

基于稀疏表征的可见光和近红外光人脸图像融合快速识别算法

赵英男 文学志 成亚萍

(南京信息工程大学 江苏省网络监控中心 南京 210044)

(南京信息工程大学计算机与软件学院 南京 210044)

摘 要 近年来,融合可见光和近红外光的人脸图像特征识别成为一个研究热点。对该领域中的快速人脸识别技术进行研究,并给出了一个具体的实现方案。该方案主要包括以下 3 种技术:原始样本的下采样;基于稀疏表征原理,选取测试样本的 M 近邻来代替原始训练样本;加权决策融合。在 CSIST 人脸库上的实验结果表明,和同类算法相比,所提算法在识别率和计算速度上均有提高。

关键词 稀疏表征,可见光图像,近红外光图像,数据融合,人脸识别

中图分类号 TP391 文献标识码 A

Fast Face Recognition of Sparse Representation Based Fusion of Visible and Near Infrared Images

ZHAO Ying-nan WEN Xue-zhi CHENG Ya-ping

(Jiangsu Engineering Center of Network Monitoring, Nanjing University of Information Science & Technology, Nanjing 210044, China)

(School of Computer & Software, Nanjing University of Information Science & Technology, Nanjing 210044, China)

Abstract Recently, the face recognition of the fusion of visible (VIS) and near infrared (NIR) images has attracted more attention. Here we studied the fast face recognition in this field and gave an applied scheme in detail, which includes three main techniques: down-sampling of the original samples, sparse representation based selection of M -neighbors to substitute the original training samples and the weighted decision fusion strategy. Comparing with the classic algorithms, the experiments on CSIST face databases show that our scheme can achieve higher classification accuracy with a lower computation load.

Keywords Sparse representation, Visual image, Near infrared image, Data fusion, Face recognition

1 引言

近几十年来,人脸识别技术从理论研究到实用系统开发,发展相当迅速,并取得一些较为满意的实用效果^[1]。但是,现有的人脸识别技术依然受到一些外部环境因素的制约,例如面部表情、光照变化、人脸的遮挡等,这时一些常用的人脸识别技术性能退化非常严重。相对于可见光(VIS)人脸图像,最近,红外(IR)人脸图像识别受到人们的日益关注^[4-6]。IR 图像具有抗干扰性强、独立于 VIS 光源、防伪装等优点。而且,人脸皮肤的热辐射系数与周围景物的热辐射系数有明显区别,所以易与周围的景物区分。这些特点可以在很大程度上弥补 VIS 人脸识别技术的不足。但是,热红外(TIR)图像丢失了原有 VIS 人脸图像具有的一些细节边缘、纹理等特征。于是,Stan Z. Li 等提出了近红外(NIR)图像成像系统^[4],即波长范围为 $0.7 \sim 1.1 \mu\text{m}$ 。NIR 图像既具有 TIR 图像的鲁棒性特点又能保持一定的细节特征。但是,NIR 图像依然存在对周围环境温度变化比较敏感、对玻璃的非穿透性等缺点^[7],导致识别性能下降,而 VIS 图像又可以克服上述

问题。因此将二者有机地融合则可以充分利用互补信息来提高识别的性能。

目前,对该领域的研究是一个新兴的热点问题,已有的研究成果可以分成两类。一类是基于融合技术的研究,可以具体分为数据级融合、特征级融合和决策级融合。文献[6]基于粒子群优化算法对两类人脸图像进行图像级融合,然后进行分类。文献[7]进行了简单的特征级融合和决策级融合。最近,文献[13]提出的一种基于复数表示的特征融合策略,取得了较好的识别率和计算速度。图像级融合策略对原始样本有一定的约束条件。如文献[6]中 NIR 和 VIS 人脸图像样本是在同一时间获取的,否则需要两类样本的对准或两类样本之间进行相互的映射转换,这就涉及到该领域的另一类研究方向,即 VIS 和 NIR 之间映射关系的研究。Stan Z. Li 领导的工作组对此进行了深入的研究。Chen 等应用局部线性嵌入方法将 NIR/VIS 人脸对转换成 NIR 和合成的 VIS 图像进行匹配^[8]。Yi 等应用标准相似分析来学习 NIR 和 VIS 之间的相似度^[9]。

文中的研究内容属于上述的第一类,并在决策级别进行

到稿日期:2011-09-01 返修日期:2011-11-28 本文受江苏高校优势学科建设工程项目,国家自然科学基金项目(61103141)资助。

赵英男(1973—),女,博士,副教授,主要研究方向为模式识别、图像处理,E-mail:yingnanzhao@163.com;文学志(1970—),男,博士,副教授,主要研究方向为图像处理、智能车辆;成亚萍(1968—),女,硕士,副教授,主要研究方向为图像处理、系统集成。

