

一种改进的全景图自动拼接算法

赵辉 陈辉 于泓

(山东大学信息科学与工程学院, 济南 250100)

摘要 提出了一种带有图像自动排序功能的全景图全自动拼接算法。首先利用相位相关法对输入图像进行排序,同时估算角点匹配区域以加快角点匹配速度,提高匹配稳健性;然后使用RANSAC算法去除外点,采用奇异值分解法配合LM非线性优化方法求解变换参数,提高配准精度;最后使用多频带融合算法进行图像混合以改善线性加权融合算法带来的高频细节模糊;算法中还设有检测机制以确保LM优化顺利进行。与现有算法相比,对于图像重叠区域较小、图像中含有运动物体、重复性纹理等较难自动拼接的情况,使用该算法可以有效提高拼接成功率,缩短总体拼接时间,而且合成的图像清晰度较高。实验结果表明该算法具有较好的实用价值。

关键词 相位相关法 角点匹配 奇异值分解最小二乘法 全景图 多频带融合

中图分类号: TP391.41 **文献标识码**: A **文章编号**: 1006-8961(2007)02-0336-07

An Improved Fully-automatic Image Mosaic Algorithm

ZHAO Hui, CHEN Hui, YU Hong

(School of Information Science and Technology, Shandong University, Jinan 250100)

Abstract In this paper, an improved fully-automatic image mosaic algorithm is presented. Firstly, phase correlation is used to sort the unordered image sequence and roughly compute the translation offset between adjacent images, which speeds up corner match procedure and improves matching stability as well. Secondly, RANSAC algorithm is used to eliminate outliers to ensure effectiveness of the matched corner pairs. Then, Singular Value Decomposition-Least Square (SVDLS) method and Levenberg-Marquardt optimization are used to robustly determine the 8 parameters transform model, which aims to reduce the registration error and also improve the stability when handling very small number of matched corners. At the last of the algorithm, a multi-band blending technique is used to generate the final panorama. Invalid parameters are verified by the translation offset to make Levenberg-Marquardt optimization more successful and robust. The experimental results show that the proposed image mosaic algorithm exceeds existing ones at matching speed, stability, and it has less blur or ghost effect after blending, especially when there are noise, moving objects, repeated texture and small overlaps presented in the images.

Keywords phase correlation, corner matching, SVDLS, panorama, multi-band blending

1 引言

近年来全景图自动拼接技术^[1-4]一直是计算机视觉领域的研究热点。图像镶嵌(image mosaic)技术在虚拟现实、视频压缩与检索等领域都有广泛应用^[2,3]。使用普通相机对周围环境进行摇移拍摄并保证相邻图像间有一定重叠,然后可利用软件算法

对图像进行自动或者手动拼接以生成全景图像。生成的全景图像可以投影到柱面或球面等曲面用以显示与浏览。

全景图拼接算法可以分为基于像素灰度差最小化的方法^[2,3]和基于特征匹配的方法^[1,4]。目前市场上已有不少全景图制作软件,有代表性的比如QuickTime Pro、Photo Stitch等。这些软件一般都要

基金项目:山东省优秀中青年科学家科研奖励基金项目(03BS001);山东省自然科学基金重点项目(Z2005G02)

收稿日期:2005-06-27;改回日期:2005-11-28

第一作者简介:赵辉(1981~),男,2006年于山东大学获信息处理专业工学硕士学位。主要研究方向为图像处理、计算机视觉。

E-mail: bugzhao@sdu.edu.cn

大量人工干预;另外对拍摄者和拍摄方式的要求也较为严格,一般要求使用三脚架将相机严格在水平面内转动拍摄而且图像间至少要有30%重叠才能取得较好拼接效果。噪声、运动物体和大量重复性纹理特征的存在将会导致合成图像清晰度较低,甚至拼接失败。在学术研究方面,文献[2]中提出了一种基于摄像机旋转矩阵表示的全景图拼接技术,同时提出图像全局优化(global alignment)和局部调整去鬼影(deghosting)算法,提高了图像拼接的清晰度,但该算法在拼接总视角大于 90° 的图像时易出现较大失真,同时不具备对光照变化的稳健性,另外整个算法需要利用全部像素灰度进行运算,速度较慢,实用性不强。文献[4]中提出了一种改进的基于角点匹配的图像自动拼接算法,在匹配特征点中使用了去外点技术以提高拼接的稳健性,但是该算法仍然需要手动指定图像间的顺序,而且在拼接含运动物体、重复性纹理的图像时往往会失败,合成图像的清晰度也不高,高频细节往往被模糊。

