

THE SPHERICAL APPROACH TO OMNIDIRECTIONAL VISUAL ATTENTION

Iva Bogdanova, Alexandre Bur, and Heinz Hügli

Institute of Microtechnology, University of Neuchâtel
A.-L. Breguet 2, 2000 Neuchâtel, Switzerland
email: {iva.bogdanova; alexandre.bur; heinz.hugli}@unine.ch
web: www2.unine.ch/imt

ABSTRACT

Computational visual attention (VA) has been widely investigated during the last three decades but the conventional algorithms are not suitable for omnidirectional images which often contain a significant amount of radial distortion. Only recently a computational approach was proposed that processes images in the spherical (non-Euclidian) space and produces attention maps with a direction independent homogeneous response. This paper investigates how this spherical approach applies to real scenes and particularly to different omnidirectional visual sensors. Reported experiments refer to omnidirectional images obtained from a multi-camera omnidirectional sensor as well as a parabolic and a hyperbolic catadioptric image sensor.

1. INTRODUCTION

1.1 Visual Attention

The human visual system makes an extensive use of the VA in order to select relevant visual information and speed up the vision process. The saliency-based computer model of VA has been widely investigated in the last three decades [8, 9, 4]. Nowadays, various such models exist as well as numerous soft- and hardware implementations [6, 10]. It is applied in color image segmentation [11], robot guidance [5], object recognition [12]. Unfortunately, in all these applications, the VA is computed in the Euclidean geometry and thus is restricted only to conventional images. One approach to VA in omnidirectional scenes is reported in [5] but the approach used, namely the panoramic representation, is still subject to radial distortions.

A natural choice of a non-distorted domain for the spherical field of view is the sphere $S^2 \in \mathbb{R}^3$. Relying on spherical geometry, in [3] the authors proposed an algorithm that produces attention maps with a direction independent homogeneous response. In this paper, we investigate how this new spherical approach applies to real scenes and particularly to different omnidirectional visual sensors. Reported experiments refer to omnidirectional images obtained from a multi-camera omnidirectional sensor as well as a parabolic and hyperbolic catadioptric image sensor.

1.2 Omnidirectional vision

Any real scene can be described by the ideal *plenoptic function*. This function is associated to the light field through any point in space, at any time, and over any range of wavelength [1]. Let an observer stand at any point in the space (x, y, z) , from which one selects any of the viewable rays by choosing

an azimuth and elevation angle $(\theta_{pl}, \varphi_{pl})$, as well as a band of wavelength λ which one wishes to consider. In addition, if this is a dynamic scene, one can choose the time t , at which the light field to be evaluated. This results in the following form of the plenoptic function: $p = P(\theta_{pl}, \varphi_{pl}, \lambda, x, y, z, t)$. One can easily measure any of the variables but an important case is where $\vec{p} = (x, y, z)$ is fixed and one records $I_{\vec{p}}(\theta, \varphi)$ that is the light incoming on a perfect punctual observer located at \vec{p} , i.e. the full sphere of view. Measuring $I_{\vec{p}}(\theta, \varphi)$ is equivalent to observing the scene in any direction (θ, φ) from a fixed viewpoint \vec{p} . This is achieved by *catadioptric imaging sensors* and the images obtained are *omnidirectional images*.

A catadioptric sensor is a combination of a curved mirror and a conventional camera [2]. The fixed 3-D point at which the catadioptric sensor samples the plenoptic function is known as *effective viewpoint*. It is highly desirable that such an imaging sensor has a single effective viewpoint, i.e. center of projection. Such are the central catadioptric sensors and particular examples are those designed with hyperbolic or parabolic mirror. Two omnidirectional images are shown on Figures 1(a), and (d). Eventhough, both are omnidirectional, they are governed by different (non-Euclidean) geometry inherited from the corresponding mirror they are obtained with: hyperbolic and parabolic, respectively.

Another way of obtaining omnidirectional image is using a multi-camera system [13] (five cameras are configured in a horizontal ring and one is pointing vertically) which collects images of approximately 75% of the full sphere. The cameras are packed tightly together such that the distance between adjacent cameras is kept minimal. The images from all the cameras in the system are stitched together, resulting to an approximative omnidirectional image.

This article is organized as follows. In Section 2 we derive how to map a hyperbolic and parabolic omnidirectional images onto the sphere. Then, in Section 3 we define the visual attention model in spherical geometry and the resulting spots of attention. The experimental results for three different types of omnidirectional images (hyperbolic [15], parabolic [14] and one approximation acquired with a multi-camera system[13]) are presented in Section 4. Finally, we conclude and expose some directions for future investigations.

