

一种优化的人脸特征数据保护算法

杨玉平^{1,2}, 陈勇²

(1. 重庆电子工程职业学院 教务处, 重庆 401331;
2. 重庆师范大学 计算机与信息科学学院, 重庆 401331)

摘要: 针对基于 SIFT 的保护算法运行效率低、算法结构庞大的问题, 提出一种基于 SIFT 人脸特征数据保护算法的优化算法. 该算法通过对原算法特征数据库中的数据存放方式、数据结构的优化, 提高了检索效率. 该算法在 AR 和 YALE 人脸库中进行实验, 结果表明, 本优化算法在保证数据安全性的前提下, 不降低原有算法识别率的同时, 使整个识别系统的运行效率更高.

关键词: 人脸识别; 特征数据保护算法; 数据安全性; SIFT

中图分类号: TN919.8 文献标志码: A

An optimized protection algorithm of face feature data

YANG Yu-ping^{1,2} CHEN Yong²

(1. Academic Affairs Office, Chongqing College of Electr. Eng., Chongqing 401331, China;
2. College of Comp. and Infor. Sci., Chongqing Normal Univ., Chongqing 401331, China)

Abstract: Aiming at the problem that the protection algorithm based on SIFT had low operating efficiency and algorithm structure was huge, an optimization algorithm based on SIFT face feature protection algorithm was presented. This algorithm improved the efficiency of retrieval by optimizing the data storage and the data structure in the feature database of the original algorithm. The experiment was carried out on the optimized algorithm in AR and YALE face databases, the experimental results showed that the optimization algorithm ensured data security of the premise, kept recognition rate of the original algorithm, at the same time, made the recognition system more efficient.

Key words: face recognition; protection algorithm of feature data; data security; SIFT

0 引言

随着社会的发展, 电子计算机技术的进步, 人们对于安全的期望也越来越高, 传统意义上的安全已不能满足人们的需求. 人脸识别系统是目前应用比较普遍的新兴的安全系统, 相比其他的新兴的识别方式, 它具有自然性和不易被识别的个体察觉的

优势. 人脸识别主要用于身份识别, 常用于考勤机以及远程视频监控, 能够远距离利用人脸检测技术定位人脸并实时与人脸库中数据进行对比, 从而实现人脸的识别^[1].

人脸识别的算法很多, 提取人脸的特征数据是算法的核心. 人脸特征数据的提取算法主要分为基于全部人脸的特征提取、基于局部人脸的特征提取

收稿日期: 2012 - 04 - 16

作者简介: 杨玉平(1984—), 男, 重庆市人, 重庆电子工程职业学院助教, 重庆师范大学硕士研究生, 主要研究方向为数字图像处理与模式识别.

和基于混合人脸的特征提取,比如主成分分析法(PCA)、线性判别分析法(LDA)、独立成分分析法(ICA)以及鲁棒性较强的SIFT特征提取法等^[2]。

人脸识别系统自身的安全性也非常重要,识别系统中已注册的人脸特征数据库是人脸识别系统安全性的重点,大多数的算法都是围绕它进行研究的。例如,随机卷积核加密算法,生物特征密码保护算法^[3],Fuzzy Commitment, Fuzzy Vau It, Biohashing 算法^[4],模糊逻辑算法以及模糊 Biohashing 算法等^[5]。这些人脸特征数据保护算法,在保护特征数据的同时,也较侧重于提高人脸识别系统的识别率。然而现在大多数的识别算法在提高识别率、提高人脸特征数据库安全性的同时,也使得算法结构趋于庞大,从而导致算法的运行效率较低,耗时过长,降低了识别系统的实时性。针对这样的情况,本文拟在SIFT的基础上提出一种优化的人脸特征数据保护算法,以期在识别率和人脸特征数据安全性不降低的情况下,提高识别系统的运行效率和实时性^[6-7]。

