

# THREE-DIMENSIONAL RECONSTRUCTION USING THE DEPTH MAP

<sup>1</sup>A.El abderrahmani , <sup>2</sup>R.Lasri and <sup>3</sup>K.Satori

<sup>1,2</sup>Advance Technology Lab, Department of Computer Sciences, Larache Poly  
Disciplinary School, Abdelmalek Essaâdi University

<sup>3</sup>LIAN, Department of Mathematic & Computer Sciences Dhar-Mahraz  
Sciences School, FEZ, MOROCCO

## *ABSTRACT*

*This paper presents an approach to reconstructing 3D objects based on the generation of dense depth map. From a two 2D images (a pair of images) of the same 3D object, taken from different points of view, a new grayscale image is estimated. It is an intermediate image between a purely 2D image and a 3D image where each pixel of this image represents a z-height according to its gray level value. Our objective therefore is to play on the precision of this map in order to prove the interest and effectiveness of this map on the quality of the reconstruction.*

## *KEYWORDS*

*Dense reconstruction; depth map; disparity map; camera parameters.*

## 1. INTRODUCTION

Obtaining 3D models of very high quality and more accurate is the main concern of researchers of 3D reconstruction domain. This is why several methods are proposed in the literature [1, 2, 3]. Most of these methods rely on a classical reconstruction process that provides us with only a scattered set of 3D points. This does not give us enough information about the scene. Hence, the interest of enriching this 3D structure by other information contained in the depth map what is the subject of this paper where we present the different steps to estimate a dense depth map. However, the quality of the depth map and the performance of the process of creating this map are our main objective in this paper.

Our work is structured around four sections. In the second section, we present a synthesis of existing approaches in the literature concerned with the estimation of the depth map. Then, we describe in the third section the different steps to build the depth map. The experiments and interpretation are the subject of the fourth section. Finally, the fifth section provides a conclusion to this work.

## 2. RELATED WORK

The depth information is a key and essential element in several fields of research such as video processing [4], visual communication [5, 6], computer vision [7, 8], and many others areas. Its importance encouraged researchers to work on the accuracy of this information by presenting the latter in the form of a map, called depth map, where the value of each pixel of this map corresponds to its depth.

In all these researches interested in the depth map, we find in the literature two categories of the researchers, the first category is interested in the exploitation of the depth map [9, 10, 11] and the second is interested only to the improvement of the quality of this map [12, 13, 15, 16], so that other researchers use it in their research.

Among the researches that used the depth map are the one based on image-based rendering (IBR) techniques [9]: in this work, Kim et al. found that they needed accurate depth information in order to generate reliable and accurate multiview intermediate images for use in a multiview 3D display system. It is for this reason that they decided to exploit the depth map. In their article, they carried out an experiment to estimate the depth map quantization for multiview intermediate image generation using depth image-based rendering (DIBR). This DIBR synthesizes several virtual views of a 3D scene from a 2D image and its associated depth map.

Using the vanishing point, Kim et al. [10] proposed a new 3D panorama system based on the generation of the depth map to restore a 3D space structure from the 2D images. The points of intersection of the vanishing lines detected indicate the vanishing points.

In order to describe the human actions Li et al. [11] also exploited the depth map, they presented a real-time human action recognition system that uses the depth map sequence as input. The proposed algorithm uses only depth information for applications where environmental illumination is weak or changing. This algorithm is based on depth continuity, which is the most essential attribute of objects in depth maps.

For the second category, several efforts have been made in recent years to estimate the depth map whose main objective is to improve the quality of this map. As an example, Pascal Fua [12] proposed a correlation algorithm that reliably produces much denser depth maps with little false correspondence and in the presence of depth discontinuities and occlusions. This algorithm aims to match each point of the image and uses a consistency criterion to reject invalid matches. This criterion is designed so that when the correlation fails, instead of producing an incorrect response, the algorithm does not return a response. Subsequently, in order to calculate dense depth maps, the author proceeds to combine the depth map produced by correlation and the gray level information present in the image itself to introduce depth discontinuities and to adjust a surface, which is smooth in pieces.

With the same objective, Zhang et al. [13] have developed a new system for the estimation of high quality and high resolution depth maps by the common fusion of stereo data and Kinect depth sensor [14]. The fusion problem is formulated as a maximum a posteriori probability (MAP) estimation problem and the MAP problem is solved using a multi scale belief propagation (BP) algorithm.

