Dense Two-Frame Stereo Correspondence by Self-Organizing Neural Network

Marco Vanetti, Ignazio Gallo, and Elisabetta Binaghi

Department of Computer Science and Communication
Università degli Studi dell'Insubria
via Mazzini 5, Varese, Italy

Abstract. This work aims at defining an extension of a competitive method for matching correspondences in stereoscopic image analysis. The method we extended was proposed by Venkatesh, Y.V. *et al* where the authors extend a Self-Organizing Map by changing the neural weights updating phase in order to solve the correspondence problem within a two-frame area matching approach and producing dense disparity maps. In the present paper we have extended the method mentioned by adding some details that lead to better results. Experimental studies were conducted to evaluate and compare the solution proposed.

Key words: Stereo matching, self-organizing map, disparity map, occlusions

1 Introduction

Stereoscopic image analysis deals with the reconstruction of the three-dimensional shape of objects in a physical scene from multiple 2-D images captured from different viewpoints [1, 2]. The accuracy of the overall reconstruction process depends on the accuracy with which the correspondence problem is solved. It concerns the matching of points or other kinds of primitives in two (or more) images such that the matched image points are the projections of the same point in the scene. The disparity map obtained from the matching stage is then used to compute the 3D positions of the scene points given the imaging geometry. A substantial amount of work has been done on stereo matching usually explored using area-based and feature-based approaches. Other types of stereo matching methods such as Bayesian, phase-based, wavelet-based and diffusion-based techniques have also been developed [3]. Despite important achievements, the high accuracy demand in diversified application domains such as object recognition, robotics and virtual reality [4] create the premise for further investigation. In previous works we investigated the potentialities of Neural Networks in Stereomatching basing our solutions on supervised learning [5, 6].

Motivated by the acknowledged biological plausibility of unsupervised neural learning, Venkatesh et al. in [7] explored the potential of Self Organizing Maps (SOM) to solve the correspondence problem conceived as imitation of the stereo-perception ability of the human visual system (HVS). In order to take care of stereo constraints, the authors introduced certain modifications within the original SOM model giving rise to the modified SOM (MSOM) model in which the estimation of the disparity map from

a stereo pair of images is obtained by computing the amount of deformation required to transform it into the other image. As seen in our experiments, the MSOM model has many special properties and potentialities, but also highlight limitations especially in dealing with occluded areas. Proceeding from these results, in this paper we proposed an extension of the MSOM for the estimation of stereo disparity. The salient main aspects of the solution proposed are the extension of matching primitives from pixel intensity to a weighted composition of RGB values, the definition of a contextual strategy within the matching cost computation task and finally the explicit handling of occlusions and direct processing within occlusion edges. The improved MSOM model was experimentally evaluated basing on the analysis of well known test images including data with true disparity maps. The aim of the experiment is twofold: to measure the performances of the model as functions of the most important parameters, and to compare performances with those obtained by well known approaches recently published in literature.

2 The MSOM model

This section describes the MSOM model proposed in [7] where the authors extend a Self-Organizing Map (SOM) neural network [8] by changing the neural weights updating phase. The base idea of the model MSOM is the following: *the matching between pixels of I_L and I_R, the left or reference image and the right or matching image in the stereo pair, is expressed in terms of the winning neurons in the network MSOM*.

The Algorithm 1 summarize the MSOM method through three main steps: *initialization*, *winner neuron selection* and *weights update*. During *initialization*, the neurons in the competitive layer are initialized with position and gray level of the pixels in the reference image I_L. During learning the input to the network is a randomly selected pixel $I_R(m, n)$ of the matching image. The learning phase proceeds searching the global minimum of a "Euclidean distance" between the input vector and the weights of the neurons (see equation (1)). If the coordinates of the winning neuron are (φ_r, φ_c), then we expect a matching between pixels $I_L(\varphi_r, \varphi_c)$ and $I_R(m, n)$. Finally, the *weights update* step, where only the first two components w_1 and w_2 of each neuron are updated (see equation (2)). From the trained MSOM model we can build the two disparity maps d^{hor} and d^{ver} of horizontal and vertical disparities respectively. In particular from the equation (4) is possible to construct the disparity values using the weights w_1 and w_2 of each neuron.

