



基于 LRC-SNN 的图像高效重建与识别

索 静¹, 宋林林², 李 强²

(1. 太原工业学院 电子工程系, 太原 030000; 2. 太原理工大学 信息与计算机学院, 山西 晋中 030600)

摘 要: 图像集分类算法种类较多,但多数存在运算繁琐、计算成本高和时效性差的问题。为此,提出一种改进的图像重建与识别算法,利用线性回归分类和共享最近邻子空间分类理论进行图像重建和分类,通过将图像下采样建立的高维空间重建为子空间,避免计算复杂度较高的训练过程。利用各个类别的图像集子空间对测试图像进行回归模型估计,根据回归模型重建测试集中的图像,基于重建图像和原始图像间重建误差最小化法,采用加权投票策略对测试集进行估计以确定图像所属的类别。在 UCSD/Honda、CMU、ETH-8 和 YouTube 数据集上进行实验,结果表明,在低分辨率采样条件下,与 ADNT 算法相比,该算法平均分类精度提高 3.6%,运算效率提高 10 倍,其最快响应时间缩短至 2.8 ms。

关键词: 图像集分类; LRC-SNN 回归模型; 误差最小化; 加权投票策略; 分类精度; 计算速度

开放科学(资源服务)标志码(OSID):



中文引用格式: 索静,宋林林,李强. 基于 LRC-SNN 的图像高效重建与识别[J]. 计算机工程, 2020, 46(7): 243-250, 259.

英文引用格式: SUO Jing, SONG Linlin, LI Qiang. Efficient image reconstruction and recognition based on LRC-SNN[J]. Computer Engineering, 2020, 46(7): 243-250, 259.

Efficient Image Reconstruction and Recognition Based on LRC-SNN

SUO Jing¹, SONG Linlin², LI Qiang²

(1. Department of Electronic Engineering, Taiyuan Institute of Technology, Taiyuan 030000, China;

2. College of Information and Computer, Taiyuan University of Technology, Jinzhong, Shanxi 030600, China)

[Abstract] Most of existing classification methods for image sets are costly, having high computational complexity and poor timeliness. To address the problem, this paper proposes an improved image reconstruction and recognition algorithm. The algorithm uses the Linear Regression Classification (LRC) and Share Nearest Neighbor (SNN) subspace classification theory for image reconstruction and classification. The high-dimensional space built by image subsampling is taken as subspace to avoid the training process with high computational complexity. Then, subspace of different categories of image sets is used to implement regression model estimation for test images. For images in the test set of regression model reconstruction, their categories are determined by using the weighted voting strategy to estimate the test set under the principle that the errors between reconstructed images and original images should be minimized. Experimental results on UCSD/Honda, CMU, ETH-8 and YouTube datasets show that under low-resolution sampling conditions, compared with the ADNT algorithm, the proposed algorithm increases the average classification accuracy by 3.6%, computational efficiency by 10 times, and shortens the fastest response time to 2.8 ms.

[Key words] image set classification; LRC-SNN regression model; error minimization; weighted voting strategy; classification accuracy; computational speed

DOI: 10.19678/j.issn.1000-3428.0055065

0 概述

图像集分类^[1]是指从多个图像中进行目标识别的过程,对于图像集分类问题而言,训练集由各个类

别中的一个或多个图像集构成,测试集包含具有相同主题的多个图像,分类算法通过某些相似度指标可对这些测试图像集与训练图像集进行匹配,进而确定测试图像的主题。与传统的基于目标识别的单

基金项目: 国家自然科学基金青年基金(61703298)。

作者简介: 索 静(1982—),女,讲师、硕士,主研方向为图像识别;宋林林,硕士研究生;李 强,讲师、博士。

收稿日期: 2019-05-30 **修回日期:** 2019-09-10 **E-mail:** 173387265@qq.com

图像分类算法相比,图像集分类算法具有多种优势,其可有效应对图像内的多种外观变化,如视角变化、遮挡、非刚性形变、光照变化及背景变化等。上述特性使得图像集分类技术在侦察、基于视频的脸部识别、摄像头网络人脸检测等领域得到广泛应用,引起了计算机视觉和模式识别领域相关学者的关注^[2]。本文利用线性回归分类(Linear Regression Classification, LRC)和共享最近邻(Share Nearest Neighbor, SNN)子空间分类理论进行图像重建和识别,并提出一种改进的基于 LRC-SNN 的图像集高效分类算法。

1 相关工作

近年来,图像分类算法成为图像处理领域的研究热点之一。文献[3]提出了一种圆周特征描述方法,该方法利用圆心在轮廓线上的圆、轮廓线、叶片形状区域两两相交所得到的 3 类叶片图像的圆周特征,通过改变圆的半径来产生由粗到细的圆周特征描述。文献[4]提出了一种基于深度卷积神经网络、可应用于大规模图像分类的深度神经网络,其在经典 AlexNet 的基础上分别从框架和内部结构 2 个方面对网络进行优化和改进。文献[5]通过在卷积结构单元中添加通道关注模块和空间关注模块,实现网络的混合关注,并提出一种递归深度混合关注网络方法。文献[6]提出一种融合 Zernike 矩全局特征和加速鲁棒性特征包 BoF-SURF 局部斑点特征的花粉图像分类识别算法。文献[7]研究深度卷积神经网络的发展及其在计算机视觉领域的应用。

图像集分类算法可分为有参数算法和无参数算法两大类。其中,有参数算法^[8-9]首先根据某种统计分布对图像集进行建模,然后计算这些分布间的相似度。然而,如果同一对象的测试图像集和训练图像集间的统计关联较弱,则参数估计值可能不够准确,导致算法的性能较差。

