



# An improvement to the SIFT descriptor for image representation and matching

Kaiyang Liao, Guizhong Liu\*, Youshi Hui

School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China

## ARTICLE INFO

### Article history:

Received 18 July 2011

Available online 3 April 2013

Communicated by S. Sarkar

### Keywords:

Local descriptor

SIFT

Point matching

## ABSTRACT

Constructing proper descriptors for interest points in images is a critical aspect for local features related tasks in computer vision and pattern recognition. Although the SIFT descriptor has been proven to perform better than the other existing local descriptors, it does not gain sufficient distinctiveness and robustness in image match especially in the case of affine and mirror transformations, in which many mismatches could occur. This paper presents an improvement to the SIFT descriptor for image matching and retrieval. The framework of the proposed descriptor consists of the following steps: normalizing elliptical neighboring region, transforming to affine scale-space, improving the SIFT descriptor with polar histogram orientation bin, as well as integrating the mirror reflection invariant. A comparative evaluation of different descriptors is carried out showing that the present approach provides better results than the existing methods.

© 2013 Elsevier B.V. All rights reserved.

## 1. Introduction

Local interest point matching is matching the corresponding points between two or more images. It has proven to be very successful in many pattern recognition and computer vision tasks such as wide baseline matching, object recognition and tracking, texture recognition, image retrieval and reconstruction, robot localization, video data mining, building panoramas, stereo correspondence, recovering camera motion, and recognition of object categories (Tuytelaars and Van Gool, 2004; Shin and Tjahjadi, 2010; Zhang and Wang, 2011; Wu and Rehg, 2011; Arican and Frossard, 2012). The two essential aspects of local interest point matching are detection and description of interest points (Li and Ma, 2009). The detection of interest points determines the reliable feature points that are used to match, and at the same time determines the proper neighboring regions that are used in computing the descriptors. The description of an interest point involves creating a distinctive and robust descriptor for it.

For a good feature descriptor, two criteria should be considered in describing the extracted feature points (Li and Ma, 2009). The first is the distinctiveness, which means that the extracted feature should have enough information to distinguish the interest points, so the descriptor of an interest point must be as discriminative as possible. The other criterion is the robustness to resist photometric and geometric deformations. These feature descriptors should be robust with respect to photometric changes such as illumination direction, color, highlight, and intensity change. The extracted features should also be invariant to different geometric variations

such as rotation, translation, scaling, mirror reflection and even viewpoint change.

Many different descriptors for interest points have been developed and have proven to be very successful in applications. Excellent reviews on the existing descriptors can be found in Li and Allinson (2008), Brown et al. (2011), Florindo et al. (2012). These descriptors can be divided into five classes:

The first class is the distribution-based descriptors. These techniques used histograms to represent different characteristics of the shape or appearance. Belongie et al. (2002) proposed a shape context descriptor, which at a reference shape point captures the distribution of the remaining points relative to it. Carneiro and Jepson (2002) proposed a phase-based local feature which is based on the phase and amplitude responses of complex-valued steerable filters. Lowe (2004) proposed a scale invariant feature transform (SIFT), which combines a scale invariant region detector and a descriptor based on the gradient distribution in the corresponding regions. Several attempts to improve the SIFT descriptor have been reported in the literature. The PCA-SIFT (Ke and Sukthankar, 2004) descriptor is an extension of the SIFT descriptor, which reduces the dimension of the SIFT descriptor vector from 128 to 36 using PCA. The GLOH (Mikolajczyk and Schmid, 2005) is also an extension of the SIFT descriptor designed to increase its robustness and distinctiveness. Morel and Yu (2009) proposed an affine SIFT, which simulates all the distortions caused by variations in the direction of the camera's optical axis, and then the SIFT is imposed on the simulated images. Guo et al. (2010) presented a mirror reflection invariant descriptor (MIFT) which is inspired from SIFT.

The second class is the differential-based descriptors. These descriptors employ a set of image derivatives computed up to a given order in a point neighborhood. Florack et al. (1991) proposed a

\* Corresponding author. Tel./fax: +86 29 82667836.

E-mail addresses: [liugz.xjtu@gmail.com](mailto:liugz.xjtu@gmail.com), [liugz@xjtu.edu.cn](mailto:liugz@xjtu.edu.cn) (G. Liu).

descriptor based on the differential invariants, which combines components of the local derivatives to obtain rotation invariance. Schmid and Mohr (1997) described the interest points using local differential gray-level invariants, and the descriptors are invariant to scale, intensity, and rotation transformations.

The third class is the filter-based descriptors. The steerable filter descriptor (Freeman and Adelson, 1991) employed quadrature pairs of derivatives of Gaussian and Hilbert transforms to synthesize any filter of a given frequency with arbitrary phase. The Gabor filter descriptor (Lee, 1996) used a set of Gabor filters tuned to various frequencies and orientations to represent the image patterns. Baumberg (2000) proposed a complex filter which uses the Gaussian derivatives. Moreno et al. (2009) improved the SIFT descriptor with the Gabor smoothing derivative filters. Gómez and Romero (2011) introduced a curvelet based descriptor which is calculated from the statistical pattern of the curvelet coefficients.

The fourth class is the color-based descriptors. It makes use of the color invariance robust against varying imaging conditions. Gevers and Smeulders (1999) proposed some new color models for the purpose of recognition of multicolored objects. Diplaros et al. (2006) described a method to merge the color and shape invariant information in the context of object recognition and image retrieval. Abdel-Hakim and Farag (2006) introduced the CSIFT as a colored local invariant feature descriptor. Stokman and Gevers (2007) proposed a generic selection model to select and weight the color invariant models for discriminatory and robust image feature detection. Verma et al. (2011) presented new color SIFT descriptors, which extended the SIFT descriptor to different color spaces.

