数, 其值可根据实际情况进行调整, σ_N 与 σ_S 分别表示噪声与无噪信号或图像特征的方差。

图 1 是信噪比为 10.89dB 与 3.98dB 的斑点图像, 分别应用 ANDI 方法与 EPANDI 方法处理后的结果。图 1 表明, EPANDI 方法对斑点噪声具备很强的抑制作用, 同时边缘与细节保留得较好。虽然 ANDI 方法能抑制大部分斑点, 但边缘与细节也随之被平滑, 图像变得很模糊。

实验采用一组共 6 幅具有不同信噪比的图像, 分别应用两种不同方法进行处理, 通过对处理结果



(a) 原始图像



(b) SNR = 10.89dB



(c) SNR = 3.98dB



(d) ANDI 方法对图 1(b) 的处理结果



(e) ANDI 方法对图 1(c) 的处理结果



(f) EPANDI 方法对图 1(b) 的处理结果



(g) EPANDI 方法对图 1(c) 的处理结果

图 1 不同信噪比仿真斑点图像及经处理后的结果

Fig. 1 Results processed on simulated speckle images with different SNR by EPANDI

的比较可知, EPANDI方法在抑制斑点与保留重要细节特征方面的性能均优于 ANDI方法。

为了进一步量化分析算法的性能, 并与标准各向异性扩散方法进行比较, 本文采用以下噪声抑制参数^[13]进行衡量:

$$\alpha = \frac{\sum (s - \bar{s}, \hat{s} - \bar{\hat{s}})}{\sqrt{\sum (s - \bar{s}, s - \bar{s}) \cdot \sum (\hat{s} - \bar{\hat{s}}, \hat{s} - \bar{\hat{s}})}} \quad (12)$$

其中, \bar{s} 与 $\bar{\hat{s}}$ 分别为原始(无斑点噪声)图像与处理后图像的灰度均值向量, $\sum(\cdot)$ 的定义如下:

$$\sum (s_1, s_2) = \sum_{(i,j) \in I} s_1(i,j) \cdot s_2(i,j) \quad (13)$$

即两幅图像对应像素点灰度乘积之总和。

边缘保留测度为

$$\beta = \frac{\sum (u - \bar{u}, v - \bar{v})}{\sqrt{\sum (u - \bar{u}, u - \bar{u}) \cdot \sum (v - \bar{v}, v - \bar{v})}} \quad (14)$$

其中, u 与 v 分别为 s 与 \hat{s} 经过 Laplacian算子进行高通滤波(窗口大小为 3×3)后的输出, \bar{u} 和 \bar{v} 分别为 u 和 v 的灰度均值向量。

α 的值越大, 说明对斑点的平滑作用越强, β 越大, 说明保留边缘与细节方面的能力越强。

表 1列出了图 1中不同信噪比斑点图像经过处理后的噪声抑制与边缘保留性能量化参数。表 1说明, 在斑点较少时, 只需经过几次迭代, 斑点即很快被抑制, 而且噪声平滑与边缘保留参数的值均较大, 此时算法的性能较好。随着信噪比的降低, 不仅需要的迭代次数增加, 而且噪声平滑与边缘保留参数的值也逐渐减小, 此时算法的性能随之下降。通过表 1中性能参数的比较进一步证明, EPANDI方法的性能明显优于 ANDI方法。

表 1 不同信噪比条件下算法的性能

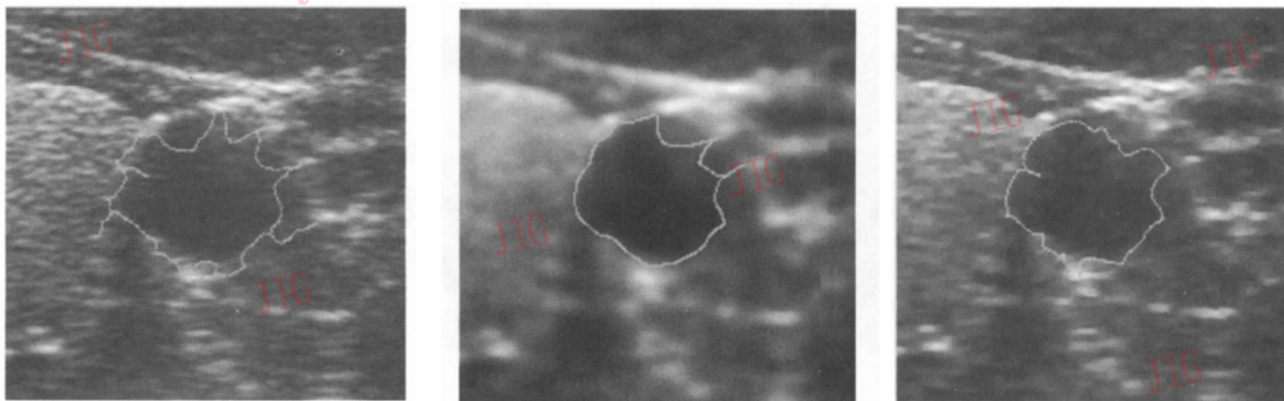
Tab 1 Performance of EPANDI applied to process simulated images with different signal-noise-ratio(SNR)

信噪比 (dB)	噪声方差	尺度 ρ	迭代次数 N	ANDI方法		EPANDI方法	
				α	β	α	β
38.43	15.94	0.39	3	0.9784	0.7830	0.9921	0.9138
24.16	25.35	0.62	4	0.9696	0.7107	0.9880	0.8791
16.90	36.23	0.89	5	0.9625	0.6638	0.9852	0.8495
10.89	56.24	0.46	8	0.9487	0.5980	0.9795	0.8103
5.94	103.18	0.25	10	0.9439	0.5908	0.9737	0.7634
3.98	153.88	0.38	12	0.9362	0.5593	0.9658	0.7166

3.2 超声斑点图像实验

选取肝脏超声图像(128 × 128), 分别应用 ANDI方法与 EPANDI方法对其进行处理, 图 2为轮廓提取实验结果。图 2(a)为原始肝脏超声图像, 其中可见大量斑点, 尤其在血管壁附近斑点的影响更为明显。图 2(b)与图 2(c)分别是肝脏超声图像经过 ANDI方法与 EPANDI方法处理后的结果, 迭代

次数均为 35, 其中 EPANDI方法的尺度 $\rho = 0.5$, $c_1 = 0.25$, $c_2 = 1.0$ 。从图 2(b)与图 2(c)可见, 虽然两种方法均对超声斑点有较强的平滑作用, 但它们在保留边缘与细节方面的性能却相差较大。图 2(c)中的边缘与局部结构比图 2(b)清晰得多。这是因为 ANDI方法在抑制斑点的同时, 随着迭代过程的继续, 边缘与细节也随之被平滑。而 EPANDI方法在



(a) 原始图像 (b) ANDI方法处理结果 (c) EPANDI方法处理结果

图 2 肝脏超声图像经 ANDI与 EPANDI方法处理后的结果及应用 GVF方法提取的血管轮廓结果比较

Fig 2 Results of ultrasonic liver images processed by the ANDI and the EPANDI contours detected by the GVF

扩散过程中, 由于引进了与图像边缘和局部结构相一致的结构张量, 因此能较好地保留图像中的重要特征。

为了进一步比较与说明上述两种方法在抑制斑点与保留细节方面的性能, 本文对处理前后的图像应用 Xu 等人提出的梯度向量流 (gradient vector flow, GVF) 方法提取血管轮廓^[12], 以便进行血管分割与精确测量。GVF 方法主要是通过引进梯度向量来对活动轮廓模型进行改进, 它对于检测“U”形或凹形轮廓具有突出的能力。在给定参数相同的条件下, 应用 GVF 方法分别对原始图像、经 ANDI 方法与 EPANDI 方法处理后的图像进行轮廓搜索。经过 60 次迭代后的轮廓提取, 结果如图 2 所示。图 2(a) 表明, 在未经任何处理的情况下, 斑点严重影响了血管轮廓的搜索过程, 其提取的轮廓并不准确。图 2(b) 中, 经过 ANDI 方法处理后, 轮廓提取结果得到很大改善, 但由于存在过平滑, 部分边缘与细节变得模糊, 因此影响了搜索轮廓的准确性。通过图 2(c) 与图 2(b) 的比较可知, 经 EPANDI 方法处理后, 应用 GVF 方法搜索得到的轮廓最准确。

4 结 论

各向异性扩散因其在抑制噪声与保留细节方面具有较好的性能, 常用于含有噪声图像的处理, 但对于特殊的超声斑点图像, 其作用并不明显。在特征检测、图像分析与处理过程中, 往往希望能检测并保留边缘与某些结构特征。由于图像的高斯卷积微分是先对图像进行高斯平滑, 再求不同方向的偏微分, 因此, 即使在存在噪声的情况下, 它仍能有效地检测图像的局部结构特征。本文设计了一种基于高斯卷积微分的结构张量, 并将其引入到各向异性扩散处理过程, 以保留图像边缘与局部结构特征。

为了验证算法的性能, 可首先生成一系列不同信噪比的仿真斑点图像, 再分别应用标准各向异性扩散与本文提出的方法对其进行处理; 然后通过处理后结果与量化性能参数的比较来说明 EPANDI 方法在平滑斑点与保留边缘方面的能力, 通过对仿真斑点图像的处理结果表明, EPANDI 方法均优于 ANDI 方法; 最后, 进行肝脏超声图像的处理实验, 实验结果表明, 这种边缘保留各向异性扩散的方法, 在抑制斑点噪声的同时, 能很好地保留图像边缘与细节, 它对临床诊断与分析、测量与图像分割等有重要的意义。

由于结构张量中 Gaussian 卷积核的尺度是决定算法性能的关键因素, 且它与图像中的噪声及特征相关, 因此, 本文也讨论了尺度的选取原则, 即可将尺度定义为噪声与图像方差 (或均值) 的函数。如果了解噪声的分布模型及图像中待检测重要特征的相关知识, 则比较容易确定尺度。如果没有任何噪声或图像特征的先验知识, 则可通过多次实验确定。

各向异性扩散是一个不断迭代的过程, 而且迭代时间或次数也对算法性能有十分重要的影响, 迭代次数过少, 噪声不能被有效地抑制; 迭代次数过多, 又存在过平滑, 细节特征会变模糊的问题, 此问题尚待进一步研究。此外, 还需采用更多的临床数据来对本文提出的方法进行验证与评价。

参考文献 (References)

- 1 Loupas T, Meddick W N, Allan P L. An adaptive weighted median filter for speckle suppression in medical ultrasonic images [J]. IEEE Transactions on Circuits and Systems, 1989, 36(1): 129~135
- 2 Dutt V, Greenleaf J F. Adaptive speckle reduction filter for log-compressed B-scan images [J]. IEEE Transactions on Medical Imaging, 1996, 15(6): 802~813
- 3 Perona P, Malik J. Scale-space and edge detection using anisotropic diffusion [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1990, 12(7): 629
- 4 You Y L, Kaveh M. Fourth-order partial differential equations for noise removal [J]. IEEE Transactions on Image Processing, 2000, 9(10): 1723~1730
- 5 Whitaker R T, Pizer S M. A multi-scale approach to nonuniform diffusion [J]. Computer Vision Graphics Image Processing Image Understand, 1993, 57(1): 99~110
- 6 Black M J, Sapiro G, Hegger D. Robust anisotropic diffusion [J]. IEEE Transactions on Image Processing, 1998, 7(3): 421~432
- 7 Weickert J. Theoretical foundations of anisotropic diffusion in image processing [J]. Computing, 1996, 11(Suppl): 221~236
- 8 You Y L, Xu W Y, Tannenbaum A, et al. Behavioral analysis of anisotropic diffusion in image processing [J]. IEEE Transactions on Image Processing, 1996, 5(11): 1539~1552
- 9 Bigun J, Granlund G H, Wiklund J. Multidimensional orientation estimation with applications to texture analysis and optical flow [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1991, 13(8): 775~790
- 10 Rao A R, Schunck B G. Computing oriented texture fields [J]. Computer Vision Graphics Image Processing Graphical Models and Image Processing, 1991, 53(2): 157~185
- 11 Weickert J, Ronen Y, Berman T H, Viergever M A. Efficient and reliable schemes for nonlinear diffusion filtering [J]. IEEE Transactions on Image Processing, 1998, 7(3): 398~410
- 12 Xu C Y, Prince J L. Snakes, shapes and gradient vector flow [J]. IEEE Transactions on Image Processing, 1998, 7(3): 359~369

一种新的 SAR 图像快速自适应去斑算法

李应岐^{1,2)} 何明一¹⁾

¹⁾ (西北工业大学电子信息学院 信息获取与处理陕西省重点实验室, 西安 710072) ²⁾ (第二炮兵工程学院, 西安 710025)

摘要 针对 SAR 图像斑点噪声的滤除, 提出了一种新的基于 Countourlet 变换的快速自适应性噪声去除方法。鉴于 SAR 图像的 Countourlet 系数主要取决于斑点噪声和信号腐化, 且呈现出很强的非高斯分布特性, 据此, 首先建立了 SAR 图像 Countourlet 系数的高斯混合分布解析模型; 然后用每个系数的邻域系数通过估计其去斑收缩因子来表现系数的自适应收缩; 最后对 Lee 滤波、Foster 滤波、Gamma 滤波、小波、Curvelet 和 Contourlet 变换的去斑性能进行了比较分析。实验结果表明, 该新方法在保留细节和锐化图像的同时, 能强有力地抑制斑点噪声。

关键词 Contourlet 变换 合成孔径雷达图像 斑点 自适应性收缩估计

中图法分类号: TP391.41 TN911.73 文献标识码: A 文章编号: 1006-8961(2008)02-0214-05

A New Fast Adaptive Algorithm for Despeckling SAR Images

LI Ying-qi^{1,2)}, HE Ming-yi¹⁾

¹⁾ (College of Electronic Information Northwestern Polytechnical University Shanxi Key Laboratory of Information Acquisition and Processing, Xi'an 710072)

²⁾ (Institute of Second Artillery Engineering, Xi'an 710025)

Abstract A new adaptive contourlet transform-based technique for speckle removal from SAR images is presented. The distribution of the SAR image's contourlet coefficients is mainly decided by speckle noise and useful signals corrupted. The analytic model for these distributions is proposed. The shrinkage factor for despeckling is estimated with the neighboring reference contourlet coefficient in every sub bands. Finally, the comparison of performance of Lee filter, Frost filter, Gamma filter, wavelet-based despeckling and contourlet transform-based despeckling is provided for both simulated and actual SAR images. It shows that the contourlet methods strongly suppress speckle, while preserving image details and sharpness.

Keywords contourlet transform, SAR image, speckle, adaptive shrinkage estimator

1 引言

合成孔径雷达 (synthetic aperture radar, SAR) 具有全天候, 全天时的成像能力, 是目前最具吸引力的遥感传感器。斑点噪声作为 SAR 图像的机理特性, 被认为是一种乘性噪声。其严重影响了图像的一致性, 使 SAR 图像自动解译变得极其困难。目前已提出的去斑方法基本分为滤波和变换两类。其中, 中值滤波、Lee 滤波、Frost 滤波、Kuan 滤波和

Gamma MAP 滤波^[1]等方法虽有一定的效果, 但其方法对滤波窗口的大小和形状很敏感, 即窗口太大 (相对于感兴趣区域), 会导致过平滑, 使图像的边缘模糊; 而窗口过小, 则会降低平滑能力和去斑效果; 而且常用的方形窗口会使非竖直、非水平和非 45° 线的直角特征钝化; 另一种基于变换的方法是先在变换域对变换系数阈值化, 然后重构图像去斑, 如小波变换、Curvelet 变换消噪^[2~5]等。其中, 小波变换在边缘附近抑制了边缘方向以外的平滑。但是 Curvelet 变换比小波变换在视觉上表现出较高的重

收稿日期: 2006-08-01; 改回日期: 2006-09-18

第一作者简介: 李应岐 (1965~), 现为第二炮兵工程学院教师, 西北工业大学在职博士研究生。主要从事图像处理、作战效能评估方面的研究。E-mail: yingqi65@126.com

构质量^[6], 特别是较好地保持了线形和曲线形边缘特征, 然而该方法计算量太大。Do 和 Vetterli 基于 Curvelet 变换思想, 提出了一种基于 2 维多尺度和多方向滤波器组的 Contourlet 变换算法^[7-9]。它不仅具有小波变换的多尺度和局部时频特征, 而且还具有多方向和各向异性, 因此能很好处理图像的轮廓。与其他多尺度多方向分析相比, Contourlet 变换在不同尺度下的分解方向可以不同。本文通过分析 Contourlet 系数的统计关系, 在研究 SAR 图像在不同尺度空间和方向子带下的 Contourlet 系数关系的基础上, 建立了有效的系数估计算法。

2 斑点噪声模型

含斑点噪声的 SAR 图像通常被认为是一个乘性模型:

$$\begin{aligned} \hat{I}(i, j) &= I(i, j)S(i, j) = I(i, j)[1 + \hat{S}(i, j)] \\ &= I(i, j) + W(i, j) \end{aligned} \quad (1)$$

其中, $\hat{S}(i, j) = S(i, j) - 1$, $W(i, j) = I(i, j)\hat{S}(i, j)$, $I(i, j)$ 表示目标的真实雷达回波值, $\hat{I}(i, j)$ 表示雷达回波的测量值, $S(i, j)$ 为斑点噪声, i 和 j 分别表示图像像素所处的行和列。若把乘性斑点噪声 $S(i, j)$ 转换为加性噪声 $W(i, j)$, 就避免了对数变换, 从而降低了处理难度。对于单视 SAR 图像, $S(i, j)$ 服从瑞利分布 (幅度图像) 或均值为 1 的负指数分布 (亮度图像); 而对于多视 SAR 图像, $S(i, j)$ 则服从均值为 1 的伽玛分布。

3 Contourlet 系数的高斯混合分布模型

Contourlet 变换包含了 2 的任意次幂个方向, 且具有灵活长宽比的基函数。与小波变换相比, Contourlet 变换在表示光滑轮廓时, 只需很少的系数。Contourlet 变换通过 2 维方向滤波器组, 可将图像在多尺度下分解成不同方向的子带。其在每一尺度下, 先由 Laplacian 塔形分解算法将逼近图像分解成低频和高频的两个部分, 然后再用方向滤波器组将高频部分分解成不同的方向子带。由于塔形分解是阶梯结构, 因此其尺度分解和方向分解是相互独立的, 即每个尺度下可以分解的方向数为 2 的任意幂次个, 而其他多尺度方向分析的方向数通常是固定的。

Contourlet 变换系数的边际分布在振幅为 0 处

表现为尖锐的顶峰, 峰的两侧有很粗的余尾。这表明自然图像的 Contourlet 系数分布稀疏, 即大多数系数接近于 0。在 Contourlet 域, 子带的边缘分布峰态比远高于正态分布的 3, 呈现出很强的非高斯分布特征^[10], 而高斯混合模型 (GSM) 则适合描述系数的边缘分布和联合分布^[11]。SAR 图像的 Contourlet 系数 x 是图像 $I(i, j)$ 的分解系数 x 与噪声 $N(i, j)$ 的分解系数 w 的和, 即 $x = x + w$ 。系数 x 可以通过其局部邻域的系数向量 X 来进行估计, X 可表示为一个高斯向量和一个无关数的乘积。一般地, 设 X 表示系数 x 的局部邻域的系数向量, \sqrt{z} 为乘性因子, U 为均值为 0、协方差为 C_U 的高斯向量, 则 X 的基本 GSM 模型可表示为 $X = \sqrt{z}U$, 其中 z 是隐性乘数, U 和 z 是独立的。系数 x 局部邻域的系数向量 X 的概率密度可表示为

$$P(X) = \int \frac{1}{(2\pi)^{N/2} |zC_U|^{1/2}} \exp\left\{-\frac{X^T C_U^{-1} X}{2z}\right\} p(z) dz \quad (2)$$

N 表示向量 U 和 X 的维数, 局部邻域通常定义为以系数 x 为中心的方形区域。

4 Contourlet 域系数的自适应去斑因子估计

假设向量 Y 是向量 X 与方差为 σ_w^2 、均值为 0 的高斯白噪声向量 W 的和, 即 $Y = X + W$ 。如果 X 是 GSM 模型的向量, 则 $Y = \sqrt{z}U + W$ 。如果 z 的值已知, 那么 Y 就是高斯分布, X 的最佳估计可用线性维纳滤波实现。

$$\hat{x} = \frac{z\sigma_U^2}{z\sigma_U^2 + \sigma_w^2} y \quad (3)$$

因此, 只要从向量 Y 中估计出 $z\sigma_U^2$, 就可以得到 X 的最佳估计 \hat{x} 。基于向量 Y 的噪声乘性因子 z 的最大似然估计为

$$\hat{z}_{\max}(Y) = (1/\sigma_U^2)(Y^T Y N - \sigma_w^2) \quad (4)$$

将式 (4) 代入数据计算, $\hat{z}_{\max}(Y)$ 经常会得到很小的负值, 其主要是由于邻域的势较小或噪声水平过高所致。这种情况下, 若令 $\hat{z}_{\max}(Y)$ 为 0 则有

$$\hat{z}\sigma_U^2 = \hat{z}_{\max}(Y)\sigma_U^2 = \max(Y^T Y N - \sigma_w^2, 0) \quad (5)$$

假定 $y_{i, j, s, d}$ 表示含噪 SAR 图像在尺度为 2^s、方向为 d 的子带上位置 (i, j) 的 Contourlet 系数, 对应的真实图像的系数记为 $c_{i, j, s, d}$, 则 $c_{i, j, s, d}$ 的值可用

$y_{i,j,s,d}$ 的邻域系数进行估计:

$$c_{i,j,s,d} = \eta y_{i,j,s,d}, \quad \eta = \frac{z\sigma_U^2}{z\sigma_U^2 + \sigma_W^2} \quad (6)$$

η 为 Contourlet 域的去斑因子, Y 表示系数 $y_{i,j,s,d}$ 邻域的 Contourlet 系数向量, σ_W^2 是图像的噪声方差, η 由系数的局部邻域和所在子带的方差确定。

基于 Contourlet 变换的自适应降噪过程可分为下面 3 个步骤:

① 对图像进行 Contourlet 多分辨率分解, 得到多尺度多方向子带;

② 对每个子带的每个系数用式 (5) 计算 $z\sigma_U^2$, 用式 (6) 进行收缩。

③ 用收缩产生的系数进行 Contourlet 逆变换, 重构图像。

为了使用式 (5) 计算 $z\sigma_U^2$, 就需要估计噪声方差 σ_W^2 , 该参数可以从子带的 Contourlet 系数的标准方差与图像的总方差中得到。

5 降斑效果评估

通常对消噪造成的图像失真进行准确评估是十分困难的, 因为简单的数学测度计算往往缺少视觉生态特征的支持, 而视觉生态的特征 (可视化特征) 又是难以表示、计算和测量的。由于 SAR 图像对于普通人是难以理解的, 而采用视觉特征进行去斑效果评估则更为困难, 因此本文采用噪声变差 (noise variance, NV)、均方差 (mean square error, MSE)、等价视数 (equivalent numbers of looks, ENL)、偏差比 (deflection ratio, DR)^[1] 等数学测度进行去斑效果评价。

5.1 噪声变差

噪声变差表示图像被斑点噪声污染的程度, NV 值越小, 图像越干净, 即图像去斑噪越彻底, 其计算公式为

$$NV = \frac{1}{R \times C} \sum_{r=1}^R \sum_{c=1}^C (f_{r,c}^{\wedge} - m)^2 \quad (7)$$

其中, R, C 为图像的行数和列数, $f_{r,c}^{\wedge}$ 为图像降噪后的像素值, m 是 $f_{r,c}^{\wedge}$ 的均值。

5.2 均方差

均方差表示了图像所有像素的平均误差。本文中的均方差不是图像的损失和误差, 而是消噪后图像和噪声图像的像素值偏差的综合 (对于光学仿真斑噪图像而言, 就是评价去噪图像和原始图像的像

素偏差总和)。MSE 值越大, 表示降噪效果越明显。其计算公式为

$$MSE = \sum_{j=1}^N (f_j - \hat{f}_j)^2 \quad (8)$$

其中, f_j 和 \hat{f}_j 分别为降噪前后的像素值, N 为图像的像素总数。

5.3 等价视数

等价视数 (ENL) 是估计 SAR 图像斑噪水平的常用指标。通常是通过图像匀质区域的等价视数来估计整幅图像的等价视数, 因此与检测的匀质区域大小有关。理论上说, 大匀质区域的等价视数高于小匀质区域的等价视数, 然而区域越大, 准确性越小。为了避免对匀质区域大小的判断, 本文先将图像分成 32×32 大小互不交叠的块, 并对每个小块求其 ENL, 然后用所有块的平均值表示整幅图像的等价视数。匀质区域的 ENL 计算如下:

$$ENL = m^2 / s^2 \quad (9)$$

m, s 分别是匀质区域的均值和标准差。

5.4 偏差比

偏差比 (DR) 是由 Guo 等人在 1994 年提出的, 其计算公式为

$$DR = \frac{1}{R \times C} \sum_{r=1}^R \sum_{c=1}^C \left| \frac{\hat{f}_{r,c} - m}{s} \right| \quad (10)$$

$\hat{f}_{r,c}$ 为图像降噪后的图像像素值, m 是 $\hat{f}_{r,c}$ 的均值, s 是 $\hat{f}_{r,c}$ 的标准差。在雷达回波强点的 DR 值比较大, 本文用它表示图像的降斑效果。DR 值越小, 表明图像消除斑噪的程度越大。

6 实验结果

为验证本文方法的去噪效果, 通过对附加乘性噪声 μ 的光学图像 I 进行去噪实验, 得到了去噪后的仿真 SAR 图像 J , 其中 μ 是均值为 0、标准差为 σ_{μ} 的均匀分布的变量。对于方差 σ_{μ} 分别为 15、25 和 35 的 Lena 仿真斑噪图像, 采用 Lee 滤波、Frost 滤波、中值滤波和 Gamma 滤波对其去斑消噪, 并用基于小波系数的隐马尔可夫模型、Curvelet 变换和基于 Contourlet 变换进行自适应收缩消噪, 其去斑效果见图 1。去斑图像的噪声方差、均方差、等效视数和偏差率见表 1。

从图 1 可明显看出, 经典滤波方法不能有效地减少斑点, 且视觉效果较差。虽然小波变换在去噪时减少了斑点, 但是细节和轮廓没有完好保存下来, 例如



图 1 原始 Lena 图像, 模拟含噪图像和不同方法去斑图像

Fig. 1 Lena image, simulated image and de-speckled image

表 1 仿真斑噪图像 (加乘性噪声的 Lena) 的不同去斑方法的去斑性能对比

Tab 1 Comparison of several de-speckled algorithms for simulated SAR image

去斑方法	NV	MSE	DR	ENL
噪声图像	45.50	675.14	0.712	10.15
Lee 滤波	38.91	169.15	0.704	35.42
增强 Lee 滤波	37.66	91.87	0.706	54.47
Frost 滤波	37.63	92.51	0.706	54.51
增强 Frost 滤波	37.86	96.14	0.705	53.42
Kuan 滤波	42.85	455.73	0.711	14.02
中值滤波	38.46	131.01	0.708	40.55
Gamma 滤波	39.45	180.68	0.691	45.73
Wavelet HMMT	36.99	81.85	0.690	148.67
Curvelet 变换	37.04	77.49	0.689	147.20
Contourlet 变换	36.05	78.54	0.686	250.11

帽子的边缘和头发的纹理相混淆。Curvelet 变换的去斑效果较好, 基于 Contourlet 变换的自适应收缩方法不仅能很好去除斑点, 同时能保持细节和轮廓。

图 2(a) 为一幅内容复杂的 RADARSAT 图像, 图 2(b) 和图 2(c) 是经 Lee 滤波和 Gamma 滤波的图

像。图 2(d), 图 2(e) 和图 2(f) 分别是使用小波隐马尔科夫模型、Curvelet 变换和 Contourlet 技术去斑的图像。从各种方法的去斑结果可以看出 Contourlet 变换的去斑效果好于其他去噪技术, 而且 Contourlet 变换的去斑没有降低图像的视觉, 而小波变换和 Curvelet 变换却减少了图像的亮度, 造成了边缘的混乱。各类去斑方法的去斑结果比较见表 2。

表 2 RADARSAT 图像不同去斑方法的去斑性能对比

Tab 2 Comparison of several de-speckled algorithms for RADARSAT image

去斑方法	NV	MSE	DR	ENL
原始图像	31.606	-	0.456	9.190
Lee 滤波	25.038	135.48	0.455	17.869
增强 Lee 滤波	22.315	358.29	0.464	25.202
Frost 滤波	22.148	377.55	0.463	25.829
增强 Frost 滤波	24.229	298.17	0.458	20.049
中值滤波	34.129	196.5	0.426	17.514
Gamma 滤波	23.366	378.66	0.461	22.025
Wavelet HMMT	24.769	179.17	0.444	18.946
Curvelet 变换	23.510	327.04	0.457	22.290
Contourlet 变换	22.090	321.14	0.457	27.036

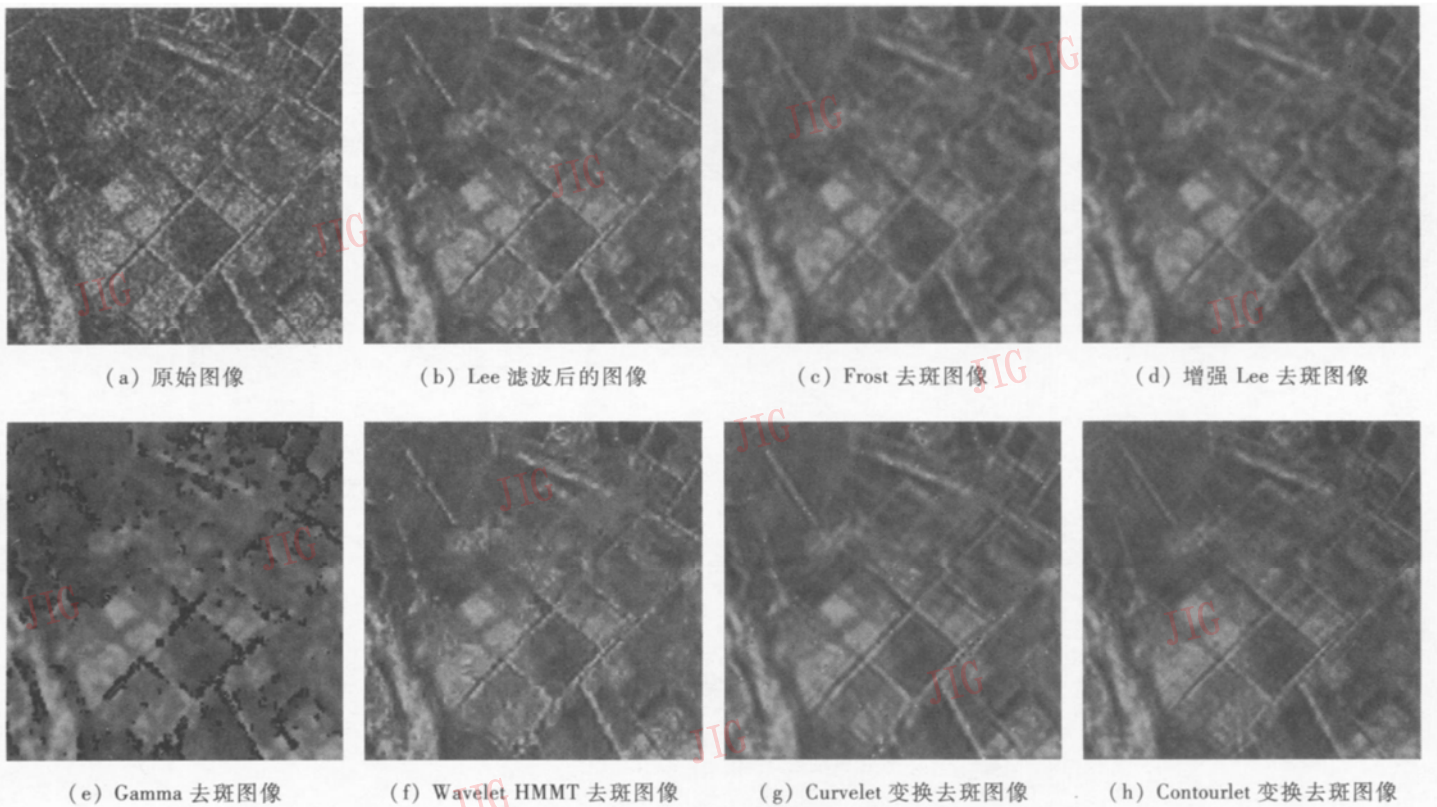


图 2 RADARSAT SAR 图像和去斑图像

Fig. 2 RADARSAT SAR image and de-speckled image

7 结 论

本文提出了一种新的快速有效的 SAR 图像去斑方法。该方法主要的创新点在于利用 Contourlet 系数的高斯混合分布模型, 在进行降斑去噪的同时, 还对轮廓和边缘信号特征提供了稀疏表示; 同时提出了针对 SAR 图像灰度分布和自动直接计算参数的解析模型。将这些分布模型应用于在不同尺度和不同方向参考系数的领域内, 不仅得到了非常好的斑点噪声变化估计和真实 Contourlet 系数的最优逼近, 而且在很好保存图像细节的同时, 实现快速去斑。

参考文献 (References)

- 1 Tan Hong Sem. Denoising of Noise Speckle in Radar Image[D]. Thesis for the Degree of Bachelor of Engineering. The University of Queensland, Australia, October 2001.
- 2 Walessa M, Datcu M. Model-based despeckling and information extraction from SAR images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2000, 38(9): 2258~2269.
- 3 Matriani M, Giraklez A E. Smoothing of coefficients in wavelet domain for speckle reduction in synthetic aperture radar images[J]. International Journal on Graphics, Vision and Image Processing, 2005, 5(8): 1~8.
- 4 Chang S G, Yu B, Vetterli M. Adaptive wavelet thresholding for image denoising and compression[J]. IEEE Transactions on Image Processing, 2000, 19(9): 1532~1546.
- 5 Argenti F, Alparone L. Speckle removal from SAR images in undecimated wavelet domain[J]. IEEE Transactions on Geoscience and Remote Sensing, 2002, 40(11): 2363~2374.
- 6 Starck J, Candes E, Donoho D. The curvelet transform for image denoising[J]. IEEE Transactions on Image Processing, 2002, 11(6): 670~684.
- 7 Do M N, Vetterli M. The Contourlet transform: an efficient directional multiresolution image representation [J]. IEEE Transactions on Image Processing, 2005, 14(12): 2091~2106.
- 8 Do M N, Vetterli M. Contourlets in Beyond Wavelets[M]. New York: Academic Press, 2003.
- 9 Eslam i R, Radha H. The Contourlet transform for image denoising using cycle spinning[A]. In: Proceedings of A Sirkar Conference on Signals, Systems, and Computers[C], Pacific Grove, 2003: 1982~1986.
- 10 Po D D-Y, Do M N. Directional multiscale statistical modeling of images [A]. In: Proceedings of SPIE Conference on Wavelet Applications in Signal and Image Processing X[C], San Diego, CA, USA, 2003: 69~79.
- 11 Portilla J, Strela V, Wainwright M, et al. Image denoising using scale mixtures of Gaussian in wavelet domain[J]. IEEE Transactions on Image Processing, 2003, 50(11): 1338~1351.

自适应的图像组合降噪

周登文 申晓留

(华北电力大学计算机科学与技术系, 北京 102206)

摘要 BayesShrink 是小波收缩降噪最好的算法之一, 而 WienerChop 方法则是利用小波域维纳滤波改进了 VisuShrink 算法。为了更好地滤除噪声, 研究了 WienerChop 组合 BayesShrink 进行降噪的方法。实验表明, 该组合算法优于 WienerChop 和 BayesShrink 算法, 其可产生更低的均方误差和更高的信噪比。它不仅综合了 WienerChop 和 BayesShrink 两种算法的优点, 而且改善了 WienerChop 算法的过光滑和 BayesShrink 算法残留较多噪声的问题, 同时可获得视觉上更为愉悦的降噪图像。

关键词 自适应 图像降噪 图像恢复 小波

中图分类号: TP391.41 TN911.73 文献标识码: A 文章编号: 1006-8961(2008)02-0219-06

Adaptive Combined Image Denoising

ZHOU Deng-wen, SHEN Xiao-liu

(Department of Computer Science and Technology, North China Electric Power University, Beijing 102206)

Abstract BayesShrink is one of the best algorithms for wavelet thresholding denoising while WienerChop improves VisuShrink by Wiener filtering in wavelet domain. We studied the denoising method uniting BayesShrink and WienerChop. The combined algorithm has smaller mean squared error (MSE) and higher signal to noise ratio (SNR) than BayesShrink or WienerChop. It integrates the advantages of the two algorithms and improves the problems which images are smoothed overly by WienerChop and BayesShrink retains some noise artifacts. It can visually obtain more pleasing denoised images.

Keywords adaptive method image denoising image restoration wavelet thresholding

1 引言

图像在获取或传输过程中, 常常污染有噪声, 在对图像做进一步处理之前, 必须去除或减少这些噪声。数字图像通常可表示为一个 2 维的灰度矩阵, 矩阵中的每一个元素值 $u(i, j)$ 表示屏幕上像素点的灰度或亮度, 而矩阵中的元素位置 (i, j) 对应屏幕上像素点的位置。由于大多数图像噪声造型为加高斯白噪声, 所以图像降噪问题可以描述为: 观察数据 $Y_{i, j}$ 为信号 $X_{i, j}$ 与噪声 $\epsilon_{i, j}$ 的叠加, 即

$$Y_{i, j} = X_{i, j} + \epsilon_{i, j} \quad i=1, \dots, I, \quad j=1, \dots, J \quad (1)$$

其中, $\{\epsilon_{i, j}\}$ 独立于信号, 且是独立恒等分布 (iid) 的高斯随机变量, 即 $\epsilon_{i, j} \sim N(0, \sigma^2)$, σ^2 是噪声的方差。

本文的目标是要基于 $Y_{i, j}$ 找到 $X_{i, j}$ 的最优估计, 而测度最优估计的客观准则是最小均方误差 (MSE)。

去除高斯噪声的传统方法是基于线性滤波的去噪方法, 例如空间域滤波方法有均值滤波、高斯滤波。维纳方法也是一种线性滤波, 它是基于傅里叶变换在频率域滤波。在线性滤波方法中, 维纳滤波是最优的^[1]。

Donoho 和 Johnstone 开创了基于小波收缩或取阈值的非线性降噪方法, 称之为 VisuShrink 方法^[2], 并证明在极小极大化均方误差意义上, VisuShrink 方法具有几乎最优的性质, 用其逼近 Besov 空间函数比线性方法具有更好的收敛率。虽然 VisuShrink 方法简单有效, 但是, 存在的问题是会过度光滑图像, 使降噪后的图像没有清晰的边。为了克服

收稿日期: 2006-08-13 改回日期: 2006-09-25

第一作者简介: 周登文 (1965~), 男, 华北电力大学计算机科学与技术系副教授。主要研究方向为图像降噪和图像压缩。E-mail: zdw@ncepu.edu.cn

VisuShrink方法的缺陷,小波收缩方法得到了广泛和深入的研究,已出现了大量的新的小波收缩方法。其中在基于抽样的正交小波变换中,SureShrink方法^[3]和 BayesShrink方法^[4]是最优的^[5]。与 SureShrink和 BayesShrink方法不同,Ghael等人构造了小波域维纳滤波器^[6],称之为 WienerChop方法,它将 VisuShrink方法用于真实信号的估计,WienerChop方法胜出了许多小波收缩的方法。本文中通过研究组合 WienerChop方法和 BayesShrink或 SureShrink方法发现,这种组合方法优于上述的任何一种去噪方法。

2 小波收缩降噪

小波收缩降噪的基本思想是:将有噪图像数据作小波变换,并将大的变换系数对应到信号,小的变换系数对应到噪声。假定式(1)中的观察数据、无噪信号数据和噪声分别用矩阵 Y, X 和 ε 表示, W 表示正交小波变换矩阵,且 $w = WY, \theta = WX, z = W\varepsilon$ 则小波降噪过程分为以下 3 步:

- (1) 计算小波系数矩阵 w ;
- (2) 将对 w 细节系数取的阈值作为 X 的小波系数估计 \hat{w} ;
- (3) 对取阈值后的系数进行逆变换,即可获得降噪后的信号数据估计 $\hat{X} = W^{-1}\hat{w}$ 。

本文使用均方误差评估降噪性能。对于 $X(i, j)$ 的降噪估计 $\hat{X}(i, j)$, 均方误差 (MSE) 定义为

$$MSE = \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J (X(i, j) - \hat{X}(i, j))^2 \quad (2)$$

在上述 3 个步骤中,第 1 步和第 3 步是平凡的,关键是第 2 步,因为它涉及到阈值如何选取和对小波系数如何收缩两个问题,也就是选择怎样的收缩函数。通常有两种常用的收缩函数,其中软收缩函数 $\delta_{\lambda}^s(w) = \text{sgn}(w) \cdot \max(|w| - \lambda, 0)$ 将绝对值比阈值大的系数向零收缩;而硬收缩函数 $\delta_{\lambda}^h(w) = w \cdot \mathbf{1}_{\{|w| > \lambda\}}$ 则保留绝对值比阈值大的系数,而将绝对值比阈值小的系数设置为零。由于硬收缩函数不连续,其不仅具有更大的方差,并且不够稳定,即对数据变化敏感,因此,软收缩函数是优选的^[2]。

VisuShrink阈值为 $\lambda = \sigma \sqrt{2 \log(M)}$, 它被称之为 universal 阈值, M 是式(1)中 I 和 J 中的较大者。本文处理 $I \times I$ 大小的图像,取 $\lambda = \sigma \sqrt{2 \log(I)}$, 其中 σ 是估计的噪声标准差。实际上,这个阈值偏

大, VisuShrink方法将这个阈值应用到小波各个分解尺度的细节子带,其虽然能够压抑噪声,但是也导致图像的过度光滑。为了保持更多的图像特征,由数据驱动的自适应的阈值选择是必要的。SureShrink阈值是数据驱动的。它是使用软收缩函数,通过最小化 Stein 的无偏风险估计来导出多个分解子带的阈值^[3]。假定 $w_s = \{w_n; n=1, \dots, N_s\}$ 表示第 s 个子带的细节系数,且噪声方差为 1, 则该子带的 SureShrink 阈值为

$$\lambda_s^S = \arg \min_{\lambda > 0} SURE^S(\lambda, w_s) \quad (3)$$

其中

$$SURE^S(\lambda, w_s) = N_s + \sum_{n=1}^{N_s} [m \ln(|w_n|, \lambda)]^2 - 2[\# \text{ of } w_n: |w_n| \leq \lambda]$$

如果噪声方差不为 1, 则要首先估计噪声方差,并将观察数据 Y 规范化,使其方差为 1。BayesShrink 阈值也是用数据驱动的,它是假定子带小波系数服从广义高斯分布 (GGD), 并使用软收缩函数通过最小化贝叶斯风险导出阈值^[4]。第 s 个子带的 BayesShrink 阈值为

$$\lambda_s^B = \frac{\sigma^2}{\sigma_x} \quad (4)$$

σ^2 是估计的噪声方差, σ_x^2 是考虑的子带上的信号方差。信号标准差 $\sigma_x = \sqrt{\max(\sigma_y^2 - \sigma^2, 0)}$, 其中 $\sigma_y^2 = \frac{1}{N_s} \sum_{n=1}^{N_s} w_n^2$ 是观察数据 Y 的变换系数在该子带上的方差估计。如果 $\sigma^2 \geq \sigma_y^2$, 则阈值 $\lambda_s^B = \max(|w_n|)$ 。

在计算任一阈值之前,都要涉及到噪声标准差的估计问题。其可以用最精细分解层的对角子带系数绝对值的中值来估计,因为该子带的系数基本上都是噪声^[2,3]。

$$\sigma = \frac{\text{median}(|w_s|)}{0.6745} \quad (5)$$

3 WienerChop 方法

小波收缩估计只是在极小极大化均方误差意义上渐近最优的^[2], 然而,对于有限长度的信号来说,使均方误差最小的最优滤波器是维纳滤波器^[6]。如前所述,假定真正的小波系数为 θ 噪声方差为 σ , 则维纳滤波器^[1,6]为

$$h_w(i, j) = \frac{\hat{\theta}^2(i, j)}{\hat{\theta}^2(i, j) + \sigma^2} \quad (6)$$

其中, $i = 1, \dots, I; j = 1, \dots, J$ 。

由于 σ 和 $\theta(i, j)$ 均是未知的, 因此需要进行估计。 σ 可用式 (5) 估计, WienerChop 滤波采用 VisuShrink 硬收缩函数来估计 $\theta(i, j)$, WienerChop 滤波过程如图 1 所示。

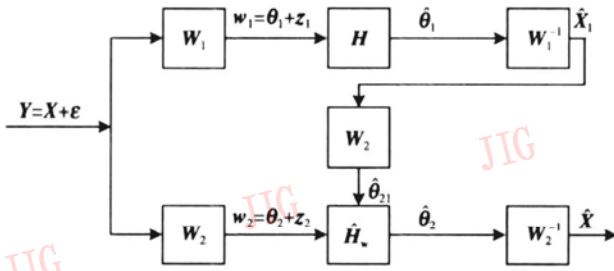


图 1 WienerChop 滤波过程

Fig 1 WienerChop filtering process

在上面的路径中, 变换 W_1 用于生成信号 X 的估计 \hat{X}_1 , H 是与硬收缩函数对应的滤波器。变换 W_2 作用于 \hat{X}_1 , 即可得到系数 $\hat{\theta}_{21}(i, j)$, 其可作为真正的小波系数 $\theta(i, j)$ 的估计, 根据式 (6) 便可得到经验的维纳滤波器

$$\hat{h}_w(i, j) = \frac{\hat{\theta}_{21}^2(i, j)}{\hat{\theta}_{21}^2(i, j) + \sigma^2} \quad (7)$$

其中, $i = 1, \dots, I; j = 1, \dots, J$ 。

再通过下面的路径对观察数据 Y 进行 W_2 变换、 \hat{H}_w 滤波和 W_2^{-1} 反变换, 便可得到真正信号 X 的估计 \hat{X} 。WienerChop 方法要求变换 W_1 和 W_2 不能相同。

WienerChop 滤波之所以胜出了许多小波收缩方法, 它可以解释为: 由于小波系数取阈值只保留了大的小波系数, 而比阈值小的信号系数均收缩为零, 因此造成了次最优的结果。由于 WienerChop 方法通过变换 W_1 和 W_2 的不匹配, 有效地伸展了较小的信号系数, 即带回了部分小的不确定的信号系数, 从而改进了估计性能^[6]。

4 自适应的组合降噪

大家知道, WienerChop 方法利用 VisuShrink 方法, 改进了小波域维纳滤波方法, 提高了估计性能。VisuShrink 方法被用于先导估计真正的信号, 以便生成维纳滤波器。很显然, 如果真正信号的估计越精确, 则维纳滤波效果应该越好。大家也知道,

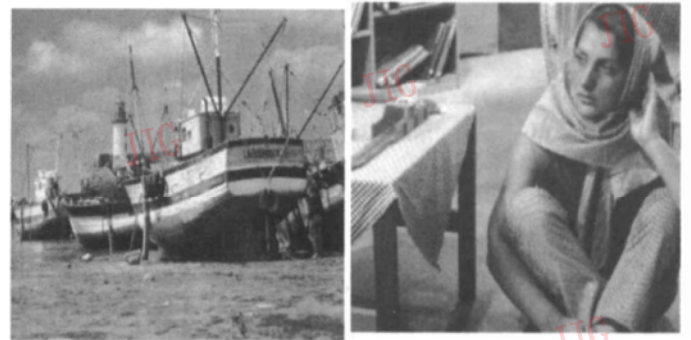
SureShrink 和 BayesShrink 方法比 VisuShrink 方法估计效果更好。一个很自然的问题是: 能不能将 WienerChop 和 BayesShrink 或 SureShrink 组合起来, 产生更好的降噪效果呢? 答案是肯定的。

为验证本文方法的降噪效果, 对一组 512×512 大小的标准测试图像 Lena、Goldhill、Boats 和 Barbara 用前述的方法进行了降噪处理效果比较。它们污染的高斯噪声标准差分别为 1Q、2Q、3Q 和 6Q。原图像如图 2 所示。



(a) Lena 图像

(b) Goldhill 图像



(c) Boats 图像

(d) Barbara 图像

图 2 原图像

Fig 2 Original images

本文采用的衡量去噪效果的客观标准是均方误差, 各种方法针对不同的图像和噪声方差进行降噪, 得到的均方误差列在表 1 中, 最小的均方误差被加粗表示。

表 1 中列出了 7 种方法去噪效果的比较, Wiener 即基于傅里叶变换的经典的维纳滤波方法, 是用 Matlab 图像处理工具箱 Wiener2 命令缺省的设置 (3×3 的局部窗口尺寸) 实现的; Visu-S 即 VisuShrink 方法, Sure-S 即 SureShrink 方法, Bayes-S 即 BayesShrink 方法, Wiener-C 即 WienerChop 方法, Wiener-S 即 WienerChop 组合 SureShrink 方法, Wiener-B 即 WienerChop 组合 BayesShrink 方法; 所有小波变换使用 4 个分解层, Visu-S、Sure-S、Bayes-S

表 1 7种去噪方法的去噪后图像均方误差比较
Tab 1 Denoising results(MSE) of 7 denoising methods

图像	σ	不同去噪方法的均方误差						
		Wiener	Visu-S	Sure-S	Bayes-S	Wiener-C	Wiener-S	Wiener-B
Lena	10	28.49	43.00	28.44	29.69	27.46	24.18	24.50
	20	81.98	87.48	59.26	61.47	58.33	51.93	53.22
	30	174.89	131.92	88.94	90.90	90.19	80.17	82.57
	60	672.49	249.55	168.19	169.13	176.74	154.47	156.18
Goldhill	10	43.00	72.92	42.42	42.03	49.37	38.33	38.04
	20	97.04	133.49	87.54	89.42	95.54	80.16	81.11
	30	189.97	183.75	125.93	126.67	135.84	117.20	118.46
	60	687.79	308.60	215.06	213.83	235.58	203.21	204.04
Boats	10	43.84	71.76	41.50	41.48	45.04	36.01	36.01
	20	98.34	140.34	88.80	90.82	94.51	79.09	79.66
	30	192.63	203.32	136.40	139.07	145.64	122.62	124.54
	60	696.32	353.61	251.97	253.56	270.64	236.75	238.56
Barbara	10	66.93	82.91	47.81	50.63	47.96	38.17	39.35
	20	135.24	192.74	118.23	120.17	128.01	101.27	103.55
	30	241.93	294.21	189.30	190.69	210.50	170.68	171.63
	60	766.15	466.47	354.49	359.69	372.33	335.67	348.95

方法和 Wiener-C 方法的变换矩阵 W_2 是使用 Daubechies symm let 8 个零矩小波, 而 Wiener-C 方法的变换矩阵 W_1 则是使用 Daubechies symm let 4 个零矩小波; 均方误差按式 (2) 计算, SureShrink 阈值按式 (3) 计算, BayesShrink 阈值按式 (4) 计算, 噪声方差按式 (5) 估计, Wiener-C 维纳滤波器按式 (7) 计算。

从表 1 可以看出, 当噪声标准差较大时, 线性降噪最优方法 Wiener 滤波的去噪结果相当坏, Visu-S 方法胜出了维纳方法。无论噪声级高低, Wiener 和 Visu-S 方法的去噪效果一致劣于其他 5 种方法, 而 Wiener-S 和 Wiener-B 方法的去噪效果都胜出了其他 5 种方法。在大多数情况下, Sure-S 和 Bayes-S 方法去噪效果优于 Wiener-C 方法, Sure-S 方法比 Bayes-S 方法的均方误差略低, 类似的结果也出现在 Wiener-S 和 Wiener-B 方法中。通过统计表 1 中均方误差的平均值发现, Wiener-S 方法较 Sure-S 方法, 均方误差降低了 9.2%; Wiener-B 方法较 Bayes-S 方法, 均方误差降低了 9.3%; Wiener-S 方法较 Wiener-C 方法, 均方误差降低了 15.3%; Wiener-B 方法较 Wiener-C 方法, 均方误差降低了 14.4%。

虽然, 大多数情况下, WienerChop 组合

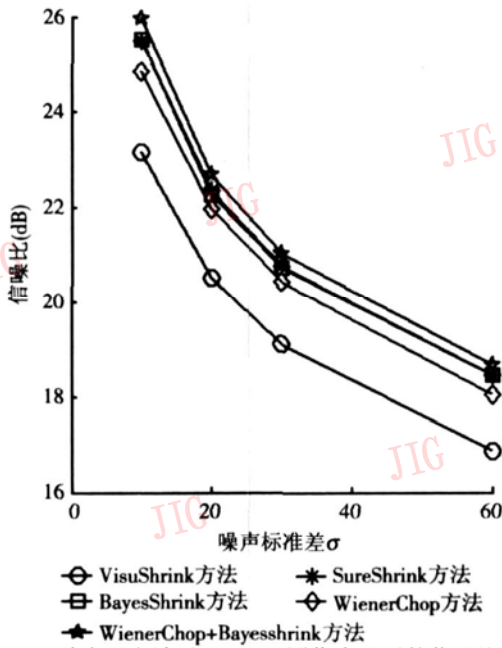
SureShrink 方法的均方误差略低于 WienerChop 组合 BayesShrink 方法, 但是, 笔者仍然推荐后者, 其原因是后者计算更简单, 计算效率更高。后者均方误差虽略高, 但可以认为属于计算中正常的波动范围。

下面以 Goldhill 图像为例, 根据表 1 中计算的均方误差, 分析除 Wiener 和 Wiener-S 方法以外的其他 5 种方法的去噪后图像的信噪比。信噪比的计算公式是

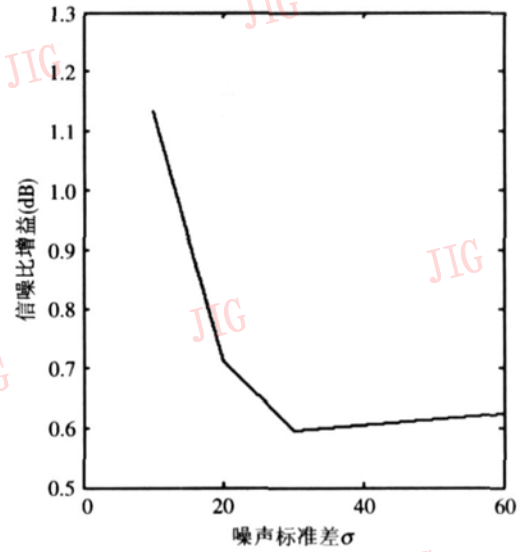
$$SNR = 10 \lg \frac{\sum_{i=1}^l \sum_{j=1}^l X(i, j)^2}{MSE}$$

5 种方法去噪后图像的信噪比曲线如图 3 所示。

从左图可以看出, Sure-S 和 Bayes-S 方法的信噪比曲线几乎重合, 这说明这两种方法去噪后图像的信噪比几乎是一样的。Wiener-C 方法去噪后图像的信噪比远高于 Visu-S 方法去噪后图像的信噪比, 但低于 Sure-S 和 Bayes-S 方法去噪后图像的信噪比, Wiener-B 方法去噪后图像的信噪比一致高于其他 4 种方法去噪后图像的信噪比。从右图可以看出, Wiener-B 方法较 Wiener-C 方法, 去噪后图像的信噪比大约提高了 0.6~1.15dB。这个分析同前面的结果是一致的。



(a) 5种去噪方法对 Goldhill图像去噪后的信噪比曲线



(b) Wiener-B较Wiener-C方法对 Goldhill图像去噪后的信噪比增益曲线

图 3 Goldhill图像去噪后信噪比分析

Fig 3 SNR analysis of the denoised Goldhill

通常均方误差越低, 信噪比越高, 图像降噪效果越好, 去噪后的图像质量越高。众所周知, 均方误差和信噪比并不总是同人的视觉一致, 还需要配合视觉观察来判断降噪效果。下面以污染噪声标准差为 20 的 Barbara 图像为例, 观察 Wiener-C、Bayes-S 和 Wiener-B 3 种方法的降噪效果。由于完整图像的降噪细节很难分辨, 因此本文截取一个局部放大观察 (如图 4 所示)。Wiener-C 方法有最大的均方误差和最小的信噪比, 它虽较好地去除了噪声, 但其缺点是图像比较模糊, 这一点只要看

去噪后 Barbara 图像的围巾和嘴唇即可, Bayes-S 方法去噪后图像有中等的均方误差和信噪比值, 虽然它较好地保留了图像的细节, 但是残留了较多的噪声。Wiener-B 方法去噪后图像有最低的均方误差和最高的信噪比, 也确实具有最高的图像质量。它综合了 Wiener-C 和 Bayes-S 方法的优点, 一方面, 它比 Wiener-C 方法去噪后的图像清晰, 且保留了更多的图像细节, 另一方面, 它比 Bayes-S 方法去噪后的图像减少了恼人的噪声, 产生了最为愉悦的效果。



(a) Wiener-C 方法去噪图像 (MSE = 128.01, SNR = 21.96dB)



(b) Bayes-S 方法去噪图像 (MSE = 120.17, SNR = 22.25dB)



(c) Wiener-B 方法去噪图像 (MSE = 103.55, SNR = 22.67dB)

图 4 $\sigma = 20$ 的 Barbara 有噪图像降噪

Fig. 4 Denoising results for Barbara while $\sigma = 20$

5 结 论

本文分析了基于小波的去噪图像中高斯随机噪声的方法。实验表明,基于小波收缩或取阈值的非线性图像降噪方法,去噪性能远优于基于傅里叶变换的传统的线性的维纳降噪方法,虽然在线性降噪方法中后者是最优的。VisuShrink方法是采用全局的 universal 阈值,由于在小波分解的各个子带上采用相同的阈值,而且 universal 阈值又偏大,导致图像过度光滑,从而损失了较多的细节。SureShrink 和 BayesShrink 方法减少了 VisuShrink 方法的缺陷,由于它们均采用数据驱动的、子带依赖或自适应的阈值,因此较好地保留了图像的细节。在正交小波变换阈值的降噪方法中, SureShrink 和 BayesShrink 方法是最优的。尽管这两种降噪方法产生的均方误差和信噪比几乎相同,但是, BayesShrink 方法计算更简单、高效,虽然 WienerChop 方法也改进了 VisuShrink 方法的缺陷,但是,改进的思路与 SureShrink 和 BayesShrink 方法不同。它是先将 VisuShrink 方法作为真正信号的先导估计,然后创建小波域的维纳滤波器,最后通过维纳滤波的有效伸展,挽回了部分被 VisuShrink 方法去掉的信号系数,从而改进了降噪效果。

本文研究了 SureShrink 或 BayesShrink 与 WienerChop 组合的降噪方法,并推荐使用 BayesShrink 与 WienerChop 组合的降噪方法(即 Wiener-B)。它优于 SureShrink、BayesShrink 和 WienerChop 方法中任何一种方法,其去噪后的图像

具有最低的均方误差和最高的信噪比。Wiener-B 方法较 WienerChop 方法,均方误差平均降低了 14.4%。在具体的 Goldhill 降噪中, Wiener-B 方法较 WienerChop 方法,去噪后图像的信噪比大约提高了 0.6~1.15dB。更重要的是, Wiener-B 方法综合了 BayesShrink 和 WienerChop 两种方法的优点,较好地平衡了 WienerChop 方法使图像过于光滑和 BayesShrink 方法残留较多噪声的问题,而且获得了视觉上最为愉悦的降噪图像。

参考文献 (References)

- 1 Mallat S. A Wavelet Tour of Signal Processing (2nd edition) [M]. New York: Academic Press, 1999.
- 2 Donoho D L, Johnstone I M. Ideal spatial adaptation via wavelet shrinkage [J]. *Biometrika* 1994, **81**(3): 425~455.
- 3 Donoho D L, Johnstone I M. Adapting to unknown smoothness via wavelet shrinkage [J]. *Journal of the American Statistical Association*, 1995, **90**(432): 1200~1224.
- 4 Chang S G, Yu B, Vetterli M. Adaptive wavelet thresholding for image denoising and compression [J]. *IEEE Transactions on Image Processing* 2000, **9**(9): 1532~1546.
- 5 Fodor I K, Kamath C. Denoising through wavelet thresholding: an empirical study [J]. *SPIE Journal on Electronic Imaging* 2003, **12**(1): 151~160.
- 6 Ghael S P, Sayeed A M, Baraniuk R G. Improved wavelet denoising via empirical wiener filtering [A]. In: *Proceedings of SPIE Conference on Mathematical Imaging* [C], San Diego, CA, USA, 1997: 389~399.
- 7 Choi H, Baraniuk R. Analysis of wavelet-domain wiener filters [A]. In: *Proceedings of the IEEE-SP International Symposium on Time-frequency and Time-scale Analysis* [C], Pittsburgh, PA, USA, 1998: 613~616.

基于边缘的实时图像缩放算法研究

刘政林 肖建平 邹雪城 郭旭

(华中科技大学电子科学与技术系, 武汉 430074)

摘要 传统的基于图像边缘的插值算法由于自身的复杂性而很少用于实时图像处理, 而不基于边缘的插值算法由于低通滤波的效应通常会使插值后的图像出现边缘模糊, 影响视觉效果。针对上述问题, 提出了一种基于边缘的自适应图像插值算法。该算法以相邻像素的二阶差分为基础, 自动选择与目标像素点相对应的源像素组进行插值运算。在 Matlab 环境中实现了各种图像插值算法, 然后对插值结果进行了主客观评价, 并对各种算法获得的图像进行了边缘检测及复杂度分析。同时, 设计了本文算法的硬件实现结构, 并用 Verilog 语言进行描述, 综合出目标代码, 最后通过 FPGA 验证。研究表明, 运用该算法插值获得了边缘清晰的目标图像, 且该算法复杂度低, 便于硬件实现。因此, 该算法能有效地实现图像插值, 适应于实时条件下的图像缩放。

关键词 实时图像处理 自适应算法 图像插值 图像边缘

中图分类号: TP391.41 文献标识码: A 文章编号: 1006-8961(2008)02-0225-05

Edge-based Algorithm of Real-time Image Resizing

LU Zheng-lin, XIAO Jian-ping, ZOU Xue-cheng, GUO Xu

(Department of Electronic Science and Technology, Huazhong University of Science and Technology, Wuhan 430074)

Abstract Due to their complexity, traditional edge-based algorithms of image interpolation are hardly used for real-time image processing. And non-edge based interpolation algorithms usually lead to the target images blurring because of their low-pass effect, so the visual effect of the interpolated images is unpleasant. In this paper, a novel adaptive edge-based algorithm of image interpolation is proposed to eliminate the defect. According to the second-order difference of the consecutive pixels, the source pixel group is automatically selected to compute the corresponding target pixel gray value. Several algorithms of image scaling are implemented in Matlab. Then the visual appearance and the edges of target images are compared, and the complexity of the algorithms is analyzed. The implementation circuit architecture for our proposed algorithm is devised and the algorithm is realized by Verilog HDL. Finally, the synthesized code is verified by downloading to FPGA on the evaluated board. The results show that our proposed algorithm achieves vivid edges of target images, and has lower computation complexity and easy for hardware implementation. So the algorithm can effectively interpolate images and adapts to real-time image resizing.

Keywords real-time image processing, adaptive algorithm, image interpolation, image edge

1 引言

图像缩放广泛应用于医学、消费电子、遥感等领域的图像处理与显示中。图像缩放的算法很多, 根据算法是否利用图像的边缘信息可将其分为两类。

第 1 类是不基于边缘的图像缩放算法, 在这类算法中, 新得到的像素值与图像的边缘信息无关, 只与其位置和其邻域的像素值有关。这类算法常用的有最近邻域法、双线性插值、双三次插值、样条插值等算法。这些经典的算法存在一些不足^[1], 通常低阶算法会出现“马赛克”现象, 而高阶算法不仅计算复

收稿日期: 2006-03-10 改回日期: 2006-10-09

第一作者简介: 刘政林 (1968~), 男, 副教授, 硕士研究生导师。2002 年获华中科技大学微电子学与固体电子学博士学位。主要研究方向为超大规模集成电路设计、数字图像处理。E-mail: liuzhenglin@21cn.com

杂,而且由于“振铃”效应会使图像边界模糊。第 2 类是基于边缘的图像插值算法,这类算法新得到的像素点的灰度值与源图像的边缘相关,可得到比第 1 类算法更满意的视觉效果,但它们算法复杂,很少应用于实时条件下的图像缩放^[2,3]。本文提出的图像插值方法是一种用于实时条件下的基于边缘的自适应图像插值算法。该算法根据相邻源像素的二阶差分值自动选取一组源像素,再根据二阶牛顿插值公式获得目标像素的值。该算法不仅获得了比不基于边缘插值算法更满意的视觉效果,而且复杂度比不基于边缘的双三次插值算法更低,远低于传统的基于边缘的插值算法。

2 基于图像边缘插值的原理

传统的基于边缘的插值算法^[3]通过检测源图像的边缘,由图像的边缘信息对图像的不同区域采用不同的插值方法。相比不基于边缘的插值算法,不但要事先对图像进行边缘检测,而且要存储图像的边缘信息,其复杂性是显而易见的。因此,这种传统的插值方法一般不用于实时图像处理。

在实时条件下对图像进行缩放处理,通常对 2 维图像依次在水平方向和垂直方向分别进行插值处理。将源图像在某一方向(如水平方向)的像素序列看作是一组等距节点,目标图像在同一方向上的任意像素点的值可由某些节点通过插值公式求得。除边界像素点外,目标图像中任意像素点与源图像像素点的位置关系如图 1 所示,任何目标像素点 N 均能找到与之对应的源像素点 f_0, f_1, f_2, f_3 , 其中 t 为目标像素点 N 与源像素点 f_0 间的距离, $1 \leq t \leq 2$ 。

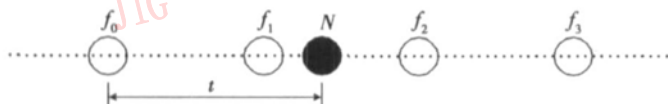


图 1 目标像素与源像素的位置关系

Fig 1 The relationship between target and source pixels

若采用不基于边缘的插值算法,如最近邻插值和双线性插值,目标像素点 N 只与源像素点 f_1, f_2 有关;若采用双三次算法,目标像素点 N 的灰度值与源像素点 f_0, f_1, f_2, f_3 有关,上述关系无论像素点 N 是否处于图像的边缘均成立。这些算法的共同点是不判断源图像中各像素点的相关性,结果不同程

度地导致了包含丰富信息的图像边缘模糊。而各像素点间的相关性隐含在图像本身,只要选择合适的参数便能得到这一重要特性。

考察源像素点 f_0, f_1, f_2, f_3 之间的相关性,将其分成两组源像素点 $I(f_0, f_1, f_2)$ 和 $II(f_1, f_2, f_3)$, 分别求出它们的二阶差分如下:

$$\Delta^2 f_0 = f_2 - 2f_1 + f_0 \quad (1)$$

$$\Delta^2 f_1 = f_3 - 2f_2 + f_1 \quad (2)$$

在此,将相邻 3 个点的二阶差分的绝对值作为衡量该 3 个点的相关性大小的标准,绝对值越小,相关性越大,反之亦然。图 1 中,目标像素点 N 的值由一组相关性较大的源像素通过插值比采用相关性较小的一组更为合理。这是因为相关性大,表明该组各像素处在同一图像区域的可能性越大。

通过比较两组源像素点的相关性,获得计算目标像素 N 的一组源像素,采用二阶插值公式即可求得目标像素 N 的灰度值。由于二阶差分已经求出,在各种经典插值公式中牛顿插值公式必须用到各阶差分,故采用二阶牛顿插值公式完成目标像素点的计算。这样可大大简化硬件电路的设计,而且不影响插值图像的视觉效果。

由此基于图像边缘的插值公式如下:

$$N_2 = \begin{cases} f_0 + \Delta f_0 t + \frac{\Delta^2 f_0}{2} t(t-1) & |\Delta^2 f_0| \leq |\Delta^2 f_1| \\ f_1 + \Delta f_1(t-1) + \frac{\Delta^2 f_1}{2} (t-1)(t-2) & |\Delta^2 f_0| > |\Delta^2 f_1| \end{cases} \quad (3)$$

$$\Delta f_i = f_{i+1} - f_i \quad i = 0, 1 \quad (4)$$

通过比较 $|\Delta^2 f_0|$ 和 $|\Delta^2 f_1|$ 的大小,选择合适的源像素点计算出目标像素的灰度值 N_2 。

3 实验结果

将本文算法及几种经典插值算法在 Matlab 环境中编程实现,对比各种插值算法处理图像的主观视觉效果、峰值信噪比及边缘信息,对各种算法进行主客观评价。

3.1 图像插值结果对比

对于灰度均一的区域,大多数算法都可获得很接近的视觉效果,而在图像边缘区域,由于灰度变化率大,不同的算法效果差别很大。为此,将着重分析图像边缘区域的缩放效果。

图 2(a)是分辨率为 512×512 的 Lena 原始图

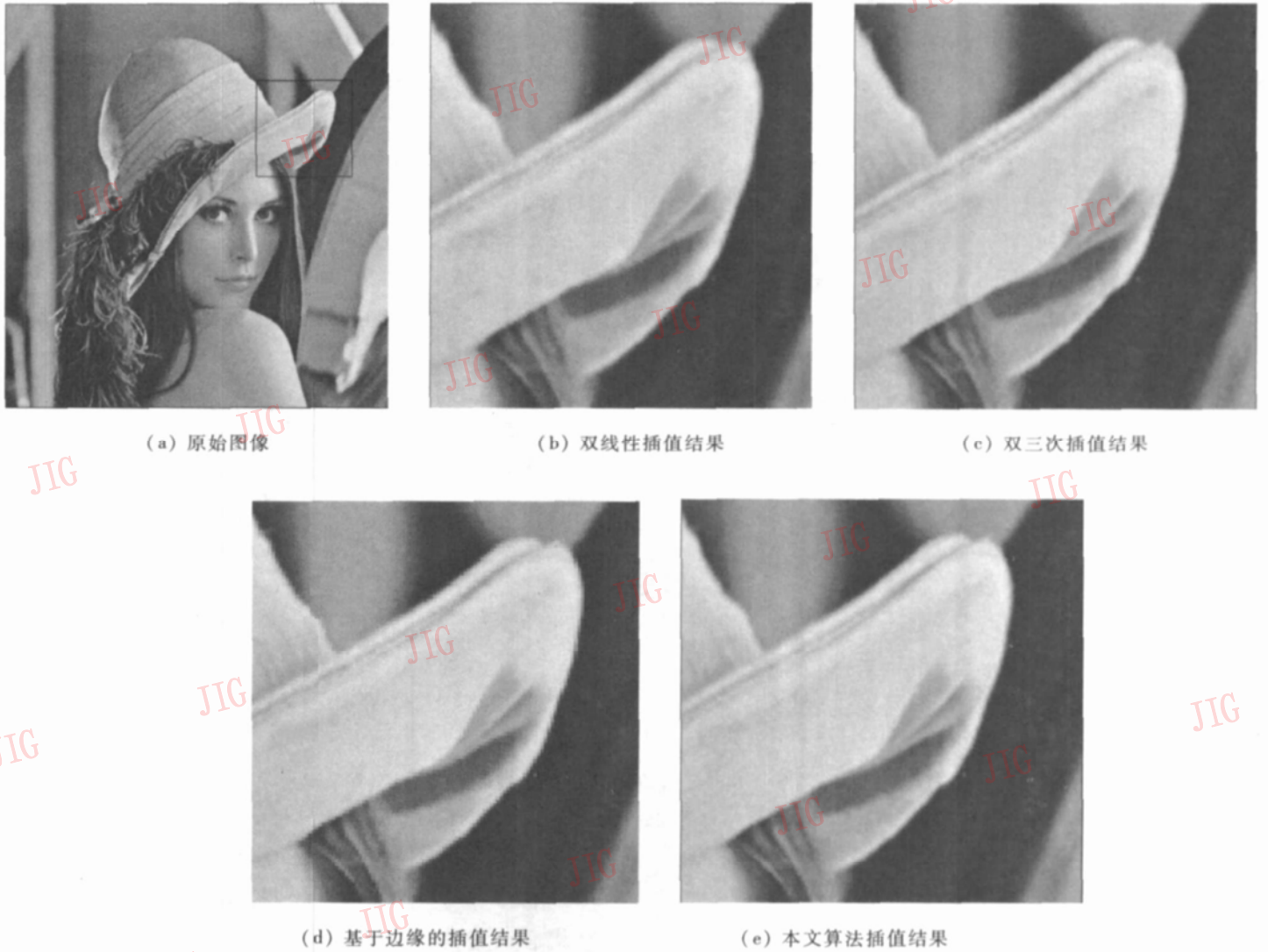


图 2 各种不同算法缩放图像效果

Fig. 2 Visual effect of resizing image by different algorithm

像, 方框标示的图像边缘区域大小为 120×120 。图 2(b) ~ (d) 分别为图 2(a) 方框区域采用不同的缩放算法得到的放大图像, 其大小均为 512×512 。由图可见, 图 2(b) 的帽沿区域出现了锯齿现象, 整体视觉效果模糊; 图 2(c) 较图 2(b) 而言, 图像整体效果平滑, 视觉效果较好, 但帽沿处出现了“振铃”现象, 轮廓不清晰, 而图 2(e) 的图像边缘轮廓清晰, 整体效果层次分明; 图 2(d) 在主观视觉效果上与图 2(e) 差异并不明显。

客观评价图像缩放算法优劣的一个重要指标是峰值信噪比 (PSNR), 其计算公式^[4]如下:

$$PSNR(F_0, F) = 10 \lg \frac{255^2}{\frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N [F(i, j) - F_0(i, j)]^2} \quad (5)$$

式中, F_0 为原始图像, F 为原始图像由某一算法放

大后, 再经相应算法缩小到与原始图像相同大小的图像。对图 2(b) ~ (d) 分别采用相对应的算法对各图像进行缩小处理, 再根据式 (5) 计算, 得到双线性插值、双三次插值及传统的基于边缘的插值处理此边缘区域的信噪比分别为 46.183 2dB、48.077 2dB、51.835 3dB, 而本文算法获得的信噪比为 49.338 6dB, 明显高于不基于边缘的双线性和双三次算法。尽管传统的基于边缘的插值获得了比本文算法更高的信噪比, 但其复杂度在实时条件下是无法接受的。

3.2 缩放图像的边缘分析

双线性插值算法有明显的低通效应, 导致高频信息丰富的图像边缘产生模糊, 而双三次插值能较好地保持图像的边缘特性, 但会出现“振铃”现象^[5]。本文提出的基于边缘的自适应图像插值算法在计算目标像素的灰度时, 根据源像素组二阶差

分最小的原则,避免了在图像边缘两侧不同灰度区域同时选取源像素参与运算,从根本上抑制了“振铃”效应的出现。

对目标图像进行边缘检测可以明显看到这一现象, Marr和 Hildreth提出的拉普拉斯边缘检测算子 $\nabla^2 G$ 被誉为最佳边缘检测器之一^[6]。对图 2(c)和图 2(d)采用 $\nabla^2 G$ 进行边缘检测,尺度空间常数均

取 2.0 所获得的边缘如图 3(a)和图 3(b)所示。图中箭头所指为 Lena 图像的帽沿。图 3(a)为双三次插值进行边缘检测的效果图,在帽沿的外侧形成了一条明显的虚边缘,而图 3(b)有效地避免了这一现象的出现(帽沿外侧有几个孤立的点,这是由于源图像的噪声所致)。

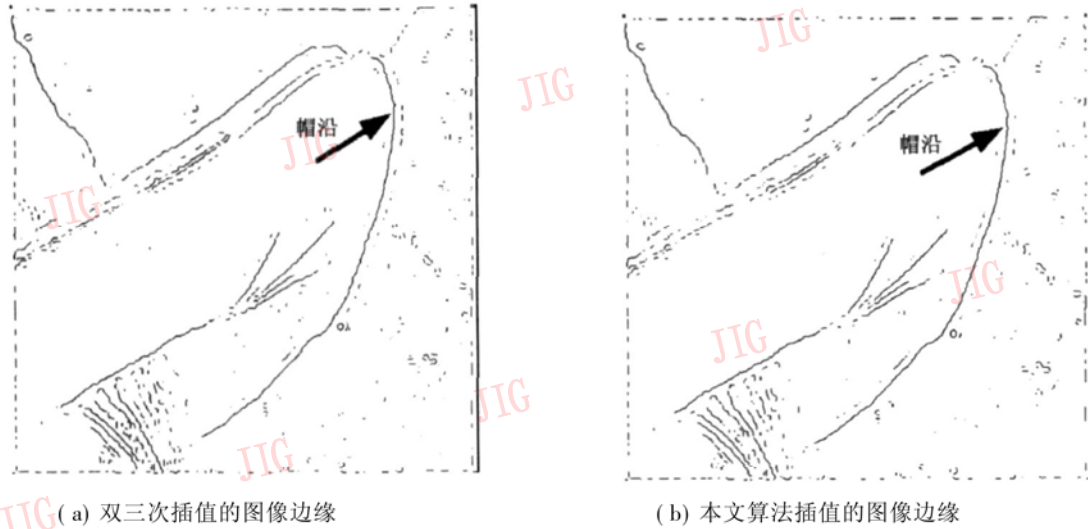


图 3 不同插值算法的图像边缘比较

Fig 3 Comparison of image edge by different algorithm

4 算法硬件实现与验证

在评价复杂算法硬件实现时,对于常数参与的加法和乘法运算占用的硬件资源可以忽略不计(例如某一未知量与一常数相乘,完全可以通过移位实现,而无需设计乘法器实现运算)。对于本文提出的自适应插值算法(如式(3)),尽管在 $|\Delta^2 f_0|$ 和 $|\Delta^2 f_1|$ 相对值不同的情况下表达式并不一样,但实际上在这两种情况下只是相应的输入条件不同,两种情况可统一按下式处理。

$$N_2 = f + \Delta f \cdot t + \frac{\Delta^2 f}{2!} \cdot t \cdot (t-1) \quad (6)$$

在 t 已知的情况下,式(6)中,计算 $\Delta f \cdot t$ 需一次乘法运算, $\frac{\Delta^2 f}{2!} \cdot t \cdot (t-1)$ 需两次乘法运算,而计算 Δf 和 $\Delta^2 f$ 分别需一次加法和两次加法运算,然后式(6)各项相加需 2 次加法运算。因此,本文方法只需 3 次乘法和 5 次加法即可完成一次插值运算。同样,双三次插值计算一目标像素点需 5 次乘法和 10 次加法运算,对比这两种算法,本文算法约减少

40% 的计算量。而最近邻域插值根据已知参数 t 直接获取像素值,双线性插值需 2 次乘法和 1 次加法运算,三次样条插值^[4]需 6 次乘法和 9 次加法运算。对于传统的基于边缘的插值算法,由于要对原图像进行边缘检测,并保存边缘信息,然后根据边缘信息在图像的不同区域选取不同的插值算法,因此,其复杂度要远远超出上述各种算法。

在实时条件下用 FPGA (field programmable gate array) 实现本文算法的数据流图如图 4 所示。系统的输入 f_0, f_1, f_2, f_3 由串联存储器 FIFO (first-in first-out) 或随机存储器提供,通过并行的加法运算即可得到一阶差分 $\Delta f_0, \Delta f_1, \Delta f_2$ 同样的方法得到二阶差分 $\Delta^2 f_0, \Delta^2 f_1$, 比较 $\Delta^2 f_0, \Delta^2 f_1$ 的大小产生的逻辑量作为 3 个多路选择器 (MUX) 提供判断条件,多路选择器自动选择一组值参与式(3)所示的运算 (\sum), 最终得到插值结果 N_2 。

用 Verilog HDL 语言对算法进行描述,在相应工具中进行综合,将得到的目标代码下载至如图 5 所示的评估板的 FPGA 中进行验证,在 LCD (liquid crystal display) 屏上得到了仿真结果完全一致的目标图像。

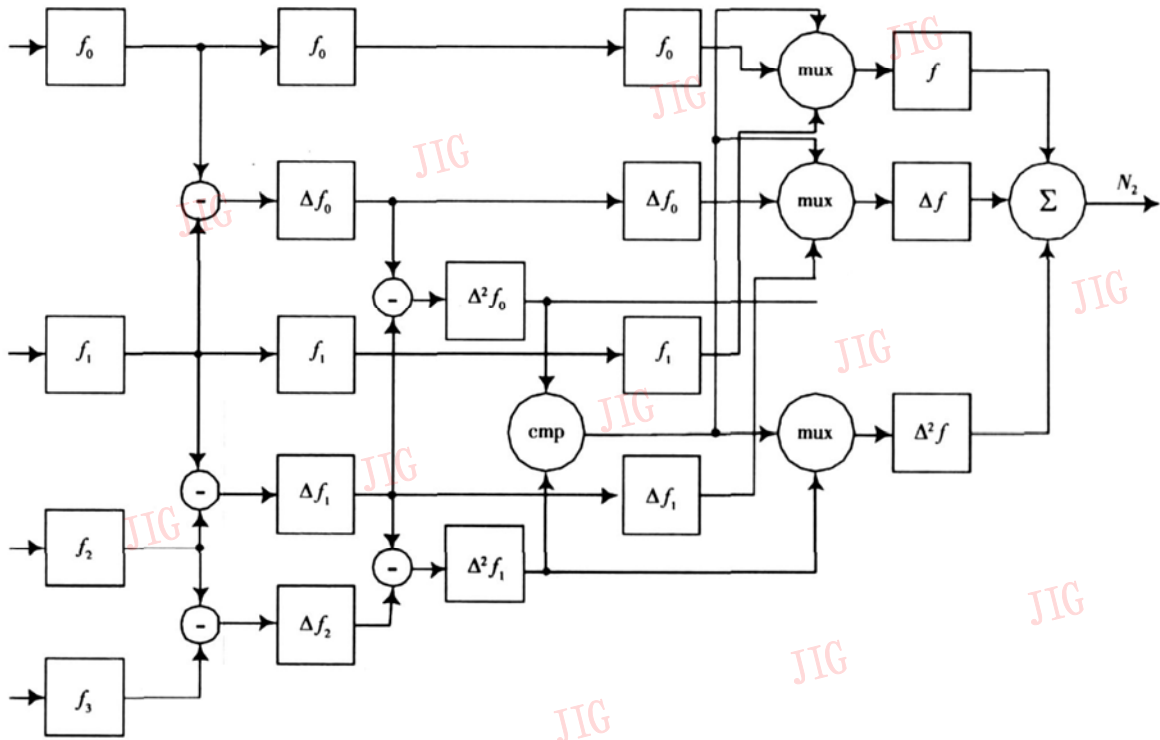


图 4 算法实现信号流程图

Fig 4 Signal flow of the algorithm's implementation



图 5 算法的 FPGA 验证

Fig 5 Algorithm's implementation by FPGA

5 结 论

提出了一种基于边缘的图像插值算法, 并采用三级流水线将该算法成功的应用于实时图像处理中。实验结果表明, 采用该算法进行图像缩放可得到比传统的不基于边缘的双三次插值算法更满意的视觉效果和更高的信噪比。而且该算法具有比双三

次插值算法更低的复杂度, 便于硬件实现。综上所述, 该算法能有效实现图像缩放, 尤其适合于实时条件下的图像处理。

参考文献 (References)

- 1 Liu Zhijun, Cai Chaq, Peng Xiaoming *et al*. A novel regularized image interpolation algorithm based on genetic algorithm [J]. Journal of Image and Graphics, 2004, 9(8): 935~ 940 [刘志军, 蔡超, 彭晓明等. 一种新颖的基于遗传算法的正则化图像插值方法 [J]. 中国图象图形学报, 2004, 9(8): 935~ 940]
- 2 Jensen K, Anastassiou D. Subpixel edge localization and the interpolation of still images [J]. IEEE Transactions on Image Processing 1995, 4(3): 285~ 295
- 3 Xin L, Orchard, M T. New edge-directed interpolation [J]. IEEE Transactions on Image Processing 2001, 10(10): 1521~ 1527.
- 4 Han J K, Baek S U. Parametric cubic convolution scaler for enlargement and reduction of image [J]. IEEE Transactions on Consumer Electronics 2000, 46(2): 247~ 256
- 5 Saif Zahir, Rabab Ward. A new edge preserving binary images resizing technique [J]. Journal of Circuits, Systems and Computers 2002, 11(3): 247~ 257.
- 6 Rafael C Gonzalez, Richard E Woods. Digital Image Processing Second Edition [M]. New Jersey, USA: Prentice Hall 2002

CAVLC 解码的一种有效方法

曹 宁 梅 侠

(河海大学电气工程学院, 南京 210098)

摘 要 已有的 CAVLC 解码方法包括二叉树解码方法、全码表解码方法和 Hashemian 解码方法等, 但是这些解码方法都只关注解码性能的一个方面: 解码速度或存储空间, 因而无法有效地提高整体性能。针对这一问题提出了一种快速的解码方法。该方法通过自动码表分配技术和码表地址转移技术来提高限定存储空间条件下的解码速度。实验结果表明, 使用相同的存储空间, 该方法的速度是传统解码方法的 1.5 倍, 更加适用于 H. 264 标准。

关键词 CAVLC 二叉树 全码表 Hashemian 方法 自动码表分配 码表地址转移

中图分类号: TN919.81 **文献标识码:** A **文章编号:** 1006-8961(2008)02-0230-04

An Effective Way of CAVLC Decoding Methods

CAO Ning MEIXia

(College of Electrical Engineering, Hohai University, Nanjing 210098)

Abstract CAVLC decoding methods known by people include bintree, full code table, and Hashemian decoding methods etc. all of which lay emphasis on only one aspect of decoding: decoding speed or memory space, and thus the general capability can not be improved effectively. In view of that, the paper puts forward a new and fast decoding method which by means of the automatic distribution of the code table and the transfer of the code table address, improves the decoding speed in the limited memory space. As the experiment shows, in the same-size memory space, the speed of this new method is 1.5 times as fast as that of the tradition ways so that the new one is more suitable to H. 264.

Keywords CAVLC, bintree, full code table, Hashemian method, the automatic distribution of the code table, the transfer of the code table address

1 引 言

ITU-T VCEG (video coding experts group) 和 ISO/IEC MPEG (moving pictures experts group) 共同组建的联合视频专家组 JVT 于 2003 年正式发布了新一代视频编码标准 H. 264/AVC。H. 264/AVC 是在 H. 26L 的基础上推出的, 它的压缩效率是 MPEG-4 或 H. 263+ 的 2 倍左右。H. 264 将主要用于视频会议、网络视频流和消费电子应用等。

在 H. 264 的 3 个应用档次中都包含了 CAVLC 编码(基于内容的自适应变长编码), 特别是基本档次和扩展档次只使用 CAVLC 一种熵编码方式, 因此

CAVLC 的编码码表和解码是现在研究的一个方向。对于 CAVLC 的解码, 已经有了二叉树、全码表和 Hashemian 等解码方法, 但是这些解码方法都只考虑了解码性能的一个方面, 即解码速度或者存储空间, 本文探讨一种基于 Hashemian 方法的改进方法, 它使用自动化的方法来获得更好的解码码表分配。

2 CAVLC

H. 264 的早期草案中使用 UVLC (统一的可变长编码) 对残差数据进行熵编码, 但是由于 UVLC 没有利用相邻的 4×4 变换块之间的关系, 因而编码效率不高, 于是在最终草案中就用 CAVLC 来代替

收稿日期: 2006-06-07; 改回日期: 2006-10-17

第一作者简介: 曹宁 (1962~), 男, 教授, 硕士生导师。1990 年于东南大学无线电系获信号处理专业硕士学位, 长期从事多媒体图像压缩与传输、数字通信以及数字信号处理方面的研究和教学工作。E-mail: caoning@vip.163.com

UVLC。CAVLC 具有以下优点^[1]:

(1) 基于上下文的自适应编码方法可全面提高编码的质量。

(2) 应用锯齿形扫描方法, 将非零系数值 (Level) 和零行程 (Run) 分开编码, 可提高编码的自适应性。

(3) 将非零系数值 (Level) 和零行程 (Run) 分开可降低存储数据的复杂度。

CAVLC 对残差数据的编码过程如下^[2]:

(1) 非零系数的数目 (TotalCoeffs) 以及拖尾系数的数目 (TrailingOnes) 进行编码。

(2) 对每个拖尾系数 (Tls) 的符号编码。

(3) 对除拖尾系数外的非零系数的幅值 (Levels) 编码。

(4) 对最后一个非零系数前的总的零的数目 (TotalZeros) 编码。

(5) 对每个非零系数前零的数目 (RunBefore) 编码。

编码过程实际上就是一个查表寻找码字的过程, 而解码过程则是一个根据码字反相查找的过程。由于码字不是统一长度的, 而是出现频率高的使用短码字, 出现频率低的则使用长码字, 这样在解码的时候就需要先确定码字的长度。

3 已经被提议的解码方法

VLC(可变长编码) 广泛应用于各种视频压缩标准, 例如 MPEG-2 MPEG-4 H. 26L 等, 因此已经有很多 VLC 的解码方法, 下面介绍 3 种已经被提议的方法。

3.1 二叉树解码方法

二叉树解码方法的原理就是将码表存储结构的码字转化为树形存储结构的码字, 即形成 1 棵二

叉树。解码过程实际上就是一个寻找叶节点的过程。图 1 为 1 棵简单的二叉解码树。

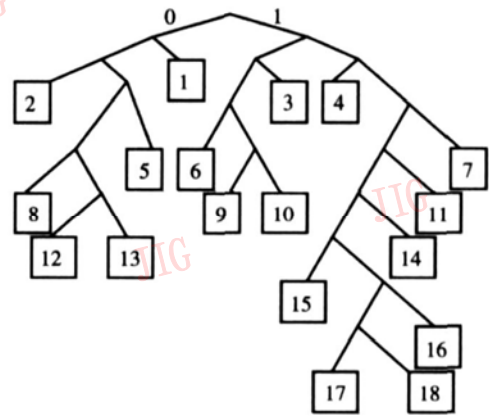


图 1 二叉解码树

Fig 1 Decoding bintree

叉树方法充分利用了码字之间的相关性, 对存储空间消耗非常少, 但是这种解码方法是以 1 个比特作为比较对象的, 所以它的解码速度很慢。

3.2 全码表解码方法

由于编码码表中的码字是以非连续的方式存在的, 因此可以通过对每个码字的扩展来构建一张大小为 2^L (L 是编码码表中码字的最大长度) 的连续的解码码表^[3]。解码的时候只需要一次性的读取 L 位比特, 按其大小到解码码表中查找结果。

由于只需要比较一次就可以得到结果, 所以这种解码方法的速度非常快, 但是它要求有极大的存储空间, 所以只能用于 L 很小的场合。

3.3 Hashemian 解码方法

Hashemian 解码方法^[4]是对二叉树和全码表解码方法的改进。它使用固定的长度 H 来分割解码二叉树, 对长度小于 H 的码字进行扩展。以图 1 所示的二叉树码表为例, 当 $H = 3$ 时, 获得图 2 所示的 Hashemian 解码码表。

码字	码长 D	结果 R	码字	码长 D	结果 R	码字	码长 D	结果 R	码字	码长 D	结果 R	码字	码长 D	结果 R	转移码表码字	读取的长度	码表地址
000	3	2	000	2	8	000	0	3	000	1	15	00	1	6	0	3	L1
001	0	0	001	2	8	001	3	14	001	1	15	01	1	6	1	2	L2
010	2	1	010	3	12	010	2	11	010	1	15	10	2	9	2	3	L3
011	2	1	011	3	13	011	2	11	011	1	15	11	2	10	3	3	L4
100	0	1	100	1	5	100	1	7	100	3	17						
101	3	3	101	1	5	101	1	7	101	3	18						
110	3	4	110	1	5	110	1	7	110	2	16						
111	0	2	111	1	5	111	1	7	111	2	16						

图 2 Hashemian 解码码表

Fig 2 Hashemian decoding table

解码过程就是读取固定长度的比特, 然后查表的过程。如果 $D > 0$ 则找到结果 R ; 如果 $D = 0$ 则使用 R 作为转移码表码字到中转码表中查找下一个码表的入口地址。

Hashemian 方法通过一次比较 H 个比特来提高解码速度, 同时使用码字分割技术来减少对存储空间的需求。但是 Hashemian 方法也有其缺点, 一是对于一个给定大小的存储空间, 它无法尽可能地使用给定的存储空间以提高解码速度; 二是中转码表的使用使得可利用的存储空间进一步减少。

4 基于自动化码表分配的解码方法

平均解码码表的查找次数与解码速度有很大的关系, 解码速度的提高也就意味着平均的解码码表查找次数的减少。下面介绍平均解码码表查找次数的计算方法^[5]。

$$\begin{aligned} T &= k_1 P_1 + k_2 P_2 + k_3 P_3 + \dots + k_1 P_{n+1} + k_2 P_{n+2} + \\ &\quad k_3 P_{n+3} + \dots + k_1 P_{m+1} + k_2 P_{m+2} + k_3 P_{m+3} + \dots \\ &= k_1 (P_1 + P_{n+1} + P_{m+1} + \dots) + k_2 (P_2 + P_{n+2} + \\ &\quad P_{m+2} + \dots) + k_3 (P_3 + P_{n+3} + P_{m+3} + \dots) + \dots \\ &= k_1 S_1 + k_2 S_2 + k_3 S_3 + \dots \end{aligned} \quad (1)$$

$$N = \sum_{i=0}^{n-1} M_i \quad (2)$$

式中, T 是平均解码码表的查找次数。 k_i 表示解码码表 t_i 中的码字需要查找的码表的个数。 P_i 是第 i 个码字的概率。 S_i 表示在 t_i 中最终可以查到结果的码字的概率和。 N 表示所需要的内存空间的总和。 M_i 表示每个解码码表所使用的存储空间。

对于一个给定大小的存储空间 A , 为了获得最快的解码速度, 则应在满足 $N \leq A$ 的条件下使得 T 最小。

T 的计算过程中需要用到每个码字的概率 P_i 。但是 H. 264 标准只是提供了 CAVLC 的编码码表, 并没有提供每个码字的概率, 这样就不能用式 (1) 来计算 T 。CAVLC 编码码表中的码字的分配是有一定的规则的, 短码字分配给使用频率高的符号, 长码字分配给使用频率低的符号。即

$$P(L_{\text{short}}) > P(L_{\text{long}}) \quad (3)$$

并且有 $\sum_{i=1}^M P_i = 1$ 其中, M 是码表中码字的最大数目。

根据上面的这个分配原则, 在限定存储空间为

A 的条件下, 使用下面的方法来代替式 (1) 获得更好的码表分配。

(1) 令 $k_1 = L$ (编码码表中码字的最大长度), 获得 $M = 2^{k_1}$, 比较 M 与 A , 如果 $M > A$ 的话就逐渐减小 k_1 使得满足存储空间的要求, 如果 $M < A$ 则分配结束。

(2) 找出第 1 层分割前缀码相同, 并且码长最短的一批码字, 获得这批码字的长度 L_1 , 令 $k_2 = L_1 - k_1$ ($k_2 \leq k_1$), 同样通过减小 k_2 来满足存储空间的要求。

(3) 找出第 2 层分割前缀码相同并且码长最短的一批码字, 获得这批码字的长度 L_2 , 令 $k_3 = L_2 - k_2$ ($k_3 \leq k_2$), 同样通过减小 k_3 来满足存储空间的要求。

(4) 使用第 3 步相同的方法获得 k_4, k_5 等, 直到分配完第 2 步所获得的码字。

(5) 返回第 2 步获得下一批前缀相同的码字, 然后使用上面的方法进行分配, 直到分配完所有的码字。

使用这种方法有以下的优点:

(1) 使用软件的方法来实现, 可以减少大量的手工计算。

(2) 不需要知道每个码字的概率也可以获得最小的平均解码码表查找次数。

(3) 不同层次使用不同长度的分割, 同一层次也使用不同长度的分割, 使得给定的存储空间得到最大的利用。

为了进一步减少对存储空间的消耗, 使用统一解码码表来代替 Hashemian 解码码表。统一解码码表的关键就是使用了码表地址偏移技术。

统一解码码表中使用了变量 Y 和 Z 。当 $Y > 0$ 时, Y 和 Z 的定义与 Hashemian 码表中 D 和 R 的定义相同, 当 $Y < 0$ 时, Y 表示下一次需要读取的比特个数的负数, Z 表示从解码码表的起始位置往后移动的个数。以图 1 为例, 令 $H = 3$ 获得如表 1 的统一解码码表。

其中坐标值等于偏移量 Z 加上读取值的大小。

使用统一解码码表来解 CAVLC 的过程如下:

(1) 从码流中读取长度为 k_1 的比特, 以其大小 b 作为坐标值到统一解码码表中查找结果。

(2) 如果 $Y > 0$ 则 Z 即为所要的结果, 解码完成; 如果 $Y < 0$ 则 $-Y$ 为下一次需要读取的比特个数, Z 为地址偏移。

表 1 统一解码码表

Tab 1 Uniform decoding table

坐标值	Y	Z	坐标值	Y	Z	坐标值	Y	Z
0	3	2	12	1	5	24	1	7
1	-3	8	13	1	5	25	1	7
2	2	1	14	1	5	26	1	7
3	2	1	15	1	5	27	1	7
4	-2	16	16	1	6	28	1	15
5	3	3	17	1	6	29	1	15
6	3	4	18	2	9	30	1	15
7	-3	20	19	2	10	31	1	15
8	2	8	20	-3	28	32	3	17
9	2	8	21	3	14	33	3	18
10	3	12	22	2	11	34	2	16
11	3	13	23	2	11	35	2	16

(3) 从码流中读取 $-Y$ 个比特, 以其大小 b 加上 Z 作为坐标值来查找结果, 并把获得的结果返回第 2 步进行比较。

5 性能分析

下面以 $0 \leq NC < 2$ 时解码总的非零系数个数 (TotalCoeff) 和拖尾系数的个数 (TrailingOnes) 为例来比较 Hashemian 方法和我们所介绍的改进方法。当限制存储空间 $A = 120$ 的时候, Hashemian 解码方法使用固定长度 $H = 4$ 分割编码码表, 而改进方法使用长度为 $5, 4, 4, 3$ 分割编码码表。两者的性能比较如表 2。

表 2 中 C 表示查表, 1C 表示查一次表, S 表示地址转移, 1S 表示一次地址的偏移。查表的次数和转移的次数越少则解码的速度越快。如果查一次表的时间和一次地址转移的时间相同, 由上表可以知道, 在码字长度小于 4 的时候, 两种方法的解码速度是相同的, 但是当码字长度大于 4 的时候, 改进方法的解码速度就比 Hashemian 方法要快, 特别是当码字长度为 5 的时候, 改进方法的速度是 Hashemian 方法的 5 倍; 其他情况下速度的提升也在 2 倍左右。因此, 相对于 Hashemian 方法, 使用改进方法可以有效提高解码速度。但同时我们发现随着 A 的提高两者之间的性能差异逐渐减小。

限定存储空间 $A = 120$ 。在 JM 模型下, 使用 JVT 提供的标准解码方法则解码器在 CAVLC 上花费的时间为 16% 左右, 使用 Hashemian 方法这个比

表 2 性能比较

Tab 2 Performance comparing

码字长度	相同长度的个数	Hashemian 方法	改进方法
1	1	1C	1C
2	1	1C	1C
3	1	1C	1C
4	0	NULL	NULL
5	1	3C 2S	1C
6	3	3C 2S	2C 1S
7	2	3C 2S	2C 1S
8	4	3C 2S	2C 1S
9	4	5C 4S	2C 1S
10	4	5C 4S	3C 2S
11	4	5C 4S	3C 2S
12	0	NULL	NULL
13	8	7C 6S	3C 2S
14	8	7C 6S	4C 3S
15	9	7C 6S	4C 3S
16	12	7C 6S	4C 3S

例下降到 7.5% 左右, 而使用改进方法则只需要 5% 左右。可以看出改进方法具有明显的优势。

6 结 论

在限定的存储空间条件下, 改进方法通过使用更好的码表分配和码表地址偏移技术, 使得解码速度得到极大的提高, 从而为 H. 264 标准应用于实时环境提供了一种可行的解决方案。

参考文献 (References)

- 1 Bjontegaard G, Lilevold Karl. Context-adaptive VLC (CVLC) coding of coefficients[R]. Fairfax Virginia: Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, 3rd Meeting, 2002.
- 2 Bi Hou-jie. A Novel Video Coding Standard H. 264/AVC [M]. Beijing: Posts and Telecom Press, 2005: 119~121. [毕厚杰主编. 新一代视频压缩编码标准——H. 264/AVC[M]. 北京: 人民邮电出版社, 2005: 119~121.]
- 3 Iyengar V, Chakrabarty K. An efficient finite-state machine implementation of Huffman decoders [J]. Information Processing Letters, 1997, 64(6): 271~275.
- 4 Hashemian R. Memory efficient and high-speed search Huffman coding [J]. IEEE Transactions on Communications, 1995, 43(1): 2576~2581.
- 5 Hong D, Eleftheriadis A. Automatic generation of C++/Java code for efficient VLC decoding [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2005, 21(3): 253~260.

