TELKOMNIKA, Vol. 11, No. 5, May 2013, pp. 2716 ~ 2722
ISSN: 2302-4046

Features Extraction for Object Detection Based on Interest Point

Amin Mohamed Ahsan*, Dzulkifli Bin Mohamad
Faculty of Computing, Universiti Teknologi Malaysia
81310, Skudai, Johor, Malaysia
Telp: +(6)07-553 3333, fax: +(6)07-556 5044/ 557 4908
*Corresponding author, e-mail: am2002as@gmail.com, dzulkifli57@gmail.com

Abstract

In computer vision, object detection is an essential process for further processes such as object tracking, analyzing and so on. In the same context, extraction features play important role to detect the object correctly. In this paper we present a method to extract local features based on interest point which is used to detect key-points within an image, then, compute histogram of gradient (HOG) for the region surround that point. Proposed method used speed-up robust feature (SURF) method as interest point detector and exclude the descriptor. The new descriptor is computed by using HOG method. The proposed method got advantages of both mentioned methods. To evaluate the proposed method, we used well-known dataset which is Caltech101. The initial result is encouraging in spite of using a small data for training.

Keywords: *Object Detection, SURF, HOG, k-NN*

1. Introduction

Nowadays, the applications of object detection and classification became one of the most leading uses in many fields such as, industries, robotics, security, mobile and internet services. In robots the object classification and localization commonly used to recognize a certain object within a scene, moreover, facial recognition play important role in the security issues.

Object detection techniques or methods are essentially for further tasks (i.e classification, categorization, analysis, etc). Yilmaz [1] categorized the object detection methods into four categories, point-based, segmentation-based, background-based and, supervised-based to detect the object.

Mean-shift [2], Graph-cut [3], and Active contour [4] are example of segment-based to detect the object. While, background modeling used to detect the object within a scene are vary; mixture of Gaussian [5], Eigenbackground [6] and Dynamic texture background [7] are the common models based on modeling the background. In the other hand, Support Vector Machine [8], Neural Network [9] and Adaptive boosting [10] used to detect the object as supervised techniques

Point-based detector is used to search for points that demonstrate quick changes in both the horizontal and vertical orientation of their intensity. Such points called the keypoints or interest points. Those points are invariant to changes in transformation and illumination.

Commonly detectors used based on the interest points include: Harris interest point detector [11], Scale Invariant Feature Transform (SIFT) [12] and Speed-Up Robust Features (SURF) [13]. While SIFT and SURF are invariant to illumination, rotation and scale, Harris interest point's detector is not invariant to scale. In the same time, Harris detector is faster than both SIFT and SURF but less accurate.

Bauer et al. [14] performed a comparison study on both SIFT and SURF regarding the invariance against: rotation, scale change, image noise, change in lighting conditions, and change of view point. During all their tests, SIFT performed little bit better than SURF, but it's slower and more complex computationally than SURF.

Received January 20, 2013; Revised March 18, 2013; Accepted March 25, 2013

In spite of, SURF is optimal in term of detecting the interest point and has a reasonable features dimension (descriptor), but, still has some drawbacks in term of rotation transformation and illumination (i.e. shadow).

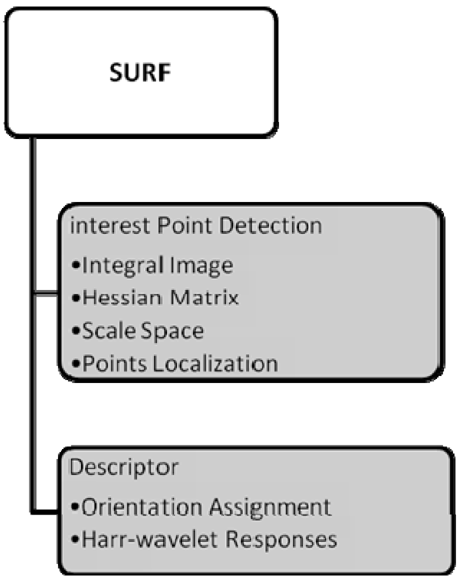
Relative to aforementioned issues, Dalal and Triggs [15] presented a method based on histogram of gradient (HOG) as descriptor or feature extractor as explained in section 2.2. HOG performed well in term of invariance against rotation and illumination, especially the shadow. But, it is not invariant to scale transformation.

