A novel image registration algorithm based on salient region and SIFT

**Abstract.** The Scale Invariant Feature Transform (SIFT) is a classical algorithm in the field of feature point detection, and is utilized widely in image registration because of its robustness. Meanwhile, there are so many essays derived from SIFT pay attention to improving its efficiency and taking advantage of other detection methods. In this paper, we focus on figuring out features which can represent more information and are more likely to be captured by eyes. So we combine SIFT with Frequency-tuned Salient Region Detection (FT) and promote performance compared to SIFT-only algorithm. What’s more, we applied Salient Region Segmentation Algorithm to image registration in order to avoid detecting unnecessary features and decrease computational complexity. This paper also proposes a modified segmentation method by padding images, which lead to better performance.

**Keywords:** SIFT; Mean-shift; saliency region; FT; Image registration; RMSE

1. Introduction

As for two or more images that describe the same scene, they may come from different sensors at different times with different view angles, so there are distortions between them. Considering follow-up image processing, it is necessary to find out matches and transform the selected image to the reference one and we call the processing as image registration. Image registration is a crucial step for image mosaic, object identification, image fusion, change detection [1], image stitching and has been widely used in medical diagnosis, satellite remote sensing and artificial intelligence. The process of image registration can be divided into feature detection, feature matching, estimation of parameter of transformation model, and image transformation. According to the selection of features in the first step, feature detection can be classified as feature detection based on feature points [1] and based on feature regions, namely feature-based and area-based methods [5]. Since features could be points, lines, contours and moments, Lowe find a way to obtain feature points which are invariant to image scale and rotation, and can provide robust matching across a substantial range of affine distortion, change in 3D viewpoint, addition of noise, and change in illumination [6]. Following Speeded Up Robust Features (SURF) algorithm [7] and KAZE algorithm[8] are both refined on the basis of SIFT, focusing on speeding up computation and non-linear transformation respectively.

In this paper, we take into account Human vision system that tends to firstly pick some interesting regions. So based on the SIFT algorithm, the significant region detection FT method [9] is added, and more feature points representing meaningful information can be extracted to improve the effect of image registration by achieving lower Root Mean Square Error (RMSE). Sometimes subject of image registration may be part of the whole image, so we don’t have to match feature points in the background, and we can extract feature points in regions of interest (ROI) by making use of Achanta’s another contribution [10] and reduce the amount of calculation. The results of experiments show that this method has good resistance to noise in the background. In addition, in the article, we refine the segmentation algorithm by expanding images and filling the rest part with mean value of pixels artificially. The modification makes it easier to distinguish important parts and regular parts.

The rest part of the paper is organized as follows. Section 2 makes a brief introduction to the SIFT algorithm. Section 3 presents several methods of salient region detection including FT algorithm especially and the approach of image segmentation based on FT. The main contribution of this work is placed in section 4, where we record details and ideas of our research. Experiment results and analysis are reported in section 5 and conclusions are located in section 6.

1. Scale-invariant feature transform

The first step is to build an image pyramid to achieve the representation of scale space of image [11]. proved that the Gaussian convolution kernel is the only linear kernel that realizes scaling, so the scale space of a two-dimensional image is defined as:

 (1)

 is scale variable Gaussian function.  is the input image.

The purpose of constructing the scale space is to detect the feature points that exist at different scales, and the operator that detects the feature points well is (Gauss Laplace):

 (2)

The amount of computation is too large, and it can be approximated by Difference of Gaussian (DoG).

 (3)

The detection of feature points is performed in a pyramid of DoG, and compare the detected point with its 8 adjacent points of the same scale and the 26 points of the 9×2 points corresponding to the upper and lower adjacent scales to get extreme points, ensuring detected points are odd both in the scale space and the two-dimensional image space.

Since the discrete image space is the result of sampling the continuous space, the extreme points found in the discrete space are not bound to be the extreme points in the true sense, so we must try to eliminate the points that do not meet the conditions. The points to be eliminated are low contrast feature points and points on the edge which are unstable.

We have to assign direction information for feature points in order to guarantee feature points are invariant to the change of angle and rotation. The distribution of directions is achieved by finding the gradient magnitude and orientation of each extreme point.

For the block of pixels around the feature points, the gradient histogram within the block is calculated, which includes the influence of the surrounding points on the feature points, and we can generate unique vectors, namely descriptors of feature points.

According to the descriptors of feature points, the matching pairs between the two images are obtained, and then the transformation model can be estimated to achieve image registration. More details please see [6].

