Gaussian Compound Feature Descriptor and Contourlet Transform for Image Registration in 3D Reconstruction

Huan Liu¹, Genfu Xiao²* and Wei Peng³

¹College of Electronic and Information Engineering, Jinggangshan University & Key Laboratory of Watershed Ecology and Geographical Environment Monitoring NASG, Jinggangshan University, Ji’an, Jiangxi 343009, P.R. China
²College of Mechanical and Electronic, Jianggangshan University, Ji’an, Jiangxi 343009, P.R. China
³College of Architectural and Civil Engineering, Jianggangshan University, Ji’an, Jiangxi 343009, P.R. China

Abstract

This paper presents a new method for multi-view images registration based on stereo vision system. Our aim is to recover the surface shape of the subject accurately and efficiently in spite of the influences derived from illumination variation, blur affection and image transformation on the 2D images. For this purpose, we devote to developing an innovative stereo registration algorithm. In the first phase, a novel feature descriptor is constructed by adding multi-scale Gaussian parameters into the illumination-robust and anti-blur combined moment invariants in fusion of the pixel gray and gradient. The new Gaussian combined moment invariants are calculated on the multi-scale low frequency sub-band by Contourlet transform. Meanwhile, grid entropy was computed on the multi-direction high frequency sub-band as to get the structure characteristics of the image. Then a novel compound feature descriptor was presented by a combination of the Gaussian moment invariants and grid entropy. It is applied to conduct the similarity measure for the initial image registration. In the second phase, the bidirectional matching strategy with strict geometric constraints composed of the distance and the slope between matching pairs is proposed for eliminating the incorrect matching pairs in the initial image registration. Consequently, the correct matching pairs are obtained at this stage. The experimental results reveal that both the accuracy and the efficiency of our approach are superior to those of SIFT and SURF. Finally, 3D cloud data and 3D model of the subject are achieved.

Key Words: Multi-scale and Multi-direction, Gaussian Combined Moment Invariant, Grid Entropy, Feature Registration, Bidirectional Geometric Constraint, 3D Reconstruction

1. Introduction

Image registration is often regarded as an important preprocessing step for many image processing applications such as computer vision [1,2], change detection [3], pattern recognition [4,5], image analysis [6], 3D reconstruction [7,8], pose detection [9] etc. Image registration should be efficient, robust and accurate [10]. Its purpose is to align two or more images taken at different times or from different viewpoints [11]. Due to illumination changes, large noise, and various transformations, an excellent image registration is still a challenging task.

Generally, image registration methods can be classified into area-based methods and feature-based methods. In area-based methods, intensity of each pixel in both images is used to compute similarity metric to find the optimized geometric transformation iteratively. The computation time should be taken care of. In feature-based methods [12], correspondence features of the images such as points, lines, edges etc. are extracted and used to find
the required geometric transformation model between two images through these correspondences. The main purpose of this paper is solving the multi-view image registration based on the feature image registration method. Multi-view image registration realizes the image registration and completes the 3D reconstruction by the images collected from the different views. Those multi-view images are good for the whole information of the object.

At first, the feature points are extracted in left reference and right target images, respectively. And then the feature descriptor and corresponding relationship between feature points are established. Next then the translation relationship is obtained by calculating the corresponding relationship. Finally the right sensed target image can be transformed to the specific form according to the above geometric transformation. Most of feature-based methods usually consist of four steps: feature detection, feature matching, transformation estimation, transformation and resampling [13]. The first two processes are extremely critical steps. Scale-invariant feature transform algorithm (SIFT) is the most common way to construct matching pairs [14,15]. Hasan et al [16] proposed spatial relationship analysis on SIFT keypoints. SIFT method and some other improved SIFT methods [17] are used for image registration. The advantages of SIFT are scale, rotation and illumination invariant and so on. But it is a complex algorithm with slow speed. So that it is difficult to satisfy the real-time applications. When an angle exceeds $35^\circ$, the invariance of multi-view transform is impossible. Speed-up Robust Feature [18] (SURF) is derived from SIFT but it is modified by using hessian matrix, integral image and Harr response. In literature [19] the improved SURF algorithm can extract more accurate matching points than the original SURF. However the stability and robustness of the improved SURF method still needs further study. Among any other algorithms such as Maravec, Harris and Susan, Harris is the best algorithm for its simple computation, stability and robustness.

