

一种基于多视点图象的可变姿态人脸识别系统*

游素亚 张永越 李武军 徐光祐

(清华大学计算机系 北京 100084)

摘要 本文报告了一种多姿态人脸图象识别原型系统,它不同于现有系统和方法,该系统可工作在合作对象下允许姿态变化(存在图象平面内旋转和深度方向上旋转,限于双眼可见)的人脸图象识别。由于对成像条件有所放松,故可望应用于身份验证、保安和视频会议等领域。对姿态可变条件下的人脸特征检测、姿态估计、识别建模以及基于模板相关的匹配等技术进行了深入研究,分析了光照、姿态及分辨率变化等因素对识别的影响程度。实验结果表明,对于30类人脸,每人18幅图象大小的测试集,达到了100%的识别率。

关键词 人脸识别,模板匹配,归一化,人脸姿态

1 引言

人脸对于人类识别身份、传递情感有很重要的作用。因此,研究利用计算机进行人脸的自动分析与识别技术一直是计算机视觉和模式识别领域的重要课题,它是实现高性能的智能接口、虚拟现实(VR)、基于模型的人脸图象压缩以及计算机支持的协同工作(CSCW)系统中不可缺少的重要部分。同时,人脸的自动识别分析在保安、身份验证、犯罪防范等领域中也有着广泛的应用前景。深入研究人类这一最重要特征的识别分析方法,不但对人脸识别本身具有重要意义,而且对有关人脸结构特征、生理、心理感知机理等基础课题的研究也具有重要意义。

但是,人脸的分析识别是一项富有挑战性的课题。因为,人脸组成了一类非常相似的物体,所有的人脸几乎由同样的几何特征所构成,人脸的分析识别需要利用脸部或特征结构的细微差别,这是极其困难的。而且,由于光照、人脸姿态和表情的改变都将使脸部三维特征的二维投影有很大变化,这更增加了人脸处理的复杂性。因此现有大多数的人脸

识别方法和系统都对工作条件作了严格的控制[1—3],使得这些系统难以在实际中广泛应用,尤其是对人脸姿态的限制更是其中的关键。例如,在身份验证和视频会议等领域,对光照、表情条件可作合理的控制,但严格限制人脸的姿态是不现实的。因此,为了达到实际应用的程度,这些条件必须有所放松,即要把工作条件的约束限制在合理范围内,以促进人脸分析识别技术逐步适应实际应用的条件。

本文给出了一种多姿态人脸图象识别原形系统,该系统可工作在合作对象下,存在图象平面内旋转和深度方向上旋转的人脸图象识别。实验结果表明,对于30类人脸,每人18幅图象大小的测试集,达到了100%的识别率。

2 人脸识别建模

正如一般三维物体识别方法一样,任意姿态下的人脸识别需要精确的三维人脸模型。三维人脸模型已经在基于模型的人脸图象压缩编码中得到了应用,但是用于压缩的三维人脸建模方法并不适合人脸识别,现有的三维建模方法还不能满足人脸识别的要求。从人脸识别的生物机理来看,人类在识别

* 国家863计划资助项目

收稿日期:1995.12.21

人脸时仿佛并不要求建立精确的三维模型,而只需从单一视点,甚至部分被遮掩的二维图象就可完成识别。因此,研究直接利用视点图象建立二维识别模型可能是目前最有效的方法。为此,我们提出了基于视点图象模板的识别建模方法,即将可视半球量化成不同视点,用可视半球的多视点图象来表示多姿态人脸。这样作的好处是:首先,从可视半球得到的多视点图象可以表达人脸的不同姿态,避免了复杂的三维建模,虽然这将减少姿态变化范围,但实验表明,在可视半球范围的姿态变化都能得到较准确的识别,可满足实际应用要求;其次,人脸建模的姿态不必十分准确,它只要在所要求的视角变化范围即可,在识别时,输入图象和模型图象的姿态差别可以用几何校准方法处理;最后,在人脸识别时,面部的各个特征对识别的影响并不是均等的,也就是说,面部特征对识别存在很大的冗余信息,提取出最有用的特征模型来表征人脸,在识别精度和效率方面都有极大的益处。

