

基于高斯混合模型的视频运动对象自动分割算法

李新仕 王天江 刘芳

(华中科技大学计算机科学与技术学院智能与分布计算实验室 武汉 430074)

摘要 提出的算法首先采用高斯混合模型依据空间属性对当前帧进行聚类分割,可以克服一般聚类算法对数据集中的噪声无法建模以及聚类数目难以确定的问题。然后依据时序属性,分割出当前帧的运动对象的初步轮廓区域。最后将初步轮廓区域和聚类分割的区域进行匹配,提取出视频运动对象。通过实验验证,算法具有较好的准确性和抗干扰性,在运动微小的情况下也能取得比较好的效果。

关键词 视频运动对象,高斯混合模型,最短描述长度,运动估计

Video Moving Object Segmentation Based on Gaussian Mixture Model

LI Xin-shi WANG Tian-jiang LIU Fang

(Intelligent and Distributed Computing Laboratory, College of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China)

Abstract An automatic and accurate video moving object segmentation algorithm was proposed which is based on the spatio-temporal property segmentation. The Gaussian Mixture Model was employed to cluster the current frame into a set of homogeneous regions based on the spacial property. Comparing with the general cluster algorithms, it could model the noise in the data set and estimate the optimal number of clusters more easily. It extracted the preliminary contour region of the moving object in the current frame based on the temporal property. The video moving object was extracted accurately by matching the homogeneous regions with the preliminary contour region. The experiment results demonstrate that the proposed algorithm is not only accurate and robust to noise, but also effective even if the movement is little.

Keywords Video moving object, Gaussian mixture model, Minimum description length (MDL), Motion estimation

1 引言

目前提出的从视频序列中分割运动对象的算法主要是以时序属性为依据^[1,2],这些算法根据视频的时序属性(如帧间差、光流等)来分割运动对象,其中时序帧差法由于分割效率高而成为最常用的方法,但仅依据时序信息却难以获得准确的对象轮廓。因此,为了获得准确的对象轮廓,一些算法还结合了基于图像空间属性的分割^[3,4],主要思想是首先根据灰度或色彩的相似性将视频的每一帧划分成若干区域,然后与基于时序属性的分割得到的运动区域进行匹配,从而实现运动对象分割。基于空间属性的区域划分主要用聚类方法来实现,但聚类方法对数据集中的噪声难以描述,且聚类的数目比较难以确定,为此许多聚类方法只能采用穷举法“自适应地”确定聚类数目^[5]。

本文提出的算法也是结合了图像的时序属性分割和空间属性分割,但是在基于空间属性进行分割时,采用高斯混合模型来实现,因为高斯混合模型本身就直接确定了噪声的模型,因此具有较好的抗干扰性。在确定聚类的数目时,依据最短描述长度准则,采用跨步法的搜索策略,能够比较快速地找到理想的聚类数目。

2 基于空间属性的分割

对像素的特征矢量进行聚类,可以将一幅图像分为若干个区域,每个区域内的像素具有较高的相似性,称之为类。一般来说,一个像素如果在某个类的质心的附近,那么它属于这个类的概率就大。因此对于每一个区域中的像素属于该类的概率分布可用一个高斯分布来描述,而对于整个图像的所有 k 个区域,则可用具有 k 个混合成分的高斯混合模型来描述。

利用高斯混合模型对特征矢量进行聚类,首先需要进行特征的提取。特征提取以像素为单位进行,每一个特征矢量包含颜色和位置的信息。由于 YUV(亮度、色度)颜色空间在视觉方面近似线性,因此选择 (Y, U, V) 作为颜色信息,同时选择每个像素的坐标 (x, y) 作为位置信息,引入位置信息的目的是为了降低过分割。将每一个分量归一化到 $[0, 1]$ 区间。通过以上方法将每一个像素映射到 5 维空间里面的一个矢量 $X = (Y, U, V, x, y)$ 中。

2.1 高斯混合模型及其 EM 求解

高斯混合模型是单一高斯概率密度函数的延伸,它能够平滑地近似任意形状的密度分布。设 X 是 d 维随机变量,若

到稿日期:2008-01-30 本文得到国家高技术研究发展(863)计划(项目编号:2007AA01Z161)的支持。

李新仕(1967-),博士研究生,副教授,主要研究方向为模式识别、计算机视觉, E-mail: xinshi_lee@126.com; 王天江(1960-),教授,博士生导师,主要研究领域为数据挖掘、知识发现、模糊推理。

x 为其中的一个采样,其高斯混合分布的密度函数为

$$P(x|\theta) = \sum_{j=1}^k \alpha_j p(x|\mu_j, \Sigma_j) = \sum_{j=1}^k \alpha_j \frac{1}{\sqrt{(2\pi)^d |\Sigma_j|}} \exp\left\{-\frac{1}{2}(x-\mu_j)^T \Sigma_j^{-1} (x-\mu_j)\right\} \quad (1)$$

