

# 大视场视频全景图快速生成方法

李庆忠, 耿晓玲, 王冰

(中国海洋大学工程学院, 青岛 266100)

**摘要:** 针对视频图像拼接中配准误差的积累问题, 提出一种适合大面积静态场景观测的视频全景图快速生成方法, 包括摄像机在一维单幅扫描方式和二维多幅扫描方式下的视频全景图的直接拼接算法。实验结果表明, 无论摄像机工作在单幅方式还是多幅方式下, 该方法都可以快速、准确地拼接出较高质量的视频全景图像, 能有效避免局部配准误差的积累与传播, 并解决一般全局配准算法复杂费时的问题。

**关键词:** 大视场全景图; 视频拼接; 全局配准误差积累

## Fast Construction Method for Large Field Video View Panorama

LI Qing-zhong, GENG Xiao-ling, WANG Bing

(College of Engineering, Ocean University of China, Qingdao 266100)

**【Abstract】** Aiming at the problem of registration error accumulation, this paper presents a fast video mosaic method especially suitable for real-time observation of large static scene. The proposed method is composed of direct registration techniques for both 1D scan mode and 2D scan mode of a camera. Experimental results show that the proposed algorithm can construct seamless video mosaic quickly both for 1D scan mode and for 2D scan mode of a camera, and it can avoid the error accumulation and propagation resulted from the concatenation of pair-wise transformations. Moreover, the proposed method can greatly reduce the complexity and time cost of the general global alignment algorithms.

**【Key words】** large field view panorama; video mosaics; global alignment error accumulation

### 1 概述

视频拼接技术是将各帧图像相互配准而生成一幅包含视频序列所有信息的全景图。该技术在大面积静态场景观测、虚拟现实、视频检索以及高分辨率图像的获取等方面有着重要应用前景。

静态场景的大面积观测主要涉及高大建筑物高分辨率全景图的获取、遥感观测以及海底观测全景图的快速生成等。为完成这些实际的全景观测任务, 摄像机一般可按照任意路线进行拍摄, 即不但可以进行一维(1D)单幅扫描拍摄, 还可以对场景进行二维(2D)多幅“ $\infty$ ”或“Z”型扫描。在这些基于视频的全景观测中, 摄像机的主要运动一般为平移运动, 且一般要求随着视频图像的不断输入, 在屏幕上重建的全景图像可以不断延展, 故要求视频拼接算法具有较好的实时性和精确性。

目前的图像拼接方法重点研究相邻2幅图像间的配准。对于单幅扫描的情况, 一般是采用对相邻图像两两配准<sup>[1-2]</sup>, 然后再将得到的帧对间的变换进行连乘, 以得到某帧图像与全景图像坐标系间的变换矩阵。对于全景观测问题, 随着大量摄取视频的输入, 会使矩阵连乘越来越多, 误差积累与传播会急剧增加, 从而拼接生成的全景图质量就会急剧下降。对于摄像机按照任意路径2D多幅扫描的视频拼接问题, 图像各帧间不仅存在时间相邻性, 还存在空间相邻性, 因此必须采用全局配准策略以有效抑制误差积累造成的全景图像不一致问题。目前已经有大量工作者对2D多幅扫描方式下全局图像配准方法<sup>[3-4]</sup>进行了研究, 这些方法的基本思路是对一个序列子集在两两配准的基础上, 利用线图技术进行最优连接路径的寻找, 并利用线性或非线性优化技术, 进行最佳变换参数的处理, 优化目标是使全局配准误差达到最小。这些

方法的一个共同特点是必须对视频序列的一个子集进行全局优化处理, 虽然可以得到质量较好的全景图, 但复杂性和实时性较差。

针对以上问题, 本文提出了一种将当前帧图像与全景图像直接进行配准的方法。该方法在一维或二维多幅视频图像拼接时, 可以同时考虑时间和空间的邻接性, 不需要对成批图像整体进行优化处理, 不存在变换矩阵连乘造成的误差积累和传播, 可快速、准确地完成当前输入图像和全景图像的拼接。

