

一种序列图像配准的计算框架

彭晓明 丁明跃 周成平 张天序

(华中科技大学图像识别与人工智能研究所 图像信息处理与智能控制教育部重点实验室, 武汉 430074)

摘要 提出了一种对来自多传感器的序列图像进行时间-空间配准的计算框架。该框架适用于摄像机静止的场合,而且所拍摄的图像序列中有运动目标存在,但在图像序列的开始阶段为静止背景。首先对静止背景进行配准,得到空间变换的初始估计;然后,利用运动目标质心间的对应关系得到时间变换的初始估计;最后,结合共同信息计算出最终结果。本框架的空间配准精度可得到亚像素级,时间配准精度可达到亚帧级。本框架已成功应用于可见光/红外图像序列的配准实验。

关键词 序列图像配准 共同信息 多传感器

中图分类号: TP391.4 **文献标识码**: A **文章编号**: 1006-8961(2005)04-0441-09

A Computational Framework for the Temporal-spatial Alignment of Multi-sensor Image Sequences

PENG Xiao-ming, DING Ming-yue, ZHOU Cheng-ping, ZHANG Tian-xu

(Institute for Pattern Recognition and Artificial Intelligence, State Education Commission Key Lab for Image Processing and Intelligent Control, Huazhong University of Science and Technology, Wuhan 430074)

Abstract A computational framework for the temporal-spatial alignment of multi-sensor image sequences is presented in this paper. The framework is suitable to the circumstance where the cameras are static; the captured sequences contain moving objects but the initial segments of the sequences are frames of the static background. The framework first registers the static backgrounds of the sequences to yield the initial spatial transformation. Then it uses the correspondence of the centroids of moving objects to estimate the initial temporal transformation. Finally, mutual information is incorporated into this framework to compute the final temporal-spatial transformations. This framework, which can obtain a sub-pixel and sub-frame registration accuracy, has been successfully applied to a visible/infrared sequence alignment experiment.

Keywords image sequence alignment, mutual information, multi-sensor

1 引言

图像配准是计算机视觉和模式识别领域中的一项重要课题,在遥感探测、医学成像和基于多传感器融合的目标识别等研究中有重大的应用价值。当前的研究主要集中在单幅图像的配准方面。有关单幅图像配准方法见文献[1]、[2]。单幅图像配准需要解决的问题是两幅图像在空间域上的对应关系。然

而在实际应用中,有时需要同时考虑两幅图像在时间域上的对应关系,这反映在对两个图像序列进行配准的问题上。例如,多传感器目标识别中,利用可见光和红外传感器对运动目标进行图像采集而分别得到两个图像序列。在进行融合目标识别之前,首先需要对这两个图像序列进行配准。与单幅图像的配准相比,序列图像的配准不但需要求解两个图像序列之间的空间变换关系,而且还要求解不同序列的两帧图像在时间上的对应关系。也就是说,序列

基金项目:国家自然科学基金重大资助项目(60135020 F F 030405)

收稿日期:2004-02-25; 改回日期:2004-09-13

第一作者简介:彭晓明(1974~),男。2000年毕业于中国科学院光电技术研究所并获得工学硕士学位,2004年于华中科技大学图像识别与人工智能研究所获工学博士学位。主要研究方向为图像处理、图像融合和目标识别等。E-mail: pxmhust@yahoo.com.cn

图像之间的配准不仅需要解决空间上的配准问题,而且还需要解决时间上的配准问题。与用于单幅图像配准的方法相比,用于序列图像的配准方法要少得多。比较典型的几种方法如下:

(1) Sharma 等人的方法^[3]

首先配准来自两个图像序列(记为 S_1 和 S_2)的第 1 帧图像(记为 f_1 和 g_1)。然后,采用图像序列帧与帧之间的图像配准方法(例如文献[4]中的方法)完成单个序列图像帧与帧之间的图像配准。对于不同序列的对应图像之间的配准问题,例如,对 S_1 的第 i 帧图像 f_i 和 S_2 的第 j 帧图像 g_j 进行配准时,则可利用 f_i 和 f_1 之间以及 g_j 和 g_1 之间的空间变换关系来解决。此方法的前提是 f_1 和 g_1 在时间域上的对应关系为已知。

(2) Stein 的方法^[5]

用两台计算机分别处理两个静止相机输出的序列图像。对于每个图像序列,采用自适应跟踪方法提取目标的质心 (x, y) ,并根据计算机时钟为每个目标质心加一个时间标志 t 。这样,生成两个列表 T 和 T' 。 T 和 T' 的每个元素分别是一个 3 元组 $T_i(x, y, t)$ 和 $T'_j(x', y', t')$ 。由于计算机时钟是已知量,所以 t 和 t' 也是已知量。将 T 和 T' 中所有满足 $|t - t'| < t_0$ (t_0 是一个小阈值)条件的元素进行组合,假设共有 M 对组合。利用 LMS(最小均方差)算法从这 M 对组合中经过一定次数的筛选,挑选出匹配最好的一部分组合计算两个图像序列的空间变换关系,作为粗配准结果,然后,采用基于区域的方法再对粗配准结果进行精配准。

(3) Caspi 等人基于特征的方法^[6]

该方法分为以下几个步骤:

①对图像序列中运动目标的特征进行提取和跟踪,生成运动轨迹(trajectory);

②将运动目标的轨迹与静止的背景区分开来;

③建立两个图像序列运动轨迹之间的初始对应关系,对运动轨迹进行配对(pairing),这可以结合运动轨迹的形状等信息来进行;

④随机选取一对对应的运动轨迹,通过迭代的方法同时计算两个图像序列在时间和空间上的对应关系 $P = (P_{\text{spatial}}, P_{\text{temporal}})$;

