People Detection and Localization in Real Time during Navigation of Autonomous Robots

Percy W. Lovon-Ramos, Yessica Rosas-Cuevas, Claudia Cervantes-Jilaja
Maria Tejada-Begazo, Raquel E. Patiño-Escarcina and Dennis Barrios-Aranibar
Grupo de Investigación en la Línea de Automatización Industrial, Robótica y Visión Computacional (LARVIC)
Centro de Investigación en Ciencia de la Computación
Universidad Católica San Pablo
Arequipa, Perú
{percylovon, yrosasc19, cj.3.cervantes}@gmail.com {maria.tejada.begazo,rpatino,dbarrios}@ucsp.edu.pe

***Abstract*—Currently the navigation involves the interaction of the robot with its environment, this means that the robot has to find the position of obstacles (natural brands and artificial) with respect to its plane. Its environment is time-variant and computer vision can help it to localization and people detection in real time. This article focuses on the detection and localization of people with respect to plane of the robot during the navigation of autonomous robot; for people detection is used Morphological HOG Face Detection algorithm in real-time; where our goal is to localization people in the plane of the robot, obtaining position information relative to the X-Axis (left, right, obstacle) and with the Y-Axis (near, medium, far) with respect robot; to identify the environment in that it's located in the robot is applied the vanishing point detection. Experiments show that people detection and localization is better in the medium region (201 to 600 cm) obtaining 93.13% of accuracy, this allows the robot has enough time to evade the obstacle during navigation; the navigation getting 97.03% of accuracy for the vanishing point detection.**

***Keywords*—People detection, Vanishing Point, HOG method, Autonomous Vehicle Navigation.**

I. INTRODUCTION

Work in structured and unstructured environments is difficult because the robot is facing hard lighting conditions that are changing over time. These environments have visual characteristics such as: Artificial marks (objects of a particular color for navigation of the robot) and natural marks (wall, floor, objects, people, stair, etc.). The visual perception of the robot can translate these visual characteristics in: Lines, curves, points and edges that will be necessary to navigation in structured environments. Also, we must keep in mind that the occlusion of objects or people during the navigation can be found.

This work proposes people detection and localization (X-Axis (left, right, obstacle) and the Y-Axis (near, medium, far)) in any position (vertical, back, front or upper half body) during navigation of autonomous robot using computer vision, using an Morphological Face HOG Detection algorithm and vanishing points detection to navigation of the robot in structured environment. In this sense, the paper is organized as follows: Section II presents the research about other approaches of people detection and locate the robot during navigation. Morphological face HOG detection algorithm and vanishing points detection are explained in section III, results are shown in section IV and finally the section V, authors present a discussion about conclusions and future works.

II. RELATED WORK

Most effective algorithm for pedestrian detection is the histogram of oriented gradients (HOG) proposed by Yao et al. [1], nevertheless, the presence of partial occlusion in human bodies causes HOG failures. Therefore, Castillo and Chang [2] present the detection of human bodies in a structured environment, through two steps: Detection of silhouette and the presence of skin, thus improving the detection of the person. Dalal and Triggs [3] show experimentally that grids of Histograms of Oriented Gradient (HOG) descriptors significantly outperform existing feature sets for human detection, for their test using MIT pedestrian database. Afterwards, Intel [4] developed an open source library for computer vision related programs, this library is the so called Open Computer Vision Library (OpenCV). The OpenCV library implements the Viola-Jones method [5] with some important modifications (based on Haar-like Features, e.g., profile face and frontal face), for images that show only parts of the person or upper body where sometimes the face is exposed.

Another proposal presented by Tejada-Begazo et al. [6] presents a evaluation of different morphological operators (erode, dilate, open, close, open-close) applied to improve the HOG method, where obtaining a $close$ (86.62%) and $erode$ (84.35%) had better results than HOG without this preprocessing (77.32%). However, authors analyzed that there are a great quantity of images that contain people with more than half of the body exposed and where the person's face is shown; is the reason that Cervantes-Jilaja et al. [7] include face Detect, in case we have not found the human detect, algorithm called"Morphological Face HOG Detection" (Alg. 1). Meanwhile Tejada-Begazo et al. [6] as Cervantes-Jilaja et al. [7], they performed their experiment through the human bodies database "IG02-v1.0-people" is used (size of 152 MB, Marszalek and Schmid [8]).