Image moment invariants have been widely applied to image registration. Classical moment invariants include Hu invariant moments, Legendre moments, Zernike moments, composite moments, etc. Dellinger [20] developed robust image features namely invariant moments which are stable to image rotation and slight scaling. However, they cannot maintain robustness to face light variance effectively. Wu [21] adopted seven moment invariants into feature extraction. And the outliers are used to remove the error pairs through the spatial restraint. All in all, most of the existing moment invariants have no consideration of the factors including both illumination changes and blur influences. In addition, those geometric moment invariants merely focus on describing the edge information of the image without the center information particularly in the higher-order geometric moment invariants. It is well known that the most important information is just from the center of the image.

Recently, researchers have gradually focused on the analysis of the multi-scale geometric in feature extraction, such as Gabor filtering [22], wavelet transform [23], Contourlet transform [24] etc. The properties of contourlet transformation are multi-direction, multi-scale, multi-resolution, anisotropic, small sampling redundancy, fast calculation speed and so on. Therefore, it has great advantages in application of image fusion [25,26], image enhancement [27] and image reconstruction [28]. Cha hira Serief [29] proposed a new feature points extraction method based on the nonsubsampled contourlet transform. The NSCT based feature detection algorithm is proposed to solve the automated image registration problem. Swapna [30] proposed an image retrieval algorithm based on contourlet transform by taking coefficients on directional sub-bands as image features. Although the higher average precision was obtained, the dimension of feature descriptor was high and the calculation was complicated. Dong et al. [31] applied C-mean clustering method to solve texture classification by Contourlet transform sub-band. The research shows that the method on the basis of the multi-scale geometric analysis has obvious advantages in feature extraction. In literature [32], with the multi-resolution Contourlet transform for image preprocessing, control points are identified well for a more reliable image registration. So contourlet is considered as a key feature extraction for enhancing image registration. All of the above studies are successful applications with the multi-scale geometric analysis of contourlet transform in feature extraction, which fully demonstrated the advantages of this method. However, all of them only extract a single vector as image feature which is invariant in some cases. And yet they are easy to increase the false matching ratio in most conditions.
In view of the above mentioned deficiencies, the primary contributions in this paper are summarized concisely as follows.

1. Contourlet is a multi-scale and multi-direction transformation in discrete domain. Contourlet can simultaneously get the outline and details of the image at different scales and directions sub-bands. The low frequency sub-band describes the general characteristic of the whole image, and the high frequency sub-band describes the details of the image. With contourlet transform, the reference image and the target image are decomposed into multi-scale low frequency sub bands and multi-direction high frequency sub bands.

2. Then Gaussian combined moment invariants (GCMs) on low frequency sub bands were defined on image gray and gradient. Image grid entropy reflects the image structure. Scale factor is added in the new proposed GCMs as to extract the statistical characteristics of different low frequency sub bands including gray and gradient. Grid entropy (GE) for each grid cell on high frequency sub bands is extracted for depicting image structure feature. The direction feature of high frequency sub-band compensates the disadvantage of GE without direction property. Then the new composed feature descriptor combines the global invariance of GCMs and GE, and local performance of the multi-scale and multi-direction of Contourlet. It has obvious superiority in image feature registration.

3. On the purpose of improving the reliability of image registration, the distance and slope of the geometric constraint between the lines linked matched pairs are involved. And then the bidirectional matching strategy is also combined to remove the false correspondences in the initial similarity measurement.

The remainder of the paper is organized as follows. Section 2 describes the related former work. In section 3, the details of the improved strategies are expounded. Section 4 illustrates image correspondences with comparisons to other approaches, followed by some concluding remarks in section 5.

2. Related Work

2.1 Area-based Methods

Area-based methods perform the image intensity values directly without image features. These methods can be classified into three categories: correlation-like methods, Fourier methods, and mutual information methods. The main idea of Correlation-like methods is to compute the similarities of window pairs in two images. And then the pair with the most similarity was considered as a correspondence. The drawbacks are high computational complexity and the flatness of the similarity measure in textureless regions. Fourier methods execute the Fourier representation on images in the frequency domain with the limitations in the two images with substantially different contents. At last, Mutual information methods provide an attractive metric for maximizing the dependence between two images. Nevertheless, they cannot obtain a global maximum. Thus they reduce the robustness inevitably.

