

# Methods for Reliable Robot Vision with a Dioptric System

E. Martínez and A.P. del Pobil

*Robotic Intelligence Lab., Jaume-I University, Castellón, Spain*

*Interaction Science Dept., Sungkyunkwan University, Seoul, S. Korea*

## 1. Introduction

There is a growing interest in Robotics research on building robots which behave and even look like human beings. Thus, from industrial robots, which act in a restricted, controlled, well-known environment, today's robot development is conducted as to emulate alive beings in their natural environment, that is, a real environment which might be dynamic and unknown. In the case of mimicking human being behaviors, a key issue is how to perform manipulation tasks, such as picking up or carrying objects. However, as this kind of actions implies an interaction with the environment, they should be performed in such a way that the safety of all elements present in the robot workspace at each time is guaranteed, especially when they are human beings.

Although some devices have been developed to avoid collisions, such as, for instance, cages, laser fencing or visual acoustic signals, they considerably restrict the system autonomy and flexibility. Thus, with the aim of avoiding those constrains, a robot-embedded sensor might be suitable for our goal. Among the available ones, cameras are a good alternative since they are an important source of information. On the one hand, they allow a robot system to identify interest objects, that is, objects it must interact with. On the other hand, in a human-populated, everyday environment, from visual input, it is possible to build an environment representation from which a collision-free path might be generated. Nevertheless, it is not straightforward to successfully deal with this safety issue by using traditional cameras due to its limited field of view. That constrain could not be removed by combining several images captured by rotating a camera or strategically positioning a set of them, since it is necessary to establish any feature correspondence between many images at any time. This processing entails a high computational cost which makes them fail for real-time tasks.

Despite combining mirrors with conventional imaging systems, known as catadioptric sensors (Svoboda et al., 1998; Wei et al., 1998; Baker & Nayar, 1999) might be an effective solution, these devices unfortunately exhibit a dead area in the centre of the image that can be an important drawback in some applications. For that reason, a dioptric system is proposed. Dioptric systems, also called fisheye cameras, are systems which combine a fisheye lens with a conventional camera (Baker & Nayar, 1998; Wood, 2006). Thus, a conventional lens is changed by one of these lenses that has a short focal length that allows cameras to see objects in an hemisphere. Although fisheye devices present several advantages over catadioptric sensors

such as no presence of dead areas in the captured images, a unique model for this kind of cameras does not exist unlike central catadioptric ones (Geyer & Daniilidis, 2000).

So, with the aim of designing a dependable, autonomous, manipulation robot system, a fast, robust vision system is presented that covers the full robot workspace. Two different stages have been considered:

- moving object detection
- target tracking

First of all, a new robust adaptive background model has been designed. It allows the system to adapt to different unexpected changes in the scene such as sudden illumination changes, blinking of computer screens, shadows or changes induced by camera motion or sensor noise. Then, a tracking process takes place. Finally, the estimation of the distance between the system and the detected objects is computed by using an additional method. In this case, information about the 3D localization of the detected objects with respect to the system was obtained from a dioptic stereo system.

Thus, the structure of this paper is as follows: the new robust adaptive background model is described in Section 2, while in Section 3 the tracking process is introduced. An epipolar geometry study of a dioptic stereo system is presented in Section 4. Some experimental results are presented in Section 5, and discussed in Section 6.

## 2. Moving Object Detection: A New Background Maintenance Approach

As it was presented in (Cervera et al., 2008), an adaptive background modelling combined with a global illumination change detection method is used to properly detect any moving object in a robot workspace and its surrounding area. That approach can be summarized as follows:

- In a first phase, a background model is built. This model associates a statistical distribution, defined by its mean color value and its variance, to each pixel of the image. It is important to note that the implemented method allows to obtain the initial background model without any restrictions of bootstrapping
- In a second phase, two different processing stages take place:
  - First, each image is processed at pixel level, in which the background model is used to classify pixels as foreground or background depending on whether they fit in with the built model or not
  - Second, the raw classification based on the background model is improved at frame level

Moreover, when a global change in illumination occurs, it is detected at frame level and the background model is properly adapted.

