

一种视频序列的全景图绘制算法

韩玉兵 吴乐南

(东南大学无线电工程系, 南京 210096)

摘要 针对帧间变换为单映变换的视频序列全景图的绘制, 提出了一种视频序列的全景图绘制算法。该算法首先采用基于模型的等级估计法对相邻帧的单映变换矩阵进行估计; 然后采用基于 RANSAC 的特征点匹配方法对非相邻帧进行带指导的改进估计, 以使其能够达到要求的子像素级精度; 最后运用时间中值滤波绘制全景图。实验结果表明, 该算法检测的全景图不仅帧间配准达到了一定的子像素级精度, 而且基本达到了无缝连接。

关键词 视频处理 全景图 单映变换 RANSAC 等级估计

中图分类号: TP391.41 **文献标识码:** A **文章编号:** 1006-8961(2006)05-0684-05

A Panorama Rendering Algorithm of Video Sequence

HAN Yu-bing, WU Le-nan

(Department of Radio Engineering, Southeast University, Nanjing 210096)

Abstract A panorama rendering algorithm of video sequence based on homography transformation between frames is presented in this paper. Firstly, the hierarchical model-based motion estimation is adopted to compute the adjacent homography matrix. Then, for non-adjacent frames, the homography matrix is estimated by guided feature registration based on RANSAC algorithm, which can obtain sub-pixel accuracy. Finally, the video sequence panorama is rendered using temporal median filter.

Keywords video processing, panorama, homography transformation, RANSAC, hierarchical estimation

1 引言

在视频序列中, 任意时刻获得的图像信息都是对现实场景的局部描述, 只有将所有局部信息有机地结合起来, 方可形成对场景的整体认识, 运动全景图就是由多帧视频序列拼接形成的一个完整全景图像^[1]。它在遥感遥测、视频监控、医学图像处理等领域已得到广泛应用。

在视频序列的全景图绘制中, 一个最主要的难点就是图像配准必须要达到适当的子像素级精度^[1-3], 如果配准不精确, 则会带来严重的运动重影。鉴于此, 本文针对一种常见视频序列, 即帧间运动模型为单映变换的情形, 提出了基于模型的等级

估计法和基于带指导的 RANSAC (random sampling and consensus) 特征匹配方法相结合的运动配准策略, 该策略在充分利用两者优点的同时, 摒弃了各自的缺点, 并将其运用到视频序列的全景图绘制中。

2 帧间单映变换的基本概念

本文采用的单映变换 (homography transformation) 是一种常见的视频帧间模型^[3]。它主要适用于以下 3 种情形: (1) 空间为平面场景和任意摄像机运动; (2) 任意空间场景, 摄像机为旋转或放缩运动; (3) 远距离场景拍摄, 即景深远小于场景到摄像机的距离 (如航拍或卫星成像等), 其中第 3 种情形近似满足适用条件。单映变换帧间关系为

收稿日期: 2005-03-07; 改回日期: 2005-07-25

第一作者简介: 韩玉兵 (1971 ~), 男, 1999 年获河海大学硕士学位, 现为东南大学无线电工程系信号与信息处理专业博士研究生。研究方向为数字图像处理、数字视频处理。E-mail: hanholly@sina.com

$$\begin{pmatrix} \hat{x} \\ \hat{y} \\ \hat{z} \end{pmatrix} = \mathbf{H} \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \mathbf{H} = \begin{bmatrix} h_{1,1} & h_{1,2} & h_{1,3} \\ h_{2,1} & h_{2,2} & h_{2,3} \\ h_{3,1} & h_{3,2} & h_{3,3} \end{bmatrix} \quad (1)$$

其中, $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ 为齐次坐标, 对应的非齐次坐标为 $\begin{pmatrix} x/z \\ y/z \\ 1 \end{pmatrix}$,

由于上式在相差一个尺度因子的意义下等价, 所以矩阵 \mathbf{H} 实际上只有 8 个自由度, 不妨设 $h_{3,3} = 1$, 其等价的非齐次关系为

$$\begin{aligned} \hat{x} &= \frac{h_{1,1}x + h_{1,2}y + h_{1,3}}{h_{3,1}x + h_{3,2}y + 1} \\ \hat{y} &= \frac{h_{2,1}x + h_{2,2}y + h_{2,3}}{h_{3,1}x + h_{3,2}y + 1} \end{aligned} \quad (2)$$

