

基于模糊聚类的运动对象分割算法研究

刘海华^① 张武^② 陈心浩^① 陈亚光^①

^①(中南民族大学电信学院 武汉 430074)

^②(中国科学院声学所 北京 100080)

摘要 为了从视频序列中分割出完整的、一致的运动视频对象,该文使用基于模糊聚类的分割算法获得组成对象边界的像素,从而提取对象。该算法首先使用了当前帧以及之前一些帧的图像信息计算其在小波域中不同子带的运动特征,并根据这些运动特征构造了低分辨率图像的运动特征矢量集;然后,使用模糊C-均值聚类算法分离出图像中发生显著变化的像素,以此代替帧间差图像,并利用传统的变化检测方法获得对象变化检测模型,从而提取对象;同时,使用相继两帧之间的平均绝对差值大小确定计算当前帧运动特征所需帧的数量,保证提取视频对象的精确性。实验结果证明该方法对于分割各种图像序列中的视频对象是有效的。

关键词 视频对象,对象分割,模糊C-均值聚类

中图分类号: TN919.8

文献标识码: A

文章编号: 1009-5896(2006)09-1689-04

Algorithm of Moving Object Segmentation Based on Fuzzy Clustering

Liu Hai-hua^① Zhang Wu^② Chen Xin-hao^① Chen Ya-guang^①

^①(School of Electronics and Information Engineering, South-Central University for Nationalities, Wuhan 430074, China)

^②(Institute of Acoustics, Chinese Academy of Sciences, Beijing 100080, China)

Abstract In order to obtain integrated and consistent segmentation of motion video objects, the segmentation algorithm based on fuzzy clustering is used to obtain pixels that constitute boundaries of motion object and extract video objects in the sequence sequentially. The motion properties of the current frame in each wavelet sub-band first are calculated using information of current frame and some frames before it in this algorithm. The set of motion eigenvectors is constructed with these properties. Then, significant change pixels are separated by fuzzy C-mean clustering algorithm based on these motion eigenvectors of low-resolution image. The change detection mask of motion object is obtained with significant change image, instead of frame difference, by conventional method of change detection and video objects are extracted. At the same time, the mean absolute different between consecutive two frames is used to determine number of frames that are used for properties calculation. It ensures accuracy of video objects obtained. The experimental results demonstrate the algorithm effective.

Key words Video object, Object segmentation, Fuzzy C-mean clustering

1 引言

基于对象的视频分割在数字视频处理和计算机视觉领域中重要的研究内容。视频对象分割的任务已经融入到许多应用中,如基于对象的视频编码(MPEG-4^[1]),基于内容的视频索引与恢复(MPEG-7^[2])等。因此,将视频序列分割为视频对象平面(VOPs)已成为数字视频处理中非常重要的问题。

目前出现了各种运动对象分割算法^[3-5],其中基于帧间差的变化检测是一种比较流行的分割方法。由于该方法实现比较简单,被广泛应用于开发完全自动的基于对象的视频处理系统。然而,基于帧间差变化检测方法也存在一些缺点。这些缺点包括噪声所引起的错误区域和

由对象运动所引起的未覆盖区域。为了解决该问题, Kim等建议了一个新的、基于帧间差图像的边界图的鲁棒的语义分割算法^[6,7]。然而,该算法仍然会受到背景噪声影响。另外,由于这些方法检测变化的标准是帧差,而帧差值又依赖于对象的运动速度。如果对象的运动速度在序列中发生较大变化,其分割质量很难保持一致。为此,人们将帧间差变化检测引入到小波域中处理^[8,9],从而提高构成对象边界像素的数量。虽然这些方法在一定程度上提高了分割质量,但运动速度所引起的分割质量差异仍然存在。同时,这些方法在分割过程中需要对每帧的每个子带图像进行显著性检验,以决定分离噪声的域值,因此增加了计算复杂性。不同的视频序列,以及同一序列不同帧中噪声程度是不同的,准确的域值很难确定,从而引起分割的不准确性。因此,本文提出了一种在小波域中基于模糊C-均值(fuzzy C-mean)聚类的方法检测

