维普资讯 http://www.cqvip.com

视频图像序列运动参数估计与动态拼接*)

汤庆阳　陆佩忠
(复旦大学计算机科学与工程系　上海200433)

摘　要　本文采用多重分层叠代算法来估计全局运动参数，并提出应用于动态拼接的运动分割新方法，实现既有摄像机运动又有物体运动的视频图像序列自动拼接。我们的方法基本步骤如下：首先进行全局运动参数的初始估计，并且在分层叠代过程中进行区域分类，得到初始运动模板。接着空间分割原始图像，先根据图像的空间属性由底向上分层合并图像空间区域，再利用视频图像时间属性进一步向上合并，得到图像空间分割结果。然后结合初始运动模板和图像空间分割结果，采用区域分类新方法重新对图像空间分割结果的每个区域进行分类。然后根据分类结果逐步精确求解全局运动参数。最后进行图像合成，得到全景拼接图像。我们的方法利用了多重分层叠代的优点，并且充分考虑到视频图像空间和时间上的属性，实现了运动物体和覆盖背景的精确分割，避免了遮挡问题对全局运动参数估计精度的影响。而且在图像合成时我们解决了拼接图可能产生模糊或某些区域不连续等问题。实验结果表明我们的方法实现了动态视频图像序列高质量的全景拼接。

关键词　全局运动估计，运动分割，静态拼接，动态拼接，运动模板

The Estimation of Motion Parameters and Dynamic Mosaic Representation for Videos

TANG Qing-Yang　LU Pei-Zhong
(Dept. of Computer Science and Engineering, Fudan University, Shanghai 200433)

Abstract　In this paper, we use a double hierarchical algorithm for estimating motion parameters, and a new motion segmentation method for dynamic mosaicking is proposed for creating mosaics with moving objects, which considers a moving camera. Our method consists of five main steps. In the first step, we get a initial motion mask as the result of an initial classification in the process of global motion estimation. By the second step, a image spatial segmentation algorithm uses a hierarchical and bottom-up strategy, and region merging is performed again according to temporal information of video. In the third step, we present a new method for region classification, which combines the initial motion mask and the image spatial segmentation result. By the fourth step, the precision of global motion parameters is improved using the result of region classification. In the final step, the images are composited into a panoramic mosaic. Our method can get accurate motion segmentation and region classifications as a result of the advantage of the double hierarchical algorithm and fully using spatiotemporal information in the video. Our method can also avoid inaccuracies in occlusion areas, and addresses the problem of getting a blurred and discontinuous mosaic. Experimental results show that our method achieves the aim of the creation of high quality panoramic mosaics for videos in the presence of moving objects.

Keywords　Global motion estimation, Motion segmentation, Static mosaicking, Dyna-mic mosaicking, Motion mask

1　引言

视频序列图像拼接在计算机图形学、计算机视觉、多媒体等领域中有着越来越普遍的应用，例如在视频编辑中，可用来移去视频对象，或人为修改场景视点和镜头焦点[1]；在处理机器人躲避障碍物、对象识别等问题时，可用来简化图像分析[2]；在视频压缩及传输中，可用来避免对每一帧的背景进行编码或传送；在新的激光视网膜手术中，首先构造拼接图像，并把它作为空间基准图，用来跟踪视网膜上激光位置[3]。

