

文章编号 1004-924X(2007)07-1117-07

# 视觉驱动的语音合成系统中唇形轮廓的正交变换描述

李 刚<sup>1</sup>, 王蒙军<sup>1</sup>, 林 凌<sup>1</sup>, 曾锐利<sup>1,2</sup>

(1. 天津大学 精密仪器与光电子工程学院, 天津 300072; 2. 军事交通学院, 天津 300161)

**摘要:** 为了能够自动而且快速地获取唇读系统中所必需的唇形轮廓特征, 提出了将正交压缩变换的方法用于唇形轮廓的特征提取, 并对得到的唇形轮廓曲线进行了分析研究。通过离散傅里叶变换(DFT)和离散余弦变换(DCT)分别得到描述唇形轮廓特征的傅里叶描述子和离散余弦变换描述子, 然后将两类描述子作为唇形轮廓的特征向量, 采用隐马尔可夫模型(HMM)进行学习和识别。基于独立汉字发音的实验表明: 在达到 40% 的识别率时, 刻画唇形轮廓特征所需的离散余弦变换描述子数目为 15 个, 傅里叶描述子数目为 20 个。在相同的识别效果时, 刻画唇形轮廓特征所需的离散余弦变换描述子数目少于傅里叶描述子, 可减少数据运算量和运算所需时间。

**关键词:** 非对称模型; 唇形轮廓; 边缘特征; 傅里叶描述子; 离散余弦变换描述子; 隐马尔可夫模型(HMM)

**中图分类号:** TN912.34 **文献标识码:** A

## Lip contour description based on orthogonal transform in visual driven speech synthesis system

LI Gang<sup>1</sup>, WANG Meng-jun<sup>1</sup>, LIN Ling<sup>1</sup>, ZENG Rui-li<sup>1,2</sup>

(1. *School of Precision Instrument and Opto-electronics Engineering, Tianjin University, Tianjin 300072, China;*  
2. *Military Traffic Institute, Tianjin 300161, China*)

**Abstract:** In order to describe the lip contours in a lip reading system automatically and fleetly, orthogonal compression transformation was applied to the feature extraction of lip contours. Discrete Fourier Transform (DFT) and Discrete Cosine Transform(DCT) were used to get the descriptors of lip contours in the asymmetrical lip contour model. Then the Hidden Markov Model (HMM) was trained using two kinds of descriptors as the eigenvectors of lip contours. The experiments based on isolated Chinese words show that the number of DCT descriptors needed is 15, while the number of DFT descriptors is 20 at the same recognition rate of 40%. Experiments also show that the computing quantity and the consuming time are reduced obviously by the DCT at the same recognition rate.

**Key words:** asymmetrical model; lip contour; boundary feature; Fourier descriptor; discrete cosine transform descriptor; Hidden Markov Model

## 1 引言

随着模式识别技术的不断发展以及这些技术在人类实际生活中的逐步应用<sup>[1]</sup>,使得改善残疾人的生活水平成为可能。在聋哑人中,有一类人因疾病而切除了喉或声带,失去了语音交流能力,但他们仍拥有或可通过训练拥有正常发音时的唇部动作,所以,可以将人工智能领域中的唇读技术应用于语言残疾者的康复工程中。设计一个视觉语言驱动的实时语音合成系统,该系统能够识别讲话者唇动过程的图像序列,并将识别结果合成语音,从而还原或是部分还原出这一类人的语言表达能力。以往唇读技术研究<sup>[2-4]</sup>表明唇形及其动态特征在人类语言的理解过程中有重要作用,通过分析唇动图像序列来识别语言的音节,完全可应用于聋哑人与正常人交流辅助工具的研究。

唇形区域和唇形轮廓特征的生成是该系统研究的基本内容,也是唇动识别的前提条件。现有的特征生成方法主要有两类:(1)基于像素的方法:该方法直接将包含唇形区域的图像进行压缩变换,生成特征,然后将这些特征送入识别器。该方法虽然含有丰富的原始信息,但该类信息受环境影响较大,并且特征量是高维和高冗余的<sup>[5-6]</sup>。(2)基于模型的方法:该方法在完成唇形区域提取之后,采用部分主要参数来描述唇形轮廓。现阶段表征嘴唇轮廓主要采用了两种方法:(1)主动轮廓模型(Active Contour Model),该方法需要事先以手工方式在唇形轮廓上标定一定数目的点,将这些点的位置关系、距离作为参数送入识别器<sup>[7-8]</sup>。(2)可变模板(Deformable Template),该方法采用几条曲线来描述唇形轮廓,根据限定条件设计费用函数,使费用函数最小化,让曲线逼近唇形轮廓,将得到的曲线的参数送入识别器<sup>[9]</sup>。

