

基于 Grassmann 流形几何深度网络的图像集识别方法

刘天赐^{1,2,3}, 史泽林^{1,3}, 刘云鹏^{1,3}, 张英迪^{1,2,3}

(1. 中国科学院沈阳自动化研究所, 辽宁 沈阳 110016;

2. 中国科学院大学, 北京 100049;

3. 中国科学院光电信息处理重点实验室, 辽宁 沈阳 110016)

摘要:近年来,深度学习以其强大的非线性计算能力在目标检测和识别任务中取得了巨大的突破。现有的深度学习网络几乎都是以数据的欧氏结构为前提,而在计算机视觉中许多数据都具有严格的流形结构,如图像集可表示为 Grassmann 流形。基于数据的流形几何结构来设计深度学习网络,将微分几何理论与深度学习理论相结合,提出一种基于 Grassmann 流形的深度图像集识别网络。同时在模型训练过程中,使用基于矩阵链式法则的反向传播算法来更新模型,并将权值的优化过程转换为 Grassmann 流形上的黎曼优化问题。实验结果表明:该方法不仅在结果上识别准确率得到了提高,同时在训练和测试速度上也有一个数量级的提升。

关键词:深度学习; Grassmann 流形; 黎曼优化; 图像集识别

中图分类号: TP391 **文献标志码:** A **DOI:** 10.3788/IRLA201847.0703002

Geometry deep network image-set recognition method based on Grassmann manifolds

Liu Tianci^{1,2,3}, Shi Zelin^{1,3}, Liu Yunpeng^{1,3}, Zhang Yingdi^{1,2,3}

(1. Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China;

2. University of Chinese Academy of Sciences, Beijing 100049, China;

3. Key Laboratory of Opto-Electronic Information Processing, Chinese Academy of Sciences, Shenyang 110016, China)

Abstract: In recent years, deep learning techniques have achieved great breakthrough for its powerful non-linear computations in the tasks of target recognition and detection. Existing deep networks were almost designed based on the precondition that the visual data reside on the Euclidean space. However, many data in computer vision have rigorous geometry of manifolds, i.e., image sets can be represented as Grassmann manifolds. The deep network was devised based on the non-Euclidean structure of the manifold-valued data, which combined the differential geometry and deep learning methods theoretically. Furthermore, a deep network for image-set recognition based on the Grassmann manifold was proposed. In the training process, the model was updated by the use of the backpropagation algorithm derived from the matrix chain rule. Learning of the weights can be transformed as the Riemannian optimization problem

收稿日期: 2018-02-10; 修订日期: 2018-03-20

基金项目: 中国科学院重点创新基金(Y6K4250401)

作者简介: 刘天赐(1993-), 男, 博士生, 主要从事目标检测与目标识别方面的研究。Email: liutianci@sia.cn

导师简介: 史泽林(1965-), 男, 研究员, 博士生导师, 主要从事光电成像、光电跟踪、图像处理、目标识别方面的研究。Email: zlshi@sia.cn

刘云鹏(1980-), 男, 研究员, 硕士生导师, 主要从事光电跟踪、图像处理、流形学习、目标识别方面的研究。Email: ypliu@sia.cn

on the Grassmannian. The experimental results show that this method not only improves the accuracy of recognition, but also accelerates the training and test process in one magnitude.

Key words: deep learning; Grassmann manifolds; Riemannian optimization; image-set recognition

0 引言

近年来,随着大数据时代的来临,以大规模训练样本库为基础的深度学习方法得到了飞速发展,使得其在语义识别、自然语言处理和计算机视觉等领域都取得了重大突破,尤其是卷积神经网络(CNN)在目标检测与识别等各种图像任务中取得了前所未有的表现^[1-4]。深度学习以其强大的学习能力超越了许多传统学习方法,这主要得益于它强大的非线性计算能力和以反向传播方法为基础的训练策略。

