Visual Attention on the Sphere

Iva Bogdanova, Alexandre Bur and Heinz Hügli

Abstract

Human visual system makes an extensive use of visual attention in order to select the most relevant informations and speed-up the vision process. Inspired by visual attention, several computer models have been developped and many computer vision applications rely today on such models. But the actual algorithms are not suitable to omnidirectional images, which contain a significant amount of radial distorsion. In this paper, we present a novel computational approach that performs in spherical geometry and thus is suitable for omnidirectional images. Following one of the actual models of VA, the spherical saliency map is obtained through multi-scale analysis on the sphere. The spots of attention on the sphere represent the consecutive maximums in the spherical saliency map. The proposed method is compared to the standard one using a synthetic and a real omnidirectional images. Then, we provide some more examples of spots detection in spherical images. Finally, the proposed algorithm is rotation-invariant and can be applied on any type of omnidirectional images that can be mapped on the sphere.

I. INTRODUCTION

A. State of the Art on Visual Attention

It is generally admitted today that the human visual system makes extensive use of visual attention (VA) in order to select relevant visual information and speed up the vision process. Visual attention represents also a fundamental mechanism for computer vision where similar speed up of the processing can be envisaged. Thus the paradigm of computational VA has been widely investigated over the past three decades and possible fields of application include image and video compression [1], [2], object recognition [3], [4], image segmentation [5] and robot localization [6], [7].

While some computational VA models concentrate on psychophysical and biologically plausible aspects [8], [9], [10], other models focus more on efficient computation for related computer vision applications. Proposed in [11], the first computational architecture of VA includes the main concepts such as the feature

The authors are with the Institute of Microtechnology, University of Neuchâtel, Rue A.-L. Breguet 2, CH-2000 Neuchâtel, Switzerland. E-mail: {iva.bogdanova, alexandre.bur, heinz.hugli}@unine.ch

integration, the saliency map, winner-take-all network (WTA), center-surround difference and inhibition of return (IOR). Several models are based on these concepts. In [12], the authors develop one of the first implementation including center-surround difference based on classical filtering and a relaxation process is used for the map integration, resulting to high computational cost. One of the most actual and used model is presented in [13] and [14]. An efficient approximation of center-surround difference is performed with gaussian image pyramid and the relaxation process is replaced by a weighting scheme for the map integration, resulting to faster computation.

There are several VA models that use classical filtering approach [14], [12], [4], [1], [15], [10], and other models that are based on neural network [16], [17]. The mentionned computer models are mainly bottom-up, i.e. focus of attention driven by a reflexive behavior, due to strong feature-related contrasts. Other approaches include top-down informations, in which attention is driven by prior knowledges, expectations or tasks. In [18] and [19], the saliency map results from the fusion of bottom-up and top-down cues.

B. Omnidirectional Vision: the sphere of view

While conventional imaging systems (like photographic or video) are severely limited in their field of view, omnidirectional imaging systems were developed so that they are capable of viewing the world in all directions from a center of projection, i.e. the entire sphere of view around a single point. An ideal omnidirectional image thus provides a full spherical field of view (4π steradian) of a given scene. The scene, as viewed by an omnidirectional imaging system, can be represented by the ideal plenoptic function which completely describes the light field [20]. In real omnidirectional systems, the field of view is somehow restricted: typical omnidirectional images will thus exhibit a full azimuthal (360^0) range but a restricted zenithal ($< 180^0$) range. Such images will be considered here.

Various kinds of omnidirectional imaging systems exist. For instance, a *rotating camera* at 360^0 can produce a sequence of images that covers the whole scene. But this kind of imaging system cannot simultaneously cover actions in all directions of a dynamic scene. To cope with this, either a *multi-camera* or a *catadioptric sensor* can be used instead. The first one simultaneously captures images that cover about 75% of the visible sphere. Actually, in this case, all individual images (from each of the cameras in the set) are stitched together in order to form a spherical image. The second one is an imaging system based on combination of a curved mirror and a lens to form a projection onto the image plane of a (video) camera [21], [22]. Such an omnidirectional imaging system offers the potential for simultaneously capturing an image with a high resolution on a target as well as a wide field-of-view periphery. Of interest are catadioptric imaging systems with a single effective viewpoint called central catadioptric sensors, as

the hyperbolic or the parabolic ones. It is important to note that there is an equivalence between the catadioptric projection and two-steps mapping onto the sphere [23].

It is clear that the omnidirectional images can be defined as non-Euclidean: spherical, hyperbolic, parabolic.

C. A word of motivation: why visual attention on the sphere?

The current VA algorithms are limited to Euclidean images and cannot process non-Euclidean images (like omnidirectional images). Therefore, only conventional images can be treated by so known visual attention implementations. In the same time, the omnidirectional sensors are more and more used nowadays because of their advantages over the conventional imaging sensors, namely they have larger field of view. One particular example of application is in surveillance [24], [25], where one would be able to track persons in heavily cluttered environments. They are required in robotics, where an autonomous robot may benefit from omnidirectional vision for collision avoidance and situational awareness [26], [27].

The intuitive approach for defining a VA on omnidirectional image is to first map the image to a panoramic image (i.e. unwrap it) and then to apply the conventional VA algorithm. In fact, it was proposed in [6] an adaptation of the actual models to omnidirectional (panoramic) images. The proposed representation basically uses a cylindrical source image and the this algorithm solves the circularity problem by introducing a modulo operation on the i-indexes in order to handle the cyclic nature of that coordinate. This approach however is bound by the cylindrical projection and therefore is limited to panoramic images with a specific and limited zenithal extension.

The images obtained by omnidirectional sensors suffer under significant deformations. The deformation has a continuous component (i.e. it would appear also on a continuous imaging surface) and a disctete component (due to decimation induced by sampling). This makes the deformations in omnidirectional imaging sensors different than the deformations in space-variant sensors like log-polar retinas where the deformations is based on non-uniform sampling. A natural choice of a non-deformed domain for the full sphere of view, where there are no limitations on the zenithal range, is the sphere $S^2 \in \mathbb{R}^3$. Therefore, we argue that detection of spots of attention in omnidirectional images has to take place in spherical geometry. Developing an VA algorithm for an omnidirectional image, providing the full sphere of view of a scene, is equivalent to detecting the spots of attention on the sphere, i.e. in all directions .

The possibility to compute VA on the sphere opens new perspectives for large-field-of-view imaging applications, where the Euclidean geometry does not hold anymore. Computing VA in spherical coordi-

nates provides a homogeneous, i.e. rotation-invariant behaviour applicable to any omnidirectional image that can be mapped on the sphere.

In this paper, we present a derivation of a new algorithm that offers a solution for processing images obtained by omnidirectional imaging systems. Inspired by a classical visual attention model applied on Euclidean images, the proposed model operates in spherical geometry and is thus applicable for any signal defined on the sphere. By computing in spherical coordinates, the attention mechanism remedies the distortions introduced by the computations as performed in the Euclidean case.

This paper is organized as follows. Section 2 presents a classical model of visual attention and its algorithm. Then the paper proceeds with pointing out data processing on the sphere and more particularly, with presenting in Section 3 the multiscale analysis of spherical data. Section 4 then presents the VA algorithm on the sphere which is then applied to omnidirectional images, first to a synthetic spherical image in Section 5 and to a real omnidirectional image from a catadioptric sensor in Section 6.

II. Visual Attention: the Euclidean case

This section presents the classical model of VA and the related equations which hold for its computation in the 2D Euclidian space.

A. The Saliency-Based Model of VA

The saliency-based model of VA, originally proposed by Koch and Ullman in [11] is now widely used. Several works have dealt with the realization of this model e.g. [13]. It is based on three major principles: VA acts on a multi-featured input; saliency of locations is influenced by the surrounding context; the saliency of locations is represented on a scalar saliency map.

The different steps are detailed in the model illustrated in Figure 1 which, for a simpler notation, has a specific number of features (7) and cues (3) although any number can be considered in general.

First, seven features $(1..j..7)$ are extracted from the scene by computing the so-called feature maps from an RGB color image. The features are:

- intensity feature F_1;
- two chromatic features based on the two color opponency filters red-green F_2 and blue-yellow F_3;
- four local orientation features $F_{4..7}$.

In a second step, each feature map is transformed into its conspicuity map C_j, which highlights the parts of the scene that strongly differ, according to a specific feature, from their surrounding. This process, relies on a multiscale center-surround filtering which is described in Section II-C.
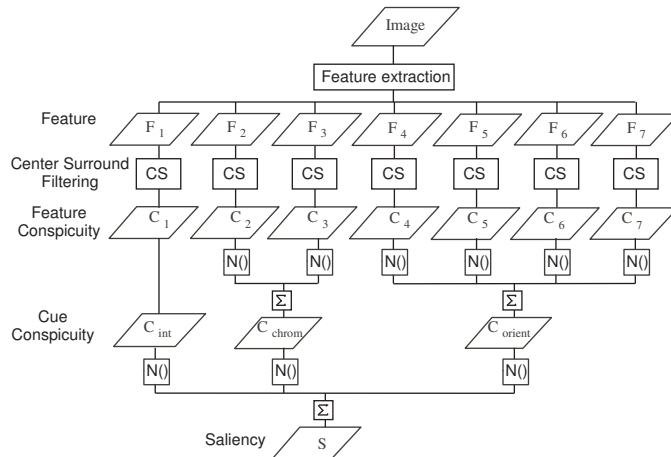
Fig. 1.   The saliency-based model of visual attention.

In the third step, the seven ($j = 1...7$) features are then grouped using a competitive map integration scheme and according to their nature into the three cues: intensity, color and orientation. The ***cue conspicuity maps*** are thus:

$$C_{int} = C_1; \tag{1}$$

$$C_{chrom} = \sum_{j\epsilon\{2,3\}} \mathcal{N}(C_j); \tag{2}$$

$$C_{orient} = \sum_{j\epsilon\{4,5,6,7\}} \mathcal{N}(C_j), \tag{3}$$

where C_{int} is the intensity conspicuity map, C_{chrom} is the chromaticity conspicuity map and C_{orient} is the orientation conspicuity map. $\mathcal{N}(.)$ refers to a normalization function defined below in Section II-B.