为了克服现有算法的种种不足,本文提出了一套完整的改进的全自动图像拼接算法,在特征匹配、变换参数计算和图像融合过程中均做出改善,提高了图像拼接的速度、稳健性、清晰度和自动化程度。

2 图像自动排序及平移参数粗估计

相位相关法^[5]使用互功率谱进行快速配准。假设图像 I_1 和图像 I_2 具备平移关系:

$$I_2(x, y) = I_1(x - x_0, y - y_0) \quad (1)$$

将其进行傅里叶变换:

$$F_2(\xi, \eta) = e^{-j2\pi(\xi x_0 + \eta y_0)} \cdot F_1(\xi, \eta) \quad (2)$$

互功率谱定义为

$$\frac{F_1^*(\xi, \eta) F_2(\xi, \eta)}{|F_1^*(\xi, \eta) F_2(\xi, \eta)|} = e^{-j2\pi(\xi x_0 + \eta y_0)} \quad (3)$$

其中, F_1^* 为 F_1 的复共轭。将上式进行傅里叶反变换得到:

$$\delta(x - x_0, y - y_0) = F^{-1}[e^{-j2\pi(\xi x_0 + \eta y_0)}] \quad (4)$$

可见寻找上式中冲激函数的峰值位置所在,即可确定 (x_0, y_0) ,从而得到平移参数。

上述结论是基于两幅图像具备简单平移关系的假设。冲激函数的峰值高低反应了两幅图像的相关性大小。如果两幅图像具备更为复杂的透视变换关系,同时被噪声干扰,甚至含有运动物体,那么冲激函数的能量将从单一峰值分布到其他小峰值,但其

最大峰值的位置具备稳定性。

本文提出了一种利用相位相关信息来对图像进行快速排序的方法。其思路是利用冲激函数的幅度值作为图像间的相关性度量以从图像集合中寻找两两相关的图像。具体分为3步:

(1) 对给定的图像集,选定任一图像 i ,计算它与图像 j 的互功率谱并求得峰值处冲激函数的幅度值作为图像 i, j 的相关度 $D_{i,j}$ 。对于 N 幅输入图像,其任一图像可以计算得到 $N-1$ 个相关度。

(2) 先来确定头图像(最左边的图像)和尾图像(最右边的图像)。把相关度进行阈值处理。对于图像 i ,如果有两个相关度大于阈值 T ,则表明该图像是中间图像,其与左边的图像、右边的图像具备较高相关性;如果只有一个相关度大于 T ,则说明其为头图像或者尾图像。对于找出的这两幅头、尾图像来进一步区分。如果该图像大于阈值 T 的那个相关度所对应图像的水平平移量为负,则是头图像;如果为正,则是尾图像。

(3) 接下来可以顺着找出的头图像进一步确定整个图像链。如果水平平移量为负,则图像排在右边。这样就可以一步步地将图像进行排序。

排序后的图像对应于相机由左向右转动拍摄的结果。同时此步粗估算了水平、竖直平移量。这两个平移参数对于后续计算过程至关重要。

3 角点检测与匹配

采用Harris算法^[6]作为角点检测算子。设 I_x 和 I_y 分别为图像 I 在水平、竖直方向的一阶偏导数, $w(x, y)$ 是2维高斯平滑函数,有:

$$M = \sum_{x,y} w(x, y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \quad (5)$$

$$R = \det M - k \cdot (\text{trace} M)^2, k = 0.04 \sim 0.2 \quad (6)$$

对于图像的每一点解得 R ,其在某一阈值之上并取得局部极大值的点即作为角点。

角点提取之后,下一步采用归一化相关法(NCC)进行角点匹配。相关系数计算如下:

$$\rho_{NCC} = \frac{\sum_i (I_1(x_i, y_i) - u_1)(I_2(x_i, y_i) - u_2)}{\sqrt{\sum_i (I_1(x_i, y_i) - u_1)^2 \sum_i (I_2(x_i, y_i) - u_2)^2}} \quad (7)$$

