doi: 10.3788/gzxb20124108.0903

Sequence Image Mosaic Based on Approximate Scale Invariant Feature Transform Descriptors

CHEN Ai-hua, YANG Ben-quan, ZHANG Shi-qing
(*Taizhou University, Taizhou, Zhejiang* 318000, *China*)

Abstract: Through analyzing SIFT process, an approximate SIFT algorithm was proposed which changes framework of SIFT algorithm and SIFT descriptors are regarded as a kind of special Harris corners to deal with. It reduces computation time and retains advantage of SIFT algorithm. In addition, in order to enhance the accuracy of matching, a new method to refine matching couples was proposed. Experiments show that the approximate SIFT algorithm reduces computation time greatly and it does not affect the accuracy of matching. After refining matching couples, the accuracy of matching enhances greatly. The final panoramic image transits gently and has no evidence boundaries in overlapped areas.

Key words: Scale Invariant Feature Transform (SIFT) descriptors; Image matching; Extracting matching couples; Image mosaic

CLCN: TP391.41 **Document Code**: A **Article ID**: 1004-4213(2012)08-0903-6

0 Introduction

Panoramic mosaic based on sequence images is a fundamental problem in computer vision. It is widely used in many areas, such as computer vision, computer graphics, remote sensing and martial reconnaissance[1]. The general methods of panoramic mosaic include two main steps[2]. First step is to detect stable features in sequence images and match the images to get the transformation parameters. Second step is to blend the intensities between the overlapping parts in order that the overlapping image contains no obstructive boundaries. In sequence image mosaic, the first stage plays a key role and its accuracy has great influence on the final results.

In recent years, a kind of feature called SIFT attracts much attention because it is invariant to rotation, translation and scale variation and partially invariant to affine distortion, illumination variance and noise. A detailed SIFT method is presented in 2004[3]. As its high performances in feature descriptor, it is used in wide field. The SIFT scale and orientation constraints are employed for matching stereo images in Ref. [4-5]. A modified version of the SIFT approach is proposed and used to solve the robot localization problem in Ref. [6]. SIFT applied to panoramic images has been shown to give good results in indoor environments[7] and also to some extent in outdoor environments[8]. In relate work Ref. [9], CSIFT as a colored local invariant feature descriptor by using a structure similar to that of the SIFT descriptors and combining color as well as geometrical information into feature description is introduced.

As SIFT algorithm is so many calculating works, up to the present, many researchers focus on how to simplify SIFT process. In the process of SIFT features detection, the stages scale-space peak selection and key points descriptor are most consumable time. It is difficult to change scale-space peak selection. So the main idea is to simplify SIFT description, i. e. to reduce dimension of SIFT vector and speed up this algorithm. For instance, PCA-SIFT concatenates the first order x and y image derivatives of every sub-region, and for reducing the feature vector dimension is performed a PCA data selection[10]. The main objective of PCA-SIFT is to keep the SIFT matching properties, reducing the descriptor size. Simplified SIFT is proposed to reduce

Foundation item: The Natural Science Foundation of Zhejiang Province (No. Y1111058)
First author: CHEN Ai-hua (1981－), female, Ph. D. degree, lecturer, mainly focuses on digital image acquisition and processing, machine vision. Email: chen_1216@163.com
Corresponding author: YANG Ben-quan (1980－), male, M. S. degree, experimentalist, mainly focuses on digital image acquisition and processing, machine vision. Email: yangbq266@sina.com
Received date: 2012-01-09 **Revised date**: 2012-04-23

descriptor size in Ref. [11] which applies local circle image region and build only 12-element vector for each key point. The Gradient location-orientation histogram (GLOH) is an extension of SIFT that computes the histogram using a log-polar spatial grid and reduces the descriptor size using PCA[12]. In addition, promising approach is the Speeded Up Robust Features (SURF) by Bay et al.[13], which has been shown to yield comparable or better results to SIFT while having a fraction of the computational cost[14-15].

