

人脸图象

人脸识别系统

人脸识别
形变匹配

④

计算机科学 1998 Vol. 25 No. 5

人脸图象识别系统及其方法^{*}

Face image recognition system and recognition methods

熊志勇 刘翼光 沈理

(中国科学院计算技术研究所 北京 100080)

43-46

TP391.41

摘要 A face image recognition system is presented in this paper. It includes the preprocessing, representation and feature-extraction of face image and face recognition. Three recognition methods are specially introduced.

关键词 Face recognition, Eigenfaces, Elastic matching, ANN

一、引言

如同人的指纹一样,人脸也具有唯一性,也可用来鉴别一个人的身份。现在已有实用的计算机自动指纹识别系统面世,并在安检等部门得到应用,但还没有通用成熟的人脸自动识别系统出现。人脸图象与指纹图象不同的是,前者受很多因素的干扰:人脸表情的多样性、以及外在的成像过程中的光照,图象尺寸、旋转、姿势变化等,即使同一个人,在不同的环境下拍摄所得到的人脸图象也不同,有时更会有很大差别。因此在各种干扰条件下实现人脸图象的识别,就更具有挑战性。国外对于人脸图象识别的研究较早,现已有实用系统面世,只是对于成像条件要求较苛刻,应用范围也就较窄。在调研中发现,国内也有许多科研机构从事这方面的研究,并已取得许多成果,但似乎还没有成熟产品出现。

人脸图象自动识别系统主要由三个功能模块组成:人脸图象预处理模块,表示与特征抽取模块,识别模块。人脸图象识别方法,常用的有模块识别法、统计结构识别法、神经网络识别法等。最近,国外则着重于研究更通用的识别方法,如本征脸(eigenface)法,弹性匹配(elastic matching)法等。本文在介绍一种人脸图象识别系统的同时,将对这些识别方法进行阐述。

二、人脸图象预处理

对人脸图象进行预处理的目的是:使外界干扰对识别目标的影响减至最小,并且按照人脸图象识别方法的要求使图象达到标准化(标准尺寸及标准位置)。预处理包括消除噪声、灰度规一化、几何校正、滤波变换等^[1]。

三、表示与特征抽取

人脸的多样性以及识别要求,使得人脸的表示应具有多样性与唯一性,严格意义上,不存在两张完全相同的人脸,因此,将现实空间的图象反映到机器空间时也应保持这种唯一性与多样性,才能做到人脸图象的准确识别。对于人脸,有两种常用表示法^[10]:

1. 基于特征表示(feature-based representation)

考虑两种特征:正面特征和侧面特征。正面特征包括各种距离比例及角度值,如两眼间距离,两眼间距离,眼大小,以及各特征点之间组合所形成的距离,角度值等。侧面特征是由人脸图象侧向投影所形成的凹凸点,以作为特征点(fiducial-point);如鼻尖,下颌等。

基于特征的表示法,其优点是:

所选取的特征较稳定,受光(lighting),旋转(rotating),形变(deforming),尺寸(scaling)等因素影响

^{*} 本文工作得到中国科学院“九五”基础性研究项目 KJ952-J1-707 的资助。熊志勇 博士生,研究兴趣:数据压缩,模式识别,神经网络。刘翼光 博士生,研究兴趣:模式识别,函数式语言,计算理论。沈理 研究员,博士生导师,研究方向为软计算,模糊系统等。

小。缺点:特征提取较难(两难问题),如果提取两眼间距,就需先定出眼睛的位置,而这本身就是图象理解及识别要解决的问题。特征提取可分两步:

①定位各特征点,可用灰度统计法,神经网络法(ANN),文法结构法,数学变换法等,关于这些传统方法及其在人脸图象识别方面的具体应用,可参考相关文献。^[1]

②取特征值,以用于识别。

2. 基于图象表示(image based representation)

