第33卷 第7期 北京工业大学学报 Vol.33 No.7
2007年 7月 JOURNAL OF BEIJING UNIVERSITY OF TECHNOLOGY Jul. 2007

基于 HASM 的口形特征点定位

孙艳丰，陈 贺，贾熹滨，李敬华
（北京工业大学 多媒体与智能软件技术北京市重点实验室，北京 100022）

摘 要：针对口形特征点定位的准确性问题，提出一种基于 HASM(hierarchical active shape model)的口形轮廓定位方法，采用不等步长、不等角度建模策略和口形聚类策略，构建局部纹理模型作为特征点搜索依据，并利用马氏距离选取最佳定位点．试验结果表明，HASM 模型的口形特征点定位方法使内唇定位和闭口口形定位的准确率达到 90％以上．

关键词：特征点定位；主动形状模型；局部纹理模型；马氏距离
中图分类号：TP 391 **文献标识码**：A **文章编号**：0254－0037(2007)07－0726－05

可视语音合成研究的重点在于获得有效的口形描述信息，口形特征点方法是一种较为典型的定位方法，如：Voice Puppetry 系统[1]采用包括口形在内的 26 个人脸特征点；Video Rewrite 系统[2]利用描述口形及下颚的 54 个特征点表示口形；Person Authentication 系统[3]采用描述口形整体轮廓的 14 个特征点表示口形．关于定位特征点的位置，目前的解决方案包括：核函数的局部模板匹配算法[4]和 Eigon Point 系统[5]中提出的图像纹理与矢量之间的映射算法，还有基于统计的主动形状模型方法[6-7]．主动形状模型(ASM)是一种基于统计模型的图像搜索方法，通过统计建模得到形状模型和纹理模型，然后利用局部搜索算法进行特征点精确定位[8]．作者在综合分析上述口形描述策略的基础上，采用 20 个特征点，并结合传统的 ASM 算法，提出了基于 HASM(hierarchical active shape model)的口形轮廓定位方法．

1 HASM 算法的整体构架

作者在传统 ASM 模型的基础上，引入了一种基于 HASM 的口形轮廓定位方法，算法流程见图 1．

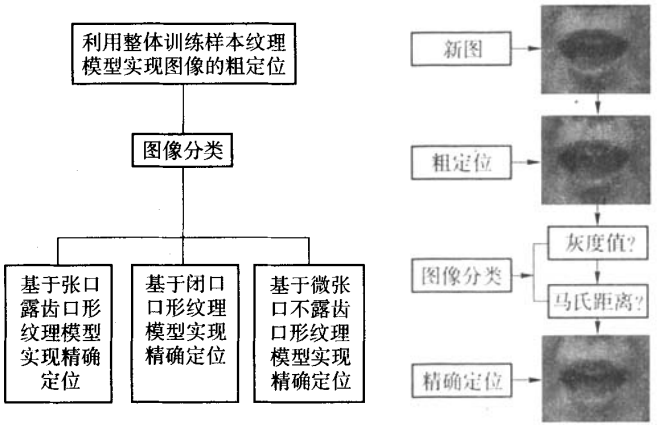


图 1 HASM 算法流程

Fig.1 The flowchart of HASM

收稿日期：2006-03-02．
基金项目：国家自然科学基金资助项目(60375007)；北京市自然科学基金资助项目(40410031)．
作者简介：孙艳丰(1964－)，女，黑龙江齐齐哈尔人，教授．

这种方法采用分层处理的方式大大提高了口形定位精度. HASM 由 2 层 ASM 组成，高层为全局分类层，底层为精确定位层. 在高层中，利用整体训练图像的局部纹理模型实现对新图像的第 1 次初步定位，定位的结果作为口形分类的依据. 在底层中，将人们说话时所涉及的所有口形分为张口露齿、微张口不露齿和闭合口形. 作者利用这 3 种不同的局部纹理模型实现了对原图像的精确定位，选取最优的轮廓作为最终的定位结果.

2　局部纹理模型的建立

局部纹理模型特征点定位算法的总体思路是以样本集中特征点为单位，通过对每个特征点灰度信息的统计学习，得到相应的统计特征，然后利用统计特征作为评判依据.