然而无需人工干预和特别的限定条件,提取唇形轮廓特征,是自动唇形识别系统实用化的发展方向。本文介绍将正交压缩变换的方法用于唇形轮廓的特征提取。首先采用运动检测和数学形态学的方法<sup>[10-11]</sup>,获取唇动图像序列中非对称唇形轮廓模型的唇形区域以及唇形轮廓;然后对轮廓序列进行正交变换,得到 DFT 和 DCT 系数,将系数进行归一化得到傅里叶描述子和离散余弦描述子,作为唇形轮廓的特征向量;最后把两类描述

子作为观测序列输入隐马尔可夫模型(HMM)进行训练和识别,分析试验结果。

## 2 提取唇形区域和唇形轮廓

分析人讲话时嘴唇的动作过程可知,在每个音节发音过程中,不但嘴唇的张开闭合反映语音变化,而且随语音的变化嘴唇还会突起,所以唇动过程是一个三维过程,有必要获取嘴唇突起的信息。因此采用了非对称唇形轮廓的描述方法<sup>[12]</sup>,该轮廓模型如图 1,其中既有正面图像中嘴唇的内外轮廓动态变化信息,又包含人脸侧面图像反映出来的嘴唇突出度动态变化信息。

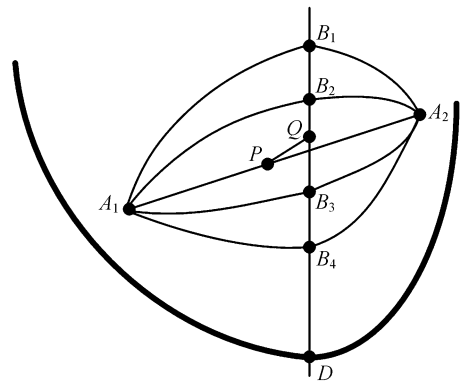


图 1 非对称唇形轮廓模型

Fig. 1 Model of asymmetrical lip contours

在讲话过程中,摄像头和人脸相对静止,所获得的唇形图像序列中,嘴唇是图像中的主要运动目标<sup>[10]</sup>,所以在提取唇形区域时采用运动检测和数学形态学的办法,对唇动灰度图像序列进行处理,处理后所得到的唇形区域虽然不是每一时刻所对应的唇形实际值,但是从中所获得的参数是随着唇动过程变化的,能够反映整个唇动的过程,用这些参数便能刻画一个音节发音的变化过程。

提取唇形区域和唇形轮廓的整个过程为<sup>[13]</sup>:

首先采用运动检测方法,通过计算间隔帧而不是相邻帧之间帧间差来进行唇形区域的定位,如公式(1):

$$\bar{\delta}_i(x, y) = |f_i(x, y) - \bar{f}_{i-k}(x, y)|, \quad (1)$$

这里  $f_i(x, y)$  为  $t$  时刻的图像帧,  $k$  为计算时帧与帧之间的间隔数目,  $\bar{f}_{i-k}(x, y)$  是为了减少图像的

噪声,利用模板  $\frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 \cdot & 1 \\ 1 & 1 & 1 \end{bmatrix}$  对前帧图像进行平滑处理。

然后将帧间差图像进行增强滤波,来获得明显的嘴唇区域,减少噪声的影响。这里先用结构元素对帧间差图像进行形态学开操作的运算,开操作的定义为(2)式:

$$\bar{\delta}_i(x, y) \circ \mathbf{B} = (\bar{\delta}_i(x, y) \ominus \mathbf{B}) \oplus \mathbf{B}, \quad (2)$$

这里  $\mathbf{B}$  代表结构元素  $[1 \ 1 \ 1 \ 1]$ 。之后可以用开闭( $O-C$ )运算或是闭开( $C-O$ )运算来消除图像中的亮细节或暗细节。

