

视频对象分割与两种面向对象的视频编码器

翁南钊, 蔡德钧

(华中理工大学电子与信息工程系, 武汉 430074)

摘 要: 在基于对象的视频编码中, 视频对象的分割是重要的任务. 本文研究一种利用位移帧差 (DFD) 的高阶统计特性和数学形态学算子的视频对象自动分割方法. 这种方法首先根据一组转移帧差的高阶矩来得到一个大体覆盖运动对象的区域 (模板), 接着用形态学的腐蚀算子从模板的边沿向内腐蚀, 直到对象的边沿. 提出一种简单而高效的基于查找人头肩部轮廓最大转折点的头肩分离算法. 在分割的基础上, 用软件实现了一种基于 MPEG-4 的视频编码系统. 提出一种面向对象分配带宽 (OOBA — object-oriented bandwidth allocation) 的极低比特率视频编码器. 该编码器与传统基于帧的视频编码器相比, 在低比特率环境下, PSNR 略为下降, 但图像的主观视觉质量得到提高.

关键词: 视频对象分割; 头肩分离; MPEG-4; 基于对象的视频编码; 低比特率视频编码

中图分类号: TN919.81 **文献标识码:** A **文章编号:** 0372-2112 (2000) 10-0106-05

Video Object Segmentation and Two Kinds of Object-Oriented Video Coder

WENG Nan-shan, CAI De-jun

(Dept of Electronic & Information Engineering, Huazhong Univ. of Scie. & Tech., Wuhan 430074, China)

Abstract: Video object segmentation is an important task in object-based video coding. In this paper, an automatic video segmentation scheme based on higher-order statistics (HOS) of displaced frame differences (DFD) and mathematical morphological operators is studied. In this approach, a map (mark) overlapping object is first gotten by thresholding a group of higher-order moment maps of DFD, and then a morphological erosion from the boundary of the mark up to the boundary of the object is performed to get the final segmentation map. A simple but efficient algorithm for separating head and shoulder is proposed based on searching for the greatest degree turning-point in head-shoulder contour curve. Based on segmentation, the MPEG-4 based video coding system we implemented with software is described. Then, a very low bit-rate video coder with object-oriented bandwidth allocation (OOBA) technology is proposed. Compared with conventional frame-based video coder, the OOBA-coder can improve subjective quality of images at low bit-rate, while decrease PSNR just a little.

Key words: Video object segmentation; head-shoulder separation; MPEG-4; object-based video coding; low bit-rate video coding

1 引言

近年视频编码的研究集中在两个方面, 一是低比特率编码, 二是基于对象的视频编码. 对前者, 人们研究的重点在于提高编码图像的主客观质量. 后者更侧重于为多媒体业务 (如, 交互电视, 网上购物, 交互游戏等) 提供更高的交互性, 它是近年研究的热点, 它的兴起在很大程度上受 MPEG-4 标准制订活动的驱使. 新的运动图像编码标准 MPEG-4^[1] 就支持基于对象的编码, 它能够对任意形状的视频对象 (VO) 进行编码, 而不仅仅对传统的矩形图像序列编码. 由于要对不规则形状对象编码, 基于对象的视频编码遇到许多新的难题, 如怎样自动分割有意义的视频对象; 怎样对任意形状的图像区域进行编码和运动估计等. 其中分割是最大的难题之一.

本文提出一种利用转移帧差 (DFD) 的高阶统计特性和数

学形态学算子的视频对象自动分割方法, 这种方法是文献 [2] 方法的改进. 提出一种简单而高效的基于查找人头肩部轮廓最大转折点的头肩分离算法. 用软件实现了一种基于 MPEG-4 的视频编码系统. 提出一种面向对象分配带宽 (OOBA) 的极低比特率视频编码器, 它把更多比特分配给主观感兴趣的对象, 以提高编码图像的主观视觉质量.

2 对象分割

2.1 帧间变化的检测

这种分割方法的基本思想是从静止的图像背景中隔离出运动对象. 首先进行帧间变化的检测, 采用文 [2] 中提出的检测方法. 在文 [2] 中, 把对帧间变化的检测建模为在高斯噪声中检测随机非高斯信号, 采用了四阶矩检测器.

