以多目标优化的轮廓定位分割视频对象 [1]

宋立锋 韦 岗 王群生

(华南理工大学电子与通信工程系 广州 510641)

摘 要 该文提出一种用于半自动分割视频对象的对象跟踪方法. 它以方块均方色差和图像像素边缘强度构造目标函数, 把对象轮廓点的运动估值和补偿表达为用带容错的分层排序法实现多目标优化, 从而确定与参考帧对象轮廓点对应的当前帧像素为轮廓点; 以边缘强度最大为目标将这些轮廓点连接成闭合曲线, 从而得到当前帧的对象分割掩膜. 用这种方法处理多个视频测试序列得到满意结果.

关键词 视频对象分割, 半自动视频对象分割, 对象跟踪, 多目标优化, 运动估值和补偿

中图号 TN919.8

1 引 言

最新的多媒体数据压缩国际标准 MPEG-4 采用基于对象的编码方法, 一方面提高编码效率, 另一方面向用户提供存取场景内容的功能, 增强了多媒体节目的娱乐性. 视频对象分割是 MPEG-4 标准的关键技术. 然而因为成像过程中的信息丢失和噪声, 使输入图像数据不满足唯一正确解的充分条件, 同时人工智能技术的现状决定计算机不具有人的观察、识别、理解图像的能力, 所以至今还没有通用的有效方法去解决这个问题.

目前视频对象分割有全自动分割 [1−3] 和半自动分割 [4−9] 两类. 全自动分割检测运动并自动生成所有分割结果, 而通常的半自动分割是人工生成初始帧分割结果, 然后通过对象跟踪自动生成后续帧分割结果. 这两类分割都是人机结合的过程. 人的参与不仅表现在半自动分割中通过图形用户界面生成初始帧分割结果, 而且表现在全自动分割中设定门限值、初始条件以及挑选适合计算机处理的视频序列. 半自动分割不仅较全自动分割更实用、结果更精确, 而且目前是获得符合人主观意图的任意对象的唯一途径. 在评价图像和视频对象分割的质量方面尚缺乏规范的标准. 现有方法包括主观视觉评价和定量评价.

对象跟踪是半自动分割的核心, 关键点是对象特征表示和对象特征定位. 在对象特征表示上, 文献 [4] 以 2D 网格为对象特征, 文献 [5] 以区域为对象特征, 而更多的是以轮廓线为对象特征 [6−9] . 在以轮廓线为对象特征的方法中又分为基于像素的方法 [6,7] 和基于闭合曲线 [8,9] 甚至是参数表示的闭合曲线 [9] 的方法. 在对象特征定位上, 以运动估值和补偿 [5−7] 、能量最小化 [8,9] 两种方法最重要, 并结合一些静止图像的处理方法如水线算法 [6] .

在这些方法中, 韩国 Electronics and Telecommunications Research Institute (ETRI) 的方法 [7] 效果最好. 它是半自动分割, 以轮廓线为对象特征, 以像素为单元, 以运动估值和补偿为主要对象特征定位方法. 但是遇到较复杂运动 (人物头部不停转动) 以及较复杂背景时出现较大误差. 对此本文提出一种新的对象跟踪方法, 以方块均方色差和图像像素边缘强度等多个图像特征构造目标函数, 采用带容错的分层排序法进行运动估值和补偿, 求出所有参考帧轮廓点在当前帧的对应点. 然后以边缘强度最大为目标把所有这样的点连接成闭合曲线, 得到当前帧分割掩膜 (Segmentation mask) . 用本文方法处理多个头肩像视频测试序列, 分割结果令人满意, 表现在对象轮廓点运动估值和补偿的精度提高, 而且对于对象的平移、缩放、旋转、变形等运动的适应性有一定增强.

[1] 2000-11-24 收到, 2002-03-25 定稿
国家自然科学基金重大项目 (69896246) 资助

2 两步完成轮廓定位

本文对象跟踪方法以轮廓线为对象特征，基于像素，用对象轮廓点的运动估值和补偿、将对象轮廓点连接成闭合曲线这两个主要步骤完成轮廓定位。

2.1 对象轮廓点的运动估值和补偿

“所有运动估值算法的基本原理 (又称为光流约束条件) 是使在真实运动路径上逐帧的图像强度保持不变 (或者以一种已知或可以预测的形式变化) ” [10]。不过，对象轮廓点在图像中具有特定的相对于前景对象和背景的位置。这个约束条件比上述光流约束条件还重要。当对象发生单纯的平移运动和缩放变化时，根据光流约束条件在当前帧找到的参考帧轮廓点的对应点，相对于对象和背景的位置不变，就是目标点。而当对象发生其它运动时，根据光流约束条件所找到的对应点如果不满足相对位置不变的约束条件，就不是目标点。同时希望当参考帧的某个轮廓点在当前帧被覆盖 (遇到遮挡) 时也能确定最佳对应的轮廓点。

