

# 人脸序列图象的头部运动参数估计

刘云海 虞 露 姚庆栋

(浙江大学信息与通信工程研究所, 杭州 310027)

**摘 要** 在人脸序列的图象编码中, 由于模型基编码方法可以获得高的主观图象质量和低的码率, 因而受到广泛重视. 但是, 其运动参数的可靠估计还是一个难点. 为此, 该文分析了头部运动的特点, 并把它分为头部刚体运动、脸部表情的简单运动和脸部表情复杂运动 3 种形式. 其中, 提取头部刚体运动参数利用了基于特征点对的运动参数估计算法, 并提出了一个线性的实现方法; 文中还指出提高运动参数估计的精度在于选择合适的特征点和建立一个和特定人脸相一致的三维线框模型; 另外, 还为脸部表情的简单运动建立了形变矩阵; 最后给出了用面积误差函数评价的运动参数估计误差.

**关键词** 模型基编码 图象分析 图象综合 3D 线框模型 运动参数估计

中图法分类号: TN919.81 文献标识码: A 文章编号: 1006-8961(2000)09-0773-06

## Motion Parameters Estimation in Facial Sequence

LIU Yun-hai, YU Lu, YAO Qing-dong

(Institute of Information and Communication Engineering, Zhejiang University, Hangzhou 310027)

**Abstract** Model-based facial image coding has received extensive attention due to its high subjective image quality and low bit-rates. But the estimation of object motion parameters is still a difficult problem. In this paper, head motion characteristics is analyzed, and it can be classified into three motion types, head rigid motion, simple motion of facial expression and complex motion of facial expression. A motion parameter estimation algorithm based on corresponding feature points is utilized to extract head rigid motion. A linear implementation algorithm is suggested. The key works to improve the accuracy of motion parameter estimation are the selecting of some proper feature points and the building of an exactly 3D wire-frame corresponding to specified face. A deformable matrix is constructed to extract simple motion parameter of facial expression. At last, the error of motion parameter is evaluated by the area-error function.

**Keywords** Model-based coding, Image analysis, Image synthesis, 3D wire-frame model, Motion parameter estimation

## 0 引 言

模型基图象编码方法在 80 年代初提出后, 由于在极低的码率条件下( $\leq 32\text{bits/s}$ )能够获得较好的图象质量, 而得到广泛的研究<sup>[1~3]</sup>. 目前, 主要针对以人肩图象为主要内容的会议电视和可视电话的图象序列进行编码研究. 其主要思想是基于发送和接收双方都有一个共同的有关某一对象的 3D 模型.

在编码过程中, 可通过场景的分析来获得对象的运动参数、色彩参数和形状参数, 而解码端就是利用这些参数来综合出图象. 由于在编码时, 充分利用了物体已有的信息(如色彩), 因而可结合运动参数和更新的参数来编码, 这样不仅大大减少了需要传输的数据量, 且解码的图象也没有传统的块编码所引入的块效应. 但编码的难点是图象分析, 即要得到一个可靠的运动参数估计还是很困难的. 常见的运动参数估计方法可以归纳为基于光流场和基于特征的估

计方法<sup>[4]</sup>. 其中基于光流场的方法计算量很大, 而且对噪声敏感, 也不一定可靠收敛; 而基于特征的方法虽计算量小, 但前后帧的特征点对估计要准确. 在用第二类方法估计脸部的运动参数时, 文献[5]是用人机交互的方法获得特征点对来估计三维刚体的运动参数, 当然, 该方法在实时图象编码过程中就不适用. 文献[2]是用块匹配的方法来找到对应点, 并通过特征点深度的修正, 用迭代的方法估计头部刚体的运动参数. 但是一旦对应点匹配有误差, 其误差就会反馈到模型的深度值上. 同时, 特征点上的深度修正后, 也并不能做到合理修正整个头部的线框模型, 而且还改变了模型原有的一致性. 文献[6]虽然对文献[2]的深度修正进行了改进, 但还是不能保证特定人脸线框模型的一致性. 也就是说, 在人脸序列图象的运动参数跟踪研究中, 必然引起线框模型产生较大的误差, 从而使运动参数的估计也产生更大的误差, 这种误差的反馈, 使系统对误差很敏感. 本文通过对头部序列图象的运动分析, 首先把整个头部的全局运动作为刚体运动模型来考虑, 还建立了一个线性的运动参数估计系统, 并指出要提高运动参数估计的精度应该进一步降低特征点对的估计误差, 也就是要选择一些合适的特征点, 同时提高特定人模型和实际头部几何特性的一致性; 然后把脸部的局部表情变化作为非刚体的运动模型, 再把它们细分为简单的表情运动和复杂的表情运动, 并对简单的表情运动参数估计作了进一步的研究.

