

用于口型识别的实时唇定位方法*

姚鸿勋¹ 高文¹ 李静梅² 吕雅娟¹ 王瑞¹

¹(哈尔滨工业大学计算机科学与工程系 哈尔滨 150001)

²(哈尔滨工程大学计算机科学与信息系 哈尔滨 150001)

E-mail: yhx@vilab.hit.edu.cn

摘要 在许多应用于有噪声环境下的语音识别系统中,唇读技术能有效地降低噪声的影响,通过视觉通道来补充仅取决于听觉通道的信息量,从而提高语音识别系统的识别率.该文提出了一种有效和稳健的唇定位跟踪方法,以满足不用特殊标识物和规范性照明就能对信息进行有效提取的应用需求.该方法首先用肤色模型查找脸;然后用迭代算法搜索脸部区域内的眼睛;再根据眼睛的位置来确定脸的大小和位置,并对脸的下半部分采用彩色坐标变换法将唇从肤色中明显地区分出来;最后,用可变模板将上下唇的内外轮廓描述出来.

关键词 唇读,唇动,肤色模型,光流,可变模板.

中图法分类号 TP391

一个说话者的视觉图像为其单纯由声波带来的听觉提供了补充信息,有助于提高识别的准确性,尤其是对那些处于有大量噪声干扰或者有众多说话者的环境更为适用.近几年,有关唇读研究在国外已经有人做了大量的工作.实验表明,通过HMM(hidden Markov model,隐马尔可夫模型)^[1]、神经网络^[2]或统计模式识别方法^[3],视觉信息可增进语音识别.然而,这些研究大多数都是运用唇部的一些特殊标记,如口红、设置若干发光二极管等来进行的,这种限制对一般的视频输入是苛刻的.由于音韵学与视觉信息的相关性大多数都来源于嘴和唇,因此,对它们进行准确而稳健的动态定位与跟踪是十分重要的.而且,无需用特殊的标识物和规范性的照明来对信息进行提取以满足应用的需求.因此,说话者无任何限定性条件的唇的定位与跟踪是亟待解决的唇读的首要问题.所谓唇读问题,即如何将一个连续的序列图像进行变换得到有效的特征值,从而进行识别的问题.目前,从图像序列中抽取特征的方法可分为如下3类:

- (1) 可变模板(deformable templates)^[1,4]和Snake算法^[5];
- (2) 光流(optical flow)分析^[6,7];
- (3) 特征唇(eigenlips)和能量序列^[8].

而对唇的定位,以往的系统^[1,9]大多通过手工来完成.只有用计算机自动完成嘴和唇的定位与跟踪,且不限定光照条件,才能真正实时地、在线地完成唇读,并与语音系统结合起来.

1 意义

在唇读系统中,话者的唇部跟踪是必要的.因为唇的形状对口型识别贡献最大.但是,直接定位和跟踪唇,有较大的困难,除非人脸在图像中的位置是事先知道的.一般的思路是,寻找唇,可以通过寻找脸部的其他明显特

* 本文研究得到国家自然科学基金(No. 69789301)和国家863高科技项目基金(No. 863-306-ZT03-01-2)资助.作者姚鸿勋,女,1965年生,副教授,主要研究领域为图像处理,模式识别,多媒体智能人机接口技术.高文,1956年生,博士,教授,博士生导师,主要研究领域为数据压缩,模式识别,计算机视觉,人工智能.李静梅,女,1964年生,讲师,主要研究领域为网络通信,多媒体技术.吕雅娟,女,1972年生,硕士生,主要研究领域为唇读,多功能感知,机器翻译.王瑞,1971年生,博士生,主要研究领域为唇读,多功能感知技术.

本文通讯联系人:姚鸿勋,哈尔滨 150001,哈尔滨工业大学计算机科学与工程系 321 信箱

本文 1999-05-17 收到原稿,1999-09-09 收到修改稿

征,如瞳孔等,然后预测唇的位置及大小.之后,再进一步精确定位及特征提取来实现.在本系统所使用的方法中,首先用肤色模型查找脸;然后搜索脸部区域内的眼睛;再根据眼睛的位置和脸的大小确定唇的位置;最后,再将上下唇的内外轮廓描述出来.这样做的最大优点是,对于各种姿态的人脸图像区域都能检测到,而且对光照变化不敏感.否则,以先检测瞳孔为例,系统很难自动确定何为瞳孔,尤其是当人的眼睛不在同一水平线上或不呈对称状态时更为困难.

