A Method of Image Registration Based On Best Similarity of Local Geometric Figure

Zetao Jiang
Guilin University of Electronic Technology, Guilin, China
Nanchang Hangkong University, Nanchang, China
Email: zetaojiang@126.com

Le Zhou and Liwen Zhang
Nanchang Hangkong University, Nanchang, China
Email: {zhoule888,zhangliwen}@126.com

Abstract—Image Registration is an important part of computer vision. We propose a method of image registration by obtaining best similarity of local geometric figure that utilizes opposite core difference (OCD) of corresponding local figure. This method gets initial matching after describing precisely SIFT points by constructing feature subspaces based on the detection of SIFT feature points. Then we describe the corresponding similarity by OCD of local figure constructed by SIFT points and choose the feature points that possess highest similarity measure as point set to compute projective transformation matrix $T_{opt}$.

Experiments have proved that the precision of the matrix $T_{opt}$ and the Image matching is at a high level.

Index Terms—SIFT feature subspace, local geometric similarity, projective, OCD

I. INTRODUCTION

Image matching is one of key technology of computer vision. There are two kinds of methods[1-5] for it: based on feature and based on density. One method based on feature extracts image feature mainly by some algorithms and the another one estimate similarity by image distribution information, the more similar, the higher correct registration is.

The goal of image registration is to get precise transformation from one image to another one. The main procedure is from Coarse to Fine[15,16,19]. Recently, a lot of methods [12-20] of image registration have existed and most of these are focused on feature point such as SIFT [6], Harris [7], SURF [8] points. Harris corner is invariant to rotation, noise and illumination changes of image. But it is not invariant for scale changes. Compared with Harris, SIFT points remedy the drawback of Harris. Thus it’s robust to projective transformation of image. And SURF point is the improvement of SIFT. The dimension of SURF descriptor is half of SIFT, but the precision is reduced. Accordingly considering of precision, we choose SIFT points to get initial matching. However, we find that some SIFT corresponding points can’t get absolute correspondences even if they are matched in vision. Actually the difference may be about 1-2 pixel. After computing transformation matrix, the image that obtained by the transformation can’t coincide with another absolutely, the ghost will appear as showed in Figure 1(a):

![Figure 1(a). The superposition from two images.](a)

(b). Four point pairs used to compute transform matrix

In Figure 1(b), we get transformation by the 4 corresponding points. Then we get the transformed image and correspond to another, as show in Figure 1(a). Hence, this paper will compute OCD of local geometry constructed by 4 corresponding points respectively and we regard OCD as similarity measure. By several iterations, we compute the best transformation matrix $T_{opt}$ through the points that possess the minimum of OCD.

Tal Hassner[9] had studied SIFT algorithm and its scale theory and proposed the concept of SIFT subspace. They constructed basis vectors spanning the scale space of descriptors and got the scale-less SIFT descriptors ultimately. This theory combines the descriptors under different scales, so it ascends the precision of descriptors. We will apply this theory to our feature point descriptor in order to get high precision before initial matching.

© 2014 ACADEMY PUBLISHER
doi:10.4304/jsw.9.4.964-969
II. MULTI-SCALE FEATURE DETECTION AND INITIAL MATCHING

A. SIFT Point Detection

The SIFT algorithm first constructs Gaussian pyramid. For example, we intend to construct $O$ octaves, $S$ levels images and appoint scale value $\delta$ for each image defined as below:

$$\delta(a,s) = 2^{a-l} \times k^s \times \delta_0$$

Where $o$ and $s$ represent $oth$ and $sth$ image. $k = 2^{1/5}$, $\delta_0$ means initial scale value.

Simply, the scale space $L(x,y,\delta)$ of image $I(x,y)$ will be constructed:

$$L(x,y,\delta) = G(x,y,\delta) \ast I(x,y)$$

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

Then, Gaussian DOG images can be obtained by DOG operator:

$$D(x,y,\delta) = [G(x,y,k\delta) - G(x,y,\delta)] \ast I(x,y)$$

From Gaussian DOG images, we can see the changes of pixels. No changes, no features. The points that change heavily may be features. DOG images portray the contour of object. Feature points are made up of 3-D local extreme points of DOG images. After detecting extreme points we can’t regard them as SIFT points until the heavy edge corresponding is eliminated.

B. SIFT Subspace And Feature Descriptor

SIFT descriptor can be obtained by gradient histogram of $16 \times 16$ around the feature point. As show in Figure 2, first we compute the main orientation of feature point and then we divide $16 \times 16$ into 16 small regions with the size $4 \times 4$. Next, we make a statistics about 8 gradient histogram of every $4 \times 4$ window and obtain the descriptor vector whose dimension is $4 \times 4 \times 8 = 128$. The descriptor describes the change of gradient orientation around the feature point. Therefore it is robust to delight, scale and rotation. Due to the elimination of edge corresponding, it is also robust to noise.

