LETTER Improving Hessian Matrix Detector for SURF

Yitao CHI^{†a)}, Zhang XIONG[†], Nonmembers, Qing CHANG^{††}, Member, Chao LI[†], and Hao SHENG[†], Nonmembers

SUMMARY An advanced interest point detector is proposed to improve the Hessian-Matrix based detector of the SURF algorithm. Round-like shapes are utilized as the filter shape to calculate of the Hessian determinant. D_{xy} can be acquired from approximate round areas, while the regions for computing D_{yy} or D_{xx} are designed with the consideration to symmetry and a balance of pixel number. Experimental results indicate that the proposed method has higher repeatability than the one used in SURF, especially in the aspects of rotation and viewpoint, due to the centrosymmetry of the proposed filter shapes. The results of image matching also show that more precision can be gained with the application of proposed detector.

key words: Hessian matrix, interest point detector, image matching

1. Introduction

Local features have been widely applied in fields of computer vision, such as image registration [1], object recognition [8], 3D object retrieval based content [9], object tracking [10] and so on. They are invariant to geometric transformation, robust to occlusion and clutter and distinctive from each other and have become increasingly popular in recent years.

There are varieties of approaches for feature detection and the features extracted by these algorithms differ in localization, scale and structure (corners, blobs, multi-junctions). Among them, the second moment matrix is an efficient way to determine an affine-invariant region. It describes the firstorder signal changes in a region surrounding a point. Distinctive features are localized on strong signal changes.

Lindeberg and Garding [4] developed a method for finding blob-like affine features. They first extract maxima of the normalized Laplacian in scale-space and then iteratively modify the scale and shape of the regions based on the second moment matrix. Baumberg [1] applies affine shape estimation for matching and recognition. He extracts Harris interest points at several scales and then adapts the shape of the point neighborhood to the local image structure using the iterative procedure proposed by Lindeberg [4]. The affine shape is estimated for a fixed scale and fixed location. Note

DOI: 10.1587/transinf.E94.D.921

that there are many points repeated at neighboring scale levels, which increases the probability of false matches as well as the complexity. Mikolajczyk and Schmid [6] proposed affine invariant Harris-Affine and Hessian-Affine detectors. They extended the scale invariant detectors [5] by the affine normalization. The location and scale of points are given by the scale-invariant Harris and Hessian detector. Those two detectors were used to search for the maximum response of Laplacian over scales to estimate the characteristic scale. Later they generalized these algorithms to the affine invariance by an iterative method [7].

SURF (Speeded-Up Robust Features) [3] employs Hessian Matrix to detect the interest points. This method is selected according to its accuracy. More precisely, bloblike structures are detected at locations where the determinants are maximum, and the determinant of the Hessian is used for the scale selection as done by Lindeberg [4]. The method utilizing approximate and simplified shapes to compute Gaussian second-order partial derivatives, gains apparent enhancement in time efficiency. However, we found that in practice the box shape of the filter is crude and reduces the repeatability, especially under conditions of rotation and viewpoint changes. Thus our proposed method is to improve the filter with more symmetric shapes in order to enhance invariance. The experimental results confirmed that the improved detector has higher repeatability and gains more accuracy on image matching than original Hessian-Matrix detector.

2. Original Hessian Matrix Detector for SURF

2.1 Algorithm Description

The SURF detector utilizes the Hessian Matrix for its good performance in accuracy to find out the locations where determinant are maximum as interest point positions.

First an integral image is calculated by convolution. Given a point \mathbf{x} in an input image, the Hessian matrix at scale σ is defined as

$$\mathbf{H} = \left(\begin{array}{cc} L_{xx}(\mathbf{x},\sigma) & L_{xy}(\mathbf{x},\sigma) \\ L_{xy}(\mathbf{x},\sigma) & L_{yy}(\mathbf{x},\sigma) \end{array}\right)$$

Where $L_{xx}(\mathbf{x}, \sigma)$ represents the convolution of Gaussian second-order derivative $\frac{\partial^2}{\partial x^2}g(\mathbf{x}, \sigma)$ with the input image at point \mathbf{x} , and similarly for $L_{xy}(\mathbf{x}, \sigma)$ and $L_{yy}(\mathbf{x}, \sigma)$.

