

一种改进的基于边缘信息的运动分割方法

An Improved Motion Segmentation Method based on Edg Information

杜敏松 苏志勋

(大连理工大学应用数学系 大连116024)

Abstract Motion segmentation is the necessary stage of most video analysis implements, and it divides frames into different regions according different motion so that those motion objects can be abstracted. Although there are many scholars working in this field nowadays, it is a challenge job to bring out an accuracy and efficient motion segmentation method. This paper presents an improved theoretical framework for motion segmentation based on the motion of tracked region edges, to dealing video sequences which include two motions. The practical result has proved the validity of the improved motion tracking and region labeling method.

Keywords Motion segmentation, Expectation-Maximization algorithm, Motion tracking, Region growing

1. 前言

MPEG-4图像压缩标准的制定使得视频分割技术显得尤为重要。视频分割是指视频序列按一定的标准分割成区域,从而从视频序列里分离出有意义的实体。在MPEG-4标准的图像编码中,为了实现基于图像内容的交互功能,将视频序列的各帧图像按照不同的视频对象面(VOP)来表示,然后对视频对象分别进行编码,这样一方面提高了编码效率,一方面实现了用户与图像场景内容的交互性,所以在编码的过程中正确地划分视频对象(video object)和层次(layer)是必不可少的环节。这也正是视频分割的目的所在,实际上只要运用适当的分割技术就可以提取出符合实际的图像对象。

目前视频分割的技术有很多,基于像素的光流场法^[6,7]是估计帧与帧之间的光流场,然后将具有相似运动的像素分组。为了得到可靠的光流场需对图像进行平滑,但这就打破了帧与帧之间的运动边界,割裂了图像的空间连续性,从而不能精确地提取图像对象。马可夫随机场法(MRF)^[3]可以加强图像的空间连续性,但它仍然只考虑像素的分组,也会导致不确定的运动边缘。其他的方法还有基于模板的块匹配法,将提取到的运动对象作为模板再对后继的视频图像做一定的度量匹配,来跟踪并提取后继帧中的运动对象。此外基于区域变化的时空法^[9]也是一种常用的方法,这种方法通过检测序列图像中的变化区域和不变区域,结合时空图像亮度和独立运动区域来对运动物体和静止背景进行分割的。基于边缘特征的分割方法已经受到人们的重视^[4,5],本文采用的基于边缘的运动分割方法是利用视频序列中的两帧图像,对当前帧的边缘进行运动跟踪,将图像分为前后景两个运动层,再利用区域标记找到合理的分割以提取图像对象。相比而言,运动分割法有较

好的鲁棒性和稳定性,同时因为计算量较小,所以还有着很高的效率。

2. 理论框架

2.1 图像边缘的运动

边缘是运动分割的基础,在图像中能够提供非常好的运动信息,所以合理地利用边缘可以得到高效、鲁棒性很好的分割方法。对于标记所有区域的运动层次来说,只知道边缘的运动标记就已经足够了。

在图像序列里所有的运动都是分层的,也就是说一个运动是完全发生在另一个运动的前面(或后面)的。本文所研究的图像序列中只包含两个运动,这两个运动分别被称为前景运动和背景运动,而且图像中所有的区域必然属于这两者中的一个。假设在图像序列中任何的截断边界(即前景的边缘)都是可见的,从而每段边缘都遵循相同的运动。以下一些定理或推论可以用来当作本文的理论依据:

1) 如果一条边缘遵循一种运动,那么这条边缘所界定的两个区域中至少有一个也遵循这种运动。

2) 两个图像对象间的截断边界一定遵循前景运动。

3) 一个区域是前景区域当且仅当其所有的边界都是前景的。

4) 边缘的交界处不可能只有一条前景边缘。两个不同的运动层相遇时的边缘交界处,一定有两条边缘属于前景运动。

2.2 运动跟踪和区域标记

确定边缘的运动和其所属运动概率是两个相互依赖的迭代问题,其中一个问题的解决是以另一个的结果为前提的,本文选用最大似然法(EM)来解决这一问题。对当前帧的所有边缘做运动跟踪后就能得到它们分属两个运动的概率,由此确定两个不同的运动层,但此时尚不知道哪一个是前景层。为了

