

# 基于单视图的多姿态人脸图像生成技术研究<sup>1</sup>

朱长仁 王润生

(国防科技大学电子科学与工程学院 ATR 国家重点实验室 长沙 410073)

**摘要** 该文提出了一种基于单视图或小样本的多姿态人脸图像生成技术。它首先利用一个特征点集表示人脸,然后基于二元高次多项式函数最小二乘方法对人脸各姿态之间的特征点集坐标变化进行拟合,形成全局的变形域,最后由单视图通过变形映射生成多姿态人脸图像。实验结果表明,利用单视图和生成的多姿态图像进行多姿态人脸识别,正确率得到大大提高,证明该文人脸图像生成技术十分有效。

**关键词** 人脸图像生成,多姿态,单视图,多项式函数,最小二乘拟合,人脸识别

**中图分类号** TP391.41

## 1 引言

人脸识别是计算机视觉、模式识别中一个重要内容,在电视监控、罪犯查询识别、安全门卫系统等领域中有着广泛的应用价值<sup>[1-3]</sup>。人脸识别按姿态可划分为前视人脸识别和多姿态人脸识别,其中前者的研究已经较为深入<sup>[1-3]</sup>,也取得很好的结果,而后者虽有一些文献<sup>[4-6]</sup>论述,但是其中仍然存在许多问题没有解决,这也是导致目前人脸识别技术没有得到足够推广的主要原因之一。

目前人脸识别方法对人脸姿态变化都较为敏感,这也是把它们推广到多姿态识别时正确率迅速下降的主要原因,尤其是基于灰度匹配的主分量分析(如特征脸)方法<sup>[5]</sup>,在前视人脸识别中因正确率较高而倍受关注,但由于该方法的实质是进行图像之间的灰度快速匹配,它随着姿态变化增大,多姿态图像之间相关性减小而识别率迅速下降。所以有人提出基于多视图的多姿态人脸识别算法,可以提高多姿态人脸识别正确率,并在一定程度上解决了多姿态人脸识别的问题。但是该方法存在最大的缺陷是:它需要以对每个人脸在多个视角点都进行拍摄采集形成多视图为前提条件。这条件不仅费事而且费时,最为关键的是,在实际许多场合中无法得到多视图。例如,在罪犯搜索工作中,已有的历史人脸库中很多只有正前视图像,而实际监视偷拍的待识别人脸图像姿态可能是随机的,这时只能基于单视图进行多姿态人脸识别。

这样基于单视图或小样本的多姿态人脸识别技术就很重要了。但是基于单视图或小样本进行多姿态人脸识别是一个极其困难的问题,它主要表现在:

(1) 人脸姿态的变化,导致人脸成像出现变形,同一人脸不同姿态的人脸图像之间相关性随着姿态变化的增加迅速减少,这给不同姿态的人脸图像匹配带来困难;亦即如何利用某一姿态的人脸图像对其它姿态的人脸进行识别,这是一个极其困难的问题。

(2) 在现实物理世界,人脸是三维目标,而现在只有某一个姿态的二维成像信息,这些二维信息不足以准确恢复三维目标信息。

(3) 人脸是一个非刚体,人脸的肌肉具有弹性,随着表情、姿态等变化,给识别增加了难度。

但是,基于单视图或小样本的多姿态人脸识别是许多场合中急需应用的技术,是人脸识别真正走向广泛应用之前必须彻底解决的问题。由此,基于单视图的多姿态人脸识别的研究就十分必要和迫切了,而基于单视图生成多姿态人脸图像技术就是其中的关键技术之一。目前关于这方面的研究起步不久,只有少量文献<sup>[7,8]</sup>出现,它们大致可分为两大类:一类是基于三维重建方法,例如文献<sup>[7]</sup>,这类方法虽然理论上是最优的,但是由于实际中许多信息的未知以及三维重建需要大量复杂的计算,其效果并不理想;另一类是基于二维图像的生成,例如文献<sup>[8]</sup>,这类方法简单可行,也不需要复杂的三维重建。但是文献<sup>[8]</sup>采用一种线性目标类方法,该方法

