

快速有效的视频图像序列拼接方法

林 武 洪景新 张 昊 李 琳

LIN Wu HONG Jing-xin ZHANG Hao LI Lin

厦门大学 计算机科学系 福建 厦门 361005

Computer Science Department of Xiamen University Xiamen Fujian 361005 China

E-mail hawk27@126.com

LIN Wu HONG Jing-xin ZHANG Hao et al. Fast and effective method for video mosaic. *Computer Engineering and Applications* 2009 45(24) :173-175.

Abstract : As the existing methods for video mosaic take high computational costs a fast and effective algorithm based on SURF feature for video mosaic is proposed. The algorithm uses SURF method with strong robustness and superior performance to extract feature instead of SIFT. At the aspect of feature matching a novel matching scheme based on hash mapping and bidirectional nearest neighbor distance ratio is presented which can quickly and effectively obtain the relationship between the features. In order to exclude the error matchings a RANSAC technique is applied to eliminate outliers to ensure effectiveness of the matched pairs and then the global motion parameters are estimated by a least-squares solution finally panorama from video sequence is achieved using the parameters. Experimental results show that the method with strong robustness performs fast and effectively and has highly valuable in practice.

Key words : video mosaic Speeded Up Robust Feature(SURF) feature matching motion parameters estimation image registration

摘 要 : 针对现有的视频图像序列拼接方法处理速度慢的问题 提出一种基于 SURF 特征的快速有效的拼接算法。该算法用鲁棒性强且计算性能优越的 SURF 算子取代传统的 SIFT 算子进行特征点提取 在特征点匹配方面 提出了一种基于哈希映射和双向最近邻距离比的匹配算法 可以快速有效地获得特征点间的对应关系。为了消除由于运动物体干扰带来的误匹配 采用随机采样一致性(RANSAC)方法来消除外点确保匹配的有效性 再通过最小二乘法估计视频帧之间的全局运动参数 最终拼接形成全景图。实验结果表明 该拼接算法快速有效 鲁棒性强 具有较高的使用价值。

关键词 : 视频拼接 SURF 特征点匹配 运动参数估计 图像配准

DOI :10.3778/j.issn.1002-8331.2009.24.051 文章编号 :1002-8331(2009)24-0173-03 文献标识码 :A 中图分类号 :TP391

1 引言

视频图像序列自动拼接是指将一组有一定重合区域的视频图像序列自动拼接为一幅能够更为全面描述场景内容的全景图像的过程 它能够弥补视频分辨率低、视野范围小的缺陷 有助于人们对视频内容有更全面、更直观的了解。视频图像序列拼接在虚拟现实、遥感遥测、军事侦察、场景监控等领域中有着越来越普遍的应用。

目前 视频图像序列拼接的方法主要可以分为两类 :一类是直接法 即利用图像的所有像素信息 通过迭代计算极小化误差平方和求得图像间的变换参数 此类方法计算量大且鲁棒性较差 ;另一类是基于特征法 通过跟踪图像特征来确定图像间的运动矢量^[1]。基于特征的方法稳定可靠 能够容忍较大的图像差异 因此可以通过提取和匹配具有良好不变性的图像特征点来准确估计视频图像间的运动参数。1999年 Lowe 提出了 SIFT 方法^[2] 它是一种基于尺度空间的 对图像缩放、旋转甚至仿射变换保持不变性的图像局部特征描述算子。鉴于 SIFT 算

子具有再现性和区分性良好等特点 文献[3-4]都采用该算子来提取特征点 很好地完成了视频图像序列的自动拼接 然而 SIFT 算子运算十分耗时 拼接效率低下 难以满足一些有实时性要求的应用。为此 对文献[3]提出的方法进行改进 采用 SURF 算子^[5]来提取特征点 SURF 特征类似于 SIFT 特征也具有尺度不变性、鲁棒性好、特征点之间区分性良好等特点 与 SIFT 算子相比它在运算速度方面有了很大的提高^[6]。