融合。该方案主要包括以下3种技术:原始样本的下采样;基于稀疏表征原理,选取测试样本的 M 近邻来代替原始训练样本;加权融合策略。在相关人脸库上的实验结果表明,和同类算法相比,所提算法在识别率和计算速度上都有所提高。

本文第2节为算法描述,从原始样本的下采样、测试样本的 M 近邻选择和分类融合3个方面进行具体叙述;第3节给出实验验证过程及数据分析;最后给出相应的结论。

2 算法描述

整个算法可以分为3部分,即样本预处理、测试样本的 M 近邻选择和分类决策融合。首先针对VIS和NIR两类图像样本进行下采样;然后选择每个测试样本的 M 近邻;最后进行分类决策融合,并产生最终的分类结果。算法的流程框架如图1所示。

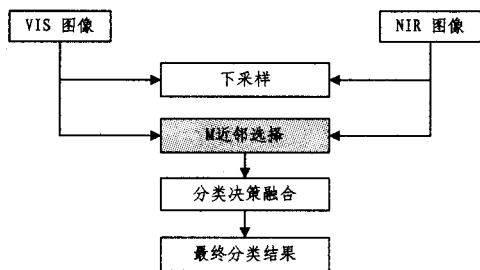


图1 算法流程框架

2.1 原始样本的下采样

这里的下采样技术采用文献[10]提出的方法,该方法能够在降低分辨率的情形下,依然保持较高的识别率。现简述如下。

一维数列 $\{x_n\}, n=1, 2, \dots, N, N+1, \dots, 2N$ 经过下采样之后,可以得到序列 $\{y_m\}, m=1, 2, \dots, N$ 和 $\{z_m\}, m=1, 2, \dots, N$ 。其满足下列公式:

$$y_m = \frac{x_{2m} + x_{2m-1}}{2} \quad (1)$$

$$z_m = \frac{x_{2m} - x_{2m-1}}{2} \quad (2)$$

而 $\{y_m\}$ 和 $\{z_m\}$ 可以重构 $\{x_n\}$ 序列,即

$$\begin{cases} x_{2m} = y_m + z_m \\ x_{2m-1} = y_m - z_m \end{cases} \quad (3)$$

式中, $m=1, 2, \dots, N$ 。而 $\{x_m\}$ 和 $\{x_{2m-1}\}$ 恰好组成 $\{x_n\}$ 序列。

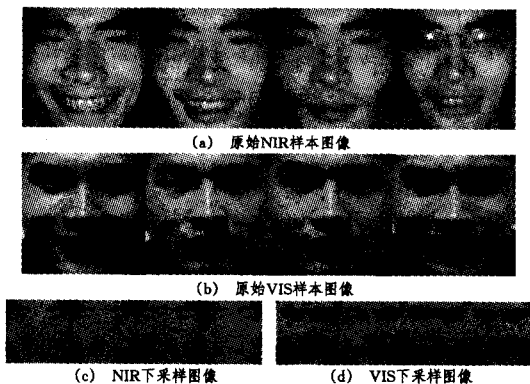


图2 NIR和VIS样本图像及其下采样图像

针对二维的图像矩阵,可以应用上述结论分别对行数据进行下采样,再对列数据进行下采样。一个 $m \times n$ 大小的图像,经过行、列两次下采样之后,得到的图像大小为 $m/2 \times n/2$,其中一个为低频图像,另一个则是差图像,类似于频域中的高频图像和低频图像,低频图像保存有大部分的图像信息,而高频图像则由一些细节和噪声组成。因此,在经过下采样之后,选择低频图像作为下一步处理的样本图像。图2给出了VIS和NIR原始样本图像以及二者经过下采样之后的低频图像。可以看出,经过下采样之后,人脸图像的大小减半,而且一些细节噪声减弱,同时一些主要的人脸特征信息得以保留。

2.2 测试样本的 M 近邻选择

最近Xu等^[11]提出一种两步人脸识别算法,该算法取得了较高的识别率,其计算速度较快。该方法基于稀疏表征原理^[12],将测试样本表示为 n 个训练样本的线性组合。首先,计算某个测试样本与所有 n 个训练样本的欧式距离,然后对其进行排序,选择 M 个最近距离的训练样本代替原训练样本对测试样本进行线性表征。在该算法中将其余对测试样本表征能力较弱的 $n-M$ 个样本的稀疏表征系数认定为零。然后再进一步进行识别,从而有效降低计算量。下面进行具体叙述。