其中, k 是混合成分的个数; $\theta = \{\alpha_j, \mu_j, \Sigma_j\}_{j=1}^k$ 为参数集; α_j 是混合成分系数,且 $\sum_{j=1}^k \alpha_j = 1$ 。若 $\{x_1, x_2, \dots, x_N\}$ 是 X 的一组样本集,参数 θ 的最大似然估计为

$$\theta_{ML} = \arg \max_{\theta} L(\theta|x_1, \dots, x_n) = \arg \max_{\theta} \sum_{i=1}^n \log p(x_i|\theta) \quad (2)$$

EM算法是一种从“不完全数据”中求解模型分布参数的极大似然估计的方法,它通过迭代来获得 θ_{ML} 。给出参数集 θ 的当前估计,则每一次迭代可按照以下两个步骤来对参数集进行重新估计:

E-Step:

$$\omega_{ij} = \frac{\alpha_j p(x_i|\mu_j, \Sigma_j)}{\sum_{i=1}^k \alpha_i p(x_i|\mu_i, \Sigma_i)} \quad j=1, \dots, k; i=1, \dots, n \quad (3)$$

ω_{ij} 表示样本 x_i 由成分 j 产生的概率,即由第 j 个高斯分布产生的后验概率,即 $p(j|x_i)$ 。

M-Step:

$$\begin{aligned} \hat{\alpha}_j &= \frac{1}{n} \sum_{i=1}^n \omega_{ij} \\ \hat{\mu}_j &= \frac{\sum_{i=1}^n \omega_{ij} x_i}{\sum_{i=1}^n \omega_{ij}} \\ \hat{\Sigma}_j &= \frac{\sum_{i=1}^n \omega_{ij} (x_i - \hat{\mu}_j)(x_i - \hat{\mu}_j)^T}{\sum_{i=1}^n \omega_{ij}} \end{aligned} \quad (4)$$

利用式(3)、式(4)进行迭代,直到相邻两次迭代的似然估计的对数的增量小于给定的阈值(本文取 0.001)为止。

2.2 模型的选择

模型选择指的是对式(1)参数 k 的选择。模型的选择对于输入样本的精确聚类是十分重要的, k 就是代表合适的聚类簇数。为了选择最优的 k 值,本文采用文献[6]中提出的最小描述长度(Minimum Description Length, MDL)准则。最小描述长度准则已经被用来在静态图像和视频序列中进行模型的选择^[7,8],其判别公式如下:

$$MDL = \log L(\theta_{ML}|x_1, \dots, x_n) - \frac{l_k}{2} \log n \quad (5)$$

θ_{ML} 是一个具有 k 个成分的最大似然高斯混合模型, L 是似然函数,见式(2)。 l_k 是需要确定的参数的个数,由以下公式给出:

$$l_k = (k-1) + kd + k\left(\frac{d(d+1)}{2}\right) \quad (6)$$

式(5)中等号右边的第一项是模型的似然度,第二项是模型的复杂度。模型的似然度和复杂度都随着 k 的增加而增加。最短描述长度判别准则就是选择 k 使得似然度尽可能大而模型的复杂度尽可能简单。为了选择合适的 k 值,本文对 MPEG4 的标准测试序列(trevor 序列和 Mother&Daughter 序列)利用高斯混合模型进行建模,并绘制出混合成分个数 k 和 MDL 值的变化曲线,如图 1 所示。

从图 1 中可以看出, MDL 随着 k 的增加,起初增加比较快,然后逐渐变慢,最后达到最高点,其中会有一些局部峰值。当曲线达到最高点以后,随着 k 的增加,模型的复杂度的增加比似然度的增加要大,因此按照最小描述长度的准则,最高点即为最优解。为了快速地找到最优解且避免陷入局部最优,

本文采用跨步走的办法取代逐步搜索办法。 k 从 2 开始,每次跨三步($k=k+3$),比较当前的 MDL 值(设为 MDL_k)和前一步的 MDL 值(设为 MDL_{k-1})。若 $MDL_k < MDL_{k-1}$,则最优解在前一步的附近,可在该点的附近进行搜索,即可找到最优的模型。采用跨步法可以避免过早选择局部最优解的问题,同时可以减少建模的次数,提高算法的效率。

