

基于特征运动的表情人脸识别

余冰 金连甫 陈平

(浙江大学计算机科学与工程系, 杭州 310027)

摘要 人脸像的面部表情识别一直是人脸识别的一个难点. 为了提高表情人脸识别的鲁棒性, 提出了一种基于特征运动的人脸识别方法. 该方法首先利用块匹配的方法来确定表情人脸和无表情人脸之间的运动向量, 然后利用主成分分析方法(PCA)从这些运动向量中, 产生低维子空间, 称之为特征运动空间. 测试时, 先将测试人脸与无表情人脸之间的运动向量投影到特征运动空间, 再根据这个运动向量在特征运动空间里的残差进行人脸识别. 同时还介绍了基于特征运动的个人模型方法和公共模型方法, 实验结果证明, 该新算法在表情人脸的识别上, 优于特征脸方法, 有非常高的识别率.

关键词 基于特征运动的方法 特征脸 特征运动 运动向量残差 特征运动空间

中图分类号: TP391.41 **文献标识码**: A **文章编号**: 1006-8961(2002)11-1139-05

Expression-invariant Face Recognition Based on Eigenmotion

YU Bing, JIN Lian-fu, CHEN Ping

(Department of computer science and engineering, Zhejiang University, Hangzhou 310027)

Abstract The difficulty of a face recognition problem is to handle different types of variations, such as facial expression, illumination and pose. In order to improve the robustness of face recognition with respect to facial expression, this paper proposes a new approach, the eigenmotion-based method, which is tolerant to large variations of facial expressions. In this new approach, first motion vectors are computed between a testing face image and a neutral training image using the block-matching method, then projected to a low dimensional subspace that is pre-trained by applying principal component analysis(PCA) to motion vectors resulting from training images with expression variations. This subspace is called an eigenmotion space. Finally the identification of the testing image is determined based on its residue to the eigenmotion space. Both the individual modeling method and the common modeling method are described in this paper. Experimental results show that the proposed eigenmotion-based method outperforms the eigenface approach in the presence of facial expression variations. The approach can be extended to model other types of variations as well, for example, illumination and pose variations.

Keywords Eigenmotion-based method, Eigenface, Eigenmotion, Motion vector residue, Eigenmotion space

0 引言

简单地说, 人脸识别就是将待识别的人脸像(也称人脸)或视频, 与数据库中已知人脸像进行比较, 从而得出该人脸像的标识. 这是模式识别领域的一个重要研究方向. 因为其在安全验证系统、视频会议、人机交互等方面有巨大应用前景, 所以人脸识别

正越来越成为当前模式识别和人工智能领域的一个研究热点.

人脸识别的方法有很多^[1]. 概括而言, 这些方法可以归为如下两种类型, 一种是基于几何和面部特征匹配的方法, 另一种是基于模板匹配的方法. 20世纪90年代初, Turk 和 Pentland 提出了基于主成分分析(PCA)的特征脸方法^[2], 它就是一种具有代表性的模板匹配方法, 其基本思想是在像素域进行

PCA,然后将高维空间的人脸投影到由特征脸构成的特征脸空间中,最后在特征脸空间里进行识别.在限制光照、人脸表情和姿态变化的情况下,PCA方法有非常不错的识别率.

针对这些限制,近年来研究者在姿态不变、光照不变的人脸识别上 (pose-invariant, illuminant-invariant face recognition) 进行了很多工作^[3,4],但是对于表情人脸,大量的工作则集中在表情的识别上,即给定任意一个人脸像,用一些方法识别出其表情^[5],而对表情不变的人脸识别在很大程度上被忽略了. Yacoob 比较了基于统计的方法(特征脸方法)和基于几何特征的方法(feature-graph)对表情人脸的识别结果,得出特征脸方法稍稍好一点的结论^[6],但是用特征脸方法识别表情人脸时,至少有如下两个问题没有很好解决:(1)构成特征脸空间的特征脸不仅表征了不同人的脸之间的区别,同时也表征了同一个人自身的多种人脸变化^[7],如由姿态,表情引起的人脸变化,而后者显然对人脸识别造成干扰,针对这一弱点,在本文的算法中,为每一个人建立了一个仅表征它自身的特征空间,由于这个空间更好地反映了这个人本身的特点,因此测试人脸像到子空间的距离就成为识别的一个很好依据;(2)特征脸方法作用在像素域上,由于其是表征人脸像各个像素的灰度变化的一种模型,因此不能很好地表征人脸像平面上的运动,可是当存在表情变化时,像素却在人脸像平面上存在着运动.针对这点,本文提出特征运动的概念,其基本思想是在人脸像平面上建立像素运动的模型.该模型描述了同一个人做出各种表情变化时的脸部像素的运动特性,相应地,在测试阶段就可用该模型来识别表情人脸.

