



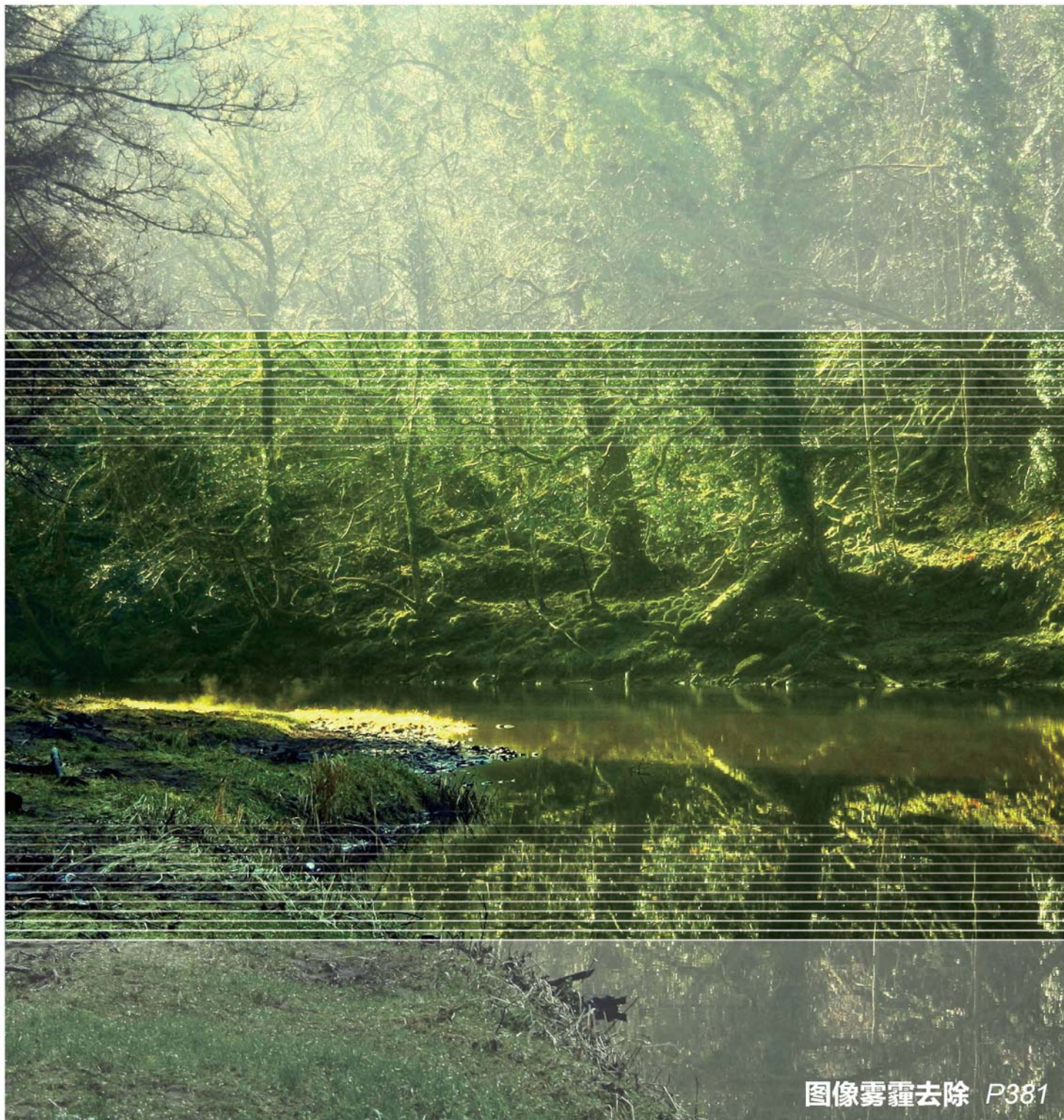
JOURNAL OF IMAGE AND GRAPHICS

主办: 中国科学院遥感与数字地球研究所
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象图形学报

2014
03
VOL.19

ISSN1006-8961
CN11-3758/TB



图像雾霾去除 P381



第19卷第3期（总第215期）

2014年3月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊

版权声明

凡向《中国图象图形学报》投稿，均视为同意在本刊网站及CNKI等全文数据库出版，所刊载论文已获得著作权人的授权。本刊所有图片均为非商业目的使用，所有内容，未经许可，不得转载或以其他方式使用。

Copyright

All rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use.

主管单位 中国科学院

主办单位 中国科学院遥感与数字地球研究所
中国图象图形学学会
北京应用物理与计算数学研究所

主 编 李小文

编辑出版《中国图象图形学报》编辑出版委员会

邮政信箱 北京9718信箱

邮 编 100101

电子信箱 jig@irsa.ac.cn

电 话 010-64807995

网 址 www.cjig.cn

广告经营许可证 京朝工商广字第0361号

总 发 行 北京报刊发行局

订 购 全国各地邮局

海外发行 中国国际图书贸易集团有限公司
(邮政信箱：北京399信箱 邮编：100048)

印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian

Monthly, Started in 1996

Superintended by Chinese Academy of Sciences

Sponsored by Institute of Remote Sensing and Digital Earth, CAS
China Society of Image and Graphics

Institute of Applied Physics and Computational Mathematics

Editor-in-Chief LI Xiaowen

Editor, Publisher Editorial and Publishing Board of Journal of
Image and Graphics