2. MAPPING OMNIDIRECTIONAL IMAGES ONTO THE SPHERE

It was shown in [7] that there is an equivalence between any central catadioptric projection and a composite mapping through the sphere. This mapping is a composition of central projection to the unit sphere followed by the projection from

The authors acknowledge the support of the Swiss NCCR IM2 and the Swiss National Science Foundation under FN-108060

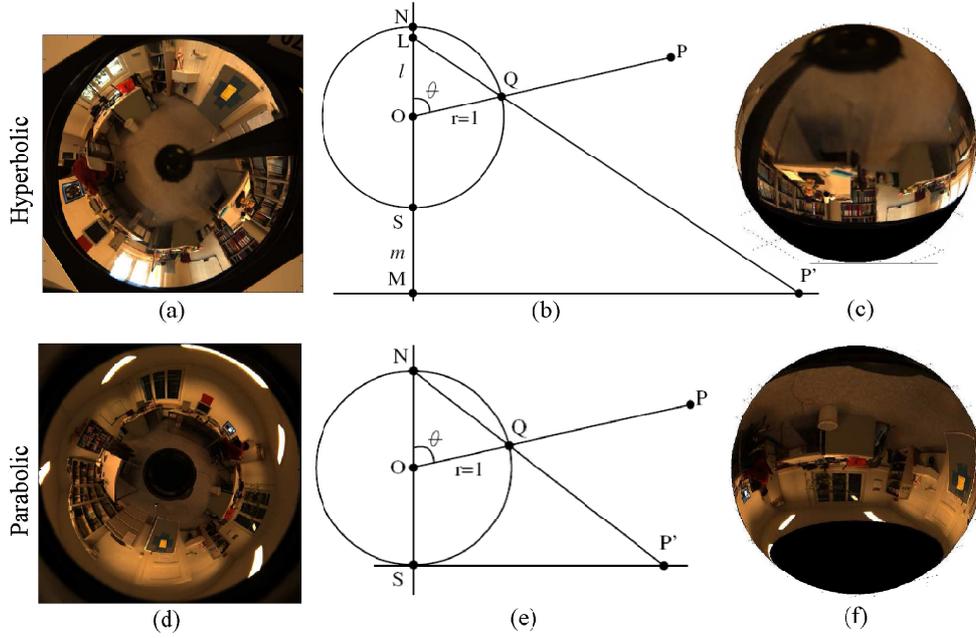


Figure 1: Hyperbolic and parabolic omnidirectional images and their mapping on the sphere.

a point on the some axis of the sphere at distance l from the sphere's center to a plane perpendicular to the axis at distance m below the center. The position of this point of projection depends on the mirror's shape.

2.1 Hyperbolic projection

According to [7], the hyperbolic projection is equivalent to the composition of normalization to the unit sphere followed by a projection from point L , as illustrated on Figure 1(b). Consequently, the mapping of a hyperbolic image onto the sphere is a projection defined in spherical coordinates

$$s_{hyp} \equiv \frac{(l+m)\sin\theta}{l-\cos\theta}(\sin\varphi, \cos\varphi), \quad (1)$$

where $\theta \in [0, \pi]$ and $\varphi \in [0, 2\pi)$. In other words, a point on the sphere Q is transformed into a point on the catadioptric sensor plane P' using Equation (1). The projection parameters l and m depend on the mirror and are obtained as:

$$l = \frac{2\varepsilon}{l+\varepsilon^2}, \quad m = \frac{2\varepsilon(b-1)}{l+\varepsilon^2}, \quad (2)$$

where the eccentricity ε of the hyperbolic mirror, with semi-axis a and b , is $\varepsilon = \frac{\sqrt{a^2+b^2}}{a}$. The points lying above the projection point L are those that would be reflected by the lower sheet of the two-sheeted hyperboloid. Actually, the lower sheet is not used in the design of the hyperbolic catadioptric sensor because this is where the camera is placed. On Figure 1(c) is shown the omnidirectional hyperbolic image mapped on the sphere.