式(7)中 u_1, u_2 为相关窗口内像素灰度的均值。

进行角点匹配时首先对于图像 1 中的任一角点,寻找在图像 2 中和它相关性最大的角点;然后对于图像 2 中的任一角点,寻找在图像 1 中和它相关性最大的角点,当双向搜寻到的最大相关性的角点彼此对应时,即认为寻找到一对匹配点。

由于角点匹配过程使用了遍历搜索算法,对于第 1 幅图像中的所有角点需要在第 2 幅图像中的全部角点中进行匹配以求得最相关的点,这在整个全景图拼接过程中占有很大的时间开销。另外,该匹配方法属于局部匹配方法,比如说一个运动的人同时出现在第 1 幅、第 2 幅图像中,那么此时应该用大背景作为匹配对象而不是运动的人,而使用遍历搜索的匹配方法很可能将运动的人也作为匹配对象,从而造成局部匹配无法满足全局匹配。同时,光照不均、图像噪声、大量重复性纹理的存在也往往会导致角点错匹配。由于大量外点的存在,即使使用去外点算法也很难成功的去除局部错配。

解决方法是,首先根据前面相位相关法估算的水平、竖直平移量来粗确定第 1 幅图像与第 2 幅图像的重叠区域,仅在此重叠区域内进行角点检测与匹配,避免了在全部图像区域内无效的计算;另外根据平移量和第 1 幅图像中的角点位置,计算出第 2 幅图像中匹配角点的估测区域(可以近似认为第 2 幅图像中对应角点的中心位置是第 1 幅图像角点位置加上平移量,匹配时,仅在该点的邻域窗口内搜索可能匹配的点。一般选择 40×40 的窗口),这就大大加快了角点匹配的速度。可见,对噪声、运动物体、光照不均等均不敏感的相位相关法求得的平移参数具有“全局指导”的优点,避免了无意义的局部匹配运算,降低了错配。

剩余的少量错配会对后续的变换模型参数估计有较大影响。所以继续采用了 RANSAC 算法^[7]去除外点(outliers),最大程度保证匹配点对的有效性,去除局部错配对全局匹配的影响。

4 变换模型参数估计与图像融合

4.1 透视变换模型参数估计

由文献[8]知相机绕近似固定点摇移拍摄的图像满足 8 参数透视变换关系,变换公式如下:

$$X'_i = \begin{bmatrix} x'_i \\ y'_i \\ w'_i \end{bmatrix} = H \cdot X_i = \begin{bmatrix} h_0 & h_1 & h_2 \\ h_3 & h_4 & h_5 \\ h_6 & h_7 & h_8 \end{bmatrix} \cdot \begin{bmatrix} x_i \\ y_i \\ w_i \end{bmatrix} \quad (8)$$

这里 X_i, X'_i 分别为相邻两幅图像中第 i 组匹配点的齐次坐标向量表示, H 为透视变换矩阵。矩阵 H 中第 9 个参数可以归一化为 1。

代入匹配点坐标到式(8)以求解 8 参数的过程是一个极小值优化求解问题。一般使用 10 组以上匹配点代入计算,这样方程数多于未知数,可以通过伪逆法来求解这个超定方程的最小二乘解。伪逆方法求的最小二乘解稳定性不佳,如果存在 3 点近似共线的情况,列秩亏损将导致所求解误差较大甚至出现发散。所以采用更加可靠的奇异值分解法^[9]来求式(8)中 8 个未知参数的最小二乘解(SVDLS)。其计算步骤如下:

对于第 i 组匹配点有

$$A_i = \begin{bmatrix} \mathbf{0} & -w'_i X_i^T & y'_i X_i^T \\ w'_i X_i^T & \mathbf{0} & -x'_i X_i^T \\ -y'_i X_i^T & x'_i X_i^T & \mathbf{0} \end{bmatrix}^T \quad (9)$$

其中, $\mathbf{0}$ 表示 1×3 的零阵。那么对于全部 N 个匹配点组有

$$A = [A_1 \ \cdots \ A_i \ \cdots \ A_N]^T \quad (10)$$

这里 A 为 $3N \times 9$ 的矩阵。将 A 进行奇异值分解:

$$UDV^T = A \quad (11)$$

V 的最后一列即为变换矩阵 H 的 9 个参数。奇异值分解法求得的最小二乘解比伪逆法更加稳定可靠,而且残差较小,即使在特征点数目较少(少于 10 组)的情况下也可以得到较为精确的变换参数。