无参数算法利用多种指标来衡量图像集间的相似性,该类算法是目前图像集分类问题中的研究热点。其中,较为典型的算法有基于时域图像序列的脸部识别算法(TIS)^[10]、判别规范关联分析识别算法(DCC)^[11]、流形-流形距离识别算法(MMD)^[12]、流形判别分析算法(MDA)^[13]、仿射壳体图像集线性距离算法(AHISD)^[14]、凸壳体图像集距离算法(CHISD)^[15]、图形嵌入判别分析算法(GEDA)^[16]、稀疏近似最近邻点算法(SANP)^[17]、协方差判别学习算法(CDL)^[18]、正规化最近邻点算法(RNP)^[19]、平均序列稀疏表示分类算法(MSSRC)^[20]、集合-集合距离度量学习算法(SSDML)^[21]等。另外,图像集还可表示为线性子空间或复杂非线性流形的组合。对于线性子空间,往往利用某个子空间中的向量与另一子空间中向量间的最小角度余弦值来衡量 2 个图像集的相

似度。文献[22]将训练和测试图像集看作高维空间的子空间,提出一种对偶线性回归分类算法(DLRC)以进行图像集分类,该算法综合利用了各个图像集的最后一幅图像、其他被比较图像集的变化以及线性回归分类算法,从而确定 2 个子空间的距离。文献[23]对文献[22]的研究进行拓展,提出一种成对线性回归模型(PLRC)以进行图像集分类,该模型利用图像集的平均图像而非最后一幅图像来确定子空间的距离。然而,上述算法要求特征向量的维度远大于画廊图像集和测试图像集中图像的总数量,此外,进行图像集合表示时往往做出部分先验假设。在许多实际应用中,这些假设可能无法成立,尤其是当集合内部存在大量复杂的数据变化时。因此,上述算法只适用于小规模测试图像集。

文献[24]利用深度学习理论提出一种自适应深度网络模板(ADNT)算法。该算法利用一种深度自动编码器来定义训练集中各个类别的模型,并通过高斯受限玻尔兹曼器(GRBM)对自动编码器的权重进行初始化。在分类时,ADNT 利用学习过的类别模型重建测试集中的各幅图像,将重建误差作为指标来确定测试图像集。仿真结果表明,ADNT 算法的性能优于多种经典算法,但其依赖于人工 LBP 特征,并且需要对多个参数进行微调才能保证算法性能。另外,ADNT 算法的训练需要大量图像,计算成本很高。

本文提出一种改进的图像重建与识别算法。对各个画廊图像集进行下采样后,每个类别的画廊图像集形成高维空间中的一个子空间。在测试阶段,测试图像集中的每个测试图像表示为每个画廊图像集中图像的线性组合。利用最小二乘算法估计各个测试图像的回归模型参数,然后利用估计后的回归模型并结合画廊子空间对测试图像进行重建,真实测试图像和重建图像间的欧氏距离作为距离度量。在此基础上,每个测试图像对画廊中的各个类别进行投票,最后累积权重最高的类别作为图像类别。

2 问题描述

设 N 表示画廊集合 K_c 中某一类别 C 的图像数量。每个图像转换为灰度图并下采样到分辨率 $a \times b$ 后表示为 $k_c^n \in \mathbf{R}^{a \times b}$, 其中, $c = 1, 2, \dots, C$ 且 $n = 1, 2, \dots, N$ 。每个画廊图像通过列的拼接转化为向量且 $k_c^n \in \mathbf{R}^{a \times b} \rightarrow q_c^n \in \mathbf{R}^{T \times 1}$, 其中, $T = ab$ 。将类别 c 的多个图像向量沿水平方向进行拼接,可构建得到类别 c 的图像模型 Q_c :

$$Q_c = [q_c^1 q_c^2 \cdots q_c^N] \in \mathbf{R}^{T \times N}, c = 1, 2, \dots, C \quad (1)$$

通过上述方式可用称为回归量的向量子空间 Q_c 来表示类别 c 。回归量 Q_c 的每个向量 q_c^n ($n = 1, 2, \dots, N$) 横跨 $\mathbf{R}^{T \times 1}$ 子空间。本文研究的问题是已知测试图像集 Y_μ 包含 M 个图像且类别 $c = 1, 2, \dots, C$, 如何确定图像的未知类别 μ 。图 1 所示为本文算法结构。

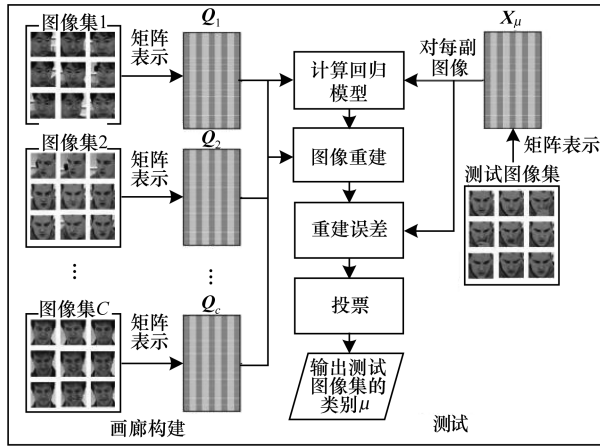


图1 本文算法结构

Fig.1 Structure of the proposed algorithm

3 图像识别算法

3.1 图像的子空间表示

与画廊图像类似,测试图像集的每个图像也转换为灰色图像,下采集为 $a \times b$,然后将图像表示为 $y_\mu^m \in \mathbf{R}^{a \times b}$,其中, μ 表示未知类别, $m=1,2,\dots,M$ 。通过列拼接将每个下采样图像转化为向量,且 $y_\mu^m \in \mathbf{R}^{a \times b} \rightarrow x_\mu^m \in \mathbf{R}^{T \times 1}$,其中, $T=ab$ 。图像向量 $x_\mu^m, m=1,2,\dots,M$ 通过水平拼接后生成测试矩阵 X_μ :

$$X_\mu = [x_\mu^1 x_\mu^2 \cdots x_\mu^M] \in \mathbf{R}^{T \times M} \quad (2)$$

