

基于改进 SURF 的实时视频拼接方法

雷 飞,王文学,王雪丽,任 维

(北京工业大学 电子信息与控制工程学院,北京 100124)

摘 要:视频拼接在现实生活中具有十分重要的意义。文中实现的功能是实时地将多路视频进行拼接以形成一个大视野的视频。为了使系统具有良好的实时性,先用 DirectShow 分离视频流、提取图像帧,然后提出了一种改进的 SURF 算法进行图像配准,最后采用渐入渐出法进行图像融合。改进的算法在特征点提取时对范围作了限制,同时改进了原有特征描述符,这样可以大大加快提取速度。在特征匹配时采用基于最近邻和次近邻距离比值的方法进行粗匹配,然后采用 RANSAC 方法进行提纯和求取变换矩阵,最后在视频拼接中,不是对每一帧图像进行配准和融合,而是只对首帧图像进行图像配准,之后的每帧图像利用第一帧计算出的变换矩阵进行图像拼接,这样可以加快视频拼接的速度。实验结果表明,该方法可以实时地生成无缝清晰的宽场景视频。

关键词:视频拼接;DirectShow;SURF;图像配准;图像融合

中图分类号:TP301

文献标识码:A

文章编号:1673-629X(2015)03-0032-04

doi:10.3969/j.issn.1673-629X.2015.03.008

Real-time Video Stitching Method Based on Improved SURF

LEI Fei, WANG Wen-xue, WANG Xue-li, REN Wei

(College of Electronic Information & Control Engineering, Beijing University of Technology,
Beijing 100124, China)

Abstract: Video stitching has great significance in real life. In this paper, achieve the function that stitches the multi-channel videos to form a large field of view video in real-time. To make the system has good real-time feature, firstly isolate the video streaming by Direct-Show and use an improved SURF algorithm for image registration. The improved SURF algorithm restricts the scope in the step of feature point extraction and improves the feature descriptor, which can greatly speed up the extraction. Match feature points based on the nearest neighbor and second-closest neighbor distance ratio method, and then use RANSAC approach to purify and calculate the transformation matrix. Finally, only register the first frame of the video image and use the calculated transformation matrix to stitch the other frames, instead of registering and fusing each frame of image, which can greatly accelerate the speed of the video mosaic. The experimental results demonstrate that this method can generate the seamless and clear video with wide scene.

Key words: video stitching; DirectShow; SURF; image registration; image fusion

0 引 言

视频拼接被广泛应用在视频会议、视频监控、卫星遥感等多个领域^[1]。单个视频采集设备获取的视频视角相对较小,而视频拼接可以获得清晰的宽视角视频。视频拼接技术是指将多个设备采集到的视频通过拼接技术合成一个大视野的视频。

通常有以下两种视频拼接方法^[2]: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;第二种方法是通过多个视频采集设备同时采集视

频图像,然后对视频图像帧进行特征匹配与图像融合后完成拼接。文中采用的是第二种方法,为了方便起见,采用两路 USB 摄像头在 PC 机上完成实验。

视频图像是由每秒若干帧的静态图像组成的,所以视频拼接的本质是图像拼接,而图像配准又是图像拼接最核心的一步,通常使用的方法是基于图像特征的配准。David G. Lowe 提出的尺度不变特征算子 SIFT^[3]在图像拼接中被经常使用,但是 SIFT 算法的计算速度较慢,不适合应用于实时视频拼接。文献[4]在 SIFT 算法基础上,提出了一种改进的尺度不变特

收稿日期:2014-04-13

修回日期:2014-07-16

网络出版时间:2015-01-20

基金项目:北京市教育科技计划面上项目(KM201210005003)

作者简介:雷 飞(1972-),男,副教授,研究方向为图像处理和实时控制。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20150120.2159.019.html>

征—SURF(Speeded Up Robust Features)。文献[5]研究表明:SURF 算法提高了特征检测的速度,同时与 SIFT 算法性能基本相同。但是视频拼接与图像拼接又有着很大的不同,最大的不同是视频拼接需要有良好的实时性。SURF 虽然速度较快,但是为了能将其应用在实时视频拼接中仍需要改进,文中采用一种改进的 SURF 算法来进行图像配准。

文中采用的视频拼接流程是:使用微软的 DirectShow^[6]实时地采集视频图像帧,将它以 25 帧/s 的速率进行解帧,对首帧视频图像利用改进的 SURF 进行配准以确定变换矩阵,利用这个变换矩阵对之后的每一帧图像进行融合,形成一个大视角的图像,最后将这些图像以视频的形式输出。

1 获取视频图像帧

DirectShow 是微软公司提供的一套 Windows 平台上进行流媒体处理的开发包,它能够高质量地对多媒体数据进行捕获和回放。DirectShow 简化了视频的捕捉工作,采用 DirectShow 技术,可以方便地进行图像帧的提取,同时它还保证了摄像头的同步启动,不需要工作者再进行帧之间的对齐操作。

