Edges detection in depth images for a gesture recognition application using a Kinect WSN

Asma Ben Hadj Mohamed, Amal Baghdadi, Thierry Val, Laurent Andrieux, Abdennaceur Kachouri

Edges detection in depth images for a gesture recognition application using a Kinect WSN

Asma Ben Hadj Mohamed[1,2], Amal Baghdadi[1], Thierry Val[2]
, Laurent Andrieux[2], Abdennaceur kachouri[1],
[1] LETI Laboratory, Sfax and Gabes Universities,
Sfax Tunisia
{abennaceur.kachouri, amal.baghdadi}@enis.rnu.tn

[2] CNRS-IRIT Lboratory
Toulouse University, Toulouse France
{thierry.val, laurent.andrieux, asma.ben-hadj-mohamed}@univ-tlse2.fr

Abstract. The detection of persons in an image has been the subject of several studies. Most of these works were done on images taken by cameras in visible light (RGB). In this paper,we are interested in people contours detection on the Kinect3D images. We investigate the application of Gradient approach and optimal filters on depth images. We also use this detection to monitor the person via her gestures. Results show that edge detection of Canny is good for people in both light condition but, the performance of Sobel algorithm was better for the images taken in the dark depths.

Keywords: Kinect;3D images, depth images, people edge detection, gestures recognition

Introduction

Images acquisition and processing are taking significant growth especially in the evolution of technologies and the audio-visual communications tools in the last years.Recognition and identification in digital images constitutes an important part which does not lack importance in this field. However, the processing of these images is not as easy as in 2D. Indeed these are less complex since each pixel has a level of color (RGB or gray level), while a pixel in a depth image has the coordinates in space of an object in the scene being shot.

The detection of persons in an image has been the subject of several studies [1] [2]. Most of these works were done on images taken by cameras RGB [3]. Methods of detection and recognition generally used to imitate the same detection process as humans [4]. They use features based on gradients or determine interest points of the image. The purpose of edge detection is to identify the points of a digital image correspond to abrupt changes in light intensity. These changes in image properties generally reflect significant events or changes in the properties of the world. These include discontinuities in depth, in the orientation of a surface, the material properties and the lighting of a scene.

The new technology of intelligent video sensors has generated depth images with the advantage of capturing in different environments where light conditions are poor or ambiguous [5]. Processing 3D images invades many fields such as, industry, military and security applications, mobile technology, video games etc.

In this paper we are interested in processing 3D images to detect person's edges using some approaches and monitoring the person detected as well as recognizing her gestures.
Actually, this work is part of a medical application of monitoringpeople with disabilities into their own homes [6] [7]. Indeed, it allows them to control their environments (state of the lights, air conditioning, opening and closing doors etc) via gestures recognition. We use Matlab to extract the depth of images and to perform transformations at these digital images.
The data acquisition system which we used is the Kinect sensor. In fact, it is an intelligent video sensor made by PrimeSense company semiconductor fables.This device is originally designed for the Microsoft console 360 which aims to control the video game controller without using joystick[8].It consists of horizontal bar connected to the base via a small motor. This allows the camera to perform small up and down movements in order to adapt the perception of the camera depending on the position in the room. The horizontal bar is the main element of the Kinect technology. It contains a series of multi-microphones,an RGB camera, and finally a 3D depthsensor which is the main interesting element in this subject (Figure1).

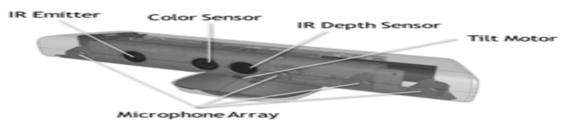

Fig. 1. Kinect sensor elements

Thus, the device can not only detect the color through the video camera, but also the depth of the data. Thanks to the infrared projector which projects a pattern into the entire room. The sensor collects at first the infrared image obtained. Thanks to the IR projector, the Kinect can work on darkness. Then, it analyses the deformation of the patterns using a highly parallel SOC called "PS1080" and finally infers a depth map of the image. This technical is called "light coding" is specification of the Kinect [9]. Therefore, the device transmits to the computer the frames containing the depth map, the color map and an audio streamall synchronized and whose size are VGA and with an average rate of 30 frames per second.
These features make the Kinect available in many applications such as robotic, industry, etc [10].
In this paper, we will talk at a first time about the construction of 3D mesh with this sensor. In a second time, we willinvestigate the application of several filters and approach to edge detection in these images in aim to detect human body edges and finally we will represent how to monitor this person through her gestures.

2 People detection in depth images

2.1 Construction of 3D mesh in Kinect

The infrared camera provides an image representing the thermal releases emitted by the observed object (Figure 2).