视频拼接可分为两类：静态拼接和动态拼接。静态拼接的关键是要精确求解全局运动参数。本文主要讨论动态拼接方法，它除了要考虑由于摄像机运动造成的全局运动外，还要涉及场景中运动物体自身的局部运动，所以我们一方面要考虑到运动物体区域和覆盖背景区域对于全局运动参数估计精度的影响，另一方面也需要采用运动分割技术将运动物体分割开来。运动分割是MPEG-4标准、视频检索及视频监测中的一项关键技术，但在这些课题中，作者往往假定摄像机处于静止状态，或者主要关心运动物体的分割及跟踪[4]。Mech 和 Wollborn[6]提出了考虑运动摄像机的情况下自动运动分割的一种方法，他们假定视频边界附近区域不存在运动物体，然后将边界附近的小块区域作为初始背景区域进行全局运动估计和补偿，该方法缺乏一定的鲁棒性而且估计出来的全局运动参数精度比较低，而在运动分割时仅仅利用初始运动模板和位移矢量域进行区域分类，最后分割得到的运动物体边缘不太准确，另外他们采用块匹配的方法只将变化区域中的一部分显露背景分离出来。Nicolas[1]提出一种新的动态拼接方法，他在运动分割方面并没有做一些新的工作，他选择了一个半自动分割算法得到第一帧图像的物体对象，在以后各帧中进行自动跟踪，他主要讨论了全局运动参数估计方法和拼接图像合成策略，在估计全局运动参数时，先提出三个假设(摄像机不动、同一运动、恒定加速度)，再选择一种假设为参数设定初始值，然后采用全搜索图像匹配技术精确求解参数，最后采用下降梯度方法循环叠代进一步精确求解，他的方法对于摄像机运动平稳且单一(例如仅作平移或仅作缩放)时能精确估计全局运动参数，而在拼接图像合成策略上，先计算每一帧的混合系数值，然后将与目标帧坐标对齐的所有像素值乘以混

*)资助项目：国家自然科学基金(10171017)；国家自然科学基金重大研究计划(90204013)，教育部全国优秀博士学位论文作者专项基金，上海市科技发展基金(01JC14056)。汤庆阳　硕士研究生，主要研究方向为图像和视频处理；陆佩忠　教授，博士生导师，主要研究方向为图像和视频处理以及信息安全。

维普资讯 http://www.cqvip.com

合系数再合成得到拼接图，该方法优于选均值的合成方法，但仍然容易产生图像模糊。

我们首先采用多重分层叠代算法估计全局运动参数[5]，该算法基于光流并且与分层叠代相结合，通过优化目标函数求解运动参数并且进行局部和全局调整优化参数值，它可以有效控制累计计算误差，并且具有很强的鲁棒性，适用于摄像机在平移、摇动、旋转和缩放等情况下拍摄的视频。我们在分层叠代过程中进行初始分类，避免运动物体在视频中的位置对估计参数和求初始运动模板的影响。接着采用由底向上分层合并的空间分割方法[7]分割原始图像，并且利用视频图序列时间属性进一步向上合并，明显减少了空间分割区域数目，这样就得到图像空间分割结果。然后结合初始运动模板和空间分割结果采用新的方法重新分类图像空间分割结果区域，并且充分考虑到视频图像空间和时间上的属性，实现了运动物体和覆盖背景的精确分割。然后根据分类结果逐步精确求解全局运动参数，避免遮挡问题对全局运动参数估计精度的影响。而当前帧的区域分类模板又可以补偿到下一帧，用于下一帧的运动分割的预处理。最后进行图像合成，图像合成时我们解决了拼接图可能产生模糊或某些区域不连续等问题，得到清晰的全景拼接图。视频图像序列拼接策略框图如图1。

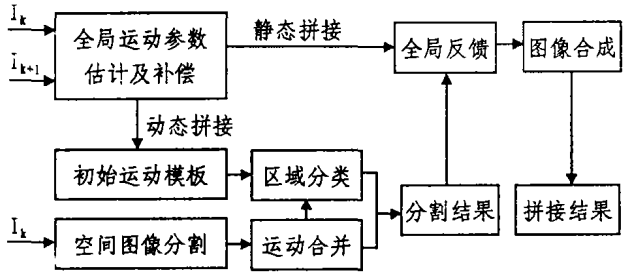


图1　视频图像序列拼接策略框图

2　全局运动估计和补偿

我们假定摄像机运动参数模型是8参数投射模型：

$$x'=\frac{a_{11}x+a_{21}y+b_1}{c_1x+c_2y+1},y'=\frac{a_{21}x+a_{22}y+b_2}{c_1x+c_2y+1} \quad (1)$$