接着用二维最大熵阈值分割算法,将嘴唇区域作为目标从背景中分割出来,选取的最佳阈值向量( $s', t'$ )应该是:  $\Phi(s', t') = \max\{\Phi(s, t)\}$ ,其中  $\Phi(s, t)$  为熵的判别函数。

最后在获取唇形区域之后,采用了数学形态学的方法提取该区域的边缘作为唇形的外轮廓。在提取图像  $A$  的边缘时,可以通过先由结构元素

$$\mathbf{B} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 \cdot & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

对  $A$  腐蚀,而后用  $A$  减去腐蚀后的结果得到。即:

$$\text{Edge}(A) = A - (A \ominus \mathbf{B}), \quad (3)$$

采用以上方法对面向残疾人设计的汉语双模态语音数据库中的唇动图像序列进行处理,均能获得满意的唇形区域和唇形轮廓,图 2 中给出了采用以上方法所获的一个唇形区域和唇形轮廓线。

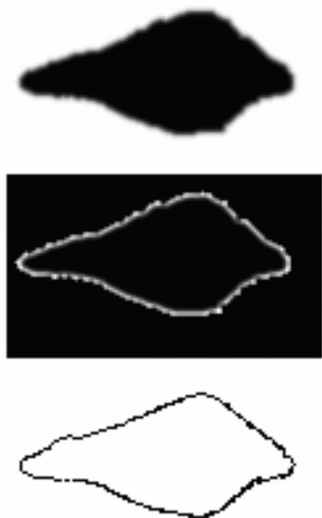


图 2 唇形区域和唇形轮廓

Fig. 2 Lip region and lip contours

在获取唇形区域后,从唇形区域提取嘴唇的宽度的投影  $W$ ,外唇线轮廓的高度  $H$ ,嘴唇突出度空间距离的投影  $F$  信息,并考察它们的差分特征参数  $dW/dt, dH/dt, dF/dt$ 。特征选择如图 3 所示。

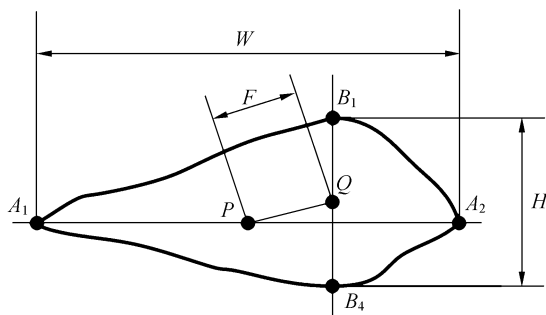


图 3 提取唇形特征

Fig. 3 Extracting lip features

### 3 唇形轮廓的边界正交变换

正交变换把空域中高度相关的像元变换为能量集中于弱相关或不相关的变换域系数,去掉了存在于相关性中的冗余度。经过正交变换,变换域里的相关性明显下降,能量明显集中于少数低频率系数,能更有效地进行图像的数据压缩和特征提取。

通常采用的正交变换有 K-L 变换,离散傅里叶变换(DFT),离散余弦变换(DCT)等。以 DCT 和 DFT 为例,变换具有能量集中于低频率系数、各系数不相关、高频系数衰减很快且能量较小等特点,略去能量较小的变换系数可以实现图像数据的压缩。

#### 3.1 边界离散傅里叶变换和傅里叶描述子

傅里叶描述子是物体形状边界曲线的傅里叶变换系数,它是物体边界曲线信号的频域分析的结果。区域形状可以看成是二维平面上的一条封闭的轮廓曲线( $x, y$ )。边界上某一点沿着边界移动的坐标变化是一个周期函数,可以展成傅里叶级数,级数的一系列系数直接和边界形状有关系。在曲线连续的条件下,当项数取得足够多时,用这些系数数值可以重建曲线的原始形状。因而可以利用其描述边界轮廓的形状,表示区域的几何特征<sup>[14]</sup>。

在数字图像中,对于边界形状是一条封闭的

曲线,沿边界曲线上所有点的序列可以表示为:  $\{x(l), y(l): l=0, 1, \dots, n-1\}$ , 在一维空间上用复数的形式表示为:

$$p(l) = x(l) + jy(l)$$

一维序列的离散傅里叶系数定义为:

$$z(k) = \frac{1}{n} \sum_{l=0}^{n-1} p(l) \exp(-j \frac{2k\pi}{n} l) \quad (k=0, 1, \dots, n-1), \quad (4)$$