2005-12-06 收到, 2006-05-24 改回
国家自然科学基金(30370393)和国家民委自然科学基金(MZZ04004)资助课题

序列中图像的变化。该方法根据对象运动速度，使用不同数量帧的信息来确定图像在小波域中的运动信息，以此构成每帧图像的运动特征图像，并利用模糊聚类的方法分离出运动变化像素；最后以变化像素代替帧间差，通过传统帧间差变化检测方法，结合图像的空间边界信息分割出精确的视频运动对象。

2 模糊聚类分割算法原理

运动对象模糊聚类分割算法的基本原理与传统的帧间差变化检测方法相似，不同之处在于不直接使用传统的小波域变化检测方法的分割判断标准——帧间差^[7,8]，而是根据图像在小波域中的变化获得其运动特征。为此使用当前帧与之前的一些帧在小波域中每个子带之间均方差构造一个运动特征矢量集，利用模糊聚类的方法分割对象。

该方法假设视频序列背景是静止的，因为在许多视频会议和监控系统中，摄像机是固定的，本文的算法是基于这种情况。对于背景是运动的情况，很多研究者提出了使用全局运动估计和补偿来处理摄像机运动所引起的背景变化^[10]。因此，本文假设输入序列是通过补偿后的序列，其背景区域是静止的。本文所提出的算法分为 4 个主要部分，小波变换与特征计算、模糊聚类与变化检测、检测运动对象边界和对对象提取，如图 1 所示。

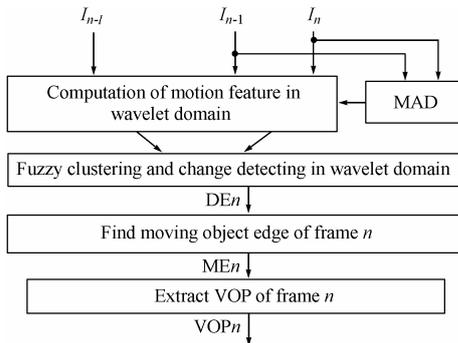


图 1 运动对象分割算法结构框图

Fig.1 Architecture of moving object segmentation algorithm

2.1 小波变换与特征计算

小波变换提供了一个完全的图像表达，它根据尺度将图像分解为 4 个不同的子带 LL, LH, HL 和 HH 等，该分解过程可以在每个尺度的 LL 子带上重复进行，获得下一个尺度下的 4 个子带。如果设 W^2 , W_H^2 , W_V^2 和 W_D^2 分别表示在尺度 i 的变换结果。对于尺度 J 的变换，原始图像可以表示为

$$(W^2, (W_H^2, W_V^2, W_D^2)_{1 \leq i \leq J}) \quad (1)$$

尺度的选择依赖于变化检测所需要的图像大小和原始序列的尺度大小。当序列图像的尺度较大时，需要选择一个大的小波变换尺度，使分割算法的操作在较低分辨率下进行，从而减少 VOP 提取的复杂度。

视频图像的特征计算是基于当前帧 I_n 和当前之前的一些帧 I_{n-j} 的子带运动信息，其中 j 为大于 1 的整数。假设计

算当前帧 I_n 的运动特征需要使用之前的 l 帧的图像信息。在 J 尺度的低分辨率下，由于每帧有 4 个子带，因此总共有 $(l+1) \times 4$ 子带图像。对于所有帧的每类子带图像，如低频分量子带 LL_n ，由于其相应的像素点上有不同小波系数，假设 $\omega_{i,j}(y)$ 为第 j 帧的第 i 类子带在像素点 y 上的小波系数，其中 $i=1,2,3,4$ 分别表示 4 个子带。那么，计算从当前帧 I_n 到前 l 帧 I_{n-l} 图像中每类子带在点 y 下的小波系数平均值 $MC_{i,j}(y)$ 为