引入一种参数校验机制。考虑最差的情况,如果代入方程的特征点数很少而且同时几乎位于一条直线的话,SVDLS 方法仍然会失败,这将导致后续 LM (levenberg-marquardt)非线性优化算法无法收敛,配准过程失败。如前所述,相位相关法估算的两个平移量相当稳健,所以解决思路就是对奇异值分解法求得的 H 中的两个平移参数与相位相关法估算的平移参数进行比较,如果二者平方差大于一阈值,则认为最小二乘解不合理,从而返回角点提取步骤,降低阈值以提取更多的角点,并重新进行匹配、计算,直至最小二乘解满足要求。这种用全局参数校验局部参数的方法更大程度上保证了算法的稳健性。

最小二乘解 H 往往含有较大的配准误差,如果直接用其进行图像融合将会出现较大的鬼影,因此对 H 继续使用 LM 算法进行非线性优化处理^[4,9]。LM 非线性优化的起始值即为前面计算的 SVDLS 最小二乘解。经 LM 优化后每像素平均配准误差可以

控制在0.1内。

4.2 图像重投影与融合

如果对图像直接进行拼接,将会破坏实际场景的视觉一致性从而无法满足实际场景中各对象间的几何关系。为此本文使用了柱面投影法,将图像先统一投影到同一个柱面上,然后再进行图像混合处理。柱面投影法参见文献[2]。

由于图像噪声、模型误差、相机晃动等因素的存在,图像重叠区域内分别处于相邻两幅图像的两个对应点并不一定能很好地重合在真实的场景点,所以图像混合算法的好坏直接关系到合成图像的清晰度。较常用的融合算法是线性插值混合算法,即在图像重叠区域内使用两幅图像共同的贡献值作为混合图像的灰度值,使用线性加权过渡函数以实现在靠近重叠区域边界的地方做平滑过渡处理。这种方法简单易行,但存在较多缺点。假设 T_w 为过渡带宽度,如果 T_w 较小,那么图像合成后将会在拼缝处产生较明显的不连续,平滑感较差;如果 T_w 较大,则会产生多重曝光效果^[10],合成后的高频细节将变得较模糊。为了解决这一矛盾,采用文献[1]、[10]中提出的基于塔式算法的多频带融合算法(multi-band blending),将图像在不同频带内进行分解,然

后取用不同的 T_w 值分别进行加权插值处理后再进行融合,以同时实现较平滑的过渡以及较好的高频细节保留能力。为了实现的简便,本文算法使用高通、低通两个频带进行多频带融合,后面实验结果证明其融合效果明显优于一般的线性加权融合算法。

5 实验结果及分析

采用佳能 S50 数码相机手持拍摄的图片进行实验。实验机器硬件配置为 CPU P4 1.5G,内存 384M,软件开发环境为 VC6.0 + Matlab7.0。注意在摇移拍摄过程中相机不能使用变焦。

首先是对两幅含有运动物体、噪声、光照不一的图片利用相位相关法估计平移参数进行实验。

图1(a)为第1幅图像,图1(b)为向右水平转动相机后拍摄的第2幅图像,并加入高斯噪声。两幅图像分辨率同为 256×256 。图1(c)为利用相位相关法检测到的峰值。最终正确估计出 x, y 两个方向上的平移量为 $(104, -5)$ 。相位相关法估计的平移量虽然不够精确,但足以满足粗配准的要求,对于指导后续步骤中的局部匹配、参数校验起着至关重要的作用。



图1 相位相关法估计平移参数

Fig. 1 Use phase correlation to estimate offset parameters

接下来的实验对5张拍于公园的图片进行自动排序、拼接。该组5幅图像分辨率均为 360×480 ,其顺序是错乱的,如图2所示。因为图片中含有草木、水波纹等难以匹配的纹理性特征,以及运动的船、树叶等运动物体,使用文献[2]、[4]中的方法均无法成功拼接。而本文算法在粗平移参数的指导下可以更准确的进行角点匹配,同时采用更为稳健的奇异值分解法求解变换参数,最终正确完成拼接,而且图像清晰度较高。下面将从图像排序、特征点匹配、图像融合几个步骤逐步进行实验并分析结果。

