A DIRECT APPROACH FOR HUMAN DETECTION WITH CATADIOPTRIC OMNIDIRECTIONAL CAMERAS

Ibrahim Cinaroglu, Yalin Bastanlar

Computer Engineering Department
İzmir Institute of Technology
İzmir, Turkey
{ibrahimcinaroglu,yalinbastanlar}@iyte.edu.tr

ABSTRACT

This paper presents an omnidirectional vision based solution to detect human beings. We first go through the conventional sliding window approaches for human detection. Then, we describe how the feature extraction step of the conventional approaches should be modified for a theoretically correct and effective use in omnidirectional cameras. In this way we perform human detection directly on the omnidirectional images without converting them to panoramic or perspective image. Our experiments, both with synthetic and real images show that the proposed approach produces successful results.

Keywords— Omnidirectional cameras; object detection; pedestrian detection; human detection

1. INTRODUCTION

Detecting people with cameras is an important task for many research and application areas such as visual surveillance, ambient intelligence and pedestrian safety. Last decade has witnessed significant advances in human detection both in terms of effectiveness and processing time.

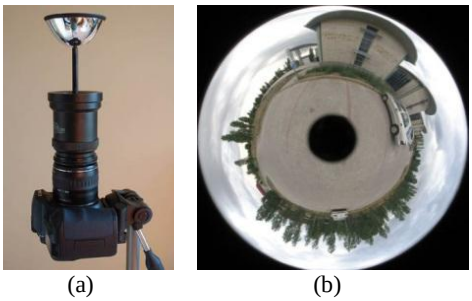
Quite a variety of approaches have been proposed for pedestrian detection and in general for human detection. A major group in these studies uses the sliding window approach in which the detection task is performed via a moving and gradually growing search window. A significant performance improvement was obtained with this approach by employing HOG (Histogram of Oriented Gradients) features. Inspired by SIFT (Scale Invariant Feature Transform) [1], Dalal and Triggs [2] proposed to use HOG for the feature extraction step and they used SVM (Support Vector Machines) for the classification step. Later on, this technique was enhanced with part based models. For instance, Felzenswalb et al. [3] proposed a method using parts of the object which are spring-like connected to each other and can move independently. Another notable enhancement was using pyramid HOG features and Intersection Kernel SVM proposed by Maji et al. [4].

Edge based features [5] and 'shapelets' [6] are examples of other features which were used for human detection. More recently, it was shown that using combinations of features outperforms the approaches that use a single type of feature [7]. For a detailed summary and comparison of methods we refer readers to [8], where an extensive evaluation of the above mentioned and many other pedestrian detection algorithms exists.

Omnidirectional cameras provide 360° horizontal field of view in a single image (vertical field of view varies). If a convex mirror is placed in front of a conventional camera for this purpose, then the imaging system is called a catadioptric omnidirectional camera (Fig. 1). With its enlarged view advantage, fewer omnidirectional cameras may substitute many perspective cameras. However, so far omnidirectional cameras have not been widely used in object detection research area and also in traffic applications like pedestrian and vehicle detection.

In a study on object recognition with omnidirectional cameras [9], a mobile robot is given the images of several objects in the environment and it is asked to recognize these objects. Actually, the omnidirectional image is warped into a cylindrical panoramic image before matching with the images of the objects. SIFT matching is employed without any modification for omnidirectional cameras. In [10], authors use Haar features to perform face detection with catadioptric omnidirectional cameras. Instead of modifying the feature extraction step, they convert the omnidirectional images into panoramic images and directly use the conventional (perspective) camera technique. In a similar manner, panoramic images are used in [11] for human detection.

A human tracking method for omnidirectional cameras is proposed in [12]. As a part of the proposed algorithm, HOG features are computed. However, a rectangular rotating and sliding window is used with no mathematical modification for the omnidirectional camera.



(a) (b)

Fig. 1. (a) A mirror apparatus is placed in front of a conventional camera to obtain a catadioptric omnidirectional camera. (b) An example image obtained by such a camera.

In this paper, we propose a direct approach to tackle human detection on catadioptric omnidirectional images. That is, we do not convert the omnidirectional images to panoramic or perspective images. To our knowledge, the proposed method is the first one to detect humans directly on omnidirectional images. In Section 2, we explain why our approach is theoretically correct. We adopt HOG+SVM [2] approach for human detection and as explained in Section 3, we modify the HOG feature extraction step for catadioptric omnidirectional cameras. With experiments, given in Section 4, we demonstrate that the adaptation of HOG features improves the performance significantly.