$$MC_{i,j}(y) = \frac{1}{l+1} \sum_{j=n-l}^n \omega_{i,j}(y) \quad (2)$$

为了描述低分辨率空间下当前帧 I_n 中每个像素的变化，使用其均方差 $\sigma_{i,n}$ 来描述每个子带图像的变化，其中

$$\sigma_{i,n} = [MC_{i,n}(y) - \omega_{i,n}(y)]^2 \quad (3)$$

由于当前帧 I_n 中视频对象的运动变化将反映到其所有的小波子带上。因此根据当前帧 I_n 在尺度 J 下的 4 个子带中每个对应像素点的运动特征 $\sigma_{i,n}(y)$ ，可以构造一个运动特征矢量，从而体现在低分辨率下图像的变化。因此，每个子带相应点的运动特征可以作为在低分辨率空间下图像运动特征矢量的一个元素。这样，低分辨率图像中每个像素的运动特征是一个 4 维的矢量，其表示为

$$\mathbf{x}_{k,n} = \{\sigma_{1,n}(y), \sigma_{2,n}(y), \sigma_{3,n}(y), \sigma_{4,n}(y)\} \quad (4)$$

其中 k 为像素的标签，它代替低分辨率图像中每个像素的位置 y 。根据图像中每个像素的运动特征描述，可以构造一个反映图像中所有像素的运动特征的矢量集 \mathbf{X} ，其表示如下：

$$\mathbf{X} = \{\mathbf{x}_{1,n}, \mathbf{x}_{2,n}, \dots, \mathbf{x}_{k,n}, \dots, \mathbf{x}_{N,n}\} \quad (5)$$

其中 N 表示低分辨率图像所有像素的值，且 k 是满足 $1 \leq k \leq N$ 的整数。

2.2 变化检测

由于在低分辨率下图像的运动特征矢量表达了序列帧间变化，那么在低分辨率下当前帧的变化检测最终归属于基于像素的矢量属性的分类问题。分类问题是根据相似性准则将一个样本集分割为 C 类。如果在低分辨率图像中的每个像素作为一个样本，那么它们的运动特征矢量所构成的运动特征矢量集 $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ 为一组有限观测样本集， $\mathbf{x}_k = \{\sigma_{k1}, \sigma_{k2}, \sigma_{k3}, \sigma_{k4}\}^T \in \mathbf{R}^4$ 为观测样本的特征矢量，其中忽略了表示当前帧的标号 n 。

由于模糊 C-均值聚类(FCM)具有很强的抗噪声性^[11]，因此对运动特征矢量集采用 FCM 算法进行划分，即通过目标函数 $J_m(\mathbf{U}, \mathbf{B}; \eta)$ 发现具有不同变化特征的像素，其中

$$J_m(\mathbf{U}, \mathbf{B}; \eta) = \sum_{i=1}^C \sum_{k=1}^N (\mu_{ik})^m D(\mathbf{x}_k, \beta_i) \quad (6)$$

而矢量 $\mathbf{B} = \{\beta_1, \beta_2, \dots, \beta_C\}$ 表示每类的特征值集，称之为类的原型，而一个 $C \times N$ 的实数矩阵 \mathbf{U} 称之为模糊划分矩阵， μ_{ik} 为第 k 个样本相对于第 i 个子集的隶属函数； m 是大于 1 的加权指数， N 是可利用的数据数量，而 $D(\cdot)$ 是数据矢量 \mathbf{x}_k 与第 i 类原型矢量 β_i 之间的距离。

FCM 算法的准则是在约束条件 $\mu_{ik} \in [0,1]$, $\sum_{i=1}^C \mu_{ik} = 1$ 和 $\sum_{i=1}^C \mu_{ik} > 0; 1 \leq i \leq C, 1 \leq k \leq N$ 下取 $J_m(\mathbf{U}, \mathbf{B}; \eta)$ 极小化, 则可以获得下式

$$\mu_{ik} = 1 / \left[\sum_{j=1}^C \left(\frac{d_{ik}}{d_{ij}} \right)^{2/(m-1)} \right], \quad 1 \leq i \leq C, 1 \leq k \leq N \quad (7)$$

$$\beta_i = \left[\sum_{k=1}^N (\mu_{ik})^m \mathbf{x}_k \right] / \left[\sum_{k=1}^N (\mu_{ik})^m \right], \quad 1 \leq i \leq C \quad (8)$$