3 相邻帧间的单映变换估计

考虑将视频序列 $I(x, y, t) (1 \leq t \leq N)$ 简记为 $I(t)$, 由于相邻帧能够很好地满足光流约束, 因此本文采用类似文献[4,5]的基于模型的等级估计法来进行单映矩阵估计, 这有以下几个优点^[5]: (1) 这是一种基于图像操作的直接方法, 且不需进行费时的特征点匹配; (2) 可以达到很高的子像素级精度; (3) 可以采取由粗至精的等级策略处理大位移矢量, 以提高收敛速度和避免陷入局部极小值。具体运算过程如下:

$$\begin{aligned} E_1(\mathbf{h}) &= \sum_{(x,y) \in \Omega} \left\{ \frac{h_{3,1}x + h_{3,2}y + 1}{\hat{h}_{3,1}x + \hat{h}_{3,2}y + 1} \left(\frac{\partial I}{\partial x} \left(\frac{h_{1,1}x + h_{1,2}y + h_{1,3}}{h_{3,1}x + h_{3,2}y + 1} - x \right) + \frac{\partial I}{\partial y} \left(\frac{h_{2,1}x + h_{2,2}y + h_{2,3}}{h_{3,1}x + h_{3,2}y + 1} - y \right) + \frac{\partial I}{\partial t} \right)^2 \right. \\ &= \sum_{(x,y) \in \Omega} \left\{ \frac{1}{\hat{h}_{3,1}x + \hat{h}_{3,2}y + 1} \left(\frac{\partial I}{\partial x} (h_{1,1}x + h_{1,2}y + h_{1,3} - x(h_{3,1}x + h_{3,2}y + 1)) + \right. \right. \\ &\quad \left. \left. \frac{\partial I}{\partial y} (h_{2,1}x + h_{2,2}y + h_{2,3} - y(h_{3,1}x + h_{3,2}y + 1)) + \frac{\partial I}{\partial t} (h_{3,1}x + h_{3,2}y + 1) \right)^2 \right. \\ &= \sum_{(x,y) \in \Omega} \left\{ \frac{1}{\hat{h}_{3,1}x + \hat{h}_{3,2}y + 1} (\mathbf{C}\mathbf{h} + \mathbf{b}) \right\}^2 \end{aligned} \quad (5)$$

其中, $\mathbf{b} = \frac{\partial I}{\partial t} - x \frac{\partial I}{\partial x} - y \frac{\partial I}{\partial y}$, $\mathbf{C} = \begin{pmatrix} x \frac{\partial I}{\partial x}, y \frac{\partial I}{\partial x}, \frac{\partial I}{\partial x}, x \frac{\partial I}{\partial y}, y \frac{\partial I}{\partial y}, x \frac{\partial I}{\partial t} - x^2 \frac{\partial I}{\partial x} - xy \frac{\partial I}{\partial y}, y \frac{\partial I}{\partial t} - xy \frac{\partial I}{\partial x} - y^2 \frac{\partial I}{\partial y} \end{pmatrix}$ 。
若将 $\mathbf{h} = \hat{\mathbf{h}} + \Delta\mathbf{h}$ 代入 $E_1(\mathbf{h})$, 则上述问题变为以下求解 $\Delta\mathbf{h}$ 使下式达到最小的问题:

$$E_1(\Delta\mathbf{h}) = \sum_{(x,y) \in \Omega} \left\{ \Delta I + \frac{1}{\hat{h}_{3,1}x + \hat{h}_{3,2}y + 1} \mathbf{C}\Delta\mathbf{h} \right\}^2 \quad (6)$$

其中, $\Delta I = I(x + \hat{u}, y + \hat{v}, t + 1) - I(x, y, t)$, $\hat{u} = \hat{u}(x, y, \hat{\mathbf{h}})$, $\hat{v} = \hat{v}(x, y, \hat{\mathbf{h}})$ 。上式关于 $\Delta\mathbf{h}$ 求导, 并令其等于零, 可得

假设 $I(x, y, t)$ 和 $I(x, y, t + 1)$ 为相邻两帧, 且 $I(x, y, t) = I(x + u, y + v, t + 1)$, 则光流方程为:

