Applied Soft Computing 46 (2016) 1056–1066

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Characterizing objects with SIKA features for multiclass classification

Siddharth Srivastava*, Prerana Mukherjee, Brejesh Lall

Indian Institute of Technology, Delhi, India

ARTICLE INFO

Article history:
Received 15 June 2015
Received in revised form
26 November 2015
Accepted 7 December 2015
Available online 23 December 2015

Keywords:
Object classification
SIKA
Minimal Complexity Machine
Support Vector Machine

ABSTRACT

This paper presents a novel approach for multiclass classification by fusion of KAZE and Scale Invariant Feature Transform (SIFT) features followed by Minimal Complexity Machine (MCM) as the classifier. Unlike the existing features, the paper proposes a new feature SIKA to represent characteristics of an object, as opposed to just forming a compendium of interest points in an image to represent the object characteristics. Further we append a strong and lightweight classifier, MCM to the technique. The resulting classifier easily outperforms existing techniques based on handcrafted features. Two new scores Keypoint Overlap Score (KOS) and Mean Keypoint Overlap Score (MKOS) have also been proposed as part of this work which are useful in establishing the strength of features for object representation.

1. Introduction

During the past few decades, object classification has remained an important research area for various real-world recognition applications like visual surveillance [1], object annotation [2], object tracking [3], obstacle avoidance/path planning [4] etc. In these tasks, the classifier has to handle the diverse nature of objects which makes it challenging to devise a single solution for all object classification problems. These challenges can be attributed to the following factors: number of classes, number of instances of each class, total number of images in the dataset, relative ratios of training and testing images, intra-class variance, ground truth annotations etc. Such solutions focus on two fundamental issues. First, the distinct characterization of objects of interest (features) and the second is the identification of objects (classification). Scale Invariant Feature Transform (SIFT) [5], Speeded Up Robust Features (SURF) [6], Histogram of Oriented Gradients (HOG) [7], Oriented and Rotated Brief (ORB) [8], KAZE [9] etc. have been widely used for solving the former issue. On the other hand, support vector machines (SVM) [10,11] have remained the most popular choice for the latter issue. Recently, Convolutional Neural Network (CNN) [12] has shown to outperform most of the traditional object classification and recognition benchmarks [13–15].

The techniques used for object classification can be broadly classified into three categories. The first set of techniques focus primarily on improving the input representation with the help of stronger features while using a simple classifier such as SVM. One such approach is linear Spatial Pyramid Matching based on Sparse Coding (ScSPM) [16] which uses sparse coding over vector quantization. It relaxes the cardinality constraints and introduces a regularization parameter to obtain a smaller number of non zero elements. This is then followed by max spatial pooling. It thus reduces the complexity of the classifier. In another related work [17], authors use a locality adaptor which allows to choose appropriate basis vectors corresponding to an input descriptor. Recently, authors in [15] perform experiments to demonstrate superiority of generic features extracted from CNN over handcrafted features for several recognition tasks. These features are based on the OverFeat [18] architecture. Later in this section we discuss that despite this overwhelming performance by generic features, there are gaps in the way these features are represented.

The second set of techniques focus on generating stronger training cases or using ensemble of classifiers. Stronger training cases allows the classifier to learn the peculiarities of the training set while a set of classifiers help in reducing bias by learning a more expressive representation. Authors in [19] illustrate this technique by formulating a latent SVM. It results in the problem being formulated as a convex training problem. The hard training examples are generated by using a feature similar to Histogram of Oriented Gradients (HOG). Recently, Regions with Convolutional Neural Network features (R-CNN) [14], a variant of CNN has been used to extract features from the region proposals. They perform

* Corresponding author. Tel.: +91 11 2659 1068.
E-mail addresses: eez127506@ee.iitd.ac.in (S. Srivastava), eez138300@ee.iitd.ac.in (P. Mukherjee), brejesh@ee.iitd.ac.in (B. Lall).

http://dx.doi.org/10.1016/j.asoc.2015.12.014
1568-4946/

supervised pre-training on a large dataset and then fine-tune this pretrained CNN on a relatively smaller target dataset. For this purpose they augment the strength of CNN by using category-specific SVM. These features are then classified into respective object categories.

Finally, the third set of techniques are those which try to balance the trade-off between classifier and feature strengths. In [20], authors propose a two stage sliding window approach for object localization. The main idea is to combine the classification and detection phases by considering latent properties of objects and scenes. Another technique, Selective Search [21] reduces the relative time for localizing objects by applying complementary grouping techniques for sampling. The reduction in localization is leveraged by constructing difficult negative examples to train the classifier.

