

# 一种全自动稳健的图像拼接融合算法

赵向阳 杜利民

(中国科学院声学研究所语音交互室, 北京 100080)

**摘要** 提出了一种全自动稳健的图像拼接融合算法。此算法采用 Harris 角检测算子进行特征点提取, 使提取的精度达到了亚像素级, 然后以特征点邻域灰度互相关法进行特征点匹配得到了初步的伪匹配集合, 并运用稳健的 RANSAC 算法将伪匹配点集合划分为内点和外点, 在内点域上运用 LM 优化算法精确地估计出了图像间的点变换关系, 最后采用颜色插值对交接处进行颜色过渡。整个算法自动完成, 它对有较大误差或错误的特征点数据迭代过滤, 并用提纯后的数据来做模型估计, 因而对图像噪声和特征点提取不准确有强健的承受能力。在参数估计时, 以特征点的坐标位置误差而不是亮度误差来构造优化函数, 克服了以往算法对光照的敏感性, 使算法更具有实用性。实验结果表明, 该算法融合效果比较理想, 鲁棒性强, 具有较高的使用价值。

**关键词** 特征点提取和匹配 内点外点 优化估计 图像拼接

中图法分类号: TP31.3 文献标识码: A 文章编号: 1006-8961(2004)04-0417-06

## An Automatic and Robust Image Mosaic Algorithm

ZHAO Xiang-yang, DU Li-min

(Institute of Acoustics, Chinese Academy of Sciences, Beijing 100080)

**Abstract** In this paper, an automatic and robust image mosaic algorithm is presented. In this algorithm, Harris corner detector is used to extract feature points, which gains sub-pixel precision for features extraction. Then, a pseudo matching set is obtained by comparing local neighborhoods of features through intensity cross-correlation method and these pseudo matches are divided into inliers and outliers using robust RANSAC algorithm. In the inliers sub-set, LM algorithm is used to estimate the point transformation matrix between two images accurately. In the end, the image color of the overlapping band is smoothed with bilinear interpolation technique. The whole algorithm is completed automatically. It filters the noisy or wrong input data iteratively, then estimates the model parameters through pure data, so it has strong error-tolerant capacity for the image noise and inaccuracy of feature extraction. When estimating the model parameters, the energy function is constructed based on the position errors of features instead of the features' intensity errors, which conquers the original methods' shortcoming of sensitivity to illuminating conditions and makes this algorithm more practical. Experimental results show the image mosaic effect is wonderful and the algorithm is stable very much. It is high valuable in practice.

**Keywords** extracting and matching feature points, inliers and outliers, optimal estimating, image mosaic

## 1 引言

由于摄像设备的机械限制, 广角的场景需要将不同角度的成像进行平滑无缝的拼接融合, 从显微镜、望远镜、手提摄像机到卫星成像系统也都需要用图像融合技术来构造由一幅图片所不能表现的全景图, 在计算机图像合成, 计算机视觉中, 也需要将融合的静态纹理图像映射到 3D 模型上, 生成栩栩如

生的 3D 实体。对于这些图像的拼接融合, 目前普遍的研究方法是通过两幅图像间的匹配特征点估计图像间的点变换关系, 然后分别通过对两幅图像重采样注册到一幅更大的空白图像中, 并在注册的过程中在两幅图像的交接边界处进行颜色平滑过渡, 最后形成一张融合的包含 2 幅图像信息的图像。整个算法主要包含 3 步: 特征点的提取和匹配; 图像间点变换的估计; 图像交接处的无缝过渡。对每步算法的研究虽然已取得了一些成果, 但基本上还没有形成

一种稳健完全自动的拼接融合技术。Richard 以手动确定至少 4 对特征点<sup>[1]</sup>,并以特征点像素的亮度误差构造优化函数来估计图像间的点变换关系,取得了比较好的平面和深度场景融合效果,但此算法以亮度误差来做点变换估计,这对图像的光照变化非常敏感且整个算法的自动化程度不够。Pollefev 详细研究了特征点自动提取和匹配问题<sup>[2]</sup>,但没有针对图像处理做更深入的研究。Peter 分别运用多分辨率样条技术和单调函数插值技术对图像的交接处进行了有效的颜色平滑过渡<sup>[3]</sup>,不过这是在特征点匹配和点变换的准确估计基础上完成的。