<sup>1</sup> 2001-09-18 收到, 2002-04-15 改回

由于采用了基于一组基的加权和实现的, 它的最大缺陷是: 生成的人脸图像忽略了几乎所有的脸部纹理细节信息, 而这些脸部纹理细节信息对一个人脸来说是很重要的, 它们对人脸识别也很有帮助。

为此本文研究并提出一种新的基于单视图的多姿态人脸生成技术。它利用一个特征点集的坐标变化表示人脸姿态的变化, 并利用二元高次多项式函数最小二乘方法进行拟合形成全局的变形域, 然后由此由单视图通过变形映射生成多姿态人脸图像。该技术既不需要复杂的三维重建, 又由于它是从原始图像上直接映射纹理的, 所以保留了所有纹理细节信息。

下面首先论述基于单视图的多姿态人脸图像生成原理和算法, 然后给出基于单视图的多姿态人脸图像生成实验结果, 文中最后给出了基于该单视图和生成的多姿态人脸图像一起形成训练样本库进行多姿态人脸识别的实验结果, 它表明识别正确率得到大大提高, 说明本文多姿态人脸图像生成技术十分有效。

## 2 基于单视图的多姿态人脸图像生成原理和算法

### 2.1 基于单视图的多姿态人脸图像生成原理

已知某一姿态的人脸图像, 如何产生其它姿态的人脸图像, 是本文要解决的问题。我们的方法是通过图像变形生成新姿态的人脸图像。由于一般情况下, 同一姿态下各个人脸图像中的器官分布大致相同, 所以我们可以对一个标准人脸在一组典型姿态进行拍摄, 获取多姿态人脸图像, 进而获取各器官的各种姿态的分布, 然后通过器官分布对输入人脸图像进行变形生成相应各个姿态的人脸图像。

由于高次多项式可以拟合几乎所有的变形变换, 变换越复杂, 需要拟合的次数就越高。同样, 人脸各个器官在姿态变化时的变形 (包括图像放缩、旋转、姿态变化引起的变形等) 也可以通过高次多项式函数拟合而成。而人脸器官可以用控制点表示, 对于二维图像中, 人脸器官可以通过控制点坐标  $(x, y)$  表示, 不同姿态之间的图像变形则可以通过二元高次多项式拟合得到。多项式的次数越高拟合效果就越好, 但随着次数增加, 计算所需的控制点数目也增加, 计算量急剧增加, 所以实际中根据需要确定次数, 并非一定次数很高。我们实验中采用了二元二次多项式拟合。

设在姿态 1 中的控制点集合为  $\Theta_1 = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$ , 需要变形到的姿态 2 的相应控制点集合为  $\Theta_2 = \{(x'_1, y'_1), (x'_2, y'_2), \dots, (x'_m, y'_m)\}$ , 各有  $m$  个控制点, 并且一一对应。二元二次多项式为

$$\begin{aligned} f_1(x, y) &= a_5y^2 + a_4xy + a_3x^2 + a_2y + a_1x + a_0 \\ f_2(x, y) &= b_5y^2 + b_4xy + b_3x^2 + b_2y + b_1x + b_0 \end{aligned}$$

两个多项式有  $(a_5, \dots, a_0, b_5, \dots, b_0)$  共 12 参数, 至少需要 6 个对应控制点才能确定这 12 参数。若控制点数目等于 6, 那么这方程组有唯一解, 但有可能拟合误差较大, 所以我们通过增加控制点数目 ( $m > 6$ ) 并用最小二乘拟合求解以减少误差。即

$$\begin{aligned} x' &= f_1(x, y) + \varepsilon_1 = a_5y^2 + a_4xy + a_3x^2 + a_2y + a_1x + a_0 + \varepsilon_1 \\ y' &= f_2(x, y) + \varepsilon_2 = b_5y^2 + b_4xy + b_3x^2 + b_2y + b_1x + b_0 + \varepsilon_2 \end{aligned}$$