通过聚类, 不同运动特征的像素被分为不同类, 然后根据每类的运动特征矢量提取运动像素。图 2(a)给出了 CIF 格式通用序列“Trevor”第 13 帧的运动像素。

2.3 对象提取

在利用运动特征矢量 $\{\mathbf{x}_{i,n}\}$ 分类获得具有显著运动变化的像素之后, 如图 2(a)所示, 以此代替帧间差(FD_n), 并利用 Canny 边缘算子, 可获得在低分辨率下的边界图 DE_n :

$$DE_n = \text{Canny_Edge_Detection}(FD_n) \quad (9)$$

然后, 使用与文献[3]类似的方法, 使用 DE_n 和当前帧的低分辨率图像的空间边界图(E_n)提取当前帧在低分辨率下的运动边界 ME_n, 其中 ME_n 是由与 DE_n 相距一个小的距离 T^c 的边界像素组成, 即

$$ME_n = \{e \subseteq E_n \mid \min_{x \in DE_n} \|e - x\| \leq T^c\} \quad (10)$$

根据从 DE_n 中检测到的运动边界图 ME_n, 对其每行和每列进行扫描, 然后使用逻辑与操作, 获得水平和垂直的交叉区域被作为初始分割变化掩膜图像, 如图 2(b)所示。由于 ME_n 中存在噪声, 因此需要使用形态学操作处理, 从而产生变化掩膜图像(CDM), 如图 2(c)。最后, 将在低分辨率下所获得的 CDM 使用小波逆变换映射到原始分辨率下, 从而提取 VOP, 如图 2(d)所示, 但该 VOP 只是一个粗糙的对象平面。

为了进一步提出在原始分辨率下的精确 VOP, 可以将低分辨率下所获得的 ME_n 使用小波逆变换映射到原始图像分辨率下, 并将映射获得的 ME_n 被作为原始分辨率下图像的 DE_n^o。根据该 DE_n^o 重复使用上述操作, 利用原始图像的边界信息, 获得原始图像分辨率下的 ME_n^o。最后, 通过形态学操作处理, 获得在原始分辨率下的 CDM, 从而提取精确 VOP。

2 试验结果与分析

为了验证本文提出的算法, 将其应用于格式为 256×256 的典型头肩视频图像序列“Trevor”, 图 3 给出了序列“Trevor”中不同帧最后分割的 VOP, 其中图 3(a)为原始分辨率下的分割结果, 而图 2(d)为同一帧图像在低分辨率下的分割结果。从如图 2(d)和图 3(a)可以发现, 在原始图像分辨率下所获得的 VOP 要比在低分辨率下直接所获得的 VOP 精确。因此,

在一些不要求较高精度 VOP 的应用中, 可直接使用在小波域中所获得的粗糙 VOP; 相反, 如果应用要求高精度的 VOP, 就需要使用在原始图像分辨率下所获得的 VOP, 但其分割需要较多的操作和计算时间。