如果 X_μ 属于第 c 个类别,则有可能将 X_μ 的图像向量表示为同一类别画廊图像的线性组合:

$$x_\mu^m = Q_c \gamma_c^m, m=1,2,\dots,M, c=1,2,\dots,C \quad (3)$$

其中, $\gamma_c^m \in \mathbf{R}^{N \times 1}$ 表示参数向量。

3.2 回归模型参数估计

为了使得式(3)具有唯一解,必须满足条件 $T \geq N$ 。如果该条件满足,则利用最小二乘法估计得到 γ_c^m :

$$\gamma_c^m = (Q_c' Q_c)^{-1} Q_c' x_\mu^m, m=1,2,\dots,M, c=1,2,\dots,C \quad (4)$$

其中, Q_c' 表示 Q_c 的转置。利用 γ_c^m 和 Q_c 可重建类别 c 的图像向量:

$$\hat{x}_c^m = Q_c \gamma_c^m, m=1,2,\dots,M, c=1,2,\dots,C \quad (5)$$

$$\hat{x}_c^m = Q_c (Q_c' Q_c)^{-1} Q_c' x_\mu^m \quad (6)$$

其中, \hat{x}_c^m 表示由 Q_c 获得的 x_μ^m 的重建图像向量,可将 \hat{x}_c^m 看成 x_μ^m 在第 c 个子空间上的投影。本文不是为了得到各个图像向量 x_μ^m 来求解式(6),而是将其表示为如下的矩阵形式从而提高计算速度:

$$X_\mu = Q_c \Gamma_c, c=1,2,\dots,C \quad (7)$$

其中, $\Gamma_c \in \mathbf{R}^{N \times M}$ 表示参数矩阵。利用最小二乘法估计即可计算得到 Γ_c :

$$\Gamma_c = (Q_c' Q_c)^{-1} Q_c' X_\mu, c=1,2,\dots,C \quad (8)$$

$$\hat{X}_c = Q_c \Gamma_c, c=1,2,\dots,C \quad (9)$$

$$\hat{X}_c = Q_c (Q_c' Q_c)^{-1} Q_c' X_\mu \quad (10)$$

其中, $\hat{X}_c \in \mathbf{R}^{T \times M}$ 表示根据 Q_c 获得的 X_μ 的重建图像向量矩阵。

3.3 加权投票策略

本文利用欧氏距离计算各个测试图像 x_μ^m 和重建图像 \hat{x}_c^m 间的重建误差,如下:

$$d_c^m = \|x_\mu^m - \hat{x}_c^m\|_2, c=1,2,\dots,C, m=1,2,\dots,M \quad (11)$$

根据重建图像和原始图像间重建误差最小化原则,采用加权投票策略对测试集进行估计。在加权投票阶段,测试图像集中的图像 m 对各个类别 c 进行投票,以确定测试图像集 X_μ 的类别。本文在 CMU、YouTube、UCSD/Honda 和 ETH-8 等数据集^[25]上对不同的投票策略进行统计分析后得出,采用欧氏距离的指数进行投票时效果最好。因此,图像 m 的投票权重 θ_c^m 定义为:

$$\theta_c^m = e^{-\alpha d_c^m}, c=1,2,\dots,C, m=1,2,\dots,M \quad (12)$$

其中, α 为常数。根据各个测试图像确定的类别 c 的累积权重可计算为:

$$\Theta_c = \sum_{m=1}^M \theta_c^m, c=1,2,\dots,C \quad (13)$$

如果类别 c 对应于根据测试图像集 X_μ 确定的所有图像 x_μ^m 的最大累积权重,则将该类别确定为测试图像集的类别,如下:

$$\mu = \arg \max_c (\Theta_c), c=1,2,\dots,C \quad (14)$$

本文提出的图像集分类算法具体描述如算法1所示。

算法1 图像集分类算法

输入 画廊图像集 $K_c, c=1,2,\dots,C$, 测试图像集 Y_μ

输出 测试图像集 Y_μ 的类别 μ

形成画廊:

For c in 1 to C do

For n in 1 to N do

$q_c^n \in \mathbf{R}^{T \times 1}$, 对图像下采样至 $a \times b$

向量化, $T=ab$

End

$Q_c \in \mathbf{R}^{T \times N} = [q_c^1 q_c^2 \cdots q_c^N]$

End

测试:

For m in 1 to M do

$x_\mu^m \in \mathbf{R}^{T \times 1}$, 图像下采样至 $a \times b$

向量化, $T=ab$

End

$X_\mu \in \mathbf{R}^{T \times M} = [x_\mu^1 x_\mu^2 \cdots x_\mu^M]$

For c in 1 to C do

For m in 1 to M do

$\gamma_c^m = (Q_c' Q_c)^{-1} Q_c' x_\mu^m$

$\hat{x}_c^m = Q_c \gamma_c^m$

$d_c^m = \|x_\mu^m - \hat{x}_c^m\|_2$

$\theta_c^m = e^{-\alpha d_c^m}$

End

```

 $\Theta_c = \sum_{m=1}^M \theta_c^m$ 
End
 $\mu = \operatorname{argmax}(\Theta_c)$ 

