

文章编号:1671-8860(2005)02-0118-04

文献标志码:A

# SVD 和 DCT 提取特征向量的方法在人脸识别中的比较

卢 健<sup>1</sup> 王 晖<sup>1</sup>

(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<sup>[1~4]</sup>。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:

$$\mathbf{A} = \mathbf{U} \mathbf{D} \mathbf{V}^T \quad (1)$$

式中,  $\mathbf{D}_{m \times n} = \begin{pmatrix} \mathbf{\Sigma}_{k \times k} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix}$ ,  $\mathbf{\Sigma}_{k \times k} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_k)$ ,  $\mathbf{U}_{m \times m} = (u_1, u_2, \dots, u_{k+1}, \dots, u_m)$ ;  $\mathbf{V}_{n \times n} = (v_1, v_2, \dots, v_k, v_{k+1}, \dots, v_n)$ 。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  $\lambda_i$  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:

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

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  $\mathbf{\Sigma}$  中主对角线上的奇异值元素  $\sigma_i$  连同  $\mathbf{D}_{m \times n}$  中剩余的( $n-k$ )个0 构成一个  $n$  维列向量:

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

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  $x_{n \times 1}$  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  $\mathbf{\Sigma}$  是唯一的,因此,原人脸图像  $\mathbf{A}$  对应于唯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:

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

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

$$\lambda = (\pi, A, \Lambda)$$

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

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

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,  $\mu_{kf}^i$  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  $\lambda = (\pi, A, \Lambda)$  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<sup>[5]</sup>。

2) 给定模型参数  $\lambda = (\pi, A, \Lambda)$  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<sup>[6]</sup>。

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<sup>[6]</sup>。

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。

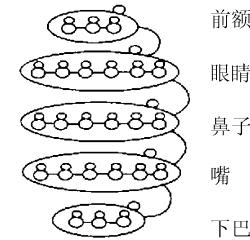


图 1 人脸图像的 EHMM

Fig. 1 EHMM for Face Image

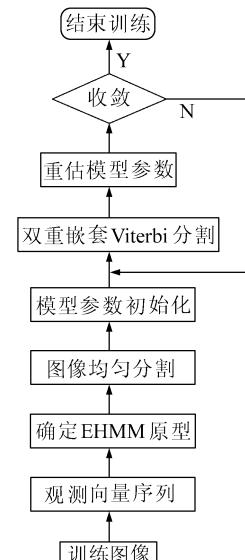


图 2 算法流程图

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 $8 \times 8$ 、 $12 \times 12$ 、 $16 \times 16$ 、 $20 \times 20$ 、 $24 \times 24$ 、 $28 \times 28$ 、 $32 \times 32$ 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 $2 \times 2$ 和 $3 \times 3$ 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

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

|                  | $8 \times 8$ | $12 \times 12$ | $16 \times 16$ | $20 \times 20$ | $24 \times 24$ | $28 \times 28$ | $32 \times 32$ |
|------------------|--------------|----------------|----------------|----------------|----------------|----------------|----------------|
| 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

|                  | $8 \times 8$ | $12 \times 12$ | $16 \times 16$ | $20 \times 20$ | $24 \times 24$ | $28 \times 28$ | $32 \times 32$ |
|------------------|--------------|----------------|----------------|----------------|----------------|----------------|----------------|
| 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 $8 \times 8$ 、 $12 \times 12$ 、 $16 \times 16$ 、 $20 \times 20$ 、 $24 \times 24$ 、 $28 \times 28$ 、 $32 \times 32$ 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。

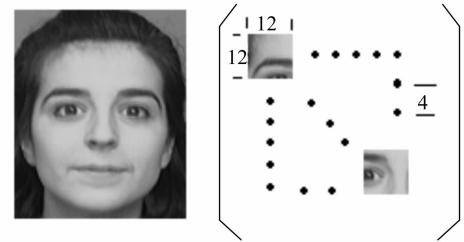


图3 观察向量序列

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。

## 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人脸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 | $8 \times 8$ | $12 \times 12$ | $16 \times 16$ | $20 \times 20$ | $24 \times 24$ | $28 \times 28$ | $32 \times 32$ |
|----|--------------|----------------|----------------|----------------|----------------|----------------|----------------|
| 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

