Image Mosaic Using Speeded Up Robust Feature Detection

Y.MURALI
M.TECH, Department Of ECE,
MITS, MADANAPALLE

V.MAHESH,
ASSO. Prof., Dept. Of ECE,
MITS, MADANAPALLE

ABSTRACT

So many methods have been proposed for the purpose of finding the feature points in the image. Methods like Harris, SUSAN, FAST, Laplacian of Gaussian, Difference of Gaussians, Determinant of Hessian Scale invariant Feature Transform (SIFT) have some defects like illumination varying, scale varying, rotation varying, more time taking. In order to recover from these problems we are discussing here the new technique, which is overcoming all the above mentioned defects. The technique is Speeded Up Robust Feature (SURF) technique. The main reason behind the huge advantages of this technique is its use of INTEGRAL IMAGES. Because of this the method performs faster than the precedent SIFT in finding the feature points and also performs faster matching. Below we have discussed about the way the SURF performs and our approach for the MOSAIC IMAGE in summary

1. INTRODUCTION

The main technique behind the image mosaic is image registration. Registration is a fundamental task in image processing used to match two or more pictures taken, for example, at different times, from different sensors, or from different viewpoints. Virtually all large systems which evaluate images require the registration of images, or a closely related operation, as an intermediate step. Specific examples of systems where image registration is a significant component include matching a target with a real-time image of a scene for target recognition, monitoring global land usage using satellite images, matching stereo images to recover shape for autonomous navigation, and aligning images from different medical modalities for diagnosis.

The automatic construction of large, high-resolution image mosaics is an active area of research in the fields of photo geometry, computer vision, image processing, and computer graphics. Image mosaics can be used for many different applications (Kumar et al.1995; Irani et al.1995a). The most traditional application is the construction of large aerial and satellite photographs from collections of images. More recent applications include scene stabilization and change detection, video compression and video indexing, increasing the field of view and resolution (Irani and Peleg1991; Chiang and Boult1996) of a camera, and even simple photo editing. A particularly popular application is the emulation of traditional film-based panoramic photography with digital panoramic mosaics, for applications such as the construction of virtual environments and virtual travel.

General detectors of feature are Harris corner detector [1], SIFT feature detector [2, 3] and SURF feature detector [4]. Harris algorithm has some characteristics, such as simple calculation and stability of extraction of corners.
However, it cannot adapt to the changes for scale which largely affects its application. The SIFT algorithm, is a good robustness of the scale invariant feature description method. But it still has some disadvantages, such as highly time complexity. Therefore, it does not favor the real-time application. So many improved versions of SIFT are proposed in order to reduce the computation time, such as PCA-SIFT and SURF. And integral images and box filters are used to improve the speed of detection in SURF. There are many research results under the SURF algorithm, such as CenSurE[6] and SUSurE[7], but they still have some shortcomings. In order to meet the real-time requirements of this article, we propose an improved SURF algorithm based on previous.

2. THE BASIC STRUCTURE OF THE IMAGE MOSAIC

The framework of the IMAGE mosaic system is showed as Figure1. The block diagram represents the step-by-step procedure in performing image mosaic using Speeded Up Robust Feature (SURF) detection. The process of image mosaic starts with the acquisition of images from the directory in our project. Then to match the images we will go for low level feature detection by using the technique called Speeded UP Robust Feature (SURF) detection and then we will match the features found by SURF algorithm by using the nearest neighbour method. After matching is performed, by using the most popular RANSAC method we will eliminate any outliers and will try to estimate HOMOGRAPHY between the pair of images. Then the final step is to align the images by using the GLOBAL ALIGNMENT technique. After global alignment if there is any misalignment then we will remove it by using Reordering step.

![Fig 1: MODEL OF IMAGE MOSAIC](image)

3. SURF FEATURE REGISTRATION

SURF, as proposed by Bay, is much faster in feature detection, and has better robustness than SIFT. Therefore, many scholars keen on it.

3.1 Feature Point Detection

Feature point detection in the SURF uses a basic second-order Hessian matrix approximation. Second-order Hessian matrix of image is defined as:

\[ H(\chi, \sigma) = \begin{bmatrix} L_{xx}(\chi, \sigma) & L_{xy}(\chi, \sigma) \\ L_{xy}(\chi, \sigma) & L_{yy}(\chi, \sigma) \end{bmatrix} \]  ......(1)

where \( L_{xx}(X,\sigma) \) is the convolution of the Gaussian second order derivative with the image \( I \) in point \( X=(x, y) \), and similarly for the \( L_{xy}(X,\sigma) \) and \( L_{yy}(X,\sigma) \). The maximum point of Hessian’s determinant in scale space and image space is identified as a feature point.

