An Effective Keypoint Selection Algorithm in SIFT

Ding Zuchun

School of Electronic and Communication,
South China University of Technology, Guangzhou, China
zucding@gmail.com

Abstract

Keypoint selection is the important step in object recognition, including general object classification, human tracing and human pose discrimination etc .This paper proposes a more accurate modified key point selection algorithm by modifying SIFT in the stage of extreme point selection. In machine vision or computer vision, including human pose recognition, to select key points, the traditional SIFT completes this according to the extremes derived from LoG (Laplacian of Gaussian) convolution with image, which provides scale invariance features for key points. The extreme points' position is the foundation of feature descriptor for the gradient calculation in the next step. But in the process of images convoluting with the difference of Gaussian function to attain the extreme point, bias is produced because the extreme points' positions aren't accurate. We modify the extreme points' selection to make key points more accurate with less bias to the theoretical points. Simulation with about 3500 images of different resolutions gives the AIPR (adjusted interest point ratio) and illustrates the universality of extreme points' selection and verifies the values of this algorithm.

Keywords: *machine vision; SIFT; key point; Extreme point; object recognition; AIPR*

1. Introduction

In the domain of machine vision or computer vision, one of the important steps is to find the key feature to descript the basic features of the object. As many object recognition algorithm, the efficient way is to employ the feature descriptor. The typical algorithms include SIFT (Scale Invariant Feature Transform) [1], PCA-SIFT (Principal Component Analysis Scale Invariant Feature Transform) [4], HOG (Histogram of Oriented Gradients) [2, 3] and so on. All the methods above are based on the kernel features presented by key points defined by SIFT. PCA-SIFT combines PCA with SIFT to extract the important key points in a lower dimension space. In fact, PCA-SIFT utilizes the dimension reduction ability of PCA, and projects gradient vectors to a compact vectors using eigenspace.

Generally speaking, SURF (Speeded-Up Robust Features) [5] is an improvement of SIFT with the same consideration of calculating the sum of response belonging to the interest point' neighbors, SIFT sums up the gradients and SURF sums up the wavelet responses. To accelerate, integral images are applied in interest point selection. The result is the high speed in real time calculation and more robust to rotation, even not in the same plane.

HOG is utilized to human discrimination which includes C-HOG (Circular HOG) and R-HOG (Rectangular HOG), the difference of which is that bins to collect HOG are not the same. C-HOG has a center bin or several center bins, and the size of the bins having the different distance from the center are different also. However, each block of R-HOG is rectangular and composed of many cells with the same rectangular size. Generally speaking, the common basis of the algorithm above is SIFT features, but in HOG these features are used differently. In this meaning, how to extract the key points which are the foundation of feature descriptors is the most important for human recognition also.

Under the constraint of edge group scale space by Sobel edge detector, [7] improves the SIFT algorithm in speed.

Other papers discuss the key points selection in SIFT and other algorithms [8-12]. We can see the importance of key point selection in object recognition.

Nowadays, SIFT is widely used because its features, *e.g.* object tracking [14] and even iris recognition [13].

In fact, SIFT is the base of many modern key point selection algorithms. The improvement in procedure of the key point selection will thereby influence the performance of the recognition algorithm which is based on SIFT key point extraction.

The performance improvement of SIFT algorithm will speedup the object recognition ability, including general objects, and has the optimistic vista in human recognition which is becoming more and more important nowadays.

This paper will analyze the inaccuracy in SIFT key point calculation and propose a useful algorithm to decrease the influence in key point selection. To attain this aim, we should select more effective extreme points which will become the key points (or the interest points in other words) in image feature selection.

In this paper, Section 2 reviews the important step in SIFT, the step is employed to produce the scale invariant features, Section 3 describes the detailed algorithm of our modified SIFT key point selection, Section 4 gives the experiment results which verify the values of this algorithm. In Section 5, summary is concluded for the algorithm and the future work is introduced.

2. Scale invariance in SIFT key point extraction

For convenience, we review briefly some important step in SIFT descriptor calculation, and deduce the modified extreme points extraction method.

