

Real-Time Video Stitching Based on Improved SURF

Zhi'ang Chen, Xiaogang Xu, Guanlei Xu

Dalian Naval Academy, Dalian Liaoning
Email: 940287607@qq.com

Received: Oct. 8th, 2015; accepted: Oct. 22nd, 2015; published: Oct. 27th, 2015

Copyright © 2015 by authors and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Video mosaic is widely used in the fields of video conferencing, video surveillance and so on; it can increase the field of view while not reducing resolution. Due to large amount of data, complex environment and other reasons, real-time video mosaic becomes an active area of research. In this paper, technique of real-time video mosaic is studied.

Keywords

Video Mosaic, Image Registration, Image Fusion, SURF

基于改进SURF算法的实时视频拼接

陈志昂, 徐晓刚, 徐冠雷

海军大连舰艇学院, 辽宁 大连
Email: 940287607@qq.com

收稿日期: 2015年10月8日; 录用日期: 2015年10月22日; 发布日期: 2015年10月27日

摘要

视频拼接可以广泛地应用在视频会议、视频监控等诸多领域,它可以在不降低分辨率的前提下扩大视野。然而由于它具有数据量大、环境复杂等特点,使得实时视频拼接成为一个研究热点。本文对实时视频拼

接技术进行了研究。视频拼接的本质是图像拼接，首先介绍了图像拼接的一般流程、图像配准的含义，并对常用的图像配准技术进行了分析，确定了采用基于特征点提取、特征点匹配和图像融合的图像拼接方法。

关键词

视频拼接, 图像配准, 图像融合, SURF

1. 绪论

视觉是人们获取现实世界客观信息的最直接的方式[1]，随着电子技术、多媒体技术的飞速发展，越来越多的人使用数码相机、摄像机、摄像头等设备来获取视频。单个设备获取的视野范围相对较小，即使可以通过调节焦距增大视野，但是这时也会降低分辨率[2]，很多时候人们需要获取大视野的视频，这时通常采取的方法有：1) 通过移动或旋转视频采集设备，比如通过旋转的云台来动态观察四周的场景，但是这样在某一时刻只能得到一个局部视野[3]；2) 通过多个视频采集设备，比如在超市等监控场合中，但是这时得到的是多个分离的界面，没有一体化的感觉；3) 通过广角镜头和鱼眼摄像机，但是这类设备通常比较昂贵，并且会出现不同程度的畸变。

由于视频实质上是由一系列静态的图像组成，所以图像拼接是视频拼接的核心，目前已经有大量关于图像拼接的研究，常规的图像拼接方法主要追求的是拼接质量，仅需满足观测的需要，对于实时性的要求并不高，但是视频拼接与图像拼接有着很大的不同，最大的不同就是对于实时性的要求[4]。所谓实时性，是指视频拼接系统能够实时的完成图像采集、图像拼接和输出显示整个过程[5]。由于获取的视频存在视角、尺度、光照等差异，因此如何保证视频拼接在具有良好实时性的同时还具有良好的拼接效果是视频拼接的研究重点。

本文围绕实时视频拼接进行了较为深入的研究。由于实时视频拼接系统要求在规定的时间内完成视频图像的拼接，同时拼接后的视频输出也必须满足特定制式的要求，这样就要求算法具有很强的实时性和运行效率。针对这个问题本文改进基于 SURF 算法的图像配准。系统地研究了 SURF 特征提取算法并对其进行了改进，增加了自适应算法，可以精确定位特征提取范围，并且改进了特征描述符，减小了维度。实验结果表明改进后的 SURF 特征提取算法可以加快特征提取速度，同时对于图像旋转、放缩、光照等条件具有不变性。

2. 视频拼接原理

2.1. 视频拼接关键技术

视频拼接的流程如图 1 所示：

由视觉暂留原理或“余辉效应”可知，当每秒连续显示的图像超过一定帧数(一般为大于 24 帧)，人眼就无法再区分这些连续图像中的每一幅静态图像[6]。由此可知，只要将连续的图像以小于 0.04 s 的显示占用时间进行顺序播放，那么人眼接受并传递到大脑中的画面就是连续画面。这时整个图像呈现的连续画面就是视频。由此可知视频是由一帧一帧的图像组成的，所以视频拼接的本质就是图像拼接[7]。