In the final step of the attention model, the cue conspicuity maps are integrated, into a ***saliency map*** S, defined as:

$$S = \sum_{cue\epsilon\{int,orient,chrom\}} \mathcal{N}(C_{cue}). \tag{4}$$

Given a saliency map, the WTA mechanism starts with selecting the location with the maximum value of the map. This selected region is considered as the most salient part of the image (winner). The ***spot of attention*** is then shifted to this location. Local inhibition is then activated in the saliency map, in an area around the actual spot. This yields dynamical shifts of the spot of attention by allowing the next most salient location to subsequently become a winner. Besides, the inhibition mechanism prevents the spots of attention from returning to previously attended locations.

B. Normalization $\mathcal{N}()$ for map fusion

The normalization function adjusts the range of maps of different nature and simulates the competition between the different maps to be integrated. Several methods were proposed, which are reviewed in [28]. Although any method would be applicable here, for simplicity, a linear weighting scheme is considered in this paper. Given the conspicuity map C(**x**), it defines $\mathcal{N}(C(\mathbf{x}))$ as:

$$\mathcal{N}(C(\mathbf{x})) = w \cdot C(\mathbf{x}) \quad \text{with} \quad w = \frac{Max(C(\mathbf{x}))}{Mean(C(\mathbf{x}))}. \tag{5}$$

C. Multiscale Center-Surround Filtering

Each feature map F_j is transformed independently into a feature conspicuity map C_j by a biologically inspired center-surround mechanism that tends to highlight the parts of the feature map that strongly differ from their surrounding. In order to highlight regions of different sizes, the mechanism applied at various map scales and thus the transform consists in a so called multiscale center-surround filtering, illustrated in Figure 2 and described as follows.

First, for a feature F_j, a gaussian pyramid is created by successively lowpass filtering the signal with a gaussian filter g and down-sampling the result by factor 2. Formally, the successive images f_k at the pyramid output are:

$$f_0 = F_j, \quad f_1 = \downarrow 2(f_0 * g), \quad ... \quad, f_k = \downarrow 2(f_{k-1} * g).$$

Then, the effective center-surround mechanism, which necessitates subtracting a surround region from a center region, is implemented by simply subtracting pairwise output images from the gaussian pyramid. According to the VA model that forsees a ratio of 8 in the relative sizes of center and surround, the scale difference of the images to be subtracted is thus 3 and 4. Accordingly, from the $f_k, k = 1..n$ maps a number of multiscale maps $\mathcal{M}_k$ are thus computed as follows:

$$\mathcal{M}_1 = |f_2 \ominus f_5|, \mathcal{M}_2 = |f_2 \ominus f_6|, \mathcal{M}_3 = |f_3 \ominus f_6|, \mathcal{M}_4 = |f_3 \ominus f_7|, \mathcal{M}_5 = |f_4 \ominus f_7|, \mathcal{M}_6 = |f_4 \ominus f_8|, \tag{6}$$

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 substraction.

Finally, the ***feature conspicuity map*** C_j is computed by combining in a competitive way the set of multi-scale maps $\mathcal{M}_k$ using the normalization function defined in Equation (5):

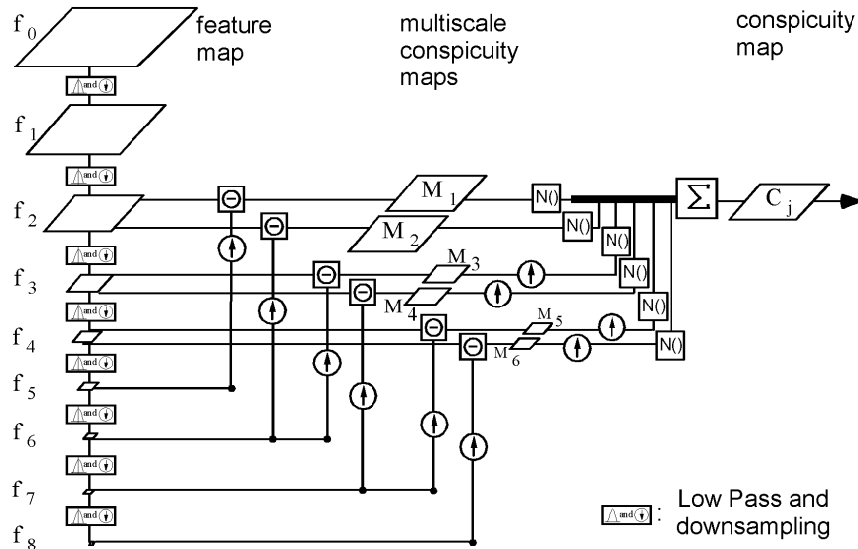$$C_j = \sum_{k=1}^{n-3} \mathcal{N}(\mathcal{M}_k). \tag{7}$$

Fig. 2. Multiscale center-surround mechanism with an $n = 8$ level pyramid and six scale maps

III. DATA PROCESSING ON THE SPHERE

This section presents data processing methods in spherical geometry and proposes a means for a new algorithm for computing in the spherical domain a conspicuity map of a single feature.

A. Spherical Geometry

The 2-sphere $(S^2 \in \mathbb{R}^3)$ is a compact manifold of constant positive curvature. In polar coordinates, each point on the sphere is a three-dimensional vector $\omega = (x_0, x_1, x_2) \equiv (r \cos \theta, r \sin \theta \sin \varphi, r \sin \theta \cos \varphi)$, with $r \in (0, \infty), \theta \in [0, \pi]$ and $\varphi \in (0, 2\pi]$ as illustated in Figure 3(a).

This Figure 3(b) also illustrates the so called stereographic projection from the South Pole, a projection that maps any point of the sphere onto a point of the tangent plane at the North Pole. If we take the sphere S^2 as the Riemannian sphere $(r = 1)$ and the tangent plane as the complex plane $\mathbb{C}^2$, then the stereographic projection is a bijection given by

$$\Phi(\omega) = 2 \tan \frac{\theta}{2} e^{i\varphi}, \tag{8}$$

where $\omega \equiv (\theta, \varphi), \ \theta \in [0, \pi], \varphi \in [0, 2\pi)$.

B. Multiscale Analysis on the Sphere S^2

The central step towards transposing VA onto the sphere concerns the multiscale center-surround filtering. As shown in Section II-C, this filtering requires to set up a Gaussian pyramid and to compute
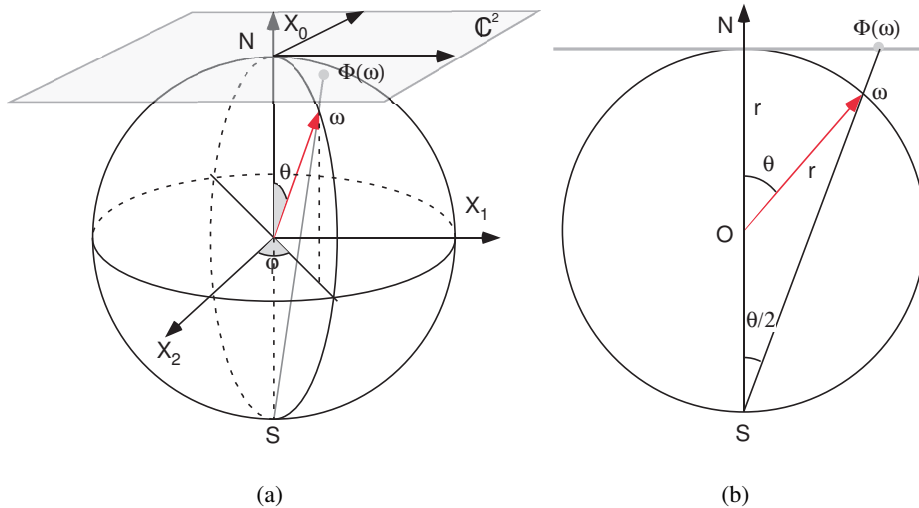
Fig. 3. Spherical geometry: (a) spherical coordinates, (b)cross-section of a stereographic projection through the South Pole.

differences of two maps of the pyramid which are each time three or four scales apart. This contrasts the implementation of anaother spherical pyramid, namely the spherical Laplacian Pyramid, published in [29]. As both problems are very similar, the procedure for VA is widely inspired from it. In fact, a Laplacian pyramid also relies on differences of maps from a Gaussian pyramid. It differs however in the scale difference of the two maps to be subtracted. The procedure is adapted accordingly. In the following we present in details the filtering on the sphere that is needed for building the Spherical Gaussian Pyramid.

C. Filtering on the Sphere

In general, the convolution on the (Euclidean) plane is defined in terms of the inner product between two functions translated relative to each other, and is parameterized by the amount of translation. On the sphere, it is more natural to use relative rotations. For a given spherical signal f and a filter g, their convolution $(g \star f)_{\alpha_2, \alpha_1, \alpha_0} \in L^2(SO(3))$ reads

$$(g \star f)_{\alpha_2, \alpha_1, \alpha_0} = \int_{S^2} [R_{\alpha_2, \alpha_1, \alpha_0} g] \, (\theta, \varphi) f(\theta, \varphi) d \cos \theta d\varphi, \qquad (9)$$

where $R_{\alpha_2, \alpha_1, \alpha_0}$ is the rotation operator that first rotates the function by α_0 about the x_0-axis, then by α_1 about the x_1-axis and finally by α_2 about x_0-axis again. These are the three (Euler) angles which define an element of $SO(3)$-group and they provide a natural parameterization of the convolution on the sphere. Actually, the convolution $(g \star f)_{\alpha_2, \alpha_1, \alpha_0}$ is the inner product of the rotated version of the filter g with the signal f, or the projection coefficient of f onto $[R_{\alpha_2, \alpha_1, \alpha_0}]$. If the filter is an axisymmetric function, i.e. $g(\theta, \varphi) = g(\theta)$, the rotation by α_0 about the x_0-axis has no effect. In other words this

reads

$$(g \star f)_{\alpha_2, \alpha_1, \alpha_0} = (g \star f)_{\alpha_2, \alpha_1},$$

and is a spherical signal parametrized by $\theta \equiv \alpha_1$ and $\varphi \equiv \alpha_2$.