3 The StereoMSOM model

The method proposed here, the StereoSOM, extends the MSOM algorithm with the main aims to speed up the convergence and increase the precision of the computed disparity map. The complete description of StereoSOM model is showed in Algorithm 2.

To reduce the convergence time of the StereoSOM algorithm we adapted the algorithm MSOM to work with epipolar images and confined the search for the corresponding pixel in a limited area $\{n + d_{min}, \ldots, n + d_{max}\}$ (see the requirements of

Algorithm 1 - The MSOM algorithm

Require: To selection a reference image I_L and the corresponding matching image I_R having dimension $W \times H$;

Require: to create a matrix of neurons $n_{rc} = [w_1^{rc}, w_2^{rc}, w_3^{rc}]$ where $r = 1 \dots H$ and $c = 1 \dots W$;

Require: $\forall r$ and $\forall c$ initialize neurons n_{rc}: $w_1^{rc} = r$, $w_2^{rc} = c$ and $w_3^{rc} = I_L(r, c)$;

1: **for** $i = 1$ to $Iterations$ **do**

2: randomly extract a pixel $I_R(m, n)$, $m = 1 \dots H$ and $n = 1 \dots W$;

3: construct the corresponding input pattern $(i_1^{mn}, i_2^{mn}, i_3^{mn})$ where $i_1^{mn} = m$, $i_2^{mn} = n$ and $i_3^{mn} = I_R(m, n)$;

4: discover the coordinates (φ_r, φ_c) of the winning neuron as follows:

$$(\varphi_r, \varphi_c) = \underset{r \in \{1,\dots,H\}, c \in \{1,\dots,W\}}{\arg\min} \sqrt{\sum_{k=1}^{3} \left(w_k^{rc} - i_k^{mn}\right)^2} \tag{1}$$

5: update the weights of neurons n_{rc} as follows:

6: **for** $r = 1$ to H and $c = 1$ to W **do**

7: **for** $k = 1$ to 2 **do**

8: update the weight w_k^{rc} of neuron n_{rc} as follows:

$$w_k^{rc} \leftarrow w_k^{rc} + h_k(r, c)\, g_k(r, c) \left(i_k^{(r-(\varphi_r - m))(c-(\varphi_c - n))} - w_k^{rc}\right) \tag{2}$$

where

$$h_k(r, c) = \eta \exp\left(-\frac{(r - \varphi_r)^2 + (c - \varphi_c)^2}{2\sigma_h^2}\right) \tag{3}$$

and $g_k(r, c) = \exp\left(-\frac{\left(w_3^{rc} - w_3^{\varphi_r \varphi_c}\right)^2}{2\sigma_g^2}\right)$

9: **end for**

10: **end for**

11: **end for**

12: compute horizontal and vertical disparity maps:

$$\begin{cases} d^{hor}(r, c) = c - w_2^{rc} \\ d^{ver}(r, c) = r - w_1^{rc} \end{cases} \tag{4}$$

Algorithm 2 and equation (5)). Moreover the learning algorithm of our method is divided into two distinct phases that we called the *ordering phase* and the *tuning phase*, characterized simply by different values of the configuration parameters. The parameters setting in the ordering phase leads to an approximate solution in a very short time leaving to the subsequent tuning phase its refinement. In particular the tuning phase improves the first rough calculation, updating the neural weights with small entities.

The StereoSOM model considers as matching primitives the color attributes. For color images encoded in RGB, color information will be presented in input to the network with $i_{C_1}^{rc} = R, i_{C_2}^{rc} = G, i_{C_3}^{rc} = B$. The overall matching primitives are processed within the network weighting the relative importance of the individual components by means of appropriate weights ρ_k (see equation (5)).