2 搜索脸

为了查找和定位脸,我们用统计的肤色模型,即由一个二维高斯正态分布的肤色构成.因为用色度表示人脸的特征,具有姿态不变性.

在输入图像中搜索肤色,肤色的最大连通区域即可能是脸,如图1所示.



Fig. 1 An example of detecting by skin color model, where the dark gray region is regarded as skin

图1 一个用肤色模型搜索的例子,图中深色区域即被认定是肤色

肤色分布仅是一个初定位,得到感兴趣区域(region of interest,简称ROI),为适应肤色的变化及实际的脸定位奠定了基础.

3 眼睛的检测

眼睛与周围面部比较,灰度差别大,容易定位.基于前面肤色的面部初定位,我们在人脸区域中寻找满足人体器官位置的测量约束的两个黑区域,它即是我们要找的瞳孔.在给定区域内,用一系列阈值反复二值化这个区域,直到有一对满足人脸中几何约束的区域出现为止.我们将此称为阈值迭代算法.如图2所示.

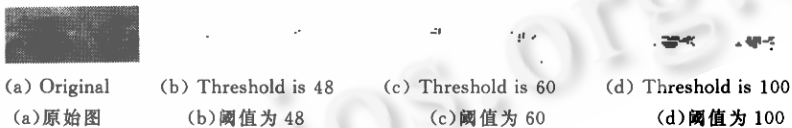


Fig. 2 Detection eyes by threshold iterative algorithm

图2 用阈值迭代算法检测眼睛

由于这个阈值是自适应的,这种方法可以适应各种光照的变化及各个不同的人,使得该方法具有很好的鲁棒性.

为了便于更精确地确定嘴的位置,我们可以结合人眼的定位^[10]来加以判断.

4 嘴的检测

有了眼睛的肤色区,我们认为那才可能是脸.在人脸初步定位以后,取脸的下半部区域作为研究对象.由于面部的主要信息颜色为肤色,为了有效地把唇色从肤色中区分出来,我们通过寻找一个合适的彩色坐标系,使肤色和唇色在该坐标系中具有较好的可分性.为此,我们进行了彩色坐标变换.

我们知道,从RGB空间向YUV空间转换可用如下公式:

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & 0.436 \\ 0.615 & -0.515 & -0.100 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

该彩色坐标系经过适当的旋转和平移,可得到易于区分肤色和唇色的灰度值表示.因为色度信号 U 和分量 V 必定可以找到一个合适的位置,将肤色和唇色两类不同色度的聚类信号分离出来.经过大量的实验,旋转和平移后的坐标轴 \tilde{U}, \tilde{V} 与 U, V 的关系可用下式表示:

$$\begin{bmatrix} \tilde{U} \\ \tilde{V} \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \cdot \begin{bmatrix} U \\ V \end{bmatrix} + \begin{bmatrix} -X_0 \\ 0 \end{bmatrix}, \quad (2)$$

其中 θ 为旋转角度,实验值为 66° , X_0 为平移量,实验值为 27. 利用式(1)和式(2)可以求得 \tilde{U}, \tilde{V} 与 R, G, B 的转换关系为

$$\begin{bmatrix} \tilde{U} \\ \tilde{V} \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \cdot \begin{bmatrix} -0.147 & -0.289 & 0.436 \\ 0.615 & -0.515 & -0.100 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} -X_0 \\ 0 \end{bmatrix}. \quad (3)$$

新坐标系的相位角 $\tilde{\varphi}$ 为

$$\tilde{\varphi} = \text{tg}^{-1} \left[\frac{\tilde{V}}{\tilde{U}} \right]. \quad (4)$$

在新坐标系下,唇色的阈值范围为 $\tilde{U} \in (\tilde{U}_{\min}, \tilde{U}_{\max})$, $\tilde{\varphi} \in (\tilde{\varphi}_{\min}, \tilde{\varphi}_{\max})$, 实验得到的经验值为 $\tilde{U}_{\min} = 5, \tilde{U}_{\max} = 45, \tilde{\varphi}_{\min} = 30, \tilde{\varphi}_{\max} = 85$.

