Neurocomputing 85 (2012) 6–10

Contents lists available at SciVerse ScienceDirect

Neurocomputing

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

Fully affine invariant SURF for image matching

Yanwei Pang [a], Wei Li [a], Yuan Yuan [b,*], Jing Pan [c]

[a] School of Electronic Information Engineering, Tianjin University, Tianjin 300072, China
[b] Center for OPTical IMagery Analysis and Learning (OPTIMAL), State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, Shaanxi, China
[c] School of Electronic Engineering, Tianjin University of Education and Technology, Tianjin 300222, China

A R T I C L E I N F O

Article history:
Received 4 July 2011
Received in revised form
23 November 2011
Accepted 14 December 2011
Communicated by Xiaofei He
Available online 10 January 2012

Keywords:
Feature extraction
Image matching
FAIR-SURF
SURF

A B S T R A C T

Fast and robust feature extraction is crucial for many computer vision applications such as image matching. The representative and the state-of-the-art image features include Scale Invariant Feature Transform (SIFT), Speeded Up Robust Features (SURF), and Affine SIFT (ASIFT). However, neither of them is fully affine invariant and computation efficient at the same time. To overcome this problem, we propose in this paper a fully affine invariant SURF algorithm. The proposed algorithm makes full use of the affine invariant advantage of ASIFT and the efficient merit of SURF while avoids their drawbacks. Experimental results on applications of image matching demonstrate the robustness and efficiency of the proposed algorithm.

1. Introduction

Image feature extraction and description are the basis of image and video analysis [1–3]. Intensive research shows that local image features are more robust and stable than the global one. Local image features are commonly employed in image matching, image retrieval [3], object detection, scene recognition, etc. All of the local image detectors, such as Harris [4], Harris-Laplace, Hessian-Laplace, DoG [7], Hessian-Affine [5], Harris-Affine, Maximally Stable extremal region (MSER) [6], are translation invariant. Among them, the first detector is also rotation invariant. The following three are scale and rotation invariant. The last three detectors are designed to be invariant to affine transformations. But they are obtained by normalizing the local regions, patches and so on. None of these approaches are yet fully affine invariant. Recently, ASIFT has been proposed [8]. It simulates all possible image views and thus is fully affine invariant. ASIFT can handle much higher transition tilts than SIFT [7], Harris–Affine, Hessian–Affine and MSER. But its complexity is about twice the complexity of SIFT. So on account of it, ASIFT does not favor the real-time application.

In order to reduce the computation time of SIFT, there were proposed many improved versions, such as PCA-SIFT [9], GLOH [10] and SURF [11]. By using integral images and box filters, SURF reduces the computation time and improves the speed of detection. Moreover, SURF's detector and descriptor are not only faster, but the detector is also more repeatable and the descriptor more distinctive. Many research results, such as CenSurE [12] and SUSurE [13], are based on the SURF algorithm. Since the best results in [12,13] were obtained using SURF, this paper focuses on that algorithm. Though SURF shows its potential in a wide range of computer vision applications, it also has some shortcomings. When 2D or 3D objects are compared, it does not work if rotation is violent or the view angle is too different. Inspired by ASIFT, to deal with the view angle problem of SURF, the proposed method in this paper simulates possible affine distortions caused by the change of camera optical axis orientation from a frontal position. And SURF algorithm is modified to speed up the process of feature descriptor extraction and further to meet the real-time requirements. We call the proposed algorithm Fully Affine InvaRiant SURF (FAIR-SURF).

Specifically, the contributions of this paper are as follows. (1) The proposed method makes full use of the affine invariant advantage of ASIFT and the efficient merit of SURF; (2) Unlike ASIFT, we carefully select the number of latitudes and longitudes not only to ensure that the simulated images keep close to any other possible view, but also speed up the feature extraction.

The remainder of this paper is organized as follows. Section 2 reviews the SURF algorithm. In Section 3 we briefly describe ASIFT and then describe our fully affine SURF in detail. Section 4 presents our experimental results comparing the new method to SURF and ASIFT on image-matching experiments. This paper is concluded in Section 5.

* Corresponding author.
E-mail address: yuany@opt.ac.cn (Y. Yuan).

0925-2312/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2011.12.006

2. Review of the SURF algorithm

SURF, as described in [11], is much faster, and more robust as opposed to other strategies. A basic second-order Hessian matrix approximation is used for feature point detection. The approximation with box filters is pushed to take place of second-order Gaussian filter. And a very low computational cost is obtained by using integral images.

In Fig. 1, the 9×9 box filters are approximation of a Gaussian with scale $\sigma = 1.2$ and represent the lowest scale for computing the blob response maps. The Hessian-matrix approximation lends itself to the use of integral images, which is a very useful technique [12,13,15]. Hence, computation time is reduced drastically.

The construction of scale image pyramid in SURF algorithm is similar to SIFT. As shown in Fig. 2, the scale space is divided into octaves, and there are 4 scale levels in each octave. Each octave represents a series of filter response maps obtained by convolving the same input image with a filter of increasing size. And the minimum scale difference between two subsequent scales depends on the length of the positive or negative lobes of the partial second order derivative in the direction of derivation. Do non-maximum suppression in a $3 \times 3 \times 3$ neighborhood to get the steady feature points and the scale of values.