其中,  $z(k)$  是  $p(l)$  的傅里叶变换, 是点序列在频域中的表示, 其傅里叶逆变换为:

$$p(l) = \sum_{k=0}^{n-1} z(k) \exp(j \frac{2k\pi}{n} l), \quad (l=0, 1, \dots, n-1) \quad (5)$$

利用傅里叶系数的性质, 在所有的傅里叶系数中, 只选取前  $p$  个, 消去从  $p+1$  以后的高频成分, 再进行傅里叶逆变换, 就可以得到和原来边界曲线近似的曲线。由于高频元素能很好地解释细节, 低频分量决定整体形状,  $p$  选择得越小, 边界细节失去的就越多, 原曲线中突变的部分将变得平滑。

由于傅里叶描述系数与边界形状的尺度、方向和曲线的起始点位置有关。为了识别具有旋转、平移和尺度不变性的形状, 需要对傅里叶描述系数进行归一化, 得到傅里叶描述子。

根据傅里叶变换的性质, 如果将边界形状的起始点位置平移  $a$  长度, 边界形状放大  $r$  倍, 旋转角度  $\Phi$  和平移位移  $(x_0, y_0)$  后, 那么所得到的傅里叶变换系数  $z'(k)$  为:

$$\begin{aligned} z'(k) &= F[r(x' + jy') \exp(j\Phi) + (x_0 + jy_0)] = \\ &= \text{rexp}(j\Phi) F(x' + jy') + F(x_0 + jy_0) = \\ &= \text{rexp}(j\Phi) F[x(l+a) + jy(l+a)] + F(x_0 + jy_0) = \\ &= \text{rexp}(j\Phi) \exp(j \frac{2k\pi}{n} a) F[x(l) + jy(l)] + F(x_0 + jy_0) = \\ &= \text{rexp}(j\Phi) \exp(j \frac{2k\pi}{n} a) z(k) + F(x_0 + jy_0), \end{aligned}$$

由上面的推导过程可知, 用傅里叶系数描述形状时, 系数的模值  $\|Z(k)\|$ ,  $k=1, 2, \dots, n-1$  具有旋转不变性和平移不变性, 并且与曲线起点的选择无关。当形状边界平移时, 仅改变  $Z(0)$  的值。把除  $Z(0)$  之外的系数模值  $\|Z(k)\|$  都除以  $\|Z(1)\|$ , 得到归一化的傅里叶描述子  $\varphi(k)$ ,

$$\varphi(k) = \frac{\|Z(k)\|}{\|Z(1)\|}, \quad k=1, 2, \dots, n-1.$$

这样得到的傅里叶描述子  $\varphi$  不会随着尺度变

化而变化, 同时具有旋转、平移和尺度不变的特性, 并且与曲线起始点的位置选择无关。

### 3.2 边界离散余弦变换和离散余弦描述子

由于 DCT 最接近于最佳变换—K-L 变换, 又具有较小的计算复杂性, 所以 DCT 在信号变换数据压缩中得到了广泛的应用<sup>[15]</sup>。目标灰度图像经分割后进行边缘提取, 获得其外形轮廓数据  $f(x[m], y[m])$ 。由于目标的边缘轮廓包含了其主要信息, 所以选择形状特征作为目标的特征矢量, 并用余弦变换描述子表达特征。

对同样的边缘曲线  $f(x[m], y[m])$  形成的一维复序列  $f(m) = x(m) + jy(m)$  ( $0 \leq m \leq N-1$ ) ( $N$  为闭合边缘的点数) 进行离散余弦变换, 有

$$\begin{aligned} z(0) &= \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} f(n), \\ z(k) &= \frac{2}{\sqrt{N}} \sum_{n=0}^{N-1} f(n) \cos \frac{(2n+1)k\pi}{2N}, \\ k &= 1, 2, \dots, N-1, \quad (6) \end{aligned}$$

和傅里叶描述子一样, 求解余弦变换描述子  $C(k)$ ,

$$\begin{aligned} X(k) &= \frac{2}{\sqrt{N}} \sum_{n=0}^{N-1} x(n) \cos \frac{(2n+1)k\pi}{2N}, \\ Y(k) &= \frac{2}{\sqrt{N}} \sum_{n=0}^{N-1} y(n) \cos \frac{(2n+1)k\pi}{2N}, \\ F(k) &= X(k) + jY(k), \\ |F(k)| &= \sqrt{X(k)^2 + Y(k)^2}, \\ C(k) &= |F(k)| / |F(1)| \quad (k \geq 2). \quad (7) \end{aligned}$$

