

基于视频拼接技术的无人实况转播系统^①

王振兴, 陈临强, 张 英

(杭州电子科技大学 计算机学院, 杭州 310018)

摘 要: 介绍了基于视频拼接技术的无人实况转播系统的实现. 首先对图像进行预处理, 然后使用 SURF 算法对视频图像进行图像配准拼接, 并对拼接图像进行融合. 提出了一种基于运动前景提取的全景视频中热点区域确定方法, 该方法可平滑追踪热点区域进行提取实现实时转播. 本系统的实现节省了视频直播过程中的人力财力.

关键词: 视频拼接; 图像配准; 视频融合; SURF 算法; 实况转播

Live Broadcasting System Based on Video Mosaic

WANG Zhen-Xing, CHEN Lin-Qiang, ZHANG Ying

(College of Computer Science, Hangzhou Dianzi University, Hangzhou 310018, China)

Abstract: This paper introduces the live broadcasting system based on video mosaic technology. First, it makes the image preprocessed. After the image pre-processing of video images this paper uses SURF algorithm for image registration stitching on the mosaic image fusion. This paper proposes a method for determining which based on foreground extraction. This method can track hot region smoothly and achieve real-time broadcasting. The system saves a lot of human and financial resources.

Key words: video mosaic; image registration; video image fusion; SURF algorithm; live broadcast

当前的电视技术正向着数字化以及多媒体方向飞速发展, 电视节目制作的智能化是目前国内外研究的热点问题^[1]. RT-SET 公司的 Larus 虚拟演播室系统、ORAD 公司的 BLUESET 系统^[2]在国内外电视台都得到了广泛的应用, 杭州电子科技大学图像所的虚拟融合系统^[3]实现了虚拟物体与视频的实时融合且应用于电视节目的实况转播. 在电视转播中, 往往需要大量的人员参与到实况现场中去, 而地方电视台及越来越多的单位、小区电视小台, 无需高质量和高标准转播, 若在现场配备专业人员, 会浪费大量的财力. 本系统主要是为了满足这样的需求, 只需要在现场架设几台摄像机, 就可以实现自动实况转播.

1 整体流程

本系统主要通过以下技术步骤实现: 视频采集、预处理; 然后通过 SURF 方法提取特征点进行视频拼接; 利用基于形态学的融合方法对图像重叠区域进行

融合处理, 进而生成全景视频; 对拼接后的全景视频利用设计的热点区跟踪方法, 按热点区域进行跟踪提取转播. 流程图如图 1 所示:

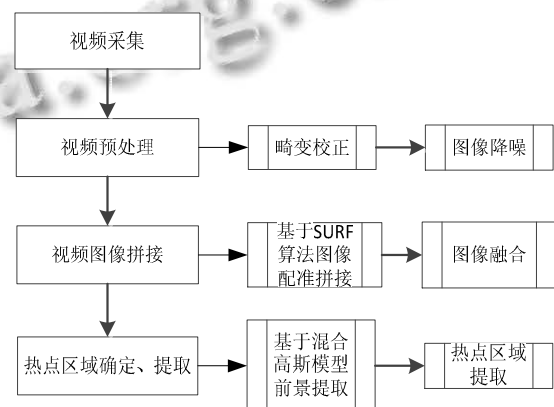


图 1 系统实现流程图

2 视频采集

系统采用多台同型号的摄像机, 架设的水平高度、

^① 收稿时间:2013-11-26;收到修改稿时间:2014-01-13

间距基本相同,且相邻两台机器所采集画面重叠的范围咱各自采集画面宽度的 1/5 以上.采用同步触发器同步启动多台摄像机,以保证视频流时间上的同步性.

3 视频的预处理

首先采用基于张正友定标原理^[4]的方法对视频进行畸变校正.

然后对视频帧进行降噪处理以降低噪声对后续的特征点提取、匹配和最终的拼接结果的影响.具体方法为:设二维图像分辨率为 $M \times N$,用 $\varphi(x,y)$ 作为亮度干扰表示噪声,噪声的强度用均值描述,分布特征用方差 σ_φ^2 描述.计算方法见公式(1)、(2)^[5].

$$\bar{E}_\varphi = E[\varphi(x,y)] = \frac{1}{M \times N} \sum_{x=1}^M \sum_{y=1}^N \varphi(x,y) \quad (1)$$

$$\sigma_\varphi^2 = E\{[\varphi(x,y) - \bar{E}_\varphi]^2\} = \frac{1}{M \times N} \sum_{x=1}^M \sum_{y=1}^N [\varphi(x,y) - \bar{E}_\varphi]^2 \quad (2)$$

4 视频拼接

4.1 基于 SURF 特征检测的视频图像拼接

图像配准是指对输入的一系列已经排序好的图像信息进行提取,在提取的信息中寻找最佳匹配,根据此信息求取相邻图像间的变换模型.视频图像配准是视频拼接中的关键步骤,图像拼接效果的优劣取决于图像的配准精度.

