ORIENTING DIGITAL STEREOPAIRS BY MATCHING FOURIER DESCRIPTORS

Yi-Hsing Tseng

Department. of Surveying Engineering
National Cheng Kung University, Taiwan, R.O.C.

Commission III

KEY WORDS: Automation, Orientation, Feature, Matching, Neural, Network, System, Design

ABSTRACT:

This paper presents a fully automatic method to reconstruct the relative orientation of a pair of stereo images. The operation of this method can be divided into two stages. First, by using the conception of feature-based matching, the Fourier descriptors of conjugate boundaries of homogeneous regions between the images are determined through a neural network system. This provides a reliable solution of approximate orientation. Second, the more accurate conjugate points are matched by using template based on the initial orientation. The relative orientation then is determined by using the photo coordinates of the matched conjugate points.

1. INTRODUCTION

The idea of fully automatic relative orientation has been proposed by Schenk, Li and Toth [1991]. They suggested to solve the initial orientation by matching edges in ψ-s domain. Then accurate orientation can be determined by using template matching, such as cross-correlation or least-squares matching. They also indicated that matching entire edges as opposed to more traditional point matching methods substantially increases the robustness of the solution. However, they also commented that it is more difficult to implement and there is still room for improvement.

Based alone on the factor of shape similarity in matching features, it tends to obtain wrong matches, if some features in an image are similar in shape. This problem can only be solved by taking the account of orientation consistency of matched features. Based on the idea of feature-based matching and considering the criterion of orientation consistency, a neural network system is implemented to obtain the optimal matching of the conjugate features of the stereo images.

The automatic process of relative orientation can be divided into two tasks (Fig. 1). The first task performs feature-based matching to solve the orientation approximately. The second task performs point matching to determine conjugate points accurately. The relative orientation then can be computed by using the photo coordinates of the matched conjugate points.

The proposed system for feature-based matching conceptually mimics the human recognition process. Conjugate features are determined not only by the condition of shape similarity but also by the fitness of orientation consistency. First, homogeneous regions are extracted from images by using region-growing method. Then, their boundaries are described by using Fourier descriptors. By applying the least-squares approach to matching Fourier descriptors [Tseng and Schenk, 1992], one can compute the shape similarity and relative orientation of matching features. All matching conditions will be combined into a cost function which can be applied by the Hopfield-Tank neural networks in determining the optimal matching. No initial values of orientation between images are required and it is expected to be adaptive to the disturbances of image distortion and noises as well as the differences of image orientation and scale.

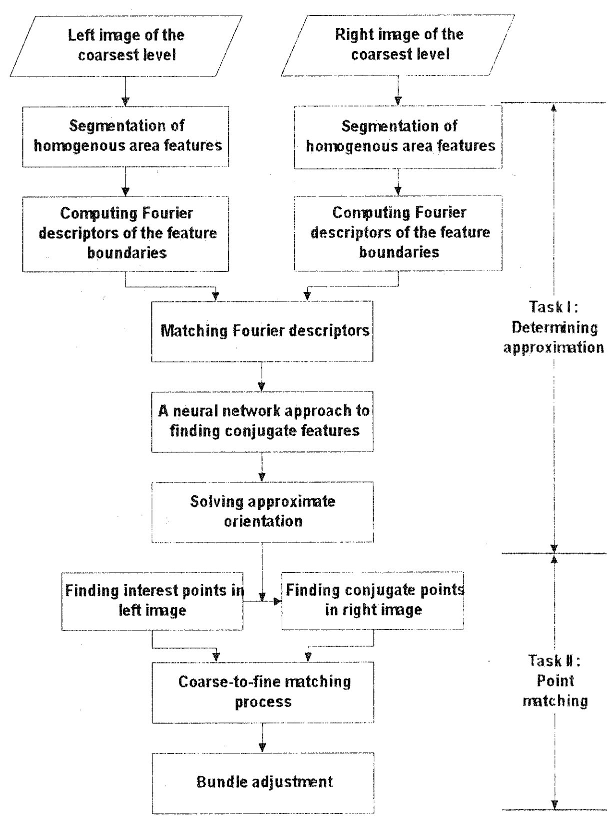

Figure 1: The work flow of the system

For the task of point matching, interest points are selected on the left image. Then, the initial conjugate locations on the right image can be obtained based on the approximate orientation. More accurate conjugate locations then can be matched by using template matching techniques. This procedure can be speeded up by using coarse-to-fine matching when images are stored in multi-resolution format.

