# MULTIRESOLUTION MOTION ESTIMATION FOR OMNIDIRECTIONAL IMAGES

*Ivana Tosic, Iva Bogdanova, Pascal Frossard and Pierre Vandergheynst*

Signal Processing Institute ITS, Ecole Polytechnique Fédérale de Lausanne (EPFL)
1015 Lausanne, Switzerland
phone: + (41) 21 693 2601, fax: + (41) 21 693 7600
email: {ivana.tosic,iva.bogdanova,pascal.frossard,pierre.vandergheynst}@epfl.ch, web: http://itswww.epfl.ch

## ABSTRACT

This paper presents a novel local motion estimation algorithm for omnidirectional images. The algorithm captures correlation between two spherical images of a scene, taken from arbitrary viewpoints, with the objective to reduce the encoding rate of these images. It first performs a multiresolution decomposition of the spherical images, in order to improve the consistency of the motion estimation, with a limited computational complexity. Then, it determines pairs of similar solid angles and matches blocks of the two omnidirectional images, directly in the spherical domain. This approach allows a simple motion estimation implementation, that avoids potential discrepancies induced while unfolding omnidirectional images to implement a classical motion estimation on images. The proposed algorithm is shown to provide a quite efficient image prediction, and the prediction error is almost exclusively composed of high frequency noise.

## 1. INTRODUCTION

Efficient representation and coding of 3-D scenes has recently gained a lot of attention from the research community, fostered by the development of emerging applications in exploration, movie production, virtual reality or even surveillance. While most of the work in this area is focusing on image-based rendering methods, this paper proposes to address the representation of the plenoptic function directly in the spherical domain, under the assumption of perfect vision sensors. This choice presents the main advantage of avoiding the potential discrepancies due to Euclidean approximations in image-based rendering.

In the proposed framework, several omnidirectional cameras capture a static 3-D scene, from arbitrary viewpoints. Each of these cameras outputs an omnidirectional image that can be mapped on a sphere, through inverse stereographic projection [1, 2]. However, the output images from multiple cameras are obviously correlated, and a rate efficient representation of the overall 3-D scene first requires the removal of redundancy between the different views. This paper proposes a local motion estimation algorithm, that captures the correlation between omnidirectional images taken from arbitrary viewpoints. The choice of local motion estimation, as opposed to global rotation estimation used in computer vision [3,4], is driven by the perspective of an efficient coding of the plenoptic function. The proposed algorithm is built on a multiresolution representation of spherical images, in order to provide a consistent motion field, even with images captured at very different viewpoints. The multiresolution coarse-to-fine motion estimation method used for classical images [5] has been adapted to the spherical framework, in order to report similarities between solid angles, instead of common blocks of pixels. The multiresolution motion estimation is shown to provide a very efficient prediction of spherical images, and the residual error is kept small and concentrated in high frequencies.

The paper is organized as follows. Section 2 overviews the framework used in this work, and the omnidirectional camera setup. Section 3 describes the multiresolution analysis for spherical images, that is used in the motion estimation algorithm. Section 4 presents the local motion estimation algorithm for omnidirectional images, and Section 5 shows experimental results.

## 2. GEOMETRY OF OMNIDIRECTIONAL IMAGES

### 2.1 Omnidirectional Imaging System

The system for obtaining omnidirectional images, in our case, is a typical parabolic catadioptric sensor. It is realized as a parabolic mirror which is placed in front of a camera approximating an orthographically projecting lens as depicted on Figure 1. In such a case, the ray of light incident with the focus of the parabola is reflected to a ray of light parallel to the parabola's axis. This construction is equivalent to a purely rotating perspective camera.

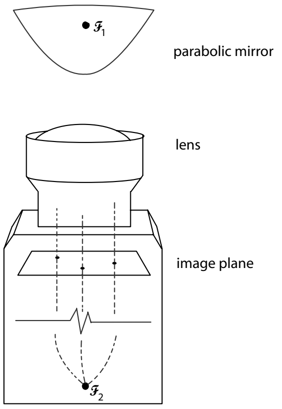

Figure 1: Omnidirectional system with parabolic mirror: the parabolic mirror is placed at parabolic focus $\mathscr{F}_1$, while the other focus $\mathscr{F}_2$ is at infinity [1].

### 2.2 Mapping of the Omnidirectional Image on the Sphere

The entire information seen by the observer can be described with the plenoptic function [6] which gives the intensity distribution of the pencil of light rays incident to the observer. Obviously, the most natural representation of this distribution is in the spherical coordinate system. Working in the natural coordinates of the observer has many advantages. It allows for directly estimating the position or direction of objects in the sensor's environment. Many Computer Vision algorithms also take advantage of geometric invariance such as, for example, the relative orientation of the sensor and objects in the scene. Thus, our goal is to recover the spherical coordinates, $\theta \in [0, \pi]$ and $\varphi \in [0, 2\pi)$, of incoming rays of light at the parabola focus $\mathscr{F}_1$, which locates our ideal observer.

