

Omnidirectional Vision for Mobile Robot Human Body Detection and Localization

Hong Liu, Zhenhua Huo, Ge Yang

Key Laboratory of Machine Perception and Intelligence,
Shenzhen Graduate School, Peking University, China

E-mail: liuh@szpku.edu.cn, huozh07115@szcie.pku.edu.cn, yangge@szpku.edu.cn

Abstract—Human body detection and localization is an essential capability of an autonomous mobile robot which works in the human-robot interaction (HRI) environments. However, due to field of view (FOV) limitations, it is hard to detect all human bodies around a mobile robot by using a conventional camera, and distances between robots and human bodies are also difficult to estimate. In this paper, we propose a novel omnidirectional visual system to locate positions of human bodies for an autonomous mobile robot. Firstly, a handy fitting shape based method (FSM) is presented to remap a omnidirectional image to a bird’s eye view image. A new bird’s eye view image segmentation algorithm, which is inspired by image pyramids, is used to split obstacle objects and ground plane. Secondly, a shape-based human body detector is implemented in unwrapped omnidirectional images to locate regions of human bodies. These human body detection results are combined with bird’s eye view image segmentation to distinguish human bodies from other obstacle objects. Experiments show that our system performs well in human-robot interaction environments.

Index Terms—Omnidirectional vision, human body detection and localization, bird’s eye view image segmentation.

I. INTRODUCTION

In recent years, the omnidirectional camera is widely used in visual based robot navigation and localization, which is due to the large field of view [1-4]. The theoretical model of omnidirectional cameras was discussed in [5], [6]. From these previous works, it can be seen that the omnidirectional image has a relationship to the ground plane, and distances from mobile robots can be estimated by using an omnidirectional camera. The bird’s eye view image is an important transformational image of an omnidirectional image. A projection model for bird’s eye view images transformation was proposed in [3]. However, it is only suitable for the single projection center omnidirectional cameras.

Motion detection and tracking by using the omnidirectional camera is also developed in these years [7], [8], [9]. Due to the omnidirectional camera with a 360-degree field of view in the ground plane, the robot which is equipped with an omnidirectional camera is able to acquire the whole visual information around itself at the same time. It is very appropriately used to detect all human beings around the robot in human-robot environments. Human bodies can be split from omnidirectional images by human motion information [7], [8]. However, the motion of omnidirectional cameras is limited in these works. Human body detection in a single frame omnidirectional image

is introduced for mobile robots. The local feature based method performs well in the task of human body detection [10], [11], [12], [13]. The feature around a point of interesting on human silhouette is transformed to a n^{th} vector by using shape-contexts descriptor [9], [12]. Some feature vectors are picked out as template through feature selection to detect regions of human bodies from a single frame image.

In this paper, a human body detection method based on local shape is implemented to detect human body regions in unwrapped omnidirectional images. A handy fitting shape based method (FSM) is presented to transform an omnidirectional image to a bird’s eye view image. A new bird’s eye view image segmentation algorithm is proposed to split obstacle objects and ground plane. Human body localization results are obtained by using human body detection results and bird’s eye view images segmentation, and human body localization results can be used to estimate distance from mobile robots in human-robot interaction environments.

The rest of this paper is organised as follows: Two important transformations of an omnidirectional are illustrated in section II. Human body detection is described in Section III. Human body localization is introduced in Section IV. Experiments analysis is in Section V, and Section VI is conclusions.

II. OMNIDIRECTIONAL IMAGE TRANSFORMATIONS

The omnidirectional camera is composed of a CCD camera and a convex mirror. Since the omnidirectional image is distorted and circular, it is difficult to directly analysis and processing [5]. Two main geometry transformations of the omnidirectional image are necessary in our omnidirectional visual system.

A. Unwrapping the omnidirectional image

The omnidirectional image is a circular image because of the shape of convex mirror. The unwrapped omnidirectional image is a rectangle image which is a normal view of the omnidirectional image. The unwrapping operation of an omnidirectional image is a coordinate transformation from polar to Cartesian. A pixel in unwrapping image has its corresponding pixel in the original omnidirectional image, $O(x_{n0}, y_{n0}) \rightarrow O(x_{m0}, y_{m0})$, $P(x_n, y_n) \rightarrow P(x_m, y_m)$, as shown in Fig.1.