2. MATCHING FEATURES

Boundary features can be modeled with Elliptic Fourier descriptors [Lin and Hwang, 1987] [Zahn and Roskies, 1972]. The best fit of shapes between features can be obtained by matching their Fourier descriptors [Tseng and Schenk, 1992]. When the best fit of features is obtained, the shape difference (the mean square difference between two fitted features) can be calculated as well.

2.1 Fourier Descriptors

A two-dimensional closed line (Fig. 2a) can be expressed by two parametric functions as Eq. (1).

$$x = f_x(t) \qquad \text{and} \qquad y = f_y(t) \tag{1}$$

In which, the t is defined as a time period from 0 to 2π. It means tracing a closed line one cycle. The functions become periodic if a line is traced repeatedly. Fig. 2b shows the periodic functions of $x(t)$ and $y(t)$.

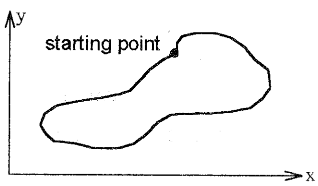

Figure 2a: A closed line feature.

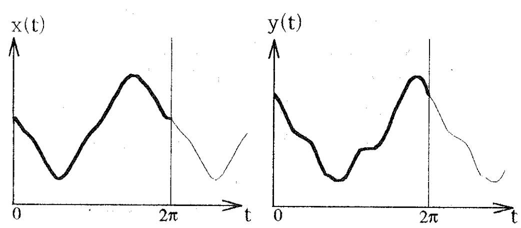

Figure 2b: The periodic forms of x(t) and y(t).

The periodic functions can be transformed into the frequency domain and expressed as Fourier series:

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} a_0 \\ c_0 \end{bmatrix} + \sum_{k=1}^{\infty} \begin{bmatrix} a_k & b_k \\ c_k & d_k \end{bmatrix} \begin{bmatrix} \cos kt \\ \sin kt \end{bmatrix} \tag{2}$$

where

$$a_0 = \frac{1}{2\pi}\int_0^{2\pi} x(t)\,dt; \qquad c_0 = \frac{1}{2\pi}\int_0^{2\pi} y(t)\,dt;$$

$$a_k = \frac{1}{\pi}\int_0^{2\pi} x(t)\cos kt\,dt; \qquad b_k = \frac{1}{\pi}\int_0^{2\pi} x(t)\sin kt\,dt;$$

$$c_k = \frac{1}{\pi}\int_0^{2\pi} y(t)\cos kt\,dt; \qquad d_k = \frac{1}{\pi}\int_0^{2\pi} y(t)\sin kt\,dt$$

In which a_k, b_k, c_k, and d_k are Fourier descriptors of the kth harmonic. The descriptors of 0th harmonic, a_0 and c_0, represent the centroid of the line feature.

2.2 Shape similarity

It is well known that a linear transformation in the spatial domain can be easily modeled in the frequency domain [Dougherty and Giardina, 1988]. As a centroid-based similarity transformation in the spatial domain is expressed as follows:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = S \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x - x_c \\ y - y_c \end{bmatrix} + \begin{bmatrix} x_c \\ y_c \end{bmatrix} + \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} \tag{3}$$

Where (x_c, y_c) are the coordinates of the centroid; Δx and Δy represent the translation of the centroid; and S and θ are scaling and rotation factors. The translation of the centroid can be modeled by using the 0th harmonic of Fourier descriptors as Eq. (4), and the transformation of scaling, rotation and phase shift (shift of starting point), Δt, can also be modeled in the frequency domain as Eq. (5).

$$\begin{bmatrix} a'_0 \\ c'_0 \end{bmatrix} = \begin{bmatrix} a_0 \\ c_0 \end{bmatrix} + \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} \tag{4}$$

$$\begin{bmatrix} a'_k & b'_k \\ c'_k & d'_k \end{bmatrix} = S \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} a_k & b_k \\ c_k & d_k \end{bmatrix} \begin{bmatrix} \cos k\Delta t & -\sin k\Delta t \\ \sin k\Delta t & \cos k\Delta t \end{bmatrix} \tag{5}$$