即将人脸图象当作矩阵进行处理,从而提取出矩阵的各种代数特征,这种表示法的优点是:不用提取特征。但缺点是:稳定性(robust)差,因为人脸图象随意性很大,受光、旋转、形变、方向、尺寸等影响大,从而要进行图象标准化处理,另外基于图象表示法,使图象内的目标的几何特征与图象数据相混淆,不能充分利用目标的几何特性。

四、人脸图象识别

模式识别可统分为类间(inter classes)识别与类内(in class)识别,人脸图象识别属于后一种^[2]。下面着重介绍几种比较重要的识别方法。

1. 本征法^[4]

较之其它方法,本征法更注重考虑在人脸图象中,什么信息对于识别最重要。在调研中发现,人脸的全局结构信息对于识别是最重要的,这与人类的人脸识别机制相符合。如果将人脸图象看作一个矩阵,则因为矩阵的本征值及相应的本征向量是矩阵的一种内在属性,具有旋转,尺度不变性,从而可作为一种信息以进行人脸图象识别^[5]。

假设人脸图象库中的图象共有 N 个,用向量表示为 I_1, I_2, \dots, I_N (假定向量维数为 $L * L$), 可得到其平均人脸图象为:

$$I_{ave} = \frac{1}{N} \sum_{i=1}^N I_i$$

由此可得到每张人脸图象相对于 I_{ave} 的均差为:

$$\Phi_i = I_i - I_{ave}$$

构造协方差矩阵:

$$C = \frac{1}{N} \sum_{i=1}^N \Phi_i \Phi_i^T$$

要求 C 的本征值 λ_k 与本征向量 μ_k , 可转化成求另一矩阵 S 的本征值与本征向量 v_k , 使得计算复杂性大大降低。其中矩阵 S 的元素为:

$$S_{ij} = \Phi_i^T \Phi_j, (i, j = 1, \dots, N)$$

求出矩阵 S 的 v_k 后, 则 C 的本征向量 μ_k 可由此求出:

$$\mu_k = \sum_{j=1}^N v_k \Phi_j, (k = 1, \dots, N-1)$$

这些求出的本征向量所形成的向量空间, 就可表示人脸图象的主要特征信息。对于图象库中所有 N 个图象都可向此空间投影, 得到投影向量为 $\Omega_1, \Omega_2, \dots, \Omega_N$ 。例如将库中任一人脸图象, 作投影, 得一向量为 $\Omega^T = [\omega_1 \omega_2 \dots \omega_N]$, 其中

$$\omega_t = \mu_t^T (I - I_{ave}), (t = 1, \dots, N)$$

则对于一待识人脸图象 (I_{new}), 得其投影向量为 P , 计算

$$e_t = \| P - \Omega_t \|^2, (t = 1, \dots, N)$$

取 e_t 最小值时所对应的 Ω_t 为所求人脸图象, 从而完成了识别^{[6][8]}。本征法原理简单, 易编程实现, 并且识别效果较好, 只是该法受光照影响大, 需要对此进行相应的预处理。

2. 弹性匹配法^[4]

此方法是模板匹配法的一种变形, 属于动态匹配法的一种^[6]。模块可有多种表示方法, 如前述的基于特征表示法, 以及基于图象表示法, 在弹性匹配法中^[11], 模板使用图(graph)来表示。对于通常的模板匹配法, 将模板 c 与待识对象 x 进行比较, 如下所示:

$$d(x, c) = \| x - c \|^2, (\| \cdot \| \text{为各种距离表示})$$

这里模板 c 中包含了各点的特征信息^[4]。可用栅格(grid)来表示模板, 对位于栅格点阵上的每一点 $p(x, y)$, 可得到其对应的特征信息, 如灰度值、傅里叶变换系数值、小波变换系数值等^{[3][6]}。如果将模板看作拓扑结构为栅格的图, 图的节点对应栅格中的格点(fiducial points), 而图的边对应栅格中的格点之间的连接关系, 图的拓扑结构可以为矩形点格阵, 亦可为三角形点格阵。如对图的节点, 计算其特征信息, 并将所得到的值赋予该节点, 这过程可称之为对节点进行标号(labeling), 这样形成的图称作标号图, 如将人脸图象库中的图象都用标号图表示, 则通过图与图之间的拓扑结构的匹配, 以及节点与节点间的特征信息值的比较, 可进行人脸图象的识别。这样就将模板的图象匹配转化成模板的图匹配^[7]。