The geometry relation of two shape image in Cartesian coordinate is illustrated as following diagrams and equations.

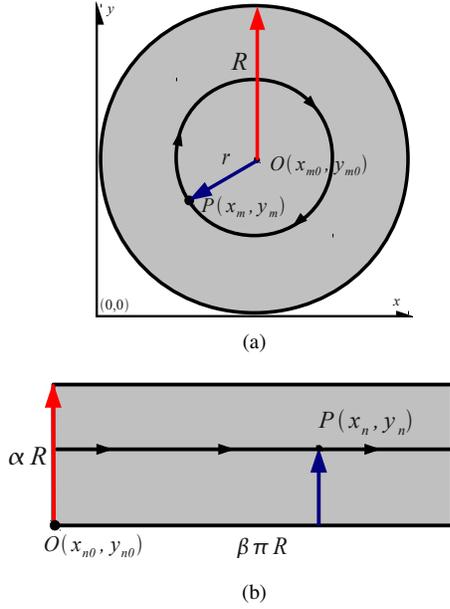


Fig.1 relation of (a) omnidirectional image and (b) unwrapped omnidirectional image

The height of the unwrapped omnidirectional image is αR and the width is $\beta \pi R$. α, β are scale ratio of width and height. $O(x_{m0}, y_{m0})$ is the center of the omnidirectional image, and following equations is described the relation of this transformation.

$$H = \alpha R \quad (1)$$

$$W = \beta \pi R \quad (2)$$

$$y_n = \alpha \sqrt{(x_m - x_{m0})^2 + (y_m - y_{m0})^2} \quad (3)$$

$$x_n = 2\pi y_n \beta \frac{1}{2\pi} \arctan\left(\frac{y_m - y_{m0}}{x_m - x_{m0}}\right) \quad (4)$$

From the transformation above, it is known that the upper part of an unwrapped omnidirectional image is clearer than the lower part. Many corresponding points of the lower part are the same point in the omnidirectional image, however this is rarely happened in the upper part of the unwrapped omnidirectional image. This is also the reason why the upper human body features are more robust than the lower body in unwrapped omnidirectional images.

B. Bird's eye view image

The omnidirectional image is seriously distorted in radial orientation. Accordingly, it is necessary to correct this distortion for distance estimation. Bird's eye view image has been implemented in vision-based navigation [1-3]. The bird's eye view image transformation model which was proposed in [3], is based on a single projection center camera. In [6] a model based on mirror shape was introduced to establish the relation of omnidirectional image space and world space. However, it is unknown that whether our omnidirectional camera is equipped with a single projection center mirror or not, and the

function of mirror shape is also unknown. Inspired by model in [6], we introduce a fitting shape based method to remapping omnidirectional images to bird's eye view images, especially all parameters of camera are unknown.

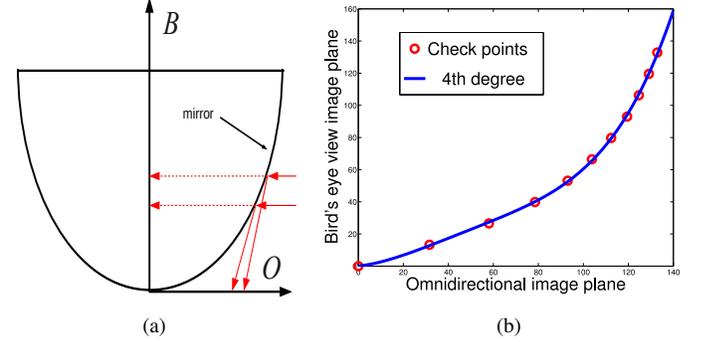


Fig.2 (a) convex mirror and (b) mirror shape function

A wide field of view is acquired by using the convex mirror to reflect light rays. However, a large amounts of visual information are compressed to a image plane, which leads to image distortion. Bird's eye view transformation is a process of decompression, which is projecting omnidirectional image to the horizontal plane.

