Journal of Computer Applications ISSN 1001-9081 2016-【-10
计算机应用,2016,【(【): 【-【 CODEN JYIIDU http://www.joca.cn

文章编号:1001-9081(2016)【-0000-00 doi:10.11772/j.issn.1001-9081.2016.【.0000

基于 SIFT 的说话人唇动识别研究

马新军，吴晨晨*，仲乾元，李园园

(1.哈尔滨工业大学(深圳) 机电工程与自动化学院，广东省 深圳市 518055；)

(*通信作者电子邮箱 870715761@qq.com)

摘　要: 针对唇部特征提取维度过高以及对尺度空间敏感的问题，提出了一种基于尺度不变特征变换算法（SIFT）做特征提取来进行说话人身份认证的技术。首先，提出了一种简单的视频帧图片规整算法，将不同长度的唇动视频规整到同一的长度，提取出具有代表性的唇动图片。然后提出一种在 SIFT 关键点的基础上，进行纹理和运动特征的提取算法，并经过 主成分分析（PCA）算法的整合，最终得到具有代表性的唇动特征进行认证。最后，根据所得到的特征，提出了一种简单的分类算法。实验结果显示，和常见的局部二元模式（LBP）特征和方向梯度直方图（HOG）特征相比较，该特征提取算法的错误拒绝率和错误接受率表现更佳。说明整个说话人唇动特征识别算法是有效的，能够得到较为理想的结果。

关键词: 唇部特征；尺度不变特征变换算法；特征提取；识别；分类

中图分类号:TP391.41　　文献标志码：A

Research on Lip Motion Recognition of Speaker Based on SIFT

MA Xinjun, WU Chenchen*, ZHONG Qianyuan, LI Yuanyuan

(1.College of Mechanical Engineering and Automation, Harbin Institute of Technology(Shenzhen), Shenzhen Guangdong 518055, China)

Abstract: Since the feature of lip has high dimension and is sensitive to scale space, a method to do feature extraction for speaker authentication based on scale-invariant feature transform (SIFT) algorithm was proposed. Firstly, a simple video frame image neat algorithm was proposed. Namely, adjusting lip motion videos with different lengths to the same length and extracting representative lip motion pictures. Then, the other algorithm based on key points of SIFT was presented, through which texture and motion features can be extracted. Integrated with principal components analysis (PCA) algorithm,, typical lip motion features can be obtained to do identity recognition. Finally, a simple classification algorithm was presented according to obtained features. Compared with the common local binary model (LBP) feature and histogram of oriental gradient (HOG) feature, experimental results show that false acceptance rate (FAR) and false rejection rate (FRR) of the proposed feature extraction algorithm peform better, proving that the whole speaker lip motion recognition algorithm is effective and can get ideal results.

Keywords: lip feature; Scale-invariant feature transform (SIFT); feature extraction; recognition; classification

0　引言

近年来越来越多的研究表明生物认证技术相比于传统的身份认证具有更好的安全性与简便性。唇动身份认证原来作为语音认证的辅助信息，现在已经独立出来成为一种新的认证手段，唯一性和准确性都得到了研究的证明[1-2]。唇动身份认证系统主要由四部分组成：在已建立的数据库的基础上，首先对获取的图像进行人脸的定位，进而做唇部定位；然后对得到的图片进行预处理；再进行特征提取；最后根据所得到的特征分类得出结果，即完成整个说话人唇动识别研究。

人脸检测方面 YANG[3]等提出了基于马赛克图进行人脸检测的方法。Kouzani[4]利用人工神经网络分别对人脸的眼睛、鼻子和嘴等器官进行检测。Sirohey[5]通过使用人脸边缘信息和椭圆拟合的方法，从复杂的背景中分割定位出人脸区域。Miao[6]等从输入图像中提取面部器官水平方向的马赛克边缘,将各段边缘的“重心”与“重心”模板进行匹配,再通过灰度和边缘特征进行验证以实现人脸的检测。梁路宏，艾海舟等[7]给出了一种基于多关联模板匹配的人脸检测方法。自 Viola 和 Jones 首次将 Adaboost 算法用于人脸检测以来，由于性能和速度优势，其便成为一种主流的人脸检测算法。由于其应用的广泛性和实用性，本文采用 Adaboost 算法作为人脸定位的算法。

收稿日期: 2016-00-00;　修回日期:2016-00-00。基金项目:国家自然科学基金资助项目（51677035）；深圳市基础研究项目（JCYJ20150513151706580）；深圳市科技计划项目（GRCK2016082611021550）

作者简介: 马新军(1972—)，男，新疆石河子人，副教授，博士，主要研究方向：图像处理及模式识别，智能汽车及智能驾驶，生物识别；吴晨晨(1993—)，女，河南濮阳人，硕士研究生，主要研究方向：模式识别；仲乾元（1990-），男，江苏徐州人，硕士研究生，主要研究方向：模式识别；　李园园（1993—)，女，河南许昌人，硕士研究生，主要研究方向：模式识别。

人脸的定位完成后，常见的唇部定位方法主要为对图像灰度投影的峰值进行分析，进而通过颜色空间变换，对唇部区域进行加强，再经过阈值的分割得到所需的唇部区域[8]。本文提出根据人脸各部分的大致比例关系给出一种唇部的粗定位算法，该算法计算简单，同时可以保证唇部边缘的一些运动与纹理特征不会被忽略。