2.2 Feature-based Methods

Feature-based methods are based on the extraction of salient features in the images. A popular strategy for solving the feature matching problem is to use a two-stage process [33–35]. In the first stage, a set of correspondences are calculated with similarity constraint. It contains abundant correct matches as well as a large number of false matches. In the second stage, the incorrect pairs are removed by utilizing geometric constraint. Random sample consensus (RANSAC) [36] is a famous robust estimator which has been utilized to calculate model parameter or search the correct correspondences broadly. RANSAC maintains robust when the outlier is a small minority. If the outliers are too many in the matched pairs, RANSAC may cause time consuming and unstable condition.

2.3 Multi-scale and Multi-direction Decomposition of Contourlet Transform

Contourlet transform is a multi-scale, multi-direction image representation method proposed by M. N. Do et al. [37]. Contourlet transformation works in discrete domain of image: it is completed by Laplacian Pramid and Directional Filter Bank. The high and low frequency of sub bands conduct with LP at first. And then the high frequency is decomposed into multi-direction sub bands by DFB. Meanwhile the low frequency continues to carry with Contourlet transform by LP. Thus it is an iterative
process. Contourlet transformation can well capture overall contour of the image as well as detailed information from multiple directions. The low frequency sub bands retain most of the energy of the image namely general outline feature. The high frequency sub bands contain the edge, texture and direction. Considering the aspects of multi-scale and multi-direction advantages, contourlet is applied to increase the different types of complementary information in order to improve the accuracy in the field of image registration.

The contourlet transformation on Pepper is displayed in Figure 1.

3. Proposed Method

3.1 Gaussian Combined Moment Invariants on Low Frequency Sub-band

New combined moment invariants with robust illumination and anti-blur are proposed in Literature [7]. However, the combination of geometric invariant moments focuses on describing the edge information of the image but losing the center information. In this paper, Gaussian moment invariants (GCMs) are presented. Scale factors are added in geometric invariant moment in order to express the feature at different scales.

Discrete GCM on order \( p + q \) is defined in image \( I(i,j) \), as follows

\[
m_{pq} = \frac{4}{(k-1)^2} \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} \left( \frac{x}{\sigma} \right)^p \left( \frac{y}{\sigma} \right)^q \exp \left( -\frac{x^2 + y^2}{2\sigma^2} \right) f(x,y)
\]

(1)

The center moment of GCMs is defined,

\[
\mu_{pq} = \frac{4}{(k-1)^2} \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} \left( \frac{x-x'}{\sigma} \right)^p \left( \frac{y-y'}{\sigma} \right)^q \exp \left[ -\frac{(x-x')^2 + (y-y')^2}{2\sigma^2} \right] f(x,y)
\]

(2)

where \((x', y')\) is the Barycentric ordinates in window \( m \times n \).

\[
\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}}
\]

(2a)

where \( r = \frac{p + q}{2} \), \( \eta_{pq} \) means the normalized central moments.

\[
\phi_1 = (\eta_{00} - 3\eta_{11})^2 + (3\eta_{11} - \eta_{00})^2
\]

\[
\phi_2 = (\eta_{00} + \eta_{11})^2 + (\eta_{11} + \eta_{00})^2
\]

\[
\phi_3 = (\eta_{00} - 3\eta_{11})(\eta_{11} + \eta_{00})[3(\eta_{11} + \eta_{00})^2 - 3(\eta_{11} + \eta_{00})^2]
\]

\[
+ (3\eta_{11} - \eta_{00})(\eta_{11} + \eta_{00})[3(\eta_{11} + \eta_{00})^2 - (\eta_{11} + \eta_{00})^2]
\]

\[
\phi_7 = (\eta_{11} - \eta_{00})(\eta_{11} + \eta_{00})[3(\eta_{11} + \eta_{00})^2 - 3(\eta_{11} + \eta_{00})^2]
\]

\[
+ (\eta_{11} - \eta_{00})(\eta_{11} + \eta_{00})[3(\eta_{11} + \eta_{00})^2 - (\eta_{11} + \eta_{00})^2]
\]

(2b)

The three Gaussian combined moments are given as follows.

\[
\zeta_1 = \frac{\phi_4}{\phi_3}, \quad \zeta_2 = \frac{\phi_7}{\phi_2}, \quad \zeta_3 = \frac{\phi_3}{\phi_1 \phi_2}
\]

(2c)

where \( \phi_n \) (\( n = 3, 4, 5, 7 \)) means the normalized central Hu moments.