One of the effective methods to recognize object is to find the key points which can present the key features of object. In SIFT algorithm, when two images are convolved with Gaussian function, the subtraction between the low pass filtered images output from Gaussian filters have the effect of Laplacian of Gaussian (LoG)[1], and are normalized by the squares of stand deviation to provide scale invariant features. For convenience, we present the related formulae here.

$$G(x,y,k\sigma)-G(x,y,\sigma)\approx(k-1)\sigma^2\nabla^2 G\ . \tag{1}$$

$$G(x,y,\sigma)=\frac{1}{2\pi\sigma^2}e^{\frac{-(x^2+y^2)}{2\sigma^2}}\ . \tag{2}$$

Eq.1 is the base of SIFT algorithm. The convolution of LoG makes the image scale invariant, and this is an important feature of SIFT. Eq.1 indicates that the difference of Gaussian convolutions with image is used only to approximate the LoG.

This procedure will continue between all the neighbor layers in an octave that is the collection of image patches convoluting with the same image by the Gaussian function with increasing deviation k. The deviation can be presented as following.

$$\sigma_k = k\sigma\ . \tag{3}$$

But there is bias inevitably. The reason to guarantee the scale invariance of interest points is convolution by LoG, but the approximating of Eq.1 is deduced by the procedure which substitutes subtraction for the partial derivative, and they are not the same. In this procedure, errors will occur, how to decrease the error and in the same time speedup the keypoint selection has its meaningful value to update the procedure of SIFT.

3. Modified feature extreme point extraction

According to Cauchy Medium Value Theory, because Gaussian function is continuous and derivative, the difference of G(x, y, $k_1\sigma$) and G(x, y, $k_2\sigma$) can be presented as

$$G(x,y,k_1\sigma)-G(x,y,k_2\sigma)=G(x,y,k'\sigma) \tag{4}$$

Where k' is between k_1 and k_2. We assume $k_1\sigma$ and $k_2\sigma$ are the consecutive derivations of neighbor layers' Gaussian filters in SIFT octave.

We know that

$$\sigma^2\nabla^2 G=\sigma\frac{\partial G}{\partial\sigma} \quad . \tag{5}$$

This is the key reason which renders SIFT interest point scale invariant. Mikolajczyk[6] proved operator in left or right of Eq.5 can produce the most stable image features.

For convenience, consider the partial derivatives at σ_0 for two neighbor layers of the octave in SIFT.

$$\sigma\frac{\partial G}{\partial\sigma}\Big|_{\sigma=\sigma_0}=\frac{G(x,y,k_1\sigma)-G(x,y,k_0\sigma)}{(k_1-k_0)}\equiv f(k_1,k_0)\,. \tag{6}$$

Similarly,

$$\sigma\frac{\partial G}{\partial\sigma}\Big|_{\sigma=\sigma_0}=\frac{G(x,y,k_2\sigma)-G(x,y,k_0\sigma)}{(k_2-k_0)}\equiv f(k_2,k_0)\,. \tag{7}$$

Here

$\sigma_0=k_0\sigma$

The right sides of Eq.6 and Eq.7 can be presented by Cauchy Medium Value Theory respectively as

$$f(k_1,k_0)=f^{'}(k')\,,\ f(k_2,k_0)=f^{'}(k'')\ . \tag{8}$$

Where

$$k_1<k'<k_0\,,\ k_0<k''<k_2\,. \tag{9}$$

Here we assume that

$k_1<k_2$.

Eq.8 means that the interest point produced at the deviation $\sigma_0=k_0\sigma$ can be presented more accurately by convolution with normalized Gaussian partial derivative function at deviation $k'\sigma$ or $k''\sigma$. In this meaning, to get more accurate interest points, the scale should be limited by Eq.9.

By Eq.6, after convolution with an image, the numerator will produce extreme points which are the interest points.

With the same reason, by Eq.7 the convolution also produces interest points of the image.

To ensure the stability of interest point, according to Eq.6 and Eq.7, the ideal interest points' positions should be the same point of an image. But we find the convolution produces different interest points for the same image in scale $k_1\sigma$ and $k_2\sigma$ respectively. In fact, the bias of position is resulted by the approximation in Eq.1.

