A new descriptor resistant to affine transformation and monotonic intensity change ☆

Zeyi Huang [a,b], Wenxiong Kang [a,*], Qiuxia Wu [a], Xiaopeng Chen [a]

[a] College of Automation Science and Engineering, South China University of Technology, Guangzhou 510641, China
[b] State Key Laboratory of Industrial Control Technology, Department of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China

A B S T R A C T

A substantial number of local feature extraction and description methodologies have been proposed as image recognition algorithms. However, these algorithms do not exhibit adequate performance with regard to repeatability, accuracy, and time consumption for both affine transformation and monotonic intensity change. In this paper, we propose a new descriptor, named Resistant to Affine Transformation and Monotonic Intensity Change (RATMIC). Unlike traditional descriptors, we utilize an adaptive division strategy and intensity order to construct the new descriptor, which is actually resistant to affine transformation and monotonic intensity change. Extensive experiments demonstrate the effectiveness and efficiency of the new descriptor compared to existing state-of-the-art descriptors.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

The task of finding correspondences between two images has been a prevalent topic for years. However, even for the simplest objects or scenes, computers still experience difficulty fulfilling their missions of recognition. With the emergence of local invariant features, it is possible to achieve better performance and improved efficiency with several successful applications, including object recognition [1], 3D object reconstruction [2], and image retrieval [3], as well as panoramic image stitching [4]. Moreover, this methodology can be divided into two main steps: the extraction of features and the description of these features. Several algorithms have been proposed for both steps; some are reviewed in the next two subsections.

1.1. Detectors

Among these invariant feature extraction methodologies, the Harris corner detector [5], which is based on the computation of the eigenvalues of the Harris matrix, is used to detect corners by curvature, and the Hessian detector, which employs the Hessian matrix, is used to retrieve conspicuous texture information from images. Nevertheless, neither is invariant to affine transformation, which reduces their performance. Thus, the Harris-Affine detector [6–8] and Hessian-Affine detector [6,7] are proposed based on the second moment matrix and automatic scale selection theory [9].

The Maximally Stable Extremal Region (MSER) [10] detects the maximally stable extreme districts to form the affine invariant areas. Based on the Accelerated Segment Test (AST) standard, the Features from AST (FAST) [11,12] utilizes machine learning to establish a decision tree for corner detection. Based on FAST, Mair et al. [13] present the Adaptive and Generic Accelerated Segment Test (AGAST), which finds the optimal decision tree in an extended configuration space and also demonstrates how to combine specialized trees to yield an adaptive and generic AST.

1.2. Descriptors

The SIFT descriptor [1,14] is the most popular descriptor that it is invariant to scale transformation by means of an image pyramid; it also undergoes robust-to-rotation transformation and intensity change through gradient information. However, it presents certain disadvantages, which will be discussed in detail in Section 2. By utilizing both the integral image and Haar wavelet transformation, the SURF descriptor [15,16] exhibits higher efficiency but lower performance than the SIFT descriptor. By combining the FAST detector with the BRIEF descriptor [17], the ORB algorithm [18] is faster than the SIFT and SURF algorithms by two orders and one order of magnitude, respectively; however, it displays poor performance when a serious affine transformation occurs. The DAISY descriptor [19] performs well for affine transformations and linearly uniform intensity changes. However, due to high dimensions and complex computations, time and space consumptions for the construction of the DAISY descriptor increased accordingly. The Rotation-Invariant Fast Feature (RIFF) [20] descriptor calculates the gradient information based on a local coordinate

---

☆ This paper has been recommended for acceptance by Tinne Tuytelaars.
* Corresponding author. Fax: +86 20 87114612.
   E-mail addresses: auwxkang@scut.edu.cn, Kangwenxiong@163.com (W. Kang).

such that it is invariant under rotation but the descriptor may result in increased computational complexity. The Compact And Real-time Descriptors (CARD) [21] makes effective use of query tables to reduce the construction time of the SIFT algorithm; however, it is still inferior to the SIFT algorithm. To cope with complicated intensity change, the Local Binary Pattern (LBP) descriptor [22] considers intensity order information but still contains high dimensions. Heikkilä et al. [23] combine the SIFT and LBP descriptors to form the Center Symmetric LBP (CS-LBP) descriptor, which can efficiently deal with complex intensity change but experiences the same disadvantages as the SIFT descriptor. For better coping with noises, Gupta et al. [24] presents the Center Symmetric Local Ternary Patterns (CS-LTP) descriptor to establish a dead zone for comparisons of pixel intensity.

As discussed previously, development of a new descriptor that can efficiently and effectively handle both affine transformation and complex intensity change is essential. Consequently, we propose the Resistant to Affine Transform and Monotonic Intensity Change descriptor (RATMIC), whose synthetical performance outperforms most of the existing state-of-the-art descriptors. The remainder of the paper is organized as follows. Section 2 discusses the shortages of SIFT-based descriptors. Section 3 introduces our proposed RATMIC descriptor in detail. Case studies are discussed in Section 4, and our conclusions are presented in Section 5.