Within the weight updating procedure, an explicit definition of the winner neuron's neighborhood is included. In particular the new function defined in (7) depends from three parameters α, β e n_{size} considering that the function $\theta\left(r,c\right)$ has the following form:
$\theta\left(r,c\right) = \alpha \exp\left(-\frac{(r-\varphi_r)^2+(c-\varphi_c)^2}{2\sigma_h^2}\right)$ with $\sigma_h^2 = \frac{n_{size}^2}{-2\ln\left(\frac{\beta}{\alpha}\right)}$. Assigning a value to these three parameters means that only an area of size $\left(2n_{size}+1\right)^2$ around the winning neuron will be updated.

The following subsections describe two relevant modifications of the MSOM algorithm. The first one concerns the introduction of a new strategy for the search of winning neuron called *Search Eye* (SE), while the second, called *Quality of Search* (QS) concerns the management of occluded areas.

3.1 Search Eye

The winning neuron strategy was updated basing on a moving window procedure: insted of comparing only the weights of the candidate neuron w^{rc}, an overall set of weights within a window is considered in the searching strategy (see (9)).

It is plausible to think that groups of neurons belonging to a single object on the scene have similar intensity and then similar disparities; adjusting the search window to neurons belonging to a single object will then improve the quality of the search process. The winning neuron search function is formalized as follows:

$$\left(\varphi_r,\varphi_c\right) = \left(m,\ \underset{c\in\{1,\ldots,W\}}{\arg\min} \sum_{\Delta r=-\xi}^{\xi}\sum_{\Delta c=-\xi}^{\xi} \sqrt{\rho_1\left(w_1^{rc}-i_1^{mn}\right)^2 + S(r,c,\Delta r,\Delta c)}\right) \quad (9)$$

with $S(r,c,\Delta r,\Delta c) =$
$\sum_{k=2}^{K+1}\left[g_s\left(r+\Delta r, c+\Delta c\right)\rho_k\left(w_k^{(r+\Delta r)(c+\Delta c)} - i_k^{(m+\Delta r)(n+\Delta c)}\right)^2\right]$
and $g_s\left(r,c\right) = \exp\left(-\frac{\sum_{k=1}^{K}\left(i_{C_k}^{rc}-i_{C_k}^{mn}\right)^2}{2\sigma_s^2}\right)$

3.2 Quality of Search

In order to deal with occlusions and false matching the StereoMSOM algorithm implements a Bidirectional Matching strategy. In the case in which the backward matching,

Algorithm 2 - The proposed StereoSOM algorithm for epipolar images

Require: To selection a reference image I_L and the corresponding matching image I_R having dimension $W \times H$;

Require: to create a matrix of neurons $n_{rc} = \left[w_1^{rc}, w_{C_1}^{rc} \ldots, w_{C_K}^{rc}\right]$ for generic images having K channels and where $r = 1 \ldots H$ and $c = 1 \ldots W$;

Require: $\forall r$ and $\forall c$ initialize neurons n_{rc}: $w_1^{rc} = c$, and $w_{C_k}^{rc} = I_L\left(r, c, C_k\right)$;

1: **for** $i = 1$ to $Iterations$ **do**

2: randomly extract a pixel $I_R\left(m, n\right)$;

3: construct the corresponding input pattern $\left(i_1^{mn}, i_{C_1}^{mn}, \ldots, i_{C_K}^{mn}\right)$ where $i_1^{mn} = n$;

4: discover the coordinates (φ_r, φ_c) of the winning neuron as follows:

$$(\varphi_r, \varphi_c) = \left(m, \underset{c \in \{n+d_{min}, \ldots, n+d_{max}\}}{\arg\min} \sqrt{\sum_{k=1}^{K+1} \left[\rho_k \left(w_k^{rc} - i_k^{mn}\right)^2\right]} \right) \tag{5}$$

5: update the weights of neurons n_{rc} as follows:

