第21卷第4期
2002年8月

红外与毫米波学报
J. Infrared Millim. Waves

Vol. 21, No. 4
August, 2002

基于双运动模型的头-肩目标的分割

林 涛 胡 波
(复旦大学电子工程系,上海,200433)

摘要 提出了一种新的头-肩目标分割方法,采用双运动模型,可以更好地描述头-肩运动,文章导出了该模型的参数估计算法,并将其运用于半自动分割系统中.实验结果表明:使用双运动模型的分割系统,可以更准确地分割头-肩目标.

关键词 目标分割,运动模型.

HEAD-SHOULDER OBJECT SEGMENTATION BASED ON BI-MOTION MODEL

LIN Tao HU Bo
(Department of Electronic Engineering, Fudan University, Shanghai 200433, China)

Abstract A new method of segmenting head-and-shoulder object was presented by using the bi-motion model, which is better in describing such objects' motion than single motion model. An algorithm was developed to estimate the model parameter and implemented in an semi-automatic segmentation system. The simulated result reveals that the segmentation system using the authors' bi-motion model can segment head-and-shoulder object more accurately.

Key words object segmentation, motion model.

引言

MPEG-4 协议代表着第二代图像编码协议,MPEG-4[1]已明确规定了形状编码及不规则块的纹理编码,但对于编码前的目标分割问题,协议没有作出规定并给出成熟的算法.同时,MPEG-4 中定义的目标是多媒体框架下可以交互操作的信息体,称之为语义目标,它与人的高级思维相对应,而不是传统意义的、与低层次图像信息相对应的区域.低层次特征定义的目标常常与高级语义目标相矛盾,例如,以色彩为依据的分割往往会把人脸和着有鲜艳衣服的身体划分为两个目标;又如,以运动为依据的分割可能将挥动的双手和移动缓慢的身体区别为不同的目标,这与一个完整的人的目标定义是不同的,这就对目标分割提出了更高的要求.

由于通过低层次信息如灰度、色彩、运动和纹理等很难精确的定义语义目标,所以实现完全的自动分割是困难的:一方面,自动分割算法往往非常复杂,需要精密的参数设置和调整;另一方面,特定的算法往往只适用解决特定的问题,在实施算法前需要对问题有一定的先验知识.

从本质上说,已有的大多数自动分割算法实现的是区域分割,或是某种特定条件下的自动分割.要想将图像或视频分割成具有丰富的语义特征的目标仍然是件富有挑战性的任务.因此,加入某种形式人工参与的分割逐渐被接受,出现了一些半自动的分割算法[2,3]因为只有人清楚的知道什么是"语义",什么是需要得到的目标,而计算机可帮助人得到精确的目标边界,完成大量的基于低层次特征的计算.

本文在实现一个半自动分割系统的基础上,提出了使用双运动模型来描述可视电话中常见的头-肩目标的运动,从而使分割系统能更准确的跟踪目标并进行精确的分割.

1 双运动模型及其模型参数的估计

运动模型是用以描述目标或背景的三维或两维运动的参数模型,常用的模型有六参数仿射模型和八参数透视模型.运动模型估计可以估计全局运动参数,这往往对应摄像机的各种操作和移动;运动模型也可以估计某个目标的运动,又称为局部运动.本

稿件收到日期 2001-06-29,修改稿收到日期 2001-08-22

Received 2001-06-29, revised 2001-08-22

文利用运动模型估计已知目标的运动，作为当前帧中目标分割的起点.

本文中，采用了八参数的透视模型作为运动模型.设(x_i, y_i)是当前帧的第 i 个的点的坐标，(x_i', y_i')是第 i 个点在前一帧对应点的坐标，则八参数透视运动模型[2]可表示为

$$x_i' = (a_0 x_i + a_1 y_i + a_2)/(a_6 x_i + a_7 y_i + 1)$$
$$y_i' = (a_3 x_i + a_4 y_i + a_5)/(a_6 x_i + a_7 y_i + 1), \tag{1}$$

其中 $a_0, a_1, \cdots a_7$ 是运动模型参数.定义两帧的平方误差和为

$$E = \sum_{i=1}^{N} e_i^2 , \tag{2}$$

其中 $e_i^2 = (f_{k-1}(x_i, y_i) - f_k(x_i' y_i'))^2$.

