

# 视频序列的全景图拼接技术

朱云芳 叶秀清 顾伟康

(浙江大学信息与通信工程研究所, 杭州 310027)

**摘要** 提出了一种对视频序列进行全景图拼接的方法。主要讨论了有大面积的非刚性运动物体出现的序列, 不过此方法也同样适用于无运动物体的纯背景序列。为计算各帧间的投影关系, 用仿射模型来描述摄像机运动, 并用特征点匹配的方法计算出模型中各参数的值。由于用相关法计算的匹配结果准确率比较低, 所以用RANSAC (Random Sampling Consensus)对匹配结果进行了筛选, 可以准确求出摄像机运动参数。利用运动参数进行投影, 然后用多帧相减并求交集, 估计出每帧图像中运动物体存在的区域, 最后计算得到了全景图。该方法的结果与前人得到的结果进行了比较, 证明用此方法能获得质量较高的全景图。

**关键词** RANSAC 全景图拼接 运动物体检测

**中图分类号**: TP391 **文献标识码**: A **文章编号**: 1006-8961(2006)08-1150-06

## Mosaic Panorama Technique for Videos

ZHU Yun-fang, YE Xiu-qing, GU Wei-kang

(Department of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027)

**Abstract** In this paper we propose a method for making panoramas from video sequence taken by moving cameras. We mainly describe the case where the camera is moving and there are large-area non-rigid moving objects in the scene. But the method also works well when there are no moving objects in videos. Affine model is used to describe the camera motion. To calculate the parameters in affine model, we use feature points matching method. Correlation method is used to match feature points in frames, but correlation method can only give a coarse result which may include some bad matching. To ensure the precision of the motion parameters, RANSAC is used to pick out bad matching. Project frames with motion parameters, and the intersection of two subtraction results of neighbor frames will give the moving object area. Last we can calculate the panorama image. Results given by this method are provided in this paper, and we compare them with other people's results. It proves that our method can make panoramas with high quality.

**Keywords** RANSAC, motion panorama, moving object detection

## 1 引言

序列图像的全景图拼接是近年来的研究热点, 但其中大多数的研究集中于静态的序列图像<sup>[1-5]</sup>。由于视频处理的普遍性和重要性, 对视频序列做全景图拼接具有重要的实用价值。将一个视频序列拼接为一幅视野范围较大的图片, 可使人们对该序列的内容有个直观的了解。因此, 该研究对视频压缩、

视频增强、视频分割等领域都具有重要意义。

相对于静态序列图像来说, 视频序列图像的全景图拼接具有其特殊性。视频序列中图像帧的数目通常要比静态序列图像多得多, 且相邻帧之间有很大的重复区域, 如何用这些重复信息生成所需的全景图? 另外, 如果视频序列中包含运动物体, 如何将其从背景中分割出来并去除其干扰, 使得到的全景图只包括纯背景。Irani对有运动物体的视频进行了研究, 但他提出的算法只能处理序列中有很小运

基金项目: 国家自然科学基金项目(60502006)

收稿日期: 2004-10-29; 改回日期: 2005-09-09

第一作者简介: 朱云芳(1980 ~ ), 女, 浙江大学信息与通信工程研究所博士研究生。主要研究方向为图像处理、计算机视觉。E-mail: ffbird@gmail.com

动物体的情况<sup>[6]</sup>。Bartoli 对 Irani 的方法进行了改进,提出了特征匹配和图像直接匹配相结合的方法,可以处理有较大运动物体的序列<sup>[7,8]</sup>。但就 Bartoli 给出的结果而言,其算法对序列的背景全景图拼接并不是十分理想,拼接而得的全景图中存在没有恢复信息的空白区域,运动物体在背景上留下痕迹,图像比较模糊。

本文用仿射变换对摄像机运动建模,并采用特征点匹配和 RANSAC<sup>[9]</sup>相结合的方法求解出模型参数。为了得到纯净的背景全景图,用相邻帧相减的方法对运动物体区域进行粗略估计,并利用视频序列相邻帧图像重叠区域所提供的信息,对有运动物体的区域进行有效填充,从而去除了运动物体的影响。

