A Fuzzy-based Approach for Approximating Depth Information in RGB-D Images

Mau Uyen Nguyen
Department of Information Systems
Faculty of Information Technology
Le Quy Don Technical University,
Hanoi, Vietnam
Email: nguyenitt2005@gmail.com

Thanh Tinh Dao
Department of Information Systems
Faculty of Information Technology
Le Quy Don Technical University,
Hanoi, Vietnam
Email: tinhdt@mta.edu.vn

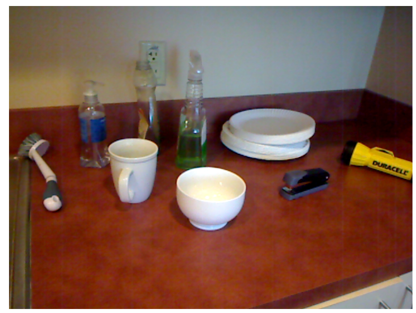
Long Thanh Ngo
Department of Information Systems
Faculty of Information Technology
Le Quy Don Technical University,
Hanoi, Vietnam
Email: ngotlong@gmail.com

***Abstract*—Robot navigation has several security and defence applications. The major technical challenges include measuring the distance between a robot and its surrounding obstacles and modelling the sensing environment. Existing methods using stereo cameras, laser sensors, and low-cost MS Kinect cameras have been suggested for the problems. In this paper, we propose a fuzzy-based approach for approximating the missing depth values in RGB-D images collected from a MS Kinect camera. By investigating different noise models, the missing information, and the relations between the depth and colour images, the proposed approach produces an accurate approximation for missing depth values, which enhances results of subsequent steps in RGB-D image processing.**

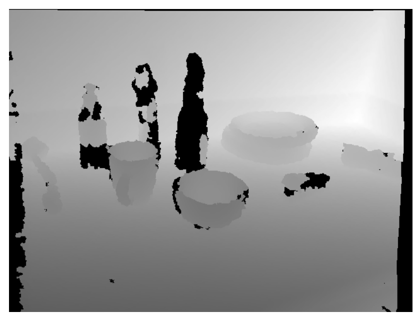
I. Introduction

Modelling three dimensional (3-D) sensing environment for robot localization and navigation is a technical challenge in computer vision. This problem consists of several tasks: capturing images, calculating the depth information between objects to the receiver, modelling environment. In reality, depth information may be determined by stereo cameras, time of flight (ToF) cameras. However, stereo cameras consume a lot of time for extracting features and mapping them to estimating distances, so it faces difficulty in the complex environment. Special ToF cameras, for instance, laser cameras, etc. are more accurate and less time for picking up depth values but the price is expensive. MS Kinect cameras are considered to solve this issue, they are low priced and provide acceptable depth information for robot navigation problems [8], [9]. Thus, MS Kinect cameras have been investigated for a lot of systems, recently. A MS Kinect system contains a colour camera (RGB-Camera) and a depth camera (an infrared projector and a receiver), and it proves a colour image (RGB-image) and a depth image (depth map - D), concurrently. Consequently, they are named as RGB-D images. In depth images, the values have ranged from 0 to 4000 according to the distance from 0 to 4000 millimetres from obstacles to the MS Kinect camera. In fact that, there are some pixels in depth images those values are not defined. The images in 1a and 1b show that a lot of pixels were black, it does mean that, the information of them are missing. Thus, approximating missing values is mentioned in the papers over the world. In [10], [11], [12] show that, various methods are proposed to denoise for color images. In relative to depth images, Frank Lenzen et al. [1] using the adaptive total variation method and edge detection to denoise missing values. Schoner, H et al. [3] cluster depth values, then nominate replacing values. Yang Lia et al. [4] propose A weighted least squares algorithm for denoising. Colour images are used as advice information for denoise progress, Kevin Lai et al. [2] mention the combining RGB-D information approach. The same idea of using RGB information for recovery missing depth pixels, Benjamin Huhle use the non - local mean filter denoising method in [5], colour images are used to identify the similarity.