6: **for** $r = \varphi_r - n_{size}$ to $\varphi_r + n_{size}$ and $r = \varphi_c - n_{size}$ to $\varphi_c + n_{size}$ **do**

7: update the weight w_1^{rc} of neuron n_{rc} as follows:

$$w_1^{rc} \leftarrow w_1^{rc} + h\left(r, c\right) g\left(r, c\right) \left(i_1^{r(c-(\varphi_c - n))} - w_1^{rc}\right) \tag{6}$$

where

$$h\left(r, c\right) = \begin{cases} \theta\left(r, c\right) & if\ \beta < \theta\left(r, c\right) < 1 \\ 1 & if\ \theta\left(r, c\right) \geq 1 \\ 0 & if\ \theta\left(r, c\right) \leq \beta \end{cases} \tag{7}$$

and $g\left(r, c\right) = \exp\left(-\frac{\sum_{k=1}^{K} \left(w_{C_k}^{rc} - w_{C_k}^{\varphi_r \varphi_c}\right)^2}{2\sigma_g^2}\right)$

8: **end for**

9: **end for**

10: to compute the disparity map:

$$d^{hor}\left(r, c\right) = c - w_1^{rc} \tag{8}$$

from I_R to I_L fails, the algorithm nullify the weight updating and continue with the next input.

4 Experiments

The experimental activity was supported by test data available on the Web at `http://vision.middlebury.edu/stereo`. We selected four test data sets named *Tsukuba*, *Venus*, *Teddy* and *Cones* including stereo image pairs and true disparity (see Fig. 1).

Among the quality measures proposed by Scharstein and Szelinski in their papers [3, 9] we adopted the percentage of bad matching pixels between the computed disparity map $d_C(x, y)$ and the ground truth map $d_T(x, y)$:

$$PBP_{\delta_d} = (\frac{1}{N}\sum(|d_C(x,y) - d_T(x,y)| > \delta_d)) \tag{10}$$

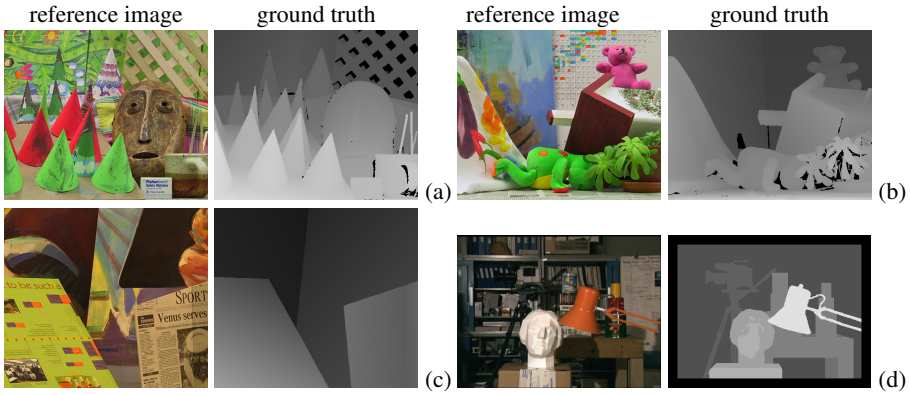

Fig. 1. Reference image and true disparity map of (a) *Cones*, (b) *Teddy*, (c) *Venus* and (d) *Tsukuba* test.

In this experiment the proposed StereoSOM algorithm was configured with the following set of parameters tuned with a trial and error procedure: for the ordering phase $\{\xi = 5, \sigma_s^2 = 700, g(r,c) = 1, \rho_1 = 0.001, \rho_2 = \cdots = \rho_{K+1} = 1, n_{size} = [80, 10], \alpha = 1, \beta = 1\}$ and for the tuning phase $\{\xi = 5, \sigma_s^2 = 700, \sigma_g^2 = 80, \rho_1 = 0.05, \rho_2 = \cdots = \rho_{K+1} = 1, n_{size} = 20, \alpha = [6, 1], \beta = [0.5, 0.005]\}$. The parameter expressed as $[a, b]$ vary linearly from a to b during the iterations.