## 2 算法描述

### 2.1 特征点检测和初始匹配

首先对视频序列的每一帧图像提取具有亚像素精度的 Harris 角点<sup>[10]</sup>作为特征点,对相邻两帧的特征点进行相关匹配得到初始匹配结果。

假设已经对第  $i$  和  $i+1$  帧图像提取了角点,用  $x_{i,m}$  表示在第  $i$  帧图像中的第  $m$  个特征点,其坐标为  $(u, v)$ 。在第  $i+1$  帧图像中以  $(u, v)$  为中心,在预先设定好大小的区域内搜索角点,作为  $x_{i,m}$  的候选匹配点。搜索区域的大小跟摄像机的运动及视频的帧率有关,通常情况下可取  $40 \times 40$ 。假设共可找到  $n$  个候选匹配点,其坐标分别为  $(u'_j, v'_j)$ ,  $j = 1, 2, \dots, n$ 。分别计算这些候选匹配点与  $x_{i,m}$  的 SAD (sum of absolute difference) 相关值,如式(1)所示。

$$d_{\text{SAD}} = \sum_{l=-N}^N \sum_{k=-N}^N |f(x_{i,m}(u+k, v+l)) - f(x_{i+1}(u'+k, v'+l))| \quad (1)$$

式中,  $f(x)$  表示点  $x$  的灰度值(若是彩色图像则分别计算 RGB 3 个分量)。

候选匹配点中 SAD 值最小的点被认为是  $x_{i,m}$  的匹配点。值得注意的是,由于相邻帧图像间提取的角点并不总是一一对应的,因此并不是所有角点都在其相邻帧中有对应的匹配角点。所以可以设置一个阈值  $T$ , 只有当求出的 SAD 最小值小于此阈值时,才认为这是一个可靠的初始匹配点对。

### 2.2 用 RANSAC 求摄像机运动参数

通常情况下,摄像机与景物之间的距离较远,可

用仿射变换对摄像机运动建模。对  $x_{i,m}$  和它在  $i+1$  帧中的对应点  $x_{i+1,m}$ , 其仿射变换模型可表示为

$$x_{i,m} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} x_{i+1,m} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \quad (2)$$

简写为  $x_i = P x_{i+1}$ 。

这里  $x_{i,m}$  表示第  $i$  帧图像中第  $m$  个点的图像坐标。此仿射模型有 6 个未知数,因此每 3 对不在同一直线上的匹配点可以确定一组摄像机运动参数。

在上节,已经用 SAD 相关的方法对提取的角点进行了初始匹配。但是由于相关匹配准确率较低,而且即使在正确的匹配结果中,有一部分会出现在运动物体上。由于运动物体和背景有不同的运动参数,所以直接用初始匹配结果结合最小二乘法计算运动参数会带来很大的误差。

采用 RANSAC 对初始匹配结果进行筛选,剔除错误以及与摄像机运动不符的匹配点。RANSAC 对样本数据进行多次随机取样,每次随机取出尽可能少但足够多个样本来确定模型参数,再根据已确定的模型对所有样本进行划分,一部分样本在此模型的一定误差范围内,一部分样本在误差范围外。因为坏样本是杂乱无章的异常数据,它所确定的模型,落在误差范围内的数据占少数,大部分数据都落在误差范围外。而对由好样本所确定的模型逼近于真实的模型,大多数样本就会落在误差范围内。经过多次随机取样试验后,RANSAC 算法找出落在误差范围内最多的点的集合,用此集合来进行最优化,最终确定模型的参数。

在初始匹配结果中随机选取 3 对匹配点,计算出一组运动参数  $P$ 。对剩下的每对匹配点,计算距离

$$d = \|x_{i,k} - P x_{i+1,k}\| \quad (3)$$

式中,  $\|\cdot\|$  表示欧氏距离,当此距离小于某经验阈值  $T$  时,认为这对匹配满足当前的运动参数,否则认为不满足。记下满足运动参数的匹配对的数目  $S$ 。经过若干次这样的随机采样后(随机采样的次数在文献[8]中有经验公式给出),得到  $S$  最大的匹配集合,再对此匹配集合用最小二乘法求摄像机运动参数。只要初始匹配中有大于 50% 的正确匹配点在背景上,就可以得到正确的摄像机运动参数。