图片的预处理工作，是在前期对图片进行处理，减少噪音，遮挡，光照不均等影响，使得特征提取能够得到更加稳定准确的特征向量。本文在 SIFT 算法[9-11]的基础上，进行算法的改进与调整。由于算法本身已包含高斯去噪功能，并且实验的光照条件变化不大，因此对于图像的预处理算法不做过多的讨论。

完成唇部定位与预处理之后，特征提取是关乎到整个认证系统稳定性与准确率的重要部分。目前的唇部特征提取主要分为三类：（1）唇部的纹理特征；（2）唇部的几何特征；（3）唇部的运动特征。纹理特征方法主要有：Scholkopf 等人用核技术将经典的 PCA 推广至核主成分分析（Kernel BasedPrincipal Component Analysis, KPCA），提取了高维特征空间中的线性鉴别特征，即就是原始输入空间中的非线性鉴别特征。Jian Yang 等人在 2004 年提出了二维主元素分析。但 PCA 存在着面对非线性特质无能为力，以及可能会忽略重要的投影方向等缺点。Timo 等人[12]使用局部二元模式(Local Binary Pattern, LBP)来提取脸部图像的纹理特征，对脸部区域进行分块计算各分块 LBP 直方图，并将它们连结起来作为表情识别的特征。LBP 特征具有较好的光照鲁棒性，但是作为一种静态特征，无法具有代表性的来表征动态的特征。H. Ertan Çetingül 等将二维图像从空间域转换到频率域，获取的特征仍然存在大量的冗余信息。几何特征主要有唇部的长宽高等人工提取的特征，对于唇部的轮廓 Kass 等人[13]在第一届国际视觉会议上提出了 Snake 模型。关于运动特征：Haralick 等人将二维物体表面划分成许多小块，假设每个小块内像素点的运动方向和速度近似且短时间内为常量，由此得到了光流的附加约束条件，光流法作为常用的运动图像处理方法，也存在着运算量大的问题。Preety Singh 等人提出[14]三正交平面窗口，唇动的运动特征能够在一个时空体积内进行表征。本文给出了一种在 SFIT 基础上的特征提取算法，既有运动的表述，又有纹理的描述，同时对于旋转变化具有一定的鲁棒性。

对所提取的特征进行分类的算法目前也有很多研究成果。高斯混合模型(Gaussian Mixture Model, GMM) [15]是唇动识别和认证领域的一种常用的分类算法，算法简单，但在数据较多的情况下分类结果不是很理想；Adaboost 和 PCA-LDA (Principal Component Analysis & Linear Discriminant Analysis)，支持向量机（Support Vector Manhine, SVM）算法在唇动认证中也是较为常用的分类算法；Yang 等人[16]提出了自调节分类面支持向量机（self-adjusting classification-plane SVM, SCSVM）方法，通过学习过完备的稀疏特征，可以在高维特征空间提高特征的线性可分性，大大降低了训练分类器的时间和空间消耗。Juan C 等将 SVM 算法用在唇动认证中，取得了较好的分类效果。基于深度学习的特征提取和分类算法是目前最为先进和火热的算法，主流的深度学习模型包括自动编码器、受限波尔兹曼机、深度信念网络、卷积神经网络等。Hinton 等人[17]通过这种方式，成功将其应用于手写数字识别、语音识别、基于内容检索等领域。本文在之前所得到的唇动特征基础上，提出一种简单的分类算法，既满足了分类的精确性，同时计算量小，实时性较好，在数据库较大时也可以和神经网络的分类算法相结合。本文的安排如下： 第一章主要描述所提出的帧图片提取算法和实验结果。第二章对 SIFT 算法进行回顾，同时提出相关的改进部分，并给出改进后的实验结果，介绍基于 SIFT 算法的特征提取算法。第三章，给出整个唇动身份认证系统的实验数据与结果。最后相关的结论将在第四章给出。

1 帧图片提取

在唇动视频中，录像的帧率一般为 30fps，如果直接将视频产生的所有帧图片都作为下一步特征提取的数据库，不仅会有大量的噪声干扰在其中，还会有大量的数据冗余，从而会加大系统的计算量并影响其鲁棒性与运算的实时性，最终降低系统认证的准确率与效率。文献[18]分析了动态时间规整算法（Dynamic Time Warping, DTW），本文给出了一种基于时间序列的动态图片提取算法，在相邻的时间段里找到帧间灰度变化最大的图片作为代表性图片。该算法不仅可以减少计算量，同时可以增强整个系统对于说话人语速变化的鲁棒性。具体实行过程如下：

（1）令唇动视频所产生的帧图片的数量为 X；

（2）如果 X 的数量小于 20，说明说话者说话的时间小于 1s，作为认证而言，说话长度明显过短，提示唇动视频所提供的帧图片数量过少，无法进行认证；

（3）如果 $20 \le X \le 60$，选取第 3 张图片作为所提取的第 1 张帧图片，选取倒数第 3 张图片作为第 12 张帧图片。$A = \lfloor (X-10)/10 \rfloor$ 和 $B = \lceil (X-10)/10 \rceil$ 将依次作为选取图片的数量间隔，每一幅帧图片的大小为 M*N，在每个间隔中，用公式（1）选取所要的帧图片：

$$Pic = \left\{ \sum_{i=1}^{M} \sum_{j=1}^{N} \left| I^{t}(i,j) - I^{t-1}(i,j) \right| + \sum_{i=1}^{M} \sum_{j=1}^{N} \left| I^{t}(i,j) - I^{t+1}(i,j) \right| \right\}_{\max} \quad (1)$$