A few checkpoints in an omnidirectional image are obtained by using check board on the horizontal plane. These points are used to establish the relationship between *omnidirectional image plane* and *bird's eye view image plane*. It is a transformation from distorted image to a plane image. The 4th degree polynomial fitting which is the best fitting of the check points is selected to construct the shape function of our omnidirectional camera mirror, as shown in Fig.2. The transformation is yield the following equations in the Polar.

$$r_b = Ar_o^4 + Br_o^3 + Cr_o^2 + Dr_o + E \quad (5)$$

$$\theta_b = \theta_o \quad (6)$$

$$0 \leq r_o \leq R_m \quad (7)$$

Here, A, B, C, D, E are constants, and R_m is the maximum radius of the check points. In practice, the maximum value R_o and the height of the omnidirectional camera is related as shown in Fig.3. The bird's eye view image can be used to measure distance from the robot to the object, when the object and the robot are on the same plane.

As shown in Fig.3, an omnidirectional camera is calibrated in the height of h , and equipped with a mobile robot in the height of H , here is $r_w = r_c \frac{H}{h}$, and $H \gg h$. In the *work plane*, the pixels which are far from the omnidirectional image center are not necessary to transformation, because they are so compressive that can not be distinguished each other. The image region of interest (ROI) of the *calibration plane* is larger than the image of the *work plane*. This fitting shape based method can be implemented to bird's eye view image transformations for human body localization.

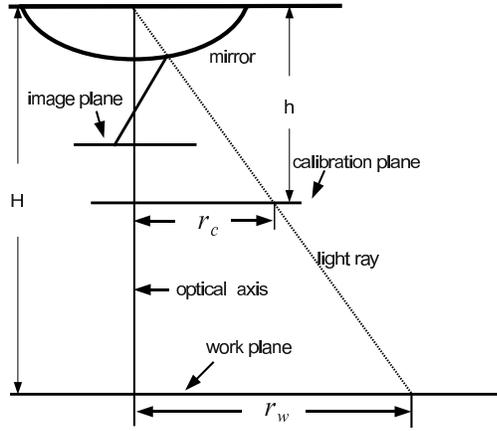


Fig.3 distance estimation model

After two main transformations, the framework of human body detection and localization is illustrated in Fig.4.

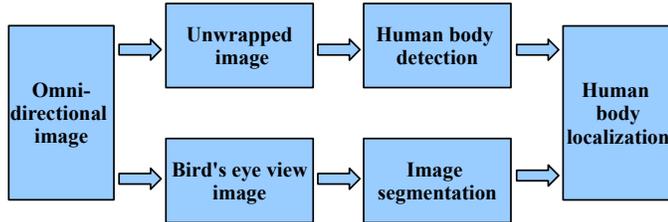


Fig.4 framework of human body detection and localization

III. HUMAN BODY DETECTION

Human body is a non-rigid object, it has many appearances at different perspectives. However, parts of human body have the characteristics of rigid object, which is relative stable in images. Hence the local feature based method performs well in the task of human body detection [11], [13]. A local shape features based detector is introduced to localize human body. When the omnidirectional camera is more higher than human bodies, the whole human body can be acquired by the omnidirectional camera. Human body detection is applied on the unwrapped omnidirectional image, which is a rectangle coordinate image transformed from the omnidirectional image as illustrated above.

A. Features description

The detail information of omnidirectional images is not obvious, and the omnidirectional image is usually distorted due to the projection theory of the omnidirectional camera [5]. The unwrapped omnidirectional image is similar with normal images, however, it is distorted serious as shown in Fig.5. Because of image distortion, human silhouettes are various in unwrapped omnidirectional image. Some stable features must be extracted to describe the human body. Shape contexts descriptor is introduced to describe human body features in the unwrapped omnidirectional image in our visual system.



Fig.5 unwrapped omnidirectional image with human bodies

The shape contexts descriptor is used to represent features around the point of interesting, as shown in Fig.6. The feature of a point on human silhouette is transformed to a N -dimensional vector (S_1, S_2, \dots, S_n) . The same scale around the feature point is selected, due to human body has not large scale changes in the unwrapped omnidirectional image. The human body silhouette has lots of shape features, and which shape features are more different from background features should be selected to detect human body. The skeleton of the unwrapped omnidirectional image is used to these features extraction.