1. Salient Region Detection and Segmentation
   1. Frequency-tuned Salient Region Detection

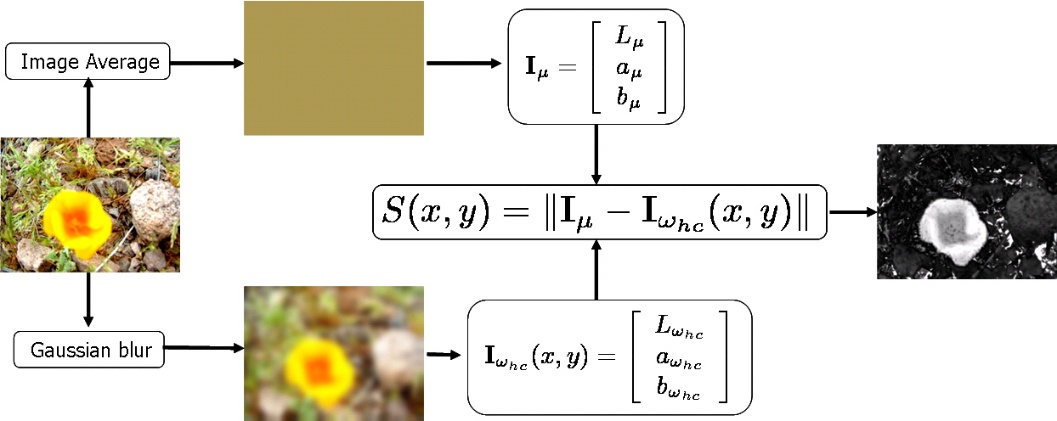
The research of salient region is inspired by the human vision system, which can actively seek interesting regions in images to reduce the search effort in tasks, such as object detection and recognition [12]. Since ltti presented a visual attention system [13], lots of work has been concentrating on saliency detection. X Hou and L Zhang eliminate background by analyzing the log-spectrum of a input image. [14]. Yun Zhai propose a spatiotemporal video attention technique while Achanta’s method outputs full resolution saliency maps with well-defined boundaries of salient objects [9]. The FT method exploits feature of color and luminance, and is simple to implement.

Like most saliency detection methods, saliency is converted to the relationship of the current pixel and the neighborhood in terms of luminance and color properties. The difference is that the FT algorithm takes the entire image as a neighborhood. Pixel saliency can be calculated using the following formula:

 (4)

Owing to the fact that *CIRLab* color model is a device independent color model and a color model based on physiological characteristics, the FT algorithm use *CIELab* as color space, so  and  are three-dimensional vectors.is the mean vector of pixels in the image,is the smooth image blurred by a Gaussian kernel, and is the L2 norm. In a word, we calculate the Euclidean distance between each point and the mean vector in Lab space as a significant value.

The FT algorithm can be divided into four parts. Gaussian blur on original image; get LAB model version of original image; calculate mean value of *CIELab* vector; calculate the L2 norm of difference value between pixel and mean value. Complete process can be seen in the Fig.1 as below.



**Fig. 1.** FT method find the Euclidean distance between the Lab pixel vector in a Gaussian filtered image with the average Lab vector for the input image

* 1. Salient Region Segmentation Algorithm

Achanta also demonstrate the use of final saliency map in segmenting whole objects with the aid of a relatively simple segmentation technique [10]. Using a simple K-means algorithm, we can get over-segmented image. Since K-means algorithm clusters pixels in the *CIELab* color model, we can get saliency value of per segments respectively. Furthermore, instead of K-means segmentation Mean-shift segmentation is used in Achanta’s later research.

The Mean-Shift algorithm has a wide range of applications in clustering, image smoothing, segmentation, and tracking. Mean-Shift algorithm generally refers to an iterative step, that is, firstly we calculate the mean value of the current point offset before moving the point to its mean value of the offset, and then take this mean value as a new starting point. The move is continuous until it meets certain conditions.