Fig. 2. Image obtained by the Kinect infra-red camera

This type of image was widely popularized by legendary movie Predator and infrared vision. However, the resulting images are not colored; it is the users who decide to add color levels based on the measured temperature. The IR or depth camera has a field of view just like any other camera. The field of view of Kinect is limited as illustrated in Figure 3. The view angles are 57 degrees horizontal and 43 vertical and the depth map returned by the Kinect associated with each pixel its distance from the sensor in mm[11]. From this distance, which is equivalent to the z coordinate of the point represented by the real pixel, we can calculate its abscissa and ordinate.

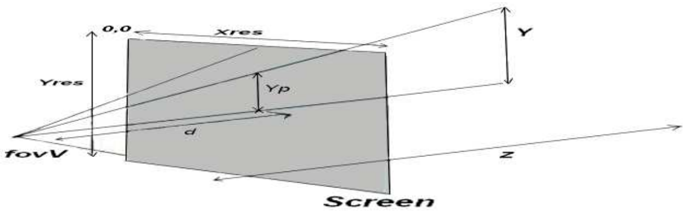

Fig. 3.. Calculation of the coordinate “y” of the point [12]

$$x = 2z\tan\left(\frac{fov_H}{2}\right)\left(\frac{x_p}{x_{res}}\right) \qquad (1)$$

$$y = 2z\tan\left(\frac{fov_V}{2}\right)\left(\frac{y_p}{y_{res}}\right) \qquad (2)$$

X_p *and* Y_p : Pixel coordinates
X_{res} and Y_{res} : The horizontal and vertical resolution of the depth map.
fov_H et fov_V :The horizontal and vertical field of view of Kinect (in radians).
Thus, the coordinates (x, y, z) corresponding to the actual points of the pixel depth map are known. This allows obtaining a 3D point cloud of the scene
seen by the Kinect whose dimensions are the same as the actual dimensions.Since weknow the depth values, we can determine the width and height of a player using trigonometry, as illustrated in figure 4 where we calculate a player’s width. The normal depth vision of the sensor ranges from around two and ahalf feet (800mm) to just over 13 feet (4000mm) [11].
The purpose of edge detection is to identify the points of a digital image that correspond to a sudden change in light intensity. These changes in image properties generally reflect significant events or changes in the properties of the world. These include discontinuities in depth, in the orientation of a surface, the material properties and the lighting of a scene.
To extract the contour of people in the 3D images we tried different approaches. However, in this second part, we will present only two filters which are the most efficient in our researches [13].

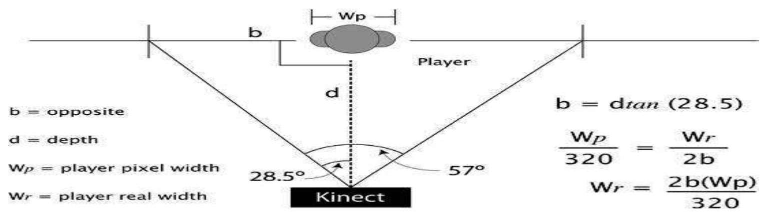

Fig. 4. Finding the player real world width [11]

2.2 Application of Edge detection Approach (The Gradient Approach)

The gradient is a vector, the more traditional approach to estimate the gradient is to choose two orthogonal preferred directions, on which we project the gradient [14]. It can be described in the following flow chart.

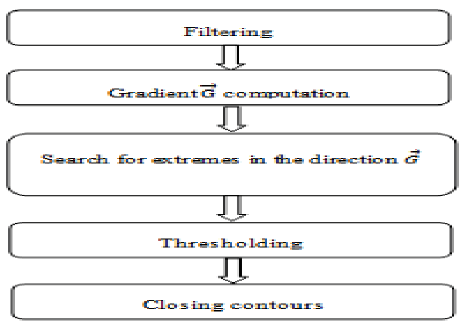

Fig. 5. Flow chart of the Gradient Approach

Applied to an image, regarded as a function of two variables, we can define two partial derivatives along x as a function of y. Let's consider an image I with 2 variables x and y; I = I (x, y)
The gradient is calculated by the following formula:

$$\vec{G} = \left(G_x, G_y\right) = \left(\frac{\partial I_{(x,y)}}{\partial_x}, \frac{\partial I_{(x,y)}}{\partial_y}\right) \quad (3)$$

Gradient filters presented below therefore possess multiple masks (several filters convoluted) according to the directions of gradient calculation.

a) Robert filter.

