

# 基于视频序列的图像拼接

李战明<sup>1,2</sup>, 施颖迪<sup>1,2</sup>

(1. 兰州理工大学 电气工程与信息工程学院, 甘肃 兰州 730050;

2. 兰州理工大学 甘肃省工业过程先进控制重点实验室, 甘肃 兰州 730050)

**摘要:** 为消除视频序列之间的冗余信息,以简单的摘要形式表达视频的主要内容,提出了一种基于视频序列的图像拼接方法。首先,采用改进的帧间聚类算法提取视频的关键帧;其次,利用 SIFT 算法提取关键帧的特征点,采用最近邻算法进行特征点匹配,通过引导互匹配法和投票过滤法提高匹配精度;再次通过 RANSAC 鲁棒估计算法得到所选帧间的单映矩阵,并使用 LM 非线性迭代算法对单映矩阵进行精炼。最后,利用级联单映矩阵结合加权融合算法实现了视频序列的无缝拼接,实验效果较为理想。

**关键词:** 关键帧; SIFT 特征点; 单映矩阵; 图像拼接; RANSAC 鲁棒估计算法

中图分类号: TN911.72

文献标识码: A

文章编号: 1674-6236(2015)12-0116-04

## Image mosaic based on video sequence

LI Zhan-ming<sup>1,2</sup>, SHI Ying-di<sup>1,2</sup>

(1. College of Electrical Engineering and Information Engineering, Lanzhou University of Technology, Lanzhou 730050, China; 2. Key Laboratory of advanced control for industrial process in Gansu Province, Lanzhou University of Technology, Lanzhou 730050, China)

**Abstract:** In order to eliminate the redundant information between video sequences, express the main content of the video in a simple abstract form, an image mosaic method based on video sequence was proposed. First of all, the key frames were extracted using the improved inter frame clustering algorithm. Secondly, the feature of key frames were extracted by using the SIFT algorithm, the nearest neighbor algorithm is used to match the feature points, guided complementary matching and voting filter is used to improve the matching accuracy. Then with the RANSAC robust estimation algorithm to get the selected frames' homography matrix. The LM nonlinear iteration algorithm was used to refine the homography matrix. Finally, cascade homography matrix and weighted fusion algorithm is used to realize the seamless splicing of video sequences. The experiment demonstrates this method is efficient.

**Key words:** key frame; SIFT feature; homography matrix; image mosaic; RANSAC robust estimation algorithm

DOI:10.14022/j.cnki.dzsjgc.2015.12.048

随着计算机视觉和微电子技术的不断发展,图像拼接已广泛应用于虚拟现实、计算机视觉、航空航天、军事应用、医学图像处理、遥感图像处理等多个领域<sup>[1]</sup>。视频序列的图像拼接是图像拼接的一个方面,主要是将多帧视频序列拼接为一幅视野范围更大的图像,从而更直观的反映视频序列的主要内容。

在视频序列的拼接中,相邻帧之间的重复率较高,若对每相邻的视频帧都进行一次拼接,不仅会耗费大量的时间,而且随着拼接帧数量的不断增加,拼接效果也会越来越差。利用关键帧拼接表示整个视频序列成为一种有效的方法。文献[2]中提出基于自适应模板匹配的视频图像拼接技术,算法的限制条件较多,容易因为外界干扰而造成拼接失败。参考文献[3]中提出将长视频分段,采用场景流算法拼接的一种视

频序列的拼接技术,但这种方法需要描绘视频流从起点到终点的最短流形,实时性不高,同时由于要对每列像素构建图卡,需要内存空间大。

针对以上问题,提出了一种视频序列拼接的新方法。首先,采用改进的帧间聚类的算法提取视频的关键帧,然后利用尺度不变特征点提取算法提取关键帧的特征点,结合最近邻算法(Nearest Neighbor-NN)进行特征点匹配,采用引导互匹配法<sup>[4]</sup>和投票过滤法提高匹配精度。再次,结合随机抽样一致性 RANSAC(Random Sample Consensus)鲁棒估计算法得到所选帧间的单映矩阵,利用 LM(Levenberg-Marquardt)非线性最小化迭代算法对得到的矩阵的进行精炼,最后,利用单映矩阵的级联性,并通过加权融合的方法完成视频序列的无缝拼接,取得了较好的效果。

