基于视觉词袋模型的人耳识别①

董 坤, 王倪传

(上海海事大学 信息工程学院, 数字影像与智能计算实验室 上海 201306)

摘 要: 人耳识别技术是生物特征识别和人工智能领域的一个重要分支. 针对人耳图像特有的纹理特征, 首先采用空间金字塔视觉词袋模型进行人耳特征提取, 该模型将人耳图像中相对低级的局部描述子特征转化为具有高级语义含义的全局特征. 最后采用支持向量机对样本向量进行训练与判别. 实验表明, 本文所采用的模型能取得较高的识别率, 可作为人耳识别方法的一种扩展与探索.

关键词: 人耳识别; 视觉词袋模型; 支持向量机

Human Ear Recognition Based on Visual Bag-of-Words Model

DONG Kun, WANG Ni-Zhuan

(Lab of Digital Image and Intelligent Computation, College of Information Engineering, Shanghai Maritime University, Shanghai 201306, China)

Abstract: Human ear recognition is one of the most important branches in biometrical recognition and artificial intelligence fields. In this paper, considering the unique texture feature of human ear image, the spatial pyramid visual bag-of-words model was adopted. It transforms the relatively low-level local descriptors of human ear images into global features to preserve the high-level semantic meanings. The support vector machine classifier is utilized to perform the training and recognition task. Experimental results demonstrate that the adopted model could achieve a better accuracy, as an extension and exploration in human ear recognition methods.

Key words: human ear recognition; visual bag-of-words model; support vector machine

1 引言

人耳识别作为一种新兴的生物特征识别技术, 越来越引起学者们的研究兴趣. 耳廓本身具有独特而丰富的结构信息, 并且具有在长期的时间内保持不变的稳定性. 研究表明, 人耳满足作为生物特征所需要的四个基本性质: 唯一性, 稳定性, 普遍性和可采集性[1]. 相比其它生物特征识别技术, 人耳识别有自己独特的优势, 与人脸识别相比, 人耳识别不受表情变化和化妆的影响, 在图像灰度化时丢失信息少; 与虹膜、视网膜识别相比, 人耳图像采集方便, 成本低, 可以在不打扰个体正常活动的情况下完成, 容易被人们接受.

当前国内外对人耳识别的研究相对较少, 主流的识别方法主要分为基于几何特征匹配的方法和基于代数统计的方法. Victor 等人比较了人脸识别和人耳识别的性能, 使用了标准的主成分分析(Principal Component Analysis, PCA)算法[2]; 徐正光等人应用独立成分分析(Independent Component Analysis, ICA)方法, 提取出一组独立基图像构成一个映射空间, 将待识别的人耳图像投影到这个映射空间, 从而可以根据投影系数进行人耳的分类和识别[3]. 鉴于上述方法对人耳姿态、大小和光照变化等条件比较敏感, 田莹等人利用当下流行的尺度不变特征(Scale Invariant Feature Transform, SIFT)和几何特征相结合的方法提高了人耳在不同角度下匹配的鲁棒性[4]. Zeng 等人利用 SIFT 描述子、全局上下文以及射影不变量相结合的方法[5], 在提升了人耳识别率的同时, 进一步丰富了人耳识别的研究方法. 本文将在图像检索领域得到广泛应用的词袋模型(Bag of Words, BOW)应用到人耳识别中, 该模型充分利用人耳独特的纹理特征, 首先使用局部不变量的 SIFT 方法, 提取人耳图像中的描述子代表特征, 进而将这些低级特征

① 收稿时间:2014-04-10;收到修改稿时间:2014-05-04

© 中国科学院软件研究所 http://www.c-s-a.org.cn

转化为具有语义含义的更高级的视觉词特征, 最终人耳图像被转化为一个个视觉词直方图表示, 紧接着采用分类器进行训练识别, 得到了可观的识别率, 并作为丰富人耳识别研究方法的一种新的探索.

2 视觉词袋模型框架与生成步骤

BOW 算法起源于基于语义的文本检索算法, 是一种有效的基于语义特征提取和描述的识别算法. 该算法忽略文本的结构信息和语法信息, 仅仅将其看做是若干个词汇的集合, 文本内的每个词的出现都是独立的, 提取其中的语义特征, 构建单词词汇表, 根据每个文本与词汇表的关系, 统计文本中相应单词的出现频率, 形成一个词典维度大小的单词直方图, 经过这样文本到向量运算问题的转化, 最后实现文本检索.

