Neurocomputing 184 (2016) 28–35

Contents lists available at ScienceDirect

Neurocomputing

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

LOAD: Local orientation adaptive descriptor for texture and material classification

Xianbiao Qi [a,b,*], Guoying Zhao [a], Linlin Shen [b], Qingquan Li [b], Matti Pietikäinen [a]

[a] Center for Machine Vision and Signal Analysis, University of Oulu, PO Box 4500, FIN-90014, Finland
[b] Shenzhen Key Laboratory of Spatial Information Smart Sensing and Services, Shenzhen University, Shenzhen 518000, China

A R T I C L E I N F O

Article history:
Received 22 January 2015
Received in revised form
20 May 2015
Accepted 13 July 2015
Available online 29 December 2015

Keywords:
Local orientation adaptive descriptor
Texture classification
Material recognition
Improved Fisher vector
Convolutional neural network

A B S T R A C T

In this paper, we propose a novel local feature, called Local Orientation Adaptive Descriptor (LOAD), to capture regional texture in an image. In LOAD, we proposed to define point description on an Adaptive Coordinate System (ACS), adopt a binary sequence descriptor to capture relationships between one point and its neighbors and use multi-scale strategy to enhance the discriminative power of the descriptor. The proposed LOAD enjoys not only discriminative power to capture the texture information, but also has strong robustness to illumination variation and image rotation. Extensive experiments on benchmark data sets of texture classification and real-world material recognition show that the LOAD yields the state-of-the-art performance. It is worth to mention that we achieve a superior classification accuracy on Flickr Material Database by using a single feature. Moreover, by combining LOAD with Convolutional Neural Networks (CNN), we obtain significantly better performance than both the LOAD and CNN. This result confirms that the LOAD is complementary to the learning-based features.

1. Introduction

Visual image classification [33,34,19,31,9,15] is a challenging problem in computer vision, especially under multiple sources of image transformations, e.g. rotation, illumination, affine and scale variations. The Bag-of-Words (BoW) [5] model, as a powerful intermediate image representation, is the most popular approach in visual categorization in the past decade. In BoW model, the low-level feature extraction and mid-level feature encoding are two most important factors. Recently, some advanced middle-level feature encoding approaches have been proposed, such as Locality-constrained Linear Coding (LLC) [37], Vector of Locally Aggregated Descriptors (VLAD) [14] and Improved Fisher Vector (IFV) [26]. These encoding methods have greatly put forward the development of BoW approach. However, on the other side, the development of low-level feature extraction is slow.

Earlier works on texture description mainly focused on capturing global texture information (e.g. GIST [25]), or fine texture micro-structure (e.g. MR8 filter bank [34], Local Binary Pattern (LBP) [24]). The global texture descriptors can well capture global texture information, but miss most of texture details. For instance, the GIST is good at capturing the spatial layout of scene, but performs poor on simple texture classification task in which the micro-structures are important. These fine texture descriptors defined on very small patches (e.g. 3×3 or 5×5) can well capture small texture structures, but ignore global texture information. For example, the LBP and MR8 perform well on simple texture data sets, but work poor on complex material data sets in which regional texture information is important. There were some works that tried to bridge the gap between these two types of features. However, as we will discuss later, these features may suffer from some limitations, such as sensitiveness to image transformations or limited discriminative power.

This paper aims to provide a powerful regional texture descriptor. To this end, we propose a novel Local Orientation Adaptive Descriptor (LOAD). The proposed descriptor has two important advantages. (i) strong regional texture discrimination: the strong texture discrimination comes from two aspects. Firstly, on single point, we adopt a binary sequence description that owns stronger discriminative power than the gradient orientation description in SIFT [22] and MORGH [7], and Local Intensity Order Pattern (LIOP [38]). Secondly, to enhance the discriminative power of the descriptor, we propose to use a multi-scale description to capture multi-scale texture information. (ii) robustness to image rotation and illumination variation: Due to that the LOAD is defined on an Adaptive Coordinate System (ACS), it is robust to image rotation. Meanwhile, the binary sequence description used in the LOAD affiliates the feature with great robustness to illumination variation.

* Corresponding author.
E-mail addresses: qixianbiao@gmail.com (X. Qi), gyzhao@ee.oulu.fi (G. Zhao), llshen@szu.edu.cn (L. Shen), liqq@szu.edu.cn (Q. Li), mkp@ee.oulu.fi (M. Pietikäinen).

http://dx.doi.org/10.1016/j.neucom.2015.07.142
0925-2312/

Our *first contribution* in this paper is to propose a novel and discriminative texture descriptor, LOAD, and demonstrate its effectiveness on two applications including texture and material classification. On the traditional texture data sets [23,16], the LOAD almost saturates the classification performance. On the real-world Flickr Material Database (FMD) [19], the LOAD achieves superior performance compared to the state-of-the-art approaches.

