Optik 126 (2015) 274–278

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

A comparison of feature description algorithms

Jinlong Hu[a,b,*], Xianrong Peng[a], Chengyu Fu[a]

[a] the Institute of Optics and Electronics, Chinese Academy of Sciences, Chengdu 610209, PR China
[b] Graduate University of Chinese Academy of Sciences, Beijing 100049, PR China

ARTICLE INFO

Article history:
Received 25 December 2013
Accepted 15 August 2014

Keywords:
Feature description
Performance evaluation
Algorithm analysis

ABSTRACT

This paper reviews and evaluates several state-of-the-art feature description algorithms. The components of each feature description method are analyzed and their applications in dealing with specific challenges are identified. In the paper, we compare state-of-the-art feature description methods including the SIFT, DAISY, HRI-CSLTP, LIOP, MROGH and MRRID with specific measurement regulation. The quantitative comparative results demonstrate these algorithms' applications in different scenes, which provide a certain guide for designing novel feature description algorithms.

1. Introduction

Feature description is widely used in computer vision, such as object recognition [1] and tracking, texture recognition [13], wide baseline matching [12], image retrieval [14] and panoramic stitching [15]. The basic idea is to firstly detect interesting points or regions of interest, then calculate invariant feature descriptors. Once feature descriptor is obtained, feature matching of different images can be determined based on certain similarity measurement.

A good feature descriptor should have high discriminability so that the descripted feature can be easily distinguished with other features. Meanwhile, it should be robust for possible transformation such as scale, rotation, blur, illumination and viewpoint variant such that correspondences captured in images can be easily matched in different conditions. Therefore, improving discriminability while maintaining robust is a primal factor considered when designing feature descriptors. During recent decades, although many researchers have been devoted to feature extraction and description methods, there is no method can be used for all scenes and tasks. Consequently, it is necessary to review recent feature description methods and evaluate their performance, which may provide certain guide for designing novel feature description method to deal with specific scenes.

Currently, there are mainly three sorts of feature description methods: one is based on intensity value, the second is based on intensity order and the third is based on combination of above two sorts. For the first one, the most famous method is SIFT (Scale Invariant Feature Transform) [1], which creates gradient orientation and location histograms, resulting in a certain scale, rotation and transform invariant. However, the algorithm's performance decreases and has high computation cost when image appears large viewpoint change. Basing on the fundamental, DAISY [2], which is different from SIFT, also depends on gradient histogram and uses Gaussian weight and circular symmetrical kernel to convolve with orientation map, thus dense computation's speed is improved largely. This kind of descriptors based on intensity gradient generally achieves good performance. In addition, there are other methods based on histograms. For example, GLOH (Gradient Location Orientation Histogram) [8] and spin image [13] create a histogram of pixel locations and intensities, and shape context [16] creates a histogram of edge point locations and orientations. However, while the above descriptors have been shown to be fully or partially robust to many of the variations and distortions, they cannot deal with more complex illumination changes including gamma correction, small specular reflections, changes in exposure time, etc. To solve these problems, some researchers have proposed to use the intensity order for feature description. Tang et al. [3] created a 2D histogram encoding both the ordinal distribution and the spatial distribution. Gupta et al. [4] presented a more robust method which contains two parts: a histogram of relative intensities and a histogram of CS-LTP (Central Symmetric Local Ternary Patterns) codes. Subsequently, Wang et al. [5] proposed a novel feature descriptor LIOP (Locality Intensity Order Pattern) based on intensity order, the basic principle of which is that the relative order of pixel intensities remains unchanged when the intensity changes are monotonic. Moreover, Fan et al. [6] developed two descriptors: MROGH (Multi-Support Region Order-based Gradient Histogram)

* Corresponding author at: The Institute of Optics and Electronics, Chinese Academy of Sciences, Chengdu 610209, PR China.

http://dx.doi.org/10.1016/j.ijleo.2014.08.166
0030-4026/

and MRRID (Multi-Support Region Rotation and Intensity Monotonic Invariant Descriptor), which have rotation invariant but not depend on the referenced orientation and have high discriminability.

In order to further exploit the above descriptors' performance, firstly, this paper reviews several state-of-the-art feature description algorithms and analyzes the components of every method and its applications in different scenes. Secondly, the methods' performance is evaluated by testing images in different scenes. Lastly, all methods' characteristics are compared and analyzed by experiments.

The rest of the paper is organized as follows. In Section 2 we introduce and compare several state-of-the-art feature description algorithms. Experiments are shown in Section 3. Section 4 evaluates the methods' performance in different scenes and concludes this paper.