2 检测 SURF 特征

2.1 快速 Hessian 检测

SURF 算法利用快速 Hessian 检测子提取特征点 Hessian 矩阵具有良好的计算时间和精度表现。对于图像 I 上一点 $X(x, y)$ 在尺度 σ 下的 Hessian 矩阵 $H(X, \sigma)$ 定义如下 :

$$H(X, \sigma) = \begin{bmatrix} L_{xx}(X, \sigma) & L_{xy}(X, \sigma) \\ L_{xy}(X, \sigma) & L_{yy}(X, \sigma) \end{bmatrix} \quad (1)$$

作者简介 林武(1986-) 男 硕士研究生 研究方向为数字图像处理与计算机视觉 洪景新(1958-) 男 高级工程师 研究方向为信号处理 张昊

(1983-) 男 硕士研究生 研究方向为数字图像处理与模式识别 李琳(1985-) 女 硕士研究生 研究方向为计算机视觉与人工智能。

收稿日期 2009-04-24 修回日期 2009-07-03

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. <http://www.cnki.net>

其中 $L_x(X, \sigma)$ 是图像 I 与二阶高斯滤波 $\frac{\partial^2}{\partial x^2} g(\sigma)$ 在 X 点的卷积 $L_y(X, \sigma)$ 和 $L_z(X, \sigma)$ 类似。在构建尺度空间时, 与 SIFT 算法不同, 这里使用一种盒子滤波(box filter)^[5]来近似二阶高斯滤波, 它在积分图的基础上计算速度快且与图像大小无关。因为使用了积分图和盒子滤波, 所以并不需要通过迭代计算滤波来创建尺度空间, 取而代之通过放大滤波的大小来计算, 从而无需进行图像降采样操作, 大大减少了计算时间。SIFT 算法和 SURF 算法创建尺度空间方式如图 1 所示:

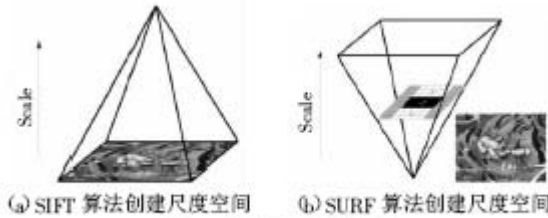


图1 创建尺度空间对比

对于每一个像素点通过非极大值抑制方法^[5]在它的领域(同尺度的 8 个相邻点和上下相邻尺度对应的 9×2 个点共 26 个点)中寻找极值点。为了确定图像的特征点, 将 Hessian 矩阵函数 $H(X, \sigma)$ 用泰勒公式展开:

$$H(X) = H + \frac{\partial H^T}{\partial X} + \frac{1}{2} X^T \frac{\partial^2 H}{\partial X^2} X \quad (2)$$

特征点的位置由式(2)的极值决定, 即:

$$\hat{X} = -\frac{\partial^2 H^{-1}}{\partial X^2} \frac{\partial H}{\partial X} \quad (3)$$

2.2 SURF 特征描述子

SURF 特征描述子的提取可以分为两步: 第一步根据特征点周围的一个圆形区域找到特征点的主方向, 即以每一个特征点为圆心, 半径为 6σ (σ 为尺度) 的邻域内的点分别计算 x 和 y 方向上的 Haar 小波响应, 得到一系列的响应向量。用 60° 的窗口绕着原点旋转并计算 Haar 小波响应的向量和, 向量和最长的方向为该特征点的主方向, 如图 2(a)。第二步在这个选定的主方向上构造一个矩形区域, 并提取出所需要的特征点描述信息。按主方向构造一个大小为 20σ 的窗口, 并将该窗口区域分成 4×4 的子区域。对每一个子区域, 按照 5×5 的大小进行取样, 分别计算相对于主方向的水平和垂直方向上的 Haar 小波响应, 记为 dx 和 dy 。每个子区域得到一个四维向量:

$$V_{sub-region} = [\sum dx, \sum dy, \sum |dx|, \sum |dy|] \quad (4)$$

因此 4×4 个子区域可以得到 64 个值, 归一化后的这 64 个值就构成 SURF-64 的特征点描述子, 如图 2(b)。它还可以扩展为

128 维(SURF-128) 根据 $dy > 0$ 和 $dy \leq 0$ 分别计算 $\sum dx$ 、 $\sum |dx|$ 、

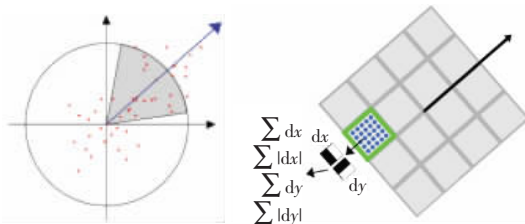


图2 SURF 特征描述子

同理根据 dx 的符号分别计算 $\sum dy$ 、 $\sum |dy|$, 这样就可以使 SURF 特征向量翻倍。SURF-128 具有更好的区分性, 采用它作为特征描述子。

2.3 优化 SURF 算法

SURF 算法鲁棒性强, 而且具有很好的计算性能^[6]。但是, 该算法提取的特征点在纹理特征丰富的区域非常多且密集, 不利于特征点匹配及变换参数的求解, 当检测出的特征点数量比较大时无疑增加了后续特征描述子提取、特征点匹配及随机采样一致性算法(RANSAC)的运算时间。视频帧间的运动参数估计并不需要对所有的特征点进行计算, 特征点太多反而增加了计算成本, 因此对 SURF 算法进行优化。原始 SURF 算法只关心特征检测的效果忽略了检测出特征点的数量及位置关系, 这里设定特征点数量阈值 S 和特征点间最小距离阈值 L , 当原有算法第一步用快速 Hessian 检测出的特征点数量大于 S 时进行特征点距离检测, 否则不进行距离检测, 距离检测, 即检测特征点与其周围特征点的距离, 当某一特征点与其周围的特征点距离小于 L 时将被遗弃, 这样使检测出的所有特征点之间的距离都大于等于 L 。距离检测拉开了特征点间的距离使检测出的特征点均匀分布在图像上, 同时由于特征点数量的减少也节省了不少 SURF 特征描述子的提取时间, 从而使优化后算法比原有算法有更好的计算性能。如图 3 SIFT 算法检测出 545 个特征点, 耗时 417 ms, SURF 算法检测出了 540 个非常密集的特征点, 耗时 131 ms, 优化后算法($S=100, L=15$)检测出 94 个特征点, 耗时 88 ms, 所有特征点均匀分布在整幅图片上。



图3 特征检测结果对比

对于视频图像序列, 用优化后的 SURF 算法对奇数视频帧进行特征点检测, 用原 SURF 算法对偶数视频帧进行特征点检测, 这样避免了所有视频帧都用原 SURF 算法进行特征点检测时, 当检测出的特征点较多时在特征点匹配及随机采样一致性操作过程中消耗太多时间, 同时也避免了都用优化后的 SURF 算法进行特征点检测时可能使检测出的特征点无法找到对应的特征点, 这是因为对应的特征点可能在距离检测中被遗弃掉了。

3 SURF 特征点匹配及消除误匹配

检测出 SURF 特征后, 要进行特征点的匹配。Lowe 提出了最近邻距离比^[7]的匹配方法, 但它只从单个方向上考虑特征量的相似性比较, 导致错误匹配的比率仍然偏高, 匹配结果经常出现一对多的情况。文献[8]提出了一种双向最近邻距离比的匹配策略, 该方法消除了一对多的误匹配对, 提高了特征点匹配的正确率, 然而却使得匹配操作的计算时间翻倍。

为此, 提出一种基于哈希映射与双向匹配相结合的方法, 该方法不仅保证了特征点匹配的正确率而且加快了运算速度。检测出的 SURF 特征点具有不同的尺度大小 S , 如图 3(c) 的特

