基于优化空间金字塔匹配模型的图像分类*

柯善武 金 聪
(华中师范大学 计算机学院 武汉 430079)

摘 要：基于空间金字塔匹配模型(SPM)的图像分类中，构建视觉词直方图时对图像中所有特征都是同等对待，没有考虑到图像中不同区域特征的影响因子。显然，图像中目标区域比背景区域的特征重要性要大，为了避免图像中不重要区域的特征给图像分类带来干扰，提出了一种优化空间金字塔模型的图像分类方法。首先利用模拟退火算法与遗传算法相结合的聚类算法(SAGA)构造视觉词典，然后利用视觉注意机制构造加权的视觉词直方图。该方法在不丢失图像的全局信息的情况下，还考虑到了图像中各个区域对图像分类的重要性。最后将图像的表示向量使用 SVM 训练和分类。实验表明，本方法能够提高图像分类的准确率。

关键词：SPM 模型；视觉词直方图；SAGA 算法；视觉注意机制
中图分类号：TP391；TN919 **文献标识码**：A **国家标准学科分类代码**：520.6040

Image classification based on optimized spatial pyramid matching model

Ke Shanwu Jin Cong
(School of Computer, Central China Normal University, Wuhan 430079, China)

Abstract: In the image classification based on the spatial pyramid matching model, all the features in the image are treated equally when the histogram of visual words are constructed, without considering the influence factors of the different regions in the image. Obviously, the target area in image is more important than the background area. In order to avoid that the features of the non-important area in the image bring interference, this paper proposes an image classification method to optimize the spatial pyramid model. Firstly, a visual dictionary is constructed by using a clustering algorithm combined with a simulated annealing algorithm and a genetic algorithm. Then, a weighted visual word histogram is constructed using the visual attention mechanism. This method also takes the importance of classifying the images in each region of the image into account without losing the global information of the image. Finally, SVM is used to train and classify the representation vectors of images. Experimental results show that the proposed method can improve the accuracy of image classification.

Keywords: SPM model; visual word histogram; SAGA algorithm; visual attention mechanism

1 引 言

视觉词袋模型(bag of visual words，BOVW)最早运用在文本分析和文本挖掘领域。近年来，BOVW 模型在图像分类中已经非常流行[1]。它从图像的区域块中提取无序的特征描述符集合，将特征量化为离散的“视觉词”，然后将每幅图像映射为视觉词直方图，从而进行目标识别和场景分类。它丢弃了局部特征描述符的空间顺序，严重限制了图像的表达能力。针对这个问题，文献[2]提出了一种空间金字塔匹配模型(SPM)，它作为 BOVW 的一种扩展，在图像分类中取得了很大的进步。文献[3]提出了一种快速低等级的表示方法在 SPM 模型中去编码特征描述子。文献[4]提出了局部约束的线性编码方法进行图像分类。上述方法中，在利用图像的局部特征描述符来构造视觉词直方图的过程中，图像中不同区域的特征描述符都是同等对待的，完全没有考虑到图像不同区域的特征作用的大小。在图像分类中，图像的目标区域是人们分类的重要依据。在 BOVW 中，构造视觉词直方图的过程中会存在图像背景区域的特征带来干扰的现象。针对这个缺陷，文献[5]提出了一种基于词袋模型的图像优化分类方法，它仅仅提取图像的感兴趣区域来构建视觉词直方图，完全忽略了图像背景区域在图像中的重要地位。

因此，本文提出了一种基于优化空间金字塔模型的图像分类方法来解决这个问题。在构造视觉词典时没有采用

收稿日期：2016-12
* 基金项目：国家社会科学基金(13BTQ050)、2016 年华中师范大学研究生教育创新项目(2016CXZZ084)资助

K-means 聚类的方法，而是采用 SAGA 聚类使得构建的视觉词典具有全局最优解。在统计视觉词视觉词直方图时并非简单的对每个视觉词进行计数，而是通过视觉注意机制的结果考虑每个特征在图像中的重要性进行加权计数。这样，最终统计出的视觉词直方图既考虑了背景区域在构建视觉词直方图带来的干扰，同时考虑了它给图像分类带来的重要作用。通过实验表明，该方法构造的视觉词直方图能更好的表示每幅图像。

