Camera Image Mosaicing Based on an Optimized SURF Algorithm

Nan Geng*1,2, Dongjian He1, Yanshuang Song2
1College of Mechanical and Electronic Engineering, Northwest A&F University
2College of Information Engineering, Northwest A&F University, Yangling, China, 86-29-87091641
e-mail: {nangeng,hdj168}@nwsuaf.edu.cn*, sysshuang@yeah.net

Abstract

For real-time and robust web camera image mosaicing, a method based on optimized Speeded up Robust Features (SURF) was proposed in this paper. Firstly, the feature points from the overlapping parts between the reference image and the target image are extracted by employing the rapid matching algorithm - Best-Bin-Fist (BBF). Then Random Sample Consensus (RANSAC) is used for mismatched features eliminating and projection matrix calculating to resample the target image and achieve the calibrated one. Finally, registered images are fused with evolutional fusion algorithm to produce image mosaicing. The resultsdemonstrate that our image mosaicing method can stitch the images captured with the web cameras with noise and different lighting effectively and efficiently while also being robust.

Keywords: Image Mosaicing, SURF, BBF, RANSAC, Web Camera

Copyright © 2012 Universitas Ahmad Dahlan. All rights reserved.

1. Introduction

Image mosaicing is the key technique of generating a panorama, which is used to create high-resolution map automatically. It is an important research branch of the field of digital image processing. It is used to match and alignment a group of overlapping image sequences and re-register after fusion, thereby produce a new image of information of the image sequences with wide viewing angle. The image mosaicing technology can enhance image resolution and reduce redundant information, thus, represent information more effectively.

Image mosaicing is the major research field in computer vision, computer graphics and vision processing [1-5], which can be used in environmental inspection and vision monitoring, mosaic large-scale scene panorama through satellite images or aerial photography for the thorough surveillance of the whole area [6, 7], such as the river drainage area, the area of the farming land and real-time crop pest.

The key technique for image mosaicing is image matching. There are two major methods according to image matching way [8]: region-based matching and feature-based matching. Region-based matching makes use of the relevance between pixels of different image regions to match [9] which have disadvantages of high calculation complexity and sensitivity to image grayscale. The feature-based matching mainly uses the points, lines and contours of images for matching [10], thereby overcomes those disadvantages mentioned above and has the advantages such as quickness and accuracy, etc.

Since the feature-based matching compresses the amount of information of images, it displays more robustness to grayscale variation, thus more convenient and reliable. The coordinates of feature points can be used directly to estimate the parameter of transformation model between the images, the major concern in the related researches at home and abroad. Currently, Harris angle point [11], SIFT feature point [12] and SURF feature points [13] are usually used for image matching. Harris presents more sensitivity to angle variation and noise, SIFT shows higher robustness to the signals' rotation, horizontal moving, noise, lighting transformation, affine and projective transformation, but requires longer processing time, difficult to realize real-time matching. The improved PCA-SIFT can reduce data dimension of SIFT, while decreases the speed due to the increasing calculation load in the meantime [14].

SURF is an improved matching algorithm proposed on the basis of SIFT, similar to SIFT in function, but obviously faster than SIFT. The paper compared SIFT and SURF to find the
SURF’s advantages over other methods directly and improved SURF accordingly, then achieved approximate matching based on the nearest neighboring algorithm. RANSAC was used to eliminate mismatching for estimation of transformation matrix; finally function verification analysis was undertaken based on the improved SURF feature point matching algorithm.

2. Methods
2.1 Key Techniques

The overall structure of the system is shown in Figure 1. Web camera is YW7600 Web intelligent even speed globe manufactured by Shenzhen YiweiRuichuang Technology Limited Company, China configured with high quality data signal processing (DSP), inside-fixed cradle head and data decoding device, which can fast locate and constantly trace and scan, realize ominous and non-blind spot surveillance, adjust automatically to changes in illumination and distance; the inside-fixed cradle head adopts delicate step motor for driving power, running steadily with sensitive response and accurate positioning. The captured video accomplishes image mosaicing in the server side, thereby a panorama is constructed.

