Neurocomputing 194 (2016) 157–167

Contents lists available at ScienceDirect

Neurocomputing

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

Local feature descriptor using entropy rate

Pu Yan, Dong Liang*, Jun Tang, Ming Zhu

School of Electronics and Information Engineering, Anhui University, Hefei 230601, China

ARTICLE INFO

Article history:
Received 22 October 2015
Received in revised form
30 November 2015
Accepted 24 January 2016
Communicated by Bin Fan
Available online 24 February 2016

Keywords:
Entropy rate
Nonsubsampled Contourlet transform
Graph
Random walk
Local descriptor

ABSTRACT

Over the past decades, an increasing number of local feature descriptors have been proposed in the community of computer vision and pattern recognition. Although they have achieved impressive results in many applications, how to find a balance between accuracy and computational efficiency is still an open issue. To address this issue, we present a local feature descriptor using entropy rate (FDER), which is robust to a variety of image transformations. We first employ the nonsubsampled Contourlet transform to produce multiple support regions and design a graph structure to describe the sub-region. We then use the entropy rate of random walks on the designed graph to build the FDER descriptor. Extensive experiments demonstrate the superiority of proposed descriptor dealing with various image transformations in comparison with the existing state-of-the-art descriptors.

1. Introduction

Local feature descriptor is vital for many computer vision tasks such as 3D reconstruction [1], texture classification [2], object recognition [3] and categorization [4], human detection [5] and action recognition [6]. The major problem for descriptor construction is how to find a balance between enhancing the discriminative power and maintaining the robustness to various image distortions. Generally speaking, a typical pipeline of feature descriptor consists of three steps: region detection, descriptor construction and matching. A large number of approaches on region detectors have been proposed and the most representative works include Harris-Affine region [7], Hessian-Affine region [8], Edge-Based region (EBR) and Intensity-Based region (IBR) [9], salient region [10] and maximally stable extremal region (MSER) [11], etc. A comprehensive survey on these detectors is referred to [12].

In order to efficiently distinguish different regions, many local feature descriptors have been presented by using different region detectors [13]. Local feature descriptors are calculated from these detected regions, enabling them to resist to geometric transformations (e.g. viewpoint changes, JPEG compression, image rotation and blur). However, the performance of local feature descriptors may be degraded if simply using one single detected region [14]. In this paper, we employ multiple support regions via nonsubsampled Contourlet transform (NSCT) to overcome this weakness. Furthermore, the entropy rate of random walks on the constructed graph shows some appealing properties. The entropy rate describes the average description length for a stationary ergodic process. The entropy rate, which is the average transition entropy, depends only on the entropy of the stationary distribution and the total number of edges on the constructed graph [15]. It is well defined for all stationary processes and is particularly easy to be calculated [15]. Therefore, we use entropy rate of the random walk to construct the discriminative feature descriptor. Our main contributions are described as follows:

Firstly, in order to further improve the discriminative ability of local feature descriptor, we employ multiple support regions via NSCT. The NSCT produces multiple support regions with different scale, resolution and direction, which contain more useful information for an interest region. Furthermore, two similar regions for non-corresponding point pair can be easily distinguished with different scale, resolution or direction.

Secondly, we use the complete graph and its adjacency matrix of each sample point to represent the relationships among the neighbors of the sample point, which are generated using a rotation invariant coordinate system. In order to improve the effectiveness and reduce the time complexity of our local descriptor, we compute the summation of all the adjacency matrices to describe the sub-region. Obviously, the subsequent operations only need to be performed for the summation of all adjacency matrices rather than all adjacency matrices.

Thirdly, considering the advantage of entropy rate in the mathematical theory of probability, we exploit the entropy rate of random walks on the constructed graph to describe the local region. The entropy rate is a useful metric for determining the complexity and the information content of the constructed graph.

* Corresponding author.
E-mail address: dongliangahu@gmail.com (D. Liang).

http://dx.doi.org/10.1016/j.neucom.2016.01.083
0925-2312/

The rest of this paper is organized as follows. Section 2 gives a brief overview on related works. Section 3 introduces the concept of graph, entropy rate and some weighting functions. Section 4 elaborates the descriptor construction via the NSCT and the entropy rate. Experiments and results on the standard Oxford dataset are reported in Section 5, and finally we conclude the paper in Section 6.

2. Related work

Recently, many researchers have proposed a variety of local feature descriptors by utilizing different image features such as image contour [16], pixel intensity [17], color [18], texture [19] and edges [20]. As described in [21], they can be roughly categorized into three groups: distribution based descriptors, spatial-frequency techniques based descriptors, and differential descriptors. In particular, the distribution based descriptors have received much more attention.

Gradient-based feature descriptors are an important branch of them, in which Scale-Invariant Feature Transform (SIFT) [22] is the most prevalent gradient-based descriptor. The Gradient Location and Orientation Histogram (GLOH) [21], PCA-SIFT [23] and DAISY [24] are the variants of the SIFT descriptor. The GLOH descriptor replaces the 4×4 spatial grid in the SIFT descriptor with a log-polar grid with three radial and eight angular bins, and then utilizes Principal Components Analysis (PCA) to reduce the descriptor size. The PCA-SIFT descriptor applies PCA to the normalized gradient patch, making it more compact and distinctive. The DAISY descriptor can be constructed more efficiently due to the usage of the convolutions of the gradients in specific directions with several Gaussian filters. In the literature [25], the authors evaluated many local feature descriptors with different combinations of local features and spatial pooling strategies. And various descriptors such as SIFT, GLOH, DAISY and spin images [26] could be incorporated into their framework.