基于以上算法的优缺点,本文给出了一种从特征点提取、匹配到图像间点变换估计和拼接融合的全自动稳健算法。该算法的根本思想是对自动提取的特征点集合,以一种非常稳健的选择策略层层过滤,保证以提纯后准确的特征点数据来做点变换估计,这样就对自动的特征点提取和匹配的准确性放松了限制条件,把核心的任务移交给了特征点的过滤策略和点变换的估计算法。

## 2 图像特征点提取和匹配

为了保证特征点的准确提取和匹配,要求每幅图像都用一样的特征提取算法,每个特征点周围应提供足够的用于判定匹配的信息(如靠近图像边缘的特征点就剔除)。采用如下的 Harris 角检测算子来做特征点提取<sup>[2]</sup>:

$$M = \begin{bmatrix} (\partial I / \partial x)^2 & (\partial I / \partial x)(\partial I / \partial y) \\ (\partial I / \partial x)(\partial I / \partial y) & (\partial I / \partial y)^2 \end{bmatrix} \quad (1)$$

其中, $I$ 为像素点亮度, $(x, y)$ 为像素点坐标。 $M$ 的两个特征值的大小反映了像素点“突显”程度,如果 $(x, y)$ 是一个特征点,那么 $M$ 的两个特征值在以 $(x, y)$ 为中心的局部范围取得极大值。因此一幅图像的所有特征点对应了一个特征点函数:

$$R = \det M - k(\text{trace} M)^2 \quad (2)$$

这里, $k$ 是一个实验参数取值(一般取 0.04),特征点就是此函数的局部极大值点。实践中,应先对图像进行低通滤波以减少图像的噪声。此方法特征点提取精度可以达到亚像素级。提取特征点后,每幅图像的特征点有相当冗余,采用以特征点为中心的灰度窗互相关法来对有冗余的特征点进行匹配计算,以像素为中心选择数据  $(2N+1) \times (2N+1)$  的灰度窗,在 2 幅图像间做互相关运算

$$C = \left( \sum_{i=-N}^N \sum_{j=-N}^N (I(x-i, y-j) - \bar{I})(I'(x'-i, y'-j) - \bar{I}') \right) \quad (3)$$

式中, $I$ 和 $I'$ 为亮度值, $\bar{I}$ 和 $\bar{I}'$ 为这个窗所有像素点的亮度平均值。通过这样的匹配可以大幅度减少不匹配的特征点,当然留下的“匹配”特征点中也存在一些伪匹配特征点。

## 3 图像间点变换的自动稳健估计

### 3.1 图像间点变换的线性求解

图像 $A$ 和 $A'$ 间的投影变换为(以齐次坐标表示)

$$k \begin{bmatrix} x'_i \\ y'_i \\ w'_i \end{bmatrix} = \begin{bmatrix} h_0 & h_3 & h_6 \\ h_1 & h_4 & h_7 \\ h_2 & h_5 & h_8 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ w_i \end{bmatrix}$$

$$\text{或向量形式: } kx'_i = Hx_i \quad (4)$$

这里, $k$ 是比例因子, $H$ 有 8 个自由度,理论上选择至少 4 对特征点就可以估计出 $H$ 。通过叉积消去比例因子得

$$x'_i \times Hx_i = 0 \quad (5)$$

令, $h_j^T$ 表示 $H$ 的 $j$ 行,那么对每一对匹配点,式(5)可以得到 2 个独立的线性方程组

$$\begin{bmatrix} 0^T & -w'_i x_i^T & y'_i x_i^T \\ -w'_i x_i^T & 0^T & -x'_i x_i^T \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix} = 0 \quad (6)$$