本文采用基于 SURF 特征的配准方法.通过提取视频图像的特征,然后根据特征来进行图像的配准和拼接,SURF 特征提取只提取待拼接图像中的关键信息.基于特征的图像配准方法有计算量小、灰度信息依赖性弱、噪声不敏感等优点.而且只在特征空间上进行配准,速度较快^[6].

拼接后的视频图像并不规则,对其进行裁剪处理得到标准的矩形拼接图像.

4.2 图像融合

图像拼接后在重叠区域往往会出现较为明显的接痕,通过融合后,图像具有较好的视觉效果.本论文使用一种基于形态学的融合方法,算法的具体步骤为:

1)对图像重叠区域内全设为 255,其余都设为 0,然后对此二值化图像进行扫描得到重叠区域边界,然后使用形态学中的腐蚀算法求得对图像的重叠区域某个范围 T 内的像素.

2)计算重叠区域的颜色差异,通过与阈值比较来得到运动物体的位置.

3)计算重叠区域的中心位置 p ,并把中心点与各边界连线,计算公式如下^[7]:

$$\begin{cases} p.x = R.width / 2 \\ p.y = R.height / 2 \end{cases} \quad (3)$$

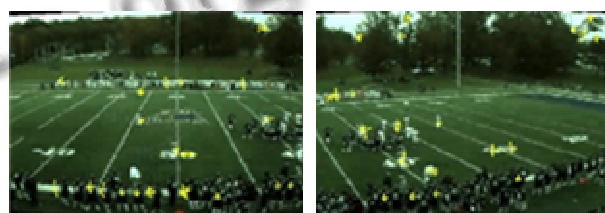
4)权值计算并进行融合,设 w 为权重.

$$\begin{cases} k = PI / T & PI = 3.1415... \\ w[i] = \frac{1 + \cos(i * k)}{2} & i = 0, 1, 2, \dots, T-1 \end{cases} \quad (4)$$

5)设 $f(I)$ 为最终融合图像, $f_1(I)$ 、 $f_2(I)$ 为拼接前图像,融合公式为:

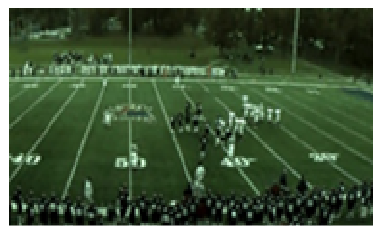
$$f(I) = w f_1(I) + (1-w) f_2(I) \quad (5)$$

视频拼接融合后的实验结果见图 2:



(a)特征点检测结果 1

(b)特征点检测结果 2



(c)拼接效果图

图 2 特征点检测及拼接结果图

5 热点区域确定

通过前面处理得到拼接后的视频,接下来确定热点区域并提取出来进行自动转播.体育比赛中关注的热点往往都在运动的目标最多区域,因此本系统中定义热点区域为:热点区域是视频中的一个子区域,在该子视频区域中运动前景目标是最多的;该子区域的高度与拼接视频相等,宽度按照指定的播出标准的宽高比来确定(本文中指定播出标准为 640*480,即宽高比为 4:3).

提取热点区域进行转播的技术步骤有:确定主区域、使用背景帧差法提取前景、去噪、确定热点区域位置、热点区域平滑追踪、热点区域提取实现转播.

5.1 主区域的提取

采用基于直方图统计的场地颜色自动提取方法来

确定主区域, 然后通过形态学的方法将提取的主区域场地补充完整.

5.2 背景帧差法提取前景

对上一步得到的主区域部分, 使用背景差分法提取前景. GRIMSON 等提出了混合高斯法^[8]建立背景模型可以很好地适应室内外场景的变化. 对每个像素点使用 K 个高斯分布来模拟像素点的背景值, 每一个像素点 x_t 的概率密度由这 K 个高斯分布的均值、协方差共同决定, 见公式(6):

$$P(X_t) = \sum_{i=1}^K \omega_{i,t} * \eta(X_t, \mu_{i,t}, \Sigma_{i,t}) \quad (6)$$

其中, K 的值一般取 3, 4, 5, $\omega_{i,t}$, $\mu_{i,t}$ 和 $\Sigma_{i,t}$ 分别为 t 时刻某个像素在对应的混合高斯模型中第 i 个高斯分布的权重, 均值和协方差矩阵; η 为高斯分布的概率密度函数:

$$\eta(X_t, \mu_{i,t}, \Sigma_{i,t}) = \frac{1}{(2\pi)^{n/2} |\Sigma_{i,t}|^{1/2}} * e^{-\frac{1}{2}(X_t - \mu_{i,t})^T \Sigma_{i,t}^{-1} (X_t - \mu_{i,t})} \quad (7)$$

t 时刻将当前图像的像素 X_t 与已排列好的混合高斯模型分布进行相似度匹配, 判别公式如下:

$$|X_t - \mu_{i,t}| \leq D * \sigma_{i,t-1} \quad (8)$$

(D 为参数, $\sigma_{i,t-1}$ 为标准差)