The fifth class is other descriptors. Apart from the above basic descriptor types, there are also other extended descriptors. Ojala et al. (2002) proposed a local binary pattern (LBP) by building statistics on the local micropattern variations. Chen et al. (2010) developed a Weber local descriptor (WLD) based on the perception of human beings, which is robust to noise. Chen and Sun (2010) presented a new image descriptor to represent the normalized region, which primarily comprises the Zernike moment (ZM) phase information.

The SIFT descriptor is one of the most successful and popular local image descriptor among all the above mentioned descriptors. It has been proven to perform better than the other local invariant feature descriptors (Mikolajczyk and Schmid, 2005) until recent time. However, the SIFT descriptor is neither mirror reflection invariant nor completely invariant to the viewpoint change, and its scale and rotation invariance is not so exact for digital images. In this paper, we propose to improve on the SIFT descriptor by considering all the above mentioned disadvantages. Firstly, a normalized elliptical neighboring region is used to enhance the invariance to viewpoint change. Secondly, the affine scale-space is applied to increase the scale invariance. Thirdly, the polar histogram orientation bin is used to improve the rotation invariance. Finally, rearranging the descriptor is used to ensure the mirror reflection invariance.

The remainder of this paper is organized as follows. Section 2 presents our proposed algorithm. Section 3 introduces the evaluation criteria and the data sets for the experiments. In Section 4 the experimental results and analysis are provided. Finally, the paper is concluded in Section 5.

## 2. An improvement to SIFT local region descriptor of images

### 2.1. Normalizing elliptical neighboring region

Recently, several researches in the literature have focused on improving local features to be invariant to the viewpoint change.

Mikolajczyk and Schmid (2004) proposed Harris-Affine and Hessian-Affine to obtain invariance to viewpoint change by the affine adaptation process based on the second moment matrix. They make effort to obtain the affine invariance in the stage of feature detection. Morel and Yu (2009) proposed an affine SIFT, which simulates all the distortions caused by variations in direction of the camera's optical axis, and then the SIFT is imposed on the simulated images. Since it has to simulate all the distortions, the affine SIFT will extract many extra feature points which reduce the efficiency of image matching and retrieval. We are going to propose a descriptor to obtain the invariance to viewpoint change by using the normalized elliptical neighboring region in the stage of feature description.

For each interest point, the SIFT descriptor computes the dominant histogram of gradient orientations in a small circular neighboring region around the point. The size of the circular region is determined by the scale of the interest point, but the shape of the neighboring region is not actually invariant to affine transformations for the different structures of neighboring regions (Li and Ma, 2009). For example, the image structure in the red circular region in Fig. 1(a) can be mapped to an elliptical region in Fig. 1(b) after an affine transformation. If we placed the same circle around the feature point in the transformed image which contains an elliptical region, there will be additional image structures in the circular region that will distort any invariant measures calculated, see the red circle in Fig. 1(b). We can see that the same image structure exists in the yellow elliptical region. Since the areas within the two red circles in Figs. 1(a) and (b) do not have the same image structures, using a circular region in the situation of large affine transformation will not produce a robust and distinctive feature descriptor. So we first get an elliptical neighboring region around the interest point, and then normalize the elliptical region to a circular one. We compute the histogram of dominant gradient orientations in the normalized circular neighboring region around the point.

The second moment matrix is often used to detect feature or to describe local image structures. We can use the second moment matrix to estimate the elliptical neighborhood of an interest point (Baumberg, 2000; Lindeberg and Gårding, 1997). After an affine transformation, the second moment matrix  $\mu$  at a given interest point  $x$  is defined by:

$$\mu \left( x; \sum_I, \sum_D \right) = \begin{pmatrix} \mu_{11} & \mu_{12} \\ \mu_{21} & \mu_{22} \end{pmatrix} = \det \left( \sum_D \right) g \left( x; \sum_I \right) * \left( \nabla L \right) \left( x; \sum_D \right) \left( \nabla L \right) \left( x; \sum_D \right)^T \quad (1)$$

where  $\sum_D$  is a symmetric positive definite  $2 \times 2$  matrix corresponding to the local scale and  $\sum_I$  is a symmetric positive definite  $2 \times 2$

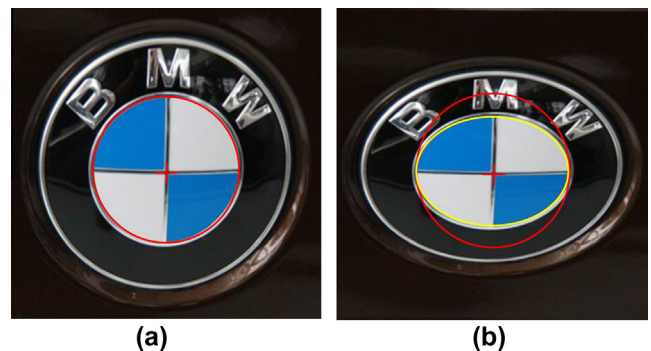


Fig. 1. Circular and elliptical neighboring regions. (a) The image structure in the circle. (b) The mismatched circular region and the matched elliptical region corresponding to the circle in (a) after an affine transformation.

Download English Version:

<https://daneshyari.com/en/article/534016>

Download Persian Version:

<https://daneshyari.com/article/534016>

[Daneshyari.com](https://daneshyari.com)