Pattern Recognition 76 (2018) 162–174

Contents lists available at ScienceDirect

Pattern Recognition

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

A rotationally invariant descriptor based on mixed intensity feature histograms

Yi Yang, Fajie Duan*, Ling Ma

State Key Laboratory of Precision Measurement Technology and Instruments, Tianjin University, No.17 building, No. 92 Weijin Road, Nankai District, Tianjin 300072, China

ARTICLE INFO

Article history:
Received 4 January 2017
Revised 21 October 2017
Accepted 30 October 2017
Available online 31 October 2017

Keywords:
Rotation invariant
Reference orientation
Mixed intensity feature histogram
SIFT
Image matching

ABSTRACT

This paper proposes a novel method to build a rotation invariant local descriptor by mixed intensity feature histogram. Most existing local descriptors based on intensity ordinal information typically encode only one local feature for each sampling point in the image patch. To address this problem, we proposed a method to encode more than one different local features for each pixel in the image patch and construct a 2D mixed intensity feature histogram, from which our proposed MIFH descriptor is then constructed. Since the rotation invariant coordinate system is adopted, the MIFH descriptor does not need to estimate the reference orientation. In order to evaluate the performance and the robustness of the MIFH with other well-known local descriptors (e.g., SIFT, GLOH, DAISY, HRI-CSLTP, LIOP, MROGH), image matching experiments were carried out on standard Oxford dataset, additional image pairs with complex illumination changes and image sequences with different noises. To further investigate the discriminative ability of the MIFH descriptor, a simple object recognition experiment was conducted on three public datasets. The experimental results demonstrate that our descriptor MIFH exhibits better performance and robustness than other evaluated local descriptors.

1. Introduction

Image matching is the key task of many computer vision applications, such as object recognition [1], texture recognition [2], wide baseline image matching [3], image retrieval [4], panoramic image stitching [5] and structure motion [6–10]. The detection of interest points or interest regions plays an important role in some image matching tasks. Therefore, many different approaches for local feature detection have been proposed [3,11–14].

A good local descriptor should have high discriminative ability and rotation invariance so that the detected local feature can be easily distinguished from other features. To achieve this goal, a number of methods for local descriptor construction have been proposed by researchers in the past decades [1,15–23]. Most descriptors are built based on the histogram, which could encode a compressed spatial information of image patch and have a certain robustness to perturbations including noises and localization error. Along this way, many floating-point descriptors [1,21,24,25] were constructed from the histograms of locations and gradient orientations. SIFT (Scale Invariant Feature Transform) [1] is probably the most famous one. It is a 3D histogram of gradient location and orientation where the contribution to the location and orientation bin is weighted by the gradient magnitude. Inspired by SIFT, many of its variants have been developed by researchers. Ke and Sukthankar [21] proposed the PCA-SIFT descriptor which uses PCA (Principal Component Analysis) method to reduce the size of SIFT. It was said to be more distinctive and compact than SIFT. The GLOH (Gradient Location and Orientation Histogram) descriptor [24] also applies PCA method to reduce the dimension of the descriptor, and uses the log-polar location grid to replace the Cartesian location grid. The SURF (Speed Up Robust Feature) descriptor [15] uses a Hessian matrix-based measure for the detector and Haar Wavelet responses for the descriptor. The computation time of SURF is significantly reduced by using integral image instead of image convolutions. Moreover, Tola et al. [25] developed a fast descriptor named DAISY for dense matching, Winder and Brown [26] proposed a framework to learn local descriptors with different combinations of local features and feature pooling schemes. However, such descriptors can only deal with linear illumination changes, and cannot handle complex illumination changes such as gamma correction and exposure time changes. In order to alleviate this problem, many methods based on intensity ordinal information have been proposed by researchers Wang and co-workers [20,19,27]. However, the methods mentioned above usually need to estimate a reference orientation to achieve its rotation invariance. And, the orientation estimation

* Corresponding author.
E-mail addresses: johnyang_tju@hotmail.com (Y. Yang), fjduan@tju.edu.cn (F. Duan), tinama_17116@tju.edu.cn (L. Ma).

https://doi.org/10.1016/j.patcog.2017.10.035
0031-3203/

is an error-prone process, which might result in the mismatch of the corresponding features. This fact has been proven in [20,28]. To address this problem, some rotation invariant local descriptors have been proposed in the literatures [20,19,28]. In their experiments, these descriptors showed better performance than those require a reference orientation. However, most intensity ordinal information based local descriptors typically encode only one local feature for sampling point in the image patch.

In this paper, we firstly proposed a new binary local feature based on intensity relative ordinal information of neighboring points around one sampling point in the image patch, namely Local Intensity Relative Order Pattern (LIROP). In order to further improve the robustness and the discriminative ability of the local descriptor, we proposed a novel method to encode more than one different local features for each sampling point in the image patch. Finally, a rotation invariant local descriptor was built from the 2D histogram, namely Mixed Intensity Feature Histogram (MIFH).

The rest of the paper is organized as follows. We firstly give a brief overview of related work in Section 2. In Section 3, our proposed methods are illustrated in detail. The experimental evaluations are given in Section 4. Finally, we conclude this paper in Section 5.