由于对象的运动而产生的两帧图像之间的变化可以反映

在差值图像 d 上. 为了判断像素 (x, y) 是属于运动对象还是静止背景, 先在差值图像 d 上计算以 (x, y) 为中心的矩形窗口 (x, y) (窗口大小为 $N \times N$) 的四阶矩

$$\hat{m}_d^{(4)}(x, y) = \frac{1}{N^2} \sum_{(s, t)} (d(s, t) - \hat{m}_d)^4 \quad (1)$$

其中, $\hat{m}_d(x, y)$ 是窗口内差值信号的平均值, 即

$$\hat{m}_d(x, y) = \frac{1}{N^2} \sum_{(s, t)} d(s, t) \quad (2)$$

接着, 把 $\hat{m}_d^{(4)}(x, y)$ 同阈值相比较, 阈值同噪声方差 \wedge_{od}^2 的平方成正比, 若大于阈值则像素 (x, y) 属于运动对象, 反之属于背景, 用式子表示如下:

$$\hat{m}_d^{(4)}(x, y) \geq c \cdot (\wedge_{od}^2)^2 \quad (3)$$

其中, c 是由实验确定的常数, H_0 表示假设“ (x, y) 点属于静止背景”, H_1 表示假设“ (x, y) 点属于前景运动对象”. 噪声方差 \wedge_{od}^2 由下式计算得:

$$\wedge_{od}^2 = \frac{1}{N_\mu^2} \sum_{(s, t)} (d(s, t) - \hat{m}_d)^2 \quad (4)$$

其中, μ 是属于静止背景的一块区域, 一般在图像的边沿地带选取, N_μ 为窗口尺寸.

2.2 预分割模板的提取

在实验中发现, 有些场景 (如视频会议) 中对象的运动较为缓慢, 而且并不是整个对象做平移运动, 可能只在某些边沿处发生形状的扭曲运动, 如果只根据某两帧图像的差值进行检测, 并不能得到完整的有意义的对象区域, 如图 1 所示. 所以, 为了增加健壮性, 本文采用了如下做法.

假设一段视频序列含 N 帧图像 ($f_0 \sim f_{N-1}$). 计算第 i 帧和第 $i+m$ 帧图像的差值 $d = f_i - f_{i+m}$ ($0 \leq i, i+m \leq N-1$), 接着对 d 进行四阶矩检测, 可以得到一个二值图像 (模板) M_i . 若 $M_i(x, y) = 0$ 则像素 (x, y) 属于静止背景; 若 $M_i(x, y) = 1$ 则像素 (x, y) 属于运动对象. 其中 m 是根据对象运动快慢确定的一个常数. 计算 $i=0, m+1, 2m+1, \dots, N-m-1$ 时的一系列模板, 把这些模板进行逻辑或操作得到最后的模板 M (图 2(左上)). 两幅二值模板的或操作定义为:

$$OR(M_i, M_j)(x, y) = \begin{cases} 0, & \text{if } M_i(x, y) + M_j(x, y) = 0 \\ 1, & \text{else} \end{cases} \quad (5)$$

为了消除小区域的影响, 或操作前用数学形态 $open$ 、 $close$ 算子^[3]来消除各个模板上可能具有的小区域. 从图 2(左上)中可以看到, 在有些场景中, 由于对象的运动更多地集中在头部和轮廓边沿处, 使产生的模板内部有洞, 但是模板最外的轮廓线已大体包围住了对象 (图 2(右上)). 通过提取模板最外的轮廓线, 就可得到一个完全覆盖对象的不含洞的模板 (图 2(左下)), 但是这个模板还不能精确地表示对象的形状.

表示对象的形状.

2.3 对模板的腐蚀

经灰度变化检测后, 有些由于对象移动而露出的背景也

会被当做运动对象, 为了消除这些背景, 文献[2]通过在阈值化后的帧间差值四阶矩图上进行块匹配搜索的方法来估计变化像素的位移, 如果位移不为 0, 则判定该像素属于运动对象, 反之属于背景. 这种方法有两个缺点: 一是搜索过程花时间, 二是无法精确得到对象的轮廓. 所以, 本文采用了简单而快速的形态学腐蚀^[3]的方法来得到对象形状的精确定义.