2. PROCESSING OF OMNIDIRECTIONAL IMAGES

Due to their nonlinear imaging geometry, working with omnidirectional cameras requires geometric transformations. At first sight, converting an omnidirectional image to a panoramic or several perspective images may seem to be a practical solution. However, it has two major drawbacks: The conversion can be computationally expensive for large frames especially when an omnidirectional image is converted to numerous perspective images to properly fit sliding windows. More importantly, the interpolation required by the image warping introduces artifacts that affect the detection performance.

Among a small number of omnidirectional object detection studies (cf. Section 1), none of them developed a method peculiar to omnidirectional cameras. On the other hand, last decade witnessed some effort on computing SIFT features in omnidirectional images. These studies consider the convolution step of SIFT and avoid warping omnidirectional images. Below, we describe these approaches and summarize their properties.

- The simplest approach would be backprojecting the image onto a sphere surface S^2 and convolving it with a spherical Gaussian function GS [14]. Since this approach requires resampling of the whole image, authors in [13] project the kernel GS into image plane instead of backprojecting the image onto S^2, and the convolution is carried directly in the image plane. This avoids image resampling but since the mapped Gaussian kernel changes at every image location it leads to an adaptive filtering. Such complexity makes the solution unsuitable.
- Another approach processes omnidirectional images on the sphere after an inverse stereographic projection [15]. Scale space is computed with Gaussian kernels on the sphere, while, the convolution is performed using the spherical Fourier transform. It was stated in [16] and [17] that this operation leads to aliasing issues due to bandwidth limitations.
- The processing on the sphere is achieved through a suitable differential operator that adapts to the non-uniform resolution, while using the original image pixel values. In [18], scale space representation is computed using the heat diffusion equation and differential operators (Laplace–Beltrami operators) on the non-Euclidean (Riemannian) manifolds. Moreover, authors in [16] tested this approach by evaluating the matching performance of SIFT on rotated and translated images. Lastly, authors in [19] compared the original SIFT with the version modified by Laplace–Beltrami operators on the Riemannian manifolds and mentioned that the modified version has a better performance. They also extend the approach to all central catadioptric systems.

Exploiting the experience gained by the summarized previous work, we decided to compute the gradients on Riemannian manifolds and adapted the HOG computation step (Section 3.1) of our algorithm accordingly.

3. THE PROPOSED HOG COMPUTATION

To detect the standing people in omnidirectional images, we rotate the rectangular sliding window around the image center. In addition, to achieve a mathematically correct detection method, we modify the image gradients. The operations that we perform can be divided into two steps:

1. Modification of gradient magnitudes using Reimannian metric.
2. Conversion of gradient orientations to form an omnidirectional (non-rectangular) sliding window.

3.1. Modification of Gradient Magnitudes Using Reimannian Metric

3.1.1. Sphere camera model

We use the sphere camera model [20] which was introduced to model central catadioptric cameras. The model comprises a unit sphere and a perspective camera. The projection of 3D points can be performed in two steps (Fig. 2). The first one is the projection of point Q in 3D space onto a unitary sphere, resulting in point r, and the second one is a perspective projection from the sphere to the image plane, resulting in point q. This model covers all central catadioptric cameras with varying ξ. $\xi = 0$ for perspective cameras, $\xi = 1$ for para-catadioptric cameras (the ones using a paraboloidal mirror), $0 < \xi < 1$ for hyper-catadioptric cameras (the ones using a hyperboloidal mirror).

A point on the sphere $\mathbf{r} = (X, Y, Z)$ can also be represented by two angles (θ, φ), the former is the vertical angle and the latter is the azimuth (Fig. 3a). In para-catadioptric case ($\xi = 1$), if we place the image plane at the south pole (which only differs the scale), $f = 2r = 2$ and the perspective projection within the sphere model corresponds to the stereographic projection (Fig. 3b).