尽管取得了这样的成果,但现有的深度学习研究主要集中在欧氏数据上,比如一维的信号数据,二维的图像数据或者三维的视频数据。但是在自然条件下,很多数据实际上具有非欧结构,利用流形处理数据已经取得了广泛的关注和理想的结果^[5-7]。在计算机视觉领域,很多处理的数据都具有非欧结构,它们的几何空间可看做是一个流形。例如,三维旋转矩阵形成了SO(3)群,线性子空间(图像集)具有Grassmann流形结构等。对于这些重要的流形数据来说,如何将深度学习方法推广至非欧结构以发挥其在识别上的优越性,这一研究近年来得到了广泛关注^[8-9]。

在图像集(视频)识别问题中(如动态表情识别,动作识别等),用线性子空间对目标图像集进行建模,在计算机视觉领域中取得了巨大成功^[10-13]。线性子空间在数学上具有Grassmann流形结构,它的优势不仅在于与传统图像表示方法相比的低计算成本,还在于其在光照、姿态等参数上的不变性^[12,14]。因此,如何在Grassmann流形上进行有效地深度学习,充分利用深度学习的强大能力,进而获得更好的图像集识别结果,是目前值得研究的问题。针对这一问题,如何将深度学习与流形几何结构相结合,进一步将深度学习拓展到非欧结构,使得其获得学习上的加速性和更高的准确性,结合图像集的流形结构来对深度学习进行目标识别研究具有重

要的理论价值和工程实际意义,并且该问题在计算机领域的顶级会议上ECCV、ICCV、CVPR等,均成立了专题workshop,且得到了广泛关注。

文中以数据的非欧结构为前提,并将微分流形的几何原理作为理论依据,设计出一种适用于非欧数据结构的深度学习网络。该网络模型与现有的流行算法相比,网络结构相对简单并且更加有效。实验证明,该模型在大规模的图像集识别任务中,不仅在精度上更加准确,而且在学习速度和测试速度上都有近一个数量级的大幅度提升。

1 基于流形结构的深度学习网络

文中首先简单介绍流形在数学意义上的基本概念。然后具体介绍所设计的深度网络结构,并对每个网络层的功能和原理做详细的描述。最后对非欧结构下的反向传播求导规则和流形优化方法进行推导,并给出其参数的更新形式。

1.1 黎曼流形及其度量

流形是局部与欧氏空间同胚的拓扑空间,其上的点与点之间是由光滑的曲线相连接的。与欧氏空间中的距离定义不同,流形上的距离定义为流形上两点之间的最短曲线的长度。具有最短长度的曲线称为测地线,该曲线的长度则称为测地线距离。

定义 1. Grassmann 流形: D 维欧氏空间 n 中的所有 n 维子空间集合构成了 Grassmann 流形 $G(n, D)$ ($0 \leq n \leq D$), 即所有的 $D \times n$ 维正交矩阵, 其 n 列正交列在正交群 $O(n)$ 下所构成的商空间:

$$G(n, D) \triangleq \{X \in R^{D \times n} : X^T X = I_n\} / O(n) \quad (1)$$

Grassmann 流形 $G(n, D)$ 是 D 维欧氏空间中的 n 维线性子空间的集合, 是一种 $n(D-n)$ 维的紧致的黎曼流形。每个 Grassmann 流形上的点都是一个线性子空间, 可表示为 $\text{span}(X)$ 。该线性子空间是由 $D \times n$ 维矩阵 X 的正交基所构成的。在 Grassmann 流形上, 任意两点间的测地线距离可以由下列公式计算:

$$d_g(p,q)=\|\Theta\|_2 \quad (2)$$

式中： p,q 为 Grassmann 流形上的任意两点； $\Theta=(\theta_1,\theta_2,\theta_k)$ 为该两点间的主角度向量，如图 1 所示。

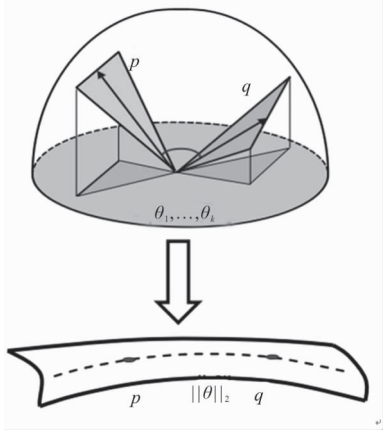


图 1 Grassmann 流形上的主角度示意图
Fig.1 Diagram of the principal angles on the Grassmann manifold

