Artificial Neural Networks as Decision-Makers for Stereo Matching

Thabit Sultan Mohammed,
ECE Department,
Dhofar University. Engineering College,
P.O.Box 2509 , 211 Salalah, Oman .
e-mail: thabitsm@yahoo.com

Nedhal Ibrahim Al-Taie,
IT Department,
Salalah College of Technology,
P.O.Box 608 , 211 Salalah, Oman .
e-mail: nedhalatai@yahoo.com

Abstract— **This paper investigates the use of artificial neural networks to help making a decision on matching of stereo images. An image matching technique based on extracting features from segmented regions is adopted in this work, and a neural network framework is applied for region matching of stereo photographs. Two types of neural networks are used, the radial basis network, (RB) for learning clustering, and the back propagation (BP) network for learning image matching. The (RB) neural network is to cluster the regions according to the locations of their centered points. For each region, the BP network uses differential features as input training data. While training and testing the system, multiple features are extracted and used for enhancing the accuracy of the matching process. Features include (compactness, Euler number, and invariant moments) for each region. Results obtained from the neural networks (namely; clustering and initial matching array) are used to select the best matching pair. Results are showing a good matching accuracy.**

Keywords-- Stereo image matching, BP neural network, Decision maker, Invariant moments, Neural learning, Classification, Aerial image, Computer vision.

I.   INTRODUCTION

Decision making problems are described as choosing an action from possible alternatives using available information. Neural networks have emerged as an important tool for decision making, where they have been tested empirically, and applied successfully to a variety of real world decision making tasks in industry, business and science [1]-[7] .

In computer vision, image matching is among the long-standing problems. In literature, techniques for image matching can be categorized into two, a category for methods based on the use of point features and/or curve fragments for image matching, while the work in the other category is considered as region-based matching. Examples of the second category are those addressing stereo matching (e.g., [8],[9]), others can be found about motion/optical-flow analysis (e.g., [10],[11]) as well as object recognition (e.g., [12], [13]).

In this work, a region-based image matching approach is adopted. The technique relies on the fact that the higher dimensionality of regions makes them richer descriptors of target objects' geometric properties, such as size and shape. Other properties of regions include photometric properties such as gray-level contrast. In addition to this, higher dimensionality of regions makes their matching more stable

to small illumination and viewpoint changes across given images.

The process in which two separate image views of an object of interest are obtained is referred to as stereo imaging. Stereo imagining has a wide application in industry throughout computer vision, where the objective is image understanding to improve the productivity and quality management, and offering great advantages of flexibility, better performance, speed, low cost, reliability and robustness. In [14] a stereo vision machine was proposed, where it was claimed to be used in an industrial environment to perform different tasks such as 3D measurements, quality control and vehicle guiding.

Arial photographs are usually adopted for the analysis of photograph and civilization level in the sensed area where the analysis of man-made objects; such as buildings, houses and fields are of interest [13].
Stereo image matching of aerial photographs find its application in urban planning, disaster prevention, and insurance business [15]. This list of applications depends on site and/or building prosperities.   Another promising application is the study of the sea surface profile by analyzing two images of the sea surface. Results of such a study, help in discussing various phenomena in coastal zone, such as sediment transport and rip current [16].

The analysis of stereo photographs involves a series of steps comprising; image acquiring, camera modeling, feature acquisitions, image matching, depth finding and interpolation [17]. Image matching may be the most difficult task among these steps, where images of the same scene may be different due to the sensor motion and intensity changes.
As mentioned earlier, the two techniques that are used in solving the problem of image matching are region-based and a technique based on the use of point features and/or curve fragments. Region-based techniques utilize correlation between the intensity patterns in the neighborhood of a pixel in the left image and those in the neighborhood of a corresponding pixel in at the right image.

