Disparity Map Estimation with Neural Network

Nadia BAHA TOUZENE
LRIA Loboratory, Computer Science Department,
USTHB,Algiers, Algeria
nbahatouzene@usthb.dz

Slimane LARABI
LRIA Loboratory, Computer Science Department,
USTHB,Algiers, Algeria
slarabi@usthb.dz

Abstract— **This work aims at defining a new approach for a dense disparity map computing based on the neural networks from a pair of stereo images. Our approach has been divided into two main tasks. The first one deals with computing the initial disparity map using a neuronal method (BP). Whereas the second one presents a simple method to refine the initial disparity map using neural refinement so that an accurate result can be acquired. In the literature, the matching score is based only on the pixel intensities. We introduce in this work two additional features: the gradient magnitude and orientation of the gradient vector of pixels which gives a true degree of similarity between pixels. Experimental results on real data sets were conducted for evaluating the proposed method.**

Keywords- Neural network;disparity; stereovision.

I. INTRODUCTION

The issue of stereo correspondence is of great importance in the field of Machine Vision. It concerns the matching of points, or any other primitive, between a pair of images of the same scene. Assuming a calibrated stereo setup, matching points reside on corresponding horizontal lines. The disparity is calculated as the distance of these points when one of the two images is projected onto other. The disparity values for all the image points constitute the disparity map. Once the stereo correspondence problem is solved the depth of the scene can be estimated. The issue is of interest in the context of 3D reconstruction, Virtual reality, robot navigation and many other domains.

A great number of approaches for disparity map estimation have been proposed in the literature, including features-based [1, 2, 3], area-based [4, 5, 6], DSI-based [7, 8] and energy-based approaches [10, 11]. A survey for the different approaches can be found in [9]

The dense matching algorithms are classified in local and global ones. Local methods (area-based) trade accuracy for speed. The disparity computation at a given point depends only on intensity values within a finite window. In addition, the size of the window can significantly affect the matching accuracy. Kanade and Okutomi [5] have presented an adaptive window algorithm with impressive results. Global methods (energy based) are time consuming but very accurate. Their goal is to minimize a global cost function that combines data and smoothness terms [10,11].

In this paper, we propose a new approach for computing a dense disparity map based on the Artificial Neural Networks from a pair of stereo images. The goal is to combine the advantages of the area-based and feature-based methods. Our approach divides the matching process in two steps: initial matching and refinement of disparity map. Initial disparity map is first approximated by neuronal method so called back-propagation (BP) neural network and then a neural refinement method is applied. This step will allow us to refine the initial disparity so an accurate result can be acquired. This paper is organized as follows: section 2 presents the stages of the proposed method. In section 3, experimental results on real images are presented. Finally, we make concluding remarks.

II. PROPOSED APPROACH

In this section, we describe the basic steps of our approach. In this work, we assume that the images pairs are rectified. Thus the search for correspondences in the images can be limited to one dimension, ensuring a fast implementation of the stereo matching algorithm.

A. Detection of the Features Points and Matching

The proposed method starts by extracting a set of feature points from each of two images by using the detector of Harris [12]. These points are then matched using the normalized correlation named ZNCC [13]. In order to validate the corresponding between points, the matching is done from the left image to the right image and from the right image to the left image. A valid match is considered only for those points that yield the best correlation score [14]. In addition, these points will be used in the training of the BP network.