2　空间金字塔的视觉词袋模型

SPM 是 BOVW 的改进[3]。它增加了图像的空间位置信息，弥补了 BOVW 模型中图像局部特征点空间位置信息丢失的缺陷。SPM 的基本思想是将一幅图像进行空间网格序列的划分，在不同的层级 $\ell = 0,1,\cdots,L$ 下，图像可以划分为 $D = 4^{\ell}$ 大小相同的网格。图 1 分别在 $\ell = 0,1,2$ 下进行网格划分的结果。假设 H_X^{ℓ} 和 H_Y^{ℓ} 表示 X 和 Y 在 ℓ 层级下的直方图特征，$H_X^{\ell}(i)$ 和 $H_Y^{\ell}(i)$ 表示图像在 ℓ 层级中 X 和 Y 落入直方图第 i 个 *bin* 的特征点的个数，那么在 ℓ 层级下匹配点的总数计算为 $I(H_X^{\ell},H_Y^{\ell}) = \sum_{i=1}^{D}\min(H_X^{\ell}(i),H_Y^{\ell}(i))$。令 $I(H_X^{\ell},H_Y^{\ell}) = I^{\ell}$，由于高层级的 *bin* 被低层级的 *bin* 所包含，为了不重复计算，每个层级的有效数定义为匹配的增量 $I^{\ell} - I^{\ell+1}$，并且不同层级下的匹配应该赋予不同的权重，显然层级越高，权重越大，因此定义权重为 $1/2^{L-\ell}$，因此两幅图像的空间匹配计算如下：

$$k^L(X,Y) = I^L + \sum_{\ell=0}^{L-1}\frac{1}{2^{L-\ell}}(I^{\ell} - I^{\ell+1}) = \frac{1}{2^L}I^0 + \sum_{\ell=1}^{L}\frac{1}{2^{L-\ell+1}}I^{\ell} \tag{1}$$

当视觉词典大小为 100，空间金字塔下的视觉词直方图如图 1 所示。

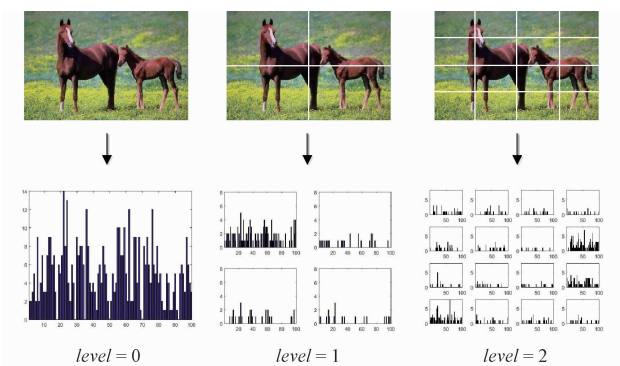


图 1　SPM 模型构造视觉词袋原理

3　视觉词直方图构造

文献[6]提出了一种基于改进 SURF 算法的遥感图像配准方法。文献[7]提出了一种基于改进的 Harris-SIFT 算子的快速图像配准方法。本文利用从训练图像库中提取 dense sift 特征使用 SAGA 算法生成视觉词典，对于给定的图像，可以将这张图像的所有特征点映射成视觉词典中的单词。因此，这张图像就是由众多单词组成的词袋，然后统计词袋中每个单词出现的次数并且组成视觉词直方图来表示这张图像。用视觉词直方图作为 SVM 分类的样本。视觉词直方图的构造步骤如下：

1）提取训练图像的 dense sift 特征，每个 dense sift 特征是 128 维的向量。

2）使用 SAGA 算法对 dense sift 特征集进行聚类，将聚类中心当做一个视觉词，因此可以生成视觉词典。

3）对照视觉词典，将训练图像和测试图像分别映射成视觉词袋。

4）构造加权的视觉词直方图。

5）训练分类器对测试图像进行分类，本文采用 LIBSVM 对图像进行分类。

3.1　模拟退火算法与遗传算法相结合的聚类算法(SAGA)

普遍采用 K-means 聚类构造视觉词典，虽然该聚类算法简单，容易实现，但是仍然存在不足：1）该算法是非数据集独立的；2）算法采用贪心搜索算法，受初始条件的影响，容易收敛于局部最优解；针对这个问题，本文采用了 SAGA 聚类算法。根据聚类问题的具体情况从而设计遗传编码方式以及适应度函数。它有效地克服了传统遗传算法的早熟现象，同时使该算法更有效、快速地收敛到全局最优解，SAGA 算法流程如下。