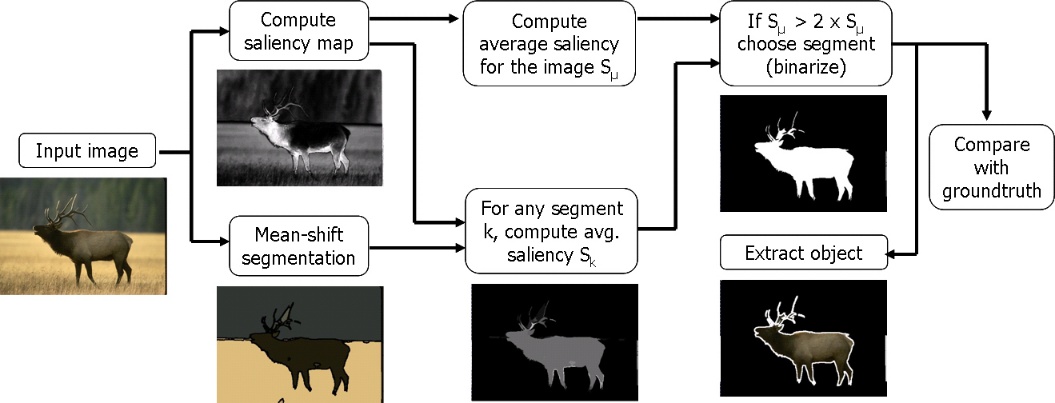
In the Mean-Shift algorithm, the most crucial thing is to calculate the average value of the offset of each point, and then update the position of the point according to the newly calculated offset mean. The basic form of Mean-Shift vector is:

 (5)

 refers to a high-dimensional sphere with radius . It’s a set of points that satisfy the following relationship.

 (6)

The general flow of segmentation is to compute saliency map and Mean-Shift segmentation of the input image, and then calculate the averages of saliency in each segmentation label. If the average value is two times greater than the average saliency of the original image, consider it as the main object region. Binary corresponded region to 255, so that we get a binary mask, and if we exert the mask on the original map, we will get the main object finally.



**Fig. 2.** The flow of segmentation. Combine two algorithms to realize image segmentation.

1. Methodology
   1. SIFT on saliency map

SIFT algorithm has hit a great success in various image processing. However, it doesn’t take human vision system into account. So it will be fine if we compute saliency map with FT firstly before detecting feature points with SIFT. It will bring two gains as below.

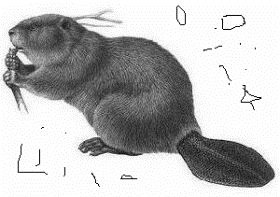
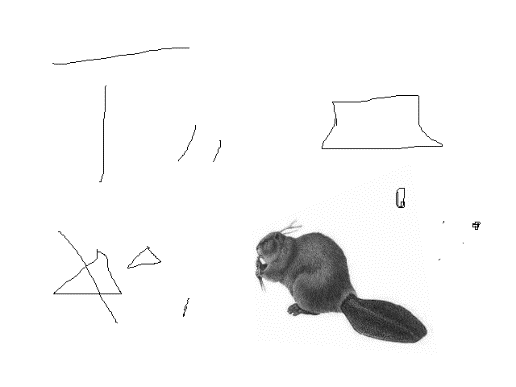
Firstly, we make use of luminance and color properties in FT and don’t increase number of channels in SIFT. SIFT algorithm requires input image to be gray image, and that’s means SIFT ignore information of colors. Saliency map not only can capture meaningful regions and make use of color information and luminance information, but also outputs gray image (normalize the salient values from 0 to 255, so the final result is a grayscale image) which is perfect for the implement of SIFT. FT is computationally efficient, so it won’t affect the speed too much.

Second, we know that the process of image pyramid is a linear operation including Gaussian convolution and downsampling while FT is nonlinear. Gauss pyramid is at the expense of accuracy, which is likely to cause blurred boundaries and missing details. SIFT on saliency map will result in nonlinear pyramid respect to original image and polish up performance in return.

* 1. SIFT on segmentation

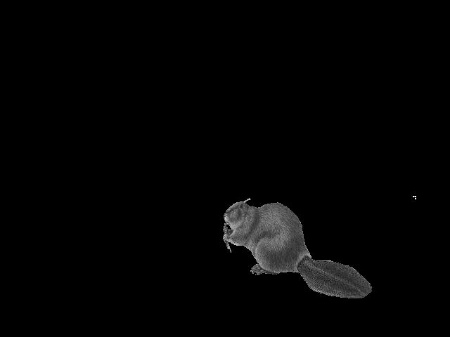
Furthermore, we apply salient region segmentation proposed by Achanta to image registration. It is helpful to reduce unnecessary feature points in the background when object we interested in is mixed with noises.

It will be illustrated with the graffiti example.

**Fig. 3.**  Graffiti on beaver.png and beaver\_xform.png provided by Rob Hess, who provide open source code of SIFT. The size of left image is  and size of right is.

(a) Graffiti on beaver.png

(b) Graffiti on beaver\_xform.png

As shown in the figure, many random pixels and lines are artificially added in the blank space of the image, and these have ill effects on image registration and increase the amount of calculation. So we can use the segmentation method mentioned above.

(a)

(b)

(a)

**Fig. 4.** The split results of images in Fig.3 with salient region segmentation algorithm. The right image is much better than the left image because its background is bigger and the contrast between object and background is higher.

