基于高斯加权和因子分析字典学习的人脸姿态估计

廖海斌[1]，邓树文[1]*，王电化[1]，范　平[1]，陈友斌[2]
(1. 湖北科技学院计算机科学与技术学院，湖北 咸宁 437100；2. 华中科技大学自动化学院，武汉 430074)

摘要：针对传统的非约束环境下人脸姿态估计方法无法在统一框架下很好地处理各种姿态相关和姿态无关因子等问题，设计了基于字典学习和稀疏表示的鲁棒性人脸姿态估计框架，提出一种新的基于鼻尖点高斯加权的人脸预处理方法. 此外，为了提高字典的鉴别性，提出一种基于姿态相关和姿态不相关因子分析的鉴别字典学习算法. 通过在公开的 XJTU、Multi-PIE、CAS-PEAL-R1 和 AFLW 人脸库实验，结果表明：该方法在具有光照、噪声和遮挡变化的人脸库上识别率均约达 95%，基本可满足实际应用的要求.

关键词：人脸姿态估计；字典学习；稀疏表示；因子分析

中图分类号：TP 391.41　　**文献标志码**：A　　**文章编号**：1007－824X(2018)04－0047－05

DOI：10.19411/j.1007－824x.2018.04.011

人类可以轻松地分辨出任何人的头部姿态，但对计算机而言并非易事. 人脸姿态识别是根据图像确定人脸在三维空间中姿态参数的过程，在智能视频监控、人脸识别、人机交互和虚拟现实领域应用前景广泛[1]. 现有的人脸姿态识别主要有基于 2D 人脸几何模型[2-3]、纹理子空间学习[4]和 3D[5]等方法. 此外，出现一些非主流方法[6-7]，只能解决人脸姿态识别中部分问题或仅应用于某些特定场合. 由于光照、噪声、遮挡、分辨率、身份和表情等因素的变化都会对姿态识别的准确性产生影响，故如何消除这些因素的影响是目前亟待解决的难题. Zhang 等[8]提出基于字典学习与稀疏表示的人脸姿态识别框架，该方法对人脸光照、噪声和分辨率变化具有鲁棒性，在 Multi-PIE 人脸库的识别率达 98.12%，但未能解决人脸遮挡和表情变化问题. Cai 等[9]采用深度学习的方法进行姿态识别，在 Multi-PIE 人脸库上取得 100%的识别率，但该方法需要人工精心地设计深度网络结构和大量的训练样本. 目前，人脸姿态识别方法在实验室环境下已达人眼识别水平，但当面对真实环境时识别率却下降至 75%左右[10]，远低于人类水平；因此，提取鲁棒性特征使得姿态因子以绝对优势压制姿态无关因子是人脸姿态估计框架设计的关键. 本文拟提出一种基于鼻尖点高斯加权的人脸预处理方法，以增强姿态相关因子抑制姿态不相关因子，同时提出一种基于姿态相关和姿态不相关因子分析的鉴别字典学习算法以期提高字典的鉴别性.

1　基于高斯加权的姿态因子增强处理

为了增强人脸姿态因子并减少姿态无关因子的干扰，笔者基于鼻尖点的高斯加权对人脸图像姿态因子进行增强处理：

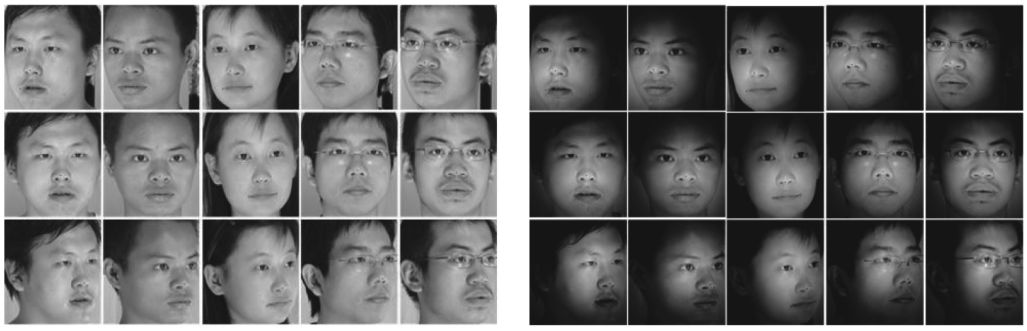
收稿日期：2018－01－30.　* **联系人**，E-mail：liao_haibing@163.com.
基金项目：国家自然科学基金资助项目(61701174)；湖北省自然科学基金资助项目(2017CFB300)；湖北省教育厅科学技术研究资助项目(Q20172805)；湖北科技学院工科硕士点建设专项科研资助项目(2018-19GZ050).
引文格式：廖海斌，邓树文，王电化，等. 基于高斯加权和因子分析字典学习的人脸姿态估计 [J]. 扬州大学学报(自然科学版)，2018，21(4)：47-51，56.