Fig.6 shape contexts descriptor

The similarity of two shape contexts features is used to describe features matching degree. The distance of two normalized features vectors is measured as follow formula.

$$D_s = \frac{1}{2} \sum_{i=1}^N \frac{[s_i - m_i]^2}{s_i + m_i} \quad (8)$$

In the formula (8), s_i is the i -th component of vector (s_1, s_2, \dots, s_n) , and m_i is the i -th component of vector (m_1, m_2, \dots, m_n) . The value of D_s is from 0 to 1, the more nearer to 1 the more similar of two vectors each other.

B. Features selection

Features which should be extracted are more similar with all human body perspectives and more different from environments. The head, shoulder, upper body are relatively stable in images by lots of observations, and features of these parts are extracted manually from human sample images.

Several positive and negative test samples are manually cropped to check the selected features, as show in the following Fig.7. The first line is negative samples which are background regions, and the second line is positive samples which are human body images. All these samples are parts of unwrapped omnidirectional skeleton image.

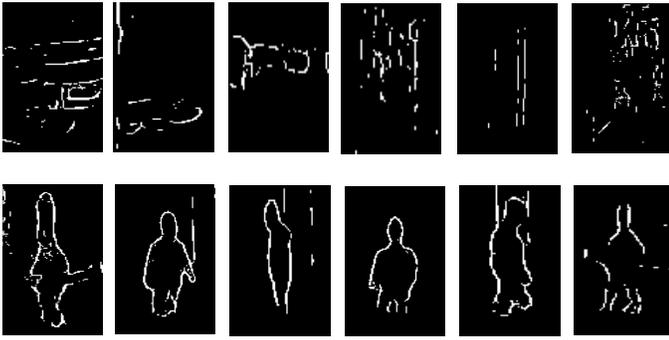


Fig.7 negative and positive test samples

A lot of shape contexts features extracted from negative and positive sample images are matched with a template feature, the results are shown in Fig.8. The template feature is selected manually, which include a priori knowledge. By testing with positive and negative samples, it is proved that the priori knowledge of human is stable as template feature to detect human body in the unwrapped omnidirectional image.

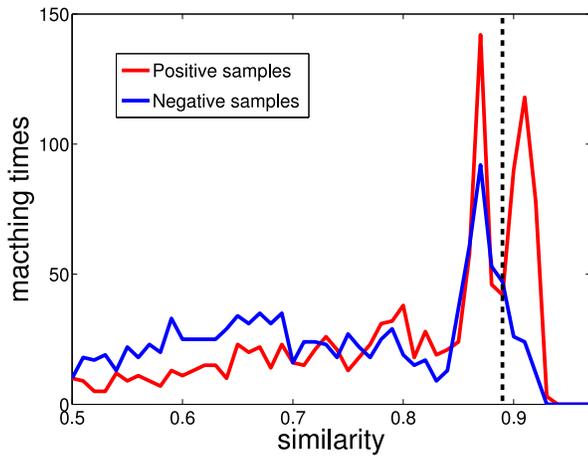


Fig.8 negative and positive test samples

As shown in Fig.8, the results of the template feature matched with positive samples are regular, the similarity is a trend to maximum. The similarity of negative samples is different from the similarity of positive samples. Accordingly, this template feature is able to detect human body from background. Several features are extracted manually, and which performs well will be selected as a template feature.

$$T_n = \{h_1, h_2, \dots, h_i, \dots, h_n\} \quad (9)$$

$$H_n = w_1 h_1 + w_2 h_2 + \dots + w_i h_i \dots + w_n h_n \quad (10)$$

Where T_n is a series of template features, H_n is the points of interesting, w_i is the threshold of similarity of template feature h_i . The equation (10) constructs a weak classifier. This weak classifier includes lots of human priori knowledge instead of a large amount of samples training. The result of human

body detection is shown Fig.9. Although, there are some negative points of interest, the positive points are concentrated in human body regions.

C. Regions of human body

The boundary of region of interesting (ROI) is split by using projection firstly. Through lots of observation and analysis, it is found that the most isolated points, which are obtained from the previous section, are background points. And most of near vertical points which are far from the other points are not points on human bodies. The others as points of interest are grouped as shown Fig.9 following.