图 2 对象分割的过程. 左上: 检测后得到的模板; 右上: 模板边界包围对象; 左下: 预分割后得到的二值图; 右下: 腐蚀后的二值图

腐蚀从图 2(左下)所示模板的最外边沿开始向内进行直到对象的边沿处. 对模板的一次腐蚀过程描述如下.

```
for (h=0; h<image.height; h++)
for (w=0; w<image.width; w++){
if (M(w, h) == 1){
if (erosion(M(w, h)) == 0) and ((x, y) is not a edge pixel)
M(w, h) = erosion(M(w, h));
}
```

不断重复上述过程, 直到没有像素被腐蚀掉为止. 这是一个迭代过程, 迭代的次数粗略取决于模板轮廓线到对象轮廓线之间的最大距离, 基本上经过 5~8 次迭代就可得到精确的对象表示. 为了判定某个像素是否为边沿像素, 在腐蚀之前需要先对原图像进行边沿检测和阈值化. 图 2(右下)是腐蚀后得到的二值分割图.

3 头肩分离

脸部在低比特率和基于对象的图像编码中是重要的区域. 在模型基图像编码^[4]中, 为了提取脸部特征, 定位脸部器官和匹配模型, 有时需要先进行人头部与肩部的分离^[5]. 这里提出一种简单然而高效的基于查找头肩轮廓最大转折点的头肩分离算法. 由于头肩部具有对称性, 所以只要查找头部与肩部的一条水平分界线 (称为分离线) 即可.

3.1 形状近似

首先, 用多边形近似法^[6]来近似对象的轮廓. 形状近似的目的有两个: 一是避免轮廓细节对分离结果的影响; 二是为了减少形状编码所消耗的比特数. 多边形近似的结果见图 3.

3.2 查找最大转折点

把头肩像的宽度投影在 x 轴上 (图 4), 即以头肩像的高度为自变量, 以宽度为应变变量建立函数 $f(x)$. 要分离头肩像即相当于找到图 4 中具有最大转折率的 A 点, 用 A 对应的水平线来分离头肩.

要判断某点是否具有最大的转折率, 只要考察这一点



图 1 只根据某两帧图像的差值进行检测的结果 (noise)



图3 多边形形状近似(左:原形状;右:近似后的形状)

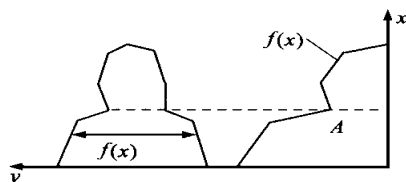


图4 宽度投影示意

+x方向的导数(正导数)与-x方向的导数(负导数)的差是否最大.对导数的计算采用了插值型的三点求导公式^[7].设已给出 $f(x)$ 上的三个点 $x_0, x_1 = x_0 + h, x_2 = x_0 + 2h, (h > 0)$,这三个点的二次插值函数为:

$$P_2(x) = \frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)}f(x_0) + \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)}f(x_1) + \frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)}f(x_2) \quad (6)$$

令 $x = x_0 + th$,上式可表示为

$$P_2(x_0 + t \cdot h) = \frac{1}{2}(t-1)(t-2)f(x_0) - t(t-2)f(x_1) + \frac{1}{2}t(t-1)f(x_2) \quad (7)$$

等式两边对 t 求导,得到

$$P_2'(x_0 + t \cdot h) = \frac{1}{2h}[2(t-3)f(x_0) - (4t-4)f(x_1) + (2t-1)f(x_2)] \quad (8)$$

由于 $h > 0$,所以把 $t=0$ 代入上式即得到 x_0 点的正导数,同时令 x 取代 x_0 则得到 $f(x)$ 上任意一点的正导数公式

$$f(x)^+ = \frac{1}{2h}[-3f(x) + 4f(x+h) - f(x+2h)] \quad (9)$$

同理,不难得到 $f(x)$ 的负导数公式

$$f(x)^- = -\frac{1}{2h}[-3f(x) + 4f(x-h) - f(x-2h)] \quad (10)$$

定义某一点的转折率为

$$K = |f(x)^+ - f(x)^-| \quad (11)$$

最大转折点 A 的 x 坐标即 K 为最大值时所对应的 x 值,即

$$x = \arg \max_x (K) \quad (12)$$

查找到最大转折点后,用它所对应的水平线即可分离头肩.