The vanishing point detection is used in structured and unstructured environments making it more difficult because the appearance changes and some artificial landmarks, however, Le et al. [9] proposes a method for detecting pedestrian lanes that painted without markers in indoor scenes and outdoors, in different lighting conditions; using the vanishing points, as a model of appearance of a region of the lane using different types of surface patterns. On the other hand, Le et al. [10] proposes a method for detecting pedestrian lane in unstructured environments, i.e. detects highway in a probabilistic framework integrating both the appearance of the region and characteristics of the edge of the lane, using vanishing points to identify of the lane boundaries based on color edge detection and the use of pedestrian detection for handling occlusion, they worked with 2,000 images collected from various scenes indoor and outdoor with different unmarked lanes.

Furthermore Tripathi and Swarup [11], proposed that the structured environment is classified into hallway, staircase, and open space by using image edge GIST descriptors and a neural network classifier, where detection of horizontal lines cluster and vanishing point is used for the navigation in staircase and hallway environment respectively, obtaining a effectively major 90%. Whenever Lu and Song [12], used heterogeneous visual features such as: points, line segments, planes, and vanishing points, and their inner geometric constraints managed by a novel multilayer feature graph (MFG), obtain KITTI dataset, this method reduces the translational error by 52.5% under urban sequences where rectilinear structures dominate the scene.

III. People Detection and Localization during Navigation of Autonomous Robot

This research attempts to use work Tejada-Begazo et al. [6] and Cervantes-Jilaja et al. [7] in real-time for people detection at any position (vertical, back, front or upper half body) using an Morphological Face HOG Detection algorithm (Alg. 1) during navigation of autonomous robot. Our goal is to locate the people in the plane of the robot, getting the position information relative to the X-Axis (left, right, obstacle) and the Y-Axis (near, medium, far) with respect the plane of the robot, the figure 3 show the full description of the people localization in the plane of the robot.

Algorithm 1 Morphological Face HOG Detection

procedure MORPHOLOGICALFACEHOGDETECTION(image)
    img = Load(image)
    imgGray = ConvertImgToGrayScale(img)
    element = CreateStructElem(cols, rows, $StructElem$)
    imgErode = erodeOperator(imgGray, $element$, $iter$)
    imgMFHD = HOG_detection(imgErode)
    numberRegion = regionHOGDetect(imgMFHD)
    if ($numberRegion$==0) **then**
        return $imgMFHD = dectectFace(imgErode)$
    else
        return $imgMFHD$
    end if
end procedure

In the Morphological Face HOG Detection algorithm (Alg. 1) 1 is observed the following functions: *CreateStructElem* which creates the structuring element based on columns, rows or shape ($StructElem$(cross, ellipse or rectangle)). The morphological operator function $erodeOperator$ presents a the structural element ($element$) and number of iterations $iter$ (0, 1, 2, 3, 4 and 5). The next step is *HOG_detection*, where *regionHOGDetect* function gives the number of human bodies detected, if there is no detected regions (*numberRegion* == *0*), means that only has the human half body, then proceed with face detection with Cascade Classifier (*detectFace*).

A serie of tests was performed to obtain the best parameters of morphological operator (erosion), these parameters are: number of iterations ($iter$) and structural elements ($StructElem$(cross, ellipse or rectangle)). Table I shows the human bodies detected by using structuring elements (*cross, ellipse and rectangle*). The choice of one of these three structural elements is based on processing time (less time) and performance. In this sense, using: *ellipse* as a structural element with *2* iterations; as shown in figures 1 and 2. Another tests can be seen in articles of Tejada-Begazo et al. [6] and Cervanteys-Jilaja et al. [7].