Similarly, Malik and Choi [15] presented a new focus measure for the estimation of the depth map using image focus. This depth map can then be used in techniques and algorithms leading to the recovery of a 3D structure of the object. This new measurement aims to determine the best number of frames for each pixel, that is to say the frame where the pixel is best focused. In other words, only the frame corresponding to the best focusing value is selected for each pixel, and all the other frames in which the pixel is less focused are ignored.

For a successful 2D-3D conversion, depth information is required. Then, for this reason Yang et al. [16] proposed an interactive method of depth map generation from a single image for 2D-3D conversion using a local depth hypothesis. The use of a depth variation hypothesis can reduce human effort to generate a depth map. The only thing required from a user is to mark some salient

areas to be distinguished with respect to depth variation. The proposed algorithm assumes hypothesis of each salient area and generates a depth map of an input image.

### 3. STEPS FOR DEPTH MAP ESTIMATION

In order to generate the depth map, the disparity map is first calculated. Given a pair of stereoscopic images, it is possible to calculate a dense disparity map, which encodes the correspondences per pixel between two views of the same scene. Given the calibration parameters of a pair of cameras, it is possible to transform a disparity map into a depth map. The steps for estimating the depth map is as in Fig. 1.

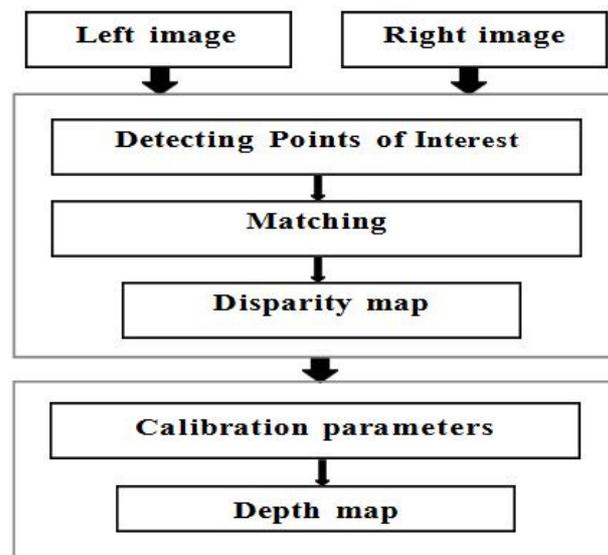


Fig. 1 Steps for estimating the depth map

#### 3.1 Calculation of the Disparity Map

A disparity map is a collection of distances of correspondence between the homologous visual indices of two images studied. We call visual index any object extracted from the image and containing in a compact way the information relevant to its analysis [17]. These indices may be points of interest, regions or contours. The points are mainly present, specific and numerous which give them more advantages over the contours and regions. In this work, we will use the primitive point.

To generate this map several steps to follow:

##### 3.1.1 Detecting Points Of Interest

This step involves extracting points of interest: Points of interest are defined as points that have characteristics that distinguish them from other points in the image. To detect these points, a point of interest detector is used. This detector consists of calculating a response value, representing the interest for each pixel of the image and then selecting the best ones. In this work, we will use the Harris detector [18], which has given better results according to this comparative study of the detectors of points of interest [19].

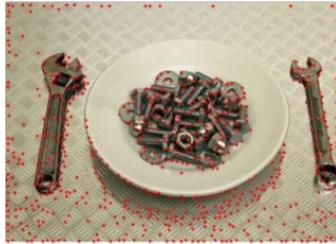


Fig. 2 The points of interest of the left image



Fig. 3 The points of interest of the right image

### 3.1.2 MATCHING

This phase consists in finding, on two images of the same scene taken from different positions, the two points corresponding to the projection of the same element of the scene. In recent years, many matching methods have been proposed [20, 21, 22, 23].

In this work we will compute the pairings of each image (left / right) using a mapping method based on the correlation measure. In this case, it is assumed that the points neighboring two homologous points have gray levels that are similar. This resemblance can be quantified by a correlation measure. Thus, only the neighborhood of a point is used to find its correspondent.

The correlation between stereoscopic images consists, from a window placed in one of the images, in calculating the degree of correlation with another window moving along the corresponding epipolar line in the other image. A sequence of correlation values along the epipolar line is thus obtained. The point corresponding to the best score will be chosen as being to match the point of the center of the fixed window in the other image.

Several families of correlation measures can be distinguished in the literature [21]. In our work, the correlation measure between the points of the images is carried out by the centered and normalized correlation function ZNC (Zero mean Normalized Cross Correlation), it is a centered version of the NCC family, which obtained the best percentages of Matching according to the evaluation carried out in [21].