⑤用 P 配准所有对应的运动轨迹。统计经过配准后二者之间的距离小于给定阈值的运动轨迹的对数(为简便起见,称满足此条件的一对运动轨迹与 P 关联),并以此作为 P 的评价标准,与 P 关联的

运动轨迹的对数越多, P 就越好;

⑥重复第 4 步和第 5 步 N 次;

⑦选择最好的 P ;

⑧利用所有与 P 关联的成对运动轨迹对 P 进行细化。

在第 4 步和第 8 步中所求的 P 满足:

$$P = \arg \min_{P_{\text{spatial}}, P_{\text{temporal}}} \sum_T \left(\sum_{t \in T} \|p'(P_{\text{temporal}}(t)) - P_{\text{spatial}}(p(t))\|^2 \right) \quad (1)$$

式中, $p(t)$ 表示 t 时刻某条运动轨迹 T 在图像中的坐标点, $P_{\text{temporal}}(t)$ 表示对 t 用 P_{temporal} 进行时间变换后的结果, $p'(t')$ 表示与 T 对应的另一条运动轨迹 t' 时刻 ($t' = P_{\text{temporal}}(t)$) 在图像中所在位置的坐标, $P_{\text{spatial}}(p(t))$ 表示对点 $p(t)$ 用 P_{spatial} 进行空间变换后的结果。

在计算 P 时,首先固定 P_{temporal} 计算 P_{spatial} ,然后固定 P_{spatial} 计算 P_{temporal} ,如此几次,得到 P 。

该方法的优点是将 P_{temporal} 和 P_{spatial} 作为一个整体同时计算,空间配准精度可到达亚像素级,时间配准精度在 1 帧以内(亚帧级精度)。第 3 步是该方法的关键。如果运动轨迹的初始配对错误,必然影响计算结果。 P_{temporal} 的数学模型可表示为

$$P_{\text{temporal}}(t) = st + \Delta t \quad (2)$$

式中, s 为两个图像序列的帧频之比,是一个常数。其中的未知量是 Δt 。在计算 P 时,首先固定 P_{temporal} ,也就是固定 Δt 。给出一个时间段,在这个时间段里取整数值赋给 Δt (穷举法)。可是,在没有先验知识的情况下选取一个合适的时间段并非易事。在文献[6]中没有提及如何选取这个时间段,然而,这个问题却十分关键:如果时间段过窄,有可能根本找不到正确的 Δt ;而如果时间段过宽,那么 Δt 的可能取值就有很多,又会影响算法的效率。

(4) Caspi 等人基于区域的方法^[6]

这种方法将用于两个单幅图像配准的基于区域的(area-based)方法推广到了序列图像上。将高斯图像金字塔(Gaussian image pyramid)^[7]推广到高斯图像序列金字塔(Gaussian sequence pyramid):输入的序列图像作为金字塔的底层,对于金字塔的每层而言,在空间域和时间域进行低通滤波并下采样(sub-sampling)生成金字塔的上一层,采样因子为 2。如果将两个序列分别标记为 $S(x, y, t)$ (参考图像序列)和 $S'(x, y, t)$ (待配准图像序列),那么算法

准则是用最小化 SSD (sum of squared brightness differences) 来计算偏移量 \mathbf{u} :

$$ERR(\mathbf{P}) = \sum_{\mathbf{x}=(x,y,t)} (S(\mathbf{x}) - S'(\mathbf{x} + \mathbf{u}(\mathbf{x}; \mathbf{P})))^2 \quad (3)$$

利用 Gauss-Newton 法求解 \mathbf{P} :

$$\begin{aligned} & \sum_{\mathbf{x}=(x,y,t)} (\mathbf{J}_p^T \nabla S' \nabla S'^T \mathbf{J}_p) \mathbf{P} \\ & = \sum_{\mathbf{x}=(x,y,t)} (S' - S) \mathbf{J}_p^T \nabla S' \end{aligned} \quad (4)$$

式中, $\nabla S' = [S'_x, S'_y, S'_t]$ 为梯度向量, \mathbf{J}_p 为 Jacobian 矩阵, 其形式与 \mathbf{P} 有关。该方法的优点是可以得到很高的配准精度(亚像素级和亚帧级)。与基于区域的 2 维图像配准方法^[8]一样, 该方法不能处理两个图像序列之间存在较大的未对准的情况, 特别是在时间域上有较大偏差(例如几百帧)的情况。此外, 在式(4)中没有结合在单幅图像配准方法中使用的“outlier rejection”技术, 因此, 无法处理多传感器的情况。

(5) Caspi 等人的相机视场互不重叠的序列图像配准方法^[9]

该方法利用单个图像序列帧与帧之间的空间变换 T_i 和 T'_i 来求解两个序列的空间和时间对应关系。其优点是可以解决相机视场互不重叠时的图像序列配准问题。然而, 该方法的应用背景是两个相机要(同时)移动。对于相机静止的情况, 该方法则无法处理。因为这时 T_i 和 T'_i 都相等, 且为常数。

2 序列图像配准算法描述

在描述具体方法之前, 首先请看问题的描述:

(1) 两个传感器(同种或不同种)的空间位置相对固定, 在采像过程中静止不动, 两个传感器的视场有部分相重合, 这是比较常见的监视(surveillance)场景, 有很广泛的应用背景;

(2) 两个传感器的启动时间可以不同, 即一个传感器可以先启动, 另一个传感器可以后启动;

(3) 两个传感器的帧频率可以不同, 例如一个传感器可以是 PAL 制式(25fps), 另一个传感器可以是 NTSC 制式(30fps);

(4) 场景中有运动目标存在, 但在开始阶段有一段静止背景。

经过图像采集过程后得到两个图像序列 S_1 和 S_2 , 记两个序列的帧频率分别为 f_1 和 f_2 。考虑问题的普遍性, 这里设 $f_1 \neq f_2$ 且有 $f_1 > f_2$; S_1 中的第 i 帧