TABLE I. Number of detected persons and processing time (num_personsDetect - ms) - Parameters of Erode Operator for 0 to 5 iterations

Structuring Element vs. N^o of Iterations of Erosion Operator			
N^o of Iterations	Structural Element		
	CROSS	RECT	ELLIPSE
0	4 - 126.25	4 - 131.57	4 - 116.37
1	3 - 98.64	2 - 99.80	3 - 98.97
2	2 - 98.98	2 - 95.93	2 - 101.95
3	2 - 100.72	1 - 100.16	2 - 96.79
4	1 - 97.11	1 - 98.52	1 - 112.59
5	0 - 98.19	0 - 94.99	0 - 98.90

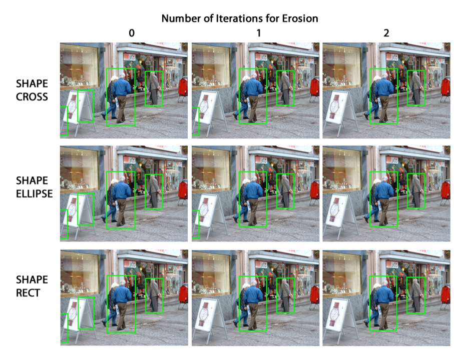

Fig. 1. Image with the structuring elements (*cross, ellipse and rect*) vs. the number of iterations $(0, 1, 2)$, applied to Morphological Face HOG Detection algorithm (Alg. 1), with the Erode operator

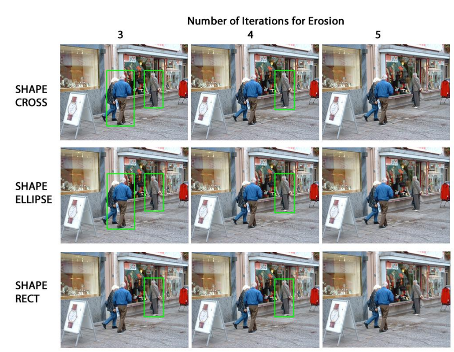

Fig. 2. Image with the structural elements (*cross, ellipse and rect*) vs. the number of iterations $(3, 4, 5)$, applied to Morphological Face HOG Detection algorithm (Alg. 1), with the Erode operator

The autonomous navigation robot is guided exclusively by the vanishing point found in the image. The following diagram (figure 3) shows the calibration camera (measurement image), people detection, detection of vanishing point for the navigation of the robot and finally localization the person in the plane of the robot, getting the position information relative to the X-Axis (left, right, obstacle) and the Y-Axis (near, medium, far) with respect the plane of the robot.

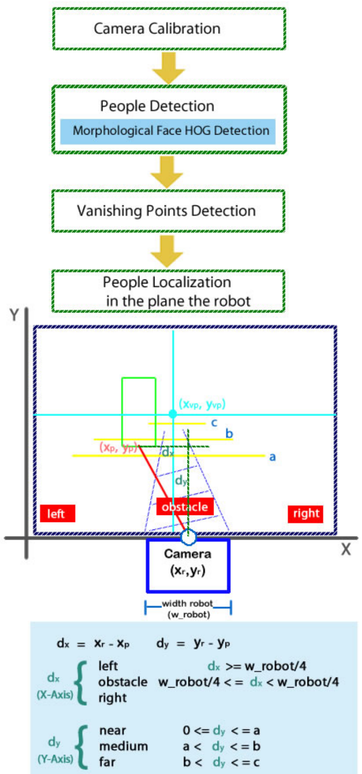

Fig. 3. People detection and localization in the plane of the robot during navigation of autonomous robot

For detection of vanishing point, firstly lines are extracted of the image (horizontal, vertical and oblique), using Hough Transform 'CV_HOUGH_STANDARD', that calculate the equation of the line with respect to the origin.

Once lines are identified, the angle obtained of each line is with respect the $X-axis$ for its subsequent classification, this classification is between a range of angles as shown in figure 4 and the equation (1) and (2), where θ_{L_i} represents the angle of the line i-th array of lines identified and m is a predetermined limit angle ($\pi/18$); return the array of lines considered valid.