利用 Hough 变换实现直线的快速精确检测

滕今朝¹⁾ 邱杰²⁾

¹⁾ (威海职业学院机电工程系, 威海 264210) ²⁾ (海军航空工程学院, 烟台 264000)

摘要 利用 Hough 变换对直线进行检测, 通常存在“速度缓慢、结果不够精确”的问题, 本文提出了“分式查表法”, 能在大幅度减少 Hough 变换的总计算量的情况下, 检测精度保持最高, 从而使超大型图像中, 直线的实时、精确检测成为可能。

关键词 Hough 变换 参数空间 精度 分式查表法

中图分类号: TP391.41 文献标识码: A 文章编号: 1006-8961(2008)02-0234-04

Fast and Precise Detection of Straight Line with Hough Transform

TENG Jin-zhao¹⁾, QU Jie²⁾

¹⁾ (Electromechanical Engineering Department, Weihai Vocational College, Weihai 264210)

²⁾ (College of Navy Aviation Engineering, Yantai 264000)

Abstract Problems as low speed or inaccurate results in the process of line detection with Hough Transform remain unsatisfactorily solved. This paper puts forward a new look-up table to decrease the computation distinctly and keeping the highest precision. It offers the possibility in real-time applications especially in large image.

Keywords Hough Transform, parameter space, precision, table look-up

1 引言

Hough 变换具有优异的鲁棒性和极佳的抗干扰能力, 利用 Hough 变换进行直线检测, 是图像分析和计算机视觉的一个重要内容。但是 Hough 变换的计算量往往非常大, 从而阻碍了其在快速、精确检测直线方面的应用。

本文提出的新方法, 不仅能大幅度减少 Hough 变换的总计算量, 而且在像素允许的情况下, 直线斜率的检测精度保持最高, 这对于超大型图像中直线的实时、精确检测, 具有重要的实用价值。

2 Hough 变换检测直线的原理

选取图像空间中一条直线 L 的某些特征, 作为参数空间的一个点 M , 并且该直线 L 上所有点, 通过

某种算法, 都能够对应着这些特征, 从而在图像空间和参数空间之间, 建立起“线-点”的对偶性。Hough 变换就是根据这种对偶性, 将图像空间中直线的检测问题, 转化为参数空间中点的检测问题, 而后的处理比前者要简单得多, 进行累加统计即可。

2.1 用极坐标建立参数空间

常用的 Hough 变换检测直线的方法, 是运用下式在图像空间和参数空间之间, 建立对偶变换。

$$\rho = x \cos \alpha + y \sin \alpha$$

式中, ρ 为极径; α 为极角, α 取 $0 \sim 180^\circ$; x 为像素点相对图像原点的行坐标; y 为像素点相对图像原点的列坐标。

为了检测出直角坐标系中, 由非零点所构成的直线, 需要根据检测分辨率的要求, 将 α 离散化为 N_α 个参数区间, 将 ρ 离散化为 N_ρ 个参数区间, 也就是说将极坐标系量化成许多小格, 建立参数空间。

这种方法被称为标准 Hough 变换方法 (standard

收稿日期: 2006-04-25 改回日期: 2006-10-17

第一作者简介: 滕今朝 (1970~), 男, 讲师。2007 年于海军航空工程学院获测试计量仪器与科学专业硕士学位。主要从事电气自动化、检测技术方面的教学和研究。E-mail: twrit@163.com

hough transform, SHT)。其优点是: 无论直线怎样变化, 参数空间中 α 和 ρ 的取值范围是有限的。所以, 目前的直线检测大多数都是基于这种方法。

但是, 这种方法在 N_α 值较大的情况下, 存在以下两个缺陷:

(1) 计算量大

N_α 越大, α 的步长越小, 计算量就越大^[1]。在要求检测精度很高的场合, N_α 的值往往非常大, 这样会使计算量大增。

(2) 需要大的存储空间

如果 α 和 ρ 都占 4 个字节, 参数空间所需要的存储空间的字节数 S 可由下式求出:

$$S = 4N_\alpha N_\rho \quad (1)$$

式中, N_α 为 α 在 $[0, \pi)$ 间取的离散值的个数; N_ρ 为 ρ 的采样个数。

对较大的图像, S 将大于数千兆字节 (GB), 单靠物理内存, 难以满足这样的要求。

为了减少这种 Hough 变换的计算量, 减小所需要的存储空间, 在此基础上出现了很多改进的 Hough 变换算法, 例如分块检测法^[1]、两次检测法^[2]、全整数 Hough 变换^[3]等, 为讨论方便, 将其统称为正弦 Hough 变换方法。但这些方法由于受 α 的步长限制, 有时候检测结果不尽人意, 难以实现对任意斜率直线的快速、精确检测, 尤其是对超大型图像 (例如像素在 2048×2048 以上)。

如果能够建立一个不受 α 步长限制的参数空间, 就有可能实现对任意斜率直线的快速、精确检测。

2.2 用直线的斜率 k 和截距 b 建立参数空间

图像空间中, 直线 $L_0: y = k_0x + b_0$ 上每一个点, 在参数空间中都代表一条直线, 这些直线都相交于一点 $M(k_0, b_0)$ 。

与用极坐标建立参数空间的方法相比, 这种方法不受 α 的步长限制, 检测了所有的可能出现的直线, 不会有任何遗漏, 在像素允许的情况下, 能精确地检测出图像中的任意直线。这种优点决定了它非常适合用来对直线进行精确检测。

为便于讨论, 将其称为 kb Hough 变换方法。

虽然这种方法在原理上非常明确, 但在具体实现过程中, 如果简单地用浮点数进行斜率和截距的计算, 还有以下 3 个问题需要解决:

(1) 如果同时计算斜率和截距, 参数空间结构数组可能异常庞大, 而且计算量非常大

1 幅像素为 $m \times n$ 的图像, 假设图像中可能出现的不重复的斜率有 u 种 (u 大于图像的像素总数)。若这些数据都占 4 个字节, 那么要求的计算机内存为

$$w_1 = 4umn \quad (2)$$

对较大的图像, 要求的计算机内存超过上千兆字节 (GB), 这显然是不现实的。

如何解决这个问题?

很多时候, 如果将一个 2 维问题分解为两个 1 维问题来解决, 往往非常方便。

参数空间数组之所以异常庞大, 是因为同时包含了待测的斜率和截距, 而斜率和截距的组合, 决定了其数组的元素个数必然非常多。

如果将斜率 k 和截距 b 分步进行检测, 先检测斜率, 找出出现次数最多的斜率 k_0 , 然后再检测截距, 找出斜率为 k_0 的, 出现次数最多的直线的截距 b_0 , 也就是说, 将一个 2 维问题分解为两个 1 维问题。由于斜率为 k_0 的二点对的数量往往非常有限, 这样就能有效地减小参数空间。

但是, 即使将斜率 k 和截距 b 分步进行检测, 如果用浮点数进行计算斜率, 并采用普通的 Hough 变换的投票方式, 即先计算完由非零点组成的所有的二点对的斜率, 再对斜率进行统计, 也可能需要很大的参数空间。1 幅 800×600 的图像, 如果非零点非常多, 需要的参数空间仍然可能超过数千兆字节 (GB)。

所以, 用浮点数计算斜率, 要减小参数空间, 只能一边计算斜率, 一边统计斜率相同的情况 (与前面的斜率进行比较)。即边投票, 边计票。但是, 这样计算量非常大。

(2) 程序的复杂程度高

根据定量度量程序的复杂程度的 McCabe 方法, 流图 (也称为程序图) 的环形复杂度可由下式求出:

$$V(G) = R + 1 \quad (3)$$

式中, $V(G)$ 为环形复杂度; R 为流图中的判定节点数目。通常 $V(G) \leq 10$ 为宜。

根据 McCabe 方法, 求出该方法的流图的判定节点数 R 为 10 由式 (3) 可知, 其环形复杂度为 11, 显然复杂程度太高。

(3) 如果直接采用浮点数计算斜率, 无法处理斜率为无穷大的垂直直线

斜率为无穷大, 给计算带来不便, 这正是极少有人采用这种方法的原因。

综上所述,除了需要分步检测斜率和截距外,还不能简单地采用浮点数计算斜率。因此采用了斜率分式查表方法来解决这些问题。

3 斜率分式查表法

斜率分式查表法具体方案如下:

将斜率用分式表示,分子、分母都是整数。但是有时候,由不同的整数表示的斜率可能是相同的,例如“3/9”和“4/12”实际是相同的,所以要进行化简等处理。

事先要建立表格,可以在初始化中进行:以一幅 $m \times n$ 的图像为例(假设 $m > n$)。

(1) 将 1~800 的整数分解,存入数组 *bd* 中。例如 14 被分解为“1×14”、“2×7”,将 10 分解为“1×10”、“2×5”。

(2) 将 1~800 的整数两两相除,利用第 1 步的结果进行化简,化简的结果存入结构数组 *reduct* 以建立表格。

例如“14/10”,利用第 1 步的结果,化简为“7/5”,记为 *reduct*[14].*sr*[10][0]=7, *reduct*[14].*sr*[10][1]=5 *sr* 为结构数组 *reduct* 内的数据成员。将来在检测斜率的时候,用数组 *reduct* 对斜率进行化简:如果遇到斜率为“14/10”,通过查表 *reduct*[14].*sr*[10] 立刻知道,该斜率的化简结果为“7/5”。

建立好表格之后,在检测斜率时,用结构数组 *reduct* 来查表、化简,根据结果,将参数空间的结构数组 *count* 中的相应计数器的值加 1,即结构数组 *count* 保存该斜率出现的次数。

对 1 幅像素为 $m \times n$ 的图像,由于结构数组 *reduct* 存储化简之后有分子、分母,所以设定其元素个数为 $2mn$ 。考虑到斜率有正负,设定 *count* 数组元素个数为图像的像素数的 2 倍,即 $2mn$ 。若忽略结构数组 *bd* 占用的存储空间(对于一幅 800×600 的图像,只有 0.256MB),则求斜率需要的内存字节数为

$$w_k = 16mn$$

求出斜率之后,再求截距。假设斜率为 k_0 的直线有 mn 条(实际上不可能有这么)。考虑到截距有正负,则设定截距数组元素个数为 $2mn$ 。求截距需要的内存字节数为

$$w_b = 8mn$$

至此,参数空间总共需要内存字节数为

$$w_2 = w_k + w_b = 24mn \quad (4)$$

这个数值与 *kb Hough* 变换方法需要的内存相比如何呢?

为讨论方便,假设 $m = n$ 。将式(4)与式(2)相比:

$$w_2 / w_1 = 3/u$$

u 大于图像的像素总数 n^2 , 即 $u > n^2$, 故

$$w_2 / w_1 < 3/(n^2)$$

斜率分式查表法需要的内存,不到 *kb Hough* 变换方法的 $3/(n^2)$ 。可见,图像越大,斜率分式查表法的优势越明显。

与标准 *Hough* 变换方法需要的内存相比,情况又如何呢?

将式(4)与式(1)相比:

$$w_2 / S = 6n^2 / (N_a N_p)$$

而 $N_a = u > n^2$, $N_p > 2n$, 故

$$w_2 / S < 3/n$$

若 $n = 1000$ 斜率分式查表法所需的内存,不足标准 *Hough* 变换方法的 $3/1000$ 优势相当明显。以像素为 800×600 的图像为例,上述参数空间总共需要的内存小于 12MB。

同时,还解决了前面 *kb Hough* 变换方法的以下两个难题:

(1) 关于减小计算量,降低程序的复杂程度。

由上述分析可知,斜率分式查表法只是在初始化时建立表格,而在此后的计算过程中,完全不需要将每一个二点对的斜率,与参数空间中的各个斜率逐一比较,极大地减小了计算量。

根据 McCabe 方法,可以求出这种方法的流图的判定节点数 R 为 6 由式(3)可知,其环形复杂度为 7,可见采用斜率分式查表法与 *kb Hough* 变换方法相比,大大降低了程序的复杂程度。

(2) 关于斜率可能为无穷大的问题。

对于垂直状态的直线斜率为无穷大的情况,只需要在参数空间结构数组中,特设一个数组元素,例如 *count*[1].*sc*[0] 作为计数器,专门对无穷大的斜率进行计数即可。

至此,斜率分式查表法解决了用直线的斜率 k 和截距 b 建立参数空间所遇到的难题,为快速、精确地检测出图像中的任意直线,提供了有力工具。

4 测试与验证

运用斜率分式查表方法,能快速、精确地检测出

图像中的任意直线。这一点已经在个人计算机上经过充分的验证, 并与标准 Hough 变换方法进行了比较。所用的软硬件环境如下:

硬件平台: CPU 为 Intel Celeron (TM), 主频 433MHz, 内存是 128MB。

软件平台: 操作系统为中文 Windows 2000, 算法程序语言是 C++, 编译器用 Visual C++ 6.0。

表 1 列出了一组各种尺寸的含有椒盐噪声图像的直线测试结果。

表 1 测试结果
Tab 1 Test result

图像尺寸 (pixel)	分辨率要求 (°)	标准 Hough 变换方法 所用时间 (s)	斜率分式 查表法 所用时间 (s)
128 × 128	0.009	6.01	0.31
256 × 256	0.00227	360.35	0.73
800 × 600	0.0003	4962.28	1.32
1024 × 1024	0.00014		2.15
2048 × 2048	0.0000353		3.73

测试表明, 当检测精度要求较低时 (α 步长大于 0.01°), 标准 Hough 变换方法能够较为迅速地为目标进行检测; 当检测精度要求较高时 (α 步长小于 0.01°), 标准 Hough 变换方法的检测速度非常慢。

以一幅 800×600 的图像为例, 要精确地检测像素为 800×600 的图像中的每一条直线, 分辨率不能低于 0.0003° 。用极坐标为参数空间的标准 Hough 变换方法, 耗时 4962.28s, 而运用斜率分式查表方法, 耗时 1.32s, 后者比前者要快上千倍。而对于在

标准 Hough 变换方法基础上改进的各种正弦 Hough 变换方法来说, 速度能提高 100 倍的, 已经算是非常快的了。

对于尺寸在 1024×1024 以上的大型图像, 分辨率不能低于 0.00014° , 用极坐标为参数空间的标准 Hough 变换方法, 运行数小时, 仍然没有结果, 已失去了实用意义; 运用斜率分式查表方法, 仍然较为迅速, 而且随着图像尺寸的增大, 所用时间增加得并不明显。

5 结 论

通过进一步优化程序, 提高计算机配置, 斜率分式查表的 Hough 变换方法, 能快速、准确地检测出目标直线, 而且图片越大, 这种查表方法的优越性就越明显, 从而使超大型图像中直线的实时、精确检测成为可能。

参考文献 (References)

- 1 Qiu Liwei, Song Zishan, Shen Weiqun. Kind of fast Hough transform used in line detection [J]. Journal of Beijing University of Aeronautics and Astronautics, 2003, 8(29): 741~744 [邱力为, 宋子善, 沈为群. 直线参数检测的快速 Hough 变换 [J]. 北京航空航天大学学报, 2003, 8(29): 741~744]
- 2 Zhang Yongzhong, Zhu Ying. Application of Hough transform in edge detection [J]. China Science and Technology Information, 2006, 13: 47~48 [张永忠, 朱英. Hough 变换在物体边缘检测上的应用 [J]. 中国科技信息, 2006, 13: 47~48]
- 3 Ohta G, Magli E. A half-integer Hough transform performance evaluation [A]. In: Proceedings of International Conference on Image Processing [C], Thessaloniki, Greece, 2001: 338~341.

基于高温辐射体颜色信息的目标图像识别方法

彭小奇 孙元 王一丁

(中南大学物理科学与技术学院, 长沙 410083)

摘要 鉴于准确识别待测目标是利用 CCD 图像传感器进行高温测量的关键, 为此在对高温辐射体彩色图像的噪声特点进行分析的基础上, 提出了一种基于高温辐射体颜色信息的目标图像识别方法。该方法首先通过对红、绿基色图像分别进行分割来减小甚至消除高温辐射体图像的各种噪声; 然后综合红、绿基色图像的分割结果实现高温辐射体的准确识别, 并运用数学形态学方法对分割结果进行后处理, 以消除游离点和孔洞, 使图像边缘平滑。实验结果表明, 该方法可以有效地实现高温辐射体的正确识别, 具有较强的实用性。

关键词 目标识别 CCD 图像传感器 温度测量

中图分类号: TP391.41 文献标识码: A 文章编号: 1006-8961(2008)02-0238-04

An Object Image Recognition Method Based on the Color Information of A High Temperature Radiator

PENG Xiao-qi, SUN Yuan, WANG Yi-ding

(School of Physics Science and Technology, Central South University, Changsha 410083)

Abstract The key of high temperature measurement by using CCD image sensor is to recognize the measured object accurately. From analyzing the characteristic of the noises in high temperature radiator color image, an image recognition method of high temperature radiator is proposed, which is based on the color information of a high temperature radiator. By segmenting the red and green color images, the various noises in high temperature radiator color image are reduced or even eliminated. And then, by synthesizing the results of the segmented red and green color images, the high temperature radiator is recognized accurately. Finally, in order to smooth the image's edge, the morphology method in mathematics is applied to process the segmented image and remove the dissociations and narrow holes. The experimental results show that the method can recognize high temperature radiator with high accuracy and practicability.

Keywords object image recognition, CCD image sensor, temperature measurement

1 引言

由于 CCD 图像传感器具有图像清晰度高、耐灼伤、耐震动冲击、工作稳定可靠、能很好地与数字图像处理技术、计算机技术相结合等优点, 使得基于 CCD 的非接触式测温技术成为高温检测领域的研究热点之一^[1,2]。在对利用 CCD 图像传感器获得的高温辐射图像进行温度计算时, 必须先从辐射图像中准确识别出待测目标, 以避免计算温度时出现奇异值^[3]。文献 [4] 提出了一种分别采用双峰法和最

大方差自动取阈值法确定分割阈值, 再根据高温辐射体的灰度值近似相等的特征选择合适的分割阈值来识别出待测目标的算法。文献 [5] 提出了一种基于 Bubble 小波变换的高温辐射体图像分割算法。由于这些算法的共同特点都是把 CCD 摄得的彩色图像转换为灰度图像后, 再进行分割, 即只利用了辐射图像的亮度信息, 因此当噪声亮度和辐射体亮度相近时, 则无法通过灰度分割准确识别高温辐射体。

由于彩色图像比灰度图像能提供更多的目标对象信息, 并可简化目标物的区分^[6]; 同时, 随着计算机处理能力的显著增强, 快速处理彩色图像的硬件

基金项目: 国家自然科学基金项目 (50374097); 博士点专项科研基金项目 (20030533008); 湖南省自然科学基金项目 (06JJ2066)

收稿日期: 2006-05-12 改回日期: 2006-09-27

第一作者简介: 彭小奇 (1962~), 男, 教授, 博士生导师, 博士。主要研究方向为系统建模与智能决策。E-mail: peng21@mail.csu.edu.cn

条件也已经成熟^[7], 基于灰度信息的图像分割算法很容易应用于彩色图像分割^[8], 因此, 近年来, 彩色图像分割正逐步引起人们的关注^[9]。

本文在对高温辐射体的 RGB 三基色图像的噪声特点进行分析的基础上, 提出首先通过对红、绿基色图像分别进行分割来减小甚至消除高温辐射体图像的各种噪声; 然后综合红、绿基色图像的分割结果实现高温辐射体的准确识别, 并运用数学形态学方法对分割结果进行后处理, 以消除游离点和孔洞, 使图像边缘平滑。实验表明, 该方法可以有效地实现高温辐射体的正确识别, 具有较强的实用性。

2 高温辐射体图像的噪声分析

在利用 CCD 图像传感器进行非接触高温测量时, 由于测量现场环境恶劣、工况复杂, 使 CCD 摄得的高温辐射体图像往往带有很多噪声, 其主要包括以下几种:

- (1) 雾状噪声: 由辐射光路上的烟雾和水汽等的折射和散射造成;
- (2) 点状噪声: 由辐射光路上的烟尘造成;
- (3) 光晕: 由辐射光在穿透观察窗和镜头时产生的反射造成;
- (4) 杂散光噪声: 由环境光源的辐射等杂散光造成。

由于雾状、点状噪声和光晕大多处于长波段, 主要干扰红基色图像; 杂散光噪声大多处于短波段, 主要干扰绿基色图像, 因此, 利用不同噪声的特点可以减小甚至消除噪声。

3 高温辐射体识别方法

3.1 直方图阈值化分割算法

利用阈值化分割算法来分割灰度图像主要有两个步骤, 首先在图像灰度值取值范围内确定一个适当的分割阈值 T , 然后按式 (1) 将原图像的各个像素分为目标像素和背景像素,

$$I(x, y) = \begin{cases} 1 & f(x, y) \geq T \\ 0 & f(x, y) < T \end{cases} \quad (1)$$

这样得到的 $I(x, y)$ 是一幅二值图像, 用原始图像 $f(x, y)$ 和它相乘就可得到分割后的前景图像^[10]

$$g(x, y) = f(x, y) \times I(x, y) \quad (2)$$

阈值 T 的选择对分割效果影响很大。因迭代法收敛速度较快, 抗噪性较强, 故经常被用于求取分

割阈值。设原始图像的共有 L 个灰度级, 则选择阈值 T 时可首先取图像灰度范围的中值作为初始阈值 T_0 , 然后按下式进行迭代:

$$T_i = \frac{1}{2} \left| \frac{\sum_{k=0}^{T_{i-1}} h_k \cdot k}{\sum_{k=0}^{T_{i-1}} h_k} + \frac{\sum_{k=T_{i-1}+1}^{L-1} h_k \cdot k}{\sum_{k=T_{i-1}+1}^{L-1} h_k} \right| \quad i = 1, 2, \dots \quad (3)$$

其中, k 是灰度值, 离散函数 h_k 是原图像中灰度值为 k 的像素个数, i 代表迭代次数, T_i 表示第 i 次迭代得到的阈值。迭代进行到 T_i 与 T_{i-1} 的差小于某个事先设定的值时结束, 并取迭代结束时的 T_i 为分割阈值 T 。因由此得到的 T 处在与两个灰度区域的重心成反比的位置, 故从路径规划的角度看, 迭代阈值 T 是一种最优阈值^[11]。

当待分割图像的直方图具有良好的双峰特征时, 直方图阈值化分割算法可取得较理想的分割效果^[12]。因高温辐射体的红、绿基色图像直方图都具有典型的双峰特征, 故直方图阈值化分割算法适用于高温辐射体的分割。

3.2 利用绿基色图像分割去除烟雾和光晕干扰

由于辐射体温度升高时, 其光谱辐射亮度曲线的峰值波长向短波方向移动, 而雾状、点状噪声和光晕则主要集中于红色波段, 因此对绿基色图像进行分割可以有效地去除这类干扰。设 $c(x, y)$ 为原图像中坐标 (x, y) 的彩色空间向量,

$$c(x, y) = \begin{bmatrix} R(x, y) \\ G(x, y) \\ B(x, y) \end{bmatrix} \quad (4)$$

对原图像的绿色分量 $G(x, y)$ 可先利用式 (3) 求取分割阈值 T^G , 然后利用式 (5)、式 (6) 得到用绿基色图像分割后的前景图像 $\hat{G}(x, y)$

$$f^G(x, y) = \begin{cases} 1 & G(x, y) \geq T^G \\ 0 & G(x, y) < T^G \end{cases} \quad (5)$$

$$c^G(x, y) = c(x, y) \times f^G(x, y) = \begin{bmatrix} R(x, y) \times f^G(x, y) \\ G(x, y) \times f^G(x, y) \\ B(x, y) \times f^G(x, y) \end{bmatrix} \quad (6)$$

3.3 利用红基色图像分割去除杂散光干扰

杂散光干扰处于电磁波谱的短波段, 其绿色分量值与高温辐射体的绿基色图像像素值相近, 因此在对绿基色图像进行分割时, 杂散光干扰会被误作为目标分割出来, 而对红基色图像进行分割却可以

有效去除杂散光干扰。

设 $c(x, y)$ 为原图像的彩色空间向量, 对原图像
 的红色分量 $R(x, y)$ 可先利用式 (3) 求取分割阈值
 T^R , 然后利用式 (7)、式 (8) 即可得到利用红基色图
 像分割后的前景图像 $c^R(x, y)$,

$$I^R(x, y) = \begin{cases} 1 & R(x, y) \geq T^R \\ 0 & R(x, y) < T^R \end{cases} \quad (7)$$

$$c^R(x, y) = c(x, y) \times I^R(x, y) \quad (8)$$

3.4 高温辐射体分割方法

对高温辐射体的红、绿基色图像进行分割后, 若
 取分割结果的交集作为高温辐射体图像的结果, 则既
 可以消除杂散光干扰, 又可以消除光晕和烟雾干
 扰。设 $c(x, y)$ 为原图像的彩色空间向量, 通过

式 (9)、式 (10) 就可以得到利用红、绿基色图像分割
 后的前景图像 $c^{RG}(x, y)$,

$$I^{RG}(x, y) = \begin{cases} 1 & R(x, y) \geq T^R \text{ 且 } G(x, y) \geq T^G \\ 0 & \text{其他} \end{cases} \quad (9)$$

$$c^{RG}(x, y) = c(x, y) \times I^{RG}(x, y) \quad (10)$$

图 1(a) 是含有光晕和杂散光干扰的黑体辐射图
 像, 当分别利用灰度双峰法和文献 [5] 中提出的基
 于 Bubble 小波的分割算法对两者进行识别, 其结果如
 图 1(b) 和图 1(c) 所示, 图 1(d) 是使用本文提出
 的方法分割的效果; 图 2(a) 是含有光晕和烟雾干
 扰的某铜冶炼厂闪速炉放铜口的铜熔体图像, 图 2(b) ~
 图 2(d) 是分别利用灰度双峰法、基于 Bubble 小波
 的分割算法和本文中的分割法进行识别的结果。

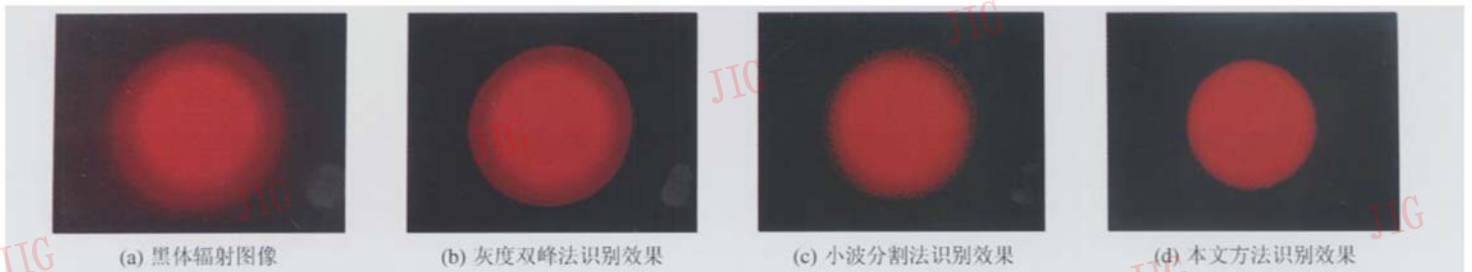


图 1 高温黑体的识别

Fig.1 Recognition of high temperature blackbody



图 2 高温熔体的识别

Fig.2 Recognition of high temperature melts

通过对比可见, 本文提出的基于高温辐射体颜
 色信息的识别方法由于有效减小了光晕、烟雾和杂
 散光的干扰, 从而提高了识别的准确度。

4 图像分割结果的后处理

由于烟雾、灰尘等吸收高温辐射体的辐射能后,
 将使辐射图像局部颜色变暗, 往往导致分割后的
 高温辐射体图像存在游离点和孔洞, 因此为不影
 响测温结果的准确性, 必须选择一种既可以去除
 游离点和孔洞, 又不改变原图像各像素颜色值分
 割结果的后处理方法。

本文采用基于数学形态学的图像处理方法首
 先将图像分割结果二值化, 然后对二值化图像进
 行开、闭操作, 并将操作结果与原图像相乘。设
 $c(x, y)$ 为原图像的彩色空间向量, 利用式 (11) ~
 (14) 可以得到对分割图像进行数学形态学 (mor
 photogy, M) 处理后的彩色空间向量 $c^M(x, y)$,

$$I^M(x, y) = \begin{cases} 1 & R(x, y) \geq T^R \text{ 且 } G(x, y) \geq T^G \\ 0 & \text{其他} \end{cases} \quad (11)$$

$$S = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \quad (12)$$

$$J = [I^M \bullet S] \circ S \quad (13)$$

$$\hat{c}^n(x, y) = c(x, y) \times J(x, y) \quad (14)$$

式中, $A \bullet S$ 表示使用结构元素矩阵 S 对矩阵 A 进行闭操作, $A \circ S$ 表示使用结构元素矩阵 S 对矩阵 A 进行开操作。

闭操作可以弥补狭窄的间断和消除小的孔洞, 并可填补轮廓线中的断裂; 开操作可以使对象的轮廓变得光滑, 并可消除细的突出物, 因此先进行闭操作, 后进行开操作可以除去比结构元素小的特定图像细节, 同时能保证不产生全局几何失真^[13]。由于仅对二值化图像进行开、闭操作, 因此不会改变原图像各像素的颜色值。图 3(a)、图 3(b) 分别是图 1(d)、图 2(d) 经过数学形态学处理后的结果, 由该图可见, 采用基于数学形态学的图像后处理可以有效地去除游离点和孔洞, 并可使图像边缘光滑。

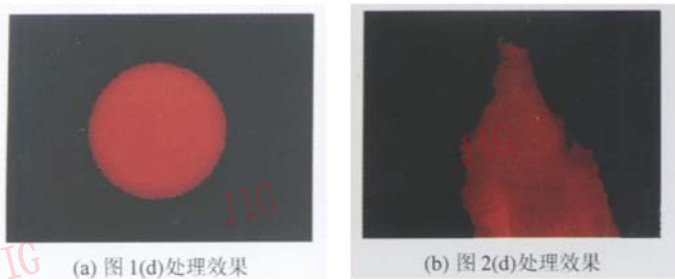


图 3 形态学处理效果

Fig 3 Result of morphology method

5 结 论

综上所述, 可得以下结论:

(1) 由于雾状、点状噪声和光晕干扰主要集中在红色波段, 杂散光干扰主要处于短波段, 因此, 通过对红、绿基色图像分别进行分割可以减小甚至消除高温辐射体图像的这类噪声;

(2) 对图像分割结果进行数学形态学处理, 可以在不引起全局几何失真的情况下消除图像干扰导致的游离点和孔洞, 并可使图像边缘光滑;

(3) 本文提出的基于高温辐射体颜色信息的目标图像识别方法, 由于可以充分利用高温辐射体彩色图像的颜色信息去除噪声干扰, 因此可使分割结果准确, 具有较强的实用性。

参考文献 (References)

1 Wang Yao-ke, Yang Hua-yong, Lu Hai-bao, *et al*. The present situation and application prospects of CCD image sensing technique [J]. *Optical Instrument*, 1996, 18(5): 32~37. [王跃科, 杨华勇, 吕海宝等. CCD 图像传感技术的现状与应用前景 [J]. 光学仪

器, 1996, 18(5): 32~37.]

- 2 Liu Zhen, Peng Xiao-qi, Ding Jian, *et al*. The application and development of CCD measurement technology in foreign industrial fields [J]. *Industrial Instrumentation & Automation*, 2005, (4): 65~69. [刘征, 彭小奇, 丁剑等. 国外 CCD 检测技术在工业中的应用与发展 [J]. 工业仪表与自动化装置, 2005, (4): 65~69.]
- 3 Zhang Yu-jie, Lu Cen. Study on measurement method of flame temperature field in hearth based on color CCD [J]. *Chinese Journal of Quantum Electronics*, 2004, 21(5): 674~678. [张玉杰, 吕岑. 基于彩色 CCD 炉膛火焰温度场测量方法研究 [J]. 量子电子学报, 2004, 21(5): 674~678.]
- 4 Song Hai-ying, Peng Xiao-qi, Liu Zhen, *et al*. Recognition method in contactless temperature measurement of high temperature melts [J]. *Journal of Central South University (Science and Technology)*, 2005, 36(3): 426~429. [宋海鹰, 彭小奇, 刘征等. 基于非接触式温度测量中的高温辐射体识别方法 [J]. 中南大学学报 (自然科学与技术版), 2005, 36(3): 426~429.]
- 5 Ding Jian, Peng Xiao-qi, Liu Zhen. The application of bubble wavelet in the images segmentation of high temperature melt [J]. *Energy Technology*, 2005, 26(4): 139~154. [丁剑, 彭小奇, 刘征. Bubble 小波在高温辐射体图像分割中的应用 [J]. 能源技术, 2005, 26(4): 139~154.]
- 6 Gonzalez Rafael C, Woods Richard E. *Digital Image Processing (Second Edition)* [M]. Ruan Qiu-qi, *et al* Translate. Beijing: Publishing House of Electronics Industry, 2003. 224. [冈萨雷斯, 伍兹. 数字图像处理 (第二版) [M]. 阮秋琦等译. 北京: 电子工业出版社, 2003. 224.]
- 7 Yin Ling, Huang Huixian. Color image segmentation: advances and prospects [J]. *Social Science Journal of Xiangtan University*, 2003, 27(May): 236~237. [尹令, 黄辉先. 彩色图像分割: 优势和展望 [J]. 湘潭大学社会科学学报, 2003, 27: 236~237.]
- 8 Wang Ze-bin, Yang Chao-hui. Study on color image segmentation technique [J]. *Digital Television and Digital Video*, 2005, (4): 20~24. [王泽兵, 杨朝晖. 彩色图像分割技术研究 [J]. 数字电视与数字视频, 2005, (4): 20~24.]
- 9 Cheng H D, Jiang X H, Sun Y, *et al*. Color image segmentation advances and prospects [J]. *Pattern Recognition*, 2001, 34(12): 2259~2281.
- 10 Cheng H D, Jiang X H, Wang Jing-li. Color image segmentation based on histogram thresholding and region merging [J]. *Pattern Recognition*, 2002, 35(2): 373~393.
- 11 Zhang Yu-jin. *Image Segmentation* [M]. Beijing: Science Publisher, 2001. [章毓晋. 图像分割 [M]. 北京: 科学出版社, 2001.]
- 12 Xi Li-chun, Yan Hong-wei. The technologies of color image segmentation [J]. *Microcomputer Development*, 2003, 13(4): 46~48. [席砺纯, 闫宏伟. 彩色图像的分割技术 [J]. 微机发展, 2003, 13(4): 46~48.]
- 13 Fecit Science and Technology Study Center. *MATLAB 6.5 Assistant Image Processing* [M]. Beijing: Publishing House of Electronics Industry, 2003. 200. [飞思科技产品研发中心. MATLAB 6.5 辅助图像处理 [M]. 北京: 电子工业出版社, 2003. 200.]

基于彩色 Gabor 特征的人脸识别技术

罗亚兰 陈锻生

(华侨大学计算机科学系, 泉州 362021)

摘要 提出了一种利用所提取的彩色 Gabor 特征来提高人脸识别系统性能的方法。首先利用四元数表示彩色信息, 考虑到 Gabor 滤波器具有空间局部性和方向选择性的特点, 将其扩展到四元数空间。然后通过人脸图像特征点与 Gabor 滤波器的卷积来提取特征, 这样就将传统的灰度 Gabor 特征拓展为彩色 Gabor 特征。最后对于所提取的特征利用 PCA 降维后送入支持向量机中分类。实验采用彩色 FERET 人脸库并利用 ROC 曲线进行交叉检验, 结果说明通过提取和利用这种彩色纹理信息能显著提高人脸识别系统性能。

关键词 彩色 Gabor 彩色纹理 特征提取 支持向量机

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2008)02-0242-05

Face Recognition Based on Color Gabor Features

LUO Ya-lan, CHEN Duan-sheng

(Department of Computer Science, Huaqiao University, Quanzhou 362021)

Abstract In this paper a method which can improve the performance of face recognition system using color Gabor features is presented. First, quaternion is used to describe the color information, considering that Gabor filters have desirable characteristics of spatial locality and orientation selectivity, and they are extended to quaternion space. Then utilizing the convolution of the key points and the Gabor filters to extract features, by doing this the gray Gabor features are extended to the color ones. In the end, for the extracted features, we used PCA for dimension reduction and SVM for recognition. The experiment carried on Color FERET Database and the result utilizing ROC curve for cross-validation show that the use of color texture information can improve the efficiency of face recognition system markedly.

Keywords color Gabor; color texture; feature extraction; SVM

1 引言

颜色是彩色图像中的重要信息, 随着多媒体技术和图像采集设备的发展, 颜色信息在图像处理中得到了广泛应用。一张图像中人可以辨识的灰度层次只有几十种, 但可以辨识的颜色达上万种。近几十年来, 人们对彩色信息的研究是通过分别处理每一个颜色分量来达到目的, 未能利用各分量之间的相关性。本文用四元数来对 RGB 颜色模型中的三分量同时进行处理, 提取人脸特征点上的四元彩色 Gabor 特征后送入支持向量机中进行训练与识别。

人脸识别实质是一个稀疏超高维空间的模式识别问题, 解决它的关键之一在于选择对分类有意义的特征。Gabor 滤波器的核类似于哺乳动物简单视觉皮层感受野, 表现出理想的局部性和方向选择性。因此, 虽然目前使用的 Gabor 特征大部分都是基于灰度的, 但已具有了较强的人脸特征表示能力, 被成功应用于许多人脸识别系统中^[1]。提取了有效特征后, 识别的关键就在于设计具有良好分类能力的分类器。支持向量机适合小样本大特征维数的统计学习和分类。一方面, 它应用结构风险最小化原则, 能在经验风险和模型复杂度之间作适当折中, 具有较好的泛化能力; 另一方面, 核函数的利用

基金项目: 福建省自然科学基金项目 (A0510020); 福建省科技计划国际合作重点项目 (20041014)

收稿日期: 2006-08-31; 改回日期: 2006-10-19

第一作者简介: 罗亚兰 (1982~), 女。2007 年于华侨大学获计算机应用专业硕士学位。主要研究方向为图像处理、模式识别等。

E-mail: rolland@hqu.edu.cn

使它具有了很强的非线性映射能力。

人脸识别的许多技术是依据生物视觉系统发展起来的, 人的视觉系统恰恰利用了颜色信息, 但以往对 Gabor小波的研究却局限于灰度图像。Gabor小波主要是边缘信息, 有人认为 90% 的边缘在灰度和彩色图像中是大致相同的, 但还有 10% 的边缘可能在灰度图像中检测不到^[2]。而 Yip 等人指出颜色信息对于人脸识别是有用的, 特别是当图像的其他特征(如分辨率)衰退的时候^[3]。当人脸识别系统性能达到一定的高水平后, 加入信息量更多的特征将是进一步提高识别系统性能的关键。虽然是否将彩色信息应用到人脸识别中仍有待商榷, 但 Jones 的研究表明彩色信息的引入的确能够提高识别效果^[4]。因此本文实现了从灰度 Gabor小波到彩色 Gabor小波的扩展, 在 FERET 彩色人脸库^[5]上的实验结果显示了它的有效性。

2 彩色图像的四元数描述

彩色信息的描述方式是阻碍其用于人脸识别领域的重要原因之一。通常情况下, 一幅彩色图像的各个分量是分别处理的, 这也就导致了人们在处理的时候认为这些分量之间是相互独立的, 然而, 对于实际的图像而言, 不同彩色分量的内容显然是相关的。

1843年哈密顿发明了四元数, 1个四元数包含4个自由度, 即 1个实部和 3个虚部, 因此它需要 4个实数变量来定义:

$$q = q_r + q_i i + q_j j + q_k k \tag{1}$$

式中, i, j, k 为四元数基矢, 运算满足右手系法则, 即

$$\begin{aligned} i^2 = j^2 = k^2 &= -1 \\ ij = -ji = k \\ jk = -kj = i \\ ki = -ik = j \end{aligned} \tag{2}$$

如果四元数中 q_r 为零, 则称其为纯四元数, 通常记为 $\underline{\mu}$ 。四元数常常被用于 3 维或 4 维向量分析, 最近更是被应用在彩色图像分析中。在本文中用四元数来表示 1 个像素, 由于四元数空间中的 3 个虚轴相互正交且并没有什么实质上的不同, 因此在 RGB 彩色空间中, 将一像素的 R、G、B 值赋给一个四元数的 3 个虚部 q_i, q_j 和 q_k (令四元数的实部 q_r 为零)。这样, 用 1 个 3 维向量表示的彩色像素 I 的形式如下:

$$I = I_R i + I_G j + I_B k \tag{3}$$

3 Gabor滤波器

Gabor滤波器的核函数 Gabor函数是通过尺度伸缩和旋转生成的一组复函数系, 含有实部和虚部。Gabor函数具有以下一些非常特殊的性质:

- (1) 具有最佳时间-频率局部化特性;
- (2) 具有旋转相位信息, 反映不同方向的纹理和边界;
- (3) 在心理学上具有接近人类视觉神经的反映特性, 并用 Gabor函数模拟视觉神经模型。

本文采用文献 [6] 中的 Gabor核函数的定义:

$$\phi_{\mu\nu} = \frac{\|k_{\mu\nu}\|^2}{\sigma^2} e^{-\frac{(\|k_{\mu\nu}\|^2 \|z\|^2 / 2\sigma^2)}{2}} [e^{i\alpha_{\mu\nu}} - e^{-(\sigma^2/2)}] \tag{4}$$

式中, μ, ν 分别是方向和尺度因子, $z = (x, y)$ 是行矢量, x, y 为 2 维坐标, $k_{\mu\nu} = \begin{bmatrix} k_\nu \cos\phi_\mu \\ k_\nu \sin\phi_\mu \end{bmatrix}$, $k_\nu = k_{max} / f^\nu$, f 为频域中核函数间的间距因子, $\phi_\mu = \pi\mu/8$ 。

式 (4) 中方括号内第 1 项是交流成分, 第 2 项是直流补偿, 当参数 σ (相对带宽) 很大的时候第 2 项可以忽略。一般取 $\sigma = 2\pi$, $k_{max} = \pi/2$, $f = \sqrt{2}$ 可获得较好的小波表征和辨识效果^[6]。Lee 从生理学的角度指出要用 Gabor离散变换无损地表示图像, 在每一个离散的位置需要 8 个离散的等间距的方向和 5 个等间距的尺度^[7]。通常是取 $\nu = 0, 1, \dots, 4$ 共 5 个不同的尺度, 和 $\mu = 0, 1, \dots, 7$ 共 8 个不同的方向。

3.1 Gabor核函数的四元数形式

Jones 提出了 4 种将 Gabor滤波器扩展到四元数空间的办法。该方法则采用他比较推崇的一种, 即通过把 2D Gabor滤波器的实部与一个单位纯四元数 $\underline{\mu}$ 相乘来得到 Gabor核函数的四元数表示形式^[4], 本文中 $\underline{\mu}$ 代表的是 3 维颜色空间中特征最明显的那一方向:

$$G_{\mu\nu} = \underline{\mu} \frac{\|k_{\mu\nu}\|^2}{\sigma^2} e^{-\frac{(\|k_{\mu\nu}\|^2 \|z\|^2 / 2\sigma^2)}{2}} \times [\cos(z \cdot k_{\mu\nu}) - e^{-(\sigma^2/2)}] \tag{5}$$

方向 $\underline{\mu}$ 的选择问题

在式 (5) 对于四元 Gabor滤波器的定义中包含了一个任意的单位纯四元数 $\underline{\mu}$, $\underline{\mu}$ 是一个 3 维向量, 它指向彩色空间中特征最明显的那一方向。很明显, 本文所选择的 $\underline{\mu}$ 要使人脸图像在其上的投影在一个较大的范围内, 或者说要使人脸图像的投影包

含最有用的信息。

使用下式来获取 $\mu^{[4]}$:

$$\beta = \begin{pmatrix} \bar{r} & \bar{g} & \bar{b} \\ -\bar{g} & \bar{r} & 0 \\ -\bar{r}\bar{b} & -\bar{g}\bar{b} & \bar{r}^2 + \bar{g}^2 \end{pmatrix} \begin{pmatrix} \mathbf{R} \\ \mathbf{G} \\ \mathbf{B} \end{pmatrix} \quad (6)$$

式中, $\mathbf{R}, \mathbf{G}, \mathbf{B}$ 分别代表由人脸图像特征点的红、蓝、绿 3 个颜色分量, 而 $\bar{r}, \bar{g}, \bar{b}$ 表示的是人脸图像特征点的 3 个颜色分量所对应的均值。由于式 (6) 中的 β 仅仅需要分别计算人脸图像特征点的 3 个颜色分量的均值, 非常方便, 故在本文中选用 β 来表示 μ 。

3.2 彩色像素的四元 Gabor 特征提取

假设彩色人脸中像素 I 用式 (3) 来表示, $\mathbf{I}(z)$ 用来表示由这些特征所构成的矩阵。用式 (5) 得到的 Gabor 滤波器对它进行特征抽取, 这个过程实际上就是一个卷积的运算, 即

$$\mathbf{O}_{\mu\nu}(z) = \mathbf{I}(z) * \mathbf{G}_{\mu\nu}(z) \quad (7)$$

式中, $\mathbf{G}_{\mu\nu}(z)$ 表示不同的参数 μ, ν 对应的 Gabor 基, $\mathbf{O}_{\mu\nu}(z)$ 表示通过这组 Gabor 基滤波后得到的四元彩色 Gabor 特征。

由于取 $\mu \in \{0, \dots, 7\}, \nu \in \{0, 1, 2, 3, 4\}$, 所以对于同一幅图像用 40 组 Gabor 基作滤波, 可以得到 40 组特征矢量 $\mathbf{O}_{00}, \dots, \mathbf{O}_{47}$, 将这些特征矢量连接起来就可以得到 Gabor 特征矢量

$$\mathbf{X} = (\mathbf{O}_{00}^T \mathbf{O}_{01}^T \dots \mathbf{O}_{47}^T)^T \quad (8)$$

式中, $\mathbf{X} \in \mathbf{R}^n$ 。

那么如何实现两个四元数之间的卷积? 卷积定理在四元数空间中是否仍然适用? Pei 等人在文献

[7] 中回答了这些问题。按以下步骤来实现两个四元数之间的卷积^[7]:

(1) 将一彩色像素的四元数表示 $I(z) = I_R(z) + I_G(z)j + I_B(z)k$ 分解为

$$I(z) = I_a(z) + I_b(z)j \quad (9)$$

式中, $I_a(z) = I_R(z) + I_G(z)j, I_b(z) = I_G(z) + I_B(z)k$ 。

(2) 而将 Gabor 核函数的四元数表示 $G(z) = G_i(z) + G_j(z)j + G_k(z)k$ 分解为

$$G(z) = G_c(z) + G_d(z)j \quad (10)$$

式中, $G_c(z) = G_i(z) + G_k(z), G_d(z) = G_j(z) - G_l(z)k$ 。

(3) $O(z) = I(z) * G(z)$, 利用 2D 复数傅里叶变换来实现两个四元数之间的卷积, 用 FT 和 IFT 分别表示 2D 复数傅里叶正、逆变换, QFT 和 $IQFT$ 分别表示四元数的傅里叶正、逆变换。其转换公式总结如下:

$$O(z) = IQFT(QFT(I(z))FT(G_c) + QFT(-z)FT(G_d)) \quad (11)$$

4 实验

4.1 彩色 FERET 人脸库简介

实验所用到的图片全部来自 Color FERET Database^[5]。该图库共有 11 338 幅人脸图像, 它们是历时 3 年 (1993~1996) 对 994 个个体从不同角度采集而来的。图库中共有 13 种不同的姿态, 头部侧转从 0° 到 90° 不等, 关于不同姿态图像集的具体信息可以如表 1 所示。该图像库中的人脸图像在年龄、姿态、发型、光照、装扮等方面都有不同程度的变化, 因此在彩色 FERET 图像库^[5]上进行的实验具有一定的通用性。

表 1 不同姿态图像集的特点
Tab 1 Images in different poses

	图像集												
	fa	fb	hl	hr	pl	pr	ql	qr	ra	rb	rc	rd	re
角度	0	0	+ 67.5	- 67.5	+ 90	- 90	+ 22.5	- 22.5	+ 45	+ 15	- 15	- 45	- 75
图像数	1 364	1 358	1 276	1 320	1 312	1 363	761	761	321	321	610	290	290
人脸数	994	993	917	953	960	994	501	501	261	261	423	236	236

注: + 表示头向左侧转, - 表示头向右侧转

4.2 特征点的选择

对于大小为 $M \times N$ 的人脸图像来说, 按照上述介绍的方法所提取的四元彩色 Gabor 特征的维数为 $M \times N \times 4 \times 40$ 。考虑到该图库中的人脸图像大小为 512×768 虽然图库也提供了较低分辨率的图像

(256×484 以及 128×192), 但一幅人脸图像所提取的特征仍然是一个庞大的数字, 并且这些特征中有很多是无用的信息, 因此不对整幅图像提取特征, 而只是针对那些包含信息丰富的特征点进行特征提取。

关于特征点选取问题不是本文研究的重点, 相

关论述可参见文献 [4]、[8]等。这里采用手工定位的方法选取了 25 个特征点。

图 1 中的 3 幅图像展示了对人脸图像的预处理过程, 首先通过旋转使原始图像中的人眼处于同一水平位置; 然后再经过缩放使两眼间的水平距离为 60 个像素, 这样做不仅保留了两眼间距这一重要特征, 同时也对由摄影造成的人脸大小不一的情况做了修正。最后对经过预处理后的人脸图像进行手工标定, 其结果如图 2 所示。



图 1 预处理效果图

Fig 1 The result of preprocessing



图 2 特征点标定图

Fig 2 The image with key points

4 3 实验结果及其分析

本文的主要目的在于研究当人脸识别系统性能达到一定的高水平后, 彩色信息的加入是否有助于识别率的进一步提高。以下用四元彩色 Gabor特征和四元灰度 Gabor特征的对比实验来说明这一点。

实验以正面人脸为研究对象, 训练及测试图像均来自彩色 FERET 的两个正面人脸图集 fa、fb。从中选取 30 人, 每人 5 幅共 150 幅图像作为训练集; 测试集则由这 30 人的其他正面图像 (fa、fb 中除去已在训练集中的图像) 共 45 幅构成, 保证了训练集和测试集无交叉。

在第 3 节中介绍了如何提取特征点的四元彩色

Gabor特征, 为了排除其他因素的干扰, 用同样的方法来提取特征点的四元灰度 Gabor特征, 只不过需要先将彩色图像转化为灰度图像, 转化是用 YUV 颜色空间中的 $Y (= 0.299R + 0.587G + 0.114B)$ 来作为灰度值的。由于要对一幅图像的 25 个点提取它的四元 Gabor特征 (无论是灰度还是彩色), 采用的 Gabor核函数大小为 3×3 则一幅图像的特征向量为 36 000 维 ($9 \times 25 \times 4 \times 5 \times 8$)。同时从表 1 可以看出 fa、fb 图集的特点是类别多, 同一类别的样本量少, 最多的一类在两个图集中一共也才 10 幅, 很明显是属于小样本大特征维数的统计学习和分类。虽然支持向量机适合于解决这类问题, 但从效率上考虑仍然决定对特征向量进行 PCA (principal component analysis) 降维。

采用林智仁提供的 C++ 库 LIBSVM^[9]。该库所使用的关键算法是修改过的序贯最小优化算法和 SVM^{light} 中的工作集选择算法。SVM 本质上是一种二分类器, 为了解决多类别问题, 在 LIBSVM 中采用“一对一”方法。它通过构造 $k(k-1)/2$ 个分类器将 k 个类别两两分开, 然后利用投票机制来决定某一样本类别, 若某一样本在不同的类别上有同样的票数, 为了简便直接将该样本判为序列号最小的那一类别。在本实验中通过比较与分析最终选定以径向基作为核函数来训练分类器。其分类效果如表 2 所示。

表 2 基于彩色 Gabor特征的人脸识别率

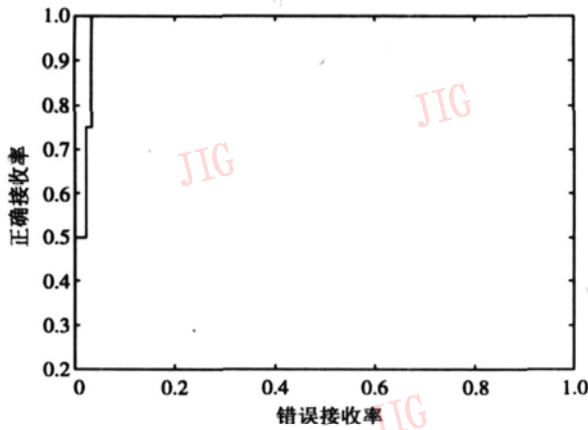
Tab 2 The face recognition rate based on color Gabor features

	彩色 Gabor特征	灰度 Gabor特征
PCA 降维后特征向量维数	30	30
径向基核函数的参数	$c = 8.0 \quad g = 0.25$	$c = 8.0 \quad g = 0.25$
识别率 (%)	97.7778	95.5556

从表 2 的实验数据可以看出, 彩色 Gabor特征的识别率在灰度 Gabor的基础上又提高了 2.222%, 由此说明了当人脸识别系统性能达到一定的高水平后, 彩色纹理信息的加入有助于识别率的进一步提高。

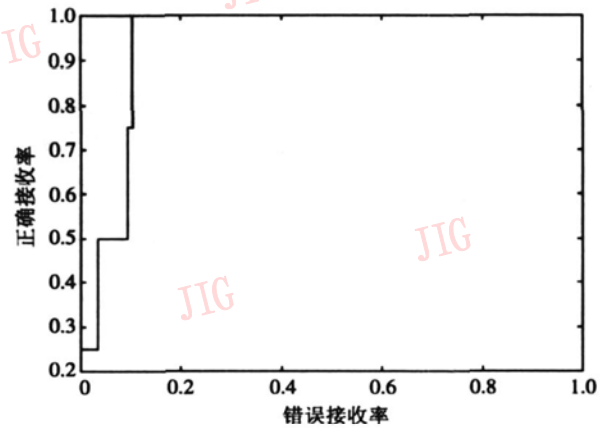
检验 SVM 分类系统的常用方法是进行 ROC (receiver operating characteristic) 曲线的交叉检验^{[10][11]}。虽然上述实验结果表明了该方法的有效性, 但对于多类识别的情况则无法直接得知这些正确接受率是在什么样的错误接受率下得到的, 又鉴于 LIBSVM 是采用“一对一”的方法来处理多类情况的, 因此通过构造一个二分类器来模拟 LIBSVM

的内部实现机制。用 ROC 曲线来评价这个二分类器的性能, 由此及彼, 从而对表 2 中的实验结果有更深入的理解。选取标号为 cfS00705 和 cfS00732 的两个类别来构造二分类器 A, 其中假定 cfS00705 是正例, 其他为反例。经典的 ROC 曲线用横轴表示错误接受率, 纵轴表示错误拒绝率, 在平面上表示的是当分类阈值变化时这两种错误率之间的消长关系。在这里借鉴人脸检测将纵轴改为正确接受率^[12],



(a) 彩色 Gabor 特征 ($AUC = 0.9856$)

ROC 曲线与横轴间所围成的面积大小表示分类系统在所有代价比上的平均性能优劣。面积越大, 表示系统的预测性能越好, 本质上反映两类样本的可分性越大。从图 3 的数据可以看出在所构造的二分类器 A 上, 彩色 Gabor 特征比灰度 Gabor 特征具有更好的 SVM 分类平均性能, 它们的 ROC 面积 (AUC) 分别为 0.9856 和 0.9425, 并且在误检率方面彩色 Gabor 特征也具有明显优势。



(b) 灰度 Gabor 特征 ($AUC = 0.9425$)

图 3 分类器 A 的 ROC 曲线

Fig. 3 The ROC curve of classifier A

5 结 论

通过四元数概念的引入, 使得彩色信息可以自然地应用到人脸识别中来。将四元数表示 3 维数据的优势同 Gabor 滤波器的局部性和方向选择性, SVM 在处理小样本问题时泛化能力强的特点以及 PCA 的降维原理结合起来, 提出了一种基于四元彩色 Gabor 特征的人脸识别方法。同灰度 Gabor 特征相比, 彩色信息的引入更进一步地提高了识别率, 在彩色 FERET 图像库上的实验验证了这一设想。目前, 有待进一步深入研究的问题是如何去除戴眼镜与眼睛的开闭对识别率的影响, 使所提取的特征具有更好的稳健性和更高的识别率; 以及是否可实现特征点的准确自动提取和定位, 以便使得这种人脸识别方法的实现得以全自动化。

参考文献 (References)

- 1 Lee T S. Image representation using 2D Gabor wavelets [J]. IEEE Transactions on PAMI, 1996, 18(10): 959~971.
- 2 Koschan A. A comparative study on color edge detection [A]. In Proceedings of the 2nd Asian Conference on Computer Vision ACCV'

- 95 [C], Singapore, 1995, 3: 574~578.
- 3 Yip A W, Sinha P. Contribution of color to face recognition [J]. Perception, 2002, 31(8): 995~1003.
- 4 Jones C F. Color Face Recognition Using Quaternionic Gabor Filters [D]. Blacksburg Virginia State, USA: Virginia Polytechnic Institute and State University, 2004.
- 5 Color FERET Database [DB]. <http://www.itl.nist.gov/iad/humanid/colorferet/>.
- 6 Lades M, Vorbruggen J C, Buhmann J *et al*. Distortion invariant object recognition in the dynamic link architecture [J]. IEEE Transactions on Computers, 1993, 42(3): 300~311.
- 7 Pei S C, Ding J J, Chang J H. Efficient implementation of quaternion fourier transform, convolution, and correlation by 2D complex FFT [J]. IEEE Transactions on Signal Processing, 2001, 49(11): 2783~2797.
- 8 Arca S, Campadelli P, Lanzarotti R. A face recognition system based on automatically determined facial fiducial points [J]. Pattern Recognition, 2006, 39(3): 432~443.
- 9 LibSVM [CP]. <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.
- 10 Chen Duan-sheng, Liu Zheng-kai. Edge features in color image and their face detection performance evaluation [J]. Journal of Software, 2005, 16(5): 727~732 [陈锻生, 刘政凯. 彩色图像边缘特征及其人脸识别性能评价 [J]. 软件学报, 2005, 16(5): 727~732].
- 11 Fawcett T. An introduction to ROC analysis [J]. Pattern Recognition Letters, 2006, 27(8): 861~874.
- 12 Yang M H. Recent Advances in Face Detection [EB/OL]. http://vision.ai.uiuc.edu/mhyang/papers/icpr04_tutorial.pdf

Panda型保偏光纤偏振轴图像识别算法

游建民 陶林

(西南交通大学峨眉校区机械工程系, 峨眉山 614202)

摘要 两 Panda型保偏光纤自动熔接时, 需要对准偏振轴, 为了高精度地探测保偏光纤的偏振轴, 针对 Panda型光纤横截面, 提出了一种偏振轴的计算机图像精确识别算法, 同时利用 LabVIEW 编程环境, 设计及开发了实现该算法的计算机程序, 并应用该程序对 Panda型保偏光纤偏振轴进行了实际图像的分析处理, 以验证该算法的有效性。通过实际运行结果的分析表明, 应用计算机图像识别和现代测控技术探测 Panda型保偏光纤的偏振轴, 可以实现保偏光纤跳线连接器自动化制造, 并可极大地提高光轴的识别精度和光纤连接器的生产效率。

关键词 Panda型保偏光纤 计算机图像识别 LabVIEW 光纤跳线连接器 边缘检测算法

中图分类号: TP391.41 **文献标识码:** A **文章编号:** 1006-8961(2008)02-0247-04

Image Identification Algorithm on Polarization Axis of Panda Polarization Maintaining Fiber

YOU Jian-min, TAO Lin

(Department of Mechanical Engineering of Emei Campus, Southwest Jiaotong University, Emeishan 614202)

Abstract Polarization axis should match when two Panda polarization maintaining fibers connect in order to getting high accuracy identification of the polarization axis. An accurate computer image identification algorithm was presented to identify the polarization axis of the Panda polarization maintaining fiber. To validate the algorithm, program was developed to perform image analysis and processing with LabVIEW software. The practical data of application indicates that they are very useful to improve the identification precision of polarization maintaining fiber and raise productivity of fiber connector with the technologies of computer-aided vision detection and modern measurement & control.

Keywords Panda polarization maintaining fiber, computer image identification, LabVIEW, fiber jumper connector, edge detection algorithm

1 引言

随着科学技术的迅猛发展, 光导纤维(简称光纤)已在通信、电子、测试传感和电力等领域得到日益广泛的应用, 成为前途广大的新型基础材料, 反过来又推动着光纤技术的深入研究。超宽带光纤通讯的发展, 导致了光信号传输过程中对偏振保持的较高要求, 这又极大地促使了保偏光纤技术的研究和

发展。由于保偏光纤在连接的时候需要对正偏振轴, 因而产生了对保偏光纤跳线端头连接器的特殊要求, 这也使得其制造工艺要比普通光纤跳线连接器的制造更复杂。

针对这一问题, 通过对目前应用较广的 Panda型(即熊猫型)保偏光纤跳线端头连接器制作工艺的分析, 提出了利用计算机图像识别技术进行熊猫型保偏光纤偏振轴识别的算法。应用结果表明, 运用计算机图像识别技术识别保偏光纤光轴是一种非

收稿日期: 2006-09-30 改回日期: 2006-10-13

第一作者简介: 游建民(1965~), 男, 美国马里兰大学博士, 西南交通大学机械工程系教授。主要从事结构胶粘剂材料机械性能的研究、自动化系统的开发, 以及计算机图像识别的研究。E-mail: yjmyq@yahoo.com

常可行的方法。通过适当地改进,该方法可推广应用到其他保偏光纤光轴的识别应用中,也可以应用到其他领域的类似应用中。

2 背景

保偏光纤是偏振保持光纤 (polarization maintaining fiber, PMF) 的简称,有时也称之为固定偏振光纤。它是一种具有双折射主轴(快/慢轴)的特制光纤,其可以维持传输信号的偏振态。保偏光纤主要类型有:熊猫型、椭圆包层型、领结型、“一”字型和椭圆芯型等。该类光纤已广泛应用于高速光通信系统中,还应用到其他诸多领域,如光纤激光器、光纤放大器、光纤传感器中,以及光纤陀螺和光纤加速度计等^[1]。由于通常熊猫型保偏光纤的偏振保持特性比一般的领结型和椭圆包层型结构光纤更好、应用较广,因此本文选用熊猫型保偏光纤作为研究对象。图 1 是通过 CCD 照相机获得的 Panda 型保偏光纤截面的放大图像。它主要具有以下结构:光纤截面外轮廓、两应力棒圆和一个纤芯圆。通过两应力棒圆圆心的直线为光纤的慢轴(如图 2 所示)。

当使用跳线连接器连接两根保偏光纤时,光轴对接的细小偏差都会大大降低消光系数,比如出现

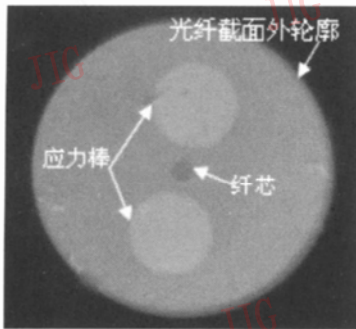


图 1 熊猫型保偏光纤截面图像

Fig 1 Cross section image of Panda polarization maintaining fiber

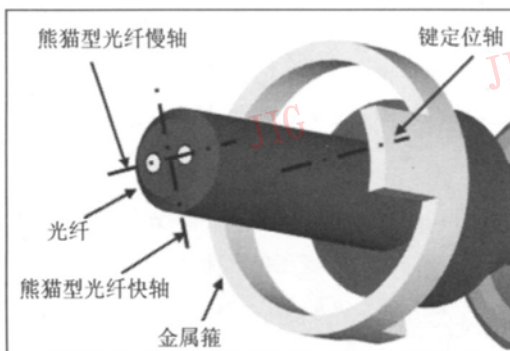


图 2 偏振光轴和连接器的定位键对准示意图

Fig 2 Diagram of alignment of polarization maintaining fiber polarization axis with connector orientation key

2°的角偏差时,消光系数就可以从 30dB 减少到 25dB^[2]。要保证保偏光纤线路机械连接时光轴对接的精度,其跳线连接器定位键与光轴的对准精度必须非常高。由此可见,用连接器的定位键来对准保偏光纤的偏振轴是首要解决的问题。鉴于当前还没有为保偏光纤而特制的光纤连接器,而且连接器套管不能通过调整其定位键来实现保偏光纤双折射轴的对准,为此,这里先使用 FC 型连接器来制作熊猫型保偏光纤连接器,然后通过旋转套管中的保偏光纤来实现保偏光纤的偏振光轴和连接器的定位键对准,其原理如图 2 所示。于是笔者开发了一种基于计算机图像识别的自动化闭环反馈控制系统,并利用美国国家仪器公司 LabVIEW 图形化编程软件^[3]开发了熊猫型保偏光纤光轴识别和其连接器自动化制造系统的分析测控程序,以实现偏振轴识别和与定位键对准的操作。

3 光纤端面图像识别

图像是一个 2 维亮度函数 $I(x, y)$: (x, y) 为像素点的空间坐标, $I(x, y)$ 定义了该点的亮度。数字图像是指图像 $I(x, y)$ 在空间坐标和亮度的数字化。一个数字图像可以看成是一个矩阵或一个 2 维数组,其基本元素称为像素。数字灰度图像由有限个像素组成。其中, 8bits 数字灰度图像是应用最早和目前使用较多的一种。它的亮度由 0 到 255 共 256 级,其中 0 为黑色, 255 为白色。

借助于显微镜,使用 CCD 照相机可获得熊猫型光纤截面的数字灰度图像,它可表示为

$$I(x, y) = (0 \sim 255) \quad (1)$$

这里, 0 为黑色, 255 为白色, $x = 0 \sim (M - 1)$, $y = 0 \sim (N - 1)$ 。

3.1 光纤截面的识别

得到光纤截面的数字灰度图像后,紧接着的任务是获得光纤截面外圆的圆心和半径。可通过“二值化”的方法将灰度图像分割成目标和背景两个区域。这里将目标区域设定为 1 背景区域设定为 0 即如果 $I(x, y) > \theta$ 则设定 $I(x, y) = 0$ 否则 $I(x, y) = 1$ 。其中, θ 为“二值化”的阈值, $I(x, y)$ 为像素坐标 (x, y) 处的灰度值。“二值化”后的图像用 $J(x, y)$ 来表示,光纤截面圆心坐标 (x_0, y_0) 及其半径 R 的计算如下:

$$x_0 = \frac{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} x \cdot J(x, y)}{A} \quad (2)$$

$$y_0 = \frac{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} y \cdot J(x, y)}{A} \quad (3)$$

$$R = \sqrt{\frac{A}{\pi}} \quad (4)$$

其中, 光纤截面面积 $A = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} J(x, y)$ 。图 3 给出了在 LabVIEW 中通过该算法得到的 Panda 型保偏光纤截面外圆周线。

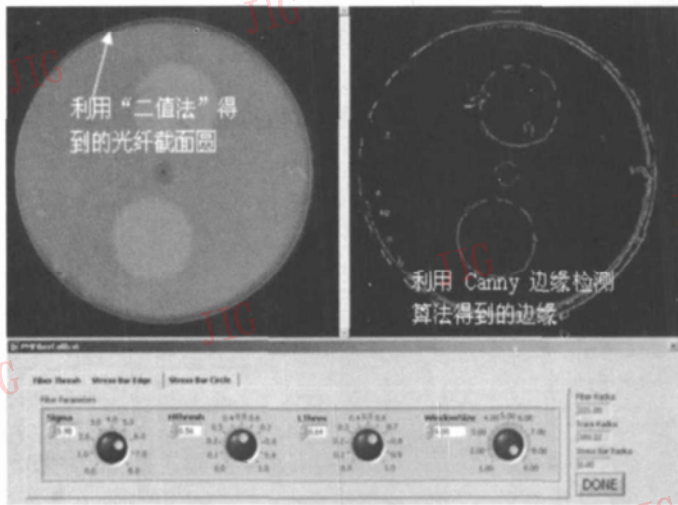


图 3 截面边界识别的软件用户界面

Fig 3 Software GUI of identification of cross section boundary

3 2 光纤应力圆棒边缘识别

计算机图像边缘检测算法有多种, 其中, Canny 提出的算法是首先对原图像进行平滑, 再进行边缘检测就能成功地检测到真正的边缘^[4], Canny 边缘检测算法是把边缘检测问题转换为检测函数极大值的问题^[4 5]。在高斯噪声中, 一个典型的边缘代表一个阶跃的强度变化。由于 Canny 边缘检测法具有定位准确、单边响应和信噪比高等优点^[6 7], 因此本文采用 Canny 边缘检测法对光纤应力圆棒边缘进行探测。本文检测法的基本步骤如下:

首先使用式 (5) 对原图像 $I_0(x, y)$ 进行平滑处理, 即

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (5)$$

其中, $G(x, y)$ 表示高斯函数滤波器。

然后, 利用两个滤波器分别沿横向和纵向求高斯函数的一阶导数, 再对平滑后的图像作卷积运算,

最后根据图像的梯度进行边缘检测。卷积运算式为

$$I(x, y) = \sum_{k=0}^{m-1} \sum_{l=0}^{n-1} I_0(x+k-1, y+l-1) \cdot g(k, l) \quad (6)$$

其中, $I(x, y)$ 表示过滤得到的图像, $I_0(x, y)$ 表示原图像, $g(k, l)$ 表示卷积核, $(m-1)$ 或 $(n-1)$ 表示卷积核的大小。本文采用 LabVIEW 编程软件实现了该算法, 图 3 给出了采用 Canny 边缘算子检测得到的 Panda 型保偏光纤截面特征边缘。

3 3 应力棒圆心的计算

通过以上的分析计算, 就可以得到光纤圆心和半径, 同时也得到了应力棒的边缘。为了抑制图像中噪声的影响, 本文提出通过沿以光纤轴心为圆心和以 $1/2$ 光纤半径为半径的圆的轨迹进行扫描来探测和计算应力棒圆, 如扫描角度可定为 1° , 这样就可以求得 360 个轨迹点, 再以这些点为圆心向外辐射多根直线, 直线与 Canny 边缘算法识别到的边缘产生交点, 然后再对这些交点进行圆拟合。根据熊猫型保偏光纤截面图像的特点, 一方面圆拟合的结果将会有两个圆出现的频率最高, 这两圆就是所要求的应力棒圆; 另一方面, 也可利用程序在显示屏幕上实时地绘制出两应力棒圆、光纤轮廓圆和慢轴, 这样操作者就可以方便地对结果进行观察及调整。最后, 通过两应力棒圆圆心来计算保偏光纤慢轴和垂直线的角度 (如图 4 所示)。

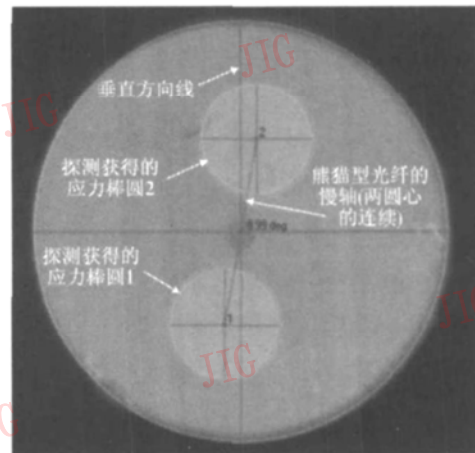


图 4 慢轴识别界面

Fig 4 Interface of slow axis identification

4 应用测试及结果

为检验该算法的有效性, 本文对一个 Panda 型保偏光纤进行了 120 次图像数据采样和慢轴角度计

算(相对于垂直线)。通过分析发现,该算法的标准偏差可以达到 0.087° ——利用六西格玛求自然极限可获得 0.26° ($6\sigma/2=0.522/2\approx 0.26$)的精度值,图 5 给出了这些数据的分布情况及分析结果。与其他的方法相比,戴维斯所介绍的方法^[8]达到的精度为 2° ,而当前较常用的方法是先利用平行光侧向照射到裸保偏光纤上(光线受应力棒及纤芯的影响会发生不同程度的折射和散射现象),然后根据光纤影像的光强分布进行对轴,其定轴精度可达到 1.67° ^[9],这优于文献[8]算法的 2° 的定轴精度。

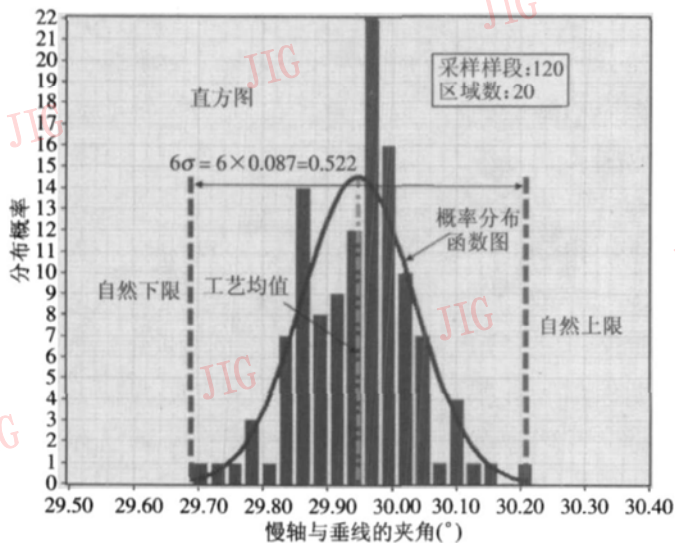


图 5 柱状图及概率正态分布函数

Fig 5 Histogram and normal probability distribution function

5 结 论

综上所述,可得以下结论:

(1) 本检测算法能有效地抑制光纤图像中噪声的影响,使探测准确有效;

(2) 应用计算机图像识别技术识别 Panda型保偏光纤的偏振轴是非常有效的,并可用于对轴精度小于 0.5° 的实际检测;

(3) 本文所提出的算法有效可行,并且很容易用 LabVIEW 编程软件实现;

(4) 该算法和软件可以推广到其他类型保偏光纤偏振轴的识别应用中和其他领域的类似应用中。

参考文献 (References)

- 1 Noda J Okamoto K, Sasaki Y. Polarization maintaining fibers and their applications[J]. Lightwave Technology, 1986, 4(8): 1071~1089
- 2 Derickson D. Fiber Optic Test and Measurement[M]. Englewood Cliffs, New Jersey, USA: Prentice Hall PTR, 1998: 642
- 3 Yang Le-ping. Program Design and Application with LabVIEW (The Second Edition) [M]. Beijing: Electronic Industry Press, 2005 [杨乐平. LabVIEW 程序设计与应用(第2版)[M]. 北京:电子工业出版社, 2005]
- 4 Canny J. A computational approach to edge detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986, 8(6): 679~698
- 5 Lin Hui, Zhao Chang-sheng, Shu Ning. Edge detection and evaluation based on Canny arithmetic operators [J]. Journal of Heilongjiang Engineering College, 2003, 17(2): 3~6 [林卉, 赵长胜, 舒宁. 基于 Canny 算子的边缘检测及评价 [J]. 黑龙江工程学院学报, 2003, 17(2): 3~6]
- 6 Gonzalez R, Woods R. Digital Image Processing (Second Edition) [M]. New York: Addison-Wesley Publishing Company, 2002, Chap 4
- 7 Haralick R, Shapiro L. Computer and Robot Vision [M]. New York: Addison-Wesley Publishing Company, 1992, Chap 7
- 8 Davies E. Machine Vision Third Edition: theory, algorithms, practicalities [M]. San Francisco, CA, USA: Academic Press, 2004
- 9 Zhen Zong-min, Yang Yuan-hong. On-line adjusting axes technology of polarization maintaining fiber [J]. Acta Photonica Sinica, 2004, 33(9): 1060~1062 [甄宗民, 杨远洪. 保偏光纤在线定轴技术研究 [J]. 光子学报, 2004, 33(9): 1060~1062]

航空图像中水面纹理的自动提取

汪海洋¹⁾ 潘德炉²⁾ 夏德深¹⁾

¹⁾ (南京理工大学计算机系, 南京 210094) ²⁾ (国家海洋局第二海洋研究所, 杭州 310012)

摘要 针对可见光航空遥感监测中, 耀斑和云阴影等强噪音的干扰使水中目标很难直接发现这一问题, 提出了一种基于 Gabor 滤波器和 BP 神经网络的尾迹纹理自动提取算法, 通过提取它们运动产生的尾迹实现对它们的准确识别。该方法分为两步: 第 1 步是选取等大小的含尾迹纹理的水面子图像和不含尾迹纹理的水面子图像, 通过一组 Gabor 滤波器得到它们的特征图像, 计算每个子图像特征图的均值和方差, 将它们作为神经网络的训练样本对 BP 网络进行训练得到用于识别的网络; 第 2 步是将待提取的整幅图像分成很多与第 1 步中子图像等大小的子图像, 分别计算它们的 Gabor 特征图像, 并得到它们的均值和方差, 把它们作为神经网络的输入, 得到它们是否是纹理区域, 由整幅子图像的分类结果得到一幅二值图像, 用 Hough 变换检测图像中的直线, 根据直线的长度判断尾迹是否存在。大量的实验结果表明, 该方法能够准确地提取运动目标产生的尾迹纹理。

关键词 Gabor 滤波器 BP 神经网络 Hough 变换 纹理提取

中图分类号: TP391.41 文献标识码: A 文章编号: 1006-8961(2008)02-0251-06

Water Texture Fetching of Airphotos

WANG Haiyang¹⁾, PAN De-lu²⁾, XIA De-shen¹⁾

¹⁾ (Nanjing University of Science and Technology, Nanjing 210094)

²⁾ (Second Institute of Oceanography of State Oceanic Administration, Hangzhou 310012)

Abstract To solve the problem that it is difficult to directly detect the object in the water due to flares and cloud shadows, this paper proposed a water wake recognition method based on Multi-Channel Gabor filters and BP neural network. First we select sample sub-images of same sizes with wake texture and without wake texture, then we obtain feature images using a group of Gabor filters and calculate the mean and variance of feature images to acquire the input vectors and train the BP network. Secondly we divide the whole image into sub-images with the same size as the first step, calculate mean and variance of Gabor feature images, calculate the input vector and judge whether the sub-image contain a wake texture by the trained BP network in the first step. We obtain a binary image by the classify results of the whole image, detect lines using Hough transform and judge whether there is a wake in the whole image. From experiment results, it is proved that the proposed algorithm can attain the wake texture precisely.

Keywords Gabor filters, BP neural network, Hough transform, texture fetching

1 引言

我国有一半以上的海域存有权益之争, 周边国家在有争议海域大肆掠夺资源, 每年的经济损失巨

大。传统的通过调查船对海域进行监测, 不仅耗时耗力而且由于我国海域很大, 很多时候无法起到有效的作用。同时由于外国的军舰, 船只等在我海域附近活动频繁, 对我国构成了很大威胁。通过遥感进行监测具有快速、全天候、实时等优点, 为维护我

基金项目: 国家高技术研究发展计划 863 项目 (2003AA131160)

收稿日期: 2006-03-20 改回日期: 2006-08-29

第一作者简介: 汪海洋 (1982~), 男, 南京理工大学模式识别与智能系统专业博士研究生。主要从事图像处理、计算机视觉和模式识别等。Email: why_haiyang@126.com

海洋权益和国家安全提供了重要保障。航空可见光遥感是遥感监测中一种非常重要的手段。但是由于耀斑,云阴影等因素的干扰,仅通过图像灰度很多时候很难直接得到有用的信息。为了解决这个问题,通过大量的实验和研究,提出了一种基于 Gabor 滤波器和 BP 神经网络的尾迹纹理自动提取算法,通过提取它们运动产生的尾迹实现对它们的准确识别,为可见光航空遥感自动监测提供了技术保障。

几十年来,人们提出了许多纹理分析的方法^[1-9],总的说来可以分为以下几类:结构模型的方法、统计模型的方法、滤波器模型的方法、随机场的方法和分形模型的方法等等。在所有的的方法中,多通道 Gabor 滤波器的方法被认为是最成功的纹理分析方法,因为生物学实验表明^[5],Gabor 滤波器和人类视觉系统的纹理认知机制相一致。Gabor 滤波器能够很好地表达和提取纹理特征,在过去的 30 年中,很多学者对它进行了研究^[14-16]。本文提出的算法就是在这些基础上提出的。

2 多通道 Gabor 滤波器

Gabor 函数是由 Gabor 首先定义的^[10],并由 Daugman 将它扩展成 2 维^[11]。2 维 Gabor 函数是一个纵横比为 λ 尺度参数为 δ 主轴与 x 轴角度为 φ 的高斯函数调制的复数正弦函数。它的一般形式为

$$h(x, y) = g(x', y') \exp[2\pi j(Ux + Vy)] \quad (1)$$

式中, (U, V) 为中心频率, $(x', y') = (x \cos\varphi + y \sin\varphi, -x \sin\varphi + y \cos\varphi)$ 为旋转坐标,而 2 维高斯函数为

$$g(x, y) = \frac{1}{2\pi\lambda\sigma^2} \exp\left\{-\frac{(x/\lambda)^2 + y^2}{2\delta^2}\right\} \quad (2)$$

式中, λ 为纵横比。对式 (1) 作傅里叶变换,可得 2 维 Gabor 函数的频域表达式:

$$H(u, v) = \exp\{-2\pi^2\delta^2[(u' - U')^2\lambda^2 + (v' - V')^2]\} \quad (3)$$

式中, $(u', v') = (u \cos\varphi + v \sin\varphi, -u \sin\varphi + v \cos\varphi)$; (U', V') 是中心频率 (U, V) 所作的相应旋转。 $H(u, v)$ 是一个主轴与 u 轴成 φ 角的带通高斯函数,其纵横比为 $1/\lambda$ (本研究中取 1),极向中心频率 $F = \sqrt{U^2 + V^2}$ (单位为 C/image CPI),方向为 $\theta = \arctg(V/U)$ 。2 维 Gabor 函数是一个位于空间域和频率域的带通滤波器函数,一般情况下,取 Gabor 函数中高斯函数部分的方向和复数调制部分函数的辐角相等 ($\varphi = \theta$)。此时 Gabor 函数的空域和频域的为

$$h(x, y) = g(x', y') \exp[2\pi jFx'] \quad (4)$$

$$H(u, v) = \exp\{-2\pi^2\delta^2[(u' - F)^2\lambda^2 + v'^2]\} \quad (5)$$

滤波器的带宽为 $B = \log_2[(\pi F\lambda\sigma + \alpha)/(\pi F\lambda\sigma - \alpha)]$, $\alpha = \sqrt{(\ln 2)/2}$ 通常取 0.6~2.0 个倍频程,本研究中取 1。

特征提取同时抽取频率信息和方向信息;文献 [5] 指出就纹理特征提取的目的,没有必要选择整个频域的滤波器参数空间。可以看出,只要选定了 F, σ, θ 3 个参数,即可确定 Gabor 滤波器。设图像大小为 $N \times N, m = \log_2 \frac{N}{4}$, 则 F 的取值依次为 $\sqrt{2} \{1, 2, 3, \dots, 2^m\}$, 这里取 4 个最高频率;取定了 F 后,根据 $B = 1, \delta$ 的值就可依次确定了; θ 的取值为 $0^\circ, 45^\circ, 90^\circ, 135^\circ$ 。确定了滤波器各个参数后,在计算各滤波器与图像卷积的能量值的基础上,计算各个子图像滤波器能量值的均值与方差作为子图像的纹理特征,这里共选取了 16 个滤波通道,共 32 个滤波特征,记为 $F = \{\mu_{00}, \sigma_{00}, \dots, \mu_{ij}, \sigma_{ij}, \dots, \mu_{33}, \sigma_{33}\}$, 设 $E_{ij}(x, y)$ 为第 (i, j) 个滤波能量值,则子图像 (大小为 $N \times N$ 像素) 的能量均值和方差为

$$\mu_{ij} = \frac{\sum_{y=0}^{N-1} \sum_{x=0}^{N-1} E_{ij}(x, y)}{N \times N} \quad (6)$$

$$\sigma_{ij} = \frac{\sum_{y=0}^{N-1} \sum_{x=0}^{N-1} \sqrt{(E_{ij}(x, y) - \mu_{ij})^2}}{N \times N} \quad (7)$$

3 BP 神经网络

Rumelhart 等人在 1986 年提出了前向多层网络的反向传播学习算法 (BP 算法)^[12] 和前馈网路模型 (BP 模型)。BP 神经网络一般由输入层、隐含层和输出层组成。它采用反向传播算法 (BP 算法) 为学习算法,因为 BP 算法功能强大,易于实现,易于理解,并且其他学习算法都可看作是对它的一种变形和修改。BP 神经网络的性能与其输入、层数、每层中的神经元数、传递函数、学习算法、初始权值及期望误差都有着密切的联系^[12-13]。图 1 是一个典型的 3 层 BP 网络结构模型,这里使用的就是这种结构模型。

(1) 学习样本数据的归一化 训练集各个样本数据的数值范围直接关系到网络学习的结果,由于航空图像的成像受外界光线的影响,当图像处于阴影之中时数值的变化范围很小,直接用 Gabor 滤波

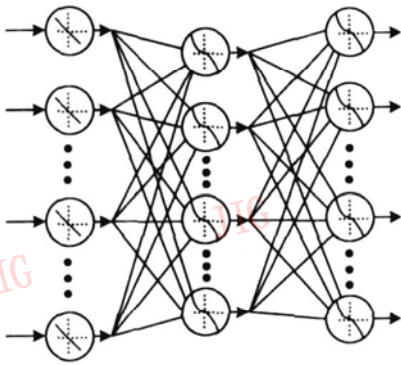


图 1 BP网络

Fig 1 BP network

器组提取它们的纹理特征并不能有效地对网络权值进行修正,所以在对图像进行纹理特征提取前必须进行图像均衡化处理。同时为了让纹理提取算法具有旋转不变性,每个样本将按照 4 个方向 (0° , 45° , 90° , 135°) 旋得到 4 个样本。

(2) 网络层数的选取 理论上已经证明: 具有偏差和至少一个 S 型 (Sigmoid 函数) 隐含层加上一个线性输入层的网络, 能够逼近任何有理函数。增加层数可以进一步地降低误差, 提高精度, 但同时也使网络复杂化。另外不能用仅具有非线性激活函数的单层网络来解决问题。因为能用单层网络完美解决的问题, 用自适应线性网络也一定能解决, 而且自适应线性网络的运算速度还要快。而对于只能用非线性函数解决的问题, 单层精度又不够高, 也只有增加层数才能达到期望的结果。根据前人的研究成果和实际应用经验, BP 神经网络在模式识别/分类中具有良好的性能, 而且一个具有无限隐含层节点的 3 层 BP 网络可以实现任意从输入到输出的非线性映射, 所以这里采用 3 层 BP 网络。

(3) 隐含层单元数 尽管输入和输出单元数分别由输入向量的维数和类别数决定, 但是隐单元个数并不简单地与此类问题的外在特性相关。隐单元个数决定了网络的表达能力, 从而决定了边界的复杂度。如果模式较易分开或线性可分, 那么仅需要较少的隐单元; 相反, 如果从具有较高分散性的复杂概率密度中抽取的, 则需要更多的隐单元。过多的隐单元可以精细地调节到特定的训练集上, 但对测试样本的误差率会高得令人无法接受。过少的隐单元将不具备足够的自由度以较好地拟和训练样本, 测试误差率依然很高。

一个基本的方法是根据训练数据相应地调节网络的复杂程度。比如, 先从一个较大的隐单元个数开

始, 然后衰减权值。由试验可知, 输入层节点数为 32 隐含层节点数为 32 输出层节点数为 2 时, 可以取得很好的效果。

(4) 传递函数及学习算法的选取 隐含层和输出层的传递函数, 均选用非线性 S 型函数, 以完成输入模式到输出模式的非线性映射。标准的 BP 学习算法通常具有收敛速度慢, 易陷入局部极小等缺点, 因此出现了许多改进的算法。快速 BP 算法从改进途径上可分为两大类: 一类是启发式学习方法, 如引入动量因子的学习算法、变学习速率学习算法和“弹性”学习算法等; 另一类则是更有效的数值优化方法, 如共轭梯度学习算法, quasiNewton 算法等。

采用引入动量因子的学习算法来加速网络的收敛速度, 试验结果表明, 该方法取得了很好的结果。

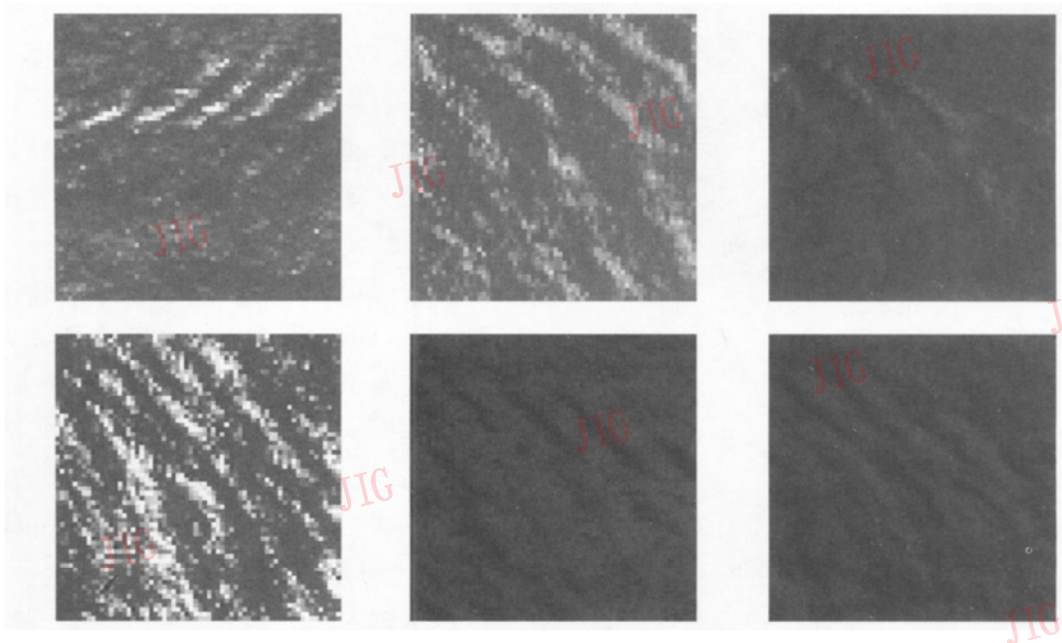
(5) 初始权值选取和期望误差选取 初始权值为 $(-0.5, 0.5)$ 之间的随机数, 误差为 0.01。

4 算法的流程和结果

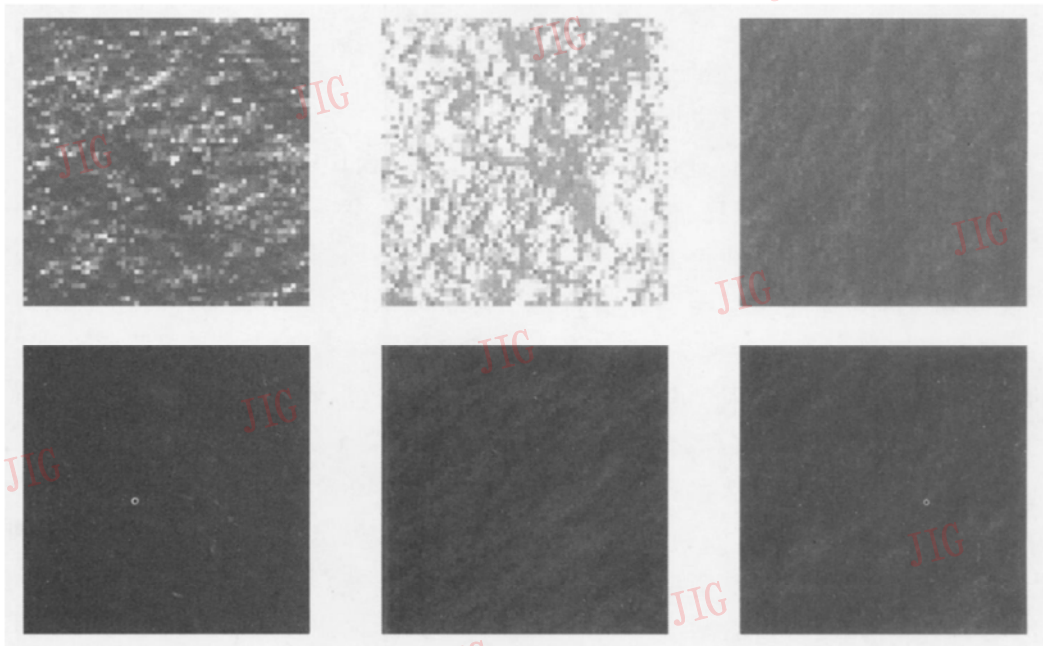
(1) 训练网络 选取训练样本 (大小为 64×64 像素的子图像) 60 个, 含有尾迹纹理图像和不含尾迹纹理图像各 30 个, 并按照 4 个方向 (0° , 45° , 90° , 135°) 进行旋转得到 240 个样本图像。图 2 是两种样本的 12 个例子。

对 240 个样本图像进行直方图均衡化处理, 用 Gabor 滤波器组卷积得到特征能量图并计算它们的均值和方差, 得到神经网络的输入向量。当含有尾迹纹理时输出向量为 $[1, 0]$; 当不含尾迹纹理时输出向量为 $[0, 1]$ 。训练图 1 所示的神经网络, 得到用于识别尾迹纹理子图像的网络。为了验证生成的神经网络性能, 选取测试样本 80 个, 其中含有尾迹纹理图像和不含尾迹纹理图像各 40 幅, 并按照 4 个方向进行旋转得到 320 个测试样本图像。表 1 给出了网络对训练样本和测试样本的识别率。

(2) 将待处理的整幅图像按照图 3 所示的 4 种方式分别分割为 64×64 的子图像, 并同时生成一幅与整幅图像大小相同的结果图像并将它赋值为 255 (即背景色), 对每个子图像直方图均衡化并用多通道 Gabor 滤波器组求得特征能量图并计算它们的均值和方差, 得到神经网络的输入向量。将神经网络的输入向量输入步骤 1 中训练好的网络, 如果输出向量的第 1 个分量大于第 2 个分量, 将结果图像中



(a) 含尾迹的样本子图像



(b) 不含尾迹的样本子图像

图 2 样本图

Fig 2 Sample images

表 1 BP神经网络识别率

Tab 1 Recognition percentage of BPNN

	训练样本	测试样本
正确识别率 (%)	96.7	87.5

对应的点赋值为 0。

(3) 用 Hough 提取结果图像中的线段, 计算线段的长度, 如果大于阈值 T (这里 T 取 200), 认为尾迹存在, 否则认为是噪音点。

为了说明整个纹理提取算法的有效性, 对 100

幅图像进行纹理提取, 其中含有尾迹 80 幅、不含尾迹 20 幅。图 4 给出了算法实验结果的 4 个例子, 其中, (a)、(c)、(e)、(g) 为原始图像; (b)、(d)、(f)、(h) 为提取的纹理区域图像 (黑色表示尾迹区域)。图像的大小为 448×448 像素, 子图像的大小为 64×64 像素。从图 (b)、(d)、(h) 可以看出, 包含尾迹纹理的区域基本都能够正确识别出来, 即使被错分为背景区域, 通过它周围正确识别为尾迹区域的子图像, 利用算法的第 3 步仍然能够正确地提取尾迹。同理如图 4(f) 即使有个别背景子图像错分为

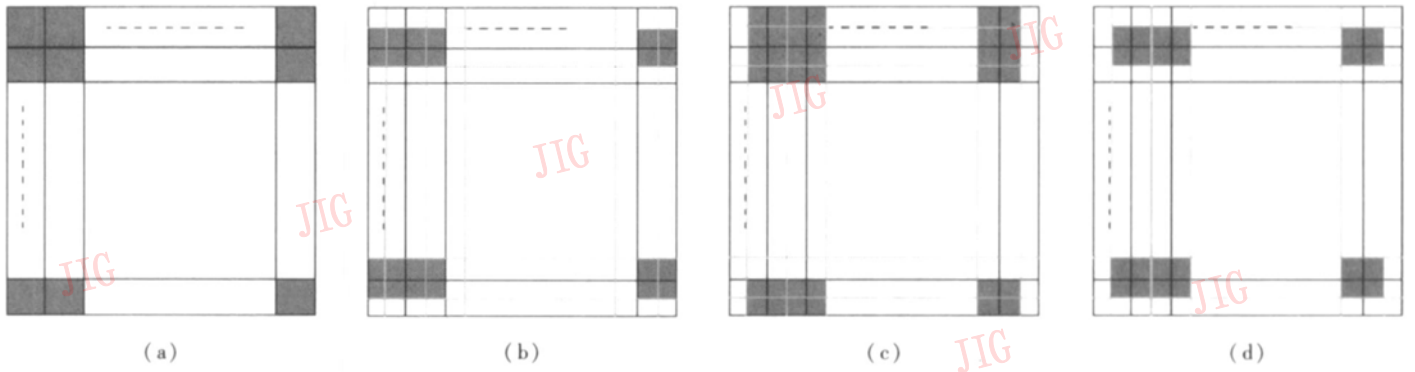


图 3 划分示意图
Fig. 3 Partition map

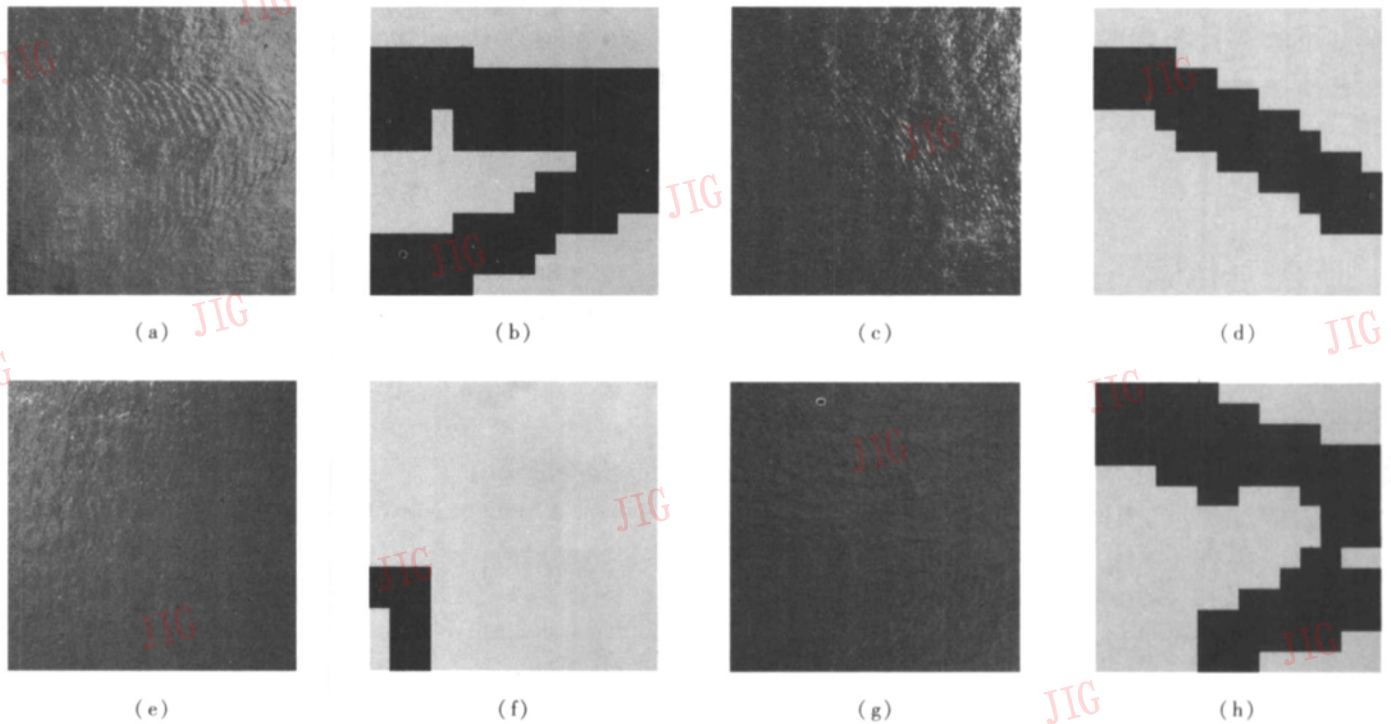


图 4 试验图
Fig. 4 Experimental pictures

尾迹区域, 通过算法的第 3 步也能够将它排除。尾迹的正确检测率为 95%、误检率为 0%: 80 幅含有尾迹的图像有 76 幅能够被准确地提取出来, 20 幅不含尾迹的图像全部识别为不含尾迹, 4 幅没有被检测尾迹图像是由于尾迹纹理信息太弱。实验结果表明, 水面尾迹纹理自动提取算法能够准确地获取尾迹区域, 具有很强的实用性。

5 与同类纹理分类方法的性能比较

为了说明本文方法对尾迹纹理分类的有效性, 分别用灰度共生矩阵^[17, 18]、小波变换^[19]两种常用的方法对尾迹纹理分类, 并将它们的分类结果和本文

方法进行性能比较。将图像的灰度量化成 16 等级, 求得它的灰度共生矩阵, 采用灰度共生矩阵的角二阶距、对比度、相关和熵 4 个统计量作为特征量, 与 4 个方向 (0°, 45°, 90°, 135°) 的灰度共生矩阵共得到 16 个纹理分类特征量。采用广泛使用的 Daubechies-4 型小波, 对图像进行 3 层分解, 分别计算低频通道, 水平方向的 3 个高频通道, 垂直方向的 3 个高频通道共 7 个通道的均值和方差作为纹理分类特征量, 共 14 个。分类器采用第 3 节描述的 BP 神经网络: 灰度共生矩阵隐含层的节点数为 16, 小波变换隐含层的节点数为 14。训练样本和测试样本与第 4 节 (1) 中的训练样本和测试样本相同。表 2 给出了不同方法对训练样本和测试样本的正确

表 2 3种方法的识别率

Tab 2 Recognition percentage of three methods

单位: %

	训练样本	测试样本
灰度共生矩阵	83.3	75
小波变换	93.3	82.5
本文方法	96.7	87.5

分类率。

从表 2 可以看出, 相对于灰度共生矩阵和小波变换, 本文算法能够更有效地提取尾迹纹理特征, 这主要是因为不同波长的尾迹纹理一定会在傅里叶变换的某个频率产生很强的响应, 本文采用的 Gabor 滤波器组能够有效地覆盖这些响应。

6 结 论

以多通道 Gabor 滤波器和 BP 神经网络理论为基础, 提出了一种水中目标运动尾迹的自动提取算法, 解决了由于信号弱, 噪音强而造成的水中目标难以识别的问题, 为可见光航空遥感监测提供了技术支撑, 对民用和军用都有着重大的意义。大量的实验结果表明, 本文算法能够准确地提取水中运动物体产生的尾迹。

参考文献 (References)

- 1 Sklansky Jack. Image segmentation and feature extraction [J]. IEEE Transactions on Systems, Man, and Cybernetics Society, 1978, 8(4): 237~247.
- 2 Haralick R M. Statistical and structural approaches to texture [J]. Proceeding of the IEEE, 1979, 67(5): 786~804.
- 3 Gool L Van, Dewaele P, Oosterlinck A. Texture analysis anno 1983 [J]. Computer Vision Graphics Image Process, 1985, 29: 336~357.
- 4 Jain A K, Farrokhnia F. Unsupervised texture segmentation using Gabor filters [J]. Pattern Recognition, 1991, 24(12): 1167~1186.
- 5 Tan T N. Texture edge detection by modeling visual cortical channels [J]. Pattern Recognition, 1995, 28(9): 1283~1298.
- 6 Bovik A C, Clark M, Geisler W S. Multichannel texture analysis using localized spatial filters [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1990, 12(1): 55~73.
- 7 Tenenier A, Pichler O, Hosticka B J. Unsupervised texture segmentation of images using tuned matched Gabor filters [J]. IEEE Transactions on Image Process, 1995, 4(11): 1549~1560.
- 8 Jin X C, Ong S H. A practical method for estimating fractal dimension [J]. Pattern Recognition Letters, 1994, 16(5): 457~464.
- 9 Pentland A P. Fractal based description of natural scenes [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1984, 6(6): 661~674.
- 10 Daugman J G. Uncertainty relation for resolution in space, spatial frequency and optimized by two-dimensional visual cortical filters [J]. Journal of the Optical Society of America A, 1985, 2(7): 1160~1169.
- 11 Pollen D A, Ronner S E. Visual cortical neurons as localized spatial frequency filters [J]. IEEE Transactions on System, Man and Cybernetics, 1983, 13(5): 907~916.
- 12 Lichard Lippmann. An introduction to computing with neural nets [J]. IEEE ASSP Magazine, 1987, 4(2): 4~22.
- 13 Richard O Duda, Peter E Hart, David G Stork. Pattern Classification (Second Edition) [M]. Brisbane Queensland Australia, John Wiley and Sons, 2001. [Richard O Duda, Peter E Hart, David G Stork 著. 模式分类 (第二版) [M]. 李宏东, 姚天翔等译. 北京: 机械工业出版社, 2003.]
- 14 David A Clausi, Jemigan M Ed. Designing Gabor filters for optimal texture separability [J]. Pattern Recognition, 2000, 33(11): 1835~1849.
- 15 Mahamadou Idrissa, Marc A cherohy. Texture classification using Gabor filters [J]. Pattern Recognition Letters, 2002, 23(9): 1095~1102.
- 16 Ramchandra Manthalkar, Biswas P K, Chatterji B N. Rotation invariant texture classification using even symmetric Gabor filters [J]. Pattern Recognition Letters, 2003, 24(12): 2061~2068.
- 17 Strand J, Taxt T. Local frequency features for texture classification [J]. Pattern Recognition, 1994, 27(10): 397~406.
- 18 Ohanian P P, Dubes R C. Performance evaluation for four classes of textural features [J]. Pattern Recognition, 1992, 25(8): 819~833.
- 19 Arivazhagan S, Ganesan L. Texture classification using wavelet transform [J]. Pattern Recognition Letters, 2003, 24(9): 1513~1521.