从边缘获得各区域的标记(即该区域是属于前景层还是后景层),我们必须首先知道运动层次。虽然从理论上说可以通过考虑两个运动层相遇时的边缘交界来决定运动层次,但在实际应用中由于噪声和边缘的错误标记的影响,其结果不太令人满意。所以另外一个途径就是同时考虑所有可能的层次和区域标记。由上述定理3)可知对于一个区域,若包围它的所有边缘都是前景边缘,则它一定是前景的,否则就是后景的。

2.3 本文结构

本文运用的是一种改进的运动分割方法,利用边缘信息来处理包含前后景两个运动的视频序列。首先,用 Canny 算法得到连续的图像边缘,由边缘的法线方向追踪出法向运动。然后对两个运动分别运用最大似然法(Expectation-Maximization algorithm, EM),交替估计出运动和各边缘所属运动的概率,反复迭代直至收敛;再以这些边缘为严格界线,用区域增长法分割出各个区域;最后由区域标记



图1 用 Canny 算子提取的当前帧的边缘图像

3.2 边缘的运动标记

首先对当前帧中的所有边缘进行运动标记(edge labeling),以确定每条边缘分别属于两个运动的概率(边缘分划)。这样会导致一个循环问题:一方面如果所有边缘所属的运动标记已知,那么两个运动就可以根据这个运动标记进一步调整,以便更加适合这两组边缘;另一方面如果两个运动是已知的,那么所有的边缘就可以根据各自最适合的运动来进一步修正运动概率。要解决这一问题,最大似然法(EM)是很好的选择。在初始化时先确定初始运动和先确定边缘的初始分划都是可以的,本文选择了前者。第一步计算运动,第二步利用得出的运动更新边缘的分划。反复迭代这两个步骤直至收敛。两个运动都要做上述的最大似然法迭代,这样可以得到两种不同的运动分割,最后由式(3.3)决定正确的分割。

3.2.1 运动的初始化

初始化运动的一般方法是假设一个运动是前景运动,可取所有边缘运动的平均值;另一个运动是后

过程交替估计出正确的运动层次,以确定最后的分割。

3. 算法实现

3.1 边缘提取

算法实现的第一步是在图像中确定边缘的位置。现在已经有很多合适的方法,本文中用到的是 Canny 边缘提取算法。其他的方法,比如 Sinclair 的分割算法^[7]也很适用。对于当前帧和后继帧均采用 Canny 算子提取边缘,其中的参数对于不同的实验序列应采用不同的值。如处理 foreman 序列时取以下的值效果较好:Low threshold=40, High threshold=80, sigma=1.5。这样两帧获得的边缘都很清晰连贯,而且孤立点较少,保留了大部分的真实轮廓信息(图1)。为了便于运动标记,需将交叉点处的边缘断开,每两个交叉点之间的部分作为一个边缘段,以增加运动跟踪的准确性。



图2 间距为5时的边缘节点图像

景运动,近似地取为零。但是这种方法会使前后景都做相似的大幅度运动,而且在整个迭代过程所有的边缘都会受这个均值运动的影响,所以这种假设是不必要的。本文用的方法是将所有的边缘随机地分成两组,分别当作前后景边缘,再通过每组边缘的样点残差估算两个运动的初始值。这种方法的好处是能提供较精确的运动概率,所有的边缘都有机会分别对两个运动做出贡献。

3.2.2 运动估算

为了减少计算量,可在运动跟踪时只考虑边缘样点的运动情况。在当前帧的每一条边缘上每隔一定的间距取一个样点,这样对边缘的运动跟踪就简化为对样点的运动跟踪了,从而减少了运动跟踪时的计算量。本文中样点的间距取为5,这样每帧典型图像中的样点数大约为1000个左右(图2)。