8参数投射模型适合于场景是一个2D平面，摄像机平移、摇动、旋转、缩放等运动。

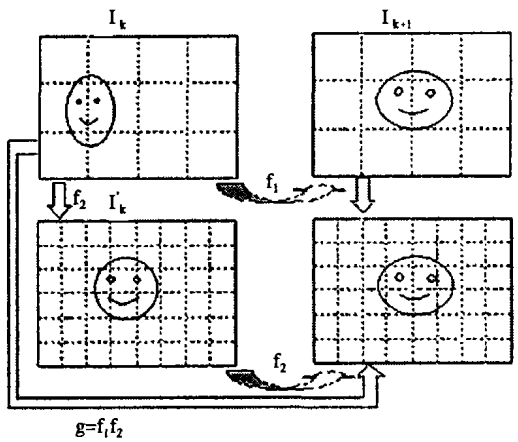


图2　多重分层叠代过程

我们采用多重分层叠代精算法[5]，该算法是基于光流的无特征点的方法，结合群论合成运算与分层叠代，通过优化目标函数来估计全局运动参数的。如图2，该算法简要过程描述如下：设前后两帧图像 I_k 和 I_{k+1}，首先将图像 I_k 和 I_{k+1}，分成 $m\times m$ 的方块作为第一层，每个方块看作一个像素。在第一层中先估计 I_k 帧和 I_{k+1} 帧之间的坐标变换 f_1；再由 I_k 经坐标变换 f_1 生成虚拟图像 I'_k，这样进行到下一层，第二层方块大小递减为 $m/2\times m/2$，在这一层上再估计出 I'_k 与 I_{k+1} 之间的坐标变换 f_2，然后将坐标变换 f_1 和 f_2 合成 $g=f_1f_2$。接着由 I_k 经坐标变换 g 生成虚拟图像 I'_k，重复上述操作直到最后一层，最后一层的块大小为1×1。这样，在由粗到细分层叠代过程中结合群论合成运算，得到 I_k 和 I_{k+1} 之间的变换参数。

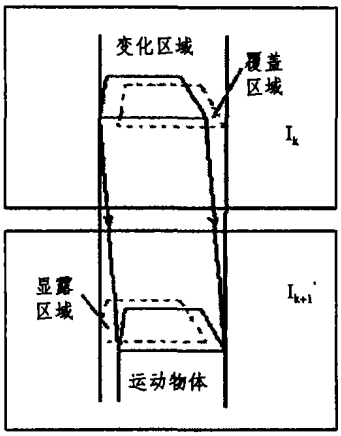


图3　区域类别举例

3　初始化运动模板

对于既有摄像机运动又有物体运动的视频，开始第一帧时由于我们不知道图像中哪部分是背景，哪部分是前景，因此我们可以先将图像中某一区域初始化为背景，例如左半边或边界邻近区域。

我们在初始化的背景区域中采用分层叠代的算法求运动变换参数，在分层叠代过程中，可以一边求运动变换参数，一边进行背景和前景的分类。我们将求初始运动模板的算法结合在分层叠代过程中，而且无论运动物体在视频区域中心或邻近边界的地方，都可以比较准确估计全局运动参数。假设分层叠代过程进行到某一层时，每一块的大小为 $m\times m$，设已求得的合成坐标变换为 g，我们将图像 I_k 经坐标变换 g 得到的虚拟图像 I'_k，这样就可以根据图像 I'_k 与 I_{k+1} 的差分图像，采用变化检测(Change Detection)方法，重新划分背景和前景，那么，当分层叠代到下一层时，只需在新划分的背景中来计算坐标变换。

设差分图像 $d(x,y)=I'_k(x,y)-I_{k+1}(x,y)$，我们将每个方块看作一个像素，先对差分图像进行量化，量化范围在 $-T$ 到 T 之间，$h(d)$ 是差分图像的直方图分布(各差分级的概率)。假定背景区域差分图像平均值 $\mu=0$，方差 δ，则划分为背景的条件概率密度函数条件分布为：

$$p(d|B)=\frac{1}{\sqrt{2\pi}\delta}e^{-\frac{1}{2}(\frac{d}{\delta})^2} \quad (2)$$

方差估计值为：

$$\delta^2=\sum_{d=-T}^{T}d^2h(d)/\sum_{d=-T}^{T}h(d) \quad (3)$$

初始运动模板定义为

$$CDM(x,y)=\begin{cases}1 & (if(d(x,y)\geqslant[2\delta]),foreground)\\ 0 & (if(d(x,y)\leqslant[2\delta]),background)\end{cases} \quad (4)$$