文中的视频拼接系统采用 DirectShow 来处理图像序列,特别是使用了 CCameraDS 类中的一些函数,例如:获取摄像头数目函数:Camerascount;启动摄像头函数:Opencamera;获取视频帧:QueryFrame;关闭摄像头:Closecamera 等。

2 改进的 SURF 图像配准算法

图像配准是为了使图像间相互重叠部分对齐,将待拼接图像转换到参考图像的坐标系^[7],构成完整的图像,它是整个图像拼接中的核心部分。图 1 是图像拼接的流程,由图可以看出图像配准包括三个方面,即特征提取、特征匹配和确定变换矩阵。文中采用改进的 SURF 图像配准算法。

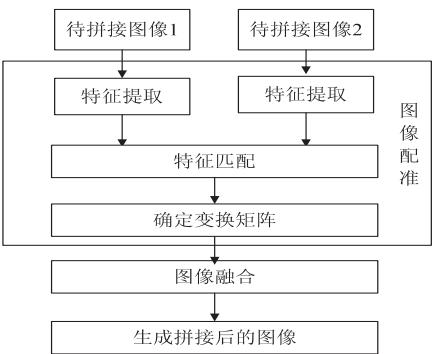


图 1 图像拼接流程

2.1 改进的 SURF 特征提取

特征提取包含特征检测和特征描述两个步骤。特

征检测包括构建尺度空间和确定极值点两个过程。

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,由于实际环境中两个摄像头拍摄的画面必然会有重叠区域的存在,所以只需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,进而减少匹配的时间并能减少误匹配。

一般要求图像的重叠区域在 30% ~ 50%,所以只需要在 30% 的范围内进行特征点提取即可。左边摄像头进行特征提取的区域将其限制在右侧 30% 的范围内,右边摄像头进行特征提取的区域将其限制在左侧 30% 的范围内。

1)特征检测。

(1)构建尺度空间。

由于 SURF 使用积分图像和盒子滤波器,所以不是像 SIFT 那样直接建立图像金字塔,而是采用不断增加盒子滤波模板尺寸的方法。通过不同尺寸盒子滤波器模板求取 Hessian 矩阵行列式的响应图像,然后在响应图像上采取非极大值抑制的方法,最终求取各种不同尺度的特征点。

文中将尺度空间划分为 4 组(Octaves)^[8],每组又由 4 层组成。每组代表了逐步增大的滤波模板对同一图像进行滤波的一连串响应图,选择 6 的倍数作为尺度间隔,下一层的尺度间隔是上一层的 2 倍。组和组之间有相互重复的值,原因是为了包含图像所有可能的尺度。将初始滤波模板设为 9×9 的滤波(其尺度近似于 σ 为 1.2 的高斯估计)。

(2)确定极值点。

在尺度空间金字塔构建好以后,就进行求取某一特定尺度下的局部极值。

在 SURF 特征检测算法中,Bay 等人采用了一个非常逼近 $\det(H)$ 的近似表达 $\Delta(H)$ ^[9],使用这样的逼近使得计算量再次减小,如式(1)所示:

$$\Delta(H) = D_{xx}D_{yy} - (0.9D_{xy})^2 \tag{1}$$

得到 $\Delta(H)$ 的值以后进行判断,如果 $\Delta(H)$ 的值为正,则该点为局部极大值点或者极小值点,虽然有了极值点,但仍然要在 3×3×3 的立体邻域进行筛选,把条件符合的极值点设定为候选极值点,并记录下尺寸和位置。

2)特征描述。

为了实现图像间的拼接,需要有一种相似性的度量方法,使用一种紧凑而完整的特征描述十分重要。

(1)确定特征点主方向。

确定特征点的主方向的目的是为了使检测出的特征点具有旋转不变性,这一步主要使用 Haar 小波来处理。它的步骤是^[10]:以特征点为圆心,以 6σ (σ 为样

本点所在空间的尺度值)为半径构成的圆形区域内,计算所有特征点在 x, y 方向的 Haar 小波响应向量(Haar 小波边长取 4σ),并同时为每个方向向量赋以不同的高斯权重,对越靠近圆心的特征点赋的权重值越大。然后以 60° 的扇形窗口围绕圆心旋转,这样可以得到 6 个扇形区域,计算每个扇形区域中各个特征点的 Haar 小波向量的矢量和,从而形成 6 个单独的方向矢量,最后选取这 6 个中幅值最大的矢量方向为主方向。