单波段单极化 SAR 图像水体和居民地 信息提取方法研究

胡德勇^{1,2)} 李京¹⁾ 陈云浩¹⁾ 蒋卫国¹⁾

¹⁾ (地表过程与资源生态国家重点实验室, 北京师范大学资源学院, 北京 100875)

²⁾ (三维信息获取与应用教育部重点实验室, 首都师范大学资源环境与旅游学院, 北京 100037)

摘要 SAR 图像上水体和居民地信息的提取在实际应用中具有重要的意义。为了更好地提取 SAR 图像上水体和居民地, 以单波段单极化 Radarsat-1 SAR 图像为研究对象, 首先利用半变异函数分析样本图像的结构特性来确定纹理信息提取的最佳参数; 然后, 在此基础上基于灰度共生矩阵计算 SAR 图像均值、角二阶矩和熵 3 种纹理测度, 建立了适于图像分类的多维特征空间, 从而有效地增强了水体和居民地信息; 最后通过样本采集, 使用支持向量机分类器进行水体和居民地信息提取, 并采用近期归一化植被指数 (NDVI) 数据和分类结果进行目标层融合来消除山体因素的影响, 信息提取的结果显示, 分类总体精度为 82.57%, Kappa 系数为 0.58 较准确地提取了水体和居民地信息。

关键词 合成孔径雷达 半变异函数图 灰度共生矩阵 支持向量机

中图法分类号: TP722.6 TP751 文献标识码: A 文章编号: 1006-8961(2008)02-0257-07

Water and Settlement Area Extraction from Single-band Single-polarization SAR Images Based on SVM Method

HU De-yong^{1,2)}, LI Jing¹⁾, CHEN Yun-hao¹⁾, JIANG Wei-guo¹⁾

¹⁾ (State Key Laboratory of Earth Surface Processes and Resource Ecology, College of Resources Science & Technology, Beijing Normal University, Beijing 100875)

²⁾ (Key Laboratory of 3D Information Acquisition and Application of Ministry of Education, College of Resource Environment & Tourism, Capital Normal University, Beijing 100037)

Abstract It is very important to extract water and settlement areas from SAR images in practical applications. In this paper the single band and single-polarization Radarsat-1 SAR image is used to water and settlement area extraction. Firstly, the statistic structure information of sample image is analyzed using semivariogram to determine the optimum parameters for textural information extraction. In order to establish the multidimension feature space for image classification, the textural measures such as mean, angle second moment and entropy have been calculated based on grey level co-occurrence matrix method. Then the water and settlement area information can be enhanced effectively using false color composite method. Three types of sample such as water, settlement and other are collected as training samples, and the image data are processed using support vector machine classification method. Finally, the image fusion on the target level between classification result image and NDVI image is conducted in order to eliminate the mountain influence, and the water and settlement areas are extracted accurately with a total classification accuracy of 82.57%, and Kappa coefficient of 0.58.

Keywords synthetic aperture radar (SAR), semivariogram, grey level co-occurrence matrix, support vector machine (SVM)

基金项目: 国家自然科学基金项目 (40671130)

收稿日期: 2006-08-30 改回日期: 2006-11-06

第一作者简介: 胡德勇 (1974~), 男, 博士。主要从事遥感与 GIS 在资源环境、自然灾害等领域的应用研究。Email: deyonghu@ires.cn

通讯作者: 李京, Email: lijing@ires.cn

1 引言

众所周知,水体和居民地信息在洪水灾害管理、环境监测及城市规划等领域具有重要的意义,如何快速有效获取这些信息,已成为人们关注的问题,但是有时往往受到观测条件限制,不能实时获得这些信息。雷达遥感由于可以全天时、全天候对地观测,具有传统光学遥感无法比拟的优点,因此其应用日益受到人们的重视。

由于地表结构复杂、地表形态多样,加上合成孔径雷达(SAR)成像原理和光学遥感有很大不同,致使雷达图像上亮度值相同的像元并不完全对应着地表同类地物,因此用基于单个像元的传统方法来提取 SAR 图像信息,受到了很大的限制^[1]。Radarsat-1 SAR 图像具有单波段、单极化特点,因此单凭它进行信息提取存在很多不足。但是由于 SAR 图像亮度范围较大,且含有较丰富的纹理结构信息,因此可以辅助原图像进行分类,以提高分类精度^[2]。纹理信息提取的窗口大小、空间距离和方向等参数对信息提取结果存在一定影响,是纹理分析研究较多的内容^[3]。许多学者在 SAR 图像信息提取方面进行过研究,例如,辅以纹理信息对居民地分类精度有明显的提高^[4];用纹理图像亮度阈值分割法提取居民地可以获得 70% 以上的正确识别率^[5];采用纹理信息复合的模糊 C 均值聚类算法进行自动分类等^[6]。另外,还有基于神经网络^[7-10]、模糊理论^[11]等方法在 SAR 图像分类中也得到了应用。SAR 图像水体和居民地的信息提取,由于受到地面特性、图像斑点噪声和分类算法等影响,目前大多采用阈值分割或者监督分类方法,因此精度难以保证,对于类似居民地的具有复杂结构的地物信息的提取方法还值得探讨。

本文选择 Radarsat-1 SAR 图像,并基于灰度共生矩阵进行了纹理结构信息的提取,首先有效地增强了 SAR 图像,然后在此基础上进行样本采集,并使用支持向量机(SVM)进行专题信息提取,以探讨单波段、单极化 SAR 图像水体和居民地信息的快速提取方法。

2 研究数据和方法

2.1 数据

在本研究中,试验区数据为 2004 年 7 月 10 日

Radarsat-1 的 ScanSAR Narrow (SNB) 降轨数据,入射角约 $39^{\circ} \sim 42^{\circ}$, 图像采样后的地面分辨率为 25m。从整景图像中裁减 $1405 \times 1412 \text{ pixels}$ 大小的区域作为研究区,范围为 $112^{\circ} 46' 50'' \sim 113^{\circ} 8' 40'' \text{ E}$, $28^{\circ} 3' 7'' \sim 28^{\circ} 21' 55'' \text{ N}$, 覆盖湖南省长沙市及其郊县地区(如图 1 所示)。该区地貌以丘陵为主,湘江自南向北贯穿整个研究区,在研究区东北、西南方向存在部分地势起伏较大地区。植被覆盖为亚热带常绿阔叶林和落叶林,主要土地利用类型为居民地、耕地、园地、林地、交通用地等。另外,为了验证 SAR 图像分类的结果,还收集到了覆盖研究区近期的 Landsat ETM+ 数据两景。

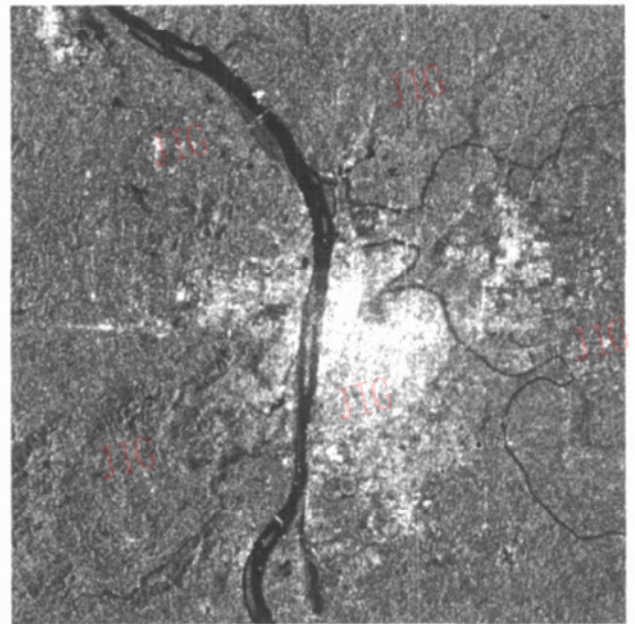


图 1 研究区 Radarsat-1 的 ScanSAR 图像

Fig 1 The Radarsat-1 ScanSAR image of the study area

2.2 处理流程

从 SAR 图像上提取水体和居民地的处理流程是:首先对 SAR 图像进行 Gamma (3×3) 滤波,同时通过半变异函数分析来分析图像的结构特征,并在此基础上提取纹理信息;然后使用 SVM 来进行图像分类,以获取水体和居民地信息。考虑到山体等因素的干扰,可通过对分类后的数据和从 ETM+ 图像提取的归一化植被指数(NDVI)数据进行目标级融合来减少错分、误分,以提高分类精度;最后是结果对比和精度评价(如图 2 所示)。

3 数据处理和结果分析

3.1 水体和居民地的微波遥感机理

众所周知,雷达图像的亮度值代表雷达回波强

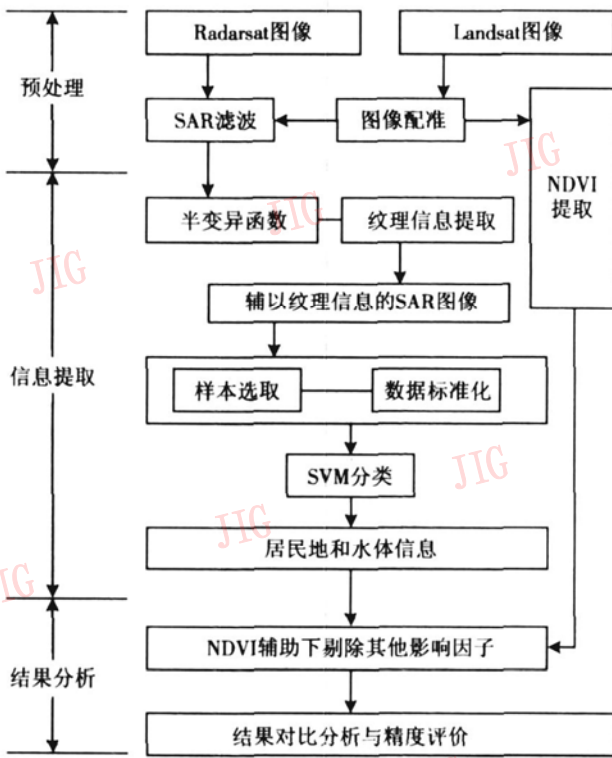


图 2 纹理复合的 SAR 图像信息提取及精度评价流程图

Fig 2 Procedure of SAR images information extraction and accuracy evaluation based on textural analysis

度的大小, 它定量地由雷达后向散射系数决定, 而雷达系统和地物目标都影响后向散射系数值大小。对于特定的波长、入射角和极化方式的雷达系统而言, 影像特征主要取决于不同的地物目标。在 SAR 图像上高密度居民地一般呈现较亮色调, 而在低密度居民地的亮色调之间则夹杂一些暗色调, 其主要由居民地内的镜面反射、漫反射、各向同性散射, 以及其角反射和谐振效应等的强度和空间配置决定^[12]; 由于水体对 Radarsat-1 SAR 的波长而言, 属于光滑表面, 因而在图像上呈暗色或黑色, 且色调较为均一。居民地和水体具有不同的纹理特征, 在进行 SAR 图像信息提取时, 可以通过加入它们的纹理特性来增强水体和居民地专题信息。

3.2 纹理测度算法

纹理测度算法可分为以下两类^[4]: 一类是结构分析方法, 它主要是提取纹理基元, 从结构组成上探索纹理的规律; 另一类是统计分析方法, 它主要是从图像有关属性的统计分析出发来计算纹理特征, 其在目前的纹理测度中占有重要的地位, 灰度共生矩阵法 (grey level co-occurrence matrix, GLCM) 就是其中常用的方法之一^[13]。它是一种基于图像灰度联合概率矩阵的方法, 即它是通过计算图像邻近像元

灰度级之间的二阶联合条件概率密度 $P(i, j | \delta, \theta)$ 来表示纹理, $P(i, j | \delta, \theta)$ 表示在给定的空间距离 δ 和方向 θ 上, 以灰度级 i 为起始点出现灰度级 j 的概率。对于具有 N 级灰度的图像, 其灰度共生矩阵 P 可以表示为

$$P = \{P_k(i, j | \delta, \theta)\}_{N \times N} \quad (1)$$

式中, N 表示图像的灰度级数, K 为像元顺序号。

由灰度共生矩阵 P 可以生成多种统计值作为纹理特征的度量。由于居民地在 SAR 图像上的表征总体上呈现亮色调, 而水体呈现暗色调, 因此根据水体和居民地的纹理特性, 选择均值 (MEA)、熵 (ENT)、角二阶矩 (ASM) 等^[14] 纹理测度都可以在 SAR 图像上将水体、居民地和其他信息区别开来, 其中

$$MEA = \sum_{i=1}^N \sum_{j=1}^N i \times P(i, j) \quad (2)$$

$$ENT = \sum_{i=1}^N \sum_{j=1}^N -P(i, j) \times \ln P(i, j) \quad (3)$$

$$ASM = \sum_{i=1}^N \sum_{j=1}^N (P(i, j))^2 \quad (4)$$

式 (2) ~ 式 (4) 中, $P(i, j)$ 为灰度共生矩阵的元素, N 为矩阵的大小。实际应用中, 常常通过计算窗口 w (大小为 $n \times n$ pixels) 的灰度共生矩阵来得到 w 的中心像元的纹理测度值 $T_{(x, y)}$ 。对于 SAR 图像的每个像素 (x, y) , 先分别计算其周围窗口 w 内的 $T_{(x, y)}$ 的值 (包括 MEA、ENT、ASM 的值), 然后在图像上移动窗口 w , 最后就可以生成纹理特征图像。它们先和 SAR 图像的灰度值组成多维空间, 然后通过假彩色合成来有效地增强单波段、单极化 Radarsat-1 SAR 影像的专题信息。

3.3 半变异函数分析与纹理信息提取

显而易见, $T_{(x, y)}$ 是 δ, θ, w 等 3 个变量的函数, 其中用于不同图像最佳纹理特征计算的参数是不同的, 这是纹理信息提取值得探讨的方面。由于半变异函数能够反映 SAR 图像数据的随机性和结构性, 且地物在图像上也表现出很好的空间自相关性和纹理特征^[15], 所以影响 $T_{(x, y)}$ 的因素 δ, θ, w 可以通过研究区样本的半变异函数 γ 的大小来分析, 本文选择 $0^\circ, 45^\circ, 90^\circ, 135^\circ$ 等 4 个方向来讨论半变异函数, 其定义分别为

东西方向 ($\theta = 0^\circ$):

$$\gamma_1 = 1/2m \sum_{i=1}^n \sum_{j=1}^n [Z(i, j) - Z(i, j + \delta)]^2 \quad (5)$$

东北-西南方向 ($\theta = 45^\circ$):

$$y_2 = 1/2m \sum_{i=1}^n \sum_{j=1}^n [Z(i, j + \delta) - Z(i + \delta, j)]^2 \quad (6)$$

南北方向 ($\theta= 90^\circ$):

$$y_3 = 1/2m \sum_{i=1}^n \sum_{j=1}^n [Z(i, j) - Z(i + \delta, j)]^2 \quad (7)$$

东南-西北方向 ($\theta= 135^\circ$):

$$y_4 = 1/2m \sum_{i=1}^n \sum_{j=1}^n [Z(i, j) - Z(i + \delta, j + \delta)]^2 \quad (8)$$

式(5)~式(8)中, δ 表示距离, n 表示样本图像大小, $Z(i, j)$ 为像素 (i, j) 的灰度值, m 是计算的数据对数目, 它取决于计算的方向和间距。为增强具有复杂结构特性的居民地信息, 可先在图像上随机采集城镇居民地样本, 然后分析样本的不同方向 ($\theta= 0^\circ, 45^\circ, 90^\circ, 135^\circ$)半变异函数分布情况(如图 3所示)。

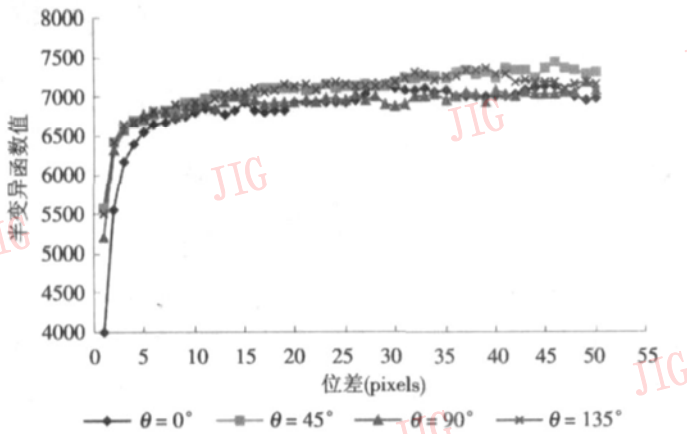


图 3 研究区城镇居民地样本的不同方向半变异函数图
Fig 3 Semivariogram of urban land sample in different directions

从图 3可以看出, 不仅 SAR 图像上的城镇居民地具有结构性, 并且不同方向的结构不一致, 即具有方向性。 $y_2、y_4$ 基台值 (sill)要高出 y_1 和 y_3 , 且 y_2 和 y_4 的变程 (range)基本吻合, 均约在 6pixels 左右曲线趋于平缓, 这说明样本在该窗口范围内的空间依赖性即达到显著。虽然 $y_2、y_4$ 在 40pixels 左右有个峰值出现, 但是由于窗口太大会将不同纹理性质均匀化, 且将导致纹理信息的平滑^[12], 不适合作为纹理提取窗口, 因此纹理信息提取的窗口大小选择为 6×6 pixels, $\delta= 1, \theta= 135^\circ$, 提取像素 (x, y) 周围 6×6 大小的范围内的 3 个纹理测度 $T_{(x, y)}$, 它和 SAR 图像灰度值一起构成了 4 维空间。

朱彩英曾经提出使用均值、方差、对比度等 3 个纹理特征来区分星载 SAR 图像上的居民地与其他地物^[5]; 杨存建提出将原始波段赋以红色, 反差分

纹理波段赋以绿色, 角二阶矩纹理波段赋以蓝色, 用来进行假彩色合成, 可以有效地增强居民地专题信息^[1]。本文选择角二阶矩、均值和熵, 分别赋为 $R、G、B$ 进行假彩色合成, 这样在假彩色合成图像上, 水体显示为红色, 居民地为黄色, 其他为蓝色(如图 5(a)所示), 这就可以有效地增强水体和居民地信息。

3.4 水体和居民地信息提取

专题信息提取可以采用图像分类来实现。目前分类算法较多, 其中最大似然法分类方法 (maximum likelihood classification method, MLC) 是较为常用的分类器, 它假设数据的分布形式是已知的, 样本是相互独立的, 但当总体分布不符合正态分布时, 其分类可靠性将下降^[16]。

SVM 是基于 Vapnik-Chervonenkis 理论的创造性机器学习方法^[17], 它是根据 Vapnik 的结构风险最小化原则, 尽量提高学习机的泛化能力, 以便由有限的训练集样本得到的小误差能够保证对独立的测试集仍保持小的误差。对于非线性分类, SVM 可通过非线性映射将样本空间转换到高维特征空间来进行线性转换, 在特征空间中引入不敏感损失函数 ξ 并通过定义最优线性回归超平面, 把寻找最优线性回归超平面的算法归结为求解一个凸约束条件下的凸规划问题, 并可以求得全局最优解^[18]。SVM 图像分类主要步骤包括样本采集、参数设置、样本集训练和应用预测模型进行图像分类等(图 4)。

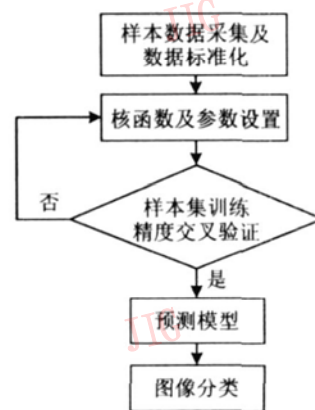


图 4 SVM 图像分类流程图

Fig 4 Procedure of image classification based on SVM

对于 SAR 图像来说, 由于其成像机理的复杂性, 加上斑点噪声影响, 致使 MLC 分类器在图像分类的时候还存在很多不足, SVM 由于适合解决本质上的非线性问题, 因此适合 SAR 图像信息提取^[19-21]。本文选择 SVM 分类器进行图像分类。分

类时,首先通过研究区的样本数据采集,获取居民地、水体和其他类型的样本,其中居民地的样本数量为 59 370 pixels 约占总体的 2.99%,水体的样本数量为 15 147 pixels 约占总体的 0.76%,其他类型为 148 000 pixels 约占总体的 7.46%。为消除不同属性数据差别很大造成的对分类结果的影响,在进行 SVM 分类前需要先进行数据标准化,其主要包括以下两个阶段:首先对训练数据进行标准化,在本研究中通过将数据标准化到 -1 到 1 之间来生成标准化模型;然后根据标准化模型来标准化用于分类的图像数据,整个 SVM 分类在 Libsvm 软件中进行^[22]。

为了对 SVM 和 MLC 分类结果进行横向对比,本文使用相同样本对纹理复合的 SAR 图像进行分类,结果见图 5(b)、图 5(c)。从图像定性分析角度可以看出, SVM 分类结果要优于 MLC 分类法,且“椒盐 (pepper and salt) 现象”明显要少。

平静的水面总是造成 SAR 图像镜面反射,在图像上呈黑色调,但是在地形起伏较大的地区,由于雷达阴影存在,对水体信息提取具有一定的干扰。考虑到研究区的具体情况,山地地区的植被覆盖较好,可以通过辅助 NDVI 数据来修正水体信息,即对 SVM 分类图像和 NDVI 数据进行目标级融合,以消除山地阴影的影响,目标层融合的准则为

$$R = W(NDVI < 0) + S + O \quad (9)$$

式中, R 为融合结果, W 、 S 、 O 分别代表水体、居民地和其他类别像素值, $NDVI$ 为归一化植被指数。

图 5(b) 为纹理复合 SVM 分类结果图,由该图可以看出,除了水体以外,西南部山体阴影部分也被作为水体信息提取出来了;图 5(d) 为消除了山地影响的效果图,由图 5(d) 可见,水体信息和实际情况比较相符,河流、坑塘在分类结果图上得到较准确的显示,但是穿过城区的河流部分仍难以准确地提取,究其原因是由于雷达图像本身特性和水体样本采集方面的原因造成的。

3.5 精度评价

为了评价研究区 SAR 图像的 SVM 分类精度,首先将 Landsat ETM+ 全色波段和多光谱数据进行融合,可得到分辨率为 15m 的假彩色合成图像,这个精度基本可以达到水体和居民地的目视解译要求,因此可以作为精度评价的参考图像;然后使用随机选点方式生成样本点 499 个,用于进行 SAR 分类结果的精度评价,其分类结果见表 1,居民地由于在 SAR 图像上的高亮特性,因此加上纹理信息后的分

表 1 纹理复合的 SAR 图像 SVM 分类精度评价

Tab 1 The accuracy evaluation of classification image using textural analysis based on SVM

类别名	参考点数	分类点数	正确点数	正确率 (%)
水体	27	25	21	84.00
其他	333	406	326	80.48
居民地	139	68	64	94.20
总数	499	499	411	

注:总体精度为 82.57%, Kappa 系数为 0.58

类精度可以达到 94.20%,水体分类精度可以达到 84.00%,而其他类分类精度为 80.48%。整个研究区图像分类精度为 82.57%, Kappa 系数为 0.58。分类之所以存在一定误差,主要是由于在水体和居民地信息提取过程中,存在少数散村、坑塘水体等的误提和漏提的缘故。

4 结论和讨论

(1) 由于单波段、单极化 SAR 图像具有较丰富的纹理信息,因此本文基于灰度共生矩阵提取其纹理结构信息,再将它和 SAR 图像复合生成多通道图像,可以有效地增强居民地和水体等专题内容,以弥补单通道 SAR 图像信息的匮乏和不足,从而为多云雨地区 SAR 图像快速获取水体和居民地信息提供支持,其对洪水灾害管理、环境监测及城市规划等领域具有重要的意义。

(2) SAR 图像加入纹理信息以后,斑点噪声对 SAR 图像分类具有减弱作用,其主要表现在分类结果的“椒盐现象”明显减少,这是因为 6×6 纹理提取窗口对图像具有一定的滤波作用,使原图像和纹理通道的复合形成多维空间,由于弱化了噪声的影响,从而有利于 SAR 图像专题信息提取;和目前使用较多的 MLC 法相比, SVM 法用于 SAR 信息提取具有较大优势,从图像定性分析来看,MLC 分类结果中,地物分布细碎,而 SVM 分类结果则多为面状分布,定性分析结果显示, SVM 的分类效果较好;本研究使用 SVM 进行水体、居民地和其他类信息提取,其分类精度分别达到 84.00%、94.20% 和 80.48%,总体分类精度达到 82.57%, Kappa 系数为 0.58。

SAR 图像分类由于受到地表特性、斑点噪声和

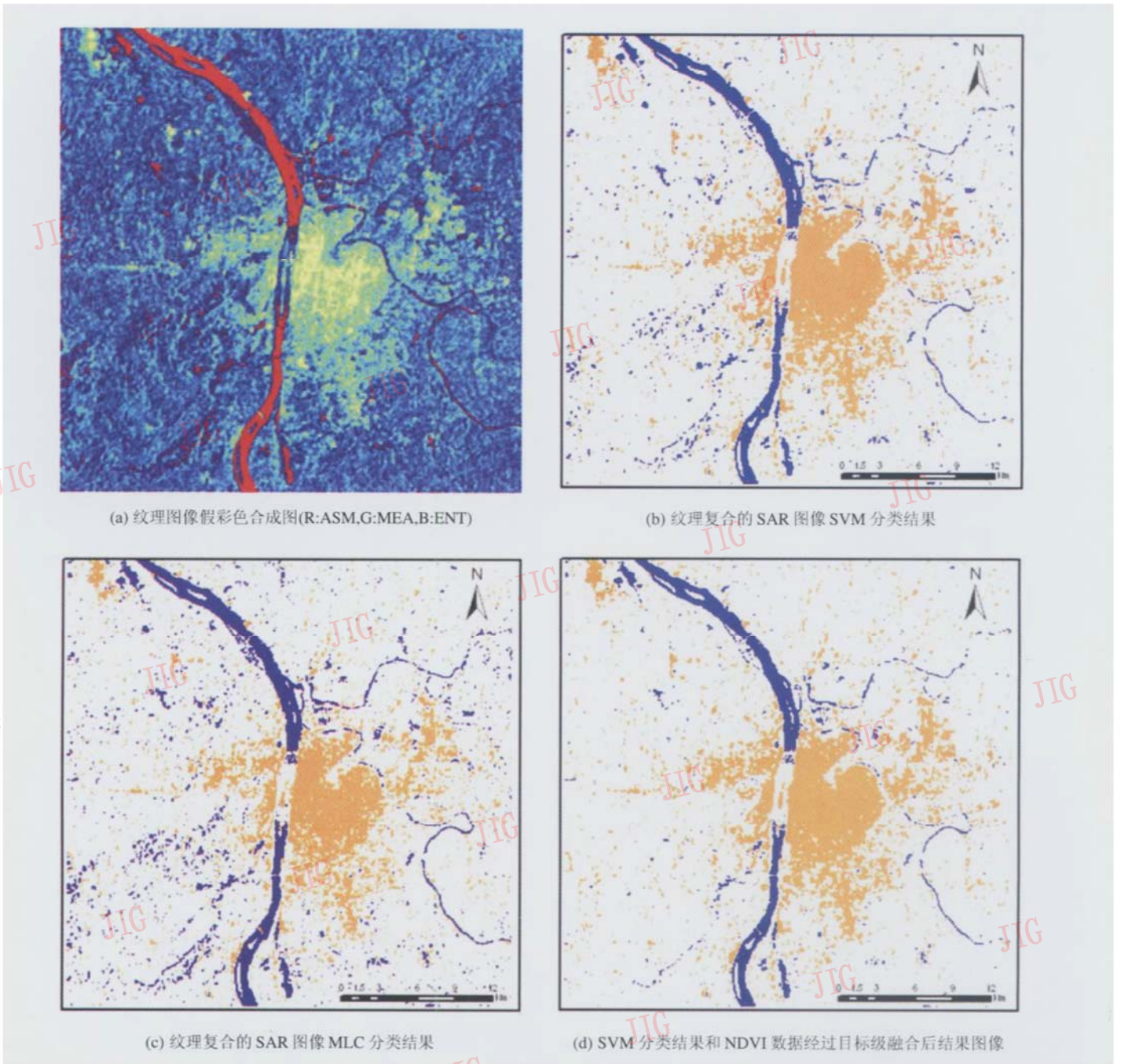


图 5 纹理图像假彩色合成图、SVM 和 SLC 分类结果以及目标层融合后结果图像

Fig 5 False colour composite image using textural channel, the classification image using SVM and MLC, the object-level fusion image

分类算法等方面的影响,致使水体和居民地信息提取还受到一定的限制,难以避免误分、误提的现象,比如机场之类的平坦地区,其较低的后向散射导致雷达影像难以和水体区别,还需要加入先验知识进行判别,以期进一步提高图像的分类精度。

参考文献 (References)

1 Yang Cun-jian, Zhou Cheng-hu, Wei Yim-ing *et al* Enhancing the settlement information in SAR imagery [J]. Remote Sensing Technology and Application, 1998, 13(4): 5~ 8 [杨存建,周成

虎,魏一鸣等.星载 SAR 图像中居民地专题信息增强的方法探讨 [J]. 遥感技术与应用, 1998, 13(4): 5~ 8]

2 Floyd M H, Xia Zong-guo SAR application in human settlement detection, population estimation and urban landuse pattern analysis a status report [J]. IEEE Geoscience and Remote Sensing, 1997, 35(1): 79~ 85

3 Ban Yifang Improving the accuracy of synthetic aperture radar analysis for agricultural crop classification [J]. Canadian Journal of Remote Sensing, 1995, 21(2): 158~ 164

4 Fang Shen-hui, Zhu Wu Classification of SAR using multipolarimetric and texture image [J]. Bulletin of Surveying and Mapping, 2001, (10): 12~ 14 [方圣辉,朱武.辅以纹理特征的 SAR 图像分类

- 研究 [J]. 测绘通报, 2001, (10): 12~ 14]
- 5 Zhu Cai-ying, Lan Chao-zhen, Jin Guo-wang. Habitation area extraction by the method of intensity separation from SAR textured image [J]. Journal of Image and Graphics, 2003, 8(6): 616~ 619. [朱彩英, 蓝朝桢, 靳国旺. 纹理图像亮度阈值法提取 SAR 图像居民地 [J]. 中国图象图形学报, 2003, 8(6): 616~ 619.]
 - 6 Hu Fu-yuan, Zhang Yan-ning, Xie Xiao-rong *et al*. A classification method for SAR image based on tree wavelet and gray-level co-occurrence matrix [J]. Systems Engineering and Electronics, 2003, 25(10): 1286~ 1288. [胡伏原, 张艳宁, 薛笑荣等. 基于树型小波和灰度共生矩阵的 SAR 图像分类 [J]. 系统工程与电子技术, 2003, 25(10): 1286~ 1288.]
 - 7 Ndi Youngui A, Tonye E, Akono A. Evaluation of speckle filtering and texture analysis methods for land cover classification from SAR images [J]. International Journal of Remote Sensing, 2002, 23(9): 1895~ 1925.
 - 8 Dell'Acqua F, Gamba P, Trianni G. Semi-automatic choice of scale-dependent features for satellite SAR image classification [J]. Pattern Recognition Letters, 2006, 27(4): 244~ 251.
 - 9 Gamba P, Dell'Acqua F. Increased accuracy multiband urban classification using a neuro-fuzzy classifier [J]. International Journal of Remote Sensing, 2003, 24(4): 827~ 834.
 - 10 Tzeng Y C, Chen K S. A fuzzy neural network to SAR image classification [J]. IEEE Transactions on Geoscience and Remote Sensing, 1998, 36(1): 301~ 307.
 - 11 Benz Ursula C. Supervised fuzzy analysis of single- and multichannel SAR data [J]. IEEE Transactions on Geoscience and Remote Sensing, 1999, 37(2): 1023~ 1037.
 - 12 Zhou Cheng-hu, Luo Jian-cheng, Yang Xiaomei *et al*. Geo-understanding and Analysis for Remote Sensing Image [M]. Beijing: Science Press, 1999. [周成虎, 骆剑承, 杨晓梅等. 遥感影像地学理解与分析 [M]. 北京: 科学出版社, 1999.]
 - 13 Raiesh K, Jawahar C V, Sengupta S *et al*. Performance analysis of textural features for characterization and classification of SAR images [J]. International Journal of Remote Sensing, 2001, 22(8): 1555~ 1569.
 - 14 Clausi David A. An analysis of co-occurrence texture statistics as a function of grey level quantization [J]. Canadian Journal of Remote Sensing, 2002, 28(1): 45~ 62.
 - 15 Miranda F P, Fonseca L E N, Carr J R. Semi-variogram textural classification of JERS-1 (Fuyo-1) SAR data obtained over a flooded area of the Amazon rainforest [J]. International Journal of Remote Sensing, 1998, 19(3): 549~ 556.
 - 16 Zhu Jun-jie, Guo Hua-dong, Fan Xiang-tao *et al*. The application of the wavelet texture method to the classification of single-band single-polarized and high-resolution SAR images [J]. Remote Sensing for Land & Resources, 2005, (2): 36~ 39. [朱俊杰, 郭华东, 范湘涛等. 单波段单极化高分辨率 SAR 图像纹理分类研究 [J]. 国土资源遥感, 2005, (2): 36~ 39.]
 - 17 Cortes C, Vapnik V. Support vector networks [J]. Machine Learning, 1995, 20(3): 273~ 297.
 - 18 Deng Nai-yang, Tian Ying-jie. Support Vector Machine: A New Method in Data Mining Field [M]. Beijing: Science Press, 2004. [邓乃扬, 田英杰. 数据挖掘中的新方法——支持向量机 [M]. 北京: 科学出版社, 2004.]
 - 19 Angiulli G, Barrile V, Cacciola M. SAR imagery classification using multi-class support vector machines [J]. Journal of Electromagnetic Waves and Applications, 2005, 19(14): 1865~ 1872.
 - 20 Su Fu-liu, Ni Liang, Li Da-fang *et al*. Classification of SAR image based on gray cooccurrence matrix and support vector machine [A]. In: Proceedings of 2004 7th International Conference on Signal Processing [C], Harbin Institute of Technology, China, 2004, 2: 1385~ 1388.
 - 21 Krishnapuram B, Sichina J, Carin L. Physics-based detection of targets in SAR imagery using support vector machines [J]. IEEE Sensors Journal, 2003, 3(2): 147~ 157.
 - 22 Zhang Jin-shui, He Chun-yang, Pang Yao-zhong *et al*. The high spatial resolution RS image classification based on SVM method with the multi-source data [J]. Journal of Remote Sensing, 2006, 10(1): 49~ 57. [张锦水, 何春阳, 潘耀忠等. 基于 SVM 的多源信息复合的高空间分辨率遥感数据分类研究 [J]. 遥感学报, 2006, 10(1): 49~ 57.]

基于各向异性 Retinex 的路面图像阴影消除

唐磊 赵春霞 王鸿南 邵文泽

(南京理工大学计算机科学与技术学院, 南京 210094)

摘要 路面图像中常常存在车辆、树木、建筑物等的阴影, 给图像的特征抽取和识别带来极大的干扰。通过分析中心/环绕 Retinex 算法 (center/surround Retinex, CSR) 以及现有的基于 Retinex 的阴影消除算法处理阴影图像的不足, 提出了一种各向异性中心/环绕 Retinex 算法 (anisotropic diffusion center/surround Retinex, ADCSR), 该算法融入了基于偏微分方程 (PDE) 的各向异性扩散, 并根据算法特点提出了基于“边界性” (edge degree, ED) 的各向异性扩散方案, 避免了梯度门限等参数选择的困难, 在消除阴影的效果上和运算效率上都取得了令人满意的结果。

关键词 路面图像 阴影消除 各向异性 Retinex 偏微分方程

中图分类号: TP391.41 文献标识码: A 文章编号: 1006-8961(2008)02-0264-05

Shadow Removal for Road Surface Images Based on Anisotropic Diffusion Retinex

TANG Lei ZHAO Chun-xia WANG Hong-nan SHAO Wen-ze

(Department of Computer Science, Nanjing University of Science and Technology, Nanjing 210094)

Abstract There are sometimes shadows, such as cars, trees and buildings, on road surface images, which make it difficult to abstract and recognize the features. After analyzing the shortages of Center/Surround Retinex (CSR) algorithm and shadow removal approaches based on Retinex in existence, an Anisotropic Diffusion Center/Surround Retinex (ADCSR) is presented to solve the problem. First anisotropic diffusion based on PDE is introduced to ADCSR, further a new anisotropic diffusion scheme based on “Edge Degree” (ED) is presented, which avoids the embarrassment to select different parameters such as gradient threshold. Theoretic analysis and experimental results show that the effectiveness of the approach.

Keywords road surface image, shadow removal, anisotropic diffusion, Retinex, PDE

1 引言

路面病害的自动检测是道路养护事业中的一个重要课题。目前的自动检测主要是通过图像采集系统采集路面图像并对图像进行特征抽取和识别来实现。但路面图像中往往会存在车辆、树木、建筑物等的阴影, 这些阴影对病害检测和分割带来极大的干扰。因此, 设计有效的算法消除图像中的阴影是很

有必要的。

Retinex 理论是由 Land 于 20 世纪 70 年代提出的一种基于颜色恒常知觉的计算理论^[1], 目前被广泛应用于图像的增强、恢复等方面, 其中一个很重要的应用就是光照补偿^[2]。通过对图像进行光照补偿, 可以平衡图像的整体灰度, 获取图像中较暗区域的细节, 同时也能保持原有较亮区域的细节。但 Retinex 理论假设空间的照度变化是缓慢的, 当图像中出现阴影时, 由于在阴影边界处照度急剧变化,

基金项目: 江苏省自然科学基金项目 (BK2004130)

收稿日期: 2006-03-23 改回日期: 2006-09-29

第一作者简介: 唐磊 (1981~), 男, 现为南京理工大学计算机系模式识别与人工智能专业博士研究生。主要研究方向为数字图像处理、智能机器人、组合导航与控制等。E-mail: tang_lei2008@163.com

Retinex 算法将导致阴影边界附近模糊, 算法失效。通过对图像的遍历并对不同亮度区域设定不同权值, Retinex 算法可以对阴影的消除取得一定效果^[3,4]。但该方法复杂度过高, 不适合数据量庞大的路面图像的处理。

中心环绕 Retinex 算法 (center/surround Retinex, CSR) 是 Retinex 理论的最新实现形式, 它计算简单, 而且关键参数容易设置, 目前得到越来越广泛的应用。本文通过深入分析 CSR 算法, 指出了其无法消除阴影的问题所在, 并针对阴影的消除设计了各向异性中心环绕 Retinex 算法 (anisotropic diffusion center/surround Retinex, ADCSR)。ADCSR 将基于偏微分方程 (PDE) 的各向异性扩散融入 Retinex 算法中, 并提出了适合算法特点的基于“边界性” (edge degree, ED) 的各向异性扩散方案, 避免了梯度门限等参数选择的困难。ADCSR 算法拓展了 Retinex 理论的适用范围, 在消除阴影的效果和运算效率上都取得了令人满意的结果。

2 Retinex 算法

2.1 中心环绕 Retinex 算法及其局限性

Retinex 理论认为, 人眼视觉系统在光照条件变化下仍然能够对场景中每一点的亮度有稳定的感知能力, 是因为人眼对物体亮度的感知不仅由物体所受到的光照强度决定, 还取决于物体本身的反射系数。根据该理论, 一幅图像 I 可以用下式表示:

$$I(x, y) = L(x, y) \times R(x, y) \quad (1)$$

L 表示入射到物体上的光源能量总和 ($0 < L(x, y) < \infty$), R 表示物体的反射系数 ($0 < R(x, y) < 1$)。两边取对数,

$$\log I = \log L + \log R \quad (2)$$

Retinex 算法的主要目的就是 1 幅给定的图像 I 分解为反射图像 R 和照度图像 L , 并以 R 作为输出, 以此来消除光照不均的影响。

CSR 算法表示式为 (以单通道图像为例):

$$R(x, y) = \log I(x, y) - \log \{k \cdot [F(x, y) * I(x, y)]\} \quad (3)$$

该算法以在 \log 域中的反射图像 R (相当于式 (2) 中的 $\log R$) 作为输出, 其中, $*$ 表示卷积运算, $F(x, y)$ 是环绕高斯函数, 其解析式为

$$F(x, y) = e^{-r^2/c^2}, r = \sqrt{x^2 + y^2} \quad (4)$$

c 为常数, 表示环绕高斯函数的尺度。 s 为归一化常数

$$s = \frac{1}{\iint (x, y) dx dy} \quad (5)$$

从式 (3) 可以看出, CSR 算法大致可分为以下 3 步:

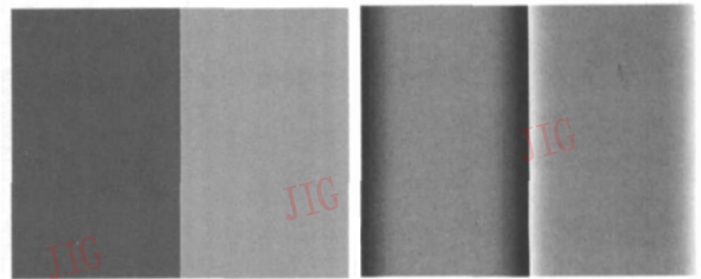
(1) 以 $s \cdot F(x, y) * I(x, y)$ 项模拟计算图像的全局照度分布。

(2) 将原图像与全局照度分布变换到 \log 域, 将乘法运算转换为加减运算。

(3) 在 \log 域中将原图像减去全局照度分布, 输出原图像在 \log 域中的反射系数, 图像的光照效果被去除。

其中第 1 步为算法的关键步骤, 它假设全局照度是缓慢变化的, 因此全局照度是图像中的低频部分, 并可以通过对原图像进行高低通滤波来模拟。

CSR 算法可以实现图像动态范围的压缩, 有效地消除图像中不均匀光照效果的影响, 但该算法在图像中包含阴影时失效。其原因在于阴影的存在使得“空间照度是缓慢变化”的假设不成立。体现在算法上的缺陷则是 CSR 算法在模拟计算图像的全局照度分布时, 所采用的高斯卷积核是一种各向同性的, 而且为防止结果图像颜色的失真, 要求采用的卷积核尺度较大。当尺度较大的卷积核通过明暗变化剧烈的阴影边界时, 将会严重模糊阴影的边界, 导致模拟全局照度分布失败, 最后将导致输出图像的阴影边界附近出现“光晕”现象, 阴影边界附近的细节也被丢失。



(a) 原图像 (b) 处理后含光晕的图像

图 1 CSR 算法的不足

Fig 1 The drawback of CSR

2.2 各向异性中心环绕 Retinex 算法

Koenderink 和 Hummel 指出, 求图像 I_0 和高斯核 G 的卷积相当于求解如下的标准热传导方程^[5,6]:

$$\begin{cases} \frac{\partial I}{\partial t} = \Delta I \\ I|_{t=0} = I_0 \end{cases} \quad (6)$$

因此通过迭代求解方程式 (6) 可以获得图像的低频部分, 可以用来模拟图像的全局照度分布 (等价于式 (3) 中 $s \cdot F(x, y) * I(x, y)$ 项)。

方程式 (6) 是一种各向同性算法, 没有考虑图像的特征区域 (如边界、角点等), 因而采用该方程直接计算图像的全局照度分布同样会模糊阴影的边界, 无法准确地模拟出全局照度。因此需要设计一种各向异性中心环绕 Retinex 算法, 在计算图像的全局照度分布时对边界采取不同的控制策略, 完好地保留图像中的边界, 这样才能真实地获取含有阴影的图像的照度分布。在计算得到准确的照度分布后, 通过从原图像中减去照度分布, 消除光照对图像的影响, 就能获得图像的本来面貌, 阴影得到了消除。

3 基于热传导方程的各向异性扩散

1990年, Perona 和 Malik 提出了标准的非线性热传导方程^[7]:

$$\begin{cases} \frac{\partial I}{\partial t} = \nabla \cdot [g(|\nabla I|) \nabla I] \\ I|_{t=0} = I_0 \end{cases} \quad (7)$$

式中, I_0 表示原始图像, $g(x)$ 表示扩散函数, 常用的形式有 $g(x) = \exp(-(x/f)^2)$ 和 $g(x) = 1/(1+(x/f)^2)$ 两种, 其中 f 是梯度门限。由于 $g(x)$ 为 x 的减函数, 因此方程式 (7) 在梯度较大的边缘处扩散较慢, 可以较好地保持图像的边缘。

由于在噪声环境下梯度函数对噪声的敏感性容易导致方程是病态的, 因此 Perona-Malik 方程的适定性不强。1992年, Catta 等人把方程修改为^[8]

$$\begin{cases} \frac{\partial I}{\partial t} = \nabla \cdot [g(|\nabla G_0 * I|) \nabla I] \\ I|_{t=0} = I_0 \end{cases} \quad (8)$$

与方程式 (7) 相比, 方程式 (8) 首先通过对图像进行卷积平滑, 在空域实现了对方程的正则化, 在一定程度上解决了方程解的适定性问题和方程的病态问题。

针对以上基本方程, 此后又有许多学者致力于对扩散模型、扩散函数等进行不同程度的改进^[9-10], 但这些方法均面临着梯度门限等关键参

数选取困难的问题。本文将提出新的基于非线性热传导方程的各向异性扩散方案, 该方案充分利用了算法特点, 定义了新的扩散函数, 避免了梯度门限等参数选择的困难, 而且在边界保持上取得了良好的效果。

4 基于各向异性 Retinex 的阴影消除

4.1 阴影检测

由于路面灰度比较均匀, 因此含有阴影的路面图像大致可分为阴影和非阴影两大块, 而且两块之间的明暗对比很明显, 因此可以采用基于阈值分割的方法来检测阴影。设

$$\begin{aligned} n_1 &= \sum_{i=0}^T H(i) \\ n_2 &= \sum_{i=T+1}^{255} H(i) \end{aligned} \quad (9)$$

式中, T 为采用 Otsu 方法计算的图像阈值^[11], $H(i)$ 表示第 i 级灰度的像素个数, 则 n_1 可以近似代表较暗区域的面积。假设阴影区域的面积不太小, 设定阴影面积的阈值 T_{shadow} (一般取为图像面积的 1/4), 当 $n_1 \geq T_{\text{shadow}}$ 时认为图像中含有阴影。在不同的环境和不同的光照条件下, 阴影的强度也不完全一致。设

$$\begin{aligned} G_{\text{dark}} &= \frac{1}{n_1} \sum_{i=1}^T i \cdot H(i) \\ G_{\text{bright}} &= \frac{1}{n_2} \sum_{i=T+1}^{255} i \cdot H(i) \end{aligned} \quad (10)$$

则 G_{dark} 和 G_{bright} 分别代表阴影区域和非阴影区域的平均明暗程度。令

$$k = G_{\text{bright}} - G_{\text{dark}} \quad (11)$$

则 k 可以表示图像中阴影区域和非阴影区域的明暗对比强度。

4.2 边界性测度

传统的基于非线性热传导方程的各向异性扩散算子目的在于保持图像中的边界以及细节, 而全局照度的目的只在于反映图像整体明暗的变化趋势, 不关心图像的细节。从而用来模拟图像全局照度的扩散算子注重保持图像陡峭的边界, 而对于图像中的一些纹理细节则不需保持。基于此, 需要定义一个合适的算子来度量图像中各点的边界性 (edge degree, ED), 并设计一个以边界性测度作为影响扩散速率重要因素的各向异性扩散方案。

结合路面图像分块均匀的特点, 用 P 表示图像中各点的边界性, 定义 (x, y) 的 ED 为

$$P(x, y) = \left| \frac{1}{(2d + 1)^2} \sum_{n=y-d}^{n=y+d} \sum_{m=x-d}^{m=x+d} (I(m, n) - M(x, y))^2 \right|^{1/2} \quad (12)$$

式中, d 表示分析窗口尺度, $M(x, y)$ 表示局部均值, 其计算公式为

$$M(x, y) = \frac{1}{(2d + 1)^2} \sum_{n=y-d}^{n=y+d} \sum_{m=x-d}^{m=x+d} I(m, n) \quad (13)$$

由于过大的 d 导致运算量增大, 而过小的 d 不能准确反映图像的局部特性, 一般取 $d = 4$ 。 P 实质上反映了图像局部方差大小。在区域的内部, 灰度变化平缓, $P(x, y)$ 值较小; 当分析窗口跨越边界时, 灰度变化强烈, $P(x, y)$ 取得较大值, 从而 P 的表达式合理。但在图像噪声较强的区域, P 也可能取得较大值。因此在计算 P 之前, 对图像进行一次小尺度的高斯平滑, 以消除噪声带来的不良影响, 同时也能增强方程解的适定性^[9]。

4.3 基于 ED 的各向异性扩散

用 I, L 分别表示原图像和照度图像, 定义求 L 的各向异性扩散方程为

$$\begin{cases} \frac{\partial L}{\partial t} = \nabla \cdot [g(p, |\nabla L|) \nabla L] \\ L|_{t=0} = I \end{cases} \quad (14)$$

式中,

$$g(p, |\nabla L|) = \exp \left\{ - \left| \frac{\beta \cdot |\nabla L|}{k^2} \right| \right\} \quad (15)$$

k 为式 (11) 定义明暗对比强度。从扩散函数 g 的表达式可以看出, g 关于 P 和 $|\nabla L|$ 均单调递减, 且 P 成为影响扩散速率的主要因素。在区域内部, 灰度变化平缓, P 较小, g 较大, 扩散速率较快; 在图像的边界附近, P 较大, g 较小, 扩散速率较慢, 有利于保持图像的边界。同时与 Perona-Malik 方程定义的扩散函数相比, 梯度对扩散速率的影响被大大削弱 (由平方项变为一次项), 图像中的一些细节可以很好地被平滑, 符合全局照度的特点。同时考虑到当对比强度较大, 即使分析窗口只有一小部分跨越边界, P 也有可能取得较大值, 导致扩散减速过早, k^{-2} 项可以对这种情况起到自适应调整的作用, 使扩散速率保持一个理想速率。同时该扩散函数不需再进行梯度门限等参数的选择, 具有很好的通用性。

综上所述, 本文设计的各向异性 Retinex 算法框

图如图 2 所示。

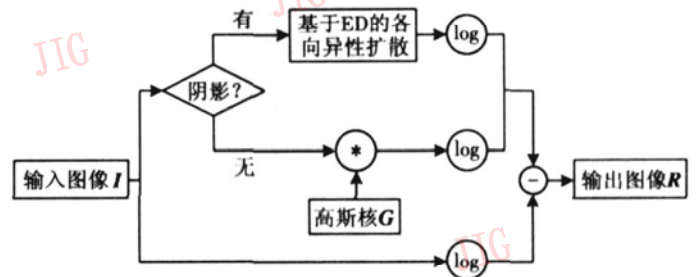


图 2 ADCSR 算法结构框图

Fig 2 The framework of ADCSR

5 实验结果及分析

为验证 ADCSR 算法的有效性, 采用 N-1 型智能路面检测车实际采集的路面图像, 在 Pentium IV 2.8GHZ 微机, Windows 2000 操作系统及 Matlab6.5 软件平台下进行了仿真实验, 并与 CSR 的实验结果进行对比。ADCSR 算法中采用 Perona-Malik 离散迭代格式^[7], 迭代 50 次; CSR 算法中采用高斯核尺度为 60。

两组实验的实验结果如图 3 所示。原始图像大小为 380×370 (图 3(a)), 且含有一块较大的阴影。CSR 算法所得的照度图像在阴影边界处被模糊 (图 3(b)), 导致输出图像中阴影边界处存在明显的光晕, 细节被模糊 (图 3(c)), 采用 ADCSR 算法所得照度图像中边界保持良好 (图 3(d)) 结果图像中光晕现象消除, 边界处的细节保留完好 (图 3(e))。

在算法效率上, ADCSR 算法在计算照度图像时需要进行方程的离散迭代求解, 而且为防止图像失真, 迭代次数不能太少 (≥ 50), 因此在时间耗费上要略多于 CSR 算法, 但优于模糊 Retinex 算法^[5]。其中, CSR 算法中的卷积运算采用快速傅里叶变换 (FFT) 进行了优化。采用模糊 Retinex 算法、CSR 算法和 ADCSR 算法处理图像 1 (第 1 组实验原始图像)、图像 2 (第 2 组实验原始图像) 的平均时间如表 1 所示。

表 1 不同算法的效率对比

Tab 1 Efficiency comparison for different algorithms 单位: s

	图像 1	图像 2
模糊 Retinex	97.73	95.56
CSR	11.14	12.07
ADCSR	17.89	17.47

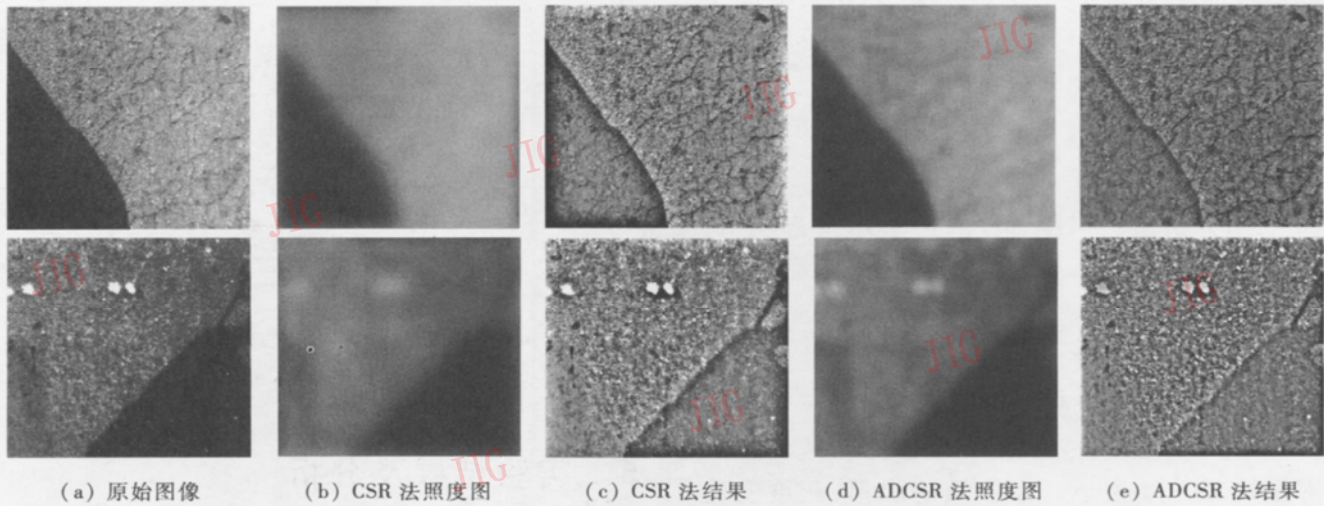


图 3 ADCSR 算法和 CSR 算法的结果对比

Fig. 3 Comparison of the results between ADCSR and CSR

6 结 论

从以上实验结果中可以看出, ADCSR 算法在保持图像边界细节、平衡图像的整体灰度等方面都优于 CSR 算法。在运算速度上, 由于 ADCSR 算法需要进行迭代运算, 在效率上略有下降, 但由于路面图像的处理往往是离线进行, ADCSR 算法的效率仍然令人满意。与传统的灰度校正算法相比, ADCSR 算法不需对阴影区域进行定位, 不需过多进行人为干预, 需要设置的参数少, 因而具有很好的通用性。

采用 ADCSR 算法消除阴影后, 阴影区域和非阴影区域并不协调, 这是由于阴影使得图像细节信息退化严重, 导致阴影区域的纹理细节不如非阴影区域的丰富。如何根据正常路面的纹理特征, 采用合适的图像修补算法, 修补阴影区域丢失的纹理细节, 使图像整体更加协调、一致, 是下一阶段需要研究的一个课题。

ADCSR 算法中原图像和照度图像均被变换到 \log 域, 有效地增强了图像中低灰度区域的细节, 如图 3(e) 所示, 原图像中模糊的裂缝能清晰地体现出来。但是, \log 变换使得图像中的噪声也得到了增强, 而且由于参数选择不恰当, 可能导致整体图像的过增强, 导致图像失真。对输出图像进行合适的非线性变换, 以抑制噪声的增强以及过增强现象的发生, 也将是下阶段的一个研究课题。

参考文献 (References)

1 Edw in H. The Retinex theory of color vision [J]. Scientific American, 1977, 237(6): 108~ 129

- 2 Jobson D J, Rahman Z. A multiscale Retinex for bridging the gap between color images and the human observation of scenes [J]. IEEE Transactions on Image Processing, 1997, 6(7): 965~ 976
- 3 Finlayson G D, Hordley S D, Drew M S. Removing shadows from images using retinex [A]. In: Proceedings of IS&T / SID Tenth Color Imaging Conference: Color Science, Systems and Applications [C], Scottsdale, Arizona, 2002, 73~ 79
- 4 Tang Liang, Xie Weixin. Removing shadows from urban aerial images based on fuzzy Retinex [J]. Acta Electronica Sinica, 2005, 33(3): 500~ 503 [唐亮, 谢维信. 城市航空影像中基于模糊 Retinex 的阴影消除 [J]. 电子学报, 2005, 33(3): 500~ 503]
- 5 Koenderink J. The structure of images [J]. Biological Cybernetics, 1984, 50(3): 363~ 370
- 6 Hummel A. Representations based on zero-crossings in scale-space [A]. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition [C], Miami Beach, FL, USA, 1986, 204~ 209
- 7 Perona P, Malik J. Scale-space and edge detection using anisotropic diffusion [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1990, 12(7): 629~ 639
- 8 Catte F, Lios P L, Morel J M, et al. Image selective smoothing and edge detection by nonlinear diffusion [J]. SIAM Journal of Numerical Analysis, 1992, 29(2): 182~ 193
- 9 Wang Y J, Zhang Liang-pei. An improved algorithm of an isotropic diffusion smoothing filter [J]. Journal of Image and Graphics, 2006, 11(2): 210~ 216 [王毅, 张良培. 各向异性扩散平滑滤波的改进算法 [J]. 中国图象图形学报, 2006, 11(2): 210~ 216]
- 10 Qian Weixin, Liu Ruigen. The anisotropic diffusion methods based on the directions of the image feature [J]. Journal of Image and Graphics, 2006, 11(6): 818~ 822 [钱伟新, 刘瑞根. 基于图像特征方向的各向异性扩散滤波方法 [J]. 中国图象图形学报, 2006, 11(6): 818~ 822]
- 11 Otsu N. A threshold selection method from a gray level histogram [J]. IEEE Transactions on Systems, Man, and Cybernetics, 1979, 9(1): 62~ 66

基于内在表示的空间曲线去噪算法

任庆军 王树艳 蔡增霞

(临沂师范学院数学系, 临沂 276000)

摘要 介绍了一种对空间离散曲线进行去噪声的新的方法。不同于一般的顶点表示方法, 这里对离散曲线采用一种新的内在表示方法, 用曲线的内在几何量——边长和边与 x 轴正方向、 z 轴正方向的夹角来表示点的位置。利用双边滤波的思想, 对边与 x 轴正方向、 z 轴正方向夹角进行去噪声。然后以去噪后的角度作为一个约束条件, 构造目标函数来反求曲线的顶点。该方法的优点主要在于解决了去噪声方法中一般都会产生的收缩问题, 并且能够保持基本的特征。

关键词 去噪 双边滤波 内在表示 曲线收缩

中图分类号: TP391 **文献标识码**: A **文章编号**: 1006-8961(2008)02-0269-08

An Intrinsic 3D Discrete Curve Smoothing Algorithm

REN Qing-jun, WANG Shu-yan, CAI Zeng-xia

(Department of Mathematics, Linyi Normal University, Linyi 276000)

Abstract In this paper, we propose a new method for 3D discrete curve smoothing by adopting an intrinsic representation for discrete curves. With this representation, a piecewise 3D curve is defined by lengths of edges, angles between edges and the positive x -axis and angles between edges and the positive z -axis. For a noisy curve, we first filter the two angle sequences of the curve by bilateral filtering method. Then we obtain the smoothed vertexes of the curve by solving an objective function under the constraint of the two angle sequences. By this algorithm, not only can main features of the original curve be preserved well, but also the smoothed curve no longer suffers from shrinkage.

Keywords smoothing, bilateral filtering, intrinsic representation, curve shrinkage

1 引言

本文所指的曲线均为空间离散多边形曲线(把一系列离散点用直线段连接起来形成的曲线), 一般我们处理的曲线都是反映物体形状信息的曲线, 比如物体轮廓线、人体切片轮廓线等。这些数据是由坐标测量机、激光扫描仪等 3 维数字仪测量得到的, 但是由于测量和计算过程中经常存在误差, 总会产生一些不可避免的数据扰动, 称之为噪声。这些数据首先需要经过去噪处理, 才能进行后续的工作, 如模式识别、边界检测、反求工程, 以及曲线信息(周

长、面积等)求解等。

曲线去噪声有很多种方法, 在早期的去噪算法中, 比较普遍的方法是将信号处理中的去噪算法应用到曲线去噪中, 其主要步骤是用傅里叶变换将空域转换到频域上, 噪声对应到频域上就是高频信号, 去噪声即是将高频信号去除。比如 Gauss 滤波^[1-3], 就是将信号与 Gauss 函数 ($g(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2}$) 进行卷积, 然后将结果转化到空域上就得到了滤波后的顶点。但是 Gauss 滤波有一个共知的缺点——产生收缩现象, 并且不能有效地保持尖锐特征。

差分光顺也是曲线去噪普遍使用的一种方法,

收稿日期: 2006-02-27; 改回日期: 2006-07-17

第一作者简介: 任庆军(1962~), 男, 教授。1984 年于曲阜师范大学获理学学士学位。主要研究方向为离散数学、计算机图形学。

E-mail: qjr288@126.com

Choi, Jerard 在 1998 年提出了进行二阶、四阶差分光顺算法^[4], 而 Cho 和 Cho 进一步证明了偶数阶差分的性质, 由原始离散曲线用添加顶点的方法拓展出两种新的曲线——反射填补曲线 (reflection-padded curve) 以及镜面延伸曲线 (mirror-extended curve)^[5], 这两种曲线分别是在原始曲线的基础上, 用不同的方法将曲线进行延伸而得到, 并证明了反射填补曲线的偶数阶差分光顺就等价于离散傅里叶变换滤波。

Li 等人利用点列离散曲率以及离散曲率的一阶差分检测出性质不好的点, 并根据曲率及曲率差分的变化将这些顶点分为两类: SBP (single bad point) 和 MBP (multiple bad point), 然后根据点的曲率性质不同采用不同的方法, 从而决定顶点去噪时应该移动的方向和距离^[6]。

曲线去噪还可以直接对顶点进行, Jones 和 Fleishman 等人利用 Tomasi, Manduchi 提出的图像双边滤波方法^[7], 在当前顶点定义一个局部坐标系, 利用双边滤波对高度函数进行去噪, 从而得到顶点的位置^[8,9]。这两种方法基本解决了保持尖锐特征的问题, 但是收缩还是它的一个缺点。

所谓收缩, 对图像来讲, 收缩就是指灰度趋于一致, 对于曲线来讲, 收缩是指曲线随着迭代次数的增加逐渐趋向于一个点。收缩问题是去噪中的一个相当重要的问题, 比如在反求工程中, 由曲线族反求曲面, 如果有曲线产生收缩, 就会引起曲面变形。为了解决收缩问题, 近年来很多人都做了这方面的研究工作。

Low 提出了一种解决曲线去噪收缩问题的修正算法^[10], 利用高斯滤波去噪后, 估计去噪后曲线的收缩量, 最后将估计的收缩量加到去噪后的曲线上, 从而人为地控制收缩。Oliensis 在 Gaussian 滤波的基础上, 重新构造了一个低通滤波函数^[11], 首先给定两个频率常量 K_L 和 K_U ($K_L < K_U$), 定义函数在低于 K_L 的范围内函数值为 1, 高于 K_U 的范围内函数值为 0 使得低频信号得到保持, 并人为地将高频信息去除, 这种方法虽然可以控制收缩, 但是却不能保持曲线的尖锐特征。Taubin 提出了一个没有明显收缩的方法^[12,13], 实际上是两次 Gaussian 滤波叠加, 只是两次滤波的参数选择不同, 对参数加以控制可以产生收缩或者扩张的效果, 两次滤波交替进行, 从而控制收缩。但是它产生了相当于低通滤波的效果, 效果仍然不够理想。

本文给出了一种空间离散曲线新的内在表示方法, 并利用这种内在表示方法, 在 Tomasi, Manduchi 提出的双边滤波的基础上^[7], 提出了一种新的去噪方法, 解决了去噪过程中的收缩问题。

2 双边滤波

Tomasi, Manduchi 提出的保形图像去噪方法——双边滤波 (bilateral filtering)^[7], 是将当前点的灰度值用周围点的灰度值的加权平均来代替, 权因子不只和两点之间的几何距离有关, 更和它们的灰度值差异有关, 所以称之为双边滤波。

一般来讲, 对连续函数 $y = f(x)$ 进行双边滤波的公式如下:

$$F(x) = k^{-1}(x) \int_{-}^{+} f(\xi) c(\xi, x) \times s(f(\xi), f(x)) d\xi \quad (1)$$

式中, $F(x)$ 为滤波后的输出函数,

$$c(\xi, x) = e^{-\|\xi-x\|^2/2\sigma_c^2}$$

$$s(f(\xi), f(x)) = e^{-\|f(\xi)-f(x)\|^2/2\sigma_s^2}$$

$$k(x) = \int_{-}^{+} c(\xi, x) s(f(\xi), f(x)) d\xi$$

式 (1) 的离散表达式为

$$F(x) = \frac{\sum_{j=-\infty}^{+\infty} c(\xi_j, x) s(f(\xi_j), f(x)) \times f(\xi_j) \Delta\xi}{\sum_{j=-\infty}^{+\infty} c(\xi_j, x) s(f(\xi_j), f(x)) \times \Delta\xi} = \frac{\sum_{j=-\infty}^{+\infty} W_c(\|\xi_j - x\|) W_s(\|f(\xi_j) - f(x)\|) \times f(\xi_j) \Delta\xi}{\sum_{j=-\infty}^{+\infty} W_c(\|\xi_j - x\|) \times W_s(\|f(\xi_j) - f(x)\|) \Delta\xi}$$

这里, $-\infty < \dots < \xi_n < \xi_{n+1} < \dots < \xi_1 < \xi_0 < \xi < \dots < \xi_{-1} < \xi < \dots < +\infty$, $\Delta\xi = \xi_j - \xi_{j-1}$, W_c, W_s 都是高斯函数: $W_c(x) = e^{-x^2/2\sigma_c^2}$, $W_s(x) = e^{-x^2/2\sigma_s^2}$, 其中, σ_c, σ_s 为常数。不妨设 $\Delta\xi$ 为常数, 可得

$$F(x) = \frac{\sum_{j=-\infty}^{+\infty} W_c(\|\xi_j - x\|) W_s(\|f(\xi_j) - f(x)\|) \times f(\xi_j)}{\sum_{j=-\infty}^{+\infty} W_c(\|\xi_j - x\|) W_s(\|f(\xi_j) - f(x)\|)} \quad (2)$$

以图像去噪为例来解释一下双边滤波。若将式 (2) 用到图像去噪中, 只需将函数 $f(x)$ 换成灰度函数, σ_c, σ_s 对于去噪效果有较大的影响。 σ_c 越大, 权因子中距离产生的影响就越大, 而 σ_s 越大, 相邻

像素的灰度差异对权因子的影响就会加强, 反之亦然。这种方法将灰度的差异引入权因子, 所以能较好地保持图像边界。当然直接对灰度进行加权平均得到去噪后点的灰度值一般都会产生收缩, 也就是灰度值趋于一致。双边滤波也不例外。

Jones和 Fleishman等人把这种方法应用到网格上, 得到了比较好的去噪声效果^[2 3]。主要思想是将所有的顶点沿着法线方向移动一段距离, 使得曲线或者曲面看起来比较平滑。它们的主要优点在于能够保持尖锐特征, 缺点仍然是会产生收缩。这种方法之所以会产生收缩, 以文献[3]为例, 它是对一阶邻域内点到曲线在当前顶点的切平面的高度函数进行双边滤波, 得到当前顶点沿着法向移动的距离, 由此求得新的顶点。鉴于双边滤波的收缩问题, 应用到网格和曲线上也会产生收缩。既然对高度函数进行滤波会产生收缩, 我们想到可以对曲线的走向进行滤波, 即对曲线内在几何量——边与 x 轴正向的夹角进行双边滤波, 这是因为曲线的噪声或者光滑程度在一定程度上可以由这些角度来反映。为此给出曲线的一种新的内在表示方法。

3 曲线的内在表示

对于空间曲线, 由以往的知识知道, 球坐标表示也能唯一地表示边向量, 由此可以得到空间曲线的内在表示。

一般情况下的空间曲线都是用 3 维点列给出的, 空间曲线的内在表示是用边长和角度表示的, 边长是指曲线边向量的长度, 而角度有两个, 分别是球坐标表示中的两个角度——一个是边向量在 xy 平面上的投影与 x 轴正方向的夹角, 另一个是边向量与 z 轴正方向的夹角, 如图 1 所示。

记 $\{p_i = (x_i, y_i, z_i); i = 0, 1, \dots, n\}$ 为空间曲线的顶点, $\{V_i; i = 0, 1, \dots, n-1\}$ 为边向量, $V_i = p_{i+1} - p_i$ 表示由顶点 p_i 指向 p_{i+1} 的向量, $\{V'_i; i = 0, 1, \dots, n-1\}$ 表示边向量在 xy 平面上的投影向量。 $\{e_i; i = 0, 1, \dots, n-1\}$ 表示多边形的边长, 且 $e_i = \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2 + (z_{i+1} - z_i)^2}$, $\{\theta_i; i = 0, 1, \dots, n-1\}$ 表示边向量在 xy 平面上的投影向量与 x 轴正方向的夹角, $\{\varphi_i; i = 0, 1, \dots, n-1\}$ 表示边向量与 z 轴正方向的夹角, 那么可以得到如下的表达式:

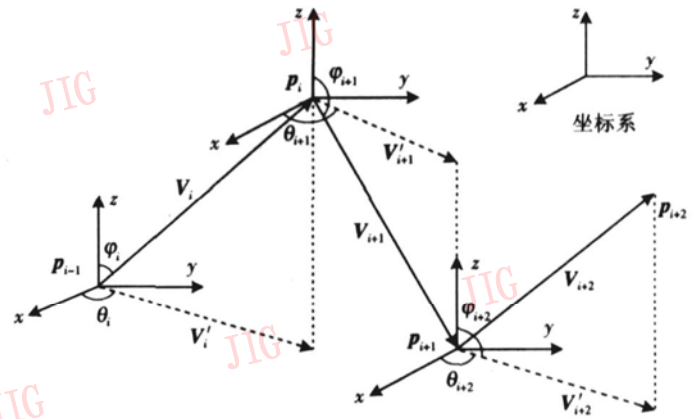


图 1 空间离散曲线的内在表示

Fig 1 Intrinsic representation of 3D discrete curve

当向量 V'_i 落在第一、四象限时

$$\theta_i = \arctan \left| \frac{y_{i+1} - y_i}{x_{i+1} - x_i} \right|$$

当向量 V'_i 落在第二象限时

$$\theta_i = \arctan \left| \frac{y_{i+1} - y_i}{x_{i+1} - x_i} \right| + \pi \tag{3}$$

当向量 V'_i 落在第三象限时

$$\theta_i = \arctan \left| \frac{y_{i+1} - y_i}{x_{i+1} - x_i} \right| - \pi$$

式中, $i = 0, 1, \dots, n-1$ 。

$$\varphi_i = \arccos \left| \frac{z_{i+1} - z_i}{e_i} \right| \tag{4}$$

式中, $i = 0, 1, \dots, n-1$ 。

由上面的表达式可以得到: $\theta_i \in [-\pi, \pi]$, $\varphi_i \in [0, \pi]$ ($i = 0, 1, \dots, n-1$), $\{(e_i, \theta_i, \varphi_i); i = 1, 2, \dots, n-1\}$ 就称为空间曲线的内在表示。那么, 由向量的球坐标表示的唯一性可以知道, 在给定初始点 p_0 的条件下, 这种空间曲线的内在表示和它的点列表示是一一对应的。

4 角度的双边滤波

空间曲线内在表示的 3 个几何量中, 两个角度量决定着曲线的整体走向, 因此, 这里采用对角度滤波的方式来达到曲线光滑的效果, 所谓对角度进行双边滤波, 是指将式 (2) 中的灰度值分别用内在表示中的两个角度代替, 从而实现对这两个角度的双边去噪。

设对角度 θ_i, φ_i 进行双边滤波得到的新的角度为 θ'_i, φ'_i , 表达式如下:

$$\theta'_i = \frac{W_c(\|p_i - p_{i-1}\|)W_s(|\theta_i - \theta_{i-1}|)\theta_{i-1} + W_c(\|p_i - p_{i+1}\|)W_s(|\theta_i - \theta_{i+1}|)\theta_{i+1} + \theta_i}{W_c(\|p_i - p_{i-1}\|)W_s(|\theta_i - \theta_{i-1}|) + W_c(\|p_i - p_{i+1}\|)W_s(|\theta_i - \theta_{i+1}|) + 1}$$

$$= \frac{W_c(e_{i-1})W_s(|\theta_i - \theta_{i-1}|)\theta_{i-1} + W_c(e_i)W_s(|\theta_i - \theta_{i+1}|)\theta_{i+1} + \theta_i}{W_c(e_{i-1}) \times W_s(|\theta_i - \theta_{i-1}|) + W_c(e_i)W_s(|\theta_i - \theta_{i+1}|) + 1} \tag{5}$$

$$\varphi'_i = \frac{W'_c(\|p_i - p_{i-1}\|)W'_s(|\varphi_i - \varphi_{i-1}|)\varphi_{i-1} + W'_c(\|p_i - p_{i+1}\|)W'_s(|\varphi_i - \varphi_{i+1}|)\varphi_{i+1} + \varphi_i}{W'_c(\|p_i - p_{i-1}\|)W'_s(|\varphi_i - \varphi_{i-1}|) + W'_c(\|p_i - p_{i+1}\|)W'_s(|\varphi_i - \varphi_{i+1}|) + 1}$$

$$= \frac{W'_c(e_{i-1})W'_s(|\varphi_i - \varphi_{i-1}|)\varphi_{i-1} + W'_c(e_i)W'_s(|\varphi_i - \varphi_{i+1}|)\varphi_{i+1} + \varphi_i}{W'_c(e_{i-1}) \times W'_s(|\varphi_i - \varphi_{i-1}|) + W'_c(e_i)W'_s(|\varphi_i - \varphi_{i+1}|) + 1} \tag{6}$$

式中, $W'_c(x) = e^{-x^2/2\sigma_c'^2}$, $W'_s(x) = e^{-x^2/2\sigma_s'^2}$, σ_c' , σ_s' 为常数。

σ_c , σ_s , σ_c' , σ_s' 分别是它们的自由参数。读者可以根据自己的需要适当选取 σ_c , σ_s , σ_c' , σ_s' 。式 (5) 式 (6) 只是一阶邻域的离散展开, 所谓一阶邻域, 就是直接与当前顶点相连的顶点。实际上, θ'_i 就是 θ_{i-1} , θ_i , θ_{i+1} 的线性组合, φ'_i 就是 φ_{i-1} , φ_i , φ_{i+1} 的线性组合。

在式 (5) 中还有一个需要注意的问题, 就是角度的范围。式 (3) 中给出的角度定义在 $[-\pi, \pi]$ 上, 那么式 (5) 中的角度之差就有可能大于 π 。两个向量的角度差有两个: θ 或 $2\pi - \theta$ ($0 \leq \theta < \pi$), 在这里统一取 θ ($0 \leq \theta < \pi$)。也就是需要改变 θ_{i-1} , θ_{i+1} 的值 (加或者减 2π 的整数倍), 使得 $|\theta_i - \theta_{i-1}|$, $|\theta_i - \theta_{i+1}|$ 的值落在 $[0, \pi)$ 之内。两个角度差中选择落在 $[0, \pi)$ 之间的那个。经过这样的改变, 最后得到的 θ'_i 就不一定会落在 $[-\pi, \pi]$ 内, 所以最后为了方便统一处理, 再将 θ'_i 的值转化到 $[-\pi, \pi]$ 的范围之内 (经过加减 2π 的整数倍)。

对于式 (6) 中的 $|\varphi_i - \varphi_{i-1}|$, $|\varphi_i - \varphi_{i+1}|$ 则不同, 这是因为 $\varphi_i \in [0, \pi]$, 所以可以得到 $|\varphi_i - \varphi_{i-1}|$, $|\varphi_i - \varphi_{i+1}| \in [0, \pi)$, 因此不需要进行这方面的处理。

不过对与 x 轴夹角进行去噪得到的结果会随着坐标系选择的不同而不同, 因为坐标系变化会导致夹角也随着变化, 从而使得去噪结果发生变化, 当然这种变化很小, 不会产生很大的影响。由式 (5) 式 (6) 得到光滑的角度之后, 并不能直接得到去噪声之后的顶点, 因为仅仅角度并不能确定顶点的位置。

5 反求顶点

对角度完成双边滤波之后, 直接的办法是再对边长进行双边滤波, 因为由边长和角度可以唯一确定一条曲线, 这时候就可以由边长和夹角反求曲线

顶点。曲线的边长经过滤波只是变得平均, 总体长度上不会发生很大的变化, 而滤波后的角度使得曲线变得光滑, 所以反求得到的曲线会比原始曲线长很多。以平面曲线为例, 一条带有噪声的直线段经过这样的去噪后会变得平直, 但是因为边长总体来说并没有发生很大的变化, 去噪相当于把带有噪声的线段拉直, 使得去噪后的曲线显得长了很多。但是因为是对边长和角度单独进行滤波, 没有任何其他的约束条件, 这样得到的曲线不可避免存在变形, 很大地偏离原来的噪声曲线。图 2 所示是对一条总体来说比较平直的平面噪声曲线去噪声。可以看到去噪之后的曲线比原始曲线长了很多, 相当于拉直了, 并且略微有变形。如果曲线是闭合曲线, 那么用这种方法滤波之后的曲线包围的面积就会变大。

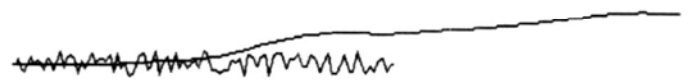


图 2 线段噪声曲线以及用边长角度直接滤波反求顶点得到的去噪曲线

Fig 2 Noisy line segment and curve denoised by filtering edges and angles

为了解决曲线的变形和偏离, 以去噪之后的角度作为一个条件, 构造一个目标函数, 求得使这个目标函数达到最小的顶点值, 从而达到对曲线去噪的目的。

设 $\{p_i = (x_i, y_i, z_i); i = 0, 1, \dots, n\}$ 表示去噪后得到的新顶点, $\{e_i; i = 0, 1, \dots, n-1\}$ 表示由新的顶点得到的边的长度, $\{\theta_i; i = 0, 1, \dots, n-1\}$ 表示过去噪之后的边向量在 xy 平面上的投影向量与 x 轴正半轴的夹角, $\{\varphi_i; i = 0, 1, \dots, n-1\}$ 表示去噪之后的边向量与 z 轴正方向的夹角。

希望去噪之后的曲线能够光滑 (角度变化比较平缓), 并且不会与原始曲线偏离太远 (去噪后的顶点与去噪之前的原始顶点距离尽可能近), 从而产生较小的变形, 综合以上两条, 去噪之后的空间曲线应该满足以下 3 个条件:

- (1) $\hat{\theta}_i$ 与 θ'_i 之差尽可能小;
- (2) φ_i 与 φ'_i 之差尽可能小;
- (3) \hat{p}_i 与 p_i 能尽可能比较接近。

为了满足这 3 个条件, 构造下面的目标函数:

$$\begin{aligned}
 f(\hat{x}_i, \hat{y}_i, \hat{z}_i) &= \|\hat{p}_i - p_i\|^2 + k|\hat{\theta}_{i-1}(\hat{p}_i) - \theta'_{i-1}|^2 + \\
 &\quad m|\varphi_{i-1}(\hat{p}_i) - \varphi'_{i-1}|^2 \\
 &= |\hat{x}_i - x_i|^2 + |\hat{y}_i - y_i|^2 + |\hat{z}_i - z_i|^2 + \\
 &\quad k|\hat{\theta}_{i-1}(\hat{p}_i) - \theta'_{i-1}|^2 + m|\varphi_{i-1}(\hat{p}_i) - \varphi'_{i-1}|^2
 \end{aligned} \tag{7}$$

这里, $k > 0, m > 0$ 均是参数, 可以自行选取。因为是逐个反求顶点, 所以 $\hat{x}_{i-1}, \hat{y}_{i-1}, \hat{z}_{i-1}$ 是已知量, 因此 $\hat{\theta}_{i-1}, \varphi_{i-1}$ 都只是 $\hat{x}_i, \hat{y}_i, \hat{z}_i$ 的函数, 所以式 (7) 也只含有 $\hat{x}_i, \hat{y}_i, \hat{z}_i$ 3 个变量。使得式 (7) 取得最小值的 $\hat{x}_i, \hat{y}_i, \hat{z}_i$ 值就是所求得的新顶点 \hat{p}_i 。只不过, $|\hat{\theta}_i - \theta'_i|^2$ 与 $|\varphi_i - \varphi'_i|^2$ 的表达式比较复杂, 不利于求导运算, 更不利于求解方程组, 所以利用近似逼近的方法来替代 $|\hat{\theta}_i - \theta'_i|^2$ 。如图 3 所示。

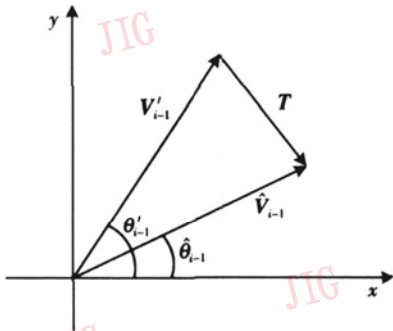


图 3 角度之差 $|\theta_i - \theta'_i|^2$ 示意图

Fig 3 Sketch map of difference of angles $|\theta_i - \theta'_i|^2$

记 $V'_{i-1} = (\cos\theta'_{i-1}, \sin\theta'_{i-1}, 0)$ 为 xy 平面内与 x 轴正方向夹角为 θ'_{i-1} 的单位向量, 设 V_{i-1} 是去噪后的边 $\hat{p}_{i-1}\hat{p}_i$ 在 xy 平面上的投影向量所对应的单位向量, 则 $V_{i-1} = \left| \frac{(\hat{x}_i - \hat{x}_{i-1})}{l_{i-1}}, \frac{(\hat{y}_i - \hat{y}_{i-1})}{l_{i-1}}, 0 \right|$, 其中,