$\frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v + \frac{\partial I}{\partial t} = 0$, 其中 $\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y}$ 为图像光强的空间导数,

$\frac{\partial I}{\partial t} = I(x, y, t + 1) - I(x, y, t)$ 为时间导数, u, v 为光流 (或位移)。

构造的误差函数为 $E(u, v) = \sum_{(x,y) \in \Omega} \left(\frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v + \frac{\partial I}{\partial t} \right)^2$, 其中, Ω 为图像区域。所谓直接法就是估计 u, v 使 $E(u, v)$ 达到最小。由于考虑帧间模型为单映变换, 所以

$$\begin{aligned} u &= \frac{h_{1,1}x + h_{1,2}y + h_{1,3}}{h_{3,1}x + h_{3,2}y + 1} - x \\ v &= \frac{h_{2,1}x + h_{2,2}y + h_{2,3}}{h_{3,1}x + h_{3,2}y + 1} - y \end{aligned} \quad (3)$$

则误差函数为

$$\begin{aligned} E(\mathbf{h}) &= \sum_{(x,y) \in \Omega} \left(\frac{\partial I}{\partial x} \left(\frac{h_{1,1}x + h_{1,2}y + h_{1,3}}{h_{3,1}x + h_{3,2}y + 1} - x \right) + \right. \\ &\quad \left. \frac{\partial I}{\partial y} \left(\frac{h_{2,1}x + h_{2,2}y + h_{2,3}}{h_{3,1}x + h_{3,2}y + 1} - y \right) + \frac{\partial I}{\partial t} \right)^2 \end{aligned} \quad (4)$$

此处 $\mathbf{h} = (h_{1,1}, h_{1,2}, h_{1,3}, h_{2,1}, h_{2,2}, h_{2,3}, h_{3,1}, h_{3,2})^T$ 。本文采用基于由粗至精等级策略的 Gaussian-Newton 方法来迭代求解上述最小值, 假设当前的 \mathbf{h} 的估计值为 $\hat{\mathbf{h}}$, 估计校正值为 $\Delta\mathbf{h}$, 即 $\mathbf{h} = \hat{\mathbf{h}} + \Delta\mathbf{h}$ 。为避免 $E(\mathbf{h})$ 非线性, 则可将误差函数修改为

$$\begin{aligned} &\left[\sum_{(x,y) \in \Omega} \left(\frac{1}{\hat{h}_{3,1}x + \hat{h}_{3,2}y + 1} \right)^2 \mathbf{C}^T \mathbf{C} \right] \Delta\mathbf{h} \\ &= - \sum_{(x,y) \in \Omega} \frac{1}{\hat{h}_{3,1}x + \hat{h}_{3,2}y + 1} \mathbf{C}^T \Delta I \end{aligned} \quad (7)$$

若将此迭代法结合由粗至精多分辨率等级策略, 则可得如下帧间单映矩阵的快速估计算法:

算法 1: 基于模型的帧间单映矩阵等级估计, 其算法步骤为:

(1) 对相邻帧 $I(t)$ 和 $I(t + 1)$ 分别构建多分辨率等级层, 由粗至精依次为 $l = L, \dots, 0$, 设置初始值 $\hat{\mathbf{h}} = \mathbf{h}_0$;

(2) 对每一等级层, $l = L, \dots, 0$

① 运用当前估计参数 \hat{h} 对 $I(t+1)$ 进行变换, 即 $\hat{I}(t) := \text{warp}(I(t+1), \hat{h})$, 并计算 $\Delta I = \hat{I}(t) - I(t)$;

② 求解式(7)得 Δh , 用于修正 $\hat{h} = \hat{h} + \Delta h$;

③ 重复步骤①、②, 直至 $\|\Delta h\| < \varepsilon$, 或达到指定迭代步数;

(3) 传递参数 \hat{h} 到上一等级层 $l-1$, 重复步骤(2), 直到最高层 $l=0$, 最后得 $h = \hat{h}$ 。

4 非相邻帧间的单映变换估计

上节已对相邻帧间单映阵 $H_{k,k-1}$ ($1 < k \leq N$) 进行了估计, 由于非相邻帧间运动模型不太满足光流限制, 故采用下述策略进行估计。假设第 t 帧为参考帧, 第 k 帧为目标帧, 若要求估计它们间的单映矩阵, 则应首先运用单映矩阵的可逆性和级联性来得到初始估值

$$\hat{H}_{k,t} = \begin{cases} \hat{H}_{k,k-1} \hat{H}_{k-1,t} & k > t \\ \hat{H}_{k,k+1} \hat{H}_{k+1,t} & k < t \end{cases} \quad (8)$$