记为 $S_{1,i}$, S_2 中的第 j 帧记为 $S_{2,j}$; S_1 的长度为 l_1 , S_2 的长度为 l_2 。

算法的具体过程如下:

(1) 分别在 S_1 和 S_2 的开始部分(静止背景)取一定长度的连续帧图像, 得到长度为 n_1 的图像序列 $\{S_{1,i}; i=1 \sim n_1\}$ 和长度为 n_2 的图像序列 $\{S_{2,j}; j=1 \sim n_2\}$ 。对 $S_{1,i}$ ($i=1 \sim n_1$) 取平均值得到图像 I_1 , 对 $S_{2,j}$ ($j=1 \sim n_2$) 取平均值得到图像 I_2 。需要指出的是, 即使在没有目标出现的情况下, 背景也并不是绝对静止的。例如, 光照条件可能出现变化; 此外, 在风的作用下, 背景中的树木还会发生轻微的移动等等。鉴于这些情况, 对一定数量的背景图像取平均值。

(2) 对 I_1 和 I_2 进行配准, 具体的配准算法需要考虑实际的空间变换模型(例如仿射变换模型、投影变换模型等等)。这里考虑仿射变换模型。采用一种基于仿射变换模型的图像配准方法^[10]对 I_1 和 I_2 进行配准, 得到 S_1 和 S_2 之间的空间变换 t 。因为在图像的采像过程中传感器的位置静止不动, 所以, 这里求得的 t 可以大致描述分别属于 S_1 和 S_2 的任意两帧图像之间的空间变换关系。

(3) 求绝对值差图像序列 S'_1 和 S'_2 。 S'_1 中的第 i 帧图像是 $S'_{1,i} = |S_{1,i} - I_1|$, S'_2 中的第 j 帧图像是 $S'_{2,j} = |S_{2,j} - I_2|$ 。利用 $S'_{1,i}$ 的前 n_1 帧图像求得一个阈值图像 I_{th_1} , 其像素 (m, n) 的值为 $I_{th_1}(m, n) = \max_{i=1 \sim n_1} S'_{1,i}(m, n)$ 。同理可以求得阈值图像 I_{th_2} 。

(4) 利用 I_{th_1} 和 I_{th_2} 对 S'_1 和 S'_2 中剩余的图像进行分割, 分别得到两个图像序列 S''_1 和 S''_2 :

$$S''_{1,i}(m, n) = \begin{cases} 1 & S_{1,i}(m, n) > I_{th_1}(m, n) + th_1 \\ 0 & \text{其他} \end{cases} \quad (5)$$

$$S''_{2,j}(m, n) = \begin{cases} 1 & S_{2,j}(m, n) > I_{th_2}(m, n) + th_2 \\ 0 & \text{其他} \end{cases} \quad (6)$$

式(5)和(6)中的 th_1 和 th_2 为两个预先设定的阈值。 S''_1 和 S''_2 的长度分别为 $l_1 - n_1$ 和 $l_2 - n_2$ 。

(5) 建立两个列数均为 3 的空数组 s_1 和 s_2 。对 S''_1 中的第 i 帧图像 ($i = n_1 + 1 \sim l_1$), 剔除面积小于阈值 s_{th_1} 的连通区域。如果经过此处理后该图像中仍然存在连通区域(可能的运动目标), 则对结果图像进行一次数学形态学膨胀和腐蚀操作(先膨胀再腐蚀)。膨胀和腐蚀的目的是改善分割后连通区域的形状。计算图像中剩余的连通区域的质心 (x, y) , 并按照 $[i, x, y]$ 的形式将帧序号 i 和质心坐标 (x, y) 添加到数组 s_1 中。显然, 如果存在多个连

通区域,在 s_1 中就会出现多个 i 相同而 (x, y) 不同的情况。此外,为了降低问题的复杂性,还要求目标距离图像的边界不能太近,也即目标上的任意一点距离图像 4 个边界的距离不能小于给定的阈值 d_1 。在实际计算时,先计算包围每一连通区域的最小矩形,然后计算这个最小矩形的 4 个顶点与图像边界的距离,以此作为判断依据。只有满足这个距离约束条件的连通区域的质心才会被记入 s_1 中。同理对 S_2'' 中的每帧图像进行类似的处理,面积阈值设为 $s_{th,2}$,得到数组 s_2 。

(6) 利用连续性约束条件对 s_1 和 s_2 进行后处理。以 s_1 为例,要求其中的任意一项 $[i, x, y]$ 处于一个长度不小于 $d_{th,1}$ 的连续序列中: $\{\dots, [i-1, x', y'], [i, x, y], [i+1, x'', y''], \dots\}$ 。在这个序列中,相邻项之间的帧序号相差为 1,质心之间的距离不超过 d_2 , d_2 为一个小阈值,即 $(\sqrt{(x-x')^2 + (y-y')^2} \leq d_2, \sqrt{(x-x'')^2 + (y-y'')^2} \leq d_2)$ 。连续性约束条件的依据是,运动目标会在一定长度的帧中连续出现,并且在相邻帧中同一目标的位移不会太大。将 s_1 中满足连续性约束条件的 $[i, x, y]$ 的集合记为 $s'_1, s'_1 \subset s_1$ 。同理,给定阈值 $d_{th,2}$ 可以得到 s'_2 。设 s'_1 和 s'_2 的长度分别为 l'_1 和 l'_2 ,于是 s'_1 的大小为 $l'_1 \times 3, s'_2$ 的大小为 $l'_2 \times 3$ 。