式(2)中 N 表示图像上象素点数或目标上的点数，f_{k-1}和 f_k 分别表示前一帧图像和当前帧图像.模型参数估计(又称运动估计)的目的就是使平方误差和最小化.

本文采用了基于梯度下降的方法[4]来求透视运动模型参数.由于 E 相对于 a 的关系是非线形的，使用迭代方程 $a^{(t+1)} = a^{(t)} + H^{-1}b$ 来求模型参数，其中 $a^{(t)}$和 $a^{(t+1)}$分别在第 t 步和第 $t+1$ 步迭代时的值，H 是 E 的 *Hessian* 矩阵的二分之一，矩阵大小为 $n \times n$，b：n 维的矢量，大小等于 E 的梯度的一半，n：模型参数个数，H 和 b 具体计算如下

$$H_{kl} = \frac{1}{2}\sum_{i=1}^{N}\frac{\partial^2 e_i^2}{\partial a_k \partial a_l} \cong \sum_{i=1}^{N}\frac{\partial e_i}{\partial a_k}\frac{\partial e_i}{\partial a_l}, \tag{3}$$

$$b_k = -\frac{1}{2}\sum_{i=1}^{N}\frac{\partial e_i^2}{\partial a_k} = -\sum_{i=1}^{N} e_i \frac{\partial e_i}{\partial a_k}. \tag{4}$$

$\frac{\partial e_i}{\partial a_k}$根据投射模型方程及 e_i 的定义求得.由于利用透视运动模型计算得到的(x_i', y_i')往往不会落在整象素点上，因此在求 $f_k(x_i', y_i')$时需要利用双线形插值.现在，我们可以描述整个迭代算法

1. 初始化 $a = (a_0, \cdots, a_7)$;

2. 对每个在目标上的点(x_i, y_i)找到其对应点(x_i', y_i')，计算相应的误差 e_i 和e_i 相对于$a_0, \cdots, a_7$的偏导，将(x_i, y_i)的贡献叠加到矩阵 H 和矢量 b 中，

3. 利用迭代方程 $a^{(t+1)} = a^{(t)} + H^{-1}b$ 更新参数;

4. 如果误差函数 E 开始增加则停止迭代，否则跳到 2.

运动模型估计描述的是刚体的运动，而实际的景物往往不是纯粹的刚体，我们只能将它们近似的看成刚体.在可视电话中，把头-肩目标看成是单一的目标进行估计将会产生很大的误差，实际上，把头和肩看成两个刚体，情况会有所好转.因此，本文提出采用双运动模型来描述头-肩目标，并在透视运动模型估计算法的基础上提出了双运动模型的估计算法.双运动模型估计算法的目的是找到头-肩目标的分离处并计算两套运动模型参数.

一般而言，头-肩目标中头部的运动较为灵活，往往包括平移、旋转以及前后深度的变化，而肩部的运动相对来说就小得多了，而且往往只包括平移.如图 1 所示，用一条与图像 x 轴平行的直线作为头、肩的分界线.用透视参数 $a^{(1)}$描述头部的运动，$a^{(2)}$描述头部以下的肩的运动.

$$S_{obj}(x, y) = \begin{cases} 1 & \text{if pixel}(x, y) \text{ belongs to object} \\ 0 & \text{otherwise} \end{cases}, \tag{5}$$

$$S_2(x, y) = \begin{cases} 1 & \text{if pixel}(x, y) \text{ belongs to object} \\ & \quad \text{and } y > ydiv \\ 0 & \text{otherwise} \end{cases}, \tag{6}$$

则平方误差函数可表示为

$$E = \sum_{i=1}^{N} S_1(x_i, y_i) e_1^2(x_i, y_i) + \sum_{i=1}^{N} S_2(x_i, y_i) e_2^2(x_i, y_i), \tag{7}$$

其中 e_1 和 e_2 分别是通过 a^1 和 a^1 计算出来的帧间对应点的差，

$$e_1(x, y) = f_k(x', y') - f_{k-1}(x, y),$$
$$e_2(x, y) = f_k(x'', y'') - f_{k-1}(x, y). \tag{8}$$

现在的问题是找到合适的 $ydiv$ 并求出 $a^{(1)}$和 $a^{(2)}$，使得平方误差函数最小.但是，平方误差函数本身与 $ydiv$ 没有解析的关系，因此很难反映 $ydiv$