### 2 一维单幅扫描方式下的直接拼接方法

在一维单幅扫描方式下, 传统的全景图拼接算法如下: 首先对相邻图像进行特征点提取, 再进行特征点匹配, 然后根据得到的匹配点集, 进行两图像间平面运动变换的鲁棒估计, 由此得到一系列局部变换矩阵 $T_{i,i+1}$ , 即 $i$ 帧图像和 $i+1$ 帧间的变换矩阵。若选择第一帧图像为参考帧(即全景图像), 则第 $k$ 帧( $2 \leq k \leq N$ )图像到全景图像间的变换矩阵可由下式得到<sup>[5]</sup>:

$$T_{ref,k} = \prod_{i=1}^{k-1} T_{i,i+1} \quad (1)$$

最后根据得到的 $T_{ref,k}$ , 将各帧图像映射到全景图像坐标系, 并利用图像融合技术形成全景图像。

由以上拼接过程可知, 相邻图像间的变换 $T_{i,i+1}$ 是通过特征点集采用非线性优化算法估计得到的, 这说明求得的变

**基金项目:** 国家“863”计划基金资助项目(2006AA09Z237)

**作者简介:** 李庆忠(1963-), 男, 教授、博士, 主研方向: 图像处理, 机器人视觉技术; 耿晓玲、王冰, 硕士研究生

**收稿日期:** 2009-06-26 **E-mail:** lqzhz@yahoo.com.cn

换只能是在最小二乘意义下的最佳运动参数，必然存在一定的误差。因此，根据式(1)，随着  $k$  值的不断增大，求得的  $T_{ref,k}$  中存在的误差积累就越大。

为克服这种误差积累，本文提出了当前图像与全景图像间的直接配准方法，算法的流程如图 1 所示。假设第一帧图像为参考帧，即全景图像坐标系。该算法主要由特征点求取、特征点匹配、当前图像与全景图像间变换计算、图像拼接融合 4 个模块组成，下面具体介绍各模块的实现。

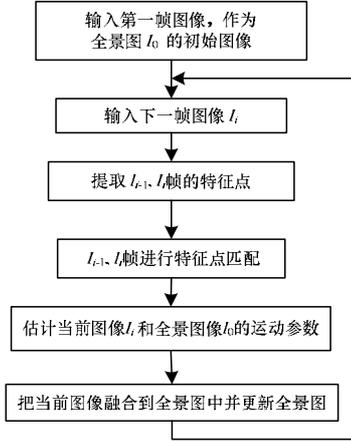


图 1 一维单幅扫描方式下视频图像拼接流程

### 2.1 相邻两图像特征点的求取

为保证视频序列全景图拼接的快速性，不是对所有视频图像帧都进行拼接，而是根据摄像机运动快慢，每隔  $m$  帧采集一幅图像进行拼接， $m$  一般由经验确定，一般要求相邻图像不少于  $1/3$  的重叠部分。对输入的要拼接的图像，本文采用基于 Harris 算子的方法提取其特征点。特征点的数量和质量严重影响后续全局运动估计的时间消耗量，为提高运算速度，必须提高特征点的质量和降低特征点的数目。为此，本文将图像分成互不重叠的子块，然后在每个子块内只选择特征点评价函数值最大的特征点为候选匹配特征点。

### 2.2 相邻两图像特征点的匹配

为提高特征点的匹配速度和准确性，本文采用指导性的匹配寻找策略。即对于当前帧中的某个特征点  $(x, y)$ ，它在前一帧图像中的搜索区域是以  $(x+x_0, y+y_0)$  为中心的  $w \times w$  区域。其中， $(x_0, y_0)$  是采用基于傅里叶变换的相位法估计得到的图像平移矢量。因为在全景观测中，图像的运动主要为平移运动，而相位相关法常用来检测 2 幅图像间的平移矢量，对噪声具有较大的容忍性，所以这种指导性的匹配寻找策略是可行的。为提高相位法的计算速度，本文首先对原始图像进行一次小波变换，然后只保留含有主要信息的低频子带 LL1。基于傅里叶变换的相位法只在大小为原图像  $1/4$  的 LL1 低频子带间进行，由此可提高计算速度。

对于特征点的匹配，本文采用不受光照变化影响的归一化相关系数作为匹配准则：

$$C = \frac{\sum_{(x,y) \in W} [I_1(x,y) \cdot I_2(x+x_0, y+y_0)]}{\sqrt{\sum_{(x,y) \in W} I_1^2(x,y)} \sqrt{\sum_{(x,y) \in W} I_2^2(x+x_0, y+y_0)}} \quad (2)$$