## 1 运动模型和运动参数的估计算法

### 1.1 头部运动模型

在笛卡儿坐标系中, 设真实世界任何物体上一点的三维坐标为  $S = (X, Y, Z)^T$ , 运动后的坐标为  $S' = (X', Y', Z')^T$ , 则物体的运动方程是

$$S' = RS + DS + T \quad (1)$$

其中,  $R$  是旋转矩阵,  $D$  是形变矩阵,  $T$  是平移矩阵,  $S$  是物体上点的位置矢量,  $S'$  是运动后该点的位置矢量,  $X, Y, Z$  是三维空间的点坐标, 这是物体运动的通用关系式. 在对人头部的运动特性进行分析后, 即可以发现头部的运动特点. 本文把头部、脸部的运动分为3种情况来考虑: (1) 定义脸(头)部的全局运动是三维刚体运动; (2) 局部表情区的简单运动, 如下巴区域的运动可以用简单的平移运动模型来表示; (3) 局部表情区的复杂运动, 如眼睛, 嘴巴是非刚

体的运动.

### 1.2 基于特征点对的头部刚体运动参数估计

由于刚体运动中没有形变运动, 所以物体运动的通用关系方程中形变矩阵  $D$  为零. 假设某时刻  $t$ , 刚体上的一点  $S = [X, Y, Z]^T$ , 在时刻  $t'$ , 运动到  $S' = [X', Y', Z']^T$  位置, 则用旋转矩阵  $R$  和平移矩阵  $T$  表示为  $S' = RS + T$ , 即

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix} \quad (2)$$

在运动量不大时, 旋转矩阵可以用转动的欧拉角表示为

$$R = \begin{bmatrix} 1 & -\Delta\phi & \Delta\varphi \\ \Delta\phi & 1 & \Delta\theta \\ -\Delta\varphi & \Delta\theta & 1 \end{bmatrix} \quad (3)$$

其中  $\Delta\theta, \Delta\varphi$  和  $\Delta\phi$  分别是关于  $X, Y$  和  $Z$  轴的逆时针方向角度位移. 根据正交投影模型的理论, 真实物体上的点坐标  $(X, Y, Z)$  与投影到图象平面的点坐标  $(x, y)$  的关系为  $x = X, y = Y$ , 若该点经运动后, 投影到图象平面的点为  $(x', y')$ , 则可以得到如下关系式

$$\begin{cases} x' = x - \Delta\phi y + \Delta\varphi z + T_x \\ y' = y - \Delta\phi x + \Delta\theta z + T_y \end{cases} \quad (4)$$

如果选择  $N$  个特征点, 那么就可以得到  $2N$  个方程, 然后就可利用最小二乘法来求解 5 个运动参数. 设方程误差的平方和  $\omega$  为

$$\omega = \sum_i ((x'_i - x_i + \Delta\phi y_i - \Delta\varphi z_i - T_x)^2 + (y'_i - y_i + \Delta\phi x_i - \Delta\theta z_i - T_y)^2) \quad (5)$$

其中, 累加的下标  $i$  为  $1, 2, \dots, N$  个特征点对. 然后即可通过求  $\omega$  对 5 个变量  $\Delta\phi, \Delta\varphi, \Delta\theta, T_x, T_y$  的偏导数为零, 来得到如下 5 个方程的方程组

$$\begin{cases} \Phi \sum_i (x_i^2 + y_i^2) - \Phi \sum_i Z_i y_i + \theta \sum_i Z_i x_i - T_x \sum_i y_i - T_y \sum_i x_i + \sum_i (x'_i y_i + y'_i x_i) = 0 \\ \Phi \sum_i y_i Z_i - \Phi \sum_i Z_i^2 - T_x \sum_i Z_i + \sum_i Z_i (x'_i - x_i) = 0 \\ \Phi \sum_i x_i Z_i - \theta \sum_i Z_i^2 - T_y \sum_i Z_i + \sum_i Z_i (y'_i - y_i) = 0 \\ \Phi \sum_i y_i - \Phi \sum_i Z_i - N T_x + \sum_i (x'_i - x_i) = 0 \\ \Phi \sum_i x_i + \theta \sum_i Z_i - N T_y + \sum_i (y'_i - y_i) = 0 \end{cases} \quad (6)$$