图像拼接过程主要由图像采集、图像预处理、图像配准图像变换和图像融合这五个部分组成[8]，其中图像配准和图像融合是其中的关键技术。图像拼接的流程如图 2 所示：

1) 图像采集：图像可以通过数码相机、数码摄像机、摄像头等设备获得。本文在视频拼接时采用的设备是普通的 USB 摄像头。

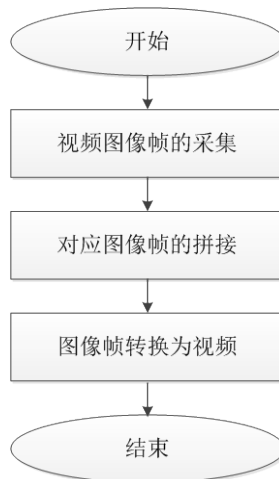


Figure 1. The flow chart of video mosaic

图 1. 视频拼接流程图

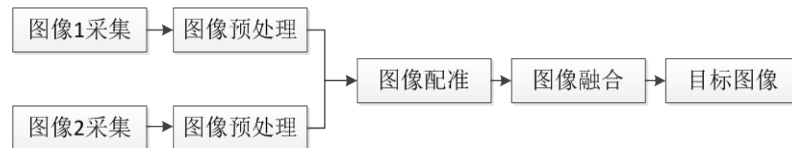


Figure 2. Process of image mosaic

图 2. 图像拼接流程

2) 图像预处理：图像预处理的目的是为了消除噪音、畸形、失真等情况，为后续的图像配准做准备[9]。由于本文采用的图像配准方法对光照、噪声等具有很好的鲁棒性且摄像头型号和参数一致，同时为了加快拼接速度，本文并未做预处理的过程。

3) 图像配准：图像配准是指把在不同条件下(比如不同时间、不同传感器、不同视角)所获取的两幅或多幅图像通过计算出变换矩阵来“对齐”具有重叠区域的图像。对于本文，不同条件主要是指视角不同。

4) 图像融合：如果将配准后图像直接拼接在一起，会出现明显的拼接痕迹，使整个图像看起来不够自然，而图像融合可以解决这一问题。

2.2. 图像配准

图像配准通俗来讲就是使两幅或多幅图像中像素点之间一一映射。而从数学上来讲，假设两幅待拼接的数字图像可以用二维矩阵表示，如果用 $I_1(x, y)$ 和 $I_2(x, y)$ 分别表示两幅图像在点 (x, y) 处的灰度值，那么图像 I_1 和 I_2 之间的配准关系[10]可以表示如下：

$$I_2(x, y) = g(I_1(f(x, y))) \tag{1-1}$$

其中， f 代表几何变换函数， g 代表灰度变换函数。配准的主要目的就是寻找最佳的几何变换关系 f 与灰度变换关系 g ，使两幅图像达到最佳对准。因为空间变换是灰度变换的前提，而且很多时候并不要求解灰度变换关系，所以配准的关键就在于寻找空间几何变换关系 f ，因此式(1-1)可以被改写为：

$$I_2(x, y) = I_1(f(x, y)) \tag{1-2}$$

目前，国内外已经有大量关于图像配准的研究，比较常用的配准方法有以下三大类：

1) 基于灰度信息的图像配准: 这种方法直接利用图像本身具有的灰度的一些统计信息来表示图像之间的相似程度, 然后采用某种搜索方法来找到使相似性度量值达到最大值或最小值的点, 以此来确定图像之间的变换关系[11]。常见的相似性度量有: ① 两幅图像的灰度的平方差之和(sum of squared different, SSD); ② 序列相似度检测(SSDA); ③ 互相关; ④ 相位相关。

2) 基于变换域的图像配准: 最常用的一种基于变换域的图像配准方法是傅里叶变换。它的原理是图像的平移、旋转和尺度等变换在频域中均有相对应的分量。相位相关法[12]是经常使用的一种基于变换域的图像配准方法, 相位相关经常被用于配准两幅图像的平移变化, 其依据的是傅里叶变换的平移特性。