征点方框半径大小,而同一特征点的 S 值在相邻视频帧之间基本保持不变或者变化很小,因此可以以 S 值为关键字创建哈希映射使整幅图像的特征点集以 S 值为关键字划分为多个不同的子集合,每个子集合内的特征点都具有相同的 S 值,这样在查找匹配点时无需遍历所有的特征点搜索对应的匹配点,只需查找与之具有相同或相近 S 值的集合即可。具体匹配步骤如下:设 $F_1=\{f_1^1, f_2^1, \dots, f_M^1\}$ 和 $F_2=\{f_1^2, f_2^2, \dots, f_N^2\}$ 分别为两帧待匹配图像的 SURF 特征点集。以特征点尺度 S 为关键字分别为 F_1 和 F_2 建立哈希映射表 P_1 和 P_2 , 则 F_1 和 F_2 被细划为 $F_1=\{A_1, A_2, \dots, A_m\}$ 和 $F_2=\{B_1, B_2, \dots, B_n\}$, 以 S 值从小到大排序 A_i 和 B_i 表示具有相同 S 值的特征点子集。首先,从 $F_1 \rightarrow F_2$ 方向查找匹配点,对于 F_1 中任意的特征点 f_k^1 根据自己的 S 值通过 P_2 映射到 S 值在 $S-\eta_1$ 与 $S+\eta_2$ 之间的子集 $T=\{B_i, B_{i+1}, \dots, B_j\}$ 上, η_1, η_2 为松弛因子用于控制搜索范围,然后在子集 T 中搜索该特征点的最近邻 f_{NN} 与次近邻 f_{SN} , 当二者的比值小于某一阈值 δ , 则认为 f_k^1 与 f_{NN} 匹配。实验表明 $\delta=0.6$ 时,匹配正确率较高。重复上面操作完成所有特征点匹配,得匹配对集合 $H_1=\{M_1, M_2, \dots, M_s\}$, $M_k=\langle f_i^1, f_j^2 \rangle$ 。同理,从 $F_2 \rightarrow F_1$ 方向匹配,得匹配对集合 H_2 。最后由 H_1 与 H_2 交集得最终的匹配对集合 H 。

上述算法可以得到比较高的正确匹配率,但是仍然存在一些误匹配,尤其在有运动物体出现时。特征点匹配的正确性直接影响到后续的变换模型的参数估计,为了确保匹配的有效性采用 RANSAC 算法^[9]来消除误匹配。

4 全局运动模型

4.1 相邻帧间的运动估计

一般图像间的变换有平移、旋转、缩放、仿射、投影等,其中,投影变换有更强的适应性,对视频序列限制少,适用于视频帧之间的各种运动变换。因此采用投影变换进行相邻帧间的运动参数估计,对于相邻帧图像上对应的两点 (x_i, y_i) 与 (x_i', y_i') , 满足如下变换关系:

$$\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \begin{bmatrix} m_0 & m_1 & m_2 \\ m_3 & m_4 & m_5 \\ m_6 & m_7 & 1 \end{bmatrix} \begin{bmatrix} x_i' \\ y_i' \\ 1 \end{bmatrix} \quad (5)$$

上述变换参数有 8 个未知变量。根据线性方程相关理论,只需 4 个匹配特征点对就能对其求解,但是由于选取的特征点往往存在误差,以及选取的 4 点中,有可能存在 3 点共线的情况,这种线性方程解很不稳定。采用最小二乘法对参数进行拟合估计,匹配特征点越多,估计值就越可靠,平均误差也越小。

4.2 任意帧间的全局运动估计

4.1 节中完成相邻帧之间的运动估计,但是创建视频图像序列的全景图需要将各个视频帧图像映射到参考坐标上。因此,需要选定一个参考帧以它的坐标作为参考坐标。任意帧可以选择为参考帧,其他所有帧以参考帧为基准进行配准。设 I_r 为参考帧,对于任意的帧 $I_s (1 \leq s \leq n)$, I_s 映射到 I_r 的关系 $H_{r,s}$ 定义如下:

$$H_{r,s} = \begin{cases} I(\text{单位矩阵}) & s=r \\ H_{r,s-1} \cdot H_{s-1,s} & s < r \\ H_{r,s+1} \cdot H_{s+1,s} & s > r \end{cases} \quad (6)$$