1 SIFT 人脸特征保护算法

现有的基于SIFT人脸局部特征保护算法的基本结构如图1所示。从图1可以看出,采用SIFT作为人脸特征提取的算法,就是把提取出来的特征向量根据用户的密钥 k 生成随机投影矩阵,然后把投影后的特征向量或放于人脸特征模板数据库中,或与人脸特征模板数据库中的数据进行比较,最终得出识别结果。

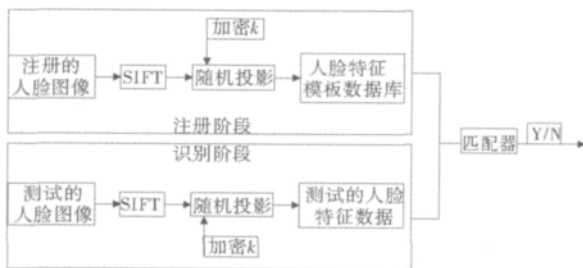


图1 SIFT 保护算法的基本结构

SIFT 特征提取算法是一种局部特征提取算法。相比 PCA 特征提取算法^[8]、LDA 特征提取算法^[9], SIFT 特征提取算法可以避免光照、姿势以及表情变化的影响,提高整个系统的识别率。该算法通过用户的密钥 k 产生的随机投影矩阵来对提取的人脸特征向量进行随机投影,从而对人脸特征模板数据库中的特征数据起到了保护的作用。这个模板数据库

中的数据一般是一个类的人脸图像的特征向量连续存放在一起。在识别阶段,当一个测试的人脸图像通过 SIFT 提取特征以后,经过注册时同样的随机投影得到一个特征向量,将此特征向量与人脸模板数据库中的每个特征向量进行比较,得出它们各自与这个向量的距离,通过距离来判断是具体的哪个人。一般来说,与特征模板数据库中的某个特征向量距离越小,说明二者指向的就是同一个人;与特征模板数据库中的某个特征向量距离越大,说明二者指向的不是同一个人。当然还有一个阈值^[10-11],当测试人的特征向量与模板数据库中各向量的距离大于这个阈值时,说明测试人并没有在识别系统中注册。

2 优化的 SIFT 特征保护算法

本文对最新保护算法的特征数据库进行研究,对其数据库的数据结构及检索方式进行优化^[12]。

优化算法的步骤如下:

1) 运用 SIFT 提取训练图像的特征向量 x_m^i , 其中 $x_m^i \in R^n$, i 为训练图像的第 i 类, m 为第 i 类训练图像的个数。

2) 通过用户的密钥 k 生成随机投影矩阵 $R_{n \times n'}$ 来对 x_m^i 进行降维,即最终存于特征数据库中的特征向量为 $X_{m \times n'} = x_m^i R_{n \times n'}$ 。其中 $R_{n \times n'}$ 中的 n' 是根据所取能量值来确定的。

3) 通过 Matlab 中的 mean() 函数,求得 $\bar{X}_{m \times n'}$, 即第 i 类的平均特征向量值,且存储于第 i 类的开头项。

4) 将第 i 类的每一个特征向量值分别与它们对应的 $\bar{X}_{m \times n'}$ 作欧氏距离计算,并把这个距离值放于各自项的最后,且以这个距离值作升序排列。

最后特征数据库中存放的数据为: 训练图像的多张图片的特征向量类内连续存放,类间分开存放;类内存放的顺序是第 1 项为平均特征向量,末尾的欧氏距离为 0,从第 2 项开始到该类后一项是依次以欧氏距离为升序排序方式存放,且各自项的末尾一维数字是相对应的欧氏距离。

3 实验结果与分析

本优化算法的实验主要用到了 AR 和 YALE 2 个人脸库。AR 人脸库中有 120 人,每人有 14 张人脸

图像,选 7 人作为注册人脸图像,另选 7 人作为待识别的人脸图像,它们的大小都是 40×50 ,如图 2 所示.