其中  $\varepsilon_1$  和  $\varepsilon_2$  是拟合误差。由最小二乘拟合得

$$\begin{aligned} E\varepsilon_1^2 &= \sum_{i=1}^m (x'_i - f_1(x_i, y_i))^2 = \sum_{i=1}^m (a_5y_i^2 + a_4x_iy_i + a_3x_i^2 + a_2y_i + a_1x_i + a_0 - x'_i)^2 = \text{Min} \\ E\varepsilon_2^2 &= \sum_{i=1}^m (y'_i - f_2(x_i, y_i))^2 = \sum_{i=1}^m (b_5y_i^2 + b_4x_iy_i + b_3x_i^2 + b_2y_i + b_1x_i + b_0 - y'_i)^2 = \text{Min} \end{aligned}$$

$\frac{\partial(E\varepsilon_1^2)}{\partial a_i} = 0, \frac{\partial(E\varepsilon_2^2)}{\partial b_i} = 0$ , 由此得到由 12 个方程组成的方程组, 设

$$X = \begin{bmatrix} 1 & x_1 & y_1 & x_1^2 & x_1y_1 & y_1^2 \\ 1 & x_2 & y_2 & x_2^2 & x_2y_2 & y_2^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_m & y_m & x_m^2 & x_my_m & y_m^2 \end{bmatrix}, X' = \begin{bmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_m \end{bmatrix}, Y' = \begin{bmatrix} y'_1 \\ y'_2 \\ \vdots \\ y'_m \end{bmatrix}, A = \begin{bmatrix} a_0 \\ a_2 \\ \vdots \\ a_5 \end{bmatrix}, B = \begin{bmatrix} b_0 \\ b_2 \\ \vdots \\ b_5 \end{bmatrix}$$

解此方程组得到解为  $A = (X^T X)^{-1} X^T X'$ ,  $B = (X^T X)^{-1} X^T Y'$ . 得到系数后就可以得到全局的变形域, 进而通过该多项式进行全图的变形变换、纹理映射, 即由姿态 1 的人脸图像变形生成姿态 2 的人脸图像. 其它姿态的人脸图像生成算法与此相同.

### 2.2 基于单视图的多姿态人脸图像生成算法

该算法可以描述为

步骤 1 训练阶段: 确定一个典型人脸的各个姿态的对应特征控制点集, 它们组成一个集合  $\{\Theta_i | 0 < i < PN\}$ , 其中 PN 为姿态数目.

步骤 2 生成阶段, 可分以下步骤进行:

(1) 确定一个输入人脸 (即待生成多姿态图像的人脸) 某一姿态的单视图的特征控制点集, 设为  $\Theta_1$ ;

(2) 并对每个需生成图像的姿态: 取典型人脸相应姿态特征控制点集为  $\Theta_2$ , 通过上一节推导的方程求解  $A$  和  $B$ , 然后根据解  $A$  和  $B$  形成的变形域, 将该单视图通过变形变换、纹理映射生成新姿态的人脸图像.

而生成的多姿态人脸图像用于人脸识别时 (即识别阶段), 则是基于已有的单视图和生成多姿态人脸图像一起形成训练库进行训练, 识别时把待识别图像与整个训练库进行匹配识别.

对于人脸识别而言, 训练阶段和生成阶段是离线工作的, 识别阶段是在线工作的. 所以, 人脸的相应特征控制点集可以用全自动方式提取, 也可以用半自动方式提取.

## 3 实验结果

本文算法对从 AT&T ORL 等人脸库中选取 10 人、60 幅不同姿态的人脸图像数据进行了生成和测试实验.

