國 立 交 通 大 學

電機與控制工程學系

碩 士 論 文

智慧型立體對應式之視差偵測法則設計

Intelligent-Stereo-Matching-Based
Disparity Detection Algorithm Design

研 究 生：何承育

指導教授：陳永平　教授

中 華 民 國 九 十 八 年 六 月

智慧型立體對應式之視差偵測法則設計
Intelligent-Stereo-Matching-Based Disparity Detection Algorithm Design

研 究 生：何承育　　　　　Student：Cheng-Yu Ho

指導教授：陳永平　　　　　Advisor：Professor Yon-Ping Chen

國 立 交 通 大 學
電機與控制工程學系
碩 士 論 文

A Thesis
Submitted to Department of Electrical and Control Engineering
College of Electrical Engineering and Computer Science
National Chiao Tung University
In Partial Fulfillment of the Requirements
For the degree of Master
In
Electrical and Control Engineering
June 2009

Hsinchu, Taiwan, Republic of China

中華民國九十八年六月

智慧型立體對應式之
視差偵測法則設計

學生 ： 何承育　　　　　　　指導教授: 陳永平 教授

國立交通大學電機與控制工程研究所

摘　要

在本篇論文中，提出處理一對左右影像特徵點立體對應問題的演算法來偵測視差，而特徵點是使用 Harris 角落偵測器所抓取而得到的。此立體對應問題可被系統化為一個最佳化問題的目標函數，該函數代表問題答案的限制和特性，而且此目標函數可轉換成二維 Hopfield 神經網路的能量函數去做最小化處理。Hopfield 神經網路是一個單層回授網路，每個神經元代表左圖與右圖各一點的對應關係，當網路裡所有的神經元輸出皆不再改變時，也就是網路達到穩定狀態，此時可獲得對應結果和對應配對之視差。本研究比較了三種不同之目標函數的效能，並用一個簡單的應用來呈現視差的偵測。

Intelligent-Stereo-Matching-Based Disparity Detection Algorithm Design

Student：Cheng-Yu Ho   Advisor：Prof. Yon-Ping Chen

Department of Electrical and Control Engineering

National Chiao–Tung University

ABSTRACT

In this thesis, an algorithm is proposed to detect the disparity by solving the stereo matching problem for a set of feature points extracted by the Harris corner detector from a pair of stereo images. The stereo matching problem is formulated as an objective function which represents the constraints and properties on the solution. Then the objective function is transferred to the energy function of 2D discrete Hopfield neural network for minimization. This neural network is a single layer feedback network and each neuron in the network represents a possible match of two feature points, one in the left image and the other in the right image. The matching result and the disparity of the matched pairs are obtained when every output of neurons are no longer changed, i.e. the network reaches its stable status. Furthermore, the performances of three kinds of objective function are compared, and a simple application is presented to show the disparity detection.

Acknowledgement

　　本論文能順利完成，首先感謝指導老師　陳永平教授引領作者進入學術殿堂並孜孜不倦的指導，讓作者在研究方法及英文論文寫作上有著長足的進步，在為學處事的態度上亦有相當的成長，謹向老師致上最高的謝意；同時，感謝口試委員　楊谷洋老師以及　林昇甫老師提供寶貴意見，使得論文能臻於完整。

　　另外，感謝可變結構控制實驗室的吉佑、新光、楊庭以及學弟們對作者的照顧與陪伴，讓作者在實驗室的研究生活充滿溫馨與快樂。此外，感謝桓展、世宏學長平日在攻讀博士學位之餘，不吝傳授知識與經驗及給予建議。最後，感謝爸爸、媽媽與弟弟們在精神上的支持和生活上的照顧，適時地給予體諒與包容讓作者能在這兩年全心投入研究之中並擁有美好的碩士生活。

謹以此篇論文獻給所有關心我、照顧我的人

何承育　2009.6

Contents

List of Figures

List of Tables

Chapter 1
Introduction
1.1 Motivation

In robot systems, the robot vision is a popular research and has been applied in many applications such as navigation, detection and recognition. Recently, some investigators [1]–[3] have paid their attentions to the stereo vision of robot, which have two cameras to build the stereo vision containing 3-D information. For a humanoid vision system (HVS) which simulates the motion of human eyes, the stereo vision is required to obtain the depth information. The HVS consists of two cameras so that two images of the same scene can be taken at the same time by the right camera and the left camera from two different perspectives, and the passive stereo vision can be implemented by this equipment. Passive stereo vision, capturing images by using two or more cameras, is a well-known technique [4], [5] to determine the depth of an object or a point in a scene and generally based on the so-called disparity to obtain 3-D depth information of objects. The disparity is the difference between the locations of the same physical point in the image pairs and can be used in the triangulation principle [6] to obtain 3-D depth information. In the procedure of obtaining the disparity, the most difficult and time consuming problem is to determine a feature extracted from one image that matches to a given feature extracted from the other image. This is called the correspondence problem or stereo matching problem. This thesis focuses on the stereo matching problem and proposes an algorithm based on a neural network structure to deal with this problem.

1.2 Related Work

After reviewing some researches about the stereo matching problem, their approaches can be divided into two types: area-based [7-9] and feature-based [10-13] methods. For the area-based method, it often calculates the correlation between two images by using a fixed-size image window. It has been known that the resulted matched pairs are enormous. For the feature-based method, it applies the feature extraction and searches the major elements in an image related to the features, such as line segments, corners, or contours. It is evident that the number of matched pairs to be processed of the feature-based method is much less than that of the area-based method. In general, the stereo analysis using the feature-based method follows four steps: image acquisition, feature extraction, image matching, and depth determination.

Recently, many researchers have proposed to turn the matching problem into an optimization task by using Hopfield neural networks (HNN). The advantage of using the neural network technique is the matching problem can be formulated as minimization of an objective function where all the constraints on the solution can explicitly be included in the objective function. The Hopfield neural networks were first proposed by Hopfield and Tank [14], and they applied the "Energy function" idea to the stability of the neural networks. In many HNN's applications of optimization, an energy function is first constructed to represent the constraints on the solution. The HNN is used to minimize the energy function by adjusting the state of neurons so that the result can be the best answer to the problem. Obviously, the energy function is the connection between the optimization task and the HNN, so how to design the energy function is a key point of the algorithm. For the HNN which uses edge segment as the feature, G. Pajares, J. M. Cruz and J. Aranda [15] used similarity, smoothness and

uniqueness to establish the energy function of the analog HNN. Y. Ruichek [16] used the multilevel analog HNN for real-time obstacle detection and the uniqueness, ordering and smoothness constraints were used for matching. Sun [17] used the discrete HNN to minimize the energy function subject to the disparity, dummy and uniqueness constraints for solving a scanline-based stereo matching problem. For the HNN which uses point as feature, N. M. Nasrabadi and C. Y. Choo [18] proposed three constraints to establish the energy function based on physical observations: (i) uniqueness, each feature point from each image maybe assigned at most one disparity value. (ii) smoothness, if the points belong to the same object, then the disparity difference should be very small. (iii) geometric constraint, the difference between the distance separating two points in the left image and the distance separating their correspondents in the right one is small if the feature points are correctly matched. K. Achour and L. Mahiddine [19] added the similarity to the work of N. M. Nasrabadi and C. Y. Choo [18] and used an update rule to perform the network evolution. When the HNN reaches its stable state, each neuron represents a possible match between a left candidate and a right one.

This thesis will adopt the feature based method and follow the same steps. In the first step, a HVS is used for image acquisition to capture images in gray level format. Then since the dense depth map is not necessary, the number of pixels to match is highly reduced, which improves the efficiency of the algorithm. Therefore, the Harris corner detector which can find points of high interest was chosen as a feature extraction technique. By using this detector, the output value of pixels on the object's corners is large, and then through a selection to choose a fixed number of pixels whose output value is larger than a threshold as the feature points. However, it is not an easy task to find the corresponding feature points between two images, especially

when not all the feature points in one image are visible in the other image. To deal with the stereo matching, a HNN is proposed in this thesis. The proposed HNN follows the work of K. Achour and L. Mahiddine [19], but modifies the similarity and adds the vertical disparity in energy function. The result of the proposed HNN will compare with the algorithm [19] and be discussion in chapter 3 and 5. After the stereo matching, the disparity of all matched pairs can be easy derived. Finally, a simple application is proposed to obtain the depth information between object.

1.3 Thesis organization

The remainder of this thesis is organized as follows. The Harris corner detector which is used to extract the feature points is introduced in chapter 2. In chapter 3, the HNN is briefly described, and the properties of the energy function in 2D HNN are discussed and applied to deal with the stereo matching problem. In chapter 4, the real-coded genetic algorithm is used to automatic find the four parameters of HNN which will affect the performance of stereo matching. In chapter 5, the result of genetic algorithm and the results of three kinds of energy function in 2D HNN are shown and compared. A simple application to detect the relative depth information is also proposed here. Finally, the conclusions and future works will be proposed in chapter 6.

Chapter 2

Feature Extraction

When the stereo image pair is captured by two cameras, the first task is to find the feature points in the image pair. The feature points will be sent to the stereo matching process and affect the performance of the proposed algorithm, so those feature points must be the high interest and stable under local and global perturbations in the image domain. Generally, those feature points of high interest often occur at the corners and edges of the object in the image.

There are many techniques to detect corners or edges, and the Moravec operator is one of those techniques to find the points on the corners or edges, but it have some problems such as only a set of shifts at every 45 degree is considered and the response for the points of the edge are too strong. To deal with the two problems, the Harris corner detector [20], which modifies the Moravec operator, is adopted for extracting the feature points. Harris corner detector uses a method to calculate the local maximum variance of intensity in the image. It denotes the image intensities at the coordinate (x, y) by $I(x, y)$, and the change H produced by a shift (u, v) is given by

$$H(u,v) = \sum_{x,y} W(x,y)[I(x+u, y+v) - I(x,y)]^2 \qquad (2\text{-}1)$$

where $W(x,y)$ is a Gaussian window function defined as below:

$$W(x,y) = \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \qquad (2\text{-}2)$$

And the shifts, (u, v), are considered comprise [(1, 0), (1, 1), (0, 1), (-1, 1)]. To deal

with the problem that only a set of shifts at every 45 degree is considered, the Harris corner detector use Taylor's expansion truncated to the first order terms for small shifts, that is

$$I(x+u, y+v) \approx I(x, y) + I_x(x, y)u + I_y(x, y)v \tag{2-3}$$

where $I_x = \partial I / \partial x$ and $I_y = \partial I / \partial y$, so that the change of image intensities H can be

$$H(u, v) \cong \sum_{x,y} W(x, y)[I_x u + I_y v]^2 = [u, v]M\begin{bmatrix} u \\ v \end{bmatrix} \tag{2-4}$$

where the 2 x 2 symmetric matrix M is

$$M = \sum_{x,y} W(x, y)\begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \tag{2-5}$$

Note that H is closely related to the local autocorrelation function with matrix M describing its shape at the origin. Let α, β be the eigenvalues of matrix M, and the image points can be classified according to the eigenvalues. The classification contains three cases as below:

1. If both α, β are small, so that the change of intensities E is flat, then the windowed image region is of approximately constant intensity.

2. If one eigenvalue is high and another is low, so that the change of intensities E is ridge shaped, then only shifts along the ridge cause little change E. This indicates an edge.

3. If both eigenvalues are high, so that the change of intensities E is sharply peaked, then shifts in any direction will increase E. This indicates a corner.