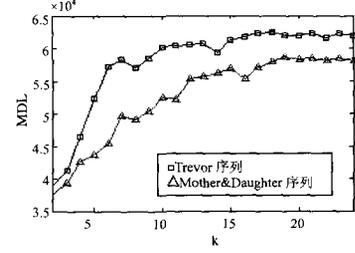


图 1 最小描述长度

2.3 样本所属类的判别

模型选择了以后,可利用 EM 算法估计出模型的参数集。那么对于任一样本 x_i ,它应当属于哪一个类?这需要判别 x_i 离哪一个类的质心最近。因为每一类中样本属于该类的概率分布可用一个高斯模型(混合模型中的一个成分)来描述,所以样本 x_i 所属的类可用以下公式进行判别:

$$Label(x_i) = \arg \max_j \alpha_j p(x_i|\mu_j, \Sigma_j) \quad j=1, \dots, k \quad (7)$$

3 基于时序属性的分割

基于时序属性的分割方法,常用的有光流法和帧间差分法。帧间差分法的缺点是易受到噪声的干扰,并且由于缺少空间边缘信息而使得对象边缘不准确。而光流法受到光流可靠性的影响,对噪声极为敏感,计算复杂。基于以上原因,本文采用块匹配法来进行运动估计,因为块匹配法计算简单,且不容易受噪声的干扰。但块匹配法得到的运动区域的边缘比较粗糙,因此本文结合边缘信息进行匹配,然后进行简单填充,估计出视频运动对象的初步轮廓区域。

3.1 运动估计

块匹配运动估计就是宏块 $T(i, j)$ 在搜索区域内与候选矢量集 CV 所对应的所有候选宏块 $S(i+m, j+n)$ 按照匹配准则进行匹配计算,其中最佳匹配块所对应的矢量即为该宏块的运动矢量,记为 $\vec{d} = (u, v)$ 。本文将当前帧划分为 8×8 的块,然后在前一帧的水平方向和垂直方向($-16, +16$)的区域内进行全搜索和匹配,以计算出最佳匹配结果作为运动估计的结果。在块匹配法中,衡量最优匹配有 3 种匹配准则:最小绝对差(MAD)、最小均方误差(MSD)和像素差分类(PDC)^[9]。本文采用像素差分类匹配准则,因为该准则计算比较简单,只需要进行加法运算。匹配准则公式如下:

$$F(i, j) = \begin{cases} 1, & |T(i, j) - S(i+m, j+n)| \leq T_m \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

$$PDC(m, n) = \sum_{i=1}^N \sum_{j=1}^N F(i, j)$$

$$\vec{d} = (u, v) | PDC(u, v) \geq PDC(m, n), \forall (m, n) \in CV$$

其中 T_m 为阈值(本文取 4), N 为块的边长。经过运动估计,得到运动区域的像素点集,设为 $D_i = \{d_1, \dots, d_m\}$ 。

3.2 初始边缘模板

选择 Canny 算子对当前帧进行边缘检测,得到当前帧的边缘点集,设为 $E_i = \{e_1, \dots, e_k\}$ 。设 $O_i = \{m_1, \dots, m_i\}$ 为视频对象初始边缘模板,则 $l \leq k$ 且 $O_i \subseteq E_i$ 。因为视频对象的边缘

模板在运动区域内或在其附近,所以可用式(9)提取出视频对象的初始边缘模板。

$$O_i = \{e \in E_i \mid \min_{x \in D_i} |e - x| \leq T_{ml}\} \quad (9)$$

T_{ml} 是距离阈值,一般可取 0~1 个像素。

3.3 简单填充

首先对二值边缘模板的每一列找到第一个和最后一个边缘点之间的所有垂直候选点,进行垂直填充。在进行水平填充时,对于每一行,选择所有有这样的一对候选点 (x, y_1) 与 (x, y_2) ,这两个候选点已被垂直填充,但它们之间的点尚未被填充,且 $y_2 - y_1 < T_F$ (T_F 为阈值,本文取 10),则将这两个候选点之间的点进行填充,否则不填充。经过垂直填充和水平填充,得到视频运动对象的初步轮廓区域。

4 提取视频运动对象

由于运动的复杂性和多样性,在利用时序属性分割得到的二值边缘模板并不是完全连续的轮廓边缘,因此不能直接用来进行视频对象的提取。如果利用简单填充后得到的轮廓区域来提取视频对象,则很不精确。基于空间属性的分割,能够比较好地将当前帧划分为若干区域,且视频对象的边缘轮廓比较精确,但由于缺乏运动信息,无法提取出运动视频对象。因此只有将二者结合起来,才能自动地提取出精确的运动视频对象,即将聚类分割得到的区域中属于视频运动对象的区域抽取出来。设 C_i 为当前帧聚类分割的某个区域, A 为当前帧的初步轮廓区域,区域匹配公式如下:

$$T_i = \frac{N(C_i \cap A)}{N(C_i)} \quad (10)$$

其中 $N(C_i \cap A)$ 为聚类区域与初步轮廓区域相交的像素数, $N(C_i)$ 为聚类区域的像素数。如果 T_i 大于阈值 T_{area} (本文取 0.6),则 $N(C_i)$ 视为运动视频对象,否则视为背景区域。