Because the coefficient k is between k_1 and k_2, to attain the robust interest point and ensure the convenience in the same time, we substitute the mid-point of two extreme points for the original extreme points in the image. The new extreme point is more stable than the original ones.

Although we calculate only two pairs of the neighbor octave layers, because the layer is the base of SIFT octave, the algorithm can be extended to all layers in octave.

In statistics, the less difference the interest points' positions in different variance are, the robust the interest points are, so that the scale invariance is guaranteed more effectively.

The importance of AIPR is that the position of interest point, namely key point will become more accurate in calculation, and that the position affluences the accuracy of feature descriptors. Because the key point are the foundation of SIFT, and the extreme point is the foundation of key point, the accurate extreme position will make the key point position, neighbor cell selection in gradient summation more accurate to result the more accurate descriptor.

In Section 3, experiment will demonstrate the position bias of extreme point in neighbor octave layer subtraction, and the ratio of the extreme point needed to be adjusted will be counted too.

4. Experiments and result

To demonstrate the improvement for the interest point selection, we use a large amount of images to calculate the position bias and their mid-position, *i.e.*, the average values of their coordinates. The experiment verifies the existence of numerous extreme points needed be adjusted to get more accurate key point positions. Thereby, the improvement of the extreme point selection, which will change and improve the key point selection ability, is verified to be very important.

Figure 1. Interest point position bias and the AIPR. The red color points are the original extreme point after subtraction of neighbor layers in SIFT octave, the yellow ones are the adjusted point positions

We can observe that the adjusted points distribute in the images where there are distinct feature change. The changes are produced maybe from the silhouettes, borders, intensive variation in brightness and so on, which indicate the feature point position in images.

4.1. Selecting dataset

For the aim of object recognition, we use the INRIA Person dataset to verify the efficiency of this algorithm. These dataset includes a large amount of human with different poses in different environment. About 3500 images are utilized in our experiments. The resolutions are divided respectively into three sets, *i.e.*, 70×134, 96×160 and mixed resolutions which include different resolutions in a dataset. In fact, the result can be extended to more general images besides human images which are utilized in our experiments.

The most important reason why we select this dataset is because its generality. Although the main objects in the images are humans, but there are also a large amount of background and some are even very complicated. In this meaning, the dataset is typical for experiment.

4.2. Interest point position bias

To illustrate the position bias, we mark the extreme points calculated by subtracting two pairs of neighbor layers in an octave of SIFT respectively with red boxes, and mark the mid-point i.e. adjusted points with the yellow ones to make them more obvious. We can find the yellow boxes are always located between the red boxes.

Figure 1 displays the extreme point positions' difference after the subtraction of two pairs of neighbor layers namely the derivations $k_0\sigma, k_1\sigma, k_2\sigma$ in an octave. For comparison, we calculate also the adjusted interest point ratio (abbreviated as AIPR from now on) of each image.

The AIPR is defined as

$$\text{AIPR} = \frac{\text{amount of adjusted extreme points}}{\text{amount of total extreme points}}. \tag{10}$$

AIPR reflects the extreme points needed to be adjusted to get more accurate extreme point positions. In other word, the more AIPR is, the more the extreme points are that their positions are needed to be adjusted. Generally speaking, AIPR reflects the degree of the need to adjust the extreme point position, thereby the adjustment of the key point position.

More information about the AIPR will be discovered hereafter. Some discuss about the AIPR will be given in Section 5.

4.3. Adjusting the interest point

We calculate the AIPR of each image group and adjust their positions according to the algorithm introduced in Section 3.

The result of AIPR is illustrated in Figure 2 and Figure 3.

The average AIPR of a dataset is displayed in red line which indicates also the computing overall load for the adjustment.

In Figure 2 to Figure 3, the abscissa is the amount of images tested, and the ordinate is the amount that the coordinates of the interest points are needed to be adjusted because the extreme positions after subtracting two octave layers aren't the same.