(7) 建立一个与 S_1 中的图像大小相同的图像 I ,假设 I 有 R 行 C 列,其所有元素的初始值设为 0。进行以下操作:

for $k = 1$ to l'_1

$$i(\lceil s'_1(k,2), s'_1(k,3) \rceil) = 1 \quad (7)$$

式中, $\lceil \cdot \rceil$ 表示取整运算,例如坐标 $(4.5, 3.2)$ 经过取整运算后的结果是 $(5, 3)$ 。计算 I 的距离变换矩阵 D 和最近邻变换矩阵 N 。距离变换矩阵 D 中的元素 $D(m, n)$ 表示点 (m, n) 到 I 中与之最近的特征点(即 I 中元素值为 1 的点)的距离。最近邻变换矩阵 N 中的元素 $N(m, n)$ 表示 I 中与点 (m, n) 最接近的特征点(在图像的行数和列数确定的情况下,其中的像素 (m, n) 可以利用一个 1 维索引来表示,例如可以表示为 $(m-1)C + (n-1)$ 。 $N(m, n)$ 中存放的数值就是 1 维索引值)。

(8) 建立一个空数组 $s'_{1,2}$,其列数为 2,进行以下操作:

for $k' = 1$ to l'_2

$$\text{if } D(t(s'_2(k',2), s'_2(k',3))) \leq d_3$$

根据 $N(t(s'_2(k',2), s'_2(k',3)))$ 的值查找 s'_1 中与 k' 对应的 k 。将 $[s'_1(k,1), s'_2(k',1)]$ 加入 $s'_{1,2}$ 中;

其中, $t(s'_2(k',2), s'_2(k',3))$ 表示用 t 对点 $(s'_2(k',2), s'_2(k',3))$ 进行空间变换的结果。这个过程如图 1 所示。其中, d_3 为一个小阈值。需要指出的是,同一个 k' 可能会对应多个 k 。如果这多个 k 的值彼此比较接近,即 $\max(k) - \min(k) \leq d_4, d_4$ 为一个小阈值,则取它们的中间值与 k' 对应;否则,无法判断对应 k' 的 k ,于是放弃 k' 。

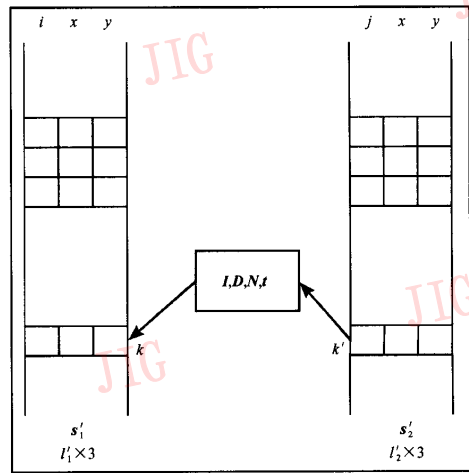


图 1 寻找 s'_1 和 s'_2 之间 k 与 k' 的对应关系

Fig. 1 Seeking the correspondence of k and k'

(9) 记向量 Y 为 $s'_{1,2}$ 的第 1 列,向量 X 为 $s'_{1,2}$ 的第 2 列。不难看出, Y 中存放的是图像序列 S_1 中的一部分图像帧的帧序号,而 X 中存放的是图像序列 S_2 中的一部分图像帧的帧序号。在图像序列帧频为确定值的情况下,可以认为在不同序列的对应帧之间存在线性关系。因此, X 和 Y 之间的关系可以利用一个线性函数来描述:

$$Y = bX + a \quad (8)$$

式中, b 和 a 为两个常数。需要求解这两个常数,并且验证这种线性关系是否成立。为此,采用数理统计中的一元线性回归模型^[11]来进行。假设 $s'_{1,2}$ 的长度为 $l_3, X = [x_1, x_2, \dots, x_{l_3}], Y = [y_1, y_2, \dots, y_{l_3}]$ 。变量 x_i 和 $y_i (i = 1 \sim l_3)$ 之间可以通过下式描述:

$$y_i = bx_i + a + \varepsilon_i \quad (9)$$

式中, $\varepsilon_i (i = 1 \sim l_3)$ 服从均值为 0, 方差为 σ^2 的正态分布,且 $\varepsilon_1 \sim \varepsilon_{l_3}$ 相互独立,则可以求得

$$b = \frac{\sum_{i=1}^{l_3} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{l_3} (x_i - \bar{x})^2} \quad (10)$$

$$a = \bar{y} - b\bar{x}$$

$$\sigma^2 = \left(\frac{1}{l_3} \sum_{i=1}^{l_3} y_i^2 - \bar{y}^2 \right) - b^2 \left(\frac{1}{l_3} \sum_{i=1}^{l_3} x_i^2 - \bar{x}^2 \right)$$

式中, $\bar{x} = \frac{1}{l_3} \sum_{i=1}^{l_3} x_i, \bar{y} = \frac{1}{l_3} \sum_{i=1}^{l_3} y_i$;

接下来,要验证线性回归是否显著。只有在线性回归是显著的条件下,计算的 b 和 a 才有意义。为此,计算

$$T = \frac{b \sqrt{\sum_{i=1}^{l_3} (x_i - \bar{x})^2}}{\sigma^*} \quad (11)$$

式中, $\sigma^* = \sigma \sqrt{\frac{l_3}{l_3 - 2}}$ 。如果 $|T| \geq t_{\frac{\alpha}{2}}(l_3 - 2)$, 则认为线性回归显著。其中, α 为显著性水平, $t(l_3 - 2)$ 表示自由度为 $(l_3 - 2)$ 的 t 分布。