In order to be invariant to image rotation, the Haar wavelet responses are calculated in x and y direction within a circular neighborhood of radius $6s$ around the feature point, s is the scale at which the feature point was detected. The Haar wavelet responses are represented as vectors. Then sum all the vector of x and y direction of the Haar wavelet responses within a sliding orientation window covering an angle of size $\pi/3$ around the feature point. The two summed response yield a new vector. And the longest vector is the dominant orientation of the feature point.

For extraction of the descriptor, construct a square region with a size of $20s$ and split the interest region up into 4×4 square sub-regions with 5×5 regularly spaced sample points inside. As shown in Fig. 2, compute the Haar wavelet response x-direction d_x and the Haar wavelet response y-direction d_y. Weight the response with a Gaussian kernel centered at the interest point. Sum the response over each sub-region for d_x and d_y separately. In order to bring in information about the polarity of the intensity changes, extract the sum of absolute value of the responses. Therefore, each sub-region is formed a 4-dimensional vector,

$$Vec = \left(\sum d_x,\ \sum d_y, \sum |d_x|,\ \sum |d_y|\right) \tag{1}$$

Finally, normalize the vector into unit length for invariance to contrast.

Concatenate the descriptor vector for all 4×4 square sub-regions, the length of a feature vector is 64. And compute the sum d_x and $|d_x|$ separately for $d_y < 0$ and $|d_y| \geq 0$, similarly for the sum of d_y and $|d_y|$, the length is doubled.

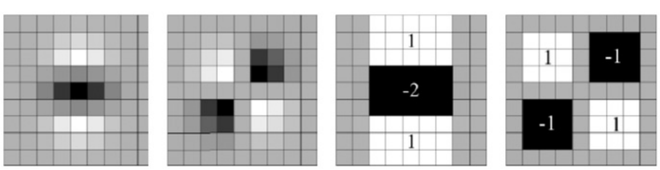

Fig. 1. Left to right: the Gaussian second order partial derivative in y and xy-direction, respectively; the approximation for the second order Gaussian partial derivative in y and xy-direction. The intensity of the gray regions is zero.[11]

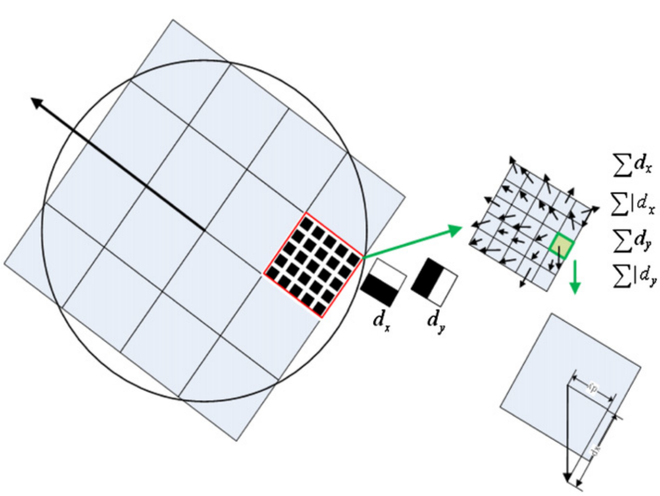

Fig. 2. Feature Descriptor of SURF-64.

3. Fully affine invariant SURF

As the comparison in [14] shows, SURF is invariant to image scaling, blur, and illumination, and partially invariant to rotation and viewpoint changes. Because 'Fast-Hessian' detector that used in SURF is three times faster than that DOG used in SIFT, SURF obtains its fast speed in the experiments. However, SURF is not fully affine invariant. For example, in Fig. 3 image matching using SURF is performed when the angle differences are 60° angle (left) and 70° angle (middle), which shows partially invariant to viewpoint changes. However, it is seen from the rightmost of Fig. 3 that the SURF based image matching fails when view variation is as large as 80°.

In this paper, we propose an effective method to overcome the above drawback of SUFR. Unlike MSER, Harris-affine, and Hessian-affine, which normalize all six affine parameters, FAIR-SURF simulates the two camera axis parameters, which is similar to ASIFT, and then applies a fast version of SURF to extract features from the generated images. In order to further speed up feature extraction, the number of latitudes and longitudes are carefully selected. As shown in Fig. 5 FAIR-SURF proceeds by the following steps.

First, transform each image by simulating all possible affine distortions caused by the change of camera optical axis orientation from a frontal position. As shown in Fig. 4, the gray image area is a flat physical object. The top right small parallelogram represents a camera. The Φ and θ are the longitude and latitude angles of the camera optical axis, respectively. The ψ angle is the camera spin, and λ represents the zoom parameter. In ASIFT [8], the longitude angle Φ is changed with the latitude angle θ, with step $\Delta\Phi = 72°/t$, $t = 1/\cos\theta$, $\Phi \in [0,\pi)$, and $\theta \in [-\pi/2, \pi/2]$. So the

Fig. 3. Image matching using SURF. From middle distance (zoom $\times 4$), the 1st magazines of all the matching results are taken at frontal view and the 2nd are taken at 60° angle, at 70° angle, at 80° angle.