The Gaussian gradient combined moment invariants (GGCMs) involve the gradient value, which is defined as follows

\[
g_{pq} = \frac{4}{(k-1)^2} \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} \left( \frac{x-x'}{\sigma} \right)^p \left( \frac{y-y'}{\sigma} \right)^q \exp \left[ -\frac{(x-x')^2 + (y-y')^2}{2\sigma^2} \right] g(x,y)
\]

(3)

where \( g(x,y) \) signifies the gradient value, it can be calculated by function

\[
g(x,y) = \sqrt{\left( \frac{\partial f(x,y)}{\partial x} \right)^2 + \left( \frac{\partial f(x,y)}{\partial y} \right)^2}
\]

(3a)
where \((\tilde{x}_g, \tilde{y}_g)\) means the Barycentric coordinates with gradient in window \(m \times n\).

\[
\tilde{x}_g = \frac{\sum_{x=1}^{m} \sum_{y=1}^{n} g(x,y) x g(x,y)}{\sum_{x=1}^{m} \sum_{y=1}^{n} g(x,y)}, \quad \tilde{y}_g = \frac{\sum_{x=1}^{m} \sum_{y=1}^{n} g(x,y) y g(x,y)}{\sum_{x=1}^{m} \sum_{y=1}^{n} g(x,y)}
\]  

(3b)

The three Gaussian gradient combined moment invariants are given as follows.

\[
\xi_1 = \frac{\phi_{s_1}}{\phi_{s_5}}, \quad \xi_2 = \frac{\phi_{s_2}}{\phi_{s_5}}, \quad \xi_3 = \frac{\phi_{s_2}}{\phi_{s_5}}, \phi_{s_1}
\]

(3c)

The final combined feature vector comprises of gray and gradient moment invariants.

\[
f = [\xi_1, \xi_2, \xi_3, \xi_4, \xi_5, \xi_6]
\]

(3d)

Due to the shrink of the low frequency sub-band image through the decomposition of Contourlet transform, the combined Gaussian combined moment invariants should be weighted on the different scales for more image information, and then the average value of those weighted combined moment invariants are taken as the final feature of the low frequency. We adopted generalized Gauss model (2GM) to calculate the scale parameter. The probability density function of 2GM is

\[
f(x, \sigma, \gamma) = \frac{\gamma}{2\sigma^2 \Gamma(1/\gamma)} e^{-\frac{\gamma}{2} x^{-\gamma}}
\]

where \(\Gamma(\sigma) = \int_0^\infty e^{-t/m^{-1}} dt\), \(\sigma\) is scale parameter, \(\gamma\) is shape parameter.

\[
F(\gamma) = \frac{\mu_2^2}{\omega_1} = \frac{\Gamma(2/\gamma)}{\Gamma(1/\gamma) \Gamma(3/\gamma)}
\]

(5)

Moment estimation of \(\sigma\)

\[
\sigma = \mu_1 \frac{\Gamma(1/\gamma)}{\Gamma(2/\gamma)}
\]

(6)

where \(\mu_1 = \frac{1}{N} \sum_{i=1}^{N} [\xi_i]_1\), \(\omega_1 = \frac{1}{N} \sum_{i=1}^{N} [\xi_i]_1^2\).

\(l_j\) means the weight of low frequency image at scale \(j\).

\[
l_j = \frac{\sigma_j}{\sum \sigma_j}
\]

(7)

where \(\sigma_j\) is parameter at scale \(j\). The large value of \(l_j\) shows the large dispersion and the strong recognition ability of sub-band, and vice versa.

Then the weighted combined moment invariants are revised as

\[
\zeta'_n = \sum_{j=0}^{\infty} l_j \zeta'_n
\]

(8)

where \(\zeta'_n\) is the \(n^{th}\) weighted combined moment invariants, \(\zeta'_n\) is the \(n^{th}\) combined moment invariants at scale \(j\).

3.2 Grid Entropy on High Frequency Sub-band

The image structure can be measured by grid entropy. Supposing the image \(I(x, y)\) has \(R\) gray levels, the probability of \(i (i = 0 \sim R – 1)\) gray level is \(p_i\), the image entropy can be defined as.

\[
H = \sum_{i=0}^{R-1} p_i \log p_i
\]

(9)

where if \(p_i = 0\), we can deem that \(p_i \log p_i = 0\).