1.2 网络结构

文中将详细介绍所设计的深度网络结构。由于 Grassmann 流形主要是用来表示图像集或视频，所以该网络主要是针对于解决图像集的分类识别问题。笔者设计网络的主要目的就是利用深度学习的优点，学习出维数更低的、分类能力更强的 Grassmann 流形，使得在相关分类任务上不仅降低了计算复杂度，而且同时也提高了识别的精度。该网络总共由 6 层结构组成，结构图如图 2 所示。上述网络主要由转换层、正交化层、映射层、全连接层和 softmax 层组成。

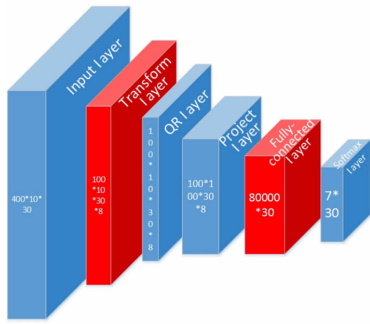


图 2 所提出的网络模型结构示意图

Fig.2 Schematic diagram of the proposed network model

1.2.1 转换层

该层是本网的核心层，其主要作用是将输入

的正交矩阵(Grassmann 流形)通过线性映射 f 转换到一个新的低维矩阵中：

$$X_{k+1}=f_{\text{trans}}^k(X_k;W_k)=W_k^T X_k \quad (3)$$

式中： $X_k \in G(n,D_k)$ 是第 k 层输入的正交矩阵； $W_k \in R^{D_{k+1} \times D_k}$ ， $D_{k+1} < D_k$ 为第 k 层的权值，即为转换矩阵； $X_{k+1} \in R^{D_{k+1} \times n}$ 为第 k 层的输出矩阵。由于 W_k 需要是列满秩的，为了防止优化时的衰减现象，给 W 矩阵赋予一个正交限制，即：

$$W_k^T W_k = I_{D_k} \quad (4)$$

式中： I_{D_k} 为 D_k 维的单位阵。

在数学观点上，带有正交限制的优化问题实际上是一个在 Stiefel 流形上的非限制性优化问题。所以， W 的正交特性使得其几何结构处在 Stiefel 流形上。那么，笔者就可以使用黎曼优化方法对 W 在 Stiefel 流形上进行梯度下降法求解。而且，笔者的目标函数对正交群 $O(n)$ 具有不变性，即对于任意的 $H \in O(n)$ ，都有 $L(W)=L(WH)$ ，所以 W 的优化空间实际上在一个 Grassmann 流形上。

1.2.2 正交化层

在转换层之后，一般都需要紧接一个正交化层。这是因为，即使 W 是正交矩阵， $W^T X$ 也无法保证一定在 Grassmann 流形上。以上结论显然成立，由 $(W^T X)^T (W^T X) = X^T W W^T X$ 可得， $W W^T$ 并不一定是单位阵。所以，只有由正交基所构成的线性子空间矩阵才能形成 Grassmann 流形。为了解决这一问题，使用 QR 分解来得到 $W^T X$ 的正交成分，即：

$$W^T X = QR \quad (5)$$

具体地说，在正交层中，有：

$$X_{k+1} = f_{QR}^k(X_k) = X_k R_k^{-1} = Q_k \quad (6)$$

式中： X_k 为第 k 层的输入； $Q_k \in R^{D_k \times n}$ 为由前 n 列组成的正交矩阵； $R_k \in R^{n \times n}$ 为可逆的上三角矩阵； X_{k+1} 为第 k 层的输出。