式(6)解得的 $H$ 就一一对应地把一幅图像的点映射到另一幅图像上了<sup>[4]</sup>。由于特征点检测的误差和选取的 4 点中有 3 点可能共线,这种线性解法的结果往往很不稳定,必须要通过一定数量的匹配特征点集合来进行非线性优化。这里以对称投影位置误差最小化来优化估计 $H$ ,其误差函数为

$$\sum_i d(x_i, H^{-1}x'_i)^2 + d(x'_i, Hx_i)^2 \quad (7)$$

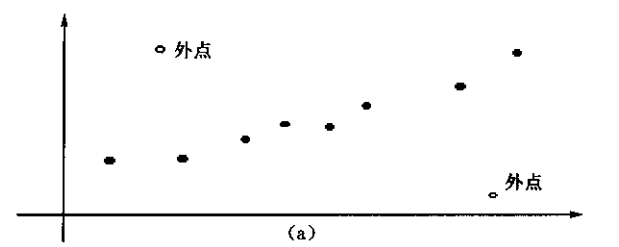
式中, $d$ 表示距离。

一般地,设特征点提取误差符合高斯分布,那么可以确定一个过滤噪声的阈值,并定义 $d_v = d(x_i, H^{-1}x'_i)^2$ 或 $d_v = d(x'_i, Hx_i)^2$ 小于等于阈值的特征点为 $H$ 的内点,大于阈值则为外点。很明显只有内点才适合做点变换估计。由于第一步自动提取和匹配得到的伪匹配特征点集合存在一些误匹配对和特征点的精度误差,因此需要一种容错能力很强的算法来提纯内点。

### 3.2 RANSAC 算法提纯数据

以直线拟合为例说明 RANSAC 算法的思想:

在数据集如图 1(a)中随机选取两点形成一条直线,并通过一定的阈值来寻找此直线的内点,由这个内点集合线性解出新的直线,再根据此新直线寻找它对应的内点,不断地重复这样的随机采样,直到某一次采样使内点数量最大,那么这次获得的直线估计就是此集合的最好估计。如图 1(b)所示,直线  $bc$  获得的内点较多,它是这个集合的最好估计,并有效地剔除了外点。此方法的特点是充分地利用了所有的



测量数据,并根据阈值把它们分成了内点和外点,利用内点数据比较准确的特点来进行参数估计而剔除了不准确的测量数据,因此,得到的结果应该是一种优化的结果<sup>[5]</sup>。针对这里的估计问题,就需要随机选取 4 对“匹配”点,每次采样用 3.1 的方法计算  $H$ ,最后根据提纯的内点集合来优化估计矩阵  $H$ 。显然 RANSAC 算法有 3 个需要确定的量:内点与外点的距离阈值  $t$ ;随机采样的次数  $N$ ;多大的内点个数。

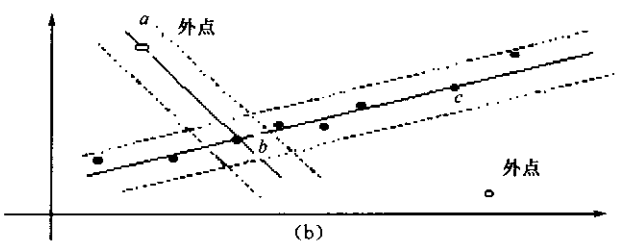


图 1 基于 RANSAC 的直线拟合

(1)距离阈值的确定 假设特征点提取误差是符合均值为 0 标准差为  $\sigma$  的高斯分布,这时, $d_v^2 = d(x'_i, Hx_i)^2$  是 高 斯 变 量 的 平 方 和,它 符 合 自 由 度 为  $m$  的  $\chi_m^2$  分布。由  $\chi_m^2$  分布的积分性质知道:符合  $\chi_m^2$  分布的随机变量的概率小于积分的上限,即

$$F_m(k^2) = \int_0^{k^2} \chi_m^2(\xi) d\xi < k^2 \tag{8}$$

那么阈值计算方法为

$$t^2 = F_m^{-1}(\alpha)\sigma^2 \tag{9}$$

则  $\begin{cases} \text{内点} & d_v^2 < t^2 \\ \text{外点} & d_v^2 \geq t^2 \end{cases} \tag{10}$

其中,  $\alpha$  表示符合内点条件的特征点对是内点的先验概率,一般取值 0.95,即符合式(10)判别条件的点为内点的概率为 95%。通过  $\chi_m^2$  分布的查值表可知  $F_4^{-1}(0.95)\sigma^2 = 5.99\sigma^2$ ,因而对提取的特征点计算出  $\sigma$  就可以完全确定阈值。