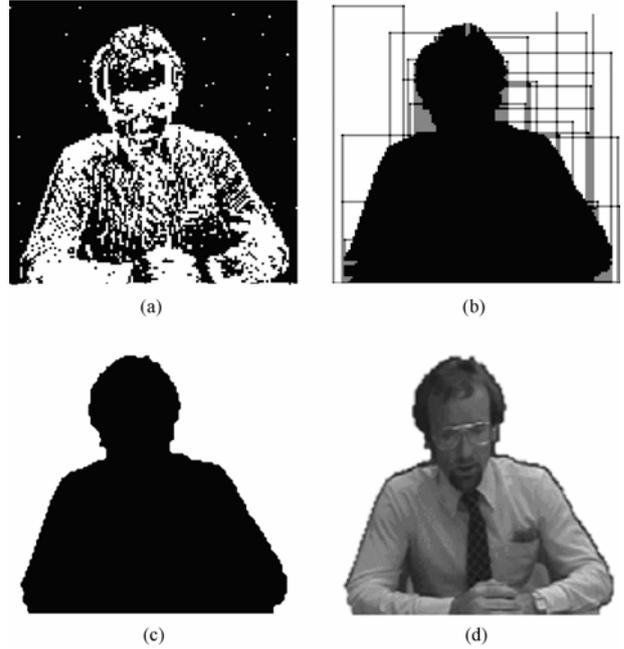


图2 小波域下VOP提取
(a) 显著运动变化像素 (b) 初始分割掩膜图像
(c) 分割掩膜图像 (d) 提取的视频对象

Fig.2 VOP extraction in wavelet domain
(a) Significant motion change pixels (b) Initial object mask
(c) Final object mask (d) Extracted VOP



图3 “Trevor” 序列不同帧分割的 VOP
(a) 第 13 帧 (b) 第 4 帧

Fig.3 VOP of different frames in “Trevor” sequence
(a) Thirteenth frame (b) Fourth frame

另外, 由于运动特征计算中使用了 l 帧的图像, 其目的是为了获得较多的对象运动变化信息, 而当帧的数量 l 较大时, 会产生较大的未覆盖区域, 因此 l 的数量必须进行控制。当对象运动变化较大时, 有足够的变化信息提取对象, 因此所需帧数量应较少; 相反, 当对象的运动变化较慢时, 变化的像素较少, 为获得足够多的构成对象边界的像素, 需要较多帧信息, 以保证对象分割的准确性。由此可见, l 的值是由对象运动变化决定, 而对象的运动变化可以通过相继两帧之间的平均绝对差(MAD)来评价。这样, 可建立 l 和 MAD

之间的关系:

$$l = \left\lceil \frac{1}{\text{MAD} - K} \right\rceil, \quad \text{MAD} = \sum_{i=1}^N \sum_{j=1}^M |I_n - I_{n-1}| \quad (11)$$

其中 M , N 为图像的尺度大小, K 为实验常数。由式(11)可知, 当对象运动较大时, 对象灰度值变化大, 所引起的 MAD 值较大, 则 l 值较小。当对象变化足够大时, MAD 值很大, l 值为 1, 表示只需要前一帧的信息。反之, 当对象运动较小时, 所引起的 MAD 的值较小, 则 l 的值较大。但为保证提取对象精度, l 值不宜太大。实现过程中, l 的值可以根据需要限定在一定范围内, 如 $l < 6$ 。实验中, 由于“Trevor”序列第 13 帧图像变化很小, 因此分割图 3(a)时 l 取值为 5, 而分割图 3(b)时, 由于第 4 帧图像的运动变化相对大, l 的选取较小, 即 $l=2$ 。

为了更好地说明算法的分割效果, 引入了对象模型的失真率 ρ , 其定义为

$$\rho = \frac{\text{error pixel count}}{\text{frame size}} \quad (12)$$

其中错误像素数量(error pixel count)是所获得的对象模型中不同于参考模型的像素数量。图 4 显示了采用本文建议算法和文献[8]建议算法分割对象的失真率。从图 4 可以看出, 在对象运动速度较慢时, 如“Trevor”的第 13 帧到 22 帧, 建议算法所得到的 VOP 失真率远小于使用文献[8]分割算法所获得的 VOP 的失真率, 提高了分割质量; 而在对象运动速度较大时, 由于都只用了前一帧的信息, 因此两种算法所获得的 VOP 的失真率相差不大, 但本文建议算法对失真率也有所改善。此外, 从图 4 的整体上可以看出, 文献[8]的分割算法所获得的 VOP 的分割质量与序列中对象的运动速度相关, 而建议算法较好地保持了分割质量的一致性。

该算法也成功地用于复杂背景的视频监控类型图像序列中。“Hall Monitor”就是一个背景较复杂, 对象较小的 CIF 格式的视频监控序列。图 5 显示了“Hall Monitor”序列中不同帧分割的 VOP, 其中图 5(a)和图 5(b)的失真率分别为 0.11% 和 0.06%, 与文献[8]建议算法分割的失真率 0.31% 和