4 基于 MPEG-4 的视频编码器

本节介绍用软件实现的一种基于 MPEG-4 的视频编码系统. MPEG-4 基于对象的视频编码器能够对任意形状的视频对象进行编码,它基本上由对象分割、形状编码、运动编码、纹理编码等模块构成^[8].

对象分割采用上面讨论的方法.形状编码采用上面提到

的多边形近似法,来减少形状信息的比特开销.运动估计方法要求能对任意形状的图像区域进行,所以对传统的块匹配算法进行扩展.当进行块匹配搜索时,需要根据绝对差值和(SAD)最小的原则在前一帧重建的图像上搜索匹配的块.现在由于对象的形状不一定是矩形,对象区域以外的像素(即透明像素)是不存在的,而搜索的范围可能超出对象区域以外,所以在搜索之前需要用某种像素值填充透明像素,把对象填充成一矩形区域.采用“外推法”进行填充,即用对象区域内所有像素的平均值来填充透明像素.当对边沿块(既包含对象区域内的像素又包含透明像素的块)进行估计时,对传统的 SAD 准则做些修改,即 SAD 的计算只对块中位于对象区域内的像素进行.形式化表示如下:设集合 E 表示块内的像素的集合,集合 F 表示对象区域内的像素的集合,则传统的 SAD 准则为:

$$SAD = \sum_{(x,y) \in E} |original(x,y) - decoded.previous(x,y)| \quad (13)$$

修改后的 SAD 准则为:

$$SAD = \sum_{(x,y) \in E \setminus F} |original(x,y) - decoded.previous(x,y)| \quad (14)$$

运动估计/预测的其它过程类似基于帧的视频编码器.对纹理编码采用 DC-SA-DCT 法^[9].形状自适应 DCT(SA-DCT)最早由 Sikora^[10~12]等人在 1995 年提出,它可以对任意形状的图像区域编码. Kauff^[9]等人于 1998 年发表了对 SA-DCT 的改进方法,称之为“基于块的 DC 隔离和 DC 纠正的 SA-DCT(DC-SA-DCT)”.在我们的视频编码系统中采用了这种改进后的 DC-SA-DCT 纹理编码方法.

5 面向对象分配带宽的极低比特率视频编码器

针对可视电话的应用领域,提出一种采用面向对象分配带宽(OOBA)技术的极低比特率视频编码器.

5.1 概述

如何提高极低比特率视频编码的图像质量是许多人感兴趣的研究课题.在极低的速率下,不应只追求图像客观评价准则(PSNR)的高低,而应充分利用人的视觉特性来提高图像的主观视觉质量.考察人的视觉特性发现:(1)人的注意力更多地集中在运动部分;(2)人的注意力更多地集中在主观感兴趣的.在可视电话的场景中,人的脸部是运动剧烈,观众最感兴趣的部分,肩部则稍次,而背景则经常被忽略.所以,把场景分为三个对象(头、肩和背景),根据重要性程度的强弱分配不同的比特数.已有的一些面向对象分配比特的方法^[13]大多是在一帧图像中根据不同的宏块通过微调量化参数(QP)来实现的,这种调节是有限的,是帧层次上的比分配.与之不同,本文采用的做法是在编码之初就给每个对象序列分配不同的带宽,每个对象序列独立地进行速率控制,这是序列层次上的比特分配.这种方法的优点在于具有面向对象的伸缩性,可以在更大程度上把比特分配给感兴趣的对象,可以更加灵活和更大范围地控制不同对象的图像质量,更加灵活地在不同对象的图像质量之间进行折衷.为了较为合理地分配带宽,提出一种以“带宽/像素(bit/s pel)”为评判准则的带宽分配策略.在比特的组织和传输仍以帧为单位传输图像,故无需编码

形状信息. 简言之, 即传输以帧为单位, 编码和速率控制以对象为单位.

