Object Recognition Based on Descriptor of Improved Histograms of Second-Order Gradients

Cuncun Wei
Faculty of Electronic and Information Engineering, Zhejiang Business Technology Institute, Ningbo, 315012, Zhejiang, China

Abstract
Object recognition is a very active research direction in the computer vision area, which has been widely used in the image field. The object recognition, in a sense, can be understood as the image recognition, namely, extract the object interested from a certain image, which takes the object feature as the emphasis and difficulty. In the actual imaging, the object to be recognized may be occluded by other objects due to the imaging angle or partly blurred by the weather, lighting and camera exposure, thus bringing difficulties to further recognition. Under these circumstances, it has gradually been a research hotspot how to utilize the limited local objects to extract the features in order to recognize the object (Hu and Peng, 2014).

In terms of image, the object recognition study starts from 1960s, and its purpose is to use the theory and method in the field of image processing and pattern recognition to determine whether there exists object of interest to determine, and also extract the useful information, with the purpose to determine the object location and realize the image description, analysis, judgment and recognition. Image object recognition algorithm includes the area method and edge method etc. Area method is less sensitive to noise, largely correct goal (object) thoroughly. The experiment result demonstrates that HSOG has better recognition ability compared with the descriptors of first-order gradient and the second-order gradient histogram, in order to improve recognition accuracy and noise robustness, an object recognition method based on an improved second-order gradient histogram descriptor is proposed in this paper. Using the absolute value to retain the first-order gradient of all directions, eliminating the defect of lost some direction when the gradient direction is lost odd, Gaussian function weighted is used to calculate second order gradient histograms, which considered the influence of the neighboring pixels to the center pixel and improved anti-noise performance. Experimental results show that the method in this paper improves the recognition accuracy and noise robustness.

Key words: Second Order Gradient, Histogram, Gaussian Weighted, Object Recognition

1. INTRODUCTION
Object recognition is a very active research direction in the computer vision area, which has been widely used in the image field. The object recognition, in a sense, can be understood as the image recognition, namely, extract the object interested from a certain image, which takes the object feature as the emphasis and difficulty. In the actual imaging, the object to be recognized may be occluded by other objects due to the imaging angle or partly blurred by the weather, lighting and camera exposure, thus bringing difficulties to further recognition. Under these circumstances, it has gradually been a research hotspot how to utilize the limited local objects to extract the features in order to recognize the object (Hu and Peng, 2014).

In terms of image, the object recognition study starts from 1960s, and its purpose is to use the theory and method in the field of image processing and pattern recognition to determine whether there exists object of interest to determine, and also extract the useful information, with the purpose to determine the object location and realize the image description, analysis, judgment and recognition. Image object recognition algorithm includes the area method and edge method etc. Area method is less sensitive to noise, largely correct goal description can also be obtained even in the strong shadow, and however, it is difficult to get accurate object location information. While, the algorithm based on the edge stresses on the gentle change of the object internal characteristics, and there are characteristic mutations between objects and non-objects. Local feature extraction has become a hot research topic, not only because it solves the problems brought in the actual imaging, but also because the local features simulate the human visual identity system and lay a foundation for the artificial intelligence and automatic object recognition. The local object features include point feature (Lai and Fang, 2002; Lowe, 2004), contour feature and region feature, etc. Although numerous local feature descriptors have been proposed in recent years, their performance varies greatly in different applications and there is no universal description algorithm. Mikolajczyk K et al have made comparative analysis to the methods, including Scale Invariant Feature Transform (SIFT), Harris-Affine Feature, Hessian-Affine Maximally Stable Extremal Regions (MSER), but each of them has their own strengths and weaknesses in the image gray-scale variation, scale variation and viewpoint changes (Mikolajczyk and Tuytelaars et al., 2005). The analysis in Literatures and has shown that SIFT descriptor has the optimal scale invariance while MROGH descriptor is the optimum in brightness transform, however, these descriptors only use first-order gradients and they can’t describe the geometric properties of the object comprehensively (Belongie and Malik et al., 2002). Dalal et al. propose the histogram (HOG) feature in the gradient direction, and such feature can characterize the appearance and shape of the image’s local object without the accurate edge location information, and make HOG descriptor equipped with the illumination invariance by adopting the overlapped local contrast normalization technique. Di and others have come up with a local feature descriptor based on histograms of second-order gradients (HSOG), which can describe the geometric features, i.e. grayscale saltation, ridges, edges, peak and valley which are relevant to the curvature of the object thoroughly. The experiment result demonstrates that HSOG has better recognition ability compared with the descriptors of first-order gradients (Di and Chao, 2014).
This paper obtains an improved descriptor of the second-order gradients on the basis of HSOG. It uses the absolute value to preserve the calculation result of the first-order gradients in all directions and eliminates the defects with the gradient direction to be the odd number and it calculates the histogram of second-order gradients through Gaussian function weighting, considers the impact of the neighbor pixels in the central pixel and improves the anti-noise performance of the algorithm. This paper applies this method in the object recognition and compares the recognition accuracy and the noise robustness of the method in this paper and the classical histogram method of second-order gradients. The experiment result proves that the method of this paper enhances the recognition accuracy and has better noise robustness.