本文对参考帧的每个对象轮廓点 (x,y) 构造一个以该点为中心的、大小为 $(N+1)\times(N+1)$ 的方块，采用最小均方彩色误差 (MSCE, Mean Square Color Error) 的块匹配法进行运动估值：

$$\mathrm{MSCE}(u,v)=\sum_{i=x-N/2}^{x+N/2}\sum_{j=y-N/2}^{y+N/2}\|\boldsymbol{I}(i+u,j+v)-\boldsymbol{I}'(i,j)\|^2 \tag{1}$$

(1) 式中，$\boldsymbol{I}'$ 和 $\boldsymbol{I}$ 分别表示参考帧、当前帧在 YUV 色度空间的矢量，取欧氏距离作范数 $\|\cdot\|$。在此基础上，本文构造含有两个图像特征的目标函数：

$$(\min \mathrm{MSCE}, \max \mathrm{EdgeStrength}) \tag{2}$$

通过彩色矢量的块匹配和最大边缘强度反映轮廓点特征、满足上述光流约束条件和轮廓点相对位置约束条件。彩色矢量块匹配比通常的亮度分量块匹配包含更多的信息，同时彩色对象特征相对于亮度对象特征对噪声更不敏感。(2) 式既有利用帧间相关性的时间分割，又有利用轮廓点空间特征的空间分割。通过引入更多的信息和特征，使轮廓点定位有较好的精度及鲁棒性。

(2) 式是多目标寻优问题，可以借鉴并引用多目标优化技术中通过排序求最优值的部分。不过对于多目标优化可以根据目标函数的性质研究解的性质 [11]，在这里却不能进行这样的研究。图像数据的随机性使图像特征、目标函数和解都是随机的，只能对随机变量具体表现的数值进行排序。因此本文只能根据实验结果，确定具有最好表现的优化方法为

$$\left.\begin{aligned}&\text{Level1:}\quad (\min \mathrm{MSCE})\\&\text{Level2:}\left(\max\frac{\mathrm{EdgeStrength}}{\sqrt{\mathrm{MSCE}}}\right)\end{aligned}\right\} \tag{3}$$

这种按重要程度分两层进行单目标优化的处理方法即分层排序法。考虑到排序时总可以输出最值，但是可能出现多个相同的数同时达到最优值的情况，本文参照文献 [11] 引入宽容量，提出带容错的分层排序法表达式：

$$\left.\begin{aligned}
&\mathrm{MSCE}(u,v)=\sum_{i=x-N/2}^{x+N/2}\sum_{j=y-N/2}^{y+N/2}\|\boldsymbol{I}(i+u,j+v)-\boldsymbol{I}'(i,j)\|^2,\\
&\qquad\text{for all}(u,v)|_{\mathrm{MSCE}(u,v)\le\min\mathrm{MSCE}(u,v)+\Delta},\qquad -p\le u,v\le p\\
&\mathrm{MV}_{\mathrm{refined}}=(u+u',v+v')\Big|_{\max\frac{\mathrm{EdgeStrength}(u+u',v+v')}{\sqrt{\mathrm{MSCE}(u+u',v+v')}}},\ -P'\le u',\ v'\le p',\ p'<N/2
\end{aligned}\right\}\tag{4}$$

式中 Δ 是第 1 层优化结果的容错量，(u,v) 和 (u',v') 分别是第 1 ，2 层优化得到的位移矢量，p 和 p' 分别决定第 1 ，2 层优化的搜索范围，$\mathrm{MV}_{\mathrm{refined}}$ 是对象轮廓点运动估值输出结果。Δ 越小，则第 1 层优化输出的匹配块数目越少，MSCE 的限制程度越强；反之，Δ 越大，则第 1 层优化输出的匹配块数目越多，边缘强度的限制程度越强。前者适应于对象发生平移或缩放的情况，而后者适应于对象发生旋转或变形的情况。调整 Δ 适应复杂情况，使分割结果更加准确。可以简单地取 Δ 为相对固定值如下：