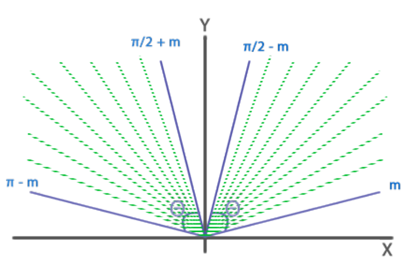

Fig. 4. Classification of lines between the ranges (θ_{L_i})

$$m < \theta_{L_i} < (\frac{\pi}{2} - m) \tag{1}$$

$$(\frac{\pi}{2} + m) < \theta_{L_i} < (\pi - m) \tag{2}$$

Then proceeds to find the crossing points with the lines classified as valid, where will only be valid intersections of lines whose difference between their angles are greater a predetermined limit ($\pi/10$) as shown the equation (3) and figure 5, further the cartesian equation described in equation (4). So, you can remove the multiple lines that appear on one edge.

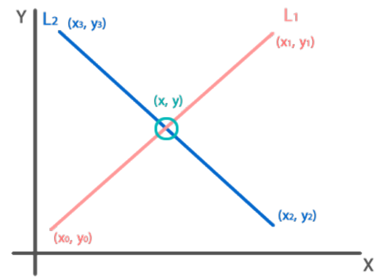

Fig. 5. Crossing point between two straight lines L_1 and L_2

$$|ang_1 - ang_2| > \frac{\pi}{10} \tag{3}$$

$$y = m * x + b \tag{4}$$

To find the crossing point, it is replaced in the equation (4) the straight lines L_1 and L_2, as shown below:

$$y = m_1 x + b_1 \tag{5}$$

$$y = m_2 x + b_2 \tag{6}$$

The crossing point between the two straight lines (L_1 and L_2) is P (x, y), then equation (5) and (6) are equalized :

$$m_1 x + b_1 = m_2 x + b_2$$

$$x = \frac{b_2 - b_1}{m_1 - m_2} \tag{7}$$

The point (x_0, y_0) which belong to the straight line L_1 is replaced in equation (5) and point (x_2, y_2) which belong to straight line L_2 in equation (6), we have.

$$b_1 = y_0 - (\frac{y_1 - y_0}{x_1 - x_0})x_0$$

$$b_2 = y_2 - (\frac{y_3 - y_2}{x_3 - x_2})x_2$$

Values are replaced in equation (7):

$$x = \frac{(y_2 - (\frac{y_3 - y_2}{x_3 - x_2})x_2) - (y_0 - (\frac{y_1 - y_0}{x_1 - x_0})x_0)}{(\frac{y_1 - y_0}{x_1 - x_0}) - \frac{y_3 - y_2}{x_3 - x_2}}$$

$$x = \frac{(x_1 - x_0)(y_2x_3 - x_2y_3) + (x_3 - x_2)(x_0y_1 - y_0x_1)}{(y_1 - y_0)(x_3 - x_2) - (y_3 - y_2)(x_1 - x_0)}$$

In order to obtain the value, $'y'$ is replaced in equation (5) or (6), thereby it find the crossing point $P(x, y)$.

Afterwards has a number of crossing points $P(x, y)$, then it proceeds to iterate each of the points and finding the distance between two points (equation (8)), where the point with the smallest distance is the vanishing point.

$$dist = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (8)$$

IV. Experiments and Results

Tests were performed in the hallways of a building, for this research we used the Komodo Robot (mobile robot with four wheels), which has a camera at 16cm from ground, with captured images of 640 x 480 pixels (px). However, before processing the image, a calibration camera process is necessary to take measurements on the image and get more information about floor, walls and people. For this we used a checkerboard pattern as figure 6 which shows the calibration of the image, to obtain an improvement in the measurement of depth as shown in figure 7.