It was shown in [1] that there is an equivalence between any central catadioptric projection and a composition of two conformal mappings on the sphere. In order to see how an omnidirectional image is mapped on the sphere, we first consider a cross-section

This work has been supported by the Swiss National Science Foundation, under grants PP-002-68737 and 200021-101880/1.

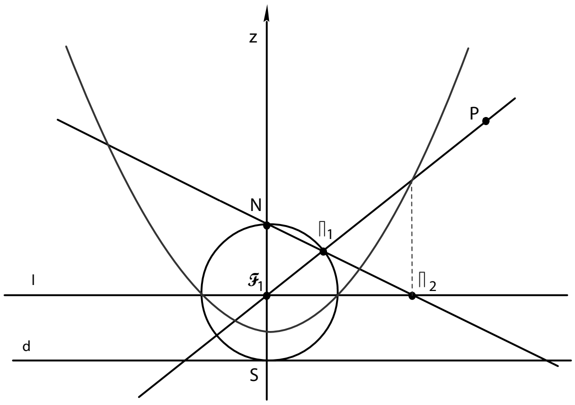

Figure 2: Cross-section of mapping the omnidirectional image on the sphere [1]

of the paraboloid. This is shown on Figure 2 . All points on the parabola are equidistant to the focus $\mathscr{F}_1$ and the directrix $d$. Let $l$ pass through $\mathscr{F}_1$ and be perpendicular to the parabolic axis. If a circle has center $\mathscr{F}_1$ and radius equal to twice the focal length of the paraboloid, then the circle and parabola intersect twice the line $l$ and the directrix is tangent to the circle. The North Pole $N$ of the circle is the point diametrically opposite to the intersection of the circle and the directrix. Point $P$ is projected on the circle from its center, which gives $\Pi_1$. This is equivalent to a projective representation, where the projective space (set of rays) is represented as a circle here. One easily sees that $\Pi_2$ is the stereographic projection of the point $\Pi_1$ to the line $l$ from the North pole $N$, where $\Pi_1$ is the intersection of the ray $\mathscr{F}_1 P$ and the circle. We can thus conclude that the parabolic projection of a point $P$ yields point $\Pi_2$ which is collinear with $\Pi_1$ and $N$. Extending this reasoning to three dimensions, the projection by a parabolic mirror is equivalent to projection on the sphere ($\Pi_1$) followed by stereographic projection ($\Pi_2$). We can thus recover the spherical coordinates of incoming light rays through a simple inverse stereographic projection of the sensor images.

Similar mapping schemes can be derived for different system constructions (with hyperbolic or elliptic mirror), by employing the inverse stereographic projection from a point specified by the chosen construction, as explained in [1].

## 3. DYADIC MULTIRESOLUTION ON $S^2$

### 3.1 Sampling and Filtering

In this section, we introduce a dyadic multiresolution representation of omnidirectional images. In the following, we will model these signals by elements of the Hilbert space of square-integrable functions on the two-dimensional sphere $L^2(S^2, d\mu)$, where $d\mu(\theta,\varphi) = d\cos\theta d\varphi$ is the rotation invariant Lebesgue measure on the sphere. These functions are characterized by their Fourier coefficients $\hat{f}(m,n)$, defined through spherical harmonics expansion :

$$\hat{f}(m,n) = \int_{S^2} d\mu(\theta,\varphi) Y^*_{m,n}(\theta,\varphi) F(\theta,\varphi),$$

where $Y^*_{m,n}$ is the complex conjugate of the spherical harmonic of order (m,n). Multiresolution is an efficient tool that allows to decompose a signal at progressive resolutions and perform coarse to fine computations on the data. The two most successful embodiments of this paradigm are the various wavelet decompositions [7] and the Laplacian Pyramid (LP) [8]. In this section, we will adapt the latter scheme to omnidirectional images.

Our omnidirectional images are mapped to spherical coordinates according to Section 2.2 and re-sampled on an equi-angular grid:

$$\mathscr{G}_j = \{(\theta_{jp}, \varphi_{jq}) \in S^2 : \theta_{jp} = \tfrac{(2p+1)\pi}{4B_j}, \varphi_{jq} = \tfrac{q\pi}{B_j}\}, \qquad (1)$$

$p,q \in \mathscr{N}_j \equiv \{n \in \mathbb{N} : n < 2B_j\}$ and for some range of bandwidth $B = \{B_j \in 2\mathbb{N}, j \in \mathbb{Z}\}$. These grids allow us to perfectly sample any band-limited function $F \in L^2(S^2)$ of bandwidth $B_j$, i.e., such that $\hat{f}(m,n) = 0$ for all $m > B_j$. Moreover, this class of sampling grids is associated to a Fast Spherical Fourier Transform [9].