2.2 Parabolic projection

The parabolic projection is equivalent to the composition of normalization to the unit sphere followed by stereographic

projection as depicted on Figure 1(e). In spherical coordinates this reads:

$$s_{par} \equiv 2 \cot \frac{\theta}{2} (\sin\varphi, \cos\varphi), \quad (3)$$

where $\theta \in [0, \pi]$, $\varphi \in [0, 2\pi)$. Using this projection we can easily map any omnidirectional image obtained with parabolic catadioptric sensor onto the sphere. One particular example is shown on Figure 1(f).

3. VISUAL ATTENTION IN SPHERICAL GEOMETRY

Inspired by the classical VA model operating on Euclidean (conventional) images, the computational VA model defined in [3] operates in spherical geometry and is thus suitable for omnidirectional images that can be mapped on the sphere. In a first step, it consists of extracting seven specific features j from a spherical image, one intensity, two chromatic (yellow-blue and red-green color opponency) and, for instance, four orientation features ($0^0, 45^0, 90^0$ and 135^0). In a second step, each feature is transformed into its conspicuity map using center-surround mechanism operating in spherical geometry, which highlights the parts of the scene that strongly differ (according to a specific feature) from their surroundings. In a third step, the conspicuity maps of the same nature are fused into three conspicuity cues namely intensity, chromatic and orientation cues. Finally, the conspicuity cues are fused in a competitive way into the *spherical saliency map* which highlight the most informative parts of the scene.

We briefly remind here the computation of the spherical conspicuity feature map. It is based on the center-surround mechanism computed from the Spherical Gaussian Pyramid (SGP). Let us start with a spherical image feature defined with the function $f(\theta, \varphi) \in L^2(S^2)$, $\theta \in [0, \pi]$, $\varphi \in [0, 2\pi)$. Its discrete version is defined on an equi-angular grid of size

$2^n \times 2^n$. For building the SGP, we use an axisymmetric spherical filter

$$\hat{g}_{\sigma_k}(l) = e^{-(\sigma_k l)^2}, \quad (4)$$

where the bandwidth parameter σ_k is chosen so that $|\hat{g}_{\sigma_k}(l)| \ll 1$. The filtering is performed in the Fourier domain and in the sense of spherical harmonics:

$$\widehat{g_{\sigma_k} \star f}(l, m) = \sqrt{\frac{4\pi}{2l+1}} \hat{g}_{\sigma_k}(l, 0) \hat{f}(l, m). \quad (5)$$

The next level, f_{k+1} , in the SGP is obtained by simply down-sampling the filtered signal, $g_{\sigma_k} \star f$, by a factor of 2. In this way, iteratively are obtained all the levels $f_k, k = 1 \dots n$ in SGP. We compute $p \equiv (n-3)$ intermediate spherical conspicuity maps $\mathcal{M}_p(\theta, \varphi)$ as follows:

$$\begin{aligned} \mathcal{M}_1(\theta, \varphi) &= |f_1(\theta, \varphi) \ominus f_4(\theta, \varphi)|, \\ \mathcal{M}_2(\theta, \varphi) &= |f_2(\theta, \varphi) \ominus f_5(\theta, \varphi)|, \\ &\dots, \\ \mathcal{M}_{n-3}(\theta, \varphi) &= |f_{n-3}(\theta, \varphi) \ominus f_n(\theta, \varphi)|, \end{aligned}$$

where \ominus refers to a cross-scale difference operator that interpolates the coarser scale to the finer one and then performs a point-by-point subtraction. The corresponding feature spherical conspicuity map is obtained by combining the set of all multiscale maps $\mathcal{M}_p(\theta, \varphi)$:

$$\mathcal{C}_j(\theta, \varphi) = \sum_{p=1}^{n-3} \mathcal{N}(\mathcal{M}_p(\theta, \varphi)), \quad (6)$$

where $\mathcal{N}(\cdot)$ refers to the non-linear spherical normalization function used in the map integration process, simulating intra-map and inter-map competition.

In order to obtain the *spherical saliency map* $\mathcal{S}_{S^2}(\theta, \varphi)$, we compute the conspicuity maps for the seven different features, then combine the features of the same nature into three conspicuity cues (namely: intensity, chromatic and orientation) according to the following equations:

$$\mathcal{C}_{int}(\theta, \varphi) = \mathcal{C}_1, \quad \mathcal{C}_{chrom}(\theta, \varphi) = \sum_{j \in \{2,3\}} \mathcal{N}(\mathcal{C}_j), \quad (7)$$

$$\mathcal{C}_{orient}(\theta, \varphi) = \sum_{j \in \{4,5,6,7\}} \mathcal{N}(\mathcal{C}_j) \quad (8)$$

and finally all the cues are combined using the same normalization function $\mathcal{N}(\cdot)$ into the saliency map:

$$\mathcal{S}_{S^2}(\theta, \varphi) = \sum_{cues} \mathcal{N}(\mathcal{C}_{cue}(\theta, \varphi)), \quad (9)$$

Based on the spherical saliency map, we can determine the spots of attention using a "winner-take-all" mechanism [8] which detects successively the most salient spots by iteratively applying maximum detection followed by local inhibition at the maximum location.

