Performance Comparison of Stereo Correspondence Algorithms in Dense Image Matching

Seyyid Ahmed-Medjahed[1,*], Fatima Boukhatem[2]

[1] University of Relizane,
Algeria

[2] University of Djillali Liabes, Sidi Bel Abbes,
Algeria

seyyidahmed.medjahed@univ-relizane.dz,
fatima.boukhatem@univ-sba.dz

Abstract. Stereo matching is one of the most active research fields in computer vision. The aim of stereo matching is to find the corresponding points in two or more images that correspond to the same physical entity in the scene. In this paper, we deal the problem of stereo vision and precisely the dense stereo matching of images using correlation measures and other algorithms. We consider eight correlation techniques, subpixel estimation, dynamic programming and Hierarchical matching. The aim of this evaluation is to show the performance of each method in dense stereo matching images and in the 3D reconstruction. We also consider noisy image pairs with different noise level. The performance evaluation of each technique is conducted in term of: computational time, mean absolute error, mean relative error and the percentage of correct matches as well as wrong matches. A 3D reconstruction will be considered for the better methods.

Keywords. Computer vision, stereo matching, correlation, correspondence.

1 Introduction

Computer vision is a process of information processing that allows the machine to understand the reality by analyzing and interpreting the information.

It is a branch of artificial intelligence whose goal is to allow a machine understands what it "sees" when it is connected to one or more cameras. Stereo matching is one of the most difficult problems in computer vision. It is a fundamental low-level vision task for many applications such as autonomous vehicle navigation, objet manipulation, robot vision and biometrics [1, 11, 6]. Indeed, stereo matching is the fundamental precursor of 3D reconstruction from a stereo pair or sequence of images.

It consists to find the corresponding points in two or more images that correspond to the same physical entity in the scene. The result of image matching is the disparity map; this map contains the spatial shift for each pixel, but, it does not allow the knowledge of the 3d structure of the observed scene [7].

Local methods are based on an analysis of the neighborhood of the points to be matched. Several methods have been developed in this context, the most used are: SAD, SSD, ZNCC, etc. These methods are very effective when it comes to doing 3D reconstruction in images with low motion and a small variation of the camera's intrinsic parameters.

The advantage of these methods lies in the fact that they require few resources and they generate images with a dense disparity. But, they have a high rate of error especially in the occlusion zones and in the texture zones.

Table 1. Some correlation methods

Methods	Formulation
Cross Correlation	
NCC	$\mathrm{NCC}(w_l, w_r) = \dfrac{w_l \cdot w_r}{\|w_l\| \, \|w_r\|}$
ZNCC	$\mathrm{ZNCC}(w_l, w_r) = \dfrac{(w_l - \overline{w_l}) \cdot (w_r - \overline{w_r})}{\|w_l - \overline{w_l}\| \, \|w_r - \overline{w_r}\|}$
Distance and Locally Centered Distance	
SSD	$\mathrm{SSD}(w_l, w_r) = \|w_l - \overline{w_l}\|^2$
SAD	$\mathrm{SAD}(w_l, w_r) = \|w_l - \overline{w_l}\|$
ZSSD	$\mathrm{ZSSD}(w_l, w_r) = \|w_l - \overline{w_l}\| - \|w_r - \overline{w_r}\|^2$
ZSAD	$\mathrm{ZSAD}(w_l, w_r) = \|w_l - \overline{w_l}\| - \|w_r - \overline{w_r}\|$
LSSD	$\mathrm{LSSD}(w_l, w_r) = \left\|w_l - \dfrac{\overline{w_l}}{\overline{w_r}} w_r\right\|^2$
LSAD	$\mathrm{LSAD}(w_l, w_r) = \left\|w_l - \dfrac{\overline{w_l}}{\overline{w_r}} w_r\right\|$

Other types of approaches are used for matching, these approaches are grouped together as global methods. The use of this type of method always leads us to a problem of minimizing an energy function defined over the entire image.

In matching, one is placed in the case of a combinatorial optimization. In this paper, we demonstrate the performances of several stereo matching algorithms: correlation techniques, subpixel estimation, dynamic programming and Hierarchical matching.

The experimental results is conducted on image pair obtained from Middlebury data base with different noise level. We consider the mean absolute error, mean relative error, the percentage of correct matches and wrong matches, computational time, as the performance measures.

Finally, we want to assess the impact of noise on these algorithms as well as algorithms parameters such as the correlation window size and the correlation function. The rest of paper is organized as follows:

In the next section, an overview of the different stereo matching algorithms is given. In section 3, we present the evaluation criteria. In section 4, we discuss the results obtained by the different approaches and finally we conclude by some perspectives.

2 Matching Algorithms

In general, the matching consists in finding in the left and right images, the homologous primitives, that is to say, the primitives that are the projection of the same entity of the scene. There are two types of matching: Dense matching: for each point of an image we calculate its correspondent.

Scattered matching: the correspondent is calculated only for points describing particular properties (the points of contours, corner, etc ...). The results of the mapping are visually represented by an image called the disparity map.