收稿日期:2014-09-26

稿件编号:201409237

作者简介:李战明(1962—),男,陕西西安人,教授。研究方向:智能信息处理与模式识别、神经模糊系统、复杂系统的建模与控制、嵌入式网络单片机系统研究与开发等。

-116-

## 1 帧间的单映变换矩阵及关键帧选取判定

### 1.1 帧间单映变换矩阵的概念

单映矩阵变换是一种常用的帧间变换模型,主要适用于任意场景空间摄像机为旋转或者缩放运动,或者空间为平面场景和任意摄像机的运动。单映矩阵变换表示为:

$$\begin{pmatrix} \hat{x} \\ \hat{y} \\ \hat{z} \end{pmatrix} = H \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \text{其中 } H = \begin{pmatrix} h_0 & h_1 & h_2 \\ h_3 & h_4 & h_5 \\ h_6 & h_7 & h_8 \end{pmatrix}, \text{其中,一般 } H \text{ 有 } 8 \text{ 个自}$$

由度,  $h_8=1$ 。

对于相邻的关键帧,可以直接采用上述帧间变换模型,但是对于非相邻的关键帧,可以利用单映矩阵的级联性质,得到非相邻关键帧之间的单映变换矩阵。

假设第  $k$  帧、第  $h$  帧为非相邻的关键帧,以第  $k$  帧作为参考帧,第  $h$  帧为目标帧,利用单映矩阵的级联性质,可以得到两非相邻关键帧的单映变换矩阵。

$$H_{k,h} = H_{h,d} H_{i,n} \cdots H_{m,l} H_{l,k} \quad (1)$$

其中,  $H_{h,d}, H_{i,n}, \cdots, H_{m,l}, H_{l,k}$  分别是第  $k$  帧、第  $h$  帧之间的相邻关键帧的单映变换矩阵,如图 1 所示。

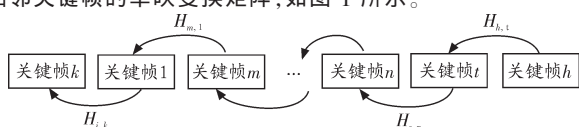


图 1 单映矩阵的级联示意图

Fig. 1 Cascade figure of homography matrix

### 1.2 关键帧的选取

在进行视频拼接的过程中,帧间重叠度比较高,数据冗余量大,而关键帧的使用很大程度上减少了视频处理的数据量。在拍摄视频影像时,由于场景中目标的运动或摄像机本身的操作(如变焦、摇镜头等)的影响,一个镜头仅用一幅关键帧不能很好地代表该镜头的内容,常需用几幅关键帧。原则上讲,关键帧应能提供一个镜头内容尽量丰富的概要。

欧式帧差法提取关键帧是通过判断视频序列之间的相似度来提取关键帧,计算帧与帧之间的帧差欧式距离,利用它来计算帧差均值,帧差方差以及帧差的差异系数来作为衡量帧与帧之间相关程度的数据量标准,根据帧差欧式距离的最大值与最小值计算出相应的平均值,并以此作为一个判别关键帧的阈值,对相应的视频帧序列进行判别,提取关键帧的帧数,得到相应视频帧序列的关键帧。借助于欧式帧差法的提取关键帧的基本思路,文中采用改进的帧差聚类的算法来提取关键帧,即将镜头中的帧计算特定的特征值,然后根据特征值的聚类将镜头分为几个关键簇,最后根据特征值的分类计算关键帧,本文是在聚类<sup>[5]</sup>的基础上实现的,文中选取帧差欧式距离、帧差均值和帧差方差作为特定的特征值,同时求取帧差欧式距离、帧差均值和帧差方差的最大值、最小值和极值,通过最大值和最小值得中间值,比较极值和中间值的大小,去掉小于中间值的极值点,剩下的点当做关

键簇,再比较关键簇和定义的关键帧的数目,若大于定义的关键帧的数目,再利用特征值选取关键帧。

算法的实现过程:

1) 假设视频  $f_1, f_2, \cdots, f_n$  代表一个视频序列,计算视频序列的帧差,即计算相邻两帧之间的每个对应的像素的灰度(亮度)变化。相邻两帧对应点的灰度(亮度)差为

$$f_d(i, j) = |f_{n+1}(i, j) - f_n(i, j)| \quad (2)$$

其中  $f_n(i, j), f_{n+1}(i, j)$  分别代表第  $n$  帧和第  $n+1$  帧像素  $(i, j)$  的灰度(亮度)。然后利用公式(1)计算相邻帧的帧差值,根据选取的视频总帧数,设定计算帧数的范围,将得到的帧差值进行比较,选取帧差值最大的相邻帧数。