通过求解上述方程组, 即可以得到 5 个运动参数. 虽然, Aizawa 等人提出的两步迭代法<sup>[2]</sup>, 及后来

Bozdagi 等人的改进算法<sup>[6]</sup>, 均试图通过迭代的方法来修正特征点对的深度, 以获得精确的运动参数估计. 本文认为该方法至少存在以下几方面的缺陷:

(1) 由于在修正特征点的深度值时, 事实上也把特征点对的估计误差转移到深度值上了, 因此虽然运动参数能收敛到某一稳定值, 但并不保证深度修正得更准确.

(2) 在局部特征点的深度值修正后, 由于在线框模型上, 仍没有合理的方法来修正特征点周围的深度值, 因此会破坏线框模型表面的平滑特性.

(3) 经过长序列的运动跟踪后, 模型变得不稳定.

所以, 我们认为在模型基图象编码中, 用特征点对的方法来估计头部的三维刚体运动参数时, 应着重于如何得到一个反映不同人特性的三维线框模型, 以及如何选择特征点, 以使在匹配特征点对时误差最小.

### 1.3 局部表情区的简单运动

众所周知, 一般头部的全局运动可以作为刚体运动来考虑, 而脸部的局部表情变化则是非刚体运动. 通过观察, 可以发现脸部表情的变化在有些区域比较复杂(如眼睛和嘴巴), 而下巴区域的表情变化则相对简单, 主要是嘴巴张大时, 下巴点向下伸张, 嘴巴闭合时, 下巴点向上收缩. 这部分区域的运动可称之为局部表情区的简单运动, 如图 1 所示的嘴巴带动下巴的运动.



图 1 嘴巴开闭带动下巴的运动

为便于叙述, 本文把下巴区域定义为鼻子以下不包括嘴巴的区域, 它包括了局部的脸颊区域, 因为嘴巴两边的脸颊运动也可以近似为和下巴一致的伸缩运动. 因此下巴的形变运动可以近似作为沿头部中心轴的伸缩运动, 可用式(7)的形变矩阵来描述.

$$D = \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (7)$$

形变系数  $\alpha$  和  $\beta$  是简单的比例参数. 例如  $\alpha(x, y, z) = k(x - x_0)$ ,  $k$  就是一个比例系数,  $x_0$  是鼻子与定义下巴区域相交界的参考点. 如果头部垂

直, 则形变只有垂直方向( $Y$  方向), 即  $\alpha$  为零, 反之,  $X, Y$  方向都有形变. 由于  $D$  是与空间位置相关的, 因此形变运动  $D$  的估计就是根据下巴到鼻孔距离按比例伸缩(如图 2 所示), 根据运动前  $AB$  点距离和运动后  $ab$  点的距离, 就可以估计  $X$  和  $Y$  方向的伸缩尺度.

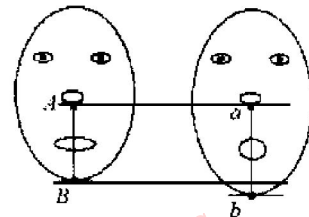


图 2 下巴点的形变(伸缩)运动

### 1.4 局部表情区的复杂运动

在模型基图象编码中, 运动参数估计的目的是希望通过运动参数用以前的图象来综合当前的图象, 以便进一步降低码率. 但由于在脸部表情变化中, 眼睛和嘴巴表情的运动比较复杂, 因此很难用闭着的眼睛通过运动参数综合出眼睛睁开的图象来, 因而本文把它们运动称之为局部表情区的复杂运动. 虽然可以用前面介绍的形变模板方法来跟踪眼睛和嘴巴变化, 但目前还没有很好的方法能够综合出逼真的眼睛、嘴巴等的表情来. 本文是采用表情粘贴的方法来重建眼睛和嘴巴的图象, 所以在此不讨论它们的运动参数估计问题. 现在看来, 用存储各种状态的眼睛和嘴巴图象的模板, 结合形变模板所提取的形状参数, 再加上简单的运动调节应该是解决这一复杂运动的较好途经.