对关心区域的每个像素点用式(3)和式(4)进行坐标变换,如果得到的 $\tilde{U}, \tilde{\varphi}$ 在指定的阈值范围内,则对相应的像素点进行增强处理,即使对应的唇色区域变换后具有较小的灰度值,非唇色区域具有较高的灰度值.变换的权值函数为:

$$f(h) = \begin{cases} K_1 \cdot \frac{(\tilde{U} - \tilde{U}_0)^2}{W_0^2} + K_2 \cdot \frac{(\tilde{\varphi} - \tilde{\varphi}_0)^2}{W_1^2}, & |\tilde{U} - \tilde{U}_0| \leq W_0, |\tilde{\varphi} - \tilde{\varphi}_0| \leq W_1, \\ 1, & \text{其他} \end{cases} \quad (5)$$

其中, $\tilde{U}, \tilde{\varphi}$ 是像素在新坐标系下的值, \tilde{U}_0 为唇色样本的 \tilde{U} 均值, $\tilde{\varphi}_0$ 为唇色样本的 $\tilde{\varphi}$ 均值. W_0, W_1 分别为 $\tilde{U}, \tilde{\varphi}$ 的滤波半径, $K_1 + K_2 = 1$, 调整 K_1, K_2 , 可以人为地控制 \tilde{U} 和 $\tilde{\varphi}$ 的权重. 实验得到的经验值为 $\tilde{U}_0 = 20, \tilde{\varphi}_0 = 50, W_0 = 30, W_1 = 40, K_1 = 0.6, K_2 = 0.4$. 增强后的灰度值 \tilde{Y} 计算表达式为

$$\tilde{Y} = f(h) \cdot 255. \quad (6)$$

图 3 为增强变换前后的唇部图像对比. 可见,通过唇色分析对其进行变换,使唇色明显增强,这使得后面的唇动定位与跟踪操作相对容易,并且提高了准确度.

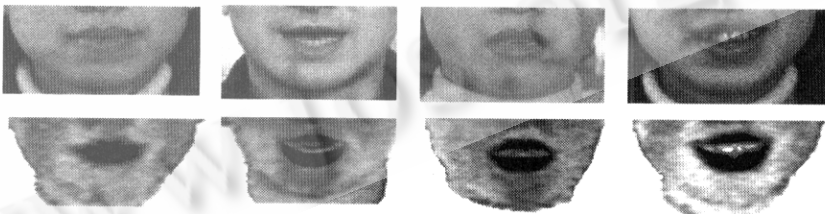


Fig. 3 The contrast of the enhanced images with the original ones

图 3 增强变换前后的唇部图像对比

5 口型轮廓定位

为了得到比较理想的口型,我们用最佳阈值算法对口型图像进行二值化.所谓最佳阈值二值化算法,即用以下条件作为二值化阈值选定的约束条件:①唇的边缘是平滑的.②左右形状对称.

我们发现,噪音多的图像其目标的轮廓边缘像素倾斜率变化较大,而当采用合适的阈值进行二值化时,所得到的图像的轮廓边缘像素倾斜率变化则基本上是一个定值.因此,边缘像素倾斜率变化趋于最小,即是理想阈值的一个约束条件.另外,理想二值化提取的口型应该是左右对称的.基于以上考虑,我们定义一个二值化的约束条件.

$$\sigma = \alpha \sqrt{\frac{1}{N} \sum_{i=1}^N (\theta_i - \bar{\theta})^2} + \beta \frac{\sum_{i=1}^M (f(x, y) - f(-x, y))^2}{M} \quad (7)$$

第1项为平滑约束条件. 其中 θ_i 定义如下: $\theta_i = \arctg(-\Delta x / \Delta y)$, $\Delta x, \Delta y$ 分别为 5×5 的模板的水平 and 垂直变化值. N 为口型轮廓总数, $\bar{\theta}$ 为 θ_i 的平均值, 即 $\bar{\theta} = \frac{1}{N} \sum_{i=1}^N \theta_i$.

水平模板:

$$\begin{bmatrix} -0.2 & -0.2 & 0 & 0.2 & 0.2 \\ -0.4 & -0.5 & 0 & 0.5 & 0.4 \\ -0.5 & -1 & 0 & 1 & 0.5 \\ -0.4 & -0.5 & 0 & 0.5 & 0.4 \\ -0.2 & -0.2 & 0 & 0.2 & 0.2 \end{bmatrix}$$

垂直模板:

$$\begin{bmatrix} 0.2 & 0.4 & 0.5 & 0.4 & 0.2 \\ 0.2 & 0.5 & 1 & 0.5 & 0.2 \\ 0 & 0 & 0 & 0 & 0 \\ -0.2 & -0.5 & -1 & -0.5 & -0.2 \\ -0.2 & -0.4 & -0.5 & -0.4 & -0.2 \end{bmatrix}$$

第2项为对称性约束条件. $f(x, y)$ 为嘴唇坐标系中的唇轮廓曲线, α 和 β 为权系数, 经验设定 $\alpha = 0.5, \beta = 0$.