Based on what mentioned, we present a method that take the advantage of SURF, just to detect the interest point. Our descriptor used HOG method instead of SURF descriptor. It computes HOG of region about each interest point that are detected by using openSURF from [19].

2. Research Method

2.1. SURF Detector and Descriptor

SURF in [13] involves two steps: first is to detect the interest point, second is to construct the descriptor to detect the key points within an image as shown in Figure 1, there are four steps involved. 1) Calculate the integral of an image. 2) Compute the Hessian matrix. 3) Construct scale space. 4) Localize the interest points. Once the interest points detected, the descriptor can be built in two steps: firstly, orientation assignment, secondly, compute sum of Harr-wavelet responses.



Figuer 1. SURF steps

To increase the performance of SURF an intermediate image representation called “Integral Image” as in [16] is used to speed up the calculation of any rectangle area by (1).

$$I_{\Sigma}(x,y) = \sum_{i=0}^{i\le x} \sum_{j=0}^{j\le y} I(x,y) \tag{1}$$

Where (x,y) is a point in the origin image I, $I_{\Sigma}(X)$ is an integral image at a location $X=(x,y)^T$, which represents the summation of all pixels in image I of a rectangular region formed by the origin and x.

To detect structure of blob-like at locations, Hessian matrix is used because of its good performance [13] as in (2).

$$H(X,\sigma) = \begin{bmatrix} L_{xx}(X,\sigma) & L_{xy}(X,\sigma) \\ L_{xy}(X,\sigma) & L_{yy}(X,\sigma) \end{bmatrix} \quad (2)$$

Where H is a hessian matrix for point X=(x,y) at scale σ in image I, and $L_{xx}(X,\sigma)$ is the convolution of the Gaussian second order derivative in point X, same thing for Lyy and Lxy.

To get an accurate approximation for the Hessian determinant, Bay [13] purposed a formula using the approximated Gaussian as in (3)

$$\det(H_{approx}) = D_{xx}D_{yy} - (0.9D_{xy})^2 \quad (3)$$

Where Dxx, Dyy, and Dxy are the approximations for the second order of Gaussian.

In contrast to Lowe [12], Bay [13] used filter increasing to build the pyramid to represent the scale-space. Instead of build a different scales of the original image, Bay built a different size of filter to apply on the original image as shown in Figure 2. So, SURF is computationally efficient and size invariant.

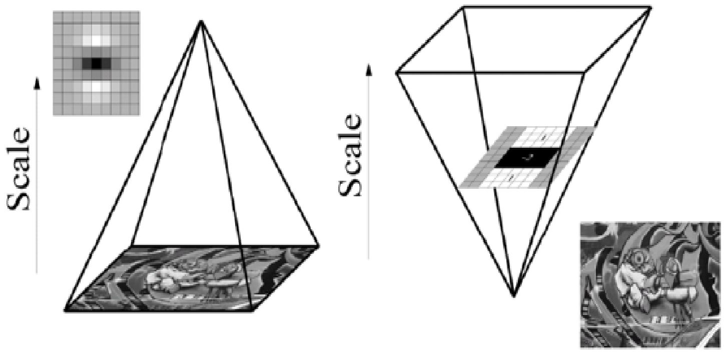


Figure 2. Scale-Space: SIFT (left), SURF (right).

Interest points localized over all scales in 3x3x3 neighborhood by applying the non-maximum suppression as in [17]. To orientation determination, the Harr-wavelet responses in x and y directions are calculated with size 4s (s: scale) and radius 6s of detected points.

To get the dominant orientation, sum of all responses within a sliding orientation window of size π/3 are calculated. The orientation of the interest point is the longest vector over all windows. Finally, the components of the descriptor are calculated by divided each window into 4x4 sub-regions, then apply the Harr-wavelet again on each sub-region to get the final vector as follow:

$$v_{subregion} = \left[\sum dx, \sum dy, \sum |dx|, \sum |dy|\right]$$

Where each sub-region gives four values, which mean 4x4x4=64 values for each interest point.

