Neurocomputing 266 (2017) 239–246

Contents lists available at ScienceDirect

Neurocomputing

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

Learning multiple local binary descriptors for image matching

Yongqiang Gao[a,b,c,*], Weilin Huang[a,d], Yu Qiao[a,d]

[a] Shenzhen Key Lab of Computer Vision and Pattern Recognition, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China
[b] Tencent Inc., China
[c] Shenzhen College of Advanced Technology, The University of Chinese Academy of Sciences, China
[d] The Chinese University of Hong Kong, Hong Kong SAR

ARTICLE INFO

Article history:
Received 11 June 2015
Accepted 17 May 2017
Available online 6 July 2017

Communicated by Prof. Qiao Yu

Keywords:
Local binary descriptors
L_1 norm
rankSVM
Convex optimization
Image matching

ABSTRACT

Binary descriptors have received extensive research interests due to their low memory storage and computational efficiency. However, the discriminative ability of the binary descriptors is often limited in comparison with general floating point ones. In this paper, we present a learning framework to effectively integrate multiple binary descriptors, which is referred as learning-based multiple binary descriptors (LMBD). We observe that previous successful binary descriptors like Receptive Fields Descriptor (RFD) which includes rectangular pooling area (RFD_R) and Gaussian pooling area (RFD_G)), BinBoost, and Boosted Gradient Maps (BGM), are highly complementary to each other. We show that the proposed LMBD can improve the discriminative ability of individual binary descriptors significantly. We formulate the fusion of multiple groups of the binary descriptors was formulated as a pair-wise ranking problem, which can be solved effectively in a rankSVM framework. Extensive experiments were conducted to evaluate the efficiency of LMBD. The proposed LMBD obtains the error rate of 12.44% on the challenging local patch datasets, which is about 2% lower than the state-of-the-art results (obtained by a learning based floating point descriptor). Furthermore, the proposed binary descriptor also outperforms other binary descriptors on image matching task.

1. Introduction

Local feature descriptor is a fundamental topic in computer vision and image processing. The performance of many computer vision applications is highly relied on an informative and robust descriptors, such as image matching [1], object recognition [2], object detection [3,4], image classification [5,6], face recognition [7], text localization [8] and so on. However, it is still highly challenging to design robust and discriminative descriptors, due to the significant variations caused by real-world changings of lighting or illumination, un-fixed viewpoints and different image qualities (blurring, noise, and low resolution), images with the same scene or object may exhibit significant appearance differences.

Current local feature descriptors can be roughly divided into two types, the floating point and binary descriptors. The former makes use of floating type values for vector feature representation. The widely-used floating point descriptors includes Scale-Invariant Feature Transform (SIFT) [1], Histograms of Gradients (HoG) [9], and GLOH [10] descriptors. In the recent years, various binary descriptors have been proposed to cater the applications in low power mobile devices and the demands of fast computation. Recent binary descriptors include Binary Robust Independent Elementary Features (BRIEF) [11], Discriminative BRIEF (D-BRIEF) [12], BinBoost [13], Binary Robust Invariant Scalable Keypoints (BRISK) [14], ORB [15], Local Difference Binary (LDB) [16], Boosted Gradient Maps (BGM) [12], Local Ternary descriptor (LTD) [17], Ring-based Multi-Grouped Descriptor (RMGD) [18] and the latest Receptive Fields Descriptor(RFD) [19]. In contrast to the floating point ones, the binary descriptors encode patch information using binary strings and hamming distance is applied for measuring the similarity between patches by using fast XOR operator.

With numerous advantages in low memory storage, fast computing and matching strategies, the binary descriptors have been attracting increasingly attention recently. Generally, there are two types of methods to compute the binary descriptors. The first approach is to explore the quantization [20,21] and hashing

* Corresponding author at: Shenzhen Key Lab of Computer Vision and Pattern Recognition, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China.
E-mail addresses: jasonyqgao@tencent.com (Y. Gao), wl.huang@siat.ac.cn (W. Huang), yu.qiao@siat.ac.cn (Y. Qiao).

http://dx.doi.org/10.1016/j.neucom.2017.05.038
0925-2312/

techniques [20–26] to binarize the existing floating point features. Obviously, the performance of this approach is significantly limited by the intermediate floating point representations. The second method is to involve binary tests by computing the intensity differences between pairs of selected pixels or defined regions, and some learning based methods have been developed to optimize the selection of binary tests [11–17,19,27,28]. For example, the BinBoost [13] and BGM [29] apply the boosting-trick to learn compact binary strings for non-linear visual representation. The Boosted Similarity Sensitive Coding (SSC) [30] is adopted to maximize the similarity between similar patches, and to minimizes it between dissimilar patches simultaneously. The RFD [19] defines a set of receptive fields in multiple gradient channels for binary tests, and generates the compact binary strings by learning to binarize the responses of defined receptive fields.

Though the binary descriptors have great advantage in speed and memory storage, they often focus overly on compact representation in the cost of significant information loss, leading to less discriminative power. This can be substantiated by the results on the challenging *Brown's* datasets [31–33] where the learning based floating point descriptor developed by Simonyan *et. al.* [34] achieves the best performance, with a large margin over the performance of existing binary descriptors.