$l_{i-1} = \sqrt{(\hat{x}_{i-1} - \hat{x}_i)^2 + (\hat{y}_{i-1} - \hat{y}_i)^2}$ 为 $\hat{p}_{i-1}\hat{p}_i$ 在 xy 平面上的投影向量的长度。记 $\hat{l}_{i-1} = \sqrt{(\hat{x}_{i-1} - x_i)^2 + (\hat{y}_{i-1} - y_i)^2}$, 那么为简化计算, 可用 \hat{l}_{i-1} 近似代替 l_{i-1} , 则 V_{i-1} 可近似表示为 $\left| \frac{(\hat{x}_i - \hat{x}_{i-1})}{\hat{l}_{i-1}}, \frac{(\hat{y}_i - \hat{y}_{i-1})}{\hat{l}_{i-1}}, 0 \right|$ 。

记 $T = V_{i-1} - V'_{i-1}$ (图 3)。因为去噪前后的角度相差不是很大, 所以有

$$\begin{aligned}
 \|T\| &\approx \|V'_{i-1}\| \sin(\theta'_{i-1} - \hat{\theta}_{i-1}) = \sin(\theta'_{i-1} - \hat{\theta}_{i-1}) \\
 &\approx |\theta'_{i-1} - \hat{\theta}_{i-1}|
 \end{aligned} \tag{8}$$

由式 (8) 得:

$$\begin{aligned}
 \|\theta'_{i-1} - \hat{\theta}_{i-1}\|^2 &\approx \left| \frac{(\hat{x}_i - \hat{x}_{i-1})}{\hat{l}_{i-1}} - \cos\theta'_{i-1} \right|^2 + \\
 &\quad \left| \frac{(\hat{y}_i - \hat{y}_{i-1})}{\hat{l}_{i-1}} - \sin\theta'_{i-1} \right|^2
 \end{aligned} \tag{9}$$

对于 $|\varphi_i - \varphi'_i|^2$, 可以用其他的量来代替, 如图 4 所示。

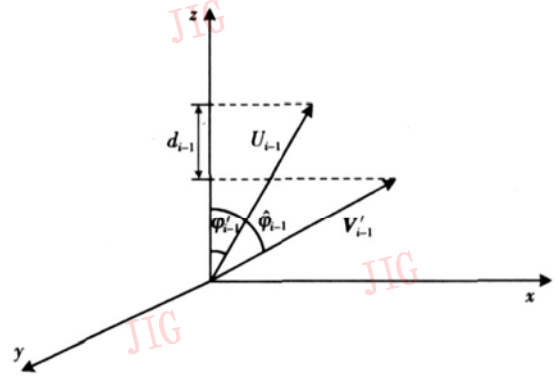


图 4 角度之差 $|\varphi_i - \varphi'_i|^2$ 示意图

Fig 4 Sketch map of difference of angles $|\varphi_i - \varphi'_i|^2$

记 $U_{i-1} = (\cos\theta'_{i-1} \sin\varphi'_{i-1}, \sin\theta'_{i-1} \sin\varphi'_{i-1}, \cos\varphi'_{i-1})$ 为与 z 轴正方向夹角为 φ'_i , 在 xy 平面投影向量与 x 轴正方向夹角为 θ'_{i-1} 的单位向量, 设 V_{i-1} 是去噪后边 $\hat{p}_{i-1}\hat{p}_i$ 所对应的单位向量, 则

$$V_{i-1} = \left| \frac{(\hat{x}_i - \hat{x}_{i-1})}{\hat{e}_{i-1}}, \frac{(\hat{y}_i - \hat{y}_{i-1})}{\hat{e}_{i-1}}, \frac{(\hat{z}_i - \hat{z}_{i-1})}{\hat{e}_{i-1}} \right|$$

记

$$\hat{L}_{i-1} = \sqrt{(x_i - \hat{x}_{i-1})^2 + (y_i - \hat{y}_{i-1})^2 + (z_i - \hat{z}_{i-1})^2}$$

则 \hat{L}_{i-1} 可以用来近似代替 \hat{e}_{i-1} , 也就是说

$$V_{i-1} \approx \left| \frac{(\hat{x}_i - \hat{x}_{i-1})}{\hat{L}_{i-1}}, \frac{(\hat{y}_i - \hat{y}_{i-1})}{\hat{L}_{i-1}}, \frac{(\hat{z}_i - \hat{z}_{i-1})}{\hat{L}_{i-1}} \right|$$

通过图 4 可以看出, 单位向量与 z 轴正方向的夹角 φ 跟向量在 z 轴上的投影——坐标分量的大小有直接的关系, z 坐标越大, φ 越大; 相反, z 坐标越小, φ 也越小。所以式 (7) 中的 $|\varphi_i - \varphi'_i|^2$ 可以用单位向量的 z 坐标之差来代替, 记

$$d_{i-1} = \frac{\hat{z}_i - \hat{z}_{i-1}}{\hat{L}_{i-1}} - \cos\varphi' \tag{10}$$

也就是说可以用 d_{i-1} 来代替目标函数中的 $|\varphi_i - \varphi'_i|^2$ 。

将式 (9) 式 (10) 代入式 (7) 可得:

$$\begin{aligned}
 f(\hat{x}_i, \hat{y}_i, \hat{z}_i) &= (\hat{x}_i - x_i)^2 + (\hat{y}_i - y_i)^2 + (\hat{z}_i - z_i)^2 + \\
 &\quad k \left| \frac{\hat{x}_i - \hat{x}_{i-1}}{\hat{l}_{i-1}} - \cos\theta_i \right|^2 + \left| \frac{\hat{y}_i - \hat{y}_{i-1}}{\hat{l}_{i-1}} - \sin\theta_i \right|^2 + \\
 &\quad m \left| \frac{\hat{z}_i - \hat{z}_{i-1}}{\hat{L}_{i-1}} - \cos\varphi' \right|^2
 \end{aligned}$$

$$m \left| \frac{\hat{z}_i - \hat{z}_{i-1}}{\hat{L}_{i-1}} - \cos\varphi_i \right|^2 \quad (11)$$

那么式 (11)取得最小值与式 (7)取得最小值是等价的, 也就是说使得式 (11)取得最小值的顶点也就是使得式 (7)取得最小值的顶点。

很明显, 式 (11)也是关于 $\hat{x}_i, \hat{y}_i, \hat{z}_i$ 的三元二次函数, 求导之后得到三元一次方程组

$$\begin{cases} \hat{x}_i - x_i + k \left| \frac{\hat{x}_i - \hat{x}_{i-1}}{\hat{L}_{i-1}} - \cos\theta_i \right| \cdot \frac{1}{\hat{L}_{i-1}} = 0 \\ \hat{y}_i - y_i + k \left| \frac{\hat{y}_i - \hat{y}_{i-1}}{\hat{L}_{i-1}} - \sin\theta_i \right| \cdot \frac{1}{\hat{L}_{i-1}} = 0 \\ \hat{z}_i - z_i + m \left| \frac{\hat{z}_i - \hat{z}_{i-1}}{\hat{L}_{i-1}} - \cos\varphi_i \right| \cdot \frac{1}{\hat{L}_{i-1}} = 0 \end{cases} \quad (12)$$

求解式 (12)就得到新的顶点

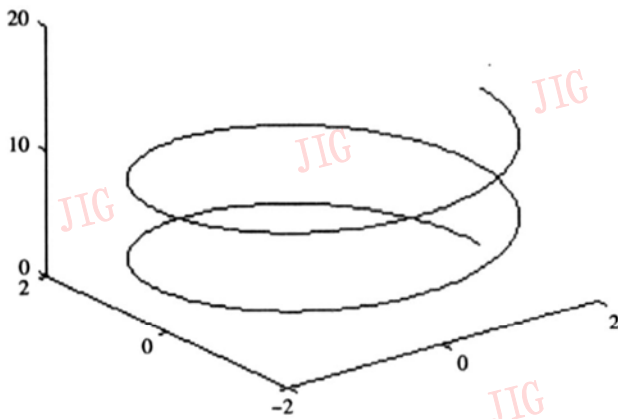
$$\hat{p}_i = (\hat{x}_i, \hat{y}_i, \hat{z}_i): \begin{cases} \hat{x}_i = \frac{\hat{L}_{i-1}^2 x_i + k \hat{x}_{i-1} + k \hat{L}_{i-1} \cos\theta_{i-1}}{k + \hat{L}_{i-1}^2} \\ \hat{y}_i = \frac{\hat{L}_{i-1}^2 y_i + k \hat{y}_{i-1} + k \hat{L}_{i-1} \sin\theta_{i-1}}{k + \hat{L}_{i-1}^2} \\ \hat{z}_i = \frac{\hat{L}_{i-1}^2 z_i + m \hat{z}_{i-1} + m \hat{L}_{i-1} \cos\varphi_{i-1}}{m + \hat{L}_{i-1}^2} \end{cases}$$

6 算法的实例

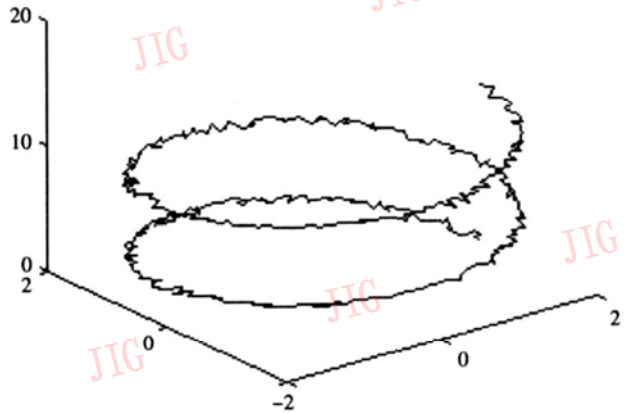
通过上面的分析, 可以看出, 本文算法主要分为 3 步进行: 首先将空间曲线的坐标表示转换成内在表示, 然后分别对两个角度进行双边滤波, 最后根据目标函数反求新的顶点序列。下面举例说明本文算法的实现效果。

本文算法中共有 6 个自由参数: 2 个高斯函数中的参数 $\sigma'_c, \sigma'_s, \sigma_c, \sigma_s$ 以及反求顶点时的参数 k, m , 在算法实现过程中需要对这些参数进行调节, 以便得到最佳的效果。

图 5(a)是一条离散的空间螺旋曲线, 由 419 个顶点组成, 图 5(b)是将图 5(a)加上随机噪声所得到的噪声螺旋曲线 (噪声是曲线法平面上的随机噪声), 也是滤波的初始顶点序列。图 6(a)是经过本文算法 5 次迭代之后得到的去噪曲线, 其中算法实现过程中参数取值分别为 $\sigma'_c = \sigma_c = 0.3, \sigma'_s = 5, \sigma_s = 10, k = 0.005, m = 0.015$ 。通过图 6 可以看到, 此算法对于一般的曲线去噪结果比较理想。



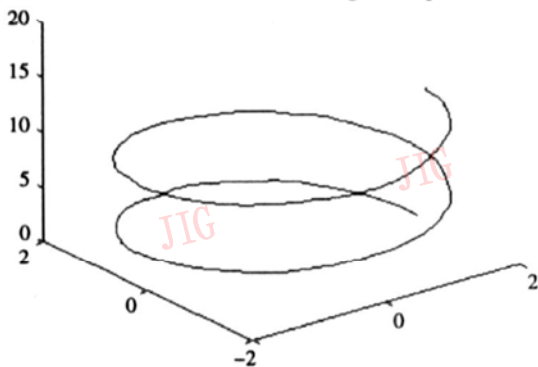
(a) 螺旋曲线的原始图形



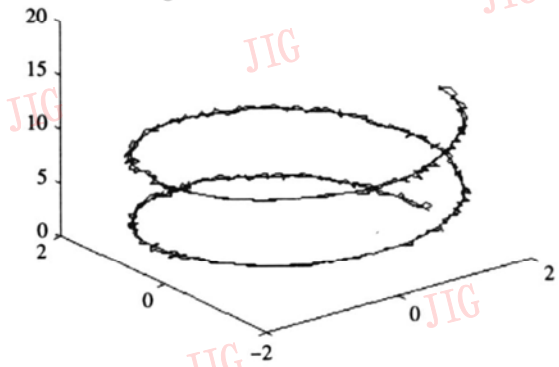
(b) 加噪声之后的图形

图 5 螺旋曲线加噪前后效果图

Fig 5 Spiral curve before and after noising



(a) 螺旋噪声曲线去噪之后



(b) 去噪前后图形叠加图

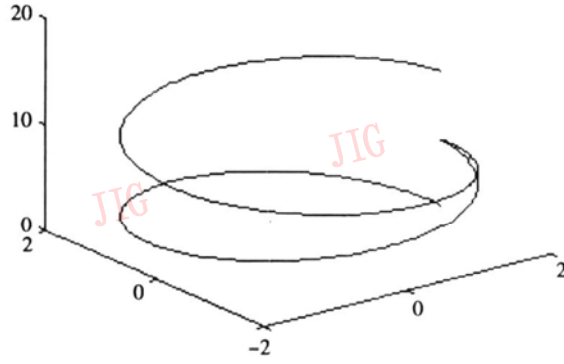
图 6 螺旋曲线去噪前后效果图

Fig 6 Spiral curve before and after denoising

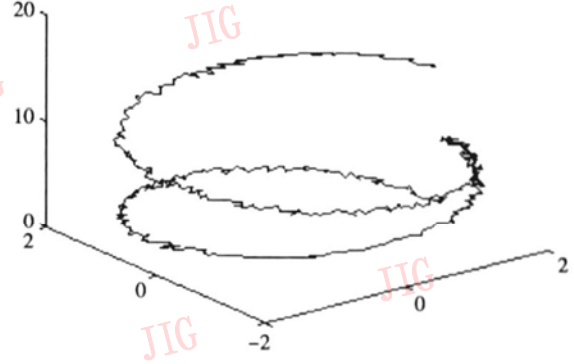
图 7 (a)是一条带有尖点的原始光滑空间曲线, 将它加上随机噪声之后得到了图 7 (b)中所显示的噪声曲线。对于这种带有尖点的空间曲线, 与平面离散曲线一样, 希望去噪之后的曲线能够保持尖锐特征 (尖点不被光滑掉), 并且希望得到的去噪曲线不会产生收缩, 但是以前的算法基本上都不能解决这个问题, 而本文算法基本上解决了这个问题。

图 8 (a)是图 7 (b)经过本文算法去噪之后得到的光滑曲线, 迭代次数是 5 次。参数的选取与图 5 (b)去噪时选择的参数相同, 也得到了比较好的结果。

图 9、图 10所示的是对长方体的部分边界线 (长方体的一半边界) 进行去噪的过程。图 9 (a)是长方体的原始边界, 加上随机噪声之后如图 9 (b)所示。利用本文算法进行去噪的时候要适当的对参数



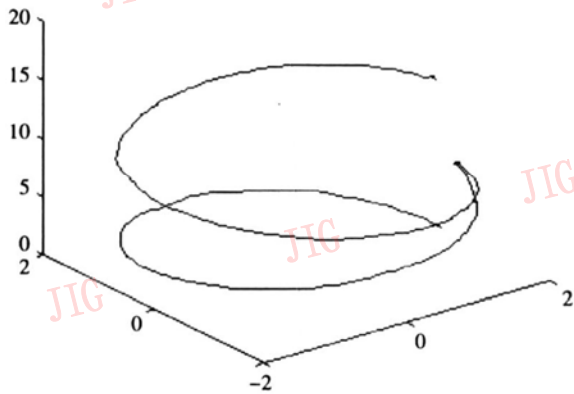
(a) 光滑的原始螺旋尖点图



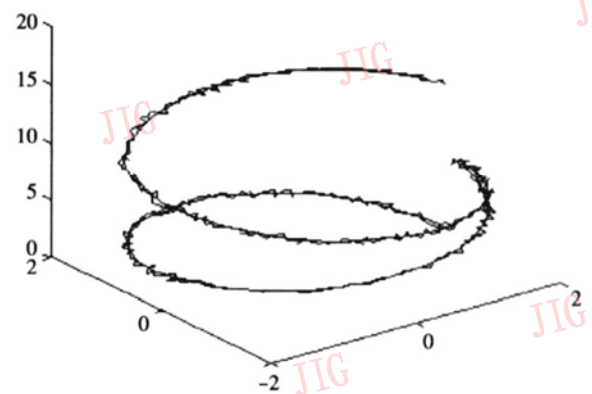
(b) 加噪声之后的螺旋尖点图

图 7. 带有尖点的螺旋曲线加噪声前后效果图

Fig 7 Spiral curve with sharp point before and after noising



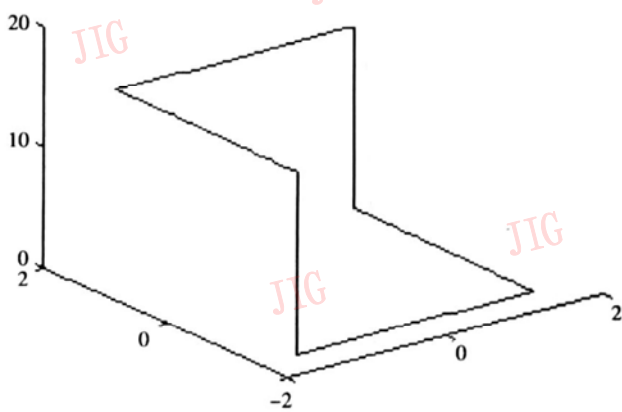
(a) 经过本文算法去噪之后的螺旋带尖点曲线



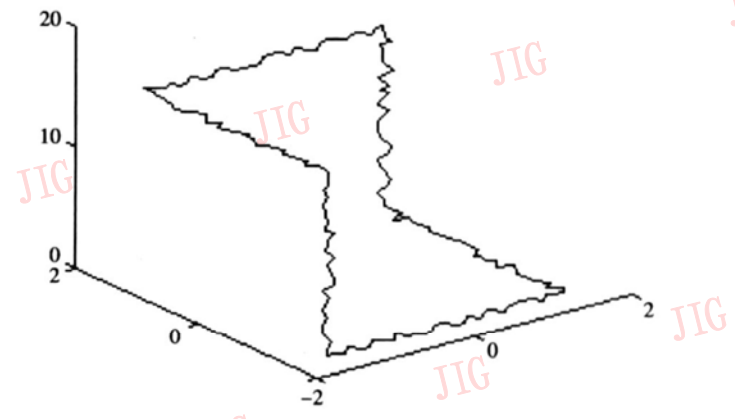
(b) 去噪前后曲线叠加图

图 8 螺旋带尖点曲线去噪效果图

Fig 8 Effect of denoising spiral curve with sharp point



(a) 长方体部分边界原始图形



(b) 加噪声之后的效果图

图 9 长方体部分边界线加噪前后效果图

Fig 9 Partial boundary of cuboid before and after noising

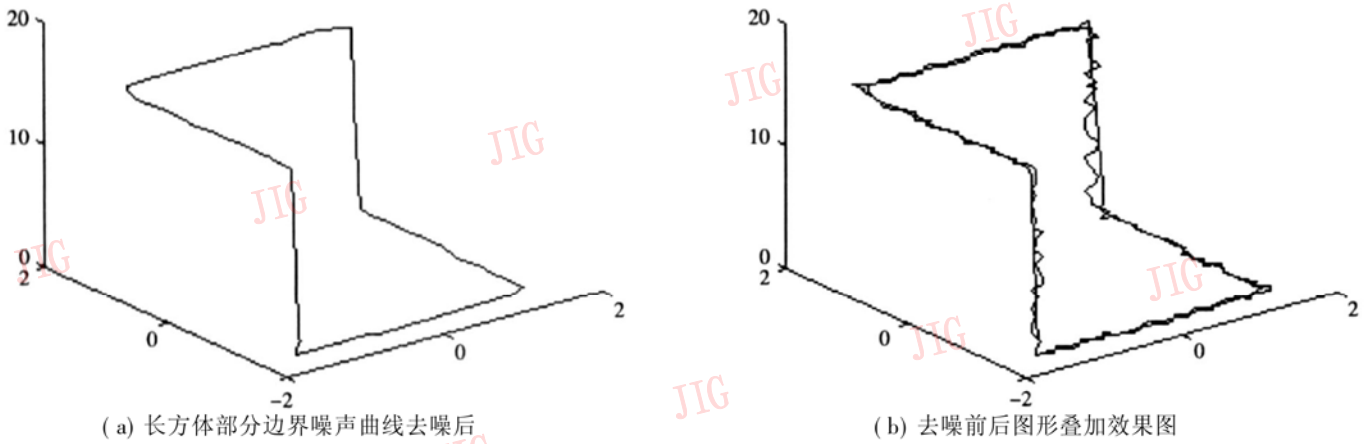


图 10 长方体部分边界去噪后效果图

Fig 10 Partial boundary of cuboid before and after denoising

进行调整。参数 σ'_c , σ'_s , σ_c , σ_s 的大小没有发生改变, 与前面两个例子中的取值相同。但是 k, m 的取值却有所变化, 如果不改变 k, m 的值, 得到的去噪结果就不理想, 这是由曲线本身的性质决定的。图 10 是对图 9(b) 利用本文算法得到的去噪结果, 算法实现过程中的 k, m 分别为 $k=0.0005$, $m=0.0015$ 。 k, m 值的变小使得算法迭代次数有所增加, 算法迭代 25 次之后得到的去噪曲线才比较平滑, 并且相应的尖点也得到了较好的保持。

本文算法是对空间曲线去噪的一大改进, 解决了曲线去噪的收缩问题。当然也有不足之处, 就是在保持尖锐特征方面不是特别好, 尤其在迭代次数增加的情况下。

7 结 论

本文将空间离散曲线采用内在表示, 对边向量和 x 轴、 z 轴正半轴的夹角进行双边滤波, 利用去噪后的角度构造了一个目标函数, 反求得到了曲线新的顶点。因为本文算法滤波的对象是表示曲线走向的角度, 并且把去噪后顶点和噪声曲线顶点的距离作为目标函数的一部分, 所以这种新的去除曲线噪声的方法不会产生一般算法的收缩现象。本文算法的一个缺点是在保持尖锐特征方面效果不是很好, 并且需要多次迭代, 当然这跟参数的选择有很大关系, 参数选择不同, 需要的迭代次数和产生的去噪效果都会发生较大变化。虽然如此, 在保持特征方面, 本文算法还是比传统的低通滤波要好。以后的工作是完善本文算法, 使得它能同时解决好收缩和保持特征问题。

参考文献 (References)

- 1 Marinont D H. A representation for image curves [A]. In Proceedings of American Association for Artificial Intelligence [C], Austin, TX, USA, 1984: 237~242
- 2 Mokhtarian F, Mackworth A. Scale-based description and recognition of planar curves and two-dimensional shapes [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986, 8(1): 34~44
- 3 Witkin A P. Scale-space filtering [A]. In Proceedings of the Eighth International Joint Conference on Artificial Intelligence [C], Karlsruhe, West Germany, 1983: 1019~1022
- 4 Choi B K, Jerard R B. Sculptured surface machining-theory and applications [M]. London: Kluwer Academic Publisher, 1998
- 5 Cho S K, Choi B K. Analysis of difference fairing based on DFT-filter [J]. Computer Aided Design, 2001, 33(1): 45~56
- 6 Liu G H, Wong Y S, Zhang Y F, et al. Adaptive fairing of digitized point data with discrete curvature [J]. Computer Aided Design, 2002, 34(4): 309~320
- 7 Thouis R, Jones F, Durand M D. Non-iterative feature-preserving mesh smoothing [A]. In Computer Graphics Proceedings Annual Conference Series ACM SIGGRAPH [C], San Diego, CA, USA, 2003: 943~949
- 8 Shashar F I, Dori D C. Bilateral mesh denoising [A]. In Computer Graphics Proceedings Annual Conference Series ACM SIGGRAPH [C], San Diego, CA, USA, 2003: 950~953
- 9 Tomasi C, Manduchi R. Bilateral filtering for gray and color images [A]. In Proceedings of the Sixth International Conference on Computer Vision [C], Bombay, India, 1998: 839~846
- 10 Lowe D G. Organization of smooth image curves at multiple scales [J]. International Journal of Computer Vision, 1989, 3: 119~130
- 11 Oliensis J. Local reproducible smoothing without shrinkage [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1993, 15(3): 307~312
- 12 Taubin G. Curve and surface smoothing without shrinkage [A]. In Proceedings of the Fifth International Conference on Computer Vision, Massachusetts Institute of Technology [C], Cambridge, Massachusetts, UK, 1995: 852~857
- 13 Taubin G. A signal processing approach to fair surface design [A]. In Computer Graphics Proceedings Annual Conference Series ACM SIGGRAPH [C], Los Angeles, USA, 1995: 51~58

一种新的基于条件数的图像配准算法

马丽涛¹⁾ 杨丹²⁾ 张小洪²⁾ 李博¹⁾

¹⁾ (重庆大学数理学院, 重庆 400030) ²⁾ (重庆大学软件学院, 重庆 400030)

摘要 提出了一种新的基于条件数的图像配准算法。该方法在 Harris 算法提取角点的基础上, 采用条件数定量地分析了噪声对确定图像间变换关系的影响程度, 通过阈值设定筛选出具有良好稳定性的角点, 克服了 Harris 角点检测可能存在的角点位置偏移和易受噪而提取出伪角点等问题。最后选择了 Random Sample Consensus (RANSAC) 匹配准则来确定匹配点对。经过实验证明了该配准算法具有精确性、抗噪性和鲁棒性。

关键词 条件数 Harris 角点检测 RANSAC 图像配准

中图分类号: TP391 文献标识码: A 文章编号: 1006-8961(2008)02-0277-07

A New Method for Image Registration Based on Condition Number

MA Li-tao¹⁾, YANG Dan²⁾, ZHANG Xiao-hong²⁾, LI Bo¹⁾

¹⁾ (College of Mathematics & Physics, Chongqing University, Chongqing 400030)

²⁾ (School of Software Engineering, Chongqing University, Chongqing 400030)

Abstract In this paper, a novel Harris corner detection algorithm based on condition number is proposed. Firstly feature points are extracted using Harris corner detector which usually leads to either missing significant corners or detecting false corners due to noise. To overcome these drawbacks, condition number, which investigates the relationship between noise and the transformation of images, is used to screen out the steady points. Then a new match method called Random Sample Consensus (RANSAC) is applied to determine matching pairs. A number of experiments demonstrate the accuracy, efficiency and robust of the approach.

Keywords condition number, Harris corner detection, RANSAC, image registration

1 引言

在计算机视觉领域中, 基于特征的图像配准方法是图像配准中最常见的方法之一。而角点作为图像的一个重要的局部特征, 在保留了图像中物体重要特征信息的同时有效地减少了信息的数据量, 成为图像理解和模式识别中重要的图像特征。基于角点的图像配准中有两个重要的环节: 一是角点特征的提取; 二是相似性度量, 即特征匹配准则。

目前的角点检测算法主要分为两类: 一类是基于边缘图像的角点提取, 即通过提取边缘线上具有最大转角的点。这类方法的主要缺点是角点对边缘线的依赖较大, 如果边缘线由于某种原因发生中断 (在实际中经常会出现这种情况), 则会对角点的提取结果造成较大的影响, 如提取出伪角点等。另一类是直接基于图像的角点检测, 它们主要通过计算曲率及梯度来达到检测角点的目的。这类方法往往用某点的梯度与曲率的某种运算结果的局部极值点作为角点的位置, 如二者乘积的局部极大值点, 但此类方法在理论上存在一些问题, 这些局部极值点并

基金项目: 重庆市自然科学基金项目 (CSTC2005BA2002)

收稿日期: 2006-06-22 改回日期: 2006-10-09

第一作者简介: 马丽涛 (1983~), 女, 重庆大学数理学院应用数学专业在读硕士研究生。研究方向为数字图像处理。Email: lmal821

不在通常所定义的角度位置上,即出现角点位置偏移。Harris角点检测是一种直接基于灰度图像的角点提取算法^[1,2],具有比较理想的结果,其应用非常广泛。但是同样存在角点位置偏移和易受噪而提取出伪角点等问题。

因此本文将条件数引入到经典的Harris算法中,通过对噪声的定量分析,获取具有良好稳定性的图像特征。从原理上分析,在基于特征的图像处理中,图像间的变换关系依赖于特征点的位置信息,这就需要特征提取算法具有良好的噪声鲁棒性。而很多传统的特征提取算法旨在提高算法的抗噪能力,没有给出对噪声影响程度的定量分析。矩阵论中矩阵的摄动对其特征值变化的影响程度是用条件数来刻划的,这正对应了图像处理中噪声对确定图像间变换关系的影响程度。新的角点检测算法弥补了Harris检测的不足,大大提升了特征点的匹配率,从而提高配准精度。

点特征的匹配方法有很多种。常见的主要有基于局部灰度信息的特征匹配方法和基于边缘形状信息的特征匹配方法。前者以局部区域灰度特性作为两点相似性的度量标准,而后者则以边缘形状信息作为两点相似性的度量标准。而在匹配过程中,通常会遇到两类错误匹配点:一是错误定位造成的错误匹配点,这往往是图像中的噪声引起的;二是错误匹配,这是由于非匹配点的局部相似性等造成的。

实际中常常发现,误匹配的存在使得这些匹配方法对变换的估计变得困难甚至失败。因此需要寻求一种方法,在考虑误配存在的情况下能够很好地消除其影响,并将它们识别出来,然后依据变换评价函数计算出最优的变换。本文选取了在这方面具有良好的表现的RANSAC(Random Sample Consensus)算法,从而提高了整个配准算法的性能。

2 条件数

2.1 概念

首先,图像的匹配问题可以转换成目标函数的最小化问题,通过求解目标函数的最优解获得图像匹配的最优变换的参数。

对给定的两幅图像 g, \hat{g} 点 (x, y) 关于变换 T 的点匹配目标函数为:

$$f(T) = \frac{1}{2} \sum (g(T(x', y')) - \hat{g}(x', y'))^2$$

式中, \sum 是在点 (x, y) 的一个特征邻域内求和; (x', y') 是点 (x, y) 邻域内的点; T 的 3 种常见变换关系如下:

(1) 平移: 参数向量是 $p = (a, b)$

$$T(x', y') = \begin{vmatrix} x' + a \\ y' + b \end{vmatrix}$$

(2) 旋转-伸缩-平移 (RST): 参数向量是 $p = (\theta, r, a, b)$

$$T(x', y') = \begin{vmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{vmatrix} \begin{vmatrix} x' - x \\ y' - y \end{vmatrix} + \begin{vmatrix} a \\ b \end{vmatrix}$$

(3) 仿射: 参数向量是 $p = (m_1, m_2, m_3, m_4, a, b)$

$$T(x', y') = \begin{vmatrix} m_1 & m_2 \\ m_3 & m_4 \end{vmatrix} \begin{vmatrix} x' - x \\ y' - y \end{vmatrix} + \begin{vmatrix} a \\ b \end{vmatrix}$$

在匹配问题中,噪声等因素的存在会对变换参数 p 产生不同程度的影响,由此引入条件数来定量地刻画这种影响程度。

定义 条件数 K 度量了参数向量 p 对扰动 $(\Delta g, \Delta \hat{g})$ 的敏感度

$$K = \lim_{\delta \rightarrow 0} \frac{\max_{\|(\Delta g, \Delta \hat{g})\| \leq \delta} \|\Delta p\|}{\|(\Delta g, \Delta \hat{g})\|}$$

式中, $\|(\Delta g, \Delta \hat{g})\| = (\|\Delta g\|^2 + \|\Delta \hat{g}\|^2)^{1/2}$, Δp 表示 p 随着扰动 $(\Delta g, \Delta \hat{g})$ 的变化量。

2.2 定理

在将条件数引入图像处理之前,介绍两个有用的定理。

定理 1^[3] 设变换加噪声问题为

$$\hat{g}(x', y') = g(T(x', y')) + \eta(x', y')$$

式中, $\|\eta\| < \delta$ (x', y') 取遍以 (x, y) 为中心的邻域。当 $\delta \rightarrow 0$ 时,得到

$$K_T^2 = \|(A^T A)^{-1}\|$$

式中, $A = \begin{vmatrix} v_1 \\ \vdots \\ v_n \end{vmatrix}$, 行向量 v_i 取决于变换类别, 平移,

RST和仿射变换对应的 v_i 分别为

平移: $v_i = (\hat{g}_x^i, \hat{g}_y^i)$

RST:

$$v_i = (\hat{g}_x^i, \hat{g}_y^i, \hat{g}_x^i(x_i - x) + \hat{g}_y^i(y_i - y), \hat{g}_x^i(y_i - y) - \hat{g}_y^i(x_i - x))$$

仿射:

$$v_i = (\hat{g}_x^i, \hat{g}_y^i, \hat{g}_x^i(x_i - x), \hat{g}_x^i(y_i - y), \hat{g}_y^i(x_i - x), \hat{g}_y^i(y_i - y))$$

式中, (x_i, y_i) 是以 (x, y) 为中心的窗口中的第 i 个点; $\hat{g}_x^i = \hat{g}_x(x_i, y_i)$, $\hat{g}_y^i = \hat{g}_y(x_i, y_i)$, 其下标表示对 $\hat{g}(x, y)$ 求偏导。

定理 2^[4] 在定理 1 的假设条件下, 有

$$K_{\text{Trans}} \ll K_{\text{RST}} \ll K_{\text{Affine}}$$

式中, K_{Trans} , K_{RST} , K_{Affine} 分别表示平移变换、RST 变换和仿射变换下对应的条件数。

由前面两个定理, 可以得到如下结论: 对平移变换病态 (对变换很敏感即条件数值很大时我们称之为病态) 的点对 RST 和仿射变换仍是病态的, 同时由定理 1 的结论可以看出, 平移变换的条件数的计算量要比 RST 和仿射变换的少很多。综合以上两点可以采用平移变换的条件数来消除图像中的病态点即不稳定点。

2.3 实现

由定理 1 中条件数的形式可以看出, 条件数主要由 \hat{g}_x 和 \hat{g}_y 组成, 并不需要知道变换关系的参数情况。对平移变换的条件数来说, 矩阵 $A^T A$ 是 2×2 的, 且

$$A^T A = \begin{bmatrix} \sum \hat{g}_x^2 & \sum \hat{g}_x \hat{g}_y \\ \sum \hat{g}_x \hat{g}_y & \sum \hat{g}_y^2 \end{bmatrix} \quad (1)$$

式中, \sum 是在点 (x, y) 的一个窗内求和。

考虑到矩阵 $A^T A$ 的特征值经常用于对局部图像结构的分析 (例如, 面、线或角点的结构) 上, 这个矩阵与条件数的联系也就不足为奇了。

对一个 2×2 的矩阵, 它的逆可以很精确地求出来。但为防止矩阵 $A^T A$ 产生奇异性有必要对它进行一个修正。

$$K^2 = \|(A^T A + \epsilon I)^{-1}\| \quad (2)$$

式中, I 是 2 阶单位矩阵, ϵ 是个很小的值, 实验中取 $\epsilon = 10^{-8}$, 使用的是 2 范数。

3 RANSAC

在实际应用中, 很多情况下很难保证匹配的结果全部是正确的, 误配的存在使得变换的估计变得困难甚至失败。我们寻求一种方法, 在考虑误配存在的情况下能够很好地消除其影响, 正确地估计变换。为了消除这些误配点的影响, 必须将它们识别出来, 然后才能在变换评价函数的帮助下计算出最优的变换。在这方面, RANSAC 算法具有良好的表现。

为了清楚解释 RANSAC 算法^[4, 5]的过程, 同时也展现其实质, 用一个简单情形的例子进行说明。例子是针对线性变换的, 但是对于仿射变换或投影变换这种算法的思想是类似的。考虑一个经典的数值分析的例子: 寻找一条直线, 使得它可以拟合一个 2 维的点集。这可以进一步叙述成: 给定一组 2 维的点, 找到一条直线使得所有点到直线的距离之和最小。可以解决这个问题的鲁棒的算法有很多, 但是它们各自有不同的适应情况, 使用哪一种依赖于无效点的比例。比如, 如果知道只有一个无效点, 那么可以通过轮流删除一个点并对剩余的点进行最优直线估计。如果无效点的比例很大, 这样的方法就不适用了。这时候就需要一个在无效点比例较大的情况下仍然有效的估计算法, 这就是 RANSAC 算法。RANSAC 算法可以处理含有很大比例无效点的情形。

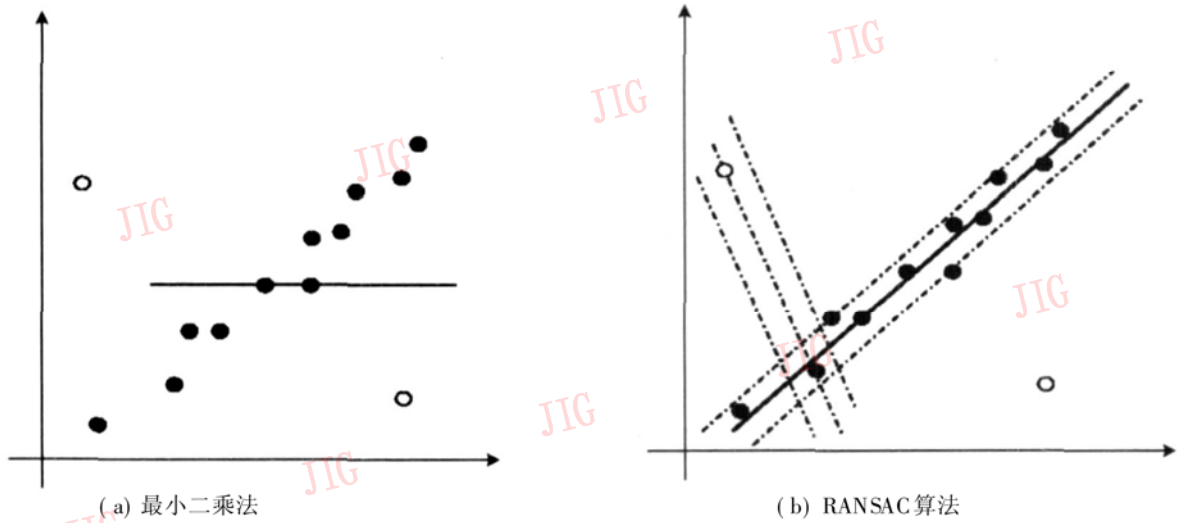
RANSAC 算法的思想简单而巧妙: 首先随机地选择两个点, 这两个点确定了一条直线。称直线的支撑为在这条直线的一定距离范围内的点的数目。这样地随机选择重复数次, 然后具有最大支撑的直线被确认为是点集的拟合。在拟合的误差距离范围内的点被认为是有效点, 它们构成所谓的一致集, 反之则为无效点。根据算法描述, 可以很快判断, 如果只有一个无效点, 那么随机选取的包含无效点的初始点集确定的直线不会获得很大的支撑。如图 1 所示。一般的最小二乘法会严重地被一个不需要的点影响; 在 RANSAC 算法中, 直线通过其在一定误差范围内的点数获得而选定, 图 1(b) 中的最优直线有 10 个支持点。

值得注意的是, 利用数据拟合得到最优的模型, 需要限定随机选取的样本包含确定模型需要数据的最小的集合, 不然无法得到确定的解。在本例子里需要确定直线, 所以需要两个点, 对于投影变换, 这样的最小集合要有 4 个点。

本文算法不同于传统的算法。RANSAC 算法过程与传统的光滑化的技术相反, 传统技术用尽可能多的数据获得初始的解决方案, 然后除去无效的点, 而 RANSAC 算法用尽可能小的初始数据并在一致性的条件下扩大这个集合^[5]。

由前面所述, RANSAC 算法的基本过程可以总结如下:

- (1) 从数据点集合 S 中随机选取一个数据点样本, 并求出这个子集初始化模型。



(a) 最小二乘法

(b) RANSAC算法

图 1 鲁棒的直线估计 (实心的点是需要点, 空心的点是不需要的点)

Fig 1 For robust estimation of line(stone points is needed and the hollow points is not)

(2) 找出按照阈值 T 成为当前模型的支撑的数据点集 S_i 。集合 S_i 就是样本的一致集 (consensus set), 被定义为有效点。

(3) 如果集合 S_i 的大小超过了某个阈值 T_1 , 用 S_i 重新估计模型并结束。

(4) 如果集合 S_i 的大小小于某个阈值 T_2 , 选取一个新的样本, 重复上面的步骤。

(5) 经过了 N 次尝试, 最大的一致集 S_i 被选中, 用它来重新估计模型, 得到最后的结果。

4 算法流程

对角点提取中出现的位置偏移、伪角点等不稳定的情况, 采用条件数来加强稳定性, 通过条件数来度量角点的稳定情况。即对每个角点提取一个特征邻域, 在此邻域内根据式 (1) 计算 $A^T A$ 从而由式 (2) 得到该点的条件数 K 。 K 越大说明该点的稳定性越差。因而设定一定的阈值 T_k 去除 $K > T_k$ 的不稳定角点, 为后一步配准提供了良好的基础。

配准时采用了可高度容忍误配存在的 RANSAC 算法, 同时为进一步提高配准效率, 实验中首先利用相关性 (RANSAC 算法本身的时间复杂度比较高。需要一个计算量较少, 较简单的算法来降低计算量) 对图像进行了初匹配, 即得到了一个比较接近最优模型的初始化模型 (称之为次优解), 保证了配准算法的有效性, 使匹配更容易更快地达到最优状态。该算法的流程图如图 2 所示。

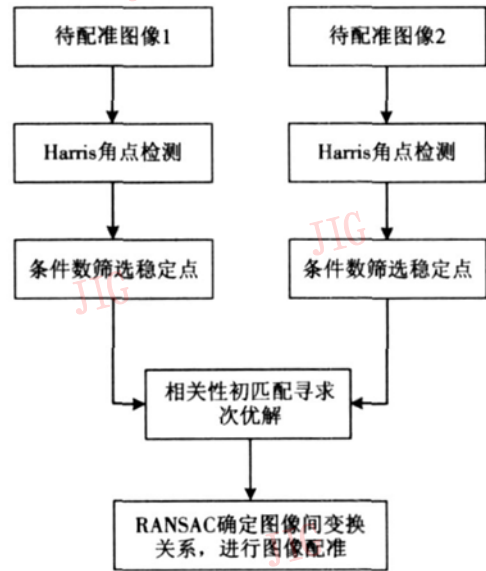


图 2 基于条件数的图像配准算法流程图

Fig 2 The flow chart of in age registration algorithm based on condition number

5 图像配准实验和结果分析

将上述方法分别应用于遥感图像和航空图像中。图 3 分别是两幅取自不同传感器的航拍图像。图 4 分别是两幅取自不同传感器的遥感监测图像。图 5 分别是两幅时间差异 4 年的图像。图 3~ 图 5 中 (c) 和 (d) 分别是利用 Harris 角点检测算法对待配准图 (a) 和 (b) 进行特征点提取的结果, 然后在这些特征点集合中, 利用条件数筛选出不稳定的特征点, 接着将 Harris 和 RANSAC 结合得到的匹配点对

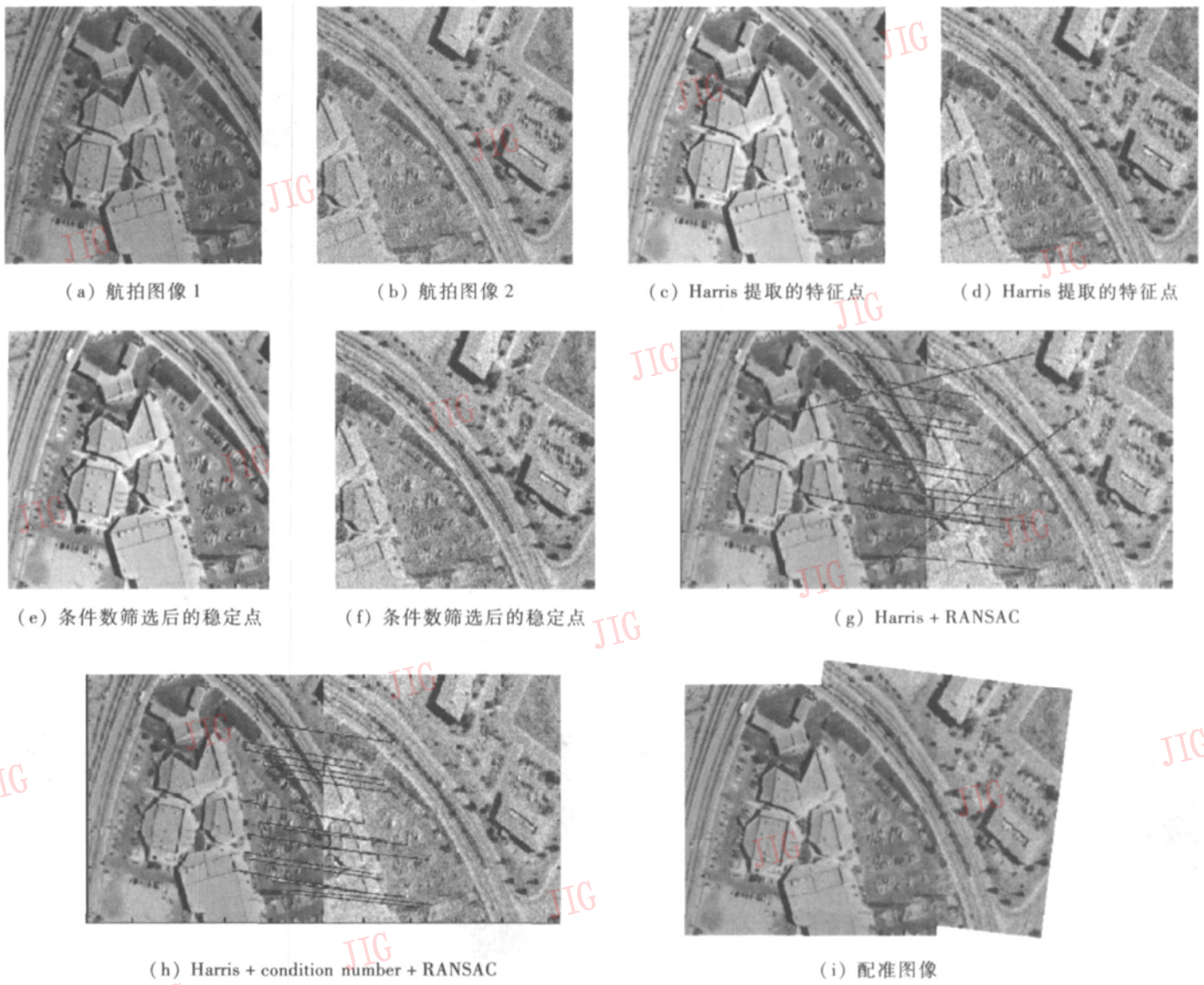
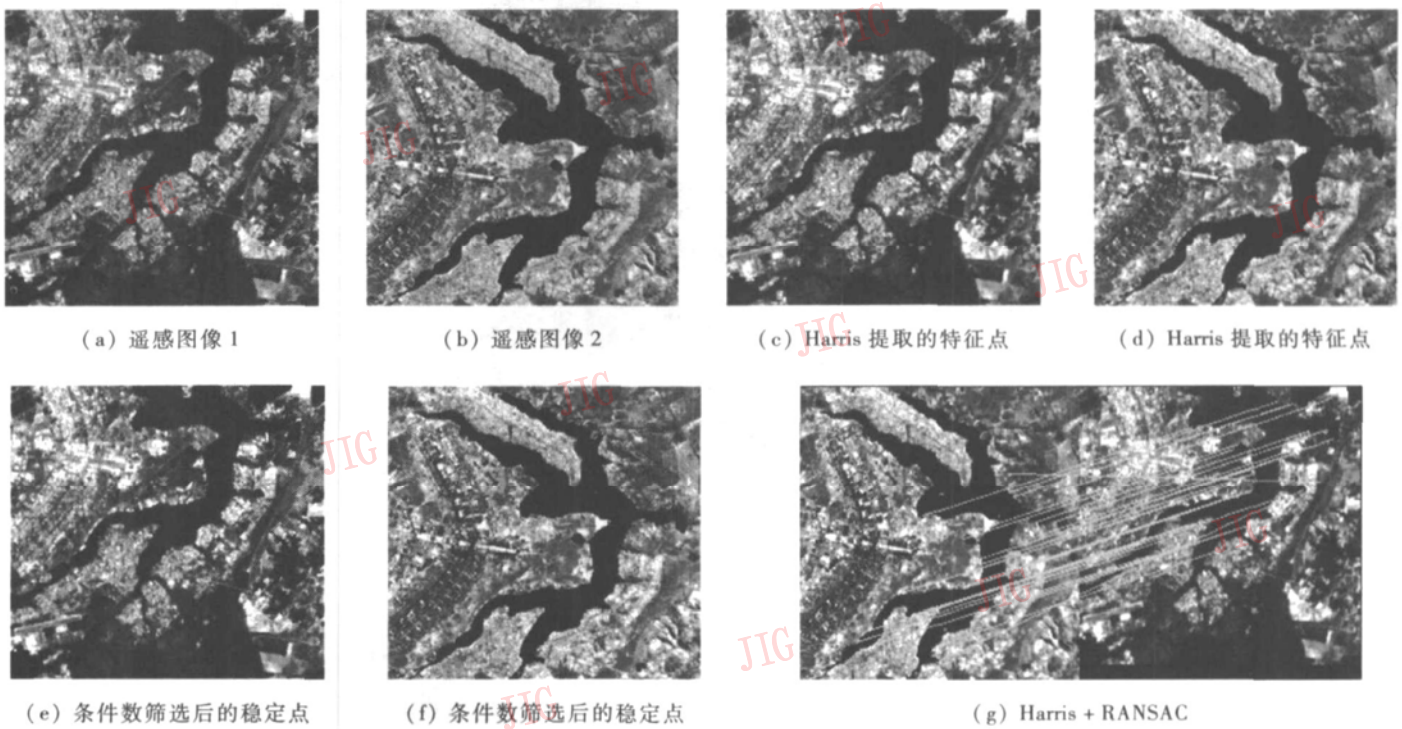
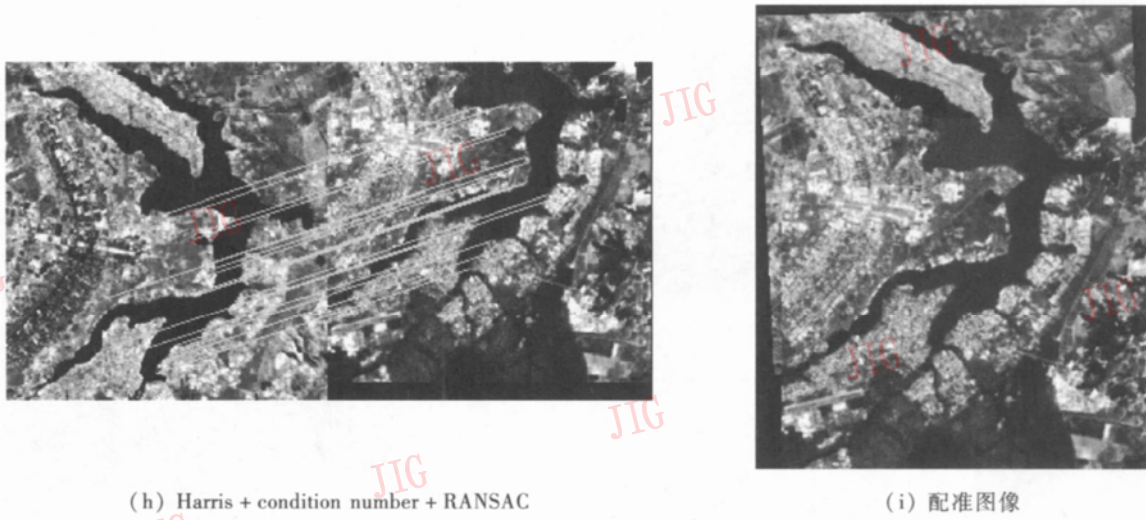


图 3 两幅航拍图像之间的配准

Fig. 3 Registration of two aerial images



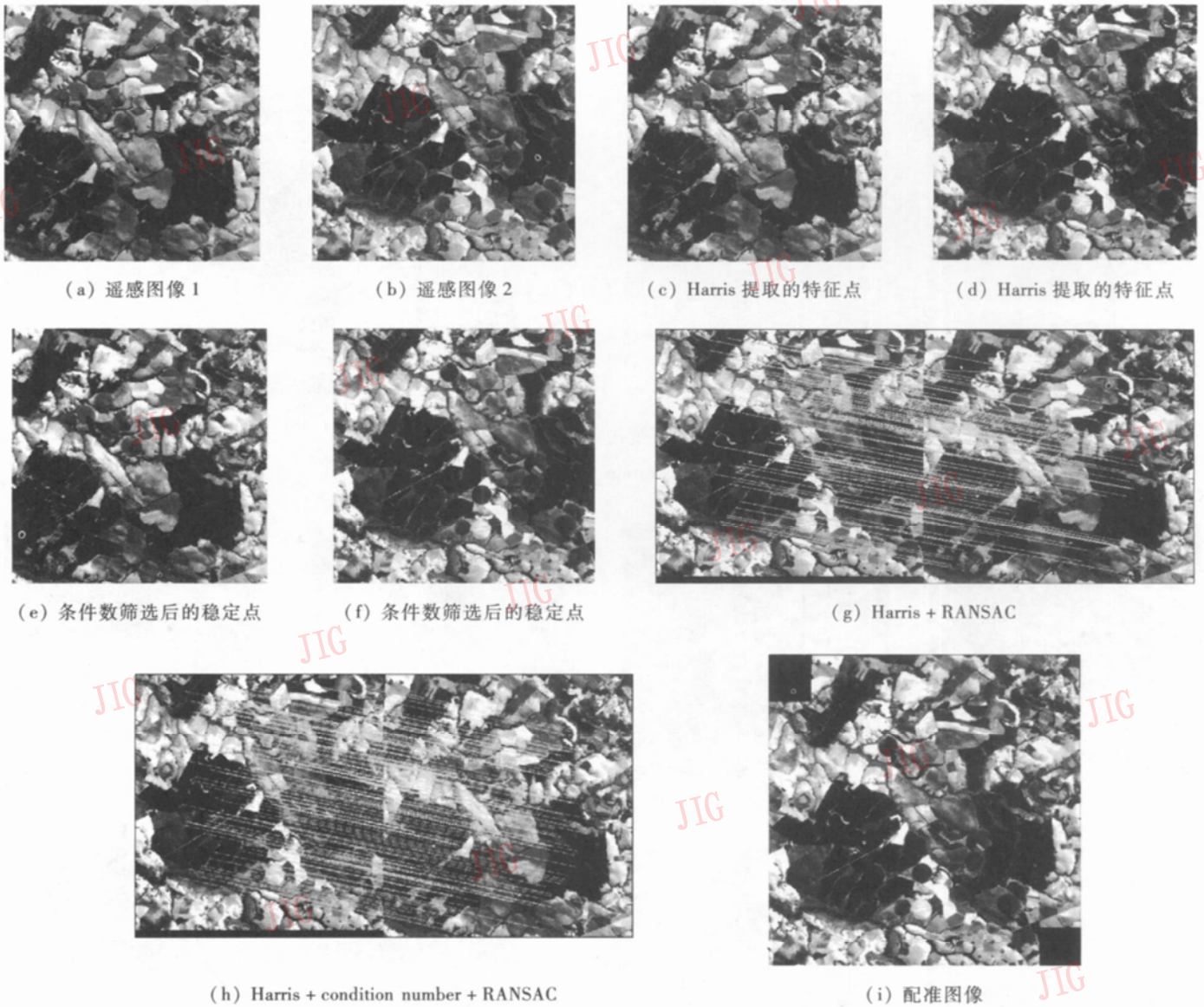


(h) Harris + condition number + RANSAC

(i) 配准图像

图 4 两幅遥感图像之间的配准

Fig. 4 Registration of two different sensor images



(a) 遥感图像 1

(b) 遥感图像 2

(c) Harris 提取的特征点

(d) Harris 提取的特征点

(e) 条件数筛选后的稳定点

(f) 条件数筛选后的稳定点

(g) Harris + RANSAC

(h) Harris + condition number + RANSAC

(i) 配准图像

图 5 两幅具有 4 年时间差的图像之间的配准

Fig. 5 Registration of two Landsat images with four year difference

图 (g) 与 Harris、condition number 和 RANSAC 结合得到的图 (h) 进行了比较, 最后给出了各图配准后的图像 (i)。

从给出的图像中可以看出条件数删掉了一些明显的伪角点和一些不太稳定的角点, 从而能进一步提高 RANSAC 的配准精度。由于实验中稳定点的个数取决于所选取的阈值, 所以可以通过调整阈值来寻找比较理想的数值。

6 结 论

基于条件数的 Harris 角点检测算法, 通过条件数来判断各个点的稳定情况, 从而剔除了不稳定点, 并利用 RANSAC 算法进行角点匹配, 两种方法的有效结合使得整个算法具有了计算量小、匹配精度高、抗噪性强和适用性广等优点。

参考文献 (References)

1 Harris C, Stephens M. A Combined Corner and Edge Detector [A].

In: Proceedings of 4th A lvey V ision Conference [C]. Manchester UK, 1988. 189~ 192

2 Overgaard N C. On a Modification to the Harris Corner Detector [A]. In: Symposium Svenska Sällskapet för Bildanalys [C], Stockholm, 2003. 6~ 7.

3 Kenney C S, Manjunath B S, Zuliani M, *et al*. A condition number for point matching with application to registration and post-registration error estimation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(11): 1437~ 1454

4 Fischler M A, Bolles R C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography [J]. Communications of ACM, 1981, 24(6): 381~ 395

5 Chen Fu-xing, Wang Rui-sheng. Fast RANSAC with preview model parameters evaluation [J]. Journal of Software, 2005, 16(8): 1431~ 1437. [陈付幸, 王润生. 基于预检验的快速随机抽样一致性算法 [J]. 软件学报, 2005, 16(8): 1431~ 1437.

基于修正 PCNN 的多传感器图像融合方法

李敏^{1,2)} 蔡骋¹⁾ 谈正¹⁾

¹⁾ (西安交通大学电信学院信息工程研究所, 西安 710049) ²⁾ (第二炮兵工程学院计算机系, 西安 710025)

摘要 多传感器图像融合技术作为信息融合的重要分支和研究热点, 已广泛应用于机器视觉、医疗诊断、军事遥感等领域。为了更好地进行多传感器图像融合, 将在图像分割、目标识别等领域具有独特优势的脉冲耦合神经网络 (pulse coupled neural network, PCNN) 引入到多传感器图像融合领域中来, 提出了一种基于修正 PCNN 的多源图像融合方法, 该方法在区域分割的基础上, 先提取区域特征, 然后由特征指导融合过程; 同时, 从目标区域相对于背景的显著性出发, 提出了一种反映目标区域突出性的新特征, 并针对传统 PCNN 参数无法自动设定的难题, 提出了基于修正 PCNN 的参数自动设定方案。实验结果表明, 该方法无论在主观视觉效果, 还是客观评价参数上均优于基于多分辨分析的融合算法, 且克服了传统像素级融合方法中融合图像模糊、对噪声敏感等不足, 尤其适用于图像不能严格配准的应用场合。这对于拓宽 PCNN 的理论研究和实际应用具有一定价值。

关键词 多传感器图像融合 脉冲耦合神经网络 参数设定 客观评价准则

中图分类号: TN911.73 TP391.41 文献标识码: A 文章编号: 1006-8961(2008)02-0284-07

Modified PCNN Based Multisensor Image Fusion Scheme

LIM in^{1,2)}, CAI Cheng¹⁾, TAN Zheng¹⁾

¹⁾ (School of Electrics and Information Engineering, Xi'an Jiaotong University, Xi'an 710049)

²⁾ (Computer Science Department, The Second Artillery Engineering College, Xi'an 710025)

Abstract Being an efficient method of information fusion, multisensor image fusion has been used in many fields such as machine vision, medical diagnosis, military applications and remote sensing. In this paper, PCNN is introduced in this research field for its interesting properties in image processing, including segmentation, target recognition et al., and a multisensor image fusion scheme based on modified PCNN is proposed. The basic idea of the scheme is to segment all different input images by PCNN and to use this segmentation to guide the fusion process. At the same time, a new region feature, which emphasized the salience of target regions and its neighbors is proposed. Focusing on the famous difficult problem of PCNN, how to determine PCNN parameters adaptively, an adaptive PCNN parameters determination algorithm is also presented in this paper. Experimental results demonstrate that the proposed fusion scheme outperforms the multiscale decomposition based fusion approaches, both in visual effect and objective evaluation criteria. It avoids some of the well-known problems in pixel-level fusion such as blurring effects and high sensitivity to noise, particularly when there is mis-registration of the source images. The research fruits have certain value on the theory research and practical application of PCNN.

Keywords multisensor image fusion, pulse-coupled neural network (PCNN), parameter determination, objective evaluation criteria

1 引言

多传感器图像融合作为信息融合的重要分支和

研究热点, 已广泛应用于机器视觉、医疗诊断、军事遥感等领域^[1]。传统的像素级多分辨图像融合方法, 包括基于拉普拉斯金字塔、比率低通金字塔、梯度金字塔和小波变换等融合方法^[2], 它们在对源图

基金项目: 军队预研项目 (403050203); 国防“十五”重点预研项目 (41322029); 国防“十一五”重点预研项目 (513220208)

收稿日期: 2006-05-20 改回日期: 2006-10-17

第一作者简介: 李敏 (1971~), 女, 副教授, 博士。研究领域为图像处理、模式识别、信息融合等。E-mail: limin@mailst.xjtu.edu.cn

像进行多尺度分解和系数选取时, 由于大都孤立地处理各像素, 从而割裂了像素间的联系。此外, 尽管基于小波变换的图像融合方法在多分辨图像融合方法中是十分有效的, 但是由于传统的小波变换不具有平移不变性, 因此在源图像不能严格配准的情况下, 将直接影响融合效果。

脉冲耦合神经网络 (pulse coupled neural network, PCNN) 是对 Eckhom 等人根据猫的视觉皮层神经元脉冲串同步振荡现象提出的神经元模型^[3, 4]加以修改和变化得到的, 因其在图像分割、目标识别等领域具有显著优势, 已成为近年来的研究热点。

本文提出了一种将 PCNN 应用于多传感器图像融合的具体方法, 即在对源图像进行配准的基础上, 对感兴趣区域进行分割, 首先提取区域特征; 然后由特征指导融合过程, 同时从目标区域相对于背景的显著性出发, 提出了一种反映目标区域突出性的新特征, 并针对传统 PCNN 参数无法自动设定的难题, 提出了基于修正 PCNN 的参数自动设定方案。文末结合同源和异源多传感器图像融合的应用实例, 给出了对比实验效果。

2 融合算法结构

本文提出的算法基本结构流程如图 1 所示。

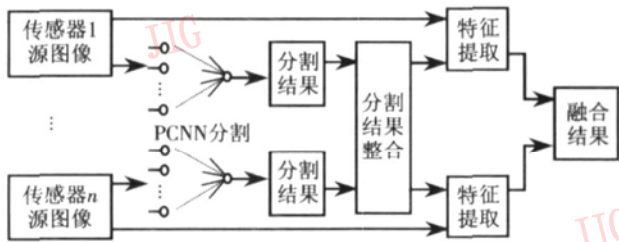


图 1 本文算法结构图

Fig 1 Schematic diagram of the proposed fusion method

具体算法步骤如下:

- (1) 对来自多传感器的源图像 I_1, I_2, \dots, I_M , 在经过配准的基础上, 利用修正 PCNN 进行区域分割, 并将各分割结果进行整合。设 $R_{m,i}$ 为第 m 个源图像 I_m ($m = 1, 2, \dots, M$) 经分割后所得的第 i 个区域;
- (2) 提取以下能反映图像分割区域 $R_{m,i}$ 性能和清晰度的特征: 显著性因子 S 和可见性因子 V ;
- (3) 将特征作为主要依据用于确定该区域在融合图像中的贡献系数, 即融合权值;
- (4) 根据像素所属区域的融合权值, 通过自适

应加权平均最终得到融合图像。

3 基于修正 PCNN 的自适应图像分割

3.1 PCNN 模型及图像分割原理

构成 PCNN 的神经元是一个综合的动态非线性系统, 它包括接受域、内部活动项 (调制部分) 和脉冲发生器 3 部分 (如图 2 所示)。

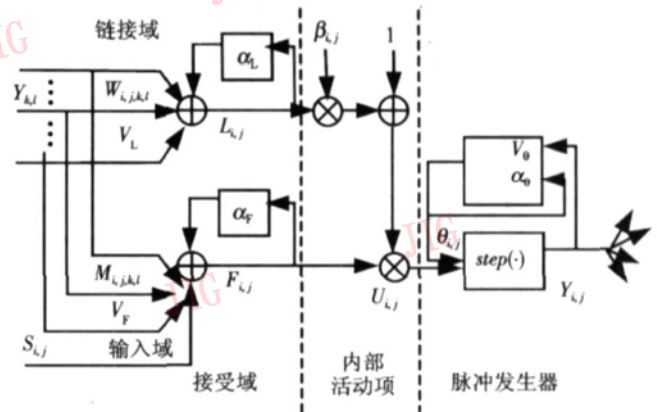


图 2 PCNN 神经元模型

Fig 2 PCNN neuron model

接受域包括输入域和链接域, 输入域接受来自外界的刺激 $S_{i,j}$ 和相邻神经元的输出 $Y_{i,j}$, 链接域仅接受相邻神经元的输出 $Y_{i,j}$ 。内部活动部分先将来自输入域和链接域的信号进行调制得到内部活动信号 $U_{i,j}$; 然后脉冲发生器将 $U_{i,j}$ 与动态门限 $\theta_{i,j}$ 进行比较, 若 $U_{i,j}$ 大于 $\theta_{i,j}$, 则脉冲发生器打开 (称该神经元点火), 输出为 1; 之后 $\theta_{i,j}$ 在阈值信号发生器中通过反馈迅速提高, 当其提高到大于 $U_{i,j}$ 时, 则脉冲发生器关闭, 输出为 0 从而形成脉冲信号。神经元 $N_{i,j}$ 的活动可由下式表述:

$$F_{i,j}(n) = \exp(-\alpha_F) \cdot F_{i,j}(n-1) + S_{i,j} +$$

$$V_F \cdot \sum M_{i,j,k,l} Y_{k,l}(n-1) \quad (1)$$

$$L_{i,j}(n) = \exp(-\alpha_L) \cdot L_{i,j}(n-1) +$$

$$V_L \cdot \sum W_{i,j,k,l} Y_{k,l}(n-1) \quad (2)$$

$$U_{i,j}(n) = F_{i,j}(n) (1 + \beta_{i,j} \cdot L_{i,j}(n)) \quad (3)$$

$$\theta_{i,j}(n) = \exp(-\alpha_\theta) \theta_{i,j}(n-1) + V_\theta Y_{i,j}(n-1) \quad (4)$$

$$Y_{i,j}(n) = \text{step}(U_{i,j}(n) - \theta_{i,j}(n)) \quad (5)$$

式中, $F_{i,j}$ 为输入项, $L_{i,j}$ 为链接项, $U_{i,j}$ 为内部活动项, $\theta_{i,j}$ 为动态门限, $Y_{i,j}$ 表示该神经元产生的输出时序脉冲序列, n 为迭代次数, $\alpha_F, \alpha_L, \alpha_\theta$ 分别为输入项、链接项、门限时间衰减常数, V_F, V_L, V_θ 为放大系

数, $M_{i j k l}$ 、 $W_{i j k l}$ 表示神经元 $N_{i+k, j+l}$ 与 $N_{i j}$ 的连接权值, $\beta_{i j}$ 是神经元 $N_{i j}$ 的连接系数, $step(\cdot)$ 表示单位阶跃函数。

$M \times N$ 大小的 2 维图像矩阵可以理解为 $M \times N$ 个 PCNN 神经元模型, 若将每个像素对应一个 PCNN 神经元, 则其每一个像素的灰度值对应为每个神经元的输入刺激 $S_{i j}$ 。若将这些神经元按一定方式连接, 就得到一个单层脉冲耦合神经网络, 当内部连接矩阵 M 、 W 所在邻域内有灰度值相近的像素存在时, 则其中某一个像素的脉动输出将会引起附近其他类似灰度像素对应神经元的激发, 并产生脉动序列输出 $Y(n)$ 。序列 $Y(n)$ 包含有图像区域、边缘、纹理等特征信息。这样由输出序列 $Y(n)$ 构成的二值图像序列就是 PCNN 输出的分割图像。

从以上分析可知, 传统的 PCNN 模型需要恰当设置其数学模型中各种放大系数、衰减时间常数、连接权值矩阵、连接系数等, 才能实现最佳分割。然而, 要自适应地确定这些参数是非常困难的。迄今为止, 解决这一问题的主要方法还停留在根据多次试验的人工调整参数阶段, 这显然不利于 PCNN 的应用。因此, 参数的自动确定问题已成为限制 PCNN 模型应用的一个难点^[4, 5]。

3.2 修正 PCNN 模型

近年来, 已有专家就 PCNN 参数自动设定问题进行了研究, 目前 PCNN 参数自动设定方法主要集中在以下两类: 一类是在固定其他参数的情况下, 只讨论某单个参数的自动设定方法^[5~7]; 另一类是采用简化 PCNN 模型, 在减少参数个数的情况下, 讨论参数的自动设定方法^[8, 9], 这也是目前的重要研究方向。然而目前此类方法要么对图像本身要求高, 缺乏自适应性; 要么如作者在论文中提到的, 需要通过许多图像数据进行大量的训练, 才能得到较好的参数定义, 进而获得满意的分割结果。

通过大量实验, 本文建立了一种基于多阈值预设的修正 PCNN 模型, 其神经元 $N_{i j}$ 的活动可用下式描述:

$$L_{i j}[n] = \sum W_{i j k l} Y_{k l}[n-1] \quad (6)$$

$$U_{i j}[n] = S_{i j} (1 + \beta_{i j}[n] L_{i j}[n]) \quad (7)$$

$$Y_{i j}[n] = \begin{cases} 1 & U_{i j}[n] > T_{i j}[n] \\ 0 & \text{其他} \end{cases} \quad (8)$$

其中, i, j 为图像中像素的位置, k, l 是指邻域内其他像素位置, n 为迭代次数。 $L_{i j}[n]$ 为像素 (i, j) 的邻

域链接项, $U_{i j}[n]$ 是像素 (i, j) 的内部活动项, 它由像素值 $S_{i j}$ 和链接项决定, $\beta_{i j}[n]$ 为链接系数, $Y_{i j}[n]$ 为 PCNN 在点 (i, j) 处的神经元输出, $T_{i j}[n]$ 是一组阈值, 其可通过 3.3 节中的水域法求得。

每个 PCNN 神经元接受其 3×3 邻域以内的神经元链接输入, 其内部链接矩阵 W 是一个 3×3 的方阵, 且每一个元素的值为对应像素到中心像素的欧几里德距离的倒数 (归一化处理)。链接系数 β_m 的值由下式决定:

$$\beta_m = \frac{T_m}{P_m} - 1 \quad (m = 1, \dots, M) \quad (9)$$

考虑到最小阈值 T_1 处, 其对应的左侧峰值处的灰度值 P_1 有可能为 0 因此可以选择 $\beta_1 = 0.1 \sim 0.3$ 。

3.3 水域法求图像多分割阈值

本文提出用水域概念来求取多分割阈值。

记 $P_{\max} = \{P_i^{\max}, i = 1, 2, \dots, K\}$ 为与图像 $f(x, y)$ 对应的灰度直方图 $hist(f(x, y))$ 中极大值点集合; $P_{\min} = \{P_j^{\min}, j = 1, 2, \dots, N\}$ 为极小值点集合; $P = \{P_m, m = 1, 2, \dots, M+1\}$ 为待求的峰值点集合, 且 $g_{P_1} < g_{P_2} < \dots < g_{P_{M+1}}$; $T = \{T_n, n = 1, 2, \dots, M\}$ 为待求的阈值点集合, 且 $g_{T_1} < g_{T_2} < \dots < g_{T_M}$, $P \subseteq P_{\max}$, $T \subseteq P_{\min}$ 。

定义 (水域 water-valley)。如图 3 所示, 对第 i 幅图像直方图 $hist(f(x, y))$ 中的两个极大值点 $P_{i_1}^{\max}$ 和 $P_{i_2}^{\max}$, 其对应的灰度值分别为 $g_{P_{i_1}^{\max}}$ 和 $g_{P_{i_2}^{\max}}$, 且 $g_{P_{i_1}^{\max}} < g_{P_{i_2}^{\max}}$, 如果在区间 $(g_{P_{i_1}^{\max}}, g_{P_{i_2}^{\max}})$ 内不存在其他极大值点或灰度均小于 $\min\{g_{P_{i_1}^{\max}}, g_{P_{i_2}^{\max}}\}$, 则两者之间可构成一个水域 R_{valley} ——即以 $hist(f(x, y))$ 包络线为底, 以 $\min\{g_{P_{i_1}^{\max}}, g_{P_{i_2}^{\max}}\}$ 为水域高度, 用假想的水充满该空间的水域, 而整个水域的蓄水量则定义为水域面积 $area(g_{P_{i_1}^{\max}}, g_{P_{i_2}^{\max}})$, 其计算公式为

$$area(g_{P_{i_1}^{\max}}, g_{P_{i_2}^{\max}}) = \frac{1}{2} \int_{g_{P_{i_1}^{\max}}}^{g_{P_{i_2}^{\max}}} \{[\min\{g_{P_{i_1}^{\max}}, g_{P_{i_2}^{\max}}\} - hist(x)] + |\min\{g_{P_{i_1}^{\max}}, g_{P_{i_2}^{\max}}\} - hist(x)|\} dx \quad (10)$$

设 P_j^{\min} 为区间 $(g_{P_{i_1}^{\max}}, g_{P_{i_2}^{\max}})$ 内的最小值点, 即对 $\forall g_x \in (g_{P_{i_1}^{\max}}, g_{P_{i_2}^{\max}})$, 均满足 $hist(g_x) \geq g_{P_j^{\min}}$ 。相应地, 可将水域表示为 R_{valley} , 其在直方图上的灰度区间为 $(g_{P_{i_1}^{\max}}, g_{P_j^{\min}}, g_{P_{i_2}^{\max}})$, 如图 3 中所示的深色区域。

峰值和阈值的求取过程如下:

- (1) 绘制图像 $f(x, y)$ 的直方图 $hist(f(x, y))$, 并根据需要对其进行平滑处理, 以减小噪声的影响。
- (2) 求出 $hist(f(x, y))$ 中所有极值点, 包括极

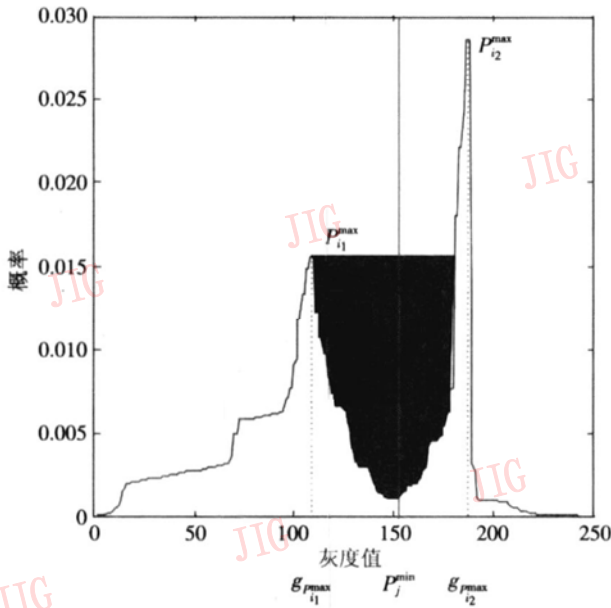


图 3 水域模型示意图

Fig 3 Figure of water valley

大值点集合 P_{max} 和极小值点集合 P_{min} 。为了构筑水域的需要, 本文规定直方图两侧的极值点必须是极大值点, 即 $K = N + 1$ 。

(3) 取最左侧极小值点 P_L^{min} 及其两侧的极大值点 P_1^{max} 和 P_2^{max} , $g_{P_1^{max}} < g_{P_L^{min}} < g_{P_2^{max}}$, 构建灰度区间为 $(g_{P_L^{max}}, g_{P_C^{max}}, g_{P_C^{max}})$ 的水域 R_{valley} , 并由式 (10) 计算水域面积 $A = area(g_{P_L^{max}}, g_{P_R^{max}})$ 。

(4) 阈值和峰值判定

1) 如果 $A > \Theta$ (Θ 为某一限值), 则认为灰度区间为 $(g_{P_L^{max}}, g_{P_C^{max}}, g_{P_R^{max}})$ 的水域 R_{valley} 为理想分割水域, P_C^{min} 为理想分割阈值, 将其存入阈值数组 T 中, 同时将相应的左侧极大值 $g_{P_L^{max}}$ 存入峰值数组 P 中; 分别取 P_R^{max} , P_R^{min} , P_R^{max} 作为新水域的左侧极大值点、最小值点和右侧极大值点。

2) 如果 $A < \Theta$, 则认为该水域为伪分割水域, 比较两侧极大值点 P_L^{max} 和 P_R^{max} 的大小:

① 如果 $g_{P_L^{max}} > g_{P_R^{max}}$, 则取 P_L^{max} 为新水域的左侧极大值点, 取 P_R^{max} 为新水域的右侧极大值点; 再比较 P_C^{min} 与 P_R^{min} 的灰度值, 取小者作为新水域的最小值点。

② 如果 $g_{P_L^{max}} \leq g_{P_R^{max}}$, 则分别取 P_R^{max} , P_R^{min} , P_R^{max} 作为新水域的左侧极大值点、最小值点和右侧极大值点。

(5) 由式 (10) 计算新水域的面积, 并重新执行步骤 (4), 直至所有的极小值点被处理完。

最终获得的阈值数组 $\{T_k, k = 1, 2, \dots, M\}$ 中就保存着图像的 M 个阈值, 并且有 $g_{T_1} < g_{T_2} < \dots < g_{T_M}$, 同

时还获得峰值数组 $\{P_m, m = 1, 2, \dots, M + 1\}$, 且 $g_{P_1} < g_{P_2} < \dots < g_{P_{M+1}}$ 。由此可知, 一个灰度区间为 $(g_{P_m}, g_{T_m}, g_{P_{m+1}})$ 的有效的水域 R_{valley} 包括两个相邻的峰值点 P_m, P_{m+1} 和一个阈值点 $T_m (g_{P_m} < g_{T_m} < g_{P_{m+1}})$ 。

3.4 自适应参数设定方法

该算法的基本思想是从最大的阈值 g_{T_M} 开始, 利用 PCNN 算法将平均值大于 g_{T_M} 的目标首先从图像中提取出来, 即保持 g_{T_M} 不变用 PCNN 算法进行迭代计算, 直至不再有新的神经元被点火。试验发现, 在同一个阈值处的迭代次数可根据图像特性不同而有一定差异, 通常为 20~70 次。在最大阈值点 T_M 处的迭代完成后, 再将所有自然点火和受激点火的神经元收集到一起, 这就是 PCNN 算法分割的第 1 个目标, 也是亮度最大的目标。接下来, 就可以在 T_{M-1} 处采用同样的算法分割出第 2 个目标, 如此反复, 直至所有的阈值被分析完毕。最后就获得了总共 $M + 1$ 个亮度不同的目标。

4 特征提取

4.1 显著性因子

通过对同源和异源图像融合的实例分析发现, 对同一目标而言, 它在不同源图像中的边缘清晰程度与相应源图像在融合过程中的贡献系数成正比。本文依据目标区域边界相对于背景邻域的突出性, 提出显著性因子作为区域的提取特征之一。

对于各分割区域 R_m , 先将其边缘进行 8 邻域扩张, 形成一条宽约为 3 个像素的边缘带, 然后求取带内所有属于和不属于该目标区域的像素平均灰度值之差, 其绝对值即定义为相应目标区域的显著性因子:

$$S = \left| \frac{1}{M} \sum_{i=1}^M f_i - \frac{1}{N} \sum_{j=1}^N g_j \right| \quad (11)$$

式中, f_i 为区域边缘带上任一属于该区域的像素灰度值, M 为其像素总个数; g_j 为区域边缘带上任一不属于该区域的像素灰度值, N 为其像素总个数。

4.2 可见性因子

根据人眼生理视觉特性, 本文采用的可见性因子 V 定义如下^[10]:

$$V_{R_m, i} = \frac{1}{H} \sum_{(x, y) \in R_{m, i}} \left| \frac{1}{m_i} \right|^\alpha \frac{|f(x, y) - m_i|}{m_i} \quad (12)$$

式中, α 为视觉常数 (0.6~0.7), H 是分割区域 $\sum R_{m, i}$ 中的像素总数, m_i 为区域中像素平均灰度值。以显著性因子和可见性因子作为区域融合依

据,通过自适应确定权值即可得到融合图像。

图 4 为未精确配准的多聚焦图像融合实例。由图 4 可见,图 4(a)所示的图片左前方的物体聚焦,

右后方的物体离焦,而图 4(b)所示的图片则刚好相反。采用人工剪切和粘贴的方法合成的聚焦清晰的参考图片如图 4(c)所示。

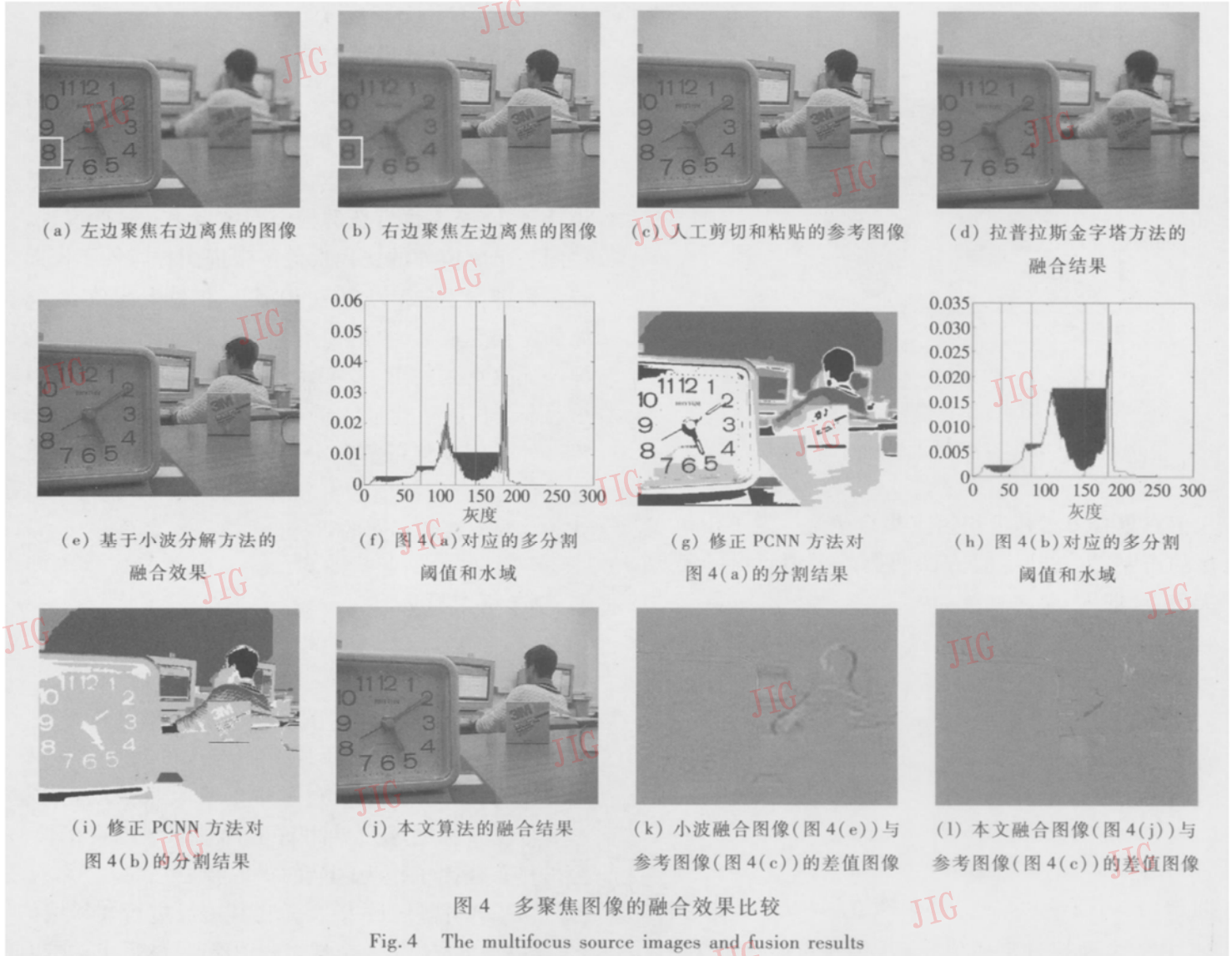


图 4 多聚焦图像的融合效果比较

Fig. 4 The multifocus source images and fusion results

本文分别对图 4(a)和图 4(b)中表盘字符“8”所在区域的 S 值和 V 值进行计算,实验结果见表 1。由表 1 可知若对应区域越清晰,则其 S 值就越大,且 S 值与视觉效果吻合良好。

表 1 显著性因子和可见性因子示例

Tab 1 S and V values for special region in Fig 4

区域	S		V	
	图 4(a)	图 4(b)	图 4(a)	图 4(b)
图 4 中表盘字符“8”	53 002	13 668	0 036 18	0 007 29

5 实验与讨论

5.1 多分辨分析融合方法

目前,基于多分辨分析的图像融合算法是多传感

器图像融合领域应用非常广泛和极其重要的一类。本文利用拉普拉斯金字塔分解融合法和小波分解融合法与本文方法进行了融合对比实验。其中小波变换采用 Daubechies 的 8 系数小波,分解层数为 3 并采用基于区域能量匹配度的融合规则^[2]进行融合, α 取值为 0.7。金字塔分解采用拉普拉斯金字塔,分解层数为 3 采用灰度值平均融合规则进行融合。

5.2 融合效果客观评价准则

为了更好地评价不同方法的融合效果,本文引入了一些用于融合图像效果客观评价的准则,其中包括均方根误差 (root mean square error, RMSE)、灰度标准差 σ_g 、交叉熵 (cross entropy, CEN) 和互信息 (mutual information, MI), 各参数定义如下:

$$RMSE = \sqrt{\frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [R(i, j) - F(i, j)]^2}$$

$$CEN(P:Q) = \sum_{g=0}^{L-1} p(g) \log_2 \frac{p(g)}{q(g)}$$

$$\sigma_g = \sqrt{\sum_{g=0}^{L-1} \left(g - \sum_{g=0}^{L-1} gp(g) \right)^2 p(g)}$$

$$MI = \sum_{i=1}^L \sum_{j=1}^L h_{R,F}(i,j) \log_2 \frac{h_{R,F}(i,j)}{h_R(i)h_F(j)}$$

其中, $R(i, j)$ 和 $F(i, j)$ 分别为理想融合图像(参考图像)和实际融合图像 (i, j) 点的值, M 和 N 为图像的尺寸, L 为图像的灰度级数, $P = \{p(0), p(1), \dots, p(g), \dots, p(L-1)\}$ 和 $Q = \{q(0), q(1), \dots, q(g), \dots, q(L-1)\}$ 分别为两幅图像的概率分布, $p(g)$ 为灰度等于 g 的像素数与图像总像素数的比值, $h_{R,F}(i, j)$ 为图像 R 和 F 间的归一化联合灰度直方图分布, $h_R(i)$ 和 $h_F(j)$ 分别为两幅源图像的归一化边缘直方图分布。均方根误差越小, 说明融合图像与理想图像越接近; 交叉熵直接反映了两幅图像对应像素的差异, 若差异越小, 则融合方法从原始图像中提取的信息越多; 灰度标准差反映了相对灰度均值的离散状况, 即标准差越大, 灰度分布越分散; 互信息衡量的是融合图像对标准参考图像不确定性的减少量, 故其值越大越好。

5.3 同源多传感器图像融合实验

以多聚焦图像为例。图 4 给出了非精确配准同源多传感器图像融合效果比较, 由图 4 可见, 图 4(a)和图 4(b)中学生的头部有相对位移。采用不同融合方法所得融合结果见图 4(d)~图 4(f)。图 4(k)、图 4(l)分别为基于小波变换的融合图像、

基于本文方法的融合图像与参考图像的差值图。

由图 4 可以看出, 利用本文算法所得的融合图像边缘最清晰、细节最丰富, 由于人的头部存在相对运动, 利用拉普拉斯金字塔和小波变换的融合方法所得的融合效果中该位置都较模糊, 而在用本文算法所得的融合图像中则基本无此现象, 这是因为本文提出的基于修正 PCNN 的融合方法充分考虑到像素间的相关性, 克服了多分辨分析方法割裂像素间联系的不足的缘故, 这一点在差值图像(图 4(k)和图 4(l))中体现得更加明显。

根据融合效果客观评价公式计算出的量化指标如表 2 所示。

表 2 图 4 中不同方法的图像融合效果对比

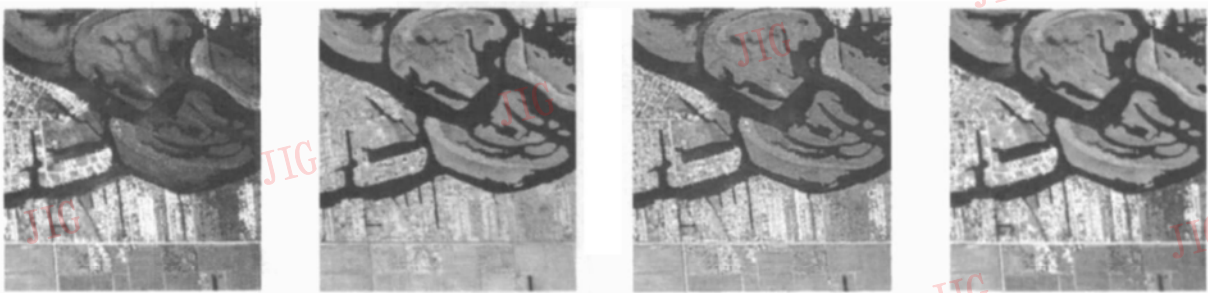
Tab 2 Performance of the different methods on Fig 4

融合方法	均方根误差	灰度标准差	互信息量
拉普拉斯金字塔	7.1126	45.5013	7.0815
小波变换	4.6622	46.7919	6.7239
本文方法	3.5002	47.7826	8.3126

从表 2 可知, 利用本文方法融合后的图像的均方根误差最小, 灰度标准差和互信息量最大, 优于基于多分辨分析的融合效果。通过人眼的主观判断也可以确定, 基于本文方法的融合结果最清晰。

5.4 异源多传感器图像融合实验

如图 5 所示, 针对不同波段的遥感图像融合示例, 融合后的图像同时具有两种不同波段源图像的信息。融合效果的客观评价参数见表 3。



(a) 遥感波段 1 图像 (b) 遥感波段 2 图像 (c) 基于小波分解的融合结果 (d) 本文算法的融合结果

图 5 某地区多波段遥感图像的融合效果比较

Fig. 5 The multi-wave band remote sensor source images and fusion results

表 3 图 5 中不同方法的图像融合效果对比

Tab 3 Performance of different fusion methods on Fig 5

融合方法	交叉熵	灰度标准差	互信息量
小波变换	2.1765	65.3594	3.6022
本文方法	0.3632	67.4870	5.8372

综上所述, 由主观视觉效果和客观评价量化指

标上都可以看出, 本文方法融合效果优于基于小波变换的融合方法。

6 结论与展望

脉冲耦合神经网络是近年来逐渐兴起的一个研

究热点,其理论模型和应用领域都有待进一步深入研究和拓展。本文提出一种将 PCNN 应用于多传感器图像融合的具体方法,并针对传统脉冲耦合神经网络参数无法自动设定的难题,提出了基于修正 PCNN 的参数自动设定方案。

结合多传感器图像融合的大量实验结果表明,本文方法充分考虑到了像素间的相关性,其无论在主观视觉效果,还是客观评价参数上均优于基于多分辨分析的融合算法,尤其适用于源图像无法严格配准的应用场合。此外,本文提出的水域面积算法对于成功解决 PCNN 参数难以自动确定的难题具有理论和实际应用价值。

参考文献 (References)

- 1 Hall D L, Llinas J. An introduction to multisensor data fusion [J]. Proceedings of the IEEE, 1997, **85**(1): 6~ 23
- 2 Zhang Z, Blum R S. A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application [J]. Proceedings of IEEE, 1999, **87**(8): 1315~ 1326
- 3 Eckhom R, ReilBoeck H J. Feature linking via synchronization among distributed assemblies: simulation of results from cat visual cortex [J]. Neural Computation, 1990, **2**(3): 293~ 307.
- 4 Kuntimad G, Ranganath H S. Perfect image segmentation using pulse coupled neural networks [J]. IEEE Transactions on Neural Networks, 1999, **10**(3): 591~ 598
- 5 Ma Y rde, Dai Ruo-lan, Li Lian. Automated image segmentation using pulse coupled neural networks and image's entropy [J]. Journal of China Institute of Communications, 2002, **23**(1): 46~ 51. [马义德, 戴若兰, 李廉. 一种基于脉冲耦合神经网络和图像熵的自动图像分割方法 [J]. 通信学报, 2002, **23**(1): 46~ 51.]
- 6 Liu Q in, Ma Y rde, Qian Zhirbai. Automated image segmentation using improved PCNN model based on cross-entropy [J]. Journal of Image and Graphics, 2005, **10**(5): 579~ 584. [刘勃, 马义德, 钱志柏. 一种基于交叉熵的改进型 PCNN 图像自动分割新方法 [J]. 中国图象图形学报, 2005, **10**(5): 579~ 584.]
- 7 Gu X D, Guo S D, Yu D H. A new approach for automated image segmentation based on unit-linking PCNN [A]. In: Proceedings of the first International Conference on Machine Learning and Cybernetics [C], Beijing, China, 2002. 175~ 178
- 8 Bi Ying-wei, Qiu Tian-shuang. An adaptive image segmentation method based on a simplified PCNN [J]. ACTA Electronica Sinica, 2005, **33**(4): 647~ 650. [毕英伟, 邱天爽. 一种基于简化 PCNN 的自适应图像分割方法 [J]. 电子学报, 2005, **33**(4): 647~ 650.]
- 9 Karvonen J A. Baltic sea ice SAR segmentation and classification using modified pulse-coupled neural networks [J]. IEEE Transactions on Geoscience and Remote Sensing, 2004, **42**(7): 1566~ 1574
- 10 Huang J W, Shi Y Q, Dai X H. A segmentation-based image coding algorithm using the features of human vision system [J]. Journal of Image and Graphics, 1999, **4**(5): 400~ 404. [黄继武, Shi Y Q, 戴宪华. 基于视觉系统特性的图像分割编码算法 [J]. 中国图象图形学报, 1999, **4**(5): 400~ 404.]

一种评价仿射不变性特征定位误差的新方法

邓宝松 高宇 魏迎梅 吴玲达

(国防科学技术大学多媒体研究开发中心, 长沙 410073)

摘要 特征点匹配是计算机视觉中的一个基本问题。不同视点图像中相应特征点邻域窗口之间存在几何上的透视畸变, 这可用平面单应映射来表示, 而目前大多匹配算法将该映射用仿射变换模型来近似, 即用具有仿射不变性的特征进行图像的匹配。仿射变换的线性特点不仅能降低算法复杂度, 还能保证迭代过程收敛的稳定性, 然而并没有人对该近似的可行性及如何减小近似误差给出定量分析。本文首先回顾了各种几何层次上的特征点匹配策略, 重点针对具有仿射不变性特征点的定位误差进行研究和定量分析, 通过本文提出的椭圆曲线规范化法推导出该近似所造成相应特征点定位误差的解析表达; 然后用真实图像的实验结果验证了本文分析方法的必要性和正确性; 最后给出相应的分析结果和结论, 并指出提高大基线图像特征点匹配精度的相应措施。

关键词 特征点匹配 单应映射 仿射不变性 椭圆曲线 定位误差

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2008)02-0291-07

A Novel Method for Evaluating the Location Error of Affine Invariant Features

DENG Bao-song GAO Yu WEI Yingmei WU Ling-da

(Multimedia Research & Development Center, National University of Defense Technology, Changsha 410073)

Abstract Feature matching is a fundamental problem in computer vision. There are geometrical deformations named plane homography between corresponding feature windows on different images of the same patch on scene surface. But most of the state-of-the-art matching schemes approximate the homography with an affine transformation model, namely affine invariant features. The linear character of affine model not only reduces the complication of matching process, but ensures a good and stable convergence of the unmanageable iteration. However, there are no quantificational discussions about this approximation by far. In this paper, an overview of invariant features under different geometric groups was given first, and then a quantificational analysis of affine invariant features was proposed. The analytic expression of location error was deduced by our novel method, named normalization of conic equations. Experimental results of real images demonstrate the correctness and necessity of our method. Finally, some valuable conclusions and suggestions based on the above analysis are deduced both on theory and in practice.