根据样本集建立特征点的标准局部纹理模型，以每个特征点为中心，沿 x 轴和 y 轴 2 个方向分别选取等距离的像素点作为灰度变化统计范围. 特征点 x 轴的左半轴为 1 号子点阵，右半轴为 2 号子点阵；特征点 y 轴的上半轴为 3 号子点阵，下半轴为 4 号子点阵，如图 2 所示. 将 1 号子点阵和 2 号子点阵组成 χ 点阵，将 3 号子点阵和 4 号子点阵组成 γ 点阵，分别表示 x 轴和 y 轴 2 个方向的灰度变化情况.

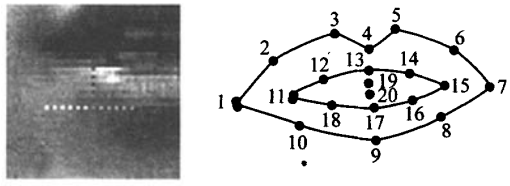


图 2　特征点局部纹理模型

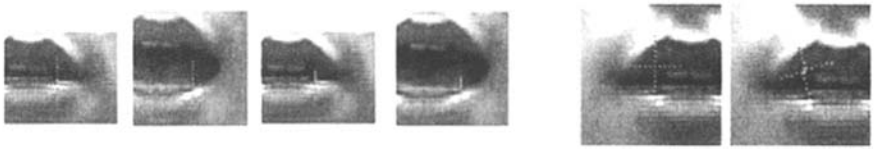
Fig. 2　Local texture model of oral features

本文采用的训练样本集为 100 幅 90×75 像素的口形样本，该样本涵盖所有发音情况下的口形信息，并按照顺时针方向由外唇到内唇的顺序标定 20 个具有代表性位置的特征点，如图 2 所示.

2.1　不等步长、不等角度建模策略

作者针对口形运动变化过程中各特征点位置变化程度的复杂性，提出了不等步长、不等角度建模策略. 步长就是指某特征点在某一方向上相邻 2 像素点的距离. 不等步长策略的核心思想为：通过调整相邻像素点的距离改变某个方向上的灰度变化统计范围，以便更有效地准确定位最佳位置点. 不等角度策略的核心思想为：通过调整旋转角度，达到更好地表示该区域纹理特点的目的.

图 3(a)表示 2 种张口程度的口形样本，图中特征点的上边界已经触及上唇，因此必然受到上唇所带来的较大影响，而在特征点选取适当步长的情况下，则能在一定程度上有效地避免其他外界因素的干扰. 对于在嘴唇内外轮廓处的特征点，从嘴唇轮廓的切线方向和法线方向建立灰度模型，可以获得表示嘴唇轮廓自身的灰度变化情况，同时又可以获得表示嘴唇轮廓运动方向的灰度变化情况，如图 3(b)图所示.



(a) 不等步长纹理模型　　(b) 不等角度纹理模型

图 3　相同特征点不等步长、不等角度纹理模型

Fig. 3　The variable step strategy and the variable angle strategy

2.2　局部纹理模型算法

作者根据本文所采用的训练样本集，以不等步长、不等角度策略为基础建立了局部纹理模型. 通过对样本集数据的统计分析，针对 20 个口形特征点分别设定步长和旋转角度. 具体计算方法如下.

首先，读取包括特征点在内的像素灰度值，并以特征点为中心建立该点的点阵和点阵的灰度值向量

$$\boldsymbol{h}_{\chi_{ki}}=(h_{\chi_{ki(-m)}},\cdots,h_{\chi_{ki0}},\cdots,h_{\chi_{ki(m)}}) \tag{1}$$

$$\boldsymbol{h}_{\gamma_{ki}}=(h_{\gamma_{ki(-m)}},\cdots,h_{\gamma_{ki0}},\cdots,h_{\gamma_{ki(m)}}) \tag{2}$$

其中，k 表示样本集中第 k 幅图像($k=1,\cdots,N-1$)；i 表示标定图像中第 i 个特征点($i=0,\cdots,n-1$)；$-m,\cdots,0,\cdots,m$ 表示以标定特征点为中心选取的像素点的标号.