2) 计算帧差的欧式距离,公式如下

$$D_1 = f_{d(i+1)}((i+1), j) - f_d(i, j) \quad (3)$$

$$D_2 = f_{d(i)}(i, (j+1)) - f_d(i, j) \quad (4)$$

$$D = \sqrt{D_1 + D_2} \quad (5)$$

3) 计算帧差之间的均值,公式如下:

$$X = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N f_d(i, j) \quad (6)$$

$$E = X/n \quad (7)$$

其中,  $M, N$  为图像的尺寸,  $n$  为总的帧数。

4) 计算帧差的方差,公式如下:

$$S_1 = \sum_{d=1}^n (X - E)^2 \quad (8)$$

$$S = \sqrt{\frac{S_1}{n}} \quad (9)$$

5) 计算帧差的欧式距离、均值和方差的最大值、最小值以及极值,利用最大值最小值求取中间值,比较中间值和极值的大小,去掉小于中间值的极值。剩下的组成关键簇,若关键簇大于定义的帧数,重复步骤 2), 3), 4), 5) 得到关键帧,反之输出关键帧。

采用改进的帧间聚类算法提取关键帧,一定程度上可以满足视频序列拼接的需要。

## 2 特征点提取与匹配

### 2.1 特征点提取

视频图像的拍摄大多是通过手持相机或者是固定在某些平台而获取的影像,由于相机的运动和外界的干扰,很容易造成视角和尺度的偏差,为了克服上述干扰,采用 SIFT (Scale Invariant Feature Transform) 算法完成图像序列特征点的提取。

SIFT 算法又称尺度不变特征提取方法,其特征的提取可以分为检测尺度空间极值、确定特征点位置、确定特征点主方向和生成 SIFT 特征描述符四步。即通过不同尺度的高斯核函数连续滤波和降采样形成高斯金字塔,再对相邻尺度的两个高斯图像相减得到 DoG (Difference of Gaussian) 金字塔的多尺度空间表示,然后取这些高斯差分图像中的局部极值,去除低对比度的点和不稳定的边缘响应点就得到尺度空间上的图像特征点。最后,用 128 维的向量(包括位置、尺度、方向等

信息)表示每个特征点,这样就生成了用于图像特征匹配的图像特征描述符。

该算法是在空间域和尺度域上同时进行特征点的计算与提取的,因此得到的特征点具有尺度不变性,能够正确的提取尺度和视角变化较大的图像序列中存在的特征点,有效地克服了输入图像中的噪声干扰。

## 2.2 特征点匹配

SIFT的匹配算法是通过计算两个特征点描述符之间的欧氏距离得到的。即找出与特征点描述符 $p_i$ 欧氏距离最近和次近的两个邻居特征点描述符 $q_i'$ 和 $q_i''$ ,然后计算 $p_i$ 与 $q_i'$ 以及 $p_i$ 与 $q_i''$ 两组描述符之间欧氏距离的比值 $r$ 。如比值 $r$ 小于规定阈值则视为匹配成功, $(p_i, q_i')$ 点则为成对匹配点,否则匹配失败。这种匹配方法简便快捷,但会产生误匹配。因此文中采用引导互匹配及投票过滤两种技术来提高匹配精度。

1)引导互匹配法:该方法的思想是缩小特征点搜索范围,并采用互映射的方法减少错误匹配。假设图像变换矩阵的当前估计为 $H$ ,根据对极几何特性, $H$ 规定了点 $x$ 在对极线 $H_x$ 周围的搜索范围。因此应在对极线 $H_x$ 周围的一定区域内搜索点 $x$ 的匹配点。此外,根据互映射原理,即匹配点对之间的映射关系的对称性,对于匹配点对 $\{p_i \leftrightarrow q_i'\}$ ,在进行匹配映射时,应存在这两点间的对应关系 $\{p_i \rightarrow q_i'\}$ 和 $\{p_i \leftarrow q_i'\}$ 。

2)投票过滤法:使用投票过滤法消除误匹配的理论依据如下:在两幅相关的图像序列中,相机的旋转和尺度变化均相对稳定,因此连接图像中各组匹配点对所形成的向量之间的偏差应符合某一规律。假设特征点 $p_i$ 和 $p_j$ 分别与特征点 $q_i$ 和 $q_j$ 相对应,即 $\{p_i \leftrightarrow q_i'\}$ 与 $\{p_j \leftrightarrow q_j'\}$ 是正确的匹配点对,则向量 $E_i = q_i - p_i$ 与 $E_j = q_j - p_j$ 应具有相同特性(长度,方向)。本文综合考虑两方面因素,首先计算图像序列中所有匹配点对的纵横坐标距离差的平均值,并将其记为该图像序列组的标准匹配阈值。然后,删除图像序列组中纵横坐标距离差与标准匹配阈值相差太远的匹配点对。即采用删除纵横坐标之差距离平均值太远的错误匹配点对的投票过滤法来提高匹配精度。

