**Paper Title**

Author Name

Affiliation

Abstract

Template matching is the process of determining the presence and the location of a reference image or an object inside a scene image under analysis by a spatial cross-correlation process. Conventional cross-correlation type algorithms are computationally expensive. In this paper, an algorithm for a robust template matching method based on the combination of the wavelet transform method and SIFT is proposed. Discrete wavelet transform is done firstly on a reference image and a template image, and low frequency parts of them is extracted, then we use harris corner detection to detect the interesting point in low frequency parts of them to determined the matching candidate region of template image in reference image, extracting SIFT features on the matching candidate region and template image, The extracted SIFT features are matched by k-d tree and bidirectional matching strategy. Experiment show that, the algorithm can improve the accuracy of matching and at the same time to reduce the computation load.

**Keywords:** Template matching, SIFT, harris corner, k-d tree, bidirectional matching, wavelet transform.

# INTRODUCTION

Template matching is a classical problem in a scene analysis: given a reference image of an object, decide whether that object exists in a scene image under analysis, and find its location if it does. The template matching process involves cross-correlating the template with the scene image and computing a measure of similarity between them to determine the displacement [[1](http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6V14-4477X94-C&_user=8734945&_coverDate=01%2F31%2F2002&_rdoc=1&_fmt=high&_orig=search&_origin=search&_sort=d&_docanchor=&view=c&_searchStrId=1568438089&_rerunOrigin=google&_acct=C000056894&_version=1&_urlVersion=0&_userid=8734945&md5=b3c791e6ebaf0ba8b5de1e505356f3da&searchtype=a" \l "bib1#bib1)]. Since the evaluation of the correlation is computationally expensive, there has been a need for low-cost correlation algorithms for real-time processing. A large number of correlation-type algorithms have been proposed [[2](http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6V14-4477X94-C&_user=8734945&_coverDate=01%2F31%2F2002&_rdoc=1&_fmt=high&_orig=search&_origin=search&_sort=d&_docanchor=&view=c&_searchStrId=1568438089&_rerunOrigin=google&_acct=C000056894&_version=1&_urlVersion=0&_userid=8734945&md5=b3c791e6ebaf0ba8b5de1e505356f3da&searchtype=a" \l "bib2#bib2)]. One of the approaches is to use an image pyramid for both the template and the scene image, and to perform the registration by a top-down search [[3](http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6V14-4477X94-C&_user=8734945&_coverDate=01%2F31%2F2002&_rdoc=1&_fmt=high&_orig=search&_origin=search&_sort=d&_docanchor=&view=c&_searchStrId=1568438089&_rerunOrigin=google&_acct=C000056894&_version=1&_urlVersion=0&_userid=8734945&md5=b3c791e6ebaf0ba8b5de1e505356f3da&searchtype=a" \l "bib3#bib3)]. Other fast matching techniques use two pass algorithms; use a sub-template at a coarsely spaced grid in the first pass, and search for a better match in the neighborhood of the previously found positions in the second pass [[4.](http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6V14-4477X94-C&_user=8734945&_coverDate=01%2F31%2F2002&_rdoc=1&_fmt=high&_orig=search&_origin=search&_sort=d&_docanchor=&view=c&_searchStrId=1568438089&_rerunOrigin=google&_acct=C000056894&_version=1&_urlVersion=0&_userid=8734945&md5=b3c791e6ebaf0ba8b5de1e505356f3da&searchtype=a" \l "bbib4) A. Rosenfeld and A. Kak. Digital Image Processing (2nd Edition, Vol. 2 ed.),, Academic Press, Orlando (1982).[4](http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6V14-4477X94-C&_user=8734945&_coverDate=01%2F31%2F2002&_rdoc=1&_fmt=high&_orig=search&_origin=search&_sort=d&_docanchor=&view=c&_searchStrId=1568438089&_rerunOrigin=google&_acct=C000056894&_version=1&_urlVersion=0&_userid=8734945&md5=b3c791e6ebaf0ba8b5de1e505356f3da&searchtype=a#bib4#bib4)].Afterwards, Jane You[5] presented a wavelet based high performance hierarchical scheme for image matching which includes dynamic detection of interesting points, adaptive thresholding selection and a guided searching strategy for best matching from coarse level to fine level.