Keywords feature point matching homography affine invariant elliptical conic location error

1 引言

特征点匹配是计算机视觉中的基本问题, 在 3 维重建、目标识别以及图像数据库检索等领域都有着重要的研究和应用价值^[1-3]。特征点是图像平面

内亮度沿不同方向变化均显著的局部极值点^[4], 作为图像中的局部低层特性, 其较小的邻域可以作为基于亮度相关性图像匹配方法的特征窗口(下称邻域窗口), 这使得它们能够对图像的几何、亮度变化以及场景的相互遮挡具有很好的鲁棒性, 因此非常适合作为图像匹配的基元。

基金项目: 国家自然科学基金项目(60473117)

收稿日期: 2006-06-30 改回日期: 2006-09-26

第一作者简介: 邓宝松(1978~), 男, 2006 年于国防科学技术大学获工学博士学位。主要研究方向为计算机视觉、虚拟现实及可视化。

E-mail: dbs310@163.com

虽然目前对小基线图像的匹配以及视频序列中的特征跟踪取得了满意的结果^[5], 然而对大基线图像的特征匹配却是目前研究的难点^[6,7], 其根源在于图像间较大的透视畸变使得很难自适应地矫正相应邻域窗口的局部形状^[8]。为了能够很好地表示特征点及其邻域在不同图像上所成的像, 使之不受视点和成像参数变化的影响, 需要给出这些特征的不变性描述和表示方法, 得到这些描述就可以在两幅或多幅图像之间进行特征的匹配^[3,9]。

为了保持本文概念的完整性以及后面论述的需要, 这里先给出如下定义:

不变性特征 是指图像上的某些局部自适应区域, 该区域在不同图像上的形状可以随视点的变化而变形, 从而保证该区域是场景中同一块物体表面所成的像。如果这些自适应区域之间可以用某种几何变换来表示, 则称为具有该层次几何不变性的特征。不变性特征按照所允许的几何变换类型可分为欧氏不变性 (euclidean invariant)、相似不变性 (similar invariant)、仿射不变性 (affine invariant) 和射影不变性 (projective invariant), 显然, 四者所涵盖的变换范围逐渐变大, 适应性也逐渐增强。

2 图像几何畸变模型

如图 1 所示, 如果将光滑场景表面上点 X_0 周围的局部邻域面片 S 近似认为平面, 其在两幅图像上所成的像窗口之间满足 2 维平面上的单应映射 (用 3×3 单应矩阵 H 表示), 而该映射将图像 I 上分布均匀的圆形邻域窗口 w 变换到图像 I' 上分布不均匀的椭圆形窗口 w' 。 H 由两相机内外参数及平面方程唯一确定, 得到该变换矩阵便可以实现相应特征点邻域窗

口的完全透视矫正, 否则传统基于窗口亮度相似性的特征点匹配方法很难直接进行^[10-11], 特别是当两幅图像之间的基线较大时, 这种几何畸变尤为明显。

设图像 I 上特征点 x_0 的邻域窗口内某点坐标为 x (以 x_0 为原点), 其对应点在图像 I' 上的坐标为 x' (以 x' 为原点), 则二者之间的映射关系为

$$\tilde{x}' \sim H\tilde{x} \sim \begin{bmatrix} A & 0 \\ v^T & 1 \end{bmatrix} \tilde{x} \quad (1)$$

式中, \tilde{x} 和 \tilde{x}' 分别表示 x 和 x' 的齐次坐标, A 为 2×2 矩阵, 表征两邻域窗口间的仿射变换; 而 $v = (v_1, v_2)^T$ 是 2 维向量, 表征两邻域窗口间射影变形大小。如果把式 (1) 表示成非齐次坐标形式:

$$x' = \frac{1}{v^T x + 1} \cdot Ax = \mu \cdot Ax \quad (2)$$

可以看出, 相对于仿射变换, 单应映射增加一个系数 μ 而 μ 本身是与坐标 x 相关的, 距离原点越远的像素点, μ 与 1 的偏差越大。单应映射本身既不是奇函数, 也不是偶函数, 它将均匀分布的圆形窗口映射为非均匀分布的椭圆形窗口。一般来讲, 目前所有特征点匹配方法 (包括大基线图像) 都是针对 v 量级较小的情况进行, 否则必将使 μ 变得非常大, 从而使两邻域窗口之间的射影畸变无法接受。虽然单应矩阵 H 能够完全表征这种映射关系, 但其具有 8 个自由度^[12], 巨大的搜索空间及其非线性特性使其难以真正用于实际问题中^[6]。

3 相关工作

特征匹配过程有两种途径实现: 一种是分别在每幅图像中单独检测具有某种不变性的特征, 然后将其变换到规范条件下, 用多维向量进行表示和描述, 匹配时通过向量间的距离衡量特征的相似性^[3,13-15]; 另一种是依据给定的几何变换模型, 直接将一幅图像上的邻域窗口变换到与另一幅图像上相应邻域窗口完全一致的形状^[5,16-18], 最后用相应位置像素的亮度相关性进行判断。这两种途径的实质是一样的, 即都要通过假定的变换模型矫正相应邻域窗口之间的透视畸变, 因此这里并没有对这两种策略进行区分。一般来讲, 第 1 种途径适用于图像检索、目标识别等非精确匹配应用中, 如需精确定位, 一般还需要将该方法的匹配结果作为初值进一步通过第 2 种途径迭代求精^[13,19]。

在没有任何先验信息的条件下, 进行特征点匹

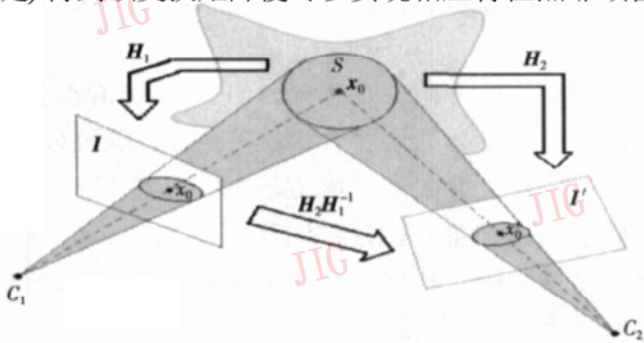


图 1 场景局部面片在两幅图像上所成像之间满足单应映射关系

Fig 1 Homography between feature windows on different images

配都是先假定相应邻域窗口之间的几何变换模型,进一步估计模型参数,从而实现相应邻域窗口的透视畸变矫正。依据几何变换的层次,相应的变换模型可分为欧式变换(3自由度)、相似变换(4自由度)、仿射变换(6自由度)和射影变换(8自由度),如图 2 所示。

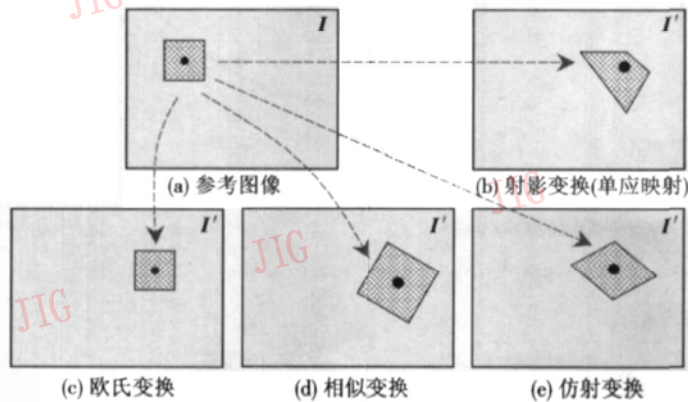


图 2 不同几何层次上特征点的匹配示意图

Fig 2 Matching process of feature windows under different geometrical groups

最早的直接匹配方法假定相应邻域窗口间仅仅是纯平移运动,而直接基于特征点周围固定大小窗口的亮度相似性进行匹配^[10-11]。由于没有考虑几何上的透视畸变,这种方法只适用于小基线或仅由相机侧移产生的图像中,如图 2(c)所示。

后来研究人员允许图像间的旋转和比例放缩,即采用自适应的旋转角度和窗口大小进行匹配。Schmid 考虑了相应邻域窗口间的旋转变换,提出具有旋转不变性(rotation invariant)的特征检测和匹配方法^[15],由于考虑的模型相对简单,该方法仅适用于图像之间透视畸变不大的情形;Dufoumaud 将该模型扩展到允许有较大比例变化的情况下,实现了不同分辨率图像之间的精确匹配^[20],被称为具有比例不变性(scale invariant)的特征检测和匹配^[13]。上述方法均可归类于具有相似不变性的特征检测和匹配方法,即用相似变换来近似相应邻域窗口间的透视畸变,如图 2(d)所示。

随着图像间透视畸变的增大,相似变换已不能胜任大基线图像匹配的要求,因此又提出具有仿射不变性的特征检测和匹配,即用仿射变换来近似单应映射,如图 2(e)所示,这也是目前该领域的研究热点。Mikolajczyk 通过对特征点邻域窗口内二阶矩矩阵的规范化给出具有仿射不变性的特征点检测方法,并用仿射自适应的 Harris 算子进行定位^[21],类

似研究还很多,这类方法都利用了二阶矩矩阵的不变性^[8-14],但这一假设对任意图像来说并不总是成立的^[13]。为确定特征区域的大小, Schaffalitzky 借鉴 Mikolajczyk 的方法^[3],计算邻域窗口在比例空间内的局部极值^[14];而 Tuytelaars 则依据场景中的特殊结构检测图像中具有仿射不变性的特征区域:局部边缘交点即角点的邻域以及亮度一致的局部区域。这些方法不能保证单独检测的相应不变性区域是完全重合的^[22],因此定位精度很低,一般还需要迭代求精^[13]。Shi 直接通过基于仿射变换模型的迭代解决了序列图像中的特征点精确跟踪^[5],而 Xiao 在此基础上用两阶段法解决了大基线图像的匹配问题并提高了定位精度^[16]。仿射变换由于其良好的线性特性以及几何上的直观性受到众多研究者的青睐,在实际应用中也不断得到证实^[9-22]。

虽然 Pritchett 使用具有射影不变性的完全单应映射进行特征点的匹配,如图 2(b)所示,但该方法要求场景中存在若干平面型结构且有特殊的四边形区域以便于计算单应矩阵,极大地限制了其适用范围^[7];用类似的方法, Lourakis 还实现了线特征的匹配^[12],同样,对场景结构的要求使得这类方法很难真正用于实际问题。

图像的匹配从侧面反映了计算机视觉的逆向性,即在未知几何变换及场景结构的前提下来估计 2 维图像平面的变换参数;然而,也正是由于特征的局部性使得以上这些几何变换模型能够很好地相互近似,而众多的特征使得这种较小的、分布均匀的近似误差可以被忽略;当然,也正是这些误差使得鲁棒性估计对计算机视觉变得异常重要^[23],因为该过程中充满了近似性条件。

4 问题描述及定位误差分析

虽然仿射变换具有众多的良好性质,但用仿射变换代替单应映射毕竟是一种近似。如图 3 所示,在单应映射和仿射变换作用下,虽然同一圆形邻域窗口均变换为椭圆形邻域窗口,然而,由于单应映射本身是非对称的,这一点从式(2)中也可看出,圆心 O 的像点 O' 并不会位于椭圆的几何中心,而基于仿射变换模型的匹配方法均是椭圆的几何中心作为特征点的正确位置^[5-6, 13-14, 16-19, 21],这显然是不合适的,其中的近似性必然会对相应特征点的定位产生一定的误差。

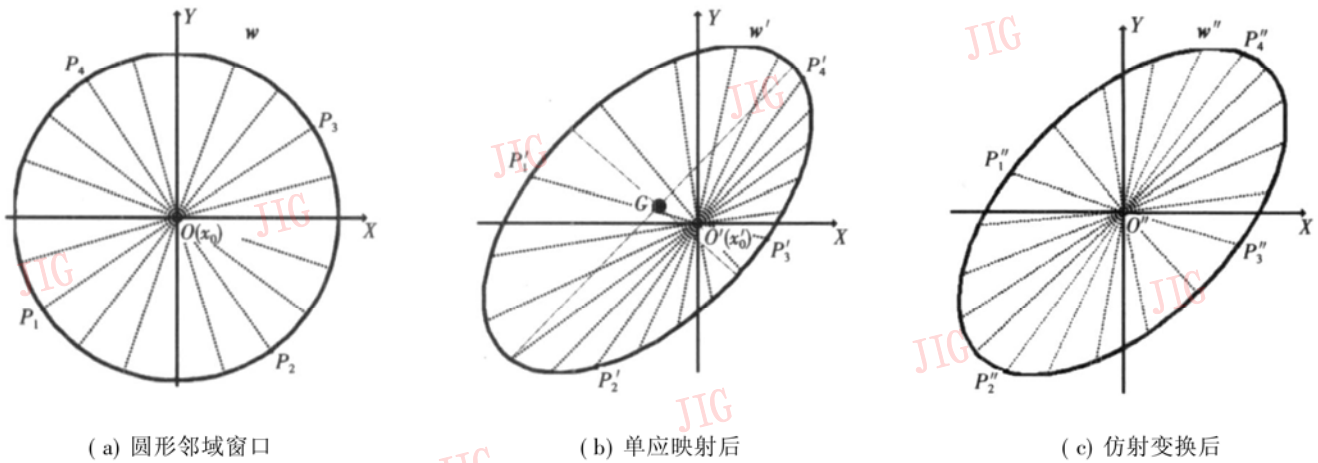


图 3 圆形邻域窗口在单应映射和仿射变换后的形状变化

Fig 3 The difference of circular feature windows under homography and affine transformation

虽然仿射变换模型良好的线性特性使其在特征匹配时成为众多研究者的首选,但这种近似导致的误差究竟有多大?这种近似关系成立的条件是什么?分析这种近似关系对我们进行基于仿射变换模型的特征匹配,特别是大基线图像的特征点匹配有什么指导意义?到目前为止,笔者并没有见到任何文献给出对这种仿射变换近似可行性的讨论,故下面即以此为背景,探讨并定量分析这种近似的可能误差,并给出相应解析表达,进一步回答上面的问题并给出相关结论和改进措施。

如图 3(a)所示,对于圆形邻域窗口 w 的边界点 $\tilde{x} = (x, y, 1)^T$ (以圆心 O 即 x_0 为原点), 满足二次曲线方程 $\tilde{x}^T C_e \tilde{x} = 0$ 其中,

$$C_e = \begin{bmatrix} 1 & & \\ & 1 & \\ & & -r^2 \end{bmatrix} \quad (3)$$

如图 3(b)所示,经过单应映射 H 后,第 2 幅图像上相应邻域窗口边界点为 $\tilde{x}' = (x', y', 1)^T$ (以相应特征点 O' 即 x'_0 为原点), 则有

$$\begin{aligned} \tilde{x}'^T C_e \tilde{x}' &= \tilde{x}'^T H^T H^{-T} C_e H^{-1} H \tilde{x}' \\ &= (H \tilde{x}')^T (H^{-T} C_e H^{-1}) (H \tilde{x}') \\ &= \tilde{x}'^T C_e \tilde{x}' = 0 \end{aligned} \quad (4)$$

式中

$$C_e = H^{-T} C_e H^{-1} = \begin{bmatrix} c_{11} & c_{12} & c_1 \\ c_{12} & c_{22} & c_2 \\ c_1 & c_2 & c_0 \end{bmatrix} \quad (5)$$

C_e 为 3×3 对称矩阵,表示第 2 幅图像上以相应特征点 x'_0 为原点的坐标系下窗口边界的曲线方

程 $c_{11}x'^2 + 2c_{12}x'y' + c_{22}y'^2 + 2c_1x' + 2c_2y' + c_0 = 0$ 此时的曲线实际是一个椭圆,但其几何中心 G 不再与原点 O' 即特征点 x'_0 重合。传统基于仿射变换模型近似单应映射的方法就是忽略了这一差异,下面就是要通过椭圆曲线规范化法定量地讨论和分析这种近似对特征点定位所造成的误差。令 C_e 左上角的 2×2 矩阵为 B , $\delta = [c_1, c_2]^T$, 则矩阵 C_e 可以分块表示为

$$C_e = \begin{bmatrix} B & \delta \\ \delta^T & c_0 \end{bmatrix} \quad (6)$$

为了消去二次方程中的交叉项,可以对曲线进行转轴操作,设需要的旋转角为 θ 则旋转矩阵为 $R = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$, 在新的坐标系下该二次曲线的方程表示为

$$C'_e = \begin{bmatrix} R^T & 0 & B & \delta \\ 0^T & 1 & \delta^T & 1 \end{bmatrix} \begin{bmatrix} R & 0 \\ 0^T & 1 \end{bmatrix} = \begin{bmatrix} R^T B R & R^T \delta \\ \delta^T R & c_0 \end{bmatrix} \quad (7)$$

矩阵 C'_e 仍为对称矩阵,经过转轴操作,新方程二次项系数只与原方程二次项系数及转角 θ 有关;新方程一次项系数只与原方程一次项系数和转角 θ 有关,由于 $\|R^T \delta\| = \|\delta^T R\| = \|R\| \cdot \|\delta\| = \|\delta\|$, 因此旋转变换不改变一次项系数的数量级且常数项保持不变。显然,可以选取适当的转角 θ 来消去交叉项。新方程交叉项系数的一半即矩阵 $R^T B R$ 的第 1 行 2 列 (或第 2 行 1 列) 元素为

$$\begin{aligned} &(c_{22} - c_{11}) \sin\theta \cos\theta + c_{12} (\cos^2\theta - \sin^2\theta) \\ &= \frac{1}{2} (c_{22} - c_{11}) \sin 2\theta + c_{12} \cos 2\theta \end{aligned} \quad (8)$$

若 $c_{12} \neq 0$ 交叉项系数为零的充要条件为

$$c \tan 2\theta = \frac{c_{11} - c_{22}}{2c_{12}} \tag{9}$$

一旦得到满足式 (9) 的转角 θ 新方程就变为

$$c'_{11}x''^2 + c'_{22}y''^2 + 2c'_1x'' + 2c'_2y'' + c'_0 = 0 \tag{10}$$

即 $R^TBR = \begin{vmatrix} c'_{11} & \\ & c'_{22} \end{vmatrix}$, 从而有

$$\begin{vmatrix} c_{11} & c_{12} \\ c_{12} & c_{22} \end{vmatrix} \begin{vmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{vmatrix} = \begin{vmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{vmatrix} \begin{vmatrix} c'_{11} & 0 \\ 0 & c'_{22} \end{vmatrix} \tag{11}$$

比较式 (11) 两边矩阵的主对角元素可以得到:

$$\begin{aligned} c'_{11} &= c_{11} + c_{12} \tan\theta \\ c'_{22} &= c_{22} - c_{12} \tan\theta \end{aligned} \tag{12}$$

由于椭圆形二次曲线方程中两个平方项系数是同号的, 此时可以将式 (10) 进一步配方得到:

$$c'_{11} \left[x'' + \frac{c'_1}{c'_{11}} \right]^2 + c'_{22} \left[y'' + \frac{c'_2}{c'_{22}} \right]^2 = \frac{c'^2_1}{c'_{11}} + \frac{c'^2_2}{c'_{22}} - c'_0 \tag{13}$$

该椭圆的几何中心为 $\left[-\frac{c'_1}{c'_{11}}, -\frac{c'_2}{c'_{22}} \right]$, 从中可以看出椭圆几何中心到原点的距离与椭圆方程中一次项系数与二次项系数的比值有关, 该方程中一次项系数的量级越小, 椭圆几何中心就会越接近坐标原点。本节的目的就是要估计椭圆几何中心 G 偏离

原点 O' 即特征点 x'_0 的距离, 从椭圆曲线规范化后的结果中即可得到相应特征点的定位误差:

$$\epsilon_i = \sqrt{\left| \frac{c'_1}{c'_{11}} \right|^2 + \left| \frac{c'_2}{c'_{22}} \right|^2} \tag{14}$$

5 实验结果

对上述算法进行实现, 以公开的测试图像为例给出了相应的实验结果, 图像之间的单应映射矩阵是由 Visual Geometry Group 提供^[24]。为了更能说明误差的量级和变化情况, 这里考虑了两种误差: 特征点定位误差, 即椭圆邻域窗口的几何中心与特征点之间的误差; 邻域窗口内部像素定位误差, 即由仿射近似所导致的邻域窗口内所有像素点的定位误差分布。

5.1 特征点的定位误差

由于版面空间的限制, 图 4 和图 5 列举了一组 (3 幅) 图像中相应特征点定位误差分布的例子。其中第 1 幅图像中所有特征点的邻域窗口半径取为 15 像素, 而后两幅图像的椭圆形邻域窗口与第 1 幅图像中的相应圆形邻域窗口对应的是同一块物理表面区域。图 5 给出了这组图像中相应特征点的定位误差分布情况。

总的来讲, 即使本文选取的 3 幅图像之间透视畸变比较大 (尤其在第 1 幅和第 3 幅之间), 但其相

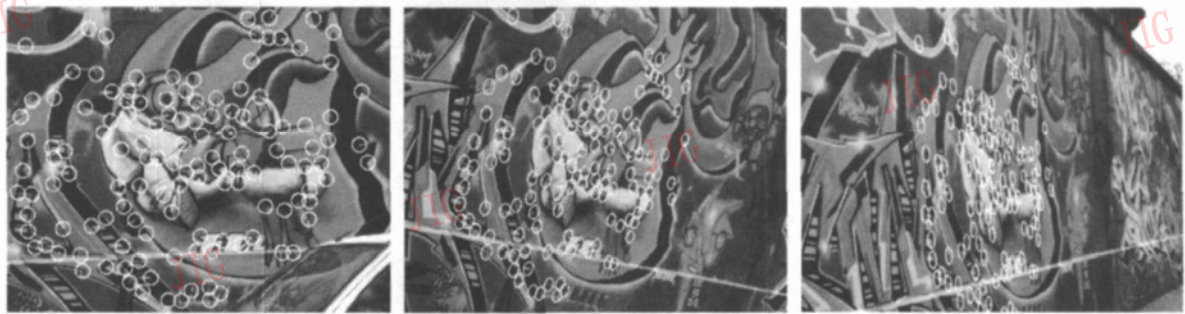


图 4 在仿射变换模型下 3 幅图像的相应特征点及其邻域窗口

Fig 4 The corresponding feature windows on three images under affine transformation

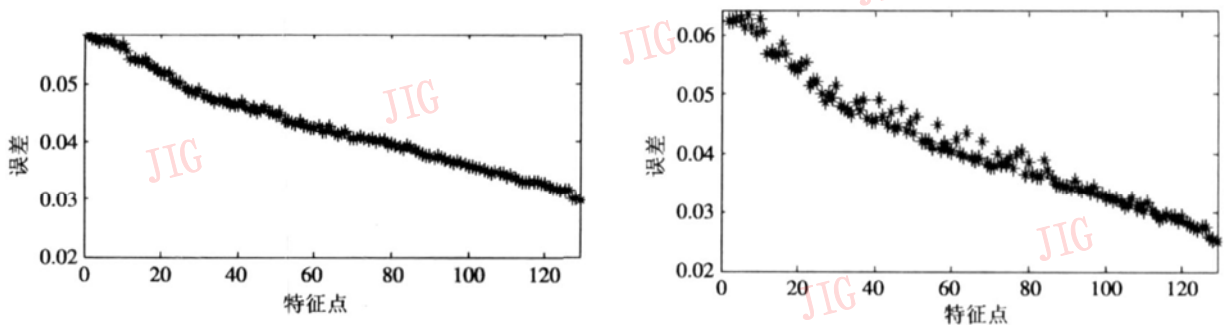


图 5 相对于第 1 幅图像, 后两幅图像的特征点定位误差分布

Fig 5 Location error of feature points relative to the first images

应特征点定位误差的量级还是非常低的,基本控制在 0.1 个像素之内。一般来讲,这种定位误差都会随着两幅图像之间透视畸变的增大而增加。更多的实验结果可以参见笔者论文的相关主页 ([http // xuefeng310 home4u china cn](http://xuefeng310.home4u.china.cn))。

5.2 邻域窗口内部的定位误差

为了进一步分析用仿射变换模型近似单应映射所产生的影响,估计整个邻域窗口内相应像素的定位误差,其计算方法是:采样圆形邻域窗口内每一个像素点,将其分别在单应映射以及仿射变换(通过二次曲线规范化法得到)的作用下变换到第 2 幅图像上的相应邻域窗口内,然后计算这两个坐标点之

间的欧氏距离作为这一像素点的定位误差,从而衡量仿射变换对整个邻域窗口内的点可能造成误差的分布情况。

如图 6 给出两组特征点邻域窗口内像素定位误差的分布示意图,其中,(a)为图 4 第 1 幅图像上特征点的圆形邻域窗口;(b)为图 4 第 2 幅图像上相应椭圆形邻域窗口;(c)为仿射变换在图 4 第 2 幅图像上对窗口内所有像素点的定位误差分布(相对于图 4 第 1 幅);(d)为图 4 第 3 幅图像上相应椭圆形邻域窗口;(e)为图 4 仿射变换在图 4 第 3 幅图像上对窗口内所有像素点的定位误差分布(相对于图 4 第 1 幅)。为使图像更清楚,这里圆形邻域窗口半径取为 25 像素。

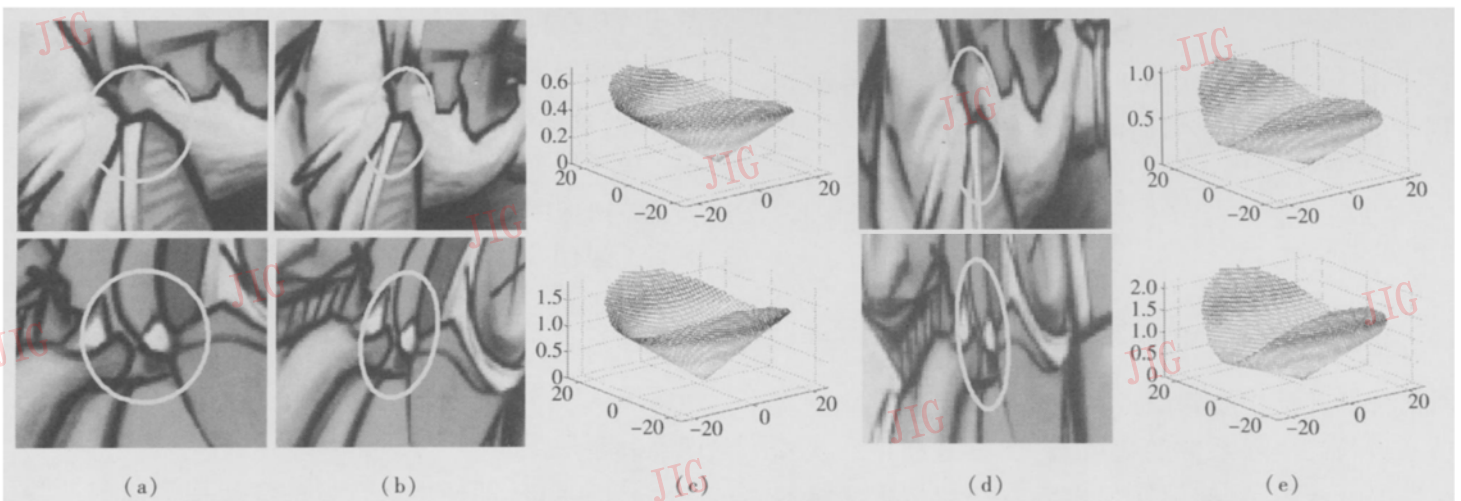


图 6 仿射变换近似对整个窗口的像素造成的定位误差

Fig.6 Location error of all pixel in the same feature windows using affine transformation approximation

从图中可以清楚地看到,随着像素点到特征点距离的增加(透视畸变也在增大),仿射变换对其造成的定位误差急剧增加,因此可以肯定地说,用仿射变换近似单应映射仅仅适用于较小的邻域窗口,随着窗口的增大,这种近似所产生的误差将不容忽视,因此在选择特征点的邻域窗口时要避免过大。当然,太小的邻域窗口又不足以区分不同的特征点,特别是在进行大基线图像的特征点匹配时。

此外还发现,坐标原点处像素点的定位误差最小,其原因主要是由于单应映射对坐标系的选择具有依赖性,因此选择合适的邻域窗口局部坐标系对提高相应特征点的定位精度将会产生一定的影响。

6 分析及结论

通过以上理论推导和定量分析,可以得出以下结论:

虽然具有仿射不变性特征的检测和匹配算法成为目前的研究热点^[2 6 8 13 14 17 18 21 25~ 27],但如果其检测的特征(区域)比较大,就不能直接应用于高精度的大基线图像特征点匹配应用中,必须在小邻域窗口的前提下通过进一步的仿射迭代方能提高其定位精度。

当邻域窗口半径不大时,用仿射变换近似单应映射对特征点产生的定位误差较小,大多为亚像素级,一般不会成为特征点匹配的主要误差来源,但在有更高精度要求的应用中,必须考虑这一近似所带来的影响。

窗口内像素局部坐标系的选择也很重要,由于定位误差与像素点本身的坐标有关,坐标原点附近的定位误差相对较小,这就要求我们通过迭代方法矫正邻域窗口形状时,最好选择以窗口中心为原点的局部坐标系;此外,规范化的像素坐标能够减弱噪声影响,改善数值求解条件,这一点与估计基本矩阵^[28]和单应矩阵是类似的^[24]。

仿射变换近似模型仅对较小的邻域窗口适用。

且窗口内像素的定位误差会随窗口半径的增加而急剧增大,因此在实际匹配应用中必须选择合适的窗口大小;而目前很多针对较大区域的仿射不变性特征检测方法均不能直接胜任大基线透视图特征匹配中的精确定位^[2 3 6 13 15 18 26]。

在用仿射变换模型矫正相应邻域窗口透视畸变的迭代过程中选择高斯函数型的加权因子能增加窗口中心像素的重要性^[5],减弱远离窗口像素的误差影响,有助于相应特征点定位误差的改善。

上述结论在定量分析的基础上得出的,在进行图像匹配尤其是大基线图像的特征匹配时更应该注意,当然,这些结论也为我们进一步提高匹配精度提供了努力的方向。

致谢 感谢牛津大学 Visual Geometry Group 提供本文所用实验数据。

参考文献 (References)

- 1 Lowe D. Distinctive image features from scale-invariant keypoints [J]. *International Journal of Computer Vision*, 2004, **60**(2): 91~110
- 2 Rothganger F, Lazebnik S, Schmid C, *et al*. 3D object modeling and recognition using affine-invariant patches and multiview spatial constraints [A]. In *International Conference on Computer Vision and Pattern Recognition [C]*, Madison, USA, 2003: 272~277.
- 3 Mikolajczyk K, Schmid C. Indexing based on scale invariant interest points [A]. In *International Conference on Computer Vision [C]*, Vancouver, Canada, 2001: 525~531.
- 4 Schmid C, Mohr R, Bauckhage C. Comparing and evaluating interest points [A]. In *International Conference on Computer Vision [C]*, Bombay, India, 1998: 230~235.
- 5 Shi J, Tomasi C. Good features to track [A]. In *International Conference on Computer Vision and Pattern Recognition [C]*, Seattle, USA, 1994: 593~600.
- 6 Georgescu B, Meer P. Point matching under large image deformations and illumination changes [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, **26**(6): 674~688.
- 7 Pritchett P, Zisserman A. Wide baseline stereo matching [A]. In *International Conference on Computer Vision [C]*, Bombay, India, 1998: 754~760.
- 8 Baumberg A. Reliable feature matching across widely separated views [A]. In *International Conference on Computer Vision and Pattern Recognition [C]*, Hilton Head Island, SC, USA, 2000: 774~781.
- 9 Mikolajczyk K, Schmid C. A performance evaluation of local descriptors [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, **27**(10): 1615~1630.
- 10 Zhang Z, Deriche R, Faugeras O, *et al*. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry [J]. *Artificial Intelligence*, 1995, **78**(1-2): 87~119.
- 11 Zhou J, Shi J Y. A robust algorithm for feature point matching [J]. *Computer & Graphics*, 2002, **26**(3): 429~436.
- 12 Lourakis M I, Tzurbakis S V, Argyros A A, *et al*. Feature transfer and matching in disparate stereo views through the use of plane homographies [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, **25**(2): 271~276.
- 13 Tuytelaars T, Gool L V. Matching widely separated views based on affine invariant regions [J]. *International Journal of Computer Vision*, 2004, **59**(1): 61~85.
- 14 Schaffalitzky F, Zisserman A. Multiview matching for unordered image sets [A]. In *European Conference on Computer Vision [C]*, Copenhagen, Denmark, 2002: 414~431.
- 15 Schmid C, Mohr R. Local grayvalue invariants for image retrieval [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, **19**(5): 530~535.
- 16 Xiao J, Shah M. Two-frame wide baseline matching [A]. In *International Conference on Computer Vision [C]*, Nice, France, 2003: 603~609.
- 17 Berthilsson R. Finding correspondences of patches by means of affine transformations [A]. In *International Conference on Computer Vision [C]*, Kerkyra, Greece, 1999: 1117~1122.
- 18 Super B J, Kluquist W N. Patch matching and stereopsis in a general stereo viewing geometry [J]. *International Journal on Pattern Analysis and Machine Intelligence*, 1997, **19**(3): 247~253.
- 19 Ferrari V, Tuytelaars T, Gool L V. Wide-baseline multiple-view correspondences [A]. In *International Conference on Computer Vision and Pattern Recognition [C]*, Beijing, China, 2003: 718~728.
- 20 Dufoumaud Y, Schmid C, Horaud R. Matching images with different resolutions [A]. In *International Conference on Computer Vision and Pattern Recognition [C]*, Hilton Head Island, USA, 2000: 612~618.
- 21 Mikolajczyk K, Schmid C. An affine invariant interest point detector [A]. In *European Conference on Computer Vision [C]*, Copenhagen, Denmark, 2002: 128~142.
- 22 Mikolajczyk K, Tuytelaars T, *et al*. A comparison of affine region detectors [J]. *International Journal of Computer Vision*, 2005, **65**(7): 43~72.
- 23 Meng Xiao-qiao, Hu Zhan-yi. Recent progress in camera self-calibration [J]. *Acta Automatica Sinica*, 2003, **29**(1): 110~124. [孟晓桥, 胡占义. 摄像机自标定方法的研究与进展 [J]. *自动化学报*, 2003, **29**(1): 110~124.]
- 24 Hartley R, Zisserman A. *Multiple view geometry in computer vision* [M]. Cambridge, UK: Cambridge University Press, 2000.
- 25 Tuytelaars T, Gool L V, *et al*. Matching of affinely invariant regions for visual servoing [A]. In *International Conference Robotics and Automation [C]*, Detroit, MI, USA, 1999: 1601~1606.
- 26 Kadir T, Zisserman A, Brady M. An affine invariant salient region detector [A]. In *European Conference on Computer Vision [C]*, Prague, Czech Republic, 2004: 345~457.
- 27 Mikolajczyk K, Schmid C. Scale & affine invariant interest point detectors [J]. *International Journal of Computer Vision*, 2004, **60**(1): 63~86.
- 28 Hartley R. In defence of the eight-point algorithm [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, **19**(6): 580~593.

基于标志点的测量数据自动拼接方法

欧阳祥波 宗志坚 熊会元

(中山大学工学院, 广州 510275)

摘要 为了对测量数据进行自动拼接, 提出了一种新的数据拼接方法。该方法根据标志点的空间拓扑关系, 先利用关系匹配自动找出对应的标志点对, 以降低数据拼接的操作复杂性; 同时应用 Rodrigues 参数表示位姿变换矩阵, 并引入中间参数进行分步求解。数据拼接实例的拼接结果表明, 该方法计算过程简单、稳定性好。

关键词 自动拼接 标志点 关系匹配 Rodrigues 参数

中图法分类号: TP391.41 文献标识码: A 文章编号: 1006-8961(2008)02-0298-04

An Automatic Registration Method for Point-clouds Based on Marked Points

OUYANG Xiang-bo ZONG Zhizhan XIONG Huiruan

(College of Engineering, Sun Yat-Sun University, Guangzhou 510275)

Abstract To register point sets automatically, a method was proposed in this paper. It used the relational matching to get the counterparts of marked points based on the topology and introduced the Rodrigues parameters to describe the transition matrix. The matrix was solved simply and steadily by using a temporary variant. Finally a registration experiment was presented.

Keywords automatic registration, marked point, relational matching, Rodrigues parameters

1 引言

在 3 维物体外形检测的过程中, 由于受设备测量范围的限制或者被测物体表面互相遮挡, 一次只能获得物体的局部数据。这就需要从不同的方位对物体进行多次测量, 然后通过对测得的数据进行拼接来生成一个统一坐标系下的 3 维数据点集。由于测量数据的拼接是一种刚性的运动, 因此可以通过求得不同测量方位的相对位姿来实现, 目前主要用以下 3 种方法解决这个问题:

(1) 精确控制测量设备或物体的运动^[1,2], 即由控制系统提供准确的旋转矩阵 R 和平移 T 。虽然该方法的拼接精度高, 但需要精密的辅助设备, 由于体积庞大、操作复杂, 因此不便于实际应用。

(2) 通过 ICP 算法 (iterative closest point

registration) 进行拼接^[3]。该方法需要相邻的两片测量数据有重叠部分以及若干明确的对应点, 拼接时, 先以这些点作为初始公共特征点, 再通过迭代算法求解坐标变换矩阵来进行拼接。该方法不但算法复杂、效率低, 而且精度也不高。

(3) 先在两次测量数据的公共区域布置若干标志点, 再通过标志点的对齐计算变换矩阵^[4]来进行拼接。这种方法操作简单, 并具有一定的拼接精度。通常由于该方法需要利用手工方式明确指定标志点的对应关系, 故操作效率比较低。

本文提出一种基于标志点的拼接方法。该方法是根据标志点的空间拓扑关系, 利用关系匹配自动找出对应的标志点对, 同时应用 Rodrigues 参数表示位姿变换矩阵, 并通过引入中间参数简化了求解过程。

基金项目: 广东省科技厅项目 (2003A1040704, 20070928021)

收稿日期: 2006-07-03 改回日期: 2006-08-25

第一作者简介: 欧阳祥波 (1973~), 男, 博士研究生。从事计算机视觉测量、逆向工程等领域的研究。E-mail: oyxb126@126.com

2 对应点查找

对应点查找就是先在被测物体表面粘贴一系列特殊的图形标记或球形标记, 然后采用各种非接触式、接触式测量方法, 来方便地获得标志点的空间坐标。但该标志点的位置应随机地分布在两次测量时公共区域的最外围, 同时需具有尽可能大的多边形面积; 标志点的位置应选择容易测量和不会发生变形的地方; 标志点的设计和获取可参看文献 [5]。本节主要讨论在已知标志点集的情况下, 如何获得标志点之间的对应关系。

2.1 标志点集的描述

任何一个标志点集均由以下 3 部分所组成^[6]:

(1) 一组基本特征元素。例如点、线和面, 在关系匹配中, 它们组成一个集合, 其中的面主要是三角形平面。例如图 1 所示的点集 A , 其特征集合 F 由 4 个点, 6 条边和 4 个面组成, 即

$$F = \{v_1, \dots, v_4; e_1, \dots, e_6; a_1, \dots, a_4\}$$

(2) 特征元素的空间关系。例如, 边元素由点对定义, 仍以图 1 为例, 边元素由以下关系定义:

$$R_{\text{edge}} = \{ \langle v_1, v_2 \rangle, \langle v_1, v_3 \rangle, \langle v_1, v_4 \rangle, \langle v_2, v_3 \rangle, \langle v_2, v_4 \rangle, \langle v_3, v_4 \rangle \}$$

面元素由关系 R_{area} 定义:

$$R_{\text{area}} = \{ \langle e_1, e_2, e_3 \rangle, \langle e_1, e_2, e_4 \rangle, \langle e_1, e_3, e_4 \rangle, \langle e_2, e_3, e_4 \rangle \}$$

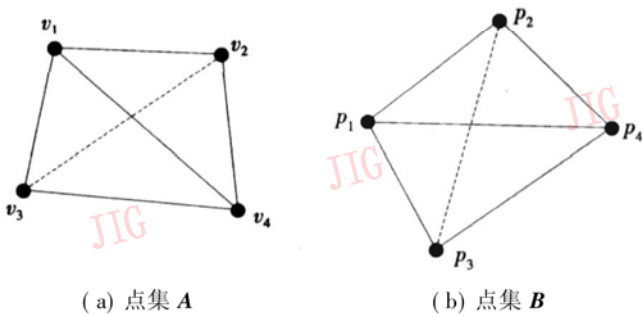


图 1 标志点集

Fig 1 Marked points

(3) 特征元素的性质, 主要指几何性质, 例如点的坐标、边的长度、面的面积等。

本文采用图 2 所示的层次结构记录一个标志点集, 其中最上层为连接任意 3 个标志点所形成的三角面, 中间层为三角面的 3 条边, 最下层为各条边的顶点。该结构基本包括了标志点集描述的 3 个部分。

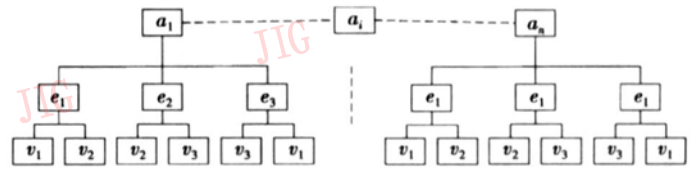


图 2 标志点集的层次结构描述

Fig 2 Description marked points using hierarchy

2.2 对应点的查找

为了获得标志点的对应关系, 需要设计一个变换 h , 将点集 A 上的点映射到点集 B , 然后比较特征元素性质的相似程度, 而具有最大相似程度的变换就确定了两个点集之间的对应关系。例如图 1 中的点集 A 与点集 B , 假设变换 h_1 为

$$\begin{aligned} v_1 &\rightarrow p_1; & v_2 &\rightarrow p_2 \\ v_3 &\rightarrow p_3; & v_4 &\rightarrow p_4 \end{aligned}$$

则点集 A 、 B 具有最大的相似程度。

对应点的查找就是在 一组变换 $H = \{h_1, h_2, \dots, h_n\}$ 中, 寻找最佳的变换, 使得两个点集具有最大的相似程度, 这个过程称为关系匹配。在两个标志点集中, 由于可能的匹配是大量的, 并且不是每一个点都能够在另一个点集中找到对应点, 因此最佳匹配的寻找时间可能会很长。为了加速计算, 本文导出以下一种分级的最佳匹配寻找方法:

(1) 比较三角面元素的面积, 根据式 (1) 判断两个面的面积是否兼容, 其中 A_1, A_2 为三角形的面积, T_A (下角 A 代表 area) 为预先设定的面积阈值; 找出所有面积兼容的面:

$$-T_A < \frac{A_1 - A_2}{\min(A_1, A_2)} < T_A \quad (1)$$

(2) 对于面积兼容的三角面, 可根据式 (2) 找出长度兼容的边的数量 N_{edge} , 其中 L_1, L_2 为三角形的边长, T_L (L 代表 length) 为预先设定的边长阈值; 如果 $N_{\text{edge}} = 3$ 则认为两个三角面兼容;

$$-T_L < \frac{L_1 - L_2}{\min(L_1, L_2)} < T_L \quad (2)$$

(3) 对兼容的三角面, 则记录 3 个点之间的对应关系; 如果没有兼容的三角面, 则这两个标志点集不能进行拼接;

面积阈值和边长阈值是寻找最佳匹配的关键参数, 它们与测量系统的测量范围、测量误差以及误差的统计学分布规律有着密切的联系, 在实际应用中需要根据设备的具体情况进行设定。在本文的拼接实例中, 边长阈值设定为 2.5%, 面积阈值设定

为 5%。

通过以上的步骤,即可以快速可靠地找到两个标志点集的对应点。对于一个精度可靠的测量系统来说,当兼容三角面的数量大于 1 的时候,一个标志点对将会在数个兼容的三角面中被找到;但一个标志点最多只能找到一个对应点。

3 变换矩阵的求解

设点 P 在坐标系 C_1, C_2 中的坐标分别为 $(x_1, y_1, z_1), (x_2, y_2, z_2)$, C_1, C_2 之间的相对旋转矩阵为 R , 相对平移向量为 $t = (t_x, t_y, t_z)^T$, 则

$$\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} = R \cdot \begin{pmatrix} x_2 + t_x \\ y_2 + t_y \\ z_2 + t_z \end{pmatrix} \quad (3)$$

旋转正交矩阵 R 可利用反对称矩阵

$$U = \begin{pmatrix} 0 & -u_z & u_y \\ u_z & 0 & -u_x \\ -u_y & u_x & 0 \end{pmatrix} \quad (4)$$

表示为

$$R = (I - U)^{-1} \cdot (I + U) \quad (5)$$

向量 $U = (u_x, u_y, u_z)$ 为坐标系 C_2 相对于坐标系 C_1 的 Rodrigues 参数^[7], 其几何意义表示刚体绕定点转动的旋转轴方向。

将式 (4)、式 (5) 带入式 (3), 并按已知量和未知量进行整理, 可得

$$\begin{pmatrix} 0 & z_1 + z_2 & -(y_1 + y_2) \\ -(z_1 + z_2) & 0 & x_1 + x_2 \\ y_1 + y_2 & x_1 + x_2 & 0 \end{pmatrix} \cdot \begin{pmatrix} u_x \\ u_y \\ u_z \end{pmatrix} + \begin{pmatrix} 1 & -u_z & u_y \\ u_z & 1 & -u_x \\ -u_y & u_x & 1 \end{pmatrix} \cdot \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix} = \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} - \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix}$$

引入辅助参数

$$\begin{pmatrix} a \\ b \\ d \end{pmatrix} = \begin{pmatrix} 1 & -u_z & u_y \\ u_z & 1 & -u_x \\ -u_y & u_x & 1 \end{pmatrix} \cdot \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix} \quad (6)$$

整理上式, 得

$$\begin{pmatrix} 0 & (z_1 + z_2) & -(y_1 + y_2) & 1 & 0 & 0 \\ -(z_1 + z_2) & 0 & x_1 + x_2 & 0 & 1 & 0 \\ y_1 + y_2 & -(x_1 + x_2) & 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} t_x \\ t_y \\ t_z \\ a \\ b \\ d \end{pmatrix}$$

$$\begin{pmatrix} u_x \\ u_y \\ u_z \\ a \\ b \\ c \end{pmatrix} = \begin{pmatrix} x_1 - x_2 \\ y_1 - y_2 \\ z_1 - z_2 \end{pmatrix}$$

简记为 $M \cdot X = \Delta$

对于 $n (n \geq 3)$ 个标志点对, 可以得到如下的方程组:

$$\begin{pmatrix} M_1 \\ M_2 \\ \vdots \\ M_n \end{pmatrix} \cdot X = \begin{pmatrix} \Delta_1 \\ \Delta_2 \\ \vdots \\ \Delta_n \end{pmatrix} \quad (7)$$

式 (7) 是一个系数矩阵列满秩的线性方程组, 若用最小二乘法求解, 则可得到 Rodrigues 参数以及引入的辅助参数。根据式 (5) 可求得旋转矩阵 R , 根据式 (6) 可求得平移向量 t 即

$$\begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix} = \frac{1}{1 + u_x^2 + u_y^2 + u_z^2} \cdot \begin{pmatrix} 1 + u_x^2 & u_x \cdot u_y + u_z & u_x \cdot u_z - u_y \\ u_x \cdot u_y - u_z & 1 + u_y^2 & u_y \cdot u_z + u_x \\ u_x \cdot u_z + u_y & u_y \cdot u_z - u_x & 1 + u_z^2 \end{pmatrix} \cdot \begin{pmatrix} a \\ b \\ d \end{pmatrix}$$

4 数据拼接实例

本文的试验数据采用双目立体视觉测量系统获得, 该系统的硬件主要包括两个 DH HV 1300FC 摄像机、M 2514MP 镜头和一个光栅投影仪, 该系统单次测量的范围为 $300\text{mm} \times 400\text{mm}$, 单次测量误差小于 0.05mm 。对于较大的物体来说, 由于单次测量只能获得被测物体的局部数据, 因此必须进行数据拼接。拼接的标志点是人工贴在物体表面的易于识别的图形, 粘贴的位置应选择曲面上相对平坦的区域。出于精度和速度方面的考虑, 在单次拼接时, 通常选择 4~6 个标志点。

图 3(a)、图 3(b) 为相邻的两片测量数据, 数据之间有一定的重叠部分, 其中的彩色十字形是测量系统检测出的标志点。利用本文方法, 首先在两个标志点集中找出对应的标志点对, 图 3(a)、图 3(b) 中颜色相同的十字形为一个标志点对; 然后以标志点对为基础, 解得刚性变换的旋转矩阵和平移向量。

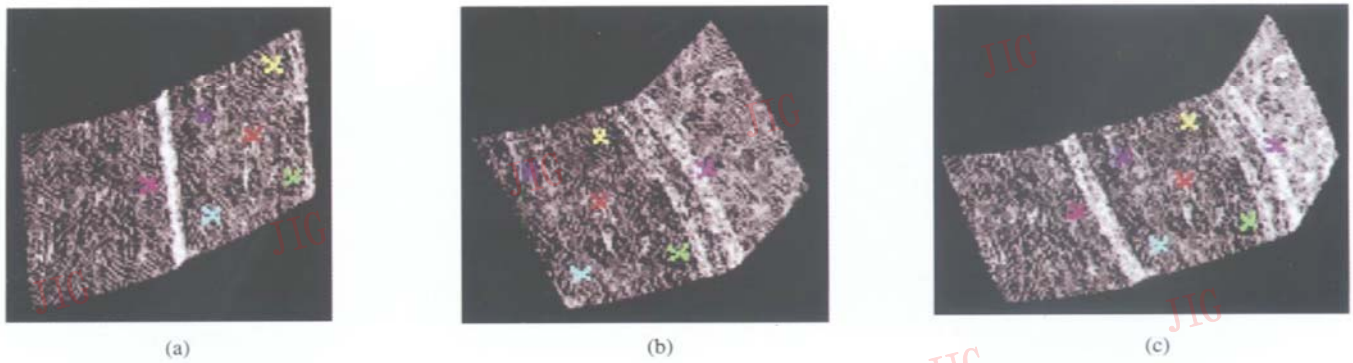


图 3 拼解实例

Fig 3 Example of registration

图 3(c)为拼接后的点云数据,图 3(c)没有对重叠区域进行均匀化处理。表 1所示为拼接后标志点对之间的距离误差,这个误差值远小于测量系统的测量误差,这说明采用本方法能够较好地实现测量数据的拼接,并且精度完全满足实际的需要。

表 1 标志点拼接误差

Tab 1 The error of mached counterpoints

点对	1-a	2-b	3-c	4-d	5-e
距离 (mm)	0.0207	0.0203	0.0165	0.0201	0.0220

5 结 论

本文给出的数据拼接方法能够自动识别标志点集之间的对应关系,并有效地提高了数据测量的操作效率; Rodrigues 参数的引入,使得位姿矩阵的求解过程主要是一些低阶向量和矩阵的代数运算,且计算简单、稳定性好。该方法既适合于 3 个对应点,也适合于 3 个以上对应点的测量数据拼接,且操作简单,精度高,非常适合于实际应用。

参考文献 (References)

- 1 Chan V H, Bradley C, Vickers G W. A multi-sensor approach to automating coordinate measuring machine based reverse engineering [J]. Computers in Industry, 2001, 44(2): 105~115.
- 2 Xie Ze-xiao, Zhang Qiu-mei, Zhang Guo-xiong. Modeling and calibration of a structured-light sensor-based five-axis scanning system [J]. Measurement, 2004, 36(2): 185~194.
- 3 Liu Yong-huai. Improving ICP with easy implementation for free form surface matching [J]. Pattern Recognition, 2004, 37(2): 211~226.
- 4 Shen Hai-ping, Da Fei-peng, Lei Jia-yong. Research of point-clouds registration based on least-square method [J]. Journal of Image and Graphics, 2004, 10(9): 1112~1116 [沈海平, 达飞鹏, 雷家勇. 基于最小二乘法的点云数据拼接研究 [J]. 中国图象图形学报, 2005, 10(9): 1112~1116].
- 5 Ahn S J, Rauh W. Circular coded target for automation of optical 3D-measurement and camera calibration [J]. International Journal of Pattern Recognition and Artificial Intelligence, 2001, 15(6): 905~919.
- 6 Zhang Zhu-xun, Zhang Jian-qing. Digital Photogrammetry [M]. Wuhan: Wuhan University Press, 1997: 198~202 [张祖勋, 张剑清. 数字摄影测量学 [M]. 武汉: 武汉大学出版社, 1997: 198~202].
- 7 Zhou Jiang-hua, Miao Yu-hong, Wang Ming-hai. Attitude representation using Rodrigues parameter [J]. Journal of Astronautics, 2004, 25(5): 514~519 [周江华, 苗育红, 王明海. 姿态运动的 Rodrigues 参数描述 [J]. 宇航学报, 2004, 25(5): 514~519].

基于整数中轴骨架的 3 维模型检索算法

唐 勇 张学范

(燕山大学信息科学与工程学院, 秦皇岛 066004)

摘 要 针对 3 维模型检索算法性能较低的问题, 提出了一种基于整数中轴骨架的 3 维模型检索算法。在对 3 维模型进行姿态调整和各向同性预处理后, 提取模型的整数中轴骨架, 并记录每个骨架点相应的几何信息, 对提取的骨架按不同的空间区域划分, 形成模型骨架二叉树。为了能够描述骨架二叉树的不同节点对模型整体相似性匹配的影响程度, 为每个节点定义一个特征权值, 其大小由该节点对应的骨架区域大小所决定。最后, 采用由粗到细逐步淘汰的策略计算不同模型的相似度。对一个标准 3 维模型测试数据库的检索实验结果表明, 由于将模型的拓扑结构和统计特征相结合, 该算法可以得到较好的检索性能。

关键词 3 维模型检索 特征变换 整数中轴骨架 骨架二叉树

中图法分类号: TP391.41 文献标识码: A 文章编号: 1006-8961(2008)02-0302-05

3D Model Retrieval Based on Integer Medial Axis Skeleton

TANG Yong ZHANG Xue-fan

(College of Information Science and Engineering, Yanshan University, Qinhuangdao 066004)

Abstract To improve the efficiency of 3D model retrieval, an algorithm for 3D model retrieval based on integer medial axis skeleton was proposed in this paper. The integer medial axis skeleton and the geometric information of skeleton point were obtained after the preprocessing of the model. The binary tree of this skeleton was acquired by decomposing the skeleton into a set of blocks by spatial region. To describe the influence of different node of the skeleton binary tree to the similarity matching, the feature weight was defined for each node. Furthermore, the weights were determined by corresponding skeleton region of the 3D model. Finally, a coarse-to-fine strategy was presented to calculate similarity between different 3D models. Differing from other algorithms applied in 3D model retrieval, this algorithm extracts statistical features as well as topological features. The experiments have been carried on a standard testing database of 3D models, and the results show that this algorithm can achieve better retrieving efficiency than other algorithms.

Keywords 3D model retrieval, feature transform, integer medial axis skeletons, skeletal binary tree

1 引言

随着激光扫描技术的发展以及计算机性能的提高, 3 维模型在很多应用领域中扮演着非常重要的角色(如工业产品的模型设计、虚拟现实系统、游戏设计、逆向工程以及仿真等), 但是构造一个真实感较强的模型是一项工作量巨大的任务, 目前在因特网和特定领域的数据库中存在数亿以兆计的模型,

且在源源不断地增加。如果能够重复利用已有模型, 就可大大减小模型设计的工作量, 因此对 3 维模型检索技术的研究已变得越来越紧迫和重要。

3 维模型检索的一个关键问题是如何快速、准确地找到所需要的模型。针对这个问题已提出了一些方法, 如形状分析方法, 将 3 维模型匹配问题简化为 2 维多边形集相似性问题^[1]; 利用直方图表示模型特征分布, 通过对面积的分布进行归一化处理并计算 L_2 差, 从而实现两模型之间的匹配^[2]; 利用数学

基金项目: 国家自然科学基金项目(60374042)

收稿日期: 2006-01-05 改回日期: 2006-10-09

第一作者简介: 唐勇(1964~), 男, 教授。2005 年于燕山大学获机械电子工程专业博士学位。研究方向为虚拟现实技术及其应用、软件工程化技术和信息安全。E-mail: tangyong@ysu.edu.cn

变换进行特征提取, 如 Hough 变换、调和映射或矩。上述方法的一个共同特点就是通过统计多个全局特征来表示 3 维形状, 比较容易实现、性能稳定, 但是没有考虑局部特征, 而且由于涉及的特征较多, 在检索速度上会出现较大的延迟。

为此, Xiao 采用了一种离散 Reeb 图 (DRG) 的方法来对人体模型的拓扑结构进行分析, 该算法采用高度函数作为 Morse 函数^[3]; Bespalov 提出了一种顶点的距离矩阵, 通过距离矩阵对 3 维模型进行分解来得到 3 维模型的 Reeb 图^[4]; 该算法与 3 维模型坐标系旋转无关, 但是这种算法计算复杂, 并且要求 3 维模型是连通模型; Foskey 采用中轴线的方法得到了手骨骼的骨架图^[5], 通过比较 3 维模型骨架图的拓扑结构获得 3 维模型的几何相似性, 但是 3 维模型的骨架图计算量很大, 计算复杂度高。

因此如何提高 3 维模型的检索性能, 就成了十分突出的问题, 文中提出一种基于整数中轴骨架的 3 维模型检索算法, 该算法的关键思想是通过整数中轴变换, 快速得到模型的中轴骨架。对骨架按区域划分, 构造骨架二叉树。同时根据该区域的大小定义每个节点的特征权值, 用于衡量其对 3 维模型整体相似性的影响程度。在对骨架二叉树匹配时, 采用由粗到细逐步淘汰的策略, 不断缩减待匹配模型的范围, 从而降低了模型匹配的时间。实验结果表明, 该算法可以得到较好的检索性能。

2 模型预处理

设两个待匹配 3 维模型 P 、 Q 的顶点分别为 $P = \{p_1, p_2, \dots, p_n\}$ 和 $Q = \{q_1, q_2, \dots, q_n\}$, 顶点序列的坐标为 $p_i = \{p_x, p_y, p_z\}$ 和 $q_i = \{q_x, q_y, q_z\}$ 。

对于 3 维模型检索算法而言, 检索的准确性是评价算法好坏的关键指标之一。3 维模型检索的准确性除了取决于 3 维模型检索算法外, 还与 3 维模型的一些预处理相关。3 维模型预处理包括 3 维模型姿态调整和各向同性预处理两个部分。

2.1 姿态调整

对于同一种检索算法而言, 处于不同坐标系下的 3 维模型应该具有相同的相似度, 因此, 检索算法在计算 3 维模型几何特征之前应该对 3 维模型进行姿态调整, 使其坐标系一致。

采用主元分析法 (principal component analysis, PCA) 对模型进行姿态调整^[6]。该方法首先根据

3 维模型点集的协方差矩阵计算出相应的特征值 $\lambda_1 > \lambda_2 > \lambda_3$, 其对应的特征矢量为 (I_1, I_2, I_3) , 以 (I_1, I_2, I_3) 为新的坐标系, 对 3 维模型进行坐标变换, 得到变换后的坐标值。处理结果如图 1 所示。

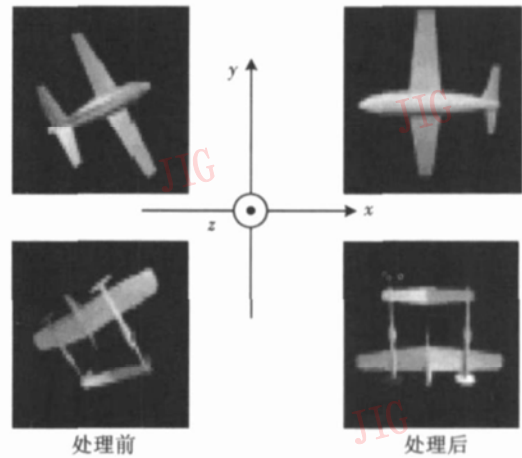


图 1 PCA 方法处理前和处理后的 3 维模型效果

Fig 1 3D models before and after PCA

2.2 各向同性预处理

在计算 3 维模型的相似度之前, 对模型进行各向同性预处理, 可以使得经过任意方向比例变换的模型比例归一化, 从而提高模型检索的准确性^[7]。采用如下变换矩阵 C_p 对 3 维模型 P 、 Q 做各向同性预处理:

$$\begin{cases} C_p = \left| \sum_{i,j=1}^n (p_i - q_j)(p_i - q_j)^T \right|^{1/2} \\ Q = C_p P \end{cases} \quad (1)$$

各向同性预处理结果如图 2 所示。

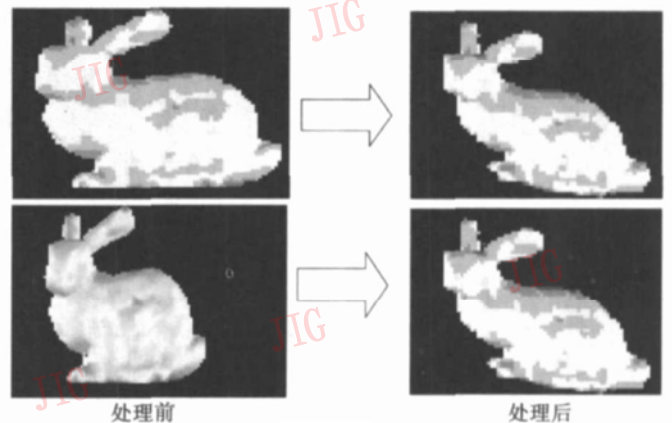


图 2 各向同性预处理的结果

Fig 2 The normalizing for isotropic scale preprocessing

3 骨架提取

设 r 是 3 维模型表面上的点, 由 Hesselink 的整数中轴算法^[8]可得:

若 $e \in E$, ($E = \{e \in I^3 \mid \|e\| = 1\}$), I^3 为模型内部的一个体素网格点, 则当 $m = r + \frac{1}{2}e$ 时

$$\|m - t(r + e)\| = \|m - t(r)\| \quad (2)$$

m 为整数中轴骨架上的一个骨架点, $t(r)$ 为点 r 的特征变换函数。

为了记录以骨架点 m 为球心的内接球的半径, 对整数中轴骨架进行改进, 定义一元函数:

$$\sigma_m = \|m - t(r)\| \quad (3)$$

σ_m 为骨架点 m 的权值。将所有骨架点连接起来形成加权骨架 H , 如图 3 所示。



图 3 骨架提取示意图

Fig 3 A schematics of extracting skeleton

4 模型匹配

4.1 生成骨架二叉树

设 Z_m 为骨架点 m 的 Z 轴坐标值。令

$$Z_{m \min} = \min\{Z_m \mid m \in H\} \quad (4)$$

$$Z_{m \max} = \max\{Z_m \mid m \in H\} \quad (5)$$

将骨架按 Z 轴坐标值划分为不同的区域 C_i , Z_{C_i} 为区域 C_i 在 Z 轴的上限 ($1 \leq i \leq n$), 如图 4(a) 所示。其中区域 C_{2i+1} 和区域 C_{2i+2} 沿 Z 轴平分区域 C_i , 如图 4(b) 所示。

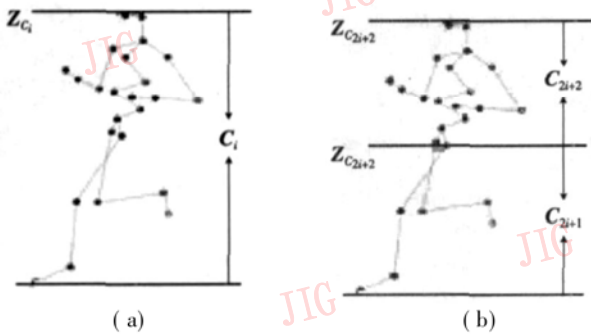


图 4 骨架区域划分示意图

Fig 4 A schematics of skeleton partition

将区域 C_i 视为二叉树节点 a_i , 其权值为

$$W_{a_i} = \sum_{m \in C_i} \sigma_m \quad (6)$$

假设二叉树根结点为 a_0 , 对于任一节点 a_i , a_{2i+1} 和 a_{2i+2} 分别为 a_i 的左右孩子节点, 且满足以下条件:

$$(1) Z_{C_{2i+1}} < Z_{C_{2(i+1)}}$$

(2) 若 $a_{2i+1} = 0$ 和 $a_{2(i+1)} = 0$ 则节点 a_i 为叶子节点;

(3) 若 $a_{2i+1} = 0$ 且 $a_{2i+2} \neq 0$ 则节点 a_i 的左子树为空; 若 $a_{2i+2} = 0$ 且 $a_{2i+1} \neq 0$ 则节点 a_i 的右子树为空。

4.2 匹配骨架二叉树

设 a_i 和 b_i ($0 \leq i \leq n$) 为 3 维模型 P, Q 对应的骨架二叉树中的节点, 则它们的相似度函数为

$$\text{sim}(a_i, b_i) = 1 - \frac{|W_{a_i} - W_{b_i}|}{W_{a_i}} \quad (7)$$

设 3 维模型 P, Q 的相似度函数为

$$SM(P, Q) = \sum_{i=0}^n \text{sim}(a_i, b_i) \quad (8)$$

但是, 在匹配过程中, 由于模型的不同部分对模型整体的相似性的影响不同, 因此对不同的 $\text{sim}(a_i, b_i)$ 赋予不同的权值 χ_i 对相似度函数加以改进, 改进的相似度函数为

$$SM(P, Q) = \sum_{i=0}^n \chi_i \cdot \text{sim}(a_i, b_i) \quad (9)$$

其中, $\chi_i = \frac{f(a_i)}{f(a_0)}$, $f(a_i)$ 表示对应节点 a_i 的区域大小, $f(a_0)$ 为整个区域的大小。

具体的匹配步骤如下:

(1) 定义域值区间 $g = [0, \beta]$, $\beta \in \mathbf{R}$; 生成两个骨架二叉树的根节点 a_0, b_0 , 若 $\text{sim}(a_0, b_0) \in g$, 则继续以下步骤, 否则匹配结束, 两个模型不相似;

(2) 若 a_i, b_i 为非叶子节点, 则生成 a_i, b_i 的左孩子节点 a_{2i+1} 和 b_{2i+1} , 若 $\text{sim}(a_{2i+1}, b_{2i+1}) \in g$, 则继续以下步骤, 否则匹配结束, 两个模型不相似;

(3) 若 a_i, b_i 为叶子节点, $\text{sim}(a_i, b_i) \in g$, 则该分支的匹配结束, 向上回溯到其父亲节点, 进行另一分支的匹配, 若 $\text{sim}(a_i, b_i) \notin g$, 则匹配结束, 两个模型不相似;

(4) 生成 a_i 和 b_i 的右孩子节点 a_{2i+2} 和 b_{2i+2} , 若 $\text{sim}(a_{2i+2}, b_{2i+2}) \in g$, 则继续以下步骤。否则匹配结束, 两个模型不相似;

(5) 若二叉树的任意节点 a_i 和 b_i 都满足: $\text{sim}(a_i, b_i) \in g$, ($0 \leq i \leq n$), 则两模型相似。

(6) 重复执行步骤 1 至步骤 5。

如图 5 所示的两个匹配二叉树, 首先 $\{a_0, b_0\}$ 构成匹配节点, 若 $sim(a_0, b_0) \in g$, 则匹配 a_1 和 b_1 , 否则匹配结束。

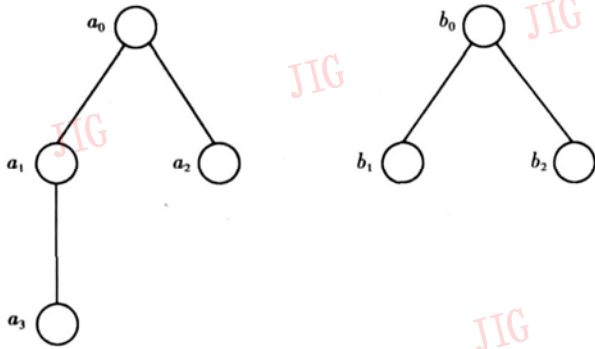


图 5 骨架二叉树匹配示意图

Fig 5 Matching between skeleton binary trees

5 实验结果与分析

为了测试算法的效果, 设计实验对本文方法、中轴骨架方法和形状分析方法的检索性能进行比较。实验在 Windows 平台上用 VC++ 6.0 语言实现, 3 维模型数据库采用普林斯顿大学形状分析小组提供的标准测试数据库^[9], 总共含有 907 个模型。采用典型的大家公认的 Precision-Recall 曲线来度量不同方法的检索性能, 3 种方法的检索性能曲线如图 6 所示。可以发现, 本文方法由于在拓扑结构的基础上融入了统计特征, 因此在检索性能上有明显提高。

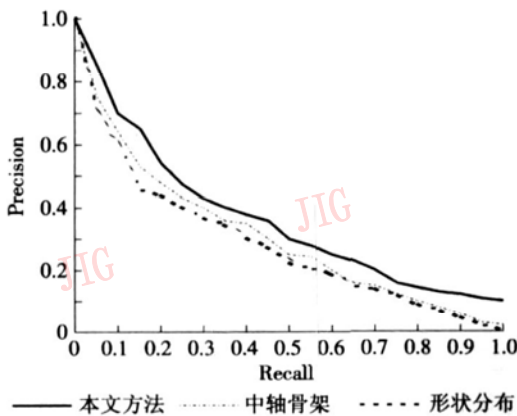


图 6 检索性能比较

Fig 6 Retrieving performance comparison of three methods

对于 3 维模型检索, 另外一个值得注意的问题是检索效率。如果检索时间过长, 导致实时性差, 即使检索准确率有了明显的改进, 那么实用性也不强。本文方法采用改进的中轴骨架提取方法, 虽然与传

统的中轴骨架方法相比降低了算法的复杂度, 但与形状分布方法相比在算法复杂度上有所增加, 因此比形状分布方法需要更多的检索时间。但是, 这种检索时间的差异很小, 不会被用户察觉。具体实验验证如下: 对一批模型数据 (40 个模型) 进行批处理, 得到总检索时间和平均检索时间, 检索时间如表 1 所示 (CPU: Pentium 4 2.4GHz 内存: 512MB, 检索时间包括打开文件读取模型数据的时间)。

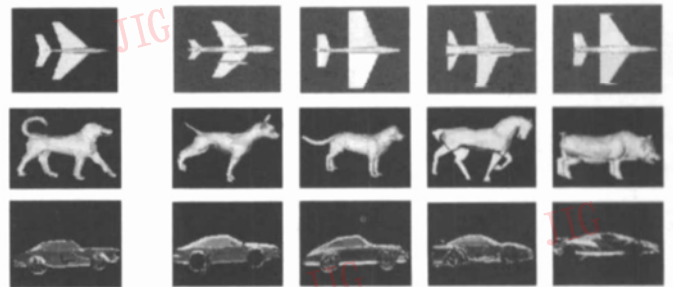
表 1 检索时间比较

Tab 1 Comparison of retrieving time

单位: s

检索方法	总检索时间	平均检索时间
本文方法	6	0.15
中轴骨架	6.96	0.174
形状分布	5.6	0.14

从表 1 中可以看出本文方法与形状分布方法相比, 虽然在检索总时间增多了 0.4s, 但是平均检索时间仅增多了 0.01s, 对于这种较小的时间差异用户是很难察觉到的。图 7 给出了部分检索结果。



(a) 输入模型 (b) 检索出的前 4 个最相似的模型

图 7 检索结果示例

Fig. 7 Results of the search experiment

3 维模型检索算法的抗干扰能力也是衡量算法性能的一个重要指标, 为了测试算法的抗干扰能力, 对 3 维模型 (图 8(a)) 做以下变形处理:

- (1) 加入噪声 3 维模型每个顶点的 x, y, z 坐标值, 加上一个随机数 (模型边界长度得 $-5\% \sim +5\%$), 如图 8(b);
- (2) 模型缺损 随机删减部分顶点 (占总顶点数的 10%), 如图 8(c)。

对变形后的 3 维模型进行检索试验, 其检索性能如图 9 所示。

由图 9 中可以看出, 本文算法在 3 维模型包含噪声或缺损部分顶点时, 仍可达到模型检索的目的。

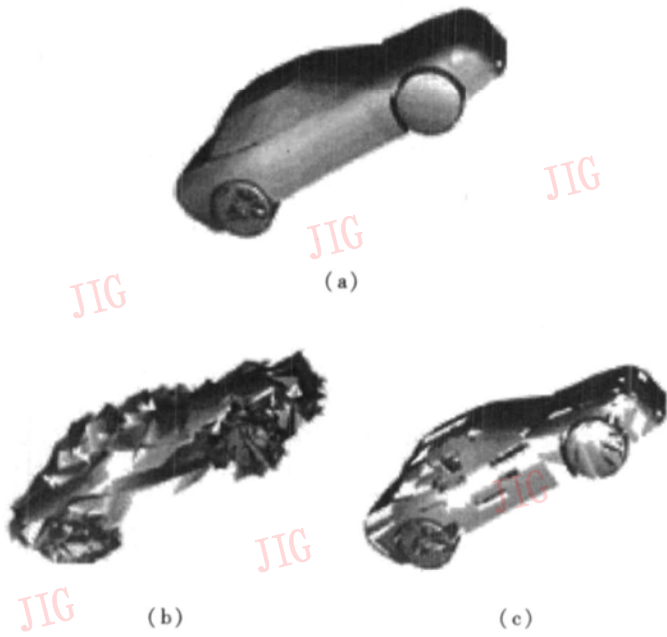


图 8 加入噪声和模型缺损

Fig. 8 Noise and decimation from original 3D model

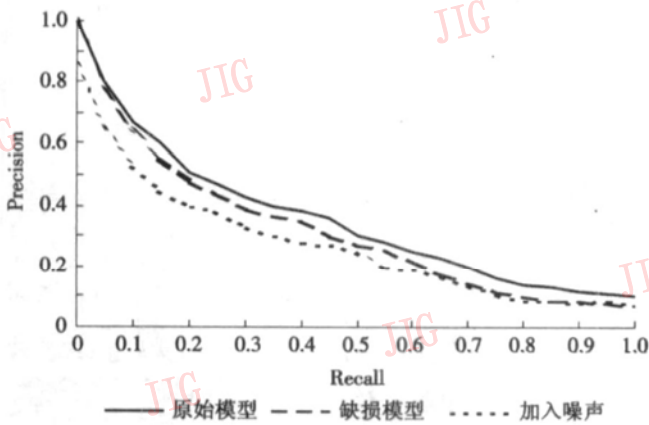


图 9 3 维模型抗噪声和缺损性能评价

Fig. 9 Robustness evaluation of noise and decimation

6 结 论

3维模型检索技术是计算机图形学研究领域的一个重要课题。本文将 3维模型的拓扑结构特征和统计特性相结合,针对 3维模型检索算法性能较低的问题,提出了一种基于整数中轴骨架的 3维模型检索算法。在对待比较模型进行姿态调整和各向同性预处理后;对整数中轴骨架生成算法进行改进,生成 3维模型的整数中轴骨架,在生成骨架的过程

中保留模型的拓扑信息和每个骨架点的几何信息;在相似性匹配时,提出了一种层次性匹配方法,对生成的模型骨架采用由粗到细逐步淘汰的策略进行匹配,逐步缩减待匹配模型的范围,从而降低了模型匹配的时间。通过设计实验对本文方法的检索性能、检索时间和抗干扰能力进行测试,实验结果表明该算法可以得到较好的检索性能。

参考文献 (References)

- 1 Pu Jian-tao, Liu Yi, Xin Gu-yu, *et al*. A 3D model retrieval method based on the similarity between 2D polygon sets [J]. Journal of Image and Graphics, 2004, 9(12): 1437~1442 [普建涛,刘一,辛谷雨等.一种基于 2 维多边形集相似性的 3 维模型检索方法 [J]. 中国图象图形学报, 2004, 9(12): 1437~1442]
- 2 Mihalek, Gabi K, Hans-Peter K, *et al*. 3D shape histogram for similarity search and classification in spatial database [A]. In: Proceedings of 6th International Symposium on Spatial Database [C], Hongkong, China, 1999, 207~228
- 3 Xiao Y J, Wenghi N, Siebert P. A topological approach for segmenting human body shape [A]. In: Proceedings of 12th International Conference on Image Analysis and Processing [C], Mantova, Italy, 2003, 82~87
- 4 Bespalov D, Ali S, William C R, *et al*. Scale-space representation of 3D models and topological matching [A]. In: 8th ACM/SIGGRAPH Symposium on Solid Modeling and Applications [C], Seattle, Washington, USA, 2003, 208~215
- 5 Foskey M, Lin M C, Manocha D. Efficient computation of a simplified medial axis [A]. In: Proceedings of the Eighth ACM Symposium on Solid Modeling and Applications [C], Washington, DC, USA, 2003, 96~107
- 6 Vranic D, Aaupe D. 3D shape descriptor based on 3D fourier transform [A]. In: Proceedings of EURASIP Conference on Digital Signal Processing for Multimedia Communications and Services [C], Budapest, Hungary, 2001, 271~274
- 7 Kazhdan M, Funkhouser T, Rusinkiewicz S. Shape matching and an isotropy [J]. ACM Transactions on Graphics (2004 SIGGRAPH), 2004, 23(3): 623~629
- 8 Hesselink W in H, Visser M enq Roerdink Jos B T M. Euclidean skeletons of 3D data sets in linear time by the integer medial axis transform [A]. In: ISMM' 2005 [C], Paris, France, 2005, 259~268
- 9 Shilane P, Michael K, Patrick M, *et al*. The princeton shape benchmark [A]. In: Proceedings of the International Conference on Shape Modeling [C], Genova, Italy, 2004, 167~178

密度分布特征及其在二值图像检索中的应用

黄春木 周利莉

(解放军信息工程大学信号分析工程系, 郑州 450002)

摘要 图像的形状是描述图像的重要视觉和语义特征, 可通过图像中像素点的区域分布表现出来。为了对二值图像进行有效检索, 提出了一种基于区域的形状特征——密度分布特征, 用来进行二值图像检索。该方法在经过形心定位和子图像区域划分后, 可得到 2 个 M 维特征向量, 其中第 1 个表示各个子图像区域的目标像素的相对密度, 第 2 个表示各个子图像区域的目标像素在极坐标方向上的相对密度的一阶数值差分。在进行相似性度量时, 首先采用 Gaussian 模型对用这 2 个特征向量计算得到的距离分别进行归一化处理; 然后综合两个特征向量的距离计算总的相似度。实验结果表明, 密度分布特征不仅能够有效地刻画二值图像的形状, 具有非常好的平移、尺度和旋转不变性, 而且检索结果优于 Hu 不变矩。

关键词 形状特征 密度分布特征 不变性 二值图像检索

中图法分类号: TP391.3 文献标识码: A 文章编号: 1006-8961(2008)02-0307-05

Density Distribution Feature and its Application in Binary Image Retrieval

HUANG Chunmu, ZHOU Lili

(Information Engineering University, Zhengzhou 450002)

Abstract Shape is a very important visual and semantic feature used to depict image and it can be revealed by image pixels' regional distribution. This paper proposes a region-based shape representation, a new "density distribution feature (DDF)". After shape center orientation and region partition, two M dimensional feature vectors are got. The first feature vector represents the relatively density of object pixels within each sub-image. And the second represents the difference of relatively density in the direction of radial coordinates. When matching the similarity, we first used the Gaussian model to normalize the two dimensional feature vectors. Then we integrated them to calculate similarity distance. The experiments results showed that this shape feature can depict the image well and is invariant to translation, scale and rotation. The paper also evaluated the effectiveness of the proposed descriptor with respect to Moment Invariants.

Keywords shape representation, density distribution feature, invariant performance, binary image retrieval

1 引言

基于形状特征的索引技术的研究目前已在基于内容的图像检索中得到充分重视。目前, 基于形状特征的图像检索大致分为以下两类: 基于图像对象区域的局部形状特征检索和基于全局形状特征的检索。其中前者是指在特征提取阶段, 先提取图像中相关对象的形状区域或边界等的完整描述, 然后利用该区域及其特征进行图像检索, 但该方法要求对

图像进行图像分割操作, 因此其检索规模和应用范围受到一定限制^[1, 2]; 而后者则无须进行图像分割, 可直接利用能反映图像形状的全局特征量来进行图像检索^[3], 其比较适合于大型的通用图像数据库。由于基于区域的形状描述方法利用了边界信息和内部信息, 并且通常不需要知道形状的轮廓, 所以可应用于一般基于形状的图像检索。

本文给出了一种基于区域的形状描述和相似性匹配算法, 并应用于二值图像检索, 且与最常用的一种方法——Hu 不变矩^[4]进行了比较。

收稿日期: 2006-03-29 改回日期: 2006-09-18

第一作者简介: 黄春木 (1981~), 男, 硕士研究生。主要研究方向为数字图像处理、图像检索技术。Email: chummu@163.com

2 密度分布特征

图像的形状是通过像素点分布在空间不同的区域而表现出来的。二值图像中,由于白色像素点(背景点)和黑色像素点(目标点)的交错分布构成了各种各样的目标形状,因此在对二值图像的形状进行描述时,目标像素点的区域分布是很重要的信息。对于两幅图案相似的二值图像来说,由于它们的目标像素点的区域分布也是相似的,而密度分布特征方法的基本思想就是通过统计图像中目标像素点在不同空间区域的分布情况来得到密度分布特征,因此这个形状特征可以用于对二值图像进行分类或检索等。下面介绍密度分布特征的提取技术。

2.1 形心定位和区域划分

一般对于一幅二值图像 $f(x, y)$ 来说,由于目标只占了其中的部分区域(如图 1 所示),因而在计算距离时,形心的选取很重要。一般选取图像的重心点 (\bar{x}, \bar{y}) 作为形心,其中

$$\bar{x} = \frac{\sum_i \sum_j i \times f(i, j)}{\sum_i \sum_j f(i, j)} \quad (1)$$

$$\bar{y} = \frac{\sum_i \sum_j j \times f(i, j)}{\sum_i \sum_j f(i, j)} \quad (2)$$

式中,

$$f(i, j) = \begin{cases} 1 & \text{if } (i, j) \in \mathbf{R} \\ 0 & \text{其他} \end{cases} \quad (3)$$

\mathbf{R} 表示二值图像中的目标像素区域。

在得到形心后,先计算图像中各目标像素点到形心的距离,并找出最大距离 D_{\max} ,然后采用圆形划分法对图像进行分块,即首先以形心为圆心,以 D_{\max}

为半径作图像中目标区域的外接圆,然后在外接圆内,由内而外,将图像划分为 M 个子图像区域 $\mathbf{R}_1, \dots, \mathbf{R}_M$, 分别为一个中心圆和多个同心圆环。子图像区域的划分方法有以下两种^[5]:

(1) 等距离区域划分法 即每个子图像区域所覆盖的区间长度相等,此时

$$\mathbf{R}_i = \{(x, y) \mid (i-1) \times D_{\max} / M < \sqrt{(x-\bar{x})^2 + (y-\bar{y})^2} \leq i \times D_{\max} / M, (x, y) \in \mathbf{I}\} \quad (4)$$

式中, $1 \leq i \leq M$ 。

(2) 等面积区域划分法 即每个子图像区域包含的面积是相等的,此时

$$\mathbf{R}_i = \{(x, y) \mid (i-1) \times D_{\max}^2 / M < [(x-\bar{x})^2 + (y-\bar{y})^2] \leq i \times D_{\max}^2 / M, (x, y) \in \mathbf{I}\} \quad (5)$$

式中, $1 \leq i \leq M$ 。

图 2 给出了一个区域划分的例子,其中,图 2(a) 是一幅二值图像,图 2(b) 给出了等距离区域划分法得到的划分结果,图中每个子图像区域(圆环或中心圆)所覆盖的区间长度(圆环的宽度或圆的半径)相同;图 2(c) 给出了等面积区域划分法的划分结果,图中每个子图像区域的面积都是相等的。

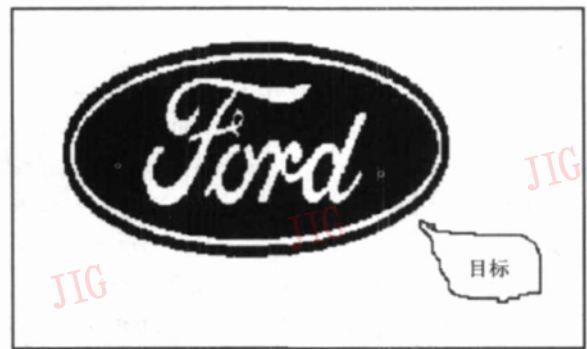
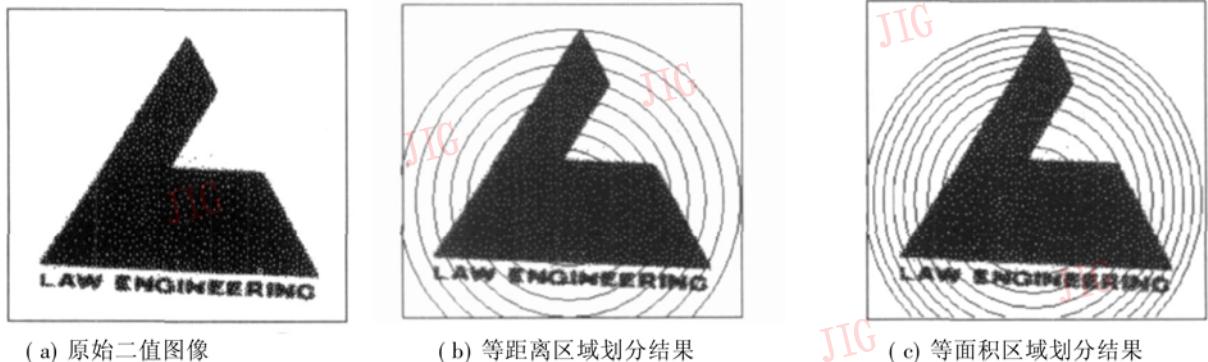


图 1 图像中的目标区域

Fig 1 Object region in binary image



(a) 原始二值图像

(b) 等距离区域划分结果

(c) 等面积区域划分结果

图 2 区域划分的一个例子

Fig 2 An instance of region partition

2.2 密度分布特征

密度分布特征 (density distribution feature, DDF) 是一种反映图像目标像素的空间分布信息的特征, 这里把它定义为

$$DDF = (r_1, \dots, r_M; dr_1, \dots, dr_M) \quad (6)$$

其中, 包含两个 M 维特征向量, 第 1 个向量 r 表示各个子图像区域的目标像素的相对密度, 第 2 个向量 dr 表示各个子图像区域的目标像素在极坐标方向上的相对密度的一阶数值差分^[6]。下面详细介绍计算密度分布特征的步骤:

- (1) 计算图像 $f(x, y)$ 的形心, 即重心 (\bar{x}, \bar{y}) ;
- (2) 计算图像 $f(x, y)$ 中, 目标像素点到形心的最大距离 D_{max} ;

(3) 先计算图像 $f(x, y)$ 中, 以形心为圆心, 以 D_{max} 为半径的目标区域外接圆, 然后在外接圆内, 使用等距离区域划分法或等面积区域划分法将图像划分为 M 个子图像区域 ($M > 0$);

(4) 对各子图像区域分别进行统计, 即计算各个子图像区域内目标像素的总数 $S_i (i = 1, \dots, M)$, 并找出 S_i 的最大值 $S_{max} = \max_{i=1, \dots, M} (S_i)$;

(5) 计算密度分布特征的第 1 个特征向量 r 的以下各个分量:

$$r_i = S_i / S_{max} (i = 1, \dots, M) \quad (7)$$

(6) 计算密度分布特征的第 2 个特征向量 dr 的以下各个分量:

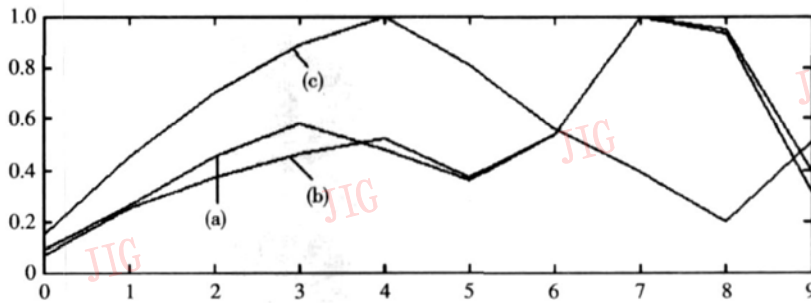
$$dr_i = \begin{cases} |r_1 - r_2| & i = 1 \\ |2r_i - r_{i-1} - r_{i+1}| & 1 < i < M \\ |r_M - r_{M-1}| & i = M \end{cases} \quad (8)$$

(7) 按照以下顺序组织密度分布特征:

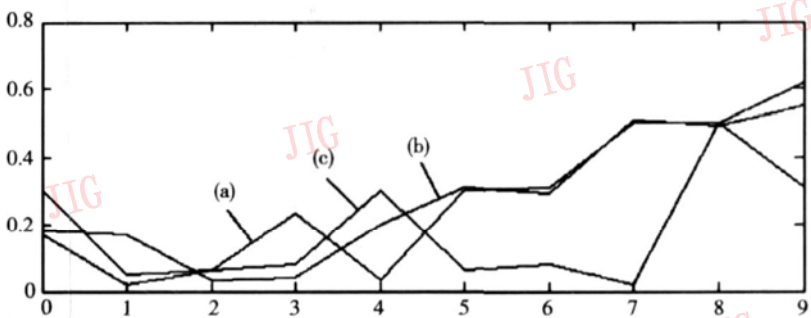
$$DDF = (r_1, \dots, r_M; dr_1, \dots, dr_M)$$

2.3 特征提取结果和分析

图 3(a) ~ 图 3(c) 的密度分布特征的第 1 维向量和第 2 维向量 dr 的对比图见图 3(d) 和



(d) 图 3(a) ~ 图 3(c) 密度分布特征的第 1 维向量



(e) 图 3(a) ~ 图 3(c) 密度分布特征的第 2 维向量

图 3 密度分布特征对比图

Fig. 3 Contrast of DDF

图 3(e), 曲线 (a) 代表图 3(a), 曲线 (b) 代表图 3(b), 曲线 (c) 代表图 3(c)。从图中可看出, 对于有相同形状区域的二值图像, 其密度分布特征大致也是相似的, 如曲线 (a) 和 (b)。这说明, 密度分布特征能够反映出二值图像的总形状信息, 是一种有效的形状特征。

用密度分布特征来描述二值图像的形状特征, 其优点是:

- (1) 密度分布特征可以抓住总体的形状信息;
- (2) 密度分布特征具有平移不变性, 因为其是选择重心作为形心, 并且子图像区域的划分是针对目标区域的, 所以目标在图像中的位置不影响密度分布特征的值;
- (3) 密度分布特征具有尺度不变性, 由于所有图像的目标区域均被划分为相同数目的子图像区域, 所以对图像进行放大或缩小后获得的图像, 其密度分布特征与原图的密度分布特征基本相同;
- (4) 密度分布特征具有旋转不变性, 因为采取的是圆形划分法, 所以旋转变换对于密度分布特征基本没有影响。

其局限性为对于总体上完全不同的图像也可能产生相似的密度分布特征。

3 在二值图像检索中的应用

为验证密度分布特征在二值图像检索中的应用效果, 用密度分布特征作为形状特征来对二值图像进行检索实验, 实验中使用 1000 幅二值商标图像组成的图像库。从实验结果可以看出, 密度分布特征确实不仅具有非常好的平移、尺度和旋转不变性, 而且具有很好的检索性能。

3.1 相似性度量

图像间的相似性可以用相应特征之间的距离进行度量, 本文采用欧氏距离进行相似性度量。度量时, 首先分别计算密度分布特征的两个特征向量的欧氏距离 d_r 和 d_{ar} ; 然后计算两个特征向量的综合距离。由于两个特征向量处于不同的特征空间, 所以在融合这两个特征前有必要对其归一化。和特征内部各分量的归一化不同, 这种归一化是对图像间某一特征的相似距离进行归一化, 在这里采用 Gaussian 模型对距离进行归一化^[7]:

计算查询图像与图像库中每幅图像对应的某个子特征的距离, 再计算出这些距离的均值 μ 和标准

差 σ 。距离归一化公式为

$$\tilde{d} = \begin{cases} 1 & d < \mu - 3\sigma \\ \frac{d - \mu}{3\sigma} + 1 & \mu - 3\sigma \leq d \leq \mu + 3\sigma \\ 1 & d > \mu + 3\sigma \end{cases} \quad (9)$$

归一化后两个特征分量的距离为 \tilde{d}_r 和 \tilde{d}_{ar} , 都归一化到 $[0, 1]$ 区间, 可以进行如下融合:

$$d_{DDF} = \alpha \cdot \tilde{d}_r + \beta \cdot \tilde{d}_{ar} \quad (10)$$

其中, α 、 β 分别为两个特征分量距离的权重, $\alpha + \beta = 1$ 。 d_{DDF} 为两幅图像密度分布特征的距离。实验中取 $\alpha = \beta = 0.5$ 。

3.2 实验结果

为了检验基于密度分布特征的检索效果, 从图像库中选出具有代表性的 20 幅图像作为查询图像进行了图像检索实验。图 4 给出了部分查询图像。对于每一幅图, 实验时, 首先对它们进行平移、旋转和缩放, 得到 5 幅图像, 然后加上从图像库中挑出的一些相似图像, 组成查询图像的相似图像集。这样, 对于这 20 幅查询图像, 图像库里总共有 1100 幅图像。

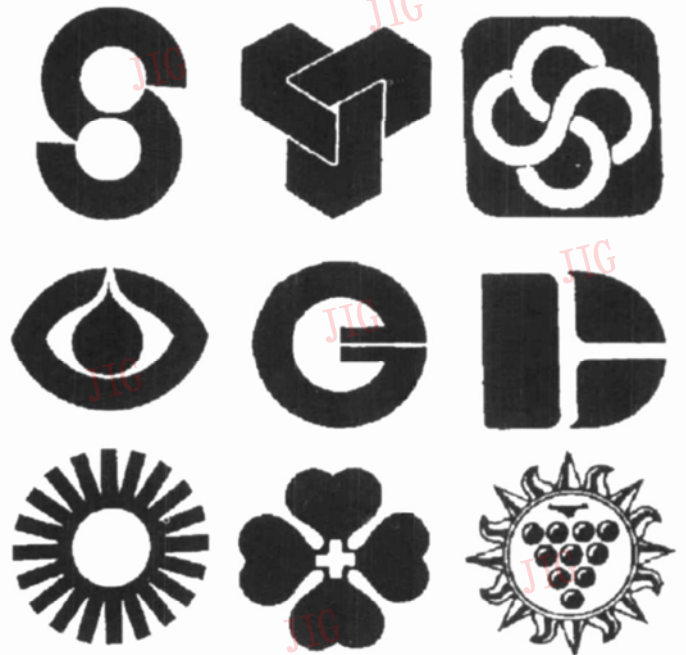


图 4 部分查询样图

Fig 4 Nine sample images in the database

在进行图像检索性能评价时, 以检索精度与检索回召率相关曲线为评价指标, 其中, 检索精度定义为检索出的有效图像数与检索出的图像总数的比率, 而检索回召率定义为检索出的有效图像数与图像库中总有效图像数的比率。检索出的有效图像是指图像库中与样本图像相似的图像。20 幅检索图

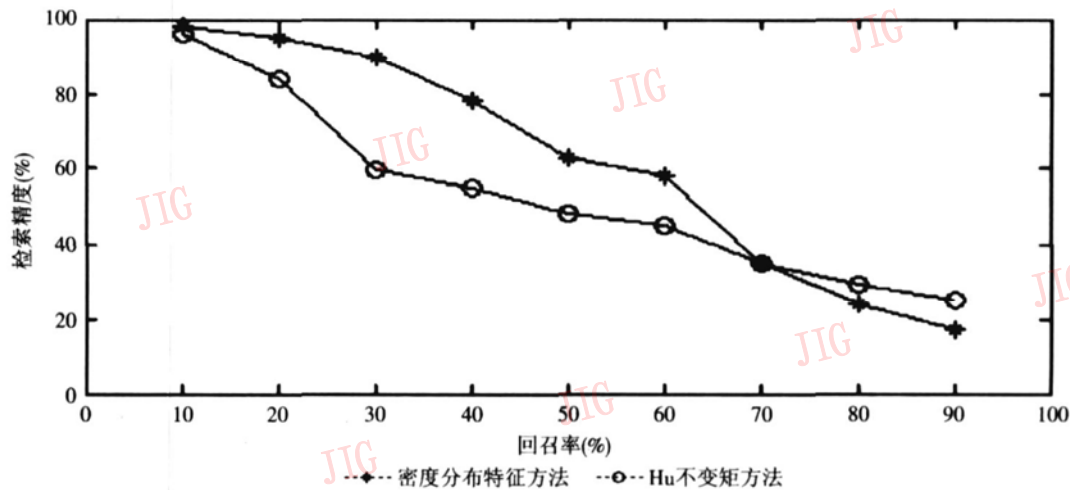


图 5 检索结果比较图

Fig 5 Recall-precision curves

像的平均检索结果如图 5 所示, 图中还给出了 Hu 不变矩的检索结果。由图 5 可知, 在相同检索回召率的情况下, 本文方法的检索精度比基于 Hu 不变矩的检索精度有一定提高。图中较高的检索精度是因为相似图像集中有 5 幅查询图像的变换图像, 这也说明了密度分布特征具有非常好的平移、旋转和缩放不变性。

由此可见, 密度分布特征不仅检索性能优于 Hu 不变矩, 而且提取方法简单, 易于计算机实现。对一幅 256×256 大小的二值图像, 在 intel P3 500PC 机上, 提取密度分布特征所需的 CPU 时间为 10ms 而提取 Hu 不变矩所需时间为 50ms。这对于大型数据库, 密度分布特征明显具有比较大的优势。

4 结 论

密度分布特征由于能够有效地刻画二值图像的形状特征, 因而在二值图像检索中取得了良好的检索结果。而且其提取方法简单, 易于计算机实现, 具有较强的实用性。但是, 由于这个特征是基于区域统计的, 对于总体上完全不同的图像也可能产生相似的密度分布特征, 少量与目标图像不相似的图像也会被检索出来, 所以今后在二值图像检索中, 需要研究如何结合其他的形状特征来对目标进行更全面的描述, 以获得更加符合人眼视觉感受的检索结果。

参考文献 (References)

- 1 Chen Y X, Wang J Z. A region-based fuzzy feature matching approach to content-based image retrieval [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(9): 1252~1267.
- 2 Wang J Z. Pathfinder: Multi-resolution region-based searching of pathology image using IRM [A]. In: Proceedings/AMIA Annual Symposium [C], Los Angeles, CA, USA, 2000: 883~887.
- 3 Jain A K, Vailay A. Shape-based retrieval: A case study with trademark image database [J]. Pattern Recognition, 1998, 31(9): 1369~1390.
- 4 Hu M K. Visual pattern recognition by moment invariants [J]. IRE Transactions on Information Theory, 1962, IT-8: 179~187.
- 5 Guo Li, Sun Xing-hua, Huang Yuan-yuan, et al. Distance distribution histogram and its application in trademark image retrieval [J]. Journal of Image and Graphics, 2002, 7A(10): 1027~1031. [郭丽, 孙兴华, 黄元元等. 距离分布直方图及其在商标图案检索中的应用 [J]. 中国图象图形学报, 2002, 7A(10): 1027~1031.]
- 6 Liu H, Feng S, Zha H. Document image retrieval based on density distribution feature and key block feature [A]. In: Proceedings of 8th International Conference on Document Analysis and Recognition [C], Seoul, Korea, 2005: 1040~1044.
- 7 Liu Ning-zhong, Yang Jing-yu. Color image retrieval using both color and texture features in object region [J]. Computer Engineering and Applications, 2003, 39(32): 187~203. [刘宁钟, 杨静宇. 综合利用目标区域颜色纹理特征的彩色图像检索 [J]. 计算机工程与应用, 2003, 39(32): 187~203.]