$$F_p=\rho\cdot\exp\left(-\frac{d(p,q)^2}{2\rho\sigma^2}\right),\tag{1}$$

其中 F_p 表示像素点 p 处新的特征值,ρ 为像素点 p 处的原始灰度值,q 为人脸鼻尖点坐标(可以根据人脸关键点定位算法求得),d 为 p 点到鼻尖点的距离,σ^2 为高斯分布方差. 由式(1)可知,p 点距离 q 越近 F_p 越大,越远 F_p 越小.

图 1 给出了人脸图像经高斯加权前后的对比图,其中每列是同一个人的 3 种不同姿态图像,代表着人脸身份;每行是同一姿态下 5 个不同身份的人脸图像,代表着人脸姿态. 由图 1 可见,人脸鼻尖点随姿态改变而变化(每行鼻尖点对比),经过预处理后新人脸的亮度分布也随姿态改变而变化,使得人脸姿态因子被增强,而姿态无关因子被减弱. 例如,假设图 1(a)中第一行第一个人脸的姿态类别为Ⅰ,则第一行剩余的人脸姿态类别也应是Ⅰ,而第一列剩余的人脸姿态类别应有别于Ⅰ;但由于身份信息的干扰,使得每一列代表的身份比每一行代表的姿态更像同一类,即第一列剩余的人脸与第一个人脸更像,导致其姿态类别更趋于一致;然而第一行剩余的人脸与第一个人脸因身份不同却有着很大的差异性,导致会误认为他们的姿态不是同一类. 在经过基于鼻尖点的高斯加权预处理后,新人脸姿态信息明显被加强,每一行代表的姿态比每一列代表的身份更像同一类.

图 2 给出了人脸姿态因子增强效果图. 由图 2 可见,原始人脸主要反映个性身份或表情等姿态无关信息,而新人脸则更多的体现了姿态信息,故本文方法预处理过程中利用该姿态信息对人脸姿态因子进行增强;本预处理方案不但增强了人脸的姿态因子,而且增加了人脸姿态的方位信息.