Operator Roberts described in a rectangular image (see Figure 5), with a contour ramp, the first derivatives, they are of limited value because the maximum slope is unlikely to be on one of the two directions considered. What matters is the length of the gradient vector of which they are components. This length is calculated in principle by the Pythagorean Theorem by computing on real and accelerating considerably using a full approximation:

$|G| = |E - W| + |N - S|$ (4)

Where: E, W, N and S are the pixels in the rectangular image

Regardless of the accuracy of the calculation, this result must be transformed into a digital filter. The physical notion of filter corresponds to the mathematical concept of convolution. In the case of digital data as in the case of image processing, the relationship between output pixel values and the input t pixels value is described by an array of numbers called convolution matrix [15]. Calculated on 4 points, this filter consists of the application of the double mask as follows:

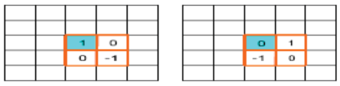

Fig. 6. Mask to apply in Robert filter

b) Sobel filter .

The Sobel filter computes an approximation rather inaccurate of intensity gradient, but enough practice in many cases. In fact, it uses a neighborhood (usually 3 × 3) around each point to calculate the gradient, and the weights used to calculate the gradient are integers [15]. Also calculated on 9 points, this filter allows focus on the calculation following several directions (horizontal, vertical, diagonal). Filters are the following:

Fig. 7. Sobel filter.

Fig. 8.. Example of edge detection using the Sobel filter

2.3 Laplacian approach

A priori, the use of second derivatives is more convenient than first derivatives: instead of seeking the maximum intensity gradient, we seek cancellation of second derivatives [14]. In the case of an image, there is no second derivative but only four partial derivatives (in x^2, y^2, xy and yx). In practice, it removes this ambiguity by using the Laplacian operator which is the sum of the two partial derivatives. Being an image I with 2 variables x and y

$$I = I(x,y)$$

The Laplacian is calculated by the following formula:

$$\nabla I^2(x,y) = (\frac{\partial^2 I_{(x,y)}}{\partial_x}, \frac{\partial^2 I_{(x,y)}}{\partial_y}) \quad (5)$$

2.4 Optimal filters

Actually, the main disadvantage of the Laplacian approach is that the derivative operators of first order exaggerate the noise effects and second derivatives are exaggerating it twice. Another type of most recent approach is based on the definition of optimality criteria of edge detection; these criteria leading to optimal smoothing filters [16].

a) Canny filter.

If we talk about the one-dimensional case, it is assumed that the detection is performed by convolving the signal by a filter with impulse response "h", the contours being characterized by the extreme of the filter output [16]. The edges are considered here types on and the noise is assumed to be white (zero mean). Optimality criteria are:

- Detection: the edge must be detected, it must minimize false answers.
- Localization: the edge must be located with accuracy; it comes to minimize the distance between the detected and the true contour.
- Single response: to minimize the number of responses for a single edge.

The minimization of these criteria in the context of the proposed model leads to a differential equation whose solution is of the form:

$$h(x) = a_1 e^{\alpha x}\cos(\omega x) + a_2 e^{\alpha x}\sin(\omega x) + a_3 e^{-\alpha x}\cos(\omega x) + a_4 e^{-\alpha x}\sin(\omega x) \quad (6)$$

Figure 9 shows an example of edge detection by Canny.

Fig. 9..Edge detection using Canny filter on grayscale image

2.5 Edge detection of depth images in light and darkness

After presenting some approach in edge detecting on color images, we will apply iton depth images. An example of 3D images produced by the Kinect sensor is shown is figure 10 and 11. We should note that color in this figure represents a distance and the user can program the Kinect to display his color choice depending on the distance. We made testes on these images in an indoor environment once when exposed to light and once in darkness. Results are illustrated in figures 10.a, 10.b, 11.a and 11.b.