Fig.9 human body detection result

IV. HUMAN BODY LOCALIZATION

A lot of human body parts are detected in the previous section, and regions of interest (ROI) are obtained. However, these regions are coarsely localize the human body position. Because the lower body is not stable in unwrapped omnidirectional image, regions of interest (ROI) is fewer than the upper human body. A new bird's eye view image segmentation is proposed in this section to locate human body positions by using the lower body information.

A. Bird's eye view image segmentation

The bird's eye view transformation of the omnidirectional image is used to detect the lower human body. Due to human foot and mobile wheels being on the same plane, the bird's eye view image is implemented to measure distance in our omnidirectional visual system. After bird's eye view transformation, the feature of the lower body is more obvious and stable, as shown in Fig.10(b).



Fig.10 (a) omnidirectional image and (b) bird's eye view image

The bird's eye view image should be a part of the whole bird's eye image, because the whole bird's eye view image is too large and a part of it is enough to human body localization

for a mobile robot in human-robot interaction environments. The other parts of bird's eye view image is considered that human body is very far from the mobile robot.

As shown in Fig.10(b), all obstacle objects on the ground have radial lines in the bird's eye view image. These radial lines are extracted to split the object regions and ground regions. However, it is difficult to split obstacle objects directly from the whole image by using these radial edges, as shown in Fig.11(a).



Fig.11 (a)radial edge image and (b) image segmentation

In this section, we propose a simple and effective method to segment bird's eye view image by using radial edges. We all know that in bird's eye view image: Firstly, the position of the obstacle object is the nearest radial line region position from bird's eye view images center. Secondly, the other radial line regions are vertical informations transformed to ground planes by bird's eye view transformation.

There are two classes images which are used to bird's eye view image segmentation. One class images are called larger scale images (LSI), which are larger than original bird's eye view images. The other class images are called double rotated images (DRI). A double rotated images is constructed by a clockwise rotated image and a counter-clockwise rotated image. The rotated angles are the same in the clockwise rotated image and the counter-clockwise rotated image. We have two conclusions from these: Firstly, a larger scale image of a origin radial edge image have parts of vertical informations. Secondly, positions of obstacle objects has not changed in double rotated images. The method is illustrated as following equations.

$$S_v = S_{v0} + S_{v1} + \dots S_{vi} \dots + S_{vn} \quad (11)$$

$$S_r = S_{r0} + S_{r1} + \dots S_{ri} \dots + S_{rn} \quad (12)$$

$$S = S_o + S_v + S_r \quad (13)$$

Where,

$$S_{vi} = \alpha S_o \quad (14)$$

$$S_{ri} = \theta_i S_o + (2\pi - \theta_i) S_o \quad (15)$$

Here, S_{vi} is a larger scale image (LSI) of a bird's eye view image, S_{ri} is a double rotated image (DRI) of a bird's eye view image, α is the scale ratio, θ is a small rotated angle. S_o is the original bird's eye view image, S is the result of image segmentation as shown Fig.11(b).

B. Human body localization

The bird's eye view segmentation is combined with results of human body detection to locate human body positions around a autonomous mobile robots. The pseudo code of human body localization is shown bellow:

Algorithm 1 Human body localization

```

while human body regions do
  if obstacle regions then
    obstacle object is human body;
    locate human body position;
  else if none of obstacle object then
    human body is far from robot;
    continue;
  end if
end while

```

Human body regions are distinguished from obstacle object regions by using human body localization algorithm, and human body positions are also obtained. Human body regions are marked by different colors in Fig.12(a), and human body positions are shown in Fig.12(b). The human body position can be used to estimate the distance from mobile robots.

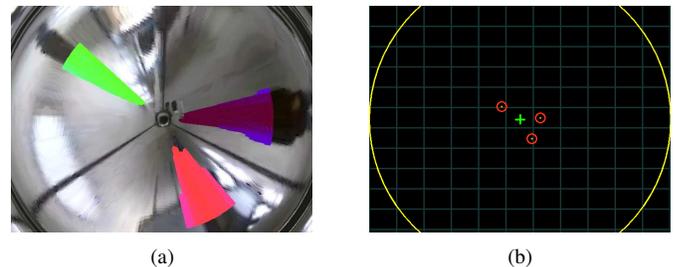


Fig.12 (a) human body regions and (b) human body localization results

The region of a human body is marked as $H(\theta_1, \theta_2, r)$, and the origin is the center of the bird's eye view image. r is the minimum radius of the human body region. θ_1 and θ_2 are the boundary of the human body region, and the same as the boundary of human body in the world space. r is proportional with the distance from a human body to the mobile robot in the world space.