### 2.3 运动物体检测

运动物体检测的重要性在于,如果不去除运动物体区域而直接进行拼接,所得到的是背景和运动物体叠加的结果;同时,由于运动物体在每帧图像中

的位置不同,会使生成的全景图杂乱而无意义。因此,首先要分割出每帧图像中包含运动物体的区域,然后用剩下的背景图像做全景图拼接。

视频处理中运动物体分割是一个比较困难的问题,许多学者对此做过相关的研究<sup>[11-13]</sup>,但至今仍没有一种非常有效的方法可以准确分割出任意帧中的运动物体。由于全景图拼接,只需将包含运动物体的大致区域分割出来,并在后续的拼接过程中剔除其干扰即可,而不需要精确分割出每帧图像中的运动物体。即使分割不十分准确,由于视频帧之间有大量的信息冗余,因此可以利用其他帧的信息对当前帧进行弥补,而不会影响最后的拼接效果。为加快算法的速度,采用相邻帧相减的方法来检测运动物体。

相邻帧相减利用了不同的运动物体具有不同的运动参数,只有静止的背景服从摄像机运动参数的性质。利用上节所得的摄像机参数投影后,背景图

像相当于静止,然后进行两帧图像相减,简单阈值化后就得到运动物体区域。

在上节中得到了相邻帧间的摄像机运动参数,用此参数可将一帧图像  $I_{i+1}$  投影到相邻帧  $I_i$  的坐标系中,用双线性插值法得到一帧新图像  $I'_{i+1}$ 。如果不考虑每帧图像的视场范围及噪声和光照带来的影响,在序列中没有运动物体时,  $I'_{i+1}$  和  $I_i$  是同样的图像;若序列中存在运动物体,  $I'_{i+1}$  和  $I_i$  会在有运动物体的区域内有差别(如图 1(e)和(f)所示)。因此,可定义  $I_i$  中的运动物体区域为

$$\{x_i | x_i \in I_i, |f(x_i) - f(x'_{i+1})| > T\} \quad (4)$$

式中,  $f(x_i)$  表示  $I_i$  中位置为  $x_i$  点的像素值,  $f(x'_{i+1})$  则是由  $I_{i+1}$  投影到  $I_i$  所生成的新图像  $I'_{i+1}$  中坐标为  $x'_{i+1}$  的像素值,  $T$  是所设定的阈值。当对应点相减的结果大于此阈值  $T$  时,认为此点是运动物体上的点,否则视为静止的背景点。