```

算法 1 流程如图 2 所示。

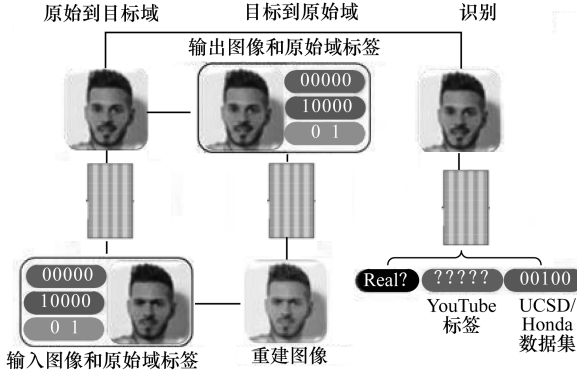


图 2 算法 1 流程

Fig. 2 Procedure of algorithm 1

3.4 奇异性问题

为了充分满足式(3)、式(7)的要求,经过下采样的画廊图像向量 q_c^n 的像素数量 $T = ab$ 必须大于或等于每个回归量 Q_c 中的画廊图像数量 N , 即 $T \geq N$ 。然而,即使满足上述要求,如果 Q_c 某些行与其他行存在线性关系,则回归量 Q_c 仍然可能奇异^[12]。此时, $r < T$, 称 Q_c 为缺秩回归量, r 表示 Q_c 的秩。因此,无法利用式(4)和式(8)计算参数向量 γ_c 或参数矩阵 Γ_c 。本文对此问题提出 2 种解决方案,分别为添加扰动与基于 QR 的分解。

3.4.1 扰动

通过添加微小扰动项对回归量 Q_c 进行正规化,便可解决 Q_c 的奇异问题。通过经验发现,添加矩阵 ϵ 并保证矩阵元素服从 $-0.5 \leq \epsilon \leq 0.5$ 随机均匀分布,可克服 Q_c 的奇异性。修正后的 Q_c 表示为:

$$Q_c^* = Q_c + \epsilon, \epsilon \in R^{T \times N}, -0.5 \leq \epsilon \leq 0.5 \quad (15)$$

式(15)需要在所有处理步骤之前实现,且矩阵 Q_c 的数值在 0 ~ 255 之间,因此,任何像素值的最大变化为 0.5,经过分析可知上述过程不会对分类精度产生影响。

3.4.2 基于 QR 分解的基本解

解向量的非 0 元素数量满足 $\gamma_c \leq r$, 因此,通过回归量 Q_c 的 QR 分解计算式(3)或式(7)的基本解便可解决奇异性问题,其中, r 表示回归量 Q_c 的秩。该方法所得的结果精度可满足本文图像重建的需要。

3.5 快速线性图像重建

与式(3) ~ 式(6)相比,使用式(7) ~ 式(10)可显著缩短算法处理时间。在形成画廊期间利用 Moore-Penrose 伪逆矩阵^[26]计算回归量 Q_c 的逆矩阵,可进一步缩短算法的处理时间。因此,本文将测试期间的计算过程分解为 2 次矩阵操作,如算法 2 所示。

算法 2 图像快速重建和分类算法

输入 画廊图像集 $K_c, c = 1, 2, \dots, C$, 测试图像集 Y_μ

输出 测试图像集 Y_μ 的类别 μ

形成画廊:

For c in 1 to C do

For n in 1 to N do

$q_c^n \in R^{T \times 1}$, 下采样图像至 $a \times b$

向量化, $T = ab$;

End

$Q_c \in R^{T \times N} = [q_c^1 q_c^2 \dots q_c^N]$

$\tilde{Q}_c = \text{pseudoinverse}(Q_c)$

End

测试:

For m in 1 to M do

$x_\mu^m \in R^{T \times 1}$, 图像下采样至 $a \times b$

向量化, $T = ab$;

End

$X_\mu \in R^{T \times M} = [x_\mu^1 x_\mu^2 \dots x_\mu^M]$

For c in 1 to C do

$\Gamma_c = \tilde{Q}_c X_\mu$

$\hat{X}_c = Q_c \Gamma_c$

$D_c = \|X_\mu^m - \hat{X}_c^m\|_2$

$\Theta_c = \sum_{m=1}^M e^{-\alpha D_c}$

End

$\mu = \operatorname{argmax}(\Theta_c)$

用 \tilde{Q}_c 表示画廊形成时计算的回归量 Q_c 的伪逆矩阵,测试期间式(7)可按如下方式求解:

$$\Gamma_c = \tilde{Q}_c X_\mu \quad (16)$$

$$\hat{X}_c = Q_c (\tilde{Q}_c X_\mu) \quad (17)$$

对于 ETH-8 数据集,快速线性图像重建算法的计算效率相对其他算法几乎提升了 2 倍。数据集规模越大,计算效率的提升幅度越明显。

4 仿真实验

本文在如下 4 个常见数据集上对算法的性能进行验证: CMU 人体识别数据集 (CMU MoBo), YouTube 名人数据集 (YTC), UCSD/Honda 人脸识别数据集, ETH-8 目标识别数据集。将本文算法与如下典型的图像集分类算法进行比较: 基于时域图像序列的脸部识别算法 (TIS)^[11], 判别规范关联分析识别算法 (DCC)^[12], 流形-流形距离识别算法 (MMD)^[13], 流形判别分析算法 (MDA)^[14], 仿射壳体图像集线性距离算法 (AHISD)^[15], 凸壳体图像集距离算法 (CHISD)^[15], 图形嵌入判别分析算法 (GEDA)^[16], 稀疏近似最近邻点算法 (SANP)^[17], 协方差判别学习算法 (CDL)^[18], 正规化最近邻点算法 (RNP)^[19], 平均序列稀疏表示分类算法 (MSSRC)^[20], 集合-集合距离度量学习算法

(SSDML)^[21], DLRC 算法^[22], PLRC 算法^[23], 自适应深度网络模板算法 (ADNT)^[24]。其中, ADNT 算法的分类精度最高, 本文采用文献[24]中的实验设置。为了便于比较, 本文给出已有文献提供的所有算法的平均识别结果。