2.2. Histogram of Gradient

Dalal and Triggs [15] presented a method based on grid of Histogram Orientation Gradient (HOG) as descriptors; those descriptors represent the features set for the object. This method involves five steps as shown in Figure 3.

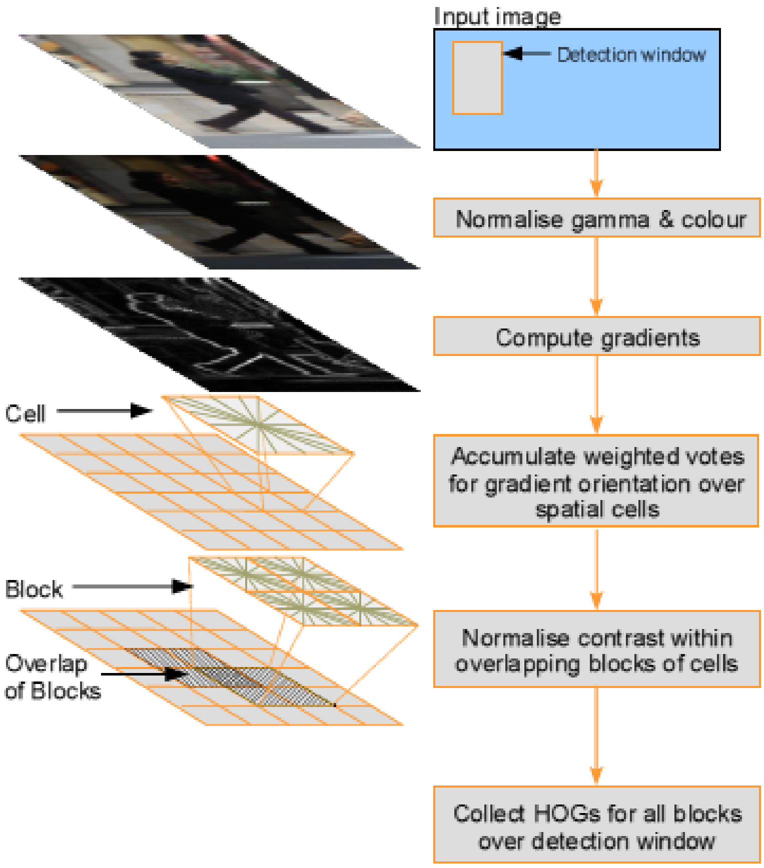


Figure 3. An overview of static HOG feature extraction, Dalal [15].

The first step is applying the normalization equalization on an image in order to reduce the effects of illumination variance and the local shadowing. Next step is to compute the first order gradients for further resistance to illumination variations. Third step involves dividing the image into small sub region called “cell’, the histogram gradient is accumulated for all pixels within each cell. Fourth step is to normalize the cell across large regions which include a group of cells called “block” to get better illumination invariance. That last step is collecting the HOG from all overlapped blocks which are considered as a descriptor.

2.3. Proposed Method

In this paper we present a new method to extract features that used to detect an object, this method based on SURF detector and HOG. The k-nearest neighbor algorithm (k-NN) is used as temporary classifier to examine the features extractor method; Figure 4 shows the purposed method.

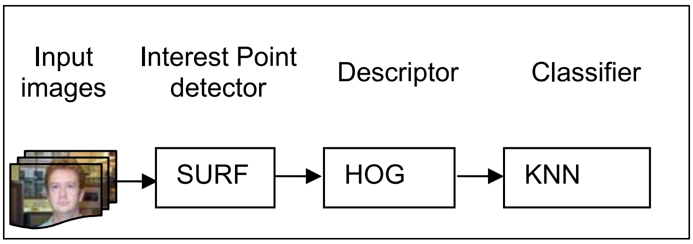


Figure 4. Purposed Method

First, input images were divided into two groups, positive group which represents the object, negative group which represents non-object. Second step is getting the interest point within images for each group, only the points that are corner being taken. It can be done by choosing the interest points that their laplacian value greater than 1 as shown in Figure 5.