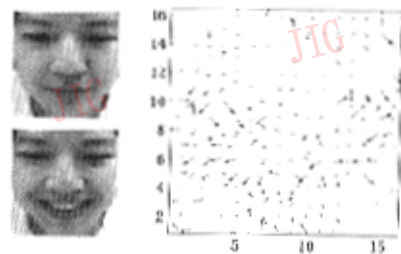
1 基于特征运动的方法

大家知道,人类所具有的丰富的表情都是由面部特征的位移所造成的,正是从表情变化的面部运动特性出发,本文提出基于特征运动的方法.该算法的主要思想是先通过表情人脸像和无表情人脸像之间的运动向量来进行人脸识别,因为运动向量表示了喜怒哀乐各种表情是如何由无表情人脸像转变而来的;然后通过对这些运动向量施以 PCA,来建立基于特征运动的特征空间,即特征运动空间;最后在这个空间里进行人脸识别.本文是采用块匹配的方法^[8]来求二维运动向量,其中,分块的思想不仅突出

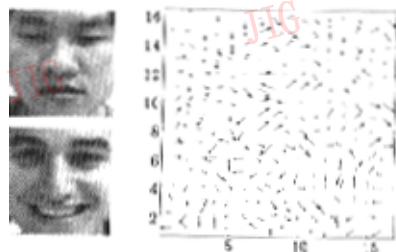
重点,即抓住了面部主要特征的运动,而且也减小了因为姿态、尺度的轻微改变所造成的噪声干扰.

1.1 基于特征运动的个人模型方法

图 1(a)和图 1(b)分别是同一个人和不同人的表情人脸像和无表情人脸像,以及它们之间的运动向量.由图 1 可以看到,不同人的表情人脸像之间的运动向量非常随机杂乱,而同一个人表情人脸像之间的运动向量则具有一定的规律性,譬如,图 1(a)的运动向量明显地表现了笑的时候,嘴角上扬的运动.其实,不仅表情人脸像如此,而且同一个人姿态不同的两幅人脸像之间的运动向量,也能够反映姿态的变化,例如,是左转,还是右转,是抬头还是低头.总而言之,只要所比较的两幅人脸像是同一个人的,那么所得到的运动向量就具有一定的规律,这就是本文用特征运动来进行人脸识别的根据.当系统的训练数据库中,包含每个用户的多个表情人脸像时,则可以为每个用户建立一个属于它自己的特征运动空间,以用于反映其自身的表情运动特性.这就是本文提出的基于特征运动的个人模型方法.



(a) 同一个人有表情和无表情人脸



(b) 不同人有表情和无表情人脸

图 1 对于不同表情的人脸像求得的运动向量

在训练阶段,假定训练数据库里有 K 个人(用户)的人脸像,每个人都有 $M+1$ 幅人脸像,其中包括 M 幅表情人脸像和 1 幅无表情人脸像.对于每个人,可利用块匹配的方法来计算 M 幅表情人脸像与

无表情人脸像之间的运动向量. 这里用 v_{ij} 表示第 i 个人的第 j 幅表情人脸像与无表情人脸像之间的运动向量, 其每个人脸像的平均运动向量 \bar{v}_i 为

$$\bar{v}_i = \frac{1}{M} \sum_{j=0}^{M-1} v_{ij} \quad (i = 0, 1, \dots, K-1)$$

由 $s_{ij} = v_{ij} - \bar{v}_i$ 组成矩阵 A_i 的列向量, $A_i = [s_{i,0}, s_{i,1}, \dots, s_{i,M-1}]$, 这里将利用 PCA 方法求出的对应于矩阵 $A_i A_i^T$ 的前 Q 个较大特征值 $\lambda_{i,n}$ 的正交特征向量 $u_{i,n}$ 称为特征运动, 并由特征运动构成特征运动空间

$$A_i A_i^T u_{i,n} = \lambda_{i,n} u_{i,n} \quad (n=0, 1, \dots, Q-1)$$

图 2 的上面为同一个人的 5 幅人脸像, 左边 1 幅是无表情人脸像, 右边 4 幅是表情人脸像, 下面是所建立的特征运动空间的前两个特征运动. 从图 2 可见, 在眼睛、嘴角等面部特征部位存在较明显的运动.