Considering type of missing values, it can be divided into 3 cases: First of all, the salt and pepper noises that are appeared in infrared equipment features; Secondly, the angle of incidence is large, it makes infrared ray cannot be reflected to the receiver; Finally, the material type, that absorbs infrared ray and there are no rays can be picked up by the receiver. Based on that investigation, we proposed logics for each type of absent information. However, we cannot point out the type of missing values, so that fuzzy logic are involved to model



(a) The colour image



(b) The gray scaled depth image

Fig. 1: The colour image and depth image

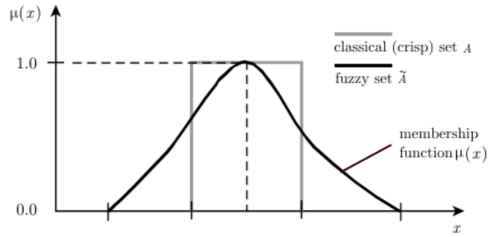


Fig. 2: Fuzzy set and crisp set

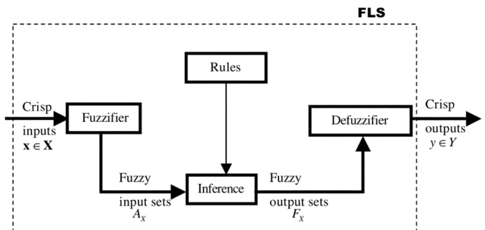


Fig. 3: Fuzzy logic system

the uncertainty.

II. Preliminaries

A. Fuzzy Sets

Fuzzy set was proposed by Zadeh, in [7], to cope with the uncertainty in data. A fuzzy set A in space X can be defined as:

$$\tilde{A} = \{(x, \mu_A(s)) | x \in X\} \tag{1}$$

Membership function $\mu_A(x)$ shows grade of membership x in X to subset A, the values of $\mu_A(x)$ in the interval $[0, 1]$. Thus, the nearer the value of $\mu_A(x)$ to unity, the higher the grade of membership of x in A. In case, $\mu_A(x)$ is only in $\{0, 1\}$, it becomes a crisp set. In [7], Zadeh mentioned all operators on fuzzy sets, for example, union, intersection, ...

B. Fuzy logic

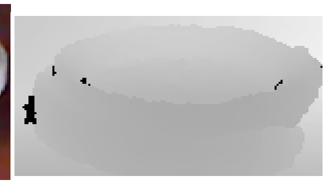
Fuzzy logic systems are inference systems, that process linguistic values based on the expert knowledge. Figure 3, shows elements in a fuzzy logic system. Crisp values are fuzzified to fuzzy sets. They are input for inference phrase, there expert knowledges are presented by rules. Then, the fuzzy sets that are output from rules are defuzzified in defuzzier, and a crisp value is produced.

III. Approximating missing values by using fuzzy logic

A. Salt and pepper noise

As the previous mention, this type of noise causes some group of small missing values. Most of the black values in those groups are on a surface of an object, so that its values can be approximated by their neighbour pixels. In this case, the median filter is the most suitable approach. However, in the area that the number of none black pixels is smaller than the number of black pixels, median filer will nominate the black value. It does mean that, some information will be lost in the median filter. To avoid the bad case, the black value that is returned from the median filter mask will be skipped.

(a) Plates in colour



(b) Plates in depth

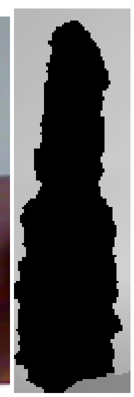
(c) The bowl in colour

(d) The bowl in depth

(e) The stapler in colour

(f) The stapler in depth

(g) The bottle in colour

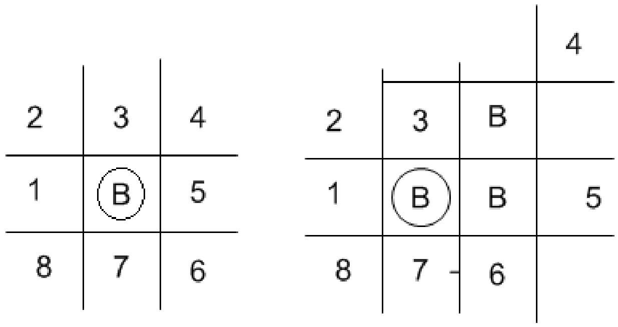


(h) The bottle in depth

Fig. 4: Types of noise

B. The angle which the incident ray makes with the normal is large

Because the angle which the incident ray makes with the normal is large, so that the receiver cannot pick up the reflect ray, and black values will be set to those pixels. In this scenario, black pixels are on the curve surface of boundary of objects, and the number of black pixels in a group is large. The missing values tend to the same group with the shorter distance groups. Assume that, the colour in one object is not changed dramatically, the values of missing pixels are the same group with area have the same colour with them. The 4a, 4b, 4c and 4d in pair show this type of noise.