求取每个特征点 χ 点阵和 γ 点阵的灰度差分

$$\Delta\boldsymbol{h}_{\chi_{ki}}=(h_{\chi_{ki(-m+1)}}-h_{\chi_{ki(-m)}},\cdots,h_{\chi_{ki1}}-h_{\chi_{ki0}},\cdots,h_{\chi_{ki(m)}}-h_{\chi_{ki(m-1)}}) \tag{3}$$

$$\Delta\boldsymbol{h}_{\gamma_{ki}}=(h_{\gamma_{ki(-m+1)}}-h_{\gamma_{ki(-m)}},\cdots,h_{\gamma_{ki1}}-h_{\gamma_{ki0}},\cdots,h_{\gamma_{ki(m)}}-h_{\gamma_{ki(m-1)}}) \tag{4}$$

分别对 $\Delta\boldsymbol{h}_{\chi_{ki}}$、$\Delta\boldsymbol{h}_{\gamma_{ki}}$ 进行标准化处理

$$\boldsymbol{g}_{\chi_{ki}}=\mathrm{d}h_{\chi_{ki}}\Big/\sum_{q=-m}^{m-1}|\boldsymbol{h}_{\chi_{ki(q+1)}}-\boldsymbol{h}_{\chi_{kiq}}| \tag{5}$$

$$\boldsymbol{g}_{\gamma_{ki}}=\mathrm{d}h_{\gamma_{ki}}\Big/\sum_{q=-m}^{m-1}|\boldsymbol{h}_{\gamma_{ki(q+1)}}-\boldsymbol{h}_{\gamma_{kiq}}| \tag{6}$$

对样本集中的 N 幅图像计算特征点标准化向量平均值

$$\bar{\boldsymbol{g}}_{\chi_i}=\frac{1}{N}\sum_{k=1}^{N}\boldsymbol{g}_{\chi_{ki}} \tag{7}$$

$$\bar{\boldsymbol{g}}_{\gamma_i}=\frac{1}{N}\sum_{k=1}^{N}\boldsymbol{g}_{\gamma_{ki}} \tag{8}$$

利用

$$\boldsymbol{C}_{\chi_i}=\frac{1}{N}\sum_{k=0}^{N-1}(\boldsymbol{g}_{\chi_{ki}}-\bar{\boldsymbol{g}}_{\chi_i})^{\mathrm{T}}(\boldsymbol{g}_{\chi_{ki}}-\bar{\boldsymbol{g}}_{\chi_i}) \tag{9}$$

$$\boldsymbol{C}_{\gamma_i}=\frac{1}{N}\sum_{k=0}^{N-1}(\boldsymbol{g}_{\gamma_{ki}}-\bar{\boldsymbol{g}}_{\gamma_i})^{\mathrm{T}}(\boldsymbol{g}_{\gamma_{ki}}-\bar{\boldsymbol{g}}_{\gamma_i}) \tag{10}$$

计算特征点在 χ 点阵和 γ 点阵标准化灰度向量的协方差矩阵，该协方差矩阵能表示出特征点处的灰度变化规律统计特性，因此将其作为所需的局部纹理模型. 其中，$\boldsymbol{g}_{\chi_{ki}}$ 和 $\boldsymbol{g}_{\gamma_{ki}}$ 表示第 i 个样本点的标准化灰度向量；$\bar{\boldsymbol{g}}_{\chi_i}$ 和 $\bar{\boldsymbol{g}}_{\gamma_i}$ 表示 k 个样本标准化灰度向量的平均值向量.

作者采用的建模方法不同于传统的仅仅依据特征点法线方向建模的纹理模型方法，而是依据特征点 x 轴和 y 轴 2 个方向并参考一定角度，建立以 χ 点阵和 γ 点阵表示的局部纹理模型. 该方法主要考虑到口形轮廓是一个变化复杂的不平滑曲线，特征点位置多在如嘴角、上唇这些不规则点处，当说话人口形发生较大的变化时，以 χ 点阵和 γ 点阵描述的灰度信息更能准确地表示嘴唇轮廓点的位置变化情况.

3 利用纹理模型特征点精确定位

在已经建立各特征点局部纹理模型的基础上，利用训练得到灰度变化规律，在一定范围内进行特征点搜索，找出与该局部纹理模型灰度变化规律统计特性差异最小的点作为最佳位置点. 特征点搜索可以具体理解为：针对某个特征点，在一定区域内对该区域中的每个点计算其规格化灰度向量与训练得到的该点平均规格化灰度导数向量的马氏距离(Mahalanobis distance)，并且从这些候选点中选取马氏距离最小的位置点作为最佳匹配点. 特征点的马氏距离是按照计算 χ 点阵和 γ 点阵规格化灰度向量协方差矩阵的平方和作为判定标准，即

$$\mathrm{d}\chi=(\boldsymbol{g}_{\chi_i'}-\bar{\boldsymbol{g}}_{\chi_i})\boldsymbol{C}_{\chi_i}^{-1}(\boldsymbol{g}_{\chi_i'}-\bar{\boldsymbol{g}}_{\chi_i})^{\mathrm{T}} \tag{11}$$