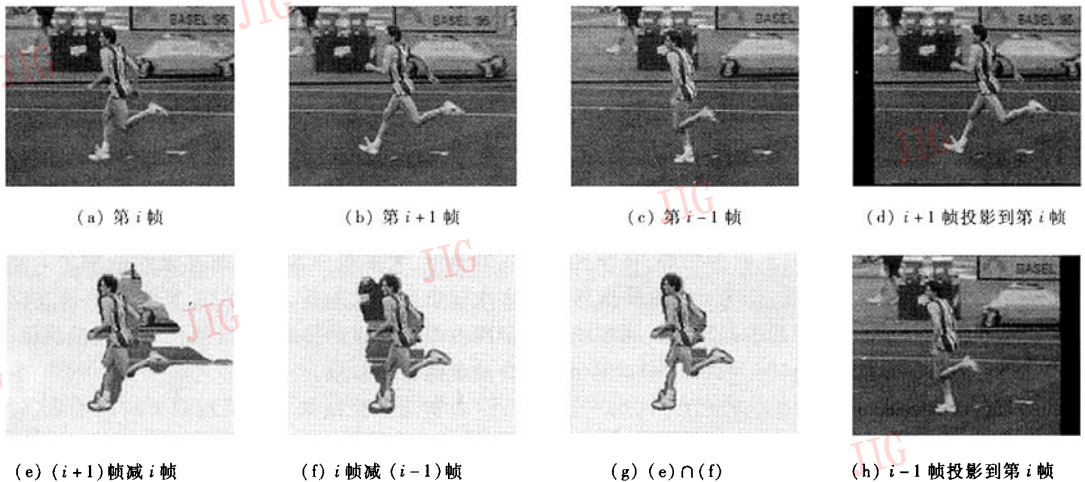


图 1 运动物体检测示意图

Fig.1 Detection of moving object

从图 1(e)和(f)中可以发现,只用两帧图像相减得到的并不是真正的运动物体区域,其中一部分是静止背景(此位置在另一帧中是运动物体)。为了尽可能多地保留每帧图像中静止背景图像的信息,可再加入第 3 帧的信息,即用前后帧图像来估计当前帧中的运动物体,因此对式(4)进行了修改,定义满足下式的点集为运动物体在当前帧的区域

$$\{x_i | x_i \in I_i, |f(x_i) - f(x'_{i+1})| > T \& |f(x_i) - f(x'_{i-1})| > T\} \quad (5)$$

其结果如图 1(g)所示。

帧相减方法的优点是计算简单,其缺点是需要预先计算运动参数,且如果运动物体和背景颜色一致,就无法检测出运动物体。此外,噪声对检测结果的影响也很大,因此有必要对帧相减的结果用形态学方法进行腐蚀和膨胀运算,把“运动物体区域”中小于某一规定大小的区域滤除。

## 2.4 全景图生成

假设  $x_i$  代表第  $i$  帧图像上点的坐标,  $P_{i,j}$  代表  $i, j$  两帧间的投影关系,即

$$\begin{aligned}
 x_1 &= P_{1,2}x_2 \\
 x_2 &= P_{2,3}x_3 \\
 &\vdots \\
 x_{n-1} &= P_{n-1,n}x_n
 \end{aligned}
 \tag{6}$$

由此可以计算出每帧和第 1 帧之间的投影关系:

$$x_1 = P_{1,2}P_{2,3}\cdots P_{n-1,n}x_n = P_{1,n}x_n \tag{7}$$

其中  $P_{1,n} = P_{1,2}P_{2,3}\cdots P_{n-1,n}$

选取第 1 帧为参考帧,就得到了每帧与全景图的投影关系,根据此关系计算出全景图的大小。

对于全景图上的点  $x$ ,通过投影的逆变换,得到此点在第  $i$  帧图像中的位置  $x_i = P_{1,i}^{-1}x$ ,在第  $i$  帧图像中用双线性插值法得到该点  $x_i$  的灰度值(彩色值)。全景图中的点可能会在序列中的  $N$  帧图像中都出现,因此在计算全景图中该点的值时,要充分利用此点在  $N$  帧中所有出现的值。考虑到运动物体的存在以及光照等原因的影响,对这些值进行简单的处理。

对于全景图上的任意点  $x$ ,求其灰度(彩色)值  $f(x)$  的步骤如下:

(1) 计算此点在序列中出现时的位置  $x_i$  和灰度,并利用上节中对运动物体检测结果对  $x_i$  标

记(静止背景或运动物体)。

(2) 如果点  $x$  共在  $N$  帧中出现,共得到此点的  $N$  个值  $(x_1, x_2, \dots, x_N)$ ,其中,有  $M$  个  $(M \leq N)$  标记为运动物体,那么将此  $M$  个值排除,用剩下的  $N - M$  个点的值进行计算。

(3) 对于剩下的点,计算下式:

$$f(x) = \frac{\sum_i w_i f(x_i)}{\sum_i w_i} \tag{8}$$

其中,  $w_i = \exp\left\{-\sum_{R(N)} |f(x_i) - f(x_{i,R(N)})|\right\}$ ,  $x_{i,R(N)}$  表示点  $x_i$  的邻域,用邻域点的值进行加权,可以去掉未检测出的运动物体点和光照带来的影响。