4.1 CMU 人体识别数据集上的分类结果

CMU 人体识别数据集包含从 6 个不同视角拍摄的 25 个人体在跑步机上的运动视频, 除了最后一个人体外的其他所有人体均有不同运动模式拍摄的 4 个视频, 即缓慢走动、快速走动、弯腰走动及手中持球走动的视频。本文采用前 24 个人体的视频序列, 每个视频的多个帧组成一个图像集合。与文献[13, 15, 17, 24]类似, 随机选择各个人体的一种步态视频作为画廊图像集, 其余 3 种步态作为测试集。从每个画廊视频中随机选择 50 个帧, 利用文献[27]中的 Viola-Jones 脸部检测算法自动检测各个图像帧中的人脸区域。与文献[18]类似, 本文按 40×40 分辨率对图像再次采样, 并转换为灰度图像。利用直方图均衡化方法增加图像对比度。与文献[22-24]不同, 本文没有使用任何 LBP 特征, 而是对原始图像进行实验, 在式(12)中设置 $\alpha = 0.2$ 。重复 10 次实验, 每轮实验随机选择不同的画廊和测试图像集, 以增加本文测试环境的多样性。表 1 所示为不同算法对 CMU 人体识别数据集的平均分类精度和标准差。可以看出, 本文算法的精度远高于其他参数化算法和非参数化算法。

表 1 各种算法对 CMU-MoBo 人体识别数据集的平均分类精度和标准差

Table 1 Average classification accuracy and standard deviation of CMU-MoBo human body recognition dataset by various algorithms

| 算法 | 分类精度和标准差 |
|-----------------------|-------------------|
| TIS ^[5] | 93.81 \pm 1.97 |
| DCC ^[6] | 88.89 \pm 2.45 |
| MMD ^[7] | 92.50 \pm 2.87 |
| MDA ^[8] | 80.97 \pm 12.28 |
| AHISD ^[9] | 92.92 \pm 2.12 |
| CHISD ^[10] | 96.52 \pm 1.18 |
| GEDA | 84.86 \pm 3.24 |
| SANP | 92.64 \pm 0.94 |
| CDL | 90.00 \pm 4.38 |
| RNP | 96.11 \pm 1.43 |
| MSSRC | 97.50 \pm 0.88 |
| SSDML | 95.14 \pm 2.20 |
| DLRC | 97.60 \pm 2.78 |
| PLRC | 97.74 \pm 4.38 |
| ADNT | 97.92 \pm 0.73 |
| 本文算法 | 98.33 \pm 1.27 |

4.2 YouTube 名人数据集上的分类结果

YouTube 名人数据集包含由 47 位名流和政客构成的 1 910 个视频片断, 如图 3 所示, 其为目前规模较大的图像集分类数据集。从 YouTube 下载的这些真

实视频噪声较大、分辨率较低、录制视频时压缩率较高。与文献[13-15, 17, 26]类似, 本文实验采用 5 折交叉验证策略: 将数据集划分为 5 个部分, 保证不同部分的重叠交叉最小化。每部分包含 423 个视频片断, 每个人体为 9 个视频片断, 其中, 随机选择 3 个片断作为画廊视频集, 其他 6 个作为单独的测试集。采用文献[24]中的方法对所有被跟踪的人脸图像按分辨率 30×30 进行重采样, 并将其转换为灰度图像, 通过直方图均衡化方法提高图像的对比度。从每部分数据集中每个人体的 3 个画廊视频中随机选择 20 个图像作为画廊数据集, 如果某个画廊视频片断的视频帧数量少于 20 个, 则该视频的所有图像均作为画廊数据集, 则每个画廊数据集最多有 60 个图像, 在式(12)中设置 $\alpha = 10.5$ 。重复 10 实验, 每次实验的画廊图像、画廊图像集和测试图像集均随机选择且各不相同。表 2 所示为不同算法对 YouTube 数据集的平均分类精度和标准差。可以看出, 本文算法的精度高于其他所有参数化算法和非参数化算法, ADNT 算法分类精度也比较理想, 但本文算法使用的训练数据量远低于 ADNT 算法, 运行速度更快。此外, 本文算法无需对任何参数进行微调, 也不需要参数训练, 因此, 其更适用于真实应用场景。



图 3 YouTube 名人数据集

Fig.3 YouTube celebrity dataset

表 2 各种算法对 YouTube 数据集的平均分类精度和标准差

Table 2 Average classification accuracy and standard deviation of YouTube dataset by various algorithms

| 算法 | 分类精度和标准差 |
|-----------------------|------------------|
| TIS ^[5] | 50.21 \pm 3.59 |
| DCC ^[6] | 51.42 \pm 4.95 |
| MMD ^[7] | 54.04 \pm 3.69 |
| MDA ^[8] | 55.11 \pm 4.55 |
| AHISD ^[9] | 61.49 \pm 5.63 |
| CHISD ^[10] | 60.42 \pm 5.95 |
| GEDA | 52.48 \pm 4.45 |
| SANP | 65.60 \pm 5.57 |
| CDL | 56.38 \pm 5.31 |
| RNP | 65.82 \pm 5.39 |
| MSSRC | 59.36 \pm 5.70 |
| SSDML | 66.24 \pm 5.21 |
| DLRC | 69.55 \pm 5.16 |
| PLRC | 70.16 \pm 5.11 |
| ADNT | 71.35 \pm 4.83 |
| 本文算法 | 70.54 \pm 5.07 |