表1是图像排序过程,即计算图像两两相关度及水平偏移量(这里取阈值 $T = 0.015$)。根据表1数据,图2中的第3幅和第5幅首先可以判定为两头的图像,然后顺着5一直到2、4、1、3,排序完毕。排序及计算平移量所用时间大约为2s。排序后的图像集合为图3所示。

接下来进行角点检测与匹配。以其中两幅图像为例,经过角点提取、匹配、RANSAC去外点后,其最终匹配角点如图4所示。

把5幅图像两两配准后,统一投影到柱面(计

表 1 图像排序的计算过程
Tab. 1 Image auto-sorting process

排序比较次数	图像序号				
	1	2	3	4	5
①	X	0.0105	0.0221*, -172→③	0.0161*	0.0105, 164
②	0.0105	X	0.0100	0.0313*, -226→④	0.0215*
③ (尾)	0.0221*, 172	0.0100	X	0.0111	0.0093
④	0.0161*, -241→①	0.0313*	0.0111	X	0.0103
⑤ (头)	0.0105	0.0215*, -164→②	0.0093	0.0103	X

注: * 表示每次找到的头图像



图 2 顺序错乱的输入图像
Fig. 2 Unordered image set

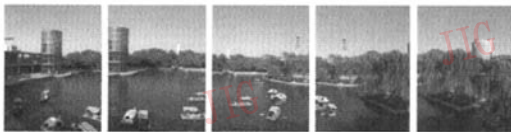


图 3 自动排序后的图像
Fig. 3 Ordered image set after auto-sorting

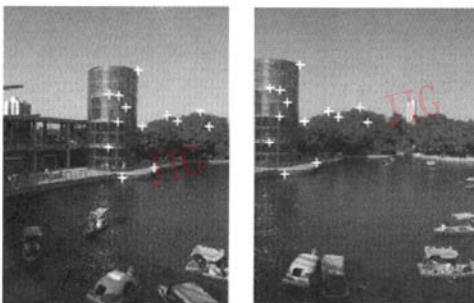


图 4 角点检测与匹配
Fig. 4 Corner extraction and matching

算出的焦距 $f = 476.0138$), 并经融合、裁剪后得到最终的拼接结果——一副合成的“超广角”图像, 如图 5(b) 所示。



(a) 使用线性加权融合算法进行图像融合的结果



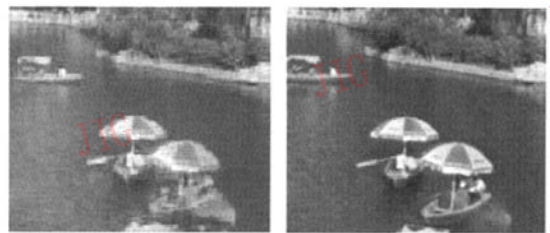
(b) 使用多频带融合算法进行图像融合的结果

图 5 最终拼接形成的全景图

Fig. 5 The final panorama

为了进行对比, 图 5(b) 为使用多频带融合算法拼接的结果, 图 5(a) 为使用一般线性加权融合算法拼接的结果。

可见, 多频带融合算法拼接的图像在清晰度和画面连贯性上要明显好于线性加权融合算法, 尤其在有运动物体的地方 (比如本例中漂泊的船只、飘荡的柳树), 线性加权融合算法不可避免地存在较大模糊。图 6 为两种融合算法拼接的图像在局部细节的对比, 可以明显看出多频带融合算法的优势所在。