和傅里叶描述子一样, 余弦变换描述子对目标具有平移、旋转及比例不变性, 对轮廓数据的起始点不敏感。余弦变换系数的低频部分反映了图像的整体轮廓, 高频部分仅刻画了外形的细节, 可以采用为数不多的余弦变换描述子来表述曲线轮廓, 达到降低数据维数的目的。

但是, 这一变换的核函数是:

$$\frac{2}{\sqrt{N}} \cos \frac{(2n+1)k\pi}{2N},$$

它为实数, 这样如果  $x(n)$  是实数, 那么它的 DCT 也是实数, 避免了 DFT 中的复数运算, 而且由于唇形轮廓变化形状、大小不同, 边沿检测后的唇形轮廓曲线上点的个数也不同, 不能采取序列补零来实现 FFT 运算, 因而不能对求解傅里叶描述子进行算法优化, 另外从 DFT 的核函数上看, 计算时要求进行复数运算, 这样会比 DCT 的运算量大很多, 所以 DCT

更适合于对实时性要求非常高,需要通过 DSP 和 CPLD 等硬件实现算法的场合,对于提高唇形检测和唇形识别的实时性,实现视觉语言驱动的声音合成系统具有积极意义。

#### 4 HMM 识别与实验结果

采用基于常见汉字的小字库,对小字库中的 30 个常见汉字发音过程的图像序列进行分析处理,所有图像分辨率为  $320 \times 240$ , 30 frame/s 拍摄,每个字由实验室的 10 名同学分别发音 10 次,其中的 8 次作为样本对隐马尔可夫模型(HMM)进行训练,2 次用来验证。采用隐马尔可夫模型(HMM)来识别唇动过程是由于隐马尔可夫模型(HMM)可以对非平稳随机序列进行随机模型估计,使其具有统计特性。隐马尔可夫模型(HMM)用隐含的状态对应于声学各相对稳定的发音单位,并通过状态转移和状态驻留来描述发音的变化,能够很好地刻画出人类语言中语音音节的发音过程<sup>[16]</sup>。

将两类描述子作为唇形轮廓的观测序列对隐马尔可夫模型(HMM)进行训练和识别。首先采用 Baum-Welch 迭代算法对训练样本集进行训练,获得模型参数的估计,训练过程对每个发音的唇动序列图像建立相应的隐马尔可夫模型(HMM)。识别过程将唇动序列图像的特征向量对每一个隐马尔可夫模型(HMM)用 Viterbi 算法计算  $P\{O|\lambda_i\}$ , 测试序列属于使其最大的那一类。

对两种描述子在选取不同数目时重建唇形轮廓进行了实验。实验结果如图 4 和图 5。

图 4 表示不同傅里叶变换描述子数目  $p$  重建唇形轮廓的效果,图 4(a)~(e)分别对应于原始轮廓, $p=5, 10, 15, 20$  时由傅里叶描述子重建唇形轮廓的效果。实验结果说明,在  $p$  选取不同的值时,对唇形轮廓的近似重建效果不同,随着  $p$  增加,逐渐接近原始轮廓。当  $p$  取值到 20 时基本上接近了原图像轮廓。

图 5 表示不同离散余弦变换描述子数目  $q$  重建唇形轮廓的效果,图 5(a)~(e)分别对应于原始轮廓, $q=5, 10, 15, 20$  时由离散余弦变换描述子重建唇形轮廓的效果。实验结果说明,在  $q$  选取不同的值时,对唇形轮廓的近似重建效果不

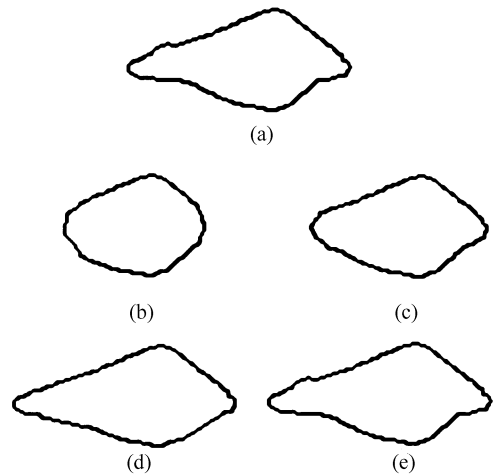


图 4 不同数目傅里叶描述子重建唇形轮廓  
Fig. 4 Rebuilding the lip contours with different amount of Fourier descriptors