B. The differential input features

Because the intensity value of each pixel is sensitive to the change in contrast and illumination, and despite the fact that the intensity is the most common information used in the matching process, the gradient magnitude and orientation of the gradient vector of each pixel deserve more detailed investigation with regard to their use as features. In this work, three pixel properties will be used in the comparison. Each pixel in the left and right image has three basic features: intensity, gradient magnitude and orientation of the gradient vector. The Sobel operator [15] is used for features extraction. When the BP network is used to compute the matching degree between two pixels from the left and right images, each input vector in fact consists of data from two 7×7 windows (one from the left image, the other from the right) in which the centers are the pixels to be matched. In particular, the differences between the three features are calculated for each 7×7 window to form a 147-Dimensional input feature vector. That is:

$$\begin{gathered}\textit{Differences of Intensity}=f_{li}-f_{ri},\\ \textit{Differences of gradient magnitude}=|\nabla f_{li}|-|\nabla f_{ri}|,\\ \textit{Differences of Orientation}=\alpha_{li}-\alpha_{ri},\end{gathered} \quad (1)$$

Where f_{li} , $|\nabla f_{li}|$, α_{li} and f_{ri} , $|\nabla f_{ri}|$, α_{ri} are the intensity, gradient magnitude and orientation of the gradient vector values of the i^{th} pixel in the left and right images, respectively. Note that these differential features constitute the actual input vectors both in the training and in matching processes.

C. *Initial Disparity Map*

The disparity map calculation is expensive both in cost and in time. To this end, we propose in this paper the use of an Artificial Neural Network (ANN) to parallelize the calculation of various costs.

1) *The BP Network architecture:*

The BP network is first used to generate an initial disparity map which will then be used as a reference map for the subsequent matching process.

Assuming that the images pair is rectified, the disparity computation concerns thus two matched points which have the same abscise. At each pixel $p_i(x_i,y_i)$ of the W_l , the matched pixel $p_j(x_j,y_j)$ will appertains to the window W^d_r of the right image centered on $p_r(x_r,y_r)$ (see figure 1). The position of W^d_r depends on the disparity d of the pair (p_l , p_r) which varies from zero to d_{max}, where d_{max} represents the highest disparity value of the stereoscopic images. The relations which bind two matched points p_i, p_j of W_l and W^d_r are:

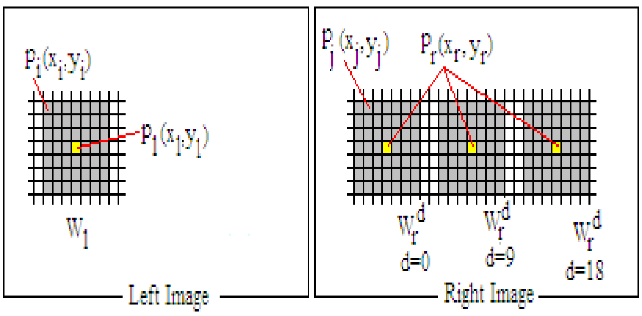

Figure 1. The windows used to the disparity computation.

$x_j = x_i+s*d$, $y_j = y_i$, where s = {-1, +1} is a sign chosen so that disparities are always positive.

This paper presents the use of the approximation capability function of the BP network to replace the traditional auto-correlation or SSD. Figure 2 shows the architecture of the three-layer BP network that we used in this paper. In our work, the correlation window size is 7x 7 in both left and right images, such as the matching is done between two pixels corresponding to the center of these two windows. In the literature, this score is based only on the pixel intensities. We introduce in this work two additional features: the gradient magnitude and orientation of the gradient vector of pixels.

Neurons in the input layer represent the absolute difference of the three features (intensity, module gradient and orientation) for each pair of pixels (see equations (1)). Therefore, we have 147 neurons (7 x 7 x 3). In order to simplify our system and to minimize the computation time, the activation function used in this layer is the linear function F (x) = x. Each neuron of the hidden layer is connected to three neurons of the input layer (intensity, module gradient and orientation) which gives us 49 neurons (147 / 3). The activation function used in this layer is the sigmoid function.

All neurons of the hidden layer are connected to the output neuron. The network gives us a score representing the degree of matching between two pixels known as: correlation score. And thanks to the ownership of the sigmoid function, the score is between 0.0 and 1.0 as the greater the degree is close to 1.0 over the two pixels are similar.