$$\Delta=\min\mathrm{MSCE}(u,v)\times 20\%\tag{5}$$

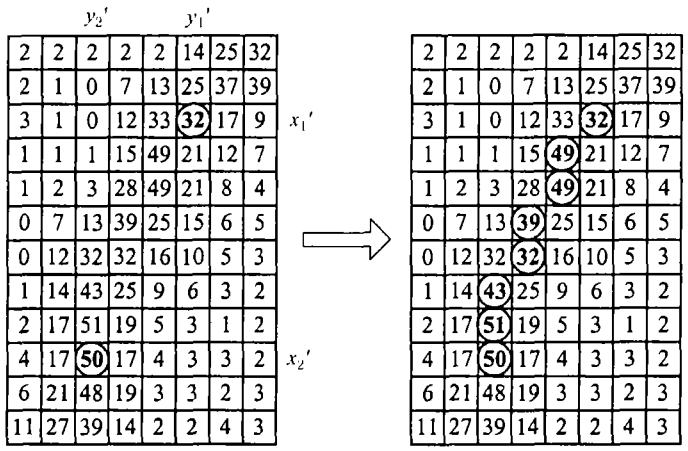
当 $\min\mathrm{MSCE}(u,v)$ 较小时，反映帧间变化不大，取少一些匹配块到第 2 层优化，MSCE 的限制程度较大；而当 $\min\mathrm{MSCE}(u,v)$ 较大时，反映帧间变化较大，可能是发生了较复杂运动或遇到复杂背景，取多一些匹配块到第 2 层优化，边缘强度的限制程度较大。从而适应复杂情况。

虽然文献 [7] 的用于处理非刚性物体的对象跟踪方法包括运动估值和补偿、对象轮廓精确化这两个步骤，所用的公式与 (4) 式相似，但是没有表现当前帧的不同方块取相同或相近的平均绝对偏差 (MAD, Mean Absolute Difference) 时如何处理。而采用 MAD 比采用 MSCE 更有可能出现不同方块取相同或相近值的情况。所以 ETRI 方法遇到较复杂运动以及较复杂背景时出现较大误差，反映在本文的图片中。

2.2 形成对象轮廓线

经过运动补偿得到的点可能是分离的，需要按像素 → 闭合曲线 → 对象的步骤形成闭合曲线并得到分割掩膜。ETRI 在文献 [7] 中未提及处理方法。本文利用参考帧轮廓点的相邻性检查当前帧估计轮廓点的相邻性，并以边缘强度最大作目标连接分离的估计轮廓点，使形成的曲线与实际轮廓线尽量吻合。

对于参考帧两个相邻的轮廓点 (x_1,y_1) 和 (x_2,y_2) ，运动补偿后在当前帧的位置分别为 (x_1',y_1') 和 (x_2',y_2') 。如果这两点同时满足 $|x_1'-x_2'|\le 1$ 和 $|y_1'-y_2'|\le 1$ 的条件，为相邻点；否则，为分离的点，需要建立两点之间的连接：如果这两点同行或同列，形成一条以这两点为端点的横线段或竖线段；如果 x_1' 、x_2' 、y_1' 、y_2' 构成一个矩形，如图 1(a) 所示，根据图像边缘强度把这两点连接起来，形成如图 1(b) 所示的由带圆圈黑体数字像素组成的曲线。图 1 取自 Claire 序列的一小块图像。每个方格为一个像素，方格中数字是图像边缘强度，其中的带圆圈黑体数字表示轮廓点。从 (x_1',y_1') 移动至 (x_2',y_2') 的过程是：每次从当前点 (x',y') 移动到相邻的 3 个点 $(x'+\Delta x,y')$ ，$(x'+\Delta x,y'+\Delta y)$ ，$(x',y'+\Delta y)$ 中的一个点。该点在矩形范围内而且边缘强度最大。Δx ，Δy 为行、列移动量。如果 $x_1'<x_2'$ ，$\Delta x=+1$ ；否则，$\Delta x=-1$ 。Δy 的设定方法相同。把这样的 3 个邻点按移动次序排序。如果 $|x_1'-x_2'|\ge|y_1'-y_2'|$ ，次序是：$(x'+\Delta x,y')$ ，$(x'+\Delta x,y'+\Delta y)$ ，$(x',y'+\Delta y)$ ；否则，次序是 $(x',y'+\Delta y)$ ，$(x'+\Delta x,y'+\Delta y)$ ，$(x'+\Delta x,y')$ 。遇到边缘强度相同的情况就取移动次序在前面的点。这样的逐点移动一直进行下去，直到移到 (x_2',y_2') 的邻点。