3.1.2. Differential operators on Riemannian manifolds

Let us briefly describe how the differential operators on the Riemannian manifolds are defined. Suppose M denotes a parametric surface on $\Re^3$ and g_{ij} denotes the Riemannian metric that encodes the geometrical properties of the manifold. In a local system of coordinates x^i on M, the components of the gradient are given by

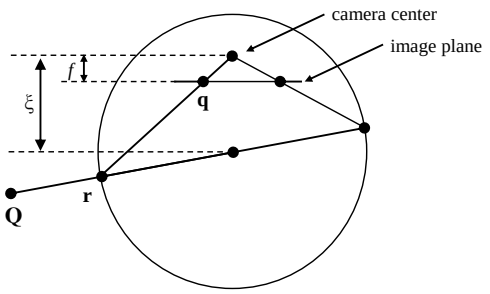


Fig. 2. Projection of a 3D point onto the image plane in the sphere camera model.

$$\nabla^{\mathrm{i}} = g^{ij}\frac{\partial}{\partial x^j} \tag{1}$$

where g^{ij} is the inverse of g_{ij}.

A similar reasoning is used in [16] and [19] to obtain the Laplace-Beltrami operator, which is the second order differential operator defined on M and used for scale space representation for SIFT. In this paper, we are working on the first derivatives. Let us briefly go over the para-catadioptric case.

Consider the unitary sphere S^2 with radius = 1 (Fig. 3a). A point on S^2 is represented in Cartesian and polar coordinates as

$$(X, Y, Z) = (\sin\theta\sin\varphi\,, \sin\theta\cos\varphi\,, \cos\theta) \tag{2}$$

The Euclidean line element in Cartesian coordinates, dl, can be expressed in polar coordinates as

$$dl^2 = dX^2 + dY^2 + dZ^2 = d\theta^2 + sin^2\theta\, d\varphi^2 \tag{3}$$

The stereographic projection of the sphere model sends a point on the sphere (θ, φ) to a point in polar coordinates (R, φ) in the image plane (plane $\Re^2$), for which φ remains the same and $\theta = 2\tan^{-1}(R/2)$ in a para-catadioptric system (Fig. 3b).

Using the identities, $\mathrm{R}^2 = \mathrm{x}^2 + \mathrm{y}^2$, $\varphi = \tan^{-1}(\mathrm{y/x})$ the line element reads

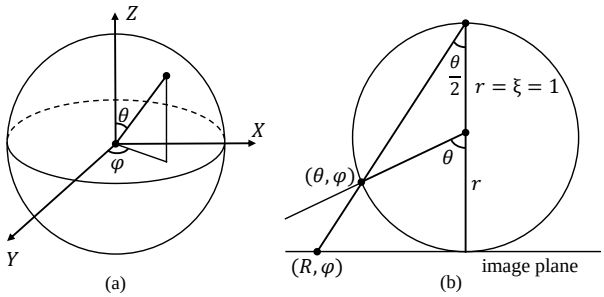


Fig. 3. (a) A 3D point on the sphere is represented by two angles (θ, φ). (b) Consider the unitary sphere $(r = 1)$. Image plane is placed at the south pole $(f = 2)$. A 3D point is first projected onto the sphere surface and then projected onto the image plane, where in this case $\xi = 1$.

$$dl^2 = \frac{16}{(4+x^2+y^2)^2}(dx^2 + dy^2) \tag{4}$$

giving the Riemannian inverse metric

$$g^{ij} = \frac{(4+x^2+y^2)^2}{16} \tag{3}$$

We refer the reader to [16] and [18] for a detailed derivation of catadioptric Reimannian metric. With this metric, we can compute the differential operators on the sphere using the pixels in the omnidirectional images. In particular, norm of the gradient reads

$$|\nabla_{S^2} I|^2 = \frac{(4+x^2+y^2)^2}{16}|\nabla_{\mathbb{R}^2} I|^2 \tag{4}$$

We see that the para-catadioptric gradients are just the scaled versions of the gradients in Euclidean domain. Therefore, we simply multiply our gradients with metric g^{ij}.

At the center of the omnidirectional image, $(x, y) = (0,0)$, Reimannian and Euclidean gradients are the same. At an image location when $\sqrt{x^2 + y^2} = 2$, which corresponds to a 3D point at the same horizontal level with the sphere center (mirror focal point), the Reimannian metric is equal to 4. Therefore the gradients are magnified as we move from the center to the periphery of the omnidirectional image. This metric is extended to all central catadioptric systems by Puig et al. [19].