(a) 原始人脸 (b) 新人脸

图 1 基于高斯加权的新、旧人脸图像

Fig. 1 Old and new face image contrast based on Gaussian weighted method

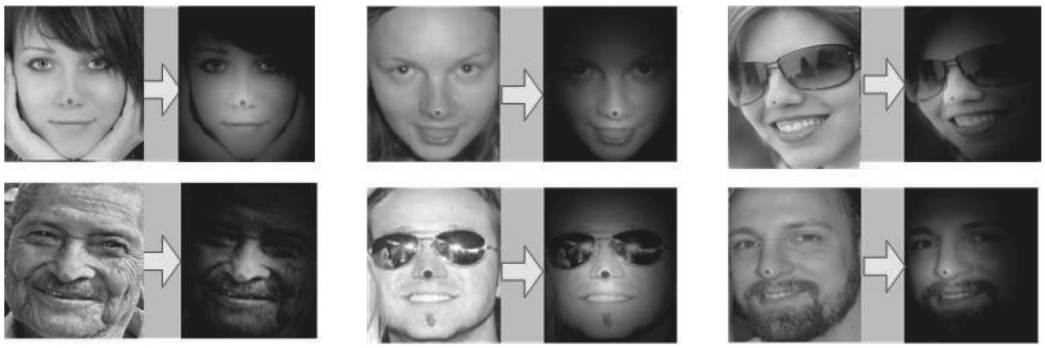


图 2 人脸姿态因子增强效果图

Fig. 2 Effect graphs of face pose factor enhancement

2 基于因子分析字典学习的人脸姿态识别

2.1 人脸姿态稀疏表示

原始人脸图像向量化后是一个高维特征向量,故须采用人脸特征降维[11]或稀疏表示方法[12]进行

姿态估计. 由于本文研究的姿态已知，故采用稀疏表示分类方法进行人脸姿态估计，原始人脸图像经过高斯加权处理后便可作为姿态估计框架的输入进行姿态识别. 将人脸姿态划分为 C 类，假设有姿态训练样本集 $\boldsymbol{A}=[\boldsymbol{A}_1,\boldsymbol{A}_2,\cdots,\boldsymbol{A}_C]$，第 $i(i=1,2,\cdots,C)$ 类训练样本采用特征向量矩阵表示为 $\boldsymbol{A}_i=[\boldsymbol{S}_{i,1},\boldsymbol{S}_{i,2},\cdots,\boldsymbol{S}_{i,n_i}]\in\mathbf{R}^{m\times n_i}$，其中 n_i 为第 i 类姿态样本的数目，m 为样本的特征维数，$\boldsymbol{S}_{i,1}$ 为第 i 类姿态中第 1 个人脸的特征向量. 根据稀疏表示分类原理，测试样本 y 可以由训练样本中少数样本线性组合表示：

$$(\ell^1)\text{：}\hat{x}_1=\arg\min\|x\|_1,\ \boldsymbol{A}x=y, \tag{2}$$

其中 x 为需要求解的稀疏表示系数. 该问题可以通过标准的线性规范方法进行求解. 理论上 $\hat{x}_1$ 应只与训练样本中的某一类姿态样本的关系密切，其对应的表征系数非零，故可以清楚地对该待测姿态进行分类.

2.2　人脸姿态字典学习

假设训练样本集 $\boldsymbol{A}$ 由人脸姿态字典 $\boldsymbol{D}$ 线性组合表示，其稀疏表示系数矩阵为 $\boldsymbol{X}$，$\boldsymbol{X}=[\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots\boldsymbol{X}_c,\cdots,\boldsymbol{X}_C]$，其中 $\boldsymbol{X}_c$ 为子集 $\boldsymbol{A}_c$ 的系数矩阵. 为了使得求取的字典 $\boldsymbol{D}$ 不但对样本集 $\boldsymbol{A}$ 具有良好的稀疏重建能力，而且具有较强的鉴别和噪声处理能力，现构建如下因子分析字典学习模型：

$$\arg\min_{(\boldsymbol{D},\boldsymbol{X})}\{\textstyle\sum_{c=1}^{C}r(\boldsymbol{A}_c,\boldsymbol{D},\boldsymbol{X}_c)+\lambda_1\|\boldsymbol{X}\|_1+\gamma\sum_{c=1}^{C}L(\boldsymbol{D}_c)+\lambda_2 f(\boldsymbol{X})\}, \tag{3}$$

其中 $\lambda_1,\lambda_2,\gamma$ 为平衡因子参数；第一项为重构保真项，$r(\boldsymbol{A}_c,\boldsymbol{D},\boldsymbol{X}_c)=\|\boldsymbol{A}_c-\boldsymbol{D}\boldsymbol{X}_c\|_2+\|\boldsymbol{A}_c-\boldsymbol{D}_c\boldsymbol{X}_{c,c}\|_2+\|\boldsymbol{D}_{c'}\boldsymbol{X}_{c,c'}\|^2$；第二项为稀疏约束项，以保证解的稀疏性；第三项为低秩正定化噪声处理项，使得学习得到的字典更纯净和紧凑；第四项为鉴别约束项，采用经典 Fisher 准则对稀疏表示系数进行姿态因子和姿态无关因子约束，并对其类间（姿态相关因子）与类内（姿态无关因子）散布矩阵加权改进使求取的字典更具鉴别性：

$$f(\boldsymbol{X})=\mathrm{tr}(S_W(\boldsymbol{X}))-\mathrm{tr}(S_B(\boldsymbol{X}))+\eta\|\boldsymbol{X}\|_2, \tag{4}$$

