Improvement on Image Rotation for Relative Self-Localization Estimation

Xing Xiong and Byung-Jae Choi

School of Electronic Engineering, Daegu University, Korea
GaleWing@gmail.com and bjchoi@daegu.ac.kr

Abstract

There are many methods to localize its position based on visual sensing schemes in indoor environment. This paper presents the problem of finding the correspondences of images feature's descriptors when the images have large rotation. SIFT and SURF have always been considered as very effective algorithms to extract interest points and their orientation and descriptors. For descriptors, one of both uses a lot of time to calculate descriptors and the other has not good performance in large rotation of the image. In this paper, we propose an improved algorithm to calculate interest points' descriptors for relative self-localization estimation. The proposed algorithm will satisfy descriptor invariant when the image rotates. Meanwhile, the proposed method reduces the calculated time as much as possible. Interest point's descriptors are formed by resampling local regions.

Keywords: *SURF, SIFT, interest points, descriptor, image rotation, subregion*

1. Introduction

With the development of science and technology, mobile robots have been employed for a wide area of industrial applications including factory automation, home automation, medical assistance, and rehabilitation for the provision of new forms of services. The robots need to know where it is and how move to the next position before it do that. Because the use of GPS (Global Positioning System) is limited or not feasible in indoor environments, the problem of calculating the self-localization of the robots become difficult. Many studies have focused on the problem. For example, in [6, 7, 8, 9], vision-based algorithms are used to estimate its self-localization.

The most of these researches uses computer vision algorithm to extract image features and match them for estimating absolute or relative self-localization of robot because the visual images provide a lot of valuable information such as color, texture, and shape of objects. Image matching algorithm plays a big role in these areas.

Image matching is to find correspondences between two images or more when there are large changes in viewpoint, scale, affine, blur, and illumination. The correspondences can be represented by similar objects, their shapes, or interest points. Interest points are more widely used to represent image correspondences because of their high accuracy. Some researches have verified that the interest points in scale space have good performance in viewpoint, scale, blur and illumination changes because interest points' local descriptors are highly discriminative for image matching.

The most successful algorithms for image matching are SIFT (Scale Invariant Feature Transform) and SURF (Speeded Up Robust Features). Both have good performance in interest points' extraction, orientation and descriptors of interest points, and image matching. There are some discussion and comparison for them and PCA (Principal Components Analysis)-SIFT in [5]. Although the results of SIFT is better than SURF, it takes four times or more as long as SURF. Just for this reason, the SIFT is more difficult in mobile robot area because it may be difficult to need real time operation. The SURF saves a lot of calculating time and has good performance in the translation and illumination changes of images, but it has not good performance in its rotation, particularly in large rotary angle. Through the analysis of SIFT and SURF, a new algorithm for solving this problem is proposed. The proposed algorithm no longer uses integral image to calculate descriptors. The descriptors are invariant as far as possible no matter what environment changes and satisfy interest points matching.

The organization of this paper is as follows: In next section we introduce some related works. We explain the description of interest points in Section 3. Some simulation examples are shown in Section 4. Finally, we present some concluding remarks

2. Related works

Feature descriptors' constitution has been researched from 1981 and a series of new algorithms were proposed. The most famous methods of all are SIFT, PCA-SIFT and SURF. In [1], the SIFT used a gradient orientation histogram in the neighborhood of the interest points to define the orientation of the interest points. The maximum in the histogram is considered as the dominant orientation of the interest points. After orientation of an interest point has been defined, a 4×4 array of histograms with eight orientation bins is extracted. Therefore, $4 \times 4 \times 8 = 128$ element descriptor vectors for each interest point is used in SIFT. It needs huge amount of computation time in SIFT. After all its application area is limited, especially for real-time computation in some low speed chips.

In [10], a new method for calculating descriptors using PCA based on SIFT was proposed. The PCA-SIFT uses 36-dimensional descriptors to replace 128-dimensional descriptors in SIFT. Through the comparison with SIFT and PCA-SIFT, PCA-SIFT is more fast for descriptors calculating and interest points matching, but causes more errors because of its less distinctive.