It is obvious that in practice we need the discrete form of the spherical convolution. For an axisymmetric filter g we can discretize Equation (9) and thus obtain

$$(g \star f)_{\alpha_2, \alpha_1} = \sum_{\theta=0}^{\pi} \sum_{\varphi=0}^{2\pi} [R_{\alpha_2, \alpha_1} g(\theta, \varphi)] f(\theta, \varphi).$$

First, we must note that the $(\theta, \varphi)-$grid is not invariant under rotation, therefore it is not usually possible to evaluate $[R_{\alpha_2, \alpha_1} g(\theta, \varphi)]$ from the samples of $g(\theta, \varphi)$ on the grid. That is why this kind of discrete implementation of convolution on the sphere is not efficient.

Let us have a spherical signal defined at scale k on a $2\beta_k \times 2\beta_k$, $(\beta_k \in \mathbb{N})$ square grid of respectively equi-angular resolution in θ and φ:

$$\omega \in \mathcal{G}_k := \left\{ \omega_{kpq} = (\theta_{kp}, \varphi_{kq}) : \theta_{kp} = \frac{(2p+1)\pi}{4\beta_k}, \varphi_{kq} = \frac{q\pi}{\beta_k}, \ p, q \in \mathbb{Z}[2\beta_k] \right\}. \tag{10}$$

This grid allows us to perfectly sample any band-limited function $f \in L^2(S^2)$ of bandwidth β_k, i.e. such that the Fourier coefficients $\hat{f}(l, m) = 0, \forall l > \beta_k$. Moreover, this class of sampling grids is associated to a Fast Spherical Transform [30].

Another method to perform spherical convolution is to first project the discretized spherical function and filter onto the span of spherical harmonics and perform the convolution in Fourier domain via simple multiplication. Of great importance is the Spherical Convolution Theorem, as derived in [31] and which we remind here for convenience: for functions $f, g \in L^2(S^2)$, the transform of the convolution is a point-wise product of the transforms:

$$\widehat{g \star f}(l, m) = \sqrt{\frac{4\pi}{2l+1}} \hat{g}(l, 0) \hat{f}(l, m), \tag{11}$$

where $\hat{f}(l, m)$ denotes the (l, m)-Fourier coefficient, which in discrete form reads

$$\hat{f}(l, m) = \frac{\sqrt{2\pi}}{2\beta_k} \sum_{p=0}^{2\beta_k - 1} \sum_{q=0}^{2\beta_k - 1} \alpha_{kp}^{\beta_k} \, f(\theta_{kp}, \varphi_{kq}) \, e^{-im\varphi_{kq}} \times P_l^m(\cos\theta_{kp}), \tag{12}$$

with P_l^m -the associated Legendre function of degree l and order m and $\alpha_{kp}^{\beta_k}$ is a weight.

It must be noted that the convolution theorem is independent of sampling. Hence, as long as we can project our samples onto the span of spherical harmonics accurately, we can perform convolution via

Fourier domain accurately, regardless of the sampling grid. These methods as implemented in *Spharmon-icKit*[1] and used together with MATLAB YAWtb toolbox[33].

D. Spherical Gaussian Pyramid

Regarding the VA mechanism as defined in Section II-A, we first need to define a Gaussian pyramid on the sphere. The filter used for smoothing the spherical data is a spherical axisymmetric Gaussian low-pass filter, defined by its Fourier coefficients:

$$\hat{g}_{\sigma_k}(l) = e^{-(\sigma_k l)^2}. \tag{13}$$

The parameter σ_k is chosen so that the filter is numerically close to a perfect half-band filter $|\hat{g}_{\sigma_k}(l)| << 1, \forall l > \beta_k$.

Let us look at a spherical signal defined on a grid of size 1024×1024 and which represents the scale level $k = 0$. The Fourier transform of this signal results in (512×512)-matrix, and thus we have $\beta_0 = 512$ and $l \in [0, \beta_0]$. Then, for designing the gaussian filter at this level, we need to define the bandwidth parameter. The choice $\sigma_0 = \frac{\sqrt{2}}{\beta_0} = 0.0028$ satisfies our requirement $|\hat{g}_{\sigma_0}(l)| \approx 0, \forall l > 512$. For purpose of visualization and for this particular value of σ_0 we can easily find the corresponding filter in spatial domain which is depicted on Figure 4.

Finally, we can define the procedure for obtaining one level in the *spherical Gaussian pyramid* that transforms f_k into f_{k+1} as follows:

1) Apply the spherical Fourier transform Equation (12) on the signal f_k and thus obtain $\hat{f}_k(l, m)$;
2) Multiply in Fourier domain with a Gaussian filter and scale as defined in Equation (11) and thus obtain $\widehat{\hat{g} \cdot f_k}(l, m)$;
3) Apply the inverse spherical Fourier transform on it and thus obtain the filtered signal $g \star f_k$;
4) Subsample this signal $g \star f_k$ by a factor of 2 and thus obtain f_{k+1}.

Subsampling on the sphere consists of reducing by a factor of 2 the spherical grid $\mathcal{G}_k$.

The full spherical Gaussian pyramid is now obtained by iteratively applying previous procedure to each scale level $k = 1 \cdots n$, thus producing the sequence of spherical signals $f_1, f_2, f_3, \cdots f_n$.

A schematic representation of the spherical Gaussian pyramid is illustrated in Figure 5 for the case of an initial spherical signal defined on a grid of size 1024×1024. Note that the grid size is reduced by a factor of 2 at each level and that a grid size of 4×4 characterizes the last 8th level. This repeated size

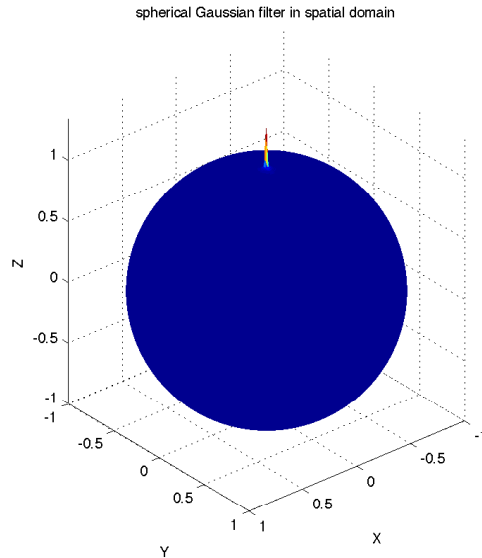---

[1]under GPL license (General Public License)[32]

Fig. 4.   Spherical Gaussian Filter.

reduction also clearly speaks in favor of an initial grid size which is sufficiently large and expressed as a power of 2: $2\beta_0 \times 2\beta_0 = 2^n \times 2^n$.

E. Spherical Gabor Pyramid

In order to deal with the orientations in spherical coordinates we proceed toward defining a Gabor Pyramid on the sphere. In fact, there are oriented spherical filters that have anisotropic nature, i.e. they are not axisymmetric. Let us remind, that we have defined in Equation 9 a rotation operator $R_{\alpha_2, \alpha_1, \alpha_0}$ that first rotates the function by α_0 about the x_0-axis, then by α_1 about the x_1-axis and finally by α_2 about x_0-axis again. If a rotation by α_0 takes place, the function is anisotropic, i.e the filter is oriented at angle $\phi \in [0, 2\pi)$ and thus depends on three rotation angles: $g(\theta, \varphi, \phi)$. In fact, on the sphere one can define orientation with respect to meridians and parallels. In other words, directions can be referred to as cardinal points: $\phi = 0^o$ corresponds to North-South direction, i.e. meridians, and $\phi = 90^0$ to East-West directions, i.e. parallels.

The intuition about what a directional filter in spherical domain looks like, is clear: if centered at North Pole, the spherical directional filter has a stereographic projection on the tangent plane that is directional in the Euclidean sense. A natural candidates for such is the Gabor spherical filter:

$$G_{gabor}(\theta, \varphi) = \frac{e^{i|k_0| \tan(\theta/2) \cos(\varphi_0 - \varphi)} e^{-(1/2) \tan^2(\theta/2)}}{1 + \cos\theta}, \tag{14}$$
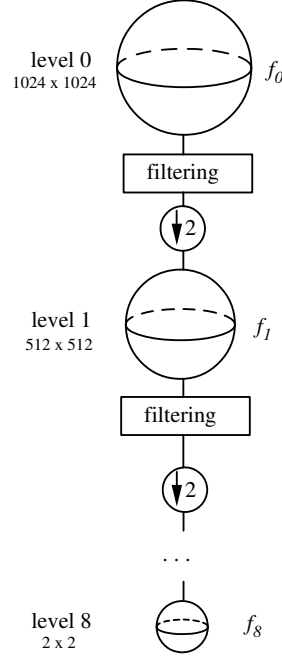
Fig. 5.   Schematic diagram for the gaussian pyramid on S^2.