Manuscript received August 17, 2010.

Manuscript revised November 22, 2010.

[†]The author are with the Faculty of State Key Laboratory of Software Development Environment in School of Computer Science and Engineering of Beihang University, Beijing, CHina.

^{††}The author is with the Faculty 202 in School of Electronic and Information Engineering of Beihang University, Beijing, China.

a) E-mail: yitaochi@gmail.com



Fig. 1 Left to right: (a) the (discretised and cropped) Gaussian filters, second-order partial derivatives; (b) the approximation for the second-order Gaussian partial derivatives in SURF; (c) our proposed filters for Gaussian partial derivative. The top lines show the derivatives for in y-direction (L_{yy}) . The bottom lines show the derivatives in xy-direction (L_{xy}) . The grey regions are equal to zero.

In SURF, box filters approximate second-order Gaussian derivatives and can be at a very low computational cost with integral image. The box filters in Fig. 1 (b) are approximation of second-order Gaussian derivative with $\sigma = 1.2$ and represent the lowest scale. They are denoted by D_{xx} , D_{xy} , D_{yy} . Hence determinant of approximate Hessian matrix is expressed as

 $\det(\mathbf{H}_{approx}) = D_{xx}D_{yy} - (wD_{xy})^2$

The weight w is used to balance the expression of the approximate Hessian determinant and w = 0.9 usually.

The scale space is divided into octaves. Every octave is subdivided into a constant number of scale levels. And box filters with different side length are applied to corresponding level to compute the Hessian determinant.

In order to detect interest point in the scale space, a non-maximum suppression in a $3 \times 3 \times 3$ neighborhood is applied. A point at a certain scale is determined as an interest point if the corresponding Hessian determinant value is the maxima in the surrounding $3 \times 3 \times 3$ space.

2.2 Problem with Original Hessian-Based Detector

The box filter is employed for Hessian matrix, which has gone further from the Lowe's LoG approximation [3]. It decreases the cost of computation for the blob values corresponding to every position at different scale. It is because that the sum of gray values of a rectangle region is independant to the size and only needs 3 addition operations, but LoG approximation convolutes the discretised and cropped second-order derivatives with the corresponding region, which means a lot multiplying and adding operations. Hence the box filters enhance the time efficiency greatly.

However, the box filters cannot ensure the repeatability of the interest points in practice. As shown in [3], repeatability curve display peaks on the rotation angle of $\pi/2$ but has lower positions especially at uneven multiples of $\pi/4$. It is mainly because of the rectangle shape of the filters. And it is apparent that when a square image rotates for an angle of uneven multiples of $\pi/4$, there is the greatest difference



Fig. 2 A quarter part of proposed filter to calculate D_{xy} .

of the regions covered by the box filters between the original and rotated one. Hence difference of the regions mainly affects invariance of the Hessian determinant greatly.

3. Improved Interest Point Detector for SURF

We improve the box filters by reselecting pixels in the regions. The basic principle in the proposed method is that if the shapes utilized are more symmetric to every orientation and more approximate to the round shape, it will be more invariant under conditions such as rotation and viewpoint changes. Because round shape can keep main characteristics of the content within its region such as color histogram, texture, etc.

3.1 Improvement for D_{xy}

Figure 2 shows a quarter part of a square inscribed by a circle, which denotes the left-top part in our proposed filter. A series of squares touching the arc *BMD* at points of M, T_1 , T_2 , T_3 , ..., U_1 , U_2 , U_3 , ... are subtracted to approach the fan area.

Given the radius value of r and the origin C which is also the center of the square, we can acquire the coordinates of $H_i(x_i, r)(x_i < 0)$ and sides of all squares s_i in the image by analytic geometry. They can be calculated out according to the geometric relationship as shown:

$$s_1 = \overline{AH_1} = \frac{\sqrt{2}}{2}\overline{AM} = \left(1 - \frac{\sqrt{2}}{2}\right)r\tag{1}$$

$$s_i = \overline{H_i H_{i-1}} = x_i - x_{i-1}, i = 2, 3, 4, \dots$$
 (2)

$$x_1 = -\frac{\sqrt{2}}{2}r\tag{3}$$

$$x_{i} = -\frac{r + x_{i-1} - \sqrt{r^{2} - 2rx_{i-1} - x_{i-1}^{2}}}{2}, i = 2, 3, 4, \dots$$
(4)

The results of s_i are $s_1 = 0.293r$, $s_2 = 0.163r$, $s_3 = 0.103r$, and so on. The sum of the fan region *MBCD* can then be approximately acquired by subtracting away the outside touching squares.