图 2 AR 人脸库

YALE 人脸库中共有 165 张人脸图像,每个人有 11 张人脸图像,选了 6 人作为注册人脸图像,5 人作为待识别的人脸图像,它们的大小都是 100×100 ,如图 3 所示.



图 3 YALE 人脸库

3.1 实验结果分析

表 1 中的数据是基于 SIFT 保护算法与基于 SIFT 的优化后的保护算法在 AR 人脸库与 YALE 人脸库上进行对比实验得到的,为了降低误差,其中的数据取的是平均值.

从表 1 可以看出,人脸图像特征提取时降的维数越多(随机投影后的维数越少),识别系统所耗时间就越少,识别率变化便很微小.另外,在不同维数情况下,优化后的 SIFT 保护算法明显比原 SIFT 保护算法所耗的时间要少,但平均识别率没有降低,

大多数还有略微的提高.所以,采用基于 SIFT 的优化后保护算法能够使系统的运行效率更高,耗时更少,并且优化后的系统的安全性并没有降低,识别率也没有降低,因此可以说改进后的算法比原算法具有更高的实时性.

3.2 计算复杂度对比

对比原 SIFT 保护算法的特征数据库中的数据结构与优化后的 SIFT 保护算法的特征数据库中的数据结构,可以看出:原 SIFT 保护算法在识别阶段需要检索特征数据库中的所有特征向量数据,如果这个数据库中的数据规模为 n ,那么原 SIFT 算法的时间复杂度是线性阶的,即 $o(n)$;而优化后的 SIFT 保护算法在识别阶段检索特征数据库时首先是与各类欧氏距离为 0 的平均特征向量进行比较,检索出与之最接近的类,然后再在类内与以欧氏距离为基准进行升序的类内特征向量的前 m 个(可自己设定)进行比较,如本实验中采用的是前 3 个类内特征向量,如果数据规模为 n ,数据库中有 m 个类,那么优化后的 SIFT 保护算法的时间复杂度就是对数阶的,即 $o(\log m^n)$.在空间复杂度方面,优化算法所用数据结构只是在原有算法所用数据结构的基础上每个类只增加一个项,且每个项后面只增加一个字节,空间复杂度变化不大.因此本优化算法的检索效率明显优于原算法,识别系统的运行效率有明显提高.

4 结论

本文提出了一种优化的人脸特征安全保护算法,即在原有 SIFT 人脸特征保护算法的基础上,优化了它的人脸特征数据库中的数据结构,提高了识别系统的检索效率.本算法在 AR 和 YALE 人脸库

表 1 基于 SIFT 保护算法与本文算法比较

人脸库	随机投影后的维数		120	110	100	90	80	70	60	50
AR	SIFT	平均识别率	0.993 9	0.992 6	0.992 4	0.993 5	0.991 8	0.993 3	0.992 0	0.991 9
		平均耗时/s	0.940 0	0.880 0	0.840 0	0.790 0	0.720 0	0.680 0	0.630 0	0.580 0
	优化后的 SIFT	平均识别率	0.994 0	0.992 6	0.992 8	0.993 2	0.991 8	0.993 2	0.992 5	0.992 0
		平均耗时/s	0.900 0	0.820 0	0.800 0	0.740 0	0.680 0	0.630 0	0.590 0	0.540 0
YALE	SIFT	平均识别率	0.987 1	0.989 6	0.992 5	0.987 5	0.988 9	0.988 3	0.991 7	0.988 5
		平均耗时/s	1.580 0	1.520 0	1.450 0	0.410 0	1.380 0	1.350 0	1.320 0	1.290 0
	优化后的 SIFT	平均识别率	0.987 3	0.989 8	0.992 6	0.987 5	0.988 9	0.988 5	0.991 6	0.988 3
		平均耗时/s	1.530 0	1.480 0	1.420 0	0.380 0	1.350 0	1.310 0	1.280 0	1.260 0