4. EXPERIMENTAL RESULTS

In this section, we first compare the results of our approach described in Section 3 with the Euclidean approach. Furthermore, we apply it on hyperbolic and parabolic omnidirectional images after they have been mapped on the sphere.

We work with omnidirectional images defined on a spherical grid of size 1024 by 1024. This permits to build $n = 8$ level SGP, and consequently $p = 5$ intermediate levels for computing the conspicuity map. In all of the following examples, we use three features j : intensity, blue-yellow and red-green, resulting into two cues, i.e. intensity and chromatic cues.

4.1 Euclidean vs. Spherical VA

We start with an image obtained by a multicamera [13] with a field of view of $\theta \in [0, 3\pi/4]$, $\varphi \in [0, 2\pi)$. The camera is fixed in the center of the ceiling of a meeting room and points down the table where a red object is placed. The original image in its unwrapped version is shown on Figure 2(b). On Figure 2(a) is shown the same image but on the sphere, where the red object is on the South pole. The saliency map calculated using the spherical approach is illustrated in Figure 2(d) and 2(e), respectively on the sphere and unwrapped. The saliency map obtained by applying the Euclidean visual attention is shown in Figure 2(f). On Figure 2 (g), (h), (i) are shown the spots of attention, where each of the spots is represented with a circle and a number corresponding to the ranking of saliency. The red object, which is expected to correspond to the most salient spot of the scene, is detected accurately only in the spherical geometry (detected as first spot) (Figure 2 (g), (h)), while this is not the case if the Euclidean visual attention were applied (Figure 2 (i)). This situation clearly illustrates the inconvenience of performing Euclidean visual attention on omnidirectional images. Actually, any salient object located in the areas of the sphere's poles can not be precisely detected by the Euclidean visual attention. Eventhough, salient objects located at the equator, are detected in both cases.

4.2 VA in parabolic omnidirectional images

The parabolic omnidirectional images [14] cover $\theta \in [40^\circ, 140^\circ]$. We use the image from Figure 1(d). The spots of attention are determined as described in Section 3 and the results are shown on Figure 3. For purpose of visualization only, we show the unwrapped version of the spherical image (Figure 3(d)). The corresponding saliency map (Figure 3(e)) is computed using Equation 9 as once again only the intensity and chromatic cues are considered. The five detected spots of attention are shown with their corresponding ranking on Figure 3(f).

4.3 VA in hyperbolic omnidirectional images

The hyperbolic image (Figure 1(a)) is obtained through a hyperbolic mirror [15]. It covers $\theta \in [0, 106^\circ]$ from the visual sphere. The results of application of VA algorithm after it has been mapped on the sphere (Equation 1) are depicted on Figure 3 at the top. The unwrapped saliency map is shown in Figure 3(b) and the corresponding spots of attention as defined in Figure 3(c).

5. CONCLUSIONS AND FUTURE WORK

In this paper, we considered the new spherical approach for computing the VA of real omnidirectional images and apply it to omnidirectional images obtained from various sensors. Three types of sensors were considered, namely a multicamera omnidirectional sensor, and two catadioptric sensors