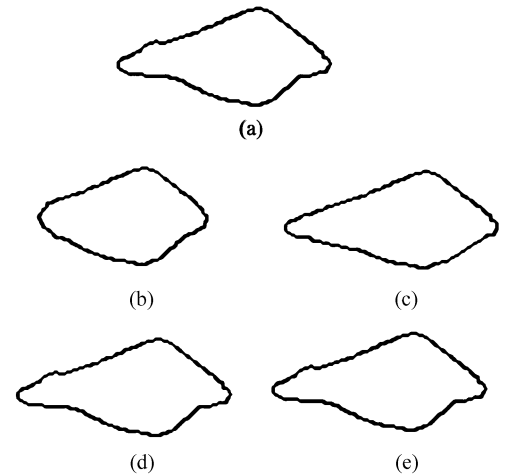


图 5 不同数目离散余弦变换描述子重建唇形轮廓  
Fig. 5 Rebuilding the lip contours with different amount of Discrete Cosine Transform descriptors

同,随着  $q$  增加,逐渐接近原始轮廓。当  $q$  取值到 15 时,就能够接近原轮廓曲线。

将两类描述子作为唇形轮廓的特征向量,应用隐马尔可夫模型(HMM)进行学习和识别,试验结果如表 1。实验表明,只要选取两类描述子中前面 15 或 20 个来刻画唇形轮廓曲线,均能完成唇形轮廓的识别,而在相同的识别效果下,所需的离散余弦变换描述子数目要少于傅里叶描述子数目。

表 1 HMM 下两种描述子的唇形识别结果

Tab.1 Recognizing rates in different descriptors based on HMM

	描述子数目			
	10	15	20	25
DFT 描述子	28.6%	34.4%	42.7%	44.2%
DCT 描述子	35.1%	43.2%	44.5%	44.3%

从实验结果中还可以看出,单纯依靠两种描述子得到的识别效果都不是很高,未来要结合从唇形区域提取嘴唇的宽度的投影  $W$ ,外唇线轮廓的高度  $H$ ,嘴唇突出度空间距离的投影  $F$ ,以及它们对时间的导数关系  $dW/dt, dH/dt, dF/dt$  等差分特征参数,来进一步提高识别效果。

## 5 结 论

采用运动检测和数学形态学的办法,自动地获得唇动图像序列中的唇形区域和唇形轮廓,该

过程不需要人工干预和特别的限定条件,避免了主动轮廓模型中手工标定的人工干预过程和可模板中费用函数的复杂计算过程,是自动唇形识别系统实用化的根本。采用正交变换的方法,对得到的唇形区域轮廓用傅里叶描述子和离散余弦变换描述子进行表示,通过将两种描述子作为观测样本序列对隐马尔可夫模型(HMM)进行训练,得到了每个发音的唇动序列图像相应的隐马尔可夫模型(HMM),能够完成唇动序列的识别验证。实验也表明,仅仅采用 15 或 20 个描述子作为唇形轮廓的特征向量就能实现唇动过程的识别,在达到 40% 的识别效果时,刻画唇形轮廓特征所需的离散余弦变换描述子数目为 15 个,傅里叶描述子数目为 20 个,即在相同的识别效果下,所需要的离散余弦变换描述子数目要少于傅里叶描述子数目。