Thus, when a human or another moving object enters in a room where the robot is, it is detected by means of the background model at pixel level. It is possible because each pixel belonging to the moving object has an intensity value which does not fit into the background model. Then, the obtained binary image is refined by using a combination of subtraction techniques at frame level. Moreover, two consecutive morphological operations are applied to erase isolated points or lines caused by the dynamic factors mentioned above. The next step is to update the statistical model with the values of the pixels classified as background in order to adapt it to some small changes that do not represent targets.

At the same time, a process for sudden illumination change detection is performed at frame level. This step is necessary because the model is based on intensity values and a change in illumination produces a variation of them. A new adaptive background model is built when an event of this type occurs, because if it was not done, the application would detect background pixels as if they were moving objects.

### 3. Tracking Process

Once targets are detected by using a background maintenance model, the next step is to track each target. A widely used approach in Computer Vision to deal with this problem is the Kalman Filter (Kalman & Bucy, 1961; Bar-Shalom & Li, 1998; Haykin, 2001; Zarchan & Mussof, 2005; Grewal et al., 2008). It is an efficient recursive filter that has two distinct phases: **Predict** and **Update**. The predict phase uses the state estimate from the previous timestep to produce an estimate of the state at the current timestep. This predicted state estimate is also known as the *a priori* state estimate because, although it is an estimate of the state at the current timestep, it does not include observation information from the current timestep. In the update phase, the current *a priori* prediction is combined with current observation information to refine the state estimate. This improved estimate is termed the *a posteriori* state estimate. With regard to the current observation information, it is obtained by means of an image correspondence approach. In that sense, one of most well-known methods is the Scale Invariant Feature Transform (SIFT) approach (Lowe, 1999; Lowe, 2004) which shares many features with neuron responses in primate vision. Basically, it is a 4-stage filtering approach that provides a *feature* description of an object. That feature array allows a system to locate a target in an image containing many other objects. Thus, after calculating feature vectors, known as SIFT keys, a nearest-neighbor approach is used to identify possible objects in an image. Moreover, that array of features is not affected by many of the complications experienced in other methods such as object scaling and/or rotation. However, some disadvantages made us discard SIFT for our purpose:

- It uses a varying number of features to describe an image and sometimes it might be not enough
- Detecting substantial levels of occlusion requires a large number of SIFT keys what can result in a high computational cost
- Large collections of keys can be space-consuming when many targets have to be tracked
- It was designed for perspective cameras, not for fisheye ones

All these approaches have been developed for perspective cameras. Although some research has been carried out to adapt them to omnidirectional devices (Fiala, 2005; Tamimi et al., 2006), a common solution is to apply a transformation to the omnidirectional image in order to obtain a panoramic one and to be able to use a traditional approach (Cielniak et al., 2003; Liu et al., 2003; Potúcek, 2003; Zhu et. al (2004); Mauthner et al. (2006); Puig et al., 2008). However, this might give rise to a high computational cost and/or mismatching errors. For all those reasons, we propose a new method which is composed of three different steps (see Fig. 1):

1. The minimum bounding rectangle that encloses each detected target is computed
2. Each area described by a minimum rectangle identifying a target, is transformed into a perspective image. For that, the following transformation is used:

$$\phi = c/R_{\text{med}} \quad (1)$$

$$R_{\text{med}} = (R_{\text{min}} + R_{\text{max}})/2 \quad (2)$$

$$c_1 = c_0 + (R_{\text{min}} + r) \sin \phi \quad (3)$$

$$r_1 = r_0 + (R_{\text{min}} + r) \cos \phi \quad (4)$$

where  $(c_0, r_0)$  represent the image center coordinates in terms of column and row respectively, while  $(c_1, r_1)$  are the coordinates in the new perspective image;  $\phi$  is the angle between Y-axis and the ray from the image center to the considered pixel; and,  $R_{\text{min}}$  and  $R_{\text{max}}$  are, respectively, the minimum and maximum radius of a torus which encloses the area to be transformed. These radii are obtained from the four corners of the minimum rectangle that encloses the target to be transformed. However, there is a special situation to be taken into account. It is produced when the center of the image is in the minimum rectangle. This situation, once it is detected, is solved by setting  $R_{\text{min}}$  to 0.