将对文本处理的词袋模型过渡到图像处理领域, 便形成了视觉词袋模型. 其实现过程大致分为四个步骤: 首先提取人耳图像中的特征描述子; 进而通过聚类算法将人耳训练图片得到特征描述子进行相似点聚类, 用得到的类心组成视觉词汇表; 然后采用基于欧式距离的最近邻算法, 将特征描述子映射到视觉词汇表维度, 生成对应的向量直方图, 从而得到一幅图像的词袋模型表示; 最后选择分类器完成训练与识别, 算法框架如图 1 所示.

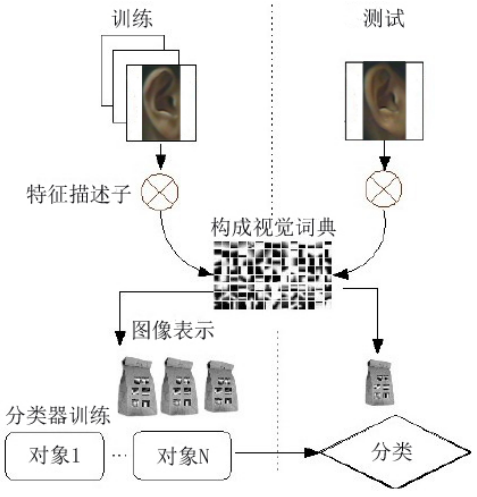

图 1 视觉词袋模型框图

2.1 图像特征提取和描述

构建视觉词汇表之前, 首先要从人耳图像中提取出具有代表性的全局特征或局部特征, 作为对该图像的“描述”. 这些被提取的特征应该具有较强的稳定性, 能够抵抗光照、视角尺度等因素带来的不利影响. BOW 通常采用局部特征来生成视觉词汇表的候选特征, Mikolajczyk 和 Schmid[6]比较了多种局部描述子, 在图像识别和物体匹配的过程中, SIFT 描述子和基于 SIFT 算法的变种描述子具有较好的匹配效果.

SIFT 是 David Lowe 在 1999 年发表, 并在 2004 年完善的一种计算机视觉算法, 用来侦测与描述影像中的局部特征, 它在尺度空间中寻找极值点, 并提取出其位置、尺度、旋转不变量[7], 下面详细介绍其提取过程.

尺度空间理论最早在1962年被提出, 其目的是模拟图像数据的多尺度特征. Koenderink和Lindeberg证明了在众多的合理假设中, 可能的尺度空间核只能是高斯函数. 因此, 一幅图像的尺度空间定义为一个函数 $L(x,y,\sigma)$, 它是由尺度可变的高斯函数 $G(x,y,\sigma)$ 和输入图像 $I(x,y)$ 卷积得到的, 二维图像的尺度空间定义为:

$$L(x,y,\sigma)=G(x,y,\sigma)*I(x,y) \tag{1}$$

其中 σ 是尺度因子, 并且有

$$G(X,Y,\sigma)=\frac{1}{2\pi\sigma^2}e^{-(x^2+y^2)/2\sigma^2} \tag{2}$$

σ 的大小决定图像的平滑程度, σ 越大, 图像被平滑的越多; 反之表示图像被平滑的越少, 大尺度对应图像的概貌特征, 小尺度对应图像的细节特征.

为了在尺度空间中进一步检测到稳定关键点的位置, Lowe在高斯尺度空间的基础上提出利用高斯差分函数(Difference of Gaussian, DoG)与图像卷积的方法生成高斯差分尺度空间. 作为LoG算子的有效线性近似, DoG算子是由两个相邻尺度且相差一个参数因子 k 的高斯函数相减得到的, 这极大的简化了尺度空间的计算. DoG算子定义为:

$$\begin{aligned}D(x,y,\sigma)&=(G(x,y,k\sigma)-G(x,y,\sigma))*I(x,y)\\&=L(x,y,k\sigma)-L(x,y,\sigma)\end{aligned} \tag{3}$$

DoG 金字塔通过高斯金字塔中相邻尺度空间函数相减得到, DoG 图像主要描绘的是目标的轮廓, 图 2 和图 3 分别为一组人耳图像的高斯尺度空间和高斯差分尺度空间图.