To distinguish the response of corners and edges, a function R which only depends on

α and β is used to measure the response as below

$$R = \det(M) - k(\mathrm{Tr}(M))^2, \qquad k = 0.04 \sim 0.06 \qquad\qquad (2\text{-}6)$$

where k is a empirical constant between 0.04 and 0.06, and det(*M*) and Tr(*M*) are

$$\det(M) = \alpha\beta \qquad\qquad\qquad\qquad (2\text{-}7)$$

$$\mathrm{Tr}(M) = \alpha + \beta \qquad\qquad\qquad\qquad (2\text{-}8)$$

The contours of constant *R*1~*R*5 and the classification are shown in Fig. 2.1 where 0<*R*1<*R*2<*R*3 and 0>*R*4>*R*5. Clearly, *R* is positive with large magnitude for a corner, negative with large magnitude for an edge, and the absolute value of *R* is small for a flat region. Practically, the flat region is specified by *R* falling below a selected threshold. In this thesis, to select a pixel as a nominated corner pixel, a threshold *T* for *R* and a limitation [N_{min}, N_{max}] for the number of the feture point are applied. The extraction process is shown as below:

1. Evaluate the change of image intensity *H* in ROI, and then calculate the response *R*.

2. To select the pixels which are the local maximum and larger than a given threshold *T* as the feature points.

3. If the number of the feature points is less than N_{min}, then the threshold of *H* will be decreased by multipling 0.8, i.e. *T* will be 0.8*T*, and then go to step 2 until the number of .the feature points satisfies the limitation.
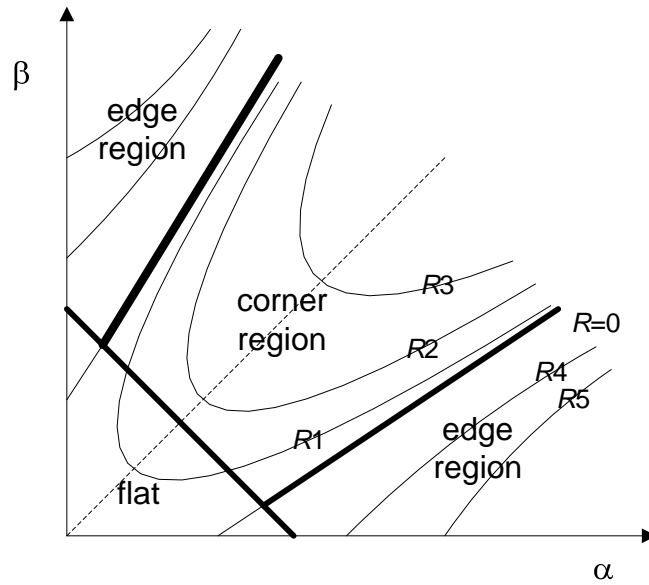
Fig. 2.1 Auto-correlation principal curvature space, heavy lines give corner/edge/flat

classification, and fine lines are equi-response contours.
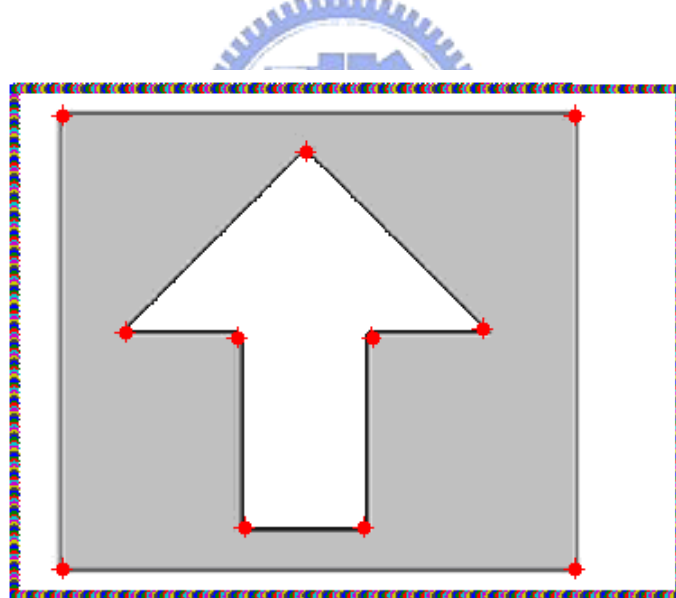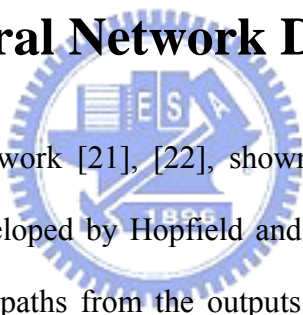
Fig. 2.2 Corners extracted by the Harris corner detector, and the yellow crosses

represent the feature points.

Chapter 3

Stereo Matching with the Hopfield Neural Network

In chapter 3, the Hopfield neural network is applied to deal with the stereo matching problem. At first, the Hopfield neural network will be introduced in the section 3.1, and then the section 3.2 contains three parts which illustrate how to use the Hopfield neural network to solve the stereo matching problem. Finally, the summary and the detail of matching process are presented in the section 3.3.

3.1 Hopfield Neural Network Description

The Hopfield neural network [21], [22], shown in Fig. 3.1, is a single layer feedback neural network developed by Hopfield and Tank. Clearly, it is a recurrent network containing feedback paths from the outputs of the neurons back into their inputs so that the response of such a network is dynamic. That means the recurrent network can perform a sequential updating process after applying an initial state. Starting from the initial state, the recurrent process is repeated again and again. Successive iterations produce smaller and smaller output change until reaching its equilibrium response.

For the recurrent process, at time step t the ith neuron in HNN has an output y_i^t and n inputs including one external constant input x_i and $n-1$ feedback signals of the form $w_{ij}y_j^t$, where $w_{ij}=w_{ji}$, $j=1, 2,\ldots, n$ and $j\neq i$. To ensure the convergence of the proposed network, the network weights are constant and symmetric and there is no self-feedback in HNN since $w_{ii}=0$. Let the total input to neuron i be denoted as u_i^t

which can be expressed as

$$u_i^t = \sum_{j=1, j\neq i}^{n} w_{ij} y_j^t + x_i \; . \tag{3-1}$$

In the recurrent process, the update rule for each neuron in a discrete HNN is

$$y_i^t = \frac{1 + \text{sgn}\left(\sum\limits_{j=1, j\neq i}^{n} w_{ij} y_j^{t-1} + x_i\right)}{2} = \frac{1 + \text{sgn}(u_i^{t-1})}{2}, i = 1, 2, ..., n. \tag{3-2}$$

where sgn(•) is the signum function, i.e. if $u_i^t > 0$, then $\text{sgn}(u_i^t)=1$, otherwise $\text{sgn}(u_i^t) = -1$. Clearly, all the outputs y_i^t are 0 or 1. The above update rule is applied in an asynchronous mode that each step time only one neuron is allowed to update its output. To describe the state of the network, its energy function is defined as

$$E^t = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} w_{ij} y_i^t y_j^t - \sum_{i=1}^{n} x_i y_i^t . \tag{3-3}$$

Since $y_i^t \in \{0, 1\}$, the energy is bounded as

$$-E_{\max} \leq E^t \leq E_{\max} \tag{3-4}$$

where the upper bound $E_{\max} = \frac{1}{2}\sum\limits_{i=1}^{n}\sum\limits_{j=1, j\neq i}^{n} |w_{ij}| + \sum\limits_{i=1}^{n} |x_i|$. If the output of neuron i is the one to be updated, then the change in energy is given by

$$\Delta E^t = E^t - E^{t-1} = -\left(\sum_{j=1, j\neq i}^{n} w_{ij} y_j^{t-1} + x_i\right)(y_i^t - y_i^{t-1}) = -(u_i^{t-1})\Delta y_i^t. \tag{3-5}$$

From (3-2), Δy_i^t is positive for $u_i^{t-1} > 0$ and nonpositive for $u_i^{t-1} \leq 0$, which implies ΔE^t is nonincreasing, i.e. $\Delta E^t \leq 0$. Furthermore, E^t is bounded in an *n*-dimensional space consisting of 2^n vertices formed a hypercube, so iterations of the update rule will lead

E^t to achieve the local minimum, one vertex of the hypercube.
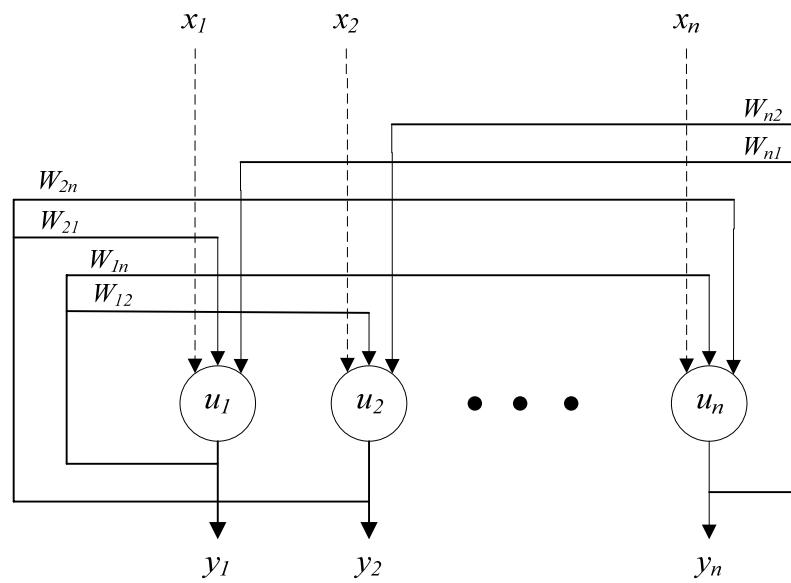

Fig. 3.1 The Hopfield neuron network with its neurons interconnection.

3.2 Applying HNN to the Stereo Matching Problem

To solve the stereo matching problem using the HNN, the first step is to decide the representation scheme which allows the outputs of the neurons to be interpreted as a solution of the problem, and the next step is to define an energy function whose minimum value corresponds to the best solution of the problem from the applicable properties. The third step is to derive the connecting weights and the external input value from the energy function. Finally, the HNN will evolve to get the solution after setting up the initial output of each neuron.

3.2.1 Interpretation of the neuron outputs

This section adopts the 2D HNN shown in Fig. 3.2 to find the correspondences of the feature points between the left and the right images, and this model consist of a set of neurons representing pairs of feature points to be matched, for example if the output (or state) of the neuron n_{ik} is 1 then the pair (i, k) of feature points is matched, otherwise the pair is not matched. The 2D HNN is in a matrix form and contains $N_l \times N_r$ neurons with a binary output 0 or 1, where N_l and N_r are respectively the total number of the feature points in the left and right images. Besides, x_{ik} and y_{ik} represent the external input and the output of neuron n_{ik} respectively, and W_{ikjl} is the connecting weight between neurons n_{ik} and n_{jl}. The energy function of 2D HNN is given by

$$E = -\frac{1}{2}\sum_{i=1}^{N_l}\sum_{k=1}^{N_r}\sum_{j=1}^{N_l}\sum_{l=1}^{N_r} W_{ikjl} y_{ik} y_{jl} - \sum_{i=1}^{N_l}\sum_{k=1}^{N_r} I_{ik} y_{ik}. \qquad (3\text{-}6)$$
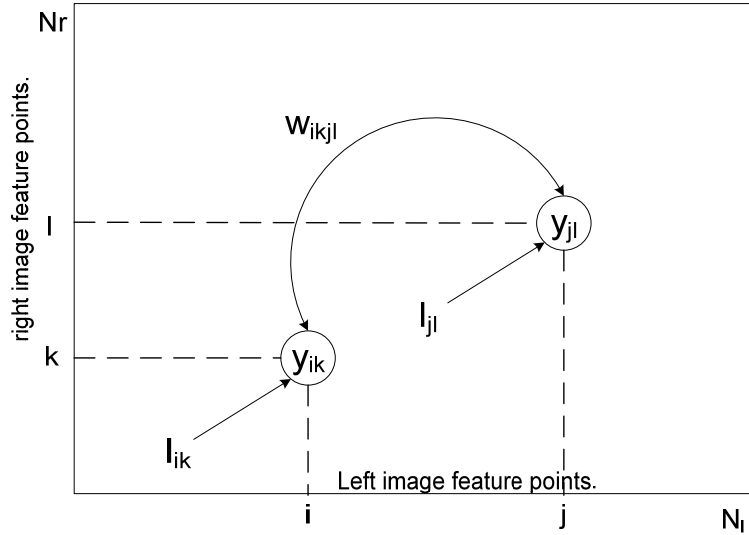
Fig. 3.2 The 2D Hopfield neuron network with its neurons interconnections.