The function ZNCC for each point (u, v) is written:

$$ZNCC(u, v) = \frac{\sum_{x,y} (I1(x,y) - \overline{I1(u,v)}) (I2(x-u, y-v) - \overline{I2})}{\sqrt{\sum_{x,y} (I1(x,y) - \overline{I1(u,v)})^2 \sum_{x,y} (I2(x-u, y-v) - \overline{I2})^2}} \tag{1}$$

With

$$\overline{I1(u,v)} = \frac{1}{M \times N} \sum_{x=u}^{u+M-1} \sum_{y=v}^{v+N-1} I1(x, y) \tag{2}$$

And

$$\overline{I2} = \frac{1}{N_x \times N_y} \sum_{x=0}^{N_x-1} \sum_{y=0}^{N_y-1} I2(x, y)$$

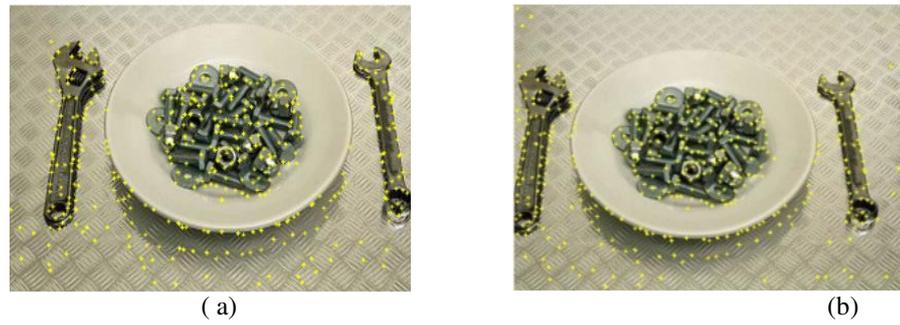


Fig. 4: Corresponding points between the couple of images (a) and (b)

### 3.2 DISPARITY MAP

After the matching step, it is possible to calculate the disparity map. For each point of the left image (right resp). The position difference is calculated with the corresponding point of the right image (left resp) on the same line. These differences are then transformed for each point into a gray level image.

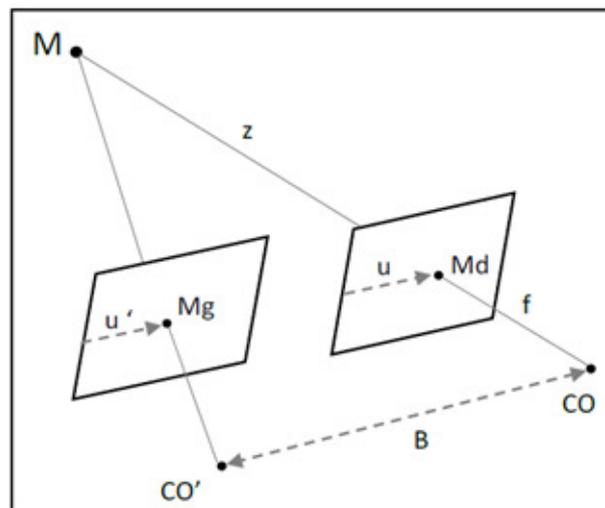


Fig. 5 Pinhole model for the projection of the point M

In this figure the measured disparity, defined as:

$$d = u - u'$$

With  $u$  and  $u'$  the horizontal position of the two image points  $Mg$  and  $Md$  corresponding to the projection of the point 3D in the two images.

The disparity map is closely related with the depth map. We can create a depth map from a disparity map if we know the camera focal length and the distance between the cameras.

The distance  $Z$  from the 3D point to the camera is inversely proportional to the disparity measured on the image as in Fig. 5:

$$Z = (fB)/d \quad (4)$$

With:

B: distance between the cameras

f: focal length

d:disparity.

To find these unknown, once the parameters of the camera are estimated then it is easy to deduce the value Z of depth.

#### 4. EXPERIMENTATIONS

To estimate the depth map, an object-oriented programming language (Java) implemented the steps of estimating this map

In the first interface as in Fig. 6, our process allows loading the left and right image and then displaying the points of interest for each image, thereafter displaying the matching points. In addition, it can generate a file of the matching points. Then the process provide the possibility of calculate disparity map and depth map as in Fig. 7.