Each pixel of the disparity map represents the distance between the pixel position of the left image and that of its corresponding in the right image. In this study, we consider the following stereo matching algorithms:

2.1 Dynamic Programming

Introduced by Bellman and Dreyfus, dynamic programming has an important role in optimization problem. In computer vision, the Viterbi algorithm is used to estimate the disparity map between two images. The purpose of this technique is to find the optimal path from one side of the image to the other by using the block matching metric as the cost function [13, 8]. Many researchers have been interested by improve the dynamic programming algorithm and defined an iterated algorithm [4], also, parallel algorithm [9, 2].
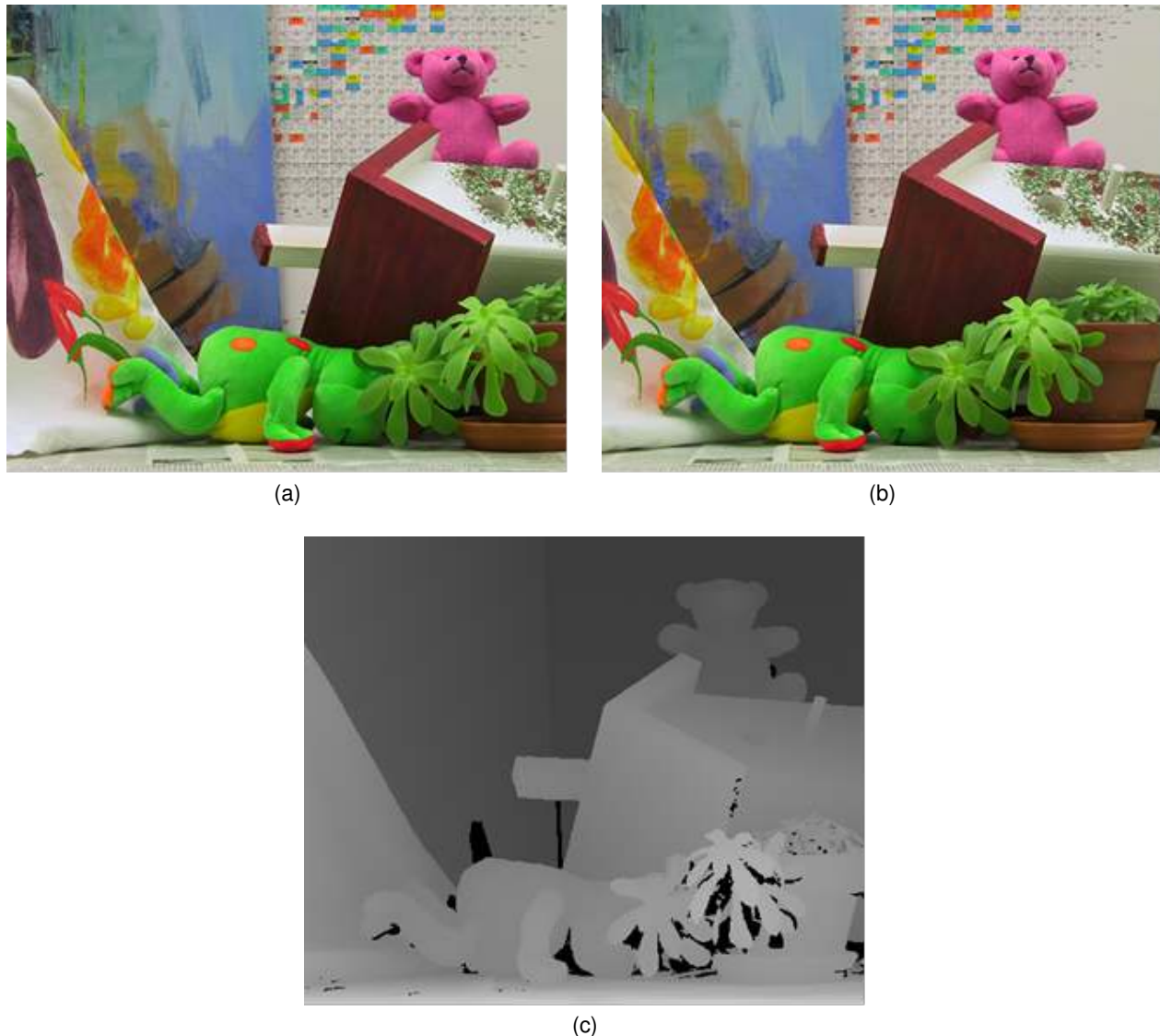
(a)

(b)

(c)

Fig. 1. The figure shows the left image (a), the right image (b) and the referenced disparity map (c)

2.2 Hierarchical Matching

Hierarchical Matching is one of the most used algorithms in stereo vision [14]. The basic idea of Hierarchical matching is the construction of pyramid of images.

Therefore, we reduce the pair of images: the top of the pyramid corresponds to the lower resolution and the base of the image corresponds to the original of image.