2. Related work

In this section, we briefly introduce the well-known methods for local feature detection and local feature description.

2.1. Local interest features

The local features from an image pair should be detectable and matchable under different capture conditions. At present, interest points and interest regions are two widely used local features. FAST (Features from Accelerated Segment Test) [29] and Harris points [12] are widely used in computer vision tasks, such as parallel tracking and mapping [30]. Meanwhile, some famous descriptors [1,15] also involve key points detector. Computer vision system using local feature point shows poor performance in the case of large viewpoint changes [31]. To alleviate this problem, many detection methods for affine covariant regions were presented in literatures [14,24]. These interest regions show good robustness in the case of the large viewpoint changes. The Harris-affine region, Hessian-affine region and MSER (Maximally Stable Extremal Regions) are three widely used local feature regions.

2.2. Local feature descriptors

Recently, many local descriptors based on the intensity ordinal information of pixels have been proposed in literatures. LBP (Local Binary Pattern) operator proposed by Ojala et al. [32] exhibits good robustness under the condition of light changes. It has been widely used in various vision tasks, such as face recognition [33] and background subtraction [34]. Inspired by LBP, Heikkila et al. [27] proposed the CS-LBP (Center-Symmetric Local Binary Pattern) operator by comparing the intensity of centrosymmetric pixels. Thus, the histogram length generated by the CS-LBP operator is only half that of the LBP operator. The CS-LBP descriptor is said to have similar performance to SIFT. The calculation process of the LBP operator and CS-LBP operator is shown in Fig. 1.

Goswami et al. [35] presented another LBP-based operator named LOCP (Local Ordinal Contrast Pattern) by computing pairwise ordinal information from adjacent circular neighborhoods to lip-based speaker authentication. Gupta et al. [17] presented a variant of the CS-LBP operator named CS-LTP (Center-Symmetric Local Ternary Patterns) descriptor, which build a histogram of central symmetric ternary pattern.

These descriptors mentioned above can only handle liner illumination changes. To address this problem, Wang et al. [19] proposed an LIOP (Local Intensity Order Pattern) operator to encode the ordinal information of neighboring points around one pixel. In fact, LIOP operator firstly sorts the neighboring points around one sampling point according to the gray scale intensity, and then obtains the LIOP feature by looking up the index table containing the sorting information of neighboring points. It can be seen from the above analysis that the LIOP operator encodes the absolute ordinal information of neighboring points. The illustrative diagram for the LIOP operator is shown in Fig. 2.

Fan et al. [20] proposed descriptors by using intensity order pooling, and applied the multiple support regions technology to enhance their robustness. Tang et al. [36] presented an OSID (Ordinal Spatial Intensity Distribution) descriptor built from 2D histogram that encoded both the spatial and ordinal distribution of pixel intensities.

In order to achieve rotation invariance, most local descriptors need to compute a reference orientation. Since the estimation error of the reference orientation might cause the corresponding point mismatch, researchers have proposed different methods to achieve rotation invariance of local descriptor without calculating the reference orientation. For example, the LBP outputs different values when image is rotated. Ojara et al. [37] obtained the smallest value of LBP outputs by rotating the image patch to achieve its rotation invariance. Wang et al. [19] proposed a rotation invariant coordinate system to achieve the rotation invariance of the local descriptor. As shown in Fig. 3, suppose P is the center pixel (interest point) of an image patch (circular region in the Fig. 3) used for constructing local descriptor, S_0 is one pixel in the image patch, and n_i are the neighboring points of S_0 . Then the local xy coordinate system can be established for S_0 in the image patch by setting $\vec{PS_0}$ as the positive x-axis. When the image patch is rotated, the position relationship between the sampling point S_0 and its neighboring point n_i is obviously unchanged. Thus, the local descriptor constructed on this coordinate system is inherently rotation invariant.

Moreover, most existing local descriptors only encode one local feature for each sampling point in the image patch. For example, in [19] Wang et al. encoded one LIOP feature for each pixel in the image patch. The LBP and its variants (e.g., CS-LBP) also encoded one local feature for each sampling point in the support region. Note that these local descriptors usually lose spatial information, thus might make the discriminative ability descend. To address this problem, OSID descriptor [36] first sorted all pixels in the imaged patch according to the their intensity order, and assigned an ordinal number for each pixel. Then, the image patch was divided into several pies. Finally, a 2D histogram encoding the ordinal and spatial information of the pixel was constructed from these pies. However, the OSID descriptor does not encode the local feature for each pixel in the image patch.

3. Our approaches

In this section, we firstly introduce our proposed LIROP operator in detail. The calculation process of MIFH (Mixed Intensity Feature Histogram) is then illustrated in Section 3.2. Finally, we show how to construct the MIFH descriptor in Section 3.3.

3.1. Local intensity relative order pattern

The LIOP operator obtains local feature by sequencing the intensity of neighboring points around one pixel. However, its discriminative ability might descend on the smoothed image patch. To address this problem, a novel local feature operator namely LIROP