The high frequency multi-direction sub-bands of image \(I(x, y)\) are constructed according to contourlet transform. \(I(x, y, O_n)\) signifies the sub-band of \(O_n\) image. The entropy of \(I(x, y, O_n)\) is marked as \(H_{I(O_n)}\). In \(I(x, y, O_n)\), the size of \(m \times n\) rectangle windows are divided by taking the feature point as center. \(I_i(x, y, O_n)\), \(i = 1, 2, \ldots, w\) (\(w\) is the number of the feature points) are marked as rectangle window sub-block. The entropy of \(I_i(x, y, O_n)\) is marked as \(H_{I_i(O_n)}\).

The entropy feature is refined as

\[
E_{I(O_n)} = \frac{H_{I(O_n)}}{2^N}
\]

(10)

High frequency sub-band increased at the speed of index 2. In order to reduce the computation, we choose the more optimal high frequency sub-bands according to the cost function.

\[
J(k) = \frac{\sigma_{max}^2(k) - \sigma_{max}^2(k)}{\sigma_{min}^2(k)}
\]

(11)

Each high frequency sub-band is divided into the same square area of \(T \times T\). In Equation (11), \(\sigma_{max}^2(k)\) and
\( \sigma^2_{\text{max}}(k) \) means the maximum and minimum variances of square area in sub-bands, respectively. According to the cost function values, we select the former four bands from the total eight high-frequency sub bands in the first order and the first band from the four high-frequency sub bands in the second order, respectively.

The frequency from different images reflects different spatial activity. So the values of grid entropy on different high frequency multi-direction bands are better to be weighted by contrast sensitivity function (CSF). It is defined as follows.

\[
A(O_n) = (2.6 \times 0.0192 + 0.114 \times f) \times e^{-0.145f^{1.1}}
\]  

where \( f = \sqrt{f_x^2 + f_y^2} \) is spatial frequency of \( I(x, y, O_n) \),

\[
f_c = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} [f(i, j) - f(i, j - 1)]^2
\]  
is the line frequency of \( I(x, y, O_n) \).

\[
f_c = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} [f(i, j) - f(i - 1, j)]^2
\]  
is the column frequency of \( I(x, y, O_n) \).

\( E_{(O_n)} \) was weighted as

\[
E'_{(O_n)} = \frac{A(O_n)}{\sum A(O_n)} \cdot E_{(O_n)}, \quad E'_{(O_n)} = \sum A(O_n) \cdot E_{(O_n)}, \quad N = 1, 2, 3...n
\]  

where \( f_1, f_2, \ldots, f_n \) mean the spatial frequencies of the \( n \) optimal high-frequency sub-bands. \( E_{(O_n)}, E'_{(O_n)}, \ldots, E'_{(O_n)} \) mean the grid entropies of the \( n \) optimal high-frequency sub-bands. \( E'_{(O_n)}, E'_{(O_n)}, \ldots, E'_{(O_n)} \) mean the weighted grid entropies.

### 3.3 Compound Feature Vector

The combined moment feature vector and grid entropy are extracted by the above methods. Considering the different types they are, standardization before fusion is required. In this paper, on the one hand, as to the low frequency sub bands, three scales sub bands from the first level to the third level are selected. On the other hand, as to the high frequency sub bands, four sub bands from eight bands in the first level and one sub-band from four bands in the second level are selected according to cost function (Eq. 11).

\[
F = [\xi_{1}', \xi_{2}', \xi_{3}', \xi_{4}', \xi_{5}', \xi_{6}', \xi_{7}', \xi_{8}', \xi_{9}', \xi_{10}', \xi_{11}', \xi_{12}', \xi_{13}', \xi_{14}', \xi_{15}', \xi_{16}']
\]

After Gauss normalization, there are

\[
F = \frac{|F - \mu_F|}{\sigma_F^2}
\]

where \( \mu_F \) and \( \sigma_F^2 \) are mean and standard deviation of feature vector \( F \), the normalized vector is

\[
F = [\xi_1', \xi_2', \xi_3', \xi_4', \xi_5', \xi_6', \xi_7', \xi_8', \xi_9', \xi_{10}', \xi_{11}', \xi_{12}', \xi_{13}', \xi_{14}', \xi_{15}', \xi_{16}']
\]

### 3.4 Matching Strategy of Bidirectional Geometric Constraint

The bidirectional matching strategy is adopted for the aim of improving the reliability of the correspondences. First, starting with \( p_l \) is in the reference image, \( p_r \) is located in the target image under similarity measurement. Conversely, starting with \( p_r \), \( p'_l \) is located in the reference image. If \( p_l \) and \( p'_l \) are the same point, \( p_l \) and \( p_r \) are judged as a correct correspondence. Otherwise it will be abandoned.