如前述, 节点的特征信息可通过离散傅里叶变换系数值得到, 如下所示:

$$F(u, v) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(m, n) \exp \left\{ -2\pi j \left(\frac{mu}{M} + \frac{nv}{N} \right) \right\}$$

由式子中可以看出傅里叶变换是一种全局变换,反映图象的全局信息,因此受图象的旋转,尺寸,变换的影响小,故此具有一定的稳定性,但也正因为它是一种全局变换,不能突出图象的局部特征,故此对于图象的形变很敏感,从而对于形变较大的人脸图象识别效果不好。小波变换是对图象局部进行变换,受图象形变干扰小,故此识别效果较好。以下考虑用小波变换对人脸图象进行处理。

设点 $p(x, y)$ 为模板栅格点阵 $\Gamma(x, y)$ 上的任一点,则由小波变换可得点 $p(x, y)$ 的特征信息为:^[12]

$$\rho_j(\bar{x}) = \int \Gamma(\bar{x}) \Psi_j(\bar{x} - \bar{x}') d^2 \bar{x}', \quad (\text{这里 } \bar{x} = (x, y))$$

其中 $\Psi_j(\bar{x}) = \frac{k_j^2}{\sigma^2} \exp\left(-\frac{k_j^2 x^2}{2\sigma^2}\right) \left[\exp(i k_j \bar{x}) - \exp(-\frac{\sigma^2}{2}) \right]$

称作 Gabor 内核(kernels)系列。式中的 \bar{k}_j 称作波向量(wave-vector), 其形式为: $\bar{k}_j = \begin{pmatrix} k_x \\ k_y \end{pmatrix} =$

$$\begin{pmatrix} k_x \cos \varphi_j \\ k_x \sin \varphi_j \end{pmatrix}, \quad \text{其中 } k_x = 2^{\frac{v+3}{2}} \pi, \varphi_j = \mu \frac{\pi}{8}, \text{ 并且 } j = \mu + 8v, \mu = 0, \dots, 7; v = 0, \dots, 4, \text{ 另外取 } \sigma = 2\pi.$$

从而在点 $p(x, y)$ 处可计算得到 40 个 ρ_j 值, 作为其特征值。同样对于模板栅格点阵上的每一点都可算出其 Gabor 特征值。

这样由模板栅格点阵所得到的标号图的节点就有 Gabor 特征值。在进行图匹配时, 通过计算节点的相似程度, 取最相似的图, 作为匹配目标。公式如下:

$$S_i(\rho, \rho') = \sum_j a_j a'_j / \sqrt{\sum_j a_j^2 \sum_j a'^2}$$

这里 ρ, ρ' 为图的节点的 Gabor 特征值; a_j, a'_j 为 ρ, ρ' 的幅度值。

小波变换克服了傅里叶变换的缺陷, 因此弹性匹配法受图象的形变影响小, 并且也不受光, 面部表情, 图象尺寸的干扰, 该法不足之处在于识别速度慢, 因为匹配过程为一反复比较过程, 计算量很大, 从而影响了识别速度。

3. 神经网络(ANN)法^[10]

ANN 由于其固有的并行运算机制以及对目标的分布式全局存储, 故此可用于模式识别, 并且不受目标形变影响。常用的网络结构为:

(a) 线性自相联网^[11]。设人脸图象库中任一图象可用一 M 维列向量 x 表示, 图象矩阵的第 i 行对应 x 中的 x_i ; 又设图象库中的图象数为 N , 则这 N 张人脸图象可用 $M * N$ 维矩阵 $X = [a_{ij}]$ 表示, 矩阵元素 a_{ij} 表示第 j 张人脸图象的第 i 点象素值。以下叙述如

何将矩阵 X 用神经网络表示, 即将矩阵 X 存入神经网络。使用自相联网络, 网络拓扑结构很简单, 只有一层(不包含隐含层), 故此学习算法也很简单, 易于编程实现。网络神经元节点之间的连接, 可用权矩阵 $W = [w_{ij}]$ 表示, w_{ij} 为节点 i, j 之间的连接权重。网络训练算法采用 Hebbian 学习规则, 所用公式如下:

$$W = \sum_{n=1}^N x_n x_n^T, \quad (\text{其中 } x_n \text{ 表示第 } n \text{ 个图象向量})$$

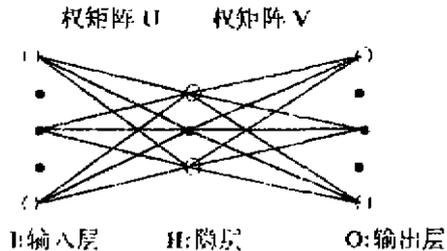
求出 W 后, 就将人脸图象矩阵 X 存入了神经网络。对一待识图象向量 x_{new} , 可作为网络的输入, 相应的输出向量为 $y_{new} = W x_{new}$ 。可通过计算此两向量间的相似程度, 来决定是否识别。所用公式如下:

$$\cos(x_{new}, y_{new}) = y_{new}^T x_{new} / \|x_{new}\| \cdot \|y_{new}\|$$

如果 $\cos(x_{new}, y_{new})$ 小于某一阈值, 则识别, 否则拒识。^[10]

使用 Widrow-Hoff 学习算法, 网络的存储性能可得到改善, 从而识别能力也得到提高。

(b) BP 网(非线性网)。网络拓扑结构如下图所示:^[12]



对网络进行训练时, 采用有教师训练算法: 即对每个人都提供几个不同姿态的人脸图象向量作为输入, 而正面姿态的图象向量作为标准输出。这里, 输入层用 M 维列向量 $x(x_1, \dots, x_M)$ 来表示输入, 隐层用 L 维列向量 $h(h_1, \dots, h_L)$ 表示, 而对于输出层则用 M 维列向量 $y(y_1, \dots, y_M)$ 表示。输入层与隐含层之间的联接用 $L * M$ 维权矩阵 U 表示, 而隐含层与输出层之间的联接用 $M * L$ 维权矩阵 V 表示。

每个隐层单元接受来自输入层的输入, 计算得到隐层输出, 以作为输出层的输入, 具体公式为: $h = f(Ux)$, 其中 $f(x)$ 为响应函数, 常用的为:

$$f(x) = \frac{1}{1 + e^{-x}}$$

同样每个输出层单元接受来自隐层的输入, 得到最终结果。所用公式为: $y = f(Vh)$ 。

进行网络训练时, 对每个人取不同姿态的人脸图象向量作为输入, 而正面姿态的人脸图象向量作为标准输出。训练集为 $(x_1, \dots, x_n | y_{norm})$, 对于输入 x_i

相应的输出为 y_t , 与标准输出的差别为: $e_t = (y_{\text{norm}} - y_t)$, BP 训练算法将 e_t 向后传播, 以对联结权矩阵 U, V 进行训练。网络训练是一动态重复过程, 所用公式为: [2][10]