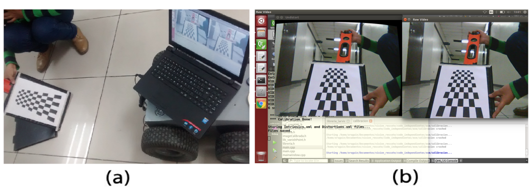

Fig. 6. Where: (a) shows a checkerboard pattern to the camera and (b) shows the calibrated and uncalibrated image

The autonomous robot presents its initial position obtained by odometry, using computer vision obtain the vanishing point and the points of the persons detected, where our goal is to locate people in the plane of the robot, obtaining position information relative to the X-Axis (left, right, obstacle) and the Y-Axis (near, medium, far) with respect robot, as shown in figure 8. These points will be updated on the map and go moving in a determined time, but remaining at a fixed distance relative to the position of the robot, i.e. until the vanishing point no longer detected (time 't'), then the robot will return to last place was mapped last (time 't-1'); for the navigation of the robot Operating System (ROS) was used.

In each test environment is performed people detection used Morphological Face HOG Detection algorithm (Alg. 1), where it detects the person in any position and for robot navigation used Vanishing Point detection for the purpose of determine what distances can be detected people during the robot navigation. Once it detected people, we proceed to the localization of people in the plane of the robot, obtaining position information relative to the X-Axis (left, right, obstacle) and the Y-Axis (near, medium, far) with respect robot.

The first test is when people is located "near region" with respect to the robot, i.e. between the distance of 50 to 200 cm. The figure 9 shows the detection of people who are near

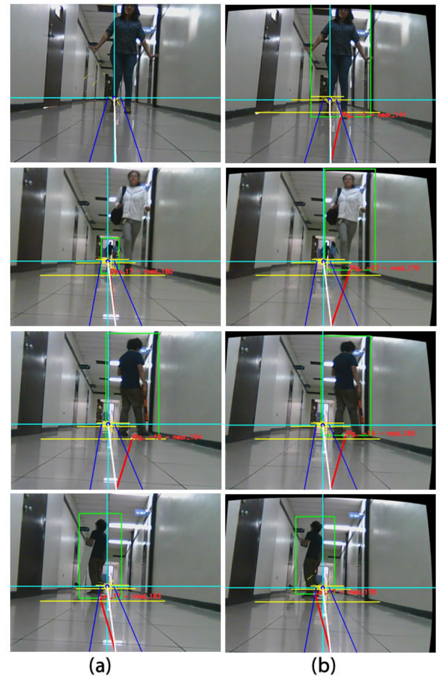

Fig. 7. Where: (a) shows the distance of the depth uncalibrated and (b) shows the distance of the depth calibrated

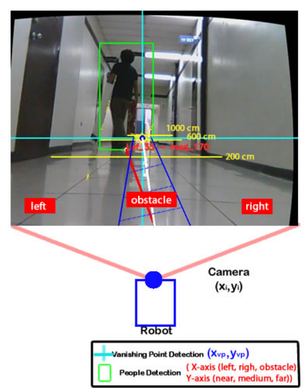

Fig. 8. Where: (x_i,y_i) initial position of the robot, (x_vp,y_vp) position of the vanishing point and position of people detected (X-Axis (left, right, obstacle) and the Y-Axis (near, medium, far))

to 200 cm, where show the entire body of people; but when is near to the robot the camera shows the lower half of the body of people making them unable to detect. Table II shows percentages of success, of people detection and localization in the plane of the robot.