3) 基于特征的图像配准法: 该方法是目前采用最多的配准方法, 它利用图像中保持不变的内部特征进行匹配。基于特征的图像配准方法的优点是大大减少了图像的信息量, 使得计算速度明显提高, 并且对图像灰度信息依赖性小、对噪声、旋转、尺度变换不是很敏感。

2.3. 图像融合

在进行图像获取的过程中, 由于图像采集设备本身参数不可能完全相同, 另外存在光照、视野、角度等差异, 使得待拼接图像在重叠区域会有差异。如果不对图像做任何处理, 只是简单的将图像进行叠加, 那么这时得到的拼接图像中就会存在非常明显的拼接缝隙, 使得图像看起来非常不自然。

图像融合目的是使配准后的图像经过拼接得到的最终图像无明显拼接缝隙, 使得图像在重叠区域实现平滑的过渡, 让人看起来觉得舒服。图像融合是图像拼接的最后一步, 它影响着图像拼接的质量同时也影响着图像拼接的速度。由于最终用于实时视频拼接中, 因此图像融合算法不能复杂, 本节对几种常用的简单图像融合方法进行分析, 并且提出一个改进的图像融合方法, 该方法在不显著增加运算量的同时, 达到了更好的融合效果。

常用的图像融合方法有平均值法、最大值法、渐入渐出法[13]。所谓平均值法是指对待融合的两幅配准图像, 将其重叠区域的像素值进行平均, 非重叠区域的值保留原图像的像素值。所谓最大值法是指重叠区域不再是对两幅待拼接的图像进行平均, 而是通过比较待拼接图像对应点像素值大小进行融合。渐入渐出法是在图像重叠区域内使用线性加权过渡函数将两幅图像的灰度值进行加权得到融合后图像的灰度值, 以实现重叠区域边界的平滑过渡。

3. 改进 SURF 的视频拼接算法

3.1. SURF 算法原理

SURF 算法实现图像配准主要有三大工序, 1) 特征点检测; 2) 对特征点附加详细的信息(局部特征)也就是所谓的描述器; 3) 特征点匹配。

特征点检测包括三个步骤, 首先建立积分图像, 然后用箱式滤波器建立图像的尺度空间, 最后在建立的尺度空间上对特征点进行定位。

特征点附加详细的信息包括两个步骤, 首先求取特征点的主方向, 这样可以保证算法的旋转不变性, 然后将特征点的邻域旋转到主方向, 对特征点进行描述。

特征点匹配就是采用最简单的两向量内积最大值为最匹配的点, 设定一阈值, 只有当这个最大值大于该阈值方可认为两特征点匹配[14]。

积分图像的建立使得 SURF 算法和其他算法相比不仅具有较好的缩放、旋转、平移等特性, 而且计算速度很快。积分图像是对原始图像进行积分计算得到的图像。积分图像的每一点表示为原图像从原点到该点的矩形区域的像素和, 积分图像的建立之所以能够加快计算速度, 是因为我们对整幅图像进行积分图像遍历后, 原始图像中的任一矩形区域的像素之和就可以通过加减运算来完成, 而与矩形的面积无

关，矩形越大，节省的计算时间越多。

SURF 算法之所以能够采用积分图像来计算，另外一个很重要的近似就是采用箱式滤波器来近似高斯核函数[15]。箱式滤波器的引入使得卷积模板都是框状模板，使用积分图像来计算就大大减少了计算量，从而提高了算法的运算效率。SURF 算法采用箱式滤波器来近似代替高斯核函数，使得卷积模板均由简单的矩形构成。积分图像的引入解决了矩形区域快速计算的问题，箱式滤波器的近似极大提升了计算速度。

因此本文采用 SURF 算法进行特征提取，为了将其运用于实时视频拼接系统中，本文对其进行了改进。

3.2. 改进的 SURF 算法

原有的 SURF 算法是对整个图像区域提取特征点，然后在整个图像范围内进行特征匹配，由于在视频拼接系统中两个摄像头拍摄的画面必然会有重叠区域的存在，所以我们只需要在重叠部分进行特征点提取就可以，这样的好处是可以减少特征提取的时间，同时减少特征匹配的时间并同时可以减少误匹配。