1）初始化控制参数：种群个体大小 *sizepop*，最大化次数 *MAXGEN*，交叉概率 P_C，变异概率 P_m，退火初始温度 T_0，温度冷却系数 k，终止温度 T_{end}。

2）随机初始化 c 个聚类中心，并生成初始种群 Chrom。对每个聚类中心用式(2)计算各个样本的隶属度，以及每个个体的适应度值 f_i，其中 $i = 1,2,\cdots,sizepop$。

$$\mu_{ik} = \frac{1}{\sum_{j=1}^{c}\left(\frac{d_{ik}}{d_{jk}}\right)^{\frac{2}{b-1}}} \tag{2}$$

3）设循环计数变量 $gen=0$。

4）对群体 *Chrom* 实施选择、交叉、变异等遗传操作，对产生的新个体使用式(2)和(3)计算 c 个聚类中心及各样本的隶属度，以及每一个体的适应度值 f_i'。若 $f_i' > f_i$，则用新个体替换旧个体；否则，以概率 $P = \exp((f_i - f_i')T)$ 去接受新个体，舍去旧个体。公式如下：

$$v_{ij} = \frac{\sum_{k=1}^{n}(\mu_{ik})^b x_{kj}}{\sum_{k=1}^{n}(\mu_{ik})^b} \tag{3}$$

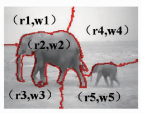
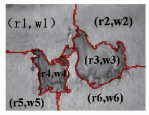
5）若 $gen < MAXGEN$，则 $gen = gen + 1$，转至步骤 4)；否则，转至步骤 6)。

6）若 $T_i < T_{end}$，则算法成功的结束，得到全局最优解；否则，执行降温操作 $T_{i+1} = kT_i$，转至步骤 3)。

3.2 图像分割后的区域视觉权值

图像的不同区域在人类视觉的权重是不一样的,利用 NCut 分割算法对图像进行分割,得到图像的多个区 $\{r_1, r_2, \cdots, r_i, \cdots, r_n\}, i \in [1,n]$。用 (r_i, w_i) 表示图像中第 i 个区域的权重为 W_i。利用 NCut 分割算法和视觉注意机制得到的图像各个区域的视觉权重值如表 1 所示。

表 1 图像的各个区域在视觉注意机制下的权值分布

原图像			
区域权值分布			

利用视觉焦点权重模型[8]来计算图像中不同区域在人类视觉中权值的大小,具体计算如下:

$$W_n = \omega_1 \cdot Area + \omega_2 \cdot Pos + \omega_3 \cdot Brightness \tag{4}$$

式中:W_n 表示第 n 个区域的视觉权重值,值越大代表该区域越重要;ω_i 代表权重系数。

1)面积 $Area$ 图像各个区域的视觉权重与该区域的大小密切相关。面积在一定的范围内会随着面积的增大更容易引起视觉注意,但是面积过大又会使该区域的显著性降低。计算公式如下:

$$Area = \frac{\max\left(\frac{s_i}{\alpha}, 1\right)}{\sum_{j \in \sum}^{l} \max\left(\frac{s_j}{\alpha}, 1\right)} \tag{5}$$

式中:s_i 表示图像分割后第 i 个区域的像素点个数;$\sum$ 表示图像分割后区域的个数;α 是一个常量值,以防区域超出饱和度,设该值为图像面积的 1%。

2)位置 pos 图像中区域的位置是引起人类视觉注意的因素。其计算如下:

$$Pos = p_i / p_{center} \tag{6}$$

式中:p_{center} 表示图像中央位置像素的个数,p_i 表示第 i 个分割区域位于图像中央位置的像素总数。

3)亮度 $Brightness$ 图像的亮度参数是引起人类视觉注意的因素。计算如:

$$Brightness = \left|\max(mean(RGB)) - \max(Globalmean(RGB))\right| \tag{7}$$

式中:$mean(RGB)$ 表示第 i 个子区域的亮度均值,$Globalmean(RGB)$ 表示整幅图像的亮度均值。

3.3 构造加权的视觉词直方图

本文使用视觉焦点权重机制来衡量图像中每个视觉词在图像中的重要程度,最终在统计视觉词出现的频数时不是简单的加 1,而是加上视觉词的权重。因此构成的视觉词直方图能够很好的反映出图像中不同区域的重要性,以免背景区域或者不重要的区域带来巨大的干扰。