$$\mathrm{d}\gamma=(\boldsymbol{g}_{\gamma_i'}-\bar{\boldsymbol{g}}_{\gamma_i})\boldsymbol{C}_{\gamma_i}^{-1}(\boldsymbol{g}_{\gamma_i'}-\bar{\boldsymbol{g}}_{\gamma_i})^{\mathrm{T}} \tag{12}$$

$$m=(\mathrm{d}\chi)^2+(\mathrm{d}\gamma)^2 \tag{13}$$

特征点相应搜索区域是通过统计训练样本中各个特征点的所有可能出现的位置得到的，因为每个特征点出现位置代表了所有发音情况下的口形位置信息. 根据口形轮廓的特殊形状以及口形变化的特殊规

律，提出了特殊点单独处理的方法以及口形聚类策略，以提高口形定位精度.

通过计算上下嘴唇间特征点的灰度值，将灰度值很低的口形提取出来，此时牙齿起到了决定性作用，所以，经过此次划分可以将一些张口露齿的口形从原样本集中分离出来单独归为一类，从而将张口不露齿或完全闭口的口形另外归为一类.

由于下唇运动范围相对较大，并且下唇没有一个明显的嘴唇边界，因此仍然存在闭口口形定位不准的问题(见图 4). 通过分析发现，这些定位不准的特征点都具有较大的马氏距离，因此，可以通过分析下唇特征点的马氏距离将一些定位不准的样本划分出来.

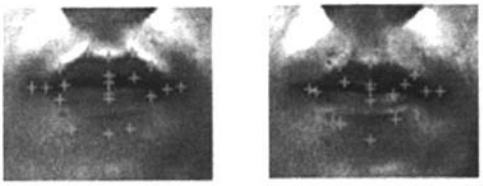


图 4　定位不准的完全闭口口形实例

Fig.4　Examples of closed mouth fitting inaccurate results

其整体构思为：在通过以上分类得到第 2 组类型样本的基础上，利用经验的灰度变化规律，针对第 9 号点在一定范围内进行特征点搜索，并计算各点与局部纹理模型灰度变化规律的马氏距离. 如果马氏距离最小点的距离数值仍然很大，就可以认定该图的类别为不可辨别的类型，因此将此样本分离出来. 通过上述分类算法可以将闭口口形从第 2 组类型样本中分离出来，从而使张口不露齿和完全闭口 2 种类型单独建模处理，提高了闭口口形定位的准确性，如图 5、图 6 所示.

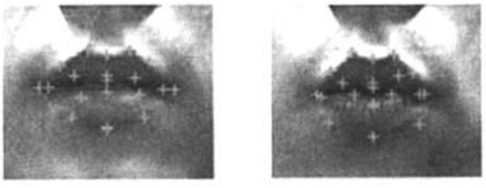


图 5　张口不露齿口形的定位实例

Fig.5　Examples of open mouth without teeth fitting results

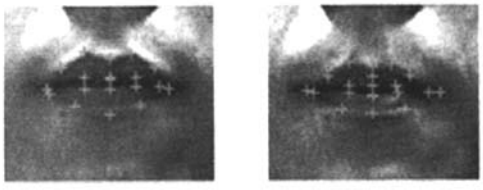


图 6　完全闭口口形的定位实例

Fig.6　Examples of closed mouth fitting results

4　试验结果分析

本文采用的训练样本集为 100 幅 90×75 像素的口形样本，试验过程中，从 36 000 多幅图像中选取出具有代表性的 150 幅样本进行标定. 为进一步验证 HASM 标定口形轮廓的准确性，将手工标定的口形形状向量 $\boldsymbol{S}_{ji}$ 及与其对应的通过算法标定的口形形状向量 $\boldsymbol{S}_{ji}^{*}$ 做差值，并通过欧氏距离比较其定位的准确性.

2 幅相同图像每个相同特征点的点欧氏距离的计算公式为

$$\Delta S_{ji} = \| \boldsymbol{S}_{ji} - \boldsymbol{S}_{ji}^{*} \| = \sqrt{(S_j(x_i, y_i) - S_j^{*}(x_i, y_i))^2}, i = 0, 1, \cdots, n-1 (i \neq j) \tag{14}$$

2 幅相同图像所有相同特征点的图像欧氏距离的计算公式为

$$\Delta S_j = \| \boldsymbol{S}_j - \boldsymbol{S}_j^{*} \| = \sqrt{\sum_{i=0}^{n-1} (S_j(x_i, y_i) - S_j^{*}(x_i, y_i))^2}, j = 0, 1, \cdots, m-1 (i \neq j) \tag{15}$$

其中，n 为手工标定特征点的数目；m 为标定的样本图像的数目.