It can be seen that the effect of (b) in Fig.4 is obviously better than that of (a). In contrast, a part of the main object is lost in (a), and some new feature points are generated which will have a bad influence on registration. We propose an improved method inspired by the segmentation result of (b). The segmentation effect is actually related to the extraction of the salient region in the first step. The greater the difference between the main part and the average of saliency map, the better the segmentation effect. Because we take the double average saliency as threshold, one of the ways to improve the difference value is to increase the proportion of the background.

Since the background of beaver image is pure white, it is clear to pad the background with white pixels. But in most cases the background is far more complicated than this, so we select the average value of image as the element to fulfill background.

1. Experimental results and analysis

We did some tests on our PC with Visual Studio2013 and OpenCV 2.4.13 on windows7. Experiments were performed on images from Rob Hess’s website of Github. We measure the speed by average run time and measure the quality of registration by Root Mean Square Error (RMSE). PMSE depends on the accuracy of feature points and choice of transform model, it may not reflect the final result of registration, so Peak Signal to Noise Ratio (PSNR) is applied to measure the similarity between reference image and registration result.

 (7)

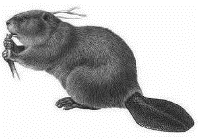
Whereand are matched feature points, is the spectra of transform model, and is the number of matches. We expect that points generated fromafter the transformation based on are close toand we can measure the precision by RMSE. RMSE is related to the precision of feature points and the accuracy of transform model.

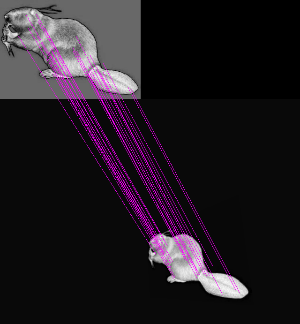
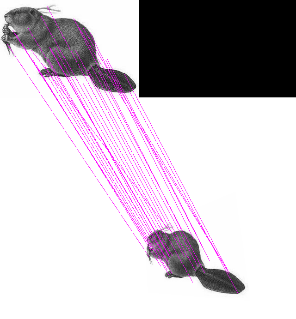
 ()

 ()

PSNR is an objective evaluation standard of image reconstruction and it is defined with mean-square error (MSE). Imageand imagein equation (8) are as big as.refers to the maximum value of pixels in images. The bigger PSNR is, the result of output image is better.

* 1. Comparison between tradition SIFT and SIFT on Ft





**Fig. 5.** Comparison of SIFT-only and SIFT on saliency map. Upper images are two input images and lines show the matches in original images and saliency maps.

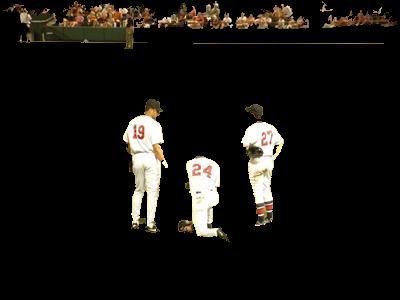
**Table 1.** Feature points detection directly on the original image and feature points detection based on their saliency maps. N1 and N2 present the number of feature points in input images.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | N1 | N2 | Time/ms | PSNR/dB | RMSE |
| Original image | 116 | 95 | 1597 | 38.719 | 18.381 |
| saliency map | 353 | 167 | 2140 | 43.920 | 0.556 |

It can be seen in experimental results that the SIFT detection based on the saliency maps obtain more feature points, and the achieved registration result has also improved by about 5dB while RMSE decrease from 18.381 to 0.556 at the cost of slowing down.

* 1. SIFT on subject in a messy background

We compared image segmentation effect based on the salient region segmentation algorithm and that of fulfilling image background proposed in this paper. The following images are from the 10K data set build by Microsoft Research Asia (MSRA).



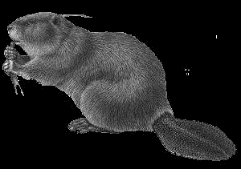
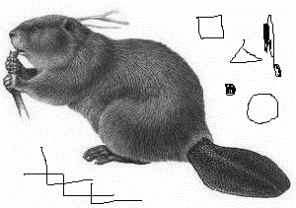




**Fig. 6.** Comparison of the two methods extracting the main part from image. The first column is the image before the segmentation, the second column is the segmentation image obtained by the method of Achanta, and the third column is the image obtained by our method.

You can see that the segmentation after filling the background retains more detail. Compared with the original method, the result of the modified method is more complete. In the first row, green belts in the distance and a head were reserved. The sign board was completely retained as well as the gaps in the woods. Rear window of the car in the third row is more complete too. In short, the improved algorithm is mainly used to retain more of the larger color difference with the background.