3.2. Conversion of Gradients for Omnidirectional Sliding Window

After the image gradients are obtained with Reimannian metric, we convert the gradient orientations to form an omnidirectional (non-rectangular) sliding window. A rectangular object in a perspective image is warped in the omnidirectional image, therefore the gradients in the sliding window should be computed as if a perspective camera is looking from the same viewpoint.

The reader should note that we train our model for human detection using INRIA perspective image dataset as described in [2], i.e. we do not train an omnidirectional HOG model. Since the shape of the non-rectangular sliding window varies according to the location in the omnidirectional image, it is not plausible to train many omnidirectional HOG models. The modifications we made for HOG computation in omnidirectional sliding window enables us to compare it with the perspective camera HOG model. Fig. 4a shows a half of a synthetic para-catadioptric omnidirectional image (400x400 pixels) where the walls of a room are covered with rectangular black and white tiles. Conventional HOG result of the marked region (128x196 pixels) in this image is given in Fig. 4b where gradient orientations are in accordance with the image. However, since these are vertical and horizontal edges in real world, we need to obtain vertical and horizontal gradients. Fig. 4d shows converted gradients for the region marked in Fig. 4c,

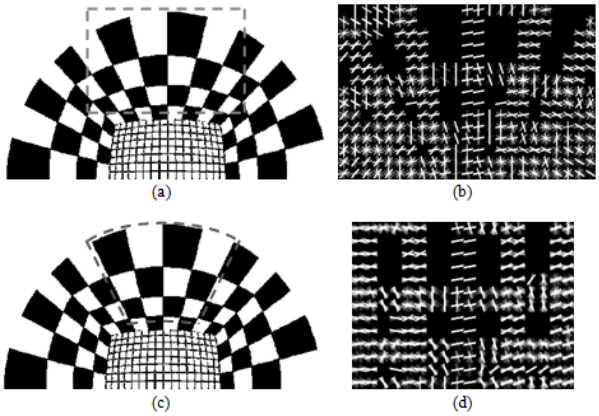


Fig. 4. Description of how the gradients are modified for an omnidirectional sliding window. Result in (b) is the regular HOG computed for the region marked with dashed lines in (a). Modified HOG computation gives the result in (d) for the region marked in (c). Vertical and horizontal edges in real world produce vertical and horizontal gradients in the modified version.

which is an example of the proposed HOG computation.

Since the Cartesian coordinates in the detection window (Fig. 4d) corresponds to a nonlinear distribution of pixels in the image (Fig. 4c), we employ bilinear interpolation with backward mapping both for gradient orientations and gradient magnitudes.

4. EXPERIMENTS

4.1. Evaluation of the Modified HOG Using SVM Scores

Let us first compare the results of the proposed HOG computation and the regular HOG computation on the omnidirectional images. Since the computed HOG features are given to an SVM trained with person image dataset, we aim to obtain higher SVM scores with the proposed omnidirectional HOG computation.

We artificially created 210 omnidirectional images containing humans. While creating this set, we followed an approach similar to [15], where images in INRIA person dataset are projected to omnidirectional images using certain projection angle and distance parameters. Fig. 5 shows an example omnidirectional image, where the regular HOG window (rectangular, 128x64 pixels) and the proposed omnidirectional HOG window (non-rectangular) are shown. The HOG features computed with the two window types are compared with their resultant SVM scores. Since the locations of projections in these images are known, no search is needed for this experiment. However, vertical position of the window affects the result. For both approaches, we chose the position that gives the highest mean SVM score. Table 1 summarizes the result of the comparison, where we see that the mean score for the proposed approach is higher than that of regular HOG window.

4.2. Experiments with Real Images

In this subsection, we present the results for a set of images

TABLE I. COMPARISON OF THE REGULAR AND PROPOSED HOG WINDOW BY THEIR SVM SCORES

	Mean SVM Score	Minimum SVM Score	Maximum SVM Score
Regular HOG window	1.69	-1.01	3.21
Proposed HOG window	1.93	-0.42	3.64

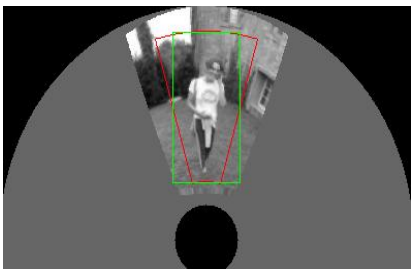


Fig. 5. Depiction of the regular HOG window (green rectangle) and the proposed window (red doughnut slice) on an omnidirectional image artificially created by projecting a perspective image from INRIA person dataset.