For example: if the original image is 2048×2048 pixels, we reduce the resolution to 16×16. In this case, the pyramid contains eight levels. To move from one level to other, we reduce the image by averaging the pixel values in the square $N \times N$. To compute the disparity map, we start from the top of the pyramid and lower resolution images are matched, we repeat this process until the higher resolution images, hence, obtaining the final disparity map [3, 12].
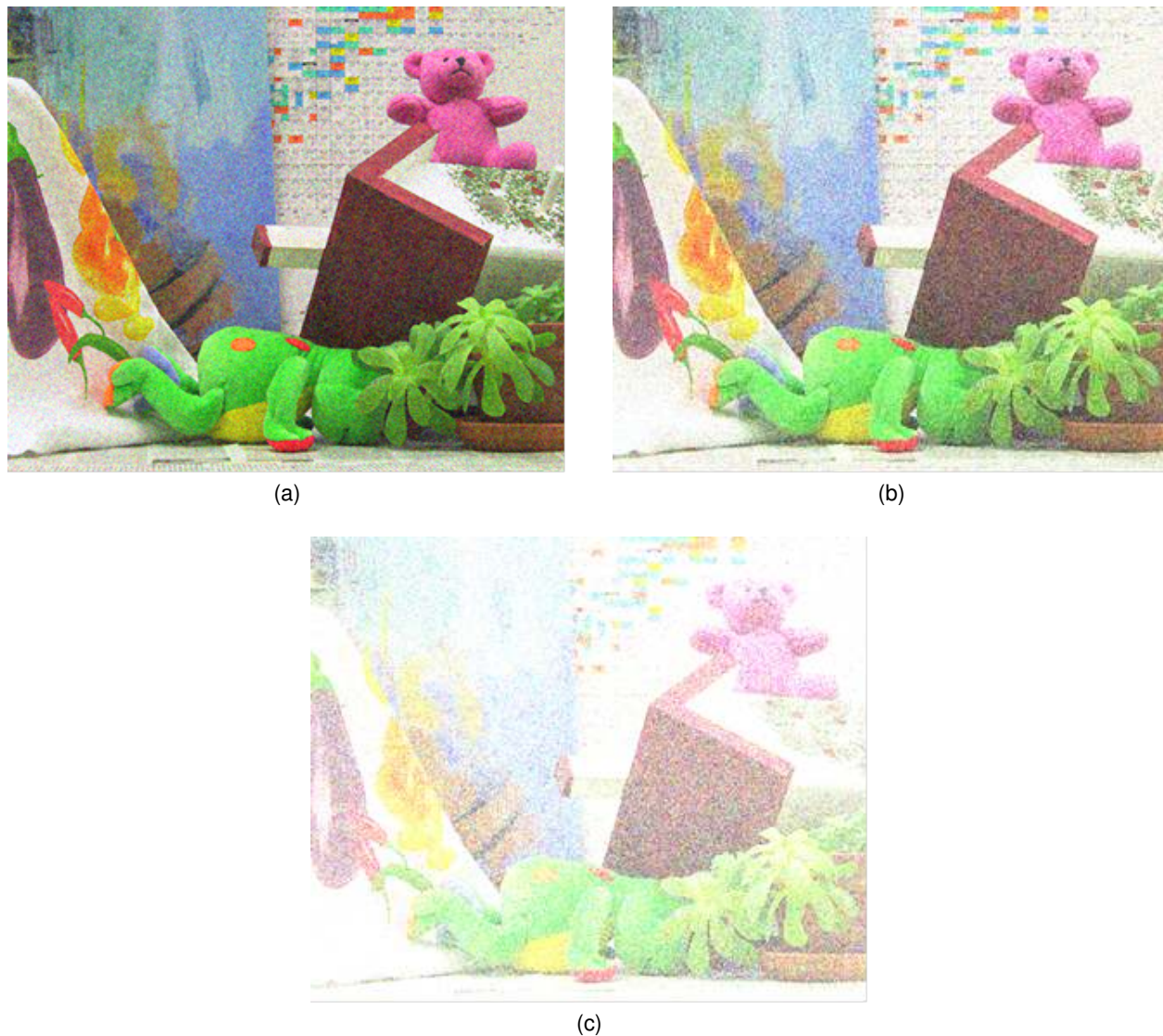
Fig. 2. The figure shows the left image (a), (b) and (c) with $0, 1$ variance, $0, 2$ variance and (c) $0, 3$ variance

2.3 Correlation Methods

The correlation techniques calculate the resemblance between two images. We consider a correlation window around the matched pixel. The correlation window is compared with a similar window in the right image at all along the epipolar line. Each movement of the window in the right image, a correlation index is calculated. The movement that minimizes the correlation index is retained as disparity. The correlation methods can be classified into two categories: Cross-Correlation and Distances measures. The values obtained by the NCC are in $[0, 1]$ and the ZNCC are in $[-1, 1]$.

ZNCC is very used in the literature and it corresponds to the classical linear correlation in statistics [5]. The distances measures are also very used. The disparity values are in $[0, \mathrm{DisparityRange}]$ [10].

Table 2. The results obtained by the correlation methods. (D. P. is the Dynamic Programming and H. M. is the Hierarchical Matching)

Methods	MAE	MRE	PCM	PWM
SAD	1,79	0,015	96,69	3,31
ZSAD	0,95	0,010	98,28	1,72
LSAD	1,03	0,011	98,11	1,89
SSD	1,58	0,013	97,06	2,94
ZSSD	1,16	0,011	98,06	1,94
LSSD	1,26	0,012	97,85	2,15
NCC	1,24	0,012	97,87	2,13
ZNCC	0,91	0,009	98,30	1,70
SubPixel	3,69	0,031	93,56	6,44
D. P.	2,61	0,021	95,56	4,44
H. M.	8,67	0,047	83,82	16,16