2. IMPROVED DESCRIPTOR OF HISTOGRAMS OF SECOND-ORDER GRADIENTS

The classical local feature descriptor such as SIFT and MSER only uses the information of the first-order gradients. First-order gradients are sensitive to such geometric features like gray-scale slope, length and area, but non-sensitive to the related features to curvature. The second-order gradients reflect the relevant geometric features to the curvature and to use second-order gradients can extract the object features in a thorough manner. HSOG extracts the object features with second-order gradients and its building process is comprised of three steps: firstly, calculate and normalize the first-order gradients of the local feature regions. Secondly, calculate the second-order gradients on the basis of the first-order gradients (Krystian and Cordelia, 2005). Thirdly, obtain the histogram of second-order gradients with spatial pooling. The specific process is indicated as Fig. 1.

![Building process of HSOG](image)

**Figure 1. Building process of HSOG**

For the given image \( I(x, y) \), calculate the gradient image of the object image in the given direction and obtain a series of first-order gradient images of different directions, as indicated in Formula (1).

\[
G_d = \text{abs} \left( \frac{\partial I}{\partial d} \right) \quad d = 1, 2, \ldots, N
\]  

In this formula, \( d \) stands for different directions and \( N \) refers to the number of directions. In this paper, \( N = 8 \) and \( \text{abs}(\cdot) \) means the absolute value. Literature reserves the positive value of the calculation result of the first-order gradients and makes the negative value 0. Different numbers of directions doesn’t equal to the difference of 180 degrees in these two directions. For example, when \( N = 9 \), the negative value of the calculation result of the first-order gradients can’t be reflected by the gradients different in 180 degrees and the negative value reflects the gray-scale variation in this direction so as to reflect the local features (Bin and Fuchao et al, 2011). Therefore, this paper takes absolute value of the calculation result of the first-order gradients and retains the calculation results of the gradients in all directions.

In the imaging process, tiny translation is inevitable in the image. The coordinate of the image centroid is required to be calculated in order to eliminate the impact of translation. Move the image center to the centroid. Human visual cells have excellent robustness to this tiny translation. In order to simulate this human visual mechanism, no gradient saltation will be caused in the small changes in the neighborhood through convolution...
between the result of the first-order gradients and Gaussian function \( G_r \). The standard deviation of the Gaussian function is in direct proportion to the neighborhood radius \( R \), as indicated by the formula below.

\[
\rho^{(R)}_d = G_r \ast G_d
\]  

Obtain the first-order gradients of all directions. Form a one-dimensional vector with the first-order gradients in all directors of Point \((x, y)\) in the feature region and normalize them to the range of \([0,1]\); therefore, the feature region can be shown by first-order gradients, as indicated by the following formula and \( \tilde{\rho}^{(R)} \) is the normalized first-order gradients.

\[
\tilde{\rho}^{(R)}(x, y) = \{ \rho^{(R)}_{x}(x, y), \rho^{(R)}_{y}(x, y), \ldots, \rho^{(R)}_{n}(x, y) \}
\]  

As the lighting change in the external environment is usually reflected by adding a constant to all the pixels in the image, to calculate the image gradient in Formula (3) can eliminate the lighting impact, since the contrast change is generally shown by multiplying a constant to all pixels, the normalization operation can remove the influence of the contrast changes and some noise impact can be dispelled after convolution with Gaussian function. Therefore, the calculation result of Formula (1) has lighting and contrast change invariance and certain noise robustness. To obtain the second-order gradients as well as the amplitude and phase from the calculation result of first-order gradients are demonstrated by Formulas (4), (5) and (6).

\[
\begin{align*}
\frac{\partial \rho^{(R)}_{x}(x, y)}{\partial x} &= \rho^{(R)}_{x}(x+1, y) - \rho^{(R)}_{x}(x-1, y) \\
\frac{\partial \rho^{(R)}_{y}(x, y)}{\partial y} &= \rho^{(R)}_{y}(x, y+1) - \rho^{(R)}_{y}(x, y-1)
\end{align*}
\]  

\[
\text{mag}_d(x, y) = \sqrt{\left( \frac{\partial \rho^{(R)}_{x}(x, y)}{\partial x} \right)^2 + \left( \frac{\partial \rho^{(R)}_{y}(x, y)}{\partial y} \right)^2}
\]  

\[
\theta_d(x, y) = \arctan \left( \frac{\frac{\partial \rho^{(R)}_{y}(x, y)}{\partial y}}{\frac{\partial \rho^{(R)}_{x}(x, y)}{\partial x}} \right)
\]  