Region-based technique is preferred since using points or lines as matching primitives may lead to too many computational loads. For images of simply structured scene and of man-made object, segmented regions are preferable primitives for matching. The challenging task left, therefore is to select useful features of those segmented regions and use the features to achieve the proper matching. Among such useful features are the invariant moments, which were firstly derived by Hu [18] according to the mathematical fundamental of algebra and invariant to translation, rotation and scaling.
Because techniques based on extracting features of segmented regions allow simple comparisons between the attributes of a feature; they are generally faster and more accurate.

Many of the processes in image matching such as region classification and comparisons between the attributes of a feature requiring a powerful tool for making the proper decision. Neural networks are proposed and used here for such decision making. A neural network drives its computing power through its massively parallel-distributed structure and its ability to learn and, therefore generalize [19]. Generalization refers to production by networks of reasonable outputs for inputs not encountered during training. Many works have adopted neural networks in stereo image matching (e.g, [20] – [26]).

The neural network architecture adopted in our work incorporates a fast clustering neural network (named probabilistic neural which is a variant of radial basis net.) along with back-propagation (BP) neural network to achieve fast and high performance image matching.

The architecture will employ the capabilities of the BP in function approximation and generalization to learn the non structured knowledge required to achieve higher accuracy in image matching compared to area-based methods.

In the next section of this paper the main features extracted from segmented regions are presented. Section III, introduces the neural-based proposed system. Section IV illustrates the BP neural network architecture and the philosophy of employing neural networks in image matching. In section V, the matching process is presented. Experimental results and concluding remarks are presented in sections VI and VII respectively.

II. SEGMENTED REGIONS AND THEIR EXTRACTED FEATURES

Among the branches of image processing is the scene analysis or image analysis. The input to the image analysis process is a picture (mostly referred to as an image or a scene), but the desired output is a description of this picture. These descriptions are parts of the image named as regions or objects, and for the descriptions to be generated, it is necessary to segment the image into these parts.

In digital image processing, there are a number of interconnected processes. For instance, thresholding is a well-known technique that combines image segmentation. Labeling comes after image segmentation, where segmented regions are isolated and given labels, so that it is easy to deal with each one as a separate image.

A. Connected Components:

A connected component is a set of connected pixels that share a specific property. Two pixels, p and $q,$ are connected if there is a path from p to q of pixels.

In applications; such as visual inspection, optical character recognition, robot vision, and face smile coding system, detection of connected components is an essential step to be performed on their images, where they are in binary form.

If we assume that $p\,(i, j)$ to be a point of a given picture then such point will have four horizontal and vertical neighbors as well as four diagonal neighbors. Such a scheme is referred to as connected components [27].

B. Euler Number, Area, and Perimeter.

Euler number is a topological property unaffected by any deformation) of a region simply defined, as the difference between number of holes H and connected components C:

$$E = C - H \qquad (1)$$

In situations, where the size of the objects of interest is invariant, the area and perimeter are used. For a given region, the area is defined as the number of pixels contained within the region's boundary. The perimeter of a region is referring to the length of region's boundary. Both area and perimeter are sometimes used as descriptors, establishing a measure of compactness CO of a region, where:

$$CO = (perimeter)^2 \, / \, area \qquad (2)$$

C. Moment Representation of Regions.

The invariant moments are moment-based descriptors of planer shapes, which are invariant under general transitional, rotational, scaling, and reflection transformation [28].

The moments of a binary image f (i, j) are given by:

$$M_{pq} = \sum\sum i^x \, j^y \qquad (3)$$

In equ. (3) x and y define the order of moment.
Using object moments, the center of gravity of this object is given by:

$$x_c = m_{10}/m_{00} \qquad y_c = m_{01} \, / \, m_{00}$$

The central moments then are defined as:

$$M_{pq} = \sum_i \sum_j (i - x_c)^p (j - y_c)^q \, f(i, j) \qquad (4)$$

- The binary object area is represented by the zero-order moment.
- The distribution of matter around the center of gravity is expressed by the second-order moments. (In the case of objects with mass, they are called moments of inertia).
- Third-order moments express the basic properties of symmetry of object.
- Moments of higher order describe more slight variations in shape, but they are more sensitive to noise [29].