(a) 连接前 (b) 连接后

图 1 连接分离的轮廓点

形成闭合曲线后可以很容易地生成当前帧的分割掩膜：从图像平面的 4 个顶点中，选择其中的非对象像素为种子，向上、下、左、右 4 个方向生长，直到遇到对象像素。所形成的区域为背景区，而余下的区域为对象区域。

3 实现流程

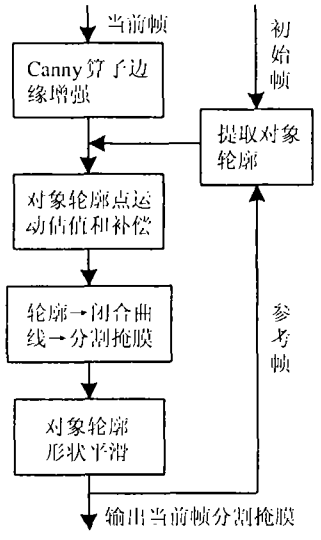


图 2 本文对象跟踪流程图

图 2 是本文方法的流程图。用 Canny 算子 [12] 对图像 Y 分量进行边缘增强。用 4 邻点 "掏空" 法从已知的参考帧分割掩膜中提取对象轮廓。执行逐个对象轮廓点的运动估值和补偿，得到当前帧的对象轮廓点，如 2.1 节所述。然后完成检查当前帧轮廓点的相邻及连接分离的点的步骤，如 2.2 节所述，从而形成一条闭合曲线，并确定以此闭合曲线为轮廓线的对象。最后用先开后闭的形态学滤波对二值的分割掩膜完成对象形状平滑。所用的结构元素是圆。平滑后的分割掩膜作为本帧的最后分割结果输出。以上各步算法除本文第 2 节所述的外，都是常用的数字图像处理算法，参见文献 [13] 的有关章节。

4 实验结果

用 Microsoft Visual C++ 6.0 工具把以上流程编成 C 程序，处理 QCIF(176×144，Y:U:V 4:1:1) 格式的彩色视频测试序列。用 Adobe Photoshop 5.0 工具对每个序列的第 1 帧图像勾画

出对象轮廓线，并生成初始帧分割掩膜。初始帧的选择是任意的，它确定了对象跟踪的起点。(4) 式的有关参数设为： N 为 8， p 为 7， p' 为 1， Δ 为 $\min \mathrm{MSCE}(u,v) \times 20\%$ 。匹配块的搜索方法是全搜索。形态学滤波的圆形结构半径为 2 。读入每帧图像后通过色度分量的双线性内插转化为 Y:U:V 4:4:4 格式。

对视频对象分割进行定量评价需要正确的对象标记图，才能统计分割结果的误标记像素数目。文献 [14] 提出了一个综合全局测度和局部测度的评价函数 F ，使评价结果接近主观视觉效果。但是，“对真实图像确定每个像素的正确标记，即使不是不可能，也是非常困难。虽然有人根据人工预先分割好的结果计算误标记率，但是只能应用于简单场景，而且难以令人信服这样的人工分割结果就一定是正确的” [14] 。而且生成正确的对象标记图需要精确的图像定位工具。所以在绝大多数情况下，图像和视频对象分割采用主观视觉评价方法，如文献 [1–9] 。为方便观察，把每帧的分割结果与每帧的原始图像结合起来，或形成对象轮廓线图像，或去除背景形成前景对象图像，或保存为一个包括多帧连续运动图像的序列的文件，或保存为多个静止图像文件。由于篇幅所限，在图 3 中只列出 5 个序列的初始帧、中间帧、末尾帧的分割结果，为灰度图中的白色轮廓线。用本文方法处理这 5 个序列的结果都准确，达到了实用的要求。

将本文方法处理的结果与 ETRI 的版本为 1.2 的演示程序 [15] 的处理结果进行比较，在图 4 列出存在差别的比较图像。 ETRI 演示程序的所有保存功能都被 ETRI 取消了。图 4 中的 ETRI 图像是经过截图等一系列处理后得到的结果。两个程序所用的初始帧分割掩膜都相同。