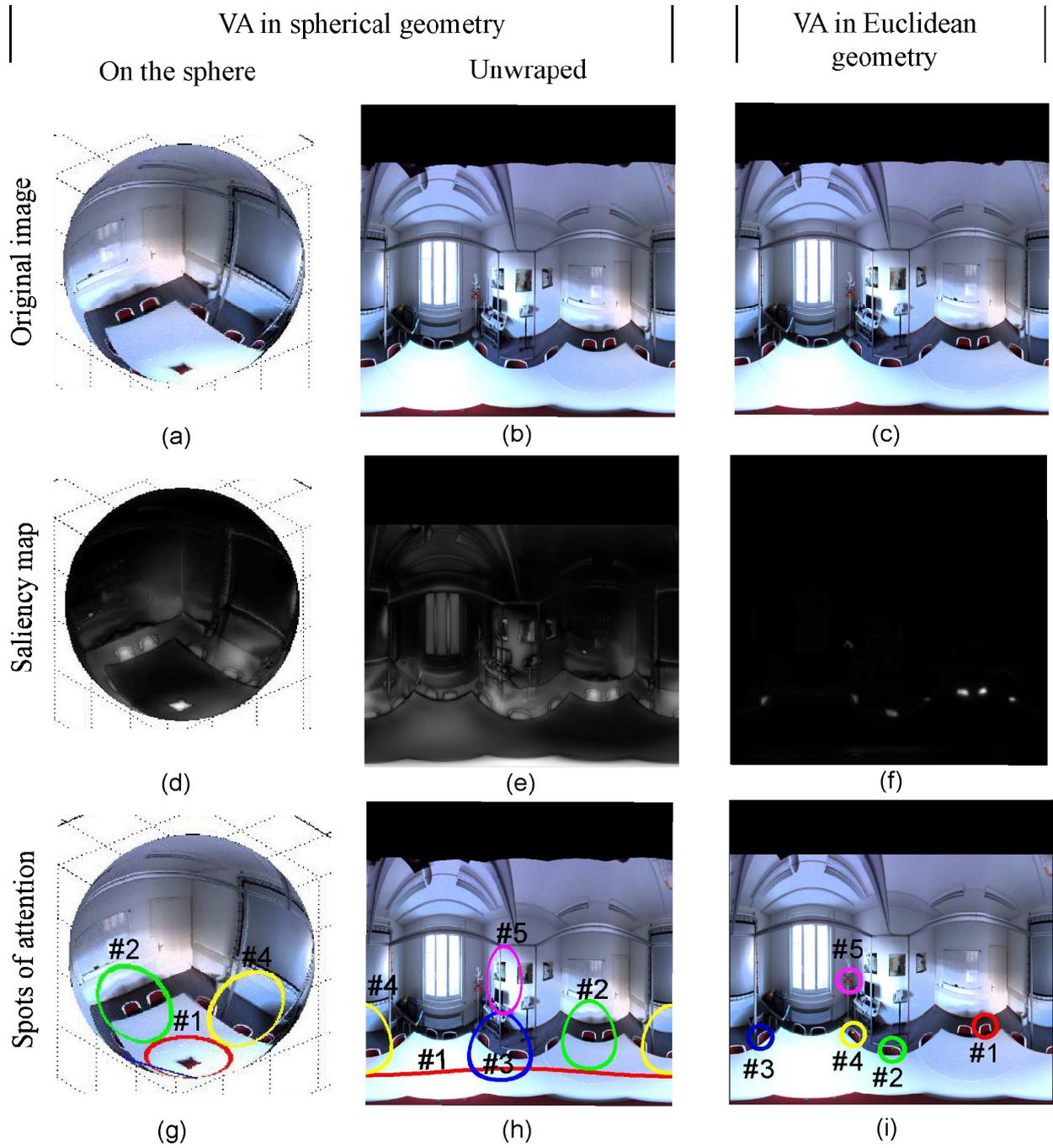


Figure 2: Comparison of saliency maps computed in Euclidean and spherical geometries: the most salient object in the scene is detected only by the spherical VA.