(2)随机采样的次数 随机采样时,理想情况是搜索所有可能的组合,但计算量非常巨大。如果模型估计需要的匹配点为  $s$  对,那么可以确定一个适当的采样次数  $N$ ,以保证此时采样的  $s$  对匹配点都是内点的概率足够高就可以了,设以  $p$  来表示此概率,那么实践中只要保证此概率超过 95%就能满足实际应用了。设  $p_1$  为任何一对匹配点是内点的概率,则  $\epsilon = 1 - p_1$  就是任何一对匹配点为外点的概率。那么采样到  $N$  次时有

$$(1 - p_1^s)^N = 1 - p \tag{11}$$

这样就可以确定需要采样的次数  $N$  为

$$N = \log(1 - p) / \log(1 - (1 - \epsilon)^s) \tag{12}$$
以本文的模型为例: $s=4$ ,令  $p=0.99$ ,则随着  $\epsilon$  的增加需要随机采样次数的变化情况如表 1 所示。

	集合中外点的比例(%)								
	5	10	20	30	40	50	60	70	80
需要的随机采样次数	3	5	9	17	34	72	178	567	2 876

(3)恰当的内点数的确定 如果总的特征点对为  $n$  对,则内点数为  $(1 - \epsilon)n$ 。在实验中也可以多次地迭代直到内点的数目基本趋于一致。

### 3.3 $H$ 的非线性估计

以式(7)作为优化目标函数 
$$F(H) = \sum_i d(x_i, H^{-1}x'_i)^2 + d(x'_i, Hx_i)^2 \tag{13}$$
需要优化的参数为  $\{h_0, h_1, \dots, h_8\}$ ,利用 LM 非线性优化算法来最小化式(13),可以求得 LM 算法所需的近似 Hessian 矩阵  $a$  和加权梯度向量  $b$ ,其各个分量元素为

$$\begin{aligned} a_{kl} &= \sum \frac{\partial F(H)}{\partial h_k} \frac{\partial F(H)}{\partial h_l} \\ b_k &= - \sum F(H) \frac{\partial F(H)}{\partial h_k} \end{aligned} \tag{13}$$

这样就可以求得  $H$  参数增量:  $\Delta H = (A + \lambda I)^{-1}b$ ,  $\lambda$  是一个随迭代步骤自动调节的因子<sup>[6]</sup>,如果误差  $F(H)$  减小则  $\lambda$  减小,否则  $\lambda$  增大,每一步更新参数  $H^{(n+1)} = H^n + \Delta H$  直到  $F(H)$  下降到指定的精度。实验表明此算法可以快速得到准确的变换矩阵  $H$ 。

综上,RANSAC 算法对  $H$  的自动稳健估计步

骤如下：

- (1) 运用 Harris 角点检测算子提取 2 幅图各自的特征点；
- (2) 用邻域像素灰度互相关法粗略估计 2 幅图像的伪匹配点；
- (3) 运用 RANSAC 估计图像间变换矩阵：重复  $N$  次随机采样
  - (a) 随机选取 4 对伪匹配点(选取的 4 点应保证  $H$  的秩为 3)，线性地计算变换矩阵  $H$ ；
  - (b) 计算每对伪匹配点离  $H$  的垂直距离  $d_v$ ；
  - (c) 根据内点距离小于阈值  $d_v < t = 2.447\sigma$  的原则计算  $H$  的内点，并在此内点域上重新估计  $H$ ；
- (4) 由 RANSAC 的  $N$  次随机采样得到了最大内点集合，运用 LM 优化算法对  $H$  进行优化估计；
- (5) 根据  $H$  再对所有数据进行内点计算，直到内点数目趋于一致。

4 图像交叠处的颜色融合

一旦准确地求得了图像间的点变换关系  $H$ ，就可以确定图像间交叠的准确位置，并对两幅图像进行重采样，注册到一幅新的空白图像中形成一幅融合图像。如图 2 所示，求得了图像间的变换矩阵  $H_{12}, H_{32}, H_{2m}$ ，就可以得到融合图像矩阵  $m$  和图像矩阵 1, 2, 3 之间的变换矩阵  $H_{1m}, H_{2m}, H_{3m}$ ，根据此变换对融合图像矩阵逐个像素地注册，如像素  $i$  在图像 1 中对应的像素  $k$  是由  $k$  周围的 4 个像素双线性插值构成的，因此注册过程是对图像的重采样过程。在融合过程中，可以对交叠处进行颜色插值或用多分辨率样条技术进行颜色过渡<sup>[3,7]</sup>。本文采用了简单的双线性插值方法，设图像  $I_1$  和  $I_2$  交叠区的