The normalized central moments denoted by η_{xy} are defined

as: $\qquad \eta_{xy} = m_{xy}/ \, (m_{00})^\gamma \qquad (5)$

Where $\qquad \gamma \qquad = (x + y/2) + 1$

III. THE NEURAL-BASED PROPOSED SYSTEM

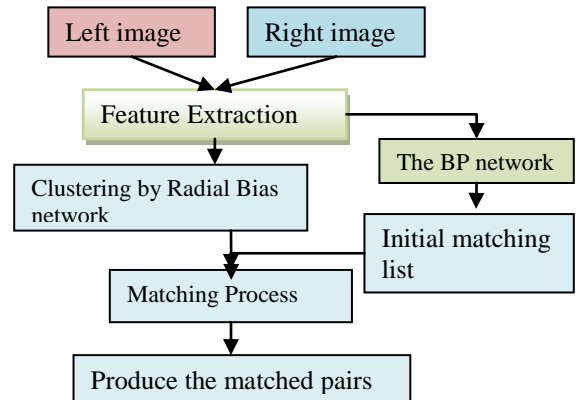The proposed system, is illustrated by the block diagram shown in Fig.1.

Figure 1. Block diagram of proposed image matching system.

In this system, the matching degree between any two regions is computed by employing a BP network. Three features for each isolated region in the image pair are extracted, including:

- Region compactness.
- Euler number.
- Sum of seven invariant moments.

These features will form the neural network input data.

IV. TRAINING A NEURAL NETWORK

Back propagation (BP) is one of the supervised learning methods, and is a generalization of the gradient descent learning rule. It requires a teacher that knows, or can calculate, the desired output for any input in the training set. The idea is to minimize the network total error by adjusting the weights.

To derive a mathematical equation for BP algorithm, whose neural network topology is shown in Fig. (2).
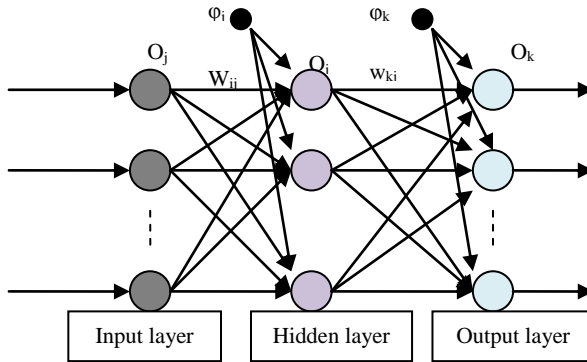

Figure 2. Back Propagation (BP) Architecture

In this figure:

- O_k, O_j, O_i = are output values at the output, hidden, and input layers.
- W_{kj} = connection weight from unit j (at hidden layer) to unit k (output layer).
- W_{ji} = connection weight from unit i (at input layer) to unit j (hidden layer).

A problem that is normally encountered with back-propagation is slow convergence or even non-convergence. To avoid this problem and to speed up the training, the learning rate is updated using the throsholding, increment-decrement method, where the idea is to decrease the learning rate by a specific factor when the error is more than a predetermined threshold, and increase it when the error is less than a threshold's value.

A. Radial Basis Network

A radial basis function network is a neural network approached by viewing the design as a curve-fitting (approximation) problem in a high dimensional space. Learning is equivalent to finding a multidimensional function that provides a best fit to the training data, with the criterion for "best fit" being measured in some statistical sense [30]. The probabilistic network is a type of RBF; this neural network is used in this paper as a decision maker for

clustering to assign each region to one cluster, depending on the location of its center of gravity with respect to the location of the centeroid of the whole image.