P.O.Box 9718, Beijing, P.R.China

Zip code 100101

E-mail jig@irsa.ac.cn

Telephone 010-64807995

Website www.cjig.cn

Distributed by Beijing Bureau for Distribution of Newspapers and Journals

Domestic All Local Post Offices in China

Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048, P.R.China))

Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB

ISSN 1006-8961

CODEN ZTTXFZ

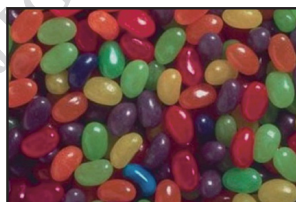
国外发行代号 M1406

国内邮发代号 82-831

国内定价 50.00元



控制关键帧选择的H.264熵编码加密算法(第358页)



显著性检测指导的高光区域修复(第393页)



由灭点进行径向畸变的自动校正(第407页)

综述

图像场景分类中视觉词包模型方法综述

赵理君, 唐娉, 霍连志, 郑柯 0333

面向对象的高分辨率SAR图像处理及应用

张红, 叶曦, 王超, 张波, 吴樊, 汤益先 0344

图像处理和编码

控制关键帧选择的H.264熵编码加密算法

张小红, 袁春经 0358

小波域视觉密码零水印算法

曲长波, 杨晓陶, 袁铎宁 0365

双重轮廓演化曲线的图像分割水平集模型

王相海, 李明 0373

暗原色先验单幅图像去雾改进算法

孙小明, 孙俊喜, 赵立荣, 曹永刚 0381

基于双边滤波的图像去雾

王一帆, 尹传历, 黄义明, 王洪玉 0386

显著性检测指导的高光区域修复

王祎璠, 姜志国, 史骏, 张浩鹏 0393

应用DCT块标准差的自适应隐写算法

凌轶华, 蔡晓霞, 陈红 0401

由灭点进行径向畸变的自动校正

刘丹, 刘学军, 王美珍 0407

图像分析和识别

黎曼流形上的保局投影在图像集匹配中的应用

曾青松 0414

结合全局和局部信息的“两阶段”活动轮廓模型

戚世乐, 王美清 0421

图像理解和计算机视觉

基于超像素的点追踪方法

罗会兰, 钟睿, 孔繁胜 0428

遥感图像处理

结合相位一致与全变差模型的高分辨率遥感图像边缘检测

黄秋燕, 肖鹏峰, 冯学智, 王珂 0439

第9届和谐人机环境联合学术会议专栏

人眼视觉感知特性的非下采样Contourlet变换域多聚焦图像融合

杨勇, 郑文娟, 黄淑英, 方志军, 袁非牛 0447

面向手绘军标图形的旋转自由识别方法

吴玲达, 张友根, 邓维, 宋汉辰 0456

基于深度相机的手腕识别与掌心估测

姚争为, 潘志庚, 滕国栋 0463

多信息融合的快速车牌定位

王永杰, 裴明涛, 贾云得 0471

H.264/AVC中基于决策树的P帧快速模式选择

王萍, 张晓丹, 张磊 0476

互信息域中的无参考图像质量评价

董宏平, 刘利雄 0484

CONTENTS

JOURNAL OF IMAGE AND GRAPHICS



H.264 video entropy coding encryption by controlling key frames(P358)



Algorithm for highlight area inpainting guided by saliency detection(P393)



Automatic approach of lens radial distortion correction based on vanishing points(P407)

Review

- Review of the bag-of-visual-words models in image scene classification
Zhao Lijun, Tang Ping, Huo Lianzhi, Zheng Ke 0333
- The processing and applications of high resolution SAR images with object-based image analysis
Zhang Hong, Ye Xi, Wang Chao, Zhang Bo, Wu Fan, Tang Yixian 0344

Image Processing and Coding

- H.264 video entropy coding encryption by controlling key frames
Zhang Xiaohong, Yuan Chunjing 0358
- Zero-watermarking visual cryptography algorithm in the wavelet domain
Qu Changbo, Yang Xiaotao, Yuan Duoning 0365
- Level set model for image segmentation based on dual contour evolutionary curve
Wang Xianghai, Li Ming 0373
- Improved algorithm for single image haze removing using dark channel prior
Sun Xiaoming, Sun Junxi, Zhao Lirong, Cao Yonggang 0381
- Image haze removal using a bilateral filter
Wang Yifan, Yin Chuanli, Huang Yiming, Wang Hongyu 0386
- Highlight area inpainting guided by saliency detection
Wang Yifan, Jiang Zhiguo, Shi Jun, Zhang Haopeng 0393
- Adaptive steganographic algorithm utilizing block standard deviation of DCT coefficients
Ling Yihua, Cai Xiaoxia, Chen Hong 0401
- Automatic approach of lens radial distortion correction based on vanishing points
Liu Dan, Liu Xuejun, Wang Meizhen 0407

Image Analysis and Recognition

- Locality preserving projection on Riemannian manifold for image set matching
Zeng Qingsong 0414
- "Two-stage" active contour model driven by local and global information
Qi Shile, Wang Meiqing 0421

Image Understanding and Computer Vision

- Method of point tracking based on superpixel
Luo Huilan, Zhong Rui, Kong Fansheng 0428

Remote Sensing Image Processing

- Edge detection from high resolution remote sensing image combining phase congruency with total variation model
Huang Qiuyan, Xiao Pengfeng, Feng Xuezhi, Wang Ke 0439