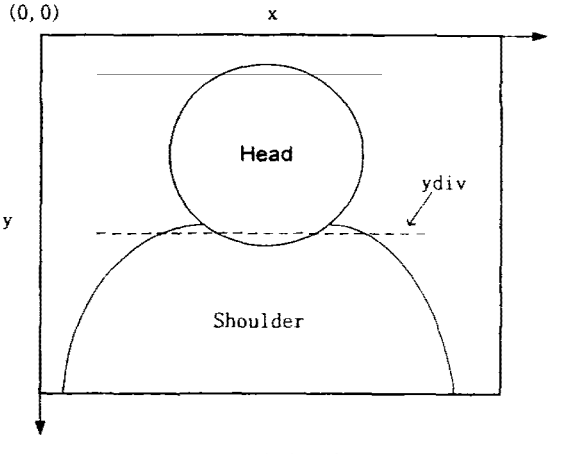

图 1 头肩目标

Fig. 1 Head-and-shoulder object

对其的影响，因此，采用了以下改进的平方误差函数来近似平方误差函数

$$E_m = \sum_{i=1}^{N} g_1(x_i, y_i) e_1^2(x_i, y_i) + \sum_{i=1}^{N} g_2(x_i, y_i) e_2^2(x_i, y_i), \quad (9)$$

其中，

$$g_1(x, y) = \frac{1}{1 + e^{a(y - ydiv)}}, \quad g_2(x, y) = \frac{1}{1 + e^{-a(y - ydiv)}}, \quad (10)$$

当 $y \ll ydiv$ 时，$g_1(x, y) = 1$，$g_2(x, y) = 0$；当 $y \gg ydiv$ 时，$g_1(x, y) = 0$，$g_2(x, y) = 1$；实际上 g_1 和 g_2 将目标分成两个部分，分别用不同的误差模型去描述. 建立了误差函数和 $ydiv$ 之间的关系，即可求出误差函数相对 $ydiv$ 的偏导，再用梯度下降法求得最佳的 $ydiv$.

$$\frac{\partial E_m}{\partial ydiv} = \sum_{i=1}^{N} e_1^2(x_i, y_i) \frac{\alpha e^{(y_i - ydiv)}}{[1 + e^{a(y_i - ydiv)}]^2} + \sum_{i=1}^{N} e_2^2(x_i, y_i) \frac{-\alpha e^{(y_i - ydiv)}}{[1 + e^{a(y_i - ydiv)}]^2}. \quad (11)$$

在上述基础上，提出了双运动模型估计算法

1. 初始化 $ydiv = top + (bottom - top)/2$，其中 top 和 $bottom$ 分别是头－肩目标的上、下边界，$a_{init}^{(1)} = a_{init}^{(2)} = (1,0,0,0,1,0,0,0)$，

2. 固定 $ydiv$，用迭代方程 $a^{k(t+1)} = a^{k(t)} + H^{-1}b$ 分别求出头、肩目标最优的运动参数 $a^{(1)}$ 和 $a^{(2)}$，

3. 固定 $a^{(1)}$ 和 $a^{(2)}$，用迭代方程 $ydiv^{(k+1)} = ydiv^{(k)} - \lambda \frac{\partial E_m}{\partial ydiv}$ 求出最优的 $ydiv^{new}$，

4. 如果 $ydiv^{new} = ydiv$，则退出，否则跳到 2.

算毕后求出最佳的 $ydiv$ 和双运动模型参数 $a^{(1)}$ 和 $a^{(2)}$. 计算表明，双运动模型的估计误差比单运动模型估计误差小. 表 1 比较了两种情况下误差

表 1　单运动模型与双运动模型估计误差

Tab.1　Estimated errors of single motion model and bi-motion model

		Ydiv	绝对值平均误差
Claire 序列	单模型	None	6.67
60、62 帧	双模型	178	5.66

情况及模型参数，采用了 Claire 序列的两帧图像，误差用目标点误差绝对值的平均值表示. 图 2 显示了两种情况下计算得到的残余图像，由此看出，使用双目标模型的残余图像值比使用单目标模型的残余图像值小，说明了双目标模型运动估计算法更能精确描述头－肩目标的运动.

2　语义目标的半自动分割系统

本文实现的语义目标的半自动分割系统是在[2]的基础上发展的针对可视电话中的头-肩目标的分割系统. 该分割系统将待分割的图像序列分成 I 帧和 P 帧，I 帧的分割是在用户的参与下完成的，其中主要由空间多值分水岭算法[5]构成. P 帧的分割是在前帧分割的基础上自动完成的，对于图像序列中运动目标进行运动估计，估计采用了双运动模型估计算法，并根据估计的运动模型参数将前帧分割得到的目标投射到当前帧，最后通过多值分水岭算法完成精确分割.