(2) 特征点描述符生成。

原有的特征描述符生成方法是^[11]:以特征点为中心,沿上一步得到的主方向,选取边长为 20σ 的正方形窗口,将这个窗口划分为 4×4 个子窗口,对于每个 $5\sigma \times 5\sigma$ 的子窗口区域,以采样间隔为 σ ,分别计算每个子窗口在水平和垂直方向上的小波响应,得到小波系数记为 d_x 和 d_y 。为了增加描述子的鲁棒性,对 d_x 和 d_y 进行加权计算,最后分别对每个子块的响应值进行统计,如式(2)所示,得到每个子块的矢量,共有 $4 \times 4 \times 4 = 64$ 个。

$$V = (\sum d_x, \sum |d_x|, \sum d_y, \sum |d_y|) \quad (2)$$

文中对其进行改进,选取边长为 $10s \times 10s$ 的正方形窗口,将窗口划分为 2×2 个子窗口,这样特征向量长度为 $2 \times 2 \times 4 = 16$,维数大大降低,可以加快运算速度。改进后的特征描述符如图 2 所示。

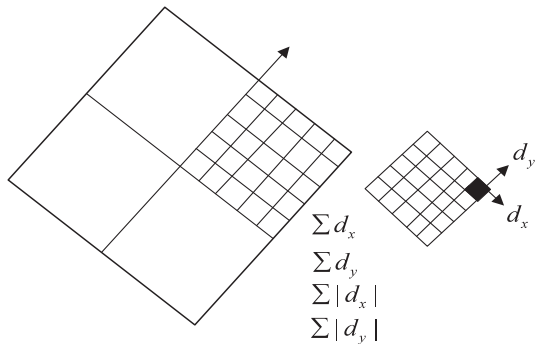


图 2 改进的特征描述算子

2.2 SURF 特征点匹配

为了加快匹配过程, SURF 在特征矢量中增加了一个新的变量,即特征点的拉普拉斯相应正负号,这两类特征点具有不同的属性,在匹配时只需对同类的特征点进行匹配,这样会大大地提高匹配速度和精度。

图像中两个特征点之间相似性的度量一般采用欧氏距离等距离函数进行度量,通过相似性的度量就可以得到待配准图像间的潜在匹配点对。这其实是一个距离检索问题,如果用穷举法则浪费太多时间。由于用标准的 K-d 树时数据集的维数不应该超过 $20^{[12]}$,而改进的 SURF 的维数为 16,所以采用 K-d 树进行特征匹配。

通常使用比值匹配法,即将其中的一幅图像作为样本,在样本的特征点中寻找在另外一幅图像中与它距离最近的特征点和次近特征点,然后计算这两个特征点与样本点之间欧氏距离的比值。对于比值小于某一阈值范围内的特征点,则认为是正确匹配的特征点。

$$\text{ratio} = \frac{\text{最近邻特征点距离}}{\text{次近邻特征点距离}} \leq \text{阈值} \quad (3)$$

2.3 确定变换矩阵

图像的变换矩阵是指两幅图像之间的坐标变换关系。图像 I_1 和图像 I_2 的对应关系可表示为 $I_2 = H \times I_1$,理论上用 4 对匹配点就可以求出 H ,但是为了使误差最小化,文中使用 RANSAC^[13]方法进行进一步的特征提纯并计算 H ,可以获取多数匹配点满足的最终解。

3 图像融合

如果对配准后的图像直接进行叠加拼接,拼接处会因为光照不同等因素存在明显的拼接缝。为消除拼接缝,需采用一定的方法来对图像进行平滑处理,常用的图像融合方法有直接平均法、加权平均法、多分辨率法等^[14]。在视频拼接中,因为要对视频的每一帧都进行融合,所以融合的方法要尽可能简单有效。直接平均法虽然简单但是效果不好,多分辨率法效果不错但是运算时间太长,因此文中采用渐入渐出^[15]的方法对图像进行融合。假设待拼接的两幅图像为 I_1 和 I_2 ,融合后的图像用 I 表示,融合公式为:

$$f(x, y) = \begin{cases} I_1(x, y) & (x, y) \in I_1 \\ d_1 I_1(x, y) + d_2 I_2(x, y) & (x, y) \in I_1 \cap I_2 \\ I_2(x, y) & (x, y) \in I_2 \end{cases} \quad (4)$$

其中, $d_1 + d_2 = 1$, $d_2 = \frac{w_i}{W}$, w_i 为当前像素点与两幅图像重合区域左边缘的横向距离, W 是两幅图像重合区域的总宽度。 d_1 从 1 均匀变化到 0, d_2 从 0 均匀变化到 1,实现了拼接图像的平滑过渡。

4 实验结果及分析

该系统实验测试平台为 VC++6.0, Intel Core2 T5870 CPU,内存为 2 GB。视频来源于普通的 USB 摄像头,图像大小为 640×480 ,速率为 25 帧/s。