## 2 头部线框模型的建立

在运动参数估计中, 首先要把真实世界物体的运动映射到该物体的线框模型上来, 然后才能从线框模型上获得物体的深度信息, 所以一个准确反映物体几何特性的线框模型, 对精确估计运动参数是很重要的, 如果模型和实际头部深度参数误差较大, 则很难得到准确的运动参数估计. 特定人头部线框模型的建立是通过一些参数来调节通用的线框模型而获得的. 本文用文献[7]、[8]介绍的特征提取算法来分析得到头部轮廓、眼睛、嘴巴、鼻孔的位置参数, 再用这些参数调节得到的线框模型就能较精确地反映特定人头部的几何特征. 图 3 是通用的线框模型, 图 4 是用特征参数来调整通用模型所得到的特定人的线框模型.

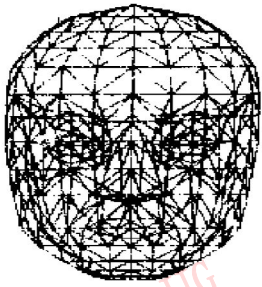


图3 通用3D线框模型 图4 特定人的3D线框模型

### 3 脸部特征点及特征点对的提取方法

#### 3.1 脸部特征点的提取

在用特征点对的方法来估计刚体的运动参数时,首先需要在前后帧寻找对应的特征点,然后通过特征点在二维图象平面的移动,来估计三维的运动参数.早期的头部运动估计就是在脸部做一些标记点,如贴上白色的标记物,随着头部的运动,在序列图象上就可以标记出这些白点的运动轨迹,从而估计出头部的三维运动.如果不用人工参与的方法来估计运动参数,则需要自动选择一些特征点,但这些特征点的选择应该是便于在后续帧中准确地搜索到相对应的特征点,所以特征点选择不应该选在亮度平坦的区域,而应该选在边沿的位置,最好是边沿的端点,这样在用块匹配法来找对应点时,才有较好的结果.本文把特征点的选择表示为边沿、亮度和端点的函数.并且,特征点不能选在表情敏感区,也就是特征点不能落在眼睛、嘴巴的这些表情变化较大的区域.因此,特征点选择函数可定义为

$$p(x, y) = f(e(x, y), g(x, y), n(x, y)) \quad (8)$$

式中,  $e(x, y)$  是边沿能量函数,该能量函数可用下述方法求得,即先用 Robert 或 Sober 算子来求得一幅边沿图象,再选择合适的阈值  $T$  将该图象二值化得  $b(i, j)$ , 则  $e(x, y)$  的定义如下

$$e(x, y) = \sum_{i, j} b(i, j) \quad (9)$$

$g(x, y)$  是亮度函数,定义为在一个方块内的平均亮度;  $n(x, y)$  是端点判决函数.如果边沿能量值主要集中在某一区域,则确定为端点.反之,若分布在二个以上区域则认为不是端点.为此,本文定义了计算边沿能量的8个区域(如图5所示).边沿能量的计算是首先以当前的边沿点为中心,把一个  $N \times N$  的块划分8个区域,  $N$  取奇数,并计算每个区域边沿的能量值,然后以每个区域的能量分布作为判决的

条件.在图5中,以  $p_1$  为当前点,先在1~4区域中计算边沿能量,若其主能量集中在一个区域,则  $p_1$  是端点.反之,则在5~8区域计算边沿能量,并判决能量是否集中在一个区域.图5中  $p_2$  在1~4区域或5~8区域能量都不集中在一个区域,因而不是端点.而且边沿点需在满足端点条件,且亮度大于某一阈值时,才选为特征点.根据亮度阈值和边沿阈值的不同,可以得到一组不同的特征点位置.图6上白十字点是在亮度阈值为80,边沿阈值  $T$  为30时所得到的不在表情区域内的特征点.

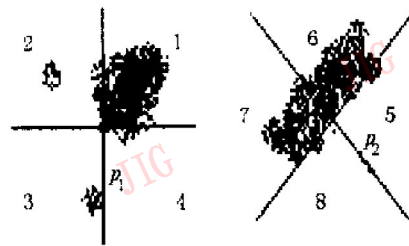


图5 8区域及能量分布示图



图6 选择得到的特征点

#### 3.2 特征点对的确定

根据2.3节描述的方法,在得到第一帧图象的特征点后,就可以用块匹配的方法求出这些特征点在后续帧中的对应点;再利用这些对应点的位移关系,就可以确定运动的参数.块匹配方法有全搜索、三步搜索等方法,但为了能有较好的匹配结果,这里采用了全搜索的方法.根据实验测试,本文选择以特征点为中心的一个  $15 \times 15$  的方块来进行实验,匹配结果比较准确、稳定.匹配的搜索范围也选在  $15 \times 15$  的块内,也就是上、下、左、右最大位移为7个像素,这对运动量比较小的脸部序列图象来说是合理的;匹配搜索的顺序是从内向外进行.