4.3 UCSD/Honda 数据集上的分类结果

UCSD/Honda 数据集包含 20 个人体的 59 个视频,每个人体有 1 个~5 个视频。该数据集的主要作用是作为标准的视频数据集,对各种人脸跟踪和识别算法进行性能评估,其所有视频均有严重的摇头和姿态变化,另外,部分视频序列存在局部遮挡现象。本次实验设置与文献[13-15,17,24]相同,利用文献[27]中的 Viola-Jones 人脸检测算法自动检测各个视频帧的人脸。与文献[24]类似,本文对检测出的人脸图像按分辨率 20×20 进行重采样并转换为灰度图像,通过直方图均衡化方法提高图像的对比度。将图像与均值图像相减然后与标准差相除,实现图像的标准化。从各个个体中随机选择一个视频作为画廊图像集,将其余视频作为测试图像集。为了保证画廊图像数量远低于像素点数,从各个画廊视频中随机选择少量视频帧,比如 50 个。在式(12)中设置 $\alpha = 0.2$ 。为了保证实验得分的稳定性,重复进行 10 实验,每次实验的画廊图像、画廊图像集和测试图像集均随机选择且各不相同。表 3 所示为不同算法对 UCSD/Honda 数据集的平均分类精度和标准差。从中可以看出,即使采用极少量的画廊图像,本文算法的分类精度也较高。

表 3 各种算法对 UCSD/Honda 数据集的平均分类精度和标准差
Table 3 Average classification accuracy and standard deviation of UCSD/Honda dataset by various algorithms

| 算法 | 分类精度与标准差 |
|-----------------------|---------------|
| TIS ^[5] | 88.21 ± 3.86 |
| DCC ^[6] | 92.56 ± 2.25 |
| MMD ^[7] | 92.05 ± 2.25 |
| MDA ^[8] | 94.36 ± 3.38 |
| AHISD ^[9] | 91.28 ± 1.79 |
| CHISD ^[10] | 93.62 ± 1.63 |
| GEDA | 91.28 ± 5.82 |
| SANP | 95.13 ± 3.07 |
| CDL | 94.97 ± 1.32 |
| RNP | 93.90 ± 2.16 |
| MSSRC | 93.95 ± 2.65 |
| SSDML | 86.41 ± 3.64 |
| DLRC | 96.31 ± 1.88 |
| PLRC | 98.74 ± 2.02 |
| ADNT | 100.00 ± 0.00 |
| 本文算法 | 100.00 ± 0.00 |

4.4 ETH-8 数据集上的分类结果

ETH-8 数据集由苹果、梨子、西红柿、奶牛、狗、马、轿车和茶杯 8 种对象组成,每种类别有 8 个不同的图像集,每个图像集包含从不同角度拍摄的 41 幅图像,如图 4 所示。本文实验只采用经过修剪后只包含对象且没有边缘区域的图像,采用文献[24]的实验设置,按照分辨率 32×32 对图像进行重采样,

并将其转换为灰度图像,通过与均值图像相减以及与标准差相除进行图像的标准化。与文献[13-15,17,26]类似,随机选择各个图像类别的 5 个图像集作为画廊图像集,其余 5 个图像集作为独立的测试图像集,在式(12)中设置 $\alpha = 0.2$ 。实验重复 10 次,每次实验均随机选择不同的画廊和测试图像集。表 4 所示为不同算法对 ETH-8 数据集的平均分类精度和标准差。从中可以看出,本文算法的性能与 ADNT 算法接近。

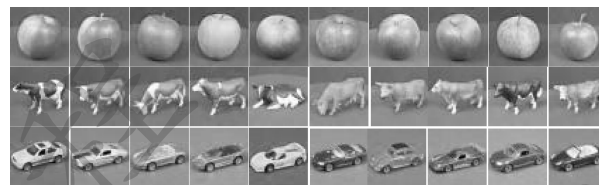


图 4 ETH-8 数据集

Fig. 4 ETH-8 dataset

表 4 各种算法对 ETH-8 数据集的平均分类精度和标准差
Table 4 Average classification accuracy and standard deviation of ETH-8 dataset by various algorithms

| 算法 | 分类精度与标准差 |
|-----------------------|--------------|
| TIS ^[5] | 75.50 ± 4.83 |
| DCC ^[6] | 91.75 ± 3.75 |
| MMD ^[7] | 77.50 ± 5.00 |
| MDA ^[8] | 77.25 ± 5.46 |
| AHISD ^[9] | 78.75 ± 5.30 |
| CHISD ^[10] | 79.53 ± 5.30 |
| GEDA | 79.50 ± 5.24 |
| SANP | 77.75 ± 7.31 |
| CDL | 77.75 ± 4.16 |
| RNP | 81.00 ± 3.16 |
| MSSRC | 90.50 ± 3.07 |
| SSDML | 81.00 ± 6.58 |
| DLRC | 95.11 ± 4.22 |
| PLRC | 96.30 ± 3.25 |
| ADNT | 98.12 ± 1.69 |
| 本文算法 | 97.75 ± 4.32 |

4.5 低分辨率条件下的实验结果

利用分辨率更低的图像进行实验,进一步评估本文算法与 ADNT 算法的性能。实验设置与 4.4 节相同,对于 ADNT 算法,采用文献[24]中相同的参数设置。实验重复 10 次,每次均随机选择不同的画廊和测试图像集。表 5 给出了 2 种算法的平均分类精度和标准差。从表 5 可以看出,对于 CMU 数据集,当图像分辨率降低时,本文算法的分类精度仍然有所提升。对于 UCSD/Honda 数据集,当图像分辨率降低时,本文算法的分类精度也非常理想。对于 ETH-8 数据集,图像分辨率为 20×20 时,本文算法

的分类精度最优。综上,虽然图像分辨率发生变化,但本文算法的分类精度没有受到显著影响。另外,对于分辨率较低的目标图像,本文算法的性能总是优于 ADNT 算法,这表明本文算法更适用于低分辨率应用场景。

表 5 2 种算法对低分辨率图像的平均分类精度和标准差
Table 5 Average classification accuracy and standard deviation of two algorithms for low resolution images

| 数据集 | 算法 | 文献[24]中的 分辨率 | 20×20 分辨率 | 15×15 分辨率 |
|----------------|---------|-----------------|--------------|--------------|
| CMU | ADNT 算法 | 97.92±0.73 | 91.81±2.40 | 90.56±3.13 |
| | 本文算法 | 98.33±1.27 | 98.75±1.38 | 99.31±1.18 |
| UCSD/ Honda | ADNT 算法 | 100.00±0.00 | 100.00±0.00 | 99.74±0.81 |
| | 本文算法 | 100.00±0.00 | 100.00±0.00 | 100.00±0.00 |
| ETH-8 | ADNT 算法 | 98.12±1.69 | 88.75±6.26 | 90.25±4.63 |
| | 本文算法 | 94.75±4.32 | 95.50±4.04 | 92.75±6.39 |