5 试验结果与分析

利用本文算法对标准序列 Trevor, Mother&Daughter 和 Clarie 序列进行了实验。为了减少计算量,本文进行了 1/4 的采样,即对于 176×144 的图像,隔行隔列进行采样。选择出最优模型 k 和求出模型参数 θ 以后,再利用模型对所有的像素进行分类。利用高斯混合模型对特征矢量进行聚类,将原图像分割为若干个区域,如图 2(b)所示。每一类可能得到一个或多个分割区域,这些区域中可能会有一些面积比较小的区域,为了减少匹配次数,可以预先将面积小于某个给定阈值的区域合并到周围较大的区域中。

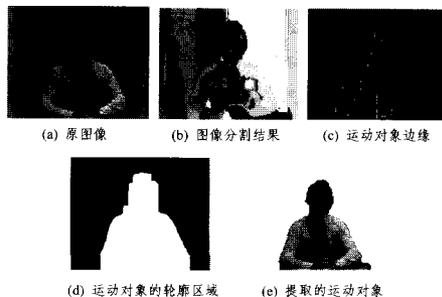


图 2 Trevor 序列实验结果

图 2 为 Trevor 序列中第 81 帧的分割结果和提取结果。图 2(a)为原图像;2(b)为利用高斯混合模型进行图像聚类分割的结果;2(c)为利用运动估计和 canny 算子提取的边缘图像进行匹配得到的运动对象边缘;2(d)为对运动对象边缘进

行简单的填充后得到的视频运动对象的初步轮廓区域;2(e)为利用聚类分割的区域和初步轮廓区域进行匹配的提取结果。

图 3 和图 4 分别为 Mother&Daughter 和 Clarie 序列原图像、运动对象的轮廓区域和最终提取结果。



图 3 Mother&Daughter 序列实验结果

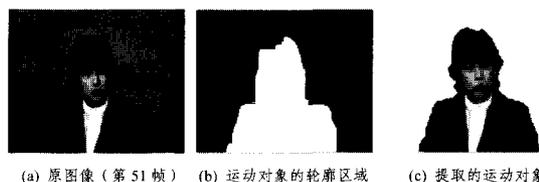


图 4 Clarie 序列实验结果

本文的算法不要求精确估计出运动对象的轮廓区域(如图 3(b)和 4(b)所示),因此在运动微小的情况下也能取得比较好的效果。例如图 3(a)中的小女孩以及图 4(a)中的播报员,头部运动明显,头部以下的运动并不明显,在利用块匹配进行运动估计时,头部以下部分只能检测出比较少的运动块。尽管如此,最后提取出的运动对象还是比较精确的,如图 3(c)和 4(c)所示。

结束语 本文提出的算法在基于时序属性进行分割时,并不要求精确地提取出运动对象的连续边缘,因此比较易于实现。算法采用块匹配法进行运动估计,不但计算量小而且不易受到噪声的干扰。在进行高斯混合模型的模型选择时,本文提出了跨步搜索和采样的方法,可以减少建模的次数和计算量,进一步提高算法的性能。

参考文献

- [1] Kim C, Hwang J N. Fast and automatic video object segmentation and tracking for content-based applications. IEEE Trans on Circuits and Systems for Video Technology, 2002, 12(2): 122-129
- [2] 杨莉,张弘,李玉山. 视频运动对象的自动分割. 计算机辅助设计与图形学报, 2004, 16(3): 301-306
- [3] Wei Wei, Ngan K N, Habibi N. Multiple feature clustering algorithm for automatic video object segmentation // Acoustics, Speech, and Signal Processing (ICASSP'04). 2004, 3: 17-21
- [4] 刘志,杨杰,彭宁嵩. 基于假设检验和区域合并的视频对象分割. 数据采集与处理, 2004, 19(2): 124-129
- [5] 向日华,王润生. 一种基于高斯混合模型的距离图像分割算法. 软件学报, 2003, 14(7): 1250-1257
- [6] Cover T M, Thomas J A. Elements of Information Theory. John Wiley and Sons, 1991
- [7] Belongie S, Carson C, Greenspan H, et al. Color and texture-based image segmentation using EM and its application to content based image retrieval // Proc. of the Int. Conference on Computer Vision. 1998: 675-682
- [8] Greenspan H, Goldberger J, Mayer A. Probabilistic space time video modeling via piecewise GMM[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2004, 26(3): 384-396
- [9] Gharvi H, Mills M. Block - matching motion estimation algorithms-new results. IEEE Trans on Circuits and Systems, 1990, 37(5): 649-651