On the other hand, note that only the detected targets are transformed into its cylindrical panoramic image, not the whole image. It allows the system to reduce the computational cost and time consumption. Moreover, the orientation problem is also solved, since all the resulting cylindrical panoramic images have the same orientation. In this way, it is easier to compare two different images of the same interest object.

3. The cylindrical panoramic images obtained for the detected targets are compared with the ones obtained in the previous frame. A similarity likelihood criterion is used for matching different images of the same interest object. Moreover, in order to reduce computational cost, images to be compared must have a centroid distance lower than a threshold. This distance is measured in the fisheye image and is described by means of a circle of possible situations of the target having as center its current position and as radius the maximum distance allowed. In that way, the occlusion problem is solved, since all parts of the cylindrical panoramic images are being analyzed.



Fig. 1. Tracking Process

This method is valid for matching targets in a monocular video sequence. Thus, when a dioptic stereo system is provided, the processing is changed. When an interest object firstly appears in an image, it is necessary to establish its corresponding image in the other camera of the stereo pair. For that, a disparity estimation approach is used. That estimation is carried out until the interest object is in the field of view of both stereo cameras. In that moment, the same identifier is assigned to both images of the interest object. Then, matching is done as a monocular system in a parallel way, but as both images have the same identifier it is possible to estimate its distance from the system whenever it is necessary. As it is performed in a parallel way and processing is not time-consuming, a real-time performance is obtained.

## 4. Experimental Results

For the experiments carried out, a mobile manipulator was used which incorporates a visual system composed of 2 fisheye cameras mounted on the robot base, pointing upwards to the ceiling, to guarantee the safety in its whole workspace. Fig. 2 depicts our experimental setup, which consists of a mobile Nomadic XR4000 base, a Mitsubishi PA10 arm, and two fisheye cameras (Sony SSC-DC330P 1/3-inch color cameras with fish-eye vari-focal lenses Fujinon YV2.2x1.4A-2, which provide 185-degree field of view). Images to be processed were acquired in 24-bit RGB color space with a 640x480 resolution.



Fig. 2. Experimental Set-up

Different experiments were carried out obtaining a good performance in all of them. Some of the experimental results are depicted in Fig. 3. where different parts of the whole process are shown. The first column represents the image captured by a fisheye camera, then the binary image generated by the proposed approach appears in the next column. The three remaining columns represent several phases of the tracking process, that is, the generation of the minimum bounding rectangle and the cylindrical panoramic images. Note that in all cases, both the detection approach and the proposed tracking process were successful in their purpose, although some occlusions, rotation and/or scaling had occurred.

## 5. Conclusions & Future Work

We have presented a dioptric system for reliable robot vision by focusing on the tracking process. Dioptric cameras have the clear advantage of covering the whole workspace without resulting in a time consuming application, but there is little previous work about this kind of devices. Consequently, we had to implement novel techniques to achieve our goal. Thus, on the one hand, a process to detect moving objects within the observed scene was designed. The proposed adaptive background modeling approach combines moving object detection with global illumination change identification. It is composed of two different phases, which consider several factors which may affect the detection process, so that constraints in illumination conditions do not exist, and neither is it necessary to wait for some time for collecting enough data before starting to process.