其中，相关计算窗口宽度  $W$  可取 5 或 7。因为只在重叠区域中的规定搜索窗口内进行匹配点寻找，所以可大大降低匹配计算量。

### 2.3 当前图像与全景图像间变换计算

本文采用 8 参数投影模型来描述两图像间的平面运动<sup>[6]</sup>：

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} a_0 & a_1 & a_2 \\ a_3 & a_4 & a_5 \\ a_6 & a_7 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (3)$$

下面以图 2 为例说明任一帧图像和全景图像间运动变换的直接计算方法。

(1)对于第 0 帧和第 1 帧，因为 0 帧被选为参考帧，即全景图像，所以 1 帧和 0 帧的变换就是 1 帧和全景图像间的变换。采用 2.2 节中的算法得到 0 帧和 1 帧重叠区域内的匹配点集，再利用 L-M 算法进行投影运动参数的优化估计，得到的变换矩阵记为  $T_{0,1}$ 。

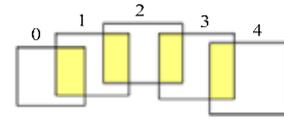


图 2 一维直接拼接过程示意图

(2)对于第 1 帧和第 2 帧，由 2.2 节中的算法可得到 1 帧和 2 帧重叠区域内的匹配点集，分别设为  $match_1$ ， $match_2$ 。不是直接在这组匹配点集间利用 L-M 算法进行 8 参数估计，而是先将图像 1 的特征点集  $match_1$  利用步骤(1)得到的  $T_{0,1}$  映射到全景图像中，设得到的 1 帧在全景图像中特征点集为  $match'_1$ ，这些特征点的坐标可以为浮点数，像素值保持不变。然后对于得到的新的匹配点集  $match'_1$ ， $match_2$ ，再利用 L-M 算法进行 8 参数估计，设得到的变换矩阵是  $T_{0,2}$ ，显然这就是 2 帧图像和全景图像间的变换矩阵。

(3)对于第  $i$  帧和第  $i+1$  帧，设它们重叠区域的匹配点集为  $match_i, match_{i+1}$ ，求得的第  $i$  帧和全景图像间的变换矩阵为  $T_{0,i}$ ，则利用  $T_{0,i}$  可得到  $match_i$  在全景图像中的对应特征点集  $match'_i$ ，然后对于新的匹配点集  $match'_i, match_{i+1}$ ，利用 L-M 算法就可以求得第  $i+1$  帧和全景图像间的直接变换矩阵  $T_{0,i+1}$ 。

(4)以此类推，重复以上步骤就可以求得任意输入图像和全景图像间的直接变换矩阵。显然，在求解过程中不存在传统两两变换矩阵连乘造成的误差积累和传播。

### 2.4 当前图像与全景图像的融合

当得到了每帧图像相对于全景图像的运动变换后，就可以将它拼接融合到全景图中了。具体方法如下：

(1)设图像  $I_k$  相对全景图像的变换矩阵为  $T_{0,k}$ ，设  $(x_i, y_i)$  为全景图像  $I_0$  中的特征点坐标， $(u_i, v_i)$  为图像  $I_k$  中与  $(x_i, y_i)$  匹配的特征点坐标，则

$$(x_i, y_i) = T_{0,k}(u_i, v_i) \quad (4)$$

根据式(4)可以算出图像  $I_k$  在全景图像中的 4 个角点坐标，由此得出其外接最小矩形的左上角点坐标  $(x_{\min}, y_{\min})$  和右下角点坐标  $(x_{\max}, y_{\max})$ 。

(2)在以上得到的外接矩形范围内，扫描全景图像中的每一个点  $(x_i, y_i)$ ，通过  $(u_i, v_i) = T_{0,k}^{-1}(x_i, y_i)$  计算在  $I_k$  中的坐标值，并通过维纳滤波插值算法可求出该点的像素值。对于彩色图像，应对  $R, G, B$  这 3 个分量分别进行插值计算。

(3)由于拍摄过程中光照存在差异，因此对于全景图像  $I_0$

与图像  $I_k$  重叠区域部分, 本文采用加权平均融合法, 对 2 幅图像的重叠区域内的对应像素值采用渐入渐出的方法平滑过渡。

### 3 二维多幅扫描方式下的直接拼接方法

二维多幅扫描方式下的拼接算法和一维情况类似, 不同之处在于进行当前图像与全景图像间的变换计算时, 不但要考虑时间相邻性, 还要考虑空间相邻性。下面以图 3 为例介绍当前图像与全景图像间变换矩阵的计算方法, 其他模块和一维相同, 不再赘述。