As comparison, Figure 2 copes with images with small size, *i.e.*, 70×134 and 96×160 respectively. Ratio and average ratio of images in different size group is displayed in Table 1. It can be observed that the AIPR varies from 3% to 58%, and the average AIPR can be up to

10%, which indicates after subtraction of two pairs neighbor layers, many extreme points' positions have to be changed.

For more general circumstance, Figure 3 illustrates the AIPR of 500 images with a variety of resolution ranged from 176X257 to 1280X791.The average AIPR is 8%, and AIPR varies between 4% and 69%. It can be found the extreme points' positions needed to be changed in consecutive subtraction calculation vary their positions around their neighbor regions locally. The result can be extended to all pairs of neighbor layers in octaves.

In these experiments, AIPR varies in a large region. In summary, the minimal AIPR is above 3% and the maximal one can approach 69%. The higher the AIPR is, the more inaccurate the interest point' position is. So the elimination of AIPR increases the accuracy of interest point's position effectively.

Although we can do experiment in more dataset, we think this is enough because the dataset we used here is selected randomly from the INRIA Person dataset, which have the feature of generality, and humans' being present in these dataset make these dataset more typical for reality life.

Table 1. AIPR of different image Groups

Image resolution	*Average ratio (%)*	*Maximum ratio (%)*	*Minimum ratio (%)*
70×134	8	36	3
96×160	10	58	4
mixture	7	69	5

From Table 1, we observe the AIPR varies from 3% to 69% in our experiments. So it can be concluded the extreme point bias is general and the affluence to accurate interest point position is large. By adjusting the extreme point's position, the bias is reduced.

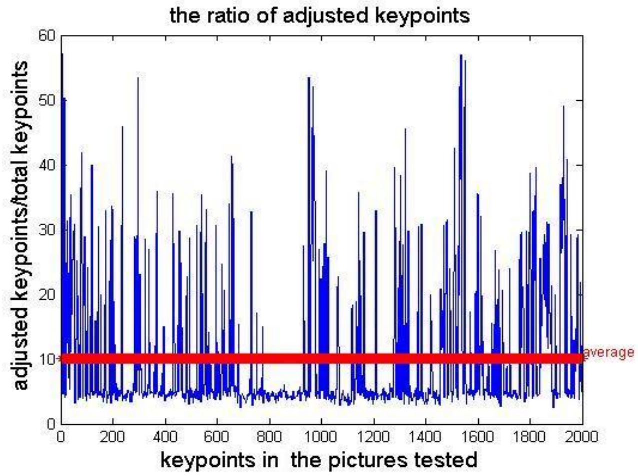
This table also reflects a phenomenon that when the image resolution is mixed, the AIPR rises to a higher level. This indicates that some images have more feature points which extreme points are necessary to be adjusted, and this comprehensible.

We can find the average AIPR of images with different resolutions varies little, although the single AIPR changes in a wide range. In these dataset, the main object is human in each image, so we can conclude that for human images the AIPR is obvious and the adjustment of extreme point selection, thereby the adjustment of key point selection is important and will have an effective improvement in key point selection.

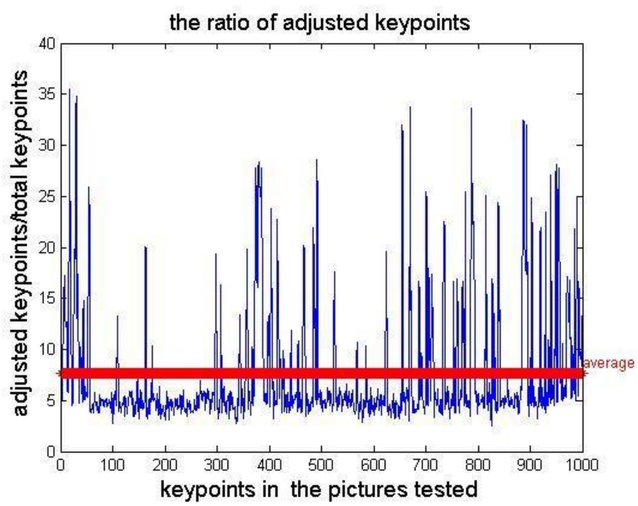
In fact, the AIPR is relative to the features in the image. The more the features are, the higher the AIPR is. In some situation, when the image is larger, the more details will be included. This can interpret the rise along with the resolution of the resolution. This is reflected by Table 1.