Another class is the intensity order-based descriptors, which use the relative order of pixel intensities to construct the local feature descriptors rather than the original intensities. The typical work is Local Binary Pattern (LBP) descriptor [19], which is built by encoding the contrast relation of the neighbors and central pixel point. The advantage of the LBP descriptor is its invariance to the illumination changes and computational simplicity. However, its dimension is extremely high and is sensitive to Gaussian noise in the flat regions. Heikkila et al. [27] proposed a texture CS-LBP (Center-Symmetric Local Binary Pattern) descriptor by considering only diagonal comparisons among the neighboring points. Gupta et al. [28] proposed a Center Symmetric Local Ternary Patterns (CS-LTP) descriptor by combining relative histogram with the LBP histogram together. Furthermore, they concatenated the histogram of relative intensities (HRI) and the CS-LTP descriptor to generate the HRI-CSLTP descriptor, which is robust to Gaussian noise. More recently, Wang et al. [29] have proposed the Local Intensity Order Pattern (LIOP) descriptor, which encodes the intensity order pattern among the neighbors obtained by using a rotation invariant coordinate system. They divided the support region into different sub-regions according to the global ordering of each pixel in the support region and calculated the LIOP in each sub-region separately.

In order to further improve the performance of local feature descriptor, we propose a novel local feature descriptor using entropy rate (FDER) in this paper, which is different from most existing methods. Firstly, multiple support regions with different scale, resolution and direction are generated via the NSCT for the descriptor construction, which can be used to improve the discriminative ability of our proposed descriptor. Secondly, the support region is divided using intensity order, which can avoid boundary errors brought by most dividing strategy based on geometrical position. Thirdly, we introduce the complete graph and its adjacency matrix [30] to represent the sample point in support region and compute the summation of all adjacency matrices to describe the sub-region, which can improve the effectiveness and reduce the time complexity of descriptor construction. Fourthly, we exploit the entropy rate of the random walk on the constructed graph to describe the local region, which can improve the performance of our proposed descriptor. Experimental results show the competitive performance of the proposed descriptor in comparison with several state-of-the-art descriptors under image blur, viewpoint changes, illumination changes, and JPEG compression.

3. Preliminaries

3.1. Graph and entropy rate

An undirected graph is denoted by $G=(V,E,W)$ where V, E and W are the vertex set, edge set and edge weight set, respectively. The notations $|V|$ and $|E|$ denote the cardinality of the vertex set and edge set, respectively. If $e_{ij} \in E$ is an edge between vertices v_i and v_j, then v_i and v_j are adjacency. If any two vertices of G are adjacent, then G is called a complete graph. The weight w_{ij} of edge e_{ij} is acquired by the nonnegative weight function $w: E \to \mathbb{R}^+ \cup \{0\}$. Obviously, the edge weight W is a symmetric matrix. Let $P = v_1 e_{12} v_2 e_{23} \cdots e_{(k-1)k} v_k$, where $e_{(k-1)k}$ is an edge between v_{k-1} and v_k. If $v_0, v_1, \ldots, v_k$ are different, then P is called a path of G, and k is the length of P.

In the information theory, entropy is a measure of unpredictability of information content. Suppose that X is a discrete random variable with a probability mass function p_X. When X is taken from a finite sample, the entropy can be explicitly written as:

$$H(X) = -\sum_{x \in \mathcal{X}} p_X(x) \log p_X(x) \tag{1}$$

Here, $\mathcal{X}$ is the support of the random variable X. Moreover, the conditional entropy of two random variables X and Y is defined as:

$$\begin{aligned} H(X|Y) &= \sum_{y \in \mathcal{Y}} p_Y(y) H(X|Y=y) \\ &= -\sum_{y \in \mathcal{Y}} p_Y(y) \sum_{x \in \mathcal{X}} p_{X|Y}(x|y) \log p_{X|Y}(x|y) \end{aligned} \tag{2}$$

where $\mathcal{Y}$ is the support of Y and $p_{X|Y}$ is the conditional probability mass function.

The entropy rate of a stochastic process is the time density of the average information in a stochastic process. For stochastic processes with a countable index, the entropy rate $\mathcal{H}(X)$ is defined as the limit of the joint entropy of n members of the process X_k:

$$\mathcal{H}(X) = \lim_{n \to \infty} \frac{1}{n} H(X_1, X_2, \ldots, X_n) \tag{3}$$

For strongly stationary stochastic processes, the entropy rate can be written as:

$$\mathcal{H}(X) = \lim_{n \to \infty} H(X_n | X_{n-1}, X_{n-2}, \ldots, X_1) \tag{4}$$

A random walk is a mathematical formalization of a path that consists of a succession of random steps. A random walk $X = \{X_t | t \in T, X_t \in V\}$ on graph G is a stochastic process, and the transition probabilities p_{ij} are defined [31] as:

$$p_{ij} = pr(X_{t+1} = v_j | X_t = v_i) = \frac{w_{ij}}{w_i} \tag{5}$$

where T is an index set and $w_i = \sum_k w_{ik}$ is the sum of incident weights of the vertex v_i. As a result, the stationary distribution is