TABLE II. Result People Detection and Localization - First Test ("Near" region)

Distance (50 to 200 cm)	Amount of analyzed frames			Percentage
	Successes	Errors	Total	Accuracy (%)
environment 1	456	42	498	91.57
environment 2	279	24	303	92.08
Average				**91.82**

The second test is when people is located in "medium region" with respect to the robot, during this test could be found all persons (full body); the figure 10 shows people

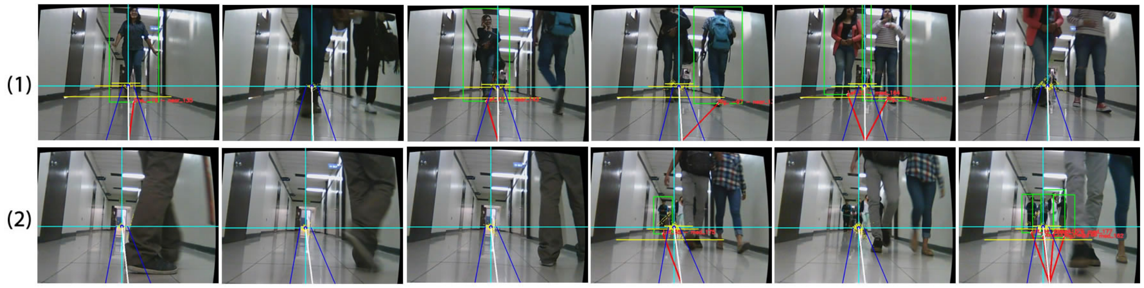

Fig. 9. First test environment (near region), the obstacle or people is situated between 50 and 200 cm of the robot. Where shown people detection and localization in the plane of the robot (X-Axis (left, right, obstacle) and the Y-Axis (near))

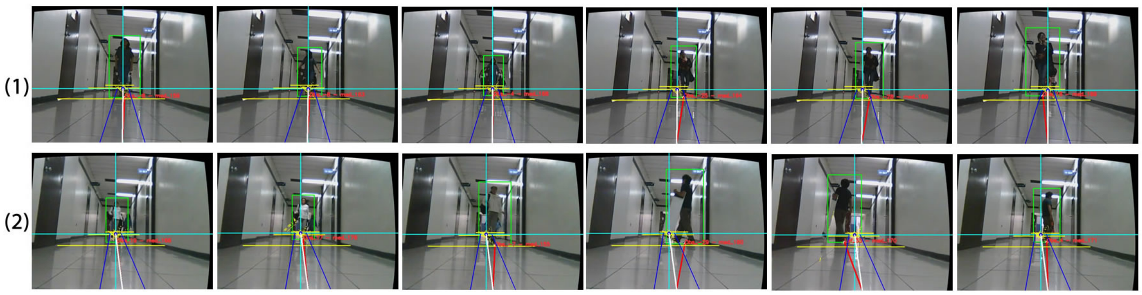

Fig. 10. Second test environment (medium region), the obstacle or people is situated between 201 and 600 cm of the robot. Where shown people detection and localization in the plane of robot (X-Axis (left, right, obstacle) and the Y-Axis (medium))

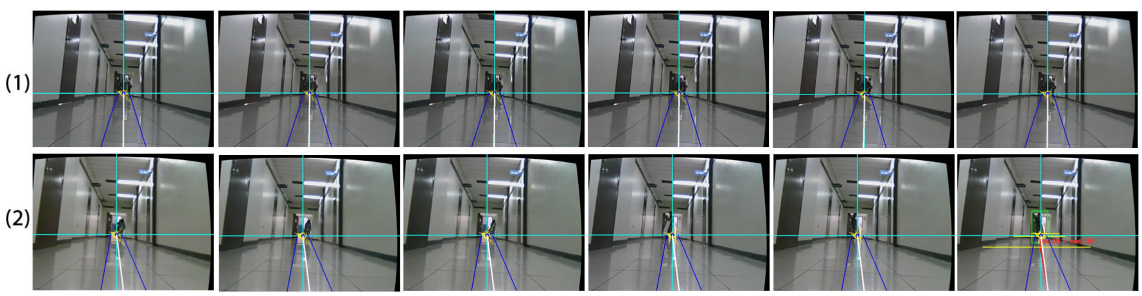

Fig. 11. Third test environment (far region), the obstacle or people is situated between 601 and 1000 cm of the robot. Where shown people detection and localization in the plane of the robot (X-Axis (left, right, obstacle) and the Y-Axis (far))

detection and localization in the plane of the robot (X-Axis (left, right, obstacle) and the Y-Axis (medium)) between the distances of 201 to 600 cm. Table III shows percentages of success, of people detection and localization of the people in the plane of the robot.