其中 $I^{t}(i,j)$ 表示第 t 帧图片 I 在点 (i,j) 处的灰度值；Pic 表示在间隔内，和相邻帧图片比较，灰度值变化最大的图片。

（4）如果 $60 < X \le 480$，选取第 10 张图片作为所提取的第 1 张帧图片，选取倒数第 10 张图片作为第 12 张帧图片。$A = \lfloor (X-40)/10 \rfloor$ 和 $B = \lceil (X-40)/10 \rceil$ 将依次作为选取图片的数量间隔，在每个间隔中，用公式（2）选取所要的帧图片：

$$Rc = \left\{ \sum_{i=1}^{M} \sum_{j=1}^{N} \left| I^{t}(i,j) - I^{t-1}(i,j) \right| + \sum_{i=1}^{M} \sum_{j=1}^{N} \left| I^{t}(i,j) - I^{t+1}(i,j) \right| \right\}_{\max} \quad (2)$$

（5）如果 $480 < X$，说明说话者说话的时间长于 16s，作为认证而言，说话时间过过长，提示唇动视频所提供的帧图片数量过多，无法进行认证。

通过上述的算法，可以从唇动视频所产生的大量帧图片中提取 12 张代表图片。

在被测试者的两段视频帧图片中用上述算法提取的代表图片如下，第一遍段视频用正常语速说‘你好’，第二遍张大嘴巴放慢语速复述。

(a)正常语速下提取的代表图片

(b)慢速大口型下提取的代表图片

图 1 视频代表图片

Fig 1 Representatives extracted from video

从图 1 中可以看出在语速和不同口型的情况下，所提取的 12 张图片其对应的序列及口型都有很强的相似性与代表性。该算法对于说话者语速的变化和口型大小的改变都有很强的鲁棒性，并能够为后面特征的提取打下良好的基础。

2　基于 SIFT 算法的特征提取和匹配模型

SIFT 算法第一次由 David Lowe[19] 提出，是一种广泛应用于图像处理的算法，具有良好的尺度不变性和对旋转的抵抗性。Sambit Bakshi 曾将 SIFT 算法用在唇印的认证与对比中，并取得了很好地效果[20]。

SIFT 算法所提出的关键点的描述方式，作为一种局部特征，对于光照，旋转，噪音与尺度的变化都不敏感，因此在这基础上进行物体的认证和识别，都具有很强的抗干扰性和针对性。这种局部特征检测算法概括的讲，就是通过在不同的尺度空间中得到关键点描述子，再对关键点进行匹配的方法，SIFT 算法的流程图如图 2 所示。

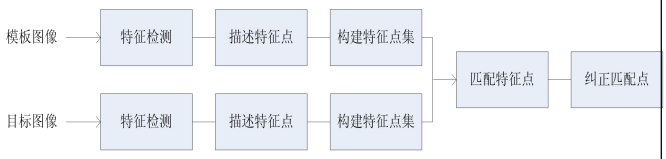
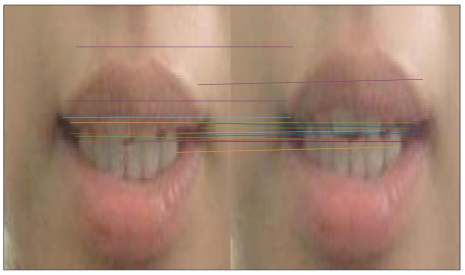


图 2 SIFT 算法流程图

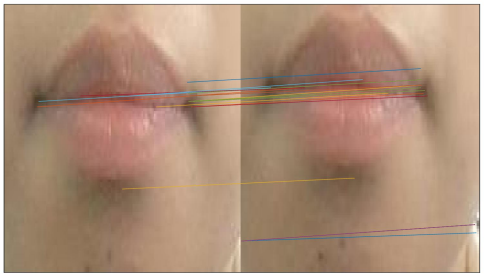
Fig 2 Flow chart of SIFT algorithm

2.1　基于 SIFT 的目标匹配与认证

SIFT 算法虽然具有很好地鲁棒性，但是计算比较复杂，在图片分辨率不高和轮廓特征不明显的情况下，表现不佳。由于数据库图像摄像头的像素为 30 万，直接用 SIFT 算法对图像进行关键点的提取与匹配，不仅实时性差，而且无法产生效果。又因为说话人距离同一摄像头的远近变化并不是很大，所以适当的减少了高斯金字塔每一塔的层数，本文中取 4-5 层。为了在较低像素的情况下，增加 SIFT 匹配点数，将对比度的条件适当放宽，取 $\left| D(\hat{X}) \right| \ge 0.01$，取 ratio=0. 4。