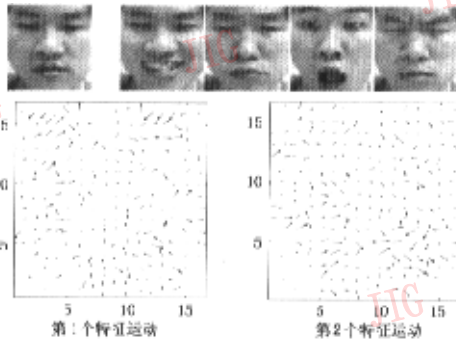


图 2 构成个人特征运动空间的训练图象和前两个特征运动

在测试阶段, 首先求出测试人脸像到第 i 个人 ($i=0, 1, \dots, K-1$) 无表情人脸像的运动向量 v_i , 然后将 v_i 投影到第 i 个人人脸像的特征运动空间, 即得到在每一个特征运动方向的系数:

$$w_{i,n} = u_{i,n}^T (v_i - \bar{v}_i) \quad (n=0, 1, \dots, Q-1; i=0, 1, \dots, K-1)$$

反过来, 也可以利用系数向量 $p_i^T = [w_{i,0}, w_{i,1}, \dots, w_{i,Q-1}]$ 来重建运动向量 \hat{v}_i :

$$\hat{v}_i = \sum_{n=0}^{Q-1} w_{i,n} u_{i,n} + \bar{v}_i \quad (i=0, 1, \dots, K-1)$$

因为在特征运动空间中只保留了矩阵 $A_i A_i^T$ 的较大特征运动, 所以在原始运动向量与重建的运动向量之间必定存在残差, 这种残差称为运动向量残差, 定义为

$$e_i = \|v_i - \hat{v}_i\|^2 \quad (i=0, 1, \dots, K-1)$$

测试一共得到 K 个运动向量残差, 而测试人脸即属于人脸数据库中与其 e_i 值最小的那个人(用

户). 因为只有当测试人脸像和其自身的无表情人脸像产生的运动向量投影到自身的特征空间时, 才有最好的重建运动向量, 且其所得到的运动向量残差包含的能量才最小.

1.2 基于特征运动的公共模型方法

尽管个人模型方法对表情人脸像有很好的识别率, 但是在实际应用中, 训练数据库中可能并没有为每个人保存多幅表情人脸像, 而仅有一幅无表情人脸像. 针对这种限制, 本文提出基于特征运动的公共模型方法.

尽管每个人的脸部外形不一, 但是当一个人做出表情的时候, 面部的肌肉运动都是类似的, 也就是说, 表情运动(即描述表情变化的运动向量)具有相当的共性, 譬如说, 人们微笑的时候, 一般是嘴角向两边, 眉头舒展. 据此, 公共模型方法就利用预先建立的一个特征运动空间来表示人们共同的表情变化. 识别时, 首先收集一个表情人脸像数据库, 其中包含了很多人的多幅表情人脸像和其相应的无表情人脸像. 然后计算同一个人的表情人脸像和其无表情人脸像之间的运动向量, 并且利用所有这些运动向量以 PCA 方法建立一个公共的特征运动空间. 由于由这些特征运动组成的特征运动空间抽取了表情运动的主要特性, 因此, 所有同一个人的合理表情运动都可以由这些特征运动的线性组合来较好地表示; 相反, 不同人之间的表情运动杂乱无章, 就不能由这些运动向量来表示. 也就是说, 从表情人脸像数据库中, 特征运动空间学习了人们是怎样做出各种表情的. 当基于公共模型的人脸识别系统交付给用户时, 只要求每个待识别的人提供一幅无表情的人脸像即可, 尽管这些人脸像已形成了一个训练数据库, 但是, 不同于通常的系统, 这里并不要对这个训练数据库进行任何的训练, 因为所有的训练在构建公共特征运动空间时, 已经完成了.

当测试人脸像提交给一个有 K 个用户的系统进行测试时, 首先将测试人脸像与训练数据库中的每个无表情人脸像进行比较, 同时计算其运动向量 v_i ($i=0, 1, \dots, K-1$); 然后利用预先建立的公共特征运动空间的特征运动来重建运动向量 \hat{v}_i , 并求出运动向量残差, 它是识别人脸的依据, 因为只有来自同一个人的规则表情运动, 在特征运动空间里, 才会得到较小的运动向量残差, 如果用不同人的表情人脸像进行比较, 则将会产生随机杂乱的运动向量, 而且其运动向量残差也会相当的大, 实验也证明了这一

点;最后,测试人脸像被识别为与其具有最小运动向量残差的那个用户。

2 实验结果

实验采用的是 Carnegie Mellon 大学的表情人脸像数据库^[9],为了重点研究本文的方法对于表情人脸像的识别能力,要求数据库中所有人脸像都是在同样光照条件下拍摄得到,同时去除了背景.由于人脸跟踪的程序可以自动跟踪眼睛的位置,所以通过定位两个眼睛,使所有的人脸像都缩放到相同的尺度,即每幅人脸像都是 64×64 个像素的灰度人脸像.另外,测试的每一幅人脸像必定属于训练用户之一,也就是说,测试集是封闭的.