Figure 5. Interest Points: edge and corner (left), corner only (right)

The fourth step is to computer the HOG for each interest point, in order to do that we compute the HOG for square area that surround the interest point, where the interest point should be center as shown in Figure 6.

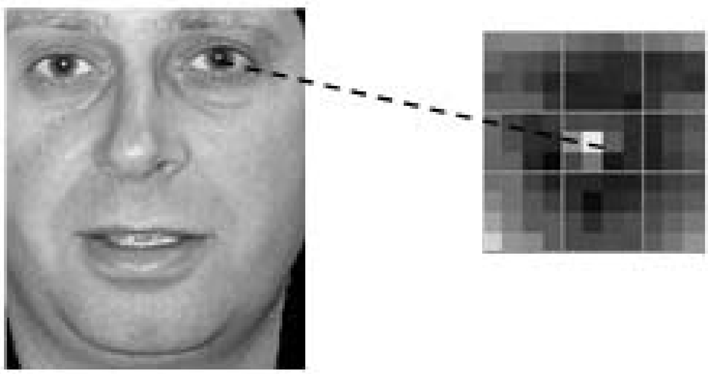


Figure 6. Region about interest point

We do not had to apply all step mentioned in [15], we compute the HOG by overlapped sliding window on the region that has been taken, that yields to get 81 features for each interest point. Before applying k-NN classifier, all positive features labeled as object and get 1 for their group, the negative features take 0 for their group. Then, all features are combined in two matrices, one for features, second for group.

The last step, applying k-NN classifier to examine the features, k-NN classifier does not require so much setting; therefore, we used it as temporary classifier. Figure 7 shows some examples.

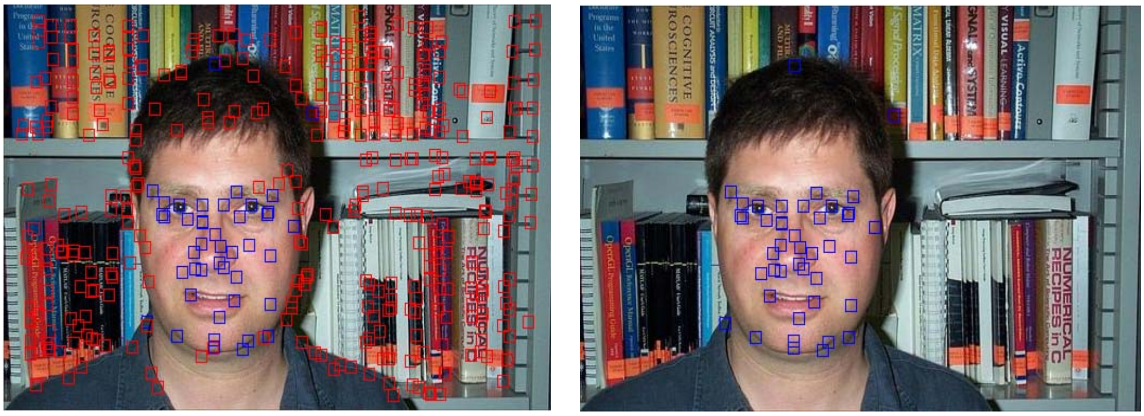


Figure 7. k-NN example result: object and non-object (left), object only (right).

Only three regions have been taken to represent the object (i.e. human face) which are the regions surround eyes, nose, and mouth.

3. Results

To evaluate the performance of the proposed method, we used a number of images from caltech101 dataset as shown in Table 1.

Table 1. Dataset used

Dataset	Phase	images	No. of images	Total
Caltech101	Train	Positive	28	52
		Negative	24	
	Test	Both	10	10

Measurements used to evaluate the performance of proposed method are: detection rate sometimes called sensitivity, specificity, and precision which are mentioned in [18].

Three measures are computed by (4), (5), and (6) respectively.

$$Detection\ Rate = \frac{TP}{TP + FN} \quad (4)$$

$$Specificity = \frac{TN}{TN + FP} \quad (5)$$

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

Where TP, TN, FP, and FN are true positive, true negative, false positive, and false negative respectively.

Table 2 shows the result of TP, TN, FP, FN, all positives, and all negatives obtained by the purposed method for each test image used in test phase. Summation also computed to be used to compute the three measures mentioned above.