水下 3 维声场体可视化两种切平面实现方法

王小林 笪良龙 杨廷武 李玉阳

(海军潜艇学院海军战术水声数据中心, 青岛 266071)

摘要 为了深入研究海洋水声环境对海军作战的影响, 在水下声场体可视化的基础上对能量场进行 3 维剖分技术研究。根据不同的战术目的, 探讨了两种不同的切平面剖分绘制方法。这两种方法相互补充, 能够使指挥员直观准确地理解水下战场环境。

关键词 水下 3 维声场 体可视化 切平面 体绘制

中图法分类号: TP391.1 文献标识码: A 文章编号: 1006-8961(2008)02-0312-04

The Realization of Two Clipping Plane Technologies in Underwater Acoustic Energy Fields Visualization

WANG Xiao-lin, DA Liang-long, YANG Ting-wu, LI Yu-yang
(Navy Submarine Academy Navy Tactical Acoustic Data Base Center, Qing Dao 266071)

Abstract This article researches the clipping plane technology during volume visualization in underwater acoustic energy fields to study the influence to naval battles and brings forward two methods to create clipping plane according to different requests of tactics. These two methods supply each other, and commanders can get the underwater sound field environment and battle information quickly.

Keywords underwater acoustic energy fields, volume visualization, clipping plane, volume rendering

1 引言

海洋水声环境对海军作战具有重要影响, 而海洋水下声传播是决定作战舰艇探测与反探测效果的直接因素, 因此如何直观、有效地刻画和描述水下声学环境对提高作战指挥人员对环境效应的理解和应用具有重要意义。同时随着我军现代化、信息化建设步伐的加速发展, 对战场环境信息保障能力提出了更高的要求, 可视化技术作为海军海洋战场环境信息化建设的重要支撑技术之一, 对水声战场环境的形象描述、直观了解以及智能化辅助决策具有重要作用, 因而得到了广泛应用和快速发展, 并在现代战争中起到了日益显著的作用。

近年来, 一些成熟的体可视化切片技术^[1]在医学上得到了广泛应用, 但在水下声学能量场的体可

视化技术, 特别是在声场可视化条件下的 3 维切片与定量分析方面的研究还较少。本文探讨了该技术在水下声场可视化切平面中的实现方法。首先利用 PE (parabolic equation) 声传播模型^[2]仿真计算海洋声学能量场, 然后, 将这些数据体绘制形成直观的 3 维水下声场, 通过利用体可视化切片, 从而可以清楚地得到声场内部结构并对其进行定量分析, 为潜艇战术行动的优化提供可靠的技术支持。同时针对水下 3 维声场可视化过程中不同的战术需求, 提出了两种能量场的切平面绘制方法, 并对其绘制效果进行了比较。

2 3 维声场 PE 模型

海洋是一个复杂的系统, 介质往往是分层的, 不同层之间由于声速或密度的不连续而产生了分

基金项目: 国防预研基金项目 (9140A0305206JB1501); 新世纪优秀人才支持计划 (NCET)

收稿日期: 2006-03-09 改回日期: 2006-09-08

第一作者简介: 王小林 (1976~), 男, 海军上尉, 2004 年军事学硕士研究生。主要从事水下 3 维可视化研究。E-mail: sr71blackbird@tom.com

界面, 在分界面上满足界面连续条件。采用 PE 模型来获得可视化 3 维声场的体数据, 优点之一就是可以求解这种声速、密度不均匀、海底地形不规则介质中的声传播问题。另一方面, 这种方法在处理声速和密度的不均匀性时, 提高了抛物方程计算精度。

在柱坐标下, 3 维 Helmholtz 方程表示为

$$\frac{\partial^2 p}{\partial r^2} + \frac{1}{r} \frac{\partial p}{\partial r} + \frac{1}{r^2} \frac{\partial^2 p}{\partial \theta^2} + \frac{\partial^2 p}{\partial z^2} + k_0^2 n^2 p = 0 \quad (1)$$

利用 Tappert 引入的分离变量法, 令

$$p = u(r, \theta, z) \cdot v(r) \quad (2)$$

其中, $v(r)$ 为 p 中随 r 显著变化部分, $u(r, \theta, z)$ 则是慢变部分, 代入式 (1), 可得

$$\left| v_{rr} + \frac{1}{r} v_r \right| + \left| u_{rr} + u_{zz} + \left| \frac{1}{r} + \frac{2}{v} v_r \right| u_r \right| + \left| \frac{1}{r^2} u_{\theta\theta} + k_0^2 n^2(r, \theta, z) u \right| = 0 \quad (3)$$

用 k_0 作为分离常数和通过远场近似, 并在考虑 1 维扩散波动方程的基础上, 由式 (3) 可得以下方程:

$$u_r = ik_0 \left(-1 + \sqrt{1 + X + Y} \right) u \quad (4)$$

其中

$$X = [n^2(r, \theta, z)] + \frac{1}{k_0^2} \frac{\partial}{\partial z^2} \quad (5)$$

$$Y = \frac{1}{(k_0 r)^2} \frac{\partial^2}{\partial \theta^2} \quad (6)$$

对于上式根号下部分采用不同近似展开, 得到不同的抛物方程, 在这里采用 LSS(lee-saad-schultz) 展开。采用线性多项式近似:

$$\sqrt{1 + X + Y} \cong 1 + \frac{1}{2}X - \frac{1}{8}X^2 + \frac{1}{2}Y \quad (7)$$

代入式 (4), 考虑 1 维扩散波动方程:

$$u_r = ik_0 \left| -1 + \left| 1 + \frac{1}{2}X - \frac{1}{8}X^2 + \frac{1}{2}Y \right| \right| u \quad (8)$$

LSS 3 维宽角波动方程, 相应地, 式 (8) 变为

$$u(r + \Delta r, \theta, z) = e^{-\delta} e^{\left| 1 + \frac{1}{2}X - \frac{1}{8}X^2 + \frac{1}{2}Y \right|} u(r, \theta, z) \quad (9)$$

采用有限差分的方法通过上式来解 LSS 方程, 就得到了水下 3 维声场体可视化数据。

3 3 维声场体可视化

首先利用基于硬件加速的 3 维纹理直接体可视化^[3]方法, 对水下 3 维声能量场进行体可视化, 其过程如图 1 所示。

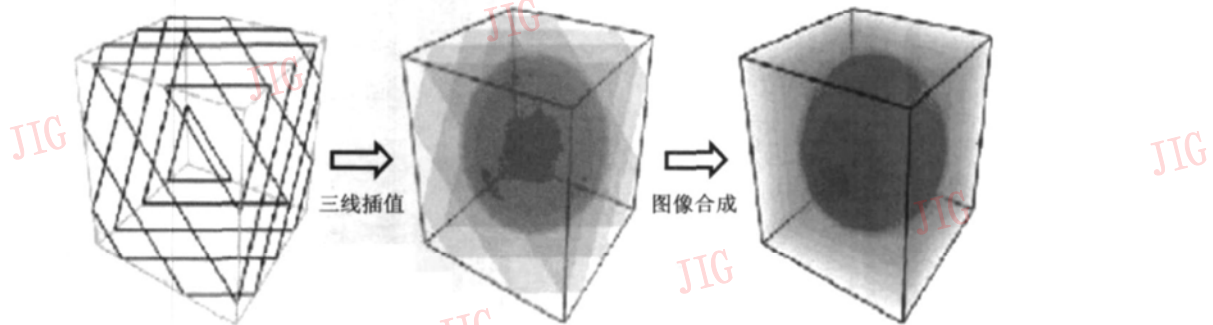


图 1 基于 3 维纹理硬件的直接体可视化方法

Fig 1 The pre-integrated volume rendering method in 3D texture graphic hardware

(1) 纹理生成 3 维数据场本身作为 3 维纹理图, 装入纹理内存。体数据的中心与物体空间的坐标原点重合, 其边界分别平行于物体空间的 3 根坐标轴。

(2) 三线性插值——重采样 在纹理空间内定义一系列采样多边形, 这些多边形彼此平行且与视线方向垂直。由于多边形的间隔及其采样密度与原始数据不同, 因此必须通过硬件的三线性插值即重采样才能获得这一系列平行多边形平面上各采样点的数值。

(3) 图像合成 首先将采样多边形上各采样点的数值按照给定的传递函数转换为相应的颜色值和不透明度值, 然后再根据适当的混合函数利用硬件的混合功能按照从后向前的顺序进行合成, 投影于视平面而形成最后的图像。图 2 为典型深海海域水下 3 维声场体可视化效果图。

从图 2 中, 可以清楚地看到处于声场中心的舰艇所产生声能量的传播情况。中心区域为声能量集中的海域, 声能量不断向外传播扩散, 渐渐减弱。声能量在大洋海底和海面经过多次反转, 并在海面产

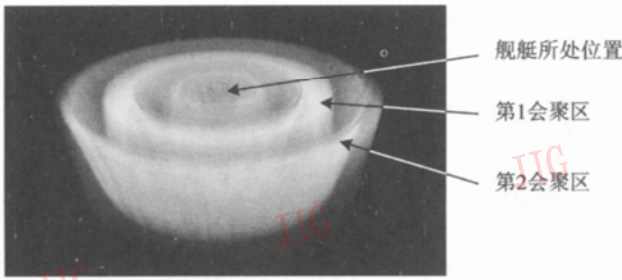


图 2 典型深海海域水下 3 维声场体可视化效果图
 Fig 2 The typical 3D visual picture of a deep-water sound field environment

生了会聚区,这就是会聚区现象。图 2 将科学计算所得的庞大无规律的体数据中隐含的科学规律和信息直观形象地展现到指挥员面前,大大提高了对作战海域声场环境的理解速度。同时,也为下一步深入研究声场内部的复杂结构奠定了基础。

4 能量场切平面

处于大洋深处的潜艇唯一能够利用的探测手段就是应用水声传播规律发现敌方目标。根据水声环境预报,详细深入地研究所处海域的声场环境及其内部结构是十分必要的。下面以会聚区声场模型为例,探讨两种对可视化声场的两种剖分方法。

4.1 切平面实现方法一

直接利用 OpenGL 切割平面生成函数 `glClipPlane` 对 3 维声场进行剖分,切片效果如图 3 所示。代码如下:

```
glClipPlane(GL_CLIP_PLANE0, clip0);
glClipPlane(GL_CLIP_PLANE0+ 1, qb); //生成切片 qb clip
glEnable(GL_CLIP_PLANE0)。
```

在眼坐标系下,人的眼睛是看向 Z 轴的负方向的,如视点 1。在坐标原点 O 处,体绘制出水下 3 维声场。在声场的周围生成一个切片 qb clip 这个切

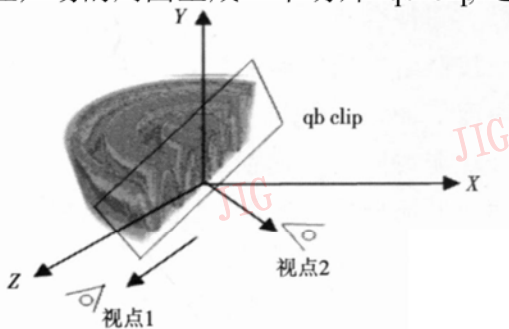


图 3 声场切平面实现方法一原理图
 Fig 3 The theory picture of sound field environment clipping planes

片是受鼠标控制的,在鼠标的拖动下,切平面一侧的纹理不再显示出来,形成一个 3 维剖面。同时,通过鼠标的拉伸, qb clip 切片可沿法线方向平移。根据这个剖面,就直观地看到声场内部的结构,而且可以看到任意方向和深度上、切面上的变化情况。

该方法随着切片的拖动,可以任意方向和深度剖分能量场,但不能正视切平面。只有当观察者在视点 2 时才能观察到切平面。在实际应用中,采用了视点转移的方法,即时获得切平面 qb clip 的法向量,将视点 2 定位在切平面的法线上。因此,随着切平面的旋转,视点也跟踪其进行旋转。正视后的效果如图 4 所示。

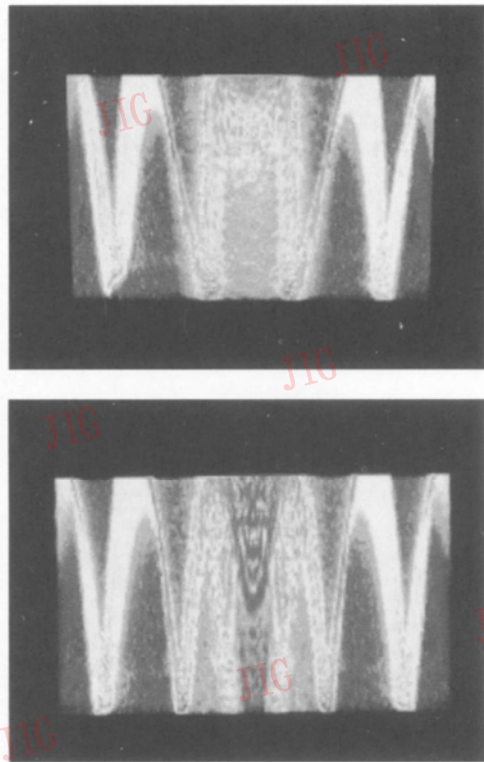


图 4 利用 OpenGL 切割平面生成切片效果
 Fig 4 The pictures of OpenGL clipping plane

该方法是应用切平面函数生成剖面,随着切平面的空间移位,可以生成空间上的任意切平面,剖分声能量场。采用这种方法,指挥官可以清楚地了解所处海域的水声环境,并对可能存在的水面舰艇的反潜防御区进行深入的剖分研究,为下一步战术行动的制定提供依据。但在实际应用中,除了要了解声能量场的空间结构,有时还需要了解切平面在空间能量场上的空间分布情况(具体区域的情况),如潜艇突破水面舰艇反潜封锁占位机动所需的具体位置声能量分布情况。因此提出第二种切平面实现方法。

4.2 切平面实现方法二

为了简化绘制进程, 直接观察到切平面, 方法二从直接体可视化的原理出发, 直接生成了正对于视点切平面。

在三线性插值——重采样过程中, 在纹理空间内定义一系列采样多边形, 这些多边形彼此平行且与视线方向垂直, 即垂直于视线的一组平面, 通过控制该平面组的绘制数量, 即可将最近一个平面作为切平面, 实现了深度上剖切。通过获得旋转纹理矩阵, 可以使用鼠标控制 3 维纹理在图像空间任意方向旋转。这样就现实在任意方向上剖切。该方法切片效果如图 5 所示。代码如下:

```
r = 0.5 + (float) i / 100.0 // 用 clip_num 控制平面数量
if (r < clip_num)
{
    glBegin(GL_TRIANGLE_STRIP);
    glVertex3f(-1.0, -1.0, r);
    glVertex3f(1.0, -1.0, r);
    glVertex3f(-1.0, 1.0, r);
    glVertex3f(1.0, 1.0, r);
    glEnd();
}
float * m; // 实现纹理旋转
m = ball.GetMatrix();
cgGLSetMatrixParameterf(Track_Matrix, m);
```

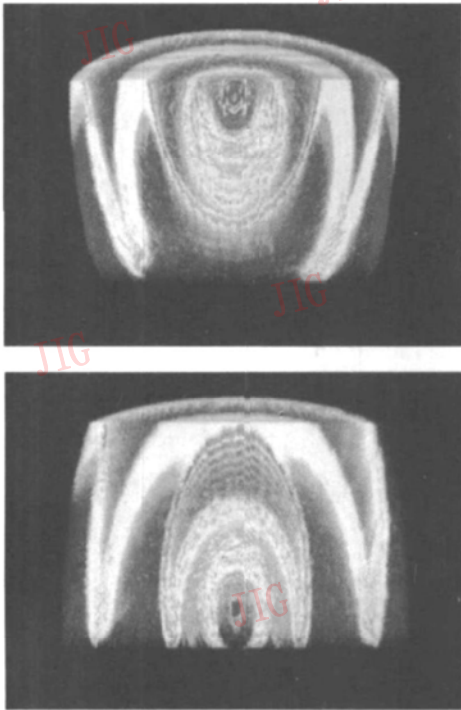


图 5 采用体绘制过程中控制切平面数量的效果图

Fig 5 The pictures acquired by controlling the clipping plane number

该方法原理简单, 切面直接正对于视点。潜艇指挥员制定接敌机动方案时, 可以以我艇为中心剖分声场, 在相对敌水面舰艇不同的方位和距离上采用不同的航线和航行状态进行机动。

5 结 论

利用两种不同切平面方法实现对水下 3 维声能量场的可视化切平面操作并获得了较好的切片效果。切平面绘制方法一可以在任意方位和深度上切割能量场, 受控的是切平面本身, 过程直观, 但正视后不能直观了解切平面在能量场中的空间位置分布。而切平面绘制方法二以控制绘制切片数量来达到剖切目的, 由于是纹理转动而切片不动, 因而解决了正视化问题, 视觉效果更好。因此, 实际应用中, 应根据需求不同而选择相应方法。

参考文献 (References)

- 1 Tang Sheng-ze, Chen Li, Deng Jun-hui. Volume Visualization of 3D Data Field[M]. Beijing: Qinghua University Publishing Company, 1999: 239. [唐泽圣, 陈莉, 邓俊辉. 三维数据场可视化[M]. 北京: 清华大学出版社, 1999: 239.]
- 2 Ding Lee, Martin H. Schulte. Numerical Ocean Acoustic Propagation in three Dimensions[M]. Singapore: World Scientific, 1995: 14.
- 3 Van Gelder, Kim K. Direct volume rendering with shading via three-dimensional textures[A]. In: ACM/IEEE Symposium on Volume Visualization[C], San Francisco, USA, 1996: 22~30.
- 4 Westermann R, Ertl T. Efficiently using graphics hardware in volume rendering applications[J]. Computer Graphics (SIGGRAPH'98), 1998, 32(4): 169~179.

基于动态纹理载入的大规模数据场体绘制

郑 杰 姬红兵

(西安电子科技大学电子工程学院, 西安 710071)

摘 要 为克服图形硬件对传统纹理映射体绘制的限制, 提出了一种在普通 PC 上进行大规模数据场体绘制的有效方法。该方法中, 体数据被划分为合适大小的数据块, 这些数据块被动态的载入图形硬件, 并利用 3 维纹理映射进行绘制。在整个绘制过程中, 仅有一个数据块存储在图形硬件上, 有效地提高了对大规模体数据的绘制能力。同时, 充分利用目前 PC 图形硬件成熟的可编程特性, 通过对梯度的实时计算来减少在传统纹理映射体绘制中巨大的内存消耗。实验结果表明, 该方法在普通 PC 上可以对超过纹理内存容量的大规模体数据进行交互式体绘制。

关键词 体绘制 3 维纹理映射 纹理分块 图形处理单元

中图法分类号: TP391.41 文献标识码: A 文章编号: 1006-8961(2008)02-0316-06

Interactive PC Texture-based Volume Rendering for Large Datasets

ZHENG Jie JI Hong-bing

(School of Electronic Engineering, Xidian University, Xi'an 710071)

Abstract A novel technique is presented for rendering large-scale volume datasets interactively on general purpose PC hardware. To circumvent the limited texture memory for texture based volume rendering, the dataset is partitioned into the bricks with reasonable size. The bricks are loaded to the graphics hardware dynamically and rendered using 3D texture mapping. During the rendering only one brick resides on the texture memory. Additionally, the sophisticated PC graphics hardware functionality is utilized to estimate the gradient on the fly avoiding the huge memory consumption in previous approaches. Using a prototype implementation of the algorithm, we are able to perform fast data loading and interactive visualization for the large datasets on a single standard PC.

Keywords volume rendering 3D texture mapping texture partitioning graphics process unit

1 引 言

体绘制对于由各种类型的 3 维设备、科学仿真等产生的体数据的可视化是一种非常重要的技术。为了克服绘制中大量的计算和对带宽的极高要求, 基于图形硬件纹理映射的体绘制逐步成为对规则网格数据场绘制的最为实际可行的方法。在体绘制中, 体数据作为 3 维纹理被载入图形硬件, 在用切片对纹理采样时, 利用图形硬件的高速并行处理能力进行 3 次线性插值, 对规模适中的体数据进行可交

互的绘制并得到较高的图像质量。

GPU (graphics processing unit) 在 1999 年首先由 NVIDIA 公司提出, 被定义为“一个单芯片的处理器, 集成了几何变换、光照、三角形构造、裁剪和绘制引擎功能, 并具有每秒至少 1 千万个多边形的处理能力”。近几年来, GPU 得到了高速的发展, 目前最新的通用图形处理器, 如 NVIDIA 公司的 GeForce 7950 具有每秒可绘制 20 亿个顶点, 填充 240 亿个像素的处理能力, 在运算速度、图像质量和可编程性方面甚至都超过了专业的图形工作站, 可以对现实世界中的各种材质和光照效果进行更真实的建模仿

收稿日期: 2006-03-17; 改回日期: 2006-08-22

第一作者简介: 郑杰 (1976~), 男, 现为西安电子科技大学电子工程学院电路与系统专业博士研究生。主要研究领域为图像处理、计算机图形学、科学可视化等。E-mail: zhengjie_ec@yahoo.com.cn

真。相对于以前采用固定绘制管道的图形硬件,新的图形处理器在顶点处理和像素处理阶段提供了灵活的可编程特性,在运算上都支持 IEEE32 位浮点运算,并且支持相关纹理采样功能,可以将纹理作为内存来使用,以方便数据的索引访问。这些新特征的引入使得大规模数据场在普通 PC 上的实时绘制成为可能。

尽管目前的 PC 图形硬件广泛提供了对 3 维纹理的支持,但基于 3 维纹理映射的体绘制仍有一些局限性。当使用纹理映射对体数据进行绘制时,硬件没有提供对梯度计算的直接支持。为了克服这个缺陷,梯度数据要预先进行计算,并和体数据一同载入图形硬件中。这就需要 4 倍于体数据量的纹理显存来完成绘制。另一方面,由于目前通用图形硬件有限的纹理内存容量及设计上的限制,如 NVIDIA 的 GeForce 系列图形硬件^[1]最大支持 $512 \times 512 \times 512$ 的 3 维纹理,这就制约了大规模体数据的体绘制。在医学领域中,随着多排螺旋 CT 的广泛应用,在一次扫描过程中可以产生多达 1 024 层的体数据,仅体数据就需要 256MB 的纹理内存空间,传统的基于纹理映射的体绘制是无法满足这样的应用。

本文基于 PC 图形硬件提出了一种对大规模数据场的体绘制算法。在绘制前,将体数据预先划分为大小合适的数据块,通过动态纹理载入技术对每个数据块进行绘制。在整个体绘制过程中,纹理内存仅存储一个数据块,有效地提高了对大规模体数据的绘制能力。同时,利用目前图形硬件的可编程特性提出了一种梯度的实时计算方法,避免了梯度的预先计算和绘制时的存储,进一步减少了纹理内存的开销,也加快了数据载入的速度。这样在普通 PC 硬件上实现了对大规模体数据的交互体绘制。

2 相关研究

最重要的基于纹理映射的体绘制方法^[2]由 Cabral 等人提出,其利用了高端图形工作站中的 3 维纹理映射特性。近年来通用图形硬件的快速发展可以达到实时的 2 维、3 维纹理映射,基于硬件加速的纹理映射逐渐成为进行逼真光照计算以及其他体绘制技术的重要手段。然而,纹理内存容量是制约纹理映射技术应用的主要瓶颈,在利用 2 维或 3 维

纹理^[3,4]进行体绘制时,有限的纹理内存限制了对大规模体数据的交互绘制。

Guthe 等人利用小波变换对体数据进行无损压缩^[5],在绘制时对编码数据实时解码。但每个纹理需要在 CPU 上完成解码后载入图形硬件进行绘制。Kraus 和 Ertl 将体数据进行分块,然后再合成为大小一致的纹理^[6]。利用图形硬件的可编程特性完成最终的解码过程,重建出绘制所需的纹理信息。Li 等人利用数据相异性进行纹理组合来减少原始体数据纹理的大小^[7]。然而梯度信息需要预先计算并存储在图形硬件中,尽管对梯度纹理也进行了组合但仍然占用了大量的纹理内存。Schneider 等人提出采用矢量量化 (vector quantization) 技术对体数据进行压缩^[8],在一定的误差内可以自动得到压缩后的数据来近似表示原始数据。但仅能采用最近邻插值进行绘制,影响了绘制图像的质量。

另外可以利用 OpenGL 和 D3D 等标准图形 APIs 提供的纹理压缩技术对体数据进行压缩,然而到目前,两种 APIs 都不支持无损压缩。由于有损压缩带来了一定的纹理误差,尤其是对梯度信息影响更大,会显著的降低绘制质量^[9]。

本文的主要贡献是提出了一种新方法,利用动态纹理载入技术实现了在普通 PC 硬件上对大规模体数据的交互体绘制。这种方法没有使用复杂的编码方式对体数据进行压缩,避免了在绘制中耗时的解码操作对交互性的影响以及编解码带来的误差。在绘制中仅有一个合适大小的体数据块被载入 3 维纹理,克服了纹理内存对体数据规模的限制。同时,通过利用目前通用 PC 图形硬件的可编程特性对光照计算中的梯度进行实时的计算和归一化处理,该方法同时降低了 CPU 和 GPU (图形处理器) 的内存消耗,也加快了体数据的预处理过程。因此,这种方法非常适用于对绘制质量有较高要求的大规模体数据可视化应用中。

3 基于 PC 硬件大规模数据场体绘制

3.1 直接体绘制

直接体绘制技术通常被分为基于像空间的绘制,如射线投射法 (ray-casting) 和基于物空间的绘制,如足迹法 (splating)、剪切-曲变法 (shearwarp) 和基于纹理映射的方法。体绘制的主要任务是在每个像素上对体绘制积分进行近似计算,即沿着入射

光线对逐渐衰减的颜色和衰减系数进行累加。

假设入射光线表示为 $\mathbf{x}(\lambda)$, 其中 λ 为到观察点的距离。对空间中的任意一点 \mathbf{x} , 颜色强度为 $\mathbf{c}(\mathbf{x})$, 衰减系数为 $\tau(\mathbf{x})$, 这样体绘制积分可以表示为

$$I = \int_0^D \mathbf{c}(\mathbf{x}(\lambda)) e^{-\int_0^\lambda \tau(\mathbf{x}(\lambda)) d\lambda} d\lambda$$

D 为最大观察距离。在空间点 \mathbf{x} 处, 体素根据函数 $\mathbf{c}(\mathbf{x})$ 进行光线发射, 同时衰减系数 $\tau(\mathbf{x})$ 在观察点和体素位置上的累加对发射的能量进行减弱。

但是, 由于颜色和衰减系数没有指定, 对于连续标量场 $s(\mathbf{x})$ 上述的积分是无法进行计算的。首先要对体数据进行分类来指定颜色和衰减系数。通过引入传递函数 (transfer function) 来将体数据强度 $s = s(\mathbf{x})$ 映射为颜色强度 $\mathbf{c}(s)$ 和衰减系数 $\tau(s)$, \mathbf{C} 一般为颜色空间中的颜色矢量函数。

这样体绘制积分可以表示为

$$I = \int_0^D \mathbf{c}(s(\mathbf{x}(\lambda))) e^{-\int_0^\lambda \tau(s(\mathbf{x}(\lambda))) d\lambda} d\lambda \quad (1)$$

沿入射光线采样, 对体绘制积分进行离散化, 引入阻光度 A , 定义为

$$A_i = 1 - e^{-\tau(s(\mathbf{x}(id)))d} \quad (2)$$

同样, 第 i 个采样光线间隔上的颜色矢量可以近似表示为

$$\mathbf{C}_i = \mathbf{c}(s(\mathbf{x}(id)))d \quad (3)$$

这样, 离散化的体绘制积分可以表示如下, 其中 $N = [D/d]$ 为采样点的个数, d 为采样点间隔。

$$\mathbf{C} = \sum_{i=0}^N \mathbf{C}_i \prod_{j=0}^{i-1} (1 - A_j) \quad (4)$$

将最大观察距离为 D 的入射光线 $\mathbf{x}(\lambda)$ 分为连续的光线段, 这样第 i 个采样光线间隔上的颜色矢量 (如式 (4) 所示) 可以由每个光线段的颜色来近似表示

$$\mathbf{C} = \sum_{i=0}^{n_0} \mathbf{C}_i \prod_{j=0}^{i-1} (1 - A_j) + \sum_{i=n_1}^{n_2} \mathbf{C}_i \prod_{j=0}^{i-1} (1 - A_j) + \dots + \sum_{i=n_{L-1}}^{n_L} \mathbf{C}_i \prod_{j=0}^{i-1} (1 - A_j) \quad (5)$$

式中, $N = \sum_{k=0}^L n_k$, 每个部分都可以通过从前向后或从后向前进行 Alpha 混合迭代进行实现。

3.2 数据分块和纹理动态载入

平行于视平面的多边形切片是在基于 3 维纹理映射的绘制中最常用的方法。体数据被装载到一个

3 维纹理中, 在切片多边形内部上进行插值得到顶点的 3 个纹理坐标, 在光栅化过程中, 这些纹理坐标被用来对 3 维纹理进行采样, 利用硬件的 3 次线形插值得到相应的体数据。

将体数据划分为 L 个连续的数据块, 对第 k 个数据块, 其对体绘制积分中第 i 个采样光线间隔上的颜色矢量 \mathbf{C} 的贡献可以表示为

$$\mathbf{C}_k = \sum_{i=n_k}^{n_{k+1}} \mathbf{C}_i \prod_{j=0}^{i-1} (1 - A_j) \quad (6)$$

采样光线在数据块上的采样为 $[n_k, n_{k+1}]$ 。在基于 3 维纹理映射的方法中, 每个数据块可以表示为一个 3 维纹理, 采用平行于视平面的多边形切片对数据块采样, 并按照由前向后或由后向前的顺序混合得到该数据块的颜色贡献, 这些临时的颜色和阻光度信息在图形硬件的帧缓存中进行累加, 当完成对所有数据块的绘制时就生成了最终的绘制结果。

在数据的预处理过程中, 体数据被划分为一系列大小不同的数据块。这些数据块都是连续却又相互不交叠的, 在一次完整的绘制中, 每个数据块最多被采样一次, 以保证没有体素被进行多次混合。

在绘制中, 由式 (5) 可知, 每次仅绘制一个数据块, 并将每个数据块的绘制结果组合起来得到的图像和同时对所有数据块进行绘制得到的结果是一致的。这样可以按照正确的绘制顺序将一个数据块动态的载入 3 维纹理进行绘制, 而将其余的数据块存放在系统内存中。每次绘制完当前的数据块时, 下一个数据块被载入图形硬件并卸载当前的数据块。在整个绘制过程中, 只需在纹理内存中驻留一个数据块。这种动态纹理载入的特性使得提出的算法克服了目前通用图形硬件对绘制数据容量的限制。整个绘制流程如图 1 所示。

在进行数据分块时, 数据块数目的确定需要特别注意。数据块越多, 用在纹理切换的开销就会越高。然而增加数据块的大小, 相当于减少数据块的数目, 由于带宽的限制会增加数据由系统向图形硬件传输的延时。对于配有 256MB 内存的图形硬件, 一般数据块的大小为 $512 \times 512 \times 256$ 。

另外, 由于图形硬件不支持跨越数据块边界的插值计算, 这会在数据块之间的边界部分产生明显的伪影。为了保证 3 次线形插值能够在边界区域正确的进行, 需要重复数据块边界的体素。在进行数据分块时, 对每个数据块, 还要同时采样其在 x , y 和

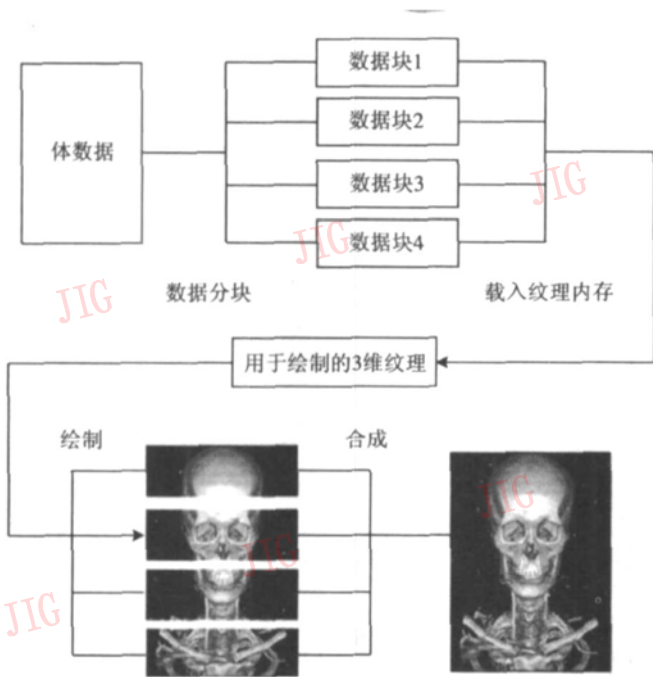


图 1 绘制流程

Fig 1 Rendering pipeline

z方向上相邻的数据块。如果数据块在整个体数据的边界上,则不在该侧进行采样。这样数据块相当于在 x、y 和 z 方向上增加了一个体素来保存边界邻域体素。在绘制中,利用这些邻域体素就可以在数据块边界处完成连续 3 次线性插值。同时由于这种技术使得数据块之间有一个体素的交叠,会带来内存使用效率的降低。

3.3 梯度的实时计算

在利用纹理映射绘制体数据时,通用图形硬件没有对梯度的计算提供支持。为了实现光照处理,通常要预先对体数据的表面梯度进行计算,梯度和体数据以 RGBA 的格式存储在纹理内存中。这样极大的占用了通用图形硬件中有限的内存空间,对一个 $512 \times 512 \times 512$ 规模的体数据,最少需要 512MB 的 3 维纹理内存来装载这样的数据,这已远远超出了目前大多数通用图形硬件的内存限制。

本文基于对梯度进行实时计算的思想来有效的减少纹理内存的巨大占用。这种方法仅在进行光照计算时,完成该点处梯度的计算。这样只需在纹理内存中存储原有传统纹理绘制 1/4 的数据,极大的增加了应用范围。同时也无需在预处理过程中进行耗时的梯度计算,直接载入体数据,提高了交互性。

为了在保证图像质量的同时兼顾运算的复杂性,采用计算简单的基于中心差分的梯度计算方法。对一个数据场 s 在体素 (x, y, z) 处的梯度为

$$g(x, y, z) = \nabla s(x_i, y_j, z_k) = \frac{1}{2} (s(x_{i+1}, y_j, z_k) - s(x_{i-1}, y_j, z_k)),$$

$$\frac{1}{2} (s(x_i, y_{j+1}, z_k) - s(x_i, y_{j-1}, z_k)),$$

$$\frac{1}{2} (s(x_i, y_j, z_{k+1}) - s(x_i, y_j, z_{k-1}))$$

近几年来,通用图形硬件在顶点处理和像素处理阶段都提供了灵活的可编程性。OpenGL 的 fragment program 特性应用在光栅化阶段,对每个像素进行处理。为了实时的计算体数据的归一化梯度,梯度“fragment program”应完成下面的任务:

- (1) 根据当前纹理坐标,得到 6 邻域纹理坐标;
- (2) 利用相关纹理采样得到所需 7 个体数据;
- (3) 计算并归一化梯度。

实现中使用了 20 条指令,其中 7 条为 3 维纹理寻址操作,6 条矢量运算和 7 条标量运算。在得到规一化的梯度后,就可以利用 Phong-Blinn 光照模型来计算采样点处的光照强度^[2]。

4 实验结果

用大量的医学临床数据对上述算法进行了测试实验,所有实验都是在一台 P4 2.8GHz, 2GB 内存,操作系统为 Windows XP 的 PC 机上完成的,显卡为配有 256MB 显存的 NVIDIA GeForce 7800 GT。采用 C++ 和 OpenGL 对提出的算法进行了实现。

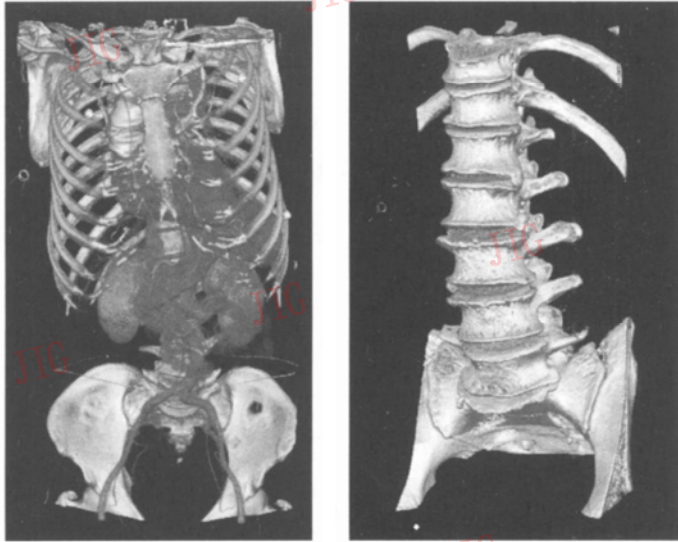
表 1 给出了实验数据的运行时间,包括体数据分块的预处理时间,数据块数目,3 维纹理消耗以及绘制的帧速率。因为在预处理过程中无需对梯度进行计算,也没有涉及数据编码过程,使得数据能够快速载入图形硬件。数据块的大小为 $512 \times 512 \times 256$ 3 维纹理占用均为 64MB。

表 1 性能参数,绘制窗口大小为 512×512

Tab 1 Performance statistics for a 512×512 viewport

体数据	数据规模	预处理 (s)	数据块数目	帧速率 (fps)
腹部	$512 \times 512 \times 527$	0.95	3	8.1
脊柱	$512 \times 512 \times 567$	0.99	3	8.1
人体-1	$512 \times 512 \times 690$	1.29	3	7.8
人体-2	$512 \times 512 \times 957$	1.71	4	5.6
下肢	$512 \times 512 \times 1156$	2.13	5	4.5
人体-3	$512 \times 512 \times 1582$	2.83	7	3.3

实验所采用的体数据均来自医学临床的 CT 扫描数据。图 2 为腹部器官和脊柱骨骼绘制效果, 可见本文提出的算法提供了理想的绘制质量。在绘制结果中可以清晰的看到血管分支, 而且在高密度部位, 如骨骼上没有出现“阶梯伪影”。

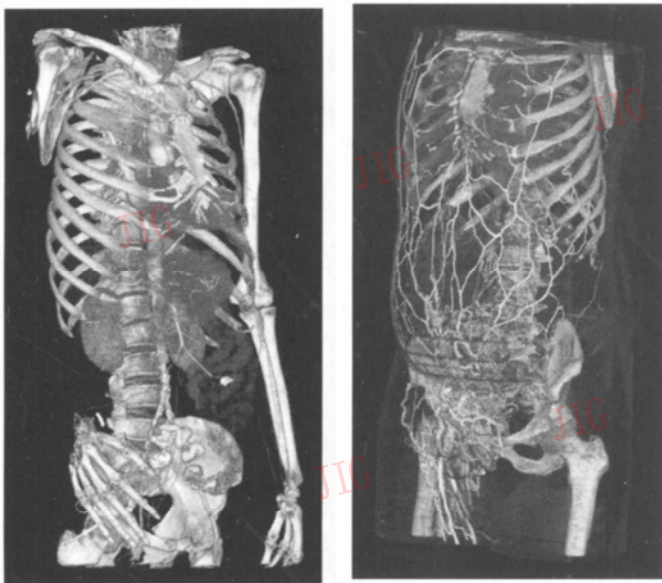


(a) 腹部 (b) 脊柱

图 2 腹部和脊柱绘制效果

Fig 2 Rendering of abdomen and spine

图 3 和图 4 为 4 个超出图形硬件限制的临床实际数据的绘制结果, 表现了算法对大规模体数据的绘制能力。从绘制图像上, 可清楚地显示出体数据中的各种内部结构, 不同器官的空间关系也非常直观的进行展示了。



(a) 人体-1 (b) 人体-2

图 3 人体-1 和人体-2 绘制效果

Fig 3 Rendering of body1 and body2



(a) 下肢 (b) 人体-3

图 4 下肢和人体-3 绘制效果

Fig 4 Rendering of lower limbs and body3

5 结 论

为克服传统基于 3 维纹理映射的体绘制方法中纹理内存对体数据规模的限制, 本文提出了一种新的体绘制算法, 在目前 PC 通用图形硬件上实现了大规模数据场的可视化。

本文提出的方法在数据预处理时, 将体数据划分为大小合适的数据块, 通过动态纹理载入技术将每个数据块按照正确的绘制顺序依次载入 3 维纹理, 并利用纹理映射进行绘制, 在图形硬件帧缓存中对数据块的绘制结果进行组合, 得到最终的绘制图像。在整个绘制过程中, 仅有一个数据块存储在图形硬件上, 实现了对超过纹理内存容量的大规模体数据的绘制。同时, 利用 OpenGL 的 fragment program 扩展将梯度的实时计算结合到绘制过程中, 明显的减少了传统纹理映射体绘制方法的巨大内存占用。实验结果表明, 利用目前通用图形硬件灵活的纹理操作和强大的光栅化能力, 可以得到具有较高质量的可交互的绘制结果。

在将来的工作中, 将利用本文提出的方法对时变 (time-varying) 体数据的可视化进行研究, 同时采

用更加灵活的数据分块算法, 提高纹理内存的利用效率。由于梯度对绘制质量非常重要, 但为了保证绘制交互的实时性, 本文采用了中心差分法来计算梯度。另一个可以进行的工作是采用更准确的梯度计算方法来提高精度, 同时可以使用目前最新的 GPU 提供的 32 位浮点运算能力进一步改善绘制质量。

参考文献 (References)

- 1 Zheng Jie, Ji Hong-bing. High quality per-pixel shading in texture-based volume rendering [A]. In Proceedings of IEEE the International Conference on Neural Networks and Brain [C], Beijing 2005. 1258~ 1261.
- 2 Cabral B, Can N, Foran J. Accelerated volume rendering and tomographic reconstruction using texture mapping hardware [A]. In Proceedings of ACM Symposium on Volume Visualization '94 [C], New York, USA, 1994. 91~ 98.
- 3 Engel K, Kraus M, Ertl T. High-quality pre-integrated volume rendering using hardware-accelerated pixel shading [A]. In Proceedings of ACM Eurographics/Siggraph Workshop Graphics Hardware 2001 [C], Los Angeles, CA, USA, 2001. 9~ 16.
- 4 Kniss J, Kindmann G, Hansen C. Multidimensional transfer functions for interactive volume rendering [J]. IEEE Transactions on Visualization and Computer Graphics, 2002, 8(3): 270~ 285.
- 5 Guthe S, W and M, Gonser J, Strasser W. Interactive rendering of large volume data sets [A]. In Proceedings of IEEE Visualization 2002 [C], Boston, MA, USA, 2002. 104~ 115.
- 6 Kraus M, Ertl T. Adaptive texture maps [A]. In Proceedings of the ACM Siggraph/Eurographics Graphics Hardware Workshop [C], Saarbrücken, Germany, 2002. 7~ 15.
- 7 Li W, Mueller K, Kaufman A. Empty space skipping and occlusion clipping for texture-based volume rendering [A]. In Proceedings of the 14th IEEE Visualization 2003 [C], Seattle, WA, USA, 2003. 317~ 324.
- 8 Schneider J, Westermann R. Compression domain volume rendering [A]. In Proceedings of the 14th IEEE Visualization 2003 [C], Seattle, WA, USA, 2003. 293~ 300.
- 9 Meißner M, Guthe S, Straßer W. Interactive lighting models and pre-integration for volume rendering on PC graphics accelerators [A]. In Proceedings of Conference on Human-Computer Interaction and Computer Graphics [C], Calgary, Alberta, Canada, 2002. 209~ 218.

一种用于半透明体绘制的光学模型

李彬¹⁾ 田联房¹⁾ 王立非²⁾ 陈萍³⁾ 毛宗源¹⁾

¹⁾ (华南理工大学自动化科学与工程学院, 广州 510640) ²⁾ (广州中医药大学祈福医院影像中心, 广州 511495)

³⁾ (广州医学院第一附属医院核医学科, 广州 510120)

摘要 为了高效率地重建 3 维医学图像, 并能充分显示内部隐含分界面及内部细节的详细信息, 提出了一种能在普通硬件条件下实现的用于半透明体绘制的光学模型 IVROM (improved volume rendering optical model)。该模型在光线吸收与发射模型的基础上, 考虑了阴影、直接散射与间接散射等因素。此外, 还详细介绍了结合该模型与 Shear-Warp 算法的半透明体绘制方法的实现。实验结果表明, 采用该模型的半透明体绘制方法不仅效率高, 而且显示效果好。

关键词 光学模型 Shear-Warp 医学图像 半透明体绘制 明暗化

中图法分类号: TP391.4 R319 文献标识码: A 文章编号: 1006-8961(2008)02-0322-06

An Optical Model for Translucent Volume Rendering

LI Bin¹⁾, TIAN Lian-fang¹⁾, WANG Li-fei²⁾, CHEN Ping³⁾, MAO Zong-yuan¹⁾

¹⁾ (College of Automation Science and Engineering, South China University of Technology, Guangzhou 510640)

²⁾ (Image Center, Clifford Hospital, Guangzhou University of Traditional Chinese Medicine, Guangzhou 511495)

³⁾ (Nuclear Medicine Department, The First Affiliated Hospital of Guangzhou Medical College, Guangzhou 510120)

Abstract In order to efficiently reconstruct 3D medical image and clearly display the detailed information of inner structures and the inner hidden interfaces between different mediums, an improved volume rendering optical model (IVROM) for translucent volume rendering is proposed which can be easily realized on a general PC. On the basis of classic absorption and emission model, the effects of volumetric shadows, direct and indirect scattering are also considered in the proposed model. Moreover, the implementation of the translucent volume rendering method incorporating the IVROM model and Shear-Warp is described detail. Experiments demonstrate the good performance of the proposed method.

Keywords optical model, Shear-Warp, medical image, translucent volume rendering, shading

1 引言

在医学图像可视化的辅助诊断和治疗中, 医生常常需要透过组织表面看清内部的病变结构和细节信息, 确定其相对位置, 从而确定诊断和治疗方案。因此, 能透过对象表面, 显示内部的隐含分界面 (内部不同介质之间的界面) 及其内部细节的半透明体绘制技术被广泛应用于医学可视化中。

目前, 结合 Blinn-Phong 面明暗化模型的光线吸

收与发射模型是应用最广泛的一种体绘制光学模型^[1]。尽管该模型能基本描述许多明显的半透明组织表面, 然而, 其难以表述半透明组织或者散射占优组织的详细信息^[2]。文献 [1]、[3] 综述了许多或简单或复杂精确的用于体绘制的光学模型。文献 [1] 与文献 [4] 证明了多重散射对于体绘制的效果显著, 其中, 文献 [1] 给出了计算多散射的方法。然而, 庞大的计算量与苛刻的内存要求使得精确的光学模型难以在实时的医学工程中得到应用。因此, 在实际的体绘制应用中, 一些相对容易实现的光学

基金项目: 国家自然科学基金项目 (30570458); 广东省自然科学基金项目 (05006530); 广州市番禺区科技攻关项目 (2005-Z-55-1)

收稿日期: 2006-08-03 改回日期: 2006-10-18

第一作者简介: 李彬 (1979~), 男, 讲师, 博士。主要研究方向或领域为医学图像处理与模式识别、科学可视化。E-mail: binli@scm.cn

经验模型被普遍采用。如文献 [5] 和文献 [6] 在基本的光线吸收与发射模型中加入直接散射部分, 并用非真实感绘制技术加大隐含分界面的显示效果。在文献 [1] 的基础上, 文献 [2]、[7]、[8] 提出了一个适用于硬件实现的体绘制的光学经验模型。而 Shear-Warp 算法^[9]是目前最快的基于 CPU 的体绘制方法^[10]。因此, 本文提出了一种能在普通硬件条件下实现的用于半透明体绘制的光学模型 IVROM (improved volume rendering optical model)。该模型在光线吸收与发射模型的基础上, 考虑了阴影、直接散射与间接散射等因素, 因此, 结合该模型与 Shear-Warp 方法的半透明体绘制技术, 能高效率地重建 3 维医学图像, 并能充分显示内部隐含分界面及内部细节的详细信息。

2 结合 Blinn-Phong 面明暗化模型的光线吸收与发射模型

标准的光线吸收与发射模型如下:

$$I(\mathbf{x}_b, \omega) = T(Q, l)I(\mathbf{x}_o, \omega) + \int_0^l (\tau(s, l))^* R(\mathbf{x}(s)) ds \quad (1)$$

式中, \mathbf{x} 为体素的空间 3 维坐标 (x, y, z), $\mathbf{x}(s)$ 指沿着视线距离 s 处的 3 维坐标。 $I(\mathbf{x}, \omega)$ 指沿着视线方向 ω 在空间坐标 \mathbf{x} 处的光线强度; $T(s, l)$ 指沿着视线从 $\mathbf{x}(s)$ 到 $\mathbf{x}(l)$ 过程中的衰减, $T(s, l) = \exp\left[-\int_s^l \tau(t) dt\right]$; τ 指光线衰减系数, $R(\mathbf{x})$ 为 \mathbf{x} 处面反射颜色值。

通常, 在方向为 ω , 空间 \mathbf{x} 处的散射光强度

$$S(\mathbf{x}, \omega) = r(\mathbf{x}, \omega, \omega')i(\mathbf{x}, \omega') \quad (2)$$

式中, $i(\mathbf{x}, \omega')$ 为沿 ω' 方向入射, 到达空间 \mathbf{x} 处的光强度。 $r(\mathbf{x}, \omega, \omega')$ 为双向反射分布函数 (bidirectional reflection distribution function, BRDF)。式 (2) 为通用的明暗化公式。

为了实现面明暗化, 引入 Blinn-Phong 模型, 其本质是表达 $r(\mathbf{x}, \omega, \omega')$ 的直接散射项。若假设沿 ω' 方向入射光没有任何阻拦, 则 $i(\mathbf{x}, \omega)$ 依然为光源亮度 L_l 。因此, 式 (1) 结合 Blinn-Phong 面明暗化模型, 得

$$I(\mathbf{x}_b, \omega) = T(Q, l)I(\mathbf{x}_o, \omega) + \int_0^l (\tau(s, l))^* R(\mathbf{x}(s))^* B_s(\mathbf{x}(s))^* L_g ds \quad (3)$$

式中, $B_s(\mathbf{x}(s))$ 指用 $\mathbf{x}(s)$ 处归一化后的梯度估算的 Blinn-Phong 面明暗化模型值; L_g 为光源亮度, 下角 g 表示沿 ω' 的光线方向。

式 (3) 便是传统的结合 Blinn-Phong 面明暗化模型的光线吸收与发射模型。

3 IVROM 光学模型

尽管结合 Blinn-Phong 面明暗化模型的光线吸收与发射模型, 能基本描述许多明显的半透明组织表面, 然而, 其难以表述半透明组织或者散射项占优组织的详细信息, 无法满足医学可视化中半透明体绘制的需要。因此, 鉴于结合 Blinn-Phong 面明暗化模型的光线吸收与发射模型存在的不足, 本文从医学工程实践出发, 提出可以在微机上实时实现的 IVROM 光学模型。

结合 Blinn-Phong 面明暗化模型的光线吸收与发射模型, 完全以空间体素的梯度作为表面明暗化的依据, 难以刻画梯度模较低但却属于不同类组织的区域。入射光在不同组织的衰减不同, 因此, 利用阴影的作用则可以适当地弥补结合 Blinn-Phong 面明暗化模型的光线吸收与发射模型所存在的不足。

实质上, 入射光在组织内, 确实存在着衰减。设亮度为 L_l 的光源沿 ω' 方向入射, 到达空间 \mathbf{x} 处的光亮度为 $i(\mathbf{x}, \omega')$, 则

$$i(\mathbf{x}, \omega') = L_g T_g(s, t') \quad (4)$$

$$T_g(s, t') = \exp\left[-\int_0^s \tau(\mathbf{x}(s) + \omega_l t) dt\right] \quad (5)$$

因此, 重写式 (3) 为

$$I(\mathbf{x}_b, \omega) = T(Q, l)I(\mathbf{x}_o, \omega) + \int_0^l (\tau(s, l))^* R(\mathbf{x}(s))^* B_s(\mathbf{x}(s))^* L_g^* T_g(s, l_g) ds \quad (6)$$

式中, l_g 表示沿着光线方向的距离。

在一随机分布介质空间 V , 当光线穿过空间 V 时, 由于光线与介质粒子的相互作用, 在光的传播方向上, 会发生吸收和散射的消光现象, 因此, 在 \mathbf{x} 处的 $I(\mathbf{x}, \omega)$ 应该由准直辐射亮度和散射辐射亮度组成。而散射辐射亮度由直接散射项与间接散射项两部分组成。设 \mathbf{x} 处的一个粒子的双向反射分布函数, 则

$$r(\mathbf{x}, \omega, \omega') = a(\mathbf{x})\tau(\mathbf{x})p(\omega, \omega') \quad (7)$$

式中, $a(\mathbf{x})$ 指该粒子的反射率, $p(\omega, \omega')$ 为相函数。

表示散射的方向性。

本质上,式(3)考虑了直接散射的影响。若令 $p_D(\omega, \omega')$ 表示直接散射的相函数,则

$$p_D(\omega, \omega') = |N \cdot \omega'| + \left| N \cdot \frac{\omega + \omega'}{|\omega + \omega'|} \right|^n \quad (8)$$

式中, N 为 r 点的梯度向量,也就是 r 点隐含分界面的法向量。考虑到隐含分量面的正反面方向,对 $|N \cdot \omega'|$ 取了绝对值。

若考虑梯度模值高的区域,则用类似 Blinn-Phong 面明暗化模型的方法,若在梯度模值较低的区域,则主要用阴影实现明暗化计算。因此,式(6)可以写为

$$I(x_b, \omega) = T(Q, l)I(x_o, \omega) + \int_0^l (s, l)^* S(s)^* L_g^* T_g(s, l_g) ds \quad (9)$$

$$S(s) = R(s) ((1-w(s)) + w(s)p_s(\omega, \omega')) \quad (10)$$

这里,可以考虑用非真实感绘制技术加大隐含分界面的显示效果。例如,给直接散射和间接散射一个权重。设直接散射 $S_D(s) = w_D S(s)$ (w_D 为权重)。直观地讲,愈接近隐含分界面,面散射亮度愈大,反之,则愈小。因此,这里的权函数本质上是一个边界检测函数,可以令该体素的 w_D 与梯度模值成正比。式(9)变为

$$I(x_b, \omega) = T(Q, l)I(x_o, \omega) + \int_0^l (s, l)^* S_D(s)^* L_g^* \exp\left[-\int_s^l \tau(t) dt\right] ds \quad (11)$$

接下来考虑间接散射的影响,令间接散射

$$S_I(s) = w_I R(s) p_I(\omega, \omega') \quad (12)$$

式中, w_I 为权值。由文献 [11] 可以得到

$$p_I(\omega, \omega') \approx \frac{1}{\sigma(x)} \sum_{p=1}^n f(x, \omega'_p, \omega) \frac{\Delta\Phi_{n,i}(x, \omega'_p)}{\frac{4}{3}\pi r^3} \quad (13)$$

式中, $\Delta\Phi_{n,i}$ 为通过间接散射影响该体素的相关的 n 个光子的亮度和, $\frac{4}{3}\pi r^3$ 为包含这 n 个的光子的最小球形体积。

则式(11)为

$$I(x_b, \omega) = T(Q, l)I(x_o, \omega) + \int_0^l (s, l)^* \left[S_D(s)^* L_g^* \exp\left[-\int_s^l \tau(t) dt\right] + S_I(s)^* L_g^* \exp\left[-\int_s^l \tau_i(x) dx\right] \right] ds \quad (14)$$

式中,类似于 τ 的定义, τ_i 指光线间接衰减系数,该间接衰减系数与物质有关。 $T_i(s, l)$ 指沿着光线从 $x(s)$ 到 $x(l)$ 过程中的间接衰减, $T_i(s, l) = \exp\left[-\int_s^l \tau_i(t) dt\right]$ 。

4 算法实现

(1) 预设置传递函数,将其设为查询表。

$T(s, l)$ 指沿着视线从 $x(s)$ 到 $x(l)$ 过程中的衰减,定义 α 为这段距离的不透明度,则 $\alpha = 1 - T(s, l)$ 。同理,定义 α_i 为这段距离的间接不透明度,则 $\alpha_i = 1 - T_i(s, l)$ 。 $\exp\left[-\int_s^l \tau_i(x) dx\right]$ 可以近似表示为

$$\exp\left[-\sum_{i=1}^n \tau(i \cdot \Delta x) \cdot \Delta x\right] = \prod_{i=1}^n \exp(-\tau(i \cdot \Delta x) \cdot \Delta x) = \prod_{i=1}^n t_i \quad (15)$$

式中, $t_i = \exp(-\tau(i \cdot \Delta x) \cdot \Delta x)$ 可以看作是沿光线方向上第 i 个区段的间接透明度。

为快速实现光学模型算法,先预设置 $\Delta x = 1$ 的间接不透明度值 α_i 的查询表。同理,预设置 $\Delta x = 1$ 的不透明度值 α 的查询表。

(2) 预设置直接散射的权重值

直观地讲,愈接近隐含分界面,直接散射亮度愈大,反之则愈小。因此,这里的权函数本质上是一个边界检测函数,因此,直接散射的权重 w_D 与梯度模值成正比。另外,若考虑到式(12)的影响,则可设置间接散射的权重 w_I 与体素值成正比。

(3) 合成中间图像

ShearWarp 体绘制主要由 3 维数据场的错切变换 (Shear) 与 2 维图像的变形 (Warp) 两个步骤实现。本光学模型将在错切变换时合成中间图像的过程中应用。3 维空间离散数据场的观察方向常常是由用户任意给定的,因而使得由物体空间到图像空间的变换也是任意的。ShearWarp 的中心思想是将 3 维离散数据场变换到一个中间坐标系,在这个中间坐标系中,观察方向与坐标系的一个轴并行。光源的方向是任意的,为叙述和执行方便,不妨将光源设于与视点处于同一侧的位置(是否在同一侧以夹角是否大于 90° 为依据)。若光源方向在另外一侧,只需要更改光线合成次序则可。

本文半透明体绘制算法将结合传统的 ShearWarp 实现 NROM 光学模型。除了原来需要的缓存

外, 还需要新开辟 3 个缓存: 当前切片的直接光照缓存、间接光照缓存以及下一个切片的间接光照缓存, 缓存大小均为中间图像大小。与 Shear-Warp 算法类似, 本半透明体绘制方法也是按由前向后序次, 循环对每个切片进行处理, 合成中间图像。在处理第 k 个切片时, 步骤如下:

(1) 计算直接散射部分

$$C_{post}^D = C_{pre}^D + \alpha_{now}^D * C_{now}^D * (1 - O_{pre}^D) * I_{pre}^D \quad (16)$$

式中, C_{pre}^D 为计算直接散射部分之前的中间图像, 在当前某一位置的 RGB 中的 R 分量的值; C_{post}^D 为计算直接散射部分之后的中间图像, 在当前同一位置的 RGB 中的 R 分量的值; O_{pre}^D 为计算直接散射部分之前的中间图像当前位置的不透明度值; I_{pre}^D 为计算直接散射部分之前的中间图像当前位置的直接光照累积光亮度; α_{now}^D 为当前体素重采样不透明度; C_{now}^D 为当前体素重采样的 R 分量值。这些变量的上标 D 表示计算直接散射部分。

$$O_{post}^D = O_{pre}^D + \alpha_{now}^D * (1 - O_{pre}^D) \quad (17)$$

式中, O_{post}^D 为计算直接散射部分之后的中间图像当前位置 pos 的不透明度值。

由于光的合成顺序由后至前, 有

$$I_{post}^D = (1 - \alpha_{now}^D) * I_{pre}^D \quad (18)$$

式中, I_{post}^D 为计算直接散射部分之后的中间图像当前位置的直接光照累积光亮度。

中间图像的 RGB 中的 G、B 分量的直接散射部分的值可以用同样的方法求得。

(2) 计算光照累积光亮度

由于光的合成顺序由后至前, 有

$$I_{post}^I = I_{pre}^I * (1 - \alpha_{now}^I) \quad (19)$$

式中, I_{pre}^I 为计算间接散射部分之前的中间图像当前位置的间接光照累积光亮度; I_{post}^I 为计算直接散射部分后的中间图像当前位置的间接光照累积光亮度; α_{now}^I 为当前体素重采样的间接不透明度值。这些变量的上标 I 表示计算间接散射部分。

(3) 计算间接散射部分

$$C_{post}^I = C_{pre}^I + \alpha_{now}^I * C_{now}^I * (1 - O_{pre}^D) * I_{post}^I * I_a^I \quad (20)$$

式中, C_{pre}^I 为添加间接散射部分前的中间图像当前位置的 RGB 中的 R 分量的值; C_{post}^I 为添加间接散射部分后的中间图像当前位置的 RGB 中的 R 分量的值; C_{now}^I 为当前体素在计算间接散射部分时重采样的 R 分量值。

考虑到 Shear-Warp 方法由前至后按次序对每

一个切片处理的特点, 式 (20) 中, I_a^I 取 6 个像素点 (下一个切片的间接光照缓存在 pos 位置的像素点、当前切片的间接光照缓存在 pos 位置的像素点以及其相邻 4 个像素点) 的平均值。这里, I_a^I 实质上是

$$\text{式 (13) 中 } \sum_{p=1}^n \frac{\Delta\Phi_{p,i}(x, \omega'_p)}{\frac{4}{3}\pi r^3} \text{ 的近似值。用最小的圆}$$

锥代替最小的球进行近似地处理间接散射部分的影响的方法与证明, 可以参考文献 [2]、[6]、[11]。

中间图像的 RGB 中的 G、B 分量的间接散射部分的值同理求得。

(4) 将下一个切片的间接光照缓存数据存放于当前切片的间接光照缓存, 并将下一个切片的间接光照缓存都设置为值 1。

(5) 处理下一个切片。

5 实验结果

采用 C++ 编程语言进行软件编写, 实验平台的机器配置为 Pentium4 3G, 512RAM, 显卡为 Geforce 6800 GT。

为了证明光学模型 IVROM 用于半透明体绘制 3 维医学图像的可行性和准确性, 用相同的观察角度 ($\alpha = -90^\circ$, $\beta = 0^\circ$), 即视点的视线绕 X 轴旋转 -90° , 绕 Y 轴旋转 0° , 相同的传递函数, 结合 Shear-Warp 体绘制算法, 分别用光学模型 IVROM 与标准的结合 Blinn-Phong 面明暗化模型的光线吸收与发射模型对医学序列 CT 图像 ($512 \times 512 \times 377$ (12 位)) 进行重建。重建结果如图 1~ 图 6 所示。

图 1 为用光学模型 IVROM 进行半透明体绘制的头部结果。图 2 为只用光学模型 IVROM 的直接



图 1 光学模型 IVROM I

Fig 1 IVROM I

散射部分进行半透明体绘制的结果。图 3 为只用光学模型 IVROM 的间接散射部分进行半透明体绘制的结果。图 4 为用标准的结合 Blinn-Phong 面明暗化模型的光线吸收与发射模型进行半透明体绘制的



图 2 IVROM 的直接散射部分
Fig 2 Direct scattering of IVROM



图 3 IVROM 的间接散射部分
Fig 3 Indirect scattering of IVROM



图 4 结合 Blinn-Phong 面明暗化模型的光线吸收与发射模型 I
Fig 4 Absorption plus emission model with Blinn-Phong shading I

结果。比较结果图 1 与图 4 显而易见, 图 1 可以突出地显示出隐含分界面, 更富有层次, 并且内部细节更清晰、细致, 对医学诊断也更有意义和价值。

图 5、图 6 也是分别用光学模型 IVROM 与标准的结合 Blinn-Phong 面明暗化模型的光线吸收与发射模型, 结合 ShearWarp 进行半透明体绘制的脑颅结果。注意到脑软组织部分, 由于其不同类物质 (如脑白质与脑灰质) 间的梯度差很低, 甚至为零。因此, 用标准的结合 Blinn-Phong 面明暗化模型的光线吸收与发射模型的半透明体绘制结果 (图 6) 很难表达该部分的细节。相比之下, 用光学模型 IVROM 的半透明体绘制结果 (图 5) 表达的内部细节就更清晰、细致。

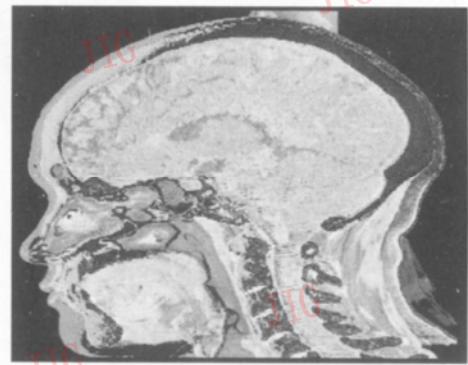


图 5 IVROM II
Fig 5 IVROM II

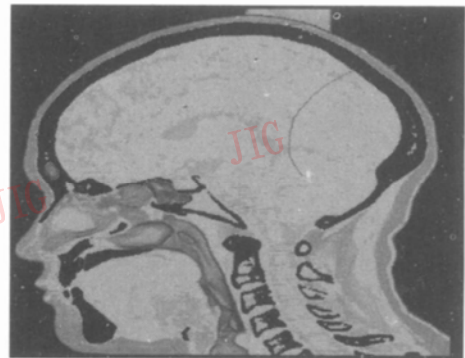


图 6 结合 Blinn-Phong 面明暗化模型的光线吸收与发射模型 II
Fig 6 Absorption plus emission model with Blinn-Phong shading II

实验证明, 采用光学模型 IVROM 的半透明体绘制对医学序列图像的重建是可行性, 较之标准的结合 Blinn-Phong 面明暗化模型的光线吸收与发射模型的半透明体绘制更准确, 对医学诊断也更有意义。

用不同算法对重建效率进行测试。实验中, 只旋转角度 β 每次增加 30° , 共 12 (360/30) 次, 求出平均时间。结果如表 1 所示。

表 1 重建效率比较

Tab 1 Comparison of reconstruction efficiency

所用模型	单位: s	
	预处理耗时	运行时间
结合 Blinn-Phong面明暗化模型的光线吸收与发射模型	16.974	1.347
光学模型 (VROM)	17.267	2.859

从表 1 可以看出,采用光学模型 VROM 的半透明体绘制的运行时间,比采用标准的结合 Blinn-Phong面明暗化模型的半透明体绘制的运行时间稍长。与采用标准的结合 Blinn-Phong面明暗化模型的半透明体绘制算法相比,采用光学模型 VROM 的半透明体绘制在合成中间图像时,增加了对非空体素进行直接与间接光照累积光亮的运算,还增加了对间接散射部分的运算(包括间接散射部分颜色以及不透明度的重采样及合成)。

此外,在传递函数预设过程中,采用光学模型 VROM 的半透明体绘制的运行时间也比采用标准的结合 Blinn-Phong面明暗化模型的半透明体绘制的运行时间略长。那是因为采用光学模型 VROM 的半透明体绘制算法增加了间接不透明度、直接散射与间接散射的权重的制表过程。但是,通过实验仍然可以得出结论,在普通硬件条件下,采用光学模型 VROM 的半透明体绘制是可以实现实时重建绘制。

6 结 论

从实际的医学工程应用出发,提出了一种能在普通硬件条件下实现的用于半透明体绘制的光学模型 VROM。该模型在光线吸收与发射模型的基础上,考虑了阴影、面散射与体散射等因素。结合该模型与 ShearWarp方法的半透明体绘制技术,能高效率地重建 3 维医学图像,并能充分显示内部隐含分界面及内部细节的详细信息,从而能满足对医学运用精确诊断的要求。因此,该算法可以应用于医学工程实践(如放疗计划的模拟手术)上。

参考文献 (References)

- 1 Nelson M ax. Optical models for direct volume rendering[J]. IEEE Transactions on Visualization and Computer Graphics 1995, 1(2): 99~ 108
- 2 Kniss Joe, Premoze Simon, Hansen Charles, et al. A model for volume lighting and modeling[J]. IEEE Transactions on Visualization and Computer Graphics 2003, 9(2): 150~ 162
- 3 Moreland Kenneth Dean. Fast High Accuracy Volume Rendering [D]. New Mexico, USA: University of New Mexico, 2004: 10~ 28
- 4 Jensen Henrik Wann, R. Marschner Stephen, Levoy Marc, et al. A practical model for subsurface light transport[A]. In: Proceedings of ACM SIGGRAPH 2001[C], USA, CA, Los Angeles, 2001: 511~ 518
- 5 Cai Wen-li, Shi Jiao-ying. Composed volume rendering model based on transfer equation [J]. Chinese Journal of Computers, 1995, 18(5): 330~ 338. [蔡文立, 石教英. 基于输运方程的混合式体绘制模型 [J]. 计算机学报, 1995, 18(5): 330~ 338]
- 6 Peng Yan-jun, Shi Jiao-ying. A method of display hide interfaces in volume rendering and its realization[J]. Journal of Software, 2002, 13(9): 1887~ 1892. [彭延军, 石教英. 体绘制中显示隐含分界面的一种方法及其实现 [J]. 软件学报, 2002, 13(9): 1887~ 1892]
- 7 Kniss Joe, Premoze Simon, Hansen Charles, et al. Interactive translucent volume rendering and procedural modeling [A]. In: Proceedings of IEEE Visualization 2002[C], USA, MA, Boston, 2002: 109~ 116
- 8 Kniss Joe, Kindhann Gordon, Hansen Charles. Multidimensional transfer functions for interactive volume rendering [J]. IEEE Transactions on Visualization and Computer Graphics, 2002, 8(3): 270~ 285
- 9 Lacroute P. Fast Volume Rendering Using a ShearWarp Factorization of the Viewing Transformation [D]. Stanford, USA: Stanford University, 1995: 29~ 43
- 10 Grimm Soren, Bruckner Stefan, Kanitsar Amin, et al. Memory efficient acceleration structures and techniques for CPU-based volume raycasting of large data [A]. In: IEEE Symposium on Volume Visualization and Graphics 2004[C], USA, Texas, Austin, 2004: 1~ 8
- 11 Jensen Henrik Wann, Per H. Christensen. Efficient simulation of light transport in scenes with participating media using photon maps [A]. In: Proceedings of SIGGRAPH 1998[C], New York: ACM Press, 1998: 311~ 320

“智能视觉监控技术及系统”专栏前言

当前,嵌入式计算、无线通信和传感器技术的迅速发展和结合,已使人们能够非常方便地采集、传输和存储视频和音频数据。海量数据及时和准确的分析与理解,将在安全、反恐、国防、智能交通、医疗监护、遥感等领域发挥重要作用。例如,在复杂的环境下检测和识别人体活动,进而对其行为和意图做出分析,以便对可能发生的异常、危险事件进行预测或判断并对可能的安全威胁进行分类、评价和预测等。因此智能视觉监控技术和系统即是计算机视觉领域极具挑战性的科学问题,又是关系国家安全、经济发展和稳定的关键问题。

基础理论研究和自主关键技术的创新,是智能视觉监控技术和系统发展的基础。“智能视觉监控技术及系统”专栏,旨在及时报道这一领域的最新进展,总结国内在智能视觉监控领域的最新热点问题和现状,展现重要科研和开发成果,开拓新的研究、发展和应用方向。

承蒙国内广大同行们的厚爱,专栏征文共收到 65 篇论文,涉及到智能视觉监控的方方面面。我们很高兴地看到,不少论文或在学术上表现出了很强的创新性,或在技术上有很强的实用性。同时,例如,来稿中对传统的基于单个摄像机的研究较多,而最新的基于分布式摄像机或传感器网络的研究很少,这反映出国内在视觉监控领域的研究总体上是规模小,方向重复、分散,我们的研究还急需进一步提高。

由于篇幅的限制,我们不得不割舍了不少高质量的论文。在此过程中难免有考虑欠周之处,敬请各位作者谅解。本专栏最终共录用了 17 篇论文,将分别在 2008 年第 2、3、4 期中刊出。我们衷心希望,专栏的推出,能促进国内同行间的交流,并为从事本领域研究和开发的学生和研究人员提供有益的参考。不妥之处,欢迎广大读者批评指正。

陶霖密 徐光祐

视觉监视中基于柯西分布的统计变化检测

明 英^{1), 3)} 蒋晶珏²⁾

¹⁾ (清华大学电子工程系, 北京 100084) ²⁾ (武汉大学计算机学院, 武汉 430079)

³⁾ (武警武汉指挥学院, 武汉 430064)

摘 要 为了更好地进行视觉监视, 该文给出了一种新的基于柯西分布的光照不变的统计变化检测算法。该算法首先将两帧图像间的灰度比值作为背景建模和剔除的特征, 并且在假定背景图像中, 当每个像素点观测的时序灰度变化由白噪声引起时, 两帧背景图像中对应像素间的灰度比值的分布符合柯西分布; 然后基于该变化检测方法, 将 YCbCr 颜色空间的亮度、色调和饱和度用来识别和消除视频序列图像中的阴影。实验结果表明, 该新算法不仅可以承受整体或局部的缓慢或突然的光线变化, 并且可以滤除由场景背景中小的扰动而导致的噪声。

关键词 变化检测 运动目标检测 背景建模 阴影 柯西分布 监视

中图法分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(2008)02-0328-07

基金项目: 国家“863”高技术研究发展计划导向类项目 (SQ2006AA12Z108506)

收稿日期: 2007-04-21; 改回日期: 2007-12-06

第一作者简介: 明英 (1970~), 男, 副教授, 博士。目前主要从事智能视觉监视和 3 维重建等研究。E-mail: whum_kk@foxmail.com

Cauchy Distribution Based on Statistical Change Detection for Visual Surveillance

MING Ying^{1, 3)}, JIANG Jing-ju²⁾

¹⁾ (Department of Electronic Engineering, Tsinghua University, Beijing 100084)

²⁾ (School of Computer Science, Wuhan University, Wuhan 430079)

³⁾ (Education department, Wuhan Command School of Armed Police Force, Wuhan 430064)

Abstract A novel illumination-invariant change detection algorithm based on Cauchy distribution is proposed. The intensity ratios between two images are used as the feature to model and subtract background. The distribution of the intensity ratios between corresponding pixels of two background images follows Cauchy distribution, assuming that some observed temporal intensity variation of each pixel in background images are caused by white noise. The intensity, hue and saturation in the YCbCr color space are employed to recognize and eliminate shadows in video sequences. Finally, experimental results demonstrate that the proposed algorithm can tolerate the whole or local quick or slow changes in illuminations, and can filter noise caused by small motion in scene background.

Keywords change detection, moving object detection, background modeling, shadows, Cauchy distribution, surveillance

1 引言

在智能视觉监控系统中, 所面临的最基础性问题就是如何检测图像序列中的前景运动目标。目前, 大多数用于动态场景分析的运动目标检测技术都是基于图像序列的变化检测。鉴于其在动态场景预先分析中的重要性, 如今人们已对它进行了许多研究^[1-6]。

文献[6]中的变化检测是基于一个小判断区域中的差分图像的噪声估计的置信检验进行的, 即在两帧图像的对应区域, 通过定义的差分阈值判断是否有变化发生。这种通过比较区域灰度分布来实现变化检测的方法对光照的变化敏感。发光模型技术 (shading model SM)^[11] 是一种主要针对时变光照情况的变化检测方法。SM 方法是利用两帧图像对应区域的灰度比值来检测发光或反光模型的变化^[7]。文献[1]中的实验结果表明, SM 方法虽可以精确定位源于场景中物体的运动、沉淀和移出等结构性变化, 然而, 由于目前这些基于 SM 的方法都是直接采用灰度比值或灰度和的比值来描述一个区域的结构特征, 因而它们可能在较暗的区域或缺乏丰富纹理的区域其检测性能变差。另外, SM 方法由于要进行大量的除法运算而费时。为了改进这些不足, 许多研究者对 SM 方法进行了改进。Duncan 和 Ebrahimi 先为参考图

像和当前图像的每个像素分配一个矢量, 然后通过检测像素周围矢量的非线性相关来发现变化像素^[8], 但对不同的场景需要取不同的阈值; Li 等人在文献[9]中提出一种集成光谱、空间和时间特征的 Bayesian 框架用来检测前景目标; Liu 和 Sun 等人提出一种新的光照不变的变化检测方法^[10], 他们将传统的发光模型方法与假设检验结合, 通过将 SM 嵌入到置信度检验模型中来降低对光照变化的敏感性, 以改进在较光滑区域的性能。

为了检测变化, 通常需要对背景场景建模。建模的典型方法是利用 Gaussian 模型 (GM), 即用单个 Gaussian 分布^[11] 或用混合 Gaussian 分布 (MoG)^[13] 对每个像素建模; 另一种办法是利用非参数估计^[5] 来获得背景模型的概率分布。这些方法虽可以成功地处理复杂背景, 但无法适应快速的光照变化和很好地处理阴影。同时, 一些非 Gaussian 的背景模型方法也被采用。在文献[12]中, 整个背景图像被用一个单宽度参数的 Laplacian 分布建模。在 SM 方法和 GM 算法的启发下, 笔者发现并证明了在场景背景图像中, 两幅差分背景图像的每个对应像素点之间的差分灰度值的比值分布可以很好地用柯西 (Cauchy) 分布建模^[13]。

为此, 本文提出了一种新的基于 Cauchy 分布的光照不变变化检测算法。通过先前工作发现, 每个背景图像像素的强度比值分布可以直接用一种 Cauchy 分布建模^[13]。YCbCr 颜色空间的亮度和色

度可以被用来识别视频图像序列中较暗的阴影种子点。

2 背景建模

如果观察一个完全静止场景中的任意一个像素强度值,假定图像噪声可以用一个标准正态分布 $N(0, 1)$ 表示,则该背景像素的强度随时间的变化可以用一个正态分布 $N(\mu, \sigma^2)$ 建模^[5,6]。这正是许多背景剔除技术中所用的正态分布模型。

若 $I = \{I_1(\mathbf{x}), I_2(\mathbf{x}), \dots, I_k(\mathbf{x})\}$ 是背景像素点 \mathbf{x} 从时刻 1到 k 的图像强度的样本集,则该点的图像强度值 $I(\mathbf{x})$ 的分布符合正态分布,即 $I(\mathbf{x}) \sim N(\mu(\mathbf{x}), \sigma^2(\mathbf{x}))$ 。若 $I_i(\mathbf{x})$ 和 $I_j(\mathbf{x})$ 分别是背景像素点 \mathbf{x} 在任意两个时刻 i 和 j 的强度值,且二者的比值为 $r_{i,j}(\mathbf{x}) = I_i(\mathbf{x}) / I_j(\mathbf{x})$,则可以证明得到同一像素点在任意两个时刻的像素强度比值 $r_{i,j}(\mathbf{x})$ 的分布符合一个位置参数为 $\mu(\mathbf{x})$ 、尺度参数为 $\lambda(\mathbf{x})$ 的 Cauchy 分布,即 $r_{i,j}(\mathbf{x}) \sim C(\mu(\mathbf{x}), \lambda(\mathbf{x}))$,

$\lambda(\mathbf{x})$, 其概率密度函数 (probability density function, PDF)为

$$P(r_{i,j}(\mathbf{x})) = \frac{1}{\pi} \cdot \frac{\lambda(\mathbf{x})}{\lambda^2(\mathbf{x}) + (r_{i,j}(\mathbf{x}) - \mu(\mathbf{x}))^2} \quad (1)$$

如图 1所示,在“Lobby”图像序列前 300帧图像序列中一个背景像素的每相邻两帧的彩色强度比值的直方图的例子。图 1(a)显示了一帧 Lobby 的图像和所选择的像素位置(白色+)。图 1(b)~图 1(d)分别是所选像素位置处 Y, Cb和 Cr的直方图。图中红色是 Cauchy分布的 PDF,绿色是 Gaussian分布。对于背景像素强度比值的时域变化, Cauchy分布(红色)的建模效果比 Gaussian分布(绿色)好。图像中每个背景像素点 \mathbf{x} 都由一个 Cauchy分布 $C(\mu(\mathbf{x}), \lambda(\mathbf{x}))$ 建模。对于彩色图像,每个颜色通道具有相应的参数 $\mu(\mathbf{x})$ 和 $\lambda(\mathbf{x})$ 。每一像素点的模型参数可以离线估计,或通过在 线迭代确定。本文变化检测是使用 YCbCr颜色空间。

图 2显示的是“Lobby”图像序列中,所有背景

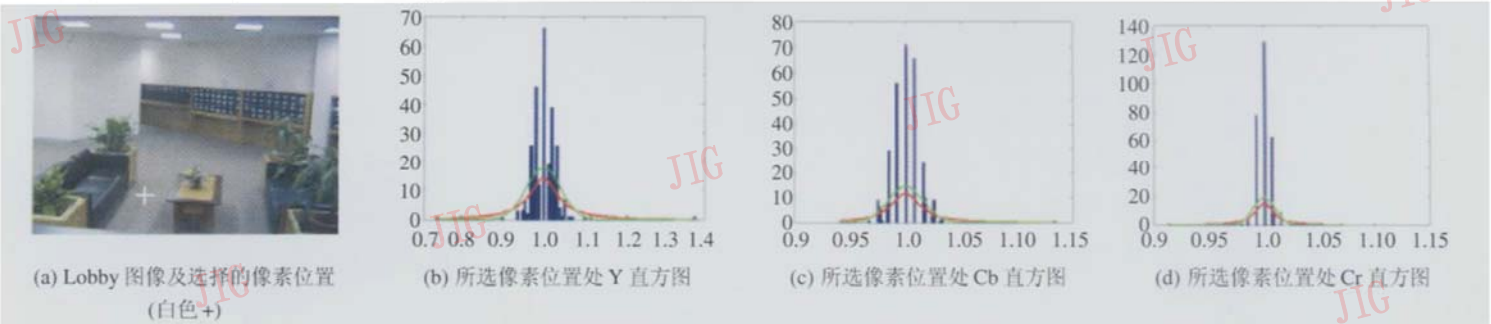


图 1 一个背景像素的每相邻两帧的 YCbCr 彩色强度比值的直方图及拟合的 Gaussian 和 Cauchy 分布

Fig.1 One example of the histogram of one background pixel's YCbCr color intensity ratio between each two consecutive frames and their fitted Gaussian or Cauchy distribution

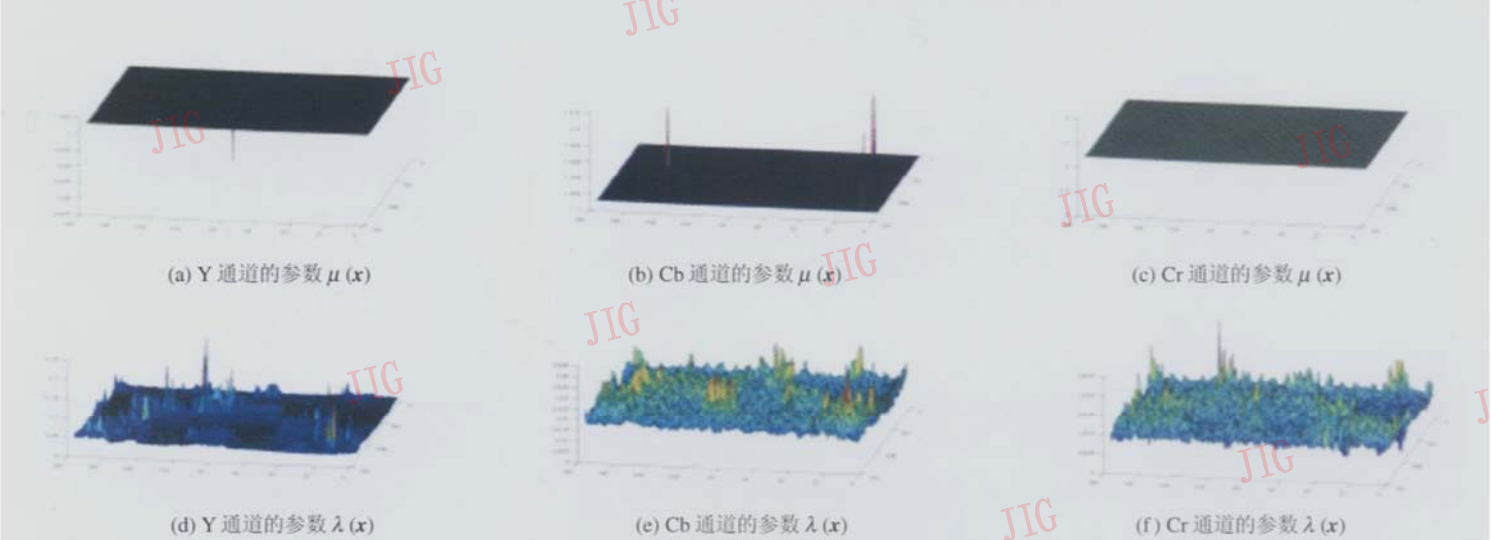


图 2 在 Lobby 图像序列中各个颜色通道中背景像素点强度比值的 YCbCr 各分量 Cauchy 分布的参数

Fig.2 The parameter $\mu(x)$ and $\lambda(x)$ of color intensity ratio values in each color channel for all pixels in the “Lobby” image sequence

像素点在不同颜色通道中颜色强度比值 Cauchy 分布参数 $\mu(\mathbf{x})$ 和 $\lambda(\mathbf{x})$ 。

3 变化检测

若 $\{r(\mathbf{x}_1), r(\mathbf{x}_2), \dots, r(\mathbf{x}_N)\}$ 是像素点 \mathbf{x}_i 周围 N 个像素强度比值。假定在像素点 \mathbf{x}_i 的邻域内没有结构性变化, 则它们的值与像素点 \mathbf{x}_i 线性相关, 即相等或近似相等^[1]。因此, 为了减小噪声影响, 可利用点 \mathbf{x}_i 及其邻域的 PDF 的均值作为像素点 \mathbf{x}_i Cauchy 分布 PDF 的观测值

$$P(\mathbf{x}_i) = \frac{1}{N} \cdot \sum_{i=1}^N K(\mathbf{x}_i) \quad (2)$$

其中, $K(\mathbf{x}_i)$ 用式 (1) 计算。

对于彩色图像, 为了简化计算, 可假定不同颜色通道间独立, 且分别具有不同的 Cauchy 模型参数, 则其 PDF 为

$$P(\mathbf{x}_i) = \prod_{j=1}^d \left| \frac{1}{N} \cdot \sum_{i=1}^N K(\mathbf{x}_i^d) \right| \quad (3)$$

式中, d 代表 Y、Cb 和 Cr。

根据每个像素点的概率估计, 若 $P(\mathbf{x}_i) < T$, 则该像素为变化像素。这里 T 是针对整个图像的全局阈值, 它可以根据期望的误警率进行计算调整^[14]。实际上, 对于大多数来源于背景图像像素的概率较精确的估计可以根据临近时间内和相邻像素的信息获得。为了对不同背景变化进行鲁棒的变化检测, 同时计算每个像素点的相邻帧间和当前帧与背景图像间的强度比值 Cauchy 分布的概率 $P_1(\mathbf{x}_i)$ 和 $P_2(\mathbf{x}_i)$ 。若 $P_1(\mathbf{x}_i) < T$ 且 $P_2(\mathbf{x}_i) < T_b$, 则认为像素 \mathbf{x}_i 为变化像素。

4 阴影的识别和消除

单摄像机运动目标检测与跟踪中的一个主要问题就是识别运动目标产生的阴影。若将这些阴影误检测为前景区域则会使后续的分析产生无法消除的错误。在最近的许多文献中已提出了一些阴影检测的方法。其中大多数算法都基于假设阴影投射在物体表面时, 仅降低该区域图像的灰度, 而不改变其色度和纹理^[15]。由于由淡的阴影造成的轻微光照变化而产生的图像色度变化很小, 也就是, 淡的阴影图像像素与背景图像像素的色度值比值与阴影覆盖前相比, 变化很小, 因此, 基于

SM 的方法比其他方法对阴影的影响较为鲁棒。然而, 基于 SM 的方法在使用较高阈值去处阴影区域时, 运动目标区域常会产生破碎。为了解决这一问题, 本文采用了一种两级阈值的阴影检测方法, 即首先使用低阈值检测出所有变化区域 R_c ; 然后, 用高阈值检测出可靠的运动目标核心区域 R_k ; 最后, 利用本文提供的基于 YCbCr 空间的颜色分析方法识别和去除候选区域的阴影。

若任意背景像素 \mathbf{x}_i 无阴影时的正常值为 $I(Y, Cb, Cr)$, 在 t 帧被阴影或反光 (高光) 覆盖时的观测值为 $I(Y_t, Cb_t, Cr_t)$ 。笔者发现, 有阴影时的观测值 Y_t 比正常值 Y 小一个值域范围; 出现反光时, 观测值 Y_t 比背景正常值 Y 大一个值域范围。

当确定了所有变化区域 R_c 和运动目标核心区域 R_k 后, 则候选阴影区域 $S_c = R_c - R_k$ 。在 S_c 中, 阴影像素点可根据以下规则确定:

- (1) $\arg \min(Y_t/Y) < Y_t/Y < 1 \mid |Y_t/Y - \mu_Y| < \tau \cdot \delta_Y$;
- (2) $k_1 < |Cb_t/Cb - \mu_{Cb}| < k_2 \cdot \delta_{Cb}$;
- (3) $r_1 < |Cr_t/Cr - \mu_{Cr}| < r_2 \cdot \delta_{Cr}$ 。

这里, μ_i 和 δ_i (i 代表 Y、Cb 和 Cr), 是背景区域像素强度比值的均值和方差。 τ , k_1 , k_2 和 r_1 , r_2 为阈值。检测时, 先以确定的阴影像素点为阴影种子点, 采用基于 K 个最近邻域^[16] 的聚类分析方法来确定 S_c 中剩余的阴影点, K 可以取 3 或 5; 而对于反光像素点的确定则采用类似的方法进行; 然后去除阴影和反光像素后, S_c 变成 \hat{S}_c ; 最后, 运动目标区域为

$$\hat{R}_c = R_k \cup \hat{S}_c$$

5 试验结果

为了检验本文算法的效果, 采用 4 组不同场景的测试数据, 如 PETS、“Hall monitor”、“Corridor”^[8] 和 “Lobby”^[9] 图像序列进行了大量的试验, 并与 Li 等人^[8] 和 MoG 算法的处理结果进行了比较。在这些测试数据中, 阴影、反光、光照变化和低分辨率造成前景检测困难。试验时用 Matlab7 实现本文算法, 在 CPU 为 Intel Pentium M 1.6G 的计算机上运行。

图 3 是一室内场景的检测结果。在文献 [9] 中, 分别给出了 Li 等人提出的方法和 MoG 方法的检测结果。在这一场景中前景检测的困难主要由灯光引起的淡阴影。试验结果表明, 本文算法比其他

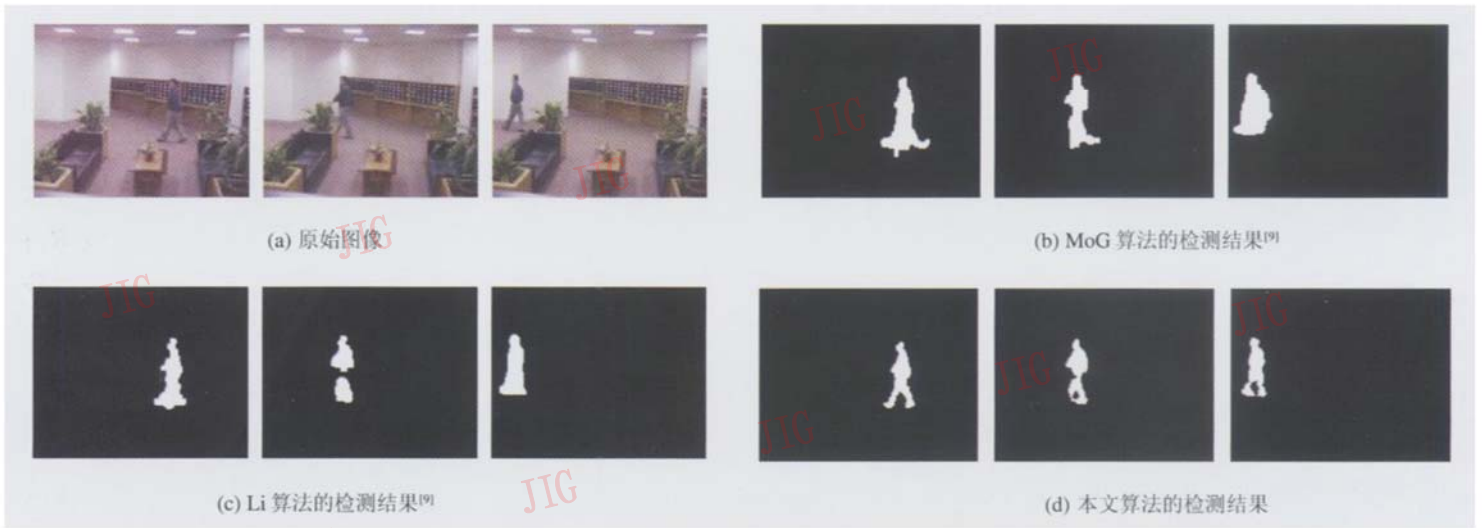


图 3 原始数据以及文献 [9] MoG 和本文算法的检测结果

Fig 3 Original images and detecting results of MoG^[9], and the proposed method in an office scene

算法对阴影的检测更鲁棒。即使没有进行滤波处理, 前景检测结果也较满意。

图 4 是“Hall monitor”图像序列的检测结果, 其中第 1 列为变化检测的结果; 第 2 列为检测的阴影区域; 第 3 列为去除阴影后的运动目标。初始背景图像是由图像序列的前 10 帧取中值得到的。在这段图像中, 由于室内有许多方向的灯光, 且有时存在着一些随机的闪烁变化, 因此目标在运动时既有投射阴影, 也有自身阴影, 并且黑衣人的裤子颜色与背景很相近, 这些都对运动目标的检测和阴影的检测带来了很大困难。但利用本文算法可取得良好的检测结果, 不仅阴影都被检测出来, 并被去除, 而且保留了运动目标的轮廓细节, 获得了较准确地运动目标图像区域。

图 5 显示了本文算法对 PETS 中户外测试数据的变化、阴影和运动目标的检测结果。PETS 的原始图像序列的数据是 JPEG 格式, 除了由于压缩引入噪声外, 还存在车窗的反光和阴影和由云层移动引起的光照变化以及植被枝叶摇动等背景噪声。图 5 中, 左边为检测的变化区域, 中间是原始数据和检测的阴影 (白色)、反光 (白色), 右边为运动目标检测结果。由图 5 检测结果可见, 图中白色和黑色小汽车、行人周围的投射阴影, 以及图中右下角黑色小汽车车窗上的反光均被成功检测和去除。

图 6 是一个建筑物内走廊场景的检测结果, 在这个图像序列中, 由于灯的开启, 引起了背景光照从暗到亮的剧烈变化。图中分别给出了原始数据、GM



图 4 室内视频图像数据的检测结果

Fig 4 Detection results of an indoor video image sequence



图 5 本文算法对 PETS 中户外测试数据的变化、阴影和目标的检测结果

Fig 5 Detection results of changes, shadows and objects for outdoor data in the PETS test data

算法、SM 算法和本文提出的算法的检测结果。由图 6 可见, 本文算法比其他算法对背景光照的变化较为鲁棒。同时, 本文算法也改善了传统的 SM 算法在场景较暗区域的检测性能。

6 结 论

本文提出了一种新的用于精确检测运动目标的 SM 方法。根据这种基于 Cauchy 分布的方法, 不仅可以实现对光照变化和背景扰动等背景变化较为鲁棒的建模, 还可以较精确地检测前景目标, 识别和去除阴影、反光。取得了优于基于 MoG 的方法、传统的 SM 方法以及 Li 等人改进的 SM 方法的检测结果, 并改善了 SM 方法在较暗区域的目标检测效果。下一阶段将进一步对 Cauchy 分布模型进行更深入的研究和简化各个阈值的确定方法。

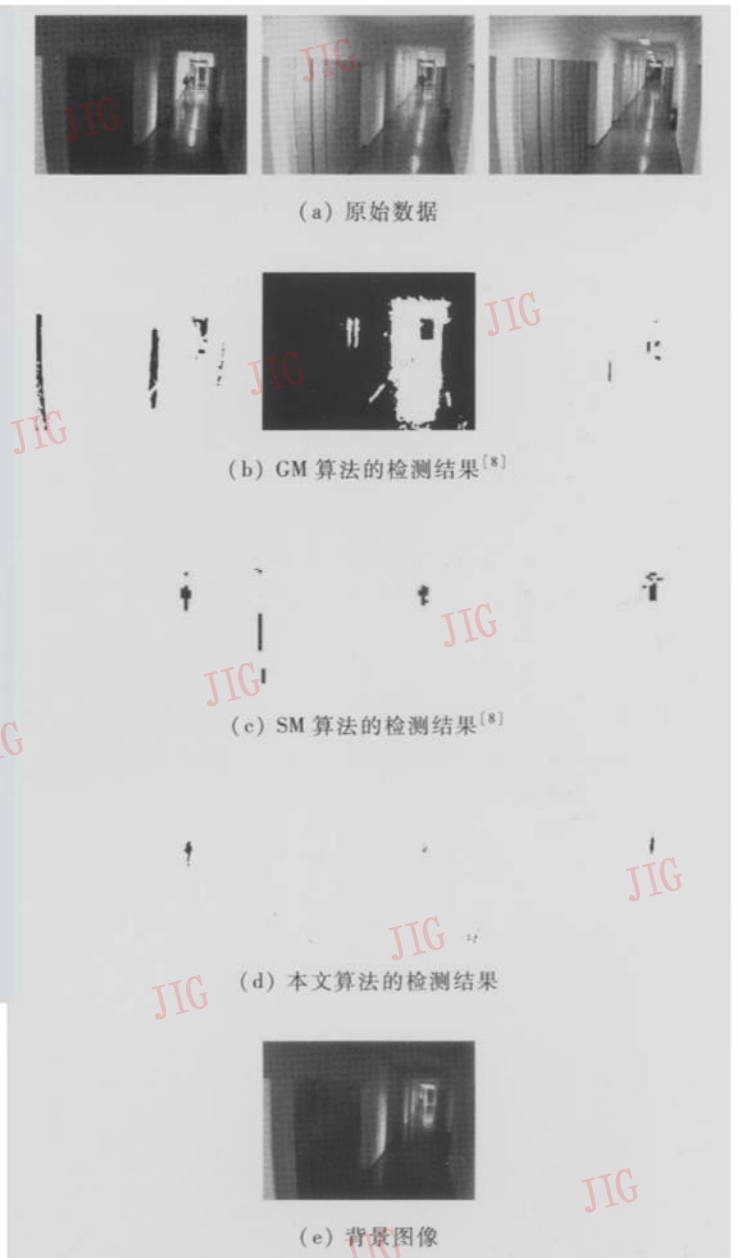


图 6 原始数据和基于 Gaussian 分布模型、基于 SM 的方法和本文算法在一个光线变化场景中的检测结果
Fig 6 Original images and detection results of a Gaussian distribution based model, a SM based method and the proposed method in an illumination change scene

参考文献 (References)

- Skifstad K, Jain R. Illumination independent change detection from real world image sequences [J]. Computer Vision, Graphic Image Process 1989, 46(3): 387~399
- Rosin P L, Ellis T. Image difference threshold strategies and shadow detection [A]. In: Proceedings of the 6th British Machine Vision Conference [C], Birmingham, U K, 1995, 1: 347 ~ 356
- Stauffer C, Grimson W. Learning patterns of activity using real-time tracking [J]. IEEE Transactions on Pattern Analysis Machine Intelligence 2000, 22(8): 747~757.