2.1 基于特征运动的公共模型和特征脸方法的比较

对于基于特征运动的公共模型方法,利用预先收集的表情人脸像数据库来构建一个基于特征运动的公共特征运动空间.当得到这个空间后,在一个 13 人的脸像数据库上进行了测试.对于每个用户,数据库中保存了 1 幅无表情训练人脸像和 70 幅测试人脸像.对每一幅测试人脸像,首先将它与数据库里 13 幅无表情训练人脸像进行比较,得到 13 个运动向量,然后将其投影到预先建立的公共特征运动空间,最后利用运动向量残差作为判定依据,进行人脸识别.对于特征脸方法,先用数据库里的 13 个无表情人脸像作为训练数据,建立特征脸空间;然后将测试人脸像投影到特征脸空间;最后利用最近邻居法则进行判断.

从表 1 中可以看到,对于同样数目的特征向量,基于特征运动的公共模型方法的识别率高于特征脸方法.此外,该算法的另一个显著优点是,识别率几乎不受特征向量数目的影响,这是因为表情运动并不是一种很复杂的运动,很少的几个特征运动就足够表示其运动特性.对于大型人脸像数据库而言,这是一个非常显著的优点,也就是说,只要取非常少的特征向量,就可以达到非常高的识别率,例如,仅选

表 1 在不同数目的特征向量下,公共模型方法和特征脸方法的识别率比较

识别方法	特征向量数目							
	1	2	3	4	5	6	7	8
特征脸方法(%)	46.6	67.9	91.8	93.2	93.0	91.7	91.8	89.9
基于特征运动的公共模型(%)	96.9	97.1	96.9	96.2	96.4	95.5	95.7	95.5

取两个特征向量,就可以达到 97.1% 的识别率;而在特征脸方法中,则无论选取多少个特征向量,都达不到这么高的识别率.当特征向量的个数为 4 时,特征脸方法有最好的识别率,其后,更多的特征向量反而会使识别率下降.表 1 中,公共模型方法选取的块大小为 8×8 ,另外,在实验中,还选择了 $2 \times 2, 4 \times 4, 8 \times 8$ 3 种块大小进行对比,它们的结果非常近似,其中又以 8×8 块大小的实验结果最好.

2.2 基于特征运动的个人模型和特征脸方法的比较

在基于特征运动的个人模型方法和特征脸方法的比较实验中,人脸像数据库包含了 13 个用户,其中,训练集每个用户有 5 个人脸像,包括 4 幅表情人脸像和 1 幅无表情人脸像;测试集中,每个用户有 70 幅具有不同表情的人脸像.在个人模型方法中,训练阶段通过对每个用户的每一幅表情人脸像和无表情人脸像进行比较,得到 4 个运动向量,然后利用这 4 个运动向量来建立一个特征运动空间.这样为 13 个用户一共建立了 13 个特征运动空间.对于每一个空间,特征运动的数目最多为 3.识别时,首先将测试人脸像分别与 13 个人的无表情训练人脸像比较,然后将得到的运动向量投影到对应的特征运动空间,最后通过比较运动向量残差进行人脸识别;而用特征脸方法识别时,先利用 13×5 幅人脸像训练一个全局特征脸空间,再通过寻找测试人脸像来对这个特征脸空间里的最近邻居进行识别.通过穷举特征脸方法所有的特征向量个数,发现当数目为 12 时,识别率达到最高.由表 2 可见,取同样数目的特征向量,基于特征运动的个人模型方法的识别率高于特征脸方法.尤其是,本文新方法选取两个特征向量就可以取得很好的识别率,而传统的特征脸方法,则要使用大量的特征向量才能获得类似的识别率.另外,与表 1 相比,个人模型方法比公共模型方法有更好的识别结果,这说明,为每一个人单独建立一个属于他自己的特征空间,比利用所有人的数据建立的全局特征空间要精确得多.