Panorama mosaic processing is shown in Figure 2. With image sequence captured by web camera rotating at even speed horizontally, the mosaic image $M_i$ ($M_0$ is the first key frame extracted, i.e. $M_0 = I_0$) is extracted and the SURF feature points of real time captured key frame, match approximately; eliminate mismatching and estimate transformation matrix with RANSAC, fuse images to get mosaic $M_i$ when the number of needed frames for panoramic mosaic is reached, the panoramic cylinder is displayed. With constant reading of key frames, the key frames are mosaic with the already-mosaic ones, renew the mosaic real-time and finally obtain panorama.
SURF algorithm

SURF was proposed by Herbert Bay in 2006. The algorithm consists of two aspects: feature point testing and feature point description. The features extracted by this way present fixed size and fixed rotation, also show partly invariance in lighting transformation, affinities and projective transformation. SURF surpassed or similar to the previous methods of the same kind in repetition uniqueness and robustness and showed obvious advantage in calculating speed. The process is as follows, shown in Figure 3.

For a point $X$ of the image $I$, the coordinate is $(x, y)$, the point’s value $I_x(X)$ is sum of point $X$ and the origin of the rectangular region formed by all points of the pixel values, as shown in (1).

$$I_x(X) = \sum_{i=0}^{(x)} \sum_{j=0}^{(y)} I(i, j)$$  \hspace{1cm} (1)

Pixels’ sum within the rectangular area in Figure 4 thus can be obtained by (2).

$$\sum = A-B-C + D$$  \hspace{1cm} (2)

Rectangle four vertices $A$, $B$, $C$, $D$, the corresponding integral image, pixels’ sum within the rectangular area can be calculated by only 3 times of adding and subtraction, and has nothing to do with the size of the rectangular area. SURF algorithm using integral images of this nature can significantly increase the calculation speed of the box-type filter.

The scale space of the SIFT algorithm according to scale the image layer by layer sampling, then the corresponding scale Gaussian filter was used to generate an image space convolution pyramid. SURF algorithm makes use of the box-type filter to build the image pyramid, there is no need to drill down-sampling the image not changing the size of image and the size of the box-type filter; different size box-type filter with the original image volume product is applied to generate the image pyramid.

In SURF algorithm, the scale space is constructed from the bottom of the box-type filter (corresponding to the beginning of the second-order Gaussian filter $\sigma = 1.2$), shown in Figure 5, when the image of the smallest scale was formed, the box-type filter grew gradually to 6 pixels, the pixel values of the convolution with the original image formed the image pyramid, shown in Figure 6.
The filter size is calculated by formula (3):

$$\text{Filter Size} = 3 \times (2^{\text{Octave}} \times \text{interval} + 1)$$

(3)

Here, Octaves is layers, interval is image sequence number within layer. The relationship between the size of a box-type filter and the Gaussian scale value is as follows:

$$\sigma = \frac{\text{Current Filter Size}}{\text{Base Filter Size}} = \frac{\text{Current Filter Size}}{\text{Base Filter Size}} \times \frac{1}{2^9}$$

Hessian matrix is the core of SURF algorithm; it performs quite well in calculating and accuracy. The second order Hessian matrix of Image is defined as follows:

$$H(x, \sigma) = \begin{pmatrix}
L_{xx}(x, \sigma) & L_{xy}(x, \sigma) \\
L_{yx}(x, \sigma) & L_{yy}(x, \sigma)
\end{pmatrix}$$

(4)

Here, $L_{xx}(x, \sigma)$ is the convolution of Gaussian second derivative $\frac{\partial^2}{\partial x^2} g(\sigma)$ and image $I$ in the X point. $L_{xx},$ $L_{yy}$ and so on. The maximum point of the determinant of Hessian matrix in the scale space and image space is determined as feature points. To quicken calculating, Bay.et proposed a square wave filter to approximately replace Gaussian second derivative using integral image to speed up convolution calculation. For a given image, approximate Hessian matrix $H_{approx}$ determinant is shown in (5).

$$\det(H_{approx}) = D_{xx} D_{yy} - (0.9 \times D_{xy})^2$$

(5)

Here, $D_{xx}$ is approximate value of convolution $\frac{\partial^2}{\partial x^2} g(\sigma)$ at point X between Gaussian second derivative and image $I$, the same for $D_{yy}, D_{xy}$.

Using feature point as the center, calculate Haar wavelet response in direction of X and Y inside the cycle of $6\sigma$ ($\sigma$ is the corresponding size of the feature point), then use Gaussian function centered on the feature points to weight and form a vector, take the longest vector as the leading direction of the feature point. Along the leading direction, select a square of $20\sigma$ neighboring region, divide it into $4 \times 4$ sub squares; calculate Haar wavelet response with Haar wavelet model ($2\sigma \times 2\sigma$) in each square; sum each square field $d_x, d_y, |d_x|, |d_y|$, get a 4-
dimensional \( \sqrt{\sum d_1^2, \sum d_2^2, \sum d_3^2, \sum d_4^2} \), the vectors of 4×4 sub squares link together and get a feature description sift for 64 dimension.

