

SVD 和 DCT 提取特征向量的方法在人脸识别中的比较

卢 健¹ 王 晖¹

(1 武汉大学遥感信息工程学院,武汉市珞喻路 129 号,430079)

摘 要:采用奇异值分解和 2D 离散余弦变换两种方法得到的特征,分别在嵌入式隐马尔可夫模型的人脸识别中进行比较,得出奇异值分解比 2D 离散余弦变换在识别率上和时间复杂度上都较差的结果。通过对结果进行分析得出,虽然奇异值分解有很多优良的特性,但是在模式识别中仅使用奇异值分解来提取特征并不是很好的方法。

关键词:SVD;DCT;EHMM;特征;人脸识别

中图法分类号:TP751.1

近年来,有许多基于奇异值分解来抽取特征向量的方法^[1~4]。本文用嵌入式隐马尔可夫模型(embedded hidden Markov model,EHMM)为分类方法,分别采用奇异值分解(singular value decomposition,SVD)和 2D 离散余弦变换(discrete cosine transform,DCT)为特征向量的提取方法进行试验,并比较了这两种方法的优劣。

1 奇异值分解

令 $\mathbf{A}_{m \times n}$ 是实矩阵(不失一般性,设 $m > n$),且 $\text{rank}(\mathbf{A}) = k$,则存在两个正交矩阵 $\mathbf{U}_{m \times m}$ 和 $\mathbf{V}_{n \times n}$ 及对角阵 $\mathbf{D}_{m \times n}$ 使下式成立:

$$\mathbf{A} = \mathbf{U} \mathbf{D} \mathbf{V}^T \tag{1}$$

式中, $\mathbf{D}_{m \times n} = \begin{pmatrix} \boldsymbol{\Sigma}_{k \times k} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix}$, $\boldsymbol{\Sigma}_{k \times k} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_k)$, $\mathbf{U}_{m \times m} = (u_1, u_2, \dots, u_{k+1}, \dots, u_m)$; $\mathbf{V}_{n \times n} = (v_1, v_2, \dots, v_k, v_{k+1}, \dots, v_n)$ 。其中, $\sigma_i = \sqrt{\lambda_i}$ ($i = 1, 2, \dots, k, \dots, n$) 称为矩阵 \mathbf{A} 的奇异值, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k > 0$ 是 $\mathbf{A} \mathbf{A}^T$ 和 $\mathbf{A}^T \mathbf{A}$ 的非零特征值的全体,而 $\lambda_{k+1} = \lambda_{k+2} = \dots = \lambda_n = 0$ 为 $\mathbf{A}^T \mathbf{A}$ 的 $n - k$ 个零特征值; u_i, v_i ($i = 1, 2, \dots, k$) 分别是 $\mathbf{A} \mathbf{A}^T$ 和 $\mathbf{A}^T \mathbf{A}$ 对应于非零特征值 λ_i 的特征向量。 u_i ($i = k + 1, \dots,$

m) 是为了表示上的方便而引入的 $(m - k)$ 个向量,可以设想它是 $\mathbf{A} \mathbf{A}^T$ 对应于 $\lambda_i = 0$ 的特征向量。同理, v_i ($i = k + 1, \dots, n$) 为 $\mathbf{A}^T \mathbf{A}$ 对应于 $\lambda_i = 0$ 的特征向量。将式(1)写成乘积的形式:

$$\mathbf{A} = \sum_{i=1}^k \sigma_i u_i v_i^T \tag{2}$$

如果矩阵 \mathbf{A} 代表一幅人脸图像,式(2)就是对应人脸图像进行正交分解。将矩阵 $\boldsymbol{\Sigma}$ 中主对角线上的奇异值元素 σ_i 连同 $\mathbf{D}_{m \times n}$ 中剩余的 $(n - k)$ 个 0 构成一个 n 维列向量:

$$\mathbf{x}_{n \times 1} = \mathbf{D}_{n \times n} \mathbf{e} = (\sigma_1, \dots, \sigma_k, 0, \dots, 0)^T \tag{3}$$