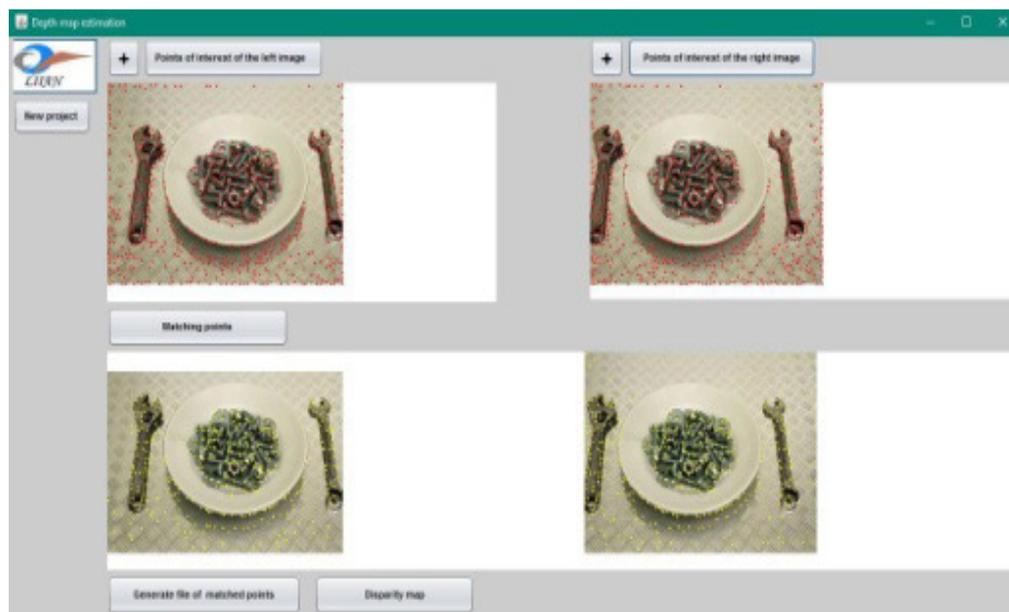


Fig. 6 The first interface of a program for estimating depth map

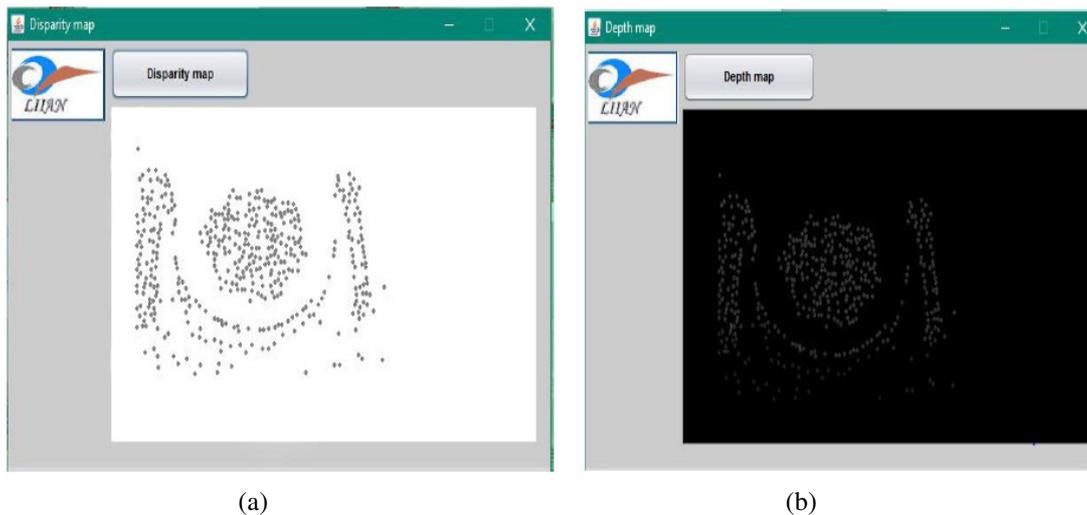


Fig. 7 Disparity map (a) and Depth map (b)

Our program is being improved, in order to generate a more precise depth map to use it later in the 3D reconstruction.

## 5. INTERPRETATION

The results obtained in Figures 6 and 7, shows the satisfaction of our depth map estimation method. In addition this result would be used later to improve the 3D reconstruction

## 6. CONCLUSION

In this article, we have presented the different steps to estimate a dense depth map for the reason that the generation of this map is a very important step in improving the quality of 3D reconstruction. Our goal later is to get other information more than depth from this map before proceeding to the reconstruction step.