where φ_0 is the argument of $|k_0|$. This filter is (numerically) admissible only for $|k_0|$ large enough, usually $|k_0| \geqslant 6$. Moreover, a spherical Gabor wavelet reads [34], [35]:

$$\psi_{gabor}(\theta, \varphi) = \lambda^{1/2}(\theta, a)e^{ik_0\frac{2}{a}\tan\frac{\theta}{2}\cos(\varphi_0 - \varphi)}e^{-\frac{2}{a^2}\tan^2\frac{\theta}{2}}\Big(1 + \frac{1}{a^2}\tan^2\frac{\theta}{2}\Big), \tag{15}$$

where $\lambda(\theta, a)$ is a normalization factor and the scale parameter $a > 0$. A particular example of Gabor spherical wavelet for two different orientations is shown in Figure 6. We can interpret $g(\theta, \varphi, \phi)$ as $g(\omega, \phi)$, where $\omega \in S^2$ is a position on the sphere and ϕ is an angle of rotation.
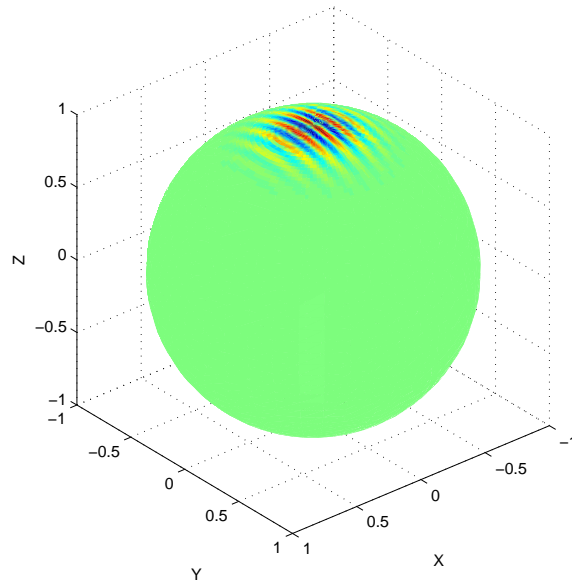
The directional correlation on the sphere is expressed in terms of the Wigner D-function coefficients which read [36]:

$$\widehat{g_{gabor} \star f}(l, m, n) = \frac{8\pi^2}{2l + 1}\hat{g}^*_{gabor}(l, n)\hat{f}(l, m), \tag{16}$$
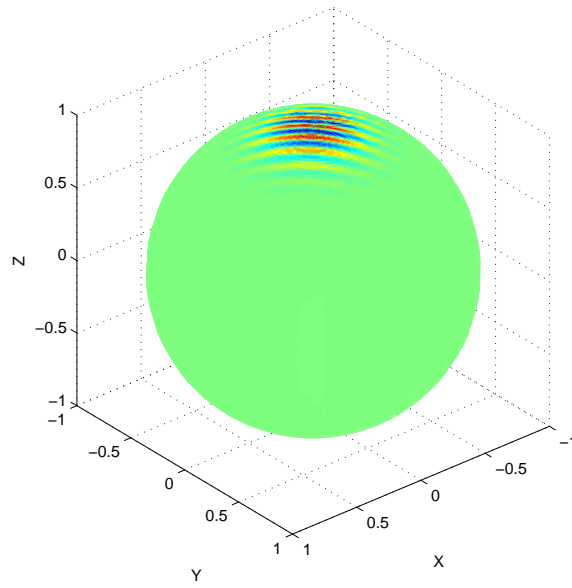
where g^* denotes the complex conjugate. In other words, the Wigner D-function coefficients of the directonal correlation are given as the pointwise product of the scalar spherical harmonics coefficients $\hat{f}(l, m)$ and $\hat{g}^*(l, n)$. This method is implemented in *Soft*.

Now, we can define the procedure for obtaining one level in the *Spherical Gabor Pyramid* which transforms r_k into r_{k+1}:

1) Apply the spherical Fourier transform Equation (12) on the signal r_k and thus obtain $\hat{f}_k(l, m)$;

(a)

(b)

Fig. 6.   Real part of the spherical Gabor wavelet: (a) $a = 0.1$ orientation at 0 ,(b) $a = 0.1$, orientaion $\frac{\pi}{4}$.
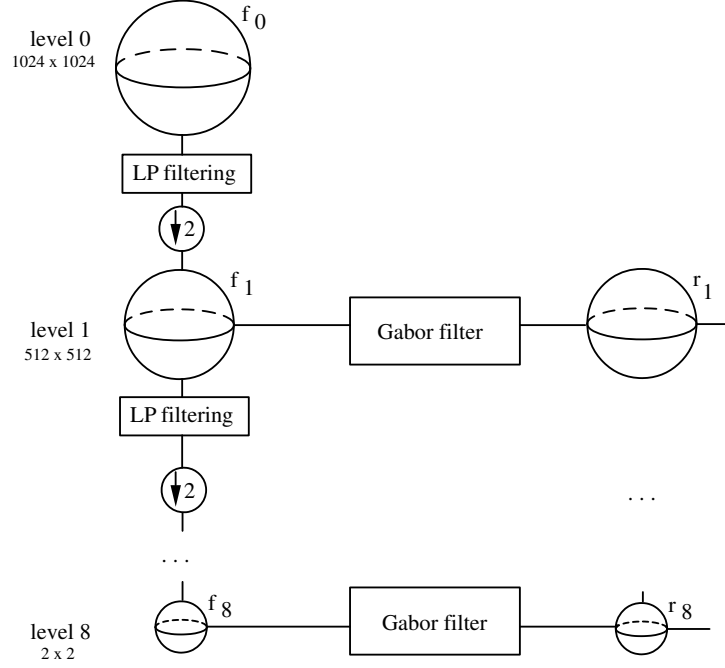
Fig. 7. Schematic diagram for the gabor pyramid on S^2.

2) Multiply in Fourier domain with a Gaussian filter as defined in Equation (11) and thus obtain $\hat{g} \cdot \hat{f}_k(l, m)$;

3) Apply the inverse spherical Fourier transform on it and thus obtain the filtered signal $g \star f_k$;

4) Subsample this signal $g \star f_k$ by a factor of 2 and thus obtain f_{k+1};

5) Apply again the spherical Fourier transform (Equation (12)) on f_{k+1} and obtain $\hat{f}_{k+1}(l, m)$;

6) Multiply in Wigner domain with a Gabor filter as defined in Equation 16 and obtain $\widehat{\hat{g}_{gabor} \cdot \hat{f}_{k+1}}(l, m, n)$;

7) Apply the inverse Wigner transform on it and obtain $r_{k+1} = g_{gabor} \star f_{k+1}$.

The full Spherical Gabor pyramid is obtained by iteratively applying the previous procedure to each scale level $k = 1 \cdots n$ and producing the sequence of spherical signals $r_1, r_2, \cdots r_n$. The schematic representation of the Spherical Gabor pyramid is illustrated in Figure 7.

F. Up-sampling on the Sphere

Up-sampling on the sphere $S^2 \in \mathbb{R}^3$ is the process of increasing the sampling rate of a spherical signal $f(\theta_k, \varphi_k)$. The sampling integer factor U multiplies the sampling rate. This process consists of two steps:

1) Add $(U-1)$ zeros between each sample in $f(\theta_k, \varphi_k)$ defined on the grid $\mathcal{G}_k$ (the spherical grid is defined in Equation (10));

2) Filter with a low-pass spherical filter;

The filtering is once again performed in the Fourier domain and a particular example is the Gaussian spherical filter as defined in Equation (13). For instance, at the lowest level of the spherical Gaussian pyramid ($k = 8$), the signal is defined on a grid of size 4×4. After introducing zeros into it, we obtain a signal defined on (8×8)-spherical grid whose FT results in 4×4-matrix, i.e. we have $\beta_8 = 4$. Consequently, $\sigma_8 = \frac{\sqrt{6}}{\beta_8} = 0.6124$, satisfies our requirement.

G. Normalization $\mathcal{N}_{S^2}()$ for spherical map fusion

The mean value of a function over the surface of a sphere, i.e. $f(\theta, \varphi) \in L^2(S^2)$, in discrete form is given by the double sum:

$$f_{mean_{S^2}} = \frac{1}{4\pi} \sum_{\theta} \sum_{\varphi} f(\theta, \varphi) \sin\theta,$$

where $\theta \in [0, \pi]$ and $\varphi \in (0, 2\pi]$.

Consequently, the normalization function for map fusion on the sphere , takes the following form:

$$\mathcal{N}_{S^2}(C(\theta, \varphi)) = w_{S^2} \cdot C(\theta, \varphi), \tag{17}$$

with

$$w_{S^2} = \frac{4\pi Max(C(\theta, \varphi))}{\sum_{\theta} \sum_{\varphi} C(\theta, \varphi) \sin\theta},$$

where $C(\theta, \varphi)$ is the corresponding spherical conspicuity map.

IV. Visual Attention Algorithm on the Sphere S^2

This section now defines the algorithm for VA in spherical geometry.

A. Spherical conspicuity map for a single feature

Let us have a spherical signal (image) f_0 define on grid of size $2^n \times 2^n$. The procedure for computing the feature spherical conspicuity map relies on the spherical Gaussian/Gabor pyramid and the center-surround mechanism. It includes the following steps:

1) construct the n-level spherical gaussian/gabor pyramid as described in Section III-D/ Section III-E;

2) compute the multiscale maps as defined in Equation (6);

3) compute the weight coefficients w_{S^2} and normalize the maps as defined in Equation (17);

4) compute the final *spherical conspicuity map* $\mathcal{C}_j$ using Equation (7).

B. Computing several features on the sphere

As a next step, we need to define the corresponding conspicuity maps C_j for each of the features of the spherical image, $j = 1 \cdots 7$ using the steps from Section IV-A:

1) intensity: C_1;

2) yellow-blue: C_2;

3) red-green: C_3;

4) orientation 0^0: C_4;

5) orientation 45^0: C_5;

6) orientation 90^0 : C_6;

7) orientation 135^0: C_7.

It sufficies to apply the algorithm as defined in Section IV-A for each of the features.

C. Spherical cue conspicuity maps

Based on the spherical feature conspicuity maps, the three spherical cue conspicuity maps are derived:

- C_{int};

- C_{chrom};

- C_{orient}

D. Spherical saliency map

The spherical saliency map is built by fusing together several previously obtained cue conspicuity maps, according to Equation (4) and using again for $\mathcal{N}()$, the spherical normalization function.

E. Spots of Attention on the Sphere

The maximum of the spherical saliency map, defines the most salient location on the sphere, which actually represents the spots of attention. The significant regions are detected successively using the "winner-take-all" (WTA) mechanism, as previously discussed in Section II-A. A local inhibition on the sphere takes place. The inhibition function used to attenuate the consecutively found spots of attention is defined by

$$F_{inh}(\omega) = 1 - G_{S^2}(\omega) \in L^2(S^2), \quad \omega \equiv (\theta, \varphi) \in S^2 \tag{18}$$

where $\theta \in [0, \pi], \varphi \in [0, 2\pi)$ and the spherical Gaussian reads [37]

$$G_{S^2}(\omega) = e^{-\eta^2 \tan^2 \frac{\theta}{2}}, \tag{19}$$

which is the inverse stereographic projection of the Gaussian in the tangent plane at the North pole of the sphere. The size of the filter depends on the parameter η.