3 Evaluation Criteria

In this study, we propose to use four performance measures to evaluate the disparity provided by the algorithms. The first one is the mean absolute error which is the mean of the absolute error between the theoretical disparity map and the computed disparity map. The mean absolute error is defined as follows:

$$\text{MAE} = \frac{1}{N_i \times N_j} \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} \left| d_c(p_l^{i,j}) - d_t(p_l^{i,j}) \right|, \quad (1)$$

where $d_c(p_l^{i,j})$ is the computed disparity map $d_t(p_l^{i,j})$ and is the theoretical disparity map. Also, we consider the mean relative error that is given by:

$$\text{MEE} = \frac{1}{N_i \times N_j} \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} \left| \frac{d_c(p_l^{i,j}) - d_t(p_l^{i,j})}{d_t(p_l^{i,j})} \right|. \quad (2)$$

Moreover, we calculate the percentage of correct matches:

$$\text{PCM} = \frac{C}{N_i \times N_j}. \quad (3)$$

And the percentage of wrong matches:

$$\text{PWM} = \frac{W}{N_i \times N_j}, \quad (4)$$

where C is the number of pixel 2 and W is the number of pixels > 2. It is very important to note that the occlusions $(d_c(p_l^{i,j}) = 0$ and $d_t(p_l^{i,j}) = 0)$ are not taken into account in the calculation of MAE, MRE, PCM and PWM. The computational time is also used as an evaluation criterion. It allows to show the efficient of each technique in term of timing.

4 Experimental Results

In this section, the performance evaluation of each stereo matching algorithm is presented. We conducted the experimentation in terms of:

– Mean absolute error,

– Mean relative error,

– Percentage of correct matches,

– Percentage of wrong matches,

– Computational time.

We consider other important evaluation criteria which is the visual aspect of disparity map. The different disparity estimate algorithms are implemented on a PC with an Intel i32.13 Ghz, 4 GB RAM by using the Matlab and we use the computer vision toolbox for the 3D reconstruction.

The experimentation is conducted on the Middlebury image called "teddy" with a size of 450×375 pixels and 72 pixels/inch for the resolution. The ground truth for the images are knows[1]. We use a 9×9 for the correlation window size and $[0, 50]$ for the disparity range.

We also consider the SAD for the dynamic programming and hierarchical matching. Figure 1 shows the left and the right images with the theoretical disparity map. The aim of this study is to evaluate the performance of matching algorithms and show their efficiency on different noise level.
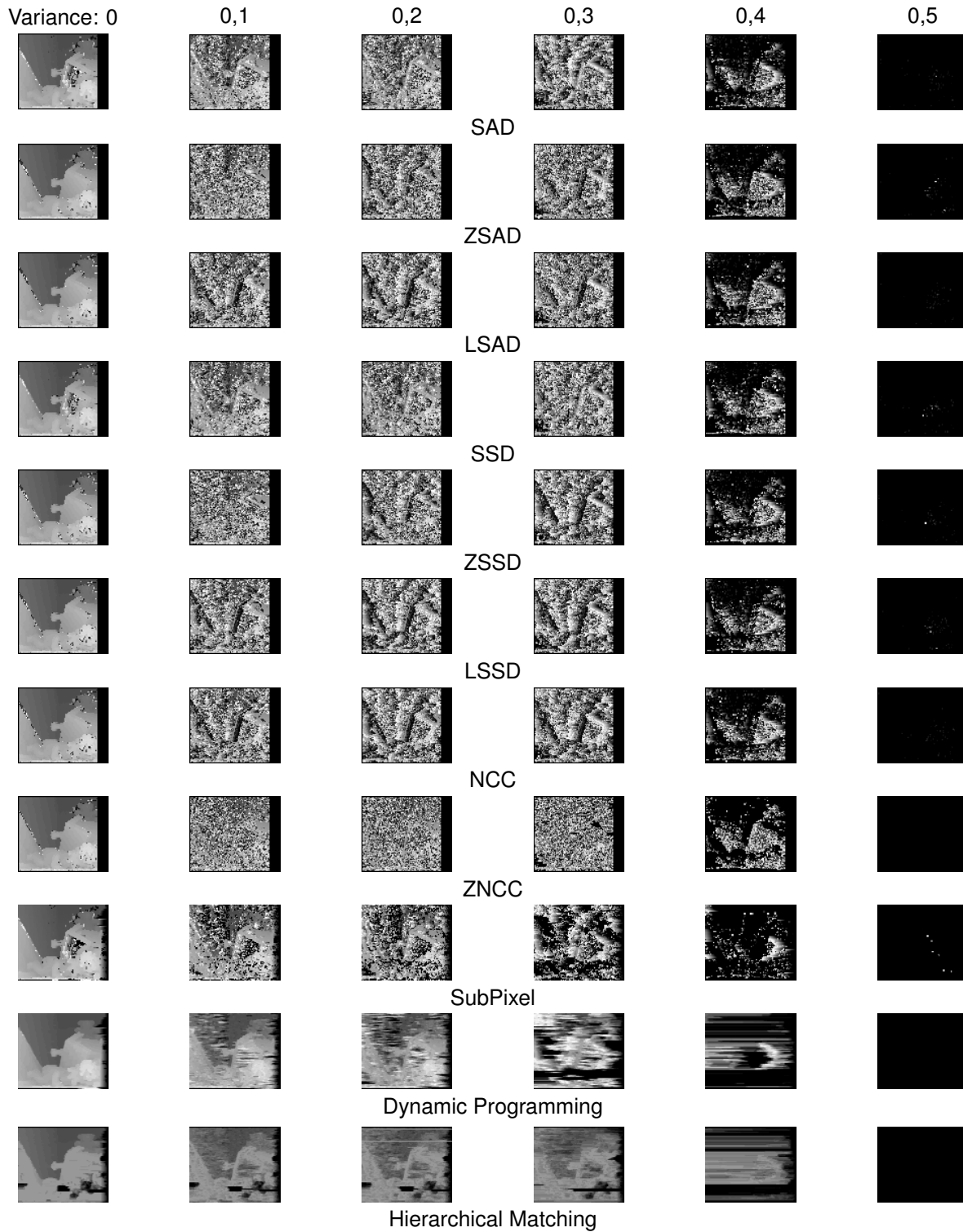
---