2. Feature description methods comparison

Generally speaking, there are three steps using feature descriptors to match points in images. Firstly, the interesting points or the regions of interest are detected in images. The detected points should be detected and matched between images in different imaging conditions. These points are called interesting points or feature points. Feature points detection usually follows extra procedure which detect affine invariant regions surround interesting points so that large viewpoint changes can be handled. Secondly, feature descriptors are built from above detected regions of interest (affine normalization) such that different features are distinguished. Thirdly, the distance between two candidate descriptors is calculated to determine whether it is a correct match. The following mainly analyze and compare a variety of feature description methods from these three aspects.

2.1. Feature extraction

Presently, many methods have been proposed to detect interest points or interest regions that are covariant with a class of transformations (e.g., affine transformation). For example, Harris corner detector [7] and DOG (Difference of Gaussian) [1] for interest point detection, and Harris-affine [9], Hessian-affine [10], MSER (Maximally Stable Extremal Region) [11] and EBR (Edge-Based Region) [12] for affine covariant region detection. In order to compare all method more fairly, the feature description methods analyzed in this paper all adopt widely used affine covariant region detectors Harris-affine and Hessian-affine to locate feature location and estimate the neighbor's affine shapes. Since the detected regions' scales and shapes are different, they are normalized into fixed diameter circular regions in the paper. To remove the noise that is introduced from difference in above normalization step, Gaussian smoothing filter is used to eliminate the noise influence and local regions are obtained.

2.2. Feature description

2.2.1. Region division

In order to improve discriminability, the above local regions are divided into several sub-regions, whose histograms are then calculated and concatenated to build descriptors. In previous time, most region division methods are based on spatial location. In SIFT, DAISY and HRI-CSLTP, spatial locations are all quantitized into 4×4 grids. The drawback of these methods is that for every local region, a local consecutive orientation must be estimated, then descriptor is built relative to this orientation to obtain rotation invariant. Thus, the performance of the method heavily depends on the accuracy of local concecutive orientation estimation, which is commonly sensitive to noise and deformation. In order to avoid local consecutive orientation estimation, spin image [13] divides the local patch into 5 rings. However, since it only quantizes the spatial location in radial direction, its discriminative power is lower than the grid-shaped region division.

The experiments have shown that orientation estimation based on histogram will result in matching error, which decreases descriptors' performance greatly. In order to improve descriptors' discriminability, LIOP [5] divides regions based on intensity order. All the pixels in the local patch are first sorted by their intensities in a nondescending order. Then, the local patch is equally quantized into B ordinal bins according to their orders. Thereby, it is not only invariant to monotonic intensities changes and image rotation, but also contains much more spatial information than the ring-shaped region division. Likewise, MROGH and MRRID [6] also divide regions based on sub-regions. In this case, the sample points in every group do not need to be neighbors, and there is no need for this adaptive division to assign a reference orientation, which improves the descriptors' discriminability greatly.

2.2.2. Descriptor construction

After region division, SIFT and DAISY calculate the pixels' gradient orientation relative to main orientation in every grid respectively, generating an 8 bins gradient histogram, then by concatenating them together to obtain 128 dimensions' descriptor vector.

HRI-CSLTP combines intensity and intensity order, adopting two technologies: histogram of relative intensity (HRI) and Central Symmetric Local Ternary Patterns (CS-LTP). For the first one, the intensity range is first divided into k equal intervals according to local patches' intensities starting-point and end-point, and every interval is obtained. Then, the local patch is divided into $s \times s$ spatial bins. For every bin, histogram is constructed according to pixels' intensity range, thus $s \times s \times k$ bins are acquired. Note that it acts on patch's global distribution and cannot capture local gradient information, which has a complementary with global order information. Thereby, the later mainly acts on local gradient information, which adopts the third value to represent the pixels having almost the same intensity values in the fundamental of CS-LBP. However, if adopting comparative method similar to CS-LBP, histogram of 81 bins will be obtained. In order to minimise histogram's size, two comparisons are considered. Because of feature normalization and common image features, diagonal comparison is only used to generate CS-LTP coding, then for every spatial bin, 9 bins histogram is obtained. For the code whose value is 1, that is to say two matching points intensities have little difference, it prefers to drifting, having much small weights. Consequently, the number of every spatial bin is reduced to 8, generating $s \times s \times 8$ dimensions CS-LTP descriptor. These two histograms are concatenated together to form final descriptor. However, since the above descriptors only compare intensities of centric symmetrical neighboring sample points, the neighboring sample points' intensities' relations are not captured. Moreover, a referenced orientation should be assigned to obtain rotation invariant, resulting in its sensitivity to orientation estimation.

In order to overcome above problems, LIOP uses the overall neighboring sample points' intensity orders to exploit local information. Furthermore, rotation invariant sampling method is used to avoid the error brought by local consecutive orientation, thus obtaining much higher discriminability. However, for a specific support region, when two non-correspondences have similar appearance models, LIOP may regard it as a correspondence, thus discriminability disappears. In order to further improve discriminability, MROGH and MRRID adopt multi-regions which have different sizes to build descriptors. Similar to LIOP, they use the method based on intensity order and rotation invariant sampling