In this paper, an approximate SIFT algorithm is proposed and is used in sequence image mosaic. The remainder of this paper is organized as follows. Section 1 reviews the relevant aspects of traditional SIFT algorithm. Section 2 details our proposed approximate SIFT algorithm. Section 3 introduces a simple method to get rid of mismatch feature couples. The experimental analysis of the proposed approach and the conclusions are given in Sections 4 and 5.

1 Review of SIFT algorithm

SIFT, as described in Ref. [3,16], consists of four major stages: 1) scale-space peak selection; 2) key point localization; 3) orientation assignment; 4) feature point descriptor. The detail process can be expressed as follows.

Stage: 1 Finding the extreme points in the scale space. Firstly, using Gaussian linear transformation of the original image gets a set of images with different scales, then search for the extreme points in scale space $\{L(x,y,\sigma)\}$. The extreme points are a point set obtained by applying different Gaussian linear transforms to the original image $I(x, y)$ with a scale factor σ. The Gaussian scale transform operator is the difference of Gaussian (DoG) and the scale space is Gaussian difference scale space $\{D(x,y,\sigma)\}$. The space can be formulated as follows.

Gaussian scale space

$$L(x,y,\sigma)=G(x,y,\sigma)I(x,y) \tag{1}$$

Gaussian difference scale space

$$D(x,y,\sigma)=G(x,y,k\sigma)-G(x,y,\sigma)I(x,y)=L(x,y,k\sigma)-L(x,y,\sigma)$$

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

Each point in the scale space is compared to its 26 adjacent points (8 points in the same space and the other $9\times2=18$ points in its neighbor scale space) and points that have the maximum or minimum value are treated as extreme points, which are invariant to scale changes.

Stage 2: Feature points localization. It includes two parts: accurate key point localization and extracting key points. Accurate key point localization uses the Taylor expansion (up to the quadratic terms) of the scale-space function, $\hat{x}=(x, y, \sigma)^{\mathrm{T}}$, shifted so that the origin is at the sample point

$$D(x,y,\sigma)=D(m)=D+\frac{\partial D^{\mathrm{T}}}{\partial m}m+\frac{1}{2}m^{\mathrm{T}}\frac{\partial^2 D}{\partial m^2}m \tag{3}$$

where D and its derivatives are evaluated at the sample point and $x=(x,y,\sigma)^{\mathrm{T}}$ is the offset from this point. The location of the extremum, $\hat{x}$, is determined by taking the derivative of this function with respect to x and setting it to zero, giving

$$\hat{x}=-\left(\frac{\partial^2 D}{\partial x^2}\right)^{-1}\frac{\partial D}{\partial x} \tag{4}$$

Extreme points which are unstable and sensitive to noise are filtered out to leave extreme points that are treated as key points. Unstable points that are sensitive to noise usually have small values and are easily dominated by noise, so may not be detected. The function value at the extremum, $D(\hat{x})$, is useful for rejecting unstable extrema with low contrast. Given a threshold t, all extrema with a value of $|D(\hat{x})|$ less than t are discarded.

$$D(\hat{x})=D+\frac{1}{2}\frac{\partial D^{\mathrm{T}}}{\partial x}\hat{x} \tag{5}$$

The difference-of-Gaussian function will have a strong response along edges, even if the location along the edge is poorly determined and therefore unstable to small amounts of noise. 2×2 Hessian matrix can be used to get rid of this kind of unstable points. The process can be formulated as follows

$$\boldsymbol{H}=\begin{bmatrix}D_{xx} & D_{xy}\\ D_{xy} & D_{yy}\end{bmatrix}$$

$$D_{xy}=0.25((D[x+1,y+1,\sigma]-D[x+1,y-1,\sigma])-(D[x-1,y+1,\sigma]-D[x-1,y-1,\sigma]))$$