To deal with the stereo matching problem, it is required to choose an appropriate energy function, commonly related to some properties, namely, uniqueness, similarity, smoothness and geometric. In addition to these properties, this thesis further introduces the vertical disparity into the energy function, which will be discussed and given in the next section.

3.2.2 Energy function for stereo matching

The energy function proposed here is mainly composed of three parts respectively related to uniqueness, similarity and the other properties such as smoothness, geometric and vertical disparity. For the first property, uniqueness, it assumes that a feature point should be in one-to-one matching between two images. To evaluate the uniqueness of a feature point, the following energy function

$$E_1 = \sum_{i=1}^{N_l}(1-\sum_{k=1}^{N_r} y_{ik})^2 + \sum_{k=1}^{N_r}(1-\sum_{i=1}^{N_l} y_{ik})^2.$$ (3-7)

13

is adopted, which sums up the output value of the neurons in each row and column. Clearly, the energy function E_1 is minimized to zero when the summation of the outputs of every row and every column equals to 1. Since the output of each neuron is either 1 or 0, the minimization of E_1 to zero implies only one output equals 1 for each row and each column, which also implies the feature points are uniquely matched between two images.

The second property, similarity, assumes that two matching points have similar neighbors with similar information in two images, such as gray level value. There is a method proposed to measure the degree of similarity for two feature points based on the following algorithm [19]:

$$M_{ik} = \frac{1}{n} \sum_{(x_l,y_l)\in W_i,(x_r,y_r)\in W_k} \left| V_i(x_l,y_l) - V_k(x_r,y_r) \right| \tag{3-8}$$

where W_i is a fixed size window centered on the ith feature point in the left image and W_k is a fixed size window centered on the kth feature point in the right image. Besides, functions $V_i(x_l,y_l)$ and $V_k(x_r,y_r)$ are defined as

$$V_i(x_l,y_l) = I_i(x_l,y_l) - \overline{I_i}$$

$$V_k(x_r,y_r) = I_k(x_r,y_r) - \overline{I_k} \tag{3-9}$$

where $I_i(x_l, y_l)$ and $I_k(x_r, y_r)$ are the gray level values of points (x_l, y_l) and (x_r,y_r) within W_i and W_k, respectively. In addition, $\overline{I_i}$ and $\overline{I_k}$ are the mean of gray level values of W_i and W_k, respectively. A larger M_{ik} implies worse similarity of the two feature points. The proposed algorithm uses above method to calculate the value M_{ik} for two feature points and a function S_{ik} which is defined below transfers M_{ik} limited to [-1, 1].

$$S_{ik} = \frac{2}{1 + e^{\lambda_s (M_{ik} - \theta_s)}} - 1. \tag{3-10}$$

The parameter θ_s represents the tolerance of M_{ik} and S_{ik} approaches to -1 for $M_{ik} \gg \theta_s$ and 1 for $M_{ik} \ll \theta_s$. The effect caused by the parameter λ_s is depicted in Fig. 3.3 which contains three curves for different values of λ_s. Commonly, when $\lambda_s > 1$ the function S_{ik} will abruptly change from $+1$ to -1 around $M_{ik} = \theta_s$, while $\lambda_s < 1$ the function S_{ik} will smoothly change from $+1$ to -1 around $M_{ik} = \theta_s$. Hence, an energy function to evaluate the similarity can be formulated as

$$E_2 = -\sum_{i=1}^{N_l} \sum_{k=1}^{N_r} S_{ik} y_{ik}. \tag{3-11}$$

Note that some of the output y_{ik} are equal to 1 and the others are equal to 0. For those outputs equal to 1, it is evident that the energy function E_2 will decrease to its minimum value when the functions $S_{ik} > 0$ corresponding to the neurons with output $y_{ik} = 1$.
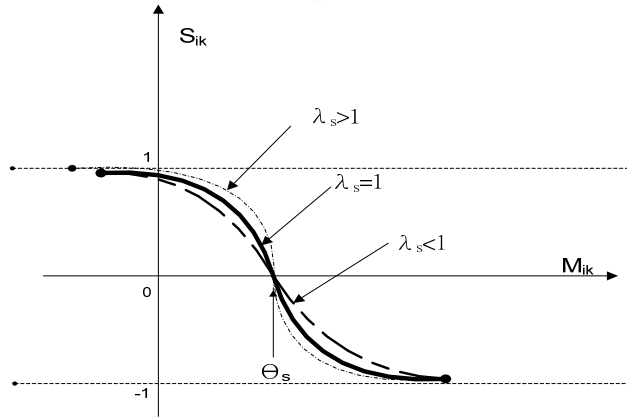

Fig. 3.3 Graph of the nonlinear function S_{ik}.

For the remaining properties, smoothness, geometric property and vertical disparity, they are adopted to represent the degree of compatibility of a match between

15

a pair of points (i, j) in the left image and a pair of points (k, l) in the right image. they are determined by the following term

$$X_{ikjl} = \Delta d + \Delta D + \Delta Y, \quad \|i(x, y), \quad j(x, y)\| < 20 \tag{3-12}$$

where Δd, ΔD and ΔY are respectively concerning the smoothness, geometric property and vertical disparity. The condition $\|i(x, y), \quad j(x, y)\| < 20$ means (3-12) is not applied to the pair of points (i, j) in the left image with distance more than 20 pixels, since such a pair of feature points may belong to different object. Besides, the term Δd is defined as the difference of the disparities between the matched pairs (i, k) and (j, l), which is related to the smoothness of neighboring feature points. The existence of depth discontinuities often leads to a lager Δd. The term ΔD is defined as the difference between the distance from i to j and the distance from k to l, which is related to the geometric property and becomes smaller for correctly matched feature points. The term ΔY is defined as the difference of the vertical disparity between the matched pairs (i, k) and (j, l), which is added to enhance the importance of the match of the feature points in vertical disparity since the two cameras are set on the same horizontal surface.

Similar to (3-10), the term X_{ikjl} is transferred into the interval [-1, 1] by the following transfer function

$$C_{ikjl} = \frac{2}{1 + e^{\lambda_c (X_{ikjl} - \theta_c)}} - 1. \tag{3-13}$$

where θ_c is the tolerance of X_{ikjl} and λ_c is a parameter to be determined. Obviously, a smaller X_{ikjl} implies better compatibility between two pairs (i, k) and (j, l). The energy function to represent smoothness, geometric and vertical disparity properties is formulated as

$$E_3 = -\sum_{i=1}^{N_l}\sum_{k=1}^{N_r}\sum_{j=1}^{N_l}\sum_{l=1}^{N_r} C_{ikjl}\, y_{ik}\, y_{jl}. \tag{3-14}$$

Clearly, the energy function E_3 becomes smaller for $C_{ikjl} > 0$ and $y_{ik} = y_{jl} = 1$, i.e., both the pairs (i, k) and (j, l) are matched and with better compatibility, a desirable case.

Next, by combining E_1, E_2 and E_3 in (3-7), (3-11), and (3-14), the total energy function for stereo matching problem is given as below

$$E = -\sum_{i=1}^{N_l}\sum_{k=1}^{N_r}\sum_{j=1}^{N_l}\sum_{l=1}^{N_r} C_{ikjl}\, y_{ik}\, y_{jl} + \sum_{i=1}^{N_l}\left(1 - \sum_{k=1}^{N_r} y_{ik}\right)^2 + \sum_{k=1}^{N_r}\left(1 - \sum_{i=1}^{N_l} y_{ik}\right)^2$$
$$- \sum_{i=1}^{N_l}\sum_{k=1}^{N_r} S_{ik}\, y_{ik}. \tag{3-15}$$

The first term is the compatibility degree between a couple of points, i and j, in the left image and a couple points, k and l, in the right one. The second and third term represent the uniqueness, and the fourth term represents the similarity. Clearly, according to the definition of E_1, E_2 and E_3, the best result will be occurred when the energy function reach its minimum. To search the minimum of the energy function, both the connection weights and the external input are required for the HNN to update the outputs of neurons. Next, the way to derive of the connection weights and the external input is shown based on (3-6) and (3-15).

3.2.3 Derivation of connection weights and external input

After obtaining the total energy function, it is necessary to transfer (3-15) to the form of (3-6) which ensures the update rule will make the energy function decreased. Referring to the work by N. M. Nasrabidi and C. Y. Choo [18], (3-15) could be written as the following form:

$$E = -(\frac{1}{2})\sum_{i=1}^{N_l}\sum_{k=1}^{N_r}\sum_{j=1}^{N_l}\sum_{l=1}^{N_r}(C_{ikjl} - \delta_{ij} - \delta_{kl})y_{ik}y_{jl} - \sum_{i=1}^{N_l}\sum_{k=1}^{N_r}(2+S_{ik})y_{ik} \qquad (3\text{-}16)$$

where δ_{ij} and δ_{kl} interpret the uniqueness property. if $i=j$, then $\delta_{ij}=1$, otherwise $\delta_{ij}=0$; similarly, if $k=l$, then $\delta_{kl}=1$, otherwise $\delta_{kl}=0$. By the comparison of (3-6) and (3-16), it is easy to derive the values of the connection weights and the external input as follows:

$$W_{ikjl} = C_{ikjl} - \delta_{ij} - \delta_{kl}$$
$$I_{ik} = 2 + S_{ik}. \qquad (3\text{-}17)$$

It is important to point out that the symmetrical inter-connection weights and no self-feedback are required to ensure the convergence of the energy function E. To substitute W_{ikjl} and I_{ik} in (3-5) by (3-17), a variation Δy_{ik} of the neuron n_{ik} leads to the energy variation, expressed as

$$\Delta E^t = -\left[\sum_{j=1}^{N_l}\sum_{l=1}^{N_r}(C_{ikjl} - \delta_{ij} - \delta_{kl})y_{jl}^{t-1} + 2 + S_{ik}\right]\Delta y_{ik}^t = -u_{ik}^{t-1}\Delta y_{ik}^t \qquad (3\text{-}18)$$

which describes the dynamic of the network and always shows non-positive when applying the following update rules at each time step t:

$$y_{ik}^t = \begin{cases} 0 & \text{if } y_{ik}^{t-1}=1 \text{ and } \left[\sum_{j=1}^{N_l}\sum_{l=1}^{N_r}(C_{ikjl} - \delta_{ij} - \delta_{kl})y_{jl}^{t-1} + 2 + S_{ik}\right] < 0 \\[2mm] 1 & \text{if } y_{ik}^{t-1}=0 \text{ and } \left[\sum_{j=1}^{N_l}\sum_{l=1}^{N_r}(C_{ikjl} - \delta_{ij} - \delta_{kl})y_{jl}^{t-1} + 2 + S_{ik}\right] > 0 \\[2mm] y_{ik}^{t-1} & \text{otherwise} \end{cases} \qquad (3\text{-}19)$$

After setting the initial state of the 2D HNN, this update rule executes repeatedly until the network outputs do not change that implies the energy function reaches its minimum, and the solution of the stereo matching problem will be obtained.