1.2.3 映射层

该层的主要作用是将流形的非欧空间结构映射到欧氏空间中，这样适用于欧氏空间结构的一些网络层，如 softmax 层，就可以应用到所设计网络

的分类任务中,以此来训练目标函数。通过微分同胚映射将 Grassmann 流形嵌入到幂等的对称矩阵空间中,即

$$X_{k+1} = f_{pro}^k(X_k) = X_k X_k^T \quad (7)$$

式中: X_k 和 X_{k+1} 是第 k 层和 $k+1$ 层的输出。

1.3 网络的参数更新与求导

文中主要基于矩阵的链式法则和反向传播方法,对所设计网络的参数更新公式进行相关推导。由于网络中的模型大多数是复杂的矩阵函数形式,其中如正交层还包含 QR 分解等矩阵因子分解运算,所以简单地使用的以元素为单位的梯度推导方法无法对网络的参数进行更新。也就是说,以矩阵为单位的反向传播方法的梯度推导,无法仅仅利用传统的以矩阵中的元素为单位的反向传播方法推导出梯度。文中使用基于矩阵的反向传播方法和矩阵链式法则^[8]来解决这一难题。

在深度网络模型中,笔者用一系列连续的函数组合来表示其学习到的函数 $f = f^{(k)} \circ f^{(k-1)} \circ f^{(k-2)} \circ \dots \circ f^{(2)} \circ f^{(1)}$, 函数中的参数为 $\{W\} = \{W_k, W_{k-1}, W_{k-2}, \dots, W_2, W_1\}$, 其中 $f^{(k)}$ 和 W_k 分别是第 k 层的函数和权值矩阵。那么,第 k 层的代价函数就可以表示为 $L^{(k)} = L \circ f^{(n)} \circ f^{(n-1)} \circ \dots \circ f^{(k+1)} \circ f^{(k)}$, 其中 L 是最后一层的输出函数, n 为网络的总层数。

1.3.1 转换层的参数更新

将使用 Grassmann 流形上的随机梯度下降法来更新转换层中的权值参数。具体地,对于 $R_w L(W)$ 的计算,可以由以下公式得到:

$$R_w L(W) = (I_D - WW^T) \nabla_w L(W) \quad (8)$$

式中: $\nabla_w L(W)$ 为 $L(W)$ 关于 W 的欧氏梯度。

该层在黎曼流形上的权值更新公式为:

$$R_w L(W) = \nabla_w L(W) - \nabla_w L(W) WW^T \quad (9)$$

$$W^{t+1} = \Gamma(W^t - \lambda R_w L(W)) \quad (10)$$

式中: W^t 为当前的权值参数; $\Gamma(\cdot)$ 为拉回映射算子; λ 为学习速率; $\nabla_w L(W) WW^T$ 为函数欧氏梯度 $\nabla_w L(W)$ 的正则成分。

根据链式法则可以求得相应的欧氏梯度 $\nabla_w L(W)$

的表达式为:

$$\nabla_w L(W) = \frac{\partial L^{(k+1)}}{\partial X_{k+1}} \frac{\partial f^{(k)}(X_k)}{\partial W} = \frac{\partial L^{(k+1)}}{\partial X_{k+1}} X_k^T \quad (11)$$

1.3.2 正交层的参数更新

正交层的计算是所有层中最复杂的一层,也是推导公式最为困难的一层。该层主要包含了公式(5)的 QR 分解,同时还包含了公式(6)的非线性计算。首先对 QR 分解层的梯度公式进行推导。结合参考文献[9],对于将 X_k 作为输入,二元组 (Q,R) 作为输出的 QR 层,其代价函数的偏导数为:

$$\begin{aligned} \frac{\partial L^{(k)} \cdot f}{\partial X_k} = & \left((I - QQ^T)^T \frac{\partial L^{(k)}}{\partial Q} + Q \left(Q^T \frac{\partial L^{(k)}}{\partial Q} \right)_{\text{btril}} \right) (R^{-1})^T \\ & + Q \left(\frac{\partial L^{(k)}}{\partial R} \left(\frac{\partial L^{(k)}}{\partial R} R^T \right)_{\text{btril}} (R^{-1})^T \right) \end{aligned} \quad (12)$$