Fig. 2 shows the final disparity after 10000 iterations of ordering phase and 500000 iterations of tuning. As regard the execution time, the average time over all the four dataset considered is approximatively 100 sec. with 50000 iterations of tuning, running the algorithm over a laptop with an AMD 1800 MHz processor.

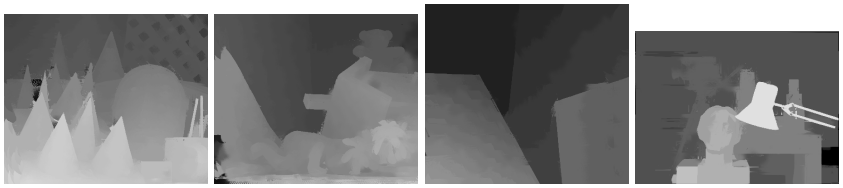

Fig. 2. Best disparity maps of the four dataset considered obtained by applying the proposed StereoSOM algorithm.

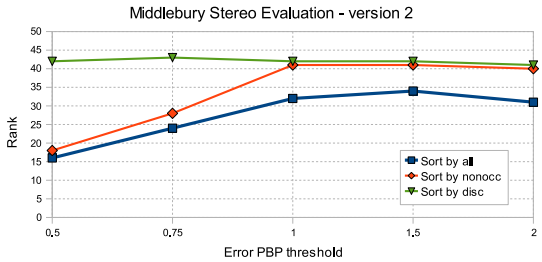

Fig. 3. Rank of StereoSOM algorithm obtained with the Middlebury Stereo Evaluation framewok. The comparison was made with the results of 54 different algorithms, already stored in the database of the framewok.

The disparity maps obtained (Fig. 2) were submitted to the Middlebury Stereo Evaluation tool, available on the Web at `http://vision.middlebury.edu/stereo/eval`, computing the PBP_{δ_d} measure over the whole image (ALL), in non occluded regions (NOCCL) and in depth discontinuity regions (DISC). The evaluation tool automatically compared other stereo matching algorithms whose performances are available on the same Web site.

The Fig. 3 shows the rank of our algorithm compared with the results of 54 different algorithms. Observing this figure we see a sharp rise in our ranking algorithm varying the threshold δ_d. This behavior leads us to believe that the proposed algorithm is particularly suitable in the management of disparity maps with real values. In order to improve furtherly the performance of the StereoSOM algorithm new solutions have to be investigated for managing discontinuity areas.

The complete set of results obtained is shown in Table 1. The StereoSOM algorithm shows a globally satisfactory competitive behavior, even if it did not prevail on some of the algorithms involved in the comparison. The comparison with the MSOM model was done using an our implementation of the algorithm and evaluating the disparity map through the same tool available from the Middlebury website. The comparison result is available on the Table 2.

Table 1. Results obtained by StereoSOM algorithm in terms of PBP_{δ_d} after 10000 iterations of ordering.