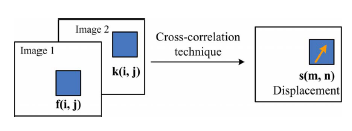
In order to improve the accuracy of matching and at the same time to reduce the computation load, In this paper, we proposed a robust image matching approach which decreases a large amount of unnecessary searches in contrast to conventional scheme and can achieve a better matching accuracy. Discrete wavelet transform is done firstly on a reference image and a scene image, and low frequency parts of them is extracted, then we use harris corner detection to detect the interesting point in low frequency parts of them to determined the matching candidate region of scene image in reference image, SIFT is used to extracting feature on the matching candidate region and scene image, The extracted features are matched by k-d tree and bidirectional matching strategy to enhance the accuracy of matching. Experiment show that, the algorithm can improve the accuracy of matching and at the same time to reduce the computation load.

# wavelet transform and harris corner detection

## Cross-correlation Algorithm

The cross-correlation technique is a method to get the displacement information of two consecutive images by comparing the similarity of a pair of image signals [6,7,8]. The traditional cross-correlation method was showed in Fig.1 and the cross-correlation coefficient was defined as formula





1. Cross-correlation method for image matching

Where  and  are the gray value of the template window and the inquiring window, both of the windows with a size .

The aimed area is the random vector which has the maximum cross-correlation coefficient . This method is univocal, easy implementing and accurate. But the large amount of calculation needs a long running time for the computer. It is difficult to realize measurement online.

## Wavelet Analysis

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

We can learn the 2-D wavelet analysis that the low frequency  is decomposed through wavelet decomposition every time decomposed into .Therefore, through db wavelet transform, the size of every part image will be reduced to 1/4 size of the original image.

In order to get an effective and accurate matching image, the size of the original image should satisfy the formula (2).





Where L is the maximum times of decomposition, the image size is .

## Harris Corner Detection

Harris corner detection algorithm is realized by calculating each pixel’s gradient [11]. If the absolute gradient values in two directions are both great, then judge the pixel as a corner. Harris corner detector is defined as followed [12]:









Where is an empirical value;  is a Gaussian function; and  are the partial derivatives of the gray scale in direction and at point , and  is the second-order mixed partial derivative;  and are the first-order directional differentials, which can be approximately calculated by convolving the gray scale and difference operators in direction and ;refers to convolution. Gaussian function is used to reduce the impact of noise, because first-order directional differentials are sensitive to noise. If exceeds certain threshold, then take the point as a corner.

# Fine Matching based on SIFT

## SIFT Detection

### 3.1.1 Keypoint extraction

The first stage of keypoint detection is to identify locations and scales that can be repeatably assigned under differing views of the same object. Detecting locations that are invariant to scale change of the image can be accomplished by searching for stable features across all possible scales, using a continuous function of scale known as scale space. It has been shown by Koenderink [14] and Lindeberg[15] that under a variety of reasonable assumptions the only possible scale-space kernel is the Gaussian function. Therefore, the scale space of an image is defined as a function, *L*(*x*, *y*, *σ*), that is produced from the convolution of a variable-scale Gaussian, *G*(*x*, *y*, *σ*), with an input image,*I*(*x*, *y*)

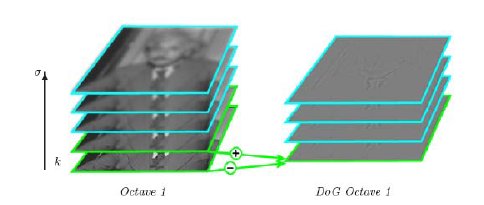
 . (7)

Where \* is the convolution operation in *x* and *y*, and



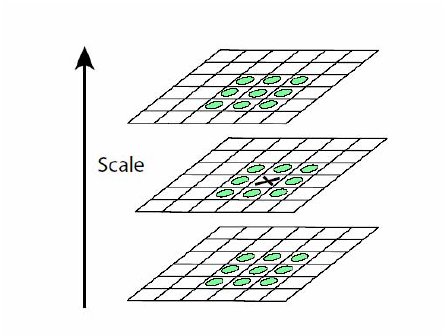
To efficiently detect stable keypoint locations in scale space, David G. Low proposed using scale-space extrema in the difference-of-Gaussian function convolved with the image





1. Gaussian scale-space pyramid create an interval in the difference-of-Gaussian pyramid.

In order to detect the local maxima and minima of *G*(*x*, *y*, *σ*), each sample point is compared to its eight neighbors in the current image and nine neighbors in the scale above and below (see Fig.3).