In order to reduce computation time drastically, it uses integral images and box filters. Box filter takes place of second-order Gaussian filter, and it improves the computing speed by using integral images to speed up the convolution. In the Figure3, the \( 9 \times 9 \) box filters are approximations of a Gaussian with \( \sigma=1.2 \) which represent the lowest scale for computing the blob response maps. They are denoted \( D_{xx}, D_{xy} \) and \( D_{yy} \).

\[ \text{det}(H_{\text{approx}}) = D_{xx}D_{yy} - (wD_{xy})^2 \]  ......(2)
where the weight w of the filter response is an important factor for balancing the determinant of Hessian. We can be solved by the following formula,

\[
w = \frac{|L_{n} (1.2) |_{f} | D_{n} (9) |_{f}}{|L_{n} (1.2) |_{f} | D_{n} (9) |_{f}} = 0.912... \approx 0.9 \quad \ldots \ldots \ldots \ldots \ldots \ldots (3)
\]

Fig: 2 : Filter Structure

Integral image is one of the main features in the SURF. It quickly calculates the convolution between the box filter and image by integral image. Integral image is defined as follows,

\[
I_{\Sigma} (X) = \sum_{i=0}^{x} \sum_{j=0}^{y} I(i,j) \quad \ldots \ldots \ldots \ldots \ldots \ldots (4)
\]

where I (x, y) represents the image, an integral image \(I_{\Sigma}(X)\) at a location \(X=(x, y)\) represents the sum of all pixels in the input image I within a rectangular region formed by the origin and x. Integral image calculation is shown as Figure 3.

Fig.3: Using Integral Images

The construction of scale image pyramid is similar to the SIFT algorithm, and it has 4 octaves. And each octave is subdivided into 4 scale levels, as shown Figure 5. The grey regions represent the size of box filters’ template. If the size of image is much larger than the size of the template, it can continue to increase the octave. If the size of box filters’ template is N×N, then the corresponding scale \(s = \frac{1.2 \times N}{9}\). And do non-maxima suppression in the 3×3×3 neighborhood of three-dimensional. Only at the biggest or smallest extreme point from 26 neighborhood points which are around the previous scale, the next scale and its scale can be considered a candidate feature point. Then it does interpolation in the scale space and the image space, and gets the steady feature points and the scale of values.

Fig 4: Construction of scale space

3.2 Descriptor Generation

In order to be invariant to image rotation, it is necessary to identify a reproducible orientation for the feature point. For this purpose, firstly it calculates the Haar wavelet responses in x and y direction within a circular neighborhood of radius 6s around the feature point, \(s\) is the scale of feature point which is detected. Secondly, it constructs a circular area around the feature point. Then a sliding orientation window of size \(\pi/3\) whirls around the feature point, and computes the vector sum of x and y direction of the Haar wavelet response. It uses these vectors to estimate the final direction, maximum value as the main orientation of the feature point.

For extraction of the descriptor, the first step is consisted of constructing a square region, which is centered the feature point and oriented along the orientation selected in the previous. The size of this window is 20s. The region is split up regularly into smaller 4×4 square sub-regions. For each sub-region, it computes Haar wavelet responses at 5×5 regularly spaced sample points, and computes the Haar
wavelet response \(x\)-direction \(dx\) and the Haar wavelet response \(y\)-direction \(dy\). Therefore, each sub-region is formed a 4-dimensional vector,

\[
V_{sub} = (\Sigma dx, \Sigma dy, |\Sigma [dy |] ) \quad \text{.........(5)}
\]

So it forms a \(4 \times 4 \times 4 = 64\) dimensional vector describing, and feature descriptor is formed of SURF-64 after vector normalization, as shown Figure 6.

**3.3 Image Matching and Gaining Image Transformation Matrix**

After extracting the SURF feature points from two video images, it needs to match these feature points. In other word, it searches the same feature points between the two images by calculating the similarity between the feature descriptor of the two images to determine whether the same point. This paper uses Nearest Neighbor (NN) [11] to match the feature points. The ratio, between the nearest neighbor point’s distance and second nearest neighbour point’s distance, is used to match the feature points. Similarity measure is taken between two feature vectors the Euclidean distance. The smaller the Euclidean distance value, the higher similarity the two vectors.