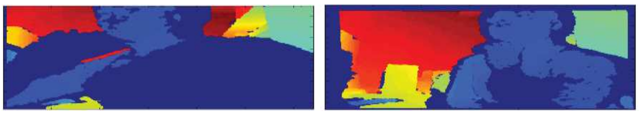

Fig. 10. Depth images of the Kinect sensor

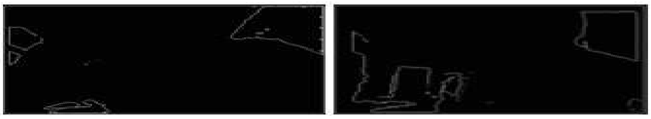

Fig. 10.a.Sobel filter applied to the depth images in light

Fig. 10.b.Canny filter applied to depth images in light

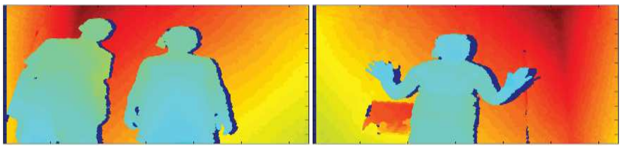

Fig. 11. Depth images of the Kinect sensor

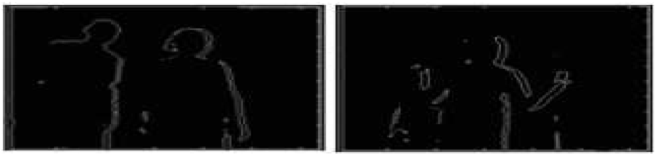

Fig. 11.aSobel filter applied to the depth images in darkness

Fig. 11.bCanny filter applied to depth images in darkness

By comparing images taken in deep darkness and those taken in good lighting conditions, we note that the quality is the same. As edge detection of Canny is good for people in both conditions. But the performance of Sobel algorithm was better for the images taken in the dark depths. Results are summarized in table I below.

Table 1.I. Table Type Styles

	Edge'sthickness	Localization	Sensibility to noise	Detectability
Sobel	Augmentation	Average	High	Low
Canny	Preserved	High	Low	High

3 Gesture detection

The tests we made on the depth images aim to detect and locate people then the monitor. We applied two approaches (Canny and Sobel filters) in the presence of light and darkness to check the efficiency of depth sensor and see the difference between the two approaches used. In this part, we made a gesture recognition using hands position. The Kinect sensor can detect 20 joints in human body.
When we calculate the coordinates (x,y,z) of each joint in the space we could make an idea about the position of each body part referred to the head or spin joint. In our application, we put sensors on the smart home building located in the University of Blagnac Toulouse such presented in figure 12.

Fig. 12. The smart house where testing our application

Therefore, we programmed the sensor to recognize hands user positions depending on the head and spin joints. Indeed, we can identify a focus gesture based on Z coordinates in the space. The sensor is being in an idle mode if the person already detected is not moving in the scene. However, when there is an action, the Kinect focus on hands position to see if there a focus gesture that exceeds a threshold distance. Thus, it triggers a recognition part. When a gesture is recognized, the monitored person can handle devices through a domestic bus where there are a large number of KNX actuators [17] such as lights, blinds, heating, air conditioning system, opening doors and windows, alarm, luminous path at night etc. Some recognized gestures are represented in figure 14.
In a previous work, we already investigate the use of a wireless network with Mesh topology[6]. Indeed, we test the application of OLSR routing protocol [18] and 802.11s standard. Besides, we demonstrate that there is difficult to transport Kinect data (depth images) through the network because of the high Baud rate required in the network which was about 180Mbits/s.

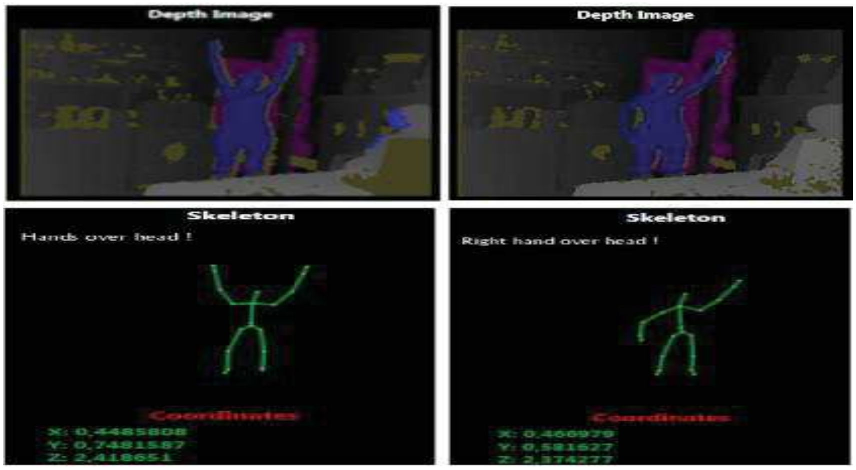

Fig. 13. Gesture recognition using Kinect

Therefore, it is easier in our application to communicate a simple text or alert. So, when we gesture is detected, we don't need to communicate all the frames through the network. The communication of a message within the network seems a more discreet and does not present a sort of spying on the patient in their own home.