When two features are expressed by using Fourier descriptors, one can use Eq. (4) to calculate the location difference of their centroids. Computing the differences of scaling, rotation and phase shift is more complicated. First, we need to reorganize Eq. (5) to be:

$$\begin{bmatrix} a'_k \\ b'_k \\ c'_k \\ d'_k \end{bmatrix} + \begin{bmatrix} v_{a'_k} \\ v_{b'_k} \\ v_{c'_k} \\ v_{d'_k} \end{bmatrix} = S \begin{bmatrix} a_k & b_k & -c_k & -d_k \\ b_k & -a_k & -d_k & c_k \\ c_k & d_k & a_k & b_k \\ d_k & -c_k & b_k & -a_k \end{bmatrix} \cdot \begin{bmatrix} \cos\theta\cos k\Delta t \\ \cos\theta\sin k\Delta t \\ \sin\theta\cos k\Delta t \\ \sin\theta\sin k\Delta t \end{bmatrix} \tag{6}$$

In which, $v_{a'_k}, v_{b'_k}, v_{c'_k}$, and $v_{d'_k}$ are the residuals of descriptors after transformation. They represent the shape differences between features. Based on the principle of least-squares solution, the transformation to obtain the best fit between features can be solved by minimizing the sum of the squares of residual.

A unique expression of the shape difference can be formed by the residuals of matching Fourier descriptors. It is the mean square difference (MSD), or say the average discrepancy, between fitted features. MSD can be computed as:

$$MSD = \frac{1}{2}\sum_{k=1}^{m} (v_{a'_k}^2 + v_{b'_k}^2 + v_{c'_k}^2 + v_{d'_k}^2) \tag{7}$$

Obviously, MSD can be used as an evaluation of the shape similarity between features. Therefore, it is used as the first criterion to find conjugate features in a pair of stereo images.

2.3 Orientation consistency

Based alone on shape similarity, it tends to fail in recognizing conjugate features when there are some similar features in an image. Under this circumstance, the condition of orientation consistency is proposed to reinforce the recognition condition. It means that the relative orientation of each pair of conjugate features should be consistent. In other words, if two features are conjugate, they not only are similar in shape but also have an orientation relation similar to the other pairs of conjugate features.

The relative orientation can be computed by using Eq. (4) and (6). It can be expressed as a similarity transformation to obtain the best fit of two features. When we calculate the MSDs of all possible matches of features, their relative orientation can be computed simultaneously.

3. A NEURAL NETWORK SOLUTION

In order to generate a flexible and powerful system for recognizing conjugate features, an artificial neural system is developed to mimic the human brain. This system is functioned based on a cost function modeled according to the shape similarity and orientation consistency. Because matching conjugate features is an optimization problem, the model of Hopfield-Tank neural network [Hopfield and Tank, 1985] is adopted.

3.1 The cost function

Using the method presented in the last section, on can evaluate the suitability of each pair of possible matches of features. However, it is necessary to combine all the evaluations into a function as a system to evaluate the overall fitness between images. This function, called a cost function, should provide a reasonable estimation of suitability to the matching status of conjugate features.

It is necessary to find a mathematical expression to present the matching status of two feature groups in order to formulate a cost function. A two-dimensional array, V, can be constructed to present all of the possible matches. If there are m features in the first group and n features in the second group, then an m by n matrix should be established. Each element in V is assigned a binary value 1 or 0 to represent one possible conjugate match. If an element, $_{ij}$, is assigned as 1, it means the ith feature in the first group and jth feature in the second group are conjugate features. On the contrary, 0 means they are not conjugate. Fig. 3 shows an example of a 5 by 6 array. Four conjugate matches, a to 1, b to 3, c to 2, and e to 4, can be identified in the array.

Since the cost function reflects the total cost of conjugate matches between two feature groups, it should be composed of factors which are thought as indices to judge the aptness of a matching status. Three factors are considered to be included in the cost function. They are shape similarity, orientation consistency, and the constraint of one-to-one match. We will discuss how each factors can be formulated respectively in the following paragraphs.

	1	2	3	4	5	6
a	1	0	0	0	0	0
b	0	0	1	0	0	0
c	0	1	0	0	0	0
d	0	0	0	0	0	0
e	0	0	0	1	0	0

Figure 3: An example of V matrix to represent a matching status.