在得到人耳图像的高斯差分尺度空间后, 下一步检测 DoG 空间金字塔内部的极值点. DoG 尺度中间层(最底层和最顶层除外)的每个像素点需和它同尺度的

© 中国科学院软件研究所 http://www.c-s-a.org.cn

8 个相邻像素点及其上下相邻两层的 9 个相邻像素点, 共 26 个相邻像素点进行比较, 以确保在尺度空间和二维图像空间都检测到局部极值点.

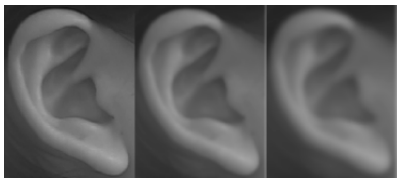

图 2　一组人耳图像高斯尺度空间图

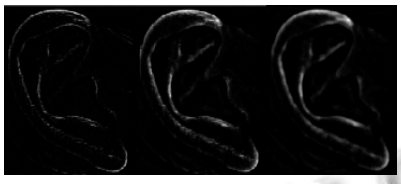

图 3　一组人耳图像高斯差分尺度空间图

然后通过拟合三位二次函数来精确定位关键点的位置和尺度, 同时滤除一些低对比度的关键点和不稳定的边缘相应点, 以增强特征点的稳定性和抗噪声能力.

在生成 SIFT 描述子之前, 需要利用特征点邻域像素的梯度方向分布特性为每个关键点赋予方向参数, 使其具有旋转不变性. 式(2-4)和(2-5)分别为(x,y)处的梯度模值和方向.

$$m(x,y)=\sqrt{(L(x+1,y)-L(x-1,y))^2+(L(x,y+1)-L(x,y-1))^2} \quad (4)$$

$$\theta(x,y)=\arctan\left(\frac{L(x,y+1)-L(x,y-1)}{L(x+1,y)-L(x-1,y)}\right) \quad (5)$$

L 为每个特征点各自的尺度, (x,y) 为指定阶层对应像素点位置, 实际计算中, 在以特征点为中心的邻域窗口内进行采样, 用梯度方向直方图统计邻域像素的梯度方向, 梯度直方图的范围是 0 度到 360 度, 其中每 10 度一个 bin, 总共 36 个 bin, 梯度方向直方图的峰值代表了该特征点邻域梯度的主方向, 即该特征点的方向, 任意其他局部峰值相当于最高封顶的 80%能量就定为该特征点的辅助方向.

最后求关键点描述子时, 首先将坐标轴旋转为关键点的主方向, 以确保旋转不变性. 在关键点所在的尺度空间, 以关键点为中心的 16×16 像素大小的邻域内均匀的划分 4×4 共 16 个子区域, 每个子区域上计算 8 个方向的梯度方向直方图, 绘制每个梯度方向的累加值, 形成一个种子点, 每个种子点有 8 个方向向量信息, 这样对于一个特征点就可以产生 4×4×8 共 128 个数据, 最终形成 128 维的 SIFT 特征向量, 如图 4 所示. 再继续将特征向量的长度归一化, 则可以进一步去除光照变化的影响.

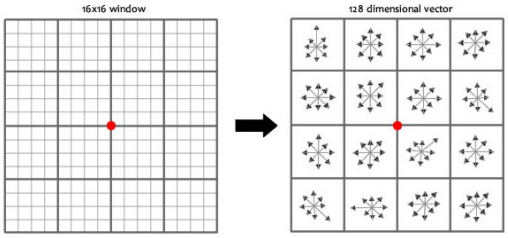

图 4　16 个种子点生成 128 维特征描述子图

人耳的 SIFT 特征图如图 5 所示. SIFT 特征作为图像的局部特征, 对旋转、尺度缩放、光照变化等保持不变性, 独特性好, 信息量丰富, 可以很方便的与其他形式的特征向量进行联合, 具有很好的可扩展性[8].