其中 $[2\delta]$ 为与方差估计值 δ 的两倍最接近的整数。

4　图像空间分割及运动合并

视频运动分割往往需要充分利用视频图像在空间和时间

维普资讯 http://www.cqvip.com

上的信息。图像空间分割可以得到每个区域的精确边缘，但分割得到的区域太多而且过于分散，而只利用时间属性进行运动分割最后得不到运动物体的精确边缘。

Wei Yu[7]用十二个容易计算的特征值来描述图像的颜色和纹理特征，采用由底向上分层合并的算法进行图像空间分割，该算法保持分割区域的空间相邻关系，且有较低的计算复杂度，我们先将图像分成2×2像素块，每一像素块作为单独的初始区域，再计算相邻区域的特征值距离，如果距离足够小就合并成一个区域。像这样再向上分层空间合并。

下面再利用视频图像时间上的运动信息进一步向上合并，因为若两个区域的运动情况相同，则它们要么同属于背景，要么同属于前景。假设 I'_{k+1} 为 I_{k+1} 帧补偿到 I_k 帧的图像，给定图像 I_k 和 I_{k+1}，我们可以在图像空间合并的基础上计算各空间合并区域 R 的平移运动参数，设 $u(x,y,t)=a$，$v(x,y,t)=d$，解如下的线性方程组即可求得平移参数 a 和 d：

$$\begin{bmatrix} \sum_{(x,y)\in R} I_x^2 & \sum_{(x,y)\in R} I_xI_y \\ \sum_{(x,y)\in R} I_xI_y & \sum_{(x,y)\in R} I_y^2 \end{bmatrix} \begin{bmatrix} a \\ d \end{bmatrix} = - \begin{bmatrix} \sum_{(x,y)\in R} I_xI_t \\ \sum_{(x,y)\in R} I_yI_t \end{bmatrix} \tag{5}$$

其中 $I_x=\frac{\partial I_k}{\partial x}$，$I_y=\frac{\partial I_k}{\partial y}$，$I_t=\frac{\partial I_k}{\partial t}=I'_{k+1}-I_k$，和号 Σ 的取值范围是区域 R。假设 R_1，R_2 是两相邻空间合并区域，平移运动参数分别为 a_1，d_1，a_2，d_2，则它们之间的距离为：

$$d_{1,2}=\sqrt{(a_1-a_2)^2+(d_1-d_2)^2}$$

我们设定合并标准，假如 $d_{1,2}<T_1$（T_1 为一门限），表明它们同属于背景或同属于前景，我们就将 R_1，R_2 合并成为一个区域。最后得到图像空间分割结果。

5 区域分类

如图3，我们给出了一个区域类别的例子，假设 I'_{k+1} 为 I_{k+1} 帧补偿到 I_k 帧的图像，需要说明的是图像 I_k 的变化区域包括运动物体及覆盖区域，图像 I'_{k+1} 的变化区域包括运动物体及显露区域。我们可以注意到覆盖区域的一些特征：(*i*) I_k 中的覆盖区域在 I_{k+1} 帧移出变化区域；(*ii*) I_k 中的变化区域的某一部分若运动后仍在同一个图像空间分割结果区域内，则一定不是覆盖区域；(*iii*) I_k 中的覆盖区域在变化区域的前方（运动物体运动方向）。我们采用如下的步骤进行区域分类：(1)首先去掉初始运动模板中面积较小的变化区域，减少噪声的影响；(2)设 R 是初始运动模板中一个变化区域，我们在 I_k 帧中计算区域 R 的特征值（与空间图像分割时计算特征值一样），采用区域三步搜索匹配法在 I'_{k+1} 补偿帧中寻找与区域 R 形状相同且特征值距离最小区域 R'，设最小特征值距离为 $dist(R,R')$，R 和 R' 区域坐标位移为 a 和 d；(3)对初始模板中的每个变化区域，根据以下规则进行区域分类，得到新运动模板：

```
if dist(R,R')>T2   则划分为前景
elseif √(a²+d²)<T3，则划分为背景区域
else   划分为前景区域
end if
if（前景区域）
    将符合覆盖区域的特征的部分区域，划分为覆盖区域
end if
end if
```