$x$  和  $y$  轴最大和最小值分别为  $x_{\max}, x_{\min}$  和  $y_{\max}, y_{\min}$ ，则融合图像  $I$  在交叠区的颜色值为

$$I = \beta I_1(x, y) + (1 - \beta) I_2(x, y) \tag{14}$$

其中， $\beta = (x_{\max} - x) / (x_{\max} - x_{\min})$ ，这种方法在多数情况下都可以取得比较好的效果，如果在交叠区 2 幅图像的亮度差别很明显时，可以采用复杂的多分辨率样条技术来解决。

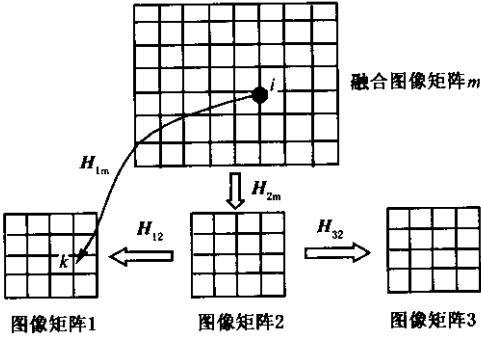


图 2 图像的重采样注册融合示意图

5 实验结果

采用  $800 \times 533$  和  $1024 \times 768$  的 24 位 RGB 图像，在特征点提取阶段，通过适当地限制特征点数目来减少 RANSAC 的随机选取次数，特征点灰度窗的大小为  $15 \times 15$  像素。内点距离阈值为  $t = 1.25$  像素，即特征点提取的标准差为 0.5107 像素。“书架”匹配出了 45 对伪匹配点，RANSAC 随机采样了 43 次，以连续 2 次的内点数目相差不超过 1 为最终迭代终止条件，LM 算法一般迭代 8 次左右就可以收敛。“河岸”是 3 张照片融合成的，“研究室”是由 8 张照片融合而成的。图 3~图 9 给出了特征点提取，特征点匹配和图像拼接融合的实验结果。