- 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
- 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
- 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
- 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
- Lawrence R. A Tutorial on Hidden Markov Models and Selected Application in Speech Recognition. IEEE,1989,77(2):257~286
- 薛斌党,欧宗瑛.嵌入式隐Markov模型和神经网络

人脸识别:[博士论文].大连:大连理工大学,2002

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

文章编号:1671-8860(2005)02-0122-04

文献标志码:A

# 数码相机的畸变差检测研究

林宗坚<sup>1</sup> 崔红霞<sup>2</sup> 孙杰<sup>1</sup> 刘丽<sup>3</sup>

(1 中国测绘科学研究院,北京市海淀区北太平路 16 号,100039)

(2 武汉大学遥感信息工程学院,武汉市珞喻路 129 号,430079)

(3 武汉大学资源与环境科学学院,武汉市珞喻路 129 号,430079)

**摘要:**提出了一种检测面阵 CCD 数码相机畸变差的方法。首先利用标准三维检校场测定面阵 CCD 数码相机的内方位元素和系统畸变系数,然后依据投影几何中空间直线经中心投影变换后仍是直线的原理,提出了对面阵 CCD 数码相机随机畸变的检测和改正方法。利用该方法对 UAVRS-II 系统中的面阵数码相机的畸变进行了检校,获取了满意的结果。

**关键词:** 相机检校; 畸变; 附加参数; 边缘检测

中图法分类号:P246.1; P231.5

一般地,数码相机的检校包括主点位置( $x_0, y_0$ )、主距( $f$ )和畸变差的测定。目前,数码相机畸变差的检测方法可以分为两大类:①建立数学畸变模型,实现“在任”或自检校改正,常规摄影测量采用该方法对相机实施检校;②建立面积不同的准二维控制场,测定各类成像系统每个像素的总畸变,建立该影像专用的数字畸变模型,以对各类成像系统各个环节产生的各种畸变进行总体补偿<sup>[1]</sup>。以上两种方法中,前者忽略了引起数码相机畸变的随机误差,如 CCD 组件引起的电学畸变等;而后者则没有对综合畸变误差中的随机畸变给出定量的分析。本文在对 UAVRS-II 系统采用的非量测用面阵数码相机检校的过程中,分别对系统畸变和随机畸变给出了定量的分析。

## 1 数码相机的检校

数码相机的畸变差分为系统畸变差和随机畸变差。光学系统畸变差和 CCD 纵横方向排列的不垂直性、像素的非正方形引起的畸变差呈现一定的系统性;电学因素等引起的畸变误差呈现较强的随机性。本文提出的检测相机畸变差方法的基本步骤是:采用附加参数的光束法平差测定内方位元素和系统畸变系数;根据空间直线经过中

心投影在像平面上的影像仍然是直线的原理,对随机畸变差进行检测并给出定量的分析。

### 1.1 附加参数光束法平差测定内方位元素和系统畸变系数

对高精度室内三维实验场进行数据采集,采用附加参数光束法平差<sup>[2]</sup>测定内方位元素和系统畸变系数。

#### 1.1.1 室内三维控制场拍摄

为使每张影像具有相同的主距  $f$ ,并且避免数码相机结构上的不严谨所造成的焦距变化以及主点位置( $x_0, y_0$ )变化,在拍摄时,将摄影方式设为手动曝光方式,并将物镜的焦距固定在无穷远处,以达到锁定主距的目的(锁定主距也就锁定了内方位元素值和物镜畸变系数)。实验使用 UAVRS-II 机载面阵数码相机,其面阵大小为 3 008 像元  $\times$  2 000 像元。采用四站对角交向摄影方式,每站曝光两次,共拍摄了 8 张影像,镜头焦距为 20 mm,光圈号数为 18,影像以 tif 格式存储输出。

超焦点距离(又称为无穷远点<sup>[3]</sup>)为  $H = \frac{f^2}{K \times E}$ 。其中  $E = 0.05$ ,  $F = 20$  mm,  $K = 18$ 。因此,当摄影距离大于 0.4 m 时,就达到了“无穷远距离”的条件。