从边缘估计运动时,样点跟踪是把当前帧中的样点按照当前运动映射到下一帧实现的,当前帧中的每条边缘按其法线方向进行一维的搜索,以像素颜色和梯度为判别条件,在下一帧中找到最匹配的

匹配边缘,如图3所示。此时搜索的距离记为 d_k 。

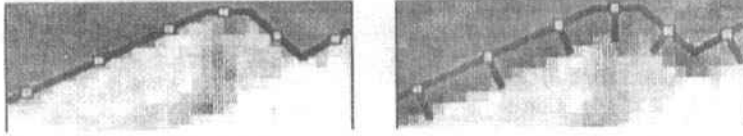


图3 边缘跟踪的示例

样点的残差 $r_{m,t}$ 定义为当前运动在边缘法线方向上的分量和 d_k 之差。在每个样点处运动 m 的参数 a_m 可由加权最小二乘法 and 边缘的运动概率联合求得,最适合的运动参数 a_m 会使下面的和式最小:

$$\min_{a_m} \sum_e \sum_{t \in e} r_{m,t}^2 P_{e,m} (m=1, 2) \quad (1)$$

其中 e 是边缘, t 为边缘 e 上的样点, $P_{e,m}$ 为边缘 e 属于运动 m 的概率。

3.2.3 运动概率估算

边缘所属运动的概率也可由残差获得,本文处理的视频序列假设均符合高斯分布,所以运动概率可由下式给出:

$$P_{e,m} = \exp(-\sum_{t \in e} r_{m,t}^2 / 2\sigma^2) \quad (2)$$

其中 $\sigma=2$ 。得出的边缘运动概率又被反馈到运动估算的过程中,用以更新运动,这两个步骤交替进行直至收敛。

3.2.4 收敛标志

最大似然法迭代的最终目的是得到当前帧中每条边缘分属两个运动的概率。例如,两个运动设为 A 和 B ,如果边缘 e 属于运动 A 的概率是 P ,则它属于运动 B 的概率为 $1-P$ 。当迭代到一定次数后,若运动概率没有明显变化则表明迭代已收敛(图4)。

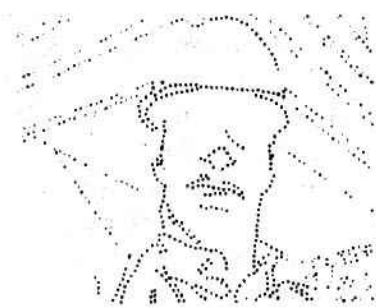


图4 分属不同运动的样点
(颜色较深的样点属于运动 A 的概率较高)



图5 用区域生长法得到的分割结果

3.3 区域分割

区域分割目前有许多方法可以实现,如分水岭(watershed)法, Sinclair 区域分割法^[9]等。本文采用的是类似于 Sinclair 区域增长法的方式,以较连贯稠密的边缘为区域间的严格界线,选取距边缘点较远的像素作为种子生长点,再以像素颜色为条件将当前帧分为若干个独立的区域。但在实际操作时由于边缘是单像素的,因此有时一个区域会穿过界限渗透到另外的区域里。为了解决这个问题,需要先将边缘加宽,即将每个边缘点的八个邻域点也当作边缘点,这样得到的新边缘就可以严格地界定区域了。得到初步的区域划分后,再以像素颜色为条件将边缘附近的剩余点加入最符合的区域中,完成区域分割过程。其中区域交界处的样点可以分别属于不同的区域。分割结果如图5。

3.3.1 区域标记

区域标记是确定每个区域是属于前景层还是后

景层的过程。决定区域的最大相似标记(maximum likelihood labeling)和运动层排序(layer ordering)同样还是个递归问题,因为只有两个运动,所以运动层的排序只有两种可能。本文的区域标记过程如下:

对于每个运动 $m(m=1, 2)$:

1)初始化区域的标记。一种合理的初始化方法就是以每个区域中主要边缘的运动标记来决定区域的标记。对于每个区域考虑其所包含的所有边缘的运动概率,累计属于前景运动边缘的样点数量,后景边缘和不确定的边缘都不计算在内。当所有前景样点的个数与样点总数之比大于一定的阈值时即认定此区域为前景区域。这样处理完所有区域后就完成了初步的区域标记。

2)重新调整边缘的运动标记。对于已经确定标记的区域(前景或后景)重新标记其从属的边缘,从而进一步标记那些先前不确定的边缘。但对于以前已经标记为前景的边缘不予更改。

3)再根据新的边缘重新标记各个区域。

反复迭代上述的2)~3)步骤大约15次左右,相对于当前运动绝大多数的区域都会被赋予正确的标记。

3.3.2 确定区域的运动层次序

由于本文研究的是包含两个运动的情形,因此对两个运动分别作两次区域分割。正确的分割是得



图6 区域标记的中间结果

到的边缘标记与实际可观测的边缘相一致的那种。为了达到这个目的,只需判定两种情形的运动残差 $r_{m,i}$,使下面的残差之和较小的那个即为最合理的分割。

$$\sum_{a \in A} \sum_{i \in e} r_{A,i}^2 + \sum_{b \in B} \sum_{i \in e} r_{B,i}^2 \quad (3)$$



图7 区域标记的最终结果

4. 实验结果

本文选取的第一个实验对象是 foreman 视频序列中的两帧图像,从当前帧到后继帧他的头部向左稍微移动了一些,虽然只是微小的运动,但是本文的方法仍然得到了很精确的结果,连帽子和左肩的部分也匹配得很好。与手动分割的结果相比大约有 98.0% 的区域都被标识出来了,同时效果也优于所

引的两篇文献结果。

选取的第二个例子是 Tennis 序列,图像中运动员在做向上击球的动作。与前例相比这个序列包含的边缘和区域要少得多,整个场景中只有运动员的前臂、球拍和球有少许的位移。本文的算法很好地将它们分割为前景,将静止的桌面作为背景,效果也要好于那两篇文献结果。



图8 P. Smith, T. Drummond^[4]



图9 Paul Alexander Smith^[5]



图10 本文的分割结果

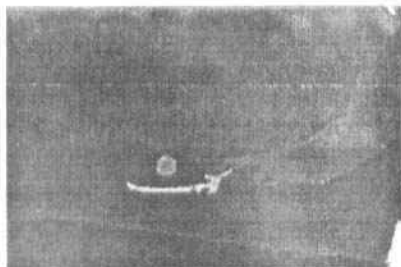


图11 Dufaux et al.^[13]

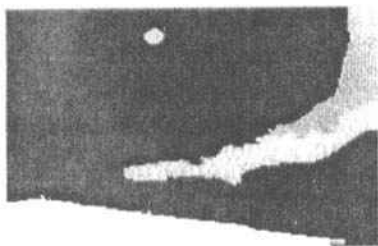


图12 Moscheni, Bhattacharjee^[14]

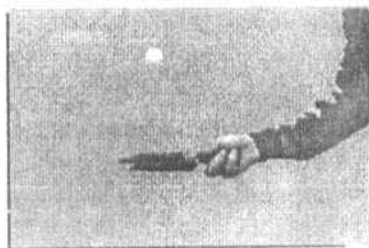


图13 本文的分割结果

结论 边缘在视频序列的运动分割中有着重要的意义,如果图像中的边缘都是可见的,那么只要简单地提取图像的边缘就可以得到很精确的分割结果。本文阐述了如何利用两帧图像之间的运动跟踪获取边缘的运动标记,从而确定区域的运动层次和区域标记的过程。运动跟踪,初始化运动和区域增长等方面的改进使得在两帧之间对边缘进行了很好的运动标记,绝大多数的样点都能匹配成功,区域增长的结果也与实际情形很一致,从而最终得到了比较令人满意的分割效果。