To calculate the primary disparity map, this process is repeated for each value of the disparity d and the disparity with the minimal cost among the various costs in the interval [0, dmax] will be chosen as the initial disparity of the pixel p_l.

The different correlation scores calculated for d_{max} disparities will be stored in a vector AC for use in the refinement stage.

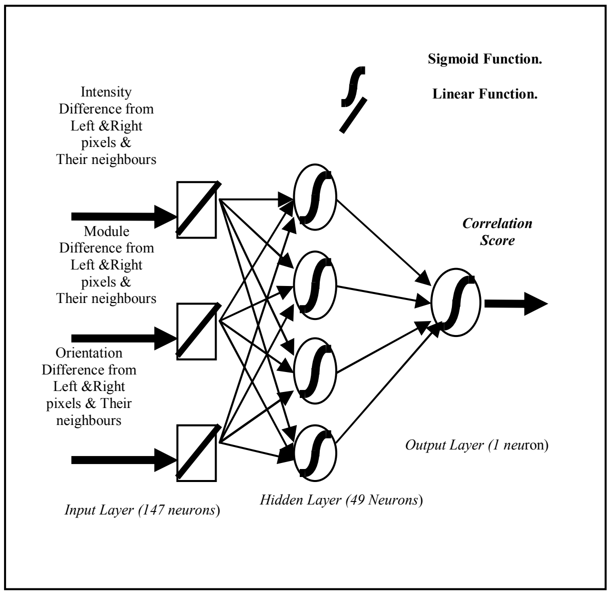

Figure 2. The BP neural network architecture.

2) *Neural Learning Procedure:*

The Neural correlation network must be trained with the learning procedure before computing the matching cost for each pixel pair. The first goal of neural learning can be formulated as the search for the most adequate weights. It achieved by applying the delta rule training algorithm [16]. To prepare the training data, hundred pairs of matched and unmatched pixels are randomly selected to offline train the BP network. During training, the differences of intensity, gradient magnitude and orientation of the gradient vector between two local windows are fed to the BP network. After the training, the BP network should have the ability to differentiate the matched pairs from unmatched pairs.

D. Disparity Map Refinement

After the initial disparity map is obtained by applying the BP network, a neuronal refinement method is used to refine this map in order to achieve more accurate disparity. In this section, we present our method to refine the initial disparity map. The neural network computes the different scores and from these scores we decide to modify the disparity or not.

Our neural network consists of two layers: input layer and output layer. The input layer has 49 neurons (7 x 7) where 7 x 7 is a shiftable window. The output layer has 1 single neuron. This architecture is illustrated by the Figure 3. We recall that in our method, each pixel has one vector AC of d_{max} values of scores previously computed (section C) such as the pixel for which we calculate its new disparity corresponds to the center of this window. Neurons in the input layer represent the values of the vector AC of pixels that belong to the shiftable window for the same disparity. All neurons in the input layer are connected to the output neuron which will give us the best score: thus the best disparity.

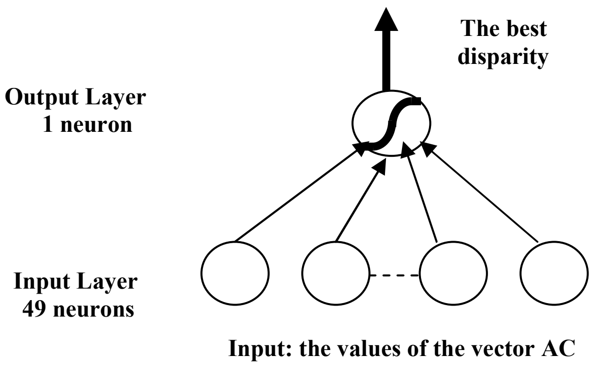

Figure 3. The refinement Neural Network architecture.

III. EXPERIMENTAL RESULTS

In the following, we will show how the proposed neural method can be applied to achieve accurate disparity map. After training, the BP network is first used to generate an initial disparity map which is refined by neural refinement method.