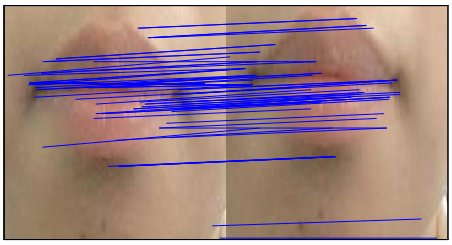
（a）示例 1



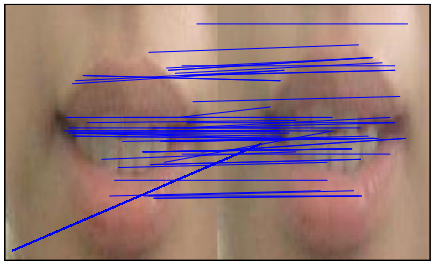
（b）示例 2

图 3 测试者未经参数调整的 SIFT 匹配示例图

Fig 3 Matching pictures through SIFT without adjusting parameter



(a) 示例 1

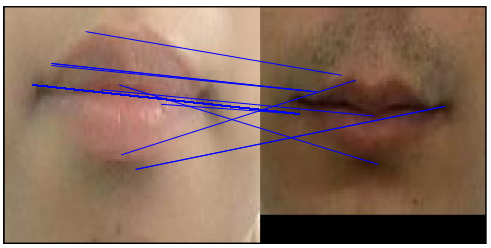


（b）示例 2

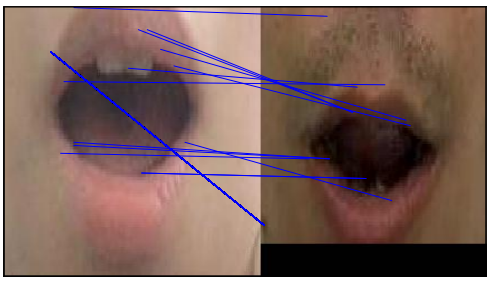
图 4 测试者经参数调整后的 SIFT 匹配示例图

Fig 4 Matching example pictures through SIFT with adjusting parameter

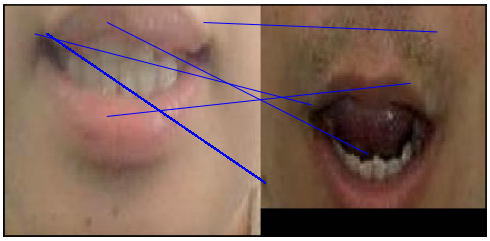
从图 3,4 中可以看出，经过参数的调整后，SIFT 算法所提取的关键点个数明显变多,但是在唇部轮廓变化较大的情况下，误匹配点的个数也有所增加。因此在后面的特征提取中增加了消除重复关键点和 PCA 降维的步骤。图 5 展现的是不同人不同尺寸的图片的 SIFT 匹配结果，可以看到两幅图片的匹配点数明显减少且明显存在匹配错误点。



(a) 示例 1



（b）示例 2



(c) 示例 3

图 5 经参数调整后的不同人唇部图片 SIFT 匹配示例图

Fig 5 Matching example pictures of different persons through SIFT with adjusting parameter

综合实验结果可以看出不同的人的唇部无论出于何种口型，能够匹配的关键点个数远少于同一个人的唇部所能匹配的关键点的个数。因此，将采用测试样本与数据库样本的关键点匹配个数的比值作为判断是否为同一个人的有效依据。

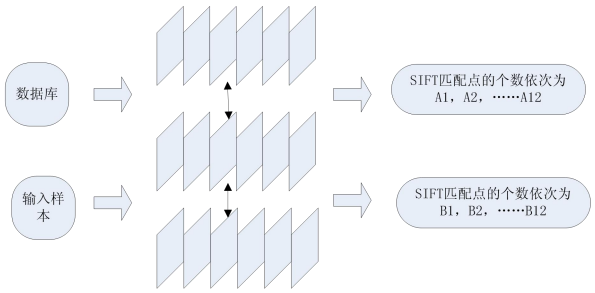


图 6 目标匹配原理图

Fig 6 Target matching schematic

如图 6 所示，用已提到过的帧图片提取算法将数据库中所存放的同一个人所说同一句话（比如说了 3 遍）的 12 幅帧图片一一进行 SIFT 匹配，共可以匹配 3 次，将匹配点的个数，求平均值得到 A1，A2，…， A12，将其存储起来。然后将测试样本与数据库中的任意样本进行 SIFT 匹配，得到匹配点的个数 B1，B2，…， B12。设置阈值θ=0.4，i=1, 2，…，12。如果 Bi/Ai<θ,则计数标志 flag+1,为了防止系统的误判断，和降低噪声图片带来的干扰，设置当 flag 的值大于 2 时，判定为不是用户本人。通过调节阈值θ的大小，可以调整错误拒