TABLE III. RESULT PEOPLE DETECTION AND LOCALIZATION - SECOND TEST ("MEDIUM" REGION)

Distance (201 to 600 cm)	**Amount of analyzed frames**			**Percentage**
	Successes	Errors	Total	**Accuracy (%)**
environment 1	770	58	828	93.00
environment 2	970	70	1040	93.27
Average				**93.13**

The third test is when people is located in the "far region" with respect to the robot, during the test it was observed that when people are at a great distance from robot (601 to 1000 cm.) are difficult to detect; the figure 11 shows people detection and localization; the table IV show percentages of people detection and localization in the plane of the robot.

TABLE IV. RESULT PEOPLE DETECTION AND LOCALIZATION - THIRD TEST ("FAR" REGION)

Distance (601 to 1000 cm)	**Amount of analyzed frames**			**Percentage**
	Successes	Errors	Total	**Accuracy (%)**
environment 1	64	280	344	18.60
environment 2	52	200	252	20.63
Average				**19.62**

After people detection and localization together with vanishing point detection for autonomous robot navigation in structured environment (fig. 12), while the robot is in movement will detect and locate people in their plane (X-Axis (left, right, obstacle) and the Y-Axis (near, medium, far)), in order to avoid obstacles (persons) during navigation, the results are shown in table V.

Corners of the structured environment must be taken into account, as it may be something different: If the corner is not as wide, can be detected good the vanishing point; but if the corner present higher amplitude, then an error occurs in the

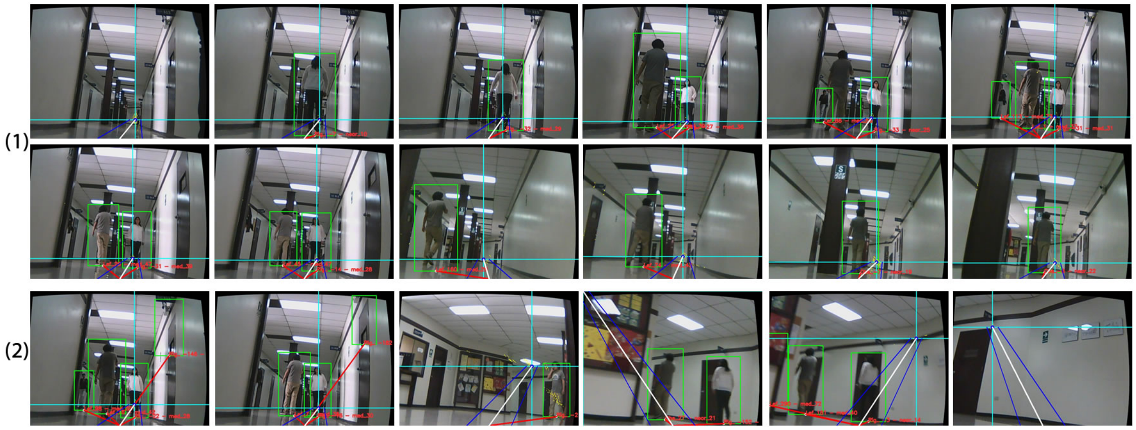

Fig. 12. Where (1) Test environment with autonomous robot in movement and (2) Error in the detection of vanishing point in corners of higher amplitude, but not in the detection of the person

vanishing point detection, therefore needs the previous position of the vanishing point for the navigation as show figure 12_2. However, the detection of persons not affected at all, since well still it detected. Table VI shows percentages of success of the vanishing points detection of the four test environments.