其中 $S_W(\cdot)=\sum_{c=1}^{C}\sum_{i=1}^{N_c-1}\sum_{j=i+1}^{N_c}w(i,j)(\boldsymbol{x}_{c,i}-\boldsymbol{x}_{c,j})(\boldsymbol{x}_{c,i}-\boldsymbol{x}_{c,j})^{\mathrm{T}}$，$\boldsymbol{x}_{c,i}$ 表示第 c 类中第 i 个样本系数，$w(i,j)$ 表示第 c 类中第 i,j 个样本系数间的权值，其目的是使得那些离得稍远的样本对得到更多关注；$S_B(\cdot)=\sum_{c=1}^{C-1}\sum_{c'=c+1}^{C}w(c,c')(\boldsymbol{\mu}_c-\boldsymbol{\mu}_{c'})(\boldsymbol{\mu}_c-\boldsymbol{\mu}_{c'})^{\mathrm{T}}$，$\boldsymbol{\mu}_c$ 为第 c 类系数 $\boldsymbol{X}_c$ 的均值；η 为常量参数；$\|\boldsymbol{X}\|_2$ 为添加的弹性项，以保证 $f(\boldsymbol{X})$ 的凸优化和稳定性.

鉴别字典学习模型的目标函数可以通过交替迭代的方法分成 2 个子问题求解：① 固定字典 $\boldsymbol{D}$，优化匹配得到系数矩阵 $\boldsymbol{X}$；② 固定系数矩阵 $\boldsymbol{X}$，优化匹配得到字典 $\boldsymbol{D}$. 交替迭代直至收敛.

2.3　人脸姿态分类

基于字典 $\boldsymbol{D}$，根据式(2)可求出稀疏表示系数 $\bar{\boldsymbol{x}}=[\bar{\boldsymbol{x}}_1,\bar{\boldsymbol{x}}_2,\cdots,\bar{\boldsymbol{x}}_C]$，其中系数向量 $\bar{\boldsymbol{x}}_i$ 对应于子字典 $\boldsymbol{D}_i$. 根据 $\bar{\boldsymbol{x}}_i$ 定义每类的残差

$$e_i=\|y-\boldsymbol{D}_i\bar{\boldsymbol{x}}_i\|_2^2+w\|\bar{\boldsymbol{x}}-\boldsymbol{\mu}_i\|_2^2, \tag{5}$$

其中第一项为第 i 类的重构误差项，第二项为稀疏表示系数 $\bar{\boldsymbol{x}}$ 与第 i 类系数均值 $\boldsymbol{\mu}_i$（在字典训练时得到）的距离，w 为预设的平衡权值. 对 e_i 进行排序，选择最小的 e_i 所对应的类别确定最终人脸姿态.

3　实验结果分析

选择 XJTU、Multi-PIE、CAS-PEAL-R1 和 AFLW 人脸数据库进行实验验证本文算法的有效性，并与主流的 PCA[4]、Gabor+LDA[4]、BME[13]、SL2[14]、CTF[15] 和 DLSR[8] 等方法比较. 设置参数 $\lambda_1=\lambda_2=0.001,\eta=1,\gamma_1=0.01,\gamma_2=0.005$，设定各子字典的个数 p_i 都相等. 为了验证各方法对人脸图像光照和噪声的鲁棒性，本文在手动对齐的人脸图像上进行训练，然后在有光照、噪声和遮挡的待

测图像上进行人脸姿态识别,分别统计不同姿态的识别准确率.