绝率和错误接受率的大小。θ值越大错误接受率越小但错误拒绝率越大，θ值越小错误接受率越大但错误拒绝率越小。

2.2 基于 SIFT 的新的特征提取算法

如图 7 所示，首先对数据库中的样本（即 12 幅帧图片）相邻的图片进行 SIFT 匹配，得到匹配的关键点。提出的特征提取算法就是在这些关键点的基础上得到的。

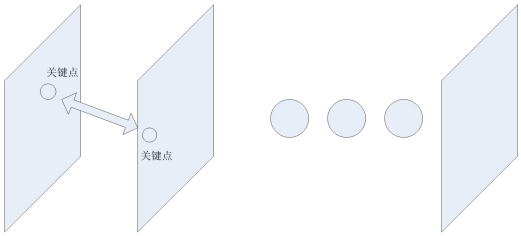


图 7 相邻序列间匹配关键点

Fig 7 Matching key points between adjacent frames

具体的特征提取算法如下：

（1）对于任意两帧图片之间匹配得到的关键点 P_1，P_2：

用公式（3）来计算关键点 P1，P2 的运动矢量幅值：

$$f_1(i,j)=\sqrt{\left(i_{p1}-i_{p2}\right)^2+\left(j_{p1}-j_{p2}\right)^2} \tag{3}$$

其中，(i_{p1},j_{p1}) 为关键点 P1 的坐标位置；(i_{p2},j_{p2}) 为关键点 P2 的位置坐标。

用公式（4）来计算关键点 P1，P2 的运动矢量的方向：

$$f_2=\tan^{-1}\left[\frac{\left(j_{p1}-j_{p2}\right)}{\left(i_{p1}-i_{p2}\right)}\right] \tag{4}$$

对于每一对匹配的关键点，通过这种方式可以得到二维的特征向量 $\boldsymbol{F}=(f_1,f_2)$。

（2）对于图像中每一个关键点，选取 4×4 的窗口，如图 8 所示。图 8 中每一个小方格代表着一个像素点，圆点代表所得到的关键点的位置，其周围 4×4 的像素点的运动特征矢量方向由箭头所表示，该矢量幅值的大小表示其矢量的大小。最后将计算所得到 16 个矢量归类统计到 8 个主要的方向上去，作为最后得到的 8 维特征向量。具体的计算方法由公式（5）和（6）给出：

梯度幅值：

$$m(x,y)=\sqrt{\left(L(x+1,y)-L(x-1,y)\right)^2+\left(L(x,y+1)-L(x,y-1)\right)^2} \tag{5}$$

其中 $L(x,y)$ 表示在点 (x,y) 处的灰度值。

梯度方向：

$$\theta(x,y)=\tan^{-1}\left[\frac{L(x,y+1)-L(x,y-1)}{L(x+1,y)-L(x-1,y)}\right] \tag{6}$$