The other famous research is SURF in [2]. The SURF shows the fastest comparing with above both, and it also has similar results with SIFT. The SURF using integral image and Haar wavelet responses to accelerated compute orientation and descriptors of interest points. For experimental results in [5], the SURF showed good performance in time, scale, and illumination changes. However, the SURF becomes difficult to match interest points and leads to a huge mistakes when the angle of image is large.

3. Description of Interest Points

In [2], the operation of descriptor extraction, the first square region was centered around the interest points and orientated along the orientation of interest points. The square region was split up to regularly into 4×4 square subregions. For each subregion, Haar wavelet filters are used to filter every sample point at x and y-direction. Then, the responses at x and y-direction are summed up over each subregion. In order to increase the distinctiveness, the absolute values of responses at x and y-direction are summed up over each subregions again. Therefore, the descriptors ares formed by 4 × 4 × 4 = 64-dimensional vectors.

After a lot of experiments, Haar wavelet response shows good performance in illumination change, but it has not good performance in rotation. Because Haar wavelet filter is rectangle filter, it is easy to overlook some specific and less distinctive.

For the reason, a circle is extracted around interest points. The size of circle is $20s$(s is the scale when the interest point was detected). Examples of such circles are illustrated in Figure 1. The extracted circle is filtered by Gaussian filter in equation (1).

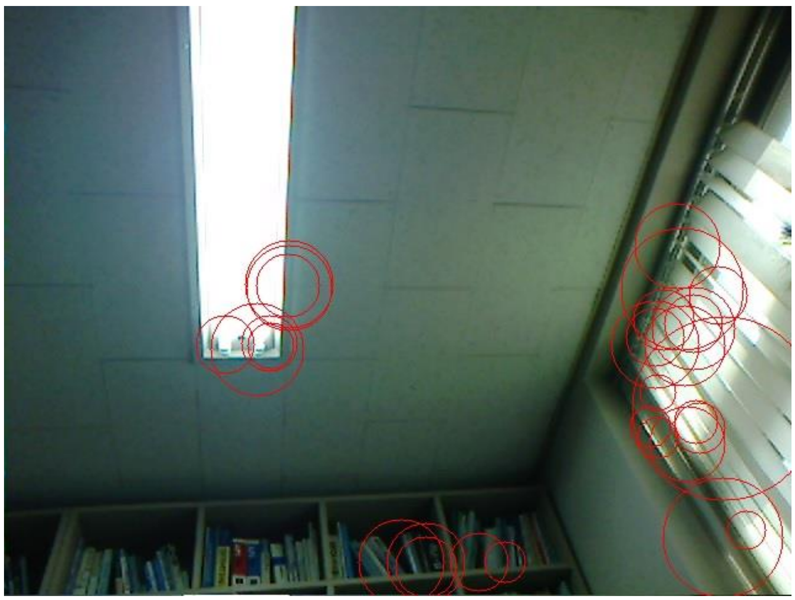


Figure 1. The size of circles at different scales

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

The circle is split up regularly into smaller 2 × 2 subregions along the orientation of interest points. An example is illustrated in Figure 2.

3.1. Edge pixels value

The pixels which not contained in circles will be removed. For reason to do that is to keep invariant local area around interest points when image rotates, the circle can satisfy all the rotation. However, in fact, because of image discreteness, some pixels are at the edge of circles. For example, the edge pixels are partly belong to the circle and other out of the circle in Figure 3. The pixels which are on the edge of circle cannot be removed directly.

For keeping invariant of descriptors, a method which is similar to bilinear interpolation is used to replace the edge pixels value, as equation (2).