Besides, there are gray squares in Fig. 2 which are neglected in our method. These gray squares will gain actual side values only when the detection of interest point is performed at a high scale. We have found these gray squares take little effect by practice and omit them.

As described in Sect. 1, the filters have a series of side length for each level. The side lengths of filters are multiples of 3, which can be expressed as 3(2k-1) and equal to ones of the original detectors in SURF. *k* is an integer not less than 2. The radius of the fan r = 2k - 1 in Fig. 2. Figure 3 shows the appearance of a series of proposed shapes. We can use linear interpolation to better the filter, but it will cost more operation and the effect of interpolation is not so obvious.

3.2 Improvement for D_{yy}

We reselect the pixels for a better distribution and a closer shape to round. And the same area of the summed region is also take into consideration for the balance of pixels number. Figure 4 (a) shows the calculation for D_{yy} when k is an odd number. Similar as the computation in SURF detector, the top and bottom regions are multiplied with +1 and the middle one is multiplied by -2 and the results of the multiplications are summed up to get D_{yy} . When k is an odd, Every region consists of three close squares. The big one has a side value of 2k - 1 and the small two have a side value of k. k has the same value as in Sect. 3.1. And the small square regions belonging to the black region have vertically middle position. When k is an even number, there exists a little difference, as shown in Fig. 4 (b). The small squares of the black region cannot locate at the precisely middle posi-



Fig.3 Several improved filter shapes for calculation of D_{xy} .



Fig.4 Improved filter shapes for calculation of $D_{yy}(D_{xx})$.

tion. So two squares around the middle position are utilized and they have locations of a pixel difference. According to the calculation method of D_{yy} , the sums of regions that are not overlapping are only added once, which are shown in Fig. 4 (b) and Fig. 4 (d) as dark gray rectangle with height of 1 pixel. k is an even number only in the first octave in SURF. Figure 4 (c) and Fig. 4 (d) show two improved filters with k = 3 and k = 4. The lengths of their sides are 15 and 27 respectively. And the improvement for D_{xx} is similar but in the horizontal direction.

4. Experiments

4.1 Performance Evaluation of the Improved Algorithm

We validated the performance of the proposed algorithm by several transforms, such as rotation, scale, blur and viewpoint.

Figure 5 shows some results obtained by the improved algorithm. The proposed method has high repeatability of



(a) Viewpoint



(b) Rotation





(c) Scale



(d) Blur

Fig. 5 The interest points detection results of the improved method.



Fig. 6 Comparison of repeatability for changes of viewpoint, rotation, scale and blur.

the interest points and is robust to viewpoint changes, rotation, scale, and blur. A series of test images of above changes were applied to evaluate the performance of the proposed algorithm respectively.

The measure of *repeatability* score is employed to evaluate the original detection method in SURF and the proposed one. We take into account only the points located in the part of the scene present in both images. Given the ground truth transformation we can transfer every point into the second image according to the scene geometry to find the corresponding points. Two points correspond if the error in point location is less than 1.5 pixels.

In Fig. 6, the comparison curves of repeatability are shown. It can be seen from Fig. 6 (a) that when the view-point change is smaller than 25, the improved detector has a higher repeatability than original one. In Fig. 6 (b), Under rotation change, the proposed one is better at any angle. High repeatability of viewpoint changes and rotation is due to the selected centrommetric shapes. On the contrary, the repeatability of the improved one is lower than the original one especially when the scale changes are not even for example, 1.5, 3, etc. The repeatability of the improved detector is higher than original one under the blur transformation.