(10) 对参数进行调整。不难看出,在第9步中求得的 b 和 a 是根据数组 $s'_{1,2}$ 中的元素求得的,而 $s'_{1,2}$ 又是从两个图像序列的运动目标质心之间的对应关系而得到的。而在通常情况下,受图像分割结果以及噪声等因素的影响,提取的质心并非十分精确可靠。此外,在第2步求得的 S_1 和 S_2 之间的空间变换 t 是以静止背景图像 I_1 和 I_2 为基准的。正如前面所分析的, I_1 和 I_2 本身也是多幅背景图像取平均值并合成后的结果;然而,两幅背景图像之间也有可能存在小的位移。这会影响到所求得的 t 的准确性。有鉴于此,必须对参数进行调整。这里需要对参数 b, a 和 t 都进行调整。设经过调整后的参数分别为 b', a' 和 t' 。其中,参数 b' 的值为固定值 f_1/f_2 。 b 与 b' 越接近越好。在实际应用中, S_1 为长波红外图像序列, S_2 为可见光图像序列, $f_1 = 30, f_2 = 25, b' = 30/25 = 1.2$ 。

下面推导一下当 $b \neq b'$ 时在时间域上的误差。假设两条直线 $y = bx + a$ 与 $y = b'x + a'$ 相交于点 (x_0, y_0) (图2)。

不失一般性,设 $b < b'$ 。由图2所描绘的关系,很容易证明

$$\Delta y = (b' - b) \Delta x \quad (12)$$

从上式可见,当 b 与 b' 很接近时,即使 Δx 的值很大, Δy 的值也会很小。也就是说,即使在远离

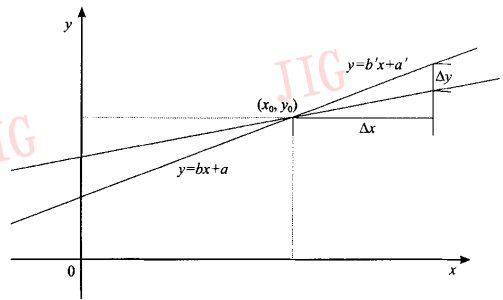


图2 时间域上的误差分析

Fig. 2 Error analysis of the time domain

(x_0, y_0) 处选择一点 (x, y) , 通过 $y = bx + a$ 计算得到的 y 与它的真实值 (利用公式 $y = b'x + a'$) 的误差 Δy 并不大。在实际应用中,无从知道 (x_0, y_0) 的实际位置,而是规定一个误差范围 $\Delta y \leq \Delta l$ 。 Δl 的确定视 b 与 b' 的接近程度而定,即两者越接近, Δl 的取值越小。一般地,取 $\Delta l = 3 \sim 5$ 。

此外,还有一个问题需要考虑:假设在某时刻 t_0 时, S_1 中的第 m 帧恰好对应于 S_2 中的第 n 帧。那么,经过时间 Δt 后,也即在 $t_0 + \Delta t$ 时刻, S_1 中的当前图像是第 $m + f_1 \Delta t$ 帧,而 S_2 中的当前图像是第 $n + f_2 \Delta t$ 帧。这便存在两个问题:一是,因为帧数都是整数,所以希望 $f_1 \Delta t$ 和 $f_2 \Delta t$ 都是整数才有意义,如果 $f_1 \Delta t$ 或 $f_2 \Delta t$ 不是整数,那么严格地说,在考察 $t_0 + \Delta t$ 时刻的两幅对应图像时,首先要通过序列图像帧间插值的办法分别内插出每一图像序列中对应 $t_0 + \Delta t$ 时刻的图像,然后在此基础上进行配准;二是,如果 $f_1 = f_2$, 当 $f_1 \Delta t$ 是整数时, $f_2 \Delta t$ 也是整数,在 $t_0 + \Delta t$ 时刻对应的图像都是整数帧,这种情形相对容易处理。但是,当 $f_1 \neq f_2$ 时,例如,所用的可见光/长波红外传感器帧频比为 $25/30, f_1 \Delta t$ 是整数时 $f_2 \Delta t$ 不一定是整数,反之亦然。严格地讲,这时也要进行图像的插值处理。为了避免引入插值运算从而增加计算负担,采取以下策略:

① 抽取向量 X 中的一部分样本 (X 中存放的是图像序列 S_2 中的一部分图像帧的帧序号,见第9步) 组成向量 X' 。设 X' 中的第1个样本是 S_2 中的第 j 帧 $S_{2,j}$, 第2个样本是 S_2 中的第 $j + 5p$ 帧 $S_{2,j+5p}$, 第3个样本是 S_2 中的第 $j + 10p$ 帧 $S_{2,j+10p}$, ..., 依此类推。其中, p 为不小于1的整数;

② 利用式(8)求得 S_1 中与 X' 中的第1个样本对应的图像帧,不妨假设该帧是 S_1 中的第 i 帧 $S_{1,i}$ 。

构造与向量 X' 长度相等的向量 Y' , Y' 中的第 1 个样本是 $S_{1,i}$, 第 2 个样本是 $S_{1,i+6p}$, 第 3 个样本是 $S_{1,i+12p}$, ..., 依此类推。不难看出, 如果 X' 中的第 1 个样本与 Y' 中的第 1 个样本相对应 (也即帧 $S_{2,j}$ 与帧 $S_{1,i}$ 在时间域上相对应), 那么显然帧 $S_{2,j+5p}$ 与帧 $S_{1,i+6p}$ 相对应, $S_{2,j+10p}$ 与帧 $S_{1,i+12p}$ 相对应, ...。显然地, 在构造 X' 和 Y' 时, 已经将 b' 的值固定为 1.2 了。

前面已经提到, 利用式 (8) 所求得的帧与帧之间的对应关系存在一定的误差, 而且误差范围在 Δl 之内。据此通过一个变量 Δj 对 X' 中的内容进行调整如下 (假设 X' (Y') 的长度为 l'_3):

$$\begin{aligned} S_{1,i} &\longrightarrow S_{2,j+\Delta j} \\ S_{1,i+6p} &\longrightarrow S_{2,j+\Delta j+5p} \\ S_{1,i+12p} &\longrightarrow S_{2,j+\Delta j+10p} \\ &\vdots \\ S_{1,i+(l'_3-1)6p} &\longrightarrow S_{2,j+\Delta j+(l'_3-1)5p} \end{aligned}$$

$$\Delta j = -\Delta l, -\Delta l+1, \dots, -1, 0, 1, \dots, \Delta l-1, \Delta l$$