(a) 8 neighbourhoods (b) 8 extended neighbourhoods

Fig. 5: 8 extended neighbourhoods

C. The absorbed material

Infrared rays are the same features with light rays, some material does not reflect the incident rays, so the receiver can not calculate the depth information. In the environment, there are objects which contain that type of material, some group of pixels will be set in the depth image. Assume that, pieces of absorbed material are parts of objects, so that we can approximate the missing values by neighbourhood values. The 4e, 4f, 4g and 4h in pair show this type of noise.

D. Solution

Considering neighbourhoods of a pixel, for example, at point (x, y) the neighbourhoods are 8 pixels, $(x+i, y+j)$ where $i, j \in \{-1, 1\}$. In a worst case, some points of $(x+i, y+j)$ are the black too, so that we must extend the neighbourhoods to 8 directions. In the figure 6, (B) are black pixels, and (i), $i \in \{1, ..., 8\}$ are non black pixels. N_8 is group of 8 extended neighbourhoods of pixel (x, y) defined above. It is easy to conclude that, the missing depths are relative with some their neighbour depths. However, It is difficult to identify neighbours that the missing pixels are relative to and how much the relative is. Based on the previous investigating in this section, these pixels are identified based on the type of missing information. In the III-A, the related neighbours are all of N_8 pixels. However, the investigating in III-B shows that, they are only some of N_8 pixels. Assuming that, the edges of objects are the curves that are approximate to the straight. Thus, number of the pixels, that are related to missing the pixel, is 4. The III-C shows that, int this case, there is not enough information to infer. Besides, It is not easy to identify grade of relation of missing value with its neighbours. To capture the uncertainty in information and use the previous investigating, we decided to use fuzzy logic for solving this issue.
Assuming that, A_i, $i \in \{1, ..., 8\}$ are fuzzy set, that present for 8 neighbourhoods of a missing pixel. The influence of a neighbour i on pixel (x, y) - $B_{(x,y)}$ is considered as the membership function $\mu_i(x, y)$. Based on grade function $\mu_i(x, y)$ and depth values at A_i we can approximate the depth value at (x, y). There are some some factor that influences on $\mu_i(x, y)$:

- The distance of coordinates from (x, y) to (x_{A_i}, y_{A_i})
- The similarity of colour at pixel (x, y) and pixel (x_{A_i}, y_{A_i}) in image colour
- The trend of increase in depth value from A_i to the missing pixel
- The similarity of colour at pixel (x, y) in image colour to the black

The logic to compute the value $\mu_i(x, y)$ may be changed depending on the values of these factors, so we defined rules for that logic. The parameters are defined as:

- $\mu_{d_i}(x, y)$: membership function presents the distance of coordinates from (x, y) to (x_{A_i}, y_{A_i})
- $\mu_{c_i}(x, y)$: membership function presents the similarity of colour at pixel (x, y) and pixel (x_{A_i}, y_{A_i}) in image colour
- $\mu_{de_i}(x, y)$: membership function presents the trend of increase in depth value
- $\mu_m(x, y)$: membership function presents the similarity of colour at pixel (x, y) in image colour to the black

1) $\mu_{d_i}(x, y)$: Because that, when pixel (x, y) is far from (x_{A_i}, y_{A_i}), the influence of A_i on pixel (x, y) is reduced. Consequently, $\mu_{d_i}(x, y)$ is defined as:

$$\mu_{d_i}(x, y) = \frac{\frac{1}{d(A_i, B_{(x,y)})}}{\sum_{k=1,8} \frac{1}{d(A_k, B_{(x,y)})}} \tag{2}$$

where $d(A_i, B_{(x,y)})$ is distance from pixel (x, y) to pixel (x_{A_i}, y_{A_i}), in the most simple case it is the Euclid distance.