[1]www.middlebury.edu/stereo/data.html

Fig. 3. The visual result of stereo matching. Disparity maps computed with different matching algorithms
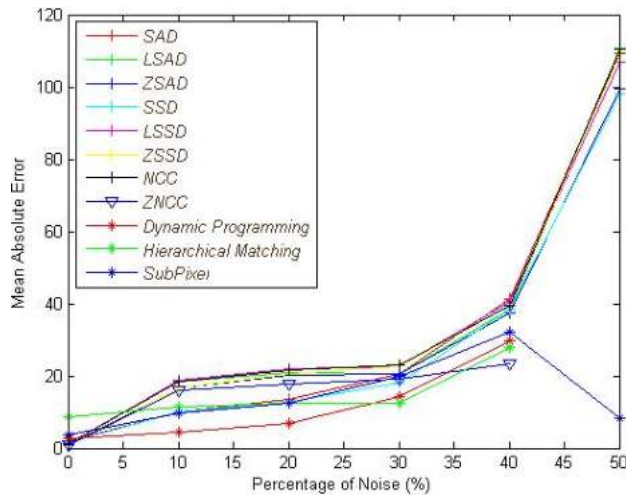
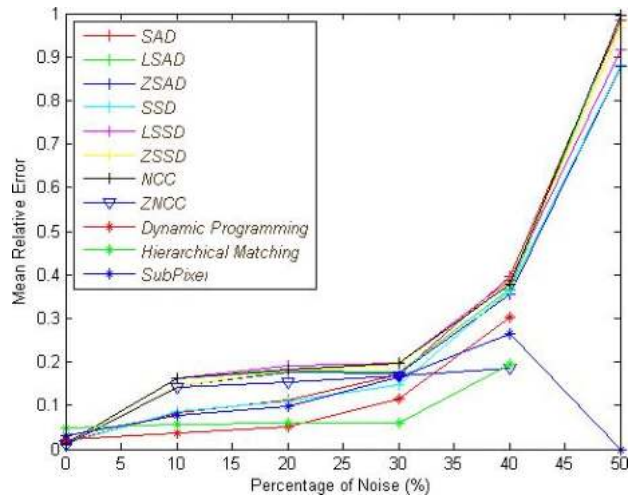**Fig. 4.** The mean absolute error VS the percentage of noise obtained by each algorithms

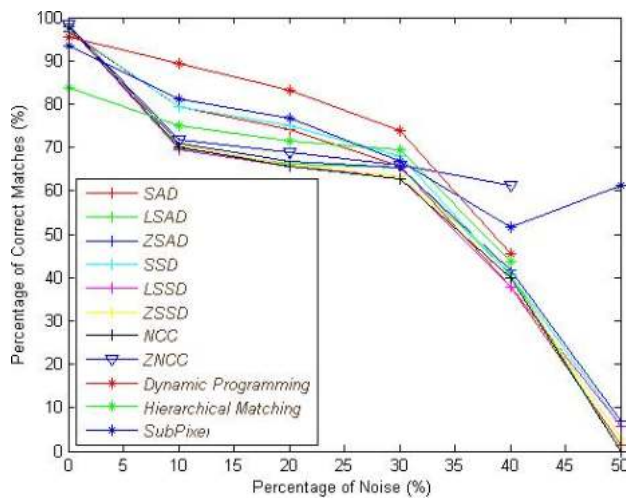**Fig. 5.** The mean relative error VS the percentage of noise obtained by each algorithms

Fig. 6. The percentage of correct matches VS the percentage of noise obtained by each algorithms
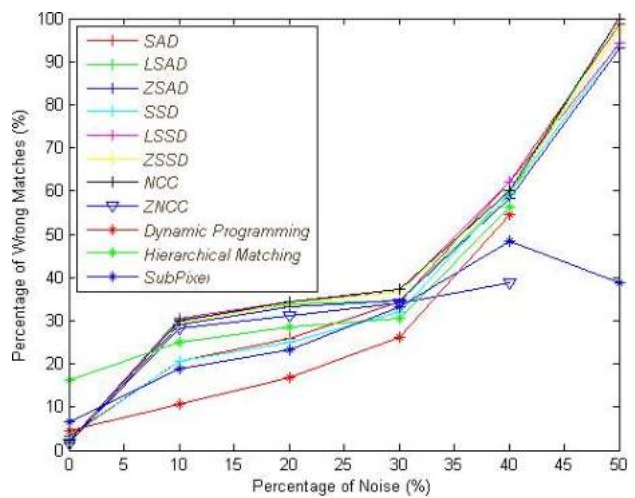

Fig. 7. The percentage of wrong matches VS the percentage of noise obtained by each algorithms