### 4 实验结果和结论

本文选用 Akiyo 序列图象进行实验,首先提取脸部特征参数,再利用这些参数调整通用的线框模型,得到 Akiyo 的三维线框模型.其头部的刚体运动参数用特征点对的方法来估计,而表情的简单运动则是通过跟踪下巴到鼻孔的距离变化,来得到伸缩运动的形变矩阵.在采用特征点对的方法来估计点的运动参数时先要在脸部区域搜索合适的特征点,并与后续帧相应的特征点匹配,组成特征点对.通过实验可以发现,若能选择得到合适的特征点,则

对应点的匹配结果就较准确, 反之, 特征点选择不合适, 对应点匹配结果就差. 那么, 如何确定匹配结果的准确性以及它对运动参数估计的影响呢? 在实际的图象序列中, 不太可能获得一个准确的参考值供比较. 为此, 我们根据模型基编码特点, 创建了面积误差函数<sup>[9]</sup>. 用以评价特征点选择、匹配和运动参数估计的准确性. 面积误差函数就是采用估计得到的运动参数来恢复当前帧图象, 然后用实际图象和当前图象比较, 来确定误差象素的多少. 显然, 如果特征点对匹配准确, 且运动参数估计也准确, 则两幅图象应该是一样的, 即面积误差是很小的; 反之, 面积误差就大. 如图 7 是 Akiyo 第 14 和 15 帧的原图象和恢复图象; 图 8 显示了选择不同的特征点对面积误差的影响. 从两图比较可见图 8(a) 选择的特征

点比较合适, 其中 80/35 (16fp) 的曲线有较低的面积误差, 它表明在亮度为 80, 边沿为 35, 16 个特征点的情况下, 结果较优. 图 8(b) 是一组不太理想的特征点选择结果, 面积误差就比较大, 且比较离散, 不太稳定.

本文针对人脸序列图象头部运动的特点, 把头部运动划分为全局的刚体运动、局部表情的简单运动和局部表情的复杂运动, 并对运动参数作了估计. 其中对刚体运动的估计提出了一个线性的方法, 并指出提高运动参数估计精度的最好途径是选择合适的特征点以及得到和实际头部相一致的三维线框模型, 另外还给出了特征点选择函数. 实验结果表明, 本文为模型基图象编码中的运动参数提取提供了一种准确、可靠的估计方法.