Special Issue on HHME 2013

- Multi-focus image fusion based on human visual perception characteristic in non-subsampled Contourlet transform domain
Yang Yong, Zheng Wenjuan, Huang Shuying, Fang Zhijun, Yuan Feiniu 0447
- Rotation free recognition of hand-drawn military marking symbols
Wu Lingda, Zhang Yougen, Deng Wei, Song Hanchen 0456
- The wrist recognition and the center of palm estimation based on depth camera
Yao Zhengwei, Pan Zhigeng, Teng Guodong 0463
- License plate detection based on multiple features
Wang Yongjie, Pei Mingtao, Jia Yunde 0471
- Fast intermode decision based on the decision tree for H.264/AVC P-frame encoding
Wang Ping, Zhang Xiaodan, Zhang Lei 0476
- No-reference image quality assessment in mutual information domain
Dong Hongping, Liu Lixiong 0484

中图法分类号: TP309.2 文献标识码: A 文章编号: 1006-8961(2014)03-0414-07

论文引用格式: 曾青松. 黎曼流形上的保局投影在图像集匹配中的应用[J]. 中国图象图形学报, 2014, 19(3): 414-420. [DOI: 10.11834/jig.20140311]

黎曼流形上的保局投影在图像集匹配中的应用

曾青松

广州番禺职业技术学院信息工程学院, 广州 511483

摘要: **目的** 提出了黎曼流形上局部结构特征保持的图像集匹配方法。**方法** 该方法使用协方差矩阵建模图像集合, 利用对称正定的非奇异协方差矩阵构成黎曼流形上的子空间, 将图像集的匹配转化为流形上的点的匹配问题。通过基于协方差矩阵度量学习的核函数将黎曼流形上的协方差矩阵映射到欧几里德空间。不同于其他方法黎曼流形上的鉴别分析方法, 考虑到样本分布的局部几何结构, 引入了黎曼流形上局部保持的图像集鉴别分析方法, 保持样本分布的局部邻域结构的同时提升样本的可分性。**结果** 在基于图像集合的对象识别任务上测试了本文算法, 在 ETH80 和 YouTube Celebrities 数据库分别进行了对象识别和人脸识别实验, 分别达到 91.5% 和 65.31% 的识别率。**结论** 实验结果表明, 该方法取得了优于其他图像集匹配算法的效果。

关键词: 黎曼流形; 集合匹配; 保局投影; 协方差矩阵

Locality preserving projection on Riemannian manifold for image set matching

Zeng Qingsong

School of Information and Technology, Guangzhou Panyu Polytechnic, Guangzhou 511483, China

Abstract: Objective This paper addresses the problem of image set matching in kernel learning approach, and a locality geometry structure preserved discriminant analysis based on a Riemannian manifold is presented for image set matching. Riemannian manifolds have been an effective way to represent image sets, which are mapped as data points on the manifold. Then, recognition can be performed by applying the discriminant analysis on such manifolds. However, the local structure of the data points is not exposed in the discriminant analysis. In computer vision applications, the multi-view facial images are nonlinearly distributed, and features often lie on Riemannian manifolds with known geometry. Set-based matching methods utilize a set of images as input and model the probe set and gallery set individually. Hence, these methods can fully utilize the information provided by multiple images to obtain better matching and recognition accuracy. However, the popular learning algorithms such as discriminant analysis and support vector machines, etc., are not directly applicable to such features due to the non-Euclidean nature of the underlying spaces. To overcome this limitation, each image set is mapped into a high-dimensional feature space, e.g. a high dimensional Reproducing Kernel Hilbert Space (RKHS), using a nonlinear mapping function, to which many Euclidean algorithms can be generalized. **Method** For set based image matching techniques, as mentioned above, the key issues can be categorized based on how to represent the image sets and how to measure the similarity between two image sets. We naturally formulate the problem of the image set matching as matching points lying on the Riemannian manifold spanned by symmetric positive definite (SPD), i.e. nonsingular covariance matrices. In general, the success of kernel-based methods is often determined by the choice of the kernel function. By exploring an effi-

收稿日期: 2013-06-25; 修回日期: 2013-09-02

基金项目: 广州市属高校科研重点项目 (2012A164)

第一作者简介: 曾青松 (1976—), 男, 副教授, 中山大学通信与信息系统专业博士研究生, 主要研究方向为模式识别与数据挖掘。

E-mail: qingsongzeng@163.com

cient metric for the symmetric positive definite covariance matrices, i. e. Log-Euclidean distance, we derive a kernel function that explicitly maps the nonsingular covariance matrix from the Riemannian manifold to a Euclidean space. Different from other methods on Riemannian manifold, the local structure of data is taking into account. With the explicit mapping, a kernel version of Locality Preserving Projection (LPP) is applied to keep the local geometry structure of the image set and an image set-based matching method is proposed. **Result** The proposed method is evaluated on set-based object classification tasks and face recognition tasks. Extensive experimental results show that the proposed method outperforms other state of the art set-based object matching and face recognition methods. In this paper, according our experimental settings, it reaches 91.5% and 65.31% on recognition rate in the public ETH80 object database and YouTube Celebrities video database respectively. **Conclusion** In this paper, we have proposed an efficient image set matching method. This method represents each image set with its covariance matrix and models the problem as matching points on the Riemannian manifold spanned by nonsingular covariance matrix. We derived a Mercer kernel function, which successfully bridges the gap between traditional learning methods operating in vector space and the learning task on a manifold. Through comparison experiments, it is shown to be generally better than other image sets matching methods.