We use the gaussian noise with different variance : $0, 1 - 0, 2 - 0, 3 - 0, 4$ and $0, 5$. Figure 2 illustrates the different left images obtained by using different noise levels. Note that the right images have the same noise levels. In the figure 3, we show the visual results of stereo matching obtained by the matching algorithms defined above. The analyse of the visual results obtained by the different stereo matching algorithms shows that the numbre of occlusions points increase when the variance of the noise change. Therefore, if the left and the right images contain a high noise level, the number of occulions point increase significatly; also; with $0, 4$ and $0, 5$ variances of gaussian noise, we observe clearly that the disparity map is black. This means that the scene of left image is occulted. Visually, the dynamic programming and the hirearchical matching algorithms provide a smooth disparity map. The smoothing of the disparity map give a good depth map with noise
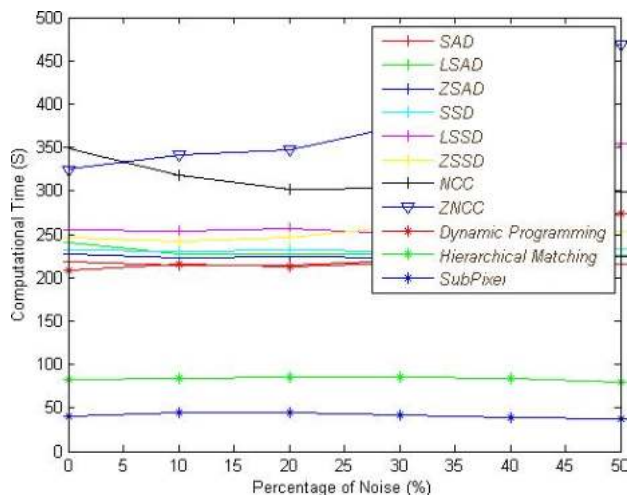
Fig. 8. The computational time VS the percentage of noise obtained by each algorithms
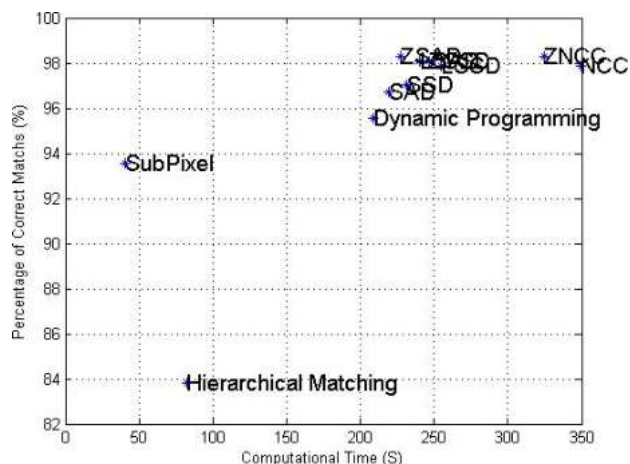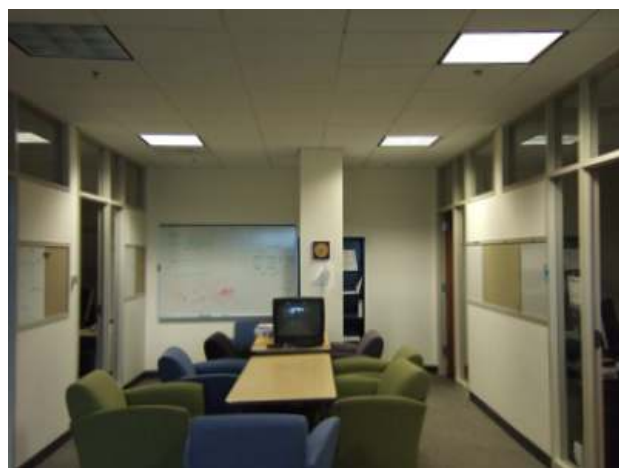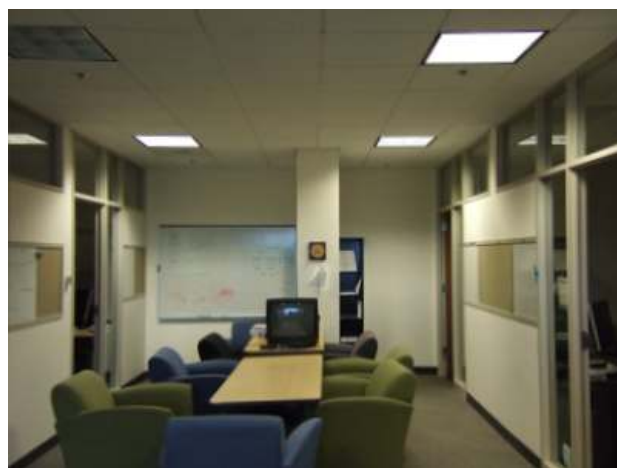

Fig. 9. The computational time VS the percentage of correct matches


Left image   Right image

Fig. 10. The left image and right image

images. ZNCC technique gives a low quality of disparity map when the variance of noise is $0, 3$ and $0, 4$ compared to others. To validate these results, numerical results are considered. Table 2 describes the numerical result obtained by the above algorithms. We calculate the MAE, MRE, PCM and PWM by using the left and the right images without noise. In term of mean absolute error, the low value is recorded for the ZNCC technique with $0, 009$ for the mean relative error and $98, 30\%$ of correct matches.