通过上述的算法，可以得到 8 维的特征向量 $\boldsymbol{R}$。

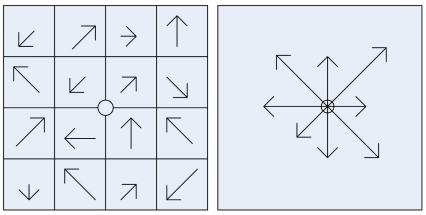


图 8 关键点周围运动矢量特征图

Fig 8　Feature extraction of motion vectors around key points

（3）对于图像中每对匹配的关键点，选取 4×4 的窗口，对 4×4 窗口中对应位置的灰度值做差取绝对值，然后将这 16 个值求和，如图 9 所示：

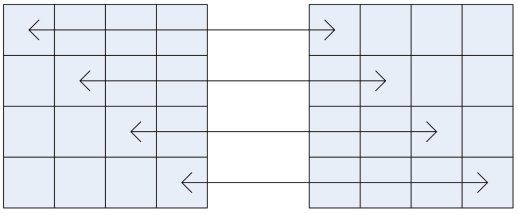


图 9 对应位置的灰度差绝对值

Fig 9 Absolute value of gray difference of corresponding position

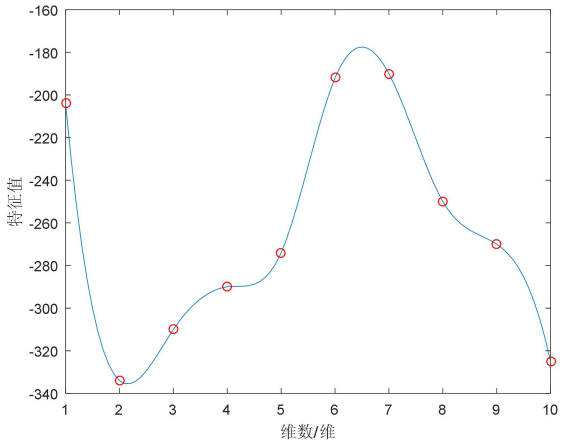
即，

$$G=\sum_{k=1}^{16}\left|I^1(i,j)-I^2(i,j)\right| \tag{7}$$

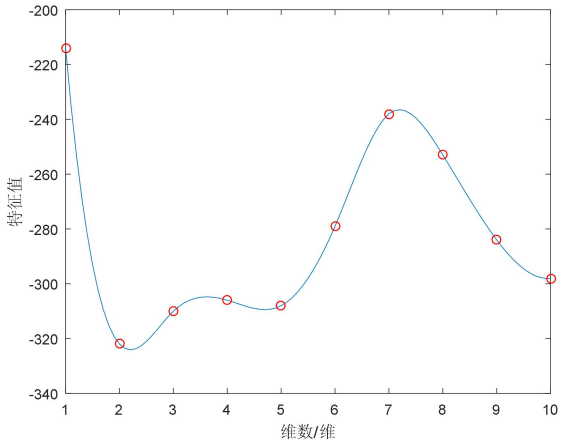
其中 I^1 和 I^2 分别代表相邻的两幅帧图片中对应点的灰度值。

（4） 综上所述，对于每一个匹配的关键点，可以得到一个 11 维的特征向量 $\boldsymbol{T}=\{\boldsymbol{F},\boldsymbol{R},G\}$。这 11 维向量中包含了唇部的运动信息 $\boldsymbol{F}$，唇部周围的纹理信息 $\boldsymbol{R}$，以及灰度的变化信息 G。假设最终得到的匹配点个数为 n，对最终得到的特征矩阵 $\boldsymbol{M}$ 采用 PCA 降维到 11 维，得到 11 维特征向量 $\boldsymbol{Z}$。$\boldsymbol{Z}$ 特征相比于 LBP 等常见的纹理信息具有更强的针对性和规律性。相比于 Snake 算法所提取的轮廓特征，具有更少的模型依赖性和更强的鲁棒性。

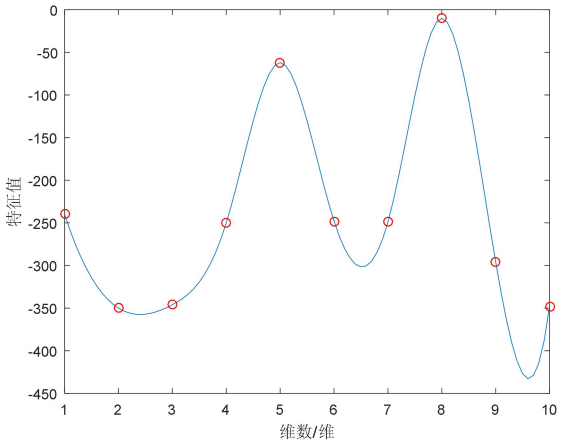
图 10 所展示的是两名测试者说同一段话所提取的特征矢量的曲线图。灰度变化累计值的值较大，为了能看出其变化趋势，因此在曲线图中只画出特征 $\boldsymbol{F}$ 和特征 $\boldsymbol{R}$。



（a）测试者第一遍说‘你好’的特征曲线图



（b）测试者第二遍说‘你好’的特征曲线图



（c）另一名测试者说“你好”的特征曲线图

图 10 特征值曲线图

Fig　10 Eigenvalue curve

从图中可以看出，本文所提出的这种特征提取方法能够很好的表征说话人的说话特征，具有很强个人特征以及区别性，可以很容易的进行分类。

2.3　基于所提取特征的分类算法

以往的分类方法，由于得到的图像特征并不明显，因此常用 SVM，神经网络，隐马尔科夫模型等算法对其进行分类，这些算法需要较大的数据库来训练，同时运算量大，计算起来十分复杂。根据前面所得到的特征，本文通过简单的比较方法进行二分类，也能得到很好地实验效果。具体的实现方法如下：

（1）通过公式（8）得到数据库中唇动视频中所提取的唇动特征的平均值：

$$Z_{mean}=\left(\sum_{i=1}^{n}Z_i\right)\Big/n \tag{8}$$

（2）再用相同的方法将测试样本的唇动特征 Z 提取出来；

（3）设置阈值 θ_1,θ_2，其中 $\theta_2=\dfrac{1}{\theta_1}$。$\theta_1$ 的值越大越大错误接受率越小但错误拒绝率越大。本论文中 θ_1 取 0.7，θ_2 取 1.42。

令 $t=z/z_{mean}$ 如果 $t<\theta_1$ 或者 $t>\theta_2$，则令计数标志 flag+1，为了防止误判断，flag 的值大于 2 时，则判断该用户所说的不是这段话。

3　说话者唇部特征识别实验结果

3.1　说话者唇部特征识别流程简介

进行说话者唇部特征识别的流程图如图 11 所示。首先对数据库样本进行人脸定位与唇部定位，然后进行帧图片的选取，对选取的帧图片进行 SIFT 匹配，在此基础上，提取特征并记录匹配结果。对测试样本采用同样的步骤，最终根据本论文提出的验证与分类方法，将输出结果与数据库中的结果比较，得出判定结果。