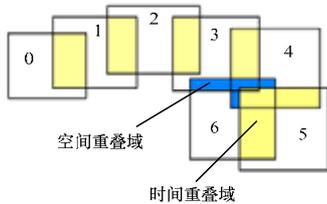


图 3 二维多幅扫描时图像拼接示意图

在图 3 中, 图像  $I_0$  至  $I_5$  的拼接方法与第 2 节介绍的一维拼接方法相同。但当拼接图像  $I_6$  时, 通过  $I_6$  的运动方向和图像中心点在全景图像中的投影位置就可以判断出  $I_6$  在配准时必须同时考虑时间相邻性和空间相邻性。  $I_6$  和时间相邻图像  $I_5$  以及和已拼接全景图像的重叠区域如图 3 所示。

$I_6$  与  $I_5$  的重叠区域以及重叠区域内的特征点匹配集和一维方法相同, 已经求出, 记为  $match_{t_6}$ ,  $match_{s_5}$ , 其中,  $match_{s_5}$  为 5 帧中特征点映射到全景图的特征点集。下面介绍图像  $I_6$  与空间相邻图像重叠区域的确定以及特征点的匹配方法。具体步骤如下:

(1) 为避免对已经拼接各帧图像变换参数的不断记录, 从而避免占用过多内存, 本文首先利用  $I_6$  和时间相邻图像  $I_5$  的匹配特征点集, 采用快速 SVD 方法计算  $I_6$  和全景图像间的 4 参数刚体图像运动模型。

(2) 利用得到的 4 参数变换矩阵, 将图像  $I_6$  的 4 个角点映射到全景图像中, 根据这 4 个角点的坐标, 就可以快速判断出图像  $I_6$  和全景图像间的空间相邻重叠区域, 然后在对应的重叠区域中, 利用前述算法进行特征点求解和匹配, 假设得到的匹配点集记为  $match_{s_6}$ ,  $match_{s_0}$ 。

(3) 综合以上 2 组匹配特征点集, 即全景图像上的特征点集为:  $match_{s_5} + match_{s_0}$ ; 图像  $I_6$  上的特征点集为  $match_{t_6} + match_{s_6}$ 。然后对这些综合匹配点集采用 L-M 算法进行 8 参数投影模型参数的估计, 因为这些特征点包括了时间和空间相邻的特征点, 所以可得到精度较高的配准变换。

由以上配准过程可知, 在拼接当前图像时, 不需要对已经拼接图像进行整批处理, 因此具有较好的实时性。此外, 在当前图像和全景图像配准时, 同时考虑了时间相邻特征点和空间相邻特征点, 所以配准精度可得到保证, 不存在配准误差的积累问题。

### 4 实验结果

采用本文提出的方法, 进行了一维以及二维的拼接实验。实验计算机为 PC: CPU 主频 1.7 GHz, 内存 256 MB, 编程环境为 Microsoft Visual C++6.0; 拍摄视频图像大小为  $256 \times 256$ 。

图 4 为一维拼接结果, 是由 28 帧图像拼接而成的。

图 4(a)为采用传统拼接方法形成的全景图像, 由于误差累积过大, 从第 8 帧开始拼接失败。图 4(b)是按照本文算法拼接而成的全景图, 由于不存在误差的积累传播, 因此质量良好。



(a)传统算法一维拼接结果



(b)本文算法一维拼接结果

图 4 一维扫描时拼接结果对比

图 5 为摄像机二维多幅扫描时拼接的视频全景图像, 共由 18 帧拼接而成, 扫描路径呈“Z”型。由图 5 中标出的方框区域可以看出, 传统拼接算法在方框区域内出现明显的拼接不一致现象, 而本文算法得到的全景图拼接一致质量良好。表 1 是用本文拼接算法拼接一帧图像所用时间的统计结果, 可见拼接一帧图像的总时间为 353 ms, 基本满足实时拼接的要求。



(a)传统算法二维拼接结果



(b)本文算法二维拼接结果

图 5 二维扫描时拼接结果对比

表 1 本文算法拼接一帧图像所用时间

本文中算法步骤	时间/ms
小波变换	15
提取特征点	47
相位法估计运动矢量	92
特征点匹配	15
求解变换参数	17
当前帧与全景图融合	167
总计	353

(下转第 184 页)