This paper uses Best-Bin-First (BBF) [3] to search the nearest neighbor. When it searches a node direction along a branch, the priority queue will be added to a member, which records information about the node. The information includes the current node position in the tree and distance between the node and the checks node. When a node is searched, it should delete it from the first team of the queue. Then it searches the other branch including the nearest neighbor node. Although the process can get higher correct matching rate, it exists some mismatch. And this paper uses RANSAC [12] to eliminate the mismatch. General image transformation includes translation, rotation, scaling and projection. [8] Projection transformation is highly adaptable and relatively few restrictions for the video image. Therefore projection transformation can represent the image transformation in this paper. The two corresponding points in the two images \((x’, y’)\) and \((x, y)\) satisfied the following transformation relation:

\[
\begin{bmatrix}
k \quad x' \\
y' \\
z \\
\end{bmatrix} = \begin{bmatrix}
h_1 & h_2 & h_3 & h_4 \\
h_5 & h_6 & h_7 & h_8 \\
h_9 & h_{10} & h_{11} & h_{12} \\
\end{bmatrix} \begin{bmatrix}
x \\
y \\
z \\
\end{bmatrix} \quad \text{or \ } \hat{X} = HX \\
\text{.........(6)}
\]

where \(k\) is the scale factor which responds to the image scaling relationship between images, \(X'=(x', y', z')^T\) is the point of the input image, and \(X=(x, y, z)^T\) is the point of the reference image. \(H\) has eight freedom degrees, so it needs to select at least four pairs of matching points to estimate \(H\). However, three points between the selected four points maybe collinear, the result would be unstable. Therefore, it must do nonlinear optimization for multi-feature points. Paper uses least squares to fit the parameter. The more the number of feature points on the estimated value, the more reliable of the estimation, and the smaller the error.

**4. GLOBAL ALIGNMENT**

In this method, the same feature point correspondences need to be identified over all views. This requires feature tracking. The \(i_{th}\) interest point on image \(k\), \(kX_i\), is a projection onto the mosaic of point \(mX_j\) which is called the pre-image point and is also usually projected in different views. All the image points that correspond to the projection...
of the same pre-image point are called N-view matches. The cost function to be minimized is defined as where $M$ is the total number of pre-image points, $\eta_j$ is the set of N-view matches and $kH_m$ is a mosaic-to-image homography $l$. In Eq. (1), both the homographies and the pre-image points are unknowns. The total number of unknowns is $\text{DOF of homography} \times \text{Number of views} + 2 \times \text{Number of pre-image points}$. Eq. (1) can be minimized iteratively by applying the non-linear least square methods.

In order to avoid the scaling effect on images the error term is defined on the image frame but also finds the position of the point on the mosaic frame.

5. RESULTS

The above figure is taken at night conditions which is given to prove that this project is efficient even under low light conditions the following table give the information

<table>
<thead>
<tr>
<th>NUMBER OF IMAGES TAKEN</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>LUMINANCE CONDITION</td>
<td>NIGHT</td>
</tr>
<tr>
<td>AVERAGE NO OF MATCHES</td>
<td>54</td>
</tr>
<tr>
<td>AVERAGE TIME TAKEN</td>
<td>6.8350s</td>
</tr>
<tr>
<td>AVERAGE MATCH TIME</td>
<td>0.0005s</td>
</tr>
</tbody>
</table>

About the number of images taken and number of matches obtained and also time taken for matching

The above figure is taken at night conditions which is given to prove that this project is efficient even under low light conditions the following table give the information about the number of images taken and the number of matches obtained and also the time taken for matching.

<table>
<thead>
<tr>
<th>NUMBER OF IMAGES TAKEN</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>LUMINANCE CONDITIONS</td>
<td>MORNING</td>
</tr>
<tr>
<td>AVERAGE NO. OF MATCHES</td>
<td>255</td>
</tr>
<tr>
<td>AVERAGE TIME TAKEN</td>
<td>8.9650s</td>
</tr>
<tr>
<td>AVERAGE MATCH TIME</td>
<td>0.103s</td>
</tr>
</tbody>
</table>
6. SUMMARY

This paper proposes IMAGE mosaic method. The method extracts feature descriptor by SURF, and implements the matching feature points by BBF. Comparing to the previous methods like Harris, SUSAN, SIFT, this method is very good, weather you take for finding the feature points, or matching the feature points between the image pairs ,or the image illumination variations or the rotation variation. The time taken for mosaicing the 12 images are approximately 8 sec, which is better than its precedent Scale invariant feature transform(SIFT), which is taking more than 20s . This also giving good results for the image rotation and image brightness variations.

ACKNOWLEDGEMENT

This work is supported by my respected guide and the project co-ordinator. Thanks to him.

REFERENCES