3.3 Summary and the Matching Procedure

Based on the energy function (3-16), the connection weights, external input and update rules of the 2D HNN are obtained. Further setting the output of each neuron in the 2D HNN by suitable initial value, the outputs of the 2D HNN can be updated to a set of fixed values, which represent matching relation of feature points. In this thesis, the initial outputs of the 2D HNN are determined by using a fixed size 30×40 window to roughly separate the possible and impossible matched pairs. For example, if a feature point i with coordinate (x_l, y_l) is selected in the left image, and a 30×40 window will be opened and centered on (x_l-20, y_l) in the right image. Every feature point k in this window will be the matching candidates and the initial output of neuron, y_{ik}, is set to 1. Clearly, the disparity range will be limited by the initialization to $\sqrt{40^2+15^2}$, that is limited to about 61.85. Besides, the center of the window in the right image is shifted left by 20 pixels, from (x_l, y_l) to (x_l-20, y_l), due to the fact that for a physical point captured by two cameras in parallel its horizontal coordinate in the left image will be larger than the one in the right image, for example the difference is 20 pixels for the two cameras system in this thesis. After that, two stereo matching classes are obtained: possible and impossible. Furthermore, the update rule is only applied to the possible matched pairs that can speed up the convergence of energy function and avoid changing the impossible matched pairs to the possible matched pairs. The initial outputs of HNN will be saved to ensure the next step, random choice, will choose the possible matched pair to be update. To calculate the compatibility term, when a possible matched pair n_{ik} is chosen to updated, according to the smoothness property, a region 40×40 pixels which is centered on feature point i is opened to search the neighboring feature points. All the points belong to this region

represent the j's of equation (3-18). For every point j, the l's of equation (3-18) are the neuron n_{jl} whose output equals to 1. Therefore, the output of the neuron n_{ik} will be adjusted by the update rule. To summarize the stereo matching, the overall flow chart is shown as Fig. 3.4, where

$$u_{ik}^{t-1} = \left[\sum_{j=1}^{N_l} \sum_{l=1}^{N_r} (C_{ikjl} - \delta_{ij} - \delta_{kl}) y_{jl}^{t-1} + 2 + S_{ik} \right]. \tag{3-20}$$

In this thesis, the random updating procedure is iterated until the outputs of network remain unchanged during 200 iterations.

The proposed algorithm ISMB refers to the work of K. Achour and L. Mahiddine [19] denoted as *Method-I*, but there are three differences between them. The first difference is that ISMB uses genetic algorithm to search the parameters of HNN instead of using try and error. The second difference is the usage of the similarity property. In *Method-I*, the degree of similarity M_{ik} is combined with smoothness and geometric properties to form a function which is similar to E_3 which requires more than one matched pairs to calculate. Therefore, other matched pairs (j, l) around the matched pair (i, k) are required, and then the M_{ik} can be used in energy function. To emphasize the similarity, ISMB measures the similarity of two feature points by the function S_{ik} which only depends on matched pair (i, k). The third difference is the usage of the vertical disparity. In *Method-I*, the vertical disparity ΔY is used as a post-processing for reducing the multiple matching cases after obtaining the result of the 2D HNN. ISMB directly combines it with other properties to form a term E_3 of the energy function in 2D HNN to reducing some incorrect matched pairs. The results of the second and third differences between them will be shown in chapter 5.
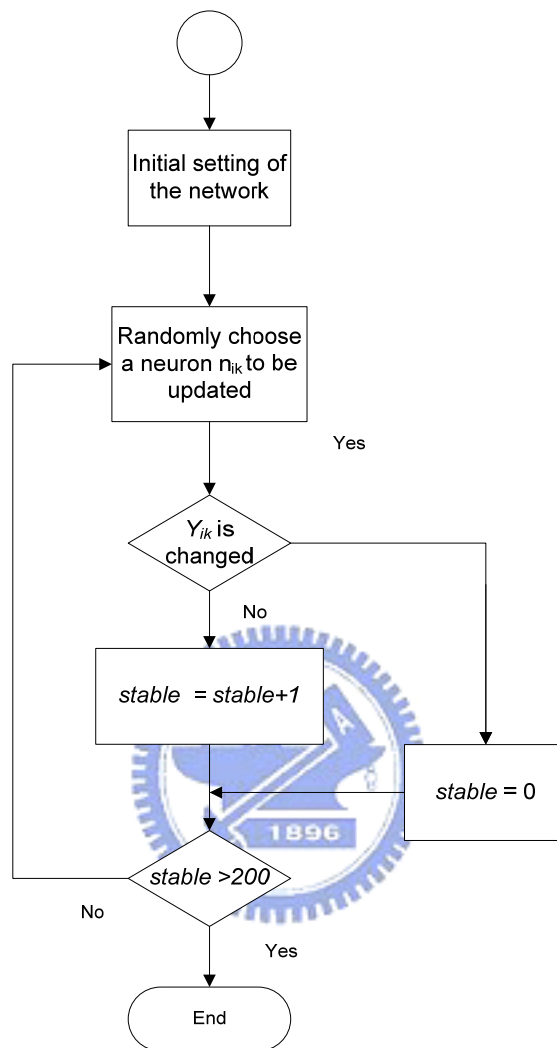
20

Fig. 3.4 The flow chart of stereo matching.

Chapter 4

Real-coded genetic algorithm

Aim to search the four parameters of the energy function in 2D HNN, the genetic algorithm is used to automatic find the parameters which make the performance of 2D HNN better. The section 4.1 will briefly introduce the real-coded genetic algorithm, and then to apply the real-coded genetic algorithm to search the parameters of the energy function, a fitness function is designed in the section 4.2

4.1 The Real-Coded Genetic Algorithms Description
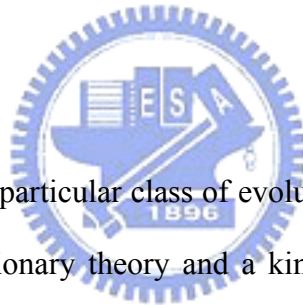
Genetic algorithms are a particular class of evolutionary algorithms based on the concepts of biological evolutionary theory and a kind of search techniques used in computing to find exact or approximate solutions to optimization and search problems. Generally, a basic genetic algorithm has the following components [23]:

- a population of chromosomes which represent potential solutions of the problem.
- a fitness function indicates the adaptation of the chromosomes.
- genetic operators (crossover and mutation).
- a selection mechanism.

At the beginning, the population is composed of randomly generated chromosomes, and then advance toward better chromosomes during successive iterations, called generations, by using the selection mechanism and applying genetic operators such as crossover and mutation operators modeled on the genetic processes in nature.

Although there are many possible variants on the basic genetic algorithm, the operation of a standard genetic algorithm is described in the following steps [24]:

1. Randomly create an initial population of chromosomes.

2. Compute the fitness of every member of the current population.

3. If there is a member of the current population that satisfies the problem requirement then stop. Otherwise, continue to the next step.

4. Create an intermediate population P' by extracting members from the current population using selection mechanism.

5. Generate a new population by applying the genetic operators of crossover and mutation to this intermediate population.

6. Go back to step 2.

Following the above steps, the chromosomes consist of a set of genes commonly are represented in binary as strings of 0s and 1s so that if a chromosome is a vector v consisting of n gene g_i, then $v = (g_1, g_2, ..., g_n)$, $g_n \in \{0,1\}$, where n is the length of the chromosome. However, it is more natural to represent the genes directly as real numbers since the representations of the solutions are very close to the natural formulation. The advantage of using real-coded genetic algorithm is increased precision since binary coding of real numbers can suffer loss of precision depending on the number of bits used to represent a number. Moreover, for real-valued optimization problems, real-value coding is much easier and more efficient to implement because it is closer to the problem space. In this thesis, real-coded genetic is used so that the chromosome is a vector of floating point numbers and the genes of the chromosome represent the variables of the problem. For the two genetic operators of real-coded genetic algorithm, the crossover operator and mutation operator used

here are arithmetic crossover and random mutation and will be introduced in next part.

For arithmetic crossover, assume that $C_1 = (g_1^1, g_2^1, \ldots, g_n^1)$ and $C_2 = (g_1^2, g_2^2, \ldots, g_n^2)$ are two chromosomes chosen for application of the crossover operator, then two offspring $H_k = (h_1^k, h_2^k, \ldots, h_n^k)$, $k = 1, 2$, are generated, where $h_i^1 = \lambda c_i^1 + (1-\lambda)c_i^2$ and $h_i^2 = \lambda c_i^2 + (1-\lambda)c_i^1$. λ is a random value limited into the interval $[0, 1]$ for each generation. For the random mutation, assume that $C = (g_1, g_2, \ldots, g_n)$ is a chromosome to mutate, one of genes in chromosome g_i will be changed by a random value g_i' in the interval $[\min(g_i), \max(g_i)]$.

4.2 Applying Real-Coded Genetic Algorithm to the 2D HNN

As mentioned above, real-coded genetic algorithm is convenient to find the appropriate four parameters for the proposed 2D HNN. The real-coded genetic algorithm used in this thesis also follows the basic genetic algorithm starts at the randomly creation of chromosomes, and each chromosome is a vector contains four elements to represent the four parameters of the 2D HNN. After the randomly creation of chromosomes, a fitness function is defined to evaluate whether the chromosome make the result of 2D HNN good or not. To define the fitness function, a target for the stereo matching problem is established by user so that the defined fitness function described below can judge the relation between the target and result of 2D HNN which use the four parameters of the chromosome.
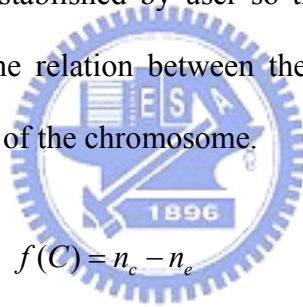
$$f(C) = n_c - n_e \tag{4-1}$$

where C is a chromosome contains four parameters, i.e. $C = [\theta_c, \theta_s, \lambda_c, \lambda_s]$, n_c is the number that the matched pairs of result are identical to the matched pairs of the target and n_e is the number that the match pairs of result are not identical to the matched pairs of the target. Follow the definition of 4-1, the best fitness value equals to the number of matched pairs in target and occurs at the result of 2D HNN is identical to the target. For the stop condition, the genetic algorithm stops when the average fitness value of all chromosome in the intermediate population equal to the maximum of the fitness value or has reached the maximum number of generations. For creating an intermediate population P', the two chromosomes with the largest fitness value will be first copy to the intermediate population to make sure that the best-performing

chromosome always survives from one generation to the next, and a competition method is used for selection mechanism by randomly choose two chromosomes and comparing their fitness value that the large one will be copy to the intermediate population. This creation process will repeat the competition until the number of chromosomes in the population reaches its maximum (30 individuals in this thesis). Once the intermediate population is created, the population of the next generation can be formed by applying the crossover and mutation operators on the chromosomes in P'. For the genetic operators, arithmetic crossover operator and random mutation are used. Two chromosomes are randomly selected from P' and serve as parents depends on the crossover rate p_c which is the probability of the two will be crossed over. If the two is crossed over, the offspring chromosome may mutates by the random mutation depends on the mutation rate p_m which is the probability of the mutation. After applying these genetic operators, the next step is to compute the fitness value of those offspring chromosomes, and then the chromosomes in p' and those offspring chromosomes will be combined to generate the new intermediate population by the process of creating an intermediate population. The genetic algorithm will continue until the stop condition is satisfied.