### 3 实验结果及分析

图 2 中给出一个跳高视频序列在做运动物体检测时的某些帧结果。

从图 2(a)的结果中可以看出,相减得到的初始结果受到噪声影响,会把一些本来为背景的图像当成运动物体。通过形态学方法处理后,图 2(c)的结



(a) 原始序列



(b) 3 帧分别相减后再求交的结果



(c) 用形态学方法进行处理后的结果(标记为白色的部分表示检测到的运动物体)

图 2 对运动物体的检测结果

Fig. 2 Results of moving object detection

果要好很多。如在图 2(b) 的第 2 列图像中出现的胳膊上的一小块被当成了静止背景, 经过填充后解决了这个问题, 而且很多噪声区域也在腐蚀时被滤掉了。但是在第一列图像中, 由于相邻两帧中运动员的右脚未移动, 所以右脚中的一部分被当成了运动物体, 这种问题用帧相减的方法无法解决。

图 3 中给出了图 2 序列的全景图拼接结果。由于拼接所得图像不是一幅规整的图像, 为了服从一般图像的格式(矩形), 实验中对摄像机未扫过的区域用黑色进行了填充。从结果中看不出拼接产生的接缝, 而且图像质量很好。为了与 Bartoli 的算法进行比较, 图 4 给出了对图 3 中圈定部分的放大细节。

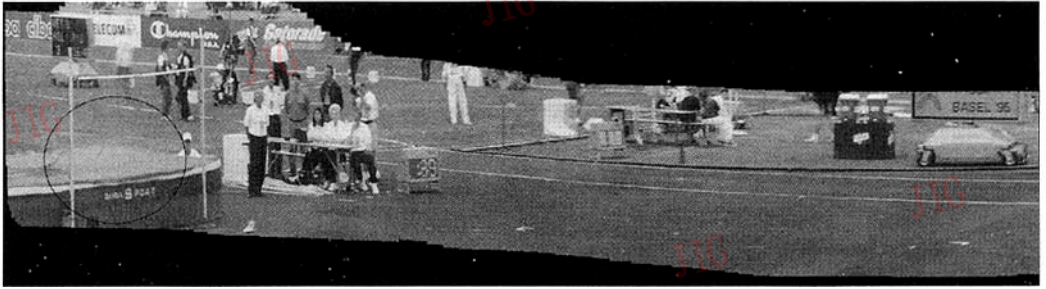
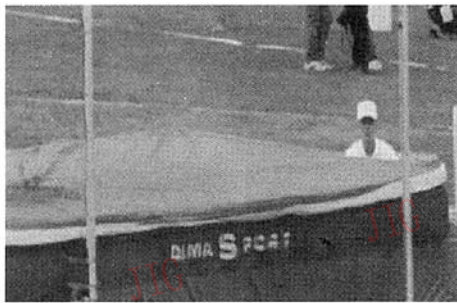
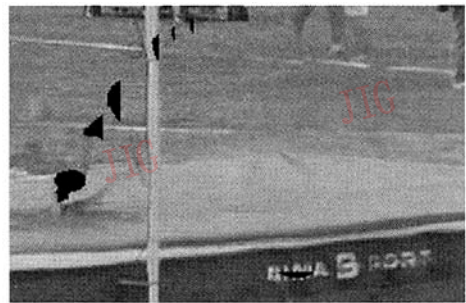


图 3 全景图拼接结果

Fig. 3 Result of frames mosaic



(a) 本文算法的结果



(b) Bartoli 算法的结果

图 4 拼接结果的比较

Fig. 4 Comparison of our result with Bartoli's

由图 4 不难看出, 本文的拼接结果要优于 Bartoli 所提出的算法。主要体现在以下 2 个方面:

(1) Bartoli 算法对运动物体检测的精度要低, 如图 4(b) 中黑色区域实际上并不是无信息区域, 而是被摄像机扫过并且背景图像可见的, 其原因是 Bartoli 在对运动物体做检测时, 将运动物体在前后 3 帧中扫过的区域都当成了当前帧的运动物体, 这很大程度上减少了可利用的静止背景的信息量。相比之下, 本文算法求取前后 3 帧结果的交集要合理得多;