Our *second contribution* is that we build a new real-world material data set from a newly introduced ETHZ Synthesizability data set that is designed to evaluate the Synthesizability of images. We name the newly introduced data set as OULU-ETHZ. We evaluate and compare the LOAD with the LBP, PRICoLBP and CNN on the OULU-ETHZ. Experiments show that our LOAD achieves promising performance on the new data set.

Our *third contribution* is to experimentally demonstrate that the LOAD shows strong complementary property with the learning based feature, such as Convolutional Neural Networks (CNN) [17,15]. For instance, on the FMD [19], our LOAD combined with the CNN achieves 72.5% that significantly outperforms the CNN (61.2%) and LOAD (65.4%). The strong complementarity is due to that the IFV representation with LOAD and CNN belong to two different approaches: non-structured and structured methods. The former is robust to image rotation and translation, but not well captures the structured information. In contrast, the latter is good at capturing the structured information because its hierarchical max-pooling strategy can preserve the structured information, but is not robust to heavy image rotation and translation.

2. Related works

Local Binary Pattern (LBP) [24] is an effective gray-scale texture operator. Each LBP pattern corresponds to a kind of local structure in natural image, such as flat region, edge, contour and so on. For a pixel (x_c, y_c) in an image I, its LBP can be computed by thresholding the pixel values of its neighbors with the pixel value of the central point:

$$LBP_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} s(g_p - g_c)2^p, s(t) = \begin{cases} 1, & t \geq 0 \\ 0, & t < 0, \end{cases}$$

where P is the number of neighbors and R is the radius. $g_c = I(x_c, y_c)$ is the gray value of the central pixel (x_c, y_c), and $g_p = I(x_p, y_p)$ is the value of its p-th neighbor (x_p, y_p).

Ojala et al. [24] also pointed out that these patterns with at most two bitwise transitions described the fundamental properties of the image, and they called these patterns as "uniform patterns". The number of spatial transitions can be calculated as follows:

$$\Phi(LBP_{P,R}(x_c, y_c)) = \sum_{p=1}^{P} |s(g_p - g_c) - s(g_{p-1} - g_c)|,$$

where g_P equals to g_0. The uniform patterns are defined as $\Phi(LBP(P,R)) \leq 2$. For instance, "11000011" and "00001110" are two uniform patterns, while "00100100" and "01001110" are non-uniform patterns.

The LBP with $P = 8$ has $2^8 = 256$ patterns, in which there are 58 uniform patterns and 198 non-uniform patterns. According to the statistics in [24], although the number of uniform patterns is significantly fewer than the non-uniform patterns, the ratio of uniform patterns accounts for 80–90% of all patterns. Thus, instead of the original 256 LBP, the uniform LBP is widely used in many applications such as face recognition.

3. Local orientation adaptive descriptor

A discriminative texture descriptor should own the following two properties:

Regional texture discrimination: Most descriptors, such as SIFT, HOG 2×2, are designed for image matching or human detection, not especially for texture description, thus their texture discrimination may be limited. Although there exist some effective texture descriptors in literature, such as GIST, LBP, Completed LBP (CLBP), most of them are constructed for a global or fine texture description, thus they ignore regional texture information. In this work, we focus on designing a discriminatively regional texture descriptor.

Robustness to image transformations: Natural images contain rich image transformations, in which rotation and illumination variations are two most common cases. Thus, when designing a feature, these two aspects should be carefully considered.

3.1. Point description

Given similar patches under different image rotations as shown in Fig. 1, our objective is to extract a kind of descriptor that is discriminative and transformation invariant. To achieve rotation invariance, the traditional methods (e.g. SIFT) firstly estimate a reference orientation (also called main orientation), and then align the patch to the reference orientation. However, estimation of the reference orientation will significantly increase the computational cost of the descriptors. Meanwhile, as indicated by [7], the descriptor is sensitive to the error brought by the orientation estimation.

As the circular patch is symmetric with respect to any line across the central point, we choose to sample a circular region around each point. Given a sampled point O (in practice, we densely sample thousands of points from each image), we can obtain a circular patch around the point O. By rotating the patch around the central point O, we can obtain a patch with arbitrary angle as shown in Fig. 1. For any one point in the patch, an Adaptive Coordinate System (ACS) can be formed for it. For example as shown in Fig. 2, for the point A, an ACS is built by the point A and the reference point O. In the ACS of the point A, its horizontal axis x_A is the line across the point A and O, and its vertical axis y_A is the line that is orthogonal to the x_A. Similarly, for the point B, its ACS is built by the point B and O. Under the ACS, the neighboring relationship between point A and its neighbors is invariant to image rotation. It means that, as shown in Fig. 2, the positions of point A's neighbors are always fixed compared to point A. Thus, the pixel values of the A's neighbors are also invariant to image rotation.

Under the ACS, any point in the patch can be encoded in a rotation invariant way. In this paper, we propose a novel Local Orientation Adaptive Descriptor (LOAD) that is built on ACS. As

Fig. 1. Sample patches under different image rotations.