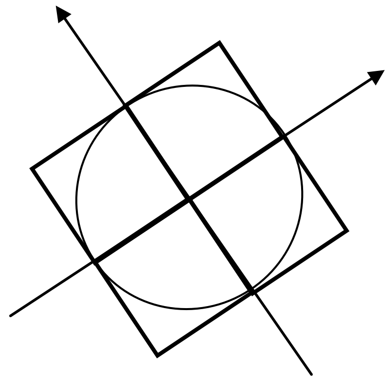


Figure 2. The 2 × 2 subregions (the direction along the orientation of interest points)

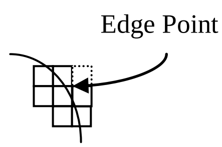


Figure 3. The Pixels on the Edge of Circles

$$F(i,j)=(1-u)(1-v)f(i,j)+(1-u)vf(i,j+1)+u(1-v)f(i+1,j)+uvf(i+1,j+1) \quad (2)$$

$$u=\frac{y^2}{T^2} \text{ and } v=\frac{x^2}{T^2} \quad (3)$$

$$T=\sqrt{x^2+y^2} \quad (4)$$

where $F(i, j)$ and $f(i, j)$ represent the value of after and before interpolation, respectively. x and y represent normalized value between the distance of the edge pixels and x-axis and y-axis, respectively. For the reason to do so is that keep $u+v=1$. The new edge pixels are replaced by their four neighborhoods.

3.2. Neighborhood influence

In SIFT [1], before calculating the orientation of interest points, for image matching more stable, the gradient magnitude and orientation for each image pixel is precomputed through Gaussian smoothed image. Then an orientation histogram is formed to find the orientation of interest points. For the approach, the details in image changes can be demonstrated more clarity. The descriptor distinctiveness for each interest points increases. Meanwhile, the computational complexity also increases (one for gradient magnitude, one for orientation of pixel).

For increasing descriptor distinctiveness and less calculating time, we considered each pixel's influence in eight difference directions. For example, calculating local neighborhood's influence in 45º and 305º is used equation (5) and (6) in Figure4. Some details in pixels in the circle are shown in the figure.

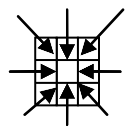


Figure 4. The influence in pixel's neighborhood

Therefore, for each pixel in the circle, it has eight values and ordered from top to bottom and from left to right.

$$L(3) = f(i, j-1) - f(i, j) \tag{5}$$

$$L(8) = f(i+1, j+1) - f(i, j) \tag{6}$$

3.3. Descriptors

The descriptor is formed by summing eight values for each pixel. For example, a subregion has 4 pixels and every pixel has eight values, sums 4 pixels all the value in different direction in Fig. 5. Therefore, one part of descriptor is formed by the sum of this subregion in different direction. One subregion contains eight values as shown in Figure 6.

According to the analysis in [2], the absolve value is useful for increase distinctive of descriptors. Then, the sum of absolve value of 4 pixels are calculated. Therefore, the descriptors are formed by 2 × 2 × 8 × 2 = 64 dimension vectors.