Key words: Riemannian manifold; set matching; locality preserving projection; covariance matrix

0 引言

随着视频监控应用的需要,基于图像集匹配的技术得到长足的发展,成为模式识别领域研究的热点问题之一。基于图像集合的匹配方法对查询图集和数据库中的图集分别建模,并利用多幅图像提供的信息以获得更好的匹配性能。通过对查询图集和数据库中的图集分别建模,两个图像集合的相似性问题转换为模型对之间的距离计算问题。

建模一个图像集合,最常见的方法是考虑使用子空间、仿射包和流形等方法。互子空间(MSM)方法^[1]将图像集投影到低维线性子空间,通过计算图像序列生成的子空间之间的夹角,利用典型相关度量两个集合的相似性。仿射包方法^[2]使用两个仿射包中距离最近的两个点之间的几何距离来度量集合之间的相似性。由于直接搜索最近邻的点,而不加任何额外约束条件,所以噪声和离群点的干扰会严重影响分类的精确度。为了克服这一缺陷,Hu^[3]使用稀疏表达的方法,提出稀疏逼近最近邻点(SANP)来度量两个仿射包之间的相似性。基于全局统一的模型来描述,没有考虑到样本的局部结构,因而不能很好地刻画那些光照、姿态、和表情等变化复杂的数据。Wang等人^[4]扩展了子空间距离,提出了流形—流形距离(MMD)的理论框架,将整个非线性流形划分成多个局部模型(MLPs)^[5],将两个流形之间的距离转换为计算成对的MLP之间的距离问题。作为MMD的扩展,通过用流形来描述每一个集合,Wang等人^[6]提出了流形鉴别分析(MDA),通

过多流形学习解决有监督的流形间距离计算问题。

最近的研究表明,如果考虑黎曼流形的几何结构可以取得更好的识别性能。Wang等人^[7]通过协方差建模图像集合,并应用线性核鉴别分析,取得很好的识别效果,但是该方法没有考虑到样本的局部结构信息。Harandi等人^[8]提出格拉斯曼流形上的图嵌入鉴别分析方法,将集合匹配问题嵌入到图嵌入鉴别分析框架^[9],该方法充分考虑样本分布的局部结构,取得很好的识别效果。

从图像集合分布的局部结构出发,利用局部保持投影技术,提出黎曼流形上的集合匹配方法,并应用到对象识别应用中。首先计算每一个图像集合的协方差矩阵建立图像集合的模型,利用非奇异协方差矩阵构成黎曼空间 Sym_d^+ ,将两个图像集合的相似性度量转换为 Sym_d^+ 上两个点之间的距离度量;进一步利用局部保持投影更好地刻画黎曼流形上样本点的局部几何关系,使得在投影空间中样本的邻域关系得以保持,同时具有更好的可分性,进一步提高鉴别性能。

1 Sym_d^+ 上的黎曼核

众所周知,非奇异协方差矩阵是对称正定矩阵,而 $d \times d$ 大小正定对称矩阵张成黎曼流形 Sym_d^+ ^[10]。因此,可以使用协方差矩阵来描述一个图像集,每一个协方差矩阵理解为 Sym_d^+ 空间上的一个点,将图像集合的相似性计算转换为 Sym_d^+ 空间上的点对的距离的计算问题。

设 $X = U\Sigma U^T$ 表示正定对称矩阵 X 的特征值分

解,矩阵 X 的对数仍然是一个对称矩阵,可以通过 $\text{LOG}(X) = U \text{LOG}(\Sigma) U^T$ 计算, $\text{LOG}(\Sigma)$ 是特征值的对数构成的对角矩阵^[7]。

在 Sym_d^+ 空间中,矩阵对数域上定义 X_i 和 X_j 之间的 Log-Euclidean 距离,即

$$d_g(X_i, X_j) = \|\text{LOG}(X_i) - \text{LOG}(X_j)\|_F \quad (1)$$

式中, $\|\cdot\|_F$ 表示 Frobenius 矩阵范数。在该空间上定义的 Log-Euclidean 距离是 Sym_d^+ 空间上的几何距离度量^[11]。式(1)中的度量可以理解为黎曼流形 M 上的点 X 通过以下对数映射投影到欧氏空间,即

$$\Phi_{\log}: M \rightarrow T_1, X \rightarrow \text{LOG}(X) \quad (2)$$

映射 $\Phi_{\log}$ 表示单位矩阵 I 所对应的点的切空间 T_1 , 这个空间由 $d \times d$ 的对称矩阵张成的向量空间。因此式(1)定义的度量转换为 $\mathbf{R}^{d \times d}$ 空间的欧氏距离的度量。通过计算切空间 T_1 中的内积,可以得到黎曼流形 M 上的黎曼核^[7], 对 $\forall X_i, X_j \in M$ 定义

$$k_{\log}(X_i, X_j) = \text{tr}[\text{LOG}(X_i) \text{LOG}(X_j)] \quad (3)$$

$k_{\log}$ 是一个实对称函数: $k_{\log}(X_i, X_j) = k_{\log}(X_j, X_i)$, 因为

$$\begin{aligned} \text{tr}[\text{LOG}(X_i) \text{LOG}(X_j)] &= \\ \text{tr}[\text{LOG}(X_i)^T \text{LOG}(X_j)] &= \\ \text{tr}[\text{LOG}(X_j) \text{LOG}(X_i)^T] &= \\ \text{tr}[\text{LOG}(X_j) \text{LOG}(X_i)] & \end{aligned} \quad (4)$$