Now making the similarity of the key points based on feature descriptor \( F \).

\[
S_{ij} = \exp \left( - \left| F_i(i) - F_r(j) \right| \right)
\]

where \( \left| F_i(i) - F_r(j) \right| \) means a distance measurement defined by

\[
\begin{align*}
\left| F_i(i) - F_r(j) \right| &= \left| \xi_{1(i)} - \xi_{1(j)} \right| \left| \xi_{2(i)} - \xi_{2(j)} \right| \left| \xi_{3(i)} - \xi_{3(j)} \right| \cdots \left| \xi_{16(i)} - \xi_{16(j)} \right| \\
&= \left| E_{1(i)} - E_{1(j)} \right| \left| E_{2(i)} - E_{2(j)} \right| \cdots \left| E_{16(i)} - E_{16(j)} \right|
\end{align*}
\]  

After the similarity measurement, there are some incorrect correspondences. So with the view of further improving of the matching accuracy, the false pairs should be eliminated continually by adding geometric constraint.

In the premise on two images with the same resolution, the idea that the lines between all the correct correspondences have the same distance and slope is reasonable. Hence the geometric constraints are applied to remove the incorrect correspondences further. The new
constraint function is redefined.

\[ CF = D \times K \]  \hspace{1cm} (19)

where \( D \) means the distance matrix and \( K \) is the slope matrix.

\[ d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \]  \hspace{1cm} (20)

\[ k_{ij} = \frac{y_j - y_i}{x_j - x_i} \]  \hspace{1cm} (21)

4. Experiments and Analysis

The outline of the multi-view images registration and the 3D reconstruction model are described as follows.

1. The feature points between the left-view and right-view images by Harris are extracted separately.
2. The initial registration by a new compound feature descriptor (CFD) comprised with Gaussian combined moment invariants (GCMs) and grid entropy (EG) is carried out.
3. The false correspondences in initial registration are removed with strategy of bidirectional geometric constraint.
4. The matching model is accomplished based on the final real correct correspondences. Lastly 3D reconstruction models are obtained.

In the tests, ceramics are taken as our research subjects. The image registration from multi-view images and 3D reconstruction are achieved. We use four CMOS cameras (Cyber-shot DSC-W30) with a resolution of 2816 \( \times \) 2112 and an effective pixel of about 6,003,000 pixels. The size of image is 140 \( \times \) 190 pixels. Two cameras are located in the front of ceramic, which are used to get the left and right view images from the front of the subject. And the others are laid out at the back of the ceramic, which are used to get the left and right view images from the rear of subject.

We devoted ourselves to do enough tests in three typical cases depending on different choices of blur and illumination darkening coefficients. The detailed instructions are expounded. For the example of the front two views images, case 1 (Figure 2) indicates the images in the normal situation, and case 2 (Figure 3a) is the condition with the blur noise parameter of 3.0 added in the normal right view image. Case 3 (Figure 3b) says the right view image added blur noise parameter of 4.5 and light darkening parameter of 6.0 with the increase of the value of blur and darkening parameters, the images become dimmer and darker. The three above mentioned situations are represented. The too large or too small coefficient has no need to study without any realistic significance for recognizing the subject difficulty.

4.1 Feature Points Extraction

Feature point extraction is an initial and important step. In this subsection, the comparison between SIFT and Harris is conducted.

In Case 1 the test is executed on two images (See Figure 2a and Figure 2b). The result is obtained by SIFT in Figure 4a and Figure 4b. As seen in the figures, 121 points in the left image and 114 points in the right image.

We devoted ourselves to do enough tests in three typical cases depending on different choices of blur and illumination darkening coefficients. The detailed instructions are expounded. For the example of the front two views images, case 1 (Figure 2) indicates the images in the normal situation, and case 2 (Figure 3a) is the condition with the blur noise parameter of 3.0 added in the normal right view image. Case 3 (Figure 3b) says the right view image added blur noise parameter of 4.5 and light darkening parameter of 6.0 with the increase of the value of blur and darkening parameters, the images become dimmer and darker. The three above mentioned situations are represented. The too large or too small coefficient has no need to study without any realistic significance for recognizing the subject difficulty.

4.1 Feature Points Extraction

Feature point extraction is an initial and important step. In this subsection, the comparison between SIFT and Harris is conducted.