其中, $\hat{H}_{k,t}$ 为当前第 t 帧到第 k 帧单映矩阵的最好估计值, 且有 $\hat{H}_{t,k} = \hat{H}_{k,t}^{-1}$; 其次, 由于估计的单映矩阵级联可能导致误差累积, 为此本文采用带指导的基于 RANSAC^[6] 的特征匹配方法来改进迭代过程。RANSAC 是一种基于假设检验策略的估计方法, 其已被广泛应用于存在许多数据出格点的情况下对模型进行可靠的拟合。

算法 2: 基于 RANSAC 特征匹配的帧间单映矩阵估计算法的步骤为:

(1) 运用特征提取算法 (如 Harris 算法^[7]) 在参考帧和目标帧中提取特征点;

(2) 对参考帧中的特征点, 以当前单映矩阵的估计值来确定相应的搜索区域, 并运用某种相似性原则计算匹配特征点对;

(3) 运用 RANSAC 鲁棒估计方法来得到最大的一致点集和矩阵 H 的估计, 即

① 随机抽取 $n \geq 4$ 个匹配点对 (因为估计矩阵 H 至少需要 4 对数据点), 用于估计矩阵 H 参数;

② 对步骤(2)中的每一个匹配点对, 计算其对矩阵 H 参数的拟合误差;

③ 设定一个门限值 T , 若拟合误差小于此门限值, 则表示此数据点对为一致点 (或内点), 否则为出格点 (或外点), 并统计一致点的数目;

④ 重复步骤①、②、③, 直至达到最大抽样次数或所有的一致点集中至少包含一个有效表征集的概率大于一定数值为止;

⑤ 选择具有最大一致点集的单映阵 H ;

(4) 重复步骤(2)和步骤(3), 直至满足一定条件 (如 H 变化很小或者内点的比例基本不变)。

注: (1) 上述算法中的主要计算集中在步骤(2)中的特征点匹配搜索上^[3], 由于当前估值 H 已达到一定的子像素级精度, 因此可以用其来指导确定搜索区域, 假设 P 为参考帧中某特征点, 则目标帧中的搜索区域可确定为以 $\hat{P} = H(P)$ 为中心的矩形邻域, 此处 $H(P)$ 为对点 P 进行单映阵 H 的单映变换, 因为其带有指导性, 所以搜索区域可以适当变小, 一般选为 5×5 或 3×3 , 这样就可大大减少搜索时间, 以提高计算速度; (2) 步骤(2)中的相似性原则可以采用规整化的邻域相关方法^[8] 来确定; (3) 步骤(3-1)中的单映阵 H 参数的估计方法是采用基于 SVD (singular value decomposition) 的方法^[8]; (4) 步骤(3-2)中的拟合误差函数为

$$R = d^2(\hat{P}, H(P)) + d^2(P, H^{-1}(\hat{P})) \quad (9)$$

其中, d 为某种距离函数。

5 视频序列全景图绘制

运动全景图绘制主要包括视频序列的帧间运动估计以及在时间轴上对调整后的图像序列进行数据融合, 其主要有以下几步: (1) 选取参考帧, 并将其余视频帧与参考帧进行运动配准, 即估计单映变换矩阵 (见本文第 3.4 节); (2) 确定全景图像的大小, 这里取能够包含所有视频帧信息的最小尺寸; (3) 运用某种插值算法, 将每个视频帧依照估计的单映矩阵 H 变换到全景图像中; (4) 对全景图像的每像素点运用某种融合算法 (时间平均或时间中值滤波) 绘制最终全景图。