参考文献

- 1 Dempster A P, Laird H M, Rubin D B. Maximum likelihood from incomplete data via the EM algorithm. *Journal of Royal Statistical Society Series B*, 1977, 39: 1~38
- 2 Moscheni F, Dufaux F. Region merging based on robust statistical testing. In: *SPIE Proc. Visual Communications and Image Processing*, Orlando, Florida, USA, March 1996
- 3 徐静波. MPEG-4标准与多媒体通讯. *信息工程学院学报*, 1999, 8 (4)
- 4 Smith P, Drummond T, Cipolla R. Edge tracking for motion segmentation and depth ordering. *Department of Engineering University of Cambridge*
- 5 Smith P A. Edge-based Motion Segmentation. *Jesus College*

Aug. 2001

- 6 Hsu S, Anandan P, Peleg S. Accurate computation of optical flow by using layered motion representations In: *Intl. Conf. on Pattern Recognition*, Oct. 1994
- 7 Ayer S, Sawhney H S. Layered representation of motion video using robust maximum-likelihood estimation of mixture models and MDL encoding. *Intl. Conf. on Computer Vision*, Cambridge, June 1995
- 8 Sinclair D. Voronoi seeded colour image segmentation. *AT&T Laboratories Cambridge*, submitted for publication, 1999
- 9 Mech R, Wollborn M. Automatic segmentation of moving objects. *MPEG98/m3187*, 1998
- 10 Weiss Y, Adelson E H. A unified mixture framework for motion segmentation; Incorporating spatial coherence and estimating the number of models. *IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, June 1996
- 11 季白杨, 陈纯, 钱英. 视频分割技术的发展. *计算机研究与发展*, 2001, 38(1)
- 12 Canny J. A computational approach to edge detection. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 1986
- 13 Dufaux F, Moscheni F, Lippman A. Spatio-temporal segmentation based on motion and static segmentation. In: *Proc. Intl. Conf. on Image Processing*, Washington DC, USA, Oct. 1995
- 14 Moscheni F, Bhattacharjee S. Robust region merging for spatio-temporal segmentation. In: *Proc. Intl. Conf. on Image Processing Lausanne, Switzerland*, Sept. 1996

(上接第106页)

IV. Discussion

This paper describes a new text location approach based wavelet. The experimental results show that our method can locate and extract text from complicated background accurately and effectively, and is not heavily dependent upon the image contrast, font type and character size. And the computation in this method has a high speed. However, there is a problem in this method that the threshold used in experiment is set by manual, which will be solved in our future studies.

Reference

- 1 Fletcher L A, Kasturi R. A Robust algorithm for text string

separation from mixed text/graphics images. *IEEE transaction on Pattern Analysis and Machine Intelligence*, 1988, 10(6): 910~918

- 2 Jain A K, Bhattacharjee S. Text Segmentation Using Gabor Filters for Automatic Document Processing. *Machine Vision and Applications*, 1992, 5: 169~184
- 3 Tan C L, NG P O. Text extraction using pyramid. *Pattern Recognition*, 1998, 31(1): 63~72
- 4 Wu V, Manmatha R, Riseman E M. *Textfinder: An automatic system to detect and recognize the text in images*. *IEEE transaction on Pattern Analysis and Machine Intelligence*, 1999, 12(11): 1224~1229
- 5 Loprestie D, Zhou J. Locating and Recognizing Text in WWW Images. *Information Retrieval*, (Kluwer Academic Publishers.), 2000, 2: 177~206
- 6 Lienhart R, Effelsberg W. Automatic text segmentation and text recognition for video indexing. *ACM/Springer Multimedia System*, 2000, 8: 69~81
- 7 Sato T, Kanade T, Hugues E K, Smith M A, Satoh S. Video OCR: Indexing digital news libraries by recognition of superimposed caption. *ACM multimedia Systems*, 1999, 7(5): 385~395