一般视频拼接的图像都会有重合部分，本文利用特征匹配的方法，在摄像头输入视频的首个处理帧进行了全图匹配，会得到特征匹配对，此时可以利用特征匹配对得到重合区域，在之后的实时视频拼接中，就不再使用全图的特征检查，而是分别在两输入帧对应的重合区域进行特征检测，这样就大大减少了 SURF 算法的时间，正常情况下，两视频图像的重叠区域在 30%~50%，通过首帧确定重叠范围的方法可以计算出来重叠范围平均会占图像大小的 30%左右，利用重叠区域的特征点可以进行后续的特征匹配。通过这种方法能够减小 SURF 特征检测的运行时间，对于提升拼接的实时性有很大的帮助。具体过程如下：

1) 首先获取两个摄像头同一时间点采集到的第一帧图像，通过 SURF 算法对两幅图像进行全局域的特征检测，其过程为建立积分图像，然后用箱式滤波器建立图像的尺度空间，最后在建立的尺度空间上对特征点进行定位。

2) 对特征点进行特征描述，首先求取特征点的主方向，这样可以保证算法的旋转不变性，然后将特征点的邻域旋转到主方向，对特征点进行描述。

3) 对两幅图像中已检测到的特征点进行粗匹配，即先根据普拉斯响应正负号将特征点分为两组，然后建立 K-d 树寻找最近邻和次近邻，最后在对应的组中分别使用最近邻和次近邻比值匹配法进行特征点匹配。

4) 使用 RANSAC 方法对特征点进行提纯，去除误匹配点，最后剩下的就是准确匹配的特征点。

5) 最后，分别选取两幅图像中与中间距离最大的特征点，特征点的位置即可确定为重叠区域最边缘。

如图 3 所示，假设待拼接的两幅首帧图像检测到(a,a')、(b,b')等共 6 对匹配点，分别选取两幅图像中与中间距离最大的匹配点，即为(c,c')与(d,d')，然后分别过 c 点和 d 点分割两幅图像，即可确定重叠区域的大致范围。