$$D_{xx}=D[x+1,y,\sigma]+D[x-1,y,\sigma]-2.0D[x,y,\sigma]$$

$$D_{yy}=D[x,y+1,\sigma]+D[x,y-1,\sigma]-2.0D[x,y,\sigma] \tag{6}$$

Let α be the eigenvalue with the largest magnitude and β be the smaller one. Then, the sum of the eigenvalues from the trace of $\boldsymbol{H}$ can be computed and their product from the determinant

$$\mathrm{Tr}(H)=D_{xx}+D_{yy}=\alpha+\beta$$

$$\mathrm{Det}(H)=D_{xx}D_{yy}-D_{xy}^2=\alpha\beta \tag{7}$$

Given $\alpha = r\beta$, then check the value of below formula to decide whether is filter out

$$\frac{\mathrm{Tr}(H)}{\mathrm{Det}(H)} < \frac{(r+1)^2}{r} \tag{8}$$

Stage 3: Quantizing the description to the key points. Lowe[3] formulated this assignment using the norm and angle in Euclidian space, with the direction of key points used as normalized the gradient direction of the key point operator in the following step. Identical directions can be extracted after an image rotation.

Stage 4: The final stage builds a local image descriptor for each feature point, based upon the image gradients in its local neighbourhood. The standard feature point descriptor used by SIFT is created by sampling the magnitudes and orientations of the image gradient in the patch around the feature point and building smoothed orientation histograms to capture the important aspects of the patch. A 4×4 array of histograms, each with 8 orientation bins, captures the rough spatial structure of the patch. This 128-element vector is then normalized to unit length and thresholded to remove elements with small values.

2 Approximate SIFT algorithm

By analyzing process of SIFT algorithm, there is redundancy in the process of detecting extreme points in the scale space, because only a few points are extreme points and the scale space has huge points. If only part points participate in this process, the computation time can be reduced greatly. So the approximate SIFT algorithm is proposed.

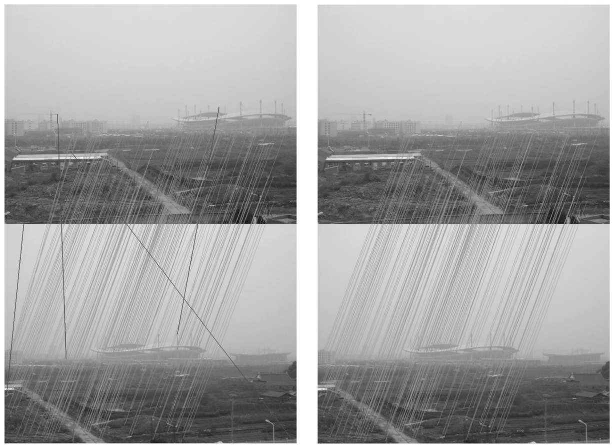
Comparison SIFT algorithm and Harris algorithm, a phenomenon that the second stage of SIFT algorithm is similar to Harris algorithm can be found easily. SIFT descriptors are extension of Harris corner descriptors. So SIFT key points can be regarded as a special kind of Harris corner which is invariant to image scale and location, and is robust to affine transformations and changes in illumination. Then a new framework for feature descriptor based on SIFT is proposed. Compared with original SIFT algorithm, the main change is first two stages. The detail process can be described as below.

Stage 1: Detecting Harris corners in original image. According to process of Harris algorithm, detect corners to set original key points set.

Stage 2: Building a difference of Gaussian (DoG) function pyramid. According to image scale, build key point set $\{x_i^{(j)}\}$ in each octave, where $i=1,2,\cdots,N_j$ stands for the number of key points in every octave, $j=1,2,\cdots,M$ stands for the number of octaves. The building key point set above is regarded as stable points. In stead of search extreme points in the scale space, only the points in the building key point set are detected whether are extreme points. In order to avoid losing stable key points, when building key points set in scale space, the points of nearby key point set should be included. The next process stages are the same as original stages, the detail process dose not mention again.