C. Initial Matching Of Feature Points

After the detection and description of features, we utilize Euclidean distance to get initial matching pairs. Let $X$ and $Y$ be SIFT point set to be matched. We get initial matching by steps:

1) For one point $i_x$ of $X$, we compute the distance $dist(x_i,y_j)$ between $x_i$ and $y_j$ of $Y$;

2) The minimum and second minimum distance $nD$ and $mD$ are chosen to compute the ratio: $\text{ratio} = \frac{nD}{mD}$;

3) We set a ratio threshold $T_{ratio}$. If $\text{ratio} < T_{ratio}$, $x_i$ and $y_j$ are matched. Otherwise, stop and compute next point $x_{i+1}$, return step 1);

4) Repeat the above steps until all points of $X$ are computed.

However, there are many wrong pairs in the corresponding points. Many factors account for it such as the same or similar object in an image. In spite of different object, the feature point descriptors may be similar. These outliers emerge mainly without considering the external figure structure of the point. Hence, this paper will take it into consideration by constructing local geometry and getting their similarity. The high similarity geometry is, the more correct the
point pair is.

III. LOCAL GEOMETRY SIMILARITY MEASURE

A. Geometry Similarity Measure

By the analysis of last part in Section II, we know the projective transformation matrix \( T \) computed by incorrect point pairs is incorrect obviously. In order to get precise matrix \( T \), we choose the most similar local points pairs that defined as:

\[
T_{\text{opt}} = \arg \max S(T(I_R), I_M)
\]

Where \( S(T(I_R), I_M) \) represents the similarity of image region \( I_R \) transformed and matched image \( I_M \).

Usually, there are a lot of methods to describe the similarity of two image regions. Jeongtae \[10\] proposed a robust correlation coefficients to describe it. He chose a non-negative density function \( f \) and weight function \( w \) and computed mean and covariance matrix and achieved the correlation coefficient \( \rho \). Then the best transform \( T \) will be obtained through maximum correlation coefficient \( \rho \). Often the methods that compute correlation coefficients focus on Sample Correlation Coefficient (SCC) and Maximum Likelihood Estimates (MLE).

SCC computes statistics directly by samples, defined as follow:

\[
\hat{\rho}(X,Y) = \frac{\hat{C}(X,Y)}{\sqrt{\hat{\sigma}_X \hat{\sigma}_Y}}
\]

Where sample means, sample variances, and sample covariance are defined:

\[
\hat{C}(X,Y) = \frac{1}{N} \sum_{i=1}^{N} (X_i - \bar{X})(Y_i - \bar{Y})
\]

\[
\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i
\]

\[
\bar{Y} = \frac{1}{N} \sum_{i=1}^{N} Y_i
\]

\[
\hat{\sigma}_X^2 = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})^2
\]

\[
\hat{\sigma}_Y^2 = \frac{1}{N-1} \sum_{i=1}^{N} (Y_i - \bar{Y})^2
\]

The process of MLE is: 1) Define correlation coefficient of \( X \) and \( Y \) as \( \rho \) whose joint distribution function is \( f_{XY}(x,y) \); 2) Compute maximum likelihood estimates according to sample parameter; 3) Get correlation coefficient through maximum likelihood estimates. Usually the result will be reached by lots of iterations.

Compared with MLE, SCC is simple and speedy. While an adequate distribution function should be chosen and many iterations are needed in SCC. However, the correlation coefficient obtained by SCC is not unbiased, even for normal distributions, it is asymptotically unbiased in that case. Therefore, it is not robust to incorrect point pairs. At the same time, it is also time-consuming.

B. Opposite Core Difference (OCD)

Based on above analysis, this paper proposes a method that can obtain best similarity by local geometry. We will describe similarity by OCD of local geometry.

Let \( T_0 \) be the transformation of \( P \), it means \( Y_i = T_0(X_i) \), thereby the cores satisfy \( T_0(\text{core}_X) \). The local regions of two images are shown in Figure 4, let the corresponding points be \( P = \{ P_1, P_2, P_3, P_4 \} \) \( (P_i = (X_i, Y_i)) \), the cores are shown below:

![Figure 4. The quadrangles constructed by four points of two images and their cores.](image)

The property called image core invariance[11]. Based on it, we define the change of cores as similarity measure:

\[
S(\text{core}_X, \text{core}_Y) = f_0(||T_0(\text{core}_X) - \text{core}_Y||)
\]

Where \( \text{core}_X \) and \( \text{core}_Y \) are cores of two images. \( f_0 \) is a non-negative monotonically increasing function. Generally exponential function whose rate of change is obvious is chosen in order to outstanding the change value.