Chapter 5

Experimental results

In this chapter, the system structure is described in section 5.1 and the algorithm is executed by the software MATLAB R2006a. The results of the real-coded genetic algorithm and stereo matching process will be shown and compared in section 5.2.

5.1 System Description

The system shown in Fig. 5.1 is established by setting two cameras on a horizontal line and their lines of vision are parallel and fixed. In addition, the distance between two cameras is set as constant equal to 10 cm and these two cameras, QuickCam^{TM} Communicate Deluxe, have specification listed below. The experimental environment for testing is our laboratory and the deepest depth of the background is 180 cm.

- 1.3-megapixel sensor with RightLight™2 Technology

- Built-in microphone with RightSound™ Technology

- Video capture: Up to 1280 x 1024 pixels (HD quality) (HD Video 960 x 720 pixels)

- Frame rate: Up to 30 frames per second

- Still image capture: 5 megapixels (with software enhancement)
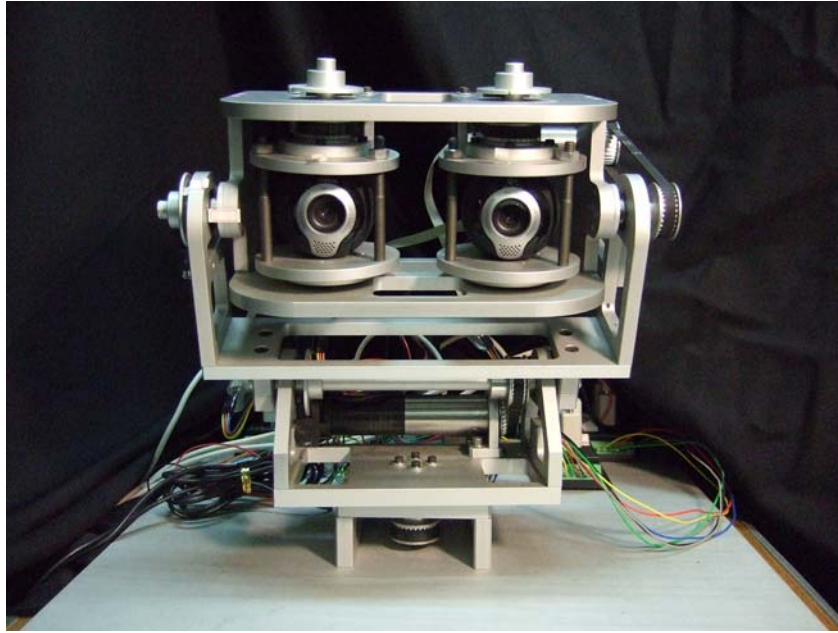
- USB 2.0 certified

- Optics: Manual focus

Fig. 5.1 The humanoid vision system.

5.2 The stereo matching results

In order to test the proposed method, there are several stereo images pairs captured from an indoor environment, and each pair consists of two left and right original images and two left and right images of labeled feature points. All the test images are 240×320 pixels in size, with 256 gray levels. Because the proposed method focuses on the center region of two images, a fixed size region of interest (ROI) whose size is one fourth of the image is allowed and labeled in the image. Fig. 5.2 show an original image pair, and Fig. 5.3 show the corresponding images with the feature points extracted with the Harris operator. Each feature point has a unique number label and denoted as + symbol, so there are 18 feature points in left image and 20 feature points in right image. For the stereo matching problem, three kinds of 2D HNN are compared in this thesis to discuss the effect of the modified similarity and

vertical disparity. The first kind 2D HNN is proposed in [19] and denoted as *Method-I*, which combines the similarity with smoothness and geometric property to form a term in energy function and does not use the vertical disparity in energy function. The second one denoted as *Method-II* and the third one denoted as *Method-III* all use the proposed 2D HNN, but the difference is the second one does not use the vertical disparity in energy function. In following section, the real-coded genetic algorithm is applied to search the four parameters of 2D HNN, and the effects of the modified similarity and the vertical disparity are presented by comparing the results of *Method-I* with *Method-II* and the results of *Method-II* and *Method-III*, respectively. Finally, a simple application is proposed to detect the relative distance information between the objects.
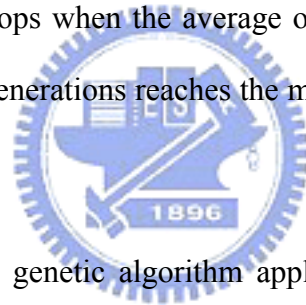
5.2.1 The result of genetic algorithm and the comparison between *Method-I* and *Method-II*

To discuss the modified similarity, the *Method-I* and *Method-II* are established to do the stereo matching process of the Fig. 5.2, and the size of each HNN is 20×20 to represent the 400 matched pairs, each neuron for one matched pair, and the initial state of the two HNN are the same. In Table 5.1, a list shows the initial state of the HNN and there are 39 pairs to be considered as the possible matched pairs after the initialization described in section 3.3. But these 39 possible matched pairs still include some incorrect matched pairs, and the HNN is used to cancel the incorrect matched pairs. After the initialization, the HNN will update the output of neurons by the update rule until the HNN satisfies the stable condition. However, before executing 2D HNN,

there are two parameters [θ_c, λ_c] of the *Method-I* and four parameters [θ_c, θ_s, λ_c, λ_s] of *Method-II* have to be determined by the proposed real-coded genetic algorithm described in section 4.2. For the real-coded genetic algorithm, some settings have to be determined as below:

1. To establish the target of the matching problem by user, and the target value will be the number of the matched pairs in the target. For this example, the target contains 13 matched pairs and is denoted as * in the Table 5.1, so the target value will be 13.

2. The maximum number of intermediate population is 30.

3. Crossover rate p_c=0.8 and mutation rate p_m=0.05.

4. The genetic algorithm stops when the average of fitness value equals the target value or the number of generations reaches the maximum number which is set as 500.

The results of the real-coded genetic algorithm applied to the two kinds HNN are respectively shown in Fig. 5.3 and Fig. 5.4. For the *Method-I*, the average of fitness value is increasing through the number of generation increases and the proposed genetic algorithm stops when the number of generation equals to 500. The chromosome with the best fitness value [41.57, 4.40] can be obtained and be used to set the parameters [θ_c, λ_c] of the *Method-I*. The stereo matching result of the *Method-I* which θ_c =41.57 and λ_c =4.40 is shown in Table 5.2. For the *Method-II*, the genetic algorithm stops when the number of generation equals to 500, and the four parameters of the *Method-II*, [θ_c, θ_s, λ_c, λ_s], can be determined by choosing one of the chromosome in the final intermediate population [7.45, 15.20, 1.47, 4.08]. The stereo matching result of the *Method-II* which θ_c =7.45, θ_s=11.78, λ_c =1.47 and λ_s =4.08 is shown in Table 5.3. Comparing Table 5.1 and Table 5.2, the *Method-I* can reduce the

incorrect matched pairs from the initial result which contains 39 matched pairs, but there are three error matched pairs whose number are 1, 2 and 5 in the result of the *Method-I* and the result does not contain a target pair whose number is 26 in Table 5.1. The Table 5.3 is the result of the *Method-II* and contains the whole target pairs, but one error matched pair still exist in the Table 5.3. Clearly, the *Method-II* can improve the performance of stereo matching by modifying the similarity in the energy function. In addition, to ensure the update rule make the energy function decrease, the Fig. 5.5 shows the average energy value of executing the *Method-II* 20 times and the energy value monotonously decreases during the update process.

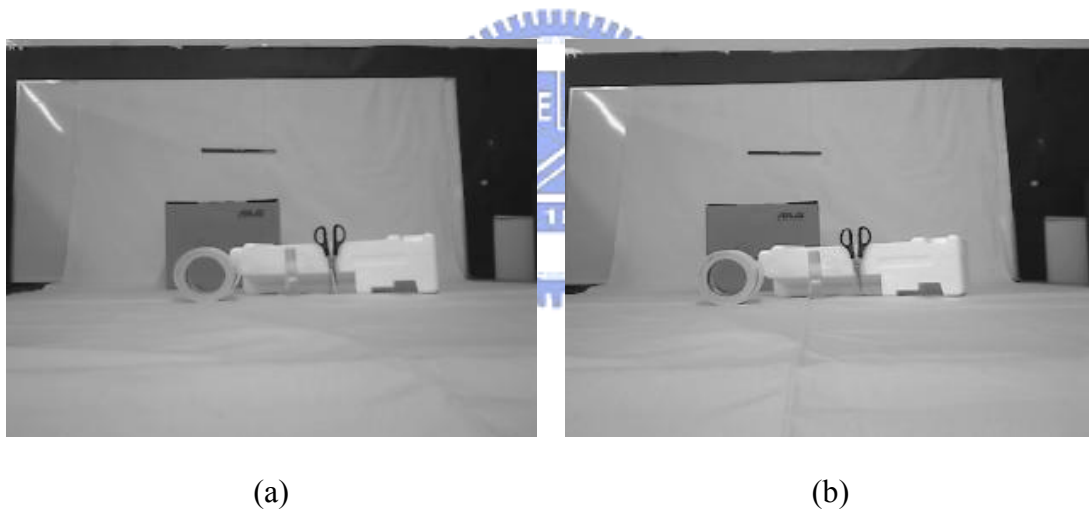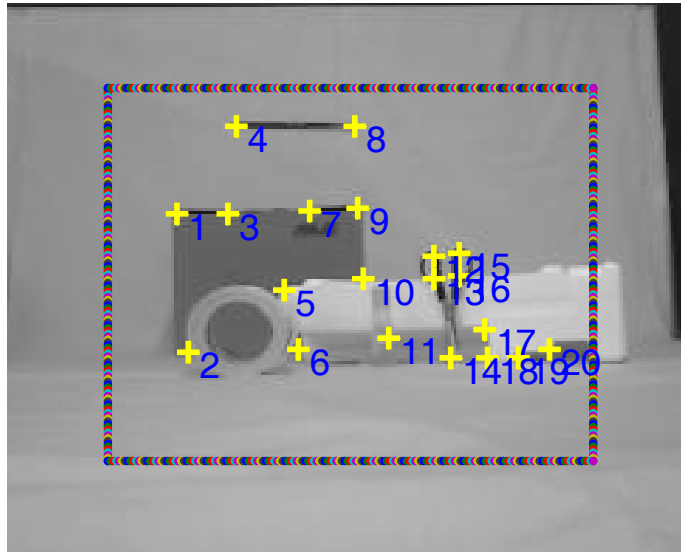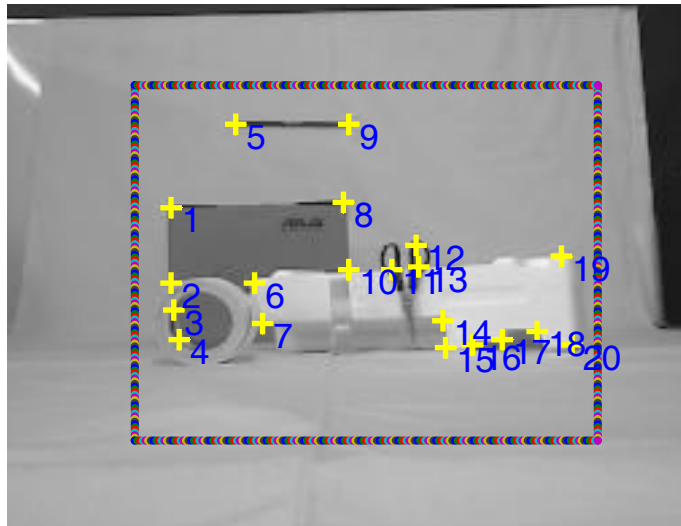(a)   (b)

Fig. 5.2 The original image pair, (a) left gray level image, (b) right gray level image.