## REFERENCES

- [1] El Hazzat, S., Saaidi, A., & Satori, K. Euclidean 3d reconstruction of unknown objects from multiple images. *Journal of Emerging Technologies in Web Intelligence*, 6(1), 59–63, 2014.
- [2] M.A.Ameller, A.Bartoli et L.Quan. Reconstruction métrique minimale à partir de trois caméras affines In 13ème Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle, Volume 2, p : 471-477, January 2002.
- [3] M.HanetT.Kanade. Multiple motion scene reconstruction from uncalibrated views.In Proc of the 8th International Conference on computer Vision, Vancouver, Canada, July 2001.
- [4] A. Fernando, S.T. Worrall, et al., 3DTV: Processing and Transmission of 3D Video Signals, John Wiley & Sons, 2013.
- [5] R. Ji, Y.Gao, R.Hong, Q.Liu, D.Tao, X.Li, Spectral-spatial constraint hyperspectral image classification, *IEEE Trans.Geosci. Remote Sens.* 52(3) (2013) 1811–1824.
- [6] Y.Liu, S.Ci, H.Tang, Y.Ye, Application adapted mobile 3d video coding and streaming a-survey, *3DR Rev.* 3 (1) (2012) 1–6.
- [7] S. Chaudhuri, V. Koltun, Data-driven suggestions for creativity support in 3d modeling, in: *ACM Transactions on Graphics (TOG)*, vol. 29, ACM, 2010, p. 183.
- [8] Y.Gao, Q.Dai, N.Y.Zhang, 3d model comparison using spatial structure circular descriptor, *Pattern Recognit.* 43 (3) (2010) 1142–1151.

- [9] Jong Chan kim et Oh Hoon Cho. Effects of Depth Map Quantization for Computer-Generated Multiview Images using Depth Image-Based Rendering. *KSII Transactions on Internet and Information Systems* vol. 5, no. 11, November 2011.
- [10] Jong Chan kim et Oh Hoon Cho. A study on 3D Panorama System using depth map Generation techniques. *International Journal of Multimedia and Ubiquitous Engineering*. Vol 11, pp 117-128, 2016.
- [11] Y. L. Li, G. J. Wang, X. G. Lin, G. Cheng, L. He, "Real-Time Human Action Recognition System Using Depth Map Sequences", *Advanced Materials Research*, Vols. 760-762, pp. 1647-1651, 2013.
- [12] Pascal Fua. A parallel stereo algorithm that produces dense depth maps and preserves image features. *Machine Vision and Applications*. 1993
- [13] S. Zhang, C. Wang, S. C. Chan. A New High Resolution Depth Map Estimation System Using Stereo Vision and Kinect Depth Sensing, 2013.
- [14] Smisek, J., Jancosek, M., Pajdla, 3D with kinect. *IEEE Workshop Consum. Depth Cameras Comput. Vision* pp. 1154–1160. 2011
- [15] Aamir Saeed Malik, Tae-Sun Choi. A novel algorithm for estimation of depth map using image focus for 3D shape recovery in the presence of noise. *Pattern Recognition* 41(7) : 2200-2225 · July 2008.
- [16] Na-Eun Yang, Ji Won Lee, and Rae-Hong Park. Depth map generation using local depth hypothesis for 2D-to-3D conversion. *International Journal of Computer Graphics & Animation (IJCGA)* Vol.3, No.1, January 2013
- [17] A. Lux. *Algorithme et contrôle en vision par ordinateur*. These de doctoral, INPG Grenoble, 1985.
- [18] C.Harris et M.Stephens. A combined Corner et Edge Detector. 4th Alvey vision Conference. pp. 147-151, 1988.
- [19] N.Elakkad, A.Baataoui, A.El abderrahmani, A.Saaidi et K.Satori. « Etude Comparative des détecteurs des points d'intérêt »WCCCS 11, 2011.
- [20] Daniel Scharstein et Richard Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. Comput. Vision*, 47 :7–42, April 2002.
- [21] S.Chambon and A. Crouzil. Similarity measures for image matching despite occlusions in stereo vision. *Pattern Recognition*. Vol. 44, No. 9, pp. 2063-2075, 2011.
- [22] S. T. Barnard. Stochastic Stereo Matching over Scale. *International Journal of Computer Vision*. Vol. 3, No. 1, pp. 17-32, 1989.
- [23] J. Shao. Generation of Temporally Consistent Multiple Virtual Camera Views from Stereoscopic Image Sequences. *International Journal of Computer Vision*. Vol. 47, No. 2, pp. 171-180, 2002.