这里利用了 $\text{LOG}(X_i)$ 与 $\text{LOG}(X_j)$ 是对称矩阵和 $\text{tr}[A \cdot B] = \text{tr}[B \cdot A]$ 。

对 $\forall X_1, \dots, X_n \in M$ 与 $b_1, \dots, b_n \in \mathbf{R}$, $k_{\log}$ 是正定的, 因为

$$\begin{aligned} \sum_{i,j} b_i b_j k_{\log}(X_i, X_j) &= \\ \sum_{i,j} b_i b_j \text{tr}[\text{LOG}(X_i) \text{LOG}(X_j)] &= \\ \text{tr} \left[\left(\sum_i b_i \text{LOG}(X_i) \right)^2 \right] &= \\ \left\| \sum_i b_i \text{LOG}(X_i) \right\|_F^2 &\geq 0 \end{aligned} \quad (5)$$

2 Sym_d^+ 上的保局投影

保局投影(LPP), 可以较好地反映样本的流形结构^[12], LPP 得到一个投影子空间, 将测试样本向训练样本进行投影, 得到相应的低维空间, 在这个子空间中原始样本点的局部信息被保存下来, 即邻近的点在新的子空间映射下仍然保持近邻。

给定黎曼流形 M 上的 N 个点 $\{(X_i, l_i)\}_{i=1}^N$,

$X_i \in \mathbf{R}^{d \times m}$ 是一个线性子空间, $l_i \in \{1, 2, \dots, C\}$ 表示类标。 M 的局部几何结构可以通过构造相似性图来表达。记 $N(X_i)$ 表示 X_i 的近邻, 相似性图矩阵 $W = [W_{i,j}]_{N \times N}$ 可以表示为

$$W_{i,j} = \begin{cases} 1 & X_i \in N(X_j) \text{ 或 } X_j \in N(X_i) \\ 0 & \text{其他} \end{cases} \quad (6)$$

目标是找到黎曼流形 M 与 M' 之间的一个映射 $\phi: X_i \rightarrow Y_i$, 该映射满足流形 M' 上局部邻域的几何结构保持不变。这个目标就是求解最优化问题

$$\min \frac{1}{2} \sum_{i,j} (Y_i - Y_j)^2 W_{i,j} \quad (7)$$

假设黎曼流形上的点可以通过其邻域上的点的

线性组合表示, 即 $\alpha_i = \sum_{j=1}^N a_{i,j} X_j$ 。记 $\langle \alpha_i, X_i \rangle = A_i^T K_i, A_i = [a_{i,1}, \dots, a_{i,N}]^T, K_i = [k_{i,1}, \dots, k_{i,N}]^T$, 可得

$$Y_i = [\langle \alpha_1, X_i \rangle, \dots, \langle \alpha_r, X_i \rangle]^T = (A_1^T, A_2^T, \dots, A_r^T) K_i = A^T K_i \quad (8)$$

式(7)可以简化为

$$\begin{aligned} \frac{1}{2} \sum_{i,j} (Y_i - Y_j)^2 W_{i,j} &= \frac{1}{2} \sum_{i,j} (A^T K_i - A^T K_j)^2 W_{i,j} = \\ \sum_i A^T K_i K_i^T A^T W_{i,i} - \sum_{i,j} A^T K_j K_i^T A^T W_{i,j} &= \\ A^T K D K^T A - A^T K W K^T A &= A^T K L K^T A \end{aligned} \quad (9)$$

式(9)利用了 $W_{ij} = W_{ji}$ 进行化简, 式中 $A = [A_1 | A_2 | \dots | A_r], K = [K_1 | K_2 | \dots | K_N], L = D - W$ 表示拉普拉斯矩阵, 对角矩阵 $D_{ij} = \sum_{j \neq i} W_{ij}$ 。增加约束条件 $A^T K D K^T A = \mathbf{1}$, 式(9)的最小化问题变为

$$\begin{aligned} \arg \min_A A^T K L K^T A \\ \text{s. t. } A^T K D K^T A = \mathbf{1} \end{aligned} \quad (10)$$

为了求最优投影矩阵, 只需解广义特征值问题, 即

$$K L K^T A = \lambda K D K^T A \quad (11)$$

最大的 r 个特征值对应的特征向量 $A_i (i=1, 2, \dots, r)$ 构成最终的投影矩阵 $A = [A_1 | A_2 | \dots | A_r]$ 。

算法1 给出了黎曼流形上保局投影的算法框架。

算法1 黎曼流形上保局投影

输入:

1) 训练集 $\{(X_i, l_i)\}_{i=1}^N$, 式中 $X_i \in \mathbf{R}^{d \times m}$ 是一个线性子空间, $l_i \in \{1, 2, \dots, C\}$ 表示类标;

2) 黎曼流形上的核函数 $k_{i,j}$, 用于计算黎曼流形上两个点之间的相似性的度量。

处理过程:

1) 对所有的图像集 X_i 和 X_j 计算核矩阵 $[K]_{i,j}$;

2) 使用式(6)构造相似性图矩阵 W , 计算度矩

阵 D 和拉普拉斯矩阵 L ;

3) 解最优化问题式 (10), 取最大的 r 个特征值对应的特征向量构成的矩阵作为投影矩阵 $A = [A_1 | A_2 | \dots | A_r]$;