分别用检测到的重叠区域的长度比图像的总长，算出百分比，取此范围作为后续帧的特征点检测范围。

6) 对于后面所有的连续帧，重复 1)~5)过程，不过对于 1)过程中的特征点提取范围限定在已经计算得到的范围内。

4. 仿真结果与分析

为了检测上文方法计算出的重叠区域的准确性，采集风景视频，然后提取首帧图像的首帧图像(图 4~6)。

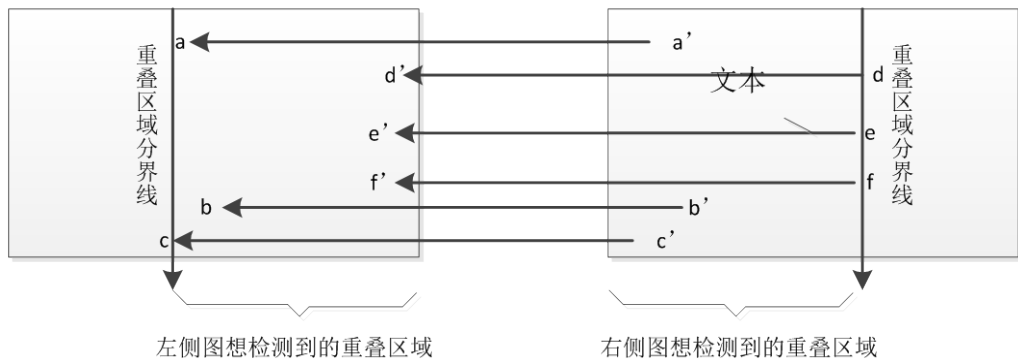


Figure 3. Feature extraction range
图 3. 特征提取范围



Figure 4. The first frame of landscape video
图 4. 风景视频首帧图像



Figure 5. The first frame of landscape video after matching
图 5. 匹配后的风景视频首帧图像

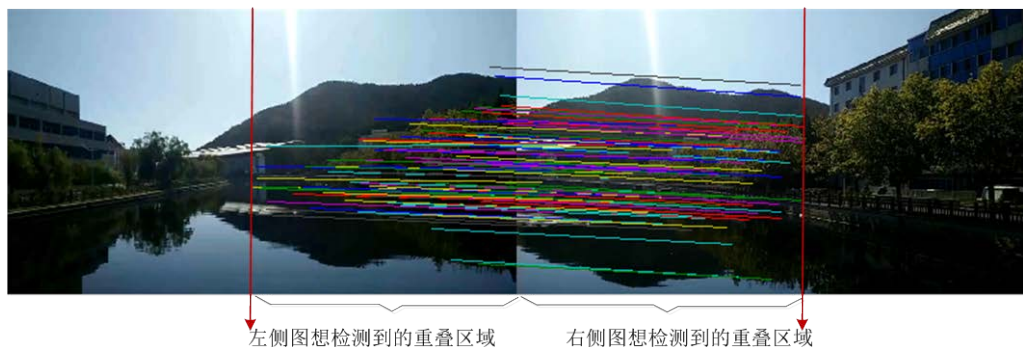


Figure 6. Overlapping area of detected
图 6. 检测到的重叠区域

通过观察图 4 与图 6 对比, 发现仿真计算出的重叠面积已经无限接近实际的重叠区域。

图 7~10 是使用原 SURF 算法和改进 SURF 算法就进行特征匹配的实验结果。

由图可以明显的看出改进 SURF 算法相比原 SURF 算法在特征粗匹配时明显的减少了误匹配。

表 1 是特征匹配的点数和时间的对比, 从中可以看出改进后的算法特征匹配时间有了很大的降低, 同时误匹配点数也降低了。

应用于实时视频拼接的改进 SURF 算法由于限定了特征提取的范围和降低了特征的维数, 使得特征提取与匹配过程的运算量取得线性的下降, 因为整个 SURF 算法的运算量大致上与其所提取特征的范围内的像素点个数成线性关系, 所以当提取范围缩小时, 能显著地线性提升算法运行速度, 提升算法在应用中的性能表现, 另外在匹配过程中, 由于维度的减小, 也线性地降低了在特征点匹配中的运算时间。

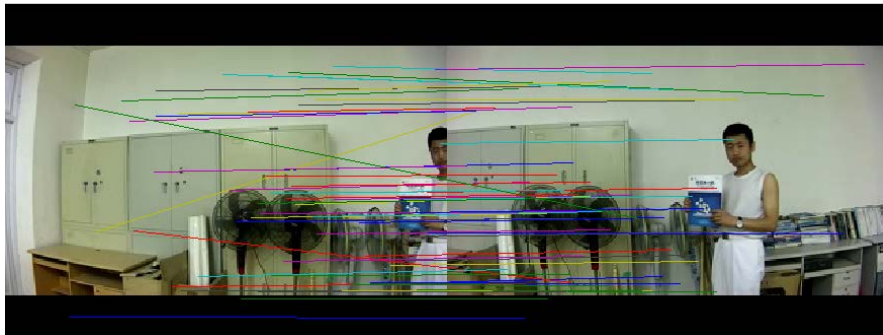


Figure 7. The original rough SURF feature points matching
图 7. 原始 SURF 特征点粗匹配



Figure 8. The original purification of SURF feature points matching
图 8. 原始 SURF 特征点提纯匹配

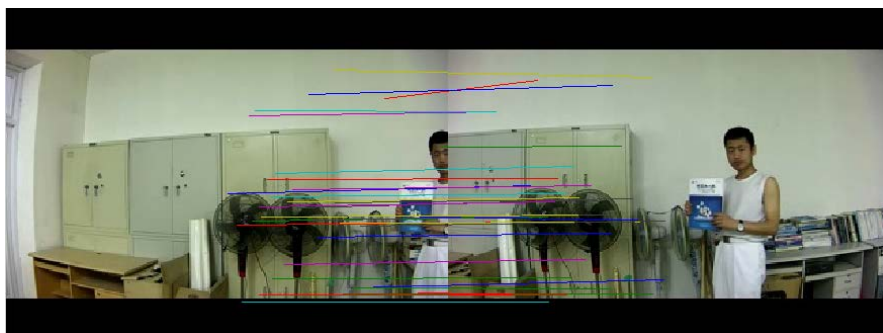


Figure 9. The improving of the rough SURF feature points matching
图 9. 改进 SURF 特征点粗匹配