(a) 线性加权融合算法 (b) 多频带融合算法

图 6 使用两种算法进行融合后图像细节对比
Fig. 6 Detail comparison using two fusion methods

下面的实验选用重叠区域仅为 15% 的图像以验证本文算法对小重叠区域图像的拼接能力。图 7 是其中两幅图像的角度检测与匹配结果, 图 8 是最终拼接结果。

图 9 给出了使用本文算法和文献 [4] 中的算法拼接结果的局部对比图, 可见图 9(b) 中树枝和汽车



图 7 角点检测与匹配

Fig.7 Corner extraction and matching

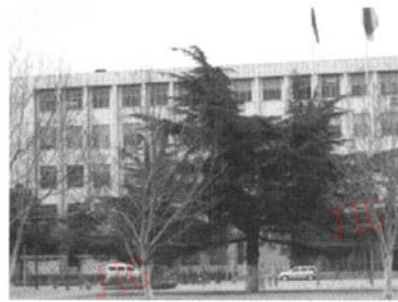


图 8 自动拼接而成的 360°校园全景图

Fig.8 The final 360° campus scene



(a) 使用文献[4]的方法拼接结果的局部



(b) 使用本文算法拼接结果的局部图

图 9 本文算法与文献[4]算法拼接图像的局部对比

Fig.9 Local comparison of the two methods

表 2 给出就本文算法与文献[2]、文献[4]的算法在拼接速度、拼接成功率上的对比。实验总共使用了 35 组图像,接近一半的图像存在运动物体和重复性纹理及一定的曝光差异。

如果仅计算角点匹配过程的时间消耗,本文算法平均所用时间为文献[4]算法的 30% 左右。

表 2 本文算法与其他代表性算法的对比

Tab.2 Comparison between our method and others'

	文献[2]算法	文献[4]算法	本文算法
平均拼接速度(s)	17.62	12.13	8.25
拼接成功率(%)	65.71	51.43	82.86

6 结 论

本文提出了一种改进的全景图全自动拼接算

处的图像清晰度较高。本文所用的奇异值分解最小二乘法,在特征点数目较少甚至存在 3 点接近共线的时候也可以准确估算出变换矩阵的参数,这就不仅提高了算法的稳健性,同时也提高了配准精度,更大程度上降低配准误差。另外在进行图像融合时,多频带融合算法可以有效地降低图像运动、参数估计误差等因素带来的高频细节模糊,减少了合成图像的鬼影,最终产生的全景图清晰度较高。

法,其充分利用了相位相关法可以稳健估计图像间平移量同时对运动物体和噪声不敏感的特性,该平移量用在估算图像重叠区域、估计角点匹配区域和验证透视变换参数最小二乘解的过程中。RANSAC 算法、LM 非线性优化保证了图像配准的稳健性与精度。在图像混合过程中采用了多频带融合算法以降低普通双线性插值融合算法带来的高频细节模糊。实验结果表明,该算法在速度、稳健性和图像拼接清晰度方面与现有代表性算法均有较大提高。

参考文献 (References)

- 1 Brown M, Lowe D G. Recognising panoramas[A]. In: Proceedings of the Ninth IEEE International Conference on Computer Vision-Volume 2[C], Washington, DC, USA, 2003: 1218 ~ 1225.
- 2 Shum H Y, Szeliski R. Panoramic Image Mosaics[R]. MSR-TR-97-23, Microsoft Research Center.
- 3 Szeliski R, Kang S. Direct methods for visual scene reconstruction

- [A]. In: IEEE Workshop on Representations of Visual Scenes[C], Cambridge, Massachusetts, USA, 1995: 26 ~ 33.
- 4 ZHAO Xiang-yang, Du Li-min. An automatic and robust image mosaic algorithm[J]. Journal of Image and Graphics, 2004, 9(4): 417 ~ 422. [赵向阳, 杜立民. 一种全自动稳健的图像自动拼接融合算法[J]. 中国图象图形学报, 2004, 9(4): 417 ~ 422.]
- 5 Reddy B S, Chatterji B N. A FFT-based technique for translation, rotation, and scale invariant image registration [J]. IEEE Transactions on Image Processing, 1996, 5(8): 1266 ~ 1271.
- 6 Harris C, Stephens M. A combined corner and edge detector[A]. In: 4th Alvey Vision Conference [C], Manchester, UK, 1988: 147 ~ 151.
- 7 Fishchler M A: Random Sample Consensus: a paradigm for model fitting with application to image analysis and automated cartography [J]. Communication Association Machine, 1981, 24(6): 381 ~ 395.
- 8 Hartley R, Zisserman A. Multiple View Geometry in Computer Vision [M]. Cambridge, UK: Cambridge Press, 2000.
- 9 Gilbert Strang. Linear Algebra and its Applications[M]. 1977.
- 10 Burt P J, Adelson E H. A multiresolution spline with application to image mosaics[J]. ACM Transactions on Graphics, 1983, 2(4): 217 ~ 236.