## 3 拼接算法的实现及相关实验

### 3.1 拼接算法的实现

1)采用改进的帧间聚类算法提取出关键帧。

2)利用SIFT算法提取关键帧中的特征点,采用最近邻算法计算特征点之间的匹配。

3)对于相邻的关键帧,利用RANSAC鲁棒估计算法<sup>[6]</sup>得到相邻关键帧之间的最大的一致点集和单映矩阵 $H$ 的估计,具体步骤为:①选择四组对应点组成一个随机样本并且估计矩阵 $H$ 的参数;②对假设的每组对应计算距离 $d$ ;③计算与 $H$ 一致的内点数;(选择具有最多内点数的 $H$ ,在数目相等时,选择内点标准方差最小的那个解)。

4)由初始 $H$ 值迭代精炼变换矩阵,这里采用了LM(Levenberg-Marquardt)<sup>[7]</sup>非线性最小化迭代算法,具体算法流程如下:

-118-

①对步骤2)中得到的每个特征点 $(x, y)$ 计算对应点 $(x', y')$ ;

②计算对应点之间的误差 $e = I'(x', y') - I(x, y)$ ;

③计算变换矩阵 $H$ 各分量 $h_i (i=0, 1, \dots, 7)$ 相对于误差 $e_i$ 的偏导数:

$$\frac{\partial e}{\partial h_0} = \frac{x}{D} \frac{\partial I'}{\partial x'}, \frac{\partial e}{\partial h_1} = \frac{y}{D} \frac{\partial I'}{\partial x'}, \frac{\partial e}{\partial h_2} = \frac{x}{D} \frac{\partial I'}{\partial y'}, \dots, \frac{\partial e}{\partial h_7} = \frac{y}{D} \frac{\partial I'}{\partial x'} \left( x' \frac{\partial I'}{\partial x'} + y' \frac{\partial I'}{\partial y'} \right) \quad (10)$$

来构造计算 $H$ 增量的函数 $(A + \lambda I)\Delta h = b$ (其中 $A$ 的分量为 $a_{ii} = \sum \frac{\partial e}{\partial h_k} \frac{\partial e}{\partial h_k}$ ,  $b$ 的分量为 $b_k = -\sum e \frac{\partial e}{\partial h_k}$ 。

④解 $H$ 增量函数得到 $\Delta h$ ,修正 $H$ 。

⑤判断误差值 $e$ ,若误差减小但未小于阈值,则继续计算新的 $\Delta h$ ,否则增大 $\lambda$ 值,重新计算 $\Delta h$ ;

⑥当误差 $e$ 小于规定阈值时,停止计算,得到精炼的 $H$ 。

5)引导匹配:用估计的 $H$ 去定义对极限附近的搜索区域,进一步确定特征点的对应。

6)反复迭代4)、5)直到对应点的数目稳定为止。

7)对非相邻关键帧利用单映矩阵的级联性进行计算,获取对应的图像变换矩阵。

8)根据获得的变换矩阵 $H$ 将对应的关键帧进行变换,确定两帧直接的重合区域,并将目标关键帧和参考关键帧映射到一幅空白图像中形成拼接图。

9)采用加权融合算法<sup>[8]</sup>处理图像拼接缝问题,对图像进行加权平滑处理以达到无缝拼接。

### 3.2 相关实验

实验采用手持数码相机采集两段视频,利用改进的帧间聚类算法提取出关键帧,并采用拼接算法实现了关键帧所表示的视频序列的无缝拼接,效果比较理想。

图2是一段82帧的视频,根据改进的帧间聚类的算法提取出的关键帧为第10帧、29帧、41帧、49帧、60帧、71帧(图2(a-f)),图3为关键帧所表示的视频序列拼接的效果图,图4为剪切之后的拼接效果图。