The process of local inhibition consists of multiplying the spherical saliency map by the so defined in (18) function which is placed at the found maximum ω_{max}:

$$S_{inh} = S \cdot F_{inh}(\omega - \omega_{max}). \tag{20}$$

Finally, the number of detected locations can be either set by the user or automatically determined through the activities of the saliency map.

V. SALIENCY MAP ON A SPHERICAL SYNTHETIC SIGNAL

A given spherical synthetic signal f_0 consists of twelve white disks distributed along a great circle passing through the poles of a black sphere. The disk sizes are same among the three disks of a same quadrant of the great circle but they differ in each quadrant. It is illustrated in Figure 8(b), where four views of the sphere in 3-D space are provided. Each of these views represents one of the four disk-groups. The signal is defined on a 1024×1024 equi-angular spherical grid (θ, φ). Its unwrapped version is shown on Figure 8(a), where $\varphi \in [0, 2\pi)$ is at the horizontal axis and $\theta \in [0, \pi]$ is at the vertical axis. The beginning and the end of the vertical axis correspond to the South and North Poles of the sphere, respectively. The four groups of disks are easily distinguished but it is clear that each of the disks is deformed when the spherical signal is unwrapped. Actually, in this and in the following examples, the unwrapped version is provided only for purpose of visualization.

According to the algorithm defined in Section IV-A, we create first the 8 levels of the spherical gaussian pyramid, then the six corresponding multiscale conspicuity maps and obtain finally, after normalization and summation, the overall saliency map.

It appears in these results that the final spherical saliency map shows each time the same type of response for all disks of a group of three disks of same size (disks in the same quadrant). This shows that the performed VA computation is independent of the feature location on the sphere and that the proposed algorithm operates thus as expected.

For better understanding the differences between the saliency map computed in the spherical geometry and the corresponding one in the Euclidean geometry, it suffices to apply the Euclidean VA algorithm on the unwrapped spherical signal f_0 (i.e. f_0 is considered to be an Euclidean signal (as if it were defined on a cartesian grid (x, y))), and then to compare both results. Figure 8 provides such a comparison by displaying the saliency maps obtained by the Euclidian (Figure 8(e)) and spherical (Figure 8(c) and (d))

approaches. A useful element for their comparison is the nature of the filtering response: depending on their size, disks are detected differently. Processed in the Euclidean way, small disks are detected as spots and large disks are detected by their contours. Now, it appears that processing in the spherical domain provides, as expected, always the same kind of spot (or contour) detection for all disks of a same quadrant. On the other hand, Euclidian processing clearly provides different type of responses for identical disks, like for the group of disks top right in the saliency map.

In conclusion, this example of a synthetic spherical signal illustrates the capacity of the proposed VA algorithm to process the features and, consequently, saliency maps in a way which is independent of the location of features on the sphere. It shows also that this property is missing when the signal is processed using the Euclidean VA.

VI. VISUAL ATTENTION ON A REAL OMNIDIRECTIONAL IMAGE

We have seen how the spherical saliency map differs from the Euclidean one while performing the experiment in the previous section. Now, we proceed toward applying our algoritm on a real omnidirectional image, which represents the entire sphere of view. We first compute the spherical saliency map and then define the spots of attention.

A. Spherical Saliency Map

Let us start with an omnidirectional (r, g, b) color image of size 1024×1024 and obtained with a multi-camera sensor [38]. The input spherical image is shown in Figure 9(a), where two views of the sphere in 3-D are represented. seven features are considered, namely intensity f_{int}, two f_{BY} and f_{RG} chrominance components derived from the color image as follows:

$$
\begin{aligned}
f_{int} = I &= 0.3r + 0.59g + 0.11b, \\
f_{BY} &= \frac{(b - r/2 - b/2)}{I}, \\
f_{RG} &= \frac{(r - g)}{I},
\end{aligned}
$$

and four orientation features, f_{0°, f_{45°, f_{90° and f_{135°. The obtained conspicuity maps for the different features, intensity, red-green, and yellow-blue, are shown on Figure 9(b), (c) and (d), and the orientation features are shown in (e), (f), (g) and (h) respectively. Concequently, three spherical cue conspicuity maps are calculated: C_{int}, C_{chrom} and $C_{orientation}$.
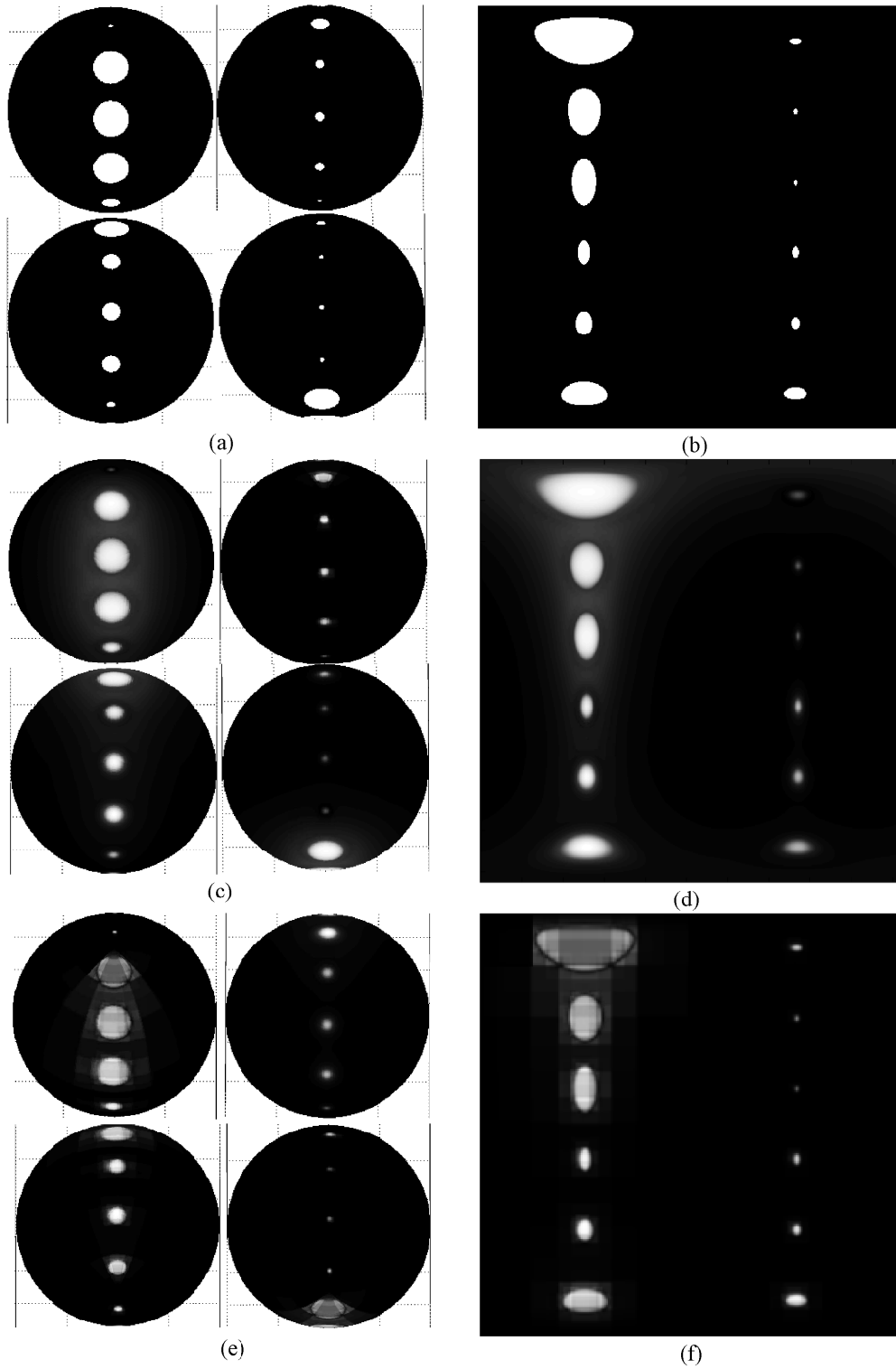
Fig. 8. Euclidean vs. spherical saliency map: (a) synthetic signal-4 views of the sphere, (b) unwrapped spherical synthetic image, (c) spherical saliency map represented on the sphere, (d) unwrapped spherical saliency map, (e) Euclidean saliency map represented on the sphere, (f) saliency map obtained after applying the Euclidean VA on the signal given in (a).

For the computation of the spherical saliency map, the procedure of Section IV-D applies. The final integration of the three cue conspicuities leads to the final spherical saliency map shown in Figure 9(c). For better visualization, we provide its unwrapped version as well, in Figure 9(d).

B. *Spots of attention*

The spots of attention are defined as described in Section IV-E. The two views of the sphere with the original image are shown in Figure 10(a) and the unwrapped version of the spherical image is illustrated in Figure 10(b). The spherical saliency map is shown in Figure 10(c) and (d) in its spherical and unwrapped versions. Finally, twelve spots of attentions are shown in Figure 10(e). Again, only for a purpose of visualization the unwrapped spherical image with the set of all spots of attention is illustrated in Figure 10(f).

VII. VA ON DIFFERENT OMNIDIRECTIONAL SCENES

In this section we experiment the spherical VA algorithm on some more omnidirectional images shown in Figure 11(a) which represent different scenes. The omnidirectional images are obtained wiha multicamera again, For computing the spherical saliency map three features are considered, i.e. intensity, red-green and yellow-blue. Consequently, two spherical cue conspicuity maps are defined and finally they are fused in order to obtain the spherical saliency map. Their saliency maps are shown in Figure 11(d).

Again, a number of spots of attention are defined based on the spherical saliency map and they are represented on Figure 11 (c).

VIII. EUCLIDEAN VS. SPHERICAL VA

For better illustrating the advantages of the visual attention on the sphere, we provide a particular example of an omnidirectional image, on which are applied both Euclidean and spherical visual attention.