### 3.1 多姿态人脸图像生成实验结果

本文实验以正前视图产生其它姿态的人脸图像为例, 图 1 列举几个典型图例. 图 1(a) 为正前视图即已拍摄的图像, 1(b) 是由 1(a) 生成的右前姿态 1 人脸图像, 1(c) 是由 1(a) 生成的左前姿态 2 人脸图像, 1(d) 拍摄的右前姿态 1 人脸图像 (与 1(b) 进行比较), 1(e) 拍摄的左前姿态 2 人脸图像 (与 1(c) 进行比较). 从图 1 中对由正前视图生成的多姿态人脸图像与相应姿态的拍摄图像进行比较分析后, 我们可以看出生成的多姿态人脸图像与相应姿态拍摄的人脸图像是很相似的, 较好地反映了人脸的相应姿态变化信息, 也保留了所有纹理细节信息.

图 2 给出了一个典型样本生成的多姿态人脸序列图. 图 2(a) 是原始图像, 2(b)~2(f) 是由 2(a) 样本生成的不同姿态人脸序列图.

### 3.2 多姿态人脸识别结果

多姿态人脸图像生成后, 是否反映姿态变化信息等, 上面是用主观评价. 其实, 多姿态人脸图像生成后主要是用于多姿态人脸识别, 它对提高识别正确率有多大帮助则是最客观的评价. 下面我们进行基于单视图的多姿态人脸识别实验予以验证.

经典算法采用文献 [5] 方法, 但是它只能基于单视图作为训练样本库进行识别. 而本文算法首先通过本文方法由单视图通过变形生成多姿态人脸图像, 然后由单视图和生成的多姿态人脸图像一起组成训练样本库进行训练, 最后基于整个训练样本库利用文献 [5] 方法进行人脸识别.

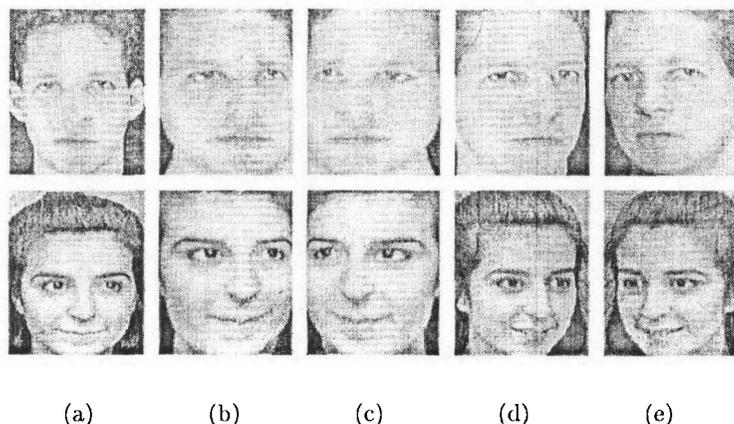


图 1 几个典型人脸的多姿态人脸生成图像与相应姿态下拍摄的真实图像的比较

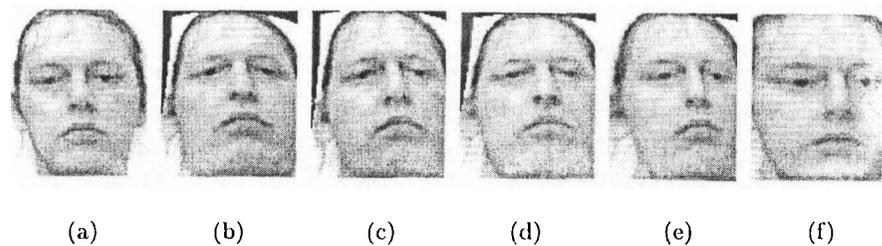


图 2 典型样本生成的多姿态人脸序列图

由于文献 [5] 方法是基于特征脸的算法实现的, 它通过计算各人脸图像在特征脸上的投影矢量之间的距离并用最近邻原则进行分类识别, 所以可以在提供人脸识别结果的同时, 也可以提供基于矢量之间匹配距离的置信度。