In order to maintain the same number of directions as the calculation of the first-order gradients, quantize the gradient directions obtained from Formula (6) according to Formula (7). In this formula, \( \text{mod}(\cdot) \) refers to modulus operation and \( \cdot \) is to round up to an integer.

\[
n_d(x, y) = \text{mod} \left( \frac{\theta_d(x, y)}{2\pi / N} + \frac{1}{2}, N \right)
\]  

In the end, count the histograms of second-order gradients in the image sub-regions. Literatures analyze the partitioning strategy and performance of various spatial sub-images and proves that the strategy in Fig.2 has the optimal performance via experiment. Therefore, HSOG also uses this strategy to conduct statistics to the histograms of second-order gradients(Zahn and Roskies et al, 1972; Matas and Chum et al, 2004).

\[\text{Figure 2. Pooling strategy of spatial sub-images}\]

As indicated in Fig.2, the image is partitioned into a series of circles of different radiuses and the circles are put in a series of concentric circular ring according to their radiuses. The radius of the circular ring is
determined the size of the region and the radius of the circle in the annulus is directly proportional to the
distance between the center of the circle to the central pixel. Therefore, for HSOG, there exist the following
parameters: the number of directions is \( N \), the radius of the sub-region is \( R \), the ring number of the concentric
circles is \( C_{\text{ring}} \) and the number of the circles in the ring is \( C_{\text{circle}} \). The standard deviation of the Gaussian
function in Formula (2) is proportional to the neighborhood radius and the value is taken according to Formula
(7). The radius of the circular ring and the position of the circle in the circular ring are shown in Formula (9): \( i \)
means the \( ith \) circular ring and \( j \) is the \( jth \) circle in the circular ring. In this paper, \( N = 8 \), \( R = 15 \), \( C_{\text{ring}} = 3 \)
and \( C_{\text{circle}} = 4 \).

\[
\sigma_d = \frac{R(i+1)}{2C_{\text{ring}}} \quad \text{(8)}
\]

\[
r_j = \frac{R(i+1)}{C_{\text{ring}}} \quad \Theta_j = \frac{2\pi j}{C_{\text{circle}}} \quad \text{(9)}
\]

The total circle in the sub-region is \( T = C_{\text{ring}} \times C_{\text{circle}} + 1 \). For every circle \( \text{CIR}_j \), \( j = 1, 2, \ldots, T \), its
histograms of second-order gradients can be obtained by the following formula.

\[
h_{2}^{(d)}(i) = \sum_{(x,y) \in \text{CIR}_i} f(n_d(x,y) == i) \ast \text{mag}_d(x,y)
\quad \text{(10)}
\]

In this formula, \( i = 0, 1, \ldots, N-1 \), \( d = 1, 2, \ldots, N \), \( j = 1, 2, \ldots, T \) and \( f(\cdot) \) is the condition judgment
function. If the condition in the bracket is tenable, the function value is 1, otherwise, it is 0. It can be seen from
Formula (10) that the method to calculate the histogram of second-order gradients doesn’t take the influence of
the distance between the neighborhood pixel to the central pixel into consideration. The influence the
neighborhood pixel plays on the central pixel is inversely proportional to the distance. This paper performs
weighted summation to the histograms of second-order gradients by using the Gaussian function form indicated
by Formula (11) and the final gradient histogram calculation is shown by Formula (12).

\[
a_{(r_{j})} = e^{-\frac{(r_{j}-r)^2}{2\sigma_{j}^2}}
\quad \text{(11)}
\]

\[
h_{2}^{(d)}(i) = \sum_{(x,y) \in \text{CIR}_i} a_{(r_{j})} \ast f(n_d(x,y) == i) \ast \text{mag}_d(x,y)
\quad \text{(12)}
\]

In this formula, \( r_j \) is the radius of the \( lth \)-layered circular ring, use the same weight \( a_{(r_{j})} \) for all the circles in
the \( lth \) layer and \( \text{CIR}_i \) means that the \( jth \) circle is located in the \( lth \)-layered circular ring. Because the
circle radius is smaller than that of the circular ring, the pixels in the same circle almost have the same impact
on the central pixel; therefore, this paper only considers the influence different circular rings have on the central
pixel of the sub-regions and takes the same weight for the circles in the same circular ring, in this way, it avoids
the influence of the calculation efficiency brought by using different weights to different circles. To count the
pixels of every circle obtains the histograms \( h_{2}^{(d)} \) of second-order gradients in every direction and to connect the
\( h_{2}^{(d)} \) in all directions obtains HSOG descriptor, namely

\[
\text{HSOG} = \left[ \hat{h}_{1}^{(d)}, \hat{h}_{2}^{(d)}, \hat{h}_{3}^{(d)}, \ldots, \hat{h}_{N_{\text{d}}}^{(d)} \right]^{T}
\quad \text{(13)}
\]