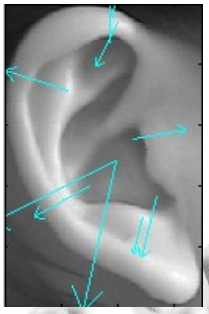

图 5　人耳的 SIFT 特征描述子图

2.2 生成视觉词直方图

在提取图像的 SIFT 描述子之后, 需要使用训练集进行视觉词典的构建. 该过程通常分为两步来完成. 首先将代表图像局部特征的描述子转换为视觉词, 一个视觉单词可以看作人耳图像中相似的特征点的集中代表, 该过程是通过聚类算法实现的. 最终得到的聚类中心就是我们所期望的视觉单词, 聚类中心的个数就是视觉词典的大小. 根据聚类的视觉词来建立每张图像的视觉词直方图, 该过程称为映射.

1) 聚类过程

BOW 算法通常采用 K-means 算法对提取的 SIFT 特征进行聚类生成视觉词典. K-means 算法是一种经典的硬聚类算法, 该算法利用函数求极值的方法得到迭代运算调整规则, 以欧氏距离作为相似性测度, 采

© 中国科学院软件研究所 http://www.c-s-a.org.cn

用误差平方和准则函数作为聚类准则函数.

该算法的主要步骤如下:

1. 给定待聚类的人耳 SIFT 描述子数据集, 随机选取 K 个对象作为初始聚类中心.

2. 求人耳 SIFT 描述子数据集中的每个数据与每个聚类中心的距离, 按照最小化原则将数据点划入最近邻聚类中心所在的类簇.

3. 重新计算每个类簇的中心.

4. 使用均方差等函数对聚类结果进行评估, 如果聚类结果已经趋于稳定则停止聚类, 输出得到的聚类中心, 否则重复到步骤 2 中, 直到评估结果显示聚类趋于稳定[9].

经过聚类生成的视觉词库的过程如图 6 所示.

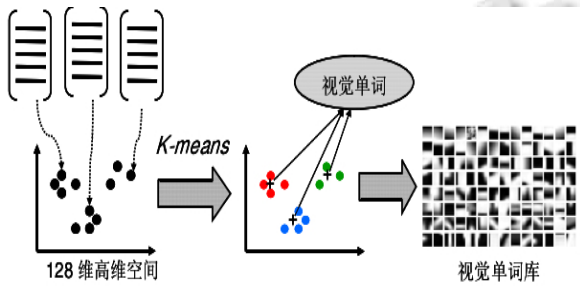

图 6 人耳 SIFT 描述子聚类成视觉词库图

2) 映射过程

映射过程是将每幅图像所有的 SIFT 特征描述子矢量分配到视觉词典维度上, 生成各自的视觉词直方图. 在分配的过程中, 采用最近邻算法, 每幅图像中的每个 SIFT 特征向量与哪一个视觉词距离最近, 就将该视觉词对应的维度高度加 1, 直到将所有的 SIFT 描述子向量分配完为止, 经过这一系列处理后, 每一幅图像都能用一个 K 维的视觉词直方图表示, 将所有人耳图像的视觉词直方图归一化处理后就可以进行下一步的训练与分类了.

3 视觉词袋模型的改进

正如词袋模型在文本检索中忽略了文本的结构信息一样, 视觉词袋模型同样忽略了视觉单词在图像上的空间位置信息. 本文借鉴文献[10]中在场景分类中取得非常好的效果的空间金字塔匹配(Spatial Pyramid Matching, SPM)模型, 对人耳图像进行不同层次的划分, 按照空间金字塔结构分成一系列子区域, 然后在每个子区域利用 BOW 模型计算特征直方图, 最后将所有区域的特征直方图组成一个向量, 过程如图 7、8, 每一层的权重分配为 $1/4, 1/4, 1/2$, 图像划分的越细, 所分配的权重越大, 假设对图像进行 L 级层次划分, 当 L=0 时, 其退化为原始的视觉词袋模型, 若字典长度为 K,最后得到的特征向量维数为