We remark that ZNCC technique provide a good results with images without noise but, it is not very perfect with noise images. The high percentage of correct matches is also recorded for the ZSAD, ZSSD with an advantage of ZSAD $98, 28\%$ of correct matches. The hierarchical matching algorithm provides a bad results, we estimate that this results is due to the fact that the low resolution of the image has given a bad disparity map. Figures 4, 5, 6 and 7 illustrate the MAE, MRE, PCM and PWM vs the percentage of
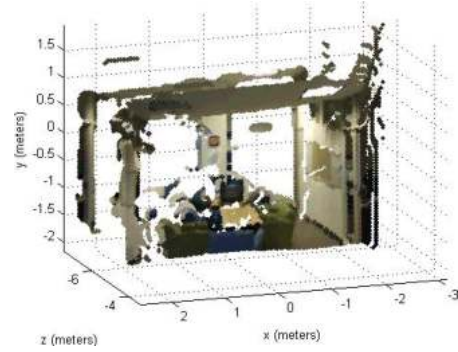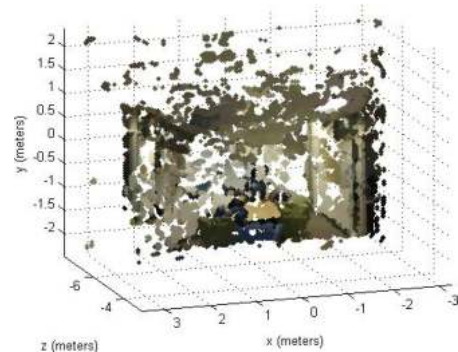
Disparity map and 3D reconstruction without noise images
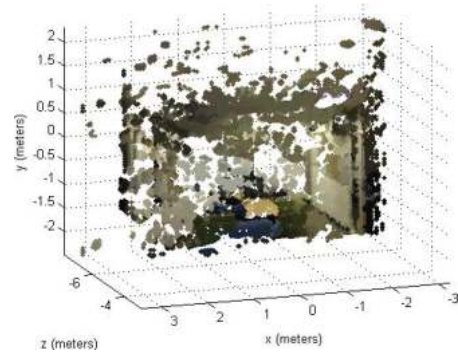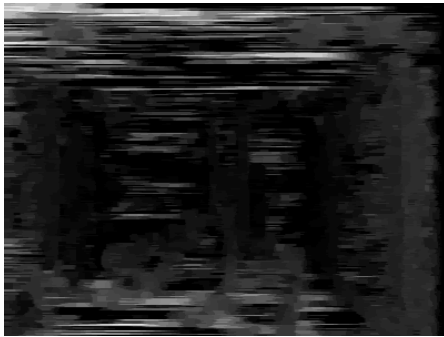
Disparity map and 3D reconstruction with variance 0,1 of noise
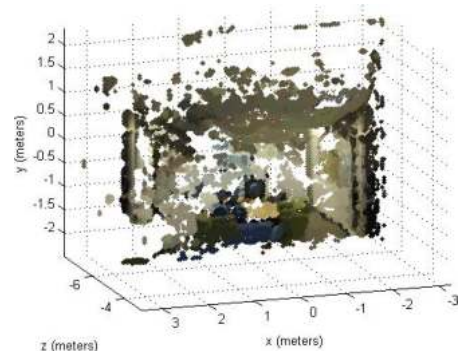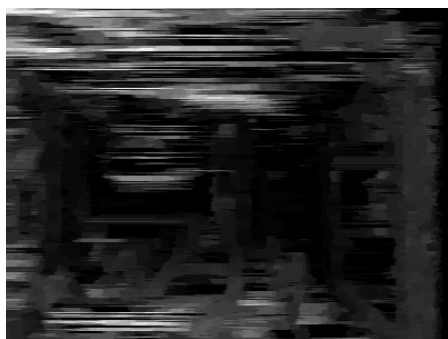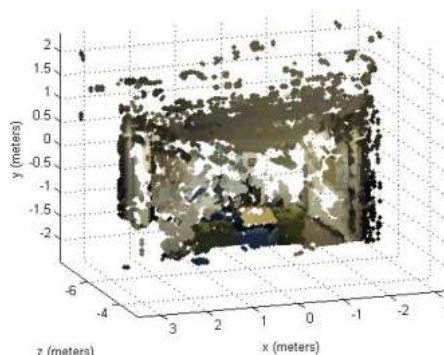

Disparity map and 3D reconstruction with variance 0,2 of noise

Disparity map and 3D reconstruction with variance 0,3 of noise

Disparity map and 3D reconstruction with variance 0,4 of noise

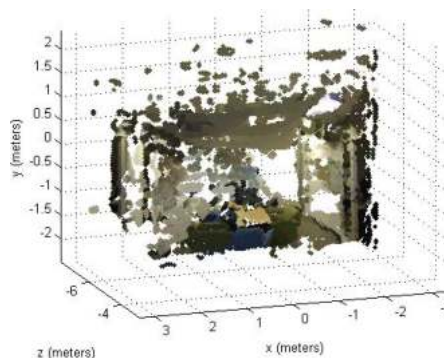Disparity map and 3D reconstruction with variance 0,5 of noise

Fig. 11. Disparity map and 3D reconstruction from the dynamic programming algorithm at different noise levels

gaussian noise applied to the left and right images. Figure 4 describes the different percentage of noise used in this study. It clearly visible that the lower value is recorded for dynamic programming in each noise level.

There is no value for hierarchical matching, ZNCC and pynamic programming for 50% percentage of noise, this means that the disparity map when the left and the right images is noised (50%) are black (occulted scene).