4.1 改进的 SURF 特征提取实验

采用改进后的 SURF 算法对原始图像进行特征点提取,实验结果如图 3 所示。

实验结果表明,改进后的 SURF 算法虽然特征点数量减少,但是提取的数量已经足够用来拼接,并且非重叠部分的特征点对特征匹配也没有作用。



图3 改进 SURF 的特征提取结果

4.2 改进的 SURF 特征匹配实验

图4是利用改进的 SURF 方法进行特征匹配和 RANSAC 方法提纯的实验结果。



图4 改进 SURF 特征匹配和 RANSAC 提纯

表1对原 SURF 和改进 SURF 算法特征点匹配的结果进行了对比,从表中可以看出改进后的 SURF 算法特征匹配时间大幅度降低,同时减少了误匹配数据。

表1 特征点匹配结果对比

算法	粗匹配点数	误匹配点数	匹配时间/ms
原 SURF	212	36	1 505
改进 SURF	35	9	278

4.3 摄像头视频拼接实验

视频拼接过程中只是用首帧进行图像特征提取和匹配,之后的每帧要按照之前计算的 H 进行融合,图像融合的时间大约为 15 ms,可以满足实时性的要求。

图5是视频第80帧的拼接效果。

5 结束语

文中提出了一种基于改进 SURF 的实时视频拼接方法,改进的SURF主要在计算速度方面有了很大提



图5 第80帧图像拼接结果

高,同时又不会影响拼接效果。实验结果表明,该方法具有良好的实用性。但是文中只是在普通 PC 机上做了研究,为了应用于嵌入式系统,应该考虑将该算法移植到 DSP、FPGA 上实现。

参考文献:

[1] 王小强,陈临强,梁 旭. 实时全自动视频拼接方法[J]. 计算机工程,2011,37(5):291-292.

[2] 首照宇,欧阳宁,张华俊,等. 基于 SURF 和动态 ROI 的实时视频拼接[J]. 计算机工程与设计,2013,34(3):998-1003.

[3] Bay H,Ess A,Tuytelaars T,et al. Speeded-Up Robust Features (SURF) [J]. Computer Vision and Image Understanding,2008,110(3):346-359.

[4] Bauer J,Sunderhauf N,Prozel P. Comparing several implementations of two recently published feature detectors [C]//Proc of the international conference on intelligent and autonomous systems. Toulouse,France:[s. n.],2007:348-350.

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

[6] 杨云涛,冯 莹,曹 毓,等. 基于 SURF 的序列图像快速拼接方法[J]. 计算机技术与发展,2011,21(3):6-9.

[7] 李士强. 基于 SURF 算法的视频拼接技术研究[D]. 杭州:浙江工商大学,2011.

[8] El-Saban M,Izz M,Kaheel A. Depth adaptive video stitching [C]//Proc of 17th IEEE international conference on image processing. [s. l.]:[s. n.],2010:1193-1196.

[9] 徐 杰,金湘亮,白瑞广. 一种改进的摄像头视频实时拼接方法[J]. 计算机工程与应用,2013,49(24):179-181.

[10] 卜 珂. 基于 SURF 的图像配准与拼接技术研究[D]. 大连:大连理工大学,2009.

[11] 谢小竹. 基于云台控制的实时视频拼接[D]. 厦门:厦门大学,2009.

[12] 蔡丽欢,廖英豪,郭东辉. 图像拼接方法及其关键技术研究 [J]. 计算机技术与发展,2008,18(3):1-4.

[13] Hsu Chiou-Ting,Cheng Tzu-Hung,Beuker R A,et al. Feature-based video mosaic [C]//Proc of IEEE international conference on image processing. [s. l.]:IEEE,2000:887-890.

[14] 李云霞,曾 毅,钟瑞艳,等. 基于 SIFT 特征匹配的图像拼接算法[J]. 计算机技术与发展,2009,19(1):43-45.

[15] 季 诚,欧阳宁,张 彤. 基于视频片断的全景图拼接[J]. 计算机技术与发展,2009,19(9):239-241.

基于改进SURF的实时视频拼接方法

作者：[雷飞](#)，[王文学](#)，[王雪丽](#)，[任维](#)，[LEI Fei](#)，[WANG Wen-xue](#)，[WANG Xue-li](#)，[REN Wei](#)
作者单位：[北京工业大学 电子信息与控制工程学院, 北京, 100124](#)
刊名：[计算机技术与发展](#)[ISTIC](#)
英文刊名：[Computer Technology and Development](#)
年，卷(期)：2015(3)

引用本文格式：[雷飞](#). [王文学](#). [王雪丽](#). [任维](#). [LEI Fei](#). [WANG Wen-xue](#). [WANG Xue-li](#). [REN Wei](#) [基于改进SURF的实时视频拼接方法](#)[期刊论文]-[计算机技术与发展](#) 2015(3)