为了建立有效、经济的多视点识别模型,我们实验测试了视点的量化间隔大小及脸部主要特征(眼、鼻、嘴)对识别的精确影响。

首先,将可视半球量化成上/下 $\pm 20^\circ$,间隔 20° ;左/右 $\pm 30^\circ$,间隔 15° ,共形成 15 视点(图 1),对每个视点图象抽取 4 个面部特征模板,分别为全脸、眼睛、鼻子、嘴巴,以此模板作为识别建模模板,分别与输入图象相应的区域进行模板匹配。实验结果表明,当视点左/右量化间隔最大为 30° ,上/下间隔为 20° 时,识别性能并未下降很多。这样我们就可将原来的 15 个视点减少为 9 个视点,即左/右 $\pm 30^\circ$,间隔 30° ;上/下 $\pm 20^\circ$ 。经过多次实验,采用上述两种间隔建模的识别结果几乎不变,文献 [4] 给出了详细的实验结果。

在人脸识别中,面部各部分特征对识别的贡献是不同的,文献 [2] 认为,其重要性是由上而下(即眼部、鼻部、嘴部等)递减的。但我们的统计实验表明,以相关为测度的模板匹配中,鼻部特征对识别的影响最大,往往起着决定性的作用,其次是眼部、全脸,最不可信的是嘴部。据此,我们在识别建模时,对这 4 种模板的重要性分别用不同的权值来表示,其中鼻部的权值最大,其次为眼部、全脸和嘴部。此权值最终将用于加权各个模板对匹配的贡献。

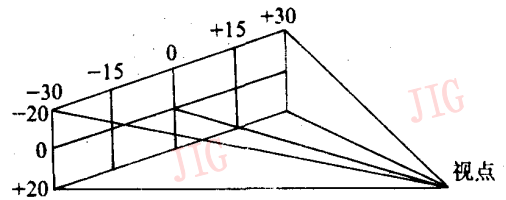


图 1 多视点人脸识别建模

Fig. 1 Face modeling based on multi-view

3 脸部特征检测

我们的特征检测算法是基于多视点模板匹配的。即根据输入姿态的可能范围和模板匹配对姿态变化的敏感程度,将可视半球量化成一系列视点,从而利用多视点图象的特征模板进行特征检测。为了提高检测精度和运算效率,整个算法是一种由粗到细引导的结构,在不同分辨率的情况下检测出不同尺度的特征,逐步细化搜索空间(姿态、特征位置),最终得到检测特征的准确定位。详细内容参见文献 [4]。

4 多姿态人脸识别

整个识别算法是基于模板匹配的识别方法,数据库中每个人的每个姿态同时存储全脸、双眼、鼻子和嘴部 4 个区域作为识别模板。考虑到姿态的可变性,必须在相关匹配前对被识别的人脸图象进行归一化处理,在我们的方法中,几何归一化是通过仿射变换和相似变换技术来实现的。除了实现尺度和旋转等不变性外,对输入图象还进行了灰度变换处理,以克服光照的影响。

4.1 几何归一化

几何归一化是通过对被识别人脸特征点的位置进行几何变换的。这里有两种情况,其一是只检测到了两眼的位置,其二是除两眼处,还得到了一个嘴角或鼻角的位置,对它们分别采用不同的归一化方法。

对于第一种情况,由于仅能得知两眼连线的水平角度和尺度,所以归一化是通过旋转变换和尺度变换来完成的。首先,计算出目标图象两眼连线角度与数据库中标准图象两眼连线角度之差,将整幅图象按此角度进行旋转变换:

$$\begin{cases} x' = x \cos \theta - y \sin \theta \\ y' = x \sin \theta + y \cos \theta \end{cases}$$

以实现角度的归一化。然后,计算出目标图象与标准图象两眼距离之比,以此为比例再进行尺度变换, $x' = ax; y' = ay$ 。几何归一化的结果首先能够保证双眼模板准确定位,但其它模板区域的准确性则要差一些,误差主要是因为图象旋转的中心点并不与全脸模板的中心点相重合,以及双眼的尺度并不能完全