Test image	δ_d	NOCCL	ALL	DISC	Iterations	NOCCL	ALL	DISC	Iterations
Tsukuba	0.5	19.99	20.52	32.26	50000	15.20	15.82	29.33	500000
	0.75	17.35	17.77	28.44		13.28	13.75	26.29	
	1	3.38	3.76	14.54		3.80	4.12	15.16	
	1.5	2.79	3.09	12.09		3.24	3.50	13.00	
	2	2.22	2.47	9.44		2.52	2.73	9.99	
Venus	0.5	6.78	7.43	19.65	50000	5.12	5.83	18.98	500000
	0.75	1.66	2.22	12.35		1.27	1.75	11.03	
	1	0.98	1.42	10.31		0.83	1.19	9.24	
	1.5	0.66	0.99	7.53		0.53	0.83	6.23	
	2	0.53	0.79	6.20		0.44	0.65	5.09	
Teddy	0.5	17.14	23.58	34.36	50000	16.24	22.68	32.50	500000
	0.75	12.50	18.49	26.28		11.94	17.92	24.95	
	1	10.41	15.73	22.39		10.26	15.54	21.30	
	1.5	7.77	12.06	17.42		8.02	12.20	16.26	
	2	6.15	9.82	13.80		6.30	9.89	12.47	
Cones	0.5	12.3	18.74	26.89	50000	10.9	17.25	22.61	500000
	0.75	7.96	14.34	20.04		6.76	12.99	16.30	
	1	6.31	12.40	16.57		5.35	11.33	13.61	
	1.5	4.91	10.51	13.32		4.15	9.71	11.10	
	2	4.11	9.35	11.35		3.46	8.65	9.30	

Table 2. Results obtained by MSOM algorithm in terms of PBP_{δ_d} after 10000 iterations of ordering and 1000000 iterations of tuning.

Test image	δ_d	NOCCL	ALL	DISC	Test image	δ_d	NOCCL	ALL	DISC
Tsukuba	0.5	43.39	43.9	60.5	Teddy	0.5	33.76	38.72	54.82
	0.75	39.94	40.41	56.61		0.75	26.66	31.54	44.7
	1	22.97	23.6	45.9		1	23.62	28.11	39.72
	1.5	19.3	19.88	38.3		1.5	20.6	24.29	33.71
	2	9.16	9.68	27.38		2	18.75	22.07	29.73
Venus	0.5	31.57	32.36	53.98	Cones	0.5	40.56	44.95	62.9
	0.75	20.66	21.48	42.45		0.75	32.44	37.24	54.35
	1	15.17	15.91	36.24		1	27.93	32.83	48.66
	1.5	10.93	11.58	29.3		1.5	22.84	27.73	41.31
	2	7.97	8.46	22.82		2	19.6	24.37	35.67

5 Conclusions and Future Works

Our objective in this study was to investigate an extension of the existing method MSOM, aimed at solving the correspondence problem within a two-frame area matching approach and producing dense disparity maps. The new StereoSOM model was tested on standard data sets and compared with several stereo algorithms available in the Middlebury Web site. The proposed StereoSOM algorithm shows globally a satisfactory competitive behavior. Salients aspects of our solution are the local processing of the stereo images, the use of a limited set of directly available features and the applicability without the image segmentation.

In future works we want to improve the behavior of the StereoSOM method in discontinuities and occluded areas. Moreover a further investigation of the robustness under non epipolar conditions will be investigated.

References

1. Ballard, D.H., Brown, C.M.: Computer Vision. Prentice-Hall, Englewood Cliffs, NJ (1982)
2. Faugeras, O.: Three-dimensional computer vision: a geometric viewpoint. MIT Press, Cambridge, MA, USA (1993)
3. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International Journal of Computer Vision **47** (2002) 7–42
4. Bishop, C.M.: Neural Networks for Pattern Recognition. Oxford University Press (1995)
5. Binaghi, E., Gallo, I., Marino, G., Raspanti, M.: Neural adaptive stereo matching. Pattern Recognition Letters **25** (2004) 1743–1758
6. Gallo, I., Binaghi, E., Raspanti, M.: Neural disparity computation for dense two-frame stereo correspondence. Pattern Recogn. Lett. **29**(5) (2008) 673–687
7. Venkatesh, Y.V., Raja, S.K., Kumar, A.J.: On the application of a modified self-organizing neural network to estimate stereo disparity. IEEE Transactions on Image Processing **16**(11) (2007) 2822–2829
8. Kohonen, T.: Self-organizing formation of topologically correct feature maps. Biological Cybernetics **43**(1) (January 1982) 59–69
9. Scharstein, D., Szeliski, R.: High-accuracy stereo depth maps using structured light. Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on **1** (2003) I–195–I–202 vol.1