3.1 鲁棒性实验

随机从 XJTU 中选择 300 个人,其中 200 人用于训练,剩余 100 人用于测试,每人包括 9 个视角(从 19 张视点图像中间隔选取).为了验证算法对噪声的鲁棒性,对测试图像进行加噪处理,噪声强度 σ 分别为 0.01 和 0.03,子字典个数 $p_i=120$,结果如图 3 所示.由图 3 可见:对于没有光照和噪声变化的人脸图像,所有的方法都能取得很好的效果,CTF、DLSR 和本文方法都能取得 99%以上的准确率;然而当人脸图像加噪处理后,PCA、Gabor+LDA 和 BME 的性能迅速下降,而本文方法识别率高且稳定.结果表明:本文方法在 XJTU 上获得了最高的准确率,尤其是在噪声测试集上也表现出很好的性能.

采用 Multi-PIE 人脸姿态库进一步验证算法对光照和表情的鲁棒性.随机从 Multi-PIE 中挑选 200 个人(每个人包括 9 种不同姿态并具有 3 种不同光照或表情变化)作为训练集,剩余的 137 个人作为测试集,子字典个数 $p_i=160$,结果如图 4 所示.由图 4 可见:当人脸出现光照或表情变化时,PCA、Gabor+LDA 和 BME 方法识别率急剧下降至 78%,SL2、CTF 和 DLSR 的识别率下降至 96%,而本文方法基本比较稳定能保持在 98%左右.结果表明:本文方法具有良好的光照和表情鲁棒性,虽然 DLSR 也是基于稀疏表示与字典学习的识别框架,但由于本文采用了高斯加权预处理、Gabor特征和因子分析字典学习方法,故本文方法的稳定性优于 DLSR.

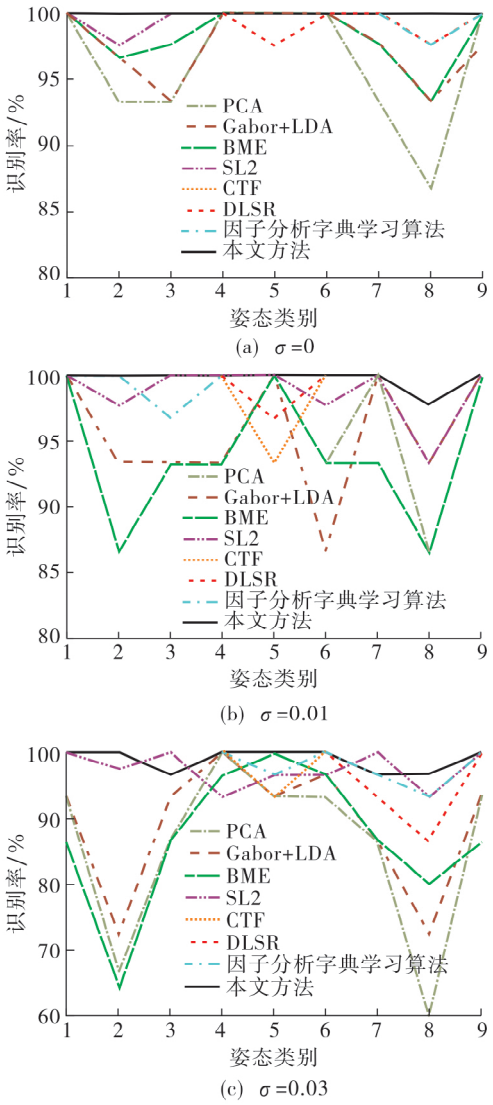


图 3 不同噪声强度下的人脸姿态识别率

Fig. 3 Face pose recognition based on different noise intensity

3.2 自然场景实验

利用 AFLW 人脸库验证不同算法对自然场景中人脸姿态识别的性能.根据 AFLW 提供的人脸关键点提取人脸区域,生成 459 张人脸区域像素不小于 64×64 的图像作为测试集.整合 CAS-PEAL-R1 和 Weizmann 人脸库构建一个包含 200 个人的训练集,子字典个数 $p_i=120$.由于训练集和测试集中的姿态并非完全一致,重新定义人脸姿态类别 $C=1,\theta<-45°$;$C=2,-45°\leqslant\theta<-10°$;$C=3,-10°\leqslant\theta<+10°$;$C=4,+10°\leqslant\theta<+45°$;$C=5,\theta\geqslant+45°$;$C=6$,抬头;$C=7$,低头,$C=1\sim7$ 分别表示左侧面、左偏转、正面、右偏转、右侧面、抬头和低头.为了验证本文方法的有效性,实验增加了采