The same remark is observed in figure 6 and 7. We cleary show that the percentage of correct matches is higher in each noise levels when we use the dynamic programming algorthim.

Generally speaking, the dymanic programming algorithm provide a very good results with noise images. It is very interesting to evaluation the efficiency of these matching algorithms in term of computational time.

Figure 8 show the computational time of each methods versus the percentage of noise. The lower computational time is recorded for the subpixel and hierarchical matching. The slow method is ZNCC and NCC. To show the efficiency of the algorithms, we plot the percentage of correct matches versus the computational time.

The most efficient algorithm is the one with small computational time and with a high percentage of correct matches. In other words, the algorithm which appear further to the top left of the figure is the better one.

Finally, it is noted that the ZNCC gives better results when used with no noisy image, whereas, the dynamic programming provides better results when used a noisy image.

This is why, we choose dynamic programming algorithm for 3D reconstruction. We use the computer vision matlab toolbox to realize the 3D reconstruction.

We consider the "hallway" images of computer vision matlab toolbox with know intrinsics matrix:

Focal_length_x	=	409,4433
Focal_length_y	=	416,0865
Camera_center_x	=	204,1225
Camera_center_y	=	146,4133

Following figures represent the left image and the right image used in the 3D reconstruction[2].

5 Conclusion

In this paper, we analyzed and evaluated the performance of some stereo matching algorithms. We proposed a comparison protocol for the efficient and comparing their efficiency and performances in term of: absolute and relative errors, accuracy of matches by calculating the percentage of correct and wrong matches.

Also, we consider the computational time and the robustness of each algorithm to noise. The experimentation has been conducted by using the teddy images obtained from middle bury data set. We used different percentage of Gaussian noise applied to the left and right images and we calculate the disparity maps. Finally, we proposed a 3D reconstruction using the better disparity map obtained from the methods.

The results has shown that the ZNCC has given a very satisfactory results when it was used with images without noise. Dynamic programming has provided a good results, it is a very robust against noisy images. In term of computational time, the low time was recorded for subpixel estimation and hierarchical methods.

References

1. **Ali, H., Nema, B. (2009).** Multi purpose code generation using fingerprint images. The International Arab Journal of Information Technology, Vol. 6, pp. 141–145.

2. **Kolesnik, M. I. (1993).** Fast algorithm for the stereo pair matching with parallel computation. Computer Analysis of Images and Patterns, pp. 533–537. DOI: 10.1007/3-540-57233-3_70.

3. **Koschan, A., Rodehorst, V., Spiller, K. (1996).** Color stereo vision using hierarchical block matching and active color illumination. Proceedings of 13th International Conference on Pattern Recognition, Vol. 1, pp. 835–839. DOI: 10.1109/ICPR.1996.546141.

4. **Leung, C., Appleton, B., Sun, C. (2008).** Iterated dynamic programming and quadtree subregioning for fast stereo matching. Image and Vision Computing, Vol. 26, No. 10, pp. 1371–1383. DOI: 10.1016/j.imavis.2007.11.013.

5. **Lhuillier, M., Guo-Quan, L. (2004).** Reconstruction quasi-dense de modeles 3d partir d'une sequence d'images. Actes du Congres AFRIF-AFIA Reconnaissance des Formes et Intelligence Articielle, RFIA, Vol. 2, pp. 895–904.

6. **Ouali, M. (2012).** A Markov random field model and method to image matching. International Arab Journal of Information Technology, Vol. 9, No. 6, pp. 520–528.

7. **Ouali, M. (2012).** Performance evaluation of stereo matching algorithms in the lack of visual features. International Journal of Computer Applications, Vol. 53, No. 5, pp. 7–11. DOI: 10.5120/8415-0636.

8. **Park, C. S., Park, H. W. (2001).** A robust stereo disparity estimation using adaptive window search and dynamic programming search. Pattern Recognition, Vol. 34, No. 12, pp. 2573–2576. DOI: 10.1016/s0031-3203(01)00016-4.

9. **Pissaloux, E. E., Le-Coat, F., Bonnin, P. J., Bezencenet, G., Durbin, F., Tissot, A. (1997).** Very fast dynamic programming-based parallel algorithm for aerial image matching. Automatic Target Recognition VII. DOI: 10.1117/12.277122.

---

[2]www.mathworks.com/products/computer-vision/

10. **Seong-Ku, J., Lee-Mu, K., Uk-Lee, S. (1998).** Multi-image matching for a general motion stereo camera model. Proceedings 1998 International Conference on Image Processing, Vol. 2, pp. 608–612. DOI: 10.1109/icip.1998.723543.

11. **Szeliski, R. (2022).** Computer vision: Algorithms and applications. Springer International Publishing. DOI: 10.1007/978-3-030-34372-9.

12. **Thevenaz, P., Ruttimann, U. E., Unser, M. (1998).** A pyramid approach to subpixel registration based on intensity. IEEE Transactions on Image Processing, Vol. 7, No. 1, pp. 27–41. DOI: 10.1109/83.650848.

13. **Veksler, O. (2005).** Stereo correspondence by dynamic programming on a tree. Vol. 2, pp. 384–390. DOI: 10.1109/CVPR.2005.334.

14. **Won, K. H., Jung, S. K. (2011).** hSGM: Hierarchical pyramid based stereo matching algorithm. pp. 693–701. DOI: 10.1007/978-3-642-23687-7_62.