利用 SAGA 聚类算法生成的视觉词典为 $vw = \{vw_1, vw_2, \cdots, vw_m\}$,假设输入的某张图像提取的 dense sift 特征 $f = \{f_1, f_2, \cdots, f_n\}$,构造加权的视觉词直方图步骤如下。

1) 初始化 m 维的视觉词直方图向量 $v[m] = [0,0,\cdots,0]$,用来保存视觉词直方图的加权频数。v[i]统计的是 $vw_i, i \in \{1,2,\cdots,m\}$ 出现的频数的加权值。

2) 对于图像中的特征 f,使用最近邻算法(nearest neighbor algorithm)找到与视觉词典中距离最近的视觉单词 vw_i,使得:

$$BIN_j = \arg\min_j \| f_i - vw_j \|^2 \quad i \in [1,n], j \in [1,m] \tag{8}$$

满足式(8),则可以将这张图像的 f_i 特征作为视觉词 vw_j。

3) 对于给定图像的特征 f_i,统计其加权频数。计算如下:

$$v[BIN_j] += \varphi(f_i) \tag{9}$$

4) 根据 3.2 节的视觉注意机制可以计算出 $v[BIN_j]$ 的值。计算公式如下:

$$\varphi(f_i) = \frac{W_q}{\sum_{j=1}^{n}(W_j)}, f_i \text{ 在第 } q \text{ 个区域} \tag{10}$$

式中:n 表示 3.2 节中图像分割后区域的总个数,$W_q, q \in [1,n]$ 表示图像分割后第 q 个区域的视觉注意权值。

4 过滤式特征选择

利用 SPM 模型构造视觉词直方图,当 $level=3$,视觉词典大小为 400 时,可得到 8 400 维的视觉词直方图。这给分类任务带来“维数灾难”,复杂度随着视觉词直方图维度的增加而增大。同时,在高维的视觉词直方图中,存在着一些与分类无效的特征,特征之间存在着巨大的冗余,去除不相关特征会降低学习任务的难度。本文利用 Relief 特征选择方法剔除视觉词直方图中一些冗余的特征,在降低视觉词直方图维度的同时,减少模型运行的时间,提高模型的准确率。Relief 的关键是确定相关统计量。给定训练集 $\{(x_1,y_1),(x_2,y_2),\cdots,(x_m,y_m)\}$,对每个示例 x_i,Relief 先在 x_i 的同类样本中寻找其最近邻 $x_{i,nh}$,称“猜中近邻”,再从 x_i 的异类样本寻找其最近邻 $x_{i,nm}$,称为“猜错近邻”,然后,相关统计量对应于属性 j 的分量为:

$$\delta^j = \sum_i -diff(x_i^j, x_{i,nh}^j)^2 + diff(x_i^j, x_{i,nm}^j)^2$$

其中 x_a^j 表示样本 x_a 在属性 j 上得取值,$diff(x_a^j, x_b^j)$ 取决于属性 j 的类型:若属性 j 为离散型,则 $x_a^j = x_b^j$ 时 $diff(x_a^j, x_b^j) = 0$,否则为 1;若属性 j 为连续型,则 $diff(x_a^j, x_b^j) = |x_a^j - x_b^j|$,注意 x_a^j,x_b^j 已规范化到[0,1]

区间。若 x_i 与猜中的近邻 $x_{i,nh}$ 在属性 j 上的距离小于 x_i 与其猜错近邻 $x_{i,nm}$ 的距离，则说明属性 j 对区分同类和异类样本是有益的，于是增大属性 j 所对应统计量分量；反之，若 x_i 与猜中近邻 $x_{i,nh}$ 在属性 j 上的距离大于 x_i 与猜错近邻 $x_{i,nm}$ 的距离，则说明属性 j 起负面作用，于是减少属性 j 所对应的统计分量。最后，对基本不同样本得到的估计结果进行平均，就得到各属性的相关统计量分量，分量值越大，则对应属性的分类能力就越强。