Figure 10. The improving of the purification of SURF feature points matching
图 10. 改进 SURF 特征点提纯匹配

Table 1. Comparison of the feature matching results
表 1. 特征匹配结果对比

算法	粗匹配点数	提纯后点数	误匹配点数	匹配时间(ms)
SURF	59	24	35	189
改进 SURF	34	17	17	93

5. 结论

随着多媒体、嵌入式、无线通讯等技术的发展,视频拼接技术在智能监控、视频检索、虚拟现实、360 度全景等领域会得到更加广泛的使用。它能够在不降低分辨率的前提下扩大视野、减少冗余。本文对实时视频拼接技术的相关理论进行了研究,并给出了一些改进的方法。主要将 SURF 算法运用于实时视频拼接中,提出一种基于改进 SURF 的特征提取算法。仿真结果表明改进的 SURF 算法在保证提取足够数量特征点的前提下加快了特征提取速度,同时对图像发生旋转、尺度和光照变换具有很好的鲁棒性。

参考文献 (References)

- [1] Agarwala, A., Agrawala, M., Cohen, M.F., Salesin, D. and Szeliski, R. (2006) Photographing long scenes with multi-viewpoint panoramas. *ACM Transactions on Graphics*, **25**, 853-861.
- [2] 黄威 (2014) 基于视频图像拼接关键技术的研究. 长春理工大学, 长春.
- [3] Kopf, J., Lischinski, D., Deussen, O., Cohenor, D. and Cohen, M.F. (2009) Locally adapted projections to reduce panorama distortions. *Computer Graphics Forum*, **28**, 1083-1089.
- [4] Gracias, N.R.E., Mahoor, M.H., Negahdaripour, S. and Gleason, A.C.R. (2009) Fast image blending using watersheds and graphcuts. *Image and Vision Computing*, **27**, 597-607. <http://dx.doi.org/10.1016/j.imavis.2008.04.014>
- [5] 王振兴 (2014) 基于视频拼接的无人实况转播系统. 杭州电子科技大学, 杭州.
- [6] Zhi, Q. and Cooperstock, J.R. (2012) Toward dynamic image mosaic generation with robustness to parallax. *IEEE Transactions on Image Processing*, **21**, 366-378.
- [7] 沈峰 (2014) 视点稀疏且可旋转的多透视视频拼接. 安徽大学, 合肥.
- [8] Zelnik-Manor, L. and Perona, P. (2007) Automating Joiners. *NPAR*, 121-131.
- [9] Efros, A.A. and Freeman, W.T. (2001) Image quilting for texture synthesis and transfer. *SIGGRAPH*, 341-346.
- [10] Hsu, C.-T. and Beuker, R.A. (2000) Multiresolution feature-based image registration. *ProcSpie on Visual Communications & Image Processing*, **4067**, 1490-1498.
- [11] Lee, J.-H., Lee, K.S., Lee, W.H., et al. (2005) Fast panoramic video generation method using morphological corner detection. *Electronic Imaging 2005 International Society for Optics and Photonics*, 2005, 547-558.
- [12] 雷中锋, 王广生, 方穗明 (2004) 可控视野全方位电视传播技术. *现代电视技术*, **2**, 28-29.

-
- [13] 班晓娟, 宁淑荣, 张亚, 等 (2008) 一种改进的自适应阈值 SSDA 的视频拼接算法——在交通场景中的研究与应用. *系统仿真学报*, **22**, 6201-6204.
 - [14] He, B., Zhao, G. and Liu, Q. (2010) Panoramic video stitching in multi-camera surveillance system. *25th International Conference of IEEE Image and Vision Computing New Zealand (IVCNZ)*, Queenstown, 8-9 November 2010, 1-6.
 - [15] Xu, W. (2012) Panoramic video stitching. Dissertations & Theses, Gradworks.