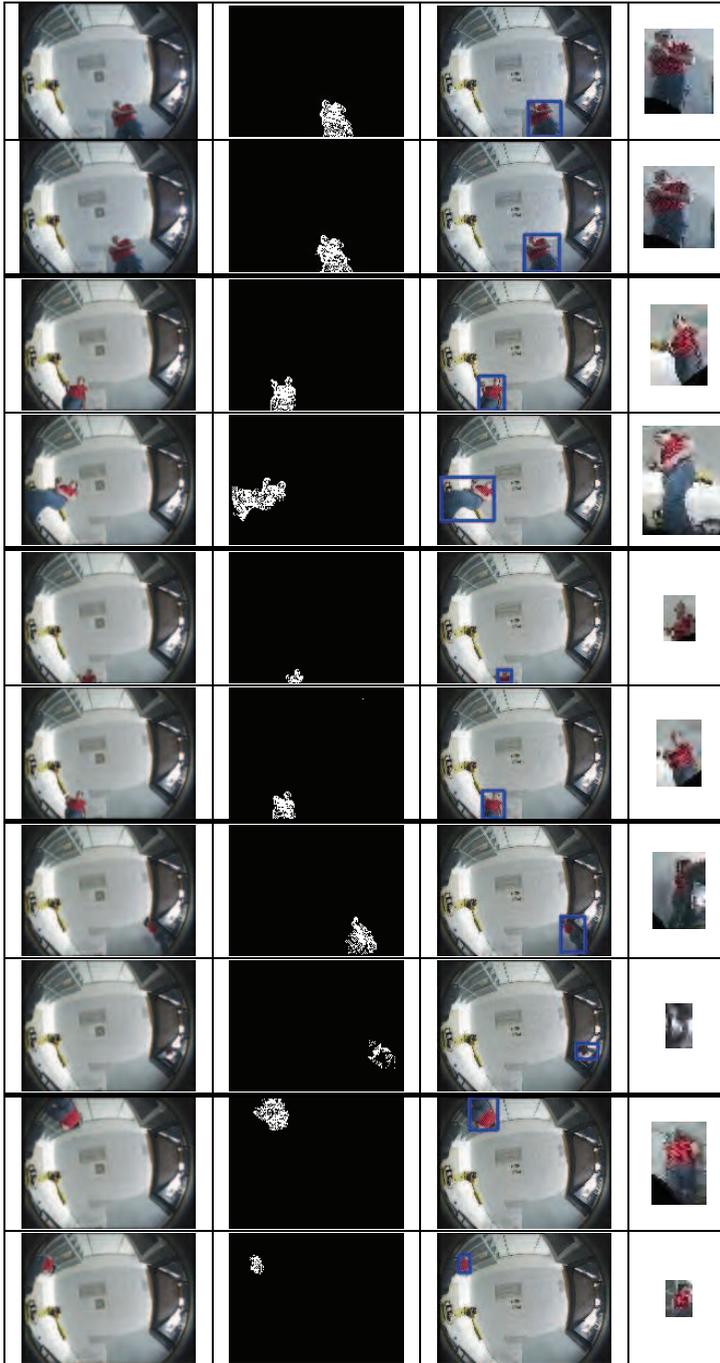


Fig. 3. Some Experimental Results

On the other hand, a new matching process has been proposed. Basically, it obtains a panoramic cylindrical image for each detected target from an image in which identified targets were enclosed in the minimum bounding rectangle. In this way, the rotation problem has disappeared. Next, each panoramic cylindrical image is compared with all the ones obtained in the previous frame whenever they are in its proximity. As all previous panoramic cylindrical images are used, the translation problem is eliminated. Finally, a similarity likelihood criterion is used for matching target images at two different times. With this method the occlusion problem is also solved. As time consumption is a critical issue in robotics, when a dioptric system is used, this processing is performed in a parallel way such that correspondence between two different panoramic cylindrical images of the same target taken at the same time for both cameras is established by means of a disparity map. Therefore, a complete real-time surveillance system has been presented.