Shape similarity is the most important factor to identify conjugate features. By matching Fourier descriptors as described in section 2.2, we can quantify the shape similarity between all possible pairs of conjugate features. For each possible pair, an MSD value can be calculated to judge the fitness of their shapes. When we want to assign a feature pair which MSD is larger than a predefined threshold as a pair of conjugate features, the cost would be increased. In this principle, the cost of shape similarity in a given status can be formulated as follows:

$$E1 = \sum_{i=1}^{m}\sum_{j=1}^{n} f\left(MSD_{ij} - t_{MSD}\right)\cdot V_{ij} = \sum_{i=1}^{m}\sum_{j=1}^{n} \theta_{ij}^{MSD}\cdot V_{ij} \quad (8)$$

In which, θ_{ij}^{MSD} denotes the cost of shape similarity and $f(x)$ is a unipolar binary function defined as $f(x) = 1$, when $x > 0$, and $f(x) = 0$, when $x <= 0$. The symbol t_{MSD} is a threshold number which can be adopted subject to the disturbance of image distortion and noises.

As described in section 2.3, the consideration of orientation consistency is required. In order to model the orientation consistency, the binary relationships of two matching pairs are examined. The idea is: if feature i and k in the left image is matched to feature j and l in the right image, then the relative orientations between i,k and j,l should be consistent. For the implementation, we can check the consistency of θ_{ij} (relative rotation between feature i and j), θ_{kl}, and $\theta_{ik,jl}$ (relative rotation between vector ik and jl as Fig. 4) as well as the consistency of S_{ij} (relative scale between feature i and j), S_{kl}, and $S_{ik,jl}$ (relative scale between vector ik and jl). This binary relationship can be modeled as:

$$E2 = \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{m}\sum_{l=1}^{n} V_{ij}\cdot W_{ijkl}^{OC}\cdot V_{kl} \quad (9)$$

In which, W_{ijkl}^{OC} denotes the cost of orientation consistency. If the consistency check is passed W_{ijkl}^{OC} is set to -1 or it is set to 1.

The constraint of one-to-one match should be taken into account. This is due to the fact that an object would only form one feature on an image. This constraint can be formulated as:

$$E3=\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{m}\sum_{l=1}^{n}V_{ij}\cdot W_{ijkl}^{col}\cdot V_{kl}$$

$$E4=\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{m}\sum_{l=1}^{n}V_{ij}\cdot W_{ijkl}^{raw}\cdot V_{kl} \qquad (10)$$

in which

$$W_{ijkl}^{col}=\begin{bmatrix}1\ ,\ k\neq i \text{ and } l=j\\ 0\ ,\ \text{others}\end{bmatrix};\quad W_{ijkl}^{raw}=\begin{bmatrix}1\ ,\ l\neq j \text{ and } k=i\\ 0\ ,\ \text{others}\end{bmatrix}$$

The cost function is formed by summing up all the factors mentioned above. One can also adjust weights of the factors by multiplying each factor by a price coefficient. By doing this, the system can be tuned subject to the needs of various applications. Therefore, the cost function can be expressed as follows:

$$E=C_1\cdot E1+C_2\cdot E2+C_3\cdot E3+C_4\cdot E4 \qquad (11)$$

In which, C_1, C_2, C_3, and C_4 are price coefficients adopted to adjust the weight of each factor.

3.2 The model of Hopfield-Tank neural network

The Hopfield-Tank neural networks are commonly used to solve an optimization problem. They are sort of feedback-style nonlinear networks. In every iteration cycle, neurons will get signals transmitted from all the neurons (including themselves) of the previous stage and change their status to decrease the system cost (or called energy). The networks will keep running until a stable condition (the lowest cost) is reached. The networks are ideally suitable for the problem of finding the optimal match of conjugate features between two feature groups.

For the application of matching conjugate features of a stereopair, the neurons should be arranged as a 2D array. If there are m and n features in the left and right image respectively, the neuron array can be shown as in Fig. 4. Values of the neurons will be 0 or 1 which represent a negative or a positive conjugation respectively. Each neuron is connected with all the other neurons due to the fact that each conjugate relation may affect the status of the other conjugate relations.

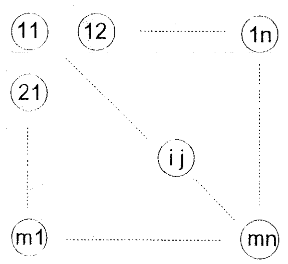

Figure 4: A 2D neuron array.