In Case 1 the test is executed on two images (See Figure 2a and Figure 2b). The result is obtained by SIFT in Figure 4a and Figure 4b. As seen in the figures, 121 points in the left image and 114 points in the right image. Most of them are overlapped, and the distribution is uneven. By Harris, 78 points and 82 points are successfully tracked in the two pairs (see Figure 5a and Figure 5b). Although the number of feature points by Harris is few, the distances between points are appropriate. The distribution of points is even not too dense, which can also

(a) Left reference image  (b) Right matching image

Figure 2. Two views images in case 1.

(a) Right image in case 2  (b) Right image in case 3

Figure 3. Right matching images in case 2 and case 3.
save time for the subsequent matching process. The distribution of these points is superior to the results from SIFT. Figure 6a and Figure 6b show the results are produced by SIFT and Harris, respectively in case 2. Figure 7a and Figure 7b are the results in case 3.

Table 1 lists the number of feature points extracted by SIFT and Harris. According to Table 1, we can observe that the number of points extracted by Harris maintains stability without large change in three cases. However, the number from SIFT reduces greatly. It can be seen that SIFT has well invariance on scale, rotation and affine transformation and ordinary performance on fuzzy and illumination variation. Harris can still extract feature points steadily even in the worse condition. Thus in our following experiments, we choose Harris as our extraction algorithm for whole 3D reconstruction.

4.2 Feature Points Matching

The tests are carried out in case 1, case 2 and case 3 based on the steps of points extraction. In this section, the tests are accomplished by SIFT, SURF and our proposed method, respectively. Despite SURF or improved SURF algorithms is modified based on SIFT, their structures are similar to them of SIFT. They all establish scale space by pyramid transformation. The slight differences are just the operation process and the dimension of descriptor. Table 2 lists the comparison of some matching parameters in three methods. initial_m means the number of initial matching, final_m says the number of final matching, correct_m means the correct number of correspondences in the final matching, ratio_m means the correct matching ratio, T says the consuming time on the construction of feature descriptor and points matching. As seen from the table, with the condition worsening, the accuracy of SIFT and SURF decrease greatly by about 55% from 80% to 25% and 57% from 79% to 22%, respectively. While the precision from our method maintains more than 85% in all three cases and the first two cases the precision are more than 90%. As to SIFT and SURF, the characteristics from adjacent scale space are correlated in virtue of pyramid transformation. Especially
when the gray difference between two images is large, and yet the characteristic difference is not obvious. In addition, it is too simply that the descriptor is merely constructed on gradient. So it is easy to cause error mapping and low registration accuracy. In the paper, contourlet can get different scale and direction sub-bands for its good directivity and anisotropy. Owing to the complementary information, Contourlet is capable of describing feature well and improving registration precision greatly. The computational time in our method with 11-dimension is superior to both SIFT with 128-dimension and SURF with 64-dimension. The efficiency of SURF is superior to SIFT. The most of time in SIFT or SURF algorithm is consumed in generation of feature vector for their high dimensions. The dimension of our new combined feature vector (i.e. CFD) is lowest in three methods. As a result, it greatly reduces the time spent in feature generation and feature matching, and the space depleted in feature storage. Thus high dimension cannot always obtain better accuracy, but the efficiency reduces on the contrary.

SURF is improved based on SIFT, so the structure and operation steps are similar to SIFT. Considering the length of the article, here merely displays the figures about matching results of SIFT and our method. By SIFT the initial matching was performed with 128-dimension feature descriptor firstly in Figure 8a. Next RANSAC was utilized to remove false correspondences in Figure 8b. Likewise, the similar process was conducted in case 2 and case 3.(see Figure 9 and Figure 10). Meanwhile, the other way is applied to implement the initial matching with Gaussian compound descriptor (GCD) a novel vector designed by combining the Gaussian combined moment invariants and Grid entropy (see Figure 11a), then the matching strategy of bidirectional geometric constraint is adopted to eliminate the incorrect pairs (see Figure 11b). The same method was utilized in case 2 and case 3 (see Figure 12, Figure 13).