V. EXPERIMENTS AND ANALYSIS

Our autonomous mobile robot is equipped with a omnidirectional camera as shown in Fig.13(a). The height is 1.90m from the ground plane to the top of the omnidirectional camera mirror. Three people and the autonomous mobile robot work in a semi-open hall which area is $8m * 8m$. The autonomous mobile works in four different regions of the semi-open hall and three people walk around it. In four different regions of the semi-open hall, the background is more different from each other in omnidirectional images, and light rays are very brighter. There are also some other obstacle objects on the ground.

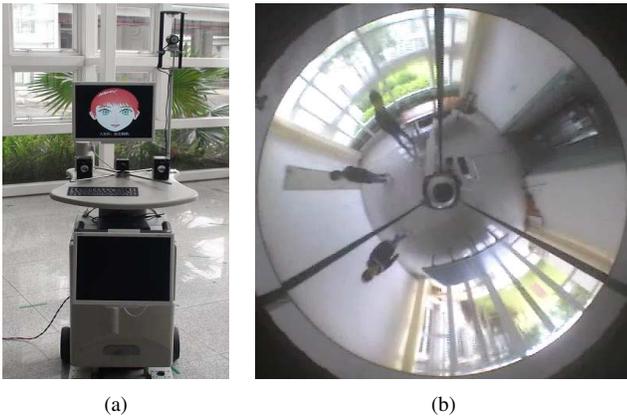


Fig.13 (a) our mobile robot and (b) experimental environment

The real position of human body in the bird's eye view image is marked manually, the measurement of experimental results is definition as the Euclidean distance between the manual marked position and detection results. if the distance is larger than five pixels, the experimental result is consider false. Experiments results are shown in Table I.

TABLE I
RESULTS OF EXPERIMENTS

| No. | Results | FN | FP | FNR | FPR |
|---------|---------|----|-----|-------|--------|
| 1 | 335 | 9 | 107 | 2.67% | 31.94% |
| 2 | 251 | 7 | 80 | 2.79% | 30.87% |
| 3 | 340 | 12 | 85 | 3.53% | 25.00% |
| 4 | 262 | 12 | 60 | 4.58% | 22.90% |
| Summary | 1188 | 40 | 332 | 3.37% | 27.95% |

From the Table I, the false positive rate (FPR) is relatively low, and the false negative rate (FNR) is a little higher. It is shown that most human bodies are detected and some other objects are considered as human body in our system. Due to the low resolution and distortion of the omnidirectional image, it is difficult to detect human bodies from other objects. However, almost objects are detected by using bird's eye view images segmentations. Due to the safety of human and robot is the first to be considered in human-robot environments, a lower false negative rate illustrated that our omnidirectional visual system performs well in these environments.

VI. CONCLUSIONS

Human body detection and localization is accomplished by using a single frame omnidirectional image in this paper. The shape-based method is used to detect human bodies from backgrounds, and the novel bird's eye view image segmentation is proposed to distinguish obstacle objects from ground areas. These cues are combined to localize human body positions. The experimental results show that our omnidirectional visual system is satisfied with the requirement of safe human-robot

interaction. Furthermore, the omnidirectional visual system may be reinforced by adding other obstacle features to distinguish human bodies from background.

ACKNOWLEDGMENT

This work is supported by National Natural Science Foundation of China (NSFC, No.60875050, 60675025), National High Technology Research and Development Program of China (863 Program, No.2006AA04Z247), Shenzhen Scientific and Technological Plan and Basic Research program (No.JC200903160369A), Natural Science Foundation of Guangdong(No.9151806001000025).