4 Conclusion

The detection of persons in an image has been the subject of several studies. Most of these works were done on images taken by cameras in visible light (RGB). In this paper, we are interested in the Kinect depth images.We first presented the video sensors considered as a smart camera as well as the construction of 3D mesh. Secondly, we applied edge detection filters on its 3D images, we extracted people contours and we did a performance comparison between filters.
Results show that edge detection of Canny is good for people in both light condition but, the performance of Sobel algorithm was better for the images taken in the dark depths. Finally, we presented the gesture recognition part based on hand motions. We develop a program which detects people gestures and to recognize it.
An automatic monitoring system seems to be interesting in the medical side. In a further work, we will implement Unistroke recognizer to use with Kinectin aim to find more gesture to command the smart house. Also, as perspectives, it is important to present a new approach for detecting people in images using depth contour models in 2D and 3D surface models.

5 References

1. A. Ben Hadj Mohamed, T. Val, L. Andrieux, and A. Kachouri, "Assisting people with disabilities through Kinect Sensors into a smart house", IEEE International Conference on Computer MedicalApplications ICCMA, Sousse, Tunisia 2013

2. A. Ben Hadj Mohamed, T. Val, L. Andrieux, and A. Kachouri, “Using a Kinect WSN for home monitoring : principle, network and application evaluation,” ICWCUCA’s, IEEE International Conference on Wireless Communications in Unusual and Confined Areas, Clermont-Ferrand, France, August 2012.
3. A.Baghdadi, "Image Processing depth from the Kinect," Master project, National School of Engineering of Sfax Tunisia, 2012.
4. A.N. Evans, Liu, X.U, “A morphological gradient approach to color edge detection”, Dept. of Electron. &Electr. Eng., Univ. of Bath, UK IEEE Transactions onImage Processing,Journals & Magazines, 2006, pp. 1454 - 1463
5. D.Demigny, “On optimal linear filtering for edge detection”, Image & Signal Process. Res. Lab., CergyPontoiseUniv, IEEE Transactions onImage Processing,Journals & Magazines, V. 11 Issue: 7, pp. 728 - 737
6. F. YLi, L. Vandoni, G. Zicca, S. Zanoli « OLSR Mesh Networks for Broadband Access: Enhancements, Implementation and Deployment», 4th IEEE International Conference on In Circuits and Systems for Communications, ICCSC 2008
7. http://www.knx.org/
8. J. Webb, J. Ashley, Beginning Kinect programming with the microsoftKinect SDK, 1St edition, New York :Spinger, 2012, pp 68-69
9. K. PAPIN,« Comparative Analysis for images processing d'algorithmes » Project graduation (ENSEM).
10. Kinect for SDK tutorial: http://msdn.microsoft.com/fr-fr/windows/sdk-kinect-tuto
11. L. Cruz, D.Lucio, L.Velho «Kinect and RGBD Images: Challenges and Applications » ,Graphics, Patterns and Images Tutorials (SIBGRAPI-T), Conference ,OuroPreto,Brazil, August 2012.
12. L. Pishchulin, A. Jain, C. Wojek, M. Andriluka, T. Thormahlen, , B. Schiele, “Learning people detection models from few training samples”, MPI Inf., Saarbrucken, MPI Inf., Saarbrucken, Germany, IEEE Conference onComputer Vision and Pattern Recognition (CVPR), 2011, pp. 1473 - 1480
13. M. Tölgyessy, P. Hubinský, P. Hubinský, “The Kinect Sensor in Robotics Education,” Institute of Control and Industrial Informatics, Faculty of Electrical Engineering and Information Technology, Slovak University of Technology in Bratislava Slovakia, 2010
14. Ni. Bingbing, Gang Wang ; P. Moulin, RGBD-HuDaAct: A color-depth video database for human daily activity recognition”, IEEE International Conference on Computer Vision Workshops (ICCV Workshops), 2011, pp.1147 - 1153
15. R. Munoz-Salinas, E.Aguirre, M.Garcia Silvente“People Detection and Trackingusing Stereo Vision and Color” Technical Superior school of Computer Engineering, Univ of Granada, DECSAI-SI-2005-04
16. S.Green, M.Blumenstein, M. Browne, and R.Tomlinson, “The detection and quantification of persons in cluttered beach scenes using neural network-based classification”, Sch. of Inf. &Commun. Technol., Griffith Univ., Gold Coast, Qld., Australia, Sixth International Conference on Computational Intelligence and Multimedia Applications, Aug. 2005. pp303 – 308
17. S.L.Phung, “Detecting People in Images: An Edge Density Approach”,Sch. of Electr., Comput. &Telecommun. Eng., Wollongong Univ., NSW, IEEE International Conference onAcoustics, Speech and Signal Processing, 2007. ICASSP 2007. I-1229 - I-1232
18. Thibaud Chrétien, Mathieu Houdeau, 2A Project speciality 2A “real time 3D Reconstruction of hands motion”, (MMIS-IRVM)