(a)

(b)

Fig. 5.3 The corresponding image pair with the feature points extracted with the Harris operator and ROI, (a) the left image with 20 feature points, (b) the right image with 20 feature points.
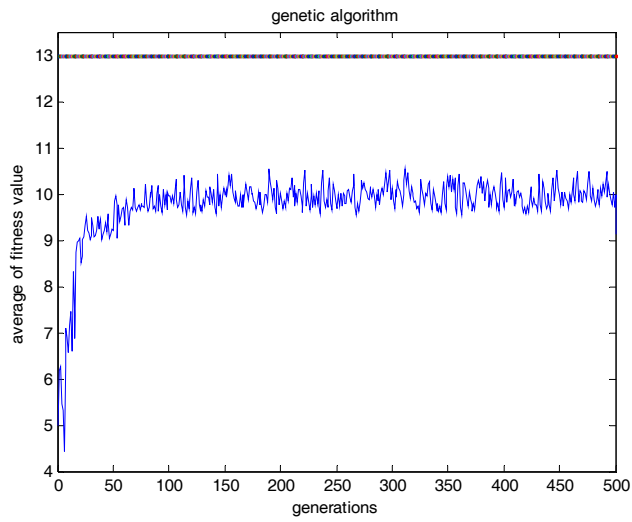
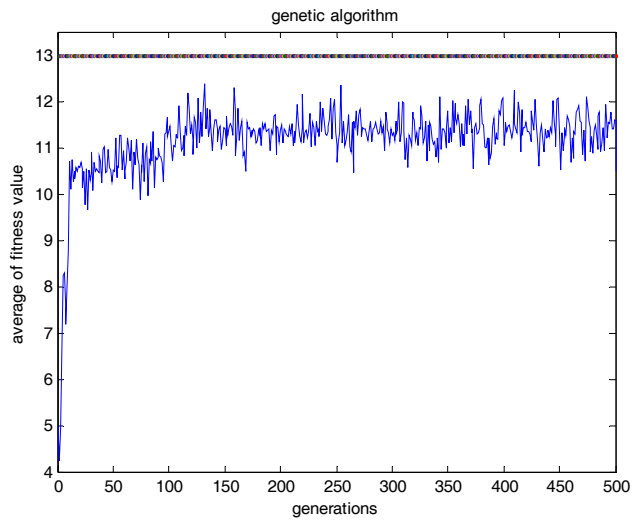Fig. 5.4 The result of genetic algorithm corresponding to *Method-I*.

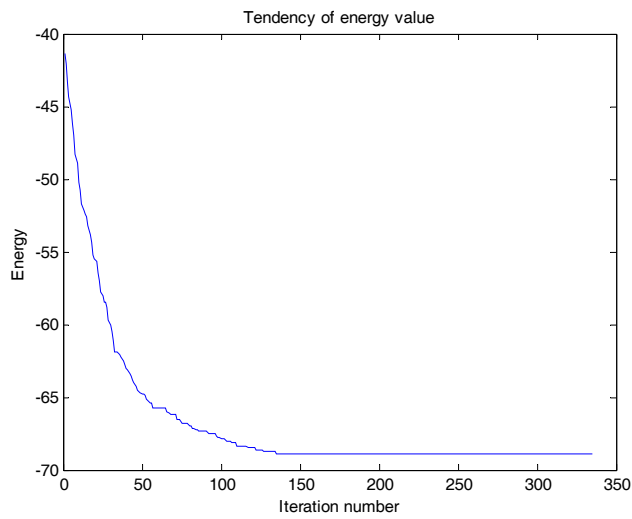Fig. 5.5 The result of genetic algorithm corresponding to *Method-II*.

Fig. 5.6 Average energy values at each iteration number.

Table 5.1 The initial state of 2D HNN applied to Fig. 5.2.

| Pair number	Left feature label	Right feature label	y_{left}	x_{left}	y_{right}	x_{right}	disparity
*1	1	1	99	99	100	88	11.05
2	2	3	148	103	138	89	17.20
3	2	4	148	103	149	91	12.04
4	3	1	99	117	100	88	29.02
*5	4	5	68	120	69	112	8.06
*6	5	6	126	137	128	119	18.11
7	6	7	147	142	143	122	20.40
*8	8	9	68	162	69	154	8.06
*9	9	8	97	163	98	152	11.05
*10	10	10	122	165	123	154	11.05
11	12	10	114	190	123	154	37.11
12	12	11	114	190	123	170	21.93
13	12	12	114	190	114	179	11.00
14	12	13	114	190	122	180	12.81
15	13	10	122	190	123	154	36.01
*16	13	11	122	190	123	170	20.03
17	13	12	122	190	114	179	13.60
18	13	13	122	190	122	180	10.00
19	14	14	150	196	142	189	10.63
20	14	15	150	196	152	190	6.32
21	15	11	113	199	123	170	30.68
*22	15	12	113	199	114	179	20.03
23	15	13	113	199	122	180	21.02
24	16	11	121	199	123	170	29.07
25	16	12	121	199	114	179	21.19
*26	16	13	121	199	122	180	19.03
*27	17	14	140	208	142	189	19.11
28	17	15	140	208	152	190	21.63
29	17	16	140	208	151	200	13.60
30	18	14	150	209	142	189	21.54
*31	18	15	150	209	152	190	19.11
32	18	16	150	209	151	200	9.06
33	19	14	150	220	142	189	32.02
34	19	15	150	220	152	190	30.07
*35	19	16	150	220	151	200	20.03
36	19	17	150	220	149	211	9.06
37	20	16	147	231	151	200	31.26
*38	20	17	147	231	149	211	20.10
39	20	18	147	231	146	224	7.07

Table 5.2 The result of *Method-I* applied to Fig. 5.2.

Pair number	Left feature label	Right feature label	y_{left}	x_{left}	y_{right}	x_{right}	Disparity
1	*2*	*4*	*148*	*103*	*149*	*91*	*12.04*
2	*3*	*1*	*99*	*117*	*100*	*88*	*29.02*
3	4	5	68	120	69	112	8.06
4	5	6	126	137	128	119	18.11
5	*6*	*7*	*147*	*142*	*143*	*122*	*20.40*
6	8	9	68	162	69	154	8.06
7	9	8	97	163	98	152	11.05
8	10	10	122	165	123	154	11.05
9	13	11	122	190	123	170	20.03
10	15	12	113	199	114	179	20.03
11	17	14	140	208	142	189	19.11
12	18	15	150	209	152	190	19.11
13	19	16	150	220	151	200	20.03
14	20	17	147	231	149	211	20.10

Table 5.3. The result of *Method-II* applied to Fig. 5.2.

| Pair number	Left feature label	Right feature label	y_{left}	x_{left}	y_{right}	x_{right}	Disparity
1	1	1	99	99	100	88	11.05
2	*2*	*4*	*148*	*103*	*149*	*91*	*12.04*
3	4	5	68	120	69	112	8.06
4	5	6	126	137	128	119	18.11
5	8	9	68	162	69	154	8.06
6	9	8	97	163	98	152	11.05
7	10	10	122	165	123	154	11.05
8	13	11	122	190	123	170	20.03
9	15	12	113	199	114	179	20.03
10	16	13	121	199	122	180	19.03
11	17	14	140	208	142	189	19.11
12	18	15	150	209	152	190	19.11
13	19	16	150	220	151	200	20.03
14	20	17	147	231	149	211	20.10

5.2.2 The comparison between *Method-II* and *Method-III*

To illustrate the effect of vertical disparity used in HNN, *Method-II* and *Method-III* are applied to a realistic image pair which is shown in Fig 5.7 and with the feature points in the ROI. The limitation for the number of the feature points is set as [20, 30] and the threshold *T* is set as 30. To deal with the stereo matching problem of Fig. 5.7, the parameters of *Method-II* and *Method-III* also have to be determined by the proposed genetic algorithm. The processes of the genetic algorithm applied to *Method-II* and *Method-III* are shown in Fig. 5.8 and Fig. 5.9, respectively. Since the number of matched pairs in target is 12, the best fitness value is also 12. When the stop condition of the genetic algorithm is satisfied, the chromosome with the maximum fitness value in the population is chosen to determine the parameters of HNN so that the parameters of *Method-II* are set as [11.51, 26.24, 8.89, 1.16] and the parameters of *Method-III* are set as [17.97, 24.02, 6.37, 2.2]. Table 5.4 is the corresponding result when *Method-II* applied to Fig. 5.7 and reaches its stable state. The pair number 11 in Table 5.4 is an error matched pair, but this pair does not occur in the Table 5.5 which is the result of the *Method-III* applied to Fig. 5.7 since the vertical disparity of point number 18 in left image and point number 13 in right image is large. From the Table 5.5, the result is one-to-one and correct matching that verifies the *Method-III* can be used to improve the performance by reducing the matched pair with the large vertical disparity. Besides, to show the performance of the parameters, the *Method-III* with the parameters [17.97, 24.02, 6.37, 2.2] is also applied to other realistic stereo image pairs shown in Fig. 5.8. The testing results are the results using the parameters [17.97, 24.02, 6.37, 2.2] and the training results are the results using another parameters which are obtained by applying the proposed genetic algorithm to each image pair. The error term represents how many matched pairs are not identical

to the target, and the percentage of the correct matching by using the *Method-III* with

the parameters [17.97, 24.02, 6.37, 2.2] is 91.24%.
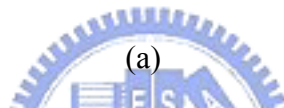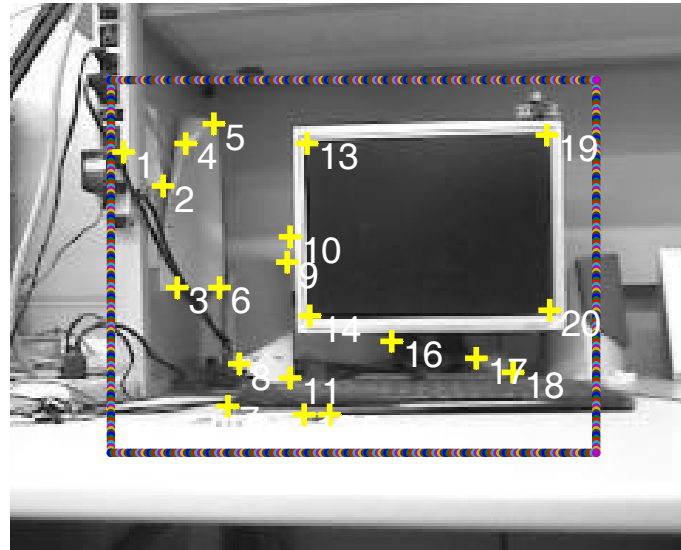
(a)
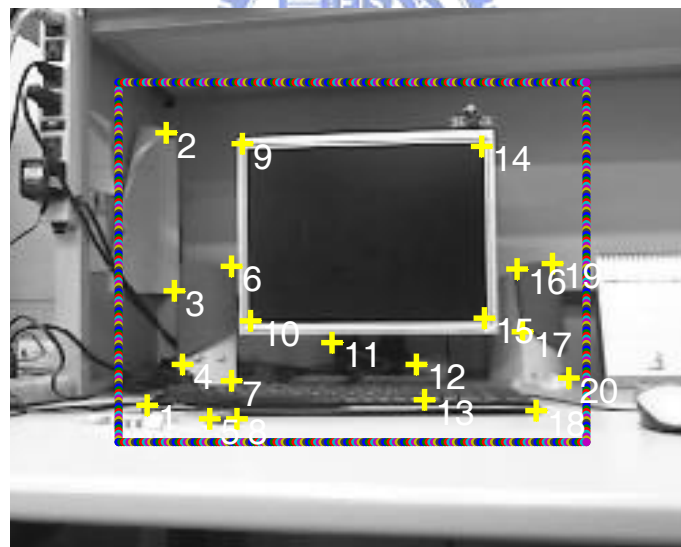
(b)