The input image is again spherical image and is shown on Figure 12 (a). Both, its unwrapped version, as well as viewed on the sphere in 3-D. Actually, we use the unwrapped version only for the purpose of visualization since this is the only way to represent everything that is on the sphere. The scene represents a meeting room. We must note as well that the omnidirectional sensor directly outputs the image in spherical coordinates (θ, φ). After applying the VA as described in Section IV, we obtain the spherical saliency map. It is illustrated in Figure 12(b). The spots of attention are shown on Figure 12(c) as only the first three among them are ranked with their corresponding number.
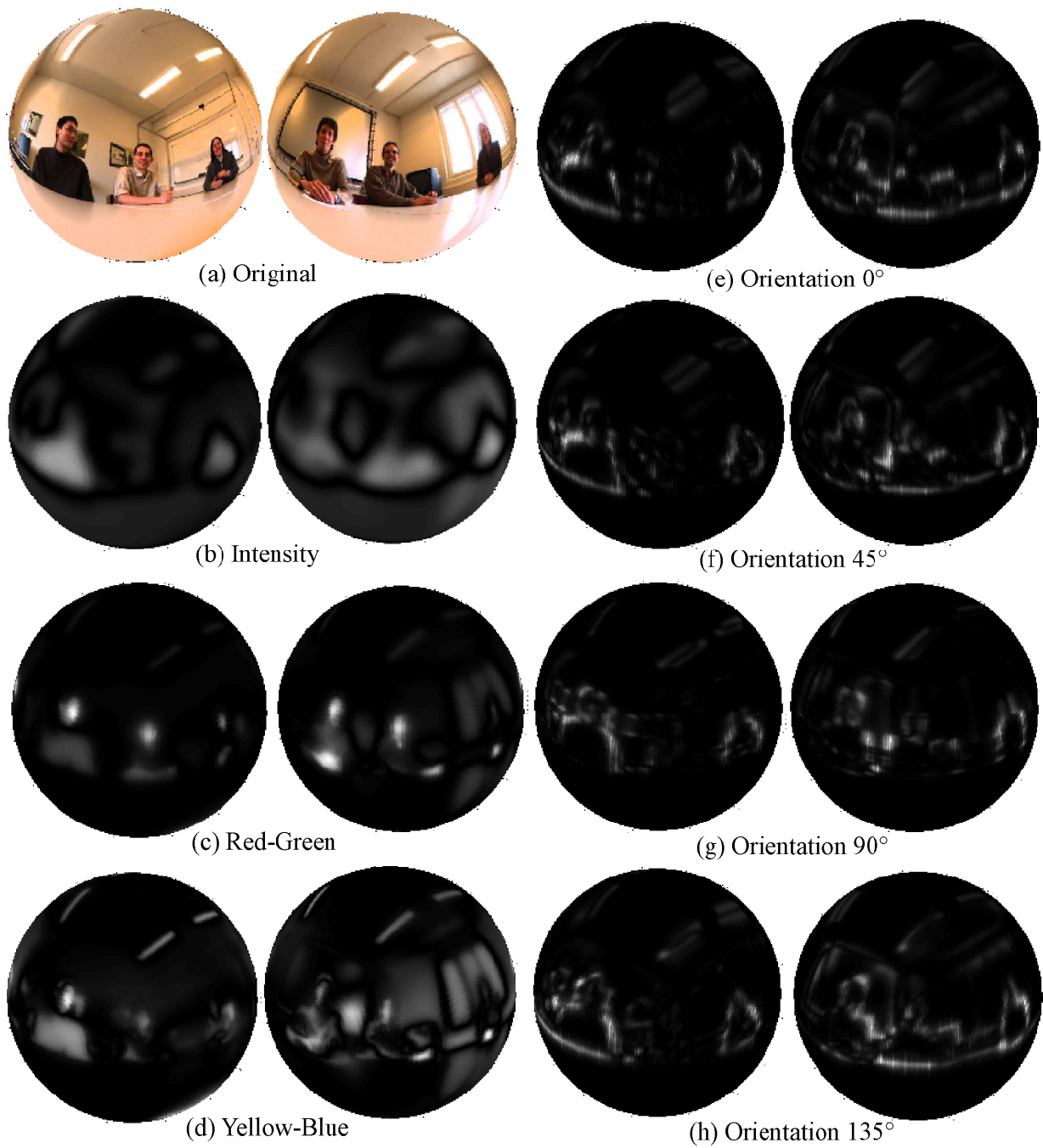
(a) Original

(e) Orientation 0°

(b) Intensity

(f) Orientation 45°

(c) Red-Green

(g) Orientation 90°

(d) Yellow-Blue

(h) Orientation 135°

Fig. 9.   Spherical saliency map of a real omnidirectional image (3-d view of the sphere): (a)original spherical image; (b) intensity conspicuity map C_{int}; (c) red-green conspicuity map C_{RG}; (d)yellow-blue conspicuity map C_{YB}; (e) orientation conspicuity map C_{0^0}, (f) orientation conspicuity map C_{45^0}, (g) orientation conspicuity map C_{90^0}, (h) orientation conspicuity map C_{135^0}.
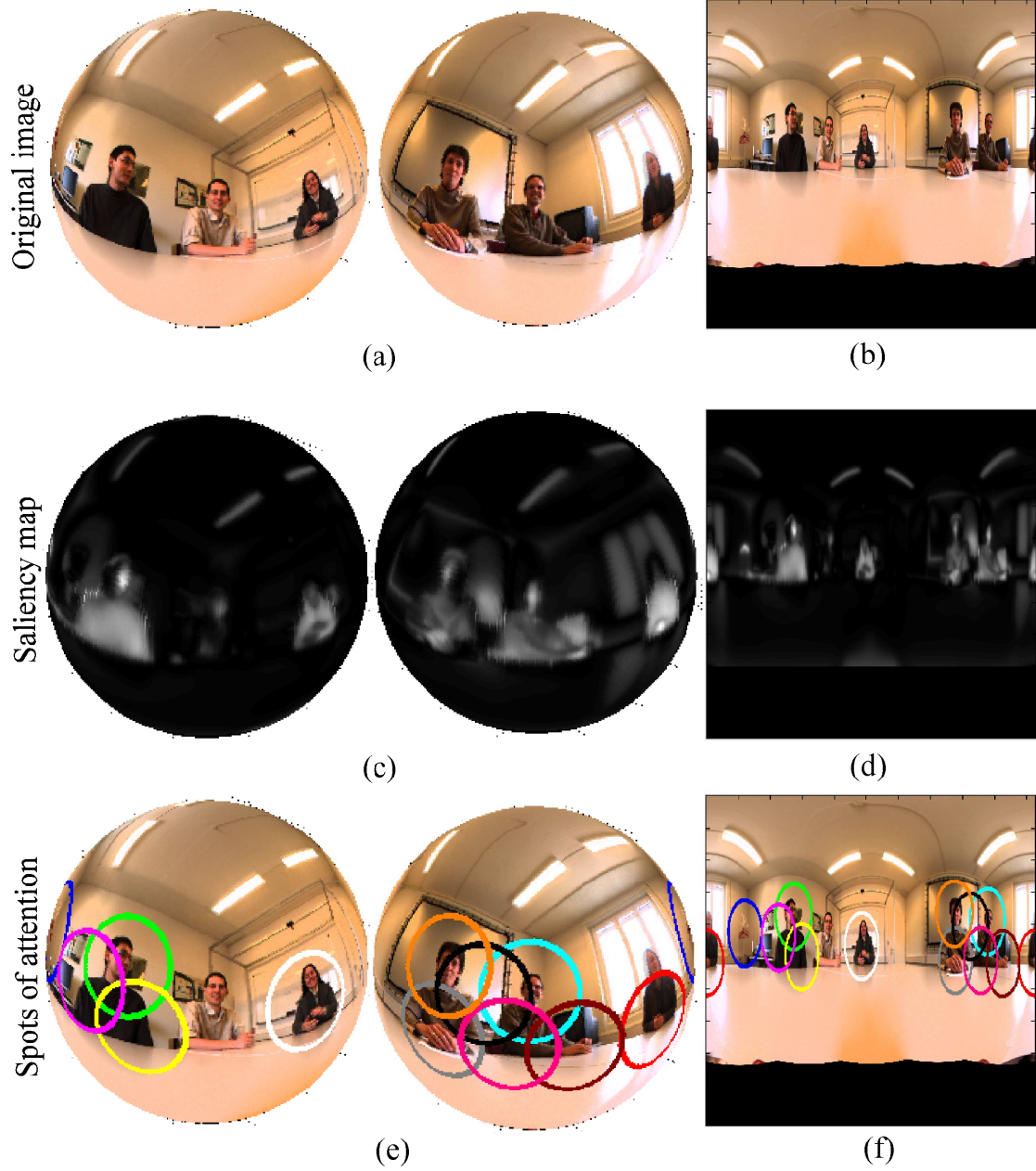
Fig. 10. Spots of attention on a real spherical image: (a) original spherical image, (b) unwrapped spherical image; (c) spherical saliency map; (d) unwrapped spherical saliency map; (e) spots of attention on the sphere; (f)unwrapped sphere with its spots of attention.
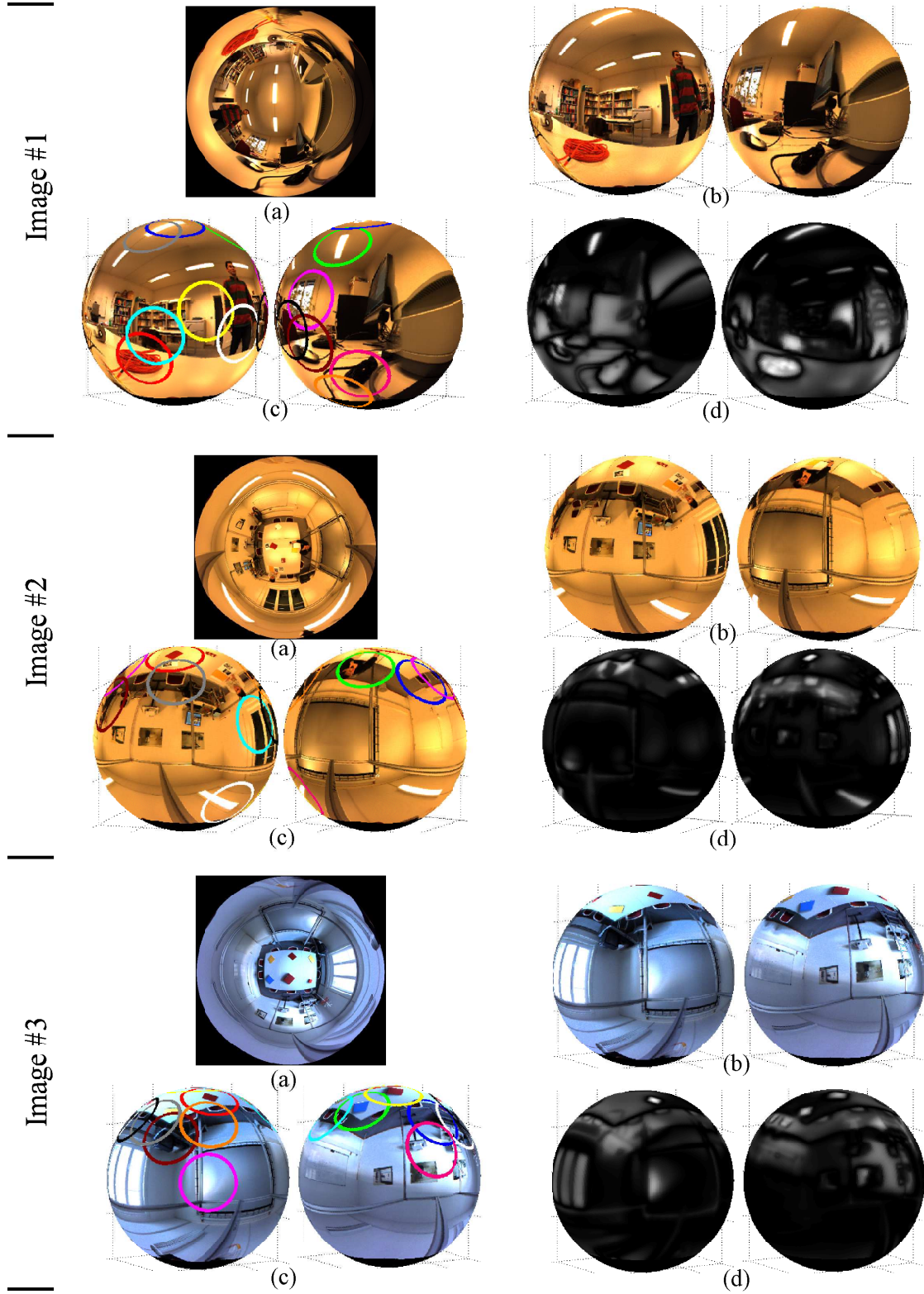
Image #1