最后,所有视频帧根据各自的变换关系 $H_{r,s}$ 映射到参考坐标上

拼接形成一幅视频全景图。

5 实验结果

该文算法用 VC++6.0 及 OPECV 库实现,实验机器硬件配置 CPU 双核 2.83 G,内存 2 G。为验证该文视频图像序列自动拼接方法的有效性,这里给出两个视频片段“road”和“sand”的拼接结果。实验中用到的这两段视频是由摄像机拍摄的,采集频率为 10 帧/s,图像大小 320×240。这两段视频序列各自包含不同的场景,视频“road”拍摄于公路上,共 12 帧,视频“sand”拍摄于沙滩上,共 25 帧,其中包含很多行走中的人。实验时,第一帧图像被选为参考帧,其他帧图像根据求得的运动变换参数映射到参考帧坐标上。

图 4、图 6 分别显示了视频“road”的第 1、5、11 帧图像和视频“sand”的第 1、7、21 帧图像。从图中可以明显看出视频图像序列的变化过程,视频“sand”的图像序列中存在很多细小的运动物体(人),部分图像帧还有旋转。视频图像序列的拼接效果如图 5 和图 7,可以看出视频帧间匹配准确得到不错的拼接效果。因为该文主要考察算法的鲁棒性和运算性能,所以并未对图像进行融合操作。



图 4 视频“road”中第 1、5、11 帧图像



图 5 视频“road”拼接效果



图 6 视频“sand”第 1、7、21 帧图像



图 7 视频“sand”拼接效果

表 1 给出了该文算法与文献[3]提出的基于 SIFT 特征的视频拼接算法的计算性能对比,由于随机采样一致性算法

表 1 算法计算性能对比

视频片段	算法	
	基于 SIFT 特征视频图像拼接算法/ms	该文算法/ms
“road”	6 238	2 146
“sand”	10 487	4 293

出了基于混合蛙跳算法的模糊 C 均值聚类算法 SFLA-FCM, 并应用于城市交通时段划分, 克服了人工时段划分的不合理性。实验结果表明, 该方法能有效地避免 FCM 易陷入局部极小的缺陷, 同时降低了 FCM 对初值的敏感度, 对解决城市交通时段划分等数据聚类问题是可行、有效的。同时, 作为一种全新的优化算法, 混合蛙跳算法将具有广泛的应用前景。

参考文献:

- [1] Hauser T, Scherer W. Data mining tools for real-time traffic signal decision support and maintenance[C]//Proc of IEEE Int Conf on Systems, Man, and Cybernetics, Tucson, USA: IEEE Press, 2001: 1471-1477.
- [2] Park B, Lee Do-Hoon, Yun Hsoo. Enhancement of time of day based traffic signal control[C]//Proc of IEEE Int Conf on Systems, Man and Cybernetics, [S.l.]: IEEE Press, 2003: 3619-3624.
- [3] 杨立才, 贾磊, 孔庆杰, 等. 基于人工免疫算法的交通时段自动划分方法[J]. 控制理论与应用, 2006, 23(2): 193-198.
- [4] 孙艺峰, 王向阳, 王春花. 一种新的快速模糊 C 均值聚类图像分割算法[J]. 小型微型计算机系统, 2008, 29(2): 320-323.
- [5] Arima C, Hakamada K, Okamoto M, et al. Modified fuzzy gap statistic for estimating preferable number of clusters in fuzzy k-means clustering[J]. Journal of Bioscience and Bioengineering, 2008, 105(3): 273-281.
- [6] 林琳, 王树勋. 基于遗传-模糊聚类的说话人识别方法及其仿真研究[J]. 系统仿真学报, 2006, 18(8): 2338-2341.
- [7] 时念云, 蒋红芬. 基于免疫单亲遗传和模糊 C 均值的聚类算法[J]. 控制工程, 2006, 13(2): 158-160.
- [8] Alexiev K M, Georgieva O I. Improved fuzzy clustering for identifi-

cation of Takagi-Sugeno model[C]//Proceedings of the 2nd International IEEE Conference on Intelligent Systems, Varna, Bulgaria, [S.l.]: IEEE Press, 2004, 1: 213-218.