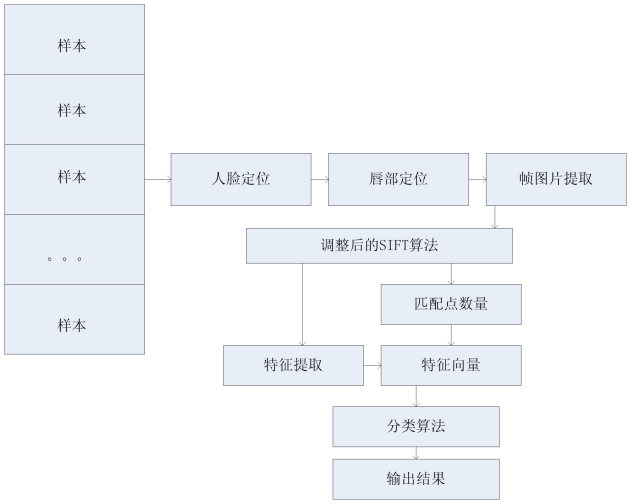


图 11 说话者唇动识别示意图

Fig　11 lip motion identification diagram of speaker

3.2　数据库的搭建

(1) 视频数据库的基本参数：

视频格式：AVI；

颜色空间：YUY2；

输出大小：640*480；

视频输出帧率：30fps；

(2) 视频数据库的搭建：

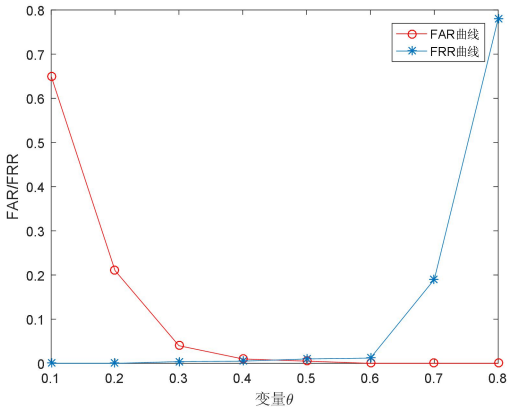
采样人数：50 人；

采样环境：正常的日光灯照明，人脸位置相对固定，无遮掩，无大角度旋转，无模糊，胡须，光照角度变化等复杂条件设置。

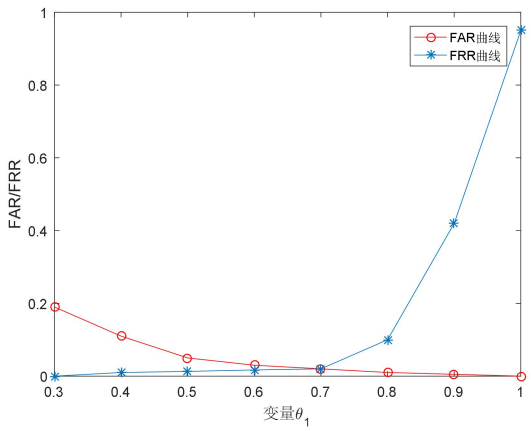
采样过程：接受采样的样本，分别以正常语速重复短句“你好”，以及数字 1 到 9 等不同的长短句各四遍，再分别以较慢语速张大口型的情况重复各四遍。

3.3　实验结果

表 1，2 展示了本实验在不同唇动视频中获得的 $\boldsymbol{Z}$ 特征的错误接受率(False Acceptance Rate, FAR)和错误拒绝率(False Rejection Rate, FRR)，以及在相同条件下 LBP 和 HOG 特征的 FAR 和 FRR 值。图 12 展示在不同的 θ 和 θ_1 值时，FAR 和 FRR 值的变化情况。



（a） θ 值不同时的 FAR 和 FRR 曲线



（b） θ_1 值不同时的 FAR 和 FRR 曲线

图 12 FAR 和 FRR 曲线

Fig. 12 FAR and FRR curves

表 1 FAR 实验结果

Tab. 1 Experimental results of FAR

说话内容	FAR（$\boldsymbol{Z}$特征）	FAR（LBP 特征）	FAR（HOG 特征）
1	0.85%	11.4%	16.33%
2	0.93%	22.13%	26.6%
4	0.74%	15%	17.62%
5	0.5%	11.3%	21.7%
9	0.94%	10.14%	18.5%
你好	1.2%	20.3%	32.6%

表 2 FRR 实验结果

Tab. 2 Experimental results of FRR

说话内容	FRR（$\boldsymbol{Z}$特征）	FRR（LBP 特征）	FRR（HOG 特征）
1	5.65%	23.4%	34.7%
2	5.88%	27.3%	23.6%
4	5.23%	32%	42%
5	4.89%	17.73%	21.46%
9	5.15%	31.4%	25.5%
你好	5.96%	29.8%	31.67%

4　结语

本文介绍了一种针对动态视频所产生的帧图片的提取算法。这种算法可以增强对于语速变化，口型大小变化以及照明变化的鲁棒性。SIFT 算法被引进到了说话者唇动识别中在通过参数的调整后有很好的表现。在 SIFT 算法的基础上提出了一种新的唇动特征提取方法，这种方法既包含纹理特征又包含运动特征，可以准确地描述说话人唇动的一系列特征。最后，在匹配点与所提取的特征的基础上，分别提出了一种验证与分类的算法，方法简单，计算量小，与常用的 LBP 和 HOG 特征相比较可以得到更为准确和有效的结果，实现说话人的唇部特征识别。在后面的实验中可以添加图像预处理的算法，将多种特征提取方法相结合以及引入神经网络来增强系统的稳定性和适应性。

说明：正文中 $\boldsymbol{F}$, $\boldsymbol{R}$, $\boldsymbol{T}$, $\boldsymbol{Z}$ 为向量，$\boldsymbol{M}$ 为矩阵

参考文献

[1] KANAK A, Erzin E, YEMEZ I, et al. Joint audio-video processing for biometric speaker identification [C]// Proceedings of the 2003 IEEE International Conference on Multimedia and Expo. Baltimore: IEEE, 2003: vol. 3, 561-564.

[2] CETINGUL H E, YEMEZ Y, ERZIN E, et al. Discriminative Analysis of Lip Motion Features for Speaker Identification and Speech-Reading[J]. IEEE Transactions on Image Processing A

Publication of the IEEE Signal Processing Society, 2006, 15(10):2879-2891...