taken with our catadioptric omnidirectional camera. We compared the proposed HOG computation not only with the regular HOG window, but also with the approach that first converts the omnidirectional image to a panoramic image and then performs HOG computation. Although it was explained in Section 2 that working on panoramic images is not a theoretically correct approach, we wanted to test its performance. Fig. 6 shows the results for one of the images in the set. SVM scores greater than 1, after non-maximum suppression, superimposed on the images with the proposed HOG window, the regular HOG window on omnidirectional image and HOG after panoramic conversion. For the humans in the scene, the average SVM scores for the proposed HOG,

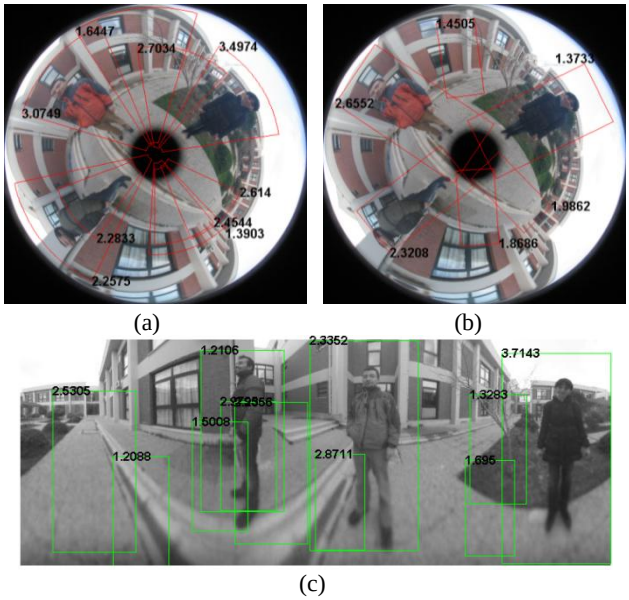


Fig. 6. Human detection results on an omnidirectional image with SVM scores (upper left corners) greater than 1. (a) Proposed sliding windows. (b) Regular (rectangular) sliding and rotating windows. (c) Regular sliding windows on panoramic image.

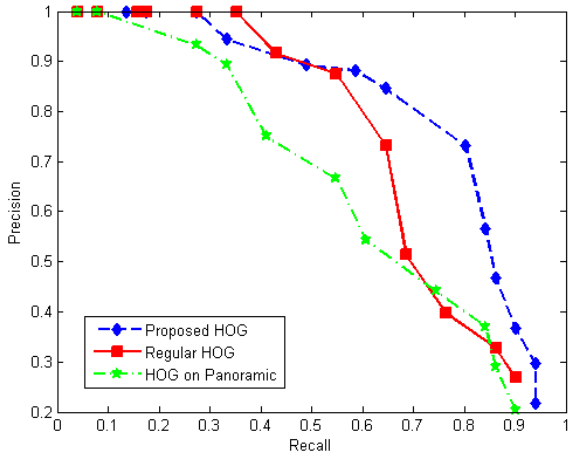


Fig. 7. Precision-Recall curves to compare the proposed HOG computation, the regular HOG and HOG after panoramic conversion. The data points in this curve correspond to the varying threshold values for the SVM score, which change from 0 to 5.

the regular HOG and HOG on panoramic image approaches are 2.94, 2.11 and 2.41 respectively.

To evaluate the overall performance of these three approaches, we plot precision-recall curves for the whole dataset which consists of 20 real images taken in different scenes including indoor and outdoor environments (Fig. 7). The aim of these curves is to show the two performance metrics together: Precision (#True positives / #Predicted positives) and Recall (#True positives / #Actual positives). The larger the area under the curve, the better the performance of the algorithm. As the threshold increases, all approaches reach Precision=1. One can observe that the performance of the proposed HOG computation is better than the others up to a threshold value of 4.0. At higher threshold values our method loses its advantage. However, since the recall comes below 0.5 for those values, it is not plausible to use them in practical systems.

A detection window is considered to be a True-positive if it overlaps an annotation by 50%, where the overlap is computed as

$$O = \frac{area(detection\ window \cap annotation)}{area(detection\ window \cup annotation)} \quad (5)$$

For a fair comparison, the annotations are separately prepared for the mentioned three methods. Annotations of the proposed HOG approach (e.g. Fig. 6a) are doughnut slices, annotations of the regular HOG approach are rectangles rotating around the omnidirectional image center. Finally, annotations of the HOG on panoramic image approach are upright rectangles.