- 4 Haritaoglu I, Hawood D, Davis L S W. Real-time surveillance of people and their activities[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, **22**(8): 809~830
- 5 Elgammal A, Hawood D, Davis L. Nonparametric model for background subtraction[A]. In: *Proceedings of the 6th European Conference on Computer Vision [C]*, Trinity College, Dublin, Ireland, 2000, **2**: 751~767.
- 6 AchT, Kaup A, Mester R. Statistical model-based change detection in moving video[J]. *Signal Processing*, 1993, **31**(2): 165~180
- 7 Phong B, Tuong I. Illumination for computer generated pictures[J]. *Communications of the ACM*, 1975, **18**(6): 311~317.
- 8 Durucan E, Ebrahimi T. Change detection and background extraction by linear algebra[J]. *Proceedings of IEEE*, 2001, **89**(10): 1368~1381.
- 9 Huang L W, Gu I Y, Tian Q. Statistical modeling of complex backgrounds for foreground object detection[J]. *IEEE Transactions on Image Processing*, 2004, **13**(11): 1459~1472.
- 10 Liu Q, Sun M G, Scabassi R I. Illumination invariant change detection model for patient monitoring video[A]. In: *Proceedings of The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society [C]*, San Francisco, CA, USA, 2004, 1782~1785.
- 11 Wren C R, Azabajejani A, Darrell T, *et al*. Pfunder: real-time tracking of the human body[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, **19**(7): 780~785.
- 12 Pece A E C. Tracking of non-Gaussian clusters in the PETS2001 image sequences[A]. In: *Proceedings of the Second IEEE Workshop on Performance Evaluation in Tracking and Surveillance (PETS' 2001) [C]*, Kauai HI, USA, 2001.
- 13 Ming Ying Jiang, Jing-jie. Background modeling and subtraction using a local linear-dependence-based Cauchy statistical model[A]. In: *Proceedings of the 7th Biennial Australian Pattern Recognition Society Conference-DICT [C]*, Sydney, Australia, 2003, **1**: 469~478.
- 14 Xiang Gao, Boulton T E, Coetzee F, *et al*. Error analysis of background adaption[A]. In: *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition [C]*, Hilton Head Island, SC, USA, 2000, 503~510.
- 15 Prati A, Mikic I, Trivedi M M, *et al*. Detecting moving shadows: Algorithms and evaluation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, **25**(7): 918~923.
- 16 Andersson M, Giesen J, Pauly M, *et al*. Bounds on the k -neighborhood for locally uniformly sampled surfaces[A]. In: *Proceedings of the 1st Symposium on Point Based Graphics (SPBG) [C]*, ETH Zurich, Switzerland, 2004, 167~171.

基于不变特征的运动视频序列自动配准算法

李 静 杨 涛 潘 泉 程咏梅

(西北工业大学自动化学院, 西安 710072)

摘 要 快速、鲁棒的图像配准是运动视频处理的基础,也是制约后继应用稳定性及可靠性的关键。针对运动视频中存在的图像平移、旋转、尺度及光照变化,提出一种基于不变特征的快速图像配准算法,包括特征点检测、描述和匹配。首先通过多层箱式滤波器构建图像多尺度空间,并同时考虑质量与空间分布检测特征点;然后用主成分分析法对 SIFT (scale invariant feature transform) 特征进行降维,用于特征描述;最后根据描述子主成分的差异设计层叠分类器,加速特征匹配。定量分析实验和对视觉监视系统中球形摄像机和无人机航拍视频的实验结果表明,该算法具有良好的匹配性能,为后继运动载体上的运动目标检测、跟踪、分类等处理提供了坚实基础。

关键词 图像配准 不变特征 多尺度空间 主成分分析 层叠分类器

中图分类号: TP391 文献标识码: A 文章编号: 1006-8961(2008)02-0335-10

Invariant Feature Based Automatic Motion Video Registration

LI Jing YANG Tao PAN Quan CHENG Yongmei

(College of Automation, Northwestern Polytechnical University, Xi'an 710072)

Abstract Fast and robust image registration is an important research problem in motion video processing. In this paper, we present a novel invariant feature based automatic image registration method to deal with the large image transformation, rotation, scale and illumination changes. The algorithm includes three parts: invariant feature detection, description and matching. First, the multi-scale space of the image is created via a multi-level box filter, and then the feature points are detected in scale space by considering the quality and special distribution simultaneously. Second, we use the Principle Component Analysis to descend dimension of SIFT (Scale invariant feature transform) for feature description. Finally, based on the principle component of the descriptors, a cascade filtering is designed to speed up the feature matching. Experiments with motion videos captured by the dome camera and the Unmanned Aerial Vehicle demonstrate that the proposed algorithm has satisfied performance, and it can provide a solid foundation for further processing such as moving object detection, tracking and classification from moving platform.

Keywords image registration, invariant feature, multi-scale space, PCA, cascade filtering

1 引言

运动视频是由搭载在运动平台上的摄像机获取的图像数据。视觉监控系统中,运动平台包括旋转云台、车辆、无人机等。相比静止摄像机采集的视频数据,运动视频能覆盖更大的场景、跟踪拍摄感兴趣的目标。因此,对运动视频的自动处理与分析已成

为智能视觉监控领域的研究热点和前沿课题。

图像配准是运动视频分析的关键技术环节,其目的是恢复图像序列帧间的空间几何关系。考虑到载台运动、摄像机自身旋转、俯仰、缩放等因素,认为针对运动视频,理想的图像配准算法应具有解决下列问题的能力:

(1) 视场旋转 载台运动或摄像机转动造成的图像旋转变化;

基金项目:国家自然科学基金重点项目(60634030);航空科学基金(2007ZC53037);高等学校博士学科点专项科研基金(20060699032)

收稿日期:2007-06-25 改回日期:2007-12-07

第一作者简介:李静(1979~),女,西北工业大学自动化学院控制理论与控制工程专业博士研究生。研究方向为计算机视觉与模式识别、图像配准、智能视频监控。E-mail: jinglinwpu@163.com

- (2) 视场缩放 载台运动或摄像机镜头缩放带来的图像尺度变化;
- (3) 仿射变化 视角变化引起的图像仿射变形;
- (4) 光照变化 不受控环境中的亮度变化;
- (5) 噪声干扰 环境噪声、运动模糊等;
- (6) 实时性 实时性是决定算法是否能够实际用的关键。

尽管图像配准算法众多^[1],但早期的序列图像配准算法多假设同一像素在相邻帧间位置变化不大,当运动视频中帧间出现较大位置差异、几何变化时算法性能会降低。

近年来,在图像匹配和目标识别^[2~4]领域取得巨大成功的不变特征(invariant feature)匹配方法得到学者的广泛关注。所谓“不变特征”是指在图像在几何变化、光照变化、噪声干扰时仍保持稳定性的特征。该类算法的关键步骤包括不变特征检测和特征描述两部分。

Lindeberg率先提出在归一化的高斯拉普拉斯(LoG)尺度空间,通过极值点检测定位出特征所在的位置和尺度^[5]。为了降低计算量,Low e用双高斯差函数(difference of Gaussian, DoG)对 LoG 进行近似^[2]。由于 LoG^[5]和 DoG^[2]都具有较强的边缘响应,M kolajczyk提出一种 Harris-Laplacian 特征检测算法,用 Harris 检测出图像中的角点,并在拉普拉斯尺度空间定位特征尺度^[6]。上述检测器均具有良好的旋转、尺度不变性,对图像的仿射变形也具有一定的适应性。但是在构建尺度空间时需要进行高斯卷积,且输出特征点数不可控,不适于对运动视频序列的快速处理。

不变特征描述子主要包括基于分布的描述子、基于时频特性的描述子和基于差分不变量的描述子等^[2,7~9]。M kolajczyk 针对光照变化、图像旋转、缩放、仿射变形、模糊、压缩等情况,对 11 种典型的描述子进行性能评估^[10],结果表明 SIFT 描述子^[2]的性能最优。

本文针对视觉监视中的运动视频自动配准进行研究。考虑到配准中存在的图像旋转、缩放、光照变化、噪声干扰、实时性等问题,提出一种基于不变特征的快速图像配准算法(见图 1),包括特征检测、描述和匹配。在检测时,首先通过多层箱式滤波器模拟高斯卷积,构建图像多尺度空间,大幅度降低计算量,然后同时考虑质量与空间分布,高效地定位出数

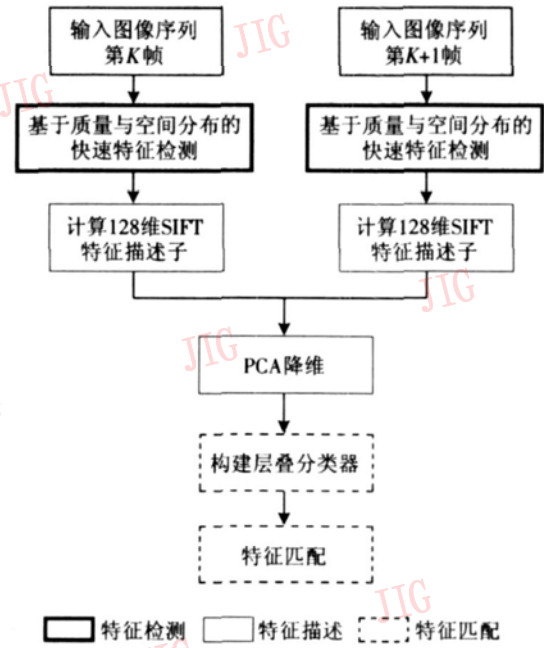


图 1 算法流程图

Fig 1 Flow chat of the algorithm

量可控的特征点;在特征描述时选用 SIFT 描述子。考虑到高维数据对后继匹配速度的影响,采用主成分分析法对 128 维 SIFT 描述子进行降维;在匹配时根据描述子主成分的差异设计层叠分类器,快速滤除大量无关特征点。

定量分析实验结果表明,检测算法在图像旋转、缩放、噪声干扰等条件下均具有良好的可重复性;与直接采用 SIFT 描述子相比,基于 PCA 降维和层叠分类器的描述与匹配算法能在保持匹配性能的同时,大幅度提高匹配速度,更适于对实时性要求较高的运动视频的分析处理。

以本算法为核心,构建了一个自动视频配准与拼接系统,对视频监视中的球形摄像机和无人机航拍获得的运动视频^[11]的实验表明,算法具有精度高、鲁棒性强等优点。

2 基于质量与空间分布的特征点检测

2.1 基于多层箱式滤波的尺度空间构建

考虑到摄像机的缩放,运动视频中图像间可能存在较大尺度变化,在特征检测时需要先构建图像的多尺度空间。文献[2]、[5]、[6]在建立尺度空间时都采用高斯核卷积,随着滤波窗宽的增大,计算量会急剧增加。Grabner 提出用窗口内像素灰度的均值来代替高斯卷积^[12]。但是, Lindeberg 证明高斯卷积核是实现尺度变换的唯一线性变换核^[13],而

均值滤波得到的尺度空间与真实值存在较大偏差。基于此, 本文用多层箱式滤波器 (multiple box filter) 模拟高斯卷积, 并通过积分图使得计算量和滤波窗宽大小无关。

假设用高斯函数 $G(x, y, \sigma)$ 对输入图像 I 进行卷积, 构建尺度空间

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (1)$$

式中, x, y 为像素点位置, σ 是高斯函数的标准差。设计一个 K 层阶梯状函数 $MB(x, y, \sigma)$ 来模拟高斯 $G(x, y, \sigma)$, 每一层的权值系数 $P(k)$ 由该层覆盖下的环形区域 C_k 中高斯核系数的均值得到:

$$P(k) = \frac{1}{Z_k} \sum_{(x, y) \in C_k} G(x, y, \sigma) \quad (2)$$

式中, Z_k 为环形区域 C_k 的面积。则尺度空间 L 可由下式构建:

$$L(x, y, \sigma) = MB(x, y, \sigma) * I(x, y) = \sum_{k=1}^K \left| P(k) \sum_{(x, y) \in C_k} I(x, y) \right| \quad (3)$$

考虑到 C_k 为环形区域, 式 (3) 求和部分不能直接用积分图计算, 因此用两个矩形区域的差求解 C_k 内的像素和:

$$\sum_{(x, y) \in C_k} I(x, y) = \sum_{(x, y) \in (C_k, C_{k+1}, \dots, C_K)} I(x, y) -$$

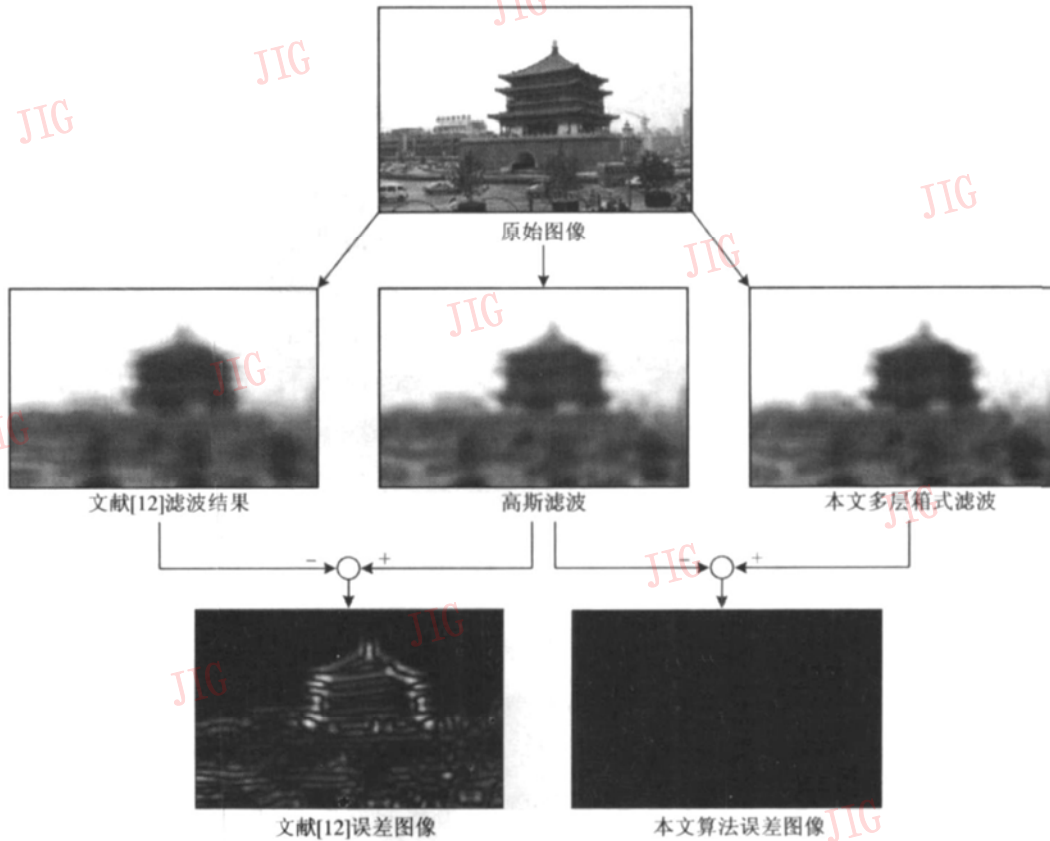


图 2 对高斯滤波模拟精度的比较

Fig 2 Comparison of filtering accuracy

$$\sum_{(x, y) \in (C_{k+1}, \dots, C_K)} I(x, y) \quad (4)$$

将式 (4) 代入式 (3) 并推导得:

$$L(x, y, \sigma) = \sum_{k=1}^K (P(k) - P(k-1)) \sum_{(x, y) \in (C_k, \dots, C_K)} I(x, y) \quad (5)$$

式 (5) 表示尺度空间 L 可以由多个不同窗宽的箱式滤波组合得到, 滤波器可用积分图快速计算。系数 $P(k)$ 可离线求解, 且有 $P(0) = 0$ 。本文实验均采用 3 层箱式滤波器构建尺度空间, 其滤波窗宽半径分别为 $1\sigma, 2\sigma, 3\sigma$ 。表 1 给出高斯滤波器、均值滤波^[12]以及本文算法在一个像素点上的计算量比较。可以看出, 本文算法和均值滤波的计算量与窗宽大小无关。图 2 给出了均值滤波和本文算法对高斯滤

表 1 滤波算法计算量比较

Tab 1 Comparison of computational time

滤波方法	加法	乘法
2维高斯滤波	$N^2 - 1$	N^2
1维高斯滤波	$2N - 2$	$2N$
均值滤波 ^[12]	3	1
本文算法	$2 + 9$	3

注: N 为滤波器窗宽

波模拟精度的比较。相比文献 [12], 多层箱式滤波能更精确地模拟高斯滤波 (图 2 差异图像), 大幅度降低构建尺度空间的误差。

本文采用多阶多层的思路构造图像多尺度空间 L 。将尺度空间分为 O 阶, 每增加一阶, 图像按 2 做降采样, O 由图像尺寸 (W, H) 和降采样率 2 求解:

$$O = \log_2(m \text{ in}(W, H)) \quad (6)$$

每阶内将尺度空间细分为 S 层。第 i 阶第 j 层图像由第 i 阶第 1 层的图像经多层箱式滤波得到, 滤波器参数由所模拟的高斯函数离线求解。令 $\sigma^{(0)}$ 为第 1 层高斯滤波的标准差, 则相应多层箱式滤波器的窗宽 $a^{(0)} = 6 \cdot \sigma^{(0)} + 1$ 第 j 层图像的滤波窗宽为 $a^{(0)} \cdot 2^{jS}$ 。实验中取 $a^{(0)} = 1.1, S = 5$ 。

2.2 基于质量与空间分布的特征点检测

图像多尺度空间建立好以后, 下一步就是如何寻找特征点。目前, 大多数算法^[2,5,6]在检测特征点时需要遍历整个尺度空间中的所有采样点, 同时根据极值、角点响应等特性进一步确定该点是否为特征点。但这种依次搜索的检测方法存在以下不足:

(1) 该方法将大部分时间消耗在非特征点的判断上, 且搜索时间随图像大小呈非线性上升;

(2) 由于本文采用仿射模型估计运动视频中帧间的几何关系, 只需要较少的匹配点即可精确估计模型参数, 而上述方法输出的特征点数量均无法精确控制。

为此, 提出一种基于质量和空间分布的搜索算法, 从根本上解决上述问题, 而且, 该算法也可以集成到现有的特征点检测算法中。搜索算法包括以下 3 个主要步骤:

(1) 构建质量空间

采用相邻尺度图像差的绝对值来构建质量空间, 并用局部极值和 Harris 角点响应判断是否为特征点。由相邻尺度的两个图像相减绝对值得到 DoMB (difference-of-multiple box) 图像:

$$I_{\text{DoMB}}(x, y, \sigma) = |L(x, y, k\sigma) - L(x, y, \sigma)| \quad (7)$$

显然, 某一点的 DoMB 取值越高, 其为局部极值点的概率越大。当给定需要检测的特征点数目后, 如果能先对 DoMB 进行排序, 然后按照取值从高到低搜索特征点, 则能够回避在大量低概率点的搜索, 快速、准确地检测到所需数量的高质量的特征点。由于整个尺度空间中像素点数量巨大, 难以快速对 DoMB 排序。为此, 在生成 DoMB 图像的同时, 根据

每一个像素点的 DoMB 值, 同步建立图像的质量直方图 $Q(n)$ 和索引表 QI 。对质量直方图中取值为 n 的点, 索引表 QI 记录了该点所在的尺度和空间位置。实验中 n 的取值范围为 $n \in [0, 60]$ 。

图 3 给出了图 2 原始图像尺度空间的质量直方图。从图中可以看出, 尺度空间中绝大部分像素点都集中在 DoMB 取值较低的部分。这样, 算法根据质量直方图由高到低依次搜索, 仅需要在最可能的点上进行少量的计算即可检测出所需数目的特征点, 大幅度减少了在非特征点上消耗的时间, 从而从根本上加快了检测特征点的速度。大量实验结果表明, 这种方法通常只需要搜索全部采样点的 2% 左右即可以找到足够的特征点。

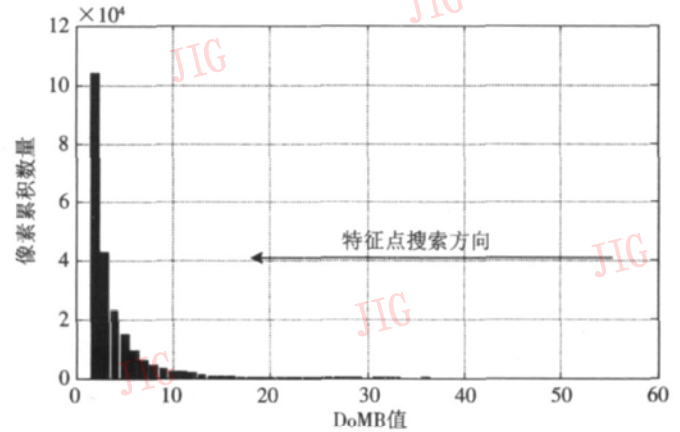


图 3 DoMB 的质量直方图
Fig 3 Histogram of DoMB

(2) 特征点定位

对待检测的采样点, 通过计算其局部极值和 Harris 角点响应进行特征点判断。局部极值检测^[5]由 Lindeberg 提出, 将采样点分别与上下相邻尺度对应位置的周围 18 个像素以及相同尺度的周围 8 个像素, 共 26 个像素比较, 确保检测到的特征点同时是尺度空间和 2 维图像空间上的局部极值。同时, 为了后继匹配的稳定性, 根据采样点局部自相关矩阵的特征值计算 Harris 角点响应, 去除不稳定边缘点。

(3) 基于动态决策图的非极大值抑制

上述基于质量空间搜索的方法具有计算量小, 可以精确控制输出点数的优点, 但是仅按照质量空间搜索的方式可能导致大量特征点集中在图像的某些纹理丰富的区域, 从而降低了特征点的空间分布性。由于载台机动, 运动视频中帧间可能出现仅有小部分重叠区域的情况, 且重叠区域的位置无法事

先得到。为了保证在该情况下仍能对图像正确匹配, 需要检测到的特征点尽可能分布在图像的各个区域。为此, 构造一个和尺度空间等大、初值为 0 的动态决策图 $D(x, y, \sigma)$ 。在按照质量空间从高到低搜索时, 先判断动态决策图中当前点的状态, 如果为 1 则表示当前点的局部邻域内已经存在检测到的特征点, 并终止对该点的进一步判断; 如果为 0 则继续判断该点局部极值和角点响应, 并将检测到的特征点在动态决策图中的邻域像素状态设为 1。其中, 关闭邻域的半径 ρ 由特征点所在的尺度图像大小确定, 且有 $\rho = \alpha \times \min(w_s, h_s)$, 式中 w_s, h_s 为特征点所在尺度图像的宽和高, 比例系数 α 用于控制关闭范围, α 越大, 特征点的空间分布性越强。这里实验中 α 取值为 0.03。

通过结合质量空间和动态决策图, 本文算法可以快速检测到具有较高质量、良好的空间分布、数量精确可控的特征点。

3 基于 PCA 和层叠分类器的特征匹配

特征点检测结束后, 下一步就是如何描述和匹配特征点。M Kolajczyk 证明了 SIFT 描述子^[2]性能最优^[10]。但由于维数较高, 用 128 维 SIFT 描述子做特征匹配时运算量大, 难以快速处理。为此, 首先用主成分分析法对 128 维 SIFT 描述子降维, 然后根据主成分的误差分布设计层叠分类器, 在搜索匹配时快速滤除大量无关特征点。

(1) 构造 SIFT 描述子

为了保证描述子的旋转不变性, 首先要计算特征点的主方向。文献[2]中主方向是指特征点邻域内梯度方向直方图中最大值所对应的方向。此外, 为了保证匹配的稳定性, Lowe 将梯度强度高于最大值 80% 的方向作为辅方向^[2], 这样一个特征点可能有多个不同方向的描述子。

构造描述子时, 对任意一个特征点, 以主方向或辅方向为参照, 首先在其所在尺度空间中取出以该点为中心的 16×16 大小的邻域, 并将该邻域均分为 4×4 个子区域, 每个子区域包含 4×4 个像素。然后计算每个子区域的 8 方向梯度方向直方图, 并根据位置将直方图依次排序, 构成一个 $4 \times 4 \times 8 = 128$ 维的向量, 即 SIFT 描述子。为了增强描述子对特征点定位误差的鲁棒性, 在匹配前用高斯核对 128 维 SIFT 描述子进行平滑。

(2) 基于主成分分析的描述子降维

主成分分析的目标是在误差平方和意义下用低维子空间表示高维数据。用主成分分析法对 128 维 SIFT 描述子降维的过程如下: 首先将两幅待匹配图像中所有 n 个特征点的 SIFT 描述子 x_1, x_2, \dots, x_n 作为样本, 计算出 128 维均值向量 $\bar{\mu}$ 和 128×128 的协方差矩阵 R 。

$$\bar{\mu} = [\bar{\mu}_1, \dots, \bar{\mu}_p], \bar{\mu}_i = \frac{1}{n} \sum_{k=1}^n x_k(i) \quad (8)$$

$$R = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1p} \\ r_{21} & r_{22} & \dots & r_{2p} \\ \vdots & \vdots & \dots & \vdots \\ r_{p1} & r_{p2} & \dots & r_{pp} \end{bmatrix} \quad (9)$$

$$r_{ij} = \frac{\sum_{k=1}^n (x_k(i) - \bar{\mu}_i)(x_k(j) - \bar{\mu}_j)}{\sqrt{\sum_{k=1}^n (x_k(i) - \bar{\mu}_i)^2 \sum_{k=1}^n (x_k(j) - \bar{\mu}_j)^2}} \quad (10)$$

其中, 维数 $p = 128$, $x_k(i)$ 表示第 k 个 SIFT 描述子的第 i 维分量。

然后计算协方差矩阵 R 的 p 个特征值 λ 和特征向量 e 并将特征值按从大到小顺序排列, 则有 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ 和对应的特征向量 $\{e_1, e_2, \dots, e_p\}$ 。选出对应最大 d 个特征值的特征向量作为主成分方向, d 的大小由下式给出:

$$v(i) = \frac{\sum_{k=1}^i \lambda_k}{\sum_{k=1}^p \lambda_k}, i = 1, 2, \dots, p \quad (11)$$

$$d = \arg \min_i (v(i) > T) \quad (12)$$

式中, $v(i)$ 为累计贡献率, 本文实验中累积贡献率的阈值 $T = 0.86$ 。最后构造一个 $p \times d$ 的矩阵 A , 它的列由 d 个特征向量组成。将原始 128 维 SIFT 描述子按照式 (13) 投影到这个 d 维子空间, 得到描述子的主成分表示 y_1, y_2, \dots, y_n 。

$$y_k = x_k \cdot A \quad (13)$$

(3) 基于层叠分类器的快速匹配

在匹配时, 对于第 1 幅图像的某一特征点, 需要遍历另一幅图像的所有特征点, 计算每个可能匹配点对描述子的欧式距离。为了快速滤除无关特征点, 进一步提高匹配速度, 用描述子主成分设计一个由 d 个判决条件构成的层叠分类器。

为构建层叠分类器, 将两幅图像正确匹配点的描述子主分量相减, 得到匹配点主成分误差图。通

过对不同场景、光照变化、图像缩放、旋转等条件下的大量测试图像的统计结果表明, 匹配点主成分的误差分布近似于一个高斯分布, 且高斯分布的标准差 σ 取值基本保持恒定, 不随测试图像改变, 有 $3\sigma = 0.15$ 。图 4 (c) 和图 4 (d) 给出一组匹配点主成分

误差曲线和分布的示意图。基于误差分布特性, 在计算欧式距离前, 首先逐一判断该特征点对的主成分差异 ΔPC_i 是否高于阈值 T_{pc} (如图 4 (e) 所示), 为了保证不遗漏真实匹配点, 在实验中 $T_{pc} = 0.25$ 略高于 3σ 。

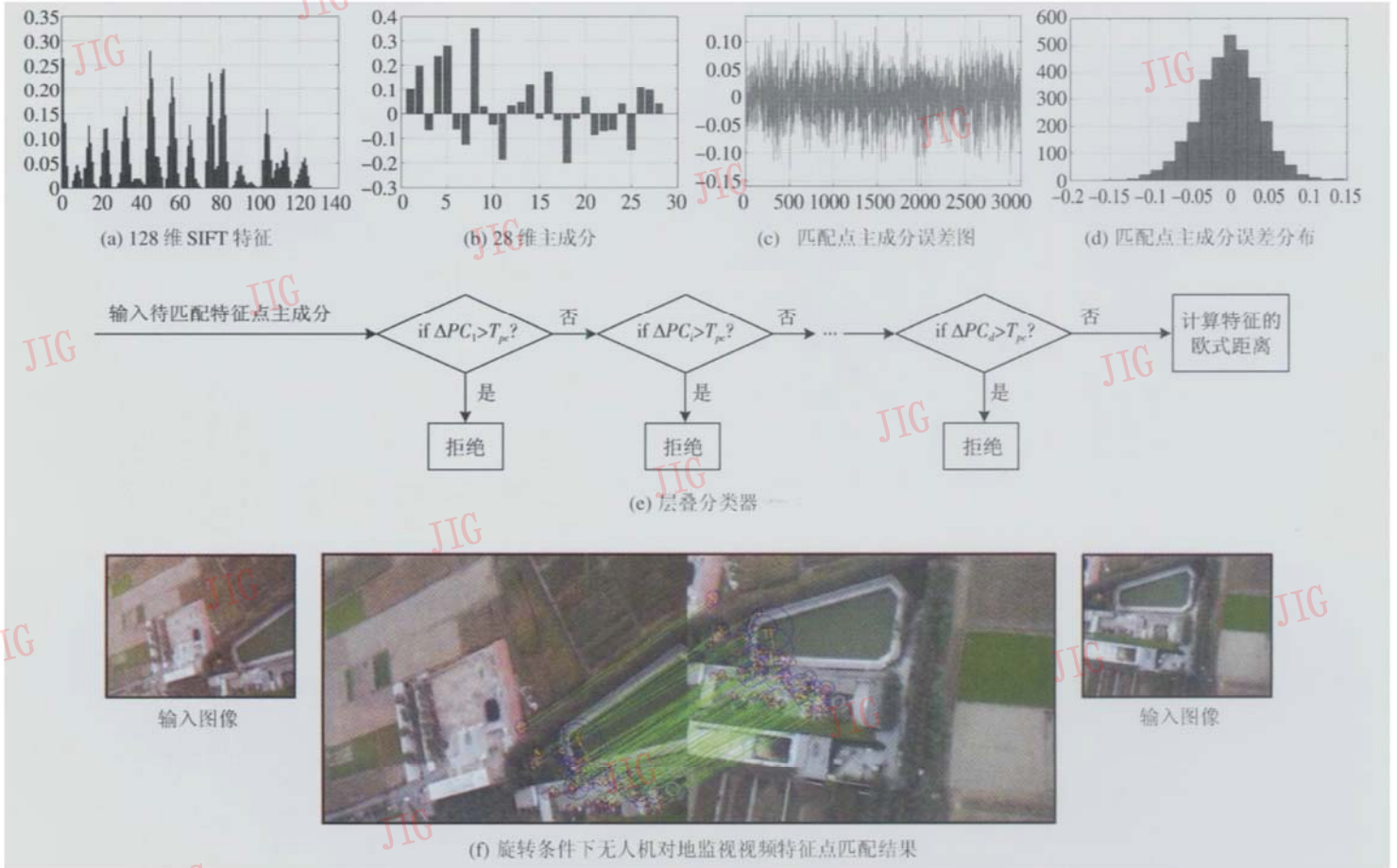


图 4 基于主成分分析和层叠分类器的特征点描述与匹配结果
Fig 4 PCA and cascade filter based feature description and matching result

在得到可能特征点对的欧式距离后, Lowe 为了保证描述子的独特性, 用所有距离中的最小值和次小值之比做度量, 如果比率小于一定阈值 T_E , 则将取得最小距离的点作为匹配点, 通常 T_E 的取值范围为 $0.4 \sim 0.6$ ^[2]。由于一个特征点可以有多个辅方向, 可能得到多个不同的描述子, 因此本文在匹配时对 Lowe 的算法^[2]进行改进, 要求距离最小值和次小值应对应于不同的特征点。

图 4 给出一个基于主成分分析和层叠分类器的特征点描述与匹配结果。测试图像为自建数据库中无人机对地监视视频, 由于载机机动, 图 4 (f) 左右两幅输入图像存在较大的位置偏差和角度旋转。采用本文第 2 节提出的基于质量与空间分布的快速特征点检测算法, 通过控制输出特征点数, 在两幅输入

图像上分别检测出 1 000 个特征点。用这 2 000 个特征点作样本, 经过主成分分析, 将原始的 128 维 SIFT 描述子 (图 4 (a)) 降为 28 维 (图 4 (b))。直接匹配的方式需要比较 1 000 000 次, 而本文提出的基于主成分差异的层叠分类器 (图 4 (e)) 拒绝了 989 604 次 (占总比较次数的 98.96%) 的无关比较, 最终找到 112 对正确匹配点。图 4 (f) 中用蓝色圆圈标明特征点所在位置和尺度, 从圆心出发的线段标明主方向。

4 实验分析

首先对算法性能进行定量分析, 然后给出算法在视频监视系统中的两类典型应用。

(1) 特征检测算法可重复性分析

可重复率 (repeatability rate) 是指算法在图像发生几何变换、光照变化以及噪声干扰前后独立检测到相同特征点的比率^[14], 它的好坏将直接影响后继匹配的性能, 是衡量检测算法性能的重要指标。实验采用文献 [6] 的方法计算可重复率

$$r_r = \frac{C(m_1, m_2)}{mean(m_1, m_2)} \quad (14)$$

假设图像间的变换矩阵 H 已知, 检测算法在两幅待匹配图像上分别检测出 m_1 和 m_2 个特征点。式中 $C(m_1, m_2)$ 表示将 m_1 个特征点经过变换矩阵 H 映射到第 2 幅图像后, 找到的可重复特征点总数。这里, 判断为相同特征点的标准是位置误差不超过 2 个像素, 尺度误差率在 20% 以内。

图 5 给出了图像在旋转、尺度变化和加噪条件下的 4 组可重复性分析实验结果。

图 5(a) 中原始图像以 $\pi/16$ 为间隔逆时针旋转一周得到 32 幅图像。其可重复率 r_r 在图 5(e) 给出。可以看到 r_r 在 4 个象限中呈周期性变化, 当旋转角度为 45° 倍数时降到最低, 但仍然达到 0.7 以上, 表明本文检测算法具有良好的旋转不变性。

图 5(b) 中尺度变化下的图像序列。图 5(f) 给出该序列的可重复率。实验结果看出, 当缩放比率为 2 时, 可重复率仍保持在 0.7 以上。文献 [6] 中给出的 Harris-Laplacian^[6] 和 DoG^[2] 在相同缩放比率下的可重复率分别为 0.67 和 0.5 而 Harris 算法在该尺度变化下仅具有 0.16 的可重复率。当缩放比率进一步增大到 4 时, 本文算法的可重复率为 0.3 Harris-Laplacian^[6], DoG^[2] 和 Harris 分别为 0.4, 0.32 和 0.08。可以看出, 本文算法在尺度变化下的可重复率同 Harris-Laplacian^[6] 和 Lowe^[2] 性能相当, 具有良好的尺度不变性, 可以满足运动视频中待配准图像可能出现的尺度变化。

图 5(c) 和图 5(d) 分别给出图像在椒盐噪声和高斯噪声下的一系列变换图像。图 5(c) 中椒盐噪声密度为 0.02~0.2 图 5(d) 中高斯噪声的方差从 2 增加到 20 从图 5(g) 和图 5(h) 可以看出, 在本文模拟的噪声强度范围内, 可重复性始终高于 0.5 表明算法对噪声也具有一定的稳定性。

(2) 特征描述与匹配算法性能分析

实验选择了 10 组典型图像 (图 6(a)), 对基于 PCA 降维和层叠分类器的匹配算法性能进行定量

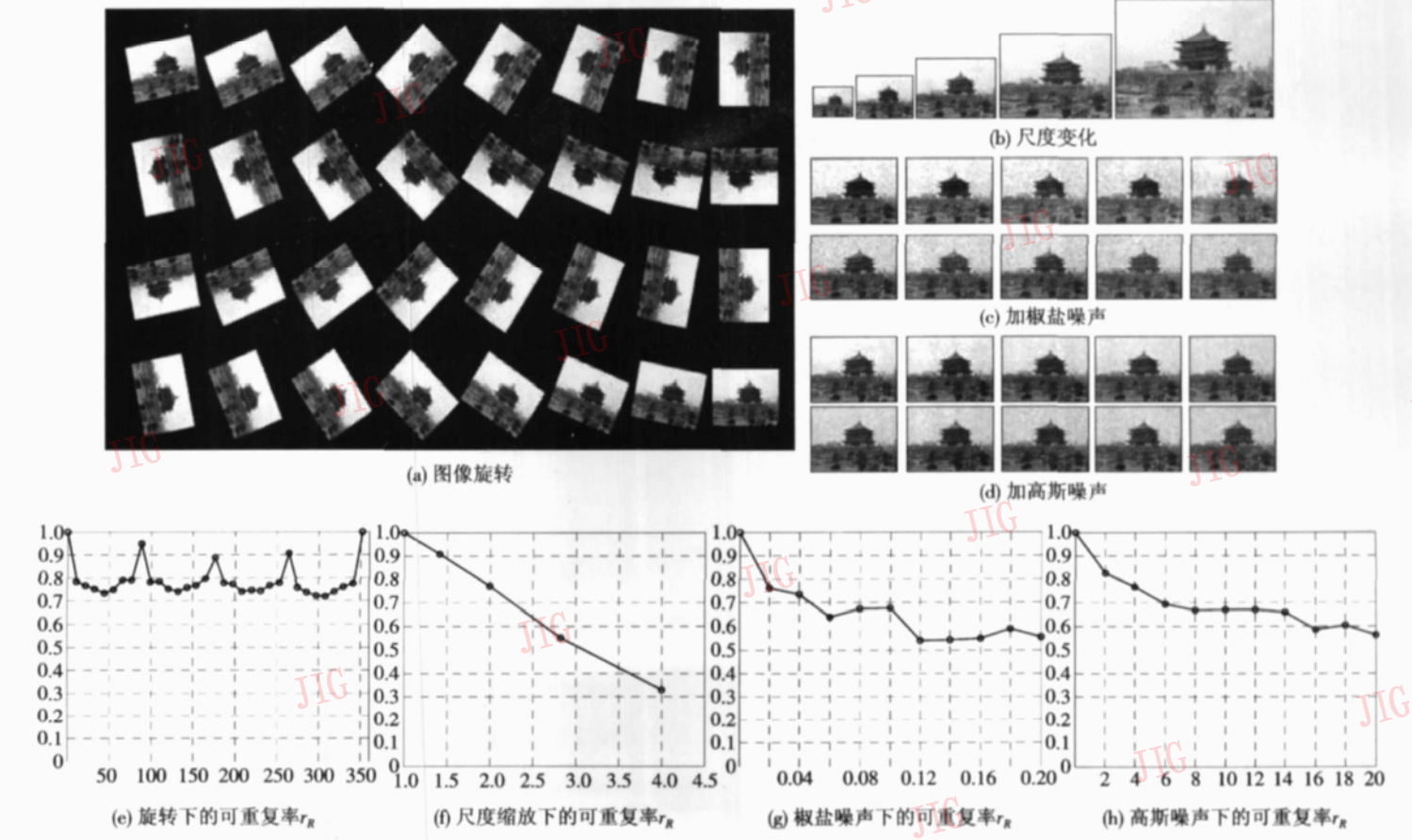


图 5 旋转、缩放、噪声条件下的特征检测算法可重复性

Fig 5 Repeatability of feature detection under rotation, scaling and noises



图 6 不变特征匹配

Fig 6 Invariant feature matching result

分析。测试图像的拍摄平台包括旋转云台 (图 6(a), No 1~ No 4) 和无人机 (图 6(a), No 5~ No 8), 图像间的变化涵盖了旋转 (图 6(a), No 1, No 2, No 6, No 8, No 9) 尺度缩放 (图 6(a), No 1~ No 4, No 9), 亮度变化 (图 6(a), No 7), 视角变化 (图 6(a), No 9, No 10) 等情况。

图 6(b) 给出部分匹配结果, 包括旋转云台 (图 6(b) 第 1 行) 和无人机 (图 6(b) 第 2 行) 图像, 图中蓝色圆圈标明特征点所在位置和尺度, 从圆心出发

的黄色线段标明特征主方向。表 2 从描述子维数、特征搜索数量和正确匹配点数 3 个方面给出 PCA 降维和层叠分类器滤波前后的性能比较。从表 2 第 2 列可以看出, 主成分分析使描述子维数从 128 维降至 30 维以下, 平均维数仅相当于降维前的 21.72%。与直接搜索特征点相比, 通过比较两个待匹配点的主成分差异, 基于层叠分类器的滤波算法快速滤除了大量无关特征点, 对 10 组测试图像的平均特征点搜索匹配次数仅为直接搜索的 5.63%

表 2 PCA 降维和层叠分类器滤波前后特征匹配性能比较

Tab 2 Comparison of matching performance with PCA and cascade filtering

图像编号	描述子维数比较			特征点搜索匹配次数比较			正确匹配点数比较		
	SIFT	本文算法	相对比率 (%)	直接搜索	本文算法	相对比率 (%)	SIFT	本文算法	相对比率 (%)
No 1	128	28	21.88	1 452 282	62 065	4.27	78	67	85.90
No 2	128	29	22.66	1 294 164	58 233	4.50	13	12	92.31
No 3	128	28	21.88	1 356 957	73 979	5.45	19	13	68.42
No 4	128	28	21.88	874 752	42 784	4.89	22	21	95.45
No 5	128	27	21.09	840 465	38 533	4.58	89	85	95.51
No 6	128	27	21.09	1 024 985	48 187	4.70	136	139	102.21
No 7	128	27	21.09	1 347 306	73 844	5.48	194	184	94.85
No 8	128	26	20.31	865 242	86 235	9.97	19	25	131.58
No 9	128	29	22.66	4 395 240	284 924	6.48	260	251	96.54
No 10	128	29	22.66	3 577 000	215 270	6.02	21	22	104.76
比率均值			21.72			5.63			96.75

(表 2 第 3 列末尾)。

正确匹配点数是衡量匹配算法性能的重要指标, 从表 2 第 4 列可以看出, 本文算法与直接采用 128 维 SIFT 描述子得到的正确匹配点数基本相当, 为后者的 96.75% (表 2 第 4 列末尾), 对部分测试图像, 本文算法的匹配性能略高于 SIFT (表 2 No 6、No 8、No 10)。

上述结果表明, PCA 降维和层叠分类器能在保持 SIFT 描述子匹配性能的同时大幅度提高匹配速度, 更适于对处理速度要求较高的运动视频分析。

(3) 景深变化下的图像匹配

图 7 给出算法对不同景深图像的匹配结果。图 7 (a) 中, 由于拍摄目标离镜头比较近, 摄像机转动时, 画面中景深出现较大变化 (图 7 (a), #6), 这些变化会使目标图像发生仿射、透视变化甚至非线性形变。图 7 (b) 和图 7 (c) 分别给出本算法对 #1、#4 和 #1、#6 的匹配结果。从结果中可以看出, 随着景深的逐渐增大, 待匹配图像间发生较大形变, 此时算法的匹配性能会下降。在进一步的研究中拟引入具有仿射或透视不变性的检测与描述方法, 提高算法的稳定性。

(4) 运动视频自动镶嵌

基于不变特征的图像镶嵌^[15]能获得大范围、高

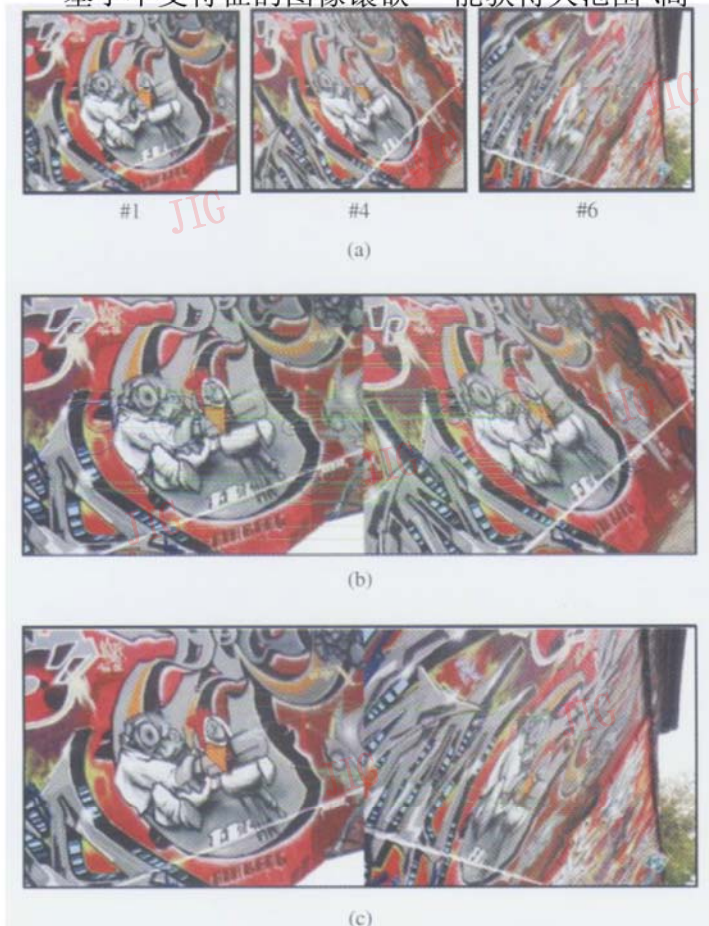


图 7 景深变化下的匹配结果

Fig 7 Matching result under large scene depth changes

分辨率的全景图, 在视频监控中具有广泛应用。以本文算法为核心, 构建了一个自动视频拼接系统。在得到匹配特征点后, 用 RANSAC (random sample consensus)^[16]去除外点, 并用最小二乘法估计帧间仿射变换参数。

图 8 给出了本文算法对视频监控中的一组运动摄像机视场拼接和定位结果。在现有监控系统中, 当操作人员控制球形摄像机跟踪某个感兴趣目标时, 往往难以定位出当前摄像机视场在整个大范围场景中的位置, 而该参数对于获取目标在整个场景的运动轨迹, 分析操作人员的感兴趣区域等都具有重要意义。图 8 第 1 列为输入视频, 第 2 列中背景为拼接得到的监控场景全景图。可以看出, 通过帧间的自动配准, 本文算法能够在摄像机缩放 (#154、#183)、转动 (#230) 等条件下准确定位出当前视场在全景图中的位置 (白色线框标明区域)。



图 8 运动摄像机自动视场定位

Fig 8 Automatic field of view location for moving camera

图 9 给出了无人机航拍运动视频的自动配准和拼接结果。测试数据来自于美国 DARPA 发布的

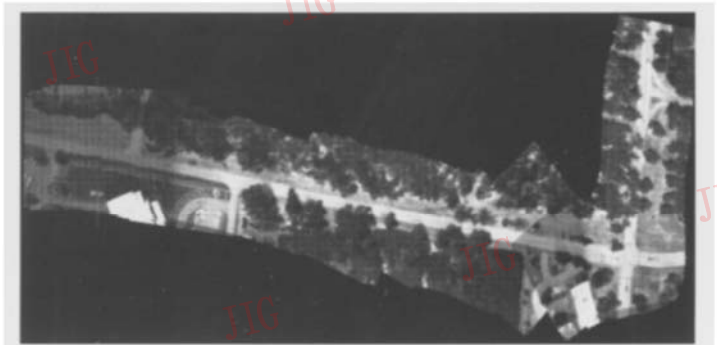


图 9 无人机航拍视频自动拼接结果

Fig 9 Automatic UAV surveillance video mosaicing result

VIVID 数据库^[11]。实验结果表明,该系统可在无人机转弯机动、光照突变和盘旋等状态下稳定、准确地进行运动视频自动配准和拼接,取得令人满意的效果。

5 结 论

提出了一种新的基于不变特征的运动视频自动配准算法,包括基于质量和空间分布的特征点检测、基于主成分分析的描述算法和基于层叠分类特征点匹配算法 3 部分。实验中,对算法性能的定量分析结果表明:(1)检测算法在图像发生旋转、尺度缩放、噪声条件下均具有良好的可重复性,且在保证特征点质量、空间分布的同时,高效检测出数量可控的特征点;(2)与直接采用 128 维 SIFT 描述子的匹配结果相比,本文算法引入 PCA 降维和层叠分类器后,不仅大幅度降低了计算量,而且仍保持了与之相同的匹配性能。

以本文算法为核心,构建了一个自动视频拼接系统,对视频监控中的球形摄像机和无人机航拍获得的运动视频进行了自动配准和拼接。实验结果表明,该算法可以解决运动视频中出现的角度旋转、尺度缩放、光照突变等问题,具有良好的鲁棒性,为后继基于运动载台的运动目标检测、跟踪、识别等处理提供了坚实的基础。

参考文献 (References)

- Zitova B, Flusser J. Image registration methods: a survey[J]. *Image and Vision Computing*, 2003, **21**(11): 977~1000
- Lowé D G. Distinctive image features from scale-invariant keypoints [J]. *International Journal of Computer Vision*, 2004, **60**(2): 91~110
- Ferrari V, Tuytelaars T, Gool L. Simultaneous object recognition and segmentation from single or multiple model views[J]. *International Journal of Computer Vision*, 2006, **67**(2): 159~188
- Zhang W, Košček J. Hierarchical building recognition[J]. *Image and Vision Computing*, 2007, **25**(5): 704~716
- Lindeberg T. Feature detection with automatic scale selection[J]. *International Journal of Computer Vision*, 1998, **30**(2): 77~116
- Mikolajczyk K, Schmid C. Indexing based on scale invariant interest points[A]. In: *Proceedings of IEEE International Conference on Computer Vision [C]*, Vancouver, BC, Canada, 2001, **1**: 525~531.
- Belongie S, Malik J, Puzicha J. Shape matching and object recognition using shape contexts[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, **24**(4): 509~522
- Ke Y, Sukthankar R. PCA-SIFT: A more distinctive representation for local image descriptors [A]. In: *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition [C]*, Washington, DC, USA, 2004, **2**: 504~513
- Bay H, Tuytelaars T, Gool L V. SURF: speeded up robust features [A]. In: *Proceedings of the 9th European Conference on Computer Vision [C]*, Graz, Austria, Springer LNCS, 2006, **3951**: 404~417.
- Mikolajczyk K, Schmid C. A performance evaluation of local descriptors[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, **27**(10): 1615~1630
- Video Verification of Identity, CMU [DB/OL], <http://www.cs.cmu.edu/~hebert/vivid.html>
- Grabner M, Grabner H, Bischof H. Fast approximated SIFT [A]. In: *Proceedings of 7th Asian Conference on Computer Vision [C]*, Hyderabad, India, 2006, 918~927.
- Lindeberg T. Scale-space theory: a basic tool for analyzing structures at different scales[J]. *Journal of Applied Statistics*, 1994, **21**(2): 224~270
- Schmid C, Mohr R, Bauckhage C. Evaluation of interest point detectors[J]. *International Journal of Computer Vision*, 2000, **37**(2): 151~172
- Brown M, Lowe D G. Automatic panoramic image stitching using invariant features [J]. *International Journal of Computer Vision*, 2007, **74**(1): 59~73
- Fischler M A, Bolles R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography[J]. *Communications of the ACM*, 1981, **24**(6): 381~395

自适应多模快速背景差算法

梁 华 刘云辉

(国防科学技术大学电子科学与工程学院, 长沙 410073)

摘 要 在多高斯模型的基础上, 从场景中模型分布不均匀性出发, 提出了一种新的快速背景差算法。该算法针对混合高斯模型中固定模型数量不足的问题, 建立了模型产生和退出的机制, 使模型数量能够自动适应场景特点, 实现了高斯模型的实时自适应分布, 即提高了准确性又有效地减少了模型的总量; 同时, 针对混合高斯模型中计算量大的问题, 对模型参数的计算进行了优化, 将耗时的浮点运算转化为整型运算, 减少了计算量; 算法中引入了生存时间和模型重现频率的概念, 通过对模型重现频率的限制有效抑制高频噪声。与混合高斯模型的实验结果对比说明, 该快速算法保持了原算法的优点, 执行速度提高 1 倍以上, 检测结果准确, 算法内存消耗小, 前景轮廓清晰, 抑制高频噪声的能力强, 整体效果优于混合高斯模型背景差算法。

关键词 视频监控 背景差算法 混合高斯模型 快速算法

中图分类号: TP391 文献标识码: A 文章编号: 1006-8961(2008)02-0345-06

Adaptive and Efficient Background Subtraction Using Multiple Models

LIANG Hua LU Yun-hui

(College of Electronic Science and Engineering, National University of Defense Technology, Changsha 410073)

Abstract This paper presents an efficient background subtraction algorithm using multiple scene models to cope with variations of noises in a background. A mechanism has been developed to add and delete scene models so that the distribution of the models is adaptive to the background characteristics. The calculation for the model parameters has been optimized so as to avoid time-consuming floating point calculation. We introduced the living time and recurrent frequency to the models so that the algorithm can suppress high frequency background noises effectively by controlling the model recurrent frequency. Experiments using video data have been conducted to compare the performance of our algorithm with that of the mixture Gaussian model algorithm. The experimental results demonstrated that our algorithm can extract the foreground contour more precisely, efficiently and with less memory, while maintaining the advantages of the mixture Gaussian model algorithm. It was also found that high frequency noises that cannot be rejected by the mixture Gaussian model can be suppressed.

Keywords VSAM (video surveillance and monitoring), background subtraction, mixture gaussian model, fast algorithm

1 引言

随着摄像设备价格下降、计算机计算能力增强、网络速度的提高, 大量视频监控设备已经在日常生活得到普及。视频监控相关技术的研究也逐渐成为计算机视觉研究领域中的热点。

视频监控中遇到的首要问题往往是提取图像序

列中感兴趣的部分, 即前景, 例如移动的人、车辆等等, 也可以称为运动检测。从 20 世纪末至今, 这方面技术的研究相当广泛, 同时产生了大量的文献。通过对这些文献的分析, 可以看出, 针对固定场景的运动检测问题, 根据对前景区域特征定义的不同, 可以将这些方法大致分为两类: 一类是基于背景差的方法, 出发点是维持监控场景的背景模型, 判定当前图像中与背景差别较大的部分为前景; 另一类是基

基金项目: 国家自然科学基金项目 (6033401Q, 60475029)

收稿日期: 2007-06-20 改回日期: 2007-12-01

第一作者简介: 梁华 (1977~), 男, 国防科学技术大学信息与通信工程专业博士研究生。主要研究方向为视频监控相关技术。E-mail

lh2078@163.com

于光流场的方法,出发点是图像序列中具有明显运动的较大面积区域作为前景。近年来也有一些学者将形态学滤波应用于图像序列的时间、空间结构上^[1,2]。

基于背景差的方法,概念清晰,不需要费时的光流场计算,速度比较快,所以此类方法的研究比较多。但随着研究不断的深入,对算法的要求也在不断的提高。特别是对较复杂场景情况下的前景检测,例如,光照变化引起的背景改变(室外的环境光、室内的灯光等),高噪声的背景区域(树木、水面、风中的旗的反复运动等),背景的改变(车辆停止、背景中物体搬动等),前景物体阴影的去除问题等,都成为研究中必须考虑的问题。

较早期的背景差方法有均值、中值、运动平均、运动高斯模型等方法。从 20 世纪末至今,出现了一批比较成熟的背景差方法。1999 年,Stauffer 等人提出了基于混合高斯模型的背景差方法^[3,4],针对高噪声的复杂背景,采用多个高斯分布描述像素过程;2000 年,Elgammal 等人提出了基于无参数模型的背景差方法^[5],该方法利用核密度估计,产生比多高斯混合分布更准确的无参分布描述像素过程,Oliver 等人提出背景特征化建模的方法^[6],对较长时间内的背景图像数据进行主分量分析,利用主特征矢量张成的子空间描述背景,用当前图像与子空间的距离作为判决前景的依据;2004 年,Han 等人在无参数模型的基础上,提出利用序列核密度进行近似的方法^[7],在效果与无参估计基本相近的基础上提高了算法的速度。

在实际监控系统的研究中,背景差往往作为像素层的方法,和区域层的许多方法相结合。1999 年, W^4 系统^[8]中采用的是最大最小值双门限背景差方法;2000~2001 年佛罗里达中央大学的 KNIGHT 系统^[9~11]中,将混合高斯模型检测出的结果与梯度检测结果结合,对剔除背景中由于物体移动和光照造成的错误前景以及前景中的阴影有良好的效果;2003 年,意大利摩德纳大学的 SAKBOT 系统^[12,13]中,提出了基于知识的方法,在中值背景差方法的基础上,通过对区域的面积、显著性和区域运动特征的分析,将前景中的阴影、鬼影以及鬼阴影区分出来。2004 年,IBM 的 Smart 系统中提出了显著性运动检测的方法^[14],对时间差分图像中较明显的区域计算光流场,将运动较明显的区域作为前景。

从背景差方法的发展可以看出,有两个主要方

向:一是对像素过程的建模不断完善,并不断发展快速算法;另一个是在基本背景差算法的基础上结合其他信息,提高检测的鲁棒性。由于算法复杂度不断增加,而实际视频监控系统对实时性又不断提高,所以一个快速的背景差算法往往是决定监控系统性能的关键。

经过分析发现,视频监控场景中多数区域的背景模型用单高斯模型近似是准确的,只有少数区域需要用多个高斯模型近似,而且模型的数量往往需要根据场景的变化而改变。所以本文提出了自适应多模快速背景差算法。

2 自适应多模快速背景差算法

对于视频监控中固定的场景,如果场景中的像素来自于固定光照条件下的一个物体表面时(即静态的背景),用一个高斯模型描述该像素过程以及图像噪声是比较合理的。而在实际监控的场景中,往往某些区域会处于多个物体表面、边缘交替变化的情况,针对这些区域,需要采用多个高斯模型近似描述。当场景的光线缓慢变化时,用参数能够自适应的高斯模型描述是合理的,但是当光线快速变化或者场景快速改变时,参数来不及适应快速变化,这时就需要删除旧模型,产生新的模型来适应场景变化。

图 1 是在不同图像序列中固定某一位置像素观测其在一段时间内蓝绿分量分布的状况,其中横轴为蓝色分量,纵轴为绿色分量。由图 1(a)可见,坐标(200, 20)处的地面和(150, 220)处的天空的像素颜色变化较小,分布用单模近似比较准确,而(150, 120)处水面的像素颜色变化较大,需要用多模近似。图 1(b)中草地(150, 30)和路面(130, 80)中像素的分布用单模近似准确,而树叶区域的像素需要用多模近似。图 1(c)中室内灯光不变时墙壁(280, 180)上像素的分布用单模近似准确。图 1(d)中关灯前后墙壁(280, 180)中像素的分布变化剧烈,必须用不同的模型进行近似。

可见,像素颜色分布在同一场景中存在不同的情况,有些区域只需要用一个高斯模型就可以准确的近似,而有些区域如图中的水面、树叶等区域需要用多个高斯模型描述。在室内场景中由于灯光变化,像素的分布发生了快速的变化,变化前后需要用不同的模型描述。针对这种情况,可对每个像素建立多个数量不定的高斯模型,并对高斯模型建立产

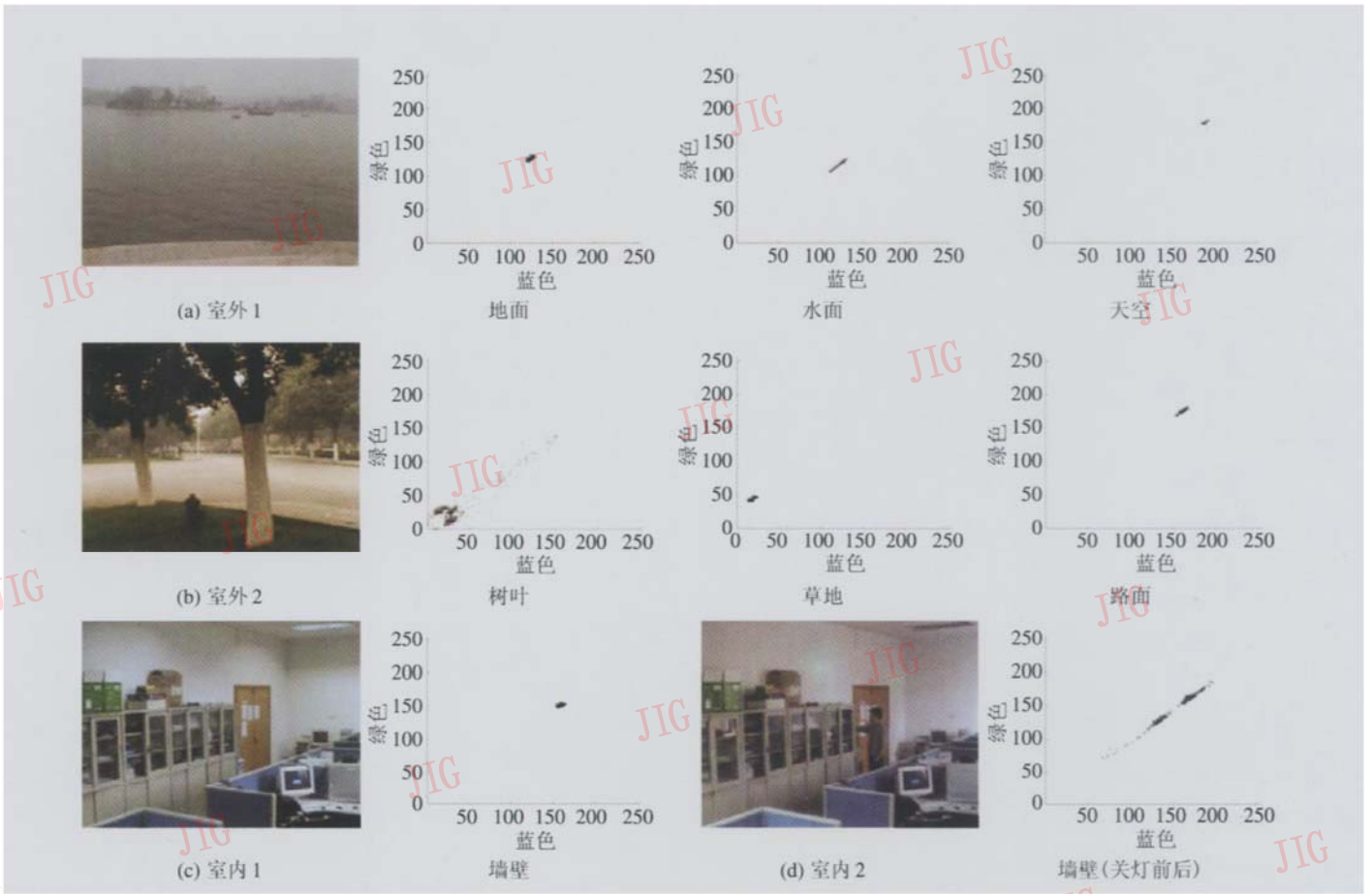


图 1 图像序列中像素颜色分布

Fig 1 Distribution of pixel process in image sequences

生和退出的机制, 让模型的数量和参数能够实时变化, 自动适应场景的不同特点。这样即能够准确描述复杂的像素过程, 又可以减少模型的总量。

设图像为 I 像素为 x , 时间为 t 每个像素 x 的过程用 N 个高斯模型 M 近似:

$$\{M_1, M_2, \dots, M_n, \dots, M_N\} \quad 1 \leq n \leq N \quad (1)$$

设每个高斯模型的均值为 μ 权值为 w , 生存时间为 l 重现频率为 f 。

首先用前若干帧图像 (一般取 255 单位) 对模型进行初始化。初始化时用第 1 帧图像为每个像素建立一个高斯模型, 模型参数 μ 设置为像素的值 (灰度或色彩), w 为 1, l 为 255 单位, f 为 0。后续图像的像素与这个模型进行匹配。像素 x 在 t 时刻与第 n 个模型的匹配定义为

$$|I_t(x) - \mu_n(x)| \leq Th_t \quad (2)$$

其中, $I_t(x)$ 为图像 I 在 t 时刻像素 x 处的像素值, $\mu_n(x)$ 表示 t 时刻像素 x 处第 n 个模型的均值, Th_t 为 t 时刻模型匹配的门限, 由下式决定:

$$Th_t = (\max(I_t) - \min(I_t)) / DIV \quad (3)$$

模型匹配的门限与当前图像中色彩变化的幅度

成正比, 其中 $\max(I_t)$ 和 $\min(I_t)$ 分别是 t 时刻图像 I 中像素值的最大值和最小值, DIV 是比例系数, DIV 值越大, 门限越低, 模型的分辨率越高, 当然也会产生较多的模型。由于监控场景的图像受光照的变化, 摄像机曝光强度、自动增益控制、白平衡等因素的影响, 整体的亮度和色彩往往会发生变化, 如果采用相同的门限, 在图像较暗时检测的分辨率会下降。采用随时间变化的门限值, 每间隔一段时间进行更新, 可以保证在不同情况下都有较好的分辨率, 所以, 这样设置门限值是合理的。

如果 t 时刻像素 x 与第 n 个模型匹配, 则模型参数更新如下:

$$w_{t+1n}(x) = w_{tn}(x) + 1 \quad (4)$$

$$\mu_{t+1n}(x) = \mu_{tn}(x) + \text{sgn}(I_t(x) - \mu_t) \quad (5)$$

$$\begin{cases} f_{t+1n}(x) = f_{tn}(x) + 1 & l_{tn}(x) \neq 255 \\ f_{t+1n}(x) = f_{tn}(x) & l_{tn}(x) = 255 \end{cases} \quad (6)$$

$$l_{t+1n}(x) = 255 \quad (7)$$

对于不匹配的模型, 参数更新如下:

$$w_{t+1n}(x) = w_{tn}(x) \quad (8)$$

$$\mu_{t+1n}(x) = \mu_{tn}(x) \quad (9)$$

$$f_{t+1n}(x) = f_{tn}(x) \quad (10)$$

$$l_{t+1n}(x) = l_{tn}(x) - 1 \quad (11)$$

其中, $w_{tn}(x)$, $\mu_{tn}(x)$, $f_{tn}(x)$, $l_{tn}(x)$ 分别为 t 时刻像素 x 处第 n 个模型的权值、均值、重现频率和生存时间, $\text{sgn}(x)$ 是符号函数。如果像素没有找到匹配的模型, 那么就产生新的模型, 新模型的初始化设置同第 1 帧。

在初始化过程完成后, 算法加入模型退出的机制。当模型的权值 w 与生存时间 l 都为 0 时将该模型删除。模型的匹配和建立以及参数的更新与式 (1)~式 (11) 基本相同, 只是对权值的更新有所变化, 如果匹配, 则

$$w_{t+1n}(x) = w_{tn}(x) + N_t(x) - 1 \quad (12)$$

如果不匹配, 则

$$w_{t+1n}(x) = w_{tn}(x) - 1 \quad (13)$$

式中, $N_t(x)$ 是 t 时刻像素 x 处模型的数量。

模型的产生、匹配、更新、退出的过程中, 像素 x 的模型权值的总和不变, 反复匹配的模型的权值不断增大, 而没有匹配的模型的权值不断减小, 权值的分配不断向匹配次数较多也就是更像背景的背景模型集中。当模型匹配时, 模型的均值随前景像素值发生改变, 能够使模型适应背景的缓慢变化。模型没有匹配时生存时间倒数计时, 匹配后重新开始计时。同时, 重现频率在模型生存时间内累计模型重复匹配的次数。只有当模型权值和生存时间均为 0 时才退出。因为, 有些背景模型存在时间短, 但是会反复出现 (例如树叶摆动反复覆盖的位置), 导致权值很小但是重现频率比较高, 所以, 不能仅根据权值判断模型是否退出。这样设计在一定程度上使背景模型能够容纳某些混合高斯模型无法容纳的高频噪声, 从而能够对这些噪声产生有效的抑制。

选择权值比较大或者重现频率比较高的模型的均值表示背景。如果像素没有找到匹配的模型, 或者虽然匹配但是匹配的模型权值较小且重现频率比较低时, 判决为前景, 模型权值和重现频率门限设置如下:

$$\begin{aligned} w_{tn}(x) &< 255N_t(x) \\ f_{tn}(x) &< F \end{aligned} \quad (14)$$

式中, 重现频率门限 F 需要根据系统的帧速率、模型生存时间长度和噪声的频率确定。

3 实验结果与分析

从 NUDT 视频和 PETS2001 视频中选取了 4 段

视频数据做实验对比。实验计算机 CPU 为 Celeron M 1.5GHz Windows XP 操作系统, 图像、视频处理平台为 Direct9.0 和 Intel OpenCV。实验中, 混合高斯模型的参数^[3]设置如下: 像素模型个数为 3 模型权值的学习速率 α 取 0.005 单位, 均值的学习因子 ρ 为 0.01, 初始化门限 σ 为 7 (等效的匹配判决门限为 $2.5\sigma = 17.5$), 背景模型权值比例设置为 0.6。自适应多模算法中判决门限系数 DIV 初始化为 13 (根据实验中场景的颜色幅度范围, 等效的匹配判决门限 Th_i 约为 17.5), 重现频率门限 F 设置为 12。这样的参数设置使二者的分辨能力和模型均值的变化速度基本相当, 实验结果对比有可信度。由于自适应多模算法中涉及模型的添加和删除, 所以, 算法实现中采用了结构链表的数据结构方式。对检测出的前景像素进行形态学开运算, 消除噪声。

实验结果如图 2 所示, 分别为混合高斯模型算法和自适应多模算法的实验结果、耗时以及后者模型数量分布图和模型总量统计。图 2 中, (a)、(b) 为 NUDT 校园视频数据的实验结果, (c)、(d) 为 PETS2001 DATASET1 视频数据的实验结果。图 2(a) 中两种算法对湖面背景的噪声抑制都比较好, 但后者的前景区域更准确, 误报率低。图 2(b) 中前者结果中对树叶摇摆的区域出现误检, 后者能够有效的抑制。图 2(c) 中后者模型分布主要集中在行人、车辆密集的区域。图 2(d) 中场景光线较暗, 汽车灯光扫过的区域出现模型数量的显著变化。

从场景中行人和车辆的检测效果来看, 自适应多模算法检测出的前景区域外形更清晰, 背景噪声相对较少。这说明该算法识别率高, 误检率低, 对场景中的噪声抑制能力强。原因来自两个方面, 一方面是由于混合高斯模型的模型数量人工设定, 模型冗余时会造成漏检, 而模型过少时会造成误报。而自适应多模算法中模型数量能适应场景特点变化, 对场景的描述更准确; 另一方面由于自适应多模算法中引入了模型重现频率的判决, 这样算法对高斯模型无法准确描述的高频噪声 (如树叶摆动) 也有很好的抑制作用。

从算法的执行速度来看, 在一般的场景复杂度情况下, 自适应多模算法的执行速度远远高于混合高斯模型算法, 原因也来自两个方面: 一方面是由于自适应多模算法的模型数量比较少; 两种算法的速度很大程度上取决于背景模型中模型总量, 混合高

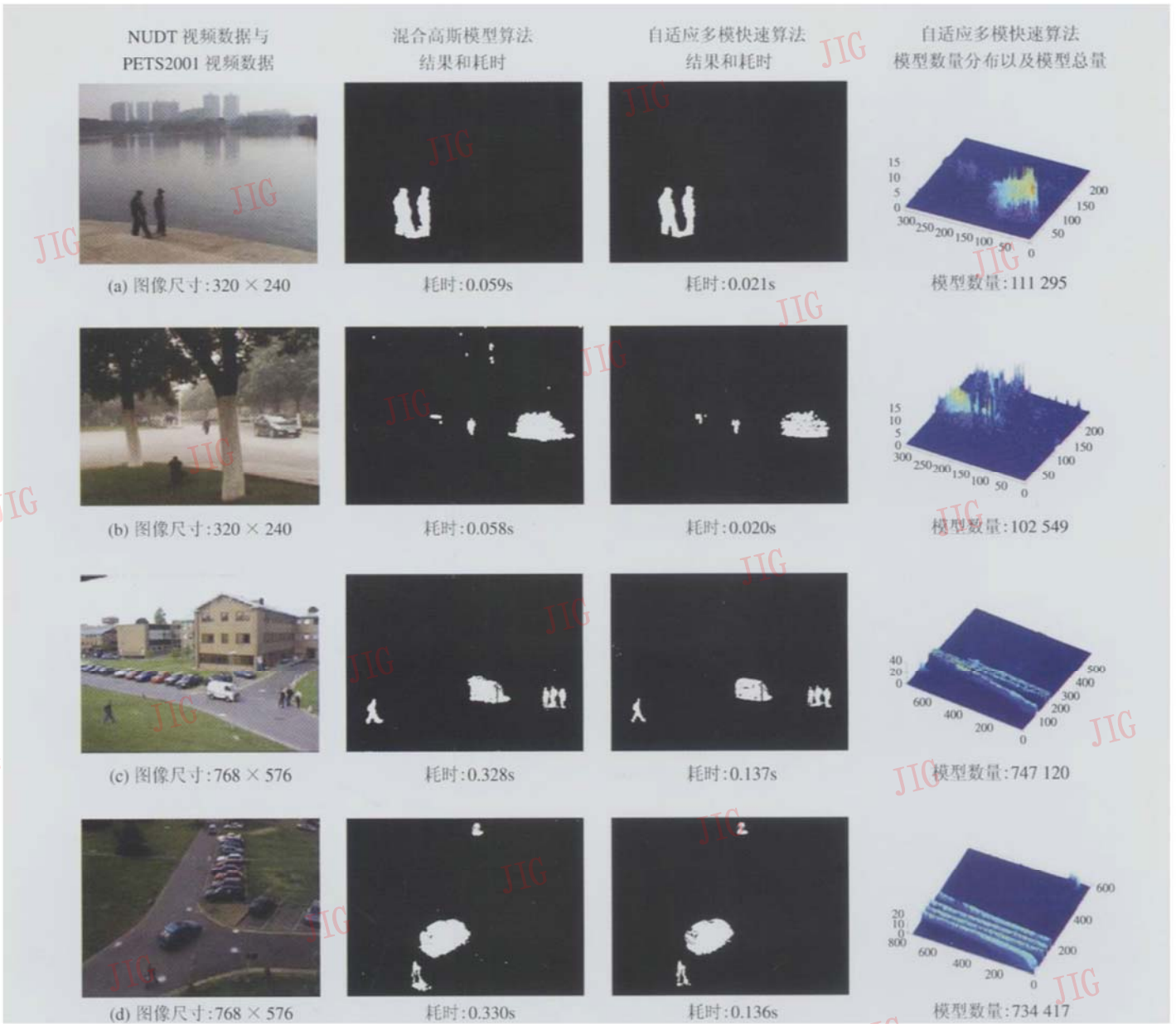


图 2 实验结果

Fig 2 Experiment results

斯模型算法的模型总量设定后执行过程中不改变 (如果像素的模型个数设为 3 那么 320×240 图像的模型总量为 230 400, 768×576 图像的为 1 327 104), 而自适应多模算法的模型数量随场景的变化而变化, 从对图像帧 (一般复杂度的场景, 背景较复杂, 前景面积占一定比例的图像帧) 的模型统计数据可以看出, 模型数量比前者减少一半左右 (4 幅实验图像中模型总量分别为前者的 48%、46%、56%、55%), 由于模型数量小, 所以, 算法内存消耗也小于前者; 另一方面是由于自适应多模算法中对数值运算做了优化, 将混合高斯模型中耗时的浮点数乘除运算和排序运算转化成了整型的加减运算和比较运算, 即使在模型数量相当的情况下, 自适应多模算法的速度也能高出混合高斯模型 20% 以上。

4 结 论

背景差算法是视频监控系统中的关键技术, 而算法的速度往往决定算法的实用价值。通过对相关文献和实验数据的分析可以看出, 当前的主流算法在普通 PC 机计算能力的条件下, 速度仍然较慢, 距离在实际系统中的应用还具有一定的差距。

本文在像素过程多高斯模型描述的基础上, 从场景中模型分布不均匀性出发, 提出了一种自适应多模快速背景差算法。该算法针对混合高斯模型中固定模型数量的不足, 建立了模型产生和退出的机制, 使模型数量能够自动适应场景特点, 实现了模型的实时自适应分布, 有效地减少了模型的数量; 针对

混合高斯模型中计算量大的问题,对模型参数计算进行优化,将耗时的浮点数乘除运算和排序运算转化成了整型数的加减运算和比较运算,并运用结构链表的数据结构方式,减少了计算量;针对多高斯模型对高频噪声描述的局限性,模型中引入了生存时间和模型重现频率,通过对模型重复出现频率的控制,有效抑制了混合高斯模型无法准确描述的高频噪声。通过与混合高斯模型算法实验比较说明,该快速算法内存消耗小,计算速度快,总体耗时在原算法的一半以下,检测结果准确,抑制高频噪声的能力强,整体效果优于混合高斯模型算法。由于该算法中没有浮点运算,所以该算法对某些浮点运算能力较弱的小型系统环境适应能力更强。

混合高斯模型算法在当前背景差算法中就其准确性而言并不是最好的,但其概念清晰、实现简单,在综合内存消耗、计算耗时等性能后,仍不失为一种优秀的算法,并且在许多系统中得到广泛的应用。所以与该算法的实验结果对比是有意义的。

参考文献 (References)

- 1 Manzanera A, Richefeu J. A robust and computationally efficient motion detection algorithm based on Σ - Δ background estimation[A]. In Proceedings of Indian Conference on Computer Vision, Graphics and Image Processing[C], Kolkata, India, 2004: 46~51.
- 2 Richefeu J, Manzanera A. A new hybrid differential Filter for motion detection[A]. In Proceedings of International Conference on Computer Vision and Graphics[C], Warsaw, Poland, 2004: 22~24.
- 3 Stauffer C, Grimson E. Learning patterns of activity using real-time tracking[J]. IEEE Transactions on Pattern Recognition and Machine Intelligence, 2000, 22(8): 747~757.
- 4 Stauffer C, Grimson W E L. Adaptive background mixture models for real-time tracking[J]. Computer Vision and Pattern Recognition, 1999, 2: 246~252.
- 5 Elgammal A, Hawwood D, Davis L. Non-parametric model for background subtraction[A]. In Proceedings of the 7th IEEE International Conference on Computer Vision Frame-Rate Workshop[C], Kerkyra, Greece, 1999: 246~252.
- 6 Oliver N M, Rosario B, Pentland A P. A bayesian computer vision system for modeling human interactions[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22: 831~843.
- 7 Haritaoglu I, Hawwood D, Davis L S W. Real-time surveillance of people and their activities[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8): 809~830.
- 8 Javed O, Shah M. KNIGHT^M: A Real Time Surveillance System[OL]. <http://www.cs.ucf.edu/~vision/papers>
- 9 Javed O, Shafique K, Shah M. A hierarchical approach to robust background subtraction using color and gradient information[A]. In Proceedings of IEEE Workshop on Motion and Video Computing[C], Washington, DC, USA, 2002: 22~27.
- 10 Javed O, Shah M. Tracking and object classification for automated surveillance[A]. In Proceedings of the Seventh European Conference on Computer Vision[C], Copenhagen, Denmark, 2002: 343~357.
- 11 Javed O, Rasheed Z, Alatas O, et al. KNIGHT^M: A real time surveillance system for multiple overlapping and non-overlapping cameras[A]. In Proceedings of IEEE International Conference on Multimedia and Expo[C], Baltimore, Maryland, USA, 2003.
- 12 Cucchiara R, Grana C, Piccardi M, et al. Detecting moving objects, ghosts and shadows in video streams[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(10): 1337~1342.
- 13 Cucchiara R, Grana C, Piccardi M, et al. Detecting objects, shadows and ghosts in video streams by exploiting color and motion information[A]. In Proceedings of IEEE International Conference on Image Analysis and Processing[C], Palermo, Italy, 2001: 360~365.
- 14 Tian Y L, Lu M, Hampapur A. Robust and efficient foreground analysis for real-time video surveillance[A]. In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition[C], San Diego, CA, USA, 2005: 1182~1187.

基于概率统计自适应背景模型的 运动目标检测方法

徐东彬^{1,2)} 刘昌平¹⁾ 黄磊¹⁾

¹⁾ (中国科学院自动化研究所, 北京 100080) ²⁾ (中国科学院研究生院, 北京 100080)

摘要 为了更好地进行运动目标检测, 提出了一种基于概率统计的自适应背景模型的运动目标检测方法。该方法能够自适应选择背景和前景阈值, 且不需要进行训练, 而且在不同的场景下能够自适应进行选择。在此基础上, 针对盲目更新和选择更新不足, 还采用了像素聚类统计和概率相结合的背景更新模型, 因为采用基于像素统计的更新机制能够适应场景中背景的局部改变(移入/移出物体), 而采用概率更新则能够降低前景污染背景的程度。实验证明, 该方法能够得到可靠的背景, 改善了运动检测效果。

关键词 运动检测 自适应背景 概率更新 统计信息

中图法分类号: TP391.4 **文献标识码:** A **文章编号:** 1006-8961(2008)02-0351-08

Adaptive Background Model for Motion Detection Based on Statistic of Probability

XU Dong-bin^{1,2)}, LU Chang-ping¹⁾, HUANG Lei¹⁾

¹⁾ (Institute of Automation, Chinese Academy of Sciences, Beijing 100080)

²⁾ (Graduate School of Chinese Academy of Sciences, Beijing 100080)

Abstract This paper proposed an adaptive background model for motion detection based on statistic information of probabilities. The approach can select thresholds of foreground and background adaptively and adapt to different scenes without training samples and human concerned. By using statistic information of historical pixels to update background, can deal with moved/inserted objects in background. Meanwhile, a background model updated according to statistical characteristic is also provided in this paper. It gets ideal background and good detection results. Experimental results demonstrate the proposed algorithms can get relative good background and improve detection results for different scenes.