3. 通过以上约束条件 σ , 使之趋于最小, 便能够得到最佳阈值. 用此阈值经二值化后, 可得到良好的口型轮廓.

在口型轮廓提取中, 可变模板方法得到了广泛的应用^[1,3], 它的优点是不受口型的变化、缩放和旋转的影响, 能够有效地描述口型轮廓和模板匹配的结果, 直接给出口型的形状特征参数, 给进一步的识别带来了方便. 但是, 可变形模板方法的缺点是, 它依靠图像二值化信息和边缘信息来实现模板匹配, 但是口型的边缘提取是非常困难的, 二值化的阈值选取也很难, 它们往往受到光照、阴影、牙齿、胡须的影响, 这使得在很多情况下变形模板方法不能达到令人满意的结果.

由于我们的方法首先对唇色信息进行了处理, 处理后的图像唇色明显增强, 这使得唇部区域的二值化处理 and 边缘提取都取得了较好的结果. 图4给出了唇色处理前后的二值化和边缘提取的结果比较, 显然, 唇色增强后的改善效果是明显的.

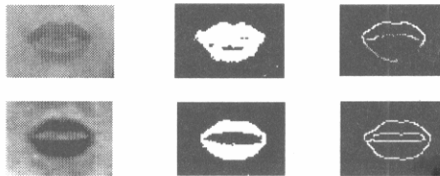


Fig. 4 The contrast of the results on the enhanced images and the original ones

图4 增强图像与原图像处理结果比较

6 口型轮廓描述与跟踪

我们用口型可变模板来描述唇的形状. 口型可变模板是一个参数化的物体模型. 它是一个用多条曲线来描述的口型, 通过最优化方法调整模型参数, 使代价函数趋于最小, 即模板逐渐靠近唇型的真实位置. 如图5所示为口型的模板, 它是根据人们日常对口型的经验信息确定的, 能够适用于大多数嘴型. 模板主要由抛物线和四次曲线这两种曲线构成. 内唇由两个抛物线描述. 外唇由四次曲线描述. 实验表明, 四次曲线能更精确地反映外唇的形状.

模型参数说明: 中心坐标 (x_c, y_c) 和转角 θ 确定嘴坐标系的位置和方向, 所有的长度和尺寸均参照此坐标系, 并且假定嘴是关于纵坐标对称的. 上唇的外边缘用两个四次曲线来描述, 它们的中心离坐标原点的偏移量为 a_{off} , 高为 h_1 , 迫使它们享有相同的参数, 这样即确定了模板的对称性. 四次曲线中参数 q_0 表示四次曲线偏离抛物线的距离, 它使模板可以更精确地接近唇的形状. 下唇的外边缘也被描述成四次曲线, 高为 h_4 , 辅助参数 q_1 所提供的可变性使之比抛物线更能准确地跟踪话者的下唇. 上下外唇具有相同的宽度 w_0 , 相交于 $(-w_0, 0), (w_0, 0)$. 两条抛物线用来描述嘴的内部边缘, 每条抛物线的高分别为 h_2, h_3 , 其宽度均为 w_1 , 相交于 $(-w_1, 0), (w_1, 0)$. 这样, 口型模型可以用以下的参数来确定: $(x_c, y_c), \theta, w_1, w_0, a_{off}, h_1, q_0, h_2, h_3, h_4, q_1$. 其轮廓的曲线方程如下:

$$\begin{aligned}
 Y_{ut} &= h_1 \times \left(1 - \frac{(x+a_{off})^2}{(w_0-a_{off})^2} \right) + 4q_0 \times \left(\frac{(x+a_{off})^4}{(w_0-a_{off})^4} - \frac{(x+a_{off})^2}{(w_0-a_{off})^2} \right), \\
 Y_{ur} &= h_1 \times \left(1 - \frac{(x-a_{off})^2}{(w_0-a_{off})^2} \right) + 4q_0 \times \left(\frac{(x-a_{off})^4}{(w_0-a_{off})^4} - \frac{(x-a_{off})^2}{(w_0-a_{off})^2} \right), \\
 Y_w &= h_2 \times \left(1 - \frac{x^2}{w_0^2} \right), \\
 Y_{li} &= -h_3 \times \left(1 - \frac{x^2}{w_0^2} \right), \\
 Y_l &= -h_4 \times \left(1 - \frac{x^2}{w_1^2} \right) - 4q_1 \times \left(\frac{x^4}{w_1^4} - \frac{x^2}{w_1^2} \right).
 \end{aligned} \tag{8}$$