在求解 t' 时, 以 t 作为它的初始值。于是, 可以采用文献 [8] 中的基于区域的图像配准算法。为了处理多幅图像之间的配准问题, 需要对原算法进行一定程度的修改。为简单起见, 设 X' 中的图像为 $X' = \{g_1, g_2, \dots, g_{l'_3}\}$, Y' 中的图像为 $Y' = \{f_1, f_2, \dots, f_{l'_3}\}$, 对文献 [8] 中的 $S_i^{(x,y)}(u,v)$ 进行修改:

$$S^{(x,y)}(u,v) = \sum_{q=1}^{l'_3} f_q(x,y) \circ_N g_q(x+u,y+v) \quad (13)$$

式中, “ \circ_N ” 代表归一化相关 (normalized correlation)。配准的具体过程请参考文献 [8]。

不同的 Δj 对应不同的 t' , 接下来需要确定哪一个 t' 是最佳的。为此, 用共同信息作为判据。给定两个随机变量集 A 和 B , 两者之间的共同信息 $I(A, B)$ 定义为^[12]

$$MI(A, B) = \sum_{a \in A, b \in B} p_{AB}(a, b) \log_2 \frac{p_{AB}(a, b)}{p_A(a)p_B(b)} \quad (14)$$

式中, $p_{AB}(a, b)$ 为 A 和 B 的联合概率分布, $p_A(a)$ 和 $p_B(b)$ 分别为 A 和 B 的边缘概率分布。

共同信息的计算量与图像的大小有关。一般来说, 图像越大, 共同信息的计算量也相应越大。为了减小计算量, 采取不在整个图像上计算共同信息, 而是只计算目标所占的部分区域 (即感兴趣的区域, region of interest-ROI) 的共同信息。

在 s'_i 中查找 $Y' = \{f_1, f_2, \dots, f_{l'_3}\}$ 中图像 f_i ($i =$

$1 \sim l'_3$) 所对应的目标的质心坐标 (x_i, y_i) , 为简单起见, 这里假设每幅图像中只有一个目标, 但事实上, 本方法也适合多目标的情况。在 f_i 中划定一个以 (x_i, y_i) 为中心, 大小为 $w \times h$ 的区域 A_i 。对于 Y' 中的图像 f_i , X' 中与之对应的图像是 g_i , 用所求得的 t' 对 g_i 进行空间变换, 得到图像 g'_i 。在 g'_i 中找到与 A_i 对应的区域, 然后对这两个区域计算共同信息 MI_i 。由于 MI_i 与 t' 有关, 记为 $MI_i(t')$ 。使得所有这些 MI_i 的总和最大的 t'^* 即为所求:

$$t'^* = \arg \max_{t'} \sum_{i=1}^{l'_3} MI_i(t') \quad (15)$$

第 1 步和第 2 步的目的是对静止背景进行对准。第 3 ~ 6 步的目的是提取连续的运动目标的质心; 第 7 步和第 8 步的目的是通过静止背景的空间配准参数来找到不同图像序列中运动目标质心的对应关系, 这两步避开了 Caspi 等人基于特征的方法中的“建立运动质心轨迹对应关系”这个比较复杂的问题; 第 9 步利用数理统计的方法给出了两个图像序列之间在时间域上的初步对应关系, 不难看出, 即使没有先验知识 (即, 知道两个传感器启动的大概时间间隔) 也无妨; 第 10 步则结合共同信息来解决多传感器之间成像性质不一致的问题。这里还要指出, 为了进一步提高速度, 在第 10 步中可以考虑引入并行处理的方法。一个简单的想法是为不同的

t' 分配一个处理器来计算 $\sum_{i=1}^{l'_3} MI_i(t')$ 。

3 实验结果

实验 1

可见光和长波红外两个序列图像的配准。实验中的长波红外图像序列由 PalmIR 250 型 7 ~ 14 微米长波红外热像仪 (视频信号帧频 30fps) 输出, 长度为 715 帧; 可见光图像序列由 Panasonic 手持式摄像机 (视频信号帧频 25fps) 输出, 长度为 440 帧。

在第 1 步中, 对长波红外图像序列前 250 帧取平均值后得到背景图像 I_1 , 可见光图像序列前 80 帧取平均值得到背景图像 I_2 。 I_1 和 I_2 分别如图 3 的 (a) 和 (b) 所示; 在第 2 步中, 对 I_1 和 I_2 进行配准, 得到空间变换 $t = (1.1388, 0.1139, -0.0085, 0.9421, -48.1908, -7.9703)$; 在第 3 步中求得的阈值图像 I_{th-1} 和 I_{th-2} 分别如图 3 的 (c) 和 (d) 所示; 在第 4 步中 th_1 和 th_2 均取为 10; 在第 5 步中面

积阈值 s_{th_1} 和 s_{th_2} 均取为 20, d_1 取为 10; 在第 6 步中 d_{th_1} 和 d_{th_2} 均取为 20, d_2 取为 5; 用 t 将 s'_2 中的质心坐标向 s'_1 的质心坐标所在的图像平面进行空间变换, 得到两者之间的空间位置关系, 如图 3(e) 所示; 在第 8 步中 d_3 和 d_4 均取为 3; 在第 9 步中求得的向量 Y 和 X 的内容绘制在图 3(f) 中, 它们的长度均为 164 帧, $b = 1.175, a = 157.64, T = 555.53$, 显著性水平 $\alpha = 0.01$ 时, 经查 t 分布上侧分位数表, 可得 $t_{\frac{\alpha}{2}}(162) \approx 2.58$, 线性回归显著; 在第 10 步中, 在构造向量 X'