$$K\sum_{0}^{L}4^{L}=K\frac{1}{3}(4^{L+1}-1) \tag{6}$$

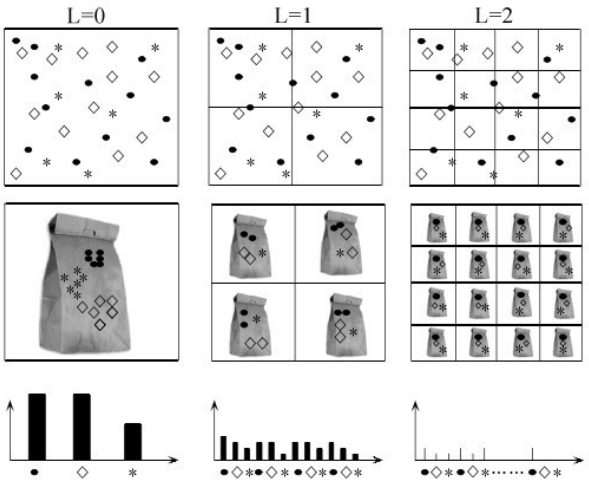

图 7 空间金字塔各层视觉词袋模型图

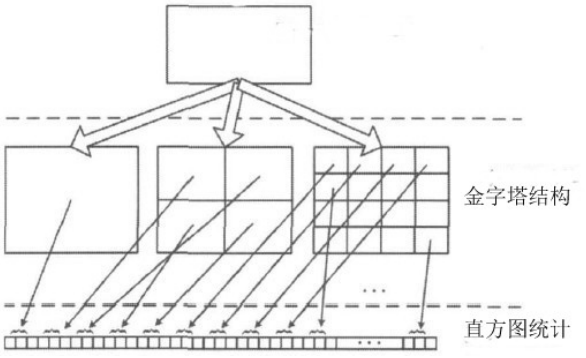

图 8 各层视觉词袋直方图联接图

4 分类器选择与实验

在视觉词袋模型中, 通常选用支持向量机(Support Vector Machine, SVM)完成图像的训练和识别任务, 支持向量机是由 Vapnik 等人通过机器学习方面的研究, 在九十年代中期提出的通用学习方法[11], 它是建立在统计学理论的 VC 维理论和结构风险最小原理基础上, 能够较好地解决小样本、非线性、高位数据和局部极小点等实际问题. 在 SVM 方法中, 可以通过定义不同的核函数, 就能实现多项式逼近、贝叶斯

© 中国科学院软件研究所 http://www.c-s-a.org.cn

分类器、径向基核函数(Radial Basic Function, RBF)、直方图交核函数(Histogram Intersection Kernel)等许多已有的学习方法[12].

本文选用台湾大学林智仁[13]教授等人开发设计的LibSVM作为实验最后的分类工具, 该软件对SVM所涉及的参数调节相对较少, 提供了很多的默认参数, 利用这些参数可以很方便的调用各种核函数, 进行对比实验和分析.

实验步骤与结果

由于当前没有标准的人耳图像库, 本实验使用的是北京科技大学提供的人耳图像库1和图像库3, 选取图像库 1 中 40 个人的人耳图像, 每人 3 幅图像, 大小为 80×150 像素, 有轻微光照和角度变化. 图像库 3 共有 79 人, 每人有 10 幅图像, 包括正面、向左旋转 5°、10°、15°、20°五个姿势, 每个姿势各有两幅图像, 大小为 768×576 像素共 790 幅图像. 对于图像库 1 中的图像, 取每个人的前 2 幅图像构成训练样本集, 剩余的构成测试样本集, 按照本文提出的方法得到每幅图像的BOW直方图表示后, 使用SVM进行训练和识别, 在聚类中心的数目, 即词典大小 K 的选择上, 本文对比了其取值 10,20,30,40,50,60, 70,80, 90,100 等分类结果, 当K值取40时识别的准确率最高, 同时本文也和文献[14]中提出的基于主轴和质心特征相结合的特征提取方法做了比较, 本文中的方法除了在识别率上有显著提高外, 还省去了前期的边缘提取等工作, 直接利用人耳图像的原始纹理特征, 实验结果如表 1.

表 1 本文方法与文献[14]方法的比较结果

采用方法	识别率
BOW+SVM 径向基核函数	77.5%
BOW+SVM 直方图交核函数	65%
BOW+Spacial Pyramid+SVM 径向基核函数	90%
BOW+Spacial Pyramid+SVM 直方图交核函数	87.5%
主轴和质心相结合 +SVM 径向基核函数	62.5%

对于人耳库 3, 按照文献[15]中的方法, 经手工分割后得到大小不同的人耳图像, 使用双三次插值方法归一化为 116×60 像素, 选取正面、向左旋转 5°、10°、15°、20°五个姿势各一幅进行训练, 剩下的进行测试, 训练和测试人耳图像各 390 张, 同样在对比了一系列K值后, 当K=100时, 取得的效果最好, 得到的实验结果如表 2.