Fig. 5.7 The realistic image pair with the feature points in the ROI, (a) left image with

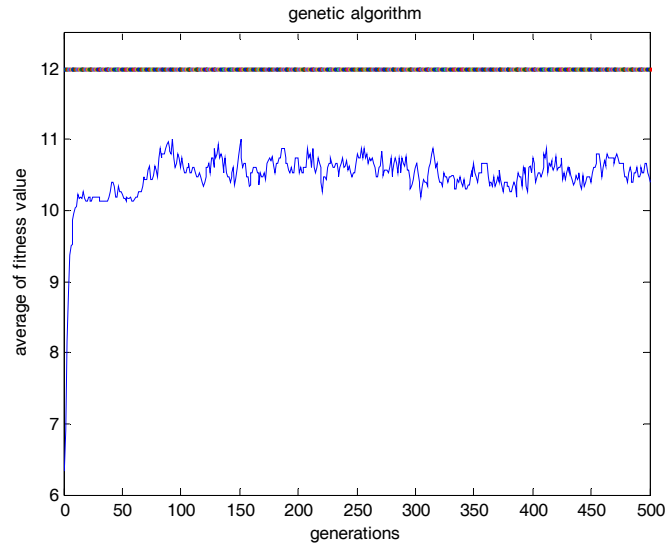21 feature points, (b) right image with 20 feature points.

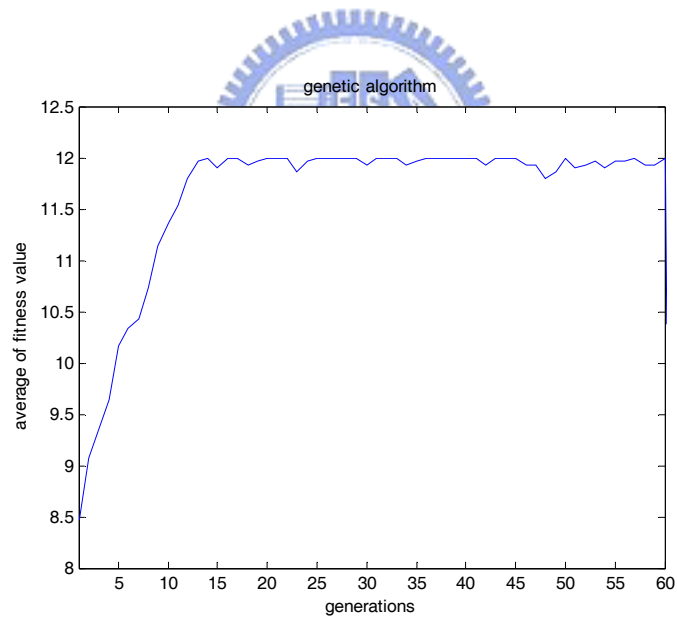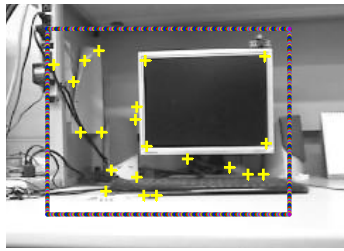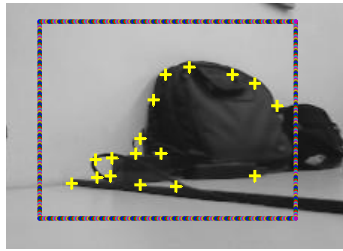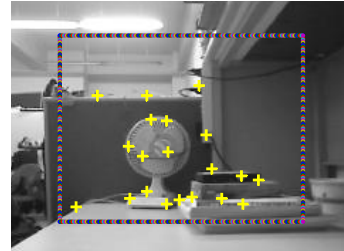Fig. 5.8 The result of genetic algorithm corresponding to *Method-II*.

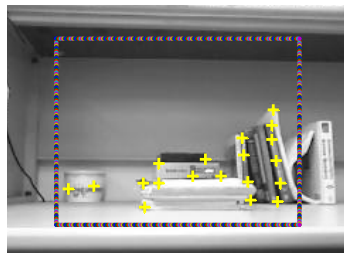Fig. 5.9 The result of genetic algorithm corresponding to *Method-III*.

Correct/Error/Target
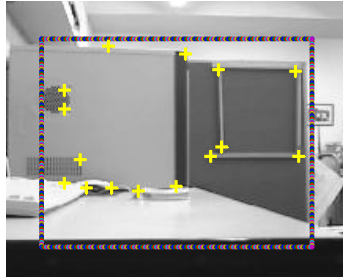Testing: 12/0/12
Training: 12/0/12

Correct/Error/Target
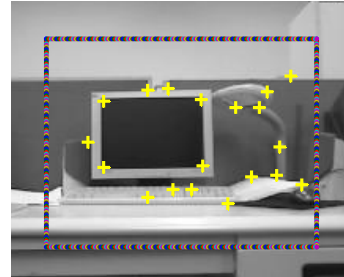Testing: 11/0+1/12
Training: 11/0+1/12

Correct/Error/Target
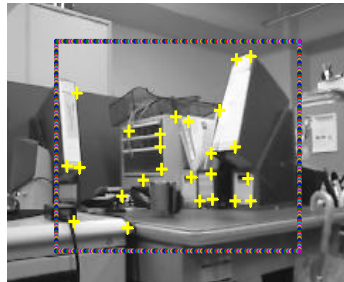Testing: 11/0+1/12
Training: 11/0+1/12

Correct/Error/Target
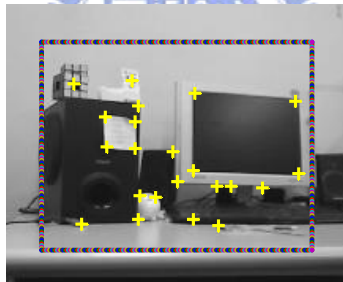Testing: 15/0/15
Training: 15/0/15

Correct/Error/Target
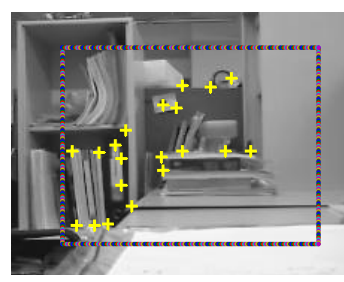Testing: 5/0+2/5
Training: 5/0/5

Correct/Error/Target
Testing: 13/0+1/13
Training: 13/0/13

Correct/Error/Target
Testing: 14/2+0/16
Training: 16/0+1/16
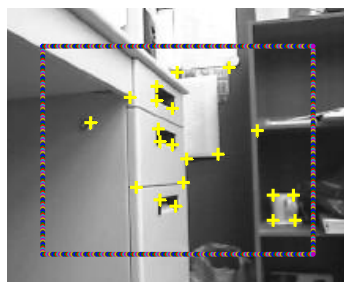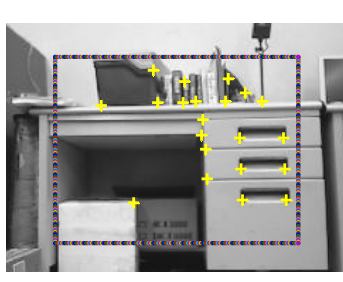
Correct/Error/Target
Testing: 14/0+1/16
Training: 15/0/16

Correct/Error/Target
Testing: 12/0+2/12
Training: 12/0+2/12

Correct/Error/Target
Testing: 16/1+1/17
Training: 17/0/17

Correct/Error/Target
Testing: 16/0+1/20
Training: 20/0/20

Total: C/E/T = 125/14/137

Accurate rate
125/137=91.24%

Fig. 5.10 The left images of the realistic image pairs in the experiment.

39

Table 5.4 The result of the *Method-II* applied to Fig. 5.7.

Pair number	Left feature label	Right feature label	y_{left}	x_{left}	y_{right}	x_{right}	Disparity
1	5	2	70	111	73	92	19.24
2	6	3	128	113	131	95	18.25
3	8	4	155	120	158	98	22.20
4	9	6	119	137	122	116	21.21
5	11	7	160	138	164	116	22.36
6	12	5	173	143	178	108	35.35
7	14	10	138	145	142	123	22.36
8	15	8	173	152	178	118	34.37
9	16	11	147	174	150	153	21.21
10	17	12	153	204	158	184	20.62
11	18	13	158	217	172	193	27.78
12	19	14	74	229	78	208	21.38
13	20	15	136	230	141	209	21.59

Table 5.5 The result of the *Method-III* applied to Fig. 5.7.

Pair number	Left feature label	Right feature label	y_{left}	x_{left}	y_{right}	x_{right}	Disparity
1	5	2	70	111	73	92	19.24
2	6	3	128	113	131	95	18.25
3	7	1	170	116	173	85	31.14
4	8	4	155	120	158	98	9.85
5	9	6	119	137	122	116	21.21
6	11	7	160	138	164	116	22.36
7	12	5	173	143	178	108	35.35
8	14	10	138	145	142	123	22.36
9	15	8	173	152	178	118	34.36
10	16	11	147	174	150	153	21.21
11	17	12	153	204	158	184	20.62
12	20	14	74	229	78	208	21.38
13	21	15	136	230	141	209	21.59

5.2.3 A simple application to obtain the relative distance between objects

A simple application to detect objects and find the relative distance between the detected objects is also proposed in the following experiment. The first step is to establish a scene as the background shown in Fig. 5.11, and the disparities of matched pairs in the background images will be derived by applying the *Method-III* and are marked in Fig. 5.12. Secondly, if there are two objects appear in the scene, such as Fig. 5.13 in which two toy cars appear, the disparities of the two objects can be acquired shown in Fig. 5.14 by applying the *Method-III* to Fig. 5.13 and deleting the feature points of background. After that the relative distance information can be obtained by comparing the disparities of background and two objects. The feature point with large disparity is more closed to the two cameras than the feature point with small disparity. From the Fig. 5.14, the feature points of the left toy car have the larger disparities so that the left car is more closed to the two cameras than the right toy car. In addition, to compare the Fig. 5.12 and Fig. 5.14, the information that two objects are in front of the box in the center of the image is also obtained.