V. THE MATCHING PROCESS

To be used as a good decision maker, the back propagation (BP) neural network can be trained to approximate any functional mapping between well-defined input vectors and output vectors. In our work, this network is architected such as three neurons are used for the input layer (feature's difference from left and right image's regions), six neurons for the hidden layer with (*Sigmoid*) activation function, while one neuron is dedicated for the output layer with (*Linear*) activation function. A real number, ranging between 0.0 and 1.0, referred to as match_scale represents the output of this neural network. Few feature vectors of matched and unmatched region pairs are sampled as a training data. During training, whenever a match pair appears in the input, the neural network is taught to output a target value of 1.0 and feedback the error (error = 1.0 − match_scale) to the hidden layer, otherwise (error = match_scale − 0.0) when a mismatch occurs.

After training, the results are used to match a pair of stereo images. An algorithm summarized by the three next steps is applied to decide on the matching.

While (Regions exist) do
Step 1: Generate a (match_array) for each region in   the left image with degree within the threshold value.

Step 2: Assign each region in the two images to a different cluster depending on the location of its center of gravity with respect to the center of the whole image. The clustering result is incorporated   with the matching algorithm. It is noted that a region in the left image might have several regions with high matching scale in the right image. Those regions with match_scale $\geq$ 0.7 are denoted as the candidate list.

Step3: Compare the cluster of left region with cluster's numbers of regions in the candidate list to achieve the best matches, hence, only the region with the same cluster's number and match_scale $\geq$ 0.7 is considered as the matched pair.

VI. EXPERIMENTAL RESULTS

Among few experiments that have been performed using the proposed system, one is illustrated and the steps are explained.   Matlab environment is used, and necessary programs are written in C language.

A gray scale image pair used in training phase is shown in Figs. (3: a, b).

In Fig. (3: c, d), the segmented images for the image pair (using a threshold value of 0.62) are shown. After applying the connected components algorithm for image segmentation, the main nine regions with the largest areas are shown in Fig. (3: e, f). These regions are labeled as (1-9).

Small differences in the shapes of the regions, reflect the exactness of the segmentation algorithm.

These regions are used as inputs to the feature extraction algorithm. The output of this algorithm, for each region, are the following; (The compactness, Euler number, and the seven invariant moments).

The results of the feature extraction step are stored into vectors. As an example of these results, Table -I illustrates the values of the computed seven invariant moments of the image pair. Vectors formed from the extracted features are used as inputs to the BP and RB neural networks during training and simulation phases. During the training phase for the clustering RB network, the networks are fed with the vectors whose elements are the distances between center of gravity of image regions and that of the whole image.

TABLE I. VALUES OF SEVEN INVARIANT MOMENTS COMPUTED FOR BOTH RIGHT AND LEFT IMAGES.

Summary of seven invariant moments.	Right image	Reg.1	Reg.2	Reg.3	Reg.4	Reg.5
		27.8120	23.5112	23.3694	30.9753	43.9004
		Reg.6	Reg.7	Reg.8	Reg.9	
		52.5985	20.3432	21.7853	18.0193	
Summary of seven invariant moments.	Left image	Reg.1	Reg.2	Reg.3	Reg.4	Reg.5
		26.3315	22.3071	22.8803	29.0976	32.8840
		Reg.6	Reg.7	Reg.8	Reg.9	
		45.0245	44.8239	19.1661	9.4635	

For the BP network-training phase, the input vector elements are the difference of the three features between matched and unmatched pairs of regions from left and right images. The number of these vectors is 32, and each vector consists of three elements, which represents the number of neurons in the input layer. The target vector consists of 32 elements with a value of 1 for matched input pairs and 0 for unmatched ones.
During the training phase for BP network, three different algorithms are tested as shown in Table II.

Results from Table II illustrates that the conjugate gradient algorithm is the fastest one among the three tested training algorithms for BP neural network as it requires the least number of epochs for the network to converge.