5 计算时间分析

以 ETH-8 数据集为实验对象,对所有算法在该数据集上的计算时间进行测试,表 6 所示为 CPU RAM 条件下各算法所需的训练时间和每个图像进行分类所需的测试时间的比较结果,其中,NR 表示不需要训练。从中可以看出,与其他算法相比,本文算法的运行效率最高,这主要是因为:

1) 本文算法不需要任何训练。

2) 虽然本文算法根据所有画廊图像集合对测试图像集中的各个图像进行重建,但是采用了高效的矩阵表示法,使得本文算法的效率优于其他算法。

表 6 不同算法进行图像集分类时的计算时间比较
Table 6 Calculation time comparison of different algorithms in image set classification

| 算法 | 总体训练时间 | 每个图像集的测试时间/s |
|----------|--------|--------------|
| TIS | NR | 0.045 0 |
| DCC | 13.26 | 0.311 0 |
| MMD | NR | 8.430 0 |
| MDA | 1.22 | 0.005 0 |
| AHISD | NR | 0.095 0 |
| CHISD | NR | 0.213 0 |
| GEDA | 2.70 | 0.068 0 |
| SANP | NR | 105.700 0 |
| CDL | 76.21 | 1.400 0 |
| RNP | NR | 0.027 0 |
| MSSRC | NR | 4.780 0 |
| SSDML | 21.92 | 0.577 0 |
| ADNT | 278.80 | 0.026 0 |
| 本文算法(平均) | NR | 0.004 6 |
| 本文算法(最优) | NR | 0.002 8 |

6 性能分析

从表 1 可以看出,DLRC 算法、PLRC 算法、ADNT 算法和本文算法的图像集分类性能远优于其他算法。其中,本文算法和 ADNT 算法的平均分类精度接近,高于 DLRC 算法和 PLRC 算法,这主要是因为:

1) 本文算法与 DLRC 算法和 PLRC 算法具有显著差异。DLRC 算法将训练和测试图像集看作高维空间的子空间,利用测试图像集和训练图像集间的距离来确定测试图像集的分类。为了确定子空间之间的距离,DLRC 算法利用各个图像集的最后一幅图像以及训练图像集和测试图像集间的变化来求解线性回归问题。PLRC 算法是 DLRC 算法的拓展,其不采用每个图像集的最后一幅图像,而是利用均值图像以及相关子空间和非相关子空间概念进行图像分类。DLRC 和 PLRC 算法均需要测试图像集和训练图像集中的图像总量远低于特征向量中的特征数量,而这一要求有时无法满足。除了画廊图像集,这 2 种算法还将测试图像集作为回归量,导致在测试期间容易发生矩阵缺秩问题。另外,这 2 种算法对部分数据集还采用了 LBP 特征。因此,DLRC 算法和 PLRC 算法采用原始图像时表现出的性能不具有普遍性。

2) 与 DLRC 和 PLRC 算法不同,本文算法对测试图像集中的各个图像独立对待,并将其看作高维空间中的点。本文算法根据画廊子空间对各个测试图像进行重建,并采用加权投票以及原始测试图像和重建测试图像间的欧氏距离进行目标分类,加权投票策略增加了本文算法面对噪声以及测试图像集中部分异常数据时的鲁棒性。本文算法对测试图像集中的图像数量不做任何要求。利用原始图像进行仿真,结果验证了本文算法的普遍性。与 DLRC 和 PLRC 算法相反,只要在构建画廊图像集时解决回归量 Q 的奇异性问题,本文算法在测试阶段便不会发生矩阵缺秩问题,原因是算法没有利用测试图像集作为回归量。另外,本文算法既可同时处理所有图像集,也可每次处理一幅图像然后对分类决策进行实时更新,因此,其适用于现场视频监控任务。本文算法的精度优于多数参数化算法和非参数化算法。

3) ADNT 算法对 YouTube 和 ETH-8 数据集的精度更高,但是其需要大量的训练数据和人工 LBP 特征。另外,ADNT 算法利用受限 Boltzman 机器方法进行参数初始化,参数微调的时间较长。相比而言,本文算法只需少量训练数据,即使采用原始图像,算法性能也与 ADNT 相当。如果图像分辨率较低,本文算法的性能优于 ADNT 算法,测试阶段本文算法效率比 ADNT 算法快 10 倍。当有新的数据需要处理时,本文算法也具有良好的可拓展性。

综上,本文算法对训练数据的要求较少,对低分

分辨率数据具有较好的处理效果,因此,其更适用于可用训练数据量有限、对决策效率要求高的应用场景。

7 结束语

本文提出一种改进的图像集分类算法。通过线性回归分类技术结合画廊图像集对测试图像集中的图像进行重建,利用累积加权重建误差确定测试图像集的分类。该算法对训练数据的需求量较少,面对不同的数据集时具有良好的普适性,在分辨率较低、训练数据较少的条件下仍然能够取得较好的分类性能。在常见的图像分类数据集上进行仿真,结果表明,本文算法测试阶段的效率比 ADNT 算法高 10 倍,且其精度与 ADNT 算法相差不大。下一步将研究并设计一种基于最优路径搜索的图像分类方法,以提高分类精度和时间效率。