Image #2

Image #3
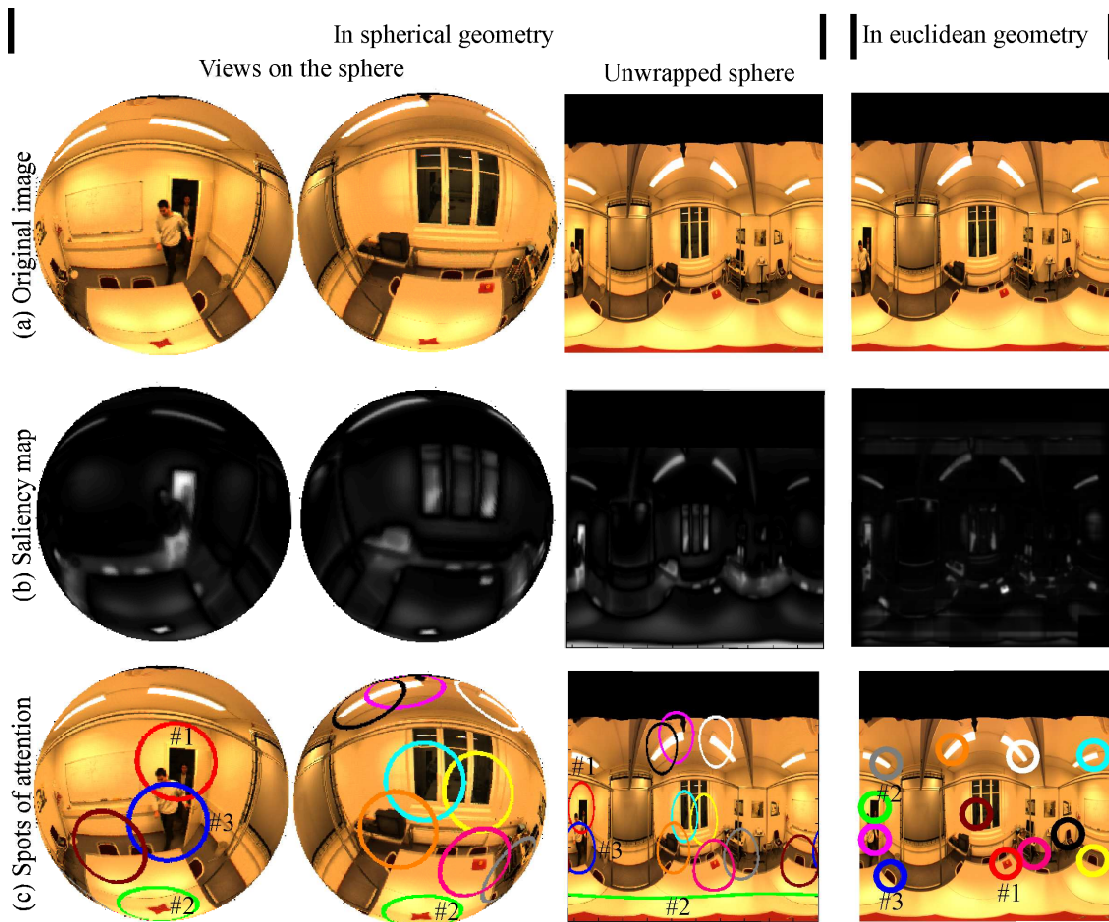
Fig. 11. Visual attention in omnidirectional scenes.

Fig. 12.   Euclidean vs. Spherical Visual Attention: (a) input signal; (b)saliency map; (c) spots of attention

Then we consider the unwrapped version as if it were a panoramic image and consequently, we apply the Euclidean VA. The Euclidean spots of attention are shown as well on Figure 12. It detects one of the red objects as most salient spot (rank 1), but does not detect the other at all.

We can see that the object situated on the pole is detected only by spherical VA. This is actually, one of the examples where the Euclidean VA would fail in correctly detecting spots of attention in omnidirectional images.

IX. ROTATION-INVARIANCE

Another experiment with the visual attention on the sphere is to check if the rotation can influence the VA on the sphere and consequently to compaire with the Euclidean case. For this purpose, a rotation by (θ, φ) on the spherical image is applied. The original image is assumed to have no rotation, i.e

$\theta = 0^0, \varphi = 0^0$ as shown on Figure 13(a). Then we consider three more posipilities: rotation only by the angle ϕ, i.e. $\theta = 0^0, \varphi = 90^0$ as shown in Figure 13(c); rotation by angle θ, i.e. $\theta = 90^0, \varphi = 0^0$ as illustrated in Figure 13(e); and rotation by both angles, i.e. $\theta = 90^0, \varphi = 90^0$. We apply the spherical VA on each of the rotated images and consider only the first 3 detected spots of attention. The spherical spots of attention with their corresponding ranking are situated in the column on the left in the same Figure. We can easily see the in all of the cases, the same objects are detected and with the same ranking. Let us now consider, the sme rotated images, as if they were Euclidean and apply the Euclidean VA. Again, we consider the first three detected spots. The obtained results for each of the cases are depicted in the right column (13(b), (d), (f), (h)). We can see, that the same objects are detected only in (b) and (d), i.e. when a rotation by φ is applied. Considering the three posibilities of rotation as in (d), (f) and (h), the detected spots of attention are never the same.

In conclusion, it is clear that the visual attention on the sphere is rotation-invariant.

CONCLUSIONS

In this paper we have defined a new visual attention algorithm for images defined on the sphere. This new algorithm operates in the spherical domain. The multiresolution analysis on the sphere is used to determing the spherical feature conspicuity maps and concequiently, the spherical cue conspicuity maps and the final spherical saliency map. The spots of attention are representing the maximums in the saliency map. Operating on the sphere, the new algorithm provides a saliency response which is homogeneous, i.e. rotation- invariant, and thus best suited for omnidirectional images. The advantage to perform homogeneously was illustrated in a comparison of results obtained by processing a synthetic spherical signal with the Euclidean VA algorithm on one hand, and with the new algorithm, on the other hand. Another comparison was performed on a real omnidirectional image obtained with a multi- camera sensor. In both comparisons it is clear that the spherical VA remedies the problem of the distortion, which is persistent in omnidirectional imaging. Another important conclusion is that objects found at sphere's poles cannot actually be detected by the Euclidean VA. In fact, the two algorithms perform in the same way only for object found at the equator. As a demonstration, the algorithm was applied with success on different omnidirectional images. Finally, the application perspectives are quite universal, as the method is basically applicable to any omnidirectional image that can be mapped onto the sphere.