## 6. Acknowledges

This research was partly supported by WCU (World Class University) program through the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology (Grant No. R31-2008-000-10062-0), by the European Commission's Seventh Framework Programme FP7/2007-2013 under grant agreement 217077 (EYESHOTS project), by Ministerio de Ciencia e Innovación (DPI-2008-06636), by Generalitat Valenciana (PROMETEO/2009/052) and by Fundació Caixa Castelló-Bancaixa (P1-1B2008-51).

## 7. References

- Baker, S. & Nayar, S.K. (1998). A Theory of Catadioptric Image Formation. *Proceedings of International Conference on Computer Vision (ICCV)*, pp. 35–42, 1998, Bombay, India
- Baker, S. & Nayar, S.K. (1999). A Theory of Single-Viewpoint Catadioptric Image Formation. *International Journal on Computer Vision*, Vol. 2, No. 2, (1999), page numbers (175-196)
- Bar-Shalom, Y. and Li X-R (1998). *Estimation and Tracking: Principles, Techniques and Software*, YBS Publishing, ISBN: 096483121X
- Cervera, E.; Garcia-Aracil, N.; Martínez, E.; Nomdedeu, L. & del Pobil, A. Safety for a Robot Arm Moving Amidst Humans by Using Panoramic Vision. *Proceedings of IEEE International Conference on Robotics & Automation (ICRA)*, pp. 2183–2188, ISBN: 978-4244-1646-2, May 2008, Pasadena, California
- Cielniak, G.; Miladinovic, M.; Hammarin, D. & Göranson, L. (2003). Appearance-based Tracking of Persons with an Omnidirectional Vision Sensor. *Proceedings of the 4<sup>th</sup> IEEE Workshop on Omnidirectional Vision (OmniVis)*, 2003, Madison, Wisconsin, USA
- Fiala, M. (2005). Structure from Motion using SIFT Features and the PH Transform with Panoramic Imagery. *Proceedings of the 2<sup>nd</sup> Canadian Conference on Computer and Robot Vision (CRV)*, pp. 506–513, 2005
- Geyer, C. & Daniilidis, K. (2000). A Unifying Theory for Central Panoramic Systems and Practical Applications. *Proceedings of 6<sup>th</sup> European Conference on Computer Vision (ECCV)*, pp. 445–461, ISBN: 3-540-67686-4, June 2000, Dublin, Ireland
- Grewal, M., Andrews, S. and Angus, P. (2008) *Kalman Filtering Theory and Practice Using MATLAB*, Wiley-IEEE Press, ISBN: 978-0-470-17366-4