本次试验中，考虑到图像的纹理质量和手工标定误差，将点欧氏距离的平方和小于 40 并且图像欧氏距离的平方和小于 150 的图像作为标定准确的图像. 试验结果为：非常准确图像 121 幅，基本准确图像 20 幅，不太准确图像 9 幅，准确率为 93.96%. 本文提出的口形特征点定位标准、不等步长搜索策略和口形初步聚类策略在一定程度上解决了内唇定位和闭口口形定位不准确的问题，在参考一定角度建立 x 轴和 y 轴的纹理模型时充分考虑到了口形变化的特殊规律，并且该纹理模型对于处理口形轮廓的特殊性存在着一定优势.

参考文献：

[1] BRAND M. Voice puppetry[C]// Proceedings of ACM SIGGRAPH 1999. Los Angeles: ACM Press, 1999: 21-28.

[2] BREGLER C, COVELL M, SLANEY M. Video rewrite: driving visual speech with audio[C]// Proc SIGGRAPH'97. Los Angeles: ACM Press, 1997: 353-360.

[3] MOK L L, LAU W H, LEUNG S H, et al. Lip features selection with application to person authentication [C/OL]// 2004 IEEE, Volume 3, Issue, 17-21 May 2004 Page(s): iii-397-400 vol. 3, Montreal, Canada, ICASSP 2004 [2006-01-10]. http://ieexplore.ieee.org/Xplore/login.jsp? url=/ie15/9248/29345/01326565.pdf.

[4] COSATTO E, GRAF H P. Sample-based synthesis of photo-realistic talking-heads[C/OL]// Proc Computer Animation, June 1998, pp. 103-110, Philadelphia, Pennsylvania, June 8-10, 1998[2006-01-10]. http://potal.acm.ofr/citation.cfm? id=791528.

[5] COVELL M. Eigen-points: control-point location using principal component analyses[C/OL]// Proceedings of Conference on Automatic Face and Gesture Recognition, P122-127, Massachusetts, USA, October 1996[2006-01-10]. http://ieeexplore.ieee.org/Xplore/login.jsp? url=/ie13/4096/12122/00557253.pdf? arnumber=557253.

[6] MAHMOODI S, SHARIF B S, CHESTER E G, et al. Bayesian estimation of growth age using shape and texture descriptors [C/OL]// Image Processing and Its Applications, Conference Publication, Volume 2, Issue, 1999 Page(s): 489-493 vol.2, India, 1999[2006-01-10]. http://ieeexplore.ieee.org/Xplore/olgin.jsp? url=/ie15/6416/17139/00791096.pdf.

[7] 谢磊，冯伟，赵荣椿. 一种基于 MASM 的口形轮廓特征提取方法及听视觉语音识别[J]. 西北工业大学学报，2004, 22(5): 38-59.
XIE Lei, FENG Wei, ZHAO Rong-chun. A lip contour extraction method based on multiple active shape model for audio visual speech[J]. Journal of Northwestern Polytechnical University, 2004, 22(5): 38-59. (in Chinese)

[8] 王巍. 人脸面部特征定位和人脸识别方法的研究[D]. 北京：北京工业大学计算机学院，2003.
WANG Wei. Research on facial features localization and face recognition[D]. Beijing: College of Computer Science and Technology, Beijing University of Technology, 2003. (in Chinese)

Oral Features Localization Based on HASM

SUN Yan-feng, CHEN He, JIA Xi-bin, LI Jing-hua

(Beijing Municipal Key Laboratory of Multimedia and Intelligent Software Technology, College of Computer Science, Beijing University of Technology, Beijing 100022, China)

Abstract: The principal idea of this research for visual speech synthesis realism is that oral features localization provides the precise geometrical information for oral images of speech sound. In this research, Hierarchical Active Shape Model (HASM) is used as local texture model. In the structuring local texture model, the local texture model is the main clue. The optimal oral features localizations are decided by Mahalanobis distance. This research utilizes variable step strategy, variable angle strategy and oral images clustering strategy to greatly improve the accuracy and efficiency of inter lip localization and special lip shape. The result shows that the accuracy and efficiency are up to 90%.

Key words: features localization; active shape mode (ASM); local texture model; Mahalanobis distance