## 参考文献:

- [1] 梁毅雄, 龚卫国, 潘英俊, 等. 基于奇异值分解的人脸识别方法[J]. 光学 精密工程, 2004, 12(5): 543-549.  
LIANG Y X, GONG W G, PAN Y J, *et al.*. Singular value decomposition-based approach for face recognition [J]. *Opt. Precision Eng.*, 2004, 12(5): 543-549. (in Chinese)
- [2] RAO R C T. Joint audio-video processing for multimedia[C]. *Proceedings of 22<sup>nd</sup> International Conference on Industrial Electronics, Control, and Instrumentation, Los Alamitos, USA; IEEE*, 1996, 1: 548-553.
- [3] ZHANG X., MERSEREAU R M, BROUN C C, *et al.*. Visual speech feature extraction for improved speech recognition[C]. *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, Piscataway, NJ, USA; IEEE*, 2002, 2: 1993-1996.
- [4] 刘鹏, 王作英. 多模式汉语连续语音识别中视觉特征的提取和应用[J]. 中文信息学报, 2004, 18(4): 79-84.  
LIU P, WANG Z Y. To extract and apply visual features in Mandarin multimodal continuous speech recognition[J]. *Journal of Chinese information processing*, 2004, 18(4): 79-84. (in Chinese)
- [5] KAYNAK M N, QI Z, CHEOK A D, *et al.*. Audio-visual modeling for bimodal speech recognition [C]. *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics, Piscataway, NJ, USA; IEEE*, 2001, 1: 181-186.
- [6] SCANLON P, REILLY R. Feature analysis for automatic speechreading [C]. *Proceedings of IEEE Fourth Workshop on Multimedia Signal Processing, Piscataway, NJ, USA; IEEE*, 2001: 625-630.
- [7] MATTHEWS I, COOTES T F, BANGHAM J A, *et al.*. Extraction of visual features for lipreading [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 24 (2): 198-213.
- [8] SEGUIER R, CLADEL N. Multiobjectives genetic snakes: application on audio-visual speech recognition [C]. *Proceedings of 4<sup>th</sup> EURASIP Conference focused on Video/Image Processing and Multimedia Communications, Croatia, Zagreb; Faculty of Electrical Engineering and Computing*, 2003, 2: 625-630.
- [9] CHANDRAMOHAN D, SILSBEE P L. A multiple deformable template approach for visual speech recognition [C]. *Proceedings of 4<sup>th</sup> International Conference on Spoken Language, Processing New York, USA; IEEE*, 1996, 1: 50-53.
- [10] LIE W N, HSIEH H C. Lips detection by morphological image processing[C]. *Proceedings of 4<sup>th</sup> International*

- Conference on Signal Processing, Piscataway, NJ, USA: IEEE*, 1998, 2: 1084-1087.
- [11] GRAF H P, COSATTO E, POTAMIANOS M. Robust recognition of faces and facial features with a multi-modal system[C]. *Proceedings of International Conference on Systems, Man, and Cybernetics, Piscataway, NJ, USA: IEEE*, 1997, 3: 2034-2039.
- [12] 李刚,王蒙军,林凌.采用非对称唇形轮廓模型提高汉语唇形识别效果[J]. *光学精密工程*, 2006, 14(3): 473-477.  
LI G, WANG M J, LIN L. Improving chinese lip-reading recognizing rate by unsymmetrical lip contour model[J]. *Opt. Precision Eng.*, 2006, 14(3): 473-477. (in Chinese)
- [13] LI G, WANG M J, LIN L. Extracting lip parameters in speech synthesis system driven by visual-speech [C]. *Proceedings of IEEE first International Conference on Innovative Computing, Information and Control, Los Alamitos, USA: IEEE-CS*, 2006, 2: 346-349.
- [14] PERSON E, FU K S. Shape discrimination using Fourier descriptors [J]. *IEEE Transactions on System, Man and Cybernetics*, 1977, 7(2):170-179.
- [15] 楚恒,朱维乐.基于DCT变换的图像融合方法研究[J]. *光学精密工程*, 2006, 14(2): 266-273.  
CHU H, ZHU W L. Image fusion algorithms using discrete cosine transform [J]. *Opt. Precision Eng.*, 2006, 14(2): 266-273. (in Chinese)
- [16] WILLIAMS J J, KATSAGGELOS A K, RANDOLPH MA. A hidden Markov model based visual speech synthesizer[C]. *Proceedings of International Conference on Acoustics, Speech, and Signal Processing, Piscataway, NJ, USA: IEEE*, 2000, 4:2393-2396.

**作者简介:**李刚(1959—),男,博士,天津大学精密仪器与光电子工程学院教授,主要研究领域为计算机视觉,生物信息检测与处理,智能仪器仪表。E-mail:ligang59@tju.edu.cn  
王蒙军(1978—),男,现为天津大学精密仪器与光电子工程学院博士研究生,主要研究方向为数字图像处理、分析,计算机视觉,模式识别。E-mail:ameng20040308@yahoo.com.cn