TABLE II. DIFFERENT ALGORITHMS BEING TESTED DURING THE TRAINING OF THE BP NEURAL NETWORK

| Algorithm	Parameters used	Number of epochs for the net to converge
The standard BP algorithm	Learning Rate (lr) = 0.05 Momentum (mc) = 0.95 Error = 0.0001	71317
The variable learning rate algorithm	Initial learning rate (lr) = 0.06 Increased Learning rate (lr_inc) = 1 Decreased Learning rate (lr_dec) = 0.8	28615
Conjugate gradient algorithm	Magnitude of gradient (min_grad) = 10×10^8 Magnitude of step (min_step) = 10×10^6	655
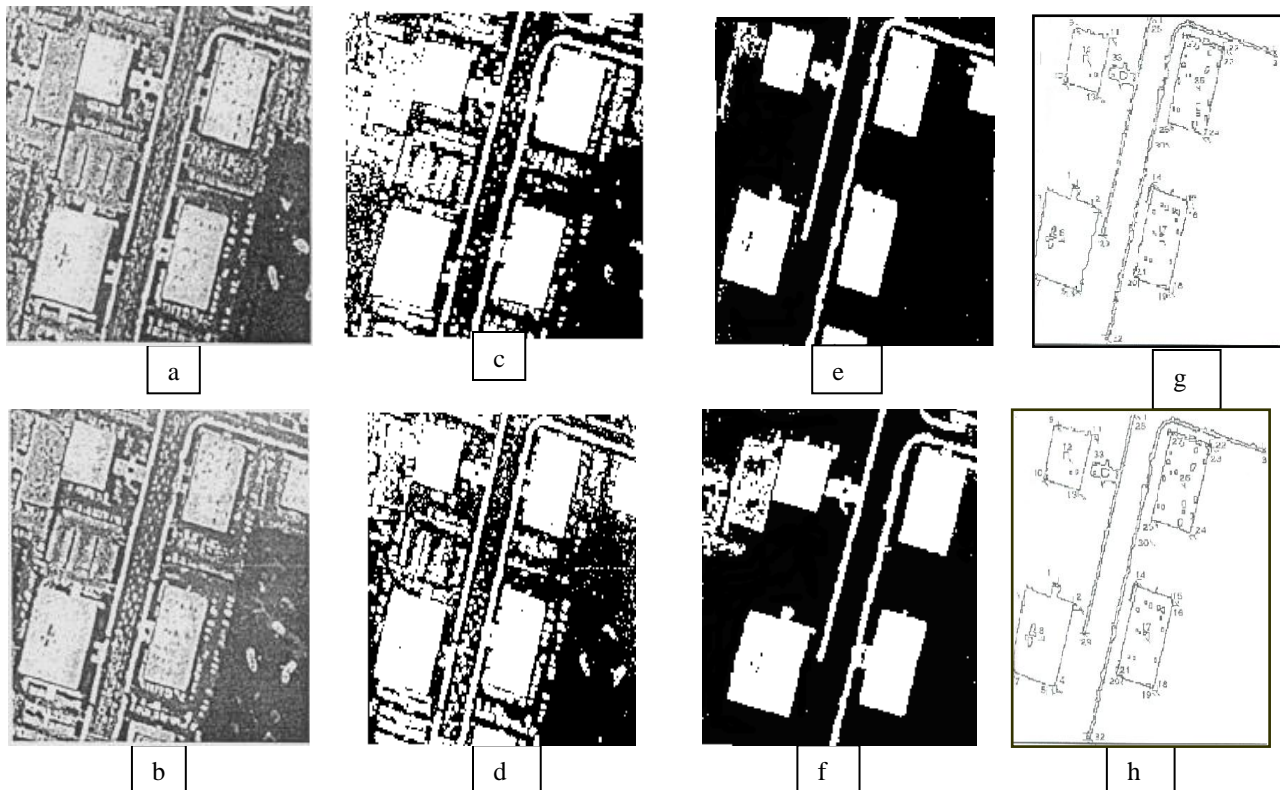
Figure 3. (a, b) The original images. (c, d) The segmentation results. (e, f) The main nine matched regions of images. (g, h) Overlapping points between the regions of the matched and unmatched pair.

After training; the trained network is simulated with the same stereo image pair used in training. Out of the nine largest pairs of regions, six are correctly matched.