In a future work, in intend to explore the dynamic visual attention on the sphere, where the motion will be considered and integrated.
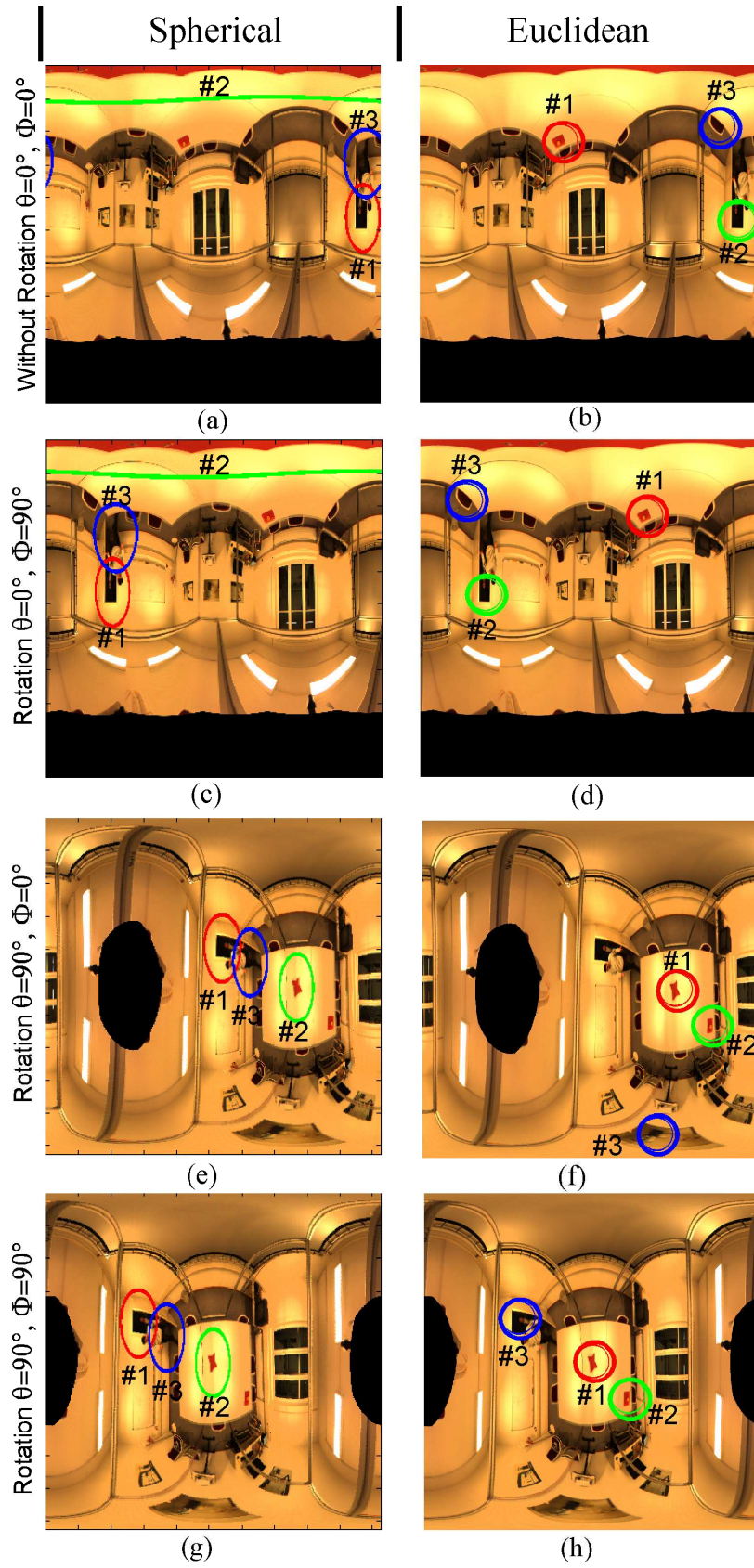
Fig. 11. Rotation invariance: Visual attention on the sphere is rotation invariant

ACKNOWLEDGMENTS

REFERENCES

[1] N. Ouerhani, *Visual Attention: from bio-inspired Modeling to Real-Time Implementation (PhD Thesis pp.42-52)*, http://www-imt.unine.ch/parlab/, 2004.

[2] L. Itti, "Automatic foveation for video compression using a neurobiological model of visual attention," *IEEE Transaction on Image Processing*, vol. 13, pp. 1304–1318, 2004.

[3] D. Walther, U. Rutishauser, C. Koch, and P. Perona, "Selective visual attention enables learning and recognition of multiple objects in cluttered scenes," in *Computer Vision and Image Understanding*, vol. 100, 2005, pp. 41–63.

[4] S. Frintrop, *VOCUS: A Visual Attention System for Object Detection and Goal-directed Search*, ser. Lecture Notes in Artificial Intelligence (LNAI). Springer, 2005.

[5] N. Ouerhani and H. Hügli, "MAPS: Multiscale attention-based presegmentation of color images," in *4th International Conference on Scale-Space theories in Computer Vision*, ser. LNCS, vol. 2695, 2003, pp. 537–549.

[6] A. Bur, A. Tapus, N. Ouerhani, R. Siegwart, and H. Hügli, "Robot navigation by panoramic vision and attention guided fetaures," in *ICPR '06: Proceedings of the 18th International Conference on Pattern Recognition*, Washington, DC, USA, 2006, pp. 695–698.

[7] C. Siagian and L. Itti, "Biologically-inspired robotics vision monte-carlo localization in the outdoor environment," in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2007.

[8] A. Treisman and G. Gelade, "A feature-integration theory of attention," *Cognitive Psychology*, vol. 12, pp. 97–136, 1980.

[9] J. Wolfe, "Guided search 2.0: A revised model of visual search," *Psychonomic Bulletin and Rewiew*, pp. 1(2):202–238, 1994.

[10] F. H. Hamker, "Modeling attention: from computational neuroscience to computer vision," in *Proc. of the 2nd International Workshop on Attention and Performance in Computational Vision (WAPCV04)*, 2004, pp. 59–66.

[11] C. Koch and S. Ullman, "Shifts in selective visual attention: Towards the underlying neural circuitry," *Human Neurobiology*, vol. 4, pp. 219–227, 1985.

[12] R. Milanese, *Detecting salient regions in an image: from biological evidence to computer implementation*, PhD thesis, University of Geneva, Switzerland, 1994.

[13] 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 (PAMI)*, vol. 20, pp. 1254–1259, 1998.

[14] L. Itti and C. Koch, "Computational modeling of visual attention," *Nature Reviews Neuroscience*, vol. 2, no. 3, pp. 194–203, 2001.

[15] O. Le Meur, and P. Le Callet, and D. Barba, and D. Thoreau, "A coherent computational approach to model bottom-up visual attention," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 5, pp. 802–817, 2006.

[16] J. Tsotsos, S. M. Culhane, W. Wai, Y. Lai, N. Davis, and F. Nuflo, "Modeling visual attention via selective tuning," *Artificial Intelligence*, vol. 78, no. 1-2.

[17] B. Olshausen, C. Anderson, and D. van Essen, "A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing information," *Journal of Neuroscience*, vol. 13, no. 11.

[18] V. Navalpakkam and L. Itti, "Modeling the influence of task on attention," *Vision Research*, vol. 45, no. 2, pp. 205–231, 2005.

[19] R. Milanese, H. Wechsler, S. Gil, J. Bost, and T. Pun, "Integration of bottom-up and top-down cues for visual attention using non-linear relaxation," in *Computer Vision and Pattern Recognition CVPR 94*, 1994, pp. 781–785.

[20] E. H. Adelson and J. Bergen, "The Plenoptic Function and the Elements of Early Vision," *Annals of Physics*, vol. 1, pp. 1–37, 1993.

[21] K. Yamazawa, Y. Yagi, and M. Yachida, "Omnidirectional imaging with hyperboloidal projection," in *Proceedings of IROS*, Jokohama, Japan, 1993.

[22] Y. Yagi, "Omnidirectionl Sensing and Its Applications," in *IEICE Trans. Inf. Syst.*, vol. 82(3), 1999, pp. 2568–578.

[23] C. Geyer and K. Daniilidis, "Catadioptric Projective Geometry," *International Journal of Computer Vision*, vol. 45(3), pp. 223–243, 2001.

[24] T. E. Boult, X. Gao, R. J. Micheals, and M. Eckmann, "Omni-directional visual surveillance." *Image Vision Comput.*, vol. 22, no. 7, pp. 515–534, 2004.

[25] H. Watanube, H. Tanabashi, Y. Satoh, Y. Niwa, and K. Yamamoto, "Event detection for a visual surveillance System Using Stereo Omni-directional System," *Knowledge-based Intelligent Information and Engineering Systems*, 2003.

[26] Y. Yagi and M. Yachida, "Real-Time Omnidirectional Image Sensors," *International Journal of Computer Vision*, vol. 58(3), pp. 173–207, 2004.

[27] E. Menegati, T. Maeda, and H. Ishiguro, "Image-based memory for robot navigation using properties of omnidirectional images," *Robotics and Autonomous Systems*, vol. 47, no. 4, pp. 251–267, 2004.

[28] A. Bur and H. Hügli, "Optimal cue combination for saliency computation: A comparison with human vision," in *Second International Work-Conference on the Interplay Between Natural and Artificial Computation (IWINAC 2007)*, ser. Lecture Notes in Computer Science, vol. LNCS 4528, 2007, pp. 109–118.

[29] I. Bogdanova, "Wavelets on non-euclidean manifolds," Ph.D. dissertation, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 2005.

[30] D. M. Healy, D. Rockmore, P. J. Kostelec, and S. S. B. Moore, "FFTs for the 2-Sphere - Implementations and Variations," vol. 9, no. 4, 2003, pp. 341–385.

[31] J. R. Driscoll and D. M. Healy", "Computing Fourier transforms and convolutions on the 2-sphere," *Adv. in Appl. Math.*, vol. 15, pp. 202–250, 1994.

[32] D.Rockmore, S. Moore, D. Healy, and P. Kostelec, *SpharmonicKit is freely available collection of C programs for doing Legendre and scalar spherical transforms developed at Dartmouth College; available at http://www.cs.dartmouth.edu/ geelong/sphere/.*

[33] YAWtb, *http://rhea.tele.ucl.ac.be/yawtb/.*

[34] J.-P. Antoine, L. Demanet, L. Jacques, and P. Vandergheynst, "Wavelets on the sphere: Implementations and approximations," *Applied and Computational Harmonic Analysis*, vol. 13, pp. 177–200, 2001.

[35] L. Demanet and P. Vandergheynst, "Gabor wavelets on the sphere," in *Proc. SPIE Wavelets X conf.*, San Diego, USA.

[36] Y. Wiaux, L. Jacques, P. Vielva, and P. Vandergheynst, "Fast directional correlation on the sphere with steerable filters," *The Astrophisical Journal*, vol. 652, pp. 820–832, 2006.

[37] J.-P. Antoine and P. Vandergheynst, "Wavelets on the n-sphere and related manifolds," *Journal of Mathematical Physics*, vol. 39, no. 8, pp. 3987–4008, 1998.

[38] LADYBUG, *http://www.ptgrey.com/products/spherical.asp.*