式中: $A_{\text{btril}} = A_{\text{btril}} - (A)_{\text{tril}}^T$, A_{tril} 为将矩阵 A 的所有上三角阵元素设为 0。

1.3.3 映射层的参数更新

相对其他层而言,映射层的梯度更新公式比较简单,但是该层却在整个网络中起到了至关重要的作用。该层可视为流形结构的线性化层,它把 Grassmann 流形的非欧结构同胚映射到对称矩阵空间的欧氏结构中,使得之后 softmax 层的代价函数训练方法可以直接适用于笔者的网络模型。虽然该层的传递过程中并没有权值参数需要训练,但是其依然参与正向传播和反向传播时的梯度更新。可求得该层的梯度更新公式为:

$$\frac{\partial L \cdot f}{\partial X_k} = \frac{\partial L}{\partial X_{k+1}} + \left(\frac{\partial L}{\partial X_{k+1}} \right)^T \quad (13)$$

2 实验结果与分析

文中通过野外动态人脸表情(AFEW)数据集^[10]来验证所提网络模型的有效性。该数据集包含了在真实野外条件下人脸面部表情的动态变化数据,是现有的人脸表情识别任务中被广泛应用的一个流行数据集。它由 330 个人拍摄的 1 345 个视频组成,这些视频是在自然环境下拍摄的,而非实验室内的可控环境下,其样本见下文。这就使得基于该数据集的人脸表情识别任务更加具有挑战性。实验

结果表明, 笔者的方法在准确度和速度上都达到了理想的结果。

2.1 实验设置

在相关实验中, 采用大量的对比算法来进行实验结果的分析与验证。首先, 选择了 4 种当下比较先进的 Grassmann 流形学习算法: DCC 方法^[15]、GDA 方法^[10]、GGDA 方法^[11]和 PML 方法^[12]。在对比实验中, 不仅选取了 4 种 Grassmann 流形学习算法^[15, 10-12], 而且还选取了与这个数据库相关的当下流行算法, 如 STM-ExpLet^[6], DeepO2P^[17], SPDNet^[8]和 GrNet^[9]。对于以上方法, 全部使用相关作者提供的源代码, 并且根据原文的策略来调整各算法中的相关参数。

如图 3 所示, AFEW 数据集的标准实验策略^[14]

是把原数据集分成训练数据集、验证数据集和测试数据集共 3 个子数据集, 在训练数据集和验证数据集中, 每个视频都被分为 7 种表情中的一种, 但是其测试数据集的标签并没有在网上公布, 所以无法直接使用其测试数据集来进行实验。基于这个原因, 笔者将验证数据集用作测试数据集来进行相关的实验分析。对于训练数据集, 将其中的视频分成 1 747 个视频片段, 而对于测试数据集, 则相应地被分成 371 个视频片段。这些视频片段中的每一帧人脸图像都被重新归一化为 20×20 尺度大小的图像。与参考文献[15]的方法类似, 将每个视频数据表示为线性子空间, 即每一个视频数据都被表示为一个 G(10,400) 的 Grassmann 流形矩阵。在训练过程中, 学习速率设置为 0.01, batch size 设为 30。



图 3 AFEW 数据库中的样本帧

Fig.3 Sample frames from the AFEW database

2.2 实验结果及分析

各种不同算法在 AFEW 数据集上的实验结果如表 1 所示。可以看出, SPDNet 和 GrNet 方法的结果在该数据集上取得了比较精确的结果, 达到了 34.23%。但是笔者的方法在所有算法中取得了最为准确的结果, 将原有的领先结果提升至 36.39%。图 4 给出了该网络在训练与测试中的收敛表现和