对于 Claire 序列，在人物头部不停转动中， ETRI 方法加强了光流约束条件的限制，却未能准确定出轮廓的位置，将人物左脸部附近的对象部分误分为背景。对于 Alexis 序列， ETRI 方法将人物右手臂的对象部分误分为背景，而把人物左手附近的背景误分为前景。这些部分没有变化或仅有不大的平移运动，只是桌面背景较复杂而且存在晃动的人影。 ETRI 方法执行块匹配的精确度不够，而且未能准确定出轮廓位置。对于 Miss America 序列， ETRI 方法处理的人物头部轮廓线形状不好。这也反映了 ETRI 方法不够精确。而本文方法得到更准确结果。

以一台配置为 Pentium II 400MHz CPU ， 64MB 内存的微机分别运行本文程序和 ETRI 演示程序，情况如表 1 。

表 1　对象跟踪运行时间

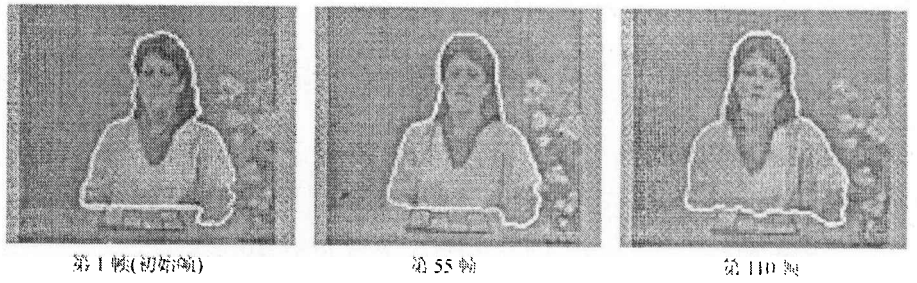
序列名称	序列长度 (帧)	ETRI 演示程序 (s)	本文程序 (s)
Claire	158	38	171
Alexis	110	27	126
Miss America	150	43	167
Mother & Daughter	100	32	132
Akiyo	50	14	56

表 1 中本文程序运行时间较长，是因为本文程序采用全搜索方法执行运动估值，而且没经过程序优化。相对于 ETRI 作为产品的程序，这样的数据仅供参考。

5 结　论

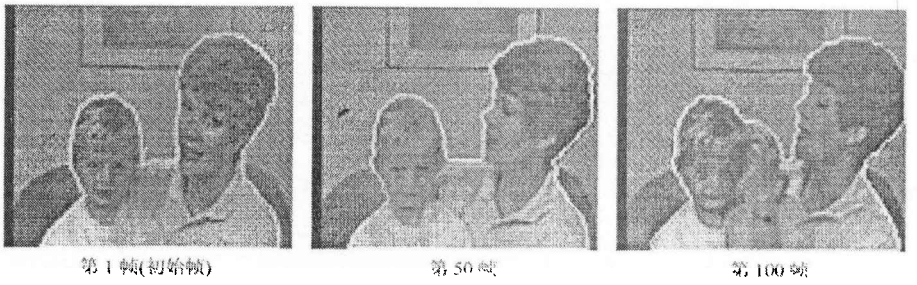
用本文方法处理多个头肩像的视频测试序列，都得到主观满意的结果，达到实用要求，比 ETRI 的演示程序分割结果更精确。虽然本文没专门对复杂运动建立数学模型求解，但是通过应用多目标优化技术，选择或调整合适的宽容量，使对象跟踪更好、更充分反映轮廓点特征，增强分割精度和对象跟踪方法的鲁棒性，从而提高了本文方法对于平移、缩放、旋转、变形等图像运动的适应性。本文方法产生像素精度的分割结果，满足 MPEG-4 标准的基于对象编码的要求。