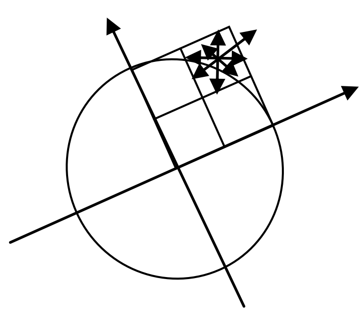


Figure 5. Every pixel in circle has eight values in eight different directions

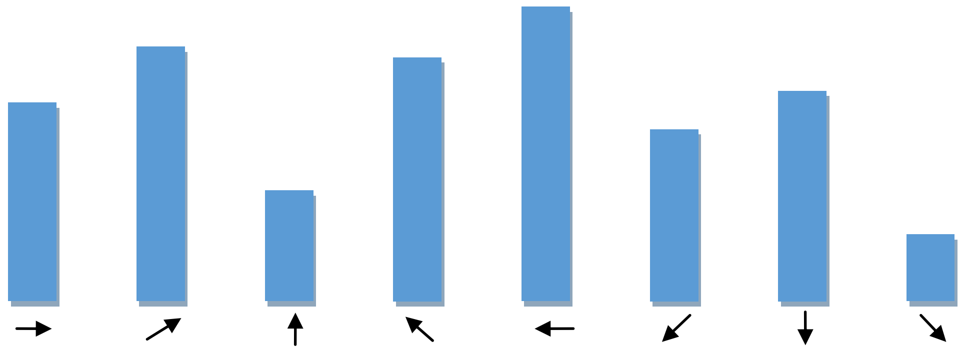
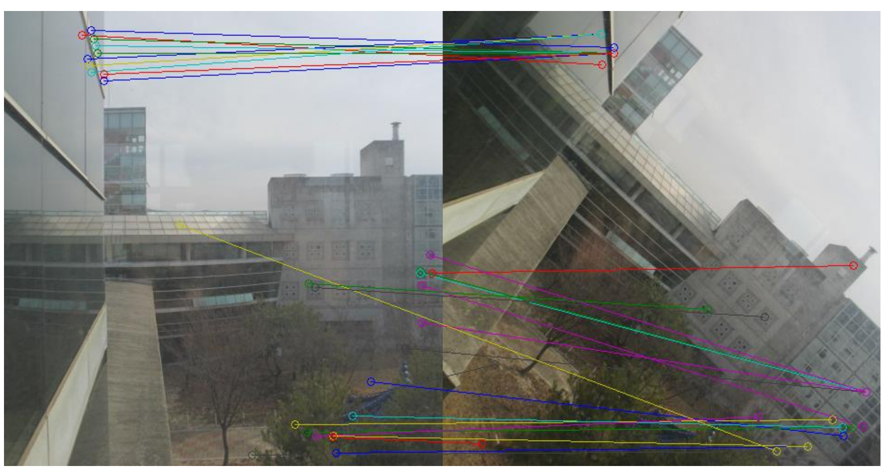


Figure 6. One subregion. Each subregion contains eight values in different direction

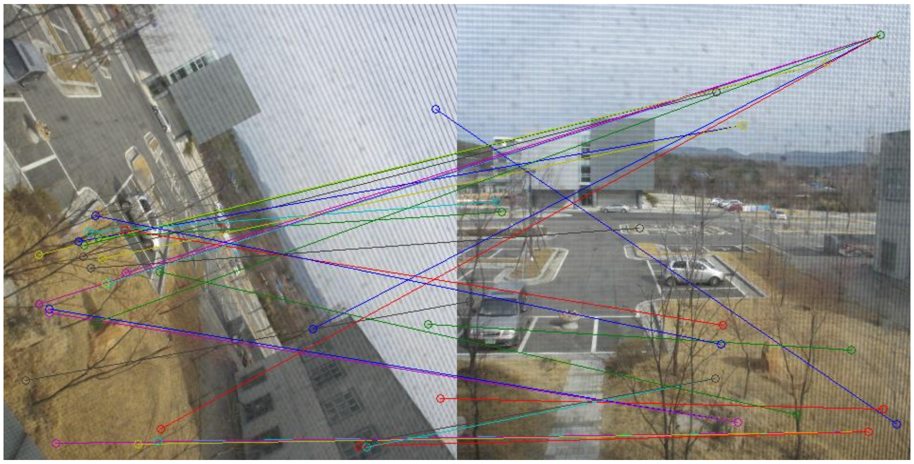
4. Simulation Results

In order to ensure the validity of the algorithm, simulation examples are presented. They are implemented in the conventional SURF and the proposed algorithm. Simulation results are shown in Figure 7 and Figure 8. The images are shot in different environment using the same camera. The Figure 7 and Figure 8 are results using SURF and the proposed algorithm, respectively. They are executed in the same hardware and software environments. For comparison, they used the same images in Figure 7 and Figure 8. Thirty of the highest similarities are connecting lines in images.