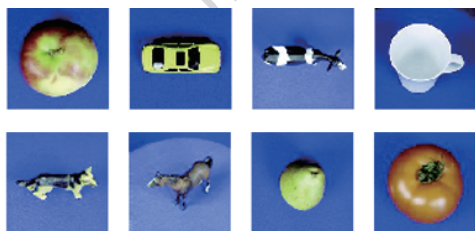
输出: 投影矩阵 $A = [A_1 | A_2 | \dots | A_r]$ 。

3 实验

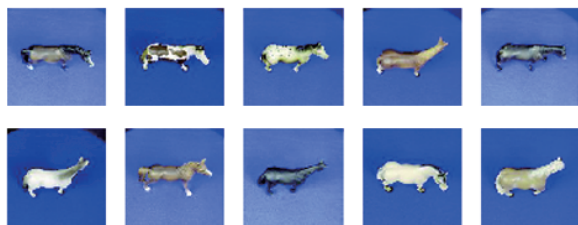
3.1 数据库

使用 ETH80 对象数据库进行分类实验以及使用 YouTube Celebrities 视频数据库进行人脸识别实验。

ETH80 数据库包含 8 类对象, 如图 1(a) 所示。每个类别包含 10 个对象, 如图 1(b) 所示。每个个体包含不同视角下的 41 幅图像, 其目的主要用于物体分类。每一个体所有 41 幅图像当成一个图像集合, 每类物体有 10 个图像集合, 整个数据库共 80 个集合。将每一个体的 10 个集合随机划分为两组, 每组 5 个集合分别用于训练和测试。实验中, 每个集合随机选择 30° 视角的图像用于构造随机训练和测试集合。



(a) 数据库包含 8 种类别的对象



(b) 同一类别包含 10 个不同视角的对象

图 1 ETH80 对象数据库中的样本

Fig. 1 Example images from the object database

YouTube Celebrities 数据库^[13] 视频文件全部来自于 YouTube 网站上的视频片段, 主要用于视频人脸跟踪与识别任务。整个库包含 47 个人的 1 910 段视频, 每个人的视频片段分为 3 个小节, 每一个小节对应不同的采集时间与场景, 每个小节包含至少 3 段视频, 大部分的视频都有多余 100 帧 (极个别视频只有 40 帧左右)。这个数据库包含大范围的姿态、光照和表情变化, 使用文献 [14] 给出的对象跟

踪算法, 提取相应的人脸图像。

图 2 给出了 AL Gore 的第 0、10、20 和 30 帧的跟踪结果。其中第 0 帧为人工标定, 后续帧由程序自动跟踪。

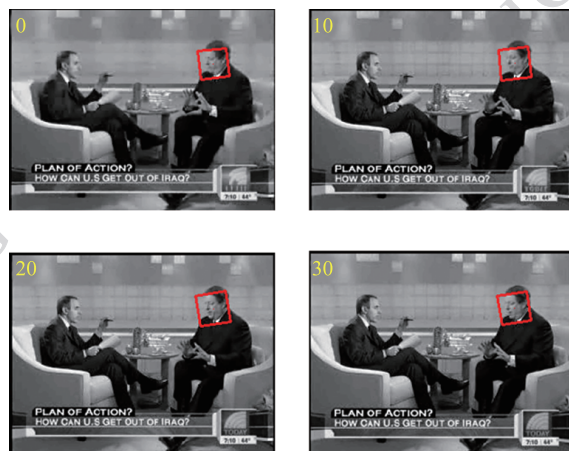


图 2 YouTube 数据库人脸跟踪的结果

Fig. 2 Face tracking results on YouTube

进行交叉验证实验, 即随机选择 10 组训练和测试集合的组合来报告识别率。使用最近邻分类器对样本进行分类。

3.2 比较算法和设置

比较了本文算法与在第 1 节中提到的另外 6 种基于集合的匹配算法。这些算法包括: 互子空间方法 (MSM)^[1]; 流形—流形距离 (MMD)^[4]; 流形鉴别分析 (MDA)^[6]; 基于仿射包的图像集距离 (AHISD)^[2]; 图像集稀疏逼近邻 (SANP)^[3] 和黎曼流形图嵌入鉴别分析 (GGDA)^[8]。

所有比较的实验基于原作者公布的代码适当的调节算法的参数取最优的结果。在 MSM 和 MMD 算法中, 使用主成分分析 (PCA) 保留 95% 的样本能量来获得子空间。对 MMD 和 MDA 算法, 参数设置与文献相同, 欧氏距离与几何距离比值设置为 2.0, 使用最大典型相关计算距离, 计算 MMD 和 MDA 时, 邻域大小设置为默认值 12。对于 AHISD, 使用其线性版本, 并通过 PCA 提取 98% 的能量。对于 SANP 算法, 使用 PCA 保留 95% 的数据能量, 使用文献相同的参数设置求解凸优化问题。GGDA 算法中, 只实现了提出的最简单的二分图, 并且使用典型相关计算黎曼核函数, 邻域大小参数设置为 5。本文算法, 在构造相似性图矩阵 W 时使用邻域大小为 3, 在求解式 (10) 时使用全部特征值对应的特征向量构成最终的投影矩阵。