式中, $\mathbf{D}_{n \times n}$ 为 \mathbf{D} 中的第一个 n 阶子式;列向量 $\mathbf{e} = (1, 1, \dots, 1)_{n \times 1}^T$,称 $\mathbf{x}_{n \times 1}$ 为 \mathbf{A} 的奇异值特征向量。对于任何实矩阵 \mathbf{A} ,在 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k$ 的限制下,奇异值对角矩阵 $\boldsymbol{\Sigma}$ 是惟一的,因此,原人脸图像 \mathbf{A} 对应于惟一的奇异值特征向量。

2 2D 离散余弦变换

2D 离散余弦变换定义为:

$$C(u, v) = a(u) a(v) \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \cdot \cos \left[\frac{(2x+1)u\pi}{2M} \right] \cos \left[\frac{(2y+1)v\pi}{2N} \right] \tag{4}$$

($u = 0, 1, 2, \dots, M-1; v = 0, 1, 2, \dots, N-1$)

式中, $C(u, v)$ 为变换结果, 也称 DCT 系数。

$a(u)$ 、 $a(v)$ 分别定义为:

$$a(u) = \begin{cases} \sqrt{1/M}, u = 0 \\ \sqrt{2/M}, u = 1, 2, \dots, M-1 \end{cases}$$

$$a(v) = \begin{cases} \sqrt{1/N}, v = 0 \\ \sqrt{2/N}, v = 1, 2, \dots, N-1 \end{cases}$$

离散余弦变换的特点是: 频域变化因子 u, v 较大时, DCT 系数 $C(u, v)$ 的值很小; 而数值较大的 $C(u, v)$ 主要分布在 u, v 较小的左上角区域, 这也是有用信息的集中区域。

3 嵌入式隐马尔可夫模型

EHMM 由一组超状态序列和多组嵌入的子状态序列组成, 每个人脸用一个 EHMM 的参数来表示, 人脸的差异就是通过模型参数的不同表现出来的。一个 EHMM 可记为:

$$\lambda = (\pi, \mathbf{A}, \mathbf{\Lambda})$$

式中, π 为超状态初始概率分布, $\pi = \{\pi_i, 1 \leq i \leq N\}$; N 为垂直方向超状态数; \mathbf{A} 为超状态转移概率矩阵, $\mathbf{A} = \{a_{ij}, 1 \leq i, j \leq N\}$; $\mathbf{\Lambda}$ 为超状态内嵌入的水平方向的子 HMM 序列, 即嵌入的 HMM, $\mathbf{\Lambda} = \{\mathbf{\Lambda}^i, 1 \leq i \leq N\}$; 第 i 个超状态可表示为 $\mathbf{\Lambda}^i = \{\pi^i, \mathbf{A}^i, \mathbf{B}^i\}$, π^i 为第 i 个超状态中嵌入状态的初始概率分布, $\pi^i = \{\pi_k^i, i \leq k \leq N^i\}$; N^i 为第 i 个超状态中嵌入状态数, 子状态 $S^i = \{S_k^i, 1 \leq k \leq N^i\}$; \mathbf{A}^i 为第 i 个超状态中嵌入状态的转移概率矩阵, $\mathbf{A}^i = \{a_{kl}^i, 1 \leq k, l \leq N^i\}$; \mathbf{B}^i 为第 i 个超状态中嵌入的子状态的观察值概率密度函数, $\mathbf{B}^i = \{b_k^i(\mathbf{O}_{t_0, t_1})\}$, 其中 \mathbf{O}_{t_0, t_1} 表示 t_0 行、 t_1 列的观察向量 ($t_0 = 1, \dots, T_0$; $t_1 = 1, \dots, T_1$), \mathbf{B}^i 通常用多维高斯概率密度函数来近似表示:

$$b_k^i(\mathbf{O}_{t_0, t_1}) = \sum_{f=1}^F C_{kf}^i N(\mathbf{O}_{t_0, t_1}, \mu_{kf}^i, \mathbf{U}_{kf}^i)$$

$$(1 \leq k \leq N^i)$$

式中, F 为组成 $b_k^i(\mathbf{O}_{t_0, t_1})$ 的混合概率密度函数的个数; C_{kf}^i 是超状态 i 中嵌入状态 k 的第 f 个组合量的组合系数; $N(\mathbf{O}_{t_0, t_1}, \mu_{kf}^i, \mathbf{U}_{kf}^i)$ 为多维高斯概率密度函数。其中, μ_{kf}^i 为超状态 i 中嵌入状态 k 的第 f 个组合量的均值向量, \mathbf{U}_{kf}^i 为超状态 i 中嵌入状态 k 的第 f 个组合量的协方差矩阵。