2) $\mu_{c_i}(x, y)$: In colour images, the RGB values change based on the light source and some other factors, so that direct comparison of colour at pixel (x, y) to colour at pixel (x_{A_i}, y_{A_i}) is not a good solution. To remove the effects of noise and show the trend of colours in pixels, we use linear regression line to estimate the nominate value, then compare to value at (x, y) to calculate the colour distance. $d_{Ri}(x, y)$, $d_{Gi}(x, y)$ and $d_{Bi}(x, y)$ are approximated line in direction i for red -R, green - G, and blue - B, respectively. $\Delta_{Ri}(x, y)$, $\Delta_{Gi}(x, y)$ and $\Delta_{Bi}(x, y)$ are distances from element R, G, B from point (x, y) to $d_{Ri}(x, y)$, $d_{Gi}(x, y)$ and $d_{Bi}(x, y)$.

$$\Delta_i(x, y) = \sqrt{\Delta_{Ri}(x, y)^2 + \Delta_{Gi}(x, y)^2 + \Delta_{Bi}(x, y)^2} \tag{3}$$

The $\mu_{c_i}(x, y)$ is defined as:

$$\mu_{c_i}(x, y) = 1 - \frac{\Delta_i(x, y)}{\sum_{k=1,8} \Delta_k(x, y)} \tag{4}$$

3) $\mu_{de_i}(x, y)$: $\alpha_i(x, y)$ is the tilt of direction from A_i to point (x, y), and $\alpha^*(x, y)$ is the $\| \min_{j=1,8} \alpha_j \|$ The $\mu_{de_i}(x, y)$ is defined as:

$$\mu_{de_i}(x, y) = \frac{\alpha_i(x, y) + \alpha^*(x, y)}{\sum\limits_{j=1,8} (\alpha_j(x, y) + \alpha^*(x, y))} \tag{5}$$

4) $\mu_m(x,y)$: Assuming that, R, G, B are values of elements Red, Green, Blue in pixel (x,y) in the color image. The $\mu_m(x,y)$ is defined as:

$$\mu_m(x,y) = 1 - \frac{\sqrt{(R^2+G^2+B^2)}}{\sqrt{3}\times 255} \qquad (6)$$

E. Verdict

In the case mentioned in the III-C, it does mean that, the $\mu_m(x,t)$ reaches 1 and the missing value is calculated based on N_8, the 4f and 4h are examples. In contrast, the angle which the incident ray makes with the normal is large, the missing depth information is influenced from 4 neighbours. In figure 4b and 4d, most black pixels belong in the plates or the bowl, so that the values of missing pixels are affected by neighbours in the plates or the bowl. They are independent of the desk surface.
To nominate the missing value based on the neighbour A_i, we build the approximate line from A_i and predict the value at $B(x,y)$, notation as $P_{A_i}(x,y)$.

1) Algorithm: approximating missing depth pixels based on RGB-D images
where f_d, f_c, f_{df} are constants to express the factor of each

---

Algorithm 1 Approximating missing depth pixels based on RGB-D images

---

Input: I_d, I_c depth and colour images
Output:I_d which are approximated missing depth information

Step 1: Run the median filter
Step 2: Foreach (x,y) in $I_d(x,y) == 0$
- Calculate $\mu_{d_i}(x,y)$ by formula 2
- Calculate $\mu_{c_i}(x,y)$ by formula 4
- Calculate $\mu_{de_i}(x,y)$ by formula 5
- Calculate $\mu_m(x,y)$ by formula 6

Step 3: Foreach (x,y) in $I_d(x,y) == 0$
- if $I_c(x,y)$ is black then $I_d(x,y) = \sum_{i=1,8} \mu_{d_i}(x,y) \times P_{A_i}(x,y)$
- else
  - Calculate $\mu_i(x,y) = f_d \times \mu_{d_i}(x,y) + f_c \times \mu_{c_i}(x,y) + f_{de} \times \mu_{de_i}(x,y)$
  - Identify i^*, $\sum_{k=0,3} \mu_{(i^*+k)\%8+1}$
  - Update $I_d(x,y) = \sum_{k=i*}^{(i*+3)\%8+1} \mu_k(x,y) \times P_{A_k}(x,y) \times \frac{1}{\sum_{h=i*}^{(i*+3)\%8+1} \mu_h(x,y)}$