决定整幅图象的尺度造成的,尤其是嘴部,由于它离双眼最远,所以误差对它造成的影响也最大。为了减少这些影响,我们在相关匹配时,把鼻子和嘴的模板区域位置,根据旋转角度作了调整,起到了一定的补偿效果(图 2)。

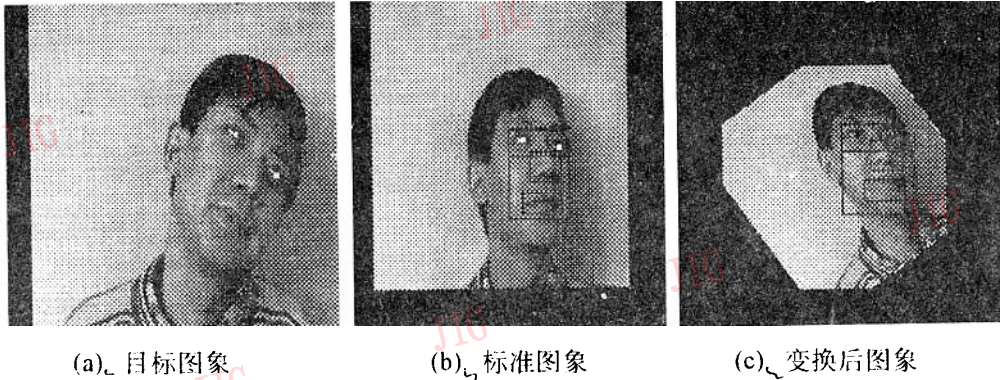


图 2 以两眼为依据进行几何归一化目标图象的结果,(a),(b)两图中所标为两眼位置,(c)图中四个黑框分别为标准图象四个模板的区域位置,其中鼻子和嘴的位置已经过调整。图象大小为 256×256 。

Fig. 2 Geometrical normalization based on the eyes. the two white points in (a) and (b) are the eyes' central points. The four regions bounded in (c) represent the four corresponding templates in the normal image in which the positions of the eye and the nose have been adjusted. The image size is 256×256 .

(a) testing image (b) normal image (c) transformed image

对于第二种情况(在实验中我们采用双眼及左嘴角的位置),可通过仿射变换将被识别人脸的目标图象按照标准图象进行几何归一化。设目标图象 3 个特征点的位置分别为 $(x_1, y_1), (x_2, y_2), (x_3, y_3)$, 标准图象相应 3 个特征点的位置为 $(x'_1, y'_1), (x'_2, y'_2), (x'_3, y'_3)$, 则所求仿射变换为:

$$\begin{cases} x' = a_{11}x + a_{12}y + a_{13} \\ y' = a_{21}x + a_{22}y + a_{23} \end{cases}, \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \neq 0;$$

图 3 为采用三个特征点进行几何归一化的结果。(a)是测试图象,很明显,头部姿态向左倾斜的很厉害,其上的 3 个白点是已经检测出的 3 个特征点。

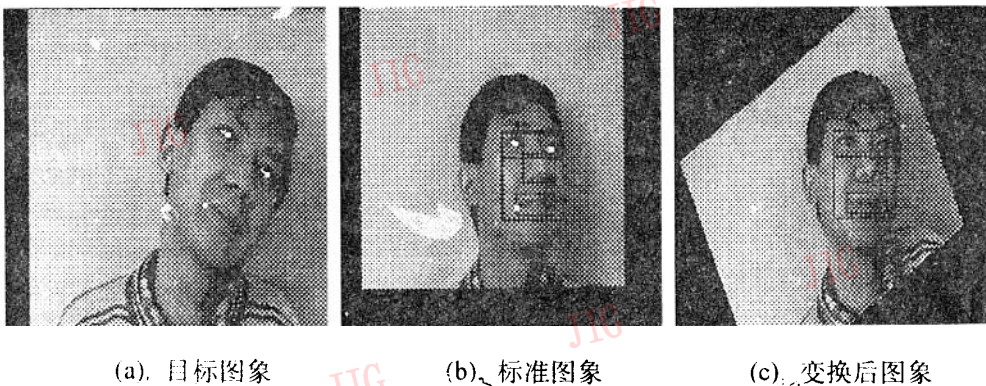


图 3 以两眼和左嘴角为依据进行几何归一化目标图象的结果。图象大小为 256×256

Fig. 3 Geometrical normalization based on two eyes' central point and the left mouth corner