图 3 图像的特征点自动提取

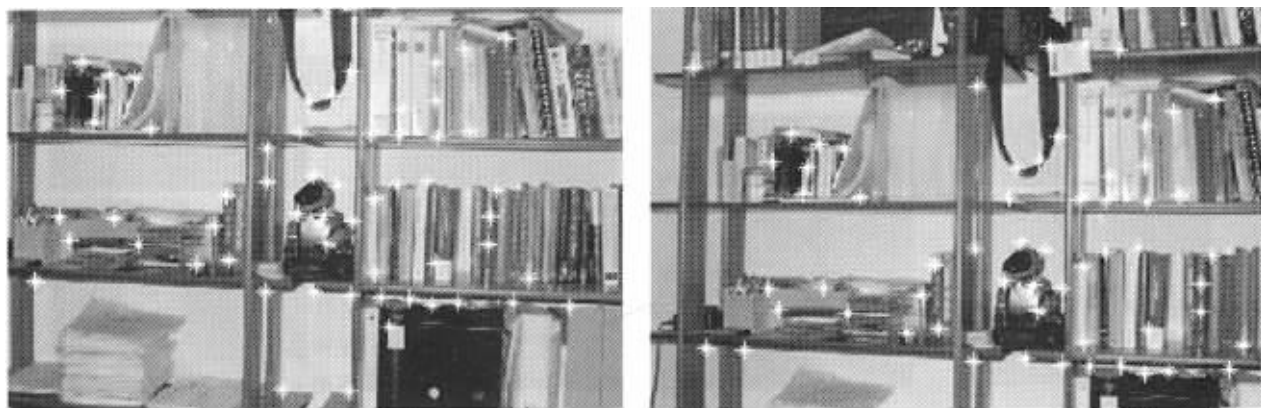


图4 基于灰度互相关法的伪特征点匹配集合



图5 自动的图像拼接融合结果



图6 河岸的3幅多角度原始图像



万方数据

图7 3张河岸图像的自动拼接融合结果



图 8 “语音交互技术研究”室的 8 张多角度原始照片

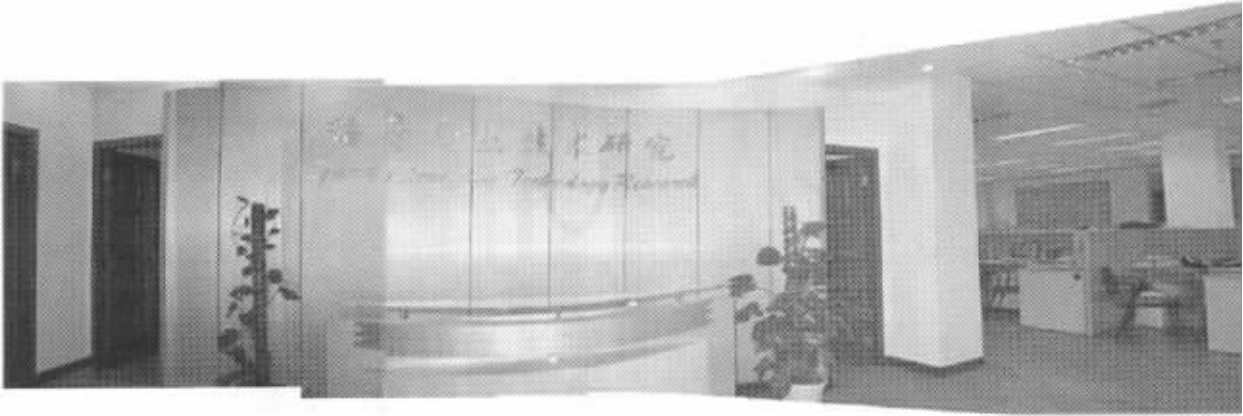


图 9 “语音交互技术研究”室图像自动融合场景

6 结 论

本文提出了一种全自动图像拼接融合算法,此算法有这几个特点:采用 Harris 角检测算子提取的特征点精度高;RANSAC 把特征点分成内点和外点的策略被理论地定量化,为图像的自动拼接融合提供了强有力的保证;算法对数据层层过滤提纯数据的过程非常稳健;LM 算法估计图像间点变换的域得到了优化,保证了 LM 算法的精确估计和快速收敛。整个算法无论是对输入数据本身还是图像的外部复杂光照条件都有很强的容错能力,是一种稳健有效的实用算法。

参 考 文 献

1 Richard Szeliski. Video mosaics for virtual environments[J]. IEEE Computer Graphics and Applications, 1996,16(2):22~30.

2 Pollefeys M. Self-Calibration and Metric 3D Reconstruction from Uncalibrated Image Sequences [D]. Belgium: K. U. Leuven,1998.

3 Peter J Burt, Edward H Adelson. A multiresolution spline with application to image mosaics [J]. ACM Transactions on Graphics, 1983,2(4):217~236.

4 Richard Hartley, Andrew Zisserman. Multiple View Geometry in Computer Vision[M]. Cambridge: The Press Syndicate of The University of Cambridge,UK,2000.

5 Fischler M A, Bolles R C. 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.

6 Press W H, Teukolsky S A, Vetterling W T, et al. Numerical Recipes in C[M]. Cambridge: Cambridge University Press,UK, 1992:681~688.

7 Richard Szeliski, Heung-Yeung Shum. Creating full view panoramic image mosaics and texture-mapped models [J]. SIGGRAPH 97 Conference Proceedings, 1997,3(1):251~258.



赵向阳 1976 年生,分别于 1999 年和 2002 年获得电子与信息工程专业学士和硕士学位,现在中国科学院声学研究所攻读信号与信息处理专业博士学位。研究方向为计算机视觉、语音驱动的人脸表情合成、科学数据可视化等。



杜利民 1957 年生,博士生导师, IEEE 高级会员,1983 年、1987 年、1991 年分别于北京大学、中国科大研究生院、中国科学院声学研究所获理学学士、工学硕士、理学博士学位。主要从事信号处理和语音交互技术的研究。