5　实验结果与分析

为了验证本文分类方法的性能，在 Caltech101、Caltech256、PASCAL VOC 2007 三个标准图像库中进行试验。在 Caltech101 图像库中，选择 airplane、butterfly、motorbike、buddha 和 watch 五类图像，在 Caltech256 图像库中我们选择 butterfly、calculator、camel、necktie、elephant 五个类别的图像，在 PASCAL VOC 2007 图像库中选择 bird、motorbike、sheep、dog 和 bus 五类图像进行试验。随机选择每类的 50 张图像作为训练图像，剩下的图像中选择 30 张作为测试图像，设置视觉词典大小为 500。使用 libsvm 进行分类试验，重复试验 10 次获得平均分类准确率结果如表 2 所示。

表 2　不同方法在不同图像集上得分类准确率比较

算法	Caltech101	Caltech256	Pascal VOC2007
	准确率/%	准确率/%	准确率/%
KCB[9]	64.14±0.53	27.17±0.95	54.60±0.63
KernelBoF[11]	45.16±0.81	11.25±0.46	34.70±2.79
ScSPM[10]	65.39±1.21	28.60±0.15	56.44±0.54
LLC[4]	67.86±1.17	29.35±0.42	59.89±1.55
LrrSPM[3]	65.89±1.03	27.43±0.98	61.42±1.46
本文算法	67.91±0.65	30.40±0.54	61.67±1.21

表 2 可以得出，本文方法的图像分类的准确率有明显的提升。实验表明，考虑图像的不同区域在人类视觉的权重问题来进行图像分类是有必要的，它考虑到图像中每个区域给图像分类带来的作用之外，还能够避免某些区域的特征信息给图像分类带来的干扰。

在 Caltech101 图像库中，设置视觉词典的大小对图像分类准确率的影响如图 2 所示。

利用 BOVW 模型进行图像分类，设置视觉词典的大小对图像分类的准确率是十分有影响的。通常需要不断的设置视觉词典的大小，使得分类的效果达到最好的效果。图 2 表明当视觉词典为 500 时可以达到很好的效果。

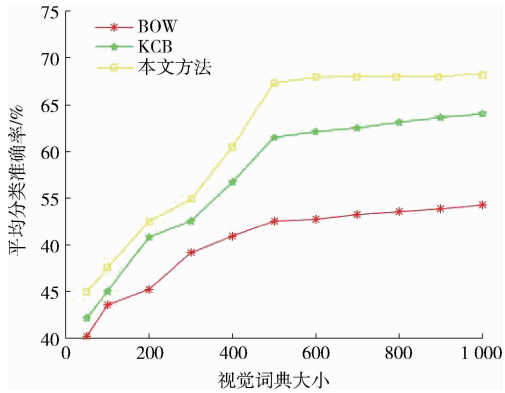


图 2　视觉词典的大小对图像分类性能的影响

6　结　论

本文提出一种基于优化空间金字塔模型的图像分类方法。它在增加图像空间分布信息的同时，还考虑图像的不同区域在人类视觉的权重。图像中某个区域越能引起人们的视觉注意，那么该区域的视觉注意权值越大，显然这个区域的特征必须得到重视，因此这个区域特征的权重也会越大。当图像中某个图像的视觉权重越小，则说明这个区域的在图像中的作用越小，因此这个区域的特征的权重也会相应的减少。将这种视觉注意机制融入到 SPM 模型中就增加了图像中重要信息的权重，同时避免了图像背景区域信息的干扰，使得图像分类的准确率明显的提高。

参考文献

[1]　樊存佳，汪友生，边航，等. 一种改进的 KNN 文本分类算法[J]. 国外电子测量技术，2015，34(12)：39-43.

[2]　LAZEBNIK S，SCHMID C，PONCE J. Beyond bags of features：Spatial pyramid matching for recognizing natural scene categories[J]. IEEE Computer Society Conference on Computer Vision & Pattern Recognition，2006：2169-2178.

[3]　PENG X，YAN R，ZHAO B，et al. Fast low rank representation based spatial pyramid matching for image classification[J]. Knowledge-Based Systems，2015，90(C)：14-22.

[4]　WANG J，YANG J，YU K，et al. Locality-constrained linear coding for image classification[J]. Computer Vision & Pattern Recognition，2010，119(5)：3360-3367.

[5]　赵春晖，王莹，KANEKO M. 一种基于词袋模型的图像优化分类方法[J]. 电子与信息学报，2012，34(9)：2064-2070.