3 Extracting matching couples

Match results of two images is shown in Fig. 1 (a). There are often many mismatch feature couples as is shown. Sometimes they may have badly influence for the following images process, so matching couples have to be refined. Fig. 1 (a), matching couples are connected by lines.



(a) Result of rough matching (b)Result of refined matching

Fig. 1 Results of features matching

Observing the character of these lines, we find lines connecting correct matching couples are parallel approximately. The slope differences in correct matching couples are very small. But the slope differences between correct matching couples and incorrect matching couples become very large. So a simple method based on slope invariant is applied to detect the incorrect matching feature points. The process of refining matched features set is described as following steps:

Step 1: Compute the slopes of feature point couples;

Step 2: Find the standard slope k_d which is average value of correct matching feature points set;

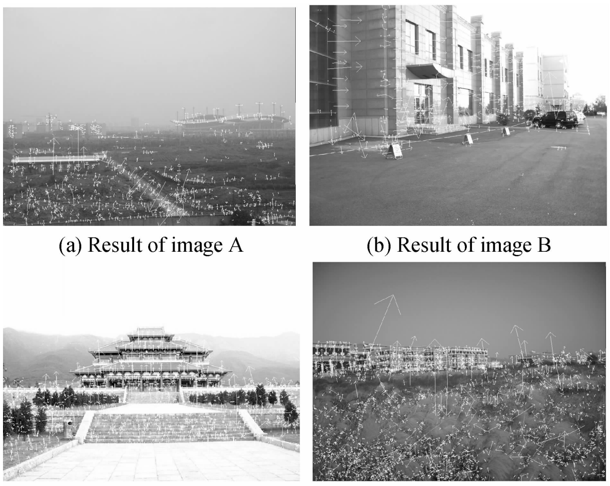
Step 3: Suppose k_i is the ith slope, the distance between k_d and k_i is $d=|k_i-k_d|$. Given a threshold ε, if d is below ε, reserves the feature point couple, otherwise discards it.

This method is very simple to refine key point couples. The matching result after refined is shown in Fig. 1 (b). Compared with Fig. 1 (a), incorrect matching couples reduce greatly.

4 Experimental results

In this section, experiments are carried out to verify the efficiency and accuracy of proposed method. The algorithm is tested using VC. NET programs on an AMD II X2 250, 1. 81 GHz machine.

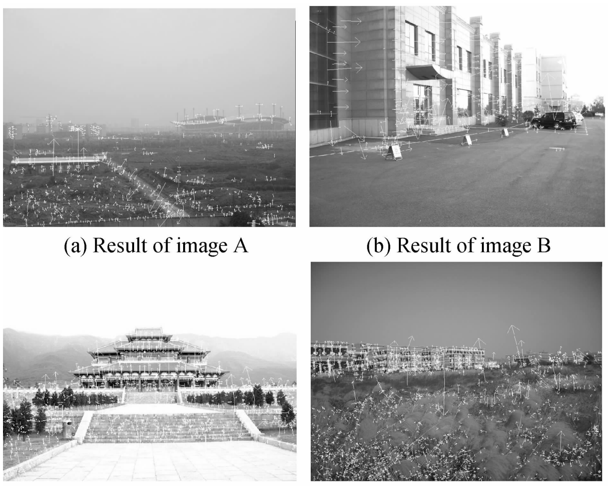
In order to prove the efficiency of proposed approximate SIFT algorithm, four images are selected to show the process result. Fig. 2 shows the results of traditional SIFT algorithm. The results of proposed approximate SIFT are shown in Fig. 3. The parameter comparisons between



(a) Result of image A (b) Result of image B

(c) Result of image C (d) Result of image D

Fig. 2 Results of traditional SIFT



(a) Result of image A (b) Result of image B

(c) Result of image C (d) Result of image D

Fig. 3 Results of proposed approximate SIFT

Table 1 Comparisons of computation time (ms)