设训练库中有  $M$  个训练样本, 训练库第  $i$  个训练样本  $C_i$  的投影权值矢量为  $W_i$  ( $0 \leq i < M$ )。待识别人脸图像投影权值矢量为  $W^0$ 。这样待识别人脸图像与训练库第  $i$  个训练样本  $C_i$  的投影权值矢量之间的欧氏距离  $d_i = \|W^0 - W_i\|$ , ( $0 \leq i < M$ )。该欧氏距离越小, 两图像越相关。由此可以定义待识别人脸图像是  $C_i$  的基于欧氏距离的人脸置信度为  $h_i = (A - d_i)/A$ , ( $0 \leq i < M$ )。其中  $A$  为一个很大的常数, 满足  $A \geq \text{Max}_{0 \leq i < M}(d_i)$ 。置信度最大值对应的为识别结果。

人脸识别实验中, 以 10 个人为对象, 分 3 个姿态: 左前、正前、右前, 每个人每个姿态取 2 幅图像, 共 60 幅图像。其中每个人取一个正前视图作为训练样本 (共 10 幅), 其它均作为测试样本 (共 50 幅)。本文算法由这 10 个前视图产生其它两个姿态的人脸图像共 20 幅, 即形成 30 幅的训练样本库, 然后进行识别。而经典算法由于无法得到其它两个姿态的人脸图像则只能以 10 幅前视图为训练样本库进行识别。

**3.2.1 本文算法演示图** 图 3 演示了本文算法与经典算法的典型识别结果比较。图 3(a) 为输入待识别的人脸图像, 3(b) 为归一化后的人脸图像, 3(c) 为经典算法识别结果, 由于姿态变化导致匹配识别, 识别结果是错的, 3(d) 是本文算法的最终结果, 由于训练库中有本文生成的相应姿态人脸图像, 所以结果是正确的。其中每个识别结果中, 第一行为识别对应的人脸图像, 第二行为对应的人名字, 第三行为输入人脸为对应人的置信度。

**3.2.2 整个库人脸识别结果** 整个库识别实验结果见表 1。表中数据均是在 PIII/800、128MB 内存微机上得到的。

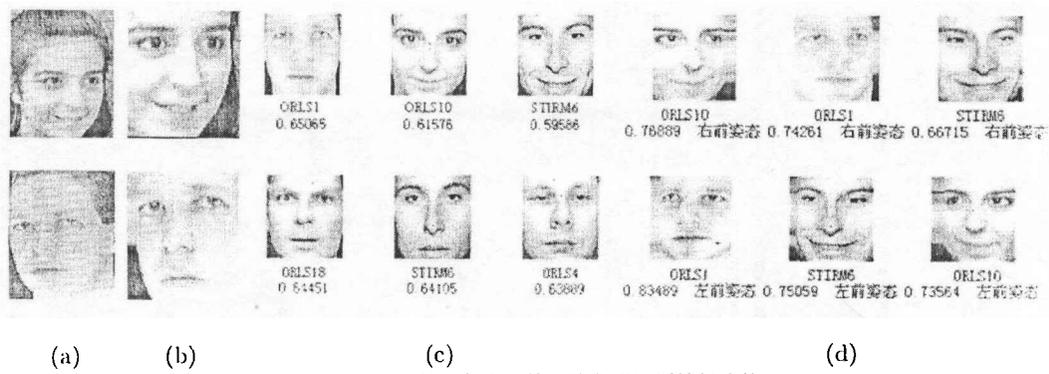


图 3 经典算法与本文算法的典型识别结果比较

表 1 人脸识别实验结果对比表

方法	经典算法 <sup>[5]</sup>	本文算法
正确率	52%	88%
平均识别一幅人脸图像耗时 (s)	0.52	0.63

表 1 数据表明基于单视图和由此生成的多姿态人脸图像一起形成训练样本库进行人脸识别, 其正确率比仅仅基于单视图进行识别的文献 [5] 经典方法有了很大地提高, 说明用本文方法生成的多姿态人脸图像对人脸识别很有帮助。