(2) Bartoli 算法的结果中还可以明显看到运动员腿部留下的痕迹, 这是因为在运动物体检测时这部分被当成了背景, 而在计算全景图中点的像素值时, 又保留了这些点的信息, 所以使运动物体在背景

图上留下明显的痕迹。而本文算法用邻域点的值进行加权, 加权后这部分点的影响大大降低了, 由此得到了较为干净的背景图像。

## 4 结论

提出了一种对视频序列图像全景图拼接的新方法。该方法对运动物体的性质(比如是否刚体等)和背景(是否复杂)没有任何约束, 故此方法能有效地应用到很多序列中, 但同时也给从序列中检测运动物体带来了困难。为了避免运动物体检测错误影响全景图拼接, 用多帧图像信息来求全景图中颜色值的方法。大量实验结果的效果都很不错。得到全

景图之后,进一步的工作就是考虑用此结果进行序列中运动物体的分割,利用全景图可以为每帧图像提供一个背景,在此条件下,可以比较精确分割出序列中的运动物体。

### 参考文献 (References)

- 1 SUN Xiao-yong, Dubois E. A novel algorithm to stitch multiple views in image mosaics [A]. In: IEEE International Conference on Acoustics Speech, and Signal Processing [C], Montreal, Quebec, Canada, 2004, 3:481 ~ 484.
- 2 LI Yong-min, XU Li-qun, Morrison G, *et al.* Robust panorama from MPEG video [A]. In: IEEE International Conference on Multimedia and Expo [C], Baltimore, MD, USA, 2003, 1:81 ~ 84.
- 3 Shum H Y, Szeliski R. Construction of panoramic mosaics with global and local alignment [J]. *International Journal of Computer Vision*, 2000, 36(2):101 ~ 130.
- 4 ArcSoft Panorama Maker. [OL/EB]. <http://www.arcsoft.com/en/products/panoramamaker>, 2001.
- 5 REALVIZ S A. *Stitcher 3.5 Software and User's Manual*, [OL/EB]. <http://www.realviz.com/>, 2002.
- 6 Irani M, Anandan P, Bergen J, *et al.* Efficient representations of video sequences and their applications [J]. *Signal Processing: Image Communication*, 1996, 8(4):327 ~ 351.
- 7 Bartoli A, Dalal N, Bose B, *et al.* From video sequences to motion panoramas [A]. In: *Proceedings of Motion and Video Computing [C]*, Orlando, Florida, 2002: 201 ~ 207.
- 8 Fischler M, Bolles R. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography [J]. *Communications of the ACM*, 1988, 24(6): 381 ~ 395.
- 9 Harris C G, Stephens M. A combined corner and edge detector [A]. In: *Proceedings of the 4th Alvey Vision Conference [C]*, Manchester, 1988: 147 ~ 151.
- 10 Goldenberg R, Kimmel R, Rivlin E, *et al.* Fast geodesic active contours [J]. *IEEE Transactions on Image Processing*, 2001, 10(10): 1467 ~ 1475.
- 11 [OL/EB] <http://www.inrialpes.fr/lear/people/Dalal/research/MotionPanorama.html#correspondingPublications>, 2002.
- 12 LIN Kai-yan, WU Jun-hui, XU Li-hong. A survey on color image segmentation techniques [J]. *Journal of Image and Graphics*, 2005, 10(1):1 ~ 10. [林开颜, 吴军辉, 徐立鸿. 视频序列图像中运动对象分割综述 [J]. *中国图象图形学报*, 2005, 10(1):1 ~ 10.]
- 13 DENG Yu-chun, JIANG Yu-ming, ZHANG Jian-rong. Review of segmentation of moving object in image sequence [J]. *Application Research of Computers*, 2005, 22(1):8 ~ 11. [邓玉春, 姜昱明, 张建荣. 视频序列图像中运动对象分割综述 [J]. *计算机应用研究*, 2005, 22(1):8 ~ 11.]