识别错误率。

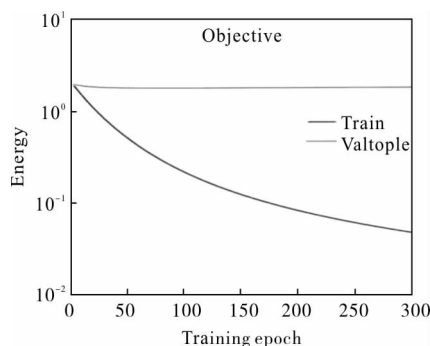
表 2 给出了所设计的网络模型在训练阶段和测试阶段与其他网络模型在运行速度上的比较结果。文中方法在训练速度和测试速度上, 较 SPDNet 和 GrNet 都有大幅度提升。其中, 与 GrNet 相比, 笔者的方法在训练速度上提升了 13 倍, 测试速度上提升了 31 倍; 而与 SPDNet 相比, 笔者的方法在训

练速度上提升了 4.4 倍,测试速度上提升了 5.3 倍。这个结果也是合理的,因为笔者的网络结构只有 6 层,而 GrNet 有 14 层,SPDNet 有 9 层。

表 1 AFEW 数据集上的表情识别结果

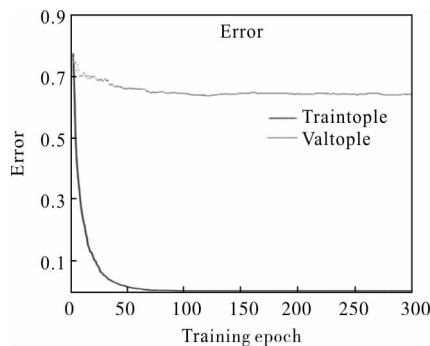
Tab.1 Emotion recognition results on the AFEW dataset

Methods	Recognition accuracies
DCC	25.78%
GDA	29.11%
GGDA	29.45%
PML	28.98%
DeepO2P	28.54%
STM-ExpLet	31.73%
SPDNet	34.23%
GrNet	34.23%
Proposed method	36.39%



(a) 目标函数的收敛结果

(a) Convergence behavior of the objective function



(b) 训练与测试中的识别错误率

(b) Error rates of the training and testing datasets

图 4 文中模型在 AFEW 数据库上的实验结果

Fig.4 Experimental results of this model

for the AFEW database

表 2 训练和测试速度结果比较

Tab.2 Comparison of running speed of the training and test data sets

Methods	Training running speed	Testing running speed
SPDNet	29.2 data/s (59.83 s)	82.6 data/s (4.53 s)
GrNet	9.5 data/s (183.89 s)	13.8 data/s (26.88 s)
Our method	127.7 data/s (13.68 s)	433.9 data/s (0.86 s)

3 结论

文中针对图像集 的识别与分类问题,基于数据的 Grassmann 流形结构,设计了一种简单且高效的深度流形几何网络。在尊重数据几何结构的同时,结合深度学习方法的强大计算能力,提出了一种面向非欧数据的深度学习方 法,使得 Grassmann 流形的学习效果在深度学习框架下得到了优化。

为了验证该方法的有效性,在 AFEW 数据集上做了相关实验,并对实验结果和网络模型进行了详细的分析。实验结果表明,所提出的网络模型不仅在准确性上较现有的先进算法有所提升,而且在学习和测试速度上也得到了大幅度提升。文中为之后将深度学习与流形结构结合的工作在理论上和方法上都打下了良好的基础。

参考文献:

[1] Luo Haibo, Xu Lingyun, Hui Bin, et al. Status and prospect of target tracking based on deep learning [J]. *Infrared and Laser Engineering*, 2017, 46(5): 0502002. (in Chinese)
罗海波, 许凌云, 惠斌, 等. 基于深度学习的目标跟踪方法研究现状与展望 [J]. 红外与激光工程, 2017, 46(5): 0502002.

[2] Luo Haibo, He Miao, Hui Bin, et al. Pedestrian detection algorithm based on dual-model fused fully convolutional networks(Invited)[J]. *Infrared and Laser Engineering*, 2018, 47(2): 0203001. (in Chinese)
罗海波, 何森, 惠斌, 等. 基于双模全卷积网络的行人检测算法(特邀)[J]. 红外与激光工程, 2018, 47(2): 0203001.