From simulation results, the proposed algorithm shows better matching of interest points when the interest point rotates.



(a)

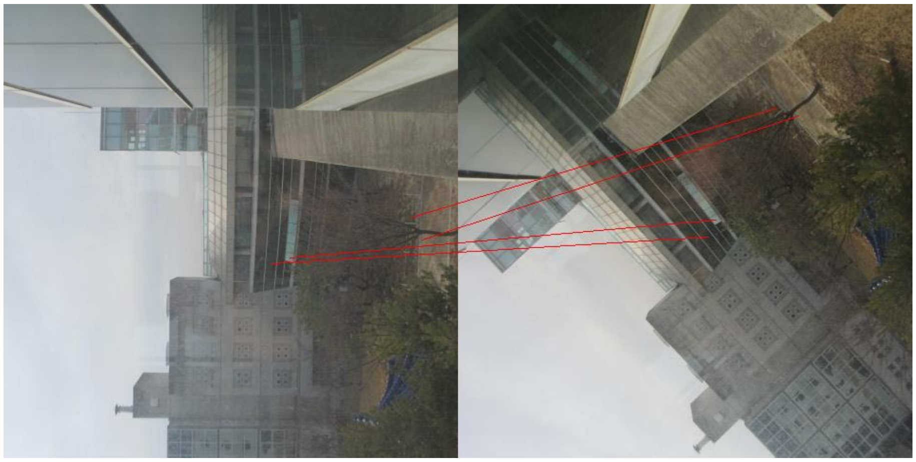


(b)

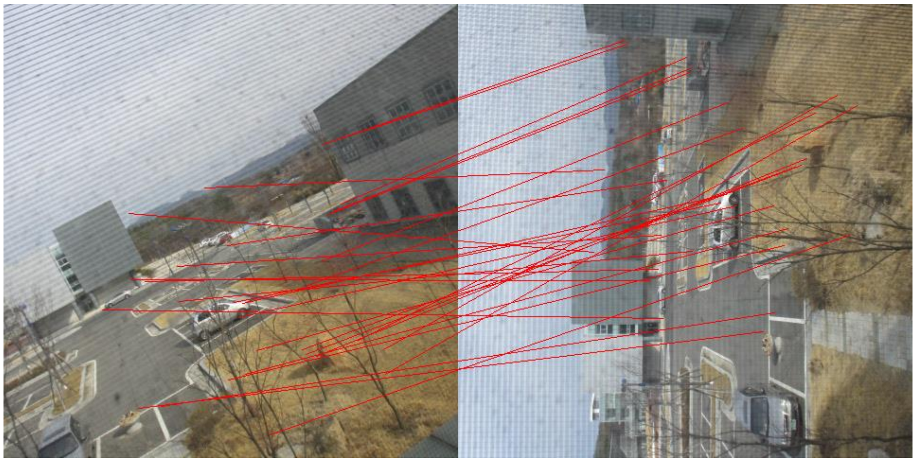
(c)

(d)

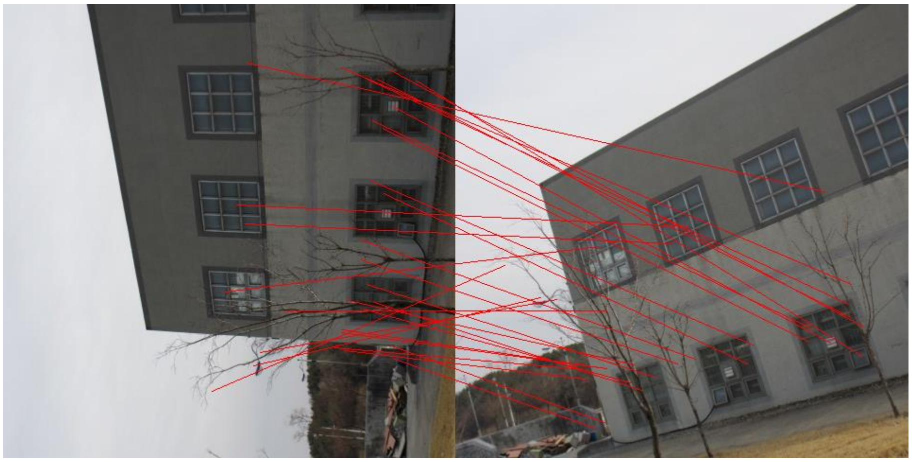
Figure 7. The simulation results using the conventional SURF