[6]　阳吉斌，胡访宇，朱高，等. 基于改进 SURF 算法的遥感图像配准[J]. 电子测量技术，2012，35(3)：

69-72.

[7] 许佳佳,张叶,张赫. 基于改进 Harris-SIFT 算子的快速图像配准算法[J]. 电子测量与仪器学报,2015,29(1):48-54.

[8] 陈祉宏,冯志勇,贾宇. 基于视觉注意权重模型的图像检索方法[C]. 图像图形技术与应用学术会议,2011.

[9] GEMERT J C V, GEUSEBROEK J M, VEENMAN C J, et al. Kernel codebooks for scene categorization[C]. European Conference on Computer Vision, Springer-Verlag, 2008:696-709.

[10] YANG J CH, YU Y, GONG Y, et al. Linear spatial pyramid matching using sparse coding for image classification[C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition DBLP, 2009:1794-1801.

[11] CAICEDO J C, CRUZ A, GONZALEZ F A. Histopathology image classification using bag of features and kernel functions [J]. Conference on Artificial Intelligence in Medicine: Artificial Intelligence in Medicine, 2009:126-135.

作者简介

柯善武,1990 年出生,硕士研究生,主要研究方向为数字图像处理。

金聪,1960 年出生,教授,博士,主要研究方向为数字图像处理。

E-mail:jinc26@aliyun.com

(上接第 94 页)

5 结 论

针对利用传统 SIFT 变换在全景图像拼接存在的效率低和误匹配较多的缺陷,本文在传统的基于 SIFT 变换的基础上,阐述了使用改进的 SIFT 算法进行图像拼接的过程。同时介绍了通过这种改进确实提高了图像拼接的效率,并且对于存在的误匹配点的情况有很大的改善,大大提高了传统 SIFT 算法拼接全景图像的效率。虽然这种改进的算法对于简单纹理图像的拼接和复杂纹理图像的拼接都有较好的效果,但在一些拼接缝隙处仍能看到一些细微的接缝,仍需要对图像融合进行改进,真正实现无缝拼接。

参考文献

[1] YU H, JIN W. Brightness adaptive algorithm for image mosaic seamless fusion[C]. Society of Photo-Optical Instrumentation Engineers, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 2010:78501N-78501N-9.

[2] PAN H, ZHANG Y, LI C, et al. An adaptive Harris corner detection algorithm for image mosaic [C]. Communications in Computer & Information Science, 2014, 484:53-62.

[3] 陈小丹,杜宇人,高秀斌. 一种基于 SURF 的图像特征点快速匹配算法[J]. 扬州大学学报:自然科学版,2012, 15(4):64-67.

[4] LOWE D G. Distinctive image featuresfrom scale-invariant keypoints [J]. International Journal of Computer Vision, 2006,9(2):91-110.

[5] 张锐娟. 图像配准理论及算法研究[D]. 西安:西安电子科技大学,2009.

[6] 骆飞,王岩飞,牛晓丽,等. 一种 SAR 高效数字去斜方法的研究[J]. 电子测量技术,2016,39(9):165-171.

[7] Mark Allen Weiss 著. 张怀勇等译. 数据结构与算法分析[M]. 北京:人民邮电出版社,2007:158-189,374-401.

[8] 任克强,胡梦云. 基于改进 SURF 算法的彩色图象配准算法[J]. 电子测量与仪器学报,2016,30(5):748-756.

[9] 李秀艳,韩倩,汪剑鸣,等. 基于改进共轭梯度法的 ERT 图像重建[J]. 仪器仪表学报, 2016,37(7):1673-1679.

[10] 吴渊凯,卞新高. 计算机视觉中摄像机标定的实验分析[J]. 电子测量技术,2016,39(11):95-99.

[11] YANG B F, LIU Y Y, LU Y, et al. Research of optical rainfall sensor based on CCD linear array [J]. Instrumentation, 2015, 2(3):27-34.

[12] 郝贵青,王冰洋. 一种基于 RGB 颜色空间的色彩还原方法[J]. 国外电子测量技术,2016,35(11):24-26.

作者简介

常伟,1991 年出生,青岛科技大学硕士研究生,研究方向为计算机视觉。

E-mail:290358191@qq.com

刘云,1962 年出生,青岛科技大学教授,研究方向为计算视觉。

E-mail: lyun-1027@163.com