(a) testing image (b) normal image (c) transformed image

现在需要把测试图象(a)校准到模板图象(b),在(b)中同样也有 3 个对应的特征点,被框出的 4 块区

域是存储的 4 个特征模板。(c)是(a)经过变换后的图象,图中 4 块区域的大小和位置与(b)中的 4 块区

域是完全相同的,下一步的匹配识别就将在这4块区域对应的两块间进行,例如,(b)中的眼睛区域与(c)中的眼睛区域。可以看到对应关系基本是吻合的。

4.2 光照归一化

一般来说,由光照产生的图象变化可看成是低频成分,对图象灰度进行适当的滤波变换就可以有效地去除光照影响。在此,我们考查了若干变换对模板相关的性能影响。包括高斯低通滤波,Sobel变换,LoG换以及基于平滑函数一次微分小波的小波变换。实验表明,在白天自然光照条件下,如果模板和实验图象同时获得,则直接采用图象灰度值进行匹配可以获得最好的效果。但如果光照不同,则先对图象进行一次微分变换,效果是最好的。表1,表2分别给出了人脸数据库中前10人的测试图象在利用灰度和梯度变换的识别结果。在这10个人中,只有第一和第二个人的测试图象与模板图象是在同一光照下拍摄的,其余人的光照条件都是不同的。

表1 利用灰度的识别性能 (识别率:60%)

Table 1 Recognition performance using gray value (recognition rate:60%)

测试人	相关值最大的结果	相关值	相关值次大的结果	相关值	正确与否
1	1	0.890	1	0.843	是
2	2	0.784	2	0.754	是
3	3	0.609	7	0.565	是
4	4	0.571	7	0.565	是
5	5	0.633	7	0.620	是
6	2	0.743	2	0.681	否
7	10	0.605	7	0.578	否
8	3	0.662	7	0.622	否
9	2	0.814	2	0.808	否
10	10	0.539	7	0.533	是

表2 利用梯度的识别性能 (识别率:100%)

Table 2 Recognition performance using gradient value (recognition rate:100%)

测试人	相关值最大的结果	相关值	相关值次大的结果	相关值	正确与否
1	1	0.641	1	0.598	是
2	2	0.543	2	0.459	是
3	3	0.682	3	0.652	是
4	4	0.703	4	0.632	是
5	5	0.604	8	0.509	是
6	6	0.622	7	0.618	是
7	7	0.632	7	0.578	是
8	8	0.668	8	0.572	是
9	9	0.728	9	0.713	是
10	10	0.583	9	0.578	是

4.3 模板匹配

在此,我们实现了两种模板匹配方法,一是基于

归一化相关测度准则,一是象素绝对差准则。归一化相关系数定义为

$$\rho_{xy} = \frac{E(XY) - E(X)E(Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

其中, E 为图象的平均值, D 为方差。

对于象素绝对差准则为

$$\delta_{xy} = \frac{\sum |H(X_i) - H(Y_i)|}{256 * N}$$

其中, $H(X)$ 为图象灰度值或梯度值, N 为图象象素点的个数。

在识别时,对于每个人的每个姿态,取4个模板的相关值之和作为它的相关结果,最后比较所有图象的相关值,最大者即为识别结果。

4.4 测试结果

我们的原形系统现有30类人脸,每个人存储视角左/右-30°-30°、上-下-20°-20°的15个姿态为建模图象,每幅模型图象含有全脸、又眼、鼻子和嘴共4幅模板。另有每人3幅测试图象,其中只有前两个人的测试图象与它们的模型图象是在同一光照条件下拍摄的,其余的测试图象均在不同光照下拍摄。在此条件下,我们实验了建模姿态数目对识别结果的影响、利用灰度相关与利用梯度的性能差异以及不同模板在识别中所起到的作用等。