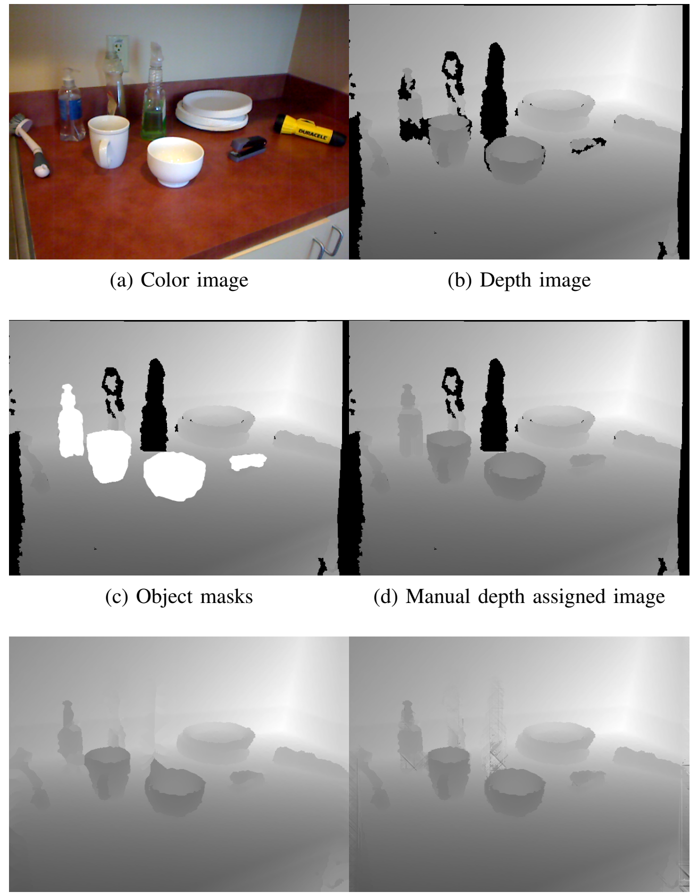
Step 4: return I_d .

---

element.
The algorithm 1 only loop on image pixels, so it stops in N steps (N number of pixels - 480 $\times$ 640). In **step 1**, **step 2** and **step 3**, each sub step does not loop over 8 times for 8 directions. Thus the complexity of the algorithm is $O(N)$.

IV. Experimental results

It is difficult to conduct the efficient experiment to show the statistics on the approximated values, because we do not have



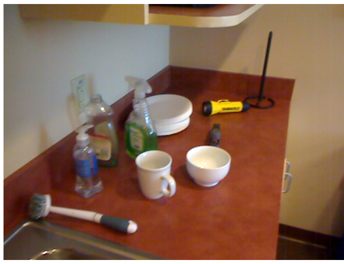
(a) Color image (b) Depth image
(c) Object masks (d) Manual depth assigned image
(e) Result based on [2] (f) Result of our method

Fig. 6: Experimental images

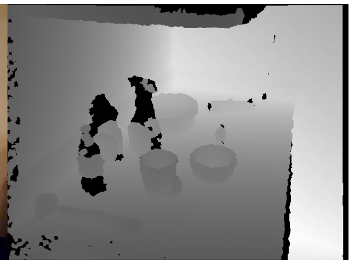
Statistics	Value
Number of calculated pixels	3987
Average	13.8
Standard deviation	69.3
Min	-345
Max	325
Less than 69	3169

TABLE I: Statistics on deviation of approximated values to manually assigned values

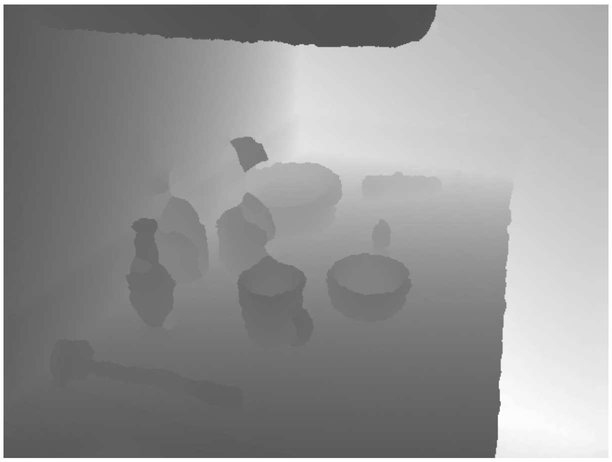
the target values for missing pixels. Thus, we build the mask for objects in an image and try to assign missing pixels with the most suitable neighbour. The table I shows the deviation of approximated values to manually assigned values. The average is 13.8, it does mean that, approximated values are balanced. Min and Max values are large, but it can be understood, because the some pixels are assigned to mistake surfaces. Standard deviation is 69.3, it is acceptable for indoor robot navigation. Number of pixels, that the differences are less than 69, is 3169, it contains about 80% percentage of considered points. It can be thought that, 80% pixels are assigned to the correct surfaces.
Beside that, the algorithm 1 was tested on RGB-D database mentioned in the paper [2], and it shows some good point in comparison with the method proposed in the paper. Intuitively,