Finally, the best transformation is got by minimizing above expression. The pseudo-code of this section described as follow:

// \( P \) \( \rightarrow \) \( \{P_1, P_2, P_3, P_4\} \) is corresponding point
// set and they are not collineation.
// \( \text{Num} \) : The total number of different choice of P.
//\( T_{\text{Iteration}} \) : Iteration threshold.
\( i = 0; \)
\( \text{Init}(S_{\text{min}}); \)

© 2014 ACADEMY PUBLISHER
while(i < min(T_{\text{max}}, Num))
{
    T_i = \text{GetTransform}(P);
    core_x = \text{GetCore}(X);
    core_y = \text{GetCore}(Y);
    S(core_x, core_y) = f(||T_i(core_x) - core_y||);
    if(S(core_x, core_y) < S_{\text{min}})
    {
        S = S(core_x, core_y);
        T_{\text{opt}} = T_i;
    } 
    i = i + 1;
    \text{Update}(P);
}

After all, the concrete steps of this paper are:
1) Feature Detection. The feature point sets $P_1$ and $P_2$ of image $I_R$ and $I_M$ are obtained from SIFT algorithm;
2) Multi-Scale Feature Description. Apply SIFT subspace theory to every SIFT point $p_{i1}(p_{i1} \in P_1)$ and $p_{i2}(p_{i2} \in P_2)$ to get precise descriptors.
3) Initial Matching. Get initial matching by Section2.3 and initial point pairs $P(P = (P_{i1}, P_{i2}))$;
4) Get Points That Possess Highest Similarity. Apply Section3.2 to 4 points that possess the highest similarity;
5) Compute Transformation Matrix $T_{\text{opt}}$ by these 4 points.

IV. EXPERIMENTS AND VERIFICATION

The source images of the paper show as follow:

SourceImage1 and SourceImage2 are the same scene that is got from different view.

We apply Matlab to prove that. In the experiment, after first iteration, the four corresponding points show in Figure 6:

From TABLE 1, the Euclidean distance is 7.2596 after first iteration. Although they are matched in vision, but the difference exists between the cores. The incorrect matching of one or more points of 4 points can account for it. So more iterations are needed to obtain best transformation $T_{\text{opt}}$. The number of iteration and their distance are:

The projective matrix when the OCD is 0.3173:

$$
T_{\text{opt}} = 
\begin{bmatrix}
1.1100 & 0.0452 & 7.6542 \\
-0.1150 & 0.9996 & -9.0902 \\
12.4259 & -32.9569 & 1.0362 \\
\end{bmatrix}
$$

Therefore, the final superposition image is shown in Figure 7:

Compared with Figure 1(a), the matching degree is enhanced conspicuously. However, the part of cock’ header is smooth in vision. The interpolation algorithm of Matlab may account for it.

Besides, another experiment results have verified the validity of our algorithm in Figure 8.
V. CONCLUSION

This paper proposes a simple and practical method that computes the transformation matrix between images. Core invariance is applied in image matching. In order to get precise descriptor, we get feature descriptor in different scales. In this paper, we chose 4 corresponding points at random and have not studied how to choose geometry points to improve speed. In theory, the quadrilateral is neither too large nor too small and more explorations will be taken.

ACKNOWLEDGEMENTS

This paper is supported by Nature Science Foundation of China (61272216) and innovation Foundation of aerospace in China (CASC201102) and Nature Science Foundation of Jiangxi Province(20114BAB201035) and Science technology cooperation of Jiangxi Province(2010EAB01700).

REFERENCES


© 2014 ACADEMY PUBLISHER
Zetao Jiang  Male, Born in 1961 in Jiangxi Province, professor, doctor. He is adjunct professor of Guilin University of Electronic Technology, PhD supervisor, academic leader in Guangxi province. He received the B.S. degree in Beijing Normal University, Beijing, China, in 1986. He received the M.S. degree in Tongji University, Shanghai, China, in 1995 and received the Doctor of Philosophy in Northwestern Polytechnical University, Xi’an, China, in 2006. Main activities were research in the field of image processing and computer vision.

He presided key technology research project “Control system based on multi-level security agent”, which was awarded the second class prize in the 2009 annual science and technology progress of Jiangxi province. He presided the research of network and information security defense technology based on digital watermarking, which was awarded the first class prize in science and technology achievements of Jiangxi province. He presided and finished Nature Science Foundation of China(60673055), two subject Foundations of province, an open foundation of measurement and control center of province and more than 10 other subjects. Currently, he is presiding Nature Science Foundation of China(60673055) and Nature Science Foundation of Province, which is in the field of 3D reconstruction.

Zhou Le  Male, Born in 1988 in Jiangsu Province, and received the B.S degree in Yancheng Teachers’ College, Yancheng,China. Currently, he is pursuing the M.S degree in Nanchang Hangkong University, and the main research area is image processing and computer vision.

Zhang Liwen  Female, Born in 1989 in Jiangxi Province, and had received the B.S degree in Nanchang Hangkong University. Currently, she is also pursuing the M.S degree in Nanchang Hangkong University, and her research area is image processing and computer vision.