虽然每人的15个姿态范围比较大,但由于我们在姿态检测的步骤中已检测出了人脸的大概姿态,所以在匹配时只需要在一定的邻域范围内进行。实验中我们取包括所检测出的姿态在内,周边左/右-15°-15°、上/下-20°-20°最多共9个姿态进行匹配。

表1,2分别给出的使用灰度和梯度(Sobel)进行匹配识别的性能分析(模板选用双眼模板和鼻子模板)的结果表明,在整体上若测试图象与建模图象是在同一光照条件下所拍摄,则利用灰度与梯度的识别性能相当;但如果模板和测试图象是在不同时刻采集的,无论在何种光照条件下,一次微分变换的性能明显优于直接的灰度相关。较好地克服了光照所带来的影响。因此,在模板相关计算前,对图象进行适当的变换预处理是非常必要的。由于对图象进行带通滤波相当于边缘增强过程,而人脸识别检测正需要利用脸部特征的细微变化,因此在边缘增强图象上进行模板检测可充分利用这些有用的信息。

需要指出的是,实验中发现,基于一次微分的匹配性能优于二次微分(如 LoG 算子)。这可能有两种解释,其一是由于二次微分的边缘增强能力太强导致噪声等无用信息过分增强所致。从人脸处理的生理实验看,在人脸图象处理中,小孩多采用抽取明显特征的方法,很容易被一些无用信息所迷惑,而成人则多采用结构和细微变化特征的策略。因此,图象特征

的无限增强未必对识别有利,关键是抽取有用信息。其二可能是由于相关算式本身所引起的,图象的二次微分结果是零交叉图。对这些正负分布的象素进行代数运算可能模糊了象素间的相关性。

此外,我们还测试了面部各部分特征对人脸识别的影响,表3是10个不同人脸测试图象的实验结果。

表3 不同模板的相关性能比较

Table 3 Recognition performance of different feature templates

测试人	全脸模板		双眼模板		鼻子模板		嘴模板	
	相关结果	相关值	相关结果	相关值	相关结果	相关值	相关结果	相关值
1	1	0.615	1	0.653	1	0.727	1	0.727
2	2	0.538	2	0.598	2	0.530	2	0.601
3	10	0.588	7	0.573	3	0.670	4	0.610
4	4	0.558	4	0.620	4	0.636	4	0.645
5	5	0.672	5	0.576	5	0.699	5	0.818
6	6	0.655	6	0.618	6	0.807	6	0.661
7	7	0.587	7	0.554	7	0.691	7	0.805
8	8	0.600	8	0.596	8	0.681	4	0.767
9	9	0.729	9	0.594	9	0.795	2	0.700
10	10	0.621	10	0.613	10	0.729	10	0.710
正确率(%)	90		90		100		70	

从表3可以看出,在4个模板中,鼻子的性能是最好的,其次是全脸、双眼、嘴部。

别分析也有意义。在此工作的基础上,进一步的研究将涉及到表情分析、系统完善等工作。

5 结束语

本文报告了我们研制的一种多姿态人脸图象识别原型系统,该系统可工作在合作对象下,允许姿态变化(存在图象平面内旋转和深度方向上旋转,限于双眼可见)的人脸图象识别。由于对成像条件有所放松,故可望应用于身份验证、保安和视频会议等领域。在研究中,我们对姿态可变条件下的人脸特征检测、姿态估计、识别建模以及基于模板相关的匹配等技术进行了深入研究,分析了光照、姿态及分辨率变化等因素对识别的影响程度。这些研究不仅对人脸图象的识别处理有益,而且对其它可形变景物的识

参考文献

- [1] A. Samal, P. A. Iyenger. Automatic Recognition and Analysis of Human Faces and Facial Expression: A Survey. *Pattern Recognition*, 1992, 25(1): 65-77.
- [2] R. Brunelli, T. Poggio. Face recognition: Features Versus Templates. *IEEE Trans. on PAMI*, 1993, 15(10): 1042-1052.
- [3] D. Valentin, et al. Connectionist Model of Face Processing: A Survey. *Pattern Recognition*, 1994, 27(9): 1209-1230.
- [4] 游素亚,徐光祐,张永越. 基于模板匹配的多姿态人脸图象特征检测. 第二届中国计算机智能接口与智能应用学术会议论文集, 北京清华大学出版社, 1995, 237-242.