Use the fast Hessian matrix to get extreme points to compare with 27 points of the three-dimensional neighborhood, use the non-maxima suppression to select extreme points as the candidate point with record location and size. Different from the SIFT algorithm, the SURF algorithm can fix the threshold, only when the determinant \( \text{det}(H_{approx}) \) of the Hessian matrix is greater than the threshold, it can be used to determine whether the candidate extreme points can control the number of feature points, increase calculation speed. In order to make the detected feature points more stable, the precise positioning of the extreme points on the scale space and image space becomes sub-pixel level by interpolation.

### 2.1.2. Key frames selection

The research selected key frames by fixing frame interval method. The specific collection method is shown in Figure 7.

![Figure 7 Key frames by fixing frame interval](image)

According to the frame speed, time consumption for the cradle head rotating one round and view angle of the camera in vision sequence capture, the interval between frames were determined, as shown in (6).

\[
\text{interval} = \frac{s \times t}{360 / (\alpha \times \eta)}
\]  

(6)

Here,
- \( s \) = frame speed
- \( t \) = time spent for pzt rotating one round
- \( \alpha \) = view angle of video camera
- \( \eta \) = overlapping ratio between key frames

### 2.1.3. Feature points approximate matching

After obtaining all feature points sifts with SURF feature point inspection algorithm, take one of which as sample image, those well-matched as to-be matched images. Select suitable similarity exponent to measure in the matching. The European measure scale is applied in the research to calculate the distance between the two vectors, as shown in (7).

\[
d(X, Y) = \sqrt{\sum_{i=1}^{N} (x_i - y_i)^2}
\]  

(7)

Here,
- \( X=(x_1, x_2, x_3, \ldots, x_N) \), \( Y=(y_1, y_2, y_3, \ldots, y_N) \), \( N \) is dimension of characteristic quantity.

The nearest neighboring point feature can be used to accomplish two images matching. If the ratio between the nearest distance and the second nearest distance is less than 60%,
take the nearest feature point as matching point, reduce that field value, the matching number would decrease correspondingly but more steady.

The searching strategy employed in this research is BBF algorithm[15], for it is approximate searching, the eventually obtained matching result might exist redundancy and mismatching, RANSAC (Random Sample Consensus) is used to sieve out the approximately matching SURF feature points to ensure accuracy and stability of final matching.

2.1.4. Feature points purification and projection transformation matrix calculation

RANSAC shows better capability in eliminating mismatching points, usually used in feature points matching of the images. After matching the corresponding feature points in the two images, the geometric model between them can be defined according to their relevance. If projection transformation relevance between the two images is 8 parameter projection transformation matrixes H, then the points in the target image can be normalized one by one to the reference coordinates.

The basic steps for eliminating mismatching with RANSAC are the following:

(1)Initialization of point set S is void, the largest iterative number is \( N = 500 \) and iterative index is \( i = 0 \);

(2)When iterative index \( i < N \), select \( k(k = 4) \) matching feature points at random, use linear transformation (DLT) algorithm to project transformation matrix \( H_i \);

(3)Calculate the distance between points on the reference image and the feature points after transformation matrix \( H_i \) is:

\[
\text{Dis}(X', HX) = \sqrt{(x' - x_h)^2 + (y' - y_h)^2}
\]

(8)

(4)Define distance field value \( T \), add feature points with values less than field value \( T \) into the current inner points set \( S_i \);

(5) If \( S_i \) is larger than \( S \), then \( S = S_i \), \( H = H_i \), return to step (2), otherwise, enter step (2) directly.

(6) Recalculate projection transformation matrix \( H \) according to the calculated largest inner points set \( S \).

2.1.5. Image fusion

The goal of image fusion is to eliminate the splice joint in the mosaic, i.e., eliminate “ghosting and explosion difference” to obtain seamless mosaic in full sense. Fade-in-fade-out is the most frequently used fusion algorithm presently. The algorithm is simple in concept, involving less calculating, easily implemented in not obvious image transition region.

For the matching points \( I_1(x, y) \) and \( I_2(x, y) \) in the overlapping region, assign 2 weight coefficients to the 2 points, which are related to the position of the point in the overlapping region, as shown in Figure 8.