5. CONCLUSION

We aimed to perform human detection directly on the omnidirectional images. As a base, we took the HOG+SVM approach which is one of the popular human detection methods. After describing how the feature extraction step of the conventional method should be modified, we performed experiments to compare the proposed method with the regular HOG computation in omnidirectional and in panoramic images. Results of the experiments indicate a performance increase for the proposed approach.

In the near future, we are planning to prepare a larger set of real omnidirectional images, and perform tests using that set.

6. REFERENCES

[1] Lowe, D., Distinctive image features from scale invariant keypoints, International Journal of Computer Vision (IJCV), 60, 91-110, (2004).

[2] Dalal, N., Triggs, B., Histograms of Oriented Gradients for Human Detection, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (2005).

[3] Felzenszwalb, P., McAllester, D., Ramanan, D., A Discriminatively Trained, Multiscale, Deformable Part Model, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (2008).

[4] Maji, S., Berg, A.C., Malik, J., Classification using Intersection Kernel Support Vector Machines is Efficient, IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), (2008).

[5] Wu, B., Nevatia, R., Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors, IEEE International Conference on Computer Vision (ICCV), (2005).

[6] Sabzmeydani, P., Mori, G., Detecting pedestrians by learning shapelet features, IEEE Conf. Computer Vision and Pattern Recognition, (2007).

[7] Walk, S., Majer, N., Schindler, K., Schiele, B., New Features and Insights for Pedestrian Detection," IEEE Conf. Computer Vision and Pattern Recognition, (2010).

[8] Dollar, P., Wojek, C., Schiele, B., Perona, P., Pedestrian Detection: An Evaluation of the State of the Art, IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(4), 743-761, (2012).

[9] Wang, M.L., Lin, H.Y., Object Recognition from Omnidirectional Visual Sensing for Mobile Robot Applications, IEEE International Conference on Systems, Man and Cybernetics, (2009).

[10] Iraqui, A., Dupuis, Y., Boutteau, R., Ertaud, J., Savatier, X., Fusion of omnidirectional and PTZ cameras for face detection and tracking, International Conference on Emerging Security Technologies, (2010).

[11] Kang, S., Roh, A., Nam, B., Hong, H., People detection method using GPUs for a mobile robot with an omnidirectional camera, Optical Engineering 50(12), 127204, (2011).

[12] Tang, Y., Li, Y., Bai, T., Zhou, X., Human Tracking in Thermal Catadioptric Omnidirectional Vision, International Conference on Information and Automation (ICIA), (2011).

[13] Daniilidis, K., Makadia, A., Bulow, T., Image Processing in Catadioptric Planes: Spatiotemporal Derivatives and Optical Flow Computation, International Workshop on Omnidirectional Vision (OmniVis), (2002).

[14] Bulow, T., Spherical diffusion for 3D surface smoothing, IEEE Trans. Pattern Anal. Mach. Intell. (PAMI), 25, 1650–1654, (2004).

[15] Hansen, P. Corke, P., Boles, W., Daniilidis, K., Scale Invariant Features on the Sphere. IEEE Int. Conference on Computer Vision, (2007).

[16] Arican, Z., Frossard, P., OMNISIFT: Scale Invariant Features in Omnidirectional Images, IEEE Int. Conf. on Image Processing (2010).

[17] Lourenço, M., Barreto, J.P., Vasconcelos, F., sRD-SIFT: Keypoint Detection and Matching in Images with Radial Distortion, IEEE Transactions on Robotics, 28(3), 752-760, (2012).

[18] Bogdanova, I., Bresson, X., Thiran, J.P., Vandergheynst, P., Scale Space Analysis and Active Contours for Omnidirectional Images. IEEE Transactions on Image Processing, 16(7), 1888-1901, (2007).

[19] Puig, L., Guerrero, J. J., Scale Space for Central Catadioptric Systems: Towards a Generic Camera Feature Extractor, International Conference on Computer Vision (ICCV), (2010).

[20] Geyer, C., Daniilidis, K., A unifying theory for central panoramic systems and practical applications, 6th European Conference on Computer Vision, 445–461, (2000).