I 帧分割就是单帧分割，即不用前一帧的分割结果来预测当前帧的分割. 这实际上是静止图像的分割，利用的主要是图像的色彩信息. 用户参与时选择目标轮廓上的关键点，得到目标的大致轮廓，然后通过腐蚀、膨胀工具得到分水岭算法所需的初始标号，在应用多值分水岭算法前，使用[6]中的腐蚀重建的开-闭滤波器简化图像，消除噪声. 分割的流程如图 3 所示.

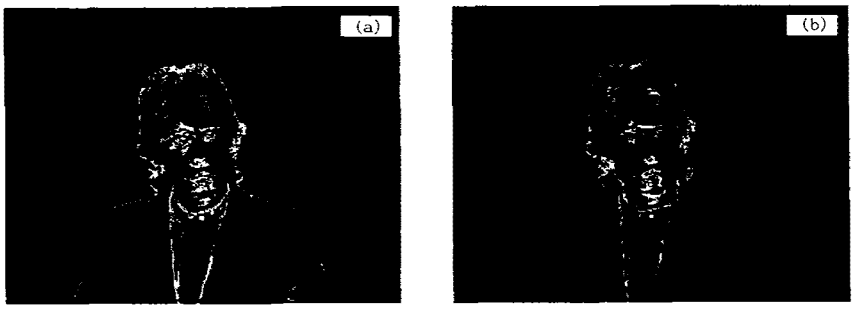

图 2　(a) 单运动模型估计的残余图像，(b) 双运动模型估计的残余图像

Fig. 2 There sidual image of single motion estimation(a) and of bi-motion estimation(b)

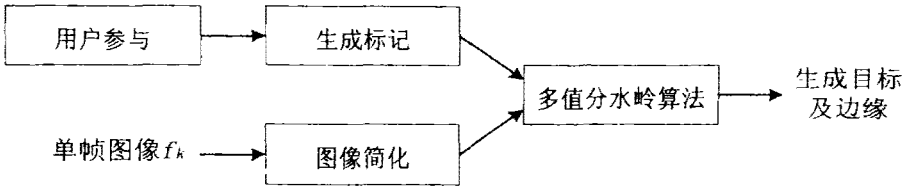

图 3 Ⅰ帧的分割

Fig. 3 Segmentation of Ⅰ frame

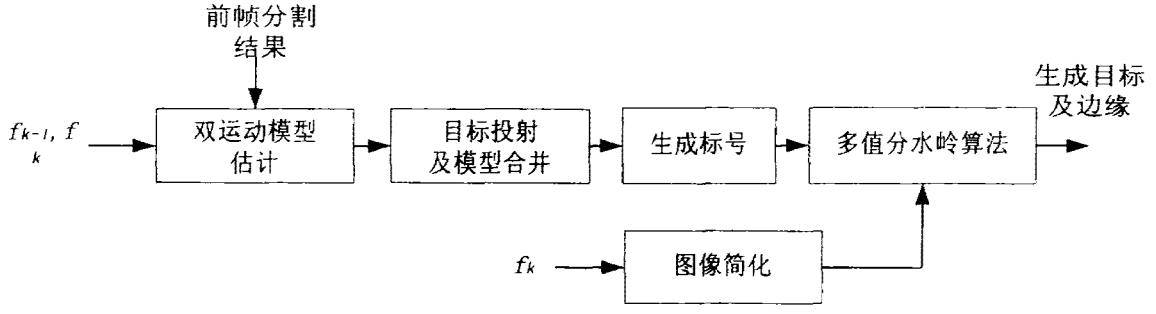

图 4 P 帧的跟踪分割

Fig. 4 Tracking and segmentation of P frame

P 帧是相对于 Ⅰ 帧而言的，P 帧中运动目标的分割不同于 Ⅰ 帧的分割：不需要人工参与，即自动分割，需要前一帧的分割结果. P 帧的跟踪分割过程如图 4 所示，根据特定的分割问题（头-肩目标分割），采用前面的双运动模型估计算法，得到两组运动模型参数，将头-肩投射后，需要将分离的目标合并，以生成单一的目标，然后如 Ⅰ 帧分割所述，生成标号并应用多值分水岭算法来得到精确的目标边缘.