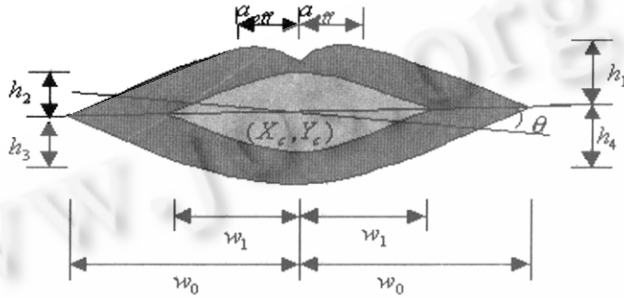


Fig. 5 Lip model: including two parabolas of inner lip, three quartic curves of outer lip and nine parameters

图 5 口型模板图:包括内唇的两条抛物线和外唇的 3 条四次式曲线及 9 个参数

模板的匹配过程也就是代价函数最小化的过程。因此,代价函数和最优化算法选取的好坏直接影响到匹配的质量和速度。

在我们的系统中,代价函数包含了唇型的边缘信息和时空约束信息以及其他一些惩罚项,对唇的变化进行了约束限制。最优化算法采用最速下降法,它使各个参数同时变化,综合全局信息使代价函数达到最小。

由于最速下降法要求函数连续变化,因此还需要对二值化后的图像进行梯度处理(blurring),我们采用 5×5 的指数 blurring 算子。该算子的中心值为 1,随着离中心点的距离按指数递减。

最速下降法的迭代公式为

$$X^{k+1} = X^k - \lambda_k \nabla f(x^k),$$

其中 $-\nabla f(x^k)$ 为代价函数的负梯度,可以证明函数沿负梯度方向下降最快; λ_k 为下降步长,可用 0.618 法求得最佳步长。如果下降幅度低于一定的阈值或超过一定的迭代次数,则停止迭代。

代价函数包括 4 个曲线积分,每个积分对应一条边,考虑到正、负边的问题,其积分形式如下:

$$\begin{aligned}
 E_1 &= -\frac{1}{|\Gamma_1|} \int_{\Gamma_1} \Phi_r(\vec{x}) ds, & E_2 &= \frac{1}{|\Gamma_2|} \int_{\Gamma_2} \Phi_r(\vec{x}) ds, \\
 E_3 &= -\frac{1}{|\Gamma_3|} \int_{\Gamma_3} \Phi_r(\vec{x}) ds, & E_4 &= \frac{1}{|\Gamma_4|} \int_{\Gamma_4} \Phi_r(\vec{x}) ds,
 \end{aligned} \tag{9}$$

其中 Γ_i 是 4 条描述口型的模板曲线, $|\Gamma_i|$ 是曲线长度, $\Phi_r(\vec{x})$ 是落在模板上的图像的灰度值。各项分别除以长度是为了对它们进行归一化处理,简化权值系数的选取。

代价函数也包括一些时空约束项和惩罚项,它们对模板变化的范围和幅度进行了约束和限制。如限制 $h_2 > h_3$,使内唇不会高于外唇;限制 $a_1 < (h_1 - h_2) < a_2$,使唇的厚度在合理的范围之内,如果超过此范围则使代价函数迅速增长作为惩罚。此外,还有其他一些惩罚项,这些约束使模板形变保持在合理的范围之内。

综上所述,代价函数可用下式表示:

$$E = \sum_{i=1}^4 C_i E_i + \sum_j K_j E_{penalty_j}, \tag{10}$$

其中 C_i, K_i 为权值系数. 适当地选取系数可使模板按要求的方向变化. 如在确定中心点后, 首先使 C_2, C_3 为 0, 使 $C_1=1, C_4=0.5$, 则先变化外唇, 并且以上唇变化为主, 当外唇基本匹配时, 再使 C_1, C_4 为 0, 变化 C_2, C_3 进行内唇匹配, 这样可提高匹配速度.

实现口型定位之后, 唇动跟踪相对容易实现. 由于相邻帧口型运动变化不大, 前一帧定位结果作为初始参数进行新的模板匹配即可实现当前帧的匹配, 如果某一帧没有成功匹配, 则开始一个新的定位过程. 图 6 为汉语口型 [a] 的部分帧的定位与跟踪结果.