2. The disadvantages of SIFT-based descriptors

As stated previously, the SIFT algorithm is the most popular algorithm in the fields of invariant feature extraction and description. Consequently, there are many new algorithms deriving from the SIFT algorithm, e.g., SURF, DAISY, and CS-LBP, that are worthy of emphasis. Based on our analyses, we present some disadvantages of SIFT-based descriptors in this section, including extensive time consumption, lack of invariance to non-uniform intensity changes, and reduced accuracy due to predefined subregions.

2.1. Extensive time consumption for descriptor construction

Fig. 1 depicts the distribution of time consumption during feature extraction and description by the SIFT algorithm. For the image with a size of 700×1000 pixels, 3354 features are extracted from the raw image. The overall process takes 13,588 ms, including 1653 ms for features extraction, 2840 ms for orientations estimation and 9095 ms for descriptors extraction in the experimental environment of Intel Core (TM) i3-2100 CPU 3.10 GHZ.

As shown in Fig. 1, the most time-consuming stages are the estimation of the principal direction and the construction of the descriptors, which together comprise about 88% of the total process time. The reasons for this proportion are fourfold. First, the gradient computation over the entire region of interest (ROI) for each feature is very large, i.e., provided that there are J ($J \in N^+$) features extracted from the raw image and ROIs centered at different

features are of an n^2 ($n \in N^+$) size, the computation complexity is up to $O(Jn^2)$, where $O(\cdot)$ is the asymptotic time complexity. Second, SIFT-based algorithms divide gradient orientation into several directions; all pixels are then pooled according to their gradient orientations. Subsequently, the direction with the largest magnitude is assigned as the principal direction. Third, to become rotation invariant, the rotation of the ROI relative to the principal direction is required, which consists of a large number of bilinear interpolations. Finally, SIFT-based algorithms divide the ROI into several predefined subregions and construct the descriptor based upon the gradient information of pixels. According to the previous analyses, Most of SIFT-based descriptors are not suitable for real-time applications because they are overly time consuming.

2.2. Lack of invariance to non-uniform intensity changes

The construction of SIFT-based descriptors depends on the gradient computation of pixels in the ROI. If the same intensity increment occurs in the image, the gradient information is invariant, i.e., all pixels in the original image change from $I(x,y)$ to $I(x,y) + \Delta(x,y)$, where $I(x,y)$ is the intensity located at (x,y) and Δ is a compatible matrix with a same element representing an equivalent intensity change. This change does not affect the gradient result for they are computed from the pixel differences. Thus, SIFT-based descriptors are invariant to the same incremental intensity change.

A change that every pixel multiplies a constant, i.e., the intensity changes from $I(x,y)$ to $k \times I(x,y)$, where k is a positive scalar, will multiply the gradient value by the identical constant. To counteract this effect, SIFT-based descriptors are unitized by the Euclidean distance.

So the SIFT-based algorithms are capable of handling linear uniform intensity change, i.e., intensity changes from $I(x,y)$ to $k \times I(x,y) + \Delta$. Nevertheless, when a non-uniform intensity change occurs, which can be represented as $f(x,y) \times I(x,y) + \Delta'(x,y)$, where the scale coefficient $f(x,y)$ is a general matrix instead of a scalar and translation coefficient $\Delta'(x,y)$ is a matrix with different elements, the gradient computation is unable to eliminate the effect caused by the non-uniform intensity changes. To reduce the influence, D. Lowe thresholds the maximum gradient value and then renormalizes the SIFT descriptor. However, such changes will still affect the construction and performance of SIFT-based descriptors.

2.3. Accuracy reduction due to predefined subregion division

To take spatial information into consideration, SIFT-based descriptors divide the ROI into several predefined subregions, e.g., the SIFT descriptor divides the ROI into 4×4 square subregions. Unfortunately, SIFT-based descriptors need to determine the principal direction and rotate the ROI relative to the principal direction to obtain rotational invariance. However, this method not only augments time consumption but also reduces accuracy. Its instability was previously pointed out by Bin et al. [25].

3. RATMIC descriptor

Aimed at overcoming the aforementioned disadvantages of SIFT-based descriptors, we propose the RATMIC descriptor, which is constructed by the intensity order instead of gradient information. It will improve repeatability and accuracy, as well as decrease time consumption. The construction process can be divided into four steps. First, we employ the affine-invariant detector to extract features from raw images. Second, we adopt a strategy, which is invariant to rotation transformation, to pool pixels in the ROI into different subregions. Third, a *ratmic* function is applied to different subregions to form their own RATMIC subdescriptors. Finally, all subdescriptors
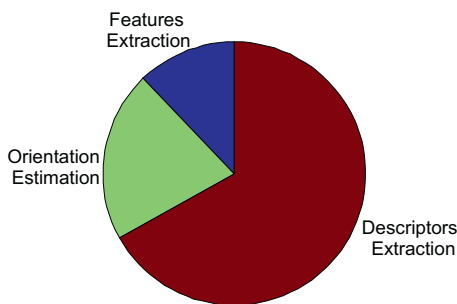

Fig. 1. Distribution of time consumption using the SIFT algorithm.