The BP network, we used had 7x7x3 input neurons, 49 neurons in the hidden layer and 1 output neuron. The network is trained to produce an output ≥ 0.9 when the two pixels are greatly matched.

In this section, we report the results of our approach using the standard data sets available at the Middlebury Web site [17]. As suggested on this site, we have tested our method on four images (Cones, Teddy, Venus, Tsukuba,). Figure 4 illustrates the results of computing a dense disparity map: from left to right: real images, ground truth, initial disparity maps, and final disparity maps. We also measured the time needed to process stereo images by our method. The timing tests were performed on a PC with a Pentium IV, 2.5 GHZ and visual studio 2005.We note, that the computation time mainly depends on the image size and it was about 109.71s per image pair having a resolution of 384 x288. The most expensive computation step is the initial disparity map calculated by the BP network.

IV. CONCLUSION

In this paper, we have presented a new approach for computing dense disparity map based on the neural network from pair of stereo images. The disparity map computing process is divided into two main steps. The first one deals with computing the initial disparity map by using a neuronal method (BP). The second one presents a method to refine the initial disparity map by using neural refinement so an accurate result can be achieved. As such, our method is more like a combination of the area-based and features-based method.

Our objective in this study was to investigate the potentialities of neural method in the field of disparity map computing. In the future work, and in order to reduce the computation time, we will implement the algorithm on a field programmable gate array: FPGA, the processing time can be further reduced.

REFERENCES

[1] W. E. L. Grimson, “Computational experiments with a feature based stereo algorithm,” IEEE Trans. Pattern Anal. Mach. Intell. 71, pp.17–34 1985.

[2] J. Y. Goulermas and P. Liatsis, “Hybrid symbiotic genetic optimisation for robust edge-based stereo correspondence,” Pattern Recognition. 34, 2477–2496,2001.

[3] C.J Harris and M.stephens, “Acombined corner and edge detector”, In Proc 4th Alvey Vision Conf,Manchester, pp 147-151.

[4] M. Gerrits and P.Bekaert, “Local stereo matching with segmentation-based outlier rejection”. In proceeding Conference on Computer and robot Vision. 2006.

[5] T. Kanade and M. Okutomi, “A stereo matching algorithm with an adaptive window: theory and experiment”. IEEE Trans. Pattern Anal. Machine Intell., 16, 920-932, 1994.

[6] S. S.Kumar and B.N. Chatterji, “Stereo matching algorithms based on fuzzy approach,” Int. Journal in Pattern Recognit. Artif. Intell. 16,7,pp. 883–899 2002.

[7] A.F Bobik and S.S. Intille, “Large occlusion stereo”, Intern. Journal on computer vision 33, pp. 181-200, 1999.

[8] E. Binaghi, I.Gallo, C.Fornasier and M.Raspanti “Growing aggregation algorithm for dense two-frame stereo correspondence”,in 1st Int.Conf. on Computer Vision Theory and Application, pp. 326-332,2006.

[9] D. Scharstein and R. Szeliski, “A taxonomy and evaluation of dense two-frame stereo correspondence algorithms,” Int. Journal on Computer Vision, 47, pp.7–42 ,2002.

[10] L. Alvarez, R. Deriche, J. Sanchez, and J. Weickert, “Dense disparity map estimation respecting image discontinuities: a PDE and scalespace based approach,” J. Visual Commun. Image Represent 13, pp.3–21,2002

[11] W.Miled and J.Pesquet , “Dense Disparity estimation from stereo images”, Int. Symposium on Image/Video Communication, 2006.

[12] C.J Harris and M. Stephens, “A combined corner and edge detector”, In Proc 4th Alvey Vision Conf., Manchester, pp 147-151.

[13] L.Zhong-Dan, “Methodes robustes en vision : application aux appariements visuels”. Thèse de Doctorat de L’INP de Grenoble 1997.