根据稀疏表征原理,测试样本可以表示成训练样本的线性组合。令 $x_1, x_2, \dots, x_n$ 为 n 个训练样本, y 为某一测试样本,则下式成立:

$$y = a_1 x_1 + a_2 x_2 + \dots + a_n x_n \quad (4)$$

式中, $a_i (i=1, 2, \dots, n)$ 是稀疏表征系数。我们也可以将式(1)写成如下形式:

$$y = XA \quad (5)$$

式中, $X = [x_1, x_2, \dots, x_n]$, $A = [a_1, a_2, \dots, a_n]^T$ 。如果 X 是非奇异的,则

$$A = X^{-1}y \quad (6)$$

否则,可以加一个小的扰动,如下:

$$A = (X^T X + \mu I)^{-1} X^T y \quad (7)$$

式中, μ 为一小的正常数, I 是单位矩阵。

式(4)显示每一个训练样本对于测试样本都有一个稀疏表征系数,该系数描述该训练样本在表达测试样本时的贡献程度。而贡献程度的大小,可以通过测试样本和该训练样本之间的距离来衡量,即

$$e_i = \|y - a_i x_i\|^2 \quad (8)$$

其中, e_i 越小,第 i 个训练样本对于测试样本的表达贡献程度越大。因此,可以将此作为度量依据,针对每个测试样本选择 M 个近邻来代替原所有训练样本。即对于每个测试样本只用 M 个近邻来表征,这样可以有效地降低计算量。图3给出VIS和NIR图像中,某个测试样本的5个近邻训练样本的选择。

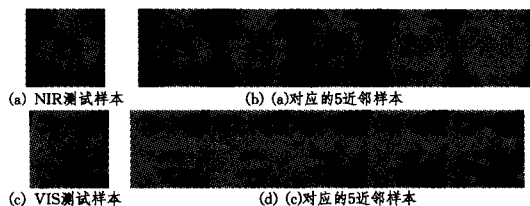


图3 NIR及VIS测试样本对应的5近邻选择样本

在该阶段将测试样本用确定的 M 近邻线性表示如下:

$$y = b_1 \tilde{x}_1 + b_2 \tilde{x}_2 + \dots + b_M \tilde{x}_M \quad (9)$$

式中, $\tilde{x}_i (i=1, 2, \dots, M)$ 为已经确定的 M 近邻。那么式(5)可以改写为:

$$y = \tilde{X}B$$

式中, $\tilde{X} = [\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_M]$, $B = [b_1, b_2, \dots, b_M]^T$ 。类似地, 如果 X 是非奇异的, 则

$$B = (\tilde{X})^{-1}y \quad (10)$$

否则, 可以加一个小的扰动, 如下:

$$B = (\tilde{X}^T \tilde{X} + \gamma I)^{-1} \tilde{X}^T y \quad (11)$$

式中, γ 为一小的正常数, I 是单位矩阵。

当每个测试样本选择了 M 近邻来表征后, 我们可以进行下一步的分类。

2.3 分类融合

在分类融合阶段, 我们提出了一种加权求和的决策融合方法。在进行匹配时, 分别在 NIR 和 VIS 人脸数据库根据置信度公式得到相应的匹配值, 然后选择最大置信度值所属的类别作为最终的分类结果。这里有两种实现方案, 一种是直接在得到的匹配值上选择最大值, 称为 Fusion-Fm 方法。另一种是对两个库上得到的匹配值进行平均加权求和, 然后在此基础上进行最大值的选择, 称为 Fusion-Fa 方法。

置信度公式如式(12)所示, 它的建立基于式(8)中的测试样本与训练样本的欧式距离。

$$h_i = \frac{A - e_i}{A} \quad (12)$$

式中, A 是一个正常数, 满足 $A \geq \max(e_i) (i=1, 2, \dots, M)$ 。由于在进行 M 近邻选择时, 已经对每个训练样本计算过 e_i , 因此很容易得到每个 M 近邻训练样本的置信度。