[3] YANG G Z, HUANG T S. Human face detection in a complex background Pattern[J]. Pattern Recognition, 1994, 27 (1):53-63.

[4] KOUZANI A Z, HE F, SAMMUT K. Commonsense knowledge-based face detection[C]// Conference on Intelligent Engineering Systems, Budapast: IEEE, 1997: 215-220.

[5] SIROHEY S A. Human face segmentation and identification[J]. Technica Report, 1993:CS-TR-3176.

[6] MIAO J, YIN B C, WANG K Q, et al. A hierachical multiscale and multiangle system for human face detection in a complex background using gravity center template[J]. Pattern Recognition. 1999, 32 (10): 1237-1248.

[7] 梁路宏，艾海舟，何克忠等.基于多关联模板匹配的人脸检测.软件学报[J],2001,12(1): 94-102.(LIANG L H, AI H Z, HE K Z, et al. Face detection based on multi-association template matching [J]. Journal of Software, 2001,12(1): 94-102.)

[8] ASHLEY D, GRITZMAN, DAVID M, et al. Comparison of colour transforms used in lip segmentation algorithms[J], Signal, Image and Video Processing. 2015, 9(4):1-11.

[9] NEERU N, KAUR L. Modified SIFT descriptors for face recognition under different emotions [J], Journal of Engineering. 2016(2):1-12.

[10] KIRCHNER M R. Automatic thresholding of SIFT descriptors [C]//IEEE International Conference on Image Processing. Phoenix,: IEEE, 2016:291-295.

[11] 许佳佳,张叶,张赫.基于改进 Harris-SIFT 算子的快速图像配准算法[J]. 电子测量与仪器学报，2015,29(1):48-54. (XU J J, ZHANG Y, ZHANG H. Fast image registration algorithm based on improved Harris-SIFT descriptor [J]. Journal of Electronic Measurement and Instrumentation, 2015, 29(1):48-54.

[12] AHONEN T, HADID A, PIETIKANINEN M. Face recognition with local binary patterns[C]// European Conference on Computer Vision, Berlin: Springer, 2004: 469-481.

[13] KASS M, WITKIN A, TERZOPOULOS D. Snakes: active contour model[C]// In Brady IM, Ro senfield A eds Proceedings of the 1st International Conference on Computer Vision. London: IEEE Computer Society Press, 1987: 259-268.

[14] SINGH P, LAXMI V, GAUR MS. Speaker identification using optimal lip biometrics[C]// Iapr International Conference on Biometrics. New Delhi: IEEE, 2012:472-477.

[15] SAEED U. Person identification using behavioral features from lip motion[C]// IEEE International Conference on Automatic Face & Gesture Recognition & Workshops. Santa Barbara,: IEEE, 2011: 131-136.

[16] YANG J C, YU K, GONG Y H, et al. Linear spatial pyramid matching using sparse coding for image classification[C]// ICCV:2009:Proceedings of the IEEE Computer Society Conference on Computer Vision & Pattern Recognition. Miami: IEEE, 2009: 1794-1801.

[17] KRIZHEVSKY A, SUTSKEVER I, HINTON GE. Imagenet classification with deep convolutional neural networks[C]// International Conference on Neural Information Processing System. Lake Tohoe: Current Associate Inc, 2012: 1097-1105.

[18] 杨洁,康宁.动态时间调整 DTW 算法的研究[J].科技与创新, 2016(4): 11-12. (YANG J, KANG N. Research on dynamic time regular DTW algorithm [J]. Science and Technology & Innovation, 2016(4): 11-12).

[19] LOWE DG. Distinctive Image features from scale-Invarient keypoints [J]. International Journal of Computer Vision, 2004, 45(32): 150-152.

[20] BAKSHI S, RAMAN R, SA P K. Lip pattern recognition based on local feature extraction [C]// IEEE India Conference. Inida: IEEE, 2011: 1-4.

国家自然科学基金项目资助项目（51677035）；深圳市基础研究项目（JCYJ20150513151706580）；深圳市科技计划项目（GRCK2016082611021550）

马新军（1972-），男，新疆石河子人，副教授，博士，主要研究方向：图像处理及模式识别、智能汽车及智能驾驶、生物识别；吴晨晨（1993-），女，河南濮阳人，硕士研究生，主要研究方向：模式识别；仲乾元（1990-），男，江苏徐州人，硕士研究生，主要研究方向：模式识别；李园园（1993-），女，河南许昌人，硕士研究生，主要研究方向：模式识别。

This work is partially supported by the National Natural Science Foundation of China (51677035), the Fundamental Research Project of Shenzhen (JCYJ20150513151706580), the Science and Technology Plan Project of Shenzhen (GRCK2016082611021550).

MA Xinjun, born in 1972, Ph. D, associate professor. His research interests include image processing and pattern recognition, intelligent vehicle and intelligent driving, biological identification

WU Chenchen, born in 1993, M. S. candidate. Her research interests include pattern recognition.

ZHONG Qianyuan, born in 1990, M. S. candidate. His research interests include pattern recognition.

LI Yuanyuan, born in 1993, M. S. candidate. Her research interests include pattern recognition.