和 Y' 时, $p = 2$; X' 的初始内容为 $X' = \{201, 211, 221, 231, 241, 251, 261, 271, 281, 291, 301, 311\}$, 得到 $Y' = \{394, 406, 418, 430, 442, 454, 466, 478, 490, 502, 514, 526\}$, 取 $\Delta l = 3$ 。经过利用共同信息进行验证后, 在 $\Delta j = 1$ 时, 得到最优的 $t'^* = (1.1344, 0.1027, -0.01306, 0.9384, -44.48, -6.16)$ 。这时 $X' = \{202, 212, 222, 232, 242, 252, 262, 272, 282, 292, 302, 312\}$, $b' = 1.2, a' = 151.6$ 。利用 b', a' 和 t' 对两个序列的对应帧进行配准, 结果如图 3(g) ~ (o)。

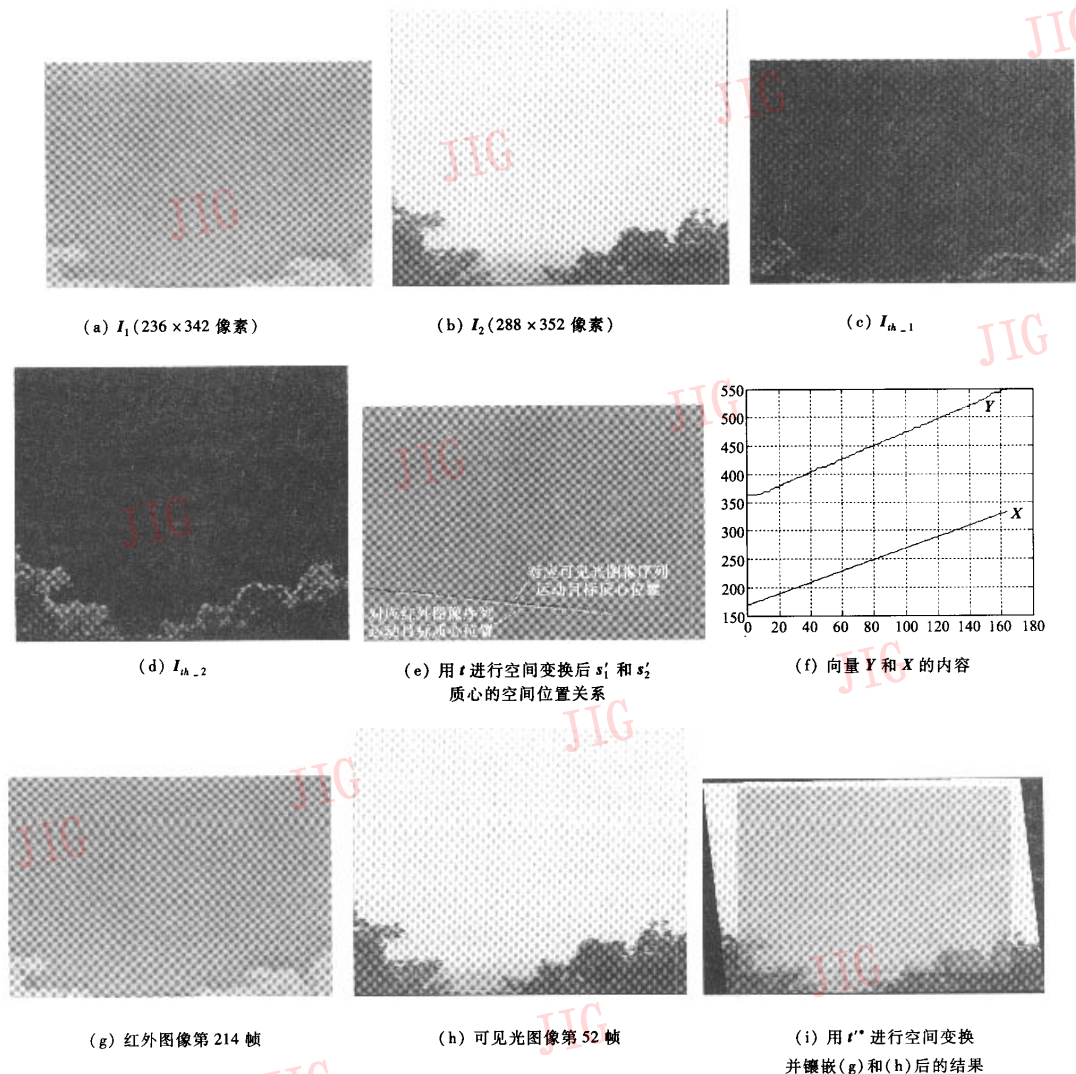
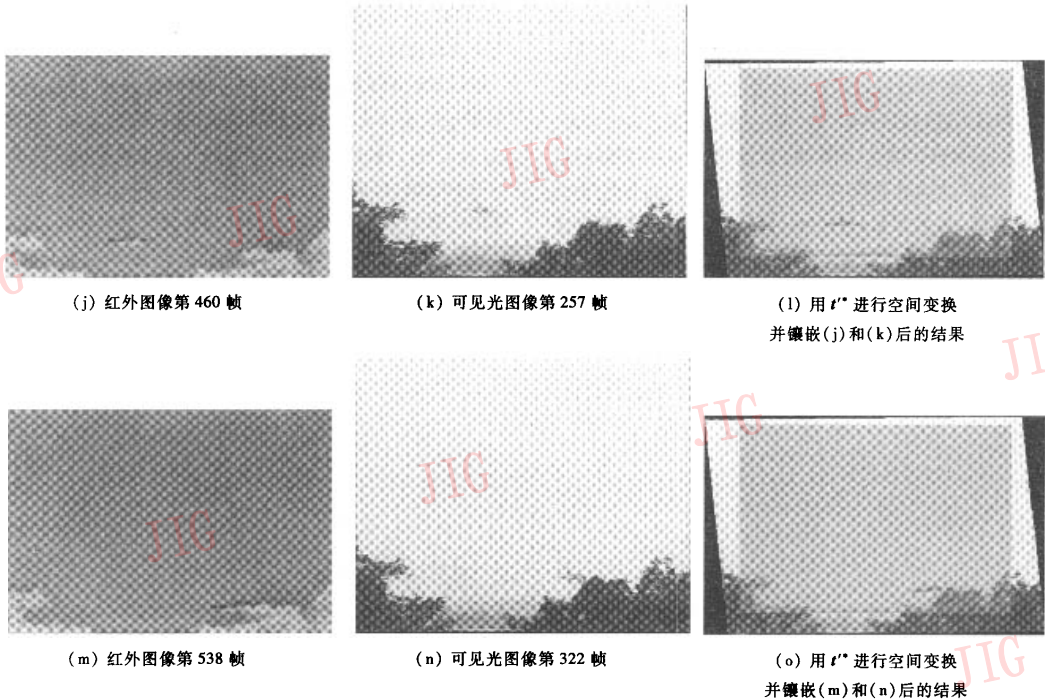