令 M 近邻训练样本所属的类别为 $\{C_1, C_2, \dots, C_r\}$ 。如果来自第 $C_j (j=1, 2, \dots, r)$ 类的近邻为 $\tilde{x}_s \dots \tilde{x}_t$, 那么第 C_j 类近邻训练样本对于测试样本表达的贡献总和如下:

$$g_{C_j} = b_s \tilde{x}_s + \dots + b_t \tilde{x}_t \quad (13)$$

则第 $C_j (j=1, 2, \dots, r)$ 类的置信度可由下式计算:

$$h_{C_j} = \frac{1}{\sum_{i=s}^t b_i} \sum_{i=s}^t b_i h_i \quad (14)$$

这样, 我们可以计算所有 M 近邻样本所属的所有类别的置信度。对于 Fusion-Fm 方法, 选择 $\max(h_{C_j}) (j=1, 2, \dots, M)$ 作为最终的分类结果。对于 Fusion-Fa 方法, 应用式(14), 分别在 VIS 和 NIR 人脸库中计算同一类别的总的置信度, 依次记为 M_V 和 M_N , 然后依据下列公式对同一类别进行最终置信度 M_F 的计算:

$$M_F = w_V M_V + w_N M_N \quad (15)$$

式中, $w_V = w_N = 0.5$ 。

下面是整体方案的主要步骤:

Step 1 根据式(1)对所有训练样本和测试样本进行先后列的下采样处理。

Step 2 根据式(7)计算表征系数矩阵, 然后根据式(8)计算每个训练样本和测试样本之间的欧式距离。最后依据排

序结果选择距离最小的 M 近邻训练样本。

Step 3 应用 M 近邻训练样本构造式(9), 并求解, 见式(11)。

Step 4 应用式(14)计算 h_{C_j} 置信度。然后采用 Fusion-Fm 或 Fusion-Fa 方案, 对该测试样本的最终类别进行判定。

3 实验与数据分析

实验中采用 CSIST 人脸数据库, 其样本类别数是 100, 每类样本包含 NIR 和 VIS 的样本个数均为 4, 大小是 128×128 。在样本图像中有光照、眼镜、表情等变化, 如图 2(a)、(b)所示。在实验中, 训练样本与测试样本的个数均为 2。那么在 NIR 和 VIS 人脸库中, 训练样本与测试样本总数均为 200。训练样本的选取采用 C_k 种抽取方法, 其余为测试样本, 然后对识别结果进行平均得到最终的识别错误率。需要提及的是, 经过下采样预处理之后, 样本大小为 64×64 。式(7)及式(11)中参数 μ 和 γ 分别取值为 0.01, 编程环境为 MATLAB 7.0。

针对测试样本的近邻选择个数对人脸识别错误率的影响进行了实验, 数据如图 4 所示。“NIR”与“VIS”代表文中算法没有应用决策融合策略分别在 NIR 和 VIS 人脸图像库进行识别的结果。从图中可以看出, 随着近邻个数选择的增加, 识别率降低。这一结果趋势和文献[11]中的结果是相似的。图 4 同时显示了 Fusion-Fa 和 Fusion-Fm 两种不同决策融合策略对识别率的影响。可以看出, 这两种方案均优于不融合时的识别率, 证明了融合策略的有效性。同时, 二者随测试样本近邻个数的增多而性能降低, 但是总体趋势要比不融合时的 VIS 和 NIR 的性能平稳。在该实验条件下, Fusion-Fa 的识别性能要高于 Fusion-Fm。文献[13]中的识别错误率为 4.9, 高于文中算法当近邻选择为 100 以内的识别错误率。

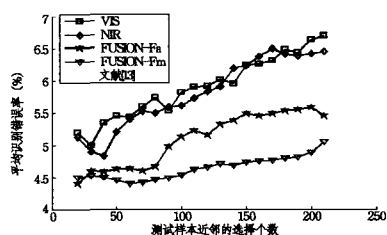


图 4 不同近邻个数对平均识别错误率的影响

在以上实验中可以看出, 当测试样本的近邻个数为 40 时, 各种算法的识别错误率较低, 因此在进行相关算法的时间比较时, 我们选择近邻个数为 40, 融合方案选择 Fusion-Fm 方法。在这里, 我们和文献[13]进行对比的数据如表 1 所列。可以看出, 文中所提算法在识别率和计算时间上均优于文献[13]的算法。

表 1 同类算法在 CSIST 数据库上的计算时间对比

算法	文献[13]	文中算法
时间(秒)	25	18

结束语 基于稀疏表征原理, 文中从实用角度出发, 提出一种 VIS 和 NIR 人脸图像融合的快速识别算法。该算法包括以下 3 种技术: 原始样本的下采样; 基于稀疏表征原理, 选

取测试样本的 M 近邻来代替原始训练样本;加权决策融合。在 CSIST 人脸库上的实验结果表明,和同类算法相比,所提算法在识别率和计算速度上均有提高。主要原因如下:

1)经过 M 近邻的选择代替原训练样本来线性表示测试样本,表达方式进行了优化。这种优化一方面可以提供更高质量的表述,表现为识别率的提高;另一方面也减少了训练样本的个数,提高了计算速度。