In this formula, \( h_{d}^{(a)} = \left[ h_{1}^{(a)}, h_{2}^{(a)} , h_{3}^{(a)}, \ldots, h_{N_{\text{d}}}^{(a)} \right]^{T} \) and \( \hat{h}_{d}^{(a)} \) is the normalized form of \( h_{d}^{(a)} \). Since the
dimension of HSOG is \( (C_{\text{ring}} \times C_{\text{circle}} + 1)N_{\text{d}}^2 \) and it is high. The experiment result shows that after PCA analysis,
the recognition accuracy is the highest when the dimension is 128. Therefore, this paper takes 128-dimensional
HSOG descriptor as the local object feature after conducting PCA analysis to HSOG.

3. EXPERIMENT RESULT AND ANALYSIS

3.1. Sample Building
As for the test image recognition algorithm and machine learning algorithm, the academic world usually
uses the standard database to facilitate the analysis of the efficiency of different algorithms. This paper uses
COIL-100 database to test the object recognition accuracy of the algorithm in this paper and a part of the object
sample images used in the experiment are shown in Fig.3. On the basis of the original image sample, this paper scales the image at the fold of 0.1-1 and obtains the object samples of different proportions. The training and test samples in this paper are divided at the standard of 1:1. In order to test the noise resistance of the method in this paper, Gaussian noises and impulse noises of different degrees have been added to the test sample image: Gaussian noises uses zero mean and the standard deviations are 5 and 10 while the impulse noise densities are 0.01 and 0.02.

![Image samples of experiment](image_url)

**Figure 3.** Image samples of experiment

### 3.2. Object Recognition

This section mainly compares the object recognition accuracy and noise robustness of the classical HSOG and the improved HSOG in this paper. For the sample set constructed in 3.1, the accuracy after taking the mean value of multiple experiments is shown in Table 1.

<table>
<thead>
<tr>
<th>Method</th>
<th>No Noise</th>
<th>Gaussian Noise (5)</th>
<th>Gaussian Noise (10)</th>
<th>Impulse Noise (0.01)</th>
<th>Impulse Noise (0.02)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSOG</td>
<td>89.79%</td>
<td>81.12%</td>
<td>70.25%</td>
<td>82.07%</td>
<td>74.93%</td>
</tr>
<tr>
<td>Method of This Paper</td>
<td>91.12%</td>
<td>87.36%</td>
<td>79.48%</td>
<td>88.49%</td>
<td>80.01%</td>
</tr>
</tbody>
</table>

It can be seen from the statistical result of Table 1 that with no noise, the method of this paper is better than HSOG and when the test samples are affected by different noises, the object recognition accuracy of HSOG fluctuates greatly and the gap between the lowest and the highest accuracy is 19% while the gap of the method in this paper is only 12%, therefore, it has better noise robustness. When calculating the first-order gradients, this paper takes the absolute value and preserves the result of all directions. In the experiment, the number of gradient directions is even, therefore, when there is no noise, the method of this paper is better than HSOG. Besides, this paper introduces Gaussian weighting in calculating histograms of second-order gradients and considers the impact of the surrounding pixels on the central pixel, therefore, it has better noise robustness.

### 4. CONCLUSIONS

The purpose of image recognition is to handle automatically image information with computer, thus replacing people to fulfill the task of image classification and recognition. In order to solve the defects of classical local feature descriptor in representing the related features to curvature as well as the deficiency that the influence the neighborhood pixels play on the central pixel is not taken into consideration when the classical histogram of second-order gradients abandons the negative gradient value in the calculation of the first-order gradients, this paper preserves the gradient calculation results of all directions by taking absolute value, eliminates the flaws of the loss of a certain direction when the number of gradient directions is odd number, conducts weighted calculation on the histogram of second-order gradients with Gaussian function, considers the influence the neighborhood pixels play on the central pixel and improves the anti-noise performance of the algorithm. Through the object classification experiment of the standard image library, the descriptor based on the improved histograms of second-order gradients enhances its recognition accuracy and noise robustness compared with the classical HSOG descriptor.

### Acknowledgements

This work was supported by 2015 Zhejiang research project for public welfare technology application (2015C33079).
REFERENCES