The other three have no matched pairs and that's could be due to one of the following reasons:

- The region may not appear due to sensor motion.
- The shape of regions from one image differs from that of the same region on the second one and this is due to intensity changes, which affects the segmentation results.
- Shapes of regions from the image pair are having noticeable differences. The features extracted will exhibit similar differences. This is effecting the value of the matching-degree.

After the matching results being obtained, the refinement step was taken place to indicate the overlapped region between each pair of matched regions.

The corresponding points between the matched pair are then calculated. Fig. (3: g, h), show locations of these points into image pairs.

VII. Conclusions

Neural networks can be viewed as an enabling tool for decision making. The best match for a pair of stereo images is a challenging issue. In this research, a neural network based architecture is developed, where two types of neural networks are used. Back propagation (BP) networks are used to take a decision on the matching list and the radial basis (RB) networks are used for clustering process.

Fairly fast and still good clustering performance is exhibited by the RB networks using literally single feature (the center of gravity of region).

Good generality results for image matching and reduction in the time required for area-based and feature-based methods demonstrated using the BP neural networks.

In addition to the broad decision on matching of an image pair. Corresponding points between image pairs are also found. Such points can be used in a list of applications such as image registration, depth finding, and applications of computer vision. Other possible applications may include, urban planning, disaster prevention, insurance business throughout an automatic detection method of building properties, and the study of the sea surface profile. … etc.

The overall accuracy of the algorithm is highly dependent on the result of image segmentation and connected components steps, so these steps are investigated thoroughly in this work.

References

[1] E. B. Barlett and R. E. Uhrig, "Nuclear power plant status diagnostics using artificial neural networks," Nucl. Technol., vol. 97, pp. 272–281, 1992.

[2] H. Bourlard and N. Morgan, "Continuous speech recognition by connectionist statistical methods," IEEE Trans. Neural Nets, vol. 4, pp. 893–909, 1993.

[3] I. Guyon, "Applications of neural networks to character recognition," Int. J. Pattern Recognit. Artif. Intell., vol. 5, pp. 353–382, 1991.

[4] J. C. Hoskins, K. M. Kaliyur, and D. M. Himmelblau, "Incipient fault detection and diagnosis using artificial neural nets," in Proc. Int. Joint Conf. Neural Nets, 1990, pp. 81–86.

[5] J. Lampinen, S. Smolander, and M.Korhonen, "Wood surface inspection system based on generic visual features," in Industrial Applications of Neural Nets, Singapore: World Scientific, 1998, pp. 35–42.

[6] R. P. Lippmann, "Review of neural networks for speech recognition," Neural Computing., vol. 1, pp. 1–38, 1989.

[7] T. Petsche, A. Marcantonio, C. Darken, S. J. Hanson, G. M. Huhn, and I. Santoso, "An autoassociator for on-line motor monitoring," in Industrial Applications of Neural Networks, Singapore: World Scientific, 1998, pp. 91–97.

[8] Cohen, L., Vinet, L., Sander, P." Hierarchical region based stereo matching." In: Proc. IEEE Conf. Computer Vision Patt. Rec. (1989) 416–421.

[9] Randriamasy, S., Gagalowicz, A. "Region based stereo matching oriented image processing." In: Proc. IEEE Conf. Computer Vision Patt. Rec. (1991) 736–737.

[10] Ming-Hsuan, Y., Ahuja, N., Tabb, M." Extraction of 2nd motion trajectories and its application to hand gesture recognition." IEEE Trans. Pattern Anal. Machine Intell. 24(8) (2002) 1061–1074.

[11] Fuh, C.S., Maragos, P. "Region-based optical flow estimation." In: Proc. IEEE Conf. Computer Vision Pattern Rec. (1989) 130–135.

[12] Basri, R., Jacobs, D. "Recognition using region correspondences." Int. J. Computer Vision 25(2) (1997) 145–166.

[13] Keselman, Y., Dickinson, S." Generic model abstraction from examples.", IEEE Trans. Pattern Anal. Machine Intell. 27(7) (2005) 1141–1156.