These works however suffer from one or more of the following shortcomings: (a) The existing interest point based feature extraction techniques focus on characterizing content based on local information. Object features are not directly targeted by these techniques and are instead a consequence of image (and not object) interest points. The features themselves are not tuned to find or represent objects. Besides, there is very little understanding on how CNN extracts features. Moreover, recently authors in [22] showed that CNN can easily be fooled even with images that are easily identified as negatives by the human vision system. However, the authors also claim that the work provides insight into two key properties of the features from CNN. First, CNN extracts low and middle level features instead of high level features such as shape, boundary etc. Second, it learns patterns in the images which is the primary reason it was fooled. In our work we attempt to address this gap by proposing features which are representative of the characteristics of an object rather than them being a compendium of abstract representation of interest points. (b) Both CNN and SVM work well in practice, but there is no theoretical explanation of their generalization ability. This makes their use primarily based on experimentation. Moreover, CNN require huge training databases which is not usually available for many domains. We therefore use handcrafted features along with a classifier which is light-weight and guarantees generalization.

In this paper, we present a novel technique for generating a stronger feature set by using a combination of KAZE and SIFT keypoints termed as SIKA features (**SI**FT-**KA**ZE). We use these features with MCM to propose a light weight yet strong object classifier. The proposed scheme outperforms most of the existing state of the art methods. The SIKA features attempt to specifically characterize the object rather than obtaining a set of interest points. SIKA keypoints are constructed from SIFT and KAZE keypoints (described in Sections 2 and 4.1). SIFT [5] and its derivatives [23–25] show good invariance to several transformations. Since SIFT is based on Gaussian Scale Space (GSS), it inherently assigns equal importance to features on the object boundaries and to those inside it. The recently proposed KAZE [9] feature is based on non linear scale space. A useful property of KAZE is that it preserves the object boundaries as it blurs the region around edges more than the edges themselves. Therefore, it assigns more importance to features on and around the boundary. Hence SIKA features obtain a good mix of boundary (drawn from KAZE) and appearance (drawn from SIFT). The classification is performed using the recently proposed Minimal Complexity Machine (MCM) [26]. It has been shown to outperform SVM in terms of accuracy, computational complexity as well as providing sparse representation of the features. The strongest argument in favor of MCM is its provably good generalization accuracy and requirement of far lesser number of support vectors as compared to SVMs. Fewer support vectors mean faster classification of test points, and consequently due to complexity and size of the object classification datasets, MCM makes a strong case for itself. Since MCM is a recent technique for the benefit of the reader, a detailed discussion about MCM is given in Section 3.1.

The key contributions of this paper can be summarized as follows:

1. We propose SIKA features that characterize properties of an object. We also establish that SIFT and KAZE are complementary features. We show that a carefully chosen combination of these (as described in Section 4) boosts the classification accuracy significantly. We achieve state of the art performance on Caltech-256 dataset while close the gap to CNN based techniques on Pascal VOC 2007 dataset.
2. We show that Minimal Complexity Machine (MCM) achieves significant improvement in classification performance over the state-of-the-art work while using fewer number of training samples. We have also implemented a improvised version of MCM on GPU. To the best of our knowledge, this is the first work to demonstrate the effectiveness of MCM on images and datasets with large number of classes.

The rest of the paper is organized as follows. In Section 2, we motivate the use of KAZE and SIFT by introducing SIKA features. We explain MCM and propose its improvised version in Section 3. Section 4 describes the proposed methodology. In Section 5, we elaborate the experimental analysis and results while Section 6 concludes the paper.

2. Keypoint selection and description

Keypoint selection aims at finding a minimum set of features which helps in achieving maximum classifier performance based on certain metrics. It helps in getting rid of redundant features, resulting in simplification of the model and reduction in training time. In this work we achieve this by combining SIFT and KAZE interest points. In the following subsection, we describe how KAZE interest points can be added to complement the information represented by SIFT interest points.

2.1. Complementing SIFT

Recent studies [27,28] indicate that SIFT is the strongest feature detector available. As discussed in Section 1, SIFT focusses on appearance/ region of the entire object using high detail interest points (not necessarily boundary points) and KAZE concentrates on the boundary information. Therefore, we aim to complement the strength of SIFT with the KAZE features. The complementarity of SIFT and KAZE is due to differences in the generation mechanism of these features. The first difference is in the construction of the scale space. KAZE is based on non-linear scale space while SIFT is based on Gaussian scale space (GSS). KAZE uses non-linear diffusion filtering as given in Eq. (1) to construct the scale space.

$$\frac{\partial L}{\partial t} = div\{(c(x, y, t) \cdot \nabla(L))\} \quad (1)$$

where div and ∇ are divergence and gradient operators respectively, c is the conductivity function and t is scale parameter. The conductivity function c, is represented as a gradient (Eq. (2)), helping in the reduction of diffusion at edges, thus resulting in more smoothening of regions as compared to edges.

$$c(x, y, t) = g(|\nabla L_\sigma(x, y, t)|) \quad (2)$$

where ∇L_σ (luminance function) is the gradient of a Gaussian smoothed original image L where σ is the amount of blur. This property of the conductivity function makes KAZE suitable for boundary representation. There are various conductivity functions defined in