中进行了实验,并对实验结果数据进行了对比分析。结果表明:在保证数据安全性的前提下,本算法不降低原有算法识别率的同时,整个识别系统的运行效率有明显提高。

参考文献:

- [1] 周玲丽,赖剑煌.人脸特征的 SIFT 保护算法[J].中国图象图形学报 2011,16(5):800.
- [2] Zhou Shang-Ming, John Q Gan, Francisco Sepulveda. Classifying mental tasks based on features of higher-order statistics from EEG signals in brain-computer interface[J]. Infor Sci 2008,178(6):1629.
- [3] Jing Xiaoyuan, Li Sheng, Lan Chao, et al. Color image canonical correlation analysis for face feature extraction and recognition[J]. Signal Processing 2011,91(8):2132.
- [4] McCool Chris, Sanchez-Riera Jordi, Marcel Sébastien. Feature distribution modelling techniques for 3D face verification[J]. Pattern Recognition Letters 2010,31(11):1324.
- [5] Ting Jo-Anne, D' Souza Aaron, Yamamoto Kenji, et al. Predicting EMG data from MI neurons with variational bayesian least squares[J]. Neural Networks,2008,21(8):1112.
- [6] Haddadnia Javad, Ahmadi Majid. N-feature neural network human face recognition[J]. Image and Vision Comp 2004,22(12):1071.
- [7] Gundimada Satyanadh, Asari Vijayan K, Gudur Neeharika. Face recognition in multi-sensor images based on a novel modular feature selection technique[J]. Infor Fusion, 2010,11(2):124.
- [8] Azmi Ida Madieha. Bioinformatics and genetic privacy: The impact of the personal data protection act 2010[J]. Comp Law & Security Review 2011,27(4):394.
- [9] Ter Haar Frank B, Veltkamp Remco C. 3D face model fitting for recognition[J]. Lecture Notes in Comp Sci 2008,5305:652.
- [10] Zhou Xiaoli, Bhanu Bir. Feature fusion of side face and gait for video-based human identification[J]. Pattern Recognition 2008,41(3):778.
- [11] Safayani Mehran, Shalmani Mohammad Taghi, Manzuri. Three-dimensional modular discriminant analysis (3DMDA): A new feature extraction approach for face recognition[J]. Comp & Electri Eng 2011,37(5):811.
- [12] Thiyagarajan R, Arulselvi S, Sainarayanan G. Gabor feature based classification using statistical models for face recognition[J]. Procedia Comp Sci 2010(2):83.

(上接第 64 页)

参考文献:

- [1] 姜谊,严京旗.车牌定位及倾斜校正方法[J].微型电脑应用 2010,26(4):45.
- [2] 李耀华,杨连贺.车牌图像倾斜的校正实现[J].仪器仪表用户 2008,15(6):81.
- [3] 刘永林,王新宇,于勇. Hough 变换在图像匹配中的应用[J].遥测遥控 2007,28(S1):139.
- [4] 武莎莎,姚敏,赵敏.基于 Hough 变换的二维条码图像倾斜校正研究[J].淮阴工学院学报 2009,18(5):54.
- [5] 梁添才,皮佑国,彭晶,等.基于 Hough 变换的列车客运票图像倾斜校正[J].华南理工大学学报:自然科学版 2007,35(5):35.
- [6] 陈圣俭,孙晋,程文刚.一种基于模板匹配的图像倾斜校正算法[J].中国电力教育 2007(S1):375.
- [7] 张顺利,李卫斌,吉军.基于投影的文档图像倾斜校正方法[J].计算机工程与应用 2010,46(3):166.
- [8] 吴成东,樊玉泉,张云洲,等.基于差分投影与优割字符的车牌字符分割[J].东北大学学报:自然科学版,2008,29(7):920.
- [9] 吴一全,丁坚.基于 K-L 展开式车牌倾斜校正方法[J].仪器仪表学报 2008,29(8):1690.