## 4 结 论

基于单视图或小样本的多姿态人脸识别技术克服了基于多视图解决多姿态人脸识别算法的最大缺陷: 需要以对每个人脸的多个视角进行拍摄采集形成多视图为前提条件。由单视图或小样本生成多姿态人脸图像是其中一个关键技术。本文提出一种基于高次多项式函数拟合方法的多姿态人脸生成技术, 它首先通过一个控制特征点集表示人脸, 并利用二元高次多项式函数最小二乘方法对人脸各姿态之间的特征点集坐标变化进行拟合, 形成全局的变形域, 最后由单视图通过变形映射生成多姿态人脸图像。该技术既不需要复杂的三维重建, 又由于它是从原始图像上直接映射纹理的, 保留了所有纹理细节信息。

另外, 在基于单视图和生成的多姿态图像进行多姿态人脸识别的实验中, 结果表明识别正确率比经典算法有了很大提高, 证明本文的多姿态人脸图像生成技术十分有效。

**致谢** 我们谨在此对为本次实验提供图像样本的 AT&T Cambridge 实验室等表示感谢! 感谢陆立明高级工程师的有益讨论!

## 参 考 文 献

- [1] R. Chellappa, C. L. Wilson, S. Sirohey, Human and machine recognition of faces: A survey, Proc. IEEE, 1995, 83(5), 705-740.
- [2] A. Samal, P. A. Iyengar, Automatic recognition and analysis of human faces and facial expressions: A survey, Pattern Recognition, 1992, 25(1), 65-77.
- [3] D. Valentin, H. Abdi, A. J. O'Toole, G. W. Cottrel, Connectionist model of face processing: A survey, Pattern Recognition, 1994, 27(9), 1209-1230.
- [4] R. Brunelli, T. Poggio, Face recognition: Feature versus templates, IEEE Trans. on PAMI, 1993, PAMI-15(10), 1042-1052.
- [5] M. Turk, A. Pentland, Eigenfaces for recognition, Journal of Cognitive Neuroscience, 1991, 3(1), 71-86.

- [6] A. Pentland, B. Moghaddam, T. Starner, View-based and modular eigenspaces for face recognition, IEEE Conf. CVPR'94, Seattle, 1994, 6, 84-91.
- [7] R. Lengagne, J. P. Tarel, O. Monga, From 2D images to 3D face geometry, IC Automatic Face and Gesture Recognition'96, Killington, Vermont, 1996, 10, 301-306.
- [8] T. Vetter, T. Poggio, Linear object classes and image synthesis from a single example image, IEEE Trans. on PAMI., 1997, PAMI-19(7), 733-741.

## RESEARCH ON MULTI-POSE FACE IMAGE SYNTHESIS FROM A SINGLE VIEW

Zhu Changren      Wang Runsheng

*(ATR National Key Lab., National University of Defense Technology, Changsha 410073, China)*

**Abstract** An algorithm that can synthesize multi-pose face images from a single view or small samples is proposed in this paper. First, the human face is represented with a dominant point set. Then, the variance of the dominant point set between different poses is fitted based on a least square fitting with a polynomial function and a global morphing field is formed. Finally, multi-pose face images are synthesized by image warping from a single view based on the global morphing field. An experiment of multi-pose face recognition is done based on the single view and synthesized multi-pose images. The results show that its performance is by far superior to those of the traditional methods. Thus, they prove the effectiveness of the algorithm in this paper.

**Key words** Face image synthesis, Multi-pose, Single view, Polynomial function, Least square fit, Face recognition

朱长仁: 男, 1968年生, 副教授、博士, 主要研究领域为图像分析理解、模式识别。

王润生: 男, 1941年生, 教授、博士生导师, 主要研究领域为图像分析理解、模式识别、信息融合等。