And we can see the performance in the artificial graffiti image (a) from Fig.4:



(a)

(d)

(c)

(b)

**Fig. 7.** The performance of the three segmentation methods. (a) is an image that artificially creates (a) graffiti background, (b) is the effect of direct segmentation,(c) is the segmentation effect after filling the background with white pixels, and (d) is the segmentation effect after the background is filled with the mean value.

It can be seen in Fig.7 that although the effect (d) is worse than being filled with white(c), it is improved compared with the direct segmentation (b). In fact, padding with white pixels in (c) takes advantage of a prior knowledge that the background is pure white, and method of padding in (d) can be applied in more situations.

**Table 2.** Distinctiveness of methods of salient region segmentation in image registration. N means the number of feature points detected by SIFT

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **(a)** | **(b)** | **(c)** | **(d)** |
| **N** | 144 | 116 | 124 | 117 |
| **RMSE** | 16.483 | 1.476 | 12.651 | 6.460 |
| **Time/ms** | 1849 | 1615 | 2258 | 1622 |
| **PSNR/dB** | 36.140 | 25.616 | 15.851 | 27.098 |

As listed in Table 2, segmentation can decrease the number of feature points, which lead to reduce the amount of calculation in SIFT. All segmentations cut down number of feature points, RMSE and run time at the cost of PSNR, but the segmentation we proposed has the highest PSNR.

1. Conclusion

The combination of the saliency region extraction method FT and feature point detection algorithm SIFT is proposed to achieve image registration. We exploit properties of color and luminance and take human vision system into account by FT, and because saliency map is grayscale image, we can exert SIFT on it directly. Experiments show that PSNR get increased and RMSE get decreased. When the image has a complex background and our region of interest is a part the image, image segmentation based on saliency detection can be used. This article uses the method of expanding the background area to improve the contrast between the subject and the background, and can retain more main parts. Experimental results turn out that it is indeed possible to reduce the RMSE of feature points.

This work is gloriously supported by the National Natural Science Foundation of China (No.61372175) and National Key Laboratory Foundation under Grant 61424110404162411004.

1. Renference
2. Ye Y, Xiong L, Shan J. Automated Multi-Source Remote Sensing Image Registration Based on Phase Congruency[J]. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2012, XXXIX-B6:189-194.
3. Knops Z F, Maintz J B, Viergever M A, et al. Normalized mutual information based registration using k-means clustering and shading correction[J]. Medical Image Analysis, 2006, 10(3):432-439.
4. Ojansivu V, Heikkila J. Image Registration Using Blur-Invariant Phase Correlation[J]. IEEE Signal Processing Letters, 2007, 14(7):449-452.
5. Lucchese L, Leorin S, Cortelazzo G M. Estimation of Two-Dimensional Affine Transformations Through Polar Curve Matching and Its Application to Image Mosaicking and Remote-Sensing Data Registration[J]. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2006, 15(10):3008-19.
6. Zitová B, Flusser J. Image registration methods: a survey[J]. Image & Vision Computing, 2003, 21(11):977-1000.
7. Lowe D G. Distinctive image features from scale-invariant keypoints [J]. International journal of computer vision, 2004, 60(2): 91-110.
8. Bay H, Tuytelaars T, Van Gool L. SURF: Speeded up robust features [M] // Computer Vision–ECCV 2006. Springer Berlin Heidelberg, 2006: 404-417.
9. Alcantarilla P F, Bartoli A, Davison A J. KAZE Features[M]// Computer Vision – ECCV 2012. Springer Berlin Heidelberg, 2012:214-227.
10. Achanta R, Hemami S, Estrada F, et al. Frequency-tuned salient region detection[C]// Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009:1597-1604.
11. Achanta R, Estrada F, Wils P, et al. Salient region detection and segmentation[C]// International Conference on Computer Vision Systems. Springer, Berlin, Heidelberg, 2008:66-75.
12. Lindeberg T. Scale-space theory: A basic tool for analyzing structures at different scales[J]. Journal of applied statistics, 1994, 21(1-2): 225-270.
13. Zhai Y, Shah M. Visual attention detection in video sequences using spatiotemporal cues[C]// ACM International Conference on Multimedia. ACM, 2006:815-824.
14. Itti L, Koch C, Niebur E. A Model of Saliency-Based Visual Attention for Rapid Scene Analysis[M]. IEEE Computer Society, 1998.
15. Hou X, Zhang L. Saliency Detection: A Spectral Residual Approach[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2007:1-8.