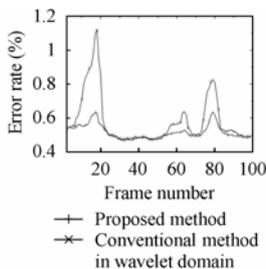


图 4 “Trevor”序列使用不同方法分割的失真率
Fig.4 Error rate of “Trevor” sequence with different segmentation method

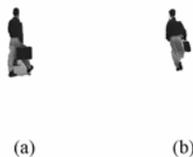


图 5 “Hall Monitor”序列不同帧分割的 VOP
(a)第 40 帧 (b)第 76 帧
Fig.5 VOP of different frames in “Hall Monitor” sequence
(a) fortieth frame
(b) seventy-sixth frame

0.22% 相比, 其分割精度同样得到改善。由于 FCM 算法需要大的计算量, 为了降低计算的复杂度, 可以使用文献[12]中的方法来减少数据, 从而提高聚类速度。由此可见, 该算法在保证对象提取准确性和一致性是非常有效。

4 结束语

本文提出了基于模糊聚类的运动对象分割算法。该算法通过对小波域中序列各子带图像的运动特征分析, 获得其运动特征矢量集, 并利用模糊 C-均值聚类算法获得低分辨率图像中发生重要变化的像素。最后以显著变化图像代替帧间差, 利用传统变化检测方法获得运动对象的边界, 从而提取对象。试验结果表明本文提出的方法能获得比较准确的视频对象的形状信息, 并可精确提取不同视频序列的运动对象。

参考文献

- [1] Sikora T. The MPEG-4 video standard verification model. *IEEE Trans.Circuit Syst. Video Technol.*, 1997, 7(1): 19–31.
- [2] Nack F, Lindsay A T. Everything you wanted to know about MPEG-7: Part 2. *IEEE Multimedia*, 1999, 6(3): 64–73.
- [3] 刘龙等. 局部运动场景中运动对象的压缩域分割算法. *电子与信息学报*, 2005, 27(8): 1249–1253.
- [4] 杨莉等. 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.
- [5] 吴思, 林守勋, 张勇东. 基于动态背景构造的视频运动对象自动分割. *计算机学报*, 2005, 28(8): 1386–1392.
- [6] Kim C, Hwang J N. Fast and automatic video object segmentation and tracking for content-based application. *IEEE Trans.on Circuits and Syst. Video Technol.*, 2002, 12(2): 122–128.
- [7] Kim C, Hwang J N. Object-based video abstraction for video surveillance system. *IEEE Trans. on Circuits and Syst. Video Technol.*, 2002, 12(12): 1128–1138.
- [8] Huang J C, Hsieh W S. Wavelet-based moving-object segmentation. *IEEE Electronics Letters*, 2003, 39(19): 1380–1382.
- [9] Huang J C, Su T S, Hsieh W S. Double-change-detection method for wavelet-based moving-object segmentation. *IEEE Electronics Letters*, 2004, 40(13): 798–799.
- [10] Dufaux F, Konrad J. Efficient, robust and fast global motion estimation for video coding. *IEEE Trans. Image Processing*, 2000, 9(3): 497–501.
- [11] Liew A W C, Shu Hung Leung, Wing Hong Lau. Segmentation for color lip images by spatial fuzzy clustering. *IEEE Trans. Fuzzy Systems*, 2003, 11(4): 542–549.
- [12] Eschrich S, Ke Jingwei, Hall L O, et al.. Fast accurate fuzzy clustering through data reduction. *IEEE Trans. Fuzzy Systems*, 2003, 11(2): 262–270.

刘海华: 男, 1966年生, 副教授, 博士生, 主要研究方向为图像处理与图像传输、计算机系统结构。

张武: 男, 1970年生, 博士生, 研究方向为多媒体技术, 信号处理。

陈亚光: 男, 1951年生, 教授, 主要研究方向为信号检测与信号处理。