游素亚, 博士后, 清华大学计算机系。

A View-based Automatic Face Recognition System

You Suya, Zhang Yongyue, Li Wujun and Xu Guangyou

(Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China)

Abstract In this paper, an automatic human face recognition system is described. Unlike most of the existing systems that can only operate on the frontal view of face or the facial profile, our system can work under varying pose with less rigid conditions. Basically, the face recognition system is based on multi-view template matching technique. In order to tackle the problems of absence invariant to illumination, translation, rotation and scaling of the faces, we conduct a preprocess stage of geometrical and illumination normalization, by which the confounding effects are first isolated, and then the faces are normalized prior to the independent classification. The performance of our system to recognize faces with varying postures has been demonstrated by operating on a database of 30 individuals, 18 images for each person, with 100% recognition ratio.

Keywords Face recognition, Template matching, Normalization, Facial pose

科技动态

INERGRAPH 快讯

Intergraph 公司向美国百老汇进军—StudioZ Workstation

据 Data Translation Inc. 报道, Intergraph 公司将向美国百老汇提供桌面式的、具有视频图象捕捉和压缩功能以及拥有 Windows NT 操作系统的 StudioZ 工作站。StudioZ 工作站的性能价格比远远优于拥有 RISC/UNIX 的其他工作站。它是新一代计算机数字动画编辑制作的工作站。

计算机动画编辑制作按磁带输出质量划分不同档次:

	磁带格式	输出
专业	D1 (Serial Digital Interface—SDI)	Component 数字
	BetaCam SP	Component 模拟
专业, 民用	Hi8	S—Video (Y/C)
	S—VHS	S—Video (Y/C)
民用	8mm	混合视频
	VHS	混合视频

捉/编码系统。StudioZ 工作站具有 JPEG 硬件压缩解压处理器、SDI 输出、文理内存和处理器、几何处理器、彩色渲染处理器和多 CPU 配置。百老汇所引进的 Intergraph 公司的高档的数字视频兼容性的 StudioZ 工作站, 具有视频专业水准, 能够随意进行捕捉, 编辑, 集成和回放。

StudioZ 开发的基础是 Intergraph TDZ 系列工作站。高质量的磁带制作输出, 结合强大的图形处理功能和丰富的应用软件, 使得 StudioZ 成为集计算机动画的创作、生成、编辑制作于一身的多功能工作站, 从而改变了传统的计算机动画制作方法, 大大提高了计算机动画制作的效率和质量。

StudioZ 系列工作站在相同磁带输出质量的前提下, 在桌边机、桌面机、单 CPU、双 CPU、四 CPU 和多种不同档次图形处理器的配置, 可供不同需求的用户选择。

能够在 StudioZ 工作站上运行的计算机动画制作软件除 Intergraph 的 ModelView 外, 还有 Microsoft Softimage 和 Autodesk 3D Studio Max 等数十种。其中有很多都和 ModelView 一样可以支持多线程处理, 因而可以充分利用双 CPU 和四 CPU 的 StudioZ 工作站的强大计算能力。

Intergraph 公司 Digital Media 部市场经理 Michael Bate 评论道, 我们之所以选择与百老汇进行有关产品集成, 是因为我们的产品无论在质量利功能方面, 还是在价格方面都是与其要求相符的, 通过这次合作, 我们发现在百老汇能够简单和快速地创造出高质量的视频效果的产品, 才是最有市场的产品。Intergraph StudioZ 工作站正在征服百老汇。

Intergraph 公司 TD 工作站的第一次演示是于 1996 年 4 月 14 日至 18 日在美国拉斯维加斯举行的 National Association of Broadcasters (美国国家广播协会工业贸易展览会) 进行的。Intergraph 公司 StudioZ 工作站是公司在数字传输领域中的最新产品, 它的生产目的是直接为新闻传播媒介和广播, 娱乐市场提供高性能并具有 Windows NT 操作系统的桌面式工作站。

适应于百老汇要求的数字传输要具有高性能的视频捕