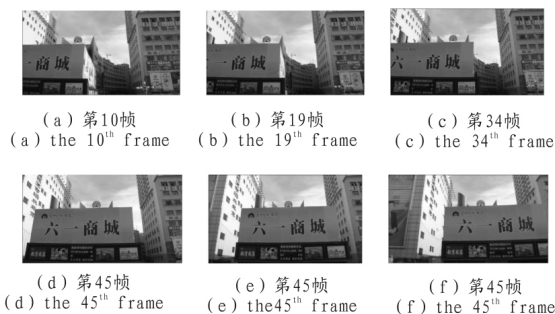


图2 关键帧所表示的视频序列拼接效果图

Fig. 2 The video mosaic based on keyframes

图5是一段118帧的视频,根据改进的帧间聚类的算法提取出的关键帧为第10帧、19帧、34帧、45帧、54帧、64帧、75帧、84帧、105帧(图5(a-i)),图6为关键帧所表示的视频序列拼接的效果图,图7为剪切之后的拼接效果图。



图3 剪切之后的拼接效果图  
Fig. 3 The result of clipped video mosaic



图4 关键帧所表示的视频序列  
Fig. 4 Keyframes sequence extraction

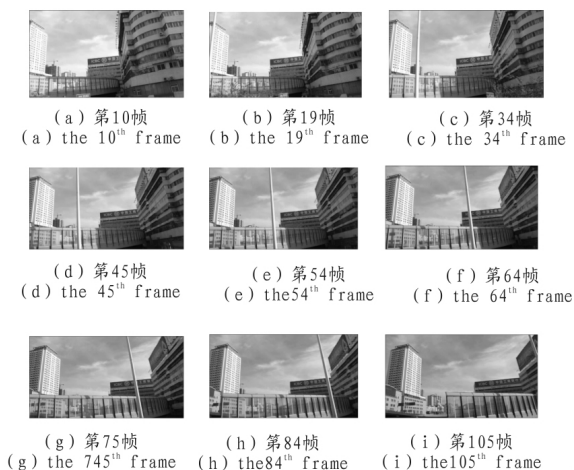


图5 关键帧所表示的视频序列拼接效果图  
Fig. 5 The video mosaic based on keyframes



图6 剪切之后的拼接效果图  
Fig. 6 The result of clipped video mosaic

#### 4 结束语

采用改进的帧间聚类算法的提取关键帧, 并利用 SIFT 算法提取特征点, 使用最近邻算法结合引导互匹配和投票过滤法提高匹配精度, 利用 RANSAC 鲁棒估计得到关键帧间的



图7 剪切之后的拼接效果图  
Fig. 7 The result of clipped video mosaic

单映矩阵  $H$ , 并使用 LM 算法精炼单映矩阵  $H$ , 获取关键帧之间的重叠区域, 结合级联单映矩阵和加权融合算法实现视频序列的无缝拼接。

参考文献:

- [1] Xiaobo Gao, Xianmei Fang. Research on Image Mosaic Algorithm Based on Computer Wizard Vector Field Algorithm[J]. Intelligent Computing Theories, 2013:156-162.
- [2] 薛艳. 基于自适应模板匹配的视频图像[D]. 安徽: 合肥工业大学, 2009.
- [3] 郭李云, 欧阳宁, 莫建文. 长视频序列拼接[J]. 计算机工程与应用, 2011, 47(14):183-185.  
GUO Li-yun, OUYANG Ning, MO Jian-wen. Long video sequence[J]. Computer Engineering and Application, 2011, 47(14):183-185.
- [4] 郭禾, 傅新元, 陈锋, 等. 任意视角的多视图立体匹配系统[J]. 计算机工程与设计, 2007, 28(12):2873-2877.  
GUO He, FU Xin-yuan, CHEN Feng, et al. Integrated multi-view stereo matching system with arbitrary viewpoints [J]. Computer Engineering and Design, 2007, 28(12):2873-2877.
- [5] Patricia Cotic, Zonko Jaglicic, Ernst Niederl-eithinger, et al. Image fusion for improved detection of near-surface defects in NDT -CE using unsupervised clustering methods [J]. Journal of Nonde structive Evaluation, 2014.
- [6] Zhaoxia Fu, Liming Wang. Optimized design of automatic image mosaic. Multimedia Tools and Applications, 2014, 72(1):503-514.
- [7] 杨敏. 基于Levenberg-Marquardt 算法的图像拼图研究 [J]. 电子工程师, 2007, 33(10):43-45.  
YANG Min. Research on Image Registration Based on Levenberg-Marquardt algorithm[J]. Electronic Engineer, 2007, 33(10):43-45.
- [8] Richard Szeliski. Computer vision: algorithms and applications[M]. German Springer, 2010.