Let alone the average AIPR for dataset varies with the resolution of images, we find the single AIPR varies acutely in every test group. The ratio of the maximum to the minimum can be up to 15. In this meaning, we can conclude for many images, the adjustment of extreme point is more important to improve the ability of SIFT, and therefore the adjustment is necessary. We can find in some images, the AIPR rise up to above 50%, these means the adjustment will influence the performance of key point selection largely and the improvement in key point selection will become more obvious.

Notice the principle that the difference of the amount pictures tested here doesn't change the up and down variation of the AIPR. This means in the same group and between groups, because the number of feature points varies according to the images themselves, the extreme point to be adjusted varies also.



(a)



(b)

Figure 2. Interest point position adjusted ratio of 2000 images (sub figure a) with resolution of 96X160 and 1000 images (sub figure b) with resolution of 70×134

It can be observed that the ratio varies sharply. The phenomenon is because the images vary sharply too. The red line is the average ratio for the group of images.

From Figure 3, we can find some images need to be adjusted for extreme points in large scale, but the average AIPR is not very large. So the computation load is low in general. The adjustment of extreme point selection simplifies the computation in keypoint selection, so the speed can rise by this way.

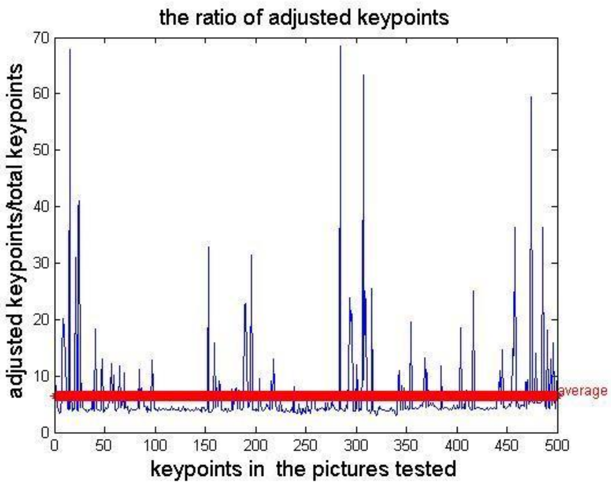


Figure 3. Interest point position adjusted ratio of 500 images with resolution from 176X257 to 1280X791

This group is more general than the ones above. It can be observed that the ratio varies sharply. The phenomenon is because the images vary sharply. The red line is the average ratio. In this experiment, the ratio of pixel number included in single image varies up to 22, so the dataset is more typical.

5. Conclusion and outlook

To increase the accuracy of the interest point's position, we use an adjusted position to approaching the ideal interest point. Because the ratio of adjusted interest points in an image's original interest points varies from 3 to 69 percent, this selection can improve the accuracy of interest points very much.

This is the typical data get from general images with human in real life, we can believe the generality of this test because the world is mainly human world and the images are representative for general life images.

We can also use the AIPR to test the complexity of images with human and change the extreme point position so as to change and speed up the keypoint for human tracing and pose recognition.

The key step in our algorithm is that we can get the extreme point more effectively in SIFT by the midpoint to speed up the key point selection.

One of the advantages of this key point selection is the calculation reduction. In the real-time signal processing, speed is the key problem in many cases.

This is important for moving object recognition, esp. human pose discrimination. When we want to find and trace a specified human in a video, the system should have the ability to trace the man as fast as possible. Otherwise, after the key points are extracted after a long time, the man has moved to other place for a long time, and the system can not find the specified man from then on because the calculation costs too much time. This work is useful for safety system.

Our work in the future will focus on the exact performance analysis in speed, because this parameter is important for real time signal processing and the application, and some comparison will be analyzed to calculate the improvements in speed more accurately. Intensive research on the key point selection method in static and moving object, which may be human and other objects, will be conducted.