In this paper, we propose a novel learning framework to effectively integrate multiple successful binary descriptors with the learnt weights, in an effort to bridge the performance gap between current binary and floating point descriptors. We refer the new descriptor as learning-based multiple binary descriptors (LMBD). Two types of binary descriptors are explored as the basic units: the boosting-trick based binary descriptors, including BinBoost [13] and BGM [12], and receptive fields based descriptors, such as the $\text{RFD}_{\mathbf{G}}$ and $\text{RFD}_{\mathbf{R}}$ described in [19]. Each binary descriptor is considered as a feature group. We develop a margin-based ranking optimization method to learn the weights optimally for multiple groups by leveraging the rankSVM algorithm [35]. Our optimization technique is inspirit similar to that of [18], but here we use different types of binary descriptors. We observe that the boosting-trick and receptive fields based descriptors are capable of capturing different characteristics of features, which can compensate strongly for each other, leading to a significant improvement on the discriminative power. The proposed LMBD was evaluated on the challenging Brown's datasets [33]. It achieves excellent results which are comparable or even better than the state-of-the-art results achieved by the floating descriptor [34]. We also evaluated the LMBD descriptor on image matching task using the Mikolajczyk datasets [10]. The experimental results show that the LMBD outperforms other binary descriptors considerably for image matching.

The rest of the paper is organized as follows. In the next section, we present the details of our basic framework for multiple binary descriptors learning, and provide an efficient algorithm for solving it optimally. Experimental results on patch and image matching are reported in Section 3, with comparisons against recent existing binary and floating point descriptors. Section 4 concludes this manuscript.

2. Learning-based multiple binary descriptors

In this section, we present the details of our framework for learning multiple binary descriptors, including the boosting-trick and receptive fields based descriptors. Then a rankSVM algorithm is presented to solve the learning problem effectively.

2.1. Basic single binary descriptors

Given an image intensity patch $\mathbf{x}$, a local descriptor $C(\mathbf{x}) = [C_1(\mathbf{x}), \ldots, C_D(\mathbf{x})]$ maps the patch to a D-dimensional vector. For a binary descriptor, $C_i(\cdot)$ denotes a binary function or a binary test. Various binary descriptors are different in computing the binary tests, C.

2.1.1. Boosting-trick binary descriptors

Both BinBoost [13] and BGM [29] learn compact binary strings using the boosting-trick. They apply the boosting to learn complex non-linear local binary representations. The employed weak learner family is capable of encoding specific design choices and meaningful descriptor properties. First, both methods construct a feature pooling $H(\mathbf{x}) = \{h_i\}_{i=1}^{M}$, which a collection of thresholded non-linear response functions of a intensity patch, $h_i(\mathbf{x}) \in \{-1, 1\}$. The size of the feature pooling (M) is generally large or possibly infinite. Then the compact binary strings and the discriminative mapping functions are learnt by minimizing the exponential loss of a defined similarity function $f(m, n)$ over a set of image patch pairs [13]

$$\mathscr{L} = \sum_{i=1}^{N} exp\Big(- l_i f(m_i, n_i)\Big), \tag{1}$$

where m_i, n_i are a pair of intensity patches, and $l_i \in \{-1, 1\}$ is a label indicating whether it is a similar (1) or dissimilar (-1) pair. The Boosted Similarity Sensitive Coding (SSC) algorithm [30] is adopted. It defines a similarity function by using a simply weighted sum of the thresholded response functions

$$f(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{M} a_i h_i(\mathbf{x}) h_i(\mathbf{y}), \tag{2}$$

which defines a weighted hash function with the importance of each dimension i given by a_i. The minimization of Eq. (1) aims to find an embedding that maximizes the similarity between similar patches, and at the same time, minimizes it between the dissimilar patches.

Computing the responses of the binary tests in multiple gradient domains leads to gradient-based BinBoost and BGM, which are different in the choice of weak learners. Each BinBoost is computed as a linear combination of many gradient orientation maps, while each BGM is constructed by a weak learner. The gradient-based weak learners applied by two descriptors are defined as

$$h(\mathbf{x}; R, e, T) = \begin{cases} 1 & \text{if } \phi_{R,e}(\mathbf{x}) \le T \\ -1 & \text{otherwise} \end{cases} \tag{3}$$

where

$$\phi_{R,e}(\mathbf{x}) = \sum_{m \in R} \xi_e(\mathbf{x}, m) / \sum_{e_k \in \phi, m \in R} \xi_{e_k}(\mathbf{x}, m), \tag{4}$$

where region $\xi_e(\mathbf{x}, m)$ is the gradient energy along an orientation e at location m, and R defines a rectangular extent within the patch ($\mathbf{x}$).

2.1.2. Receptive fields' binary descriptors

The RFD [19] is computed by thresholding the responses of a set of defined receptive fields. It defines a feature pooling, which may also be large or possibly infinite. Specifically, the is computed in three steps: primary feature extraction, receptive fields pooling and binarization. Firstly, a patch $\mathbf{x}$ is mapped into 8 feature channels with the same size as the patch, by soft assigning the gradient orientation of each pixel into 8 orientated bins: $\{0, 1 \times \frac{\pi}{4}, 2 \times \frac{\pi}{4}, \ldots, 7 \times \frac{\pi}{4}, \}$, each of which corresponds to a feature channel. Secondly, the floating-point responses of the receptive fields defined within a geometric area are calculated on each