![Figure 8 Fade-in-fade-out scheme](image)

Fused image \( I(x, y) \), shown in (9).
\[
I(x, y) = \begin{cases} 
I_1(x, y) & (x, y) \in I_1 \\
W_1 I_1(x, y) + W_2 I_2(x, y) & (x, y) \in (I_1 \cap I_2) \\
I_2(x, y) & (x, y) \in I_2 
\end{cases}
\]

(9)

Here, the sum of weight coefficient \(W_1\) and \(W_2\) is shown in equation (10).

\[
W_1 = \frac{x_r - x}{x_r - x_L} \\
W_2 = \frac{x - x_L}{x_r - x_L}
\]

(10)

Here, \(x_r\) and \(x_L\) are the right and left boundary in the overlapping region of the image \(I_1(x, y)\) and \(I_2(x, y)\), \(W_1 + W_2 = 1\).

2.2. SURF Algorithm Enhancements

The estimation of transformation matrix in image matching comes merely from the matching point of the images; the quality of transformation matrix is related to the number and position of testing feature points in the images to be examined, thus the key factor to secure transformation matrix quality is even distribution of feature points.

The feature points extracted through SURF algorithm were obtained by first getting part external with Fast-Hessian matrix, then judge whether it is feature point or not with non-external within the 3x3x3 cubic neighboring area, the inspected feature points might exist in one region, thus results in clustering, in this case, which probably reflecting image features of few objects. In addition, extracting more feature points can increase feature description and matching calculation time. Given that the feature points for image matching must reflect the image features on the whole, in order to capture more even distributed feature points, the feature point testing region can be extended.

\[
\min_i \min_j \left| x_i - x_j \right|, x_i \in O
\]

(11)

Here, \(x_i\) - the position of feature point \(i\); \(O\) - the set subtracted by \(x_i\) from the part external with the same measure.
On the foundation of consistency, decrease the testing times for feature points, define a tag for each interest point when testing, through defining the tag's value; estimate approximately the position of external point around. If the tag is 1, its pixel is part external point; if the tag is 0, the pixel is non-external point, then adjust dynamically the iterative step according to the tag's value. If the pixel is part external point, then $4r^2+4r+18$ pixels around and 18 pixels in the neighboring are not external points, at the moment, define iterative step as $2r$; if the pixel is not part external point, the iterative step is 1, examine the pixels one by one. Adjust by this method, the testing times are effectively reduced, particularly for images with more part external points, the more tests for iterative step $2r$, the less testing times for the entire examination, the less time it may take.

2.3. Image Mosaicing

Image mosaicing includes such steps as image capturing, image pre-resolution, image matching and image fusion. The process is shown in

![Figure 10. The process flow of image stitching](image)

Image preprocessing includes noise elimination, image enhancing, grayscale transformation, after preprocessing, the improved SURF is used to extract feature points of grayscale image. Use BBF (Best-Bin-First) algorithm, find the nearest neighboring feature in the image sequence to realize feature points matching. Use RANSAC to eliminate mismatching and estimate transformation matrix, use fade-in-fade-out algorithm to accomplish image fusion.

In order to verify reliability and feasibility of the improved SURF algorithm, choose the two images changing in size and resolution with 400x320 pixel resolution and 800 x 640 pixel resolution from the image store developed by KristianMikolajczk, verified with SIFT and SURF, as shown in

![Figure 11.](image)

As found in the matching result, SIFT and SURF both overcome effect of size and viewport transformation and realize image matching.

![Figure 12.](image)

Figure 12 presents the results with SURF and the improved SURF. Altogether 430 feature points were found with SURF, 225 feature points with the improvedSURF with field value S=200. The result showed feature points obtained by the improved SURF distribute more evenly, image features were displayed more extensively.
Camera Image Mosaicing Based on an Optimized SURF Algorithm (Nan Geng)

(a) Distribution of feature points with SURF algorithm

(b) Distribution of feature points with the improved SURF algorithm

Figure 12 The comparisons of feature points distribution with SURF and the improved one.

Comparison of the three algorithms is shown in Table I, where I1 is the number of extracted feature points from 1st image, I2 is the number of extracted feature points extracted from 2nd image, T1 is time-consuming for detecting feature points from 1st image, T2 is time-consuming of detecting feature points from 2nd image, PP is the final matching feature points-pair and T3 is time-consuming of matching. As discovered, for images with variation in size, viewport and rotation, there were good matching with SIFT, SURF and the improved SURF. Though feature points were fewer with SURF, it didn’t affect the final matching effect, the matching time for processing is reduced by one quarter, compared to the time taken with the original algorithm.