(4)接着对每个图像空间分割结果区域 R，计算其中属于新运动模板中的前景区域 R_m 所占的比例，假如 $(Area(R))/(Area(R_m))>T_4$（我们取0.4），就将整个空间分割区域 R 划分为前景，否则划分为背景，其中新运动模板中的覆盖背景部分保持不变，这样做的目的是利用空间分割区域使得到的分类结果具有精确的边界，避免了将一个空间分割区域的部分划分为背景而部分划分为前景；(5)考虑到在上一步骤中由于没有充分纹理或噪声等原因，可能将运动物体的部分区域错误地划为背景，设在 R_m 是一个划分为前景的区域，R 是与 R_m 相邻且划分为背景的区域，并且它的区域面积比较小，对这种区域采用第(2)(3)步的方法再作判断。

已经得到区域分类之后，对于图像 I_k 和 I_{k+1}，将原先求得的全局运动参数作为初始值，在除了前景和覆盖区域以外的区域内精确估计全局运动参数，这样很好地消除了遮盖问题对求运动参数的影响，提高了参数估计精度。接着就可以对上一帧的分类模板进行全局运动补偿，将其中的除了前景和覆盖背景区域以外的区域补偿到下一帧后作为下一帧的初始背景区域。

6 图像拼接

在多帧图像拼接时，我们用群合成的方法将源帧与目的帧通过全局运动参数联系起来。全局反馈算法能很好地控制将多帧图像拼接到同一目的图像过程中去时所产生的累计误差。在各帧之间的坐标与目标参考帧的全局坐标对齐之后，问题是选取哪个帧上的对应的像素作为拼接图像的像素。图像合成的方法是多种多样的，其中选先者的方法是选取与目标帧参考帧对齐且最近的像素作为拼接图像的像素，而且一旦选取像素值就不再改变，这种方法可以保证拼接图像的每个像素仅选自其中一帧，但可能受噪声影响，视频的边框上有黑点噪声，如果两帧间的运动比较小的话，则很可能选的全是一些边缘，造成拼接图像有明显的接缝。而混合合成方法是取与目标参考帧坐标对齐的所有像素的平均值或者所有像素分别乘以混合系数后的和值作为拼接图像中对应像素的像素值，这种合成方法会容易使图像产生模糊。我们的方法保证拼接图像每一像素仅仅选自其中一帧，可以得到清晰的拼接图像，而且尽量减少噪声的影响，使拼接图像无明显的接缝。

图4是我们的合成方法的示例图，设 M_{t-1} 是由补偿帧 I'_0，…，I'_{t-1} 合成得到的图像，当合成补偿帧 I'_t 时，我们的目标是寻找一条分界线，将重叠区域分成两部分。我们首先假设 S_0 为重叠区域边界上的某一点，将 S_0 设为分界线上的一点，假设 S_0 的坐标 (x_1,y_1)，将 S_0 作为起始点，再在坐标为 (x_1-1,y_1-1)，(x_1-1,y_1)，(x_1-1,y_1+1) 的三点中寻找 M_{t-1} 和 I'_t 灰度差最小的一点，将这一点设为分界线上的一点，并且作为新的起始点向左继续寻找分界线上的点。当到达横坐标为 x_2 的点时，设 S_1 的坐标为 (x_2,y_2)，在坐标为 (x_2-1,y_2+1)，(x_2,y_2+1)，(x_2+1,y_2+1) 的三点中寻找 M_{t-1} 和 I'_t 灰度最小的一点并设为分界线上的一点，继续向下寻找。我们将分界线上的点限制在区域 R 内，最后求得一条分界线。然后在合成图像时，拼接图像坐标在分界线左边、左上边、左下边的区域的像素取 I'_t 中的像素，其它区域的像素取 M_{t-1} 的像素。