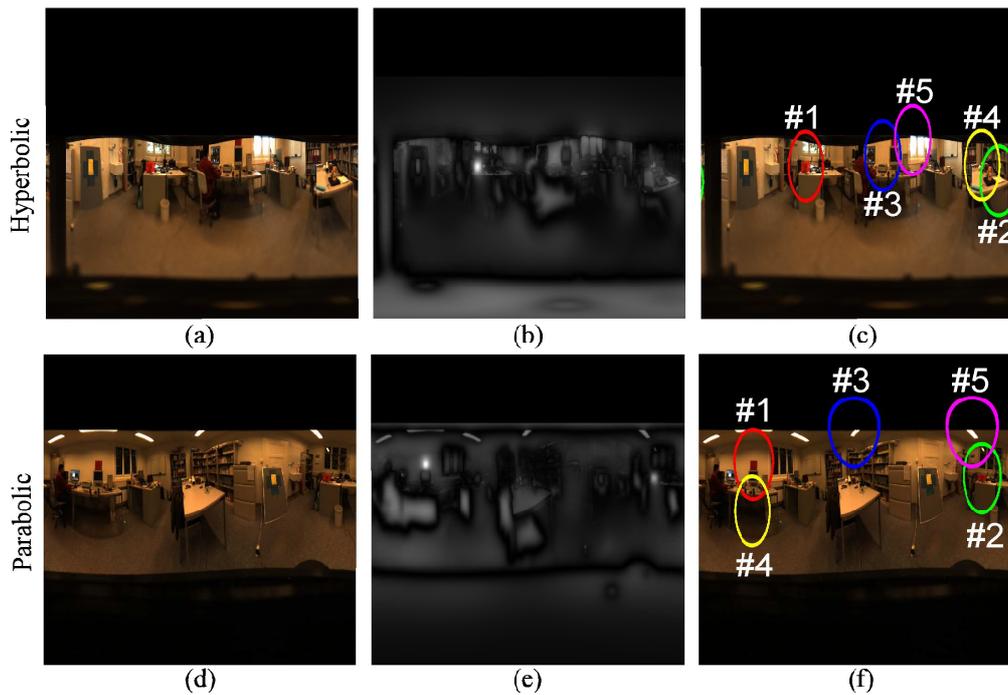


Figure 3: Visual attention in spherical geometry applied to parabolic and hyperbolic images.

with parabolic and hyperbolic mirrors. After a review of the spherical computing approach, the paper provided the required transformation for mapping the different sensor images onto the sphere for further processing. Then we presented a series of experimental results. Multi-camera omnidirectional images were used to test the capability of the new approach to improve spot detection in comparison to the conventional approach. The comparison illustrates the capability of the spherical approach to provide saliency maps with homogeneous response on the sphere and therefore shows its advantages for detecting spots of attention in omnidirectional scenes. Finally, the experiments with the catadioptric sensors demonstrate the feasibility of the approach for these types of simpler and cheaper sensors which are expected to be used in many future applications.

REFERENCES

- [1] E.H. Adelson and J. Bergen, "The Plenoptic Function and the Elements of Early Vision", *Annals of Physics*, vol.1, pp.1–37,1993.
- [2] S. Baker, and S. Nayar , "A theora of single-viewpoint catadioptric image formation", *Intrenational Journal of Computer Vision*, vol.35(2), pp 175–196, 1999.
- [3] I. Bogdanova, A. Bur, and H. Hügli, "Visual attention on the sphere [submitted]" Nov. 2007.
- [4] A. Bur, and H. Hügli, "Optimal Cue Combination for Saliency Computation: A Comparison with Human Vision", *Second International Work-Conference on the Interplay Between Natural and Artificial Computation*, LNCS 4528, pp 109-118, 2007.
- [5] A. Bur, A. Tapus, N. Ouerhani, R. Siegart and H. Hügli, "Robot navigation by panoramic vision and attention guided features," in *Proc. ICRP 2006*, Washington DC, USA, 2006, pp. 695–698.
- [6] L. Itti, C. Koch and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.20, pp 1254–1259, 1998.
- [7] C. Geyer and K. Daniilidis, "Catadioptric projective geometry," *International Journal of computer vision*, vol. 45(3), pp. 223–243, 2001.
- [8] C. Koch, and S. Ullman, "Shifts in Selective Visual Attention: Towards the underlying Neural Circuitry", *Human Neurobiology*, vol.4, pp 219-227, 1985.
- [9] O. Lemeur, P. Le Callet, D. Barba, and D. Thoreau, "A Coherent Computational Approach to Model Bottom-Up Visual Attention", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.5, pp 802-817, 2006.
- [10] N. Ouerhani and H. Hügli, "A model of dynamic visual attention for object tracking in natural image sequences," in *Proc. of International Conference on Artificial and Natural Neural Network*, Springer, 2003, pp.702–709.
- [11] N. Ouerhani and H. Hügli, "Multiscale attention-based presegmentation of color images," in *4th International Conference on Scale-Space Theories in Computer Vision*, ser. LNCS, 2003, vol. 2695, pp.537–549.
- [12] D. Walther, U. Rutishauser, C. Koch, and P. Perona, "Selective visual attention enables learning and recognition of multiple objects in cluttered scenes," *Computer Vision and Image Understanding*, vol. 100, pp. 41–63, 2005.
- [13] <http://www.ptgrey.com/products/ladybug2/index.asp>
- [14] <http://www.kaidan.com/Detail.bok?no=101>
- [15] <http://www.neovision.cz/prods/panoramic/h3s.html>