(a) There are only four interest points to be considered as the same



(b)



(c)

(d)

Figure 8. The simulation results using the proposed algorithm

5. Concluding Remarks

In this paper, for the condition of descriptor invariance when image rotates very large, a new algorithm was proposed. For avoiding the error caused by rectangle filter and increased distinctiveness of descriptors, Haar wavelet filters did not used in the neighborhood of interest points in the proposed algorithm. The neighborhood changes of pixel's are considered as descriptors. And a new method is proposed for solving the ascription problem of the pixel in the edge of neighborhood around interest points. The simulation results showed that the performance of the proposed algorithm is better than that of the SURF in image rotation.

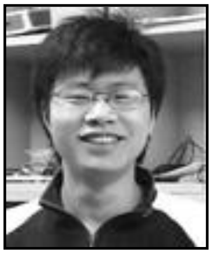
Although the proposed algorithm solves the problem of image rotation, the calculating time is added because integral image is disabled. In the future, we will improve the computation speed.

References

[1] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision, **(2004)**.
[2] H. Bay, A. Ess, T. Tuytelaars and L. V. Gool, "Speeded-Up Robust Features (SURF)", **(2008)**.
[3] D. C. K. Yuen and B. A. MacDonald, "Vision-Based Localization Algorithm Based on Landmark Matching, Triangulation, Reconstruction, and Comparison", IEEE Transactions on Robotics, vol. 21, no. 2, **(2005)**, pp. 217-226.
[4] A. Kitanov, "Mobile robot self-localization in complex indoor environments using monocular vision and 3D model".
[5] L. Juan and O. Gwun, "A Comparison of SIFT, PCA-SIFT and SURF", International Journal of Image Processing (IJIP), vol. 3, no. 4, **(2010)**.
[6] N. A. Andersen, L. Henriksen and O. Ravn, "Visual positioning and docking of non-holonomic vehicles", Proc. 4th Int. Symp. Experimental Robot, vol. 223, **(1995)**, pp. 140–149.
[7] K. -C. Chen and W. -H. Tsai, "Vision-Based Autonomous Vehicle Guidance for Indoor Security Patrolling by a SIFT-Based Vehicle-Localization Technique", IEEE Transactions On Vehicular Technology, vol. 59, no. 7, **(2010)**.
[8] S. -Y. An, J. -G. Kang, L. -K. Lee and S. -Y. Oh, "SLAM with Visually Salient Line Features in Indoor Hallway Environments", Journal of Institute of Control, Robotics and Systems, vol. 16, no. 1, **(2010)**.

[9] Pablo Sala, Robert Sim, Ali Shokoufandeh and Sven Dickinson, "Landmark Selection for Vision-Based Navigation", IEEE Transactions on Robotics, vol. 22, no. 2, **(2006)**.
[10] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors", In CVPR, vol. 2, **(2004)**, pp. 506-513.

Authors



Xing Xiong received his B.S. degree in automation engineering at NanJing Institute of Technology, and M.S. degree and Ph.D from Daegu University. He is currently Doctor Course in the School of Electronic Engineering from Daegu University. His research interests include intelligent control and embedded systems.

Byung-Jae Choi received his B.S. degree in Electronic engineering at Kyungpook National University, and M.S. degree and Ph.D from Korea Advanced Institute and Technology. He is currently Professor in School of Electronic and Electrical Engineering from Daegu University. His research interests include intelligent control theory and its applications.