5.2 对象定义

首先, 采用本文前面讨论的方法分割出头肩部, 然后用最少数目的宏块覆盖头肩部(图 5). 接着, 进行头肩隔离得到头肩的分离线, 位于分离线以上(包括分离线上)的宏块定义为头部对象, 分离线以下的宏块定义为肩部对象, 从而把场景分为图 5 所示三个对象. 每个对象以宏块为单位构成, 对象之间的区域不互相覆盖.



图 5 头、肩和背景三个对象的定义

5.3 带宽分配

本文认为给对象分配带宽至少应该考虑四个因素: 对象面积大小、对象空域细节的丰富程度、对象时域上的变化程度、希望对象达到的图像质量. 由于对象的大小是不同的, 所以用“带宽/像素 (B/pel 或 bit/s pel)”来做为分配的准则是合适的. 由于实时编码时并不知道后面要编码的图像是什么样的, 所以要事先对空域细节和时域变化这两个因素做出判断是比较困难的. 当然可以先存储一段时间内的图像序列进行分析后再分配带宽, 但这样做复杂而耗时, 对这种方法还需进一步的研究. 而对第四个因素, 要做出精确的定量分析也是困难的. 所以, 对后三个因素需要考虑主观判断. 本文采用了一种简单的比值加权法.

假设分别用 VO_h 、 VO_s 、 VO_b 表示头、肩和背景三个视频对象, 用 B_h 、 B_s 、 B_b 表示分配给每个对象的带宽, 用 S_h 、 S_s 、 S_b 表示每个对象的面积 (即所包含的像素数目或宏块数目). 每个对象的 B/pel 比值如下:

$$(B/pel)_h : (B/pel)_s : (B/pel)_b = W_h : W_s : W_b \quad (15)$$

其中 w 为加权因子, 实验中, $W_h : W_s : W_b = 5 : 2 : 1$. 各个对象的带宽比为:

$$B_h : B_s : B_b = W_h S_h : W_s S_s : W_b S_b \quad (16)$$

在给出了编码总带宽 (速率) 的情况下, 根据上式给对象分配带宽. 每个对象独立地进行速率控制.

6 实验结果和讨论

6.1 对象分割与头肩分离

采用 QCIF (176 × 144) 的 Claire 序列进行实验. 分割中采用的参数如表 1 所示, 参数的取值通过实验确定. 形态学的腐蚀算子采用 3 × 3 的矩形窗口作为结构元素. 分割结果见图 6. 头肩分离中参数 h 的取值为 5. 形状近似和头肩分离的结果见图 7.

表 1 对象分割采用的参数

N	C	N_μ	N	m
9	25	9	25	5

6.2 基于 MPEG4 的视频编码器

MPEG4 视频编码器可以对任意形状的视频对象进行编码. 在编码端可以把一幅场景分为几个对象, 每个对象独立地进行编码、传输, 每个对象编码后的比特流形成独立的层. 解

码端可以解码某个对象的比特流层, 重建某个对象. 也可以重建所有的对象, 编辑后重建原场景. 图 8 是对象的速率为 24 kbit/s, 帧率为 12.5 fm/s 时的编码结果.



图 6 分割结果 (第 1 帧). 左: 原图像; 右: 分割出的对象



图 7 形状近似和头肩分离的结果 (第 1 帧). 左: 形状近似结果; 右: 头肩分离结果



图 8 MPEG4 编码结果. 第 1 列: 第 13 帧; 第 2 列: 第 21 帧; 第 1 行: 原图像; 第 2 行: 重建的背景对象; 第 3 行: 重建的头肩对象; 第 4 行: 经对象编辑后重建的帧

6.3 面向对象分配带宽的视频编码器

实验仍然采用 Claire 序列, 经对象定义后, 头、肩和背景对象包含的宏块数目分别为 21、23、55, $W_h : W_s : W_b = 5 : 2 : 1$, 所以 $B_h : B_s : B_b = 105 : 46 : 55$. 图 9 是速率为 8 kbit/s, 帧率为 15 fm/s 时传统的未采用 Ooba 和本文提出的采用 Ooba 的两种方法编码重建图像的比较. 前者平均 PSNR 为 32.81 dB, 后

者平均 PSNR 为 32.23dB,后者略有下降,但是从图 9 可见后者的图像质量要好于前者.



图 9 解码结果比较(Claire,8kbit/s,第 142 帧).