On aspect of detection efficiency, the proposed algorithm needs more than twice the time of the original algorithm. We utilize an Intel(R) Core(TM)2 Duo CPU of a 2.93 GHz frequency and 2 GB RAM for experiments. With same input images, original algorithm consumes about 200 ms to detect 100 feature points, while the proposed detector requires about 450 ms. For the original one, filters with different sizes need same number of operations to compute Hessian matrix. For the proposed one, calculation for D_{xy} needs more operations as *k* increases because more squares should be substracted. The time cost for calculating D_{yy} or D_{xx} is more than twice of the original one, because every non-zero part of filters consists of 2 or 3 rectangles. Although the proposed algorithm costs more time, it is still much faster than the Gaussian filters shown in Fig. 1 (a) be-



(a) left: SURF: 14/72 error; right, proposed method: 7/62 error



(b) top: SURF: 3/68 error; bottom, proposed method: 1/62 error Fig.7 Image matching results.

cause the latter requires convolution with floats.

4.2 Images Matching Results

We carried out the image matching with set of images by the Hessian detector of SURF and the proposed method, using the SURF descriptor. After the interest point detection, the matching pairs between two images are determined by the similarity of their descriptors. An effective measure is obtained by comparing the distance of the closest neighbor of the descriptor and the second-closest one of the two interest points sets, which was put forward by Lowe [2].

We have found that with the same descriptor algorithm, the error matching rate of proposed one is lower. Figure 7 shows the matching results of two pairs of images. Generation of error matches of our method is from the similar information of the regions around the interest points that may generate similar descriptors at different scales. Under SURF, some locations surrounded by very different appearance information are matched. Thus the interest points detected by the proposed method are more distinctive because region informations around these points have higher uniqueness and are more robust to viewpoint changes and rotation.

5. Conclusion

We proposed an improved Hessian-matrix based detector. Round-like shapes are employed to calculate the elements of Hessian determinant, D_{xy} and $D_{yy}(D_{xx})$. The actual size of the proposed filters is determined based on the geometric relationship. Experimental results show that the proposed method is better than the original one especially in the aspects of viewpoint changes and rotation. And the image matching results also indicate that the proposed method can obtain higher precision than the original one.

Acknowledgements

This work is granted by Natural Science Foundation of China (60872062) and Independent Exploration Project (No. SKLSDE-2010ZX-04) of the State Key Laboratory of Software Development Environment of China.

References

- A. Baumberg, "Reliable feature matching across widely separated views," Proc. Conference on Computer Vision and Pattern Recognition, pp.774–781, Hilton Head Island, South Carolina, USA, 2000.
- [2] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," Int. J. Comput. Vis., vol.2, pp.91–110, 2004.
- [3] H. Bay, T. Tuvellars, and L.V. Gool, "SURF: Speeded up robust

features," Proc. European Conference on Computer Vision, pp.404–417, 2006.

- [4] T. Lindeberg and J. Garding, "Shape-adapted smoothing in estimation of 3-D shape cues from affine deformations of local 2-D brightness structure," Image Vis. Comput., vol.15, no.6, pp.415– 434, 1997.
- [5] K. Mikolajczyk and C. Schmid, "Indexing based on scale invariant interest points," Proc. 8th International Conference on Computer Vision, pp.525–531, Vancouver, Canada, 2001.
- [6] K. Mikolajczyk and C. Schmid, "An affine invariant interest point detector," Proc. 7th European Conference on Computer Vision, vol.I, pp.128–142, Copenhagen, Denmark, May 2002.
- [7] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," Proc. Conference on Computer Vision and Pattern Recognition, Madison, Wisconsin, USA, June 2003.
- [8] A. Qamra and E.Y. Chang, "Scalable landmark recognition using EXTENT," Multimed. Tools Appl., vol.38, no.2, pp.187–208, 2008.
- [9] J. Tangelder and R.C. Veltkamp, "A survey of content based 3D shape retrieval methods," Multimed. Tools Appl., vol.39, no.3, pp.441–471, 2008.
- [10] H. Zhou, Y. Yuan, and C. Shi, "Object tracking using SIFT features and mean shift," Comput. Vis Image Underst., vol.113, no.3, pp.345–352, 2009.