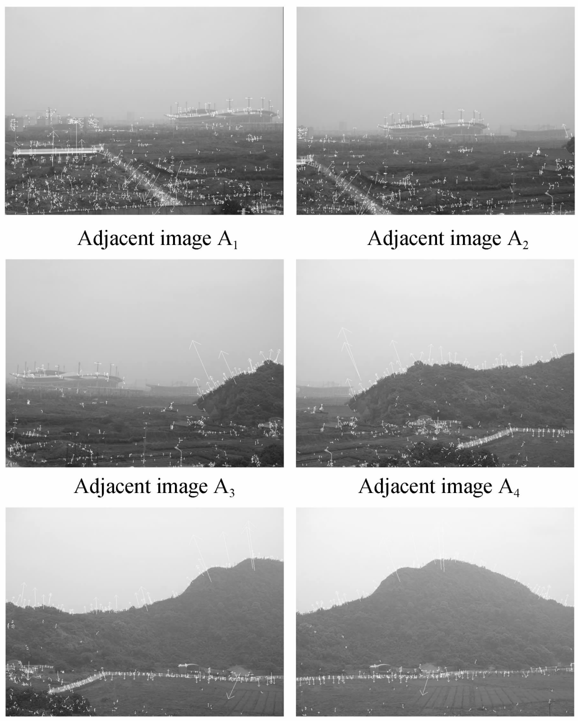
Experiment images	A	B	C	D
Approximate SIFT	1 026	1 398	1 182	1 694
Traditional SIFT	2 531	3 075	2 645	3 887

Table 2 Comparisons of number of key points

Experiment images	A	B	C	D
Approximate SIFT	695	1 018	1 644	2 869
Traditional SIFT	756	1 122	1 820	3 151

traditional SIFT and proposed approximate SIFT are shown in Table 1 and Table 2. The results show that the proposed approximate SIFT algorithm reduces computation time greatly. Now many methods are proposed to reduce dimension of SIFT vector . They can combine with proposed method in this paper, and then the computation time will reduce more. From Table 2, the drawback compared with traditional SIFT is shown. That is the number of key points reduces a little. Adjusting parameters of Harris algorithm can improve this condition. As SIFT algorithm generates large numbers of features, too many features are not necessary in image mosaic. So the number of features reduces a little does not influence result of image mosaic.

There are six adjacent images elected to carry out image mosaic. The size of sequence images is 912×684. Fig. 4 (a) shows results of proposed approximate SIFT algorithm. The final panoramic mosaic image is shown in Fig. 4 (b). It is clear that



Adjacent image A_1 Adjacent image A_2

Adjacent image A_3 Adjacent image A_4

Adjacent image A_5 Adjacent image A_6

(a) Results of proposed approximate SIFT

(b) Results of panoramic image

Fig. 4 Results of emulational experiments

there is no evident discontinuity in panoramic image and the proposed algorithm performs a perfect result.

5 Conclusions

This paper has studied sequence image mosaic technology based on SIFT descriptors. In order to speed up SIFT algorithm, an approximate SIFT algorithm which regards SIFT descriptors as a special kind of Harris corners is proposed. Experiments show that proposed approximate SIFT algorithm reduces computation time greatly. After feature points matching, there are often have incorrect matching couples, in order to improve the matching accuracy, a new method to refine matching couples is proposed. Experiments show that proposed approximate SIFT algorithm reduces computation time greatly and no influence on following process. Proposed extracting matching couple method filters out most of mismatch feature couples. The final panoramic image transits gently and has no evidence discontinuity in overlapped areas.

References

[1] FIALA M. Structure from motion using SIFT features and the PH transform with panoramic imagery [C]. Canadian Conference on Computer and Robot Vision(CRV), Ottawa, Canada, 2005: 506-513.