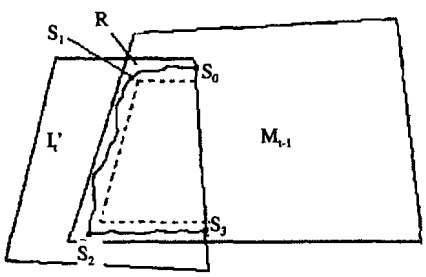


图4 图像合成方法示例

维普资讯 http://www.cqvip.com

7 实验结果

我们在 Pentium III 800MHz 的 PC 机上用 Borland C++ builder 5.0实现了我们的算法。在我们的程序中，可以直接读取视频中的 I 帧，或者先保存为 jpeg 或 bmp 图像序列后，再进行操作。测试视频是作者用手持式普通摄像机拍摄的。为了测试多重分层叠代估计运动参数的有效性和鲁棒性，我们采用一段作者在运动的火车上摇动、旋转、缩放镜头拍摄的阿尔卑斯山风景视频(如图5)，进行静态拼接试验。结果表明我们的全局运动参数估计方法适用于摄像机平移、摇动、旋转、缩放等时候拍摄得到的视频，参数估计累计误差得到了有效控制，实现了大范围的场景的拼接。

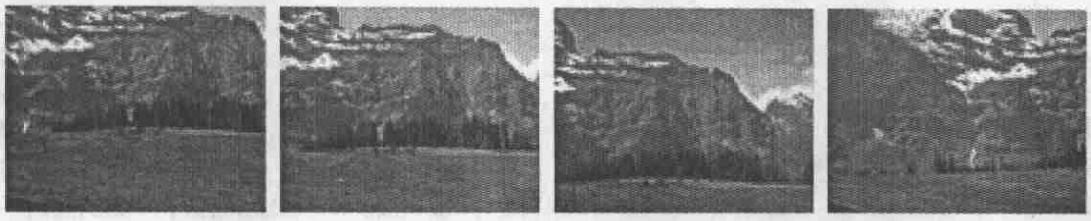


图5 阿尔卑斯山风景视频，平移、摇动、旋转和缩放摄像机时的视频片断，从第1000帧到1400帧(共401帧)选取的其中4帧，大小320×240。

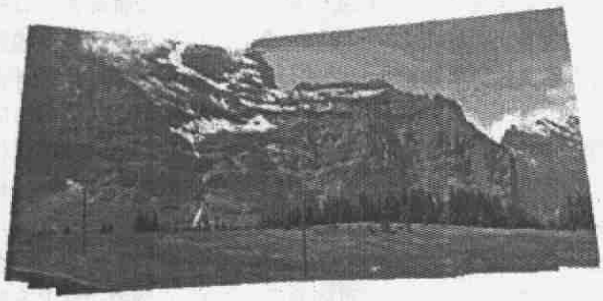


图6 阿尔卑斯山风景视频静态拼接结果，大小768×358，参考帧为第1200帧。

图7是作者在校园里通过平移、摇动和旋转摄像机拍摄的一段校园视频。运动分割过程如图(8)所示。如图8(b)，初始运动模板的变化区域包含覆盖区域、大部分运动对象以及一些噪声；图像空间合并结果如图8(c)(我们只进行了三层合并)，底层块大小为2×2，再经过运动合并，空间区域数目明显减少，如图8(d)。图8(e)是分类得到的覆盖背景区域，分类算法得到最后的具有精确边缘的运动对象，如图8(f)所示。图9显示运动对象在靠近边界的位置的情况下进行运动分割，可以看出运动对象的位置对求初始运动模板和最后的分类结果的不会产生影响。校园视频动态拼接结果如图9和图10所示。

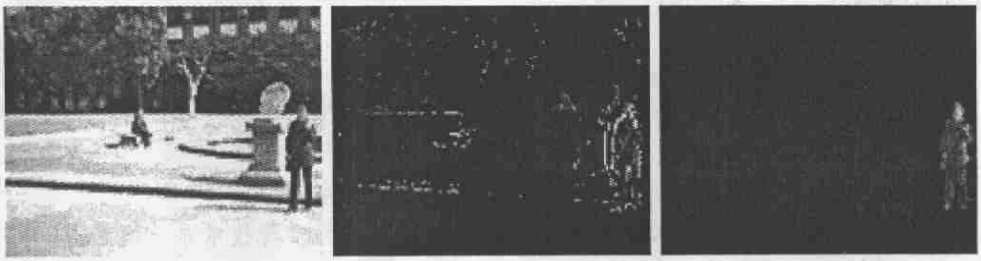
图7 校园视频，从第1030到1370帧(共341帧)中选取其中的4帧，大小320×240，在该段视频中，有一学生在镜头前走动，树从中还有一人骑车经过，另有一人坐在石凳上仅有细微的运动。