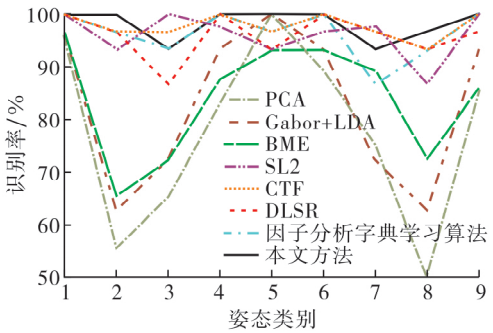


图 4 不同光照或表情变化下的人脸姿态识别率

Fig. 4 Face pose recognition with illumination or expression

用传统的基于奇异值分解(singular value decomposition，SVD)的字典学习与本文基于因子分析的字典学习方法的对比，结果如表 1 所示. 由表 1 可见，除了在第 4 类和第 6 类中 CTF 方法略优于本文方法外，在其他类别中本文方法性能均最优；基于因子分析的字典方法比基于 SVD 的字典学习方法的识别率高 1%～2%；高斯加权预处理可使识别率提高 1%～5%. AFLW 中人脸具有极大的光照、表情和噪声变化，是一个极具挑战的测试库，本文方法在此测试集上依然可保持较优的性能.

表 1　光照、表情和噪声下各类人脸姿态识别率

Tab. 1　The face poses recognition with illumination, expression and noise

算法	第 1 类	第 2 类	第 3 类	第 4 类	第 5 类	第 6 类	第 7 类
PCA	0.74	0.71	0.65	0.67	0.78	0.82	0.77
Gabor+LDA	0.74	0.73	0.73	0.68	0.85	0.86	0.81
BME	0.78	0.73	0.73	0.79	0.83	0.86	0.82
SL2	0.82	0.79	0.76	0.85	0.87	0.84	0.79
CTF	0.94	0.83	0.76	0.95	0.90	0.95	0.93
DLSR	0.86	0.80	0.69	0.89	0.85	0.79	0.90
SVD 字典学习	0.89	0.81	0.75	0.89	0.86	0.83	0.90
因子分析字典学习	0.91	0.85	0.80	0.92	0.87	0.84	0.92
高斯加权－SVD 字典学习	0.93	0.87	0.82	0.94	0.90	0.85	0.94
本文方法	0.95	0.87	0.85	0.94	0.92	0.93	0.95

参考文献：

[1] 唐云祁，孙哲南，谭铁牛. 头部姿势估计研究综述 [J]. 模式识别与人工智能，2014，27(3)：213-223.

[2] WANG Jiangang，SUNG E. EM enhancement of 3D head pose estimated by point at infinity [J]. Image Vision Comput，2007，25(12)：1864-1874.

[3] MORENCY L P，WHITEHILL J，MOVELLAN J. Monocular head pose estimation using generalized adaptive view-based appearance model [J]. Image Vision Comput，2010，28(5)：754-761.

[4] FOYTIK J，ASARI V K. A two-layer framework for piecewise linear manifold-based head pose estimation [J]. Int J Comput Vision，2013，101(2)：270-287.

[5] GE Liuhao，LIANG Hui，YUAN Junsong，et al. Robust 3D hand pose estimation from single depth images using multi-view CNNs [J]. IEEE Trans Image Proc，2018，27(9)：4422-4436.

[6] QIN Zhen，SHELTON C. Social grouping for multi-target tracking and head pose estimation in video [J]. IEEE Trans Pattern Anal Mach Intell，2016，38(10)：2082-2095.

[7] MA Bingpeng，LI Annan，CHAI Xiujuan，et al. CovGa：a novel descriptor based on symmetry of regions for head pose estimation [J]. Neurocomputing，2014，143(16)：97-108.

[8] ZHANG Yuyao. Non-linear dimensionality reduction and sparse representation models for facial analysis [D]. Lyon：Université de Lyon，2014.