[3] Pei Xiaomin, Fan Huijie, Tang Yandong. Action recognition method of spatio-temporal feature fusion deep learning network [J]. *Infrared and Laser Engineering*, 2018, 47(2): 0203007. (in Chinese)
裴晓敏, 范慧杰, 唐延东. 时空特征融合深度学习网络人

- 体行为识别方法 [J]. 红外与激光工程, 2018, 47(2): 0203007.
- [4] Ke Yuyang, Yang Xunzheng, Xiong Yan, et al. Power generation dispatching for environmental protection based on recursive neural network and ant colony optimization algorithm [J]. *Information and Control*, 2017, 46(4): 415–421. (in Chinese)
柯余洋, 杨训政, 熊焰, 等. 基于递归神经网络和蚁群优化算法的发电环保调度[J]. 信息与控制, 2017, 46(4): 415–421.
- [5] Wu Peiliang, Li Ya'nan, Yang Fang, et al. A CLM-based method of indoor affordance areas classification for service robots [J]. *Robot*, 2018, 40(2): 188–194. (in Chinese)
吴培良, 李亚南, 杨芳, 等. 一种基于 CLM 的服务机器人室内功能区分类方法[J]. 机器人, 2018, 40(2): 188–194.
- [6] Xu Peng, Zhao Dongbiao, Cheng Jinxiang, et al. Multi-objective optimization for inverse kinematics solution manifolds of redundant robots [J]. *Robot*, 2016, 38(6): 704–710. (in Chinese)
徐朋, 赵东标, 程锦翔, 等. 冗余机器人逆运动学解流形的多目标优化[J]. 机器人, 2016, 38(6): 704–710.
- [7] Xu Tao, Jia Songmin, Zhang Guoliang. Fast spatial object location method for service robot based on co-saliency [J]. *Robot*, 2017, 39(3): 307–315. (in Chinese)
徐涛, 贾松敏, 张国梁. 基于协同显著性的服务机器人空间物体快速定位方法[J]. 机器人, 2017, 39(3): 307–315.
- [8] Huang Z, Gool L V. A Riemannian network for SPD matrix learning [J]. *Association for the Advancement of Artificial Intelligence*, 2017, 2(4): 2036–2042.
- [9] Huang Z, Wu J, Gool L V. Building deep networks on Grassmann manifolds [J]. *Association for the Advancement of Artificial Intelligence*, 2018: 3279–3286.
- [10] Huang Z, Wang R, Shan S, et al. Projection metric learning on Grassmann manifold with application to video based face recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 140–149.
- [11] Harandi M, Sanderson C, Shirazi S, et al. Graph embedding discriminant analysis on grassmannian manifolds for improved image set matching [C]//IEEE Conference on Computer Vision and Pattern Recognition, 2011: 2705–2712.
- [12] Hamm J, Lee D D. Grassmann discriminant analysis: a unifying view on subspace-based learning [C]// Proceedings of the 25th international conference on Machine learning, ACM, 2008: 376–383.
- [13] Liu T, Shi Z, Liu Y. Kernel sparse representation on grassmann manifolds for visual clustering [J]. *Optical Engineering*, 2018, 57(5): 053104.
- [14] Dhall A, Goecke R, Joshi R, et al. Emotion recognition in the wild challenge 2014: Baseline, data and protocol [C]// Proceedings of the 16th International Conference on Multimodal Interaction, ACM, 2014: 461–466.
- [15] Kim T K, Kittler J, Cipolla R. Discriminative learning and recognition of image set classes using canonical correlations [J]. *IEEE T-PAMI*, 2007, 29(6): 1005–1018.
- [16] Liu M, Shan S, Wang R, et al. Learning expressionlets on spatio-temporal manifold for dynamic facial expression recognition [C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014: 4321–4328.
- [17] Ionescu C, Vantzos O, Sminchisescu C. Matrix backpropagation for deep networks with structured layers [C]// Proceedings of the IEEE International Conference on Computer Vision, 2015: 2965–2973.