[14] C. Garcia, " A Neural Network Based Stereo Vision Machine", Proc. of the International Conference on Engineering Applications of Neural Networks (EANN'96), pages 371-374, Londres, GB, 1996.

[15] Keiko Kokubu, and M. Kohiyama: "Automatic Detection of Building Properties From Arial Photographs Using Color and 3D Config", the 22nd Conference on Remote Sensing 5-9, November, 2001.

[16] Mamoru Arita, "Measurement of water surface using Stereo Matching", Proceedings of the 16th International Offshore and Polar Engineering Conference, Lisbon, Portugal, July 1-6, 2007.

[17] His-Jian, and Wen-Ling, "Region matching and depth finding for 3D objects in stereo Arial photographs", Patt. Recog., vol. 23, No. 12, 1990.

[18] Hu M-K: "Visual pattern recog. by moment invariants". Trans. Inf. Theory., 1962, IT-8: 179–187.

[19] Simon Haykin, "Neural Networks: a comprehensive foundation", 2nd ed. Prentice Hall, 1999, ISBN: 0-13-273350-1.

[20] Nidhal khalaf, "Automatic Stereo Image Matching via a Neural Network Framework", MSc Thesis, University of Technology, Baghdad, 1999.

[21] Jung-Hua Wang and Chih-Ping Hsiao, "On Disparity Matching in Stereo Vision via a Neural Network Framework", Proc. Natl. Sci. Counc. ROC(A) Vol. 23, No. 5, 1999. pp. 665-678.

[22] J. Neubert, T. Hammond, N. Guse, Y. Do , Y. Hu and N. Ferrier, "Automatic Training of a Neural Net for Active Stereo 3D Reconstruction", Proceedings of the 2001 IEEE Intrl. Confer. on Robotics and Automation, Korea, 2001.

[23] Elisabetta Binaghi , Ignazio Gallo, Giuseppe Marino, Mario Raspanti, "Neural adaptive stereo matching", Patt. Rec. Letters 25 (2004) 1743–1758.

[24] Ahmed F. Elaksher: "Multi-Image Matching Using Neural networks and Photogrammetric Conditions", The International Archives of the Photogrammetry, Remote

Sensing and Spatial Information Sciences. Vol. XXXVII. Part B3a. Beijing 2008.

[25] Te-Hsiu Sun, "Stereo Matching Using Synchronous Hopfield Neural Networks" , Journal of the Chinese Institute of Industrial Engineers, Vol. 26, No. 4, pp. 276-288 (2009).

[26] Houman Rastgar, "Stereo Matching by Neural Networks", MSc Thesis, ECE Department, University of Windsor, 2005.

[27] Rosenfield, A., and Kak, A. C: "Digital Picture processing", Academic press, New York, 1976.

[28] Price, K, and Reddy, R., "Matching segments of images", IEE Trans. Patt. Analysis Mach., 1979.

[29] Xiaplong Dai, and Siamak, K, "A feature-based image registration algorithm using improved chain-code representation combined with invariant moments" IEE Trans. Geo. and Rem., vol. 37, 1990.

[30] Broom Head, D.S., and Lowe, D. "Multivariable Functional Interpolation and Adaptive Networks Complex Systems", Complex Systems, Vol. 2, p. 321- 355, 1988.

Thabit Sultan Mohammed received his BSc degree in Electrical and Electronic Engineering form Bagdad, Iraq in 1983, MSc , and PhD degrees from Cranfield University, England in 1989, and 1992 respectively.

He is an assistant professor in Electrical and Computer Engineering Department, Dhofar University, Oman . his research interests are in speech and image recognition, information theory and coding, reliable software, and distributed computing systems.

Nidhal Khlaf recived the BS and MS degree in computer engineering in 1993 and 2003 from the University of Technology, control and systems department, Baghdad. From 1993 to 2000 , she was computer engineer in space research center, remote sensing department.

From 2007 till now she is working as lecturer in Salalah College of technology, information system department. Her current research interests include image processing, image matching, and edge detection.