参考文献

- [1] LÜ Wei, WANG Zhicheng, ZHAO Weidong, et al. An image set classification method based on low rank description[J]. Journal of Tongji University (Natural Science), 2013, 41(2): 271-276. (in Chinese)
吕焯, 王志成, 赵卫东, 等. 一种基于低秩描述的图像集分类方法[J]. 同济大学学报(自然科学版), 2013, 41(2): 271-276.
- [2] HAYAT M, BENNAMOUN M, AN S J. Deep reconstruction models for image set classification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(4): 713-727.
- [3] WANG Bin, HUANG Zhuqin, CHEN Liangxiao. Circular features description: effective method for leaf image retrieval and classification[J]. Journal of Software, 2019, 30(4): 1148-1163. (in Chinese)
王斌, 黄竹芹, 陈良宵. 圆周特征描述: 有效的叶片图像分类和检索方法[J]. 软件学报, 2019, 30(4): 1148-1163.
- [4] BAI Cong, HUANG Ling, CHEN Jianan, et al. Optimization of deep convolutional neural network for large scale image classification[J]. Journal of Software, 2018, 29(4): 1029-1038. (in Chinese)
白琮, 黄玲, 陈佳楠, 等. 面向大规模图像分类的深度卷积神经网络优化[J]. 软件学报, 2018, 29(4): 1029-1038.
- [5] GUI Jiangsheng, MA Chenfei, BAO Xiaolan, et al. Fine-grained image classification method for recurrent deep hybrid attention network[J]. Computer Engineering, 2019, 45(5): 205-209. (in Chinese)
桂江生, 麻陈飞, 包晓安, 等. 递归深度混合关注网络的细粒度图像分类方法[J]. 计算机工程, 2019, 45(5): 205-209.
- [6] XIE Yonghua, ZHU Yangang, ZHAO Xianguo. Classification and identification of pollen images based on Zernike moment and BoF-SURF feature fusion[J]. Computer Engineering, 2018, 44(7): 259-263, 270. (in Chinese)
谢永华, 朱延刚, 赵贤国. 基于 Zernike 矩与 BoF-SURF 特征融合的花粉图像分类识别[J]. 计算机工程, 2018, 44(7): 259-263, 270.
- [7] ZHANG Shun, GONG Yihong, WANG Jinjun. The development of deep convolution neural network and its applications on computer vision[J]. Chinese Journal of Computers, 2019, 42(3): 453-482. (in Chinese)
张顺, 龚怡宏, 王进军. 深度卷积神经网络的发展及其在计算机视觉领域的应用[J]. 计算机学报, 2019, 42(3): 453-482.
- [8] LU J W, WANG G, ZHOU J. Simultaneous feature and dictionary learning for image set based face recognition[J]. IEEE Transactions on Image Processing, 2017, 26(8): 4042-4054.
- [9] HUANG Z W, WANG R P, SHAN S G, et al. Face recognition on large-scale video in the wild with hybrid Euclidean-and-Riemannian metric learning[J]. Pattern Recognition, 2015, 48(10): 3113-3124.
- [10] 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.
- [11] KIM T K, KITTLER J, CIPOLLA R. Discriminative learning and recognition of image set classes using canonical correlations[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(6): 1005-1018.
- [12] HUANG Z W, WANG R P, SHAN S G, et al. Projection metric learning on Grassmann manifold with application to video based face recognition[C]//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2015: 140-149.
- [13] FENG J, WANG J, ZHANG H G, et al. Fault diagnosis method of joint fisher discriminant analysis based on the local and global manifold learning and its kernel version[J]. IEEE Transactions on Automation Science and Engineering, 2016, 13(1): 122-133.
- [14] LU J W, WANG G, MOULIN P. Localized multifeature metric learning for image-set-based face recognition[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2016, 26(3): 529-540.
- [15] HARANDI M T, SANDERSON C, SHIRAZI S, 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. Washington D. C., USA: IEEE Press, 2011: 2705-2712.
- [16] 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.
- [17] 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 2012 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2012: 2496-2503.

(上接第 250 页)

- [18] YANG M, ZHU P, GOOL L V, et al. Face recognition based on regularized nearest points between image sets [C]// Proceedings of IEEE International Conference and Workshops on Automatic Face and Gesture Recognition. Washington D. C., USA; IEEE Computer Society, 2017: 1-7.
- [19] ORTIZ E G, WRIGHT A, SHAH M. Face recognition in movie trailers via mean sequence sparse representation-based classification [C]// Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA; IEEE Press, 2013: 3531-3538.
- [20] ZHU P F, ZHANG L, ZUO W M, et al. From point to set; extend the learning of distance metrics [C]// Proceedings of 2013 IEEE International Conference on Computer Vision. Washington D. C., USA; IEEE Press, 2013: 2664-2671.
- [21] CHEN L. Dual linear regression based classification for face cluster recognition [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA; IEEE Press, 2015: 2673-2680.
- [22] FENG Q X, ZHOU Y C, LAN R S. Pairwise linear regression classification for image set retrieval [C]// Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA; IEEE Press, 2016: 4865-4872.
- [23] HAYAT M, BENNAMOUN M, AN S J. Learning non-linear reconstruction models for image set classification [C]// Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA; IEEE Press, 2014: 1915-1922.
- [24] HAYAT M, BENNAMOUN M, AN S. Deep reconstruction models for image set classification [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37 (4): 713-727.
- [25] SHORE A, HARTMANN W M. Improvements in transaural synthesis with the Moore-Penrose pseudoinverse matrix [J]. The Journal of the Acoustical Society of America, 2018, 143(3): 1938-1940.
- [26] CHIRANJEEVI P, GOPALAKRISHNAN V, MOOGI P. Neutral face classification using personalized appearance models for fast and robust emotion detection [J]. IEEE Transactions on Image Processing, 2015, 24(9): 2701-2711.
- [27] VIOLA P, JONES M J. Robust real-time face detection [J]. International Journal of Computer Vision, 2004, 57: 137-154.

编辑 吴云芳