[2] SU Ming-shing, HWANG Wen-liang, CHENG Kuo-young. Analysis on multire solution mosaic images[J]. *Transactions on Image Processing*, 2004, **13**(7): 952-959.

[3] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. *International Journal of Computer Vision*, 2004, **60**(2): 91-110.

[4] SE S, LOWE D, LITTLE J. Vision-based mobile robot localization and mapping using scale-invariant features[C]. Proceedings of the IEEE International Conference on Robotics and Automation(ICRA), Seoul, Korea, 2001: 2051-2058.

[5] SE S, LOWE D, LITTLE J. Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks[J]. *International Journal of Robotics Research*, 2002, **21**(8): 735-758.

[6] ANDREASSON H, TREPTOW A, DUCKETT T. Localization for mobile robots using panoramic vision, local features and particle filter[C]. IEEE International Conference on Robotics and Automation(ICRA), Barcelona, Spain, 2005: 3348-3353.

[7] ANDREASSON H, DUCKETT T. Topological localization for mobile robots using omni-directional vision and local features [C]. the 5th IFAC Symposium on Intelligent Autonomous Vehicles, Lisbon, Portugal, 2004: 3348-3353.

[8] VALGREN C, DUCKETT T, LILIENTHAL A. Incremental spectral clustering and its application to topological mapping [C]. Proceedings of the IEEE International Conference on Robotics and Automation(ICRA), Rome, Italy, 2007: 4283-4288.

[9] ABDEL-HAKIM A E, FARAG A A. CSIFT: a SIFT descriptor with color invariant characteristics[C]. Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), New York, NY, USA, 2006: 1978-1983.

[10] KE Y, SUTHANKAR R. PCA-SIFT: a more distinctive representation for local image descriptors[C]. Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), Washington, DC, USA, 2004: 511 - 517.

[11] LIU Li, PENG Fu-yuan, TIAN Yan, *et al*. Fast image matching for localization in deep-sea based on the simplified SIFT (scale invariant feature transform) algorithm [C]. *SPIE*, 2007, **6795**: 67953A.

[12] MIKOLAJCZYK K, SCHMID C. A performance evaluation of local descriptors [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, **27**(10): 1615-1630.

[13] BAY H, ESS A, TUYTELAARS T, *et al*. SURF: speeded up robust features [J]. *Computer Vision and Image Understanding*, 2008, **110**(3): 346-359.

[14] BAY H, FASEL B, GOOL L. V. Interactive museum guide: Fast and robustrecognition of museum objects [C]. Proceedings of International Workshop on Mobile Vision, Graz, Austria, 2006.

[15] MURILLO A C, GUERRERO J J, SAGUES C. SURF features for efficient robot localization with omnidirectional images [C]. Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Rome, Italy, 2007: 3901-3907.

[16] DENG Cheng, GAO Xin-bo. Geometrically robust image watermarking based on SIFT feature regions [J]. *Acta Photonica Sinica*, 2009, **38**(4): 1005-1010.

基于近似尺度不变特征转换的序列图像融合算法

陈爱华，杨本全，张石清

（台州学院，浙江 台州 318000）

摘　要：在分析了尺度不变特征转换算法特点的基础上提出了一种近似尺度不变特征转换算法，该算法改变了传统尺度不变特征转换算法的框架，将尺度不变特征转换描述符看作是一种特殊的 Harris 算子，该算法既保留了尺度不变特征转换描述符的优点又降低了计算量. 另外，为了提高特征匹配的精度，给出了一种新的特征点对提纯算法. 实验结果表明，近似尺度不变特征转换算法在不影响匹配性能的前提下大大缩短了处理时间，经过提纯后的特征点对匹配性能显著提高，最终得到的全景图像过渡平稳，重叠区域没有显著的痕迹.

关键词：尺度不变特征转换描述符；图像匹配；特征点对提纯；图像融合