表 2 在不同数目的特征向量下,个人模型方法和特征脸方法的识别率比较

识别方法	特征向量数目									
	1	2	3	4	5	6	8	10	12	14
特征脸方法(%)	61.2	80.1	89.3	97.4	96.0	98.4	99.0	98.9	99.7	99.7
基于特征运动的个人模型(%)	99.3	99.5	99.3							

3 总 结

本文介绍了一种用于人脸识别的新算法,它对于表情人脸像有非常高的识别率.该算法是首先将人脸像分割为若干块,并利用块的运动向量来描述表情人脸像和无表情人脸像之间的运动,同时对运动向量进行主成分分析,以产生特征向量(也就是本文提出的特征运动);然后以特征运动作为基元所建立的特征运动空间就表征了人脸像的普遍表情变化;最后将测试人脸像和不同无表情人脸像之间的运动向量投影到特征运动空间,再利用运动向量残差来进行人脸识别.考虑到实际应用中所能拥有的训练数据情况,还提出了个人模型方法和公共模型方法,识别时,应根据不同的情况选择不同的模型,即当训练数据中,包含了每个用户的一个无表情人脸像和多个有表情人脸像时,则使用基于特征运动的个人模型方法;当训练数据只包含每个用户的一个无表情人脸像时,则使用基于特征运动的公共模型方法.正是由于特征运动很好地表示了人脸的表情变化,这两种基于特征运动的方法都比特征脸方法有更高的识别率.

研究发现,本文方法的一个要求是测试数据和训练数据要在相同的光照条件下采集.而在很多的应用场合下,由于光照基本保持恒定,因此该条件还是可以被满足的,比如,安置于机场大厅的人脸识别系统.同时,笔者也正在研究利用图象处理技术进行预处理来去除光照的影响.

本文的核心思想是应用 PCA 方法来建立图象之间“差别”的模型,并用该模型进行人脸识别,同时以图象之间的运动向量来描述由人脸表情引起的图象“差别”.另外,这一思想还可以有别的应用,比如,用 affine 运动模型来描述由人脸不同姿态(pose)引起的图象“差别”,就可以让系统在一定范围内,对任何姿态的人脸都能进行识别.为了获取更好的识别率,本文的方法作为一种仅仅依赖于图象“差别”的识别算法,还可以和以前的依赖于图象自身信息进行识别的系统(如利用脸部特征分布的方法和利用脸部灰度的特征脸方法)相结合来进行人脸识别.这也是正在进行的进一步研究的课题.

参 考 文 献

1 Chellappa R, Wilson C L, Sirohey S. Human and machine

- recognition of faces: a survey [J]. Proceedings of the IEEE, 1995, 83(5): 705~741.
- 2 Turk M, Pentland A. Eigenfaces for recognition [J]. Journal of Cognitive Neuroscience, 1991, 3(1): 71~86.
- 3 Pentland A, Moghaddam B, Starner T. View-based and modular eigenspaces for face recognition [A]. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition 1994 [C], Seattle, WA, USA, 1994: 84~91.
- 4 Gao Y, Leung M K H, Wang W *et al.* Fast face identification under varying pose from a single 2-D model view [J]. Proceedings of the IEEE, Vision, Image and Signal Processing, 2001, 148(4): 248~253.
- 5 Lien J J, Kanade T, Cohn J F *et al.* Automated facial expression recognition based on FACS action units [A]. In: Proceedings of Third IEEE International Conference on Automatic Face and Gesture Recognition [C], Nara, Japan, 1998: 390~395.
- 6 Yacoob Y, Lam H, Davis L S. Recognizing faces showing expressions [A]. In: International Workshop on Automatic Face-and Gesture-Recognition [C], Zurich, 1995: 278~283.
- 7 Belhumeur P N, Hespanha J P, Kriegman D J. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection [J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 1997, 19(7): 711~720.
- 8 Tekalp A M. Digital video processing [M]. Englewood Cliffs, NJ: Published by Prentice Hall, Inc., a Simon & Schuster Company, 1995: 95~116.
- 9 Carnegie Mellon University face expression database [DB/OL]. <http://amp.ece.cmu.edu/downloads.htm>.



余 冰 1978 年生,浙江大学计算机科学与工程系硕士研究生.主要研究方向为模式识别和计算机视觉.



金连浦 1947 年生,浙江大学计算机科学与工程系副教授.主要研究多媒体技术及信息处理.



陈 平 1964 年生,浙江大学计算机科学与工程系副教授.主要研究多媒体技术及计算机网络技术.