Keywords motion detection, adaptive background, updated by probability, statistic information

1 引言

实时检测和分割出运动目标, 是计算机视觉中的重要研究内容, 其可以应用到安全监控、交通监测、行为识别和人机交互等领域。运动目标检测方法主要分为以下几种: 基于时间差分^[1]、基于背景减除^[2-7]、基于统计特征^[8]、基于光流等。其中背景减除技术(background subtraction)是在摄像机静止条件下应用广泛的一种运动检测技术, 即根据当前帧和背景帧的差异来检测运动目标。

研究人员提出了不同的运动目标检测方法^[3], 其中时间平均(time average)^[4]是最简单的方法, 它对内存要求低, 且处理速度快, 但这类方法对于场景中存在多个运动目标或目标运动缓慢时, 前景会污染背景。Chien Shao-Yi等人利用帧间差分(inter-frame difference)信息^[5]来生成背景, 该方法虽能够应用于多个运动目标的场合, 但更新率不容易选择。Stauffer和Grinson用 K 个高斯混合分布(Gaussian mixture model, GMM)对每个像素进行建模^[2], 解决了像素的多峰分布和缓慢变化的背景。Zoran Zivkovic提出了一种改进方法来自适应选择高斯模型的数目, 这在一

收稿日期: 2007-04-25 改回日期: 2007-12-06

第一作者简介: 徐东彬(1976~), 男, 在读博士研究生。主要方向为智能监控、视频行为分析。E-mail: dongbin_xu@ia.ac.cn

一定程度上提高了运算速度^[9]。但是由于这类方法假设像素符合高斯分布,而实际像素分布可能不满足该假设。因此对于频繁变化的像素,需要多个高斯混合分布才能反映这种变化。Kin 等人根据对像素的统计,通过建立 Codebook 来表示压缩背景^[6],但是该法的检测效果对参数的依赖性较强。Elgamma 等人提出了基于内核密度估计(kernel density estimation, KDE)的非参数背景模型^[7],该方法能够适应不同的场景,不同于 GMM,由于它充分利用了最近的历史帧信息来更新背景模型,不仅能够适应复杂的像素分布密度,而且能够克服像素值在短时间内发生的频繁变化,因此能够得到较准确的估计,但是由于该方法通过给定的假正率(percentage of false positives)来选择阈值,这需要有场景的先验知识,因此对于不同的场景需要人工干预,重新选择。Anurag Mittal 和 Nikos Paragios 根据设定的误报率和漏报率,通过训练样本调整阈值^[10],使用该阈值在给定的条件下虽能够获得相对好的分类结果,但由于该方法采用单一阈值,其很难解决误报率和漏报率之间的矛盾,因此不可避免会带来分类误差。特别是,当场景改变时,还需要重新选择样本进行训练,才能得到适用于该场景的阈值。为了克服上述方法的不足,本文提出了一种自适应选择前景和背景阈值的方法。同时针对背景更新机制提出了基于像素聚类统计和概率选择性更新相结合的更新机制,用来更新发生显著变化时(如物体的移入、移出等)的背景和缓慢变化的背景,以克服由于盲目更新而带来的前景目标漏检测和由于选择更新而带来的死锁。本文系统框图如图 1 所示。

2 内核密度估计运动检测方法

内核密度估计方法是模式识别领域的一种应用广泛的非参数密度估计方法,该方法能够估计任何形式的概率密度。基于内核密度估计的运动检测方法是采用最近历史帧数据来估计当前像素的概率密度,由于其能够适应复杂背景的场景,因此表现出良好的检测性能^[3]。假设视频帧中有 M 个像素点,每个像素点有 N 个背景样本,则在 t 时刻视频帧中第 i 个像素的像素值为 $x_{t,i}$ 与该像素对应的第 j 个背景样本的像素值为 $x_{t,i,j}$, $i=1, \dots, M$; $j=1, \dots, N$, 则 t 时刻视频中第 i 个像素的概率 $P(x_{t,i})$ 可以通过下式进行估计:

$$P(x_{t,i}) = \frac{1}{N} \sum_{j=1}^N K(x_{t,i} - x_{t,i,j}) \quad (1)$$

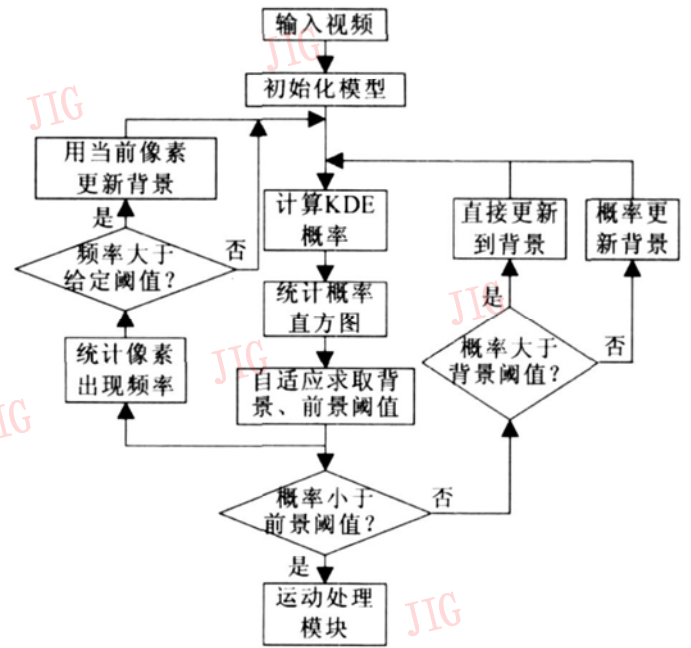


图 1 系统框图

Fig 1 System architecture

其中, K 为核估计子, 如果 K 取正态分布, 式 (1) 变为

$$p(x_{t,i}) = \frac{1}{N} \sum_{j=1}^N \prod_{m=1}^d \frac{1}{\sqrt{2\pi}\sigma_{im}} e^{-\frac{(x_{t,i,m} - x_{t,i,m,j})^2}{2\sigma_{im}^2}} \quad (2)$$

其中, d 为特征维数, σ_j 为相应的核宽, $x_{t,i,m}$ 表示第 i 个像素的第 m 个特征, 采用式 (2) 得到的概率如果满足式 (3), 则相应的像素点处于运动状态, T 为整幅图像下的全局阈值。

$$P(x_{t,i}) < T \quad (3)$$

文献 [7] 的背景模型采用的是 Long-term 模型和 Short-term 模型, 在 Long-term 模型中采用盲目更新机制, 即不管分类的结果, 都以速率 W/N 把最近的采样直接更新到背景中去, 其中 N 为采样点数, W/N 为更新率。在 Short-term 模型中则采用选择更新机制, 只有检测为前景的像素才被认为属于背景模型, 即当来自于两个模型的综合结果的掩码 $m(x_{t,i})$ 取 1 时, 才进行更新, 其中 M_S 表示 Short-term 模型, M_L 表示 Long-term 模型。

$$m(x_{t,i}) = \begin{cases} 1 & x_{t,i} \in M_L, x_{t,i} \in M_S \\ 0 & \text{否则} \end{cases} \quad (4)$$

如式 (3) 所示, 可采用单一阈值 T 进行前景、背景的分类。如果要降低误报率, 则必定会导致漏报率的增加; 反之, 如果要降低漏报率, 则会导致误报率的增加; 因此采用单一阈值 T 很难协调这一对矛盾。由于根据误报率调整阈值, 需要场景的先验知识, 当场景改变时, 则需要人工干预, 为此, 本文从像素概率直方图分析出发, 给出了自适应双阈值选择方法。

3 自适应前景、背景阈值选择方法

尽管不同的场景运动目标差别很大,但大量实验表明, KDE 概率直方图呈现一定的规律性,而利用这种规律性就可以自适应选择前景、背景阈值。

3.1 概率直方图分析

按照式 (2) 计算图像中 M 个像素的概率,为了计算方便可将式 (2) 修改为计算 N 个采样的概率和:

$$P(x_{t_i}) = \sum_{j=1}^N \prod_{m=1}^d \frac{1}{\sqrt{2\pi}\sigma_{i_m}} e^{-\frac{(x_{t_i m} - x_{t_i m j})^2}{2\sigma_{i_m}^2}} \quad (5)$$

当核宽的最大可能取值 σ_{max} 和最小可能取值 σ_{min} 确定后,在任意时刻 t 对于 N 个采样概率之和的最大值可能取值 P_{max} 可由下式确定:

$$P_{max} = N \times (1 / \sqrt{2\pi}\sigma_{min})^d \quad (6)$$

其最小可能取值 P_{min} 不仅取决于 σ_{max} 还由式 (5) 中指数项决定,可以取 0 可以将 $[0 P_{max}]$ 乘以大于 1 的比例因子 β 量化为 $[0 \beta P_{max}]$,遍历所有像素的概率,即形成概率分布直方图 H_p 。

$$H_p(j) = \begin{cases} H_p(j) + 1 & \text{如果 } \beta P(x_{t_i}) = j \\ H_p(j) & \text{否则} \end{cases} \quad (7)$$

如图 2 所示,图 2(a) 为原始图像 A,场景中的行人处于运动状态,图 2(b) 为原始图像 B,场景中没有运动目标,图 2(c) 为图像 A 的概率直方图,图 2(d) 为图像 B 概率直方图。其中图 2(c)、图 2(d) 的水平轴表示像素概率经过量化后的大小,垂直轴表示取该概率的像素数目。

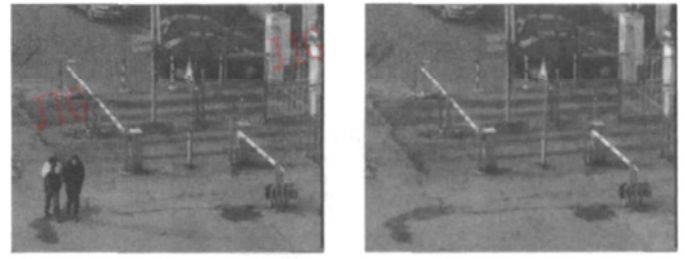
由于像素的概率分布反映了其与背景的相似程度,对照图 2(a)~图 2(d) 及式 (5) 可以得出以下规律:

- (1) 具有相同运动特性的像素应当具有相同或相近的概率;
- (2) 和背景存在较大差异的像素应当具有较低的概率,和背景相近的像素应当具有较高的概率;
- (3) 场景中包含前景目标时,概率直方图的较低的 bin 会出现峰值。反之,概率直方图的较高的 bin 会出现峰值。

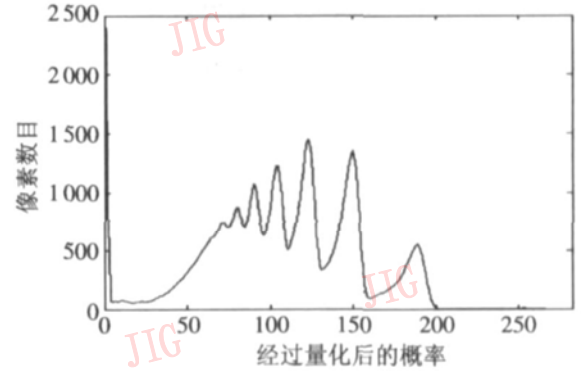
3.2 自适应前景阈值选择

如 3.1 节所示,概率直方图在背景和前景上的规律性可以用来指导选择前景和背景阈值。

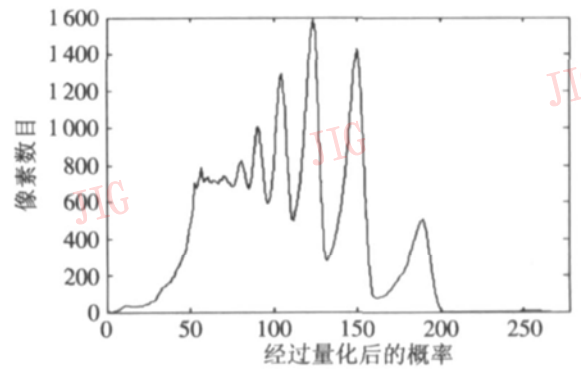
为了便于分析,本文用高斯核函数对 H_p 进行平滑得到 \hat{H}_p 。对 \hat{H}_p 相邻的 bin 按照下式进行差分,在此用小阈值 T_{small} 来进一步去除干扰,得到以



(a) 原始图像 A (b) 原始图像 B



(c) 图像 A 概率直方图



(d) 图像 B 的概率直方图

图 2 像素的概率分布直方图

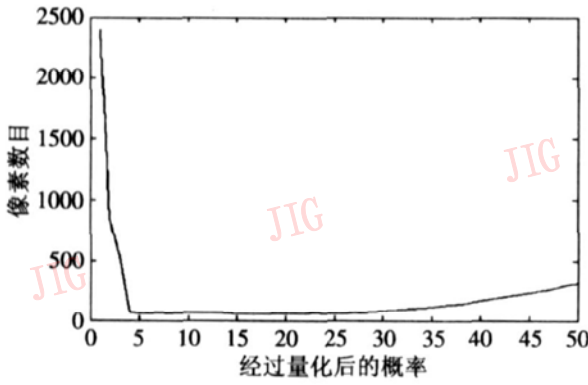
Fig.2 Probability distribution histograms of pixels

下差分直方图:

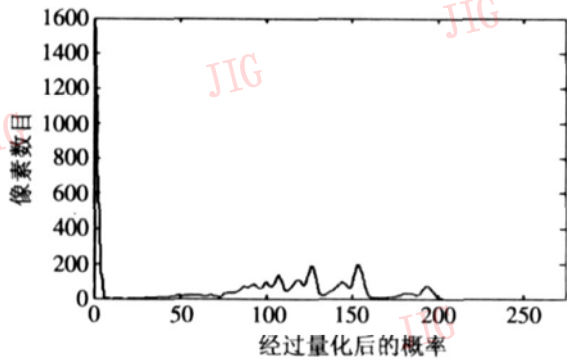
$$H_{diff} = \begin{cases} d & \text{如果 } d > T_{small} \\ 0 & \text{否则} \end{cases} \quad (8)$$

其中 $d = \text{abs}(\hat{H}_p(i+1) - \hat{H}_p(i))$ 同样用高斯核函数平滑 H_{diff} 即可得到 \hat{H}_{diff} 。

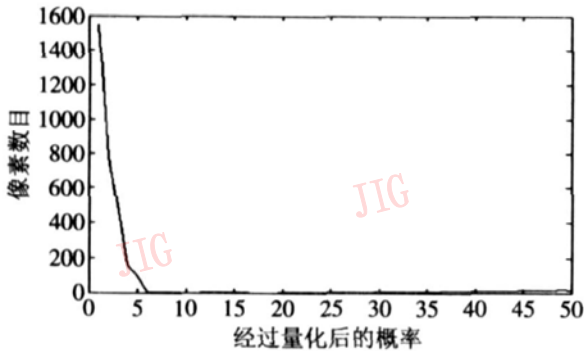
如图 3 所示,图 3(a) 为图 2(c) 的局部放大,图 3(b) 用式 (8) 计算得到的与图 2(c) 对应的差分直方图,图 3(c) 为图 3(b) 的局部放大,其中水平轴表示像素概率经过量化后的大小,垂直轴表示取该概率的像素的数目。结合图 3(a) 及图 2(c) 可以发现,直方图的变化趋势能分成如下几个部分: A 段(图中量化后概率为 0~4)出现峰值并迅速下降,且直方图 bin 数目相差很大; B 段(图中量化后概率为 4~25)该部分直方图变化不明显,直方图 bin 的数



(a) 图像 A 概率直方图的局部放大



(b) 图像 A 差分直方图



(c) 图像 A 差分直方图的局部放大

图 3 概率直方图、差分直方图及其局部放大

Fig. 3 Local histogram and differential histograms

目相差很小; C 段 (图中量化后概率为 25~300) 直方图曲线开始上升, 达到局部峰值后, 像素数目又出现多个局部峰值。

从 3.1 节分析可知, 属于 A 段的像素值的概率低, 应当属于前景像素。C 段像素概率较高, 属于背景像素, B 段介于两者之间, 其处于由一种状态向另一种状态的过渡, 这里对应于由前景到背景之间的过渡过程, 如原来是背景, 现改变为运动目标, 或原来的运动目标变为背景的一部分。为此选取如下前景阈值 T_F :

(1) 通过差分直方图 H_{diff} 寻找直方图由陡变缓的转折点 P_Z , 即 P_Z 满足下式:

$$P_Z = \arg \left| \operatorname{abs} \left| \frac{H_{diff}(i) - H_{diff}(i-1)}{H_{diff}(i+1) - H_{diff}(i)} \right| \right| \quad (9)$$

(2) 在差分直方图 H_{diff} 上从 P_Z 开始向右侧寻找变化趋势变缓的起点 P_L ;

(3) 在差分直方图 H_{diff} 上从 P_L 开始向右侧寻找缓慢变化的终点 P_R , 直到 P_{max} (为前景的概率取最大值) 止;

(4) 当 T_F 取 P_L 时会带来漏掉部分前景目标, 当 T_F 取 P_R 时会把部分背景误检测为前景, 为此做如下处理:

$$T_F = P_L + (P_R - P_L) / \lambda \quad (10)$$

其中, 引入 $(P_R - P_L) / \lambda$ 是为了降低漏报率, 同时减少检测误差, 其中 λ 为大于 1 的系数。

图 4 为取不同概率范围的分割结果, 图 4(a) $0 \leq P(x_{i,i}) < P_L$ 的结果, 从图中可以看出, 人脸部位像素出现缺失, 但噪声很小, 对应上述直方图的 A 段。图 4(b) 为 $P_L \leq P(x_{i,i}) < P_R$ 时的分割结果, 原来由运动到静止的栏杆也被检测出来, 且存在较多噪声, 即对应上述直方图的 B 段。图 4(c) 为 $P_R \leq P(x_{i,i})$ 时的分割结果, 从图中可以看出, 只有少量属于前景的像素被分到背景中。图 4(d) 为采用式 (10) 进行分类的结果, 图中人脸漏检测像素很少, 只有少量属于背景中的像素被当作前景检测出来。

3.3 自适应背景阈值选择

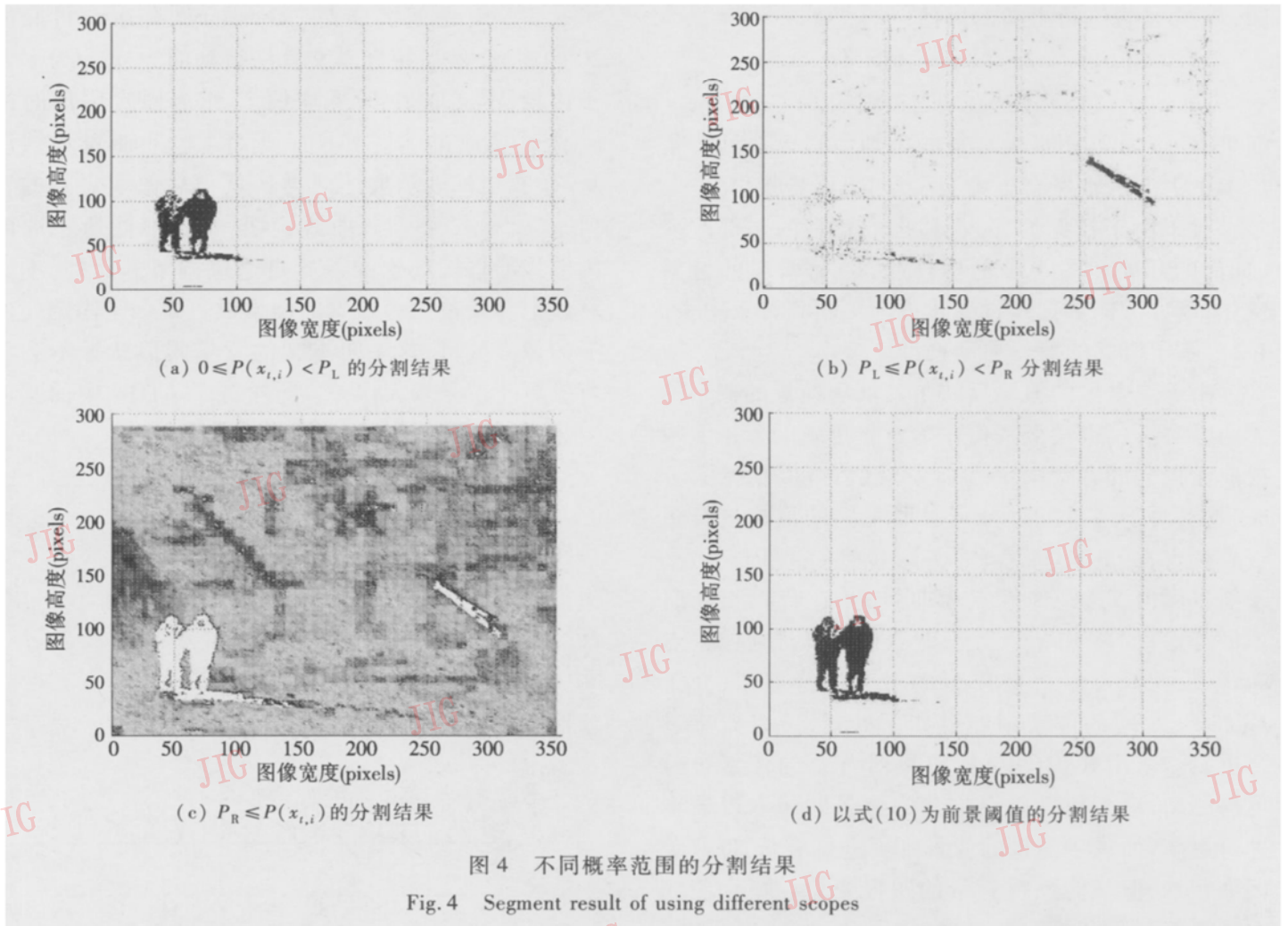
对于自适应背景阈值 T_B 的选取, 如果过大, 就会把原本属于背景的像素误判断为非背景像素, 从而更新不到背景中去, 如果选取过小, 则会把一些前景像素作为背景, 从而带来检测误差。为了分割出合理的背景, 需要适当提高判断为背景像素的阈值, 但又不能过大。因此本文把 P_R 作为背景阈值, 即

$$T_B = P_R \quad (11)$$

图 4(c) 为取 P_R 为背景阈值时的分割结果, 其中白色区域对应前景区域, 由该图可以看出被判断为背景的区域更接近于真实背景。

4 基于像素统计和概率的背景模型

对于因移入、移出物体造成的背景变化, 可通过对像素的历史统计进行更新。由于像素值的概率反应了该像素与背景点的相似程度, 因此可以用概率来更新背景模型, 就能很好地反应背景的缓慢变化。



4 1 基于像素统计的背景更新模型

当背景发生变化时(如物体的移入或移出),则该部分在 KDE 检测中被判断为前景,从而不会更新到背景中去,即会出现文献 [7] 指出的死锁情况 (deadlock situations), 文献 [7] 采用盲目更新来消除这种影响,但是由于这样会把前景更新到背景样本中,从而会带来前景的漏检测。作为背景的像素应该以较大的概率出现,受文献 [2] 的启发,本文先对像素进行分类,并统计其出现的概率,将具有较大概率的像素作为背景更新到背景样本中,这样就能够很好地解决死锁^[7]的情况。

假设 t 时刻视频中的像素为 $x_{t,i}$ $i = 1, \dots, M$, 则 $x_{t,i}$ 可以划分到 C_j 类中, $j = 1, \dots, m$, m 为最大分类数目,其最小类中心距离为 d_{min} 。算法如下:

(1) 在 t 时刻判断 $x_{t,i}$ 是否属于已有类别 C_j , 如果属于,则该类别样本数目为 $n_j = n_j + 1$, 否则按照下式调整类别中心:

$$V_j = V_j \times \alpha + (1 - \alpha) \times x_{t,i} \quad (12)$$

其中, V_j 为第 j 类的中心, α 为更新速率因子。如果类别中心发生改变,则转步骤 (4)。如果不属于任

何已有类别则转步骤 (2)。

(2) 判断当前类别总数目 N_c , 如果 $N_c < m$, 则直接建立一个新类,当前的像素值作为类别中心,属于该类别的像素数目为 1, 同时类别总数为 $N_c = N_c + 1$ 。否则转步骤 (3)。

(3) 从所有类别中找出包含像素数目最少的类别 C_{min} , 用当前像素值替换类 C_{min} 的中心,属于该类像素数目为 1。

(4) 依据下式计算现有类别的中心 V_k 与当前中心 V_j 之间的距离,进行如下判断:

$$B_{merge} = \begin{cases} 1 & \text{abs}(V_j - V_k) < d_{min} \\ 0 & \text{否则} \end{cases} \quad (13)$$

如果 B_{merge} 为 1 则进行如下合并, 否则转步骤 (5),

$$V_j = \frac{V_k \times n_k + V_j \times n_j}{n_k + n_j} \quad (14)$$

$$n_j = n_k + n_j$$

若第 k 个类别中心取值为 $V_k = 0$ 属于该类别的像素数目为 $n_k = 0$ 则当前总的类别数为 $N_c = N_c - 1$ 。

(5) 到达 T_{inter} 时刻,先求出所有类中包含像素数目最多的计数 n_{max} , 再依据下式判断是否进行更

新, T_U 为是否进行更新的阈值。

$$B_{\text{update}} = \begin{cases} 1 & n_{\text{max}} > T_U \\ 0 & \text{否则} \end{cases} \quad (15)$$

如果 B_{update} 为 1, 用最大计数的类别中心更新到背景样本中去。清空所有类中心和计数, 回步骤 (1)。

对于由于目标的移入、移出或背景的局部改变 (如门的开关和图 4 中的栏杆起落) 而引起的背景变化, 该方法能够把这种变化逐渐更新到背景中去。

4.2 基于概率的背景更新模型

由于像素值的概率反映了其属于前景的概率的大小, 因此, 可以用概率值作为背景更新的依据。根据前面得到的背景阈值和前景阈值, 可以把像素分成 3 类, 对于这 3 类可分别采取不同的策略进行更新。考虑在 t 时刻概率 $P(x_{t,i})$, 显然如果 $P(x_{t,i})$ 小于前景阈值 T_F , 则像素点为前景, 其在背景更新时不做考虑; 如果 $P(x_{t,i})$ 大于背景阈值 T_B , 则像素点为背景, 此时可以直接按照概率 1 进行更新, 介于 T_F 和 T_B 之间的像素则经过如下处理, 可以概率进行更新。采用式 (16) 更新可以对背景的像素进行可靠更新, 对于前景而误检测为背景的像素以概率进行更新, 而属于前景的像素则不进行更新, 以避免污染背景, 但又能反映背景的缓慢变化。概率值较大意味着属于背景的概率也大, 因此用式 (16) 就能够得到合理的背景。

$$P_B(x_{t,i}) = \begin{cases} 1 & \text{如果 } P(x_{t,i}) > T_B \\ 0 & \text{如果 } P_B(x_{t,i}) \leq T_F \\ \frac{P(x_{t,i}) - T_F}{T_B - T_F} & \text{否则} \end{cases} \quad (16)$$

5 实验结果及分析

实验环境为 CPU P4/3.0GHz、内存 512MB、Windows XP 系统、VC 开发环境。输入图像选择 ATON^[11] 中的视频数据为测试样本, 图像大小为 320×240 pixels。实验中背景样本为 50 帧灰度图像, 处理帧率大于 20fps。

为了评价本文的方法, 使用如下指标^[12]:

$$r_1(i) = N_1(i) / N_3(i) \quad (17)$$

$$r_2(i) = N_2(i) / N_3(i) \quad (18)$$

式中, N_1 为把背景检测为前景的像素数目; N_2 为把前景检测为背景的像素数目; N_3 为整个前景的像素数目。

实验时, 首先对样本进行标注, 然后选择有运动目标的视频帧, 计算第 i 帧图像的比例。

实验中先随机选择“laboratory”有运动目标的 20 帧视频, 然后统计本文算法得到的 r_1 、 r_2 。为了便于比较, 本文取 r_1 的平均值 \bar{r}_1 作为固定阈值时的 \hat{r}_1 , 统计相应的 \hat{r}_2 。如图 5 所示, 水平轴表示帧序号, 垂直坐标表示上述两类像素与总像素之间的比值 r_1 、 r_2 和 \hat{r}_1 、 \hat{r}_2 。该曲线说明, 本文算法所选取的阈值能够根据场景的差异自适应调节。此外, 本文算法把背景检测为前景的像素数目要低于固定阈值下的像素数目, 这说明本文的背景阈值选择不会给背景更新带来较大误差。同样对于 ATON 中的其他视频序列, 实验都能得到类似的结果。

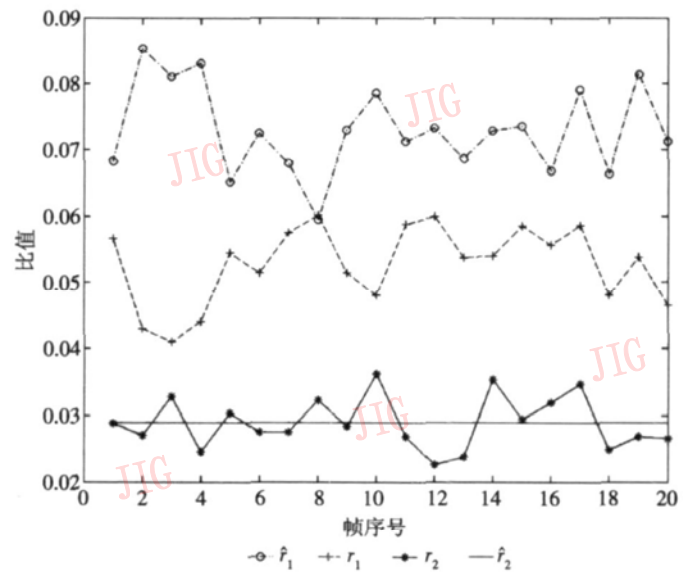
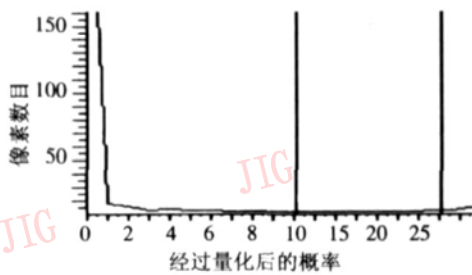
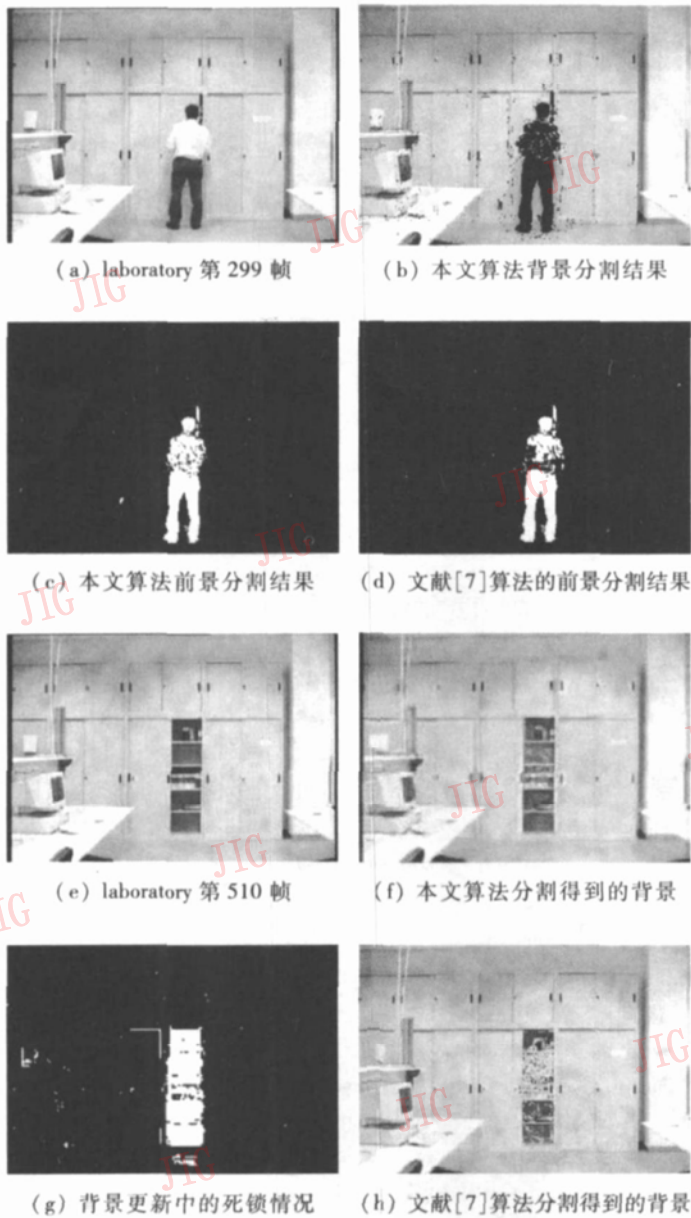


图 5 在固定和自适应阈值下比值曲线图

Fig 5 Curves of ratio at fixed and adaptive thresholds

图 6 (a) 为“laboratory”中的第 299 帧图像, 图 6 (b) ~ (c) 分别为与图 6 (a) 对应的背景分割结果和前景分割结果。图 6 (i) 为概率直方图, 为了方便显示本文对横轴做了分段拉伸, 其中从左往右第 1 条垂直于水平轴的直线所在的位置为前景分割阈值, 第 2 条垂直线所在的位置为背景分割阈值。图 6 (b) 为采用本文算法得到的背景阈值进行分割的结果。图 6 (c) 为采用本文自适应前景阈值的进行分割的结果, 从图中可看出, 只有少量前景像素被漏检测。本文采用概率更新背景, 即使前景像素被误检测为背景, 由于这些像素在概率上和前景像素概率更相近, 因此在更新中会以较小的概率更新到背景中去, 从而可减轻了前景污染背景的程度。如果属于背景的像素没有被检测为背景, 则由于这些像素的分布概率和背景像素的分布概率更接近, 因此会以较大的概率更新到背景中去, 由于本文采用概率更新背景能够得到较好的背景, 从而有利于前



(i) 图 6(a) 的概率直方图

图 6 基于像素统计和概率的背景更新

Fig. 6 Background updated by statistic of pixels and probability

景的检测。图 6(d)为采用文献[7]方法分割得到的前景,由于采用盲目更新会把前景像素更新到背景中去,因此会造成前景像素的漏检测,如图 6(d)中的行人出现部分像素的缺失。

图 6(e)为 ATON 中的第 510 帧图像。图 6(f)为采用本文基于像素统计的方法得到的背景,

图 6(g)为只采用选择更新下的分割结果,该图上门被打开后的区域一直被当作前景来检测,即会带来背景更新中的死锁情况,图 6(h)为采用文献[7]方法得到的背景。当背景像素误检测为前景时,由于采用了基于像素统计的策略来更新背景,属于背景的像素会以较大的概率出现,因此这些像素能够很快更新到背景模型中去,从而解决了背景更新中的死锁问题。如图所示,本文算法能够解决背景更新中的死锁情况,并能得到好于采用盲目更新和选择更新得到的背景。

实验中用另一段拍摄于室内多目标缓慢运动的视频来验证本文算法。场景中的目标(行人)在系统初始化阶段静止如图 7(a)所示,此时算法会把行人当作背景样本。当目标运动到图 7(b)所示的位置时,为背景的移出,则会在原来的位置留下空洞(ghost)区域(如图 7(c)中的白色方框内所示)。如

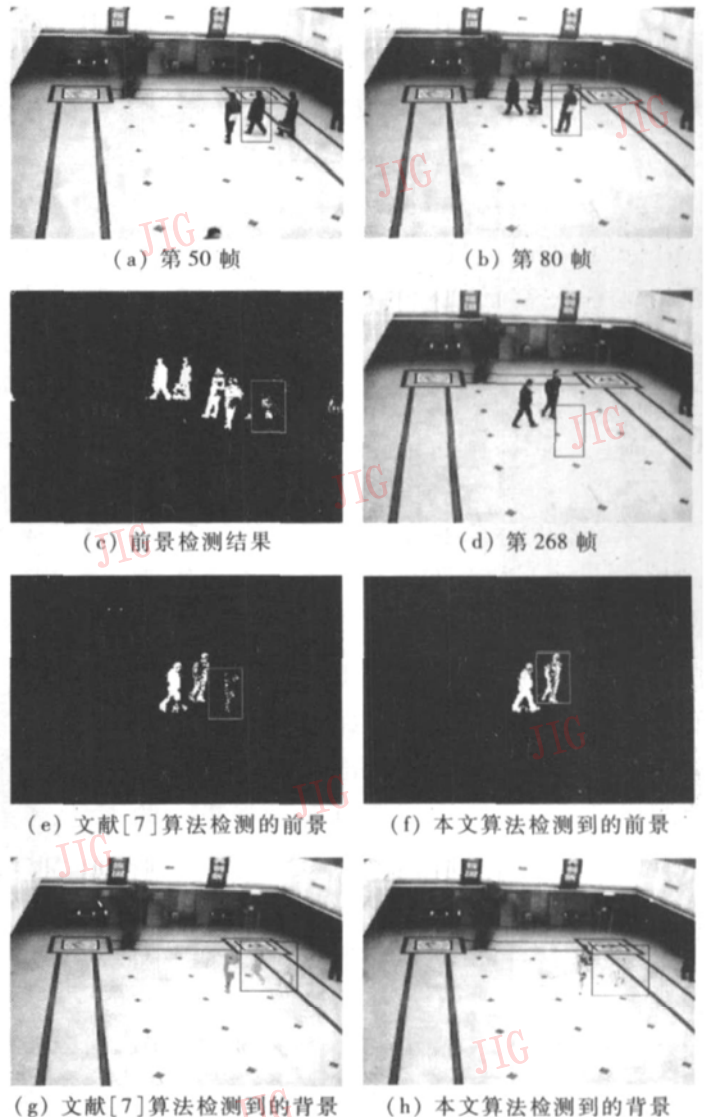


图 7 缓慢运动目标的背景和前景检测结果

Fig. 7 Results of updating background and foreground detection of slowly moving objects

果只依据选择更新,则图 7(c)中的空洞区域一直会当作前景区域,即发生死锁。图 7(d)为视频序列中的第 268 帧,图 7(e)和图 7(g)分别为文献 [7] 算法检测到的前景和背景,图 7(f)和图 7(h)分别为用本文算法分割得到的前景和背景。

对比图 7(g)和图 7(h)中的黑色方框区域,用本文算法进行检测,经过背景更新,初始背景中的行人逐渐从背景中消失,而文献 [7] 的算法则更新得不充分,这是由于本文算法采用像素统计分类的方法进行更新,能够把属于背景的像素快速更新到背景模型中去的缘故。采用文献 [7] 的算法,由于背景得不到及时的更新,原来的空洞区域仍然存在(如图 7(e)的方框所示),从而造成前景目标的误检测,而本文算法则能够解决死锁情况。

当运动目标缓慢运动时,由于盲目更新会把前景像素更新到背景中去,因此会造成前景像素的漏检测,即目标(图中方框左边的第 1 个行人)部分像素会当作背景。由于本文算法避免了盲目更新把前景更新到背景中的不足,因此前景目标很少出现在背景模型中,从而减少了前景像素的漏检测,例如图 7(f)方框内所示的目标被漏检的像素要少于图 7(e)中的对应目标被漏检的像素,这说明本文算法既能解决背景的死锁,又能够避免盲目更新的不足。

6 结 论

本文提出了一种基于概率统计的自适应背景运动检测方法。该方法通过对像素的概率直方图进行分析,提出了自适应前景、背景阈值的双阈值选择方法。这种选择阈值的方法不仅不需要对样本进行训练,也不需人工干预,而且在不同的场景下能够自适应进行。采用双阈值较好地解决了误报率和漏报率之间的矛盾。针对盲目更新和选择更新的不足,本文采用了像素聚类统计和基于概率相结合的背景更新模型。由于根据像素统计分类信息来更新背景,能够避免因盲目更新把前景更新到背景中而污染背景的问题;同时,由于结合前景和背景阈值的分类结果能把像素按照不同的概率更新到背景中,这不仅减轻了前景污染背景的程度,还能够得到更为合理的背景模型,从而有利于运动目标的检测。实验证明,本文算法由于能够得到可靠的背景,因此有利于运动目标检测。运动检测系统一般要求有较高的实

时性,但是 KDE 算法相对复杂。因此,如何提高 KDE 算法的实时性,也是本文今后的研究方向。此外,像素的概率计算是建立在历史样本基础上的,由于需要存储大量历史数据,内存消耗较大,因此在不影响算法效果的情况下,如何降低内存消耗也是需要解决的问题。

参考文献 (References)

- 1 Lipton A J, Fujiyoshi H, Patil R S. Moving target classification and tracking from real-time video [A]. In: Proceedings of IEEE Workshop on Applications of Computer Vision [C], Princeton, NJ, USA, 1998: 8~14.
- 2 Stauffer C, Grimson W E L. Adaptive background mixture models for real-time tracking [A]. In: Proceedings of Computer Vision and Pattern Recognition [C], Fort Collins, Colorado, USA, 1999: 246~252.
- 3 Piccardi M. Background subtraction techniques: a review [A]. In: Proceedings of IEEE Conference on Systems Man and Cybernetics [C], Hague, Netherlands, 2004: 3099~3104.
- 4 Wren C, Azahajehani A, Darrell T, et al. Pfinder: real-time tracking of the human body [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7): 780~785.
- 5 Chien S Y, Ma S Y, Chen L G. Efficient moving object segmentation algorithm using background registration technique [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2002, 12(7): 577~586.
- 6 Kim K, Chalidabhongse T H, Hanwood D, et al. Real-time foreground-background segmentation using codebook model [J]. Real-Time Imaging, 2005, 11(3): 172~185.
- 7 Elgammal A, Hanwood D, Davis L S. Non-parametric model for background subtraction [A]. In: Proceedings of ECCV [C], Dublin, Ireland, 2000: 751~767.
- 8 Gao D, Zhou J, Xin L. SVM-based detection of moving vehicles for automatic traffic monitoring [A]. In: Proceedings of IEEE Conference on Intelligent Transportation Systems [C], Oakland, CA, USA, 2001: 745~749.
- 9 Zivkovic Z. Improved adaptive Gaussian mixture model for background subtraction [A]. In: Proceedings of International Conference on Pattern Recognition [C], Cambridge, UK, 2004: 28~31.
- 10 Anurag Mittal, Nikos Paragios. Motion-based background subtraction using adaptive kernel density estimation [A]. In: Proceedings of Computer Vision and Pattern Recognition [C], Washington, DC, USA, 2004: 2302~2309.
- 11 Shadow detection. Autonomous Agents for On-scene Networked Incident Management (ATON) [DB/OL]. <http://cvrr.ucsd.edu/aton/shadow/index.html> 2007-6-1/2007-11-27.
- 12 Georis B, Brémond F, Thonnat M, et al. Use of an evaluation and diagnosis method to improve tracking performances [EB/OL]. In: Proceedings of the 3rd IASTED Visualization, Imaging and Image, 2003. <http://www-sop.inria.fr/orion/Publications/Articles/vip03.pdf> 2007-6-1/2007-11-27.

摄像机自由运动环境下的背景建模

金侯欣 陶霖密 徐光祐 彭玉新

(清华大学计算机科学与技术系, 北京 100084)

摘要 提出了在摄像机运动情况下使用多层 Homography 匹配算法进行背景建模的方法。该方法中, 场景可以被看作由多个平面所组成, 使用 RANSAC 方法找到场景中不同的平面, 即多层 Homography。每个像素点肯定在某个平面上, 通过所属平面相应的 Homography 变换, 就能使相邻两帧重叠视野中的像素点进行匹配, 这样就能对场景进行背景建模。实验结果表明, 该方法能有效地在摄像机运动环境中进行像素点级别的背景建模。

关键词 Homography 背景建模 对应点

中图分类号: TP391 文献标识码: A 文章编号: 1006-8961(2008)02-0359-06

Background Modeling under Free Moving Camera Environment

JIN Yu-xin TAO Lin-mi XU Guang-you PENG Yu-xin

(Department of Computer Science and Technology, Tsinghua University, Beijing 100084)

Abstract In this paper a novel Multi-layer Homography algorithm for background modeling under free moving camera environment is proposed. Background is composed of many planes. RANSAC is used to find these different planes, called Multi-Layer Homography. Each pixel definitely belongs to certain plane. Transformed by the corresponding Homography, each pixel in each frame can find its match in the subsequent frame if it occurs in the shared view of these two frames. Thus, background model can be built. Experiment shows it is effective for background modeling under free moving camera environments.

Keywords Homography, background modeling, correspondences

1 引言

理解场景中发生的事件是计算机视觉的一项基本目标。为了实现这个目标, 首先需要获得前景物体的信息; 其次才能去识别该前景物体以及理解与之相关的事件, 诸如识别人、车辆以及人的行为。在实现以上这些高层目标之前, 必须面临的问题是如何把前景物体从背景中分割出来。

显然, 如何建立一个稳定的背景模型和如何判断图像中哪些区域是前景是一项非常基础的工作。传统的背景建模方法来自于对每个像素点建立一个混合高斯模型进行描述, 如文献 [1] ~ [3], 建立起了较好的背景模型, 但他们是建立在当前帧的背景像素点能在下一帧中相同位置上找到对应, 因此, 在背景本身有一定扰动的前提下, 该假设显然不成立。

如无法适用于波纹、微风、光照变化、阴影等情况。

Mittal 等人针对动态背景, 建立了一个自适应核的密度估计模型来处理动态背景差分问题^[4]。可以很好地处理复杂背景, 但是其运算代价是十分庞大的。对此, Tian 等人在处理动态背景的同时, 采用了一些方法减少计算代价^[5], 使用对应位置上像素点领域的梯度相似性来处理光照变化问题以及用光强的协方差来处理阴影问题。

以上这些方法都是建立在假设摄像机固定情况下的, 当摄像机运动时, 找到背景像素点位置的对应关系就变成了比较棘手的问题。文献 [6] 提出了处理缓慢的摄像机运动情况下的对应点匹配问题。首先找一个最能符合两张图的 Homography 变换, 对其中一张图进行 Homography 矫正。然后将图像中每个像素点视为一个节点, 利用最小生成树 (MST) 来约束对应点的搜索空间。通过匹配对应点, 就能对

基金项目: 国家自然科学基金项目 (60673189, 60433030)

收稿日期: 2007-06-25 改回日期: 2007-12-07

第一作者简介: 金侯欣 (1982~), 男, 清华大学计算机科学与技术专业硕士研究生。主要研究方向为计算机视觉、模式识别。E-mail

jhx05@mails.tsinghua.edu.cn

这些点建立多高斯模型,从而建立一张全景图的背景模型。但它仍然存在很多不足。当背景中不同层次的平面过多时,就无法找到对应点。因为不同的背景层次有自己的一个 Homography 变换。

本文提出一种多层 Homography 匹配算法来解决以上的问题。首先,在两幅图中找到一些特征点(角点),并找到潜在的对应关系。其次,多次使用 RANSAC (random sample consensus) 的方法找到这些已经被匹配起来的角点的多个 Homography 变换,它们将对应整张图中处于不同层次背景 Homography 映射关系,从而解决层次丰富的图像序列的对应点匹配问题。实验表明,该方法有效。

2 角点检测及初匹配

运动摄像机中最重要的问题之一就是寻找对应点。而往往在纹理比较丰富的区域容易找到对应点,也就是找到一些角点。这里使用 Harris 算法寻找角点。Harris 角点检测算子是由 Harris 等人于 1988 年中提出的^[7]。它最主要的优点是具有旋转

$$s_{ij}(p_i^{(1)}, p_j^{(2)}) = \frac{\sum_{-w \leq \Delta x, \Delta y \leq w} I(x_i^{(1)} + \Delta x, y_i^{(1)} + \Delta y) I(x_j^{(2)} + \Delta x, y_j^{(2)} + \Delta y)}{\sqrt{\sum_{-w \leq \Delta x, \Delta y \leq w} I(x_i^{(1)} + \Delta x, y_i^{(1)} + \Delta y)^2} \sqrt{\sum_{-w \leq \Delta x, \Delta y \leq w} I(x_j^{(2)} + \Delta x, y_j^{(2)} + \Delta y)^2}}$$

实际上就是找两个视角中的一组点,这组点满足双向都是最优的,也就是说在点集 $p_j^{(2)}$ 中与 $p_m^{(1)}$ 最相似的点是 $p_m^{(2)}$,并且在点集 $p_i^{(1)}$ 中与 $p_m^{(2)}$ 最相似的点是 $p_m^{(1)}$ 。

当然这样做并不能保证找到的对应点就一定是正确的。尤其在纹理比较丰富(这样是保证能找到角点)并且有一大片区域中都具有相似的纹理时,会很容易找到错误的匹配。图 1 是两个视角图像之间角点的对应关系,匹配得到的对应点使用直线连



图 1 对应点初试匹配结果,存在明显错误匹配

Fig 1 Initialized Correspondences obvious mismatches exist

不变性,尺度不变性,对光照变化和图像噪声较鲁棒。Harris 角点检测算子是基于图像局部自矫正函数的,该自矫正函数衡量某个像素点在各个方向上微小偏移所造成的局部范围内的像素变化。简单而言,Harris 算子是寻找两个正交方向上的剃度变化较大的点。

使用 Harris 算子得到两个视角图像中的角点后,还需要将可能的对应点找到。假设已经得到第 1 个视角中的角点集 $p_i^{(1)} = (x_i^{(1)}, y_i^{(1)})$, $i = 1, 2, \dots, n_1$ 和第 2 视角中的角点集 $p_j^{(2)} = (x_j^{(2)}, y_j^{(2)})$, $j = 1, 2, \dots, n_2$, 找到这样一组角点 $(p_m^{(1)}, p_m^{(2)})$ 使得

$$\begin{cases} m = \arg_{i=1, 2, \dots, n_1} \max(s_{ij}(p_i^{(1)}, p_m^{(2)})) \\ j = \arg_{j=1, 2, \dots, n_2} \max(s_{ij}(p_m^{(1)}, p_j^{(2)})) \end{cases}$$

式中, $s_{ij}(p_i^{(1)}, p_j^{(2)})$ 表示点 $p_i^{(1)}$ 和点 $p_j^{(2)}$ 的相似度。当两个点之间的距离过大,则该相似度被设为 $-\infty$; 否则各取一个邻域窗口(窗口为 $(x-w, y-w)$ 到 $(x+w, y+w)$ 的一个矩形区域),窗口所涵盖的像素点的亮度 $I(\cdot, \cdot)$ 被展开为一个向量,两个点的邻域窗口构成的向量之间的相关系数即为相似度。即

接。可以明显看到,在纹理重复性比较大的树叶部分有相当多的匹配错误。

3 Homography

3.1 Homography 概念

Homography 是一种空间中某平面在两个不同的视角下成像之间的投影变换关系。给定空间中某平面上的一组点 $X_i = [x_i \ y_i \ z_i \ w_i]^T = [x_i \ h_i \ y_i / w_i \ z_i / w_i \ 1]^T$ (这里使用的是齐次坐标,空间的点被表示为 4 维向量,平面上的点为 3 维向量 $[x \ y \ w]^T = [x \ h_i \ y / w \ 1]^T$), 它们投影在第 1 视角的成像平面 $P^{(1)}$ 上的镜像为 $x_i^{(1)}$ 以及投影在第 2 视角的成像平面 $P^{(2)}$ 上的镜像 $x_i^{(2)}$ 之间通过某个 Homography 矩阵 H (3×3 的矩阵)构成投影变换 $x_i^{(1)} \leftrightarrow x_i^{(2)}$ 。因此 $Hx_i^{(1)}$ 和 $x_i^{(2)}$ 为相等的齐次坐标,即两者的差积为 0 即 $x_i^{(2)} \times Hx_i^{(1)} = 0$ 。

3.2 求解 Homography 矩阵

计矩阵 H 的第 i 行的行向量为 H^T , 有

$$Hx_i^1 = \begin{bmatrix} h^{1T} x_i^{(1)} \\ h^{2T} x_i^{(1)} \\ h^{3T} x_i^{(1)} \end{bmatrix}$$

将上式展开, 并由 $h^{jT} x_i^{(1)} = x_i^{(1)T} h^j$ 可得:

$$x_i^{(2)T} \times Hx_i^{(1)} = \begin{bmatrix} y_i^{(2)} x_i^{(1)T} h^3 - w_i^{(2)} x_i^{(1)T} h^2 \\ w_i^{(2)} x_i^{(1)T} h^1 - x_i^{(2)T} x_i^{(1)T} h^3 \\ x_i^{(2)T} x_i^{(1)T} h^2 - y_i^{(2)} x_i^{(1)T} h^1 \end{bmatrix} = 0$$

改写为

$$\begin{bmatrix} 0^T & -w_i^{(2)} x_i^{(1)T} & y_i^{(2)} x_i^{(1)T} \\ w_i^{(2)} x_i^{(1)T} & 0^T & -x_i^{(2)T} x_i^{(1)T} \\ -y_i^{(2)} x_i^{(1)T} & x_i^{(2)T} x_i^{(1)T} & 0^T \end{bmatrix} \begin{bmatrix} h^1 \\ h^2 \\ h^3 \end{bmatrix} = 0$$

$$h = \begin{bmatrix} h^1 \\ h^2 \\ h^3 \end{bmatrix}, H = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix}$$

这实际上是一个有 9 个未知数构成的线性方程组, 而这 9 个未知数就是 H 的元素。实际上, 这里的方程组中需要注意, 有一个方程是可由另外两个线性组合得到的, 所以, 一组对应点只提供了两个方程组。记

$$A_i h = \begin{bmatrix} 0^T & -w_i^{(2)} x_i^{(1)T} & y_i^{(2)} x_i^{(1)T} \\ w_i^{(2)} x_i^{(1)T} & 0^T & -x_i^{(2)T} x_i^{(1)T} \end{bmatrix} \begin{bmatrix} h^1 \\ h^2 \\ h^3 \end{bmatrix} = 0$$

其中, A_i 为一个 2×9 的系数矩阵, 是由第 i 组对应建立起来的方程。虽然 H 有 9 个参数, 但因为在齐次坐标系下, 尺度是可以任意的, 实际上只有 8 个未知参数需要求解, 并限定 $\|h\| = 1$ 。因此, 为了求解 8 个未知参数的唯一解, 至少需要 8 组线性无关的方程, 这就意味着必须要存在 4 组精确的对应点, 且每 3 个都是不共线的, 才能得到唯一解。

但在实际情况中, 找到的可能的对应点必定存在一些偏差, 这种偏差可能来自匹配算法的不精确, 也可能来自像素本身的不精确性。因为图像中像素坐标都是离散的整数, 而真实的坐标应该是连续的实数, 所以需要更多的点来减少这种偏差。当方程组数量超过未知数的数量时就成为了一个超定问题, 并且很可能由于偏差而导致方程组中出现矛盾方程。因此, 就需要通过最小二乘法来求解超定方程组。

由 n 组对应点组成的系数矩阵 A , 目标是最小化 $\|Ah\|$, 最小值所对应的 h 就是所求解。即求解特征方程 $A^T A h = \lambda h$ 的最小特征根所对应的解向量。

3.3 使用 RANSAC 算法进行鲁棒估计

前面已经介绍了寻找角点和角点匹配的方法, 也介绍了在找到对应点的情况下求解该组对应点所在平面的 Homography 变换。现在需要求解一个最能符合已经匹配的角点的 Homography 变换。最容易的做法就是直接使用这组角点的坐标建立一个超定方程组, 随后求最小二乘解。但存在两个问题: 一是对应的角点未必在同一个平面上, 这样显然会得到一个非常不精确的最小二乘解。另一个是寻找对应点的方法不能保证找到的对应点是真实的对应点, 图 1 中很显然的存在一些匹配错误的对应点, 可能对求解方程组造成毁灭性的影响。

所以, 需要一些更鲁棒的估计, 而这种估计方法能够识别出不符合所求解的对应点, 可将这类点称为外点, 相对的符合所求解的对应点称为内点。RANSAC 的方法可以解决这个问题。

给定角点及潜在的对应关系后, 使用 RANSAC 求解 Homography 变换的具体算法步骤如下, 两个视角中能符合最多对应点的 Homography 变换如图 2 所示。

(1) 使用 RANSAC 鲁棒估计: 迭代 N 次, N 由人工指定, 重复做:

① 随机采 4 组不退化的对应点作为样本, 计算 Homography 变换 H 。退化情况为 4 组点中只要存在 3 组在同一直线上。

② 对每一组对应点计算距离:

$$d^2 = \|x_i^{(1)} - H^{-1}(x_i^{(2)})\|^2 + \|x_i^{(2)} - H(x_i^{(1)})\|^2$$

这里 H^{-1} 为 H 的逆变换。

③ 确定符合当前 H 的对应点, 它们的距离应该小于某个阈值 $d_threshold$ 。这样即为内点。

当前内点数量足够多的时候, 退出循环进入步



图 2 RANSAC 找到的最佳 Homography 对应的内点

Fig 2 The correspondences conforming to the optimal Homography by RANSAC

骤 (2) 否则返回步骤 (1)。

(2) 对得到的内点做一次最小二乘法求解最终的 H 。

4 多层 Homography 提取

计算 Homography 的目的是为了将某个视角下的画面转化到另外一个视角下, 从而能够在像素点级别上进行配准, 这样就能建立背景模型。但是, 由于实际上背景的层次感较强, 通过一层 Homography 变换后进行配准的过程中, 很多点是无法在矫正后的坐标邻域中找到另外一个视角中与之相似的点。为了能让其他 (平面) 层次上的像素点做到较好的配准, 本文提出一种新的方法: 多层 Homography 提取算法。这种算法可以使得在不同平面上的点根据自己平面的 Homography 转换进行配准, 从而有效降低了像素级别精细配准时搜索邻域的大小。

4.1 多层 Homography 算法描述

实际上, 多层的 Homography 算法是基于一个非常简单的思想: 假设通过 Harris 角点匹配后得到的对应点中大多数都是匹配正确的, 而这些匹配正确的点根据所处的层次都能服从各层次的 Homography 变换, 找到一个能令最多的对应点服从的 Homography 变换后, 将这些点从对应点中去除, 并将它们的层标记为 1, 随后在剩余的对应点中继续寻找最优的 Homography 变换, 将符合这个变换的点标记为 2 并去除, 依此类推。

一般的, 通过 n 组对应点为 $\{(\mathbf{x}_i^{(1)}, \mathbf{x}_i^{(2)})\}, i = 1, 2, \dots, n$, 寻找 m 层 Homography 变换 $\{H_j\}, j = 1, 2, \dots, m$, 使得

$$\mathbf{x}_{jk_j}^{(2)} = H_j \mathbf{x}_{jk_j}^{(1)}, k_j = 1, 2, \dots, n_j, k_{out} = 1, 2, \dots, n_{out}$$

$$\{(\mathbf{x}_i^{(1)}, \mathbf{x}_i^{(2)})\} = \left(\bigcup_{j=1}^m \{(\mathbf{x}_{jk_j}^{(1)}, \mathbf{x}_{jk_j}^{(2)})\} \right) \cup \{(\mathbf{x}_{k_{out}}^{(1)}, \mathbf{x}_{k_{out}}^{(2)})\}$$

$$\forall j_1 \neq j_2 \in \{1, 2, \dots, m\}, \{(\mathbf{x}_{jk_{j_1}}^{(1)}, \mathbf{x}_{jk_{j_1}}^{(2)})\} \cap \{(\mathbf{x}_{jk_{j_2}}^{(1)}, \mathbf{x}_{jk_{j_2}}^{(2)})\} = \emptyset$$

$$\sum_{j=1}^m n_j + n_{out} = n$$

式中, $\{(\mathbf{x}_{k_{out}}^{(1)}, \mathbf{x}_{k_{out}}^{(2)})\}$ 为不服从任何 H_j 的外点集。

多层 Homography 算法的完整描述如下:

$$s(\mathbf{x}_i^{(1)}, \mathbf{x}_i^{(2)}, H_j) = \frac{\sum_{-w \leq \Delta x, \Delta y \leq w} I(x_i^{(1)} + \Delta x, y_i^{(1)} + \Delta y) \hat{I}(\hat{x}_i^{(1)} + \Delta \mu_x + \Delta x, \hat{y}_i^{(1)} + \Delta \mu_y + \Delta y)}{\sqrt{\sum_{-w \leq \Delta x, \Delta y \leq w} I(x_i^{(1)} + \Delta x, y_i^{(1)} + \Delta y)^2} \sqrt{\sum_{-w \leq \Delta x, \Delta y \leq w} \hat{I}(\hat{x}_i^{(1)} + \Delta \mu_x + \Delta x, \hat{y}_i^{(1)} + \Delta \mu_y + \Delta y)^2}}$$

第 1 层已经找到对应点的第 2 视角中的点将不会在接下去的层中寻找其在第 1 视角中的对应点,

(1) 初始化 初始化当前剩余的对应点集 $\{(\tilde{\mathbf{x}}_0^{(1)}, \tilde{\mathbf{x}}_0^{(2)})\} \doteq \{(\mathbf{x}_i^{(1)}, \mathbf{x}_i^{(2)})\}, i = 1, 2, \dots, n$ 。

(2) 计算当前最佳的 Homography 迭代 m 次:

- ① 使用 RANSAC 算法计算当前最佳的 H_j 。
- ② 找到当前的内点集 $\{(\mathbf{x}_{jk_j}^{(1)}, \mathbf{x}_{jk_j}^{(2)})\}, k_j = 1, 2, \dots, n_j$ 。
- ③ 去处这些内点得到更新后的当前剩余对应点集 $\{(\tilde{\mathbf{x}}_j^{(1)}, \tilde{\mathbf{x}}_j^{(2)})\} = \{(\tilde{\mathbf{x}}_{j-1}^{(1)}, \tilde{\mathbf{x}}_{j-1}^{(2)})\} - \{(\mathbf{x}_{jk_j}^{(1)}, \mathbf{x}_{jk_j}^{(2)})\}$ 。
- ④ 如果剩余对应点集中点数不足某个阈值 $p_threshold$, 则中止算法。否则迭代。

由以上算法, 可求最多 m 层 Homography 变换。

4.2 使用多层 Homography 进行全图对应点匹配

前文提到的对应点匹配只适用于角点, 也就是纹理丰富梯度变化较大的点, 而图像中大多数的点都是处在各个相对平滑的区域中, 因此, 以上所述的方法无法适应全图的对应。而使用多层 Homography 变换算法则可以对绝大多数的点进行匹配。

上一节中, 求得 m 层 Homography 变换, 它们表示背景中的 (平面) 层次被分成了 m 层, 而图像中绝大多数的像素点都分别属于这 m 层中的一层或若干层。为解决图像中全像素点的匹配问题, 按 $j = 1, 2, \dots, m$ 依次对第 2 视角作 Homography 变换, 在第 j 层变换中, 找到这样一些点, 使它们满足:

$$\exists (\Delta \mu_x, \Delta \mu_y), -\mu \leq \Delta \mu_x, \Delta \mu_y \leq \mu$$

$$\text{let } |I(x_i^{(1)}, y_i^{(1)}) - \hat{I}(\hat{x}_i^{(1)} + \Delta \mu_x, \hat{y}_i^{(1)} + \Delta \mu_y)|^2 < d_threshold$$

$$\hat{\mathbf{x}}_i^{(1)} = H_j^{-1} \mathbf{x}_i^{(2)}$$

式中, $\hat{I}(\cdot, \cdot)$ 表示第 2 视角作 Homography 变换后的在点 (\cdot, \cdot) 上的亮度信息, $\hat{x}_i^{(1)}, \hat{y}_i^{(1)}$ 为向量 $\hat{\mathbf{x}}_i^{(1)}$ 在 x, y 轴上的分量。即找到第 2 视角中经过 Homography 变换后坐标的微小邻域内如果存在亮度与其在第 1 视角中的对应点亮度相似的像素点, 则认为它是服从该 Homography 变换, 并找到了对应点。当然, 可以将单点的亮度信息扩展到考虑某个邻域中的亮度信息, 可以使用类似第 2 章中的方法来进行衡量。假设邻域大小为 $(2w + 1) \times (2w + 1)$ 则可将相似度定义为

因此, 在寻找第 2 层的匹配关系时只需要考虑不符合第 1 层 Homography 的那些点。依此类推。

一旦有了对应关系, 就能使用类似于静止摄像机中背景建模的方法来进行像素级别的背景建模。

5 实 验

图 3 图 4 显示了在室外和室内场景中第 1 层 ~ 第 3 层的 Homography 变换后得到的视差图, 视差图中蓝色圆圈标注的区域就是服从该层 Homography 变换的稠密点区域。

实验中, 将 RANSAC 算法的最大循环次数设为 5 000 次, 内点的阈值 $p_threshold$ 设为 0.1, 内点数需占本层剩余点的 50%。图 3 中提取了 5 层 Homography, 图 4 中提取了 4 层。

根据实验发现, 层数大概为 5 层比较适宜, 因为层数越高, 剩余的对应点中匹配失误的比例就越高。

图 5 显示了经过 Homography 配准后建立起的背景模型中存在前景 (运动物体: 人) 的情况。通过实验可见, 本文的方法能够将前景分割出来。



(a) 前 3 层 Homography 的角点对应点



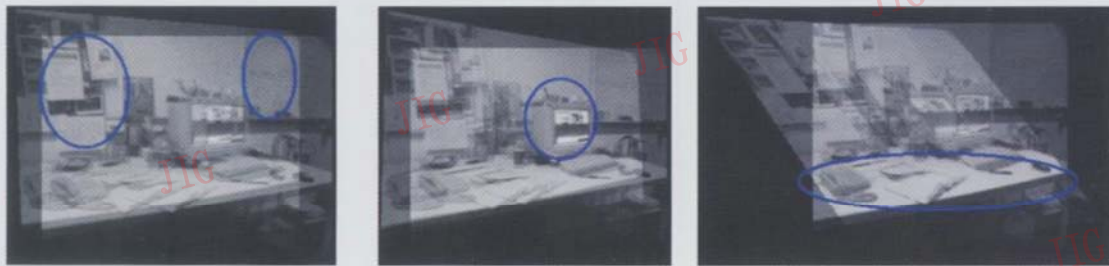
(b) 符合各层的稠密对应点

图 3 实验结果 1

Fig.3 Experiment result 1



(a) 前 3 层 Homography 的角点对应点



(b) 符合各层的稠密对应点

图 4 实验结果 2

Fig.4 Experiment result 2



图 5 包含前景的结果

Fig 5 Foreground segmentation

6 结 论

虽然背景建模已经是一个比较陈旧的话题,但以前的工作基本都是针对静止摄像机来实现的。在运动摄像机环境下的背景建模问题中,文献[6]做过一些探索性的尝试,但由于其无法解决背景层次较多的问题而导致了一些区域无法被更新或者更新错误,最终导致了模糊效应。

本文通过 Harris 算子在两个视角中寻找纹理信息丰富的角点,随后使用一些简单的匹配算法找到对应点。在给定对应点的情况下,通过多次使用 RANSAC 鲁棒估计算法求解不同层次的 Homography 变换,从而使得两个视角中绝大多数重叠的像素点都能找到较好的匹配。实验结果表明,该方法能有效地在摄像机运动环境中进行像素点级别的背景建模。

多层 Homography 提取方法已经在摄像机运动环境下做出了贡献,但还有一些不足及有待突破之处。两个视角中不是所有区域都有重叠,新的视角中有一部分是老视角中没有的,因此,如何判断这些新的区域该属于哪个 Homography 变换将是今后工作的重点。

目前的想法是:

(1) 对整张图计算所有的平滑区,新视角中的新增像素点根据所属的平滑区中其他与老视角有匹配的像素点所符合的 Homography 变换进行变换。与老视角进行拼接。运动摄像机在运动的过程中就相当于在做全景图的拼接,同时建立一个背景模型。

(2) 在寻找角点匹配时使用更好的算法进行改进,毕竟角点的匹配是否精确对计算多层 Homography 变换具有至关重要的影响。尤其在寻找比较靠后的层时,如果错误匹配的角点数量较多,则可能掩盖较好的 Homography,甚至可能会基于错误的对应点求得错误的 Homography 变换。

参考文献 (References)

- 1 Cristani M, Bicegi M, Murino V. Integrated region-and pixel-based approach to background modeling [A]. In: Proceedings of IEEE Workshop on Motion and Video Computing [C], Orlando, Florida, USA, 2002: 3~8.
- 2 Stauffer C, Grimson W. Learning patterns of activity using real-time tracking [J]. IEEE Transactions Pattern Analysis and Machine Intelligence, 2000, 22(8): 747~757.
- 3 Friedman N, Russell S. Image segmentation in video sequences: a probabilistic approach [A]. In: Proceedings of the 13th Conference on Uncertainty in Artificial Intelligence [C], Providence, Rhode Island, USA, 1997: 175~181.
- 4 Mittal A, Paragios N. Motion-based background subtraction using adaptive kernel density estimation [A]. In: Proceedings of Computer Society Conference on Computer Vision and Pattern Recognition [C], Washington, DC, USA, 2004: 302~309.
- 5 Tian Ying-li, Lu M ax, Han papur A. Robust and efficient foreground analysis for real-time video surveillance [A]. In: Proceedings of Computer Society Conference on Computer Vision and Pattern Recognition [C], San Diego, CA, USA, 2005: 1182~1187.
- 6 Rao N I, Di Hu rjun, Xu Guang-you. Joint correspondence and background modeling based on tree dynamic programming [A]. In: Proceedings of International Conference on Pattern Recognition [C], Hongkong, China, 2006: 425~428.
- 7 Harris C, Stephens M. J. A combined corner and edge detector [A]. In: Proceedings of Avey Vision Conference [C], Plessey, UK, 1988: 147~152.
- 8 Schmid C, Mohr R, Bauckhage C. Evaluation of interest point detectors [J]. International Journal of Computer Vision, 2000, 37(2): 151~172.
- 9 Forsyth D A, Ponce J. Computer Vision: A Modern Approach [M]. Englewood Cliffs, NJ, USA: Prentice-Hall, 2003: 238~299, 377~430.
- 10 Hartley R, Zisserman A. Multiple View Geometry in Computer Vision [M]. UK: Cambridge University Press, 2000: 320~344.

基于码书的运动目标检测方法

甘新胜

(中国船舶重工集团公司江苏自动化研究所, 连云港 222006)

摘要 运动目标检测是视频监控系统需要解决的关键问题之一。背景差法是固定的单摄像头监控中常用的一种运动目标检测方法,其核心是背景的构造。提出了一种基于码书的背景构造方法,它能在有限的存储空间开销下使用长时间的图像序列估计背景模型。该方法首先对每一个像素点的抽样进行聚类分析,即构造初始的码书;然后根据背景模型的条件从中挑选出合格的码字构造背景码书;最后通过判断当前的像素值是否可以划归于背景码书以区分背景点和前景点,同时做相应的更新。实验结果表明,即使背景本身存在运动和光照条件发生变化,该方法构造的背景也能有效检测运动目标。

关键词 运动目标检测 背景差法 码书

中图分类号: TP391 文献标识码: A 文章编号: 1006-8961(2008)02-0365-07

Moving Targets Detection Using Codebook

GAN Xin-sheng

(Jiangsu Automation Research Institute of CSIC, Lianyungang 222006)

Abstract Moving target detection is one of the critical issues in video surveillance. Background subtraction is a typical approach to detect moving targets in surveillance using only one static camera. The performance of background subtraction algorithms depends on how to construct background model. In this paper, codebook is used to estimate the background model using a long clip under limited memory. First, an initialization codebook is constructed for each pixel. Second, codewords in the initialization codebook are selected to construct background codebook based on the assumptions about the background model. The image can be partitioned to foreground and background according to the background codebook, and in the meanwhile codebook is updated correspondingly. Experimental results demonstrate the effectiveness of the method presented in this paper, even when there are moving clutters in the background and illumination variations.

Keywords moving targets detection, background subtraction, codebook

1 引言

运动目标检测是视频监控系统需要解决的一个关键问题,也是当前计算机视觉领域的研究热点之一。视频监控系统的典型应用是,使用固定的单摄像头和计算机系统对办公室、公园、道路、停车场等进行监控。针对这一类应用,现有的运动目标检测方法有:时间差分法、光流法、背景差法等^[1]。时间差分法通过比较相邻的 2 帧或 3 帧图像的差异

实现场景变化检测,它对动态环境具有较强的自适应性,但一般不能完整提取运动目标,在运动实体内部容易产生空洞现象。光流法利用运动目标随时间变化的光流特性,优点是在摄像机运动的情况下也能检测出运动目标,缺点是大多数的光流法计算方法相当复杂,且抗噪性能差,如果没有特别的硬件装置,则无法使用光流法进行全帧视频流的实时处理。

背景差法是当前常用的一种运动目标检测方法,思路是将视频图像与背景参考图像进行差分,差分的结果将图像分为背景点和前景点,前景点即是

收稿日期: 2007-06-20 改回日期: 2007-12-02

第一作者简介: 甘新胜(1979~),男,现为江苏省自动化研究所系统工程专业硕士研究生。研究方向为计算机视觉。Email: aganxs@

运动目标。背景差法的关键不在于差分运算,而是背景参考图像的构造即背景建模。背景模型与监控场景的复杂情况和系统的要求有关。Wren 等人使用单个高斯分布对每个像素点 x 在 YUV 空间构造背景: $I(x) \sim N(u(x), \Sigma(x))$, 其中 $u(x)$, $\Sigma(x)$ 是均值和方差, 它们由过去观察值的无限滤波器 (IR) 方式计算得到; 背景建立后, 通过比较当前的观察值与均值的偏离距离来区分背景点与前景点^[2]。

在预先能够得到静态背景的室内场景下单高斯背景模型具有很好的检测效果, 但是它不能解决户外场景中背景存在扰动以及背景物体的阴影等问题。Stauffer 和 Grimson 使用混合高斯模型 (MOG)

对每个像素点构造背景^[3]: $P(I_t) = \sum_{i=1}^K \bar{\omega}_i \cdot \eta(I_t; u_i, \Sigma_i)$, 其中, K 是高斯分布的数目, I_t 是 t 时刻的像素值, η 是高斯分布, u_i, Σ_i 为第 i 个高斯分布的均值和协方差; $\bar{\omega}_i$ 是估计的权值, 它反映该高斯分布出现的比例。背景是从 K 个高斯分布中选取符合条件的一个或多个分布。前景点的检测是通过比较 I_t 与背景所有分布的均值 u_i , 如果偏离都比较远, 该点就是前景点, 否则是背景点。若是背景点就把相应的高斯分布的权值、均值、方差均作更新。如果是前景点只有权值更新, 其他保持不变。混合高斯模型可以解决背景存在扰动、物体进入场景后持续停留、光照的变化等问题, 但是在处理背景存在的快速变化 (如树林场景) 和运动缓慢的目标时, 当背景的更新速度太高, 它会把运动缓慢的目标融入背景中, 导致较高的“漏报”, 反之, 背景的更新速度太低, 扰动的背景物体会产生过多的“虚警”。为了克服这种情况, Elgammal 等人提出了使用非参数的内核估计函数对每个像素点构造背景^[4], 首先设计一个内核估计函数 $P(I_t) = \frac{1}{n} \sum_{i=1}^n K(I_t - I_i)$, 其中,

$K(\cdot)$ 是内核估计函数, 它必须满足当 $I_t - I_i$ 接近零时函数值大, 而当 $I_t - I_i$ 偏离零时相应的函数值小, 如 $N(0, \Sigma)$, I_t 为当前的像素值, n 是存储的抽样的观察值数目。如果 $P(x_t) < th$ (th 为阈值), 那么该点就是前景点。此方法假定在 n 个像素值中, 属于背景类别的像素值比属于前景类别的像素值数要多, 因而当 I_t 靠近大的类别时, 它就有较大的估计函数值。显然, 存储的视频图像数目 n 的设定很关键, n 太大时, 存储空间开销和计算时间都会增大, n 太小时, 又不能有效地估计出背景。

本文提出一种基于码书的背景构造方法。首先对历史图像序列进行量化, 由此每一像素点得到一本码书, 其次根据前景点和背景点在图像序列中的分布特性选择合适的码字构造背景。该方法能在有限的存储空间下表示长时间的图像序列, 从而有效实现扰动的背景物体地融合, 同时克服运动缓慢的目标融入背景, 并能适应缓慢的光照变化。

2 背景建模

背景是从过去的图像序列中估计得到。设 $X = \{x_{t-T+1}, x_{t-T+2}, \dots, x_t\}$ 是某个像素点的最近 T 个历史值, 其中, x_i ($i = t - T + 1, \dots, t$) 是 RGB 向量或者单色灰度值。通过对 X 聚类分析得到一本码书 C , $C = \{c_1, \dots, c_L\}$, 其中, c_j ($j = 1, \dots, L$) 是码字^[5]。码书包含一个或多个码字, 每一个码字是样本集的一个聚类的形心, 码字可以近似地表示聚类。聚类分布可以是高斯分布, 也可以是其他的参数分布。样本与聚类的关系通过样本与聚类的形心的距离来判断。

码书只是对样本聚类的表示, 而如何使用码字构造背景就涉及到码字的统计信息。因此, 表示一个码字需要 5 个参量: $\langle v, \sigma, f, \lambda, T_{\text{last}} \rangle$, 其中, v 是聚类形心; σ 是聚类距离偏差, 是聚类成员对形心的最大距离偏差, 它是判别一个样本是否属于该聚类的唯一参考量, 训练阶段采用公共的距离偏差, 初始背景 (由训练得到的背景) 的 σ 由训练得到, 检测过程中实时创建的码字的 σ 由先验知识赋予初值, 并按属于聚类的样本与码字的距离关系实时更新; f 是码字出现的频率, 记录自从码字出现以来码字出现的次数, 初始背景的码字的频率为训练过程中的频率; λ 是码字相邻两次出现的最长时间间隔, 训练阶段由 λ 和 f 生成初始的背景, 初始背景码字的 λ 全部赋为 1 (这是为了在检测过程中更真实地反映该码字的 λ); T_{last} 是码字最后出现的时间, 是为了统计 λ 而使用的参量, 初始背景码字的 T_{last} 全部赋为 1。

通过观察知道, 对一个像素点, 在足够长的图像序列下, 属于背景的像素值出现的数目总比属于前景的要多, 而且前景的出现总是在零散的时间里, 同一个目标多次经过该像素点的次数总是比较少而且他们的时间间隔也是比较长的 (跟背景的像素值相比)。由此得到背景是一本码书, 它包含一个或多

个码字, 每个码字都具有出现频率较高、相邻两次出现的时间间隔都比较短的特点。

2.1 初始码书的训练

初始码书是通过对一段图像序列训练得到, 训练算法采用两次扫描的方法, 首先使用渐进构造聚类算法构造初始的码书, 然后合并距离相近的码字, 再由此对样本重新分类, 同时计算各个码字的参量。

渐进构造聚类算法的基本思想是对训练样本只进行一次扫描, 渐进的产生码字。算法首先把第 1 个样本作为码书, 其次依序输入其他样本, 对每一个样本 x , 计算它到所有码字的距离, 取其中的最小距离 $E = \min(d(x, c_i), c_i \in C)$, 然后比较 E 与预先设置的距离偏差 σ 的关系。如果 $E > \sigma$ 就把 x 作为新的码字的形心 $v = x$, 记 $f = 1$ 否则把 x 判为该聚类的成员, 并调整 $v = v \times f / (f + 1) + x \times 1 / (f + 1)$; $f = f + 1$ 。

当样本集存在大的聚类时, 渐进构造聚类算法可能把大的聚类划分为多个聚类, 这就需要把它们合并成一个聚类。为了更真实地统计合并后的聚类参量, 把所有码字除了形心 v 以外的参数归零, 然后在此基础上依序判别每一个样本的归属, 同时计算码字的参量。对任意 t 时刻的样本 x , 计算 x 与每一个码字的距离并取最小距离 $E = \min(d(x, c_i), c_i \in C)$, 若 $E < \sigma$, x 就判属码字 c_i , 如果 x 是 c_i 的首个样本, 则赋值 $\langle c_i, E, 1, 1, t \rangle$; 否则, $f = f + 1, \lambda = \max(\lambda, t - T_{\text{last}}), \sigma = \max(\sigma, E), T_{\text{last}} = t$ 。

以上介绍的训练算法能保证每一个样本到聚类的距离最小, 而且克服了一遍扫描生成的冗余码字, 接下来介绍使用的距离计算公式 $d(x, c)$ 。

2.2 距离 $d(x, c)$

像素值变化的主要原因一是场景物体的变化, 也就是反射面的变化; 二是场景内光照条件的变化; 三是传感器的噪声。场景物体的变化是运动目标检测所关注的焦点。光照变化、噪声都是干扰运动目标检测的因素, 它们是要抑制的对象, 也是背景需要融合的因素。由 RGB 模型和 HSI 模型的关系^[6] 构造一种距离计算模型, 如图 1 所示。

任意一个码字 c (形心为 v) 和样本 x , 设 v, x 是 $O(0, 0, 0)$ 到 v, x 的向量。定义强度轴 g 是点 O 到点 $(255, 255, 255)$ 。 g' 是过 v 与 g 同方向的向量。距离 $d(x, c)$ 可以分为两个分量, 一个是 g' 方向的距离 d_1 , 另一个是垂直 g' 的距离 d_2 。令 $x' = x - v, \theta$ 为 x' 与 g' 的夹角, 则

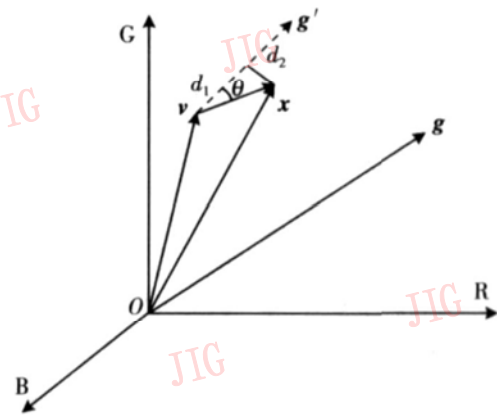


图 1 距离计算模型

Fig 1 Model of computing distance

$$d_1 = \|x'\| \cos\theta \tag{1}$$

$$d_2 = \|x'\| \sin\theta \tag{2}$$

其中, $\|x'\|$ 是向量 x' 的模, $\cos\theta = \frac{x' \cdot g'}{\|x'\| \times \|g'\|}$ 。

$d(x, c)$ 定义为

$$d(x, c) = \sqrt{a \times d_1^2 + b \times d_2^2} \tag{3}$$

其中, a, b 是权值, $a + b = 1, a > 0, b > 0, a$ 和 b 反映了两个方向的距离在总距离中的分量。由此聚类的判别式 $d(x, c) < \sigma$ 的点的轨迹描述了一个以 v 为中心平行强度轴的实心 3 维椭球。

2.3 初始的背景构造

背景码书是由出现频率较高、相邻两次出现的时间间隔比较短的码字构成。背景码书 B 满足:

$$B = \{c_i | f_{c_i} > \xi \cap \lambda_{c_i} > \epsilon_\lambda, c_i \in C\} \tag{4}$$

其中, C 是初始训练得到的码书; ξ, ϵ_λ 分别是码字的频率和码字相邻两次出现的最长时间间隔的阈值, 它们的设置与场景有关, ξ 如果取值过大而 ϵ_λ 取值太小, 将会无法融合周期运动的背景物体, 甚至出现没有满足条件的码字; 反之, ξ 如果取值过小而 ϵ_λ 取值太大, 背景融入过多的前景物体, 也会影响检测工作。

2.4 背景的更新

监控场景是变化的, 因此背景模型也要随着监控场景的变化相应地更新, 即具有自适应能力。每个像素点都有两本码书: 背景码书和前景码书。背景的更新包括把背景码书内不合格码字移出背景码书和原先不合格的前景码字满足背景条件了也要加入背景码书, 而且每一个码字都具有自适应能力。背景码字的删除通过按式 (4) 检验码字的统计信息, 而码字的更新只限于当样本被判归它时。样本 x 被判属码字 c 的判别式为 $d(x, c) < \sigma$, 若满足, 则

码字 c 进行如下更新:

$$v = (1 - \alpha) \times v + \alpha \times x \quad (5)$$

$$\sigma = (1 - \rho) \times \sigma + \rho \times d(x, c) \quad (6)$$

其中, α 是初始设定的背景更新速率, ρ 是 $d(x, c)$ 和 σ 计算得到, 同时 T_{last} 设置为当前时刻 t 频率 $f=f+1$ 。如果不属于码字 c , 仅修改 $\lambda = \max(\lambda, t - T_{\text{last}})$ 。

如果样本 x 在背景码书找不到匹配的码字, 就到前景码书中寻找匹配的码字, 过程与背景码字相同。如果都找不到就创建一个前景码字, 码字的参数设置为 $\langle \alpha, \sigma_{\text{init}}, 1, 1, t \rangle$ 。最后检测前景码书中是否有满足式 (4) 的码字, 如果有就将它移入背景码书。

3 前景点检测

前景点的检测只需要对每一个像素点的像素值逐个判断它们是否属于它们的背景码书。设任意的像素点的像素值为 x , 该点背景码书的任一个码字 c , 首先计算它们之间的距离 $d(x, c)$, 如果 $d(x, c) < \sigma$, 那么该点就是背景, 如果没有一个码字满足, 那么该点就是前景点。

为了消除孤立的噪声点和填充空洞, 使用形态学图像处理的开操作和闭操作进行处理包含前景点和背景点二类的二值图像。

4 实验分析

为了验证基于码书的背景构造方法及相应的运动目标检测算法的有效性, 利用标准测试数据实现本文算法, 并与基于混合高斯模型 (MOG) 的背景构造方法进行了比较。使用的视频数据来源 PET2001 (<http://fig.pets2001.cs.rdg.ac.uk>)。视频包含训练部分和测试部分, 训练部分用于每一个方法初始背景模型的训练, 测试部分用于测试算法的性能, 训练视频共 2989 帧, 测试视频共 2823 帧。

4.1 初始化设置

混合高斯模型 (MOG) 的初始高斯参数采用期望最大化 (EM) 算法^[7]估计, 每一个像素点的 RGB 3 个颜色通道是独立分布的, 方差不等, 背景按 $\bar{\omega}/\sigma$ 排序, $\sigma = \sigma_R + \sigma_G + \sigma_B$ 。MOG 分两组, 一组背景的自适应率较高, 另一组较低, 参数设置: 第 1 组 $K = 3$ $\lambda = 2.5$ $\alpha = 0.01$ $T = 0.75$ $\bar{\omega}_{\text{init}} = 0.1$ $\sigma_{\text{init}} = 20$ 视频率为 8 帧 /s 第 2 组为 $K = 3$ $\lambda = 2.5$ $\alpha = 0.05$

$T = 0.75$ $\bar{\omega}_{\text{init}} = 0.1$ $\sigma_{\text{init}} = 20$ 视频率为 8 帧 /s 其中 ρ 取文献 [8] 中修正的 $\rho \approx \alpha/\bar{\omega}$ 。

码书的初始训练算法采用 2.1 节介绍的方法, 初始的距离标准差 $\sigma = 16$ 初始背景的本码书平均包含的 1.36 个码字, 图 2 是训练得到的背景, 每一个像素点取频率最大的码字的形心。每本码书至少有一个码字, 最多的是 3 个。检测过程中新创建的码字的距离偏差 $\sigma = 16$ 视频率为 8 帧 /s。



图 2 训练得到的初始背景

Fig 2 The initial background

4.2 各种情况的比较分析

分析使用码书和混合高斯模型 (MOG) 在背景存在扰动、运动缓慢目标的检测、光线缓慢变化、场景物体运动变化等方面的检测效果。

(1) 背景存在扰动的情况

MOG 对扰动物体的融合通过背景自适应率对各个分布的权值进行修正, 分布命中越多, 相应的权值也越大。码书方法是通过统计码字的出现次数与出现的时间间隔, 当物体扰动的周期较小时, 它在出现一定的次数后就很快能融入背景, 当物体扰动周期较大时, 只要每两次出现的间隔在一定的时间内, 即 λ 小于设定的门限, 那么在足够长的时间后也能融入背景模型中。图 3 是码书方法和 MOG 检测在背景存在扰动的检测结果, 图 3(b) 是当时对应背景 (取频率最大的码字的形心), 由图 3(d) ~ (e) 可以看出由于树木摆动导致的误检, 而在图 3(c) 中没有。

(2) 运动缓慢目标的检测

运动缓慢的目标在背景的更新速率较高时, 目标会部分或全部的融入背景, 从而导致检测准确率的下降, 这在 MOG 常遇到。图 4 是 MOG 和码书方法对这种情况的检测效果, 第 2 组 MOG 在车子减速到速度很小时车子部分融入了背景 (图 4(d) 第 2

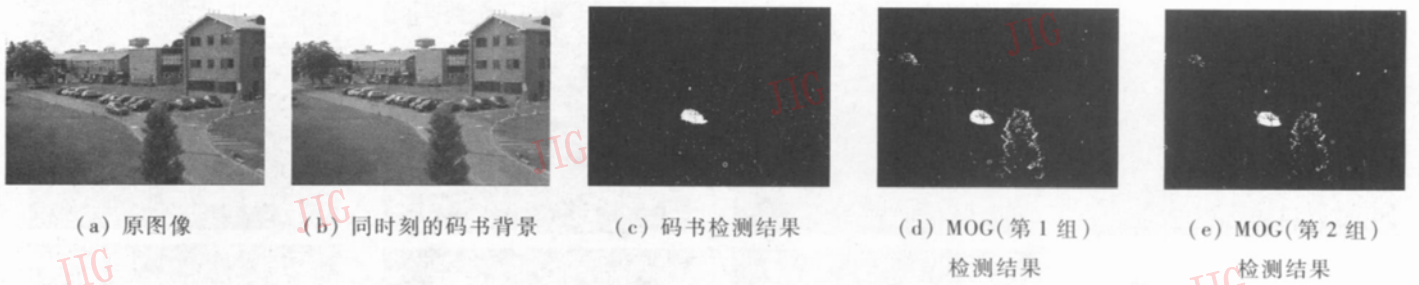


图 3 背景存在扰动的检测结果

Fig. 3 Detection results of moving background

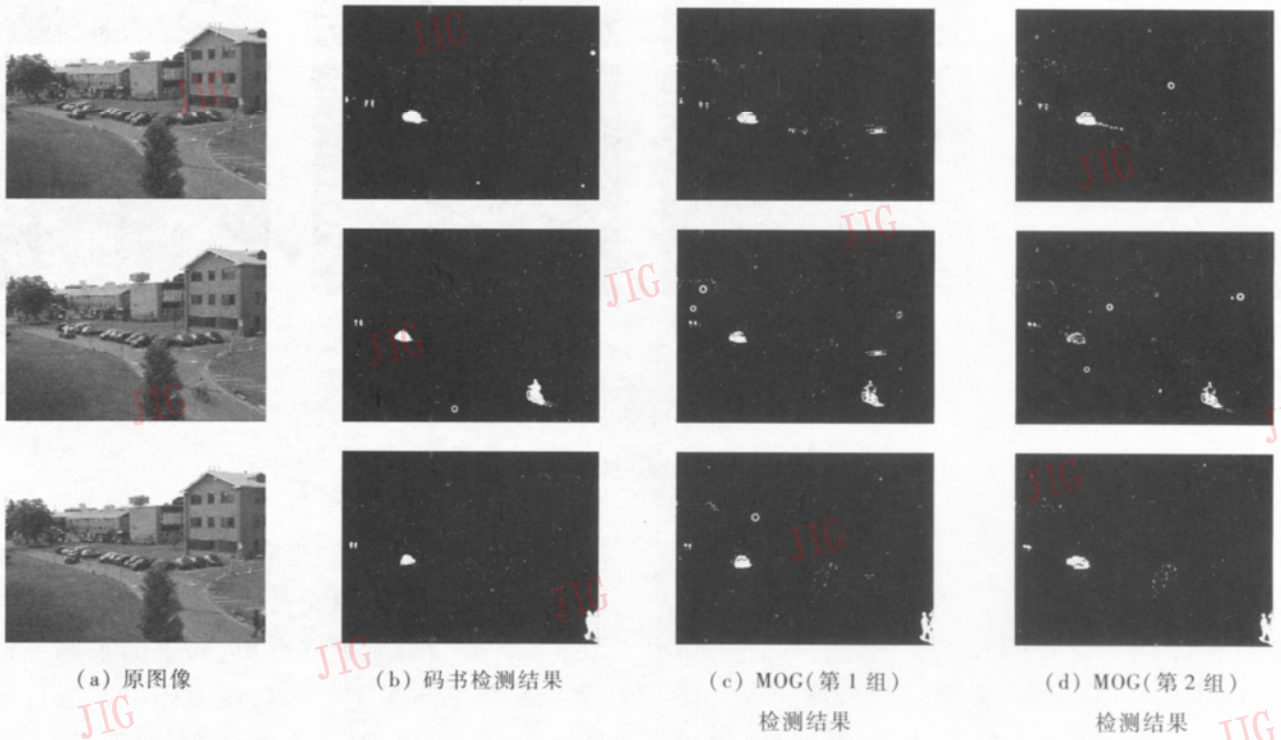


图 4 缓慢运动目标检测结果

Fig. 4 Detection results of slowly moving targets

行), 当车辆转弯进入停车位时产生“误检”(图 4(d)第 3 行), 而且在车辆速度较慢时在车的后面留下一条“尾巴”(图 4(d)第 1 行)。第 1 组 MOG 和码书没有以上的情况。码书对目标融入背景是通过判断是目标在原地停留的时间与设定的时间门限。如果停留的时间超过门限, 相应的码字频率增大, 目标就会融入背景。时间门限可以由先验知识估计得到。

(3) 光照缓慢变化

两种方法都采用自适应的方式因此都具有融合光线缓慢变化的能力, MOG 采用欧几里德距离, 码书构造采用 2.2 节介绍的分解后加权和。当场景由光线由阴暗转为比较亮时, 码书方法不受影响, MOG 在自适应率较高时也能不受影响。图 5 是光线由暗缓慢转为明亮的 MOG 和码书检测效果。图

5 第 1 行是光线比较暗的场景, 第 2 行是光线由暗缓慢变亮的场景, 第 3 行是场景光线变为更加明亮, 局部有太阳照射。

(4) 场景的运动变化

目标进入场景后持续停留在某处或者原先静止的物体突然移动, 这些情况在监测过程中都常常发生。码书方法对此的处理是由码字的频率决定, 当目标停留超过设置的门限即码字的频率大于门限时, 目标就能融入背景。对静止目标突然运动的处理与目标停留的时间有关, 如果该目标停留时间相对较短, 原先的背景还停留在背景码书中, 此时不会出现误检, 如果原先的背景已经删除, 将会出现误检。MOG 的处理是通过背景更新速率的修改, 都会出现误检。码书方法可以对初始训练得到的背景进行标识, 它们作为永久码字停留在背景中, 从而避免

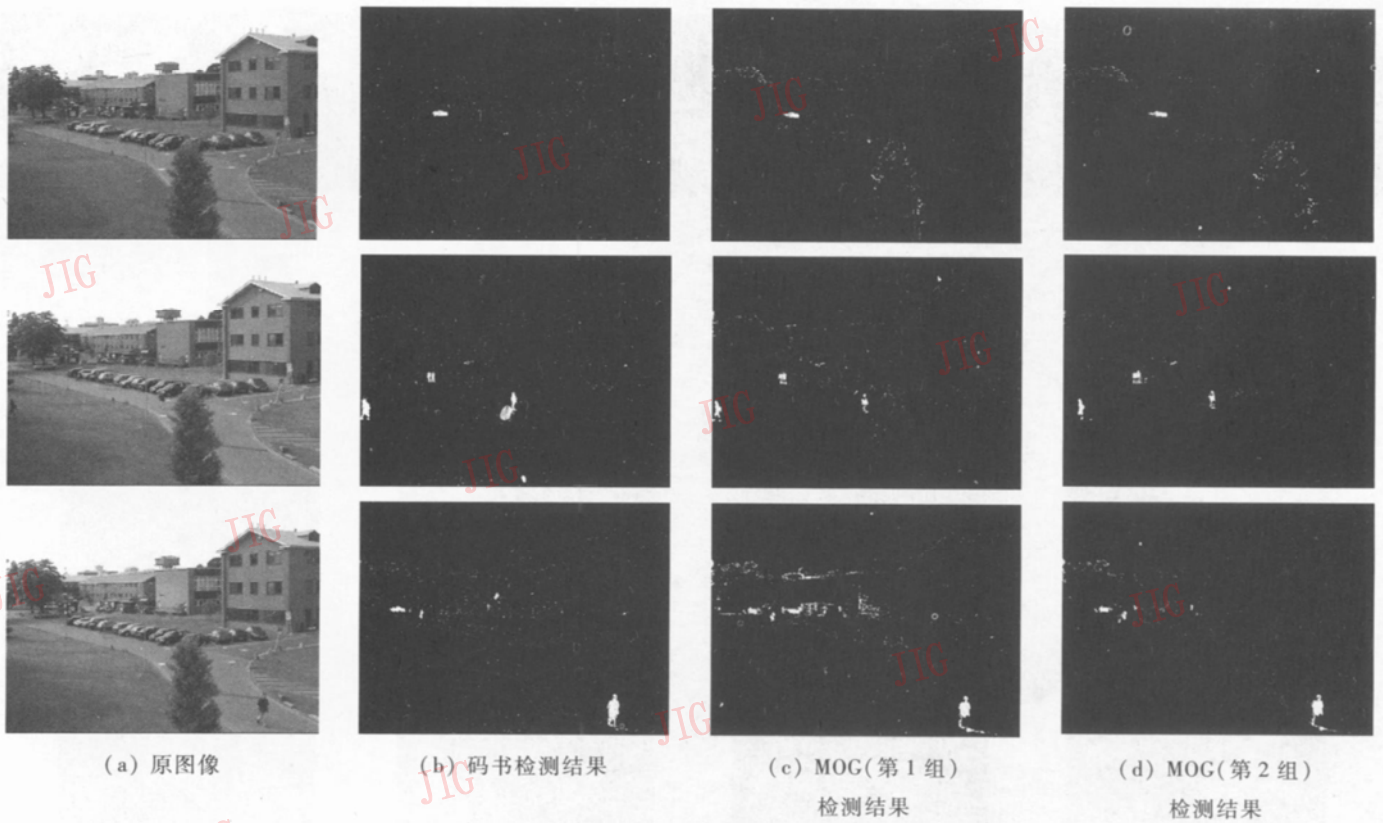


图 5 光照缓慢变化的检测结果

Fig. 5 Detection results of illumination changes slowly

这类误检的出现。

4.3 定量分析

为了得到一个有效的评估,手动生成真实值 (ground-truth), 并以此作为参照分析检测实验。比较的实验数据都经过相同的后处理 (开操作、闭操作)。真实值是从测试视频中选取 10 帧图像, 手动区分出前景点和背景点, 如图 6 所示。

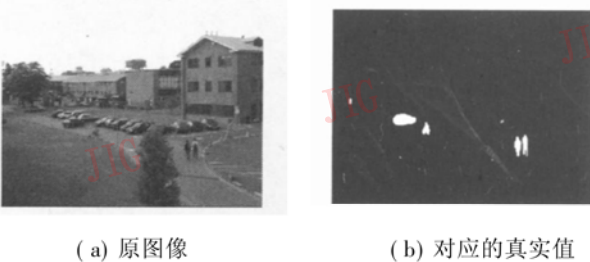


图 6 真实值

Fig 6 The ground truth

两个分析的量分别是正确率和准确率^[9]:

$$\text{正确率} = \frac{\text{算法正确检测到的前景点数目}}{\text{真实值中前景点数目}} \quad (7)$$

$$\text{准确率} = \frac{\text{算法正确检测到的前景点数目}}{\text{算法检测到的前景点总数}} \quad (8)$$

正确率和准确率是一对相互制约的参量, 当提

高准确率时, 就会导致正确率的下降, 而正确率提高准确率也会下降。表 1 是 MOG 和码书方法的检测结果统计。由表 1 可以看出, 码书的准确率都比较高, 而正确率与 MOG 相近, 说明码书在融合干扰信息的能力比较强, 但是检测的灵敏度不够高; 前 3 个抽样 MOG 的准确率比较低, 原因是初始运动目标

表 1 码书和 MOG 的检测结果统计

Tab 1 The Stat of detection results using codebook and MOG

抽样	码书		第 1 组		第 2 组	
	正确率	准确率	正确率	准确率	正确率	准确率
1	0.9784	0.67447	0.86111	0.30693	0.84568	0.34422
2	0.96544	0.82931	0.88697	0.38178	0.88913	0.47209
3	0.7656	0.91663	0.87823	0.61893	0.87976	0.58193
4	0.78238	0.70214	0.78187	0.8056	0.78238	0.81357
5	0.9109	0.73913	0.90336	0.75559	0.9017	0.81342
6	0.77671	0.88039	0.83041	0.61703	0.83041	0.5542
7	0.85425	0.77177	0.75879	0.71611	0.74459	0.79276
8	0.72143	0.85122	0.86184	0.79772	0.75634	0.72747
9	0.71589	0.83356	0.73627	0.59541	0.57826	0.75715
10	0.70182	0.70458	0.73633	0.36816	0.74544	0.61034

少而背景物体运动(树木摆动)比较明显; MOG 的第 1 组和第 2 组在前 7 个抽样的正确率都很相近, 第 8、9 个抽样第 1 组优于第 2 组, 原因是这两个抽样出现存在缓慢运动的目标, 自适应率高的把目标融入背景; 码书的正确率、准确率都比较平稳, 而 MOG 两组数据的波动比较大。

5 结 论

背景的构造是使用背景差法进行运动目标检测的难点。本文使用矢量量化的方法实现对长时间图像序列的压缩表示, 并根据背景和前景在图像序列的分布特点构造背景。实验结果表明, 该方法构造背景能够克服扰动的背景物体, 并能够很好的检测出缓慢运动的目标, 同时融合了场景的缓慢的光线变化。研究工作的下一步是如何把算法应用于实时环境中。

参考文献 (References)

- 1 Hu Weiming, Tan Tieniu, Wang Liang *et al*. A survey on visual surveillance of object motion and behaviors [J]. IEEE Transactions on Systems Man and Cybernetics-Part C: Applications and Reviews, 2004, 34(3): 334~352
- 2 Wren C, Azarbayejani A, Darrel T, *et al*. Pfister. Realtime tracking of the human body [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7): 780~785
- 3 Stauffer C, Grimson W. Adaptive background mixture models for real-time tracking [A]. In Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition [C], Fort Collins, Colorado USA, 1999, 246~252
- 4 Elgammal A, Harwood D, Davis L. Non-parametric model for background subtraction [A]. In Proceedings of International Conference on Computer Vision [C], Kerkyra Greece, 1999, 751~767
- 5 Shen F, Hasegawa O. An adaptive incremental LBG for vector quantization [J]. Neural Networks, 2006, 19(5): 694~704
- 6 Gonzalez R C, Woods R E. Digital Image Processing [M]. Beijing: House of Electronics Industry, 2003, 233~239 [阮秋琦, 阮宇智. 数字图像处理 (第二版) [M]. 北京: 电子工业出版社, 2003, 233~239]
- 7 Tomasi C. Estimating Gaussian mixture densities with EM—a tutorial [EB/OL]. <http://www.cs.duke.edu/courses/spring04/cps196/1/handouts/EM/tomasiEM.pdf> 2004-03-18
- 8 Power P W, Schoonees J A. Understanding background mixture models for foreground segmentation [A]. In Proceedings of Image and Vision Computing [C], Auckland, New Zealand, 2002, 267~271
- 9 Cheung S C, Kamnath C. Robust techniques for background subtraction in urban traffic video [A]. In Proceedings of SPIE Electronic Imaging, Visual Communications and Image Processing [C], San Jose, California, USA, 2004, 881~892



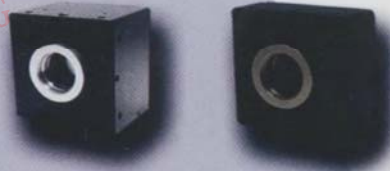
微视图像
Microview

GiGE™
VISION
兼容GenICam

北京微视新纪元科技有限公司

Beijing Microview Science and Technology Co., Ltd

微视图像千兆网产品系列—高清晰CMOS摄像机



推荐型号	分辨率	帧率
MVC300SA-GE200	640*480	200 fps
MVC360SA-GE60	752*480	60 fps
MVC1000SA-GE30	1280*1024	30 fps
MVC2000SA-GE20	1600*1200	20 fps
MVC3000SA-GE12	2048*1536	12 fps
MVC1280SA-GE30	1280*1024	27 fps
MVC6600SA-GE5	2210*3002	5 fps

微视图像千兆网产品系列—高灵敏CCD摄像机

推荐型号	分辨率	帧率
MVC600DA-GE60	659*494	60 fps
MVC610DA-GE110	659*494	110 fps
MVC800DA-GE45	1024*1024	45 fps
MVC900DA-GE30	1392*1036	30 fps
MVC1800DA-GE30	1380*1034	30 fps
MVC1830DA-GE30	1024*1024	30 fps
MVC2000DA-GE16	1620*1036	16 fps
MVC2048DLM (线阵)	2048*1	15-19k



图像传输距离受限? 传输线笨重? 性能不稳定? 成本太高?
所有问题, 让微视图像千兆网产品系列助您轻松解决!

微视图像千兆网产品系列—千兆网转换盒

推荐型号
MVCL2GE-B
MVLV2GE-S
MVDIG2GE-S
QUAD MVCL2GE-B



三维立体相机

推荐型号
MVDC360SAM-GE60ST (合体式)
MVC360SAM-GE60ST (分体式)
MVC1000SAM-GE30ST (分体式)



地址: 北京市海淀区中关村东路18号财智国际大厦A座10层
网址: www.microview.com.cn, www.microview.cn
电话: 010-82600088 (中继线) 传真: 010-82600088-6600
邮箱: sales@microview.com.cn support@microview.com.cn

一流产品 超一流服务

ISO9001: 2000质量管理体系认证企业

Journal of Image and Graphics

中国图象图形学报

标准刊号: CN11-3758/TB 邮发代号: 82-831 定价: 35元

ISSN 1006-8961



9 771006 896003