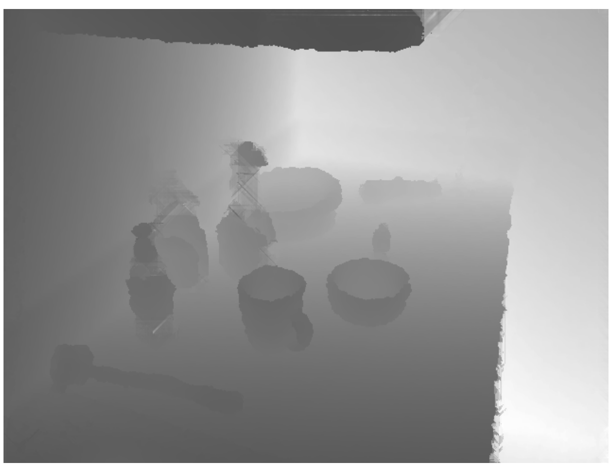
(a) Color image



(b) Depth image



(c) Method in [2]



(d) Our proposed method

Fig. 7: Experimental results on the RGB-D database mentioned in paper [2]

the figure 7 shows that, our proposed method is better than the method mentioned in [2] in case the black points near the boundary of objects, for example, the bottle, cup, etc. To consider on the boundary, our method is worse, and the approximated values in a missing area can be changed dramatically.

V. Conclusion

In this paper, the authors investigated the features of missing values in depth images, that are captured by MS Kinect cameras, and proposed a method to approximate missing depth values based on the fuzzy logic approach. The proposed algorithm 1 shows that, the result is acceptable for the next phases in processing RGB-D images. Although, the complexity of algorithm 1 is $O(N)$, it must be improved to meet real time processing requirements.

Acknowledgment

This work was funded by the Research Fund RFIT@LQDTU, Faculty of Information Technology, Le Quy Don Technical University.

References

[1] Frank Lenzen, Henrik Schfer, Christoph Garbe, *Denoising Time-Of-Flight Data with Adaptive Total Variation*, 7th International Symposium, ISVC 2011, Las Vegas, NV, USA, September 26-28, 2011. Proceedings, Part I.

[2] Kevin Lai, Liefeng Bo, Xiaofeng Ren, and Dieter Fox, *Detection-based Object Labeling in 3D Scenes*, Robotics and Automation (ICRA), 2012 IEEE International Conference.

[3] Schoner, H. Moser, B. ; Dorrington, A.A. ; Payne, A.D. ; Cree, M.J. ; Heise, B. ; Bauer, F., *A Clustering Based Denoising Technique for Range Images of Time of Flight Cameras*, Computational Intelligence for Modelling Control & Automation, 2008 International Conference.

[4] Yang Lia, Jianing Lia, Lianghao Wanga, Junfei Zhanga, Dongxiao Lia, Ming Zhanga, *A weighted least squares algorithm for time-of-flight depth image denoising*, Optik - International Journal for Light and Electron Optics, 2014

[5] Benjamin Huhle, Timo Schairer, Philipp Jenke, Wolfgang Straer, *Robust Non-Local Denoising of Colored Depth Data*, Computer Vision and Pattern Recognition Workshops, 2008. CVPRW '08. IEEE Computer Society Conference.

[6] Mendel, J.M., *Fuzzy logic systems for engineering: a tutorial*, Proceedings of the IEEE Volume:83 , Issue: 3, 1995.

[7] L. A. Zadeh, *Fuzzy sets*, Information and control 8, p 338-353, 1965.

[8] Khoshelham, K. (2011), *Accuracy analysis of kinect depth data*, In: ISPRS workshop laser scanning 2011, 29-31,2011.

[9] Kourosh Khoshelham, Sander Oude Elberink, *Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications*, Sensors, 2012.

[10] A. Buades , B. Coll , J. M. Morel, *A review of image denoising algorithms, with a new one*, Simul, v.4 , 2005.

[11] Thanh, Nguyen Minh; Mu-Song Chen, *Image Denoising Using Adaptive Neuro-Fuzzy System* , International MultiConference of Engineers & Computer Scientists;2006, p74.

[12] Stefan Schulte, Bruno Huysmans, Aleksandra Piurica, Etienne E. Kerre, Wilfried Philips , *A New Fuzzy-Based Wavelet Shrinkage Image Denoising Technique*, Advanced Concepts for Intelligent Vision Systems;2006.