EHMM 有三个基本算法。

1) 给定模型参数 $\lambda = (\pi, \mathbf{A}, \mathbf{\Lambda})$ 和观察向量序列 \mathbf{O} , 如何计算概率 $p(\mathbf{O}|\lambda)$, 即观察向量序列与

模型吻合的概率。它是一个评估问题, 可决定出在多个模型中选择与观察向量序列最佳匹配的模型。本文采用向前向后算法^[5]。

2) 给定模型参数 $\lambda = (\pi, \mathbf{A}, \mathbf{\Lambda})$ 和观察向量序列 \mathbf{O} , 选择最佳状态序列来解释观察向量序列, 即揭开模型的隐含部分, 在优化准则下找到最优状态序列。本文采用双重嵌套的 Viterbi 算法^[6]。

3) 给定观察向量序列 \mathbf{O} , 如何调整模型参数, 使 $p(\mathbf{O}|\lambda)$ 最大, 即模型参数估计问题。本文采用 Baum-Welch 算法^[6]。

根据人脸五官特征, 人脸图像的 EHMM 选择如图 1 所示的拓扑结构。文献[7]认为, 取 5 个超状态对应人脸的前额、眼睛、鼻子、嘴和下巴, 分别描述和代表人脸的宏观特征。每个超状态内嵌入的状态分别提取人脸局部区域特征, 嵌入状态数取 (3, 6, 6, 6, 3) 的结构时具有较好的识别精度。因此, 训练方法和识别方法采用文献[7]中的方法, 算法流程图如图 2 所示。

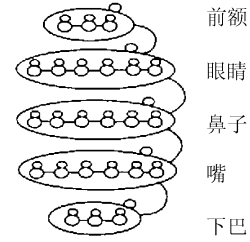


图 1 人脸图像的 EHMM

Fig. 1 EHMM for Face Image

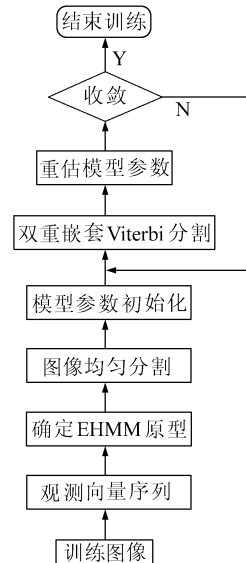


图 2 算法流程图

Fig. 2 Algorithmic Flow

4 试验分析

本文采用奇异值分解和离散余弦变换来进行观察向量的选择。具体做法是:在人脸图像平面上分别用一个像素点数为 8×8 、 12×12 、 16×16 、 20×20 、 24×24 、 28×28 、 32×32 的采样窗,从左到右、从上向下滑动,相邻窗口在垂直方向和水平方向分别有 4 行、4 列重叠来获取观察图像块,如图 3。然后对获得的观察图像块分别采用离散余弦变换,取变换后的矩阵左上角的 2×2 和 3×3 元素作为特征向量和采用奇异值分解得到向量的前 4 个和 9 个元素分别作为两种特征向量。

试验在 ORL (Olivetti research laboratory) 上进行。ORL 人脸库包含了 40 个人每人 10 幅共 400 幅人脸图片,这些人脸图片是在不同时间、不同光照条件、不同表情的情况下拍摄的面部正

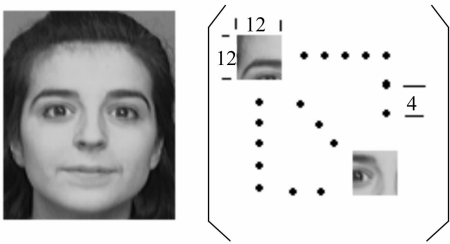


图 3 观察向量序列
Fig. 3 Observation Sequence

面或者近似正面的照片。每个人取第一到第五幅照片作为样本进行训练,第六到第十幅照片进行识别。总共进行了 26 600 次试验。试验平台为迅驰 1.4G CPU, 256M 内存, 30G 硬盘, IBMR50 53C 笔记本电脑, 操作系统采用 Windows XP Home 版, 开发平台为 VC++ 6.0。结果如表 1~表 2 所示。