According to the theory of the Hopfield-Tank model, if a cost function can be transformed to a special function called the Liapunov function, it can drive the neural system to the lowest cost status [Freeman and Skapura, 1992]. The Liapunov function for a 2D neuron array [Hopfield and Tank, 1985] can be expressed as:

$$E=-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{m}\sum_{l=1}^{n}V_{ij}\cdot W_{ijkl}\cdot V_{kl}+\sum_{i=1}^{m}\sum_{j=1}^{n}\theta_{ij}\cdot V_{ij} \qquad (12)$$

Where V denotes the status of the neural array; W is the connection strength between neurons; and θ represents a constant input.

The system operates the neurons one by one. In each operation, it examines the net gain of the cost if the status of the neuron status is altered. If the net gain is negative which means the value of the cost function can be lower by changing the status. This operation will be recursively performed on the neurons until there are no changes needed in a complete cycle. There are four cases may happen:

(1) The current status is 0 and the net gain is negative, change the status to 1;
(2) The current status is 0 and the net gain is positive, the status remain unchanged;
(3) The current status is 1 and the net gain is negative, change the status to 0;
(4) The current status is 1, and the net gain is positive, the status remain unchanged.

By examining those cases, one can discover that the judgment can be simplified if a net value of the ith column and jth row neuron is computed as follows:

$$net_{ij}=-\sum_{k}\sum_{l}W_{ijkl}\cdot V_{kl}-\frac{1}{2}W_{ijij}+\theta_{ij} \qquad (13)$$

The judgment will become as easy as the follows:

if $net_{ij}>0$, set the output of $_{ij}$ to 1
if $net_{ij}<0$, set the output of $_{ij}$ to 0
if $net_{ij}=0$, $_{ij}$ remains unchanged

The initial state of the network can be arbitrary. The system promises to reach the lowest cost with any given initial state.

3.3 Matching conjugate features

In the section 3.1, we have defined the cost function for the problem of matching conjugate features. In order to apply the Hopfield-Tank neural network, the cost function should be transformed into the format of the Liapunov function. By comparing Eqs. (11) and (12), one can obtain the following relationships:

$$\begin{cases}W_{ijkl}=-2(c_2\cdot W_{ijkl}^{OC}+c_3\cdot W_{ijkl}^{col}+c_4\cdot W_{ijkl}^{raw})\\ \theta_{ij}=c_1\cdot\theta_{ij}^{MSD}\end{cases} \qquad (13)$$

By substituting Eq. (13) into Eq. (12), we can compute the net value of the cost to determine the output in the operation of each neuron. After the system reaches the lowest cost, the final status of the neuron array represents the conjugate relationship between the feature groups.

Adopting proper price coefficients for the cost function is of importance to obtain reliable results. As shown in Eq. (13), the price coefficients form the strength of connections between neurons and the constant input of each neuron. Conceptually, one can imagine them as the weights can be adjusted to adapt the system to various conditions of applications. For example, if we think the factor of shape similarity is much more important than the other factors, we can give a large value to C_1. However, although the concept is plausible, there are no rigorous rules to tune the system to be optimal.

The settings of the thresholds in Eqs. (8) and (9) are also important. They can be adjusted to fit the conditions of the applied images. The threshold of shape similarity is set for the disturbances of image distortion and noises, and the thresholds of orientation consistency is set for the geometric changes due to the different view angles of cameras.

4. DETERMINING CONJUGATE POINTS

After conjugate features are matched, the approximate relative orientation can be solved by using the coordinates centroids of conjugate features. This allows us to narrow down the search windows when we apply template matching to determine conjugate points. In this application, the normalized cross-correlation (NCC) is used to determine conjugate points up to one-pixel accuracy. Then sub-pixel accuracy is reached by using the least-squares matching (LSM)) [Ackermann, 1984].

In order to obtain reliable matches, each template should contain enough gray-level changes. Also evenly distributed conjugate points on the overlapped image area is required to obtain reliable relative orientation. These requirements can be achieved by using interest operator to locate locally most interest points as the template locations.

5. RESULTS

Several pairs of aerial stereo photographs have been tested on the system. Fig. 5 shows one of the pairs of the test photographs. The photos were digitized in the resolution of 600 dpi which is corresponding to the pixel size of 42.3μm square.