(a) 原始帧　(b) 初始运动模板　(c) 空间分层合并

(d) 运动合并　(e) 覆盖背景　(f) 分类得到的运动对象

图8 第1135帧运动分割过程

维普资讯 http://www.cqvip.com



(a) 原始帧　　(b) 初始运动模板　　(c) 分类得到的运动对象

图9　第1365帧分割过程，运动对象在靠近视频边界的位置

图10　校园视频动态拼接结果，仅仅背景部分，参考帧第1155帧，大小636×274

图11　校园视频动态拼接结果，放置上参考帧的运动对象

结论　本文采用多重分层叠代算法，有效地估计全局运动参数，并得到初始运动模板，并且利用了图像的颜色和纹理特征和视频运动信息，采用结合空间合并和时间运动合并的有效方法以及鲁棒的且可以得到精确分割边缘的区域分类方法，一方面剔除运动目标附近的背景边缘，又避免摭挡问题对估计全局运动参数精度的影响，为拼接图的精确性得到了有力的保证。而且在图像合成时我们解决了拼接图可能产生模糊或某些区域不连续等问题。最后实现了动态视频图像序列高质量的全景图像拼接。

参考文献

1　Nicolas H. New Methods for Dynamic Mosaicking. IEEE Trans. Image Processing, 2001, 10(8): 1239～1251

2　Candocia F M. Synthesizing a Panoramic Scene with a Common Exposure via the Simultaneous Registration of Images. In: Proc. of the 15th Florida Conf. on Recent Advances in Robotics (FCRAR 2002), Miami, FL., May 2002

3　Can A, Stewart C V, Roysam B. Robust Hierarchical Algorithm for Constructing a Mosaic from Images of the Curved Human Retina. In: Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, Fort Collins, Colorado, June 1999. 286～292

4　Zhao T, Nevatia R. Stochastic Human Segmentation from a Static Camera. In: Proc of IEEE Workshop on Motion and Video Computing (WMVC'02), Orlando, Florida, 2002. 9～14

5　Lu Peizhong, Wu Lide. Double Hierarchical Algorithm for Video Mosaics. Lecture Notes in Computer Sciences 2195, Springer, Oct. 2001. 206～213

6　Mech R, Wollborn M. A Noise Robust Method for 2D Shape Estimation of Moving Objects in Video Sequences Considering a Moving Camera. Signal Processing, 1998, 66(2): 203～217

7　Yu W, Fritts J, Sun F. A Hierarchical Image Segmentation Algorithm. In: IEEE Intl. Conf. on Multimedia and Expo, Lausanne, Switzerland, Aug. 2002

(上接第158页)

结束语　本文设计实现了传统基于批量模式的 K-Means 算法和基于序列模式的 K-Means 算法，并通过实验对两种算法的结果和效率进行了详细的比较分析，实验结果表明序列模式在效率上要优于传统 K-Means 采用的批量模式。

序列模式的不足之处在于它对记录输入顺序的依赖性很强，虽然在每一趟扫描开始之前打乱数据的输入顺序可以部分解决这个问题，但是随之而来的另外一个问题是当数据量很大，需要存储在外部存储器上的时候，这种做法的有效性受到了很大的限制。

参考文献

1　Han Jiawei, Micheline Kamber 著，范明等译. 数据挖掘：概念与技术. 机械工业出版社，2001

2　Haykin S. Neural Networks: A Comprehensive Foundation, 2nd Ed. 1999, Prentice-Hall: Upper Sadle River, New Jersey

3　An Introduction to Cluster Analysis for Data Mining, 2000. http://www.cs.umn.edu/~han/dmclass/

4　李飞，薛彬，黄亚楼. 初始中心优化的 K-Means 聚类算法. 计算机科学，2002，29(7)：94～96

5　Baraldi A, Blonda P. A Survey of Fuzzy Clustering Algorithms for Pattern Recognition. Parts I and II. IEEE Trans. on Systems, Man and Cybernetics, 1999, 29: 778～785, 786～801

6　Krishnapuram R, Keller J M. A Possibilistic Approach to Clustering. IEEE Transactions on Fuzzy Systems, 1993, 1: 98～110