表 1 四维特征向量 SVD 和 DCT 识别率、训练和识别时间的比较

Tab. 1 Comparison Between SVD and DCT at 4D Vectors in
Recognition Ratio, Training and Recognition Time Cost

	8×8	12×12	16×16	20×20	24×24	28×28	32×32
识别率 DCT	97.5	97.5	98	97	96.5	93.5	90.5
SVD	85.5	81.5	87	86.5	82.5	80.5	82
平均训练和识别时间/ms DCT	242/569	216/488	218/401	195/363	183/323	174/289	164/276
SVD	484/464	727/476	1 068/521	1 431/559	1 884/608	5 505/715	2 874/575

表 2 九维特征向量 SVD 和 DCT 识别率、训练和识别时间的比较

Tab. 2 Comparison Between SVD and DCT at 9D Vectors in
Recognition Ratio, Training and Recognition Time Cost

	8×8	12×12	16×16	20×20	24×24	28×28	32×32
识别率 DCT	99	97.5	98.5	98.5	98.5	96.5	95.5
SVD	—	87	89.5	90.5	88.5	84.5	85.5
平均训练和识别时间/ms DCT	271/596	268/558	248/499	226/466	212/415	200/366	184/355
SVD	—	745/613	1 102/640	1 472/674	1 919/703	5 570/762	2 935/813

从表 1、表 2 可以看出,在所有其他条件都不变的情况下,用 2D 离散余弦变换作为特征量,不管从识别率、训练时间、识别时间上,都比用奇异值分解作为特征量的效果好。

表 3 反映了奇异值方法中,窗口大小分别是 8×8 、 12×12 、 16×16 、 20×20 、 24×24 、 28×28 、 32×32 时,特征向量从 4 维递增到 32 维所有的识别率,其中最高的也只有 92%,比余弦变换最高为 99% 的识别率要低。

5 结 语

通过上述的试验比较和分析可知,奇异值分解用来提取特征向量在人脸识别中是可行的,但不是最优的。相比之下,离散余弦变换来提取特征向量,不管从识别率和速度上都是更优的选择。

本文所用的方法除了在水印识别上能够有较好的应用以外,也适用其他一些近景模式的识别问题,如手势识别。由于平台的限制,目前的卫星遥感图像的分辨率和时间连续性还不能满足要

求,但是不排除在遥感技术高度发展的未来,用这种方法进行遥感目标模式识别的可能性。

表 3 SVD 的识别率
Tab. 3 Recognition Ratio of SVD

维数	8×8	12×12	16×16	20×20	24×24	28×28	32×32
4	85.5	81.5	87	86.5	82.5	80.5	82
5	89.5	86	88	88.5	84.5	81.5	83.5
6	87	86.5	89.5	88.5	86	83.5	84
7	88.5	87	89.5	89.5	87	84.5	83
8	88	87	89	88.5	87.5	84	86
9	—	87	89.5	90.5	88.5	84.5	85.5
10	—	87	89.5	90.5	89	86	86
11	—	87	91	90	89	87.5	85.5
12	—	87	90.5	90.5	90	87.5	86.5
13	—	—	90.5	90.5	90.5	88	87.5
14	—	—	90	90	90	88.5	88.5
15	—	—	90	90.5	91	88	89
16	—	—	90	91	92	88	87.5
17	—	—	—	91.5	92	88	88
18	—	—	—	91	91.5	88.5	88.5
19	—	—	—	90.5	91.5	88.5	88.5
20	—	—	—	90.5	91.5	89	88.5
21	—	—	—	—	91.5	89	88.5
22	—	—	—	—	91	88	88.5
23	—	—	—	—	91	88.5	89
24	—	—	—	—	91	88.5	89.5
25	—	—	—	—	—	88.5	88.5
26	—	—	—	—	—	88.5	89
27	—	—	—	—	—	88.5	89.5
28	—	—	—	—	—	88.5	89.5
29	—	—	—	—	—	—	90
30	—	—	—	—	—	—	90
31	—	—	—	—	—	—	90
32	—	—	—	—	—	—	90