Fig. 6 The result of locating and tracking Chinese vowel [a]

图 6 汉语口型[a]的定位与跟踪结果

7 实验结果及结论

在我们的实验中, 采用的是 CPE-1000 图像卡、JVC TK-1070 彩色摄像机和 MIMTRON MTV-3301CB 彩色摄像机, 采集图像为 24 位真彩色图像, 采集速度为每秒 25 帧, CPU 为 Pentium-II 300. 实验结果表明, 本文提出的唇色模型和可变模板相结合的唇动定位和跟踪方法在自然条件下取得了令人满意的结果. 系统对人脸图像实时进行唇动定位和跟踪, 粗定位准确率达 95% 以上, 跟踪速度为 14~16 帧/秒, 口型定位的准确率为 80%, 跟踪速度为 8~10 帧/秒左右. 文中所提出的分割唇色与肤色的彩色坐标变换方法对其他领域的色彩分割问题同样适用. 今后的工作是希望进一步提高口型定位的精度和跟踪速度, 以期满足实时识别的要求.

致谢 特别感谢张洪明同志, 他为肤色和唇色的 RGB 变换做了大量实验, 并给出了一组有意义的可供肤色、唇色增强的变换参数.

参考文献

- 1 Hennecke M, Prasad K, Stork D. Using deformable templates to infer visual speech dynamics. In: Proceedings of the 28th Annual Asilomar Conference on Signals, Systems and Computers, Vol 1. Pacific Grove: IEEE Computer Society Press, 1994. 578~582
- 2 Wolff G, Prasad K, Stork D *et al.* Lip-reading by neural networks: visual preprocessing, learning and sensory integration. In: Cowan J, Tesauro G, Alspector J. eds. Proceedings of the Neural Information Processing Systems-6. San Mateo, CA: Morgan Kaufmann Publishers, Inc., 1994. 1027~1034
- 3 Petajan E D. Automatic lip-reading to enhance speech recognition [Ph. D. Thesis]. University of Illinois at Urbana-Champaign, 1984
- 4 Coianiz T, Torresani L, Caprile B. 2D deformable model for visual speech analysis. In: Stork D, Hennecke M eds. Speechreading by Humans and Machines: Models, Methods, and Applications. Volume 150, NATO-ASI Series, Series F: Computer and Systems Sciences. Berlin: Springer-Verlag, 1995
- 5 Kass M, Witkin A, Terzopoulos D. Snakes: active contour models. In: Proceedings of the 1st International Conference on Computer Vision. New York: IEEE Computer Society Press, 1987. 259~268
- 6 Finn K, Montgomery A. Automatic optically-based recognition of speech. Pattern Recognition, 1988, 8(3): 159~164
- 7 Mase K, Pentland A. Automatic lip-reading by optical flow analysis. Systems and Computers in Japan, 1991, 22(6): 67~76
- 8 Bregler C, Konig Y. "Eigenlips" for robust speech recognition. In: Bongner R E ed. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. Adelaide: Adelaide Convention Center, 1994. 667~674
- 9 Bregler C, Hild H *et al.* Improving connected letter recognition by lip-reading. In: Kaveh M ed. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. Minnesota: Minneapolis Convention Center, 1993. 557~560

10 Kinmanlam, Yan H. Locating and extracting the eye in human face images. Pattern Recognition, 1996,29(5):771~779

Real-Time Lip Locating Method for Lip-Movement Recognition

YAO Hong-xun¹ GAO Wen¹ LI Jing-mei² LÜ Ya-juan¹ WANG Rui¹

¹(Department of Computer Science and Engineering Harbin Institute of Technology Harbin 150001)

²(Department of Computer Science and Information Harbin Shipbuilding Engineering Institute Harbin 150001)

Abstract For speech recognition systems under noisy environment, lip-reading technique can effectively reduce the influence of noise and improve the accurate rate of speech recognition system by adding visual information to acoustic channel. In this paper, an effective and robust approach for lip and mouth locating and tracking is presented to enable the information extraction under abnormal illumination and without special marks. This approach first locates face region with skin-color model, then finds the eyes from the face region with iterative algorithm, modifies the position and size of face according to the position of eyes, transforms the lower part of face by specific color coordinators to clearly distinguish lip color from skin color, and finally describes the outline of upper lip and lower lip with deformable template.

Key words Lip-reading, lip-movement, skin-color model, optical flow, deformable template.