表 2 当 K=100 时的实验结果

采用方法	识别率
BOW+SVM 径向基核函数	97.1795%
BOW+SVM 直方图交核函数	96.1538%
BOW+Spacial Pyramid+SVM 径向基核函数	98.4615%
BOW+Spacial Pyramid+SVM 直方图交核函数	97.6923%
主轴和质心相结合+SVM 径向基核函数	75.6410%

通过上述表中可以看出, 不同的人耳库得到的实验结果存在一定的差异, 这和人耳图像的尺寸大小以及拍摄时的角度变化有很大的联系, 通过实验结果可以看出, 在相同条件下, SVM 径向基核函数比直方图交核函数表现的识别效果好; 样本库比较充足的图像库 3 比样本数较少的图像库 1,在识别效果上有非常大的改善; 而增加了图像空间信息的空间金字塔视觉词袋模型相比原始的视觉词袋模型, 识别率有显著的提高.

5 结语

本文提出了一种使用视觉词袋模型进行人耳识别的方法, 将SIFT相对低级的局部特征经过聚类转化后, 映射为具有语义含义的全局特征, 经过实验分析, 该方法能达到较好的识别效果, 但是本方法在人耳样本图像旋转角度比较大的情况下, 例如有部分人耳图像从人耳库中标准的竖直状态变为水平状态, 识别效果将会下降, 如何在图像特征中增加更多的信息和选取更出色的局部描述子特征, 以增强对人耳图像旋转的免疫性将是作者今后努力的方向.

目前的人耳识别仍处于研究探索阶段, 虽然在实际应用中相对较少, 但是作为一种其他生物特征识别技术的补充, 倘若将人脸、人耳等方法组合成多模式识别系统, 可以更好的达到识别个体的目的, 所以人耳识别存在着巨大的研究潜力和发展空间.

致谢　感谢北京科技大学人耳识别实验室提供的人耳图像库.

参考文献

1 Burge M, Burger W. Using ear biometrics for passive identification. Proc. of the IFIP TC11 14th International Conference on Information Security (SEC). 1998, 98. 139−148.

2 Victor B, Bowyer K, Sarkar S. An evaluation of face and ear

© 中国科学院软件研究所 http://www.c-s-a.org.cn

biometrics. Proc. 16th International Conference on Pattern Recognition, 2002. IEEE. 2002, 1. 429−432.

3 田莹,苑玮琦.尺度不变特征与几何特征融合的人耳识别方法.光学学报,2009,28(8):1485−1491.

4 徐正光,武楠,穆志纯.基于独立分量分析的人耳识别方法.计算机工程,2006,32(19):178−180.

5 Zeng H, Mu Z, Yuan L, et al. Ear recognition based on the SIFT descriptor with global context and the projective invariants. 5th International Conference on Image and Graphics (ICIG'09). IEEE. 2009. 973−977.

6 Mikolajczyk K, Schmid C. A performance evaluation of local descriptors. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2005, 27(10): 1615−1630.

7 Lowe DG. Object recognition from local scale-invariant features. Proc. of the 7th IEEE International Conference on. Computer Vision. IEEE. 1999, 2. 1150−1157.

8 Lowe DG. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 2004, 60(2): 91−110.

9 Selim SZ, Ismail MA. K-means-type algorithms: A generalized convergence theorem and characterization of local optimality. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1984, (1): 81−87.

10 Lazebnik S, Schmid C, Ponce J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. IEEE Computer Society Conference on. Computer Vision and Pattern Recognition. IEEE. 2006, 2. 2169−2178.

11 晏庆华.支持向量机算法综述.2008'中国信息技术与应用学术论坛论文集(二),2008.

12 张学工.关于统计学习理论与支持向量机.自动化学报,2000,26(1):32−42.

13 Hsu CW, Lin CJ. A comparison of methods for multiclass support vector machines. IEEE Trans. on Neural Networks, 2002, 13(2): 415−425.

14 Xu Y, Zeng W. Ear recognition based on centroid and spindle. Procedia Engineering, 2012, 29: 2162−2166.

15 王瑜,穆志纯,付冬梅,等.基于小波变换和规范型纹理描述子的人耳识别.电子学报,2010,38(1):239−243.

www.c-s-a.org.cn

© 中国科学院软件研究所 http://www.c-s-a.org.cn