1. Maxima and minima detection in the difference-of-Gaussian image.

It is selected only if it is larger than all of these neighbors or smaller than all of them. To improve the stability of matching, we must reject the points that have low contrast(and are therefore sensitive to noise) or are poorly localized along an edge. The Taylor expansion of the scale-space function, *D*(*x*, *y*, *σ*), shifted so that the origin is at the sample point is



Where *D* and its derivatives are evaluated at the sample point and *X* = (*x*, *y*, *σ*)T is the offset from this point. The location of the extremum,, is determined by taking the derivative of this function with respect to *X* and setting it to zero, giving

 

The function value at the extremum is

. 

The value is useful for rejecting unstable extrema with low contrast. Usually the extema with a value of less than 0.03 should be discarded. The difference-of-Gaussian function will have a strong response along edges, even is the location along the edge is poorly determined and therefore unstable to small amounts of noise.

A 2×2 Hessian matrix computed at the location and scale of the keypoint is



Let α be the eigenvalue with the largest magnitude and β be the smaller one.



Let *α* = *rβ*, then



This quantity is at a minimum when the two eigenvalues are equal and it increases with *r*. David G. Low proposed using *r* = 10.

### 3.1.2 Descriptorrepresentation

The previous operations have assigned an image location, scale, and orientation to each keypoint. A keypoint descriptor is created by first computing the gradient magnitude and orientation at each image sample point in a region around the keypoint location. These samples are then accumulated into orientation histograms summarizing the contents over 4×4 subregions. David G. Low proposed using a 4×4 array of histograms with 8 orientation bins in each. Therefore, the feature vector has 4×4×8 = 128 elements. The descriptor is formed from a vector containing the values of all the orientation histogram entries. Finally, the vector is normalized to unit length to reduce the effects of illumination change.

## Feature matching

### 3.2.1 Matching By k-d tree

In this step, the correspondence between feature points, P, in the reference image and feature points, *P*, in the input image will be evaluated. The best candidate match for each feature point is found by identifying its nearest neighbor in the data set of feature vectors from input image. The nearest neighbor is defined as the feature point with minimum Euclidean distance for the invariant descriptor vector. However, the descriptor has a high dimension, matching the feature point by comparing the feature vector one by one will have high complexity in *o*(*n*2) time. This can be done in *o*(*n*log*n*) time by using a k-d tree [9] to find all matching pairs.

### 3.2.2 Matching By Bidirection

Many approaches have been proposed to match corresponding features between two images. Lowe proposed the using of the ratio between the Euclidean distance to the nearest and the second nearest neighbors [17].However, the method only match feature points from a single direction. The error rate of matching is still high. One-to-many matching are often appeared in results. Literature [18] proposed an improved matching strategy by matching feature points from bidirectional directions, which eliminates one-to-many matching and improve the accuracy of matching. We use the method of [18] to match the points.

# conclusion

In this paper, an algorithm for a robust template matching method based on the combination of the wavelet transform method and SIFT is proposed. Discrete wavelet transform is done firstly on a reference image and a template image, and low frequency parts of them is extracted, then we use harris corner detection to detect the interesting point in low frequency parts of them to determined the matching candidate region of template image in reference image, extracting SIFT features on the matching candidate region and template image, The extracted SIFT features are matched by k-d tree and bidirectional matching strategy. Experiment show that, the algorithm can improve the accuracy of matching and at the same time to reduce the computation load.

An open problem is still precision, consistency and efficiency of the algorithm. Future work will be focused on the development of a sophisticated method of deriving an optimized high-precision matching result under the influence of noise.

References

1. J.K. Aggarwal, L.S. Davis and W.N. Martin,“Correspondence process in dynamic scene analysis,”Proc IEEE 695, 562-572, (1981)
2. J.P. Secilla and N. Garcia,“Template location in noisy pictures,”Signal Process*.* **14**, 347-361, (1987).
3. R.Y. Wong and E.L. Hall, “Sequential hierarchical scene matching.,”IEEE Trans. Comput. **27** , 359-366,(1978)
4. A. Rosenfeld and A. Kak. Digital Image Processing (2nd Edition, Vol. 2 ed.),Academic Press, Orlando (1982)
5. Jane You and Prabir Bhattacharya, “A Wavelet-Based Coarse-to-Fine Image Matching Scheme in A Parallel Virtual Machine Environment, ”IEEE Trans. Image Proc*.* 9, 1547-1559, (2000).

Authors’ background  
\*This form helps us to understand your paper better, the form itself will not be published. You can delete it when you prepare the final paper if it’s accepted.

|  |  |  |  |
| --- | --- | --- | --- |
| Your Name | Title\* | Research Field | Personal website |
|  |  |  |  |
|  |  |  |  |

\*Title can be chosen from: master student, Phd candidate, assistant professor, lecture, senior lecture, associate professor, full professor