隐层与输出层间的权为:

$$V^{t+1} = V^t + \Delta V^{t+1}$$

$$\Delta V^{t+1} = \eta \delta_{\text{out},j} h_i^T,$$

其中 $\delta_{\text{out},j} = f'(V h_j) e_j$

同样输入层与隐层之间的权为:

$$U^{t+1} = U^t + \Delta U^{t+1}$$

$$\Delta U^{t+1} = \eta \delta_{\text{hidden},i} x_i^T,$$

其中 $\delta_{\text{hidden},i} = f'(U x_i) (V^T \delta_{\text{out},j})$

上式中, η 为常数, 上标 $(t+1)$ 表示第 $t+1$ 次迭代运算, f' 表示响应函数 f 的微分。

网络训练完后, 即可用于人脸图象识别。输入一待识图象向量 x_{new} , 经网络运算, 得到输出 y_{new} ; 如果 y_{new} 与 x_{new} 之间的相似程度小于某一阈值, 则得到识别, 否则拒识。

(c) 其它网络结构。如多层感知器模型, Hopfield 模型, 及认知器模型等, 都可用于人脸图象识别。具体应用, 可参考相关文献。[2]

小结 神经网络识别方法因为其内在的并行运算机制, 从而人脸图象的匹配过程是并行进行的, 不似其它非联结结构识别方法, 待识人脸图象要与已存储的所有图象进行逐一匹配。因此采用 ANN 方法识别速度很快。对于人脸图象库较小的识别要求, 能得到很好的实用效果。国外已有这方面的实用系统, 应用于银行, 安检等部门。

4. 其它识别方法

如统计结构法、模板识别法等。最近, 因为形变匹配技术的引入, 使得形变模板识别法也在人脸图象识别中得到应用。具体内容可参看相关文献。[1][2][9]

总结 人脸图象识别系统由图象预处理, 表示与特征抽取, 识别等部门组成。这三者相互影响, 依据不同的识别方法, 前两部分所采用的算法与形式都会不同。在介绍的识别方法中, 弹性匹配 (elastic matching) 法识别率最高, 且适用性很强, 只是识别速度较慢, 今后的工作在于提高其匹配速度。神经网络 (ANN) 方法受环境影响很大, 尤其是受光 (lighting) 的影响; 其识别率也不是很好; 只是对于特定环境应用效果较好。本征脸 (eigenfaces) 识别法的识别率界于上述两者之间, 因其简单而得到较广泛的应

用。形变模板 (cdeformable template) 识别方法也将成为今后研究的热点。

参考文献

- [1] 沈清、汤霖, 模式识别导论, 国防科技大学出版社, 1991
- [2] 庄镇泉等, 神经网络与神经计算机, 科学出版社, 1994
- [3] 李世雄、刘家琪, 小波变换和反演数学基础, 地质出版社, 1994
- [4] Jun Zhang et al., Face Recognition: Eigenface, Elastic Matchong, and Neural Nets, Proc. of The IEEE, Vol. 85, No. 9, 1997
- [5] Matthew A. Turk and Alex P. Pentland, Face Recognition Using Eigenfaces, Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition, Maui, Hawaii, 1991
- [6] M. Lades et al., Distortion invariant object recognition in the dynamic link architecture, IEEE Trans. Comput., Vol. 42, No. 3, 1993
- [7] Christoph von der Malsburg, Pattern Recognition by Labeled Graph Matching, Neural Networks, Vol. 1, pp. 141-148, 1988
- [8] W. K. Konen et al., A Fast Dynamic Link Matching Algorithm for Invariant Pattern Recognition. Neural Networks, Vol. 7, pp. 1019-1039, 1994
- [9] M. A. Shackleton and W. J. Welsh, Classification of Facial Features for Recognition, Same to [5]
- [10] D. Valentin et al., Connectionist Models of Face Processing: a Survey, Pattern Recognition, Vol. 27, pp. 1209-1230, 1994
- [11] D. J. Burr, Elastic Matching of Line Drawings, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. Pami-3, No. 6, 1981
- [12] L. Wiskott et al., Face Recognition by Elastic Bunch Graph Matching, IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7), 1997