2)在分类融合过程中,巧妙利用 M 近邻选择过程的欧式距离计算结果,使该过程的速度加快。

对于 VIS 和 NIR 人脸图像的融合,通常我们希望在光照变化比较剧烈的时候,NIR 图像能够占优,在温度变化比较剧烈或人脸佩戴眼镜的时候,VIS 图像可以占优,但在决策融合阶段目前很难量化区分这几种情况。因此如何提出一种更有效的融合策略将是下一步的研究工作。

参 考 文 献

- [1] 赵铭. 机场安检人相识别系统的设计与实现[D]. 成都:电子科技大学,2011
- [2] Adini Y, Moses Y, Ullman S. Face Recognition: The Problem of Compensating for Changes in Illumination Direction[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1997, 19(7):721-732
- [3] Xie Xu-dong, Lam K-M. Face Recognition under Varying Illumination based on 2D Face Shape Model[J]. Pattern Recognition, 2005, 38(2):221-230
- [4] Li S Z, Chu Ru-feng, Liao Sheng-cai, et al. Illumination Invariant Face Recognition Using Near-Infrared Images[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2007, 29(4):627-639
- [5] Singh R, Vatsa M, Noore A. Integrated Multilevel Image Fusion and Match Score Fusion of Visible and Infrared Face Images for Robust Face Recognition[J]. Pattern Recognition, 2008, 41(3):

880-893

- [6] Raghavendra R, Dorizzi B, Rao A, et al. Particle Swarm Optimization based Fusion of Near Infrared and Visible Images for Improved Face Verification[J]. Pattern Recognition, 2011, 44(2):401-411
- [7] Heo Jin-gu, Kong S G, Abidi B R, et al. Fusion of Visual and Thermal Signatures with Eyeglass Removal for Robust Face Recognition[C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. Washington, D. C., USA, 2007:6-6
- [8] Chen Jie, Yi Dong, Yang Ji-mei, et al. Learning Mappings for Face Synthesis from Near Infrared to Visual Light Images[C]//IEEE Conference on Computer Vision and Pattern Recognition. 2009:156-163
- [9] Yi Dong, Liu Rong, Chu Ru-feng, et al. Face Matching between Near Infrared and Visible Light Images[C]//Advances in Biometrics; International Conference on Advances in biometrics. Seoul, Korea, 2007:523-530
- [10] Xu Yong, Jin Zhong. Down-sampling Face Images and Low-resolution Face Recognition[C]//The third international conference on innovative computing information and control. Dalian, China, 2008:392-395
- [11] Xu Yong, Zhang D, Yang Jian, et al. A Two-phase sample sparse representation method for use with face recognition[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2011, 21(9):1255-1262
- [12] Wright J, Yang A Y, Ganesh A, et al. Robust Face Recognition via Sparse Representation[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2009, 31(2):210-227
- [13] Xu Yong, Zhong Ai-ni, Yang Jian, et al. Bimodal Biometrics based on a Representation Recognition Approach[J]. Optical Engineering, 2011, 50(3):037202

(上接第 250 页)

- [7] Hinckley K, Dixon M, Sarin R, et al. Codex: a dual screen tablet computer[C]//Conference on Human Factors in Computing Systems (CHI). ACM, Boston, MA, USA, 2009:1933-1942
- [8] 廖立君,何频捷,王志兵,等. 多方互动式移动学习系统的设计与实现[J]. 湖南工业大学学报,2010,24(6):38-41
- [9] 金贻,林筑英. 移动支持协同学习[C]//中国人工智能学会计算机辅助教育专业委员会(全国 CBE 学会)第十三届学术年会. 2008:303-306
- [10] 滕文星,李士宁,尹小燕. 移动协同感知研究综述[J]. 计算机科学,2008,35(6):25-27
- [11] 冷德宏. 基于多代理的计算机支持移动协同学习关键技术研究[D]. 上海:复旦大学,2004

- [12] 李浩然,张倩,张慧. 儿童从同伴相互作用中获得知识研究的新进展——Granott 的相互作用模型述评[J]. 山东师大学报:人文社会科学版,2001(3):16-18,24
- [13] Granott N. Patterns of interaction in the co-construction of knowledge: separate minds, joint effort, and weird creature[C]//Wozniak R H, Fischer K W, eds. Development in context: Acting and thinking in specific environments. Hillsdale, New Jersey:LEA,1993:173-207
- [14] Granott N, Gardner H. When minds meet: interactions, coincidence, and development in domains of ability[M]//Sternberg R J, Wagner R K, eds. Mind in context: Interactionist Perspectives on human intelligence. Cambridge university press. 1994:171-201