3.3 实验结果与分析

在 ETH80 和 YouTube 数据库上测试了本文算法的识别性能并与其他典型的基于图像集合的匹配算法进行比较。实验中本文算法采用最近邻分类器进行分类。随机实验一共运行 10 次。比较实验结果和原文存在一些差异,主要由于不同文献对数据库中样本的处理方式存在一定的差异,使用文献[4,6]等相同的处理方式,每一幅图像转换为灰度图然后归一为 20×20 像素,并做直方图均衡化处理,直接使用灰度特征,因此每一幅图像向量化为一个 400 维的列向量。

表 1 列出了 10 次随机实验,不同算法的平均识

别率、方差、最高(最低)识别率。表 1 的实验结果显示,在 ETH80 数据库上,MSM、MMD、MDA、SANP 和 AHISD 算法取得基本一致的结果,而本文算法结果比这几个算法的结果要高接近 5%~9%,平均达到 91.5% 的识别率,10 次实验中最高达到 95%。这可能是由于数据库的结构和采用局部保持投影导致的结果。ETH80 数据库每类物体只包含 41 幅图像,图像虽然包含物体的不同视角,但是图像是非连续拍摄,因此 MDA、MMD 等算法中的流形分布假设不能得到很好的满足,另外 SANP 和 AHISD 都是基于近邻比较的方法,集合中可能存在的离群点对识别率会有很大的影响。

表 1 ETH80 数据库上各种算法识别率和方差

Table 1 Recognition rate and their standard deviations of different methods on ETH80 data set

识别率	MSM ^[1]	AHISD ^[2]	SANP ^[3]	MMD ^[4]	MDA ^[6]	GGDA ^[8]	本文
均值/%	86.25	83.5	84.5	82.5	82	83.57	91.5
方差	2.43	1.29	2.58	3.33	2.58	5.80	2.68
最大/%	90.0	85.00	87.5	90.00	85.0	95.0	95.0
最小/%	82.50	82.50	80.00	77.50	77.50	75.00	87.50

图 3 显示了 ETH80 数据库上 10 次实验各个算法的识别率,图 3 的最后一列显示了 10 次实验的平均识别率。从图 3 中可以看出,GGDA 算法的 10 次单独实验的识别率变化范围在 72.5%~95%,随机实验的结果表明算法对样本的选择依赖性较强,该算法基于近邻比较,较大的识别率变化范围可能是由于离群点的干扰所致。

在 YouTube 数据库测试了各个算法的识别性

能,对比结果如表 2 所示。本文算法取得最好的结果略高于 SANP 和 AHISD,但是比 MSM、MMD 和 MDA 算法识别率都要高出接近 10%,而 GGDA 算法在这个数据库上表现最差。所有算法在 YouTube 数据库上的识别率都比较低,主要因为这个数据库视频文件来自真实的环境,视频包含大范围的姿态、光照和表情变化,而且分辨率都非常低。

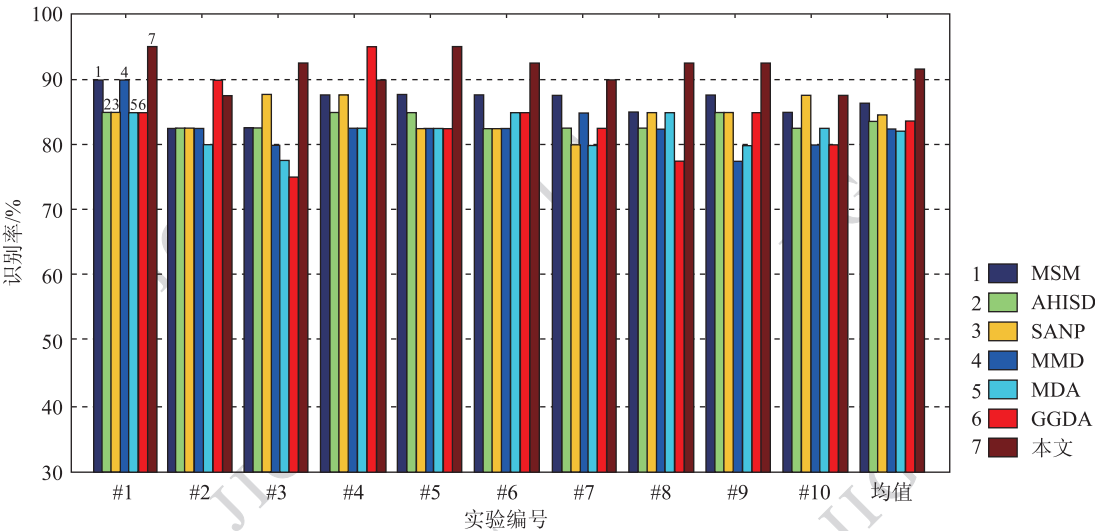


图 3 各种算法在 ETH80 数据库上运行 10 次的识别率对比

Fig. 3 Recognition rates of tested method for each trial on the ETH80 object dataset

表 2 YouTube 数据库上各种算法识别率和方差

Table 2 Recognition rate and standard deviations of different methods on YouTube data set

识别率	MSM ^[1]	AHISD ^[2]	SANP ^[3]	MMD ^[4]	MDA ^[6]	GGDA ^[8]	本文
均值/%	53.62	63.09	64.47	55.21	52.77	49.68	65.31
方差	0.55	1.01	1.21	1.17	1.68	2.70	1.05