[14] N. Baha and S. Larabi, “Obstacle detection from uncalibrated cameras”, In Proc of PCI2008, Greece, pp 152-157, 27-29 August 2008.

[15] R.C Gonzalez and R.E. Woods,” Digital image processing”: Addison-Wesley Reading, MA, USA, 1992.

[16] H . Rumelhart, G.E.Hinton and R.J.Williams: “Learning Internal Representation by Error Propagation”, in Rumelhart H., McClelland J.L. (eds.), Parallel Distributed Processing, 318-362. MIT Press, Cambridge, MA. 1986.

[17] http://vision.middlebury.edu/stereo.

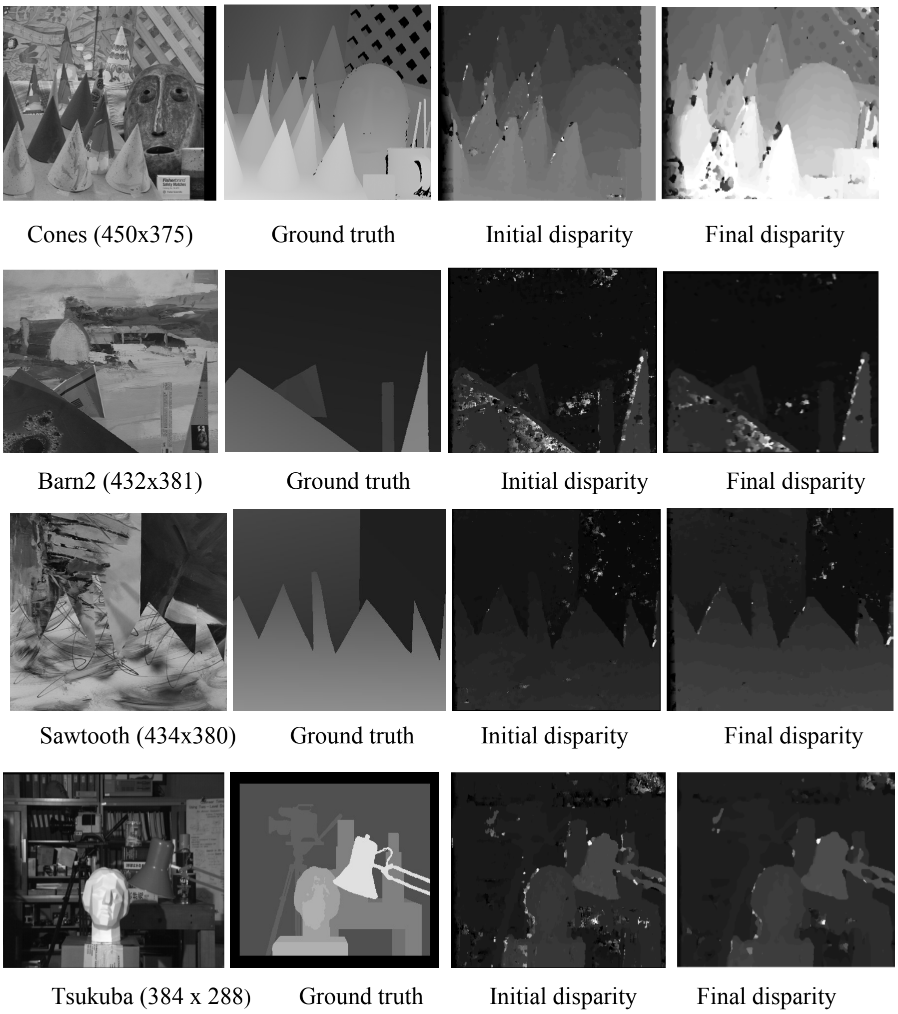

Figure 4. Results of final disparity map: from left to right, reference images, ground truth images,initial disparity maps, and final disparity maps.