左上:基于帧的编码器解码图像;右上:左上图的脸部放大图;
左下:OOBA 编码器解码图像;右下:左上图的脸部放大

7 结束语

在基于对象的视频编码中,视频对象的分割是关键技术之一.本文提出一种利用帧间差值的高阶统计特性和数学形态学算子的视频对象自动分割方法,提出一种基于查找头肩轮廓最大转折点的头肩分离算法,实验表明了上述算法的有效性.实现了一种基于 MPEG-4 的视频编码软件系统.提出一种采用面向对象带宽分配技术的极低比特率视频编码器,该编码器与传统的基于帧的视频编码器相比,能提供更佳的主观视觉质量.

参考文献:

- [1] Sikora T. The MPEG-4 video standard verification model [J]. IEEE Trans CSVT,1997,7(1):19-31.
- [2] Neri A et al. Automatic moving object and background separation [J]. Signal Processing,1998,66(2):219-232.
- [3] Salembier P. Morphological multiscale segment for image coding [J]. Signal Processing,1994,38(3):359-386.

- [4] 李海波.模型基图像编码[J].通信学报,1993,14(2):69-77.
- [5] Hu Tianjian,Cai Dejun. A method for head-shoulder segmentation and human facial feature position [J]. Journal of China Institute of Communications, 1998,19(5):28-33.
- [6] Ballard D H,Brown C M 著,王东泉,徐心平,赵经伦译.计算机视觉[M].北京:科学出版社,1987.
- [7] 李庆扬,王能超,易大义编.数值分析[M].武汉:华中理工大学出版社,1986.
- [8] 翁南钊,蔡德钧.基于内容的视频编码系统[J].无线电工程,1999,29(1):41-45.
- [9] Kauff P,Sch Üß K. Shape-adaptive DCT with block-based DC separation and DC correction [J]. IEEE Trans CSVT,1998,8(3):237-242.
- [10] Sikora T,Makai B. Shape-adaptive DCT for generic coding of video [J]. IEEE Trans CSVT,1995,5(1):59-62.
- [11] Sikora T. Low complexity shape-adaptive DCT for coding of arbitrarily shaped image segments [J]. Signal Processing: Image Communication, 1995,7(4-6):381-395.
- [12] Sikora T,Bauer S,Makai B. Efficiency of shape-adaptive transforms for coding of arbitrarily shaped image segments [J]. IEEE Trans CSVT, 1995,5(3):254-258.
- [13] Lin C,Wu J. Content-based rate control scheme for very low bit-rate video coding [J]. IEEE Trans CE,1997,43(2):124-133.

作者简介:



翁南钊 1975 年出生,1997 年 7 月于福州大学无线电工程系获学士学位,2000 年 6 月于华中理工大学电信系获工学硕士学位.主要研究领域为:视频编码、图像处理、多媒体通信、无线通信等.



蔡德钧 1938 年 11 月出生,华中理工大学电子与信息工程系教授,中国电子学会高级会员,主要研究领域为随机信号处理、图像压缩及多媒体通信.

(上接第 113 页)

参考文献:

- [1] B. M. Macq,J. J. Quisquater. Cryptology for digital TV broadcasting [J]. Proc. IEEE,1995,83(6):954-957.
- [2] C. T. Hsu,J. L. Wu. Hidden digital watermarks in images [J]. IEEE Trans. image Processing,1999,8(1):58-68.
- [3] C. T. Hsu,J. L. Wu. Multiresolution watermarking for digital images [J]. IEEE Trans. Circuits Syst. II,1998,45(8):1097-1101.
- [4] N. B. Nill. A visual model weighted cosine transform for image compression and quality assessment [J]. IEEE Trans. COM,1985,33(6):551-557.

- [5] S. H. Tan,K. N. Ngan. Classified perceptual coding with adaptive quantization [J]. IEEE Trans. Circuits Syst. for Video Technology,1996,37(6):375-388.
- [6] 张元亮,郑南宁.基于感知度量的分形编码[J].中国图像图形学报,1998,3(9):721-726.
- [7] 丁玮,齐东旭.数字图像变换及信息隐藏与伪装技术[J].计算机学报,1998,21(9):838-843.