3 模拟结果

我们将展示利用基于数学形态学和双运动模型的半自动分割算法的一些模拟结果. 采用的图是标

图 5 Ⅰ帧的分割，第一排是 Claire 序列的第 50 帧，第二排是 Forman 序列的第 70 帧，从左至右依次为：原始图像、初始标号图像和标有边缘的图像

Fig. 5 Segmentation of Ⅰ frame. In the first row is the 50^{th} frame of Claire sequences. In the second row is the 70^{th} frame of Forman sequences. Images from left to right are original images, images with initial markers and images with boundary

图 6　P 帧的分割结果,显示的是分割后的目标图像,上排从左至右显示了 Claire 序列的第 53、55 和 58 帧,下排从左至右显示 Forman 的序列第 74、77 和 79 帧

Fig. 6　Segmantation of P frame, displayed are segmented object images. From left to right in the upper row are the 53^{th}, 55^{th} and 58^{th} frames of Claire sequences. From left to right in the lower row are 74^{th}, 77^{th} and 79^{th} frames of Forman Sequences

准图像序列 Claire 和 Forman 的 YUV 图像,大小是 CIF(352 × 288),各分量的采样率比例是 4:1:1,因此,在转化成等采样率的 RGB 图像时需要插值.图像序列的频率是 30Hz.两组图像序列在分割时具有不同的难易程度,Claire 图像序列背景简单且固定不变,人的运动范围和幅度也较小,因此分割起来比较容易;Forman 图像序列背景比较复杂,且背景是在不断移动的,人的运动也比较复杂,速度也较快,因此分割起来较难,较易出错.

图 5 显示了 I 帧的分割过程,从左至右依次是始终图像、标有初始标号的图像、分割边缘图像和目标二值图像.Claire 使用的是第 50 帧,froman 使用的是第 70 帧.一般情况下,在目标周围选择 10 ~ 15 个关键点就足够了.

图 6 显示了采用双运动模型估计的 P 帧分割结果,显示的是分割后的目标图像,上排从左至右显示了 Claire 序列的第 53、55 和 58 帧,下排从左至右显示 Forman 的序列第 74、77 和 79 帧.可以看到,分割得到的图像边缘是比较精确的.

图 7 显示了采用单运动模型描述头-肩目标分割的结果,采用与图中的相同的序列,可以看到随着帧数的增加,分割出现了较大的误差,如帽子处和领口处分割都不正确,而在图 6 的右图中这些误差没有出现或不明显,所以采用双运动模型的分割系统更能稳定的分割头-肩目标.

4　结语

本文在实现针对单一目标的半自动分割系统基础上,提出了针对特定的头-肩目标的双运动模型及其模型参数的估计算法,然后将该双运动模型估计

图 7　采用单运动模型的 P 帧的分割结果,3 幅图像依次为 Forman 序列的第 74、77 和 79 帧

Fig. 7　Segmentation results of P frame by using the single motion model. The three images are 74^{th}, 77^{th} and 79^{th} frames of Forman sequences.

算法应用在半自动分割系统中.模拟结果表明,对于可视电话中的头-肩模型,基于双运动模型的半自动分割系统能准确的分割出用户关心的语义目标即头-肩目标,其稳定性和准确度比单运动模型的分割要高,因此对于特定的头-肩目标,采用双运动模型是一种很有效的方法.

REFERENCES

[1] MPEG-4 Video Group. Information technology – genetic coding of audio-visual objects: Part 2 – Visual, Nov. 1998

[2] Chuang Gu, Ming-chieh Lee. Semiautomatic segmentation and tracking of semantic video objects. *IEEE Trans. on CSVT*, 1998, **8**(5): 572—584

[3] Kim Munchurl, Jeon J G. Moving object segmentation in video sequences by user interaction and automatic object tracking. *Image and vision Computing*, 2001, **19**: 245—260

[4] Frédéric Dufaux, Janusz Konrad. Efficient, robust, and fast global motion estimation for video coding. *IEEE Trans. on Image Processing*, 2000, **9**(3): 497—501

[5] Chuang Gu. Multi-valued morphology and segmentation-based coding. Ph.D. dissertation

[6] Philippe Salembier, Montse Pardàs. Hierarchical morphological segmentation for image sequence coding. *IEEE Trans. on Image Processing*, 1994, **3**(5): 639—651