表 3 ETH80 数据库上各种算法运行时间比较

Table 3 Average computation time of different methods on ETH80 data set

	MSM ^[1]	AHISD ^[2]	SANP ^[3]	MMD ^[4]	MDA ^[6]	GGDA ^[8]	本文
时间/s	1.85	14.63	6331	6.41	2.04	8.99	66.33

在一台 Intel(R) Xeon(R) E7-4807 双 1.87 GHz CPU(2×6 核), 64 GB 内存的服务器上测试了所有的算法。表 3 列出了个算法运行的时间, 结果显示, 本文算法低于 SANP, 但是高于其他算法, 这主要是由于在计算核矩阵时需要计算奇异值分解问题, 这一过程是非常耗时。

4 结 论

比较两个集合的一种途径是将图像集合嵌入到流形, 把图像集当成流形上的一个点。从黎曼流形的角度出发, 利正定对称的非负协方差矩阵构成黎曼几何上的流形, 通过计算图像集合的协方差矩阵, 有效地将图像集合转换为黎曼流形上的点。传统的学习算法不能直接用于流形上点的分类, 为这个问题通过计算矩阵的对数, 构造一个黎曼核, 从而把黎曼流形上的点的匹配投影到欧几里德空间。进一步从样本分布的局部邻域结构出发, 利用保局投影技术刻画黎曼流形上样本点的局部几何关系, 使得在投影空间样本的邻域关系得以保持, 同时样本之间具有更好的可分性。

在基于图像集合的对象匹配和人脸识别任务上测试了本文算法, 并与经典的基于集合的匹配算法进行比较。需要特别指出的是, GGDA 算法提出了格拉斯曼流形上的图嵌入框架来保持样本分布的局部邻域关系, 并通过经典的子空间来建模图像集合, 学习基于典型相关的核将图像集合投影到核再生希尔伯特空间, 计算核函数的时候只用到了最大的主角, 因此可能丢失一些鉴别信息。本文则从黎曼流行的角度, 通过协方差学习利用核函数将图像集合投影到核再生希尔伯特空间, 由于协方差矩阵与图像集合的所有的样本有关, 因此一定程度上增加对

样本缺失和噪声的容错能力, 更具鲁棒性。在相同的实验环境下, 本文算法的识别率超越了经典的基于图像集合的匹配方法, 在 ETH80 和 YouTube Celebrities 数据库分别进行的对象识别和人脸识别实验中, 分别达到 91.5% 和 65.31% 的识别率。对象识别实验中, 比其他算法最好的要高出 5%, 比最差的要高出接近 9%, 具有明显的优势。

本文算法还是存在一些缺陷, 由于本文算法需要构造邻域图, 并且在计算核矩阵时需要计算奇异值分解问题, 因此在时间性能上与其他算法没有优势, 仅仅比 SANP 要好。另外在样本数量有限的情况下, 图像集合的协方差矩阵可能是奇异的, 通过增加正则项的方法避免求解奇异值分解时出现的问题。下一步将对其进行改进, 考虑构造更加高效的核函数, 进一步提高识别精度的同时提高算法运行的效率。

参考文献 (References)

- [1] Yamaguchi O, Fukui K, Maeda K. Face recognition using temporal image sequence[C]//Proceedings of IEEE Conference on Automatic Face and Gesture Recognition. Nara, Japan: IEEE Computer Society, 1998:318-323.
- [2] Cevikalp H, Triggs B. Face recognition based on image sets [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, CA, USA: IEEE Computer Society, 2010:2567-2573.
- [3] Hu Y Q, Mian A S, Owens R. Face recognition using sparse approximated nearest points between image sets[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(10):1992-2004.
- [4] Wang R P, Shan S G, Chen X L, et al. Manifold-manifold distance with application to face recognition based on image set [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, Alaska, USA: IEEE Computer

- Society, 2008:1-8.
- [5] Wang R P, Shan S G, Chen X L, et al. Maximal linear embedding for dimensionality reduction[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33:1776-1792.
- [6] Wang R P, Chen X L. Manifold discriminant analysis[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Miami, Florida, USA: IEEE Computer Society, 2009:429-436.
- [7] Wang R P, Guo H M, Davis L S, et al. Covariance discriminative learning: a natural and efficient approach to image set classification[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Providence, RI, USA: IEEE Computer Society, 2012:2496-2503.
- [8] Shirazi S, Alavi A, Harandi M T, et al. Graph embedding discriminant analysis on Grassmannian manifolds for improved image set matching[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Colorado Springs, CO, USA: IEEE Computer Society, 2011:2705-2712.
- [9] Yan S C, Xu D, Zhang B Y, et al. Graph embedding and extensions: a general framework for dimensionality reduction[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29:40-51.
- [10] Jayasumana S, Hartley R, Salzmann M, et al. Kernel methods on the riemannian manifold of symmetric positive definite matrices[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Portland, USA: IEEE Computer Society, 2013.
- [11] Arsigny V, Fillard P, Pennec X, et al. Geometric means in a novel vector space structure on symmetric positive-definite matrices[J]. SIAM Journal on Matrix Analysis and Applications, 2007, 29(1):328-347.
- [12] He X F, Niyogi P. Locality preserving projections[C]//Neural Information Processing Systems. Vancouver and Whistler, British Columbia, Canada: MIT Press, 2003,16:153-161.
- [13] Kim M, Kumar S, Pavlovic V, et al. Face tracking and recognition with visual constraints in real-world videos[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, Alaska, USA: IEEE Computer Society, 2008:1-8.
- [14] Ross D, Lim J, Yang M H. Adaptive probabilistic visual tracking with incremental subspace update[C]//Proceedings of European Conference on Computer Vision. Prague, Czech Republic: Springer, 2004:470-482.