[9] CAI Ying，YAG Menglong，LI Jun. Multiclass classification based on a deep convolutional network for head pose estimation [J]. Front Inf Technol Electr Eng，2015，16(11)：930-939.

[10] DEMIRKUS M，PRECUP D，CLARK J J，et al. Hierarchical temporal graphical model for head pose estimation and subsequent attribute classification in real-world videos [J]. Comput Vision Image Understand，2015，136(1)：128-145.

[11] LI Zechao，TANG Jinhui，HE Xiaofei. Robust structured nonnegative matrix factorization for image representation [J]. IEEE Trans Neural Networks Learn Syst，2018，29(5)：1947-1960.

[12] BANERJEE M，NIKHIL R P. Unsupervised feature selection with controlled redundancy (UFeSCoR) [J]. IEEE Trans Knowl Data Eng，2015，27(12)：3390-3403.

[13] BALASUBRAMANIAN V N，KRISHNA S，PANCHANATHAN S. Person-independent head pose estimation using based manifold embedding [J]. Eurasip J Adv Signal Process，2007，2008(1)：1-15.

[14] LIU Yong，WANG Qicong，JIANG Yi，et al. Supervised locality discriminant manifold learning for head pose estimation [J]. Knowl Based Syst，2014，66(1)：126-135.

[15] PENG Xi，HUANG Junzhou，HU Qiong，et al. From circle to 3-sphere：head pose estimation by instance parameterization [J]. Comput Vision Image Understand，2015，136(3)：92-102.

（下转第 56 页）

A low-complexity image registration approach based on SIFT

ZHANG Chenguang, ZHOU Quan*, HUI Zheng
(Xi'an Institute of Space Radio Technology, Xi'an 710100, China)

Abstract: In order to reduce the computational complexity in large-size image registration, the author proposes a new image registration method based on the SIFT feature point detection algorithm. Using this method, the original image is down-sampled firstly as preprocessing before detecting feature points by SIFT. Therefore, the complexity of operation, especially Gaussian-kernel convolution computation in the image pyramid building phase is reduced. In addition, the algorithm makes use of BBF algorithm to realize k neighbor points search in the k-d tree and gets initial matches of the corresponding feature points quickly. RANSAC algorithm in iteration is also used to eliminate false matches and obtain the optimal solution of the parameters of transform model that can fit all the inner-points. Finally through coordinate transformation and interpolation, image registration is realized. The similarity between the registrated image and the reference image is measured by the peak signal-to-noise ratio. The experimental results show that, the proposed method can shorten the running time and guarantee good registration performance simultaneously compared with the traditional direct registration.

Keywords: image registration; scale-invariant feature transform(SIFT); down sample; pyramid of images; peak signal to noise ratio(PSNR); computational complexity

(责任编辑 秋 实)

(上接第 51 页)

Face poses estimation based on Gaussian weighted and factors analysis dictionary learning

LIAO Haibin[1], DENG Shuwen[1]*, WANG Dianhua[1], FAN Ping[1], CHEN Youbin[2]
(1. School of Computer Science & Technology, Hubei University of Science & Technology, Xianning 437100, China;
2. School of Automation, Huazhong University of Science & Technology, Wuhan 430074, China)

Abstract: Accurate estimation of facial pose in an uncontrolled environment presents a great challenge. However, extant conventional approaches lack the capability to deal with multiple pose-related and -unrelated factors in a uniform way. This paper proposes a robust pose estimation framework based on dictionary-learning and sparse representation. With the guide of this framework, a novel face image pre-processing algorithm based on Gaussian weighted and tip of nose is designed to enhance pose-related factors. Further, a new insight on discriminative dictionary learning is also provided. Specifically, the author formulates the discrimination term based on pose-related and-unrelated factors analysis. Several experiments are performed on XJTU, CMU MultiPIE CAS-PEAL-R1 and AFLW databases. Recognition results show that the proposed method can achieve recognition rate about 95% under illumination, noises and occlusion variations.

Keywords: face pose estimation; dictionary learning; sparse representation; factors analysis

(责任编辑 林 子)