- [9] Eusuff M M, Lansey K E. Optimization of water distribution network design using the shuffled frog leaping algorithm[J]. Water Resour Plan Manage, 2003, 129(3): 210-225.
- [10] Rahimi-Vahed A, Mirzaei A H. A hybrid multi-objective shuffled frog-leaping algorithm for a mixed-model assembly line sequencing problem [J]. Computers & Industrial Engineering, 2007, 53(4): 642-666.
- [11] 高新波. 模糊聚类分析及其应用[M]. 西安: 西安电子科技大学出版社, 2004.
- [12] Maitra M, Chatterjee A. Hybrid multiresolution Slantlet transform and fuzzy c-means clustering approach for normal-pathological brain MR image segregation[J]. Medical Engineering & Physics, 2008, 30(5): 615-623.
- [13] 李英海, 周建中, 杨俊杰, 等. 一种基于阈值选择策略的改进混合蛙跳算法[J]. 计算机工程与应用, 2007, 43(35): 19-21.
- [14] Elbehairy H, Elbeltagi E, Hegazy T, et al. Comparison of two evolutionary algorithms for optimization of bridge deck repairs [J]. Computer-Aided Civil and Infrastructure Engineering, 2006, 21(8): 561-572.
- [15] 孔庆杰. 计算智能理论在交通流诱导系统中的应用[D]. 济南: 山东大学, 2006.
- [16] Zhang Yun-jie, Wang Wei-na, Zhang Xiao-na, et al. A cluster validity index for fuzzy clustering[J]. Information Sciences, 2008, 178(4): 1205-1218.
- [17] 吴黎, 杨东勇. 基于改进的模糊 C 均值聚类算法的人像照片背景替换[J]. 计算机应用, 2006, 26(2): 424-426.

(上接 175 页)

(RANSAC)具有随机性, 表中给出的是多次实验的平均值, 可以明显看出该文算法具有更好的计算性能。值得一提的是, 基于 SIFT 特征的视频拼接算法对视频“road”和视频“sand”耗费的时间并不与它们的视频长度基本成正比。其中的原因就在于视频“road”帧图像相对于视频“sand”纹理特征要丰富得多, 检测出的特征点非常多, 而大量的特征点进行匹配和随机采样一致性操作消耗了很多时间。该文算法对 SURF 算子增加了特征点距离检测, 减少了特征点数, 因而不会出现上述情况。以上实验表明, 该文视频拼接算法具有很好的鲁棒性, 可以在不同的场景下进行快速稳定的拼接。

6 结论

提出一种快速有效的视频图像序列拼接算法。基于特征的视频图像拼接算法需要经常用到特征检测算法, 几乎每一帧视频图像都要进行一次特征检测操作, 因此一个鲁棒性强且运算速度快的特征检测算子对整个视频拼接算法的效率至关重要。传统的 SIFT 算子虽然具有良好的检测效果, 但是计算速度慢。对于视频图像序列, 采用具有优越计算性能的 SURF 算法来提取特征点, 大大降低了计算成本。同时, 该文基于哈希映射和双向最近邻距离比的匹配算法也在一定程度上提高了整个算法的运算速度。

参考文献:

- [1] Li Jing, Pan Quan, Yang Tao, et al. Automated feature points management for video mosaic construction[J]. IEEE Information Technology and Applications, 2005, 1: 760-763.
- [2] Lowe D G. Object recognition from local scale-invariant features[C]//Proceeding of ICCV, Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers Inc, 1999, 2: 1150-1157.
- [3] 高超, 张鑫, 王云丽, 等. 一种基于 SIFT 特征的航拍图像序列自动拼接方法[J]. 计算机应用, 2007, 27(11).
- [4] 张朝伟, 周焰, 王耀康, 等. 基于 SIFT 特征跟踪匹配的视频拼接方法[J]. 计算机工程与应用, 2008, 44(10): 169-172.
- [5] Bay H, Tuytelaars T, van Gool L J. SURF: Speeded up robust features[C]//European Conference on Computer Vision, 2006, 1: 404-417.
- [6] Valgren C, Lilienthal A. SIFT SURF and seasons: Long-term outdoor localization using local features[C]//Proc of 3rd European Conference on Mobile Robots, Freiburg, Germany, 2007.
- [7] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.
- [8] Zhang Xiao-miao, Zhang Xiao-hu, Ji Fang-fang, et al. Known landing area rough locating from far distance based on DDM-SIFT[J]. IEEE Image and Signal Processing, 2008, 1: 686-690.
- [9] 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).