- Haykin, S. (2001). *Kalman Filtering and Neural Networks*. John Wiley & Sons, Inc. ISBN: 978-0-471-36998-1
- Kalman, R.E. and Bucy, R.S. (1961). New Results in Linear Filtering and Prediction Theory. *Journal of Basic Engineering* (1961)
- Liu, H.; Wenkai, P.I. & Zha, H. (2003) Motion Detection for Multiple Moving Targets by Using an Omnidirectional Camera. *Proceedings of IEEE International Conference on Robotics, Intelligent Systems and Signal Processing*, pp. 422–426, October 2003, Chansgsha, China
- Lowe, D.G. (1999). Object recognition from local scale-invariant features, *Proceedings of International Conference on Computer Vision*, pp. 1150–1157, September 1999, Corfu, Greece
- Lowe, D.G. (2004). Distinctive image features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, Vol. 60, No. 2, (2004), page numbers (91–110)
- Mauthner, T.; Fraundorfer, F. & Bischof, H. (2006). Region Matching for Omnidirectional Images Using Virtual Camera Planes. *Proceedings of Computer Vision Winter Workshop*, February 2006, Czech Republic
- Murillo, A.C.; Sagüés, C.; Guerrero, J.J.; Goedemé, T.; Tuytelaars, T. and Van Gool, L. (2006). From Omnidirectional Images to Hierarchical Localization. *Robotics and Autonomous Systems*, Vol. 55, No. 5, (December 2006), page numbers (372 – 382)
- Potúček, I. (2003). Person Tracking using Omnidirectional View. *Proceedings of the 9<sup>th</sup> International Student EEICT Conference and Competition*, pp. 603–607, ISBN: 80-214-2370, 2003, Brno, Czech Republic
- Puig, L.; Guerrero, J. & Sturm, P. (2008). Matching of Omnidirectional and Perspective Images using the Hybrid Fundamental Matrix. *Proceedings of the Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras*, October 2008, Marseille, France
- Svoboda, T.; Pajdla, T. & Hlavác, V. (1998). Epipolar Geometry for Panoramic Cameras. *Proceedings of the 5<sup>th</sup> European Conference on Computer Vision (ECCV)*, pp. 218–231, ISBN: 3-540-64569-1, June 1998, Freiburg, Germany
- Tamimi, H.; Andreasson, H.; Treptow, A.; Duckett, T. & Zell, A. (2006). Localization of Mobile Robots with Omnidirectional Vision using Particle Filter and Iterative SIFT. *Robotics and Autonomous Systems*, Vol. 54 (2006), page numbers (758–765)
- Wei, S.C.; Yagi, Y. & Yachida, M. (1998). Building Local Floor Map by Use of Ultrasonic and Omni-directional Vision Sensor. *Proceedings of IEEE International Conference on Robotics & Automation (ICRA)*, pp. 2548–2553, ISBN: 0-7803-4300-X, May 1998, Leuven, Belgium
- Wood, R.W. (1906). Fish-eye Views. *Philosophical Magazine*, Vol. 12, No. 6, (1906), page numbers (159–162)
- Zarchan, P. and Mussof, H. (2005) *Fundamentals of Kalman Filtering: A Practical Approach*, American Institute of Aeronautics & Ast (AIAA), ISBN: 978-1563474552
- Zhu, Z.; Karuppiyah, D.R.; Riseman, E.M. & Hanson, A.R. (2004). Keeping Smart, Omnidirectional Eyes on You [Adaptive Panoramic StereoVision]. *IEEE Robotics and Automation Magazine – Special Issue on Panoramic Robotics*, Vol. 11, No. 4, (December 2004), page numbers (69–78), ISSN: 1070-9932



## **Robot Vision**

Edited by Ales Ude

ISBN 978-953-307-077-3

Hard cover, 614 pages

**Publisher** InTech

**Published online** 01, March, 2010

**Published in print edition** March, 2010

The purpose of robot vision is to enable robots to perceive the external world in order to perform a large range of tasks such as navigation, visual servoing for object tracking and manipulation, object recognition and categorization, surveillance, and higher-level decision-making. Among different perceptual modalities, vision is arguably the most important one. It is therefore an essential building block of a cognitive robot. This book presents a snapshot of the wide variety of work in robot vision that is currently going on in different parts of the world.

### **How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

E. Martinez and A.P. del Pobil (2010). Methods for Reliable Robot Vision with a Dioptric System, Robot Vision, Ales Ude (Ed.), ISBN: 978-953-307-077-3, InTech, Available from: <http://www.intechopen.com/books/robot-vision/methods-for-reliable-robot-vision-with-a-dioptic-system>

# **INTECH**

open science | open minds

### **InTech Europe**

University Campus STeP Ri  
Slavka Krautzeka 83/A  
51000 Rijeka, Croatia  
Phone: +385 (51) 770 447  
Fax: +385 (51) 686 166  
[www.intechopen.com](http://www.intechopen.com)

### **InTech China**

Unit 405, Office Block, Hotel Equatorial Shanghai  
No.65, Yan An Road (West), Shanghai, 200040, China  
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元  
Phone: +86-21-62489820  
Fax: +86-21-62489821

© 2010 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](#), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.