Then we selected thirty points at random. The differences between real values and calculated values in three

<table>
<thead>
<tr>
<th>methods</th>
<th>Parameter</th>
<th>Case 1</th>
<th>Case 2</th>
<th>Case 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIFT</td>
<td>initial_m</td>
<td>109</td>
<td>105</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>final_m</td>
<td>61</td>
<td>48</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>correct_m</td>
<td>49</td>
<td>25</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>ratio_m(%)</td>
<td>80.3</td>
<td>56.3</td>
<td>25.0</td>
</tr>
<tr>
<td></td>
<td>T(s)</td>
<td>0.906</td>
<td>0.878</td>
<td>0.799</td>
</tr>
<tr>
<td>SURF</td>
<td>initial_m</td>
<td>106</td>
<td>97</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>final_m</td>
<td>59</td>
<td>57</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>correct_m</td>
<td>47</td>
<td>31</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>ratio_m(%)</td>
<td>79.7</td>
<td>54.4</td>
<td>22.6</td>
</tr>
<tr>
<td></td>
<td>T(s)</td>
<td>0.586</td>
<td>0.572</td>
<td>0.498</td>
</tr>
<tr>
<td>Our method</td>
<td>initial_m</td>
<td>122</td>
<td>104</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td>final_m</td>
<td>97</td>
<td>89</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>correct_m</td>
<td>92</td>
<td>81</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>ratio_m(%)</td>
<td>94.8</td>
<td>91.0</td>
<td>87.2</td>
</tr>
<tr>
<td></td>
<td>T(s)</td>
<td>0.579</td>
<td>0.551</td>
<td>0.487</td>
</tr>
</tbody>
</table>
cases by SIFT and our method are displayed in Figure 14. The coordinates marked next to the points in the chart are the corresponding coordinates in the reference image. As can be observed from the chart, the coordinates calculated by our proposed method are closer to the real values. While the differences between the computed values by SIFT and real values are larger.

Table 3 lists the root mean square error between feature points (root mean square error, RMSE) which is utilized to evaluate the registration accuracy.

$$RMSE = \frac{1}{m} \sqrt{\sum (x_i - x'_i)^2 + (y_i - y'_i)^2}$$  \hspace{1cm} (22)

where $m$ is the quantity of feature point. $x_i$ and $y_i$ are the coordinates of $i$ in the reference image. $x'_i$ and $y'_i$ are the coordinates of $i$ in the transformed image. As seen in Table 3.

The lower of RMSE is, the higher the accuracy is. In all the cases, the values from SIFT are larger than those from our proposed method. From the view of precision and robustness, SIFT is better than SURF. But with the worse of the condition, the both values increase greater than our method. The values from our proposed approach maintain stability even in worse condition. The 3D models with our method are shown in Figure 15 and Figure 16.

5. Conclusion

In this paper, we propose a new method of 3D reconstruction based on stereo vision system included only four cameras. With Contourlet transform, the image was decomposed into several multi-scale and multi-direction sub bands. The novel Gaussian compound descriptor was
constructed by moment invariant added Gaussian scale parameter on low frequency multi-scale sub bands and grid entropy on high frequency multi-direction sub bands. In the image registration, a bidirectional geometric constraint was proposed in order to remove the false correspondences in the initial matching with similarity measurement. Experiments demonstrate that a highly accurate image registration can be obtained successfully with the help of the robust stereo matching between images even in the blurred and light changing condition.

This method fully fuses the global feature and local structure character of the images. It also effectively overcomes the shortcoming from a single feature which is difficult to describe the multiform characteristics. Our method is superior to other algorithms on comprehensive performance with strong effectiveness and applicability. Given that several false correspondences still exist in the final matching model after bidirectional geometric constraint, especially in the worse conditions. Further research will be conducted in the future for the purpose of improving accuracy in consideration of multiple factors synthetically. In our experiments, the ceramics are relative simply. The study on the more complicated subjects is another focus in our next research work.

Acknowledgement

This work was supported by the Project of the National Nature Science Foundation of China (Nos. 61462046, 61640412), the Natural Science Foundation of Jiangxi Province (Nos. 20161BAB202049, 20161BAB204172), the Bidding Project of the Key laboratory of watershed ecology and geographical environment monitoring, NASG (Nos. WE2016003, WE2016013, WE2016015), the Science and Technology Research Projects of Jiangxi Province Education Department (Nos. GJJ160741, GJJ170632, GJJ170633), the art planning project of Jiangxi province (YG2016250, YG2017381), Jinggangshan University Doctoral Scientific Research Foundation (Nos. JZB1311, JZB15016, JZB15009).

References


[36] Fischler, M. A. and Bolles, R. C., “Random Sample Consensus: a Paradigm for Model Fitting with Appli-


*Manuscript Received: Apr. 11, 2017*

*Accepted: Apr. 27, 2018*