然后对混合高斯模型中的均值和协方差进行更新, 更新公式如下:

$$\begin{cases} \omega_{i,t} = (1 - \alpha)\omega_{i,t-1} + \alpha \\ u_{i,t} = (1 - \beta)u_{i,t-1} + \beta X_t \\ \sigma^2 = (1 - \beta)\sigma^2 + \beta(-u_{i,t-1})^T (X_t - u_{i,t-1}) \\ \beta = \alpha \eta(X_t | u_{i,t}, \sigma_{i,t}) \end{cases} \quad (9)$$

其中 α 是用户定义的学习率, 且 0 从 $a-1$, a 的大小决定着背景更新的速度, α 越大, 更新速度越快, 反之更新速度越慢; β 是参数学习率. 前景提取结果见图 3:



图 3 前景提取效果图

5.3 形态学方法降噪

对提取出的前景使用形态学降噪的方法, 即腐蚀和膨胀操作去除噪点. 效果如图 4 所示:



图 4 前景提取结果降噪后的效果图

5.4 确定热点区域位置

实况转播的视频中容易出现运动员相互遮挡、重叠的情况, 尤其场地较大而目标相对较小, 难以区分运动的个体. 通过实验发现多目标跟踪、统计目标个数的方法并不能达到热点区域确定和跟踪的目的. 而实况比赛视频中的前景主要为运动员, 其个体大小基本相同, 故本文采用计算前景面积的方法来代替统计运动员的个数. 即认为前景面积最大的窗口, 就是运动目标最多的窗口.

对于降噪后的拼接视频图像, 先检测前景目标的轮廓, 然后计算每个前景的面积.

接下来设置一个热点区域检测窗口, 检测窗口宽度与热点区域定义的宽度相同, 每次以步进为 S (本文 S 的值为检测窗口宽度的 $1/20$) 滑动检测窗口, 计算窗口中所有前景的面积之和, 得到一个面积和的序列 $SUM\{sum1, sum2, sum3 \dots sumN\}$, 滑动遍历整个拼接后视频帧, 找到具有最大面积的窗口 $MAX(SUM)$, 将该窗口区域视为热点区域以检测窗口的左边界作为其位置标示.

5.5 热点区域跟踪

如果直接按照上述热点区域确定方法得到的热点区域进行提取转播, 会出现抖动的现象. 由于运动员的运动是连续性的, 因此提出下面的热点区域平滑跟踪方法.

每隔 N 帧 (本文中 $N=30$) 选取一帧进行热点区域检测, 检测到的热点区域的位置设为 $newPosition$, 上一次检测到的热点区域的位置为 $oldPosition$, 两帧各自热点区域的位置的距离为:

$$posiontGap = |newPosition - oldPosition| \quad (10)$$

对于两个检测帧之间的各个视频帧, 其各自热点区域位置 $tmpPosition$ 的确定公式为:

$$tmpPosition[i] = oldPosition + i * posiontGap / N \quad (11)$$

其中 i 表示自上次被检测的帧开始计数的第 i 帧, $i \in [1, 29]$.

5.6 实验结果

实验环境: CPU: Intel Core i7 860; 内存: DDR3 4GB; 系统: Windows 7 32 位旗舰版.

摄像机的采集分辨率为 640*480, 相邻两帧的重叠区域百分比为 39.8%, 共采集拼接 5 路视频流, 对 5 路视频流拼接每一帧得到全景视频帧消耗的时间为 31ms, 对拼接后的每一帧全景视频检测热点区域位置平均消耗时间为 11ms, 实验效果见图 5.



(a)拼接后视频热点区域检测结果



(b)转播画面

图 5 热点区域检测及提取转播效果图

本系统实现后, 成功转播了学校运动会 110 米跨栏比赛以及足球比赛, 可以满足 640*480 分辨率下的实时播出. 现在正在寻求地方电视台合作, 以进行进一步的性能完善和功能的扩展.

参考文献

- 1 Kaneko H, Hamaguchi N, Inoue S. Engaging viewers through the connected studio: virtual participation in TV programs. *IEEE Consumer Electronics Magazine*, 2012: 30-39.
- 2 ORAD 公司.Datasheet. <http://www.orad.tv/zh-cn/blueset>. [2013-08-30].
- 3 杨建君.基于 GPU 的高清视频虚拟融合技术研究[硕士学位论文].杭州:杭州电子科技大学,2013.
- 4 Zhang ZY. A Flexible New Technique for Camera Calibration. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2000, 22(11): 1330-1334.
- 5 关新平,赵立兴,唐英干.图像去噪混合滤波方法.中国图象图形学报,2005,3(3):333-337.
- 6 卜珂.基于 SURF 的图像配准与拼接技术研究[硕士学位论文].大连:大连理工大学,2009.
- 7 王小强.大场景图像自动拼接算法研究[硕士学位论文].杭州:杭州电子科技大学,2010.
- 8 Grimson WEL, Stauffer C, Romano R, Lee L. Using adaptive tracking to classify and monitor activities in a site. *Proc. of the IEEE Workshop on Nonlinear Signal and Image Processing*. 1998. 22-29.