REFERENCES

- [1] J. Gaspar, N. Winters and J. S. Victor, "Vision-based Navigation and Environmental Representations with an Omni-directional Camera", IEEE Transactions on Robotics and Automation, Vol. 16(6), pp:890-898, 2000.
- [2] J. Gaspar and J. S. Victor, "Visual Path Following with a Catadioptric Panoramic Camera", International Symposium on Intelligent Robotic Systems(SIRS), pp:139-147, 1999.
- [3] R. F. Vassallo, L. F. Encarnacao, J. S. Victor and H. J. Schneebeli, "Bird's Eye View Remapping and Path Following Based on Omnidirectional vision", Processing of the XV Automatic Brazilian Conference, 2004.
- [4] H. Koyasu, J. Miura and Y. Shirai, "Recognizing Moving Obstacles for Robot Navigation using Real-time Omnidirectional Stereo Vision", Journal of Robotics and Mechatronics, Vol. 14, No. 2, pp. 147-156, 2002.
- [5] S. K. Nayar, "Catadioptric Omnidirectional Camera", IEEE Conf. on Computer Vision and Pattern Recognition, pp. 482-488, 1997.
- [6] R. A. Hicks and R. Bajcsy, " Reflective surfaces as computational sensors", Image and Vision Computing, Vol. 19, pp.773-777, 2001.
- [7] H. Liu, W. Pi and H. Zha, "Motion Detection for Multiple Moving Targets by Using an Omnidirectional Camera", IEEE Conf. on Robotics, Intelligent Systems and Signal Processing, pp. 422-426, 2003.
- [8] H. Liu, N. Dong, H. Zha, Omni-directional Vision based Human Motion Detection for Autonomous Mobile Robots, Proc. IEEE International Conference on System Man and Cybernetics (SMC'05), Hawaii, USA, pp. 2236-2241, 2005.
- [9] S. Belongie and J. Malik, "Matching with Shape Contexts", Content-based Access of Image and Video Libraries, pp. 20-26, 2000.
- [10] B. Leibe, A. Leonardis, and B. Schiele, "Robust Object Detection with Interleaved Categorization and Segmentation", International Journal of Computer Vision Special Issue on Learning for Recognition and Recognition for Learning, Vol. 77, No. 1-3, pp:259-289, 2008.
- [11] J. Park, Y. Luo, H. Wang, and Y. L. Murphey, "Pedestrian detection by modeling local convex shape features", IEEE Conf. on Pattern Recognition, pp. 1-4, 2008.
- [12] M. Barnard and J. Heikkila, "On Bin Configuration of Shape Context Descriptors in Human Silhouette Classification", Advanced Concepts for Intelligent Vision Systems, pp: 850-859, 2008.
- [13] A. Broggi, M. Sechi, M. Bertozzi, and A. Fascioli, "Shape-based Pedestrian Detection", IEEE Conf. on Intelligent Vehicles Symposium, pp:215-220, 2000.
- [14] D. Gavrilu, "Multi-cue Pedestrian Detection and Tracking from a Moving Vehicle" IJCV, Vol. 73(1), pp:41-59, 2007.
- [15] M. Jogan and A. Leonardis, "Robust Localization using Panoramic View-based Recognition", CVPR, pp:136-139, 2000.
- [16] N. Bird, S. Atev, N. Caramelli, R. Martin, O. Masoud, and N. Papanikolopoulos, "Real Time Online Detection of Abandoned Objects in Public Areas", ICRA, pp:3775-3780, 2006.
- [17] K. Kim, T. Chalidabhongse, D. Harwood, , and L. Davis, "Background Modeling and Subtraction by Codebook Construction", IEEE Conf. on Image Processing, pp:3061-3064, 2004.
- [18] P. Viola, M. Jones, and D. Snow, "Detecting Pedestrians using Patterns of Motion and Appearance", ICCV, pp:734-741, 2003.
- [19] B. Wu and R. Nevatia, "Detection of Multiple, Partially Occluded Humans in a Single Image by Bayesian Combination of Edgelet Part Detectors", ICCV, pp:90-97, 2005.
- [20] G. Cielniak, M. Miladinovic, D. Hammarin, L. Goranson, A. Lilienthal, and T. Duckett, "Appearance-based Tracking of Persons with an Omnidirectional Vision Sensor", CVPRW, pp:1063-1069, 2003.