Because of the relationship of AIPR to the change of image features which we have indicated in Section 4.3, in some case, we can also use AIPR as a parameter to reflect the complexity of an image and use the average AIPR for dataset, and this can be another application for the AIPR.

The key point selection is the kernel problem in many computer vision systems. To represent the features of objects, many algorithms based on SIFT have been proposed. The improvement in key point selection will influence the ability of object recognition in a wide scope.

For further research, we will consider the experiment of AIPR for video images, which frames are continuous and correlative. In this case, we will take account of the relationship of neighboring frames, and prediction is a factor to influence the adjustment and therefore the AIPR will have some new features.

Acknowledgements

In the experiments, we have used thousands of INRIA Person dataset. These dataset are arranged in different resolution and the class is abundant. We use these to verify our algorithm in humans. We thank sincerely the researchers in INRIA for their image collection work, which makes our experimental verification work more convenient. These dataset are important for our experiments.

References

[1] D. G. Lowe, “Distinctive image features from scale-invariant keypoints”, International Journal of Computer Vision (IJCV), vol. 60, no. 2, **(2004)**, pp. 91-110.

[2] N. Dalal, “Finding People in Images and Videos”, Ph.D. thesis, Institute National Polytechnic de Grenoble, **(2006)**.

[3] N. Dalal and B. Triggs, “Histograms of Oriented Gradients for Human Detection”, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition CVPR, vol. II, **(2005)** June, pp. 886-893.

[4] Y. Ke1 and R. Sukthankar, “PCA-SIFT: A More Distinctive Representation for Local Image Descriptors”, the IEEE Conference on Computer Vision and Pattern Recognition CVPR, vol. 2, **(2004)**, pp. 506-513.

[5] H. Bay, A. Ess, T. Tuytelaars and L. V. Gool, “Speeded-Up Robust Features (SURF)”, European Conference on Computer Vision ECCV, **(2006)**, pp. 404-417.

[6] K. Mikolajczyk and C. Schmid, “An affine invariant interest point detector”, European Conference on Computer Vision ECCV, **(2002)**, pp. 128-142.

[7] Y. Li, L. Liu, L. Wang, D. Li and M. Zhang, “Fast SIFT Algorithm based on Sobel Edge Detector, Consumer Electronics, Communications and Networks (CECNet)”, **(2012)**, pp. 1820-1823.

[8] T. Chaves, L. Figueiredo, A. Da Gama, C. de Araújo and V. Teichrieb, “Human Body Motion and Gestures Recognition Based on Checkpoints”, 14th Symposium on Virtual and Augmented Reality, **(2012)**, pp. 271-278.

[9] F. Y. Qing, S. B. Lin and W. J. Fang, “An improved SIFT algorithm”, Journal of Harbin Engineering University, vol. 31, **(2010)**, pp. 560-564.

[10] M. Wen and C. Zhong, “Application of Sobel Algorithm in Edge Detection of Image”, China High-tech Enterprise, **(2008)**, pp. 57-62.

[11] Z. Hou, L. Zhao, L. Gu, G. Lv and M. Wang, “An improved Image Registration Method Based on SIFT”, 2009 International Asia Symposium on Intelligent Interaction and Affective Computing, **(2009)**, pp. 87-90.

[12] K. Mikolajczyk and C. Schmid, “Indeximg Based on Scale Invariant Interest Points, Proc. Conf. Computer Vision, vol. 1, **(2001)**, pp. 525-531.

[13] C. Belcher and Du Yingzi, "Region-based SIFT approach to iris recognition", Optics and Lasers in Engineering, vol. 47, **(2009)**, pp. 139-147.
[14] Z. Huiyu, Y. Yuan and S. Chunmei, "Object tracking using SIFT features and mean shift", Computer Vision and Image Understanding, no. 13, **(2009)**, pp. 345-352.

Authors

Ding Zuchun is currently doing his exertions to research on computer vision, human pose discrimination and communication network also. He is always hardworking at his research work, which is proved to be useful for him to study for BSc, MSc and PhD from Zhejiang University, National University of Defense Technology and South China University of Technology respectively. He has been in charge of several national research projects and taken part in some application research work on relative domain, including signal processing and communication network successfully.