图 3 红外和可见光图像序列配准结果

Fig. 3 Results of aligning infrared and visual image sequences



续图 3

Continue Fig. 3

实验 2

用一段可见光图像序列 (25fps) 经过帧间插值的办法生成一段帧频为 30fps 的另一段序列, 来模拟与之对应的红外图像序列。虽然模拟的图像并不能代替真实的红外图像, 但它毕竟可在某种程度上为本文算法提供一个客观的评价。生成的图像序列与原图像序列在时间上的对应关系为 $Y = 1.2X - 0.2$ 。接着, 对生成的图像序列经过帧间插值在时间轴上进行一个 0.4 帧的位移, 得到一个新的图像序列。这个新的图像序列与原图像序列在时间上的对应关系为 $Y = 1.2X - 0.6$ 。利用本文算法, 得到的最终结果为 $Y = 1.2X - 0.2$, 与实际结果有 0.4 帧的偏差, 究其原因, 是该算法的第 10 步中没有进行帧间的插值, 而是寻找两个图像序列在完整帧基础上的最大可能的对应关系。不难看出, 尽管如此, 该算法在时间域上的误差在理论上不会超过 0.5 帧。在该实验中, 计算得到的空间变换与实际结果相比, 误差约为 0.3 像素。

4 结论

在摄像机静止条件下序列图像配准的计算框架,

有效地克服了现存方法中的一些问题。虽然文中所考虑的空间变换模型为仿射变换模型, 但其实质同样适用于其他的空间变换模型 (例如, 透视变换模型)。换言之, 对于不同的空间变换模型只需要对框架的相应部分进行修改即可。该框架能够处理多传感器图像序列的配准问题。在两个传感器的启动时间相差很大的情况下 (这时两个图像序列之间有几百帧的偏差或更多), 该框架也能很好地处理。

本文方法是在整数帧基础上 (未经过帧间插值) 求解两个图像序列帧与帧之间最可能的对应关系。从理论上讲, 在时间域上的最大误差为 0.5 帧。在这种情况下, 如果目标的实际运动速度超过 2 像素/秒, 则对于运动目标的配准误差会超过 1 个像素。因此, 为了获得更好的配准结果, 需要在原框架的基础上引入帧间插值的办法。

参考文献 (References)

- 1 Zitová B, Flusser J. Image registration methods: a survey[J]. Image And Vision Computing, 2003, 21(11): 977 ~ 1000.
- 2 Brown L G. A survey of image registration techniques [J]. ACM Computing Surveys, 1992, 24(4): 325 ~ 376.
- 3 Sharma R K, Pavel M. Registration of video sequences from multiple

- sensors[A]. In: Proceedings of Image Registration Workshop[C], NASA Goddard Space Flight Center, Greenbelt, Maryland, USA, 1997; 361 ~ 366.
- 4 Bergen J R, Anandan P, Hanna K J, *et al.* Hierarchical model-based motion estimation [A]. In: Proceedings of The 2nd European Conference on Computer Vision[C], Santa Margherita, Italy, 1992; 237 ~ 252.
- 5 Stein G P. Tracking from Multiple View Points: Self-Calibration of Space and Time[A]. In: Proceedings of IEEE Computer Vision and Pattern Recognition Conference[C], Fort Collins, Colorado, USA, 1999; 521 ~ 527.
- 6 Caspi Y, Irani M. Spatio-temporal alignment of sequences[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(11): 1409 ~ 1419.
- 7 Burt P J, Adelson E H. The laplacian pyramid as a compact image code[J]. IEEE Transactions on Communications, 1983, 31(4): 532 ~ 540.
- 8 Irani M, Anandan P. Robust multi-sensor image alignment[A]. In: Proceedings of 6th International Conference of Computer Vision[C], Bombay, India, 1998; 959 ~ 966.
- 9 Caspi Y, Irani M. Aligning non-overlapping sequences [J]. International Journal of Computer Vision, 2002, 48(1): 39 ~ 51.
- 10 Peng Xiaoming, Ding Mingyue, Zhou Chengping, *et al.* A practical two-step image registration method for two-dimensional images[J]. Information Fusion, 2004, 5(4): 283 ~ 298.
- 11 Wang Rongxin. Mathematical Statistics[M]. Xi'an, China: Xi'an Jiaotong University Press, 1986; 174 ~ 193. [汪荣鑫. 数理统计[M]. 西安:西安交通大学出版社,1986;174 ~ 193.]
- 12 Maes Frederik, Collignon Andre, Vandermeulen Dirk, *et al.* Multimodality image registration by maximization of mutual information [J]. IEEE Transactions on Medical Imaging, 1997, 16(2):187 ~ 198.