TABLE V. RESULT PEOPLE DETECTION AND LOCALIZATION - FOURTH TEST (AUTONOMOUS ROBOT IN MOVEMENT)

Distance (cm)	**Amount of analyzed frames**			**Percentage**
	Successes	Errors	Total	**Accuracy (%)**
environment 1	796	94	890	89.44
environment 2	1242	108	1350	92.00
Average				**90.72**

TABLE VI. RESULT THE VANISHING POINTS DETECTION

Distance (cm)	**Amount of analyzed frames**			**Percentage**
	Successes	Errors	Total	**Accuracy (%)**
Near Region (50 to 200)	778	23	801	97.13
Medium Region (201 to 600)	1812	56	1868	97.00
Far Region (601 to 1000)	579	17	596	97.15
Robot in Movement	2169	71	2240	96.83
Average				**97.03**

V. CONCLUSION

During the testing can be seen that people detection is a complex problem due to the different positions of human bodies, lighting and background. However for the navigation of autonomous robot, first we need to locate the position of the obstacles (people or object) in the plane of the robot, is why we perform tests at different distance about the Y-Axis as: near (50 to 200 cm), medium (201 to 600 cm) and far (601 to 1000 cm), furthermore the X-Axis (left, right, obstacle) with respect to the robot; the experiments show that people detection and localization is better in the medium region (201 to 600 cm) with respect to the robot, getting 93.13% of accuracy. And the navigation through the vanishing point detection is 97.03% of accuracy; this allows the robot has enough time to evade the obstacle during navigation. But also we have seen cases where it has not vanishing point detection as in the case of occlusion of objects or people (objects that are very near to the robot) and corners more amplitude, then need the previous position of the vanishing point for continuing with the navigation, however this not affect people detection.

ACKNOWLEDGMENT

This work was supported by the *Fondo para la Innovación, Ciencia y Tecnología* (FINCyT), Perú, under the contract 216-FINCyT-IA-2013.

REFERENCES

[1] S. Yao, S. Pan, T. Wang, and C. Zheng, "A new pedestrian detection method based on combined hog and lss features," *Neurocomputing*, vol. 151, pp. 1006–1014, 2015.

[2] C. Castillo and C. Chang, "A method to detect victims in search and rescue operations using template matching," *Safety, Security and Rescue Robotics, Workshop, IEEE International*, pp. 201–206, June 2006.

[3] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, pp. 886–893, 2005.

[4] I. Corporation, "Open computer vision library reference manual," *Intel Corporation, USA*, 2001.

[5] P. Viola and M. Jones, "Rapid object detection using boosted cascade of simple features," *IEEE Conference on Computer Vision and Pattern Recognition*, 2001.

[6] M. Tejada-Begazo, C. Cervantes-Jilaja, R. E. Patiño-Escarcina, and D. Barrios-Aranibar, "Morphological operators applied to human body detection hog method improvement," in *12th IEEE Latin American Robotics Symposium (LARS 2015)*. Uberlândia-Brasil: IEEE, 2015.

[7] C. Cervantes-Jilaja, M. Tejada-Begazo, R. E. Patiño-Escarcina, and D. Barrios-Aranibar, "A new improvement of human bodies detection," in *12th IEEE Latin American Robotics Symposium (LARS 2015)*. Uberlândia-Brasil: IEEE, 2015.

[8] M. Marszalek and M. Schmid, "Accurate object localization with shape masks," *IEEE Conference on Computer Vision & Pattern Recognition*, 2007.

[9] M. Le, S. Phung, and A. Bouzerdoum, "Lane detection in unstructured environments for autonomous navigation systems," *12th Asian Conference on Computer Vision, Singapore*, pp. 414–429, 2014.

[10] ——, "Pedestrian lane detection in unstructured environments for assisted navigation," *International Conference on Digital Image Computing: Techniques and Applications*, 2014.

[11] T. A.K. and S. S., "Environment interpretation for autonomous indoor navigation of micro air vehicles," *IEEE TechSym 2014 - 2014 IEEE Students' Technology Symposium*, pp. 87–92, 2014.

[12] Y. Lu and D. Song, "Visual navigation using heterogeneous landmarks and unsupervised geometric constraints," *IEEE Transactions on Robotics*, vol. 31, pp. 736–749, 2015.