(a)Claire 序列



(b)Alexis 序列

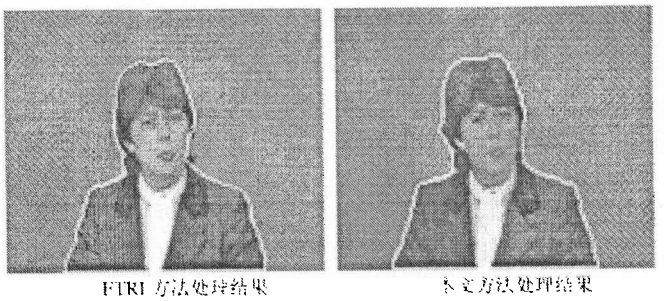
(c)Miss America 序列



(d)Mother & Daughter 序列

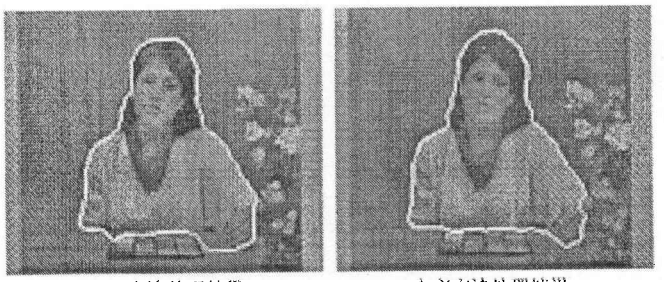
(e)Akiyo 序列

图 3 本文方法处理结果



ETRI 方法处理结果　　本文方法处理结果

(a)Claire 序列第 40 帧



ETRI 方法处理结果　　本文方法处理结果

(c)Alexis 序列第 80 帧

ETRI 方法处理结果　　本文方法处理结果

(e)Miss America 序列第 100 帧

图 4　ETR1 方法和本文方法的处理结果比较

参 考 文 献

[1] R. Mesh, M. Wollborn, 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.

[2] A. Neri, *et al.*, Automatic moving object and background separation, Signal Processing, 1998, 66(2), 219–232.

[3] M. Kim, *et al.*, A VOP generation tool: automatic segmentation of moving objects in image sequences based on spatio-temporal information, IEEE Trans. on CAS VT, 1998, 9(8), 1216–1226.

[4] C. Toklu, *et al.*, Semiautomatic video object segmentation in the presence of occlusion, IEEE Trans. on CAS VT, 2000, 10(4), 624–629.

[5] D. Zhong, S. F. Chang, An integrated approach for content-based video object segmentation and retrieval, IEEE Trans. on CAS VT, 1998, 9(8), 1259–1268.

[6] C. Gu, M. C. Lee, Semiautomatic segmentation and tracking of semantic video objects, IEEE Trans. on CSVT, 1998, 8(5), 572–584.

[7] ISO/IEC JTC1/SC29/WG11 14496-2 FDAM 1. Information Technology-Coding of Audio-Visual Objects-Part 2: Visual/Amendment 1. N3056, Maui, Dec. 1999, Annex F: 452–462.

[8] F. Leymarie, M. D. Levine, Tracking deformable objects in the plane using an active contour model, IEEE Trans. on PAMI, 1993, 15(6), 617–634.

[9] 赵雪春，戚飞虎，用可变形模板进行基于内容的图像分割算法，电子学报，2000, 28(4), 69–72.

[10] A. M. Tekalp, Digital video processing, Prentice Hall, Inc., 1998, Chapter 7: 118.

[11] 林锉云，董加礼，多目标优化的方法与理论，长春，吉林教育出版社，1992, 第四章，83–87.

[12] J. Canny, A computational approach to edge detection, IEEE Trans. on PAMI, 1986, 8(6), 679–698.

[13] 吕凤军，数字图像处理编程入门—做一个自己的 Photoshop，北京，清华大学出版社，1999, 第 6 章，第 7 章.

[14] J. Liu, Y. H. Yang, Multiresolution color image segmentation, IEEE Trans. on PAMI, 1994, 16(7), 689–700.

[15] ISO/IEC JTC1/SC29/WG11, ETRI's user-assisted video object segmentation tool. M4479, Seoul, Mar. 1999.

SEGMENT VIDEO OBJECTS BY CONTOUR LOCATING OF MULTIOBJECTIVE OPTIMIZATION

Song Lifeng Wei Gang Wang Qunsheng

(*Dept. of Electron. and Comm. Eng., South China Univ. of Tech., Guangzhou* 510641, *China*)

Abstract An object tracking method for semiautomatic video object segmentation is proposed in this paper. An objective function is constructed with mean square color error of blocks and edge strength of pixels. Then a multilevel ordering method with permissive error is used to accomplish the programming of this objective function so as to realize motion estimation and compensation of contour pixels. Thus the pixels in the current frame, which correspond to the contour pixels in the reference frame, are found out. As the contour pixels in the current frame, these pixels are connected to form a close curve by an objective of maximal edge strength. Thus the segmentation mask of the current frame is generated. The experimental results of this method on several test sequences are satisfactory.

Key words Video object segmentation, Semiautomatic video object segmentation, Object tracking, Multiobjective optimization, Motion estimation and compensation

宋立锋： 男，1967 年生，博士生，从事图像和视频信号处理、压缩以及通信等研究.

韦 岗： 男，1963 年生，博士生导师，教授，从事信号处理、数字通信等教学及研究工作.

王群生： 男，1939 年生，教授，从事图像压缩编码、数字电视和图像通信等教学和研究工作.