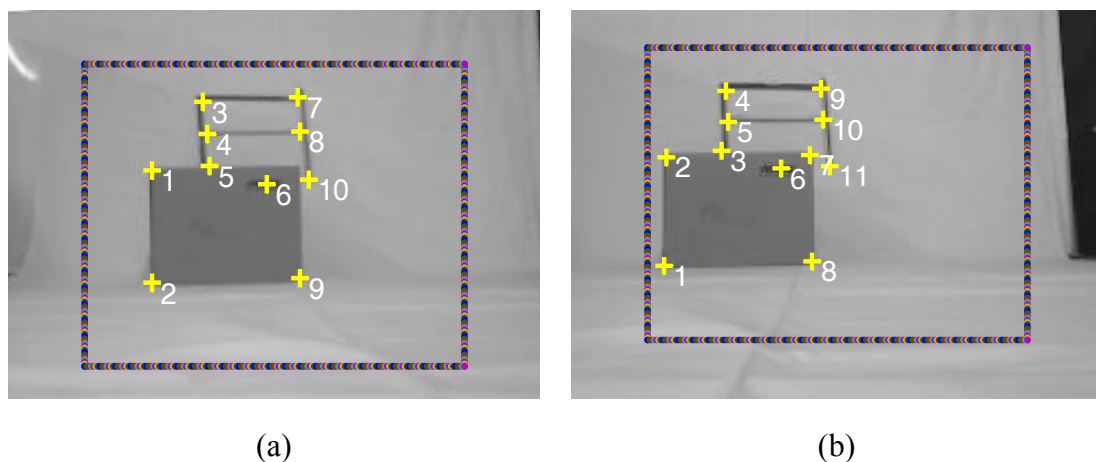(a)   (b)

Fig. 5.11 The background image pair with the feature points in the ROI, (a) left image with 21 feature points, (b) right image with 20 feature points.
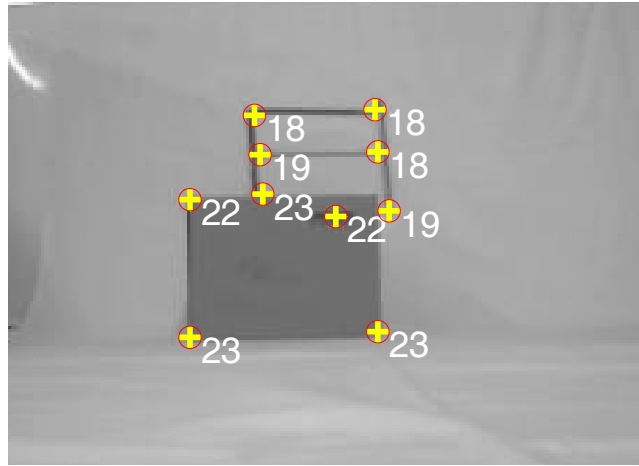
Fig. 5.12 The feature points with disparities in the background image.
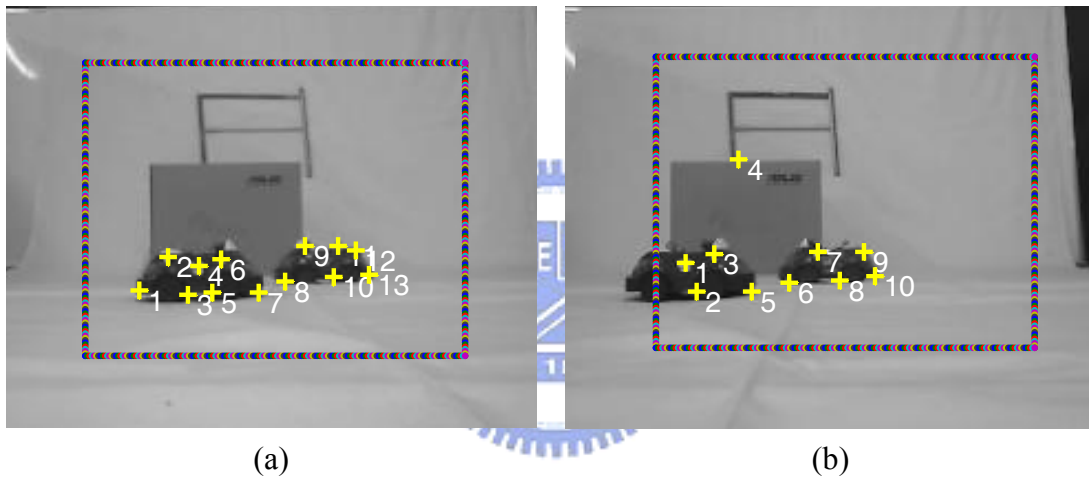

| (a) | (b) |

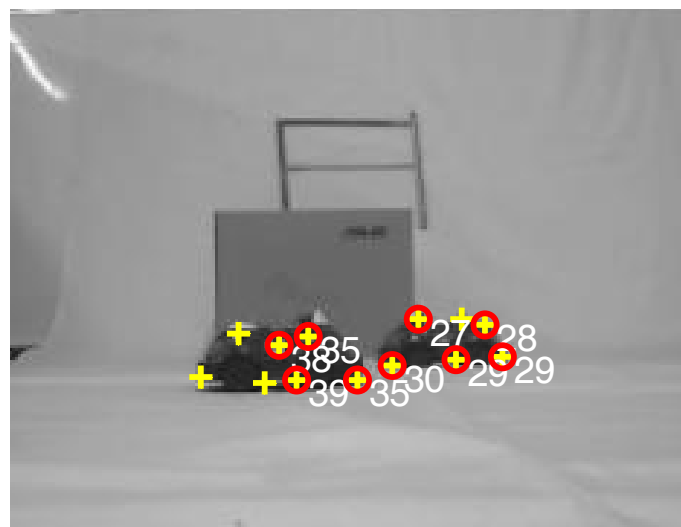Fig. 5.13 The feature points of the two objects which appear in the ROI of image pair.

Fig. 5.14 The feature points of the two objects and with their disparities.

Chapter 6

Conclusions and Future Works

An intelligent algorithm to detect the disparity of feature point by solving the stereo matching problem is proposed in this thesis. The algorithm consists of feature extraction, stereo matching process and a searching parameters algorithm. The feature extraction is implemented by the Harris corner detector, which extracts the corners of objects as the feature points. The stereo matching problem is formulated as a minimization of a neural network's energy function which represents the constraints on the solution. The neural network is a 2D Hopfield neural network with connection weights between neurons to minimize the energy function. Since there are four parameters in the energy function have to determined, a real-coded genetic algorithm is applied to automatic search those parameters which make the performance of the stereo matching process better. The contribution of this thesis is to design the energy function, and the experimental results show that using this energy function provides good matching results. Besides, the execution time of the test image pairs on a PC with AMD 1.92 GHz CPU is estimated between 2 and 3 seconds. Finally, after solving the stereo matching problem, the disparity of the matching points can be obtain.

Although the proposed ISMB have some achievements of stereo matching, it does not yet provide a full performance and several improvements to the ISBM may be possible. To enhance the performance, two works are considered in the future works: (1) using the analog Hopfield neural network to avoid the local minimum of the energy function. (2) to extract the edges as complementary features.

Reference

[1] K. Nishihara and T. Poggio, "Stereo vision for robotics," *Proc. Robotics Research, First Int. Symp. ,* pp. 489–505, 1984.

[2] D. J. Kriegman, E. Triendl and T. O. Binford, "Stereo vision and navigation in buildings for mobile robots," *IEEE Trans. Robotics and Automation,* vol. 5, No. 6, 1989.

[3] D. Murray and J. J. Little, "Using Real-Time Stereo Vision for Mobile Robot Navigation," *Autonomous Robots,* vol. 8, No. 2, pp. 161–171, 2000.

[4] M.Bertozzi and A. Broggi, "GOLD: A parallel realtime stereo vision system for genetic obstacle and lane detection, " IEEE Trans. Image Processing, vol. 7, pp. 62-81, Jan. 1998.

[5] L. Zhao, and C. E., Thorpe, "Stereo and Neural Based Pedestrian Detection," *IEEE Trans. Intelligent Transportation Systems*, Vol. 1, No. 3, pp. 148–154, Sep. 2000.

[6] W. Y. Yau and H. Wang, "Fast Relative Depth Computation for an Active Stereo Vision System," *Real-Time Imaging*, vol. 5, No. 3, pp.189–202, June, 1999.

[7] G. Medioni and W. H. Tsai, "Segment based stereo matching," *Computer Vision Graphics Image Process*, Vol. 31, pp. 2–18, 1985.

[8] Y. Shirai, *Three−Dimensional Computer Vision*, Springer−Verlag, New York, 1987.

[9] J. J. Lee, J. C. Shim and Y. H. Ha, "Stereo correspondence using the Hopfield

neural network of a new energy function," *Pattern Recognition* vol. 27, No. 11, pp. 1513−1522, Nov. 1994.

[10] W. E. L. Grimson, "Computational experiments with a feature based stereo algorithm," *IEEE Trans. Pattern Anal. Mach. Intell.* , vol. PAMI−7, pp. 17−34, 1985.

[11] G. Medioni and R. Nevatia, "Segment based stereo matching," *Comput. Vision Image Processing*, vol. 31, pp. 2−18, 1985.

[12] T. Ozanian, "Approaches for stereo matching − A review," *Modeling Identification Control*, vol. 16, pp. 65-94, 1995.

[13] G. Pajares and J. M. Cruz, "Local stereovision matching through the ADALINE neural network," *Pattern Recognition Letters*, vol. 22, pp. 1457−1437, 2001.

[14] J. Hopfield, and D. W. Tank, "Neural computation of decisions in optimization problems," *Biological Cybernetics*, Vol. 52, pp. 141−152, 1985.

[15] G. Pajares, J. M. Cruz and J. Aranda, "Relaxation by Hopfield network in stereo image matching," *Pattern Recognition*, vol. 31, No. 5, pp. 561−574, 1998.

[16] Y. Ruichek, "Multilevel- and Neural-Network-Based Stereo-Matching Method for Real-Time Obstacle Detection Using Linear Cameras," *IEEE Trans. Intelligent Transportation Systems*, Vol. 06, No. 1, pp. 54 − 62 , Mar. 2005.

[17] T. H. Sun, "Improving stereo matching quality with scanline-based asynchronous Hopfield neural networks," *Journal of the Chinese Institute of industrial Engineers*, Vol. 24, No. 1, pp. 50−59, 2007.

[18] N. M. Nasrabidi and C. Y. Choo, "Hopfield network for stereovision correspondence," *IEEE Trans. Neural Network*, Vol. 03, No. 1, pp. 123−135, Jan. 1992.

[19] K. Achour and L. Mahiddine, "Hopfield Neural Network Based Stereo Matching Algorithm," *Journal of Mathematical Imaging and Vision*, Vol. 16, No. 1, pp. 17−29, Jan. 2002.

[20] C. Harris and M. Stephens, "A combined corner and edge detector," *Fourth Alvey Vision Conference*, Manchester, UK, pp. 147−151, 1988.

[21] C. T. Lin and C. S. George Lee, Neural Fuzzy Systems A Neural−Fuzzy Synergism to Intelligent Systems, Prentice−Hall Inc, 1994.

[22] S. S. Yu and W. H. Tsai, "Relaxation by the Hopfield neural network," *Pattern Recognition*, Vol. 25, No. 2, pp. 197−209, 1992.

[23] M. Srinivas and M. Patnaik, "Genetic algorithms: a survey," *IEEE Trans. Computer*, Vol. 27, No. 6, pp. 17−26, 1994.

[24] A. Blanco, M. Delgado and M. C. Pegalajar, "A real-coded genetic algorithm for training recurrent neural networks," *Neural Networks*, Vol. 14, No. 1, pp. 93-105, 2001.