Table 1. The performance comparisons of the three algorithms

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>I1</th>
<th>I2</th>
<th>T1/sec</th>
<th>T2/sec</th>
<th>PP</th>
<th>T3/sec</th>
<th>Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIFT</td>
<td>937</td>
<td>3855</td>
<td>1.675</td>
<td>4.036</td>
<td>106</td>
<td>0.313</td>
<td>SIFT</td>
</tr>
<tr>
<td>SURF</td>
<td>430</td>
<td>1385</td>
<td>0.123</td>
<td>0.408</td>
<td>41</td>
<td>0.124</td>
<td>SURF</td>
</tr>
<tr>
<td>Our</td>
<td>225</td>
<td>770</td>
<td>0.099</td>
<td>0.313</td>
<td>23</td>
<td>0.096</td>
<td>Our</td>
</tr>
</tbody>
</table>

3. Results and Analysis

Based on the improved SURF, the productive image sets taken from the two fields were tested and analyzed. The images from the first set transformed in illumination, the second set displayed fuzzy variation and similar texture. The two sets of images showed similar texture and variation in many aspects influenced by wind in the open air, which enhanced the algorithm’s ideal overall matching capability. The results showed the improved SURF algorithm can achieve vision streams matching captured by outdoors monitoring with lighting transformation, fuzzy variation and similar texture, as shown in Figure 13.

Cylinder projection is one unavoidable step in correct and more extensive image mosaicing; Figure 14 (a) showed the result tested with the improved SURF after cylinder projection, the mismatching was clearly shown in the picture, which would influence estimation accuracy of transformation matrix at the later stage.

Figure 14(b) showed the results after elimination of mismatching with RANSAC, apparently, mismatching was reduced significantly.

Through a large amount of experiments, the research defined field value S=200, the ratio of the nearest neighboring distance and the second nearest neighboring distance is 0.5, slope variation, $\xi=0.15$; the largest iteration number, 200; the searching field value for BBF, 200, the better results can be achieved.

Figure 15 shows part of panoramic mosaic result based on video using the algorithm proposed in the paper.
Figure 13. Robustness verification of the improved SURF algorithm

(a) The matching results with lighting transformation

(b) The matching results with fuzzy variation

Figure 14. The effects of feature points matching

(a) The results with the improved SURF

(b) The results with RANSAC after mismatching elimination

(a) The panoramic of Shanyang

(b) The panoramic of Baishui

(c) The panoramic of Qingjian

(d) The panoramic of Xixiang
TABLE II is the result analysis of 8 sets of vision stream image mosaicing in field production obtained by WY7600 web-camera in different remote vision capture system in experimental sites of Northwest Sci-tech University of Agriculture and Forestry.

Figure 15 displayed the part of mosaic. It was found the improved SURF algorithm can realize image matching with size and vision variation, also showed good performance in matching images with similar texture and displayed better real time performance.

Table 2. The performance comparisons of the three algorithms

<table>
<thead>
<tr>
<th>Index</th>
<th>Key frame (Numbs)</th>
<th>Size</th>
<th>Field of View (degree)</th>
<th>T(sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9</td>
<td>1280 x 720</td>
<td>360</td>
<td>23.6</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>1280 x 720</td>
<td>300</td>
<td>18.9</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>704 x 576</td>
<td>300</td>
<td>13.5</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>688 x 544</td>
<td>360</td>
<td>24.8</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>704 x 576</td>
<td>360</td>
<td>12.8</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>1280 x 720</td>
<td>360</td>
<td>25.1</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>336 x 272</td>
<td>300</td>
<td>8.6</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>1280 x 720</td>
<td>360</td>
<td>19.6</td>
</tr>
</tbody>
</table>

4. Conclusion

Through ensuring more even distribution of the non-external feature points, our image matching and mosaicing method based on an optimized SURF is more efficient with only one quarter of processing time compared to that with the original algorithms. And the basic parameters for RANSAC algorithm was defined for images matching and elimination based on the experimental results and analyses. Our method has been gracefully applied for vision stream image mosaicing in field production with the defined parameters in the experimental sites of Northwest Sci-tech University of Agriculture and Forestry. The results show that vision stream image mosaicing, based on the improved SURF features, is suitable for panoramic mosaicing of field production vision stream with strong lighting and high noise while being efficient and robust. In addition, the proposed method can be extended and introduced into such field remote real-time surveillance as medical image capturing besides farming activities.

References