图 7 Akiyo 第 14、15 帧的结果, 左是原图象, 右是恢复的图象

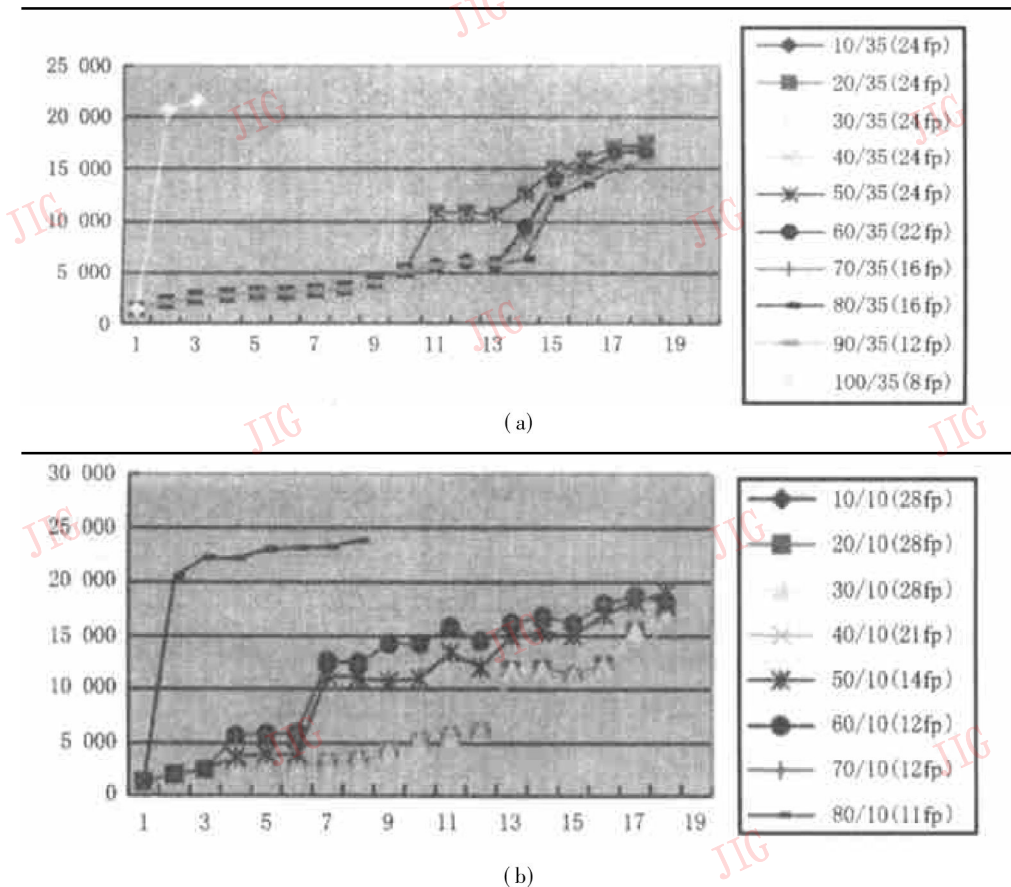


图 8 与特征点选择相关的运动参数误差与面积误差

## 参 考 文 献

- 1 Forchheimer R, Fahlander O. Low bit-rate coding through animation. PCS'83, 1983: 113~ 114.
- 2 Aizawa K, Huang Thomas S. Model-based omage coding: advanced video coding techniques for very low bit-rate applications. Proc. of the IEEE, 1995, 83(2): 259~ 271.
- 3 Musmann H G, Hoetter M, Ostermann J. Object-oriented analysis-synthesis coding of moving images, Image Commun., 1987, 1(2): 117~ 138.
- 4 Aggarwal J K, Nandhakumar N. On the computation of motion from sequences of image—a review, Proc. of the IEEE, 1988, 76(8): 917~ 935.
- 5 Bozdagi G, Tekalp A M, Onural L. An improvement to MBASIC algorithm for 3-D motion and depth estimation, IEEE Trans. on Image Processing, 1994, 3(5): 711~ 716.
- 6 Yu Lu, Yao Qingdong. Model-based image coding for video conference. Workshop on image processing and coding'98, Hong Kong, 1998(2): 139~ 144.
- 7 Yu Lu, Liu Yunhai, Yao Qingdong, Estimation of feature parameters in model-based image coding system for video conference, In: SPIE-VCIP'99, (2): 465~ 472.
- 8 Liu Yunhai, Yu Lu, Yao Qingdong, Automatic extraction of facial features using deformable templates, PCS'99, 1999: 189~ 192.

- 9 Yunhai Liu, Lu Yu, Qingdong Yao. Facial motion parameter estimation and error criteria in model-based image coding. SPIE, San Joes, California, 2000: 807~ 815.



刘云海 1966年生, 博士, 浙江大学信电系教师. 主要研究领域为图象处理、图象编码、语音信号处理.



虞 露 1996年获浙江大学博士学位, 现为浙江大学信电系副教授. 主要研究领域为图象处理、图象编码、高清晰度电视.



姚庆栋 浙江大学信电系教授, 博士生导师. 研究方向为图象编码、实时图象处理、高清晰度电视、并行处理、数字通信.

## 新 书 邮 购

由中国图象图形学学会和科学出版社联合推出的“图象图形丛书”现已出版6部, 若有需要, 请邮汇书款至: 北京海淀区花园路6号, 《中国图象图形学报》编辑部, 邮编: 100088. 汇款单上务必注明所购新书之序号, 并附加书款15%的挂号邮费.

序号	书 名	作 者	价 格
1	机器视觉	贾云得	32.00元
2	视频数据库管理系统导论	周洞汝	18.00元
3	遥感图象信息系统	张永生	36.00元
4	遥感图象获取与分析	朱述龙	18.00元
5	图象处理与分析——数学形态学方法与应用	崔 屹	23.00元
6	面向智能体的视觉信息处理	刘健勤 盛津芳 魏敏洁	20.00元