6 实验研究

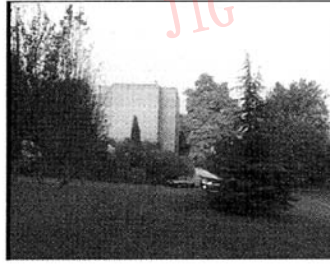
实验 1 采用大小为 360×288 的普通场景序列 Surrey car (URL: <http://www.robots.ox.ac.uk/~vgg/data/>), 共 50 帧。实验 2 采用大小为 512×480 的航空序列 Town06 (URL: <http://vasc.ri.cmu.edu/idb/html/motion/>), 共 32 帧。这里选取第 1 帧为参考帧, 以便随着时间的推移逐渐绘制全景图, 相

邻帧间的运动估计采用基于模型的等级估计方法(等级层数 L 为 4),非相邻帧间的运动估计采用基于带指导的 RANSAC 特征匹配方法,插值算法采用双三次插值,融合算法采用时间中值滤波。图 1 给出了两

个序列的各自第 1、15 和 30 帧图像。图 2 为用本文方法绘制的两个序列的全景图像。实验结果表明:本文的运动估计或帧间配准不仅达到了一定的子像素级精度,而且绘制的全景图基本达到了无缝连接。



(a) Surrey car 序列第 1 帧



(b) Surrey car 序列第 15 帧



(c) Surrey car 序列第 30 帧



(d) Town06 序列第 1 帧



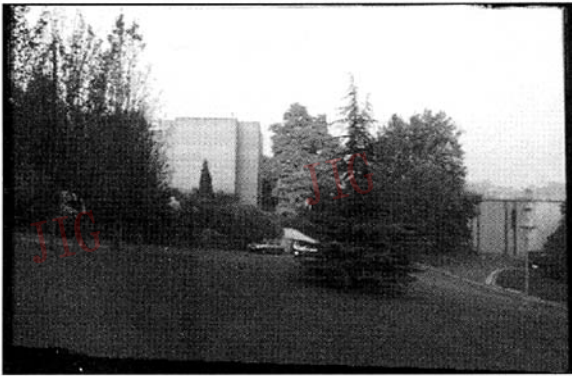
(e) Town06 序列第 15 帧



(f) Town06 序列第 30 帧

图 1 Surrey car 和 Town06 序列第 1、15 和 30 帧图像

Fig. 1 The 1st, 15th and 30th frames of Surrey car and Tow. n06 sequence



(a) Surrey car 序列



(b) Town06 序列

图 2 Surrey car 和 Town06 序列全景图

Fig. 2 The panoramas of Surrey car and Town06 sequence

7 结 论

本文针对帧间变换为单映变换的视频序列,提

出了一种全景图绘制算法,该算法主要分为以下两个部分:(1)帧间运动估计,即首先采用基于模型的等级估计方法对相邻帧的单映变换进行估计,这是一种快速的,无需特征点匹配的直接方法,然后采用

基于 RANSAC 的特征点匹配方法来对非相邻帧进行带指导的改进估计,由于这种估计带有指导性,所以可以极大地减少匹配搜索时间和提高计算速度;(2)运用时间中值滤波进行全景图的绘制。实验结果表明,该全景图绘制算法可以用于视频序列全景图的绘制。

参考文献 (References)

- 1 Shum H Y, Szeliski R. Panoramic image mosaics [R]. Technical Report, MSR-TR-97-23, Microsoft Research, Redmond, WA, USA 1997.
- 2 Kang S B. A survey of image-based rendering techniques [R]. Technical Report CRL 97/4, Digital Equipment Corporation, Cambridge Research Laboratories, Cambridge, UK August 1997.
- 3 Capel D P. Image mosaicing and super-resolution [D]. University of Oxford, UK, 2001.
- 4 Bergen J R, Anadan P, Hanna K, *et al.* Hierarchical model-based motion estimation [A]. In: Proceedings of European Conference on Computer Vision [C], Santa Margherita Ligure, Italy, 1992: 237 ~ 252.
- 5 Irani M, Anandan P. All about direct methods [A]. In: Proceedings of International Workshop Vision Algorithms [C], Corfu, Greece, 1999: 267 ~ 277.
- 6 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 Association for Computing Machinery, 1981, 24(6): 381 ~ 395.
- 7 Harris C J, Stephens M. A combined corner and edge detector [A]. In: Proceedings of Alvey Vision Conference [C], Manchester, England, 1988: 147 ~ 151.
- 8 Hartley R, Zisserman A. Multiple view geometry in computer vision [M]. Cambridge, UK: Cambridge University Press, 2000: 101 ~ 113.