Detection rate, specificity, and precision which are obtained using our method are, 0.85%, 97.8%, and 90.5% respectively.

Initial result shows a good performance although a few numbers of images have been used for training. Moreover, k-NN classifier used in our work is only to examine the proposed extractor, so, it is features extraction issue rather than classification.

Table 2. Result

Image No.	TP	TN	FP	FN	All_P[a]	All_N[b]
1	41	357	2	7	43	364
2	41	228	7	8	48	236
3	34	117	6	4	40	121
4	27	131	7	2	34	133
5	34	72	1	11	35	83
6	41	161	1	6	42	167
7	39	194	5	7	44	201
8	35	133	4	4	39	137
9	37	201	3	6	40	207
10	36	167	2	8	38	175
Sum	365	1761	38	63	403	1824

a. all positive points, b. all negative points

3. Conclusion

In this paper, we have introduced a method to extract new features for object detection based on SURF detector and HOG descriptor with some modification in aforementioned methods. Initial results are encouraging.

Currently, we are working on others classifiers such as SVM and ANN in line with continued enhancement of features extractor. Another interest point detector such as SIFT will be taken into our consideration.

References

[1] Yilmaz A, O Javed, and M Shah. *Object tracking: A survey. Acm Computing Surveys*. (CSUR). 2006; 38(4): 13.
[2] Comaniciu D and Meer P. *Mean shift analysis and applications*. In IEEE International Conference on Computer Vision (ICCV).1999; 2: 1197–1203.
[3] Shi J and Malik J. *Normalized cuts and image segmentation*. IEEE Trans. Patt. Analy. Mach. Intell. 2000; 22(8): 888–905.
[4] Caselles V, Kimmel R, and Sapiro G. *Geodesic active contours*.1995. In IEEE International Conference on Computer Vision (ICCV). 694–699.
[5] Stauffer C and Grimson W. *Learning patterns of activity using real time tracking*. IEEE Trans. Patt. Analy. Mach. Intell. 2000; 22(8): 747–767.
[6] Oliver NM, B Rosario and AP Pentland. *A Bayesian computer vision system for modeling human interactions*. Pattern Analysis and Machine Intelligence. IEEE Transactions on, 2000; 22(8): 831-843.
[7] Monnet A, Mittal A, Paragios N and Ramesh V. *Background modeling and subtraction of dynamic scenes*. In IEEE International Conference on Computer Vision (ICCV). 2003: 1305–1312.
[8] Papageorgiou C, Oren M and Poggio T. *A general framework for object detection*. In IEEE International Conference on Computer Vision (ICCV). 1998: 555–562.
[9] Rowley H, Baluja S, and Kanade T. *Neural network-based face detection*. IEEE Trans. Patt. Analy. Mach. Intell. 1998; 20(1): 23–38.
[10] Viola P, Jones M and Snow D. *Detecting pedestrians using patterns of motion and appearance*. In IEEE International Conference on Computer Vision (ICCV). 2003: 734–741.
[11] Harris C and M Stephens. *A combined corner and edge detector*. Manchester, UK. 1988.
[12] Lowe DG. *Distinctive image features from scale-invariant keypoints*. International journal of computer vision, 2004; 60(2): 91-110.
[13] Bay H, et al. *Speeded-up robust features (SURF)*. Computer Vision and Image Understanding. 2008; 110(3): p. 346-359.
[14] Bauer J, N Sunderhauf, et al. *Comparing several implementations of two recently published feature detectors*. 2007, unpublished.
[15] Dalal N and B Triggs. *Histograms of oriented gradients for human detection*. in CVPR. 2005. San Diego, CA, USA: Ieee.
[16] Paul Viola and Michael Jones. *Rapid object detection using a boosted cascade of simple features*. Cvpr. 2001; 1: 511.
[17] A Neubeck, L Van Gool. *Effecient non-maximum suppression*. in: ICPR. 2006.
[18] Fawcett T. *An introduction to ROC analysis*. 2006, Pattern Recognition Letters. 27: 861–874.
[19] www.chrisevansdev/opensurf.