参 考 文 献

1 王蕴红,谭铁牛,朱勇. 基于奇异值分解和数据融合的脸像识别. 计算机学报, 2000,20(3):629~653

2 周德龙,高文,赵德斌. 基于奇异值分解和判别式 KL 投影的人脸识别. 软件学报, 2003,14(4):783~789

3 荆晓远,郭跃飞,杨静宇. 基于奇异值特征的图像预处理及人脸识别. 信息与控制,1999(2):116~120

4 刘小军. 一种基于奇异值分解和隐马尔可夫模型的人脸识别方法. 计算机学报, 2003,26(3):340~344

5 Lawrence R. A Tutorial on Hidden Markov Models and Selected Application in Speech Recognition. IEEE,1989,77(2):257~286

6 薛斌党,欧宗瑛. 嵌入式隐 Markov 模型和神经网络

人脸识别:[博士学位]. 大连:大连理工大学,2002

7 Ara V, Monson H. Face Recognition Using an Embedded HMM. IEEE International Conference Audio Video Biometric Based Person Authentication, Washington D. C. ,1999

8 Samaria F, Harter A. Parameterization of a Stochastic Model for Human Face Identification. The 2nd IEEE Workshop on Applications of Computer Vision, Sarasota,Florida,1994

第一作者简介:卢健,教授,博士生导师。研究方向为数字信号分析和图像处理、模式识别、计算机图形学和计算机视觉。
E-mail:jianlu2003@tom.com

4

王青,王融清,鲍虎军,等. 散乱数据点的增量快速曲面重建算法. 软件学报, 2000, 11 (9) : 1 221~1 227

5

Zheng S Y, Zhan Z Q, Zhang Z X. A Flexible and Automatic 3D Reconstruction Method. 2004 ISPRS Congress, Istanbul, Turkey, 2004

6

梁荣华,陈纯,潘志庚,等. 一种面向三维点集的快速表面重构算法. 中国图像图形学报, 2003, 8A(1): 63~67

第一作者简介:

郑顺义,副教授,博士。主要研究方向包括计算机视觉、数字摄影测量和虚拟现实等。
E-mail:syzheng@263.net

Automatic Reconstruction of 3D Surface Model with 3D Irregular Points Based on Projection Constrain

ZHENG Shunyi^{1, 2} SU Guozhong¹ ZHANG Zuxun¹

(1 School of Remote Sensing and Information Engineering, 129 Luoyu Road, Wuhan 430079, China)

(2 Geomatics and Applications Laboratory, Liaoning Technical University, 47 Zhonghua Road, Fuxin 123000, China)

Abstract: Surface model reconstruction from 3D irregular points is of great importance in variable fields such as photogrammetry, 3D reconstruction based on images, scientific computing visualization, etc. 3D points can be obtained with many mature methods such as photogrammetry, laser scanning, etc. While the surface model reconstruction from 3D irregular points is still a difficult problem. In this paper, a novel algorithm based on projective constrain for automatic generation of 3D surface model is proposed. It can make full use of projective constrain of points and make surface reconstruction easy. Experiment results show its reliability.

Key words: 3D-TIN(triangulated irregular network); 3D irregular points; projective constrain; 3D surface reconstruction

About the first author: ZHENG Shunyi, associate professor, Ph. D. His research focuses on computer vision, digital photogrammetry and virtual reality.
E-mail: syzheng@263.net

(责任编辑: 晓晨)

(上接第 121 页)

Comparison Between SVD and DCT Feature Extraction Methods in Face Recognition

LU Jian¹ WANG Hui¹

(1 School of Remote Sensing and Information Engineering, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)

Abstract: This paper compares SVD (singular value decomposition) with DCT (discrete cosine transform) in the face recognition, and concludes that SVD is less effective than DCT in recognition ratio and time cost. Experimental results show that SVD is an available feature extraction method, but is not a best one for pattern recognition.

Key words: SVD; DCT; EHMM; feature; face recognition

About the first author: LU Jian, professor, Ph.D supervisor. His major research orientations are digital signal analysis and image processing and computer vision.
E-mail: jianlu2003@tom.com

(责任编辑: 涓涓)