By using the technique of region growing, homogenous areas are segmented from each image. There are 23 features derived from the left image and 34 features derived from the right image (Fig. 6).

Conjugate features are determined by using the technique of matching Fourier descriptors described in section 2 and 3. After 5 iterations of neural network computation, the final status was reached and 10 pairs of conjugate features were discovered. The matched pairs are 4-4, 6-9, 8-11, 9-12, 11-15, 12-17, 19-22, 20-23, 21-24, and 22-28. By checking the images visually, we can discover that they are all correct matches.

By using interest operator and template matching techniques 150 conjugate points were discovered. After the computation of relative orientation, 16 conjugate pairs were eliminated due to their residuals of y parallax are three times larger than the mean square error. Finally 20μm of the root-mean-square error of y parallax is obtained. The overall computation time was about 30 minutes when it is executed in a Pentium 90 computer.

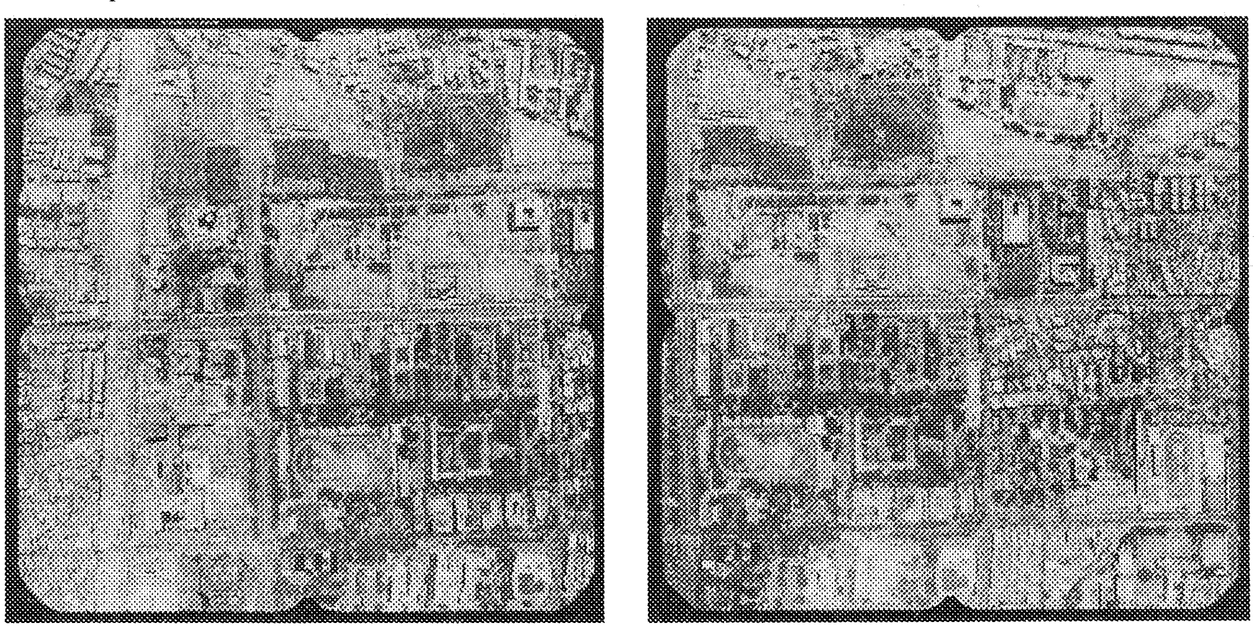

Figure 5: An example of test image pair

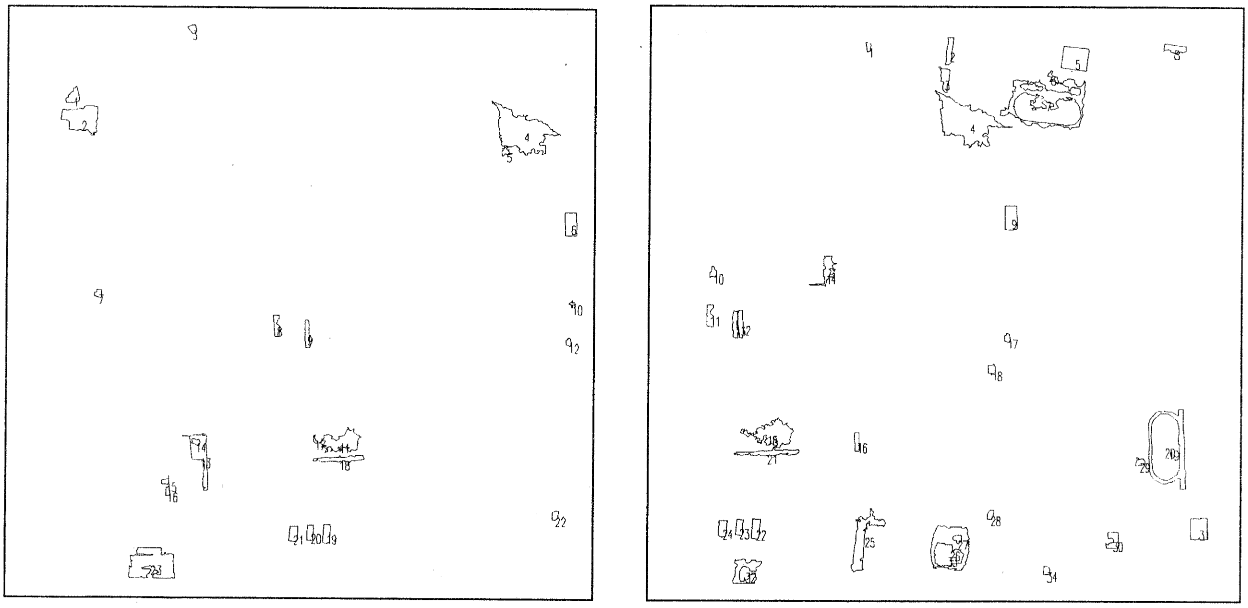

Figure 6: The boundaries of segmented homogenous areas from the test images

6. CONCLUSIONS

A fully automatic system is designed to orient digital stereopairs of images. This system solves the initial orientation by matching boundary features of homogenous regions. The merit of matching features is offering reliable initial orientation which provides the possibility of finding accurate conjugate positions by using template matching techniques.

For the matching of features, a neural network system is implemented to match Fourier descriptors of features with the consideration of shape similarity and orientation consistency. This system works well even if there are serveral similar features distributed in the images. It is also adaptive to the disturbances of image distortion and noises as well as the differences of orientation and scale between images. We expect this system will be useful in the fields of digital photogrammetry, pattern recognition and computer vision.

The sucessful experiments which we have tested on several stereopairs of varying scale and ground coverage encourage us to extend the system to orient a block of photographs for automatic aerial triangulation. The system should also be useful for the applications of close-range photogrammetry. Future experiments will be directed towards these two applications.

We expect that the method will work as long as the boundaries of homogeneous regions in both images are still similar in shape. Matched features should similar in whole shape. If conjugate features are only partially matched in shape, they will not be detected by the system. However, this does not degrade the reliability of the overall results.

ACKNOWLEDGMENTS

This research project was sponsored by the National Science Council of the Republic of China under the grant No. NSC84-2211-E006-045.

REFERENCES

Dougherty, E. R. and Giardina, C. R., 1988. *Mathematical Methods for Artificial Intelligence and Autonomous Systems*, Prentice Hall Inc.

Freeman, J. A. and Skapura, D. M., 1992. *Neural Networks Algorithms, Applications, and Programming Techniques*, Addison Wesley.

Hopfield, J. J. and Tank, D. W., 1985. Neural Computation of Decisions in Optimization Problems, *Biological Cybern*, 52, pp. 141-154.

Lin, C. S. and Hwang, C. L., 1987. New Forms of Shape Invariants from Elliptic Fourier Descriptors, *Pattern Recognition*, Vol. 20, No 5, pp. 535-545.

Schenk, T., Li, J. C. and Toth, C., 1991. Towards an Autonomous System for Orienting Digital Stereopairs, *Photogrammetry Engineering & Remote Sensing*, Vol. 57, No 8, pp. 1057-1064.

Tseng, Y. H. and Schenk T., 1992. A Least-Squares Approach to Matching Lines with Fourier Descriptors, *International Archives of Photogrammetry and Remote Sensing*, Vol. 29, Part B2, Commission III, pp. 469-475.

Zahn, C. T. and Roskies, R. Z.,1972. Fourier Descriptor for Plane Closed Curves, *IEEE Trans. on Computers*, Vol. 21, No. 3, pp. 269-281.