### 3.2 Spherical Laplacian Pyramid

The first step in our algorithm consists in low pass filtering the data, an operation we perform in the Fourier domain for speeding up the computations. We use an axisymmetric low-pass filter defined by its Fourier coefficients :

$$\hat{h}_{\sigma_0}(m) = e^{-\sigma_0^2 m^2}. \qquad (2)$$

Suppose the original data $F_0$ is bandlimited, i.e, $\hat{f}_0(m,n) = 0$, $\forall m > B_0$, and sampled on $\mathscr{G}_0$. The bandwidth parameter $\sigma_0$ is chosen so that the filter is numerically close to a perfect half-band filter $\hat{H}_{\sigma_0}(m) = 0$, $\forall m > B_0/2$. The low pass filtered data is then downsampled on the nested sub-grid $\mathscr{G}_1$, which gives the low-pass channel of our pyramid $F_1$. The high-pass channel of the pyramid is computed as usual, that is by first upsampling $F_1$ on the finer grid $\mathscr{G}_0$, low-pass filtering it with $H_{\sigma_0}$ and taking the difference with $F_0$. Coarser resolutions are computed by iterating this algorithm on the low-pass channel $F_l$ and scaling the filter bandwidth accordingly, i.e., $\sigma_l = 2^l \sigma_0$.

It should be noted that we used the LP for ease of implementation, but any other multiresolution representation could be used. For example one could compute successive low resolution image approximations by hard thresholding in a spherical wavelet frame [10].

## 4. MULTIRESOLUTION MOTION ESTIMATION ALGORITHM

Due to the distortion introduced in the unwrapped images, we choose to implement the local motion estimation algorithm directly in the spherical domain. The algorithm is based on a L-level multiresolution approach, that pairs solid angles from two spherical images (see Figure 3). Assume that the motion estimation aims at computing a prediction $\widetilde{G_0}$ of the spherical image $G_0$ from $F_0$, that is an image of the same scene, but captured from a different (arbitrary) viewpoint. Both spherical images are first filtered and downsampled, to generate a multiresolution representation of the scene, as described before. The multiresolution approach clearly limits the complexity of the motion estimation, and improves the consistency of the motion field.

The local motion estimation performs as follows. The lowest resolution spherical image $G_{L-1}$ is divided into uniform solid angles, of size $M\delta_\theta^{L-1} \times N\delta_\varphi^{L-1}$. The predicted blocks $g^i_{L-1}$ in $G_{L-1}$ are then paired with the best matching blocks with the same size in the reference image $F_{L-1}$, within a search window of $S\delta_\theta^{L-1} \times S\delta_\varphi^{L-1}$, around the location of the $g^i_{L-1}$. A full search for each block $g^i_{L-1}$ determines the best predictors in a MSE sense, $f^i_{L-1}$, and the corresponding motion vectors. Note that, even if the blocks $g^i_{L-1}$ are all distinct, the blocks $f^i_{L-1}$ may be overlapping. The implemented block-matching algorithm also takes into account the periodicity in the azimuthal direction.

The motion estimation is then iteratively refined at successive resolution levels. The blocks at resolution $l$, $b^i_l$, are divided into four sub-blocks of size $M^{l-1}\delta_\theta^{l-1} \times N^{l-1}\delta_\varphi^{l-1}$ at the next resolution level $l-1$, with $2\delta_\theta^{l-1} = \delta_\theta^l$ and $2\delta_\varphi^{l-1} = \delta_\varphi^l$, due to the change in the resolution level. The motion vectors from the lower resolution level $l$ serve as initial estimations of the motion vectors of the four sub-blocks corresponding to the block $b^i_l$. These estimations are then refined based on the spherical images at resolution $l-1$, with a full search in a window of size $S^{l-1}\delta_\theta^{l-1} \times S^{l-1}\delta_\varphi^{l-1}$ around the location specified by the motion vector from the lower resolution $l$,

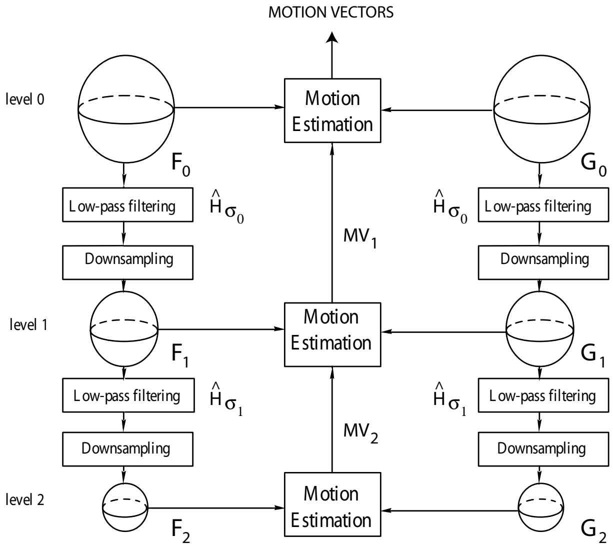

Figure 3: Algorithm for local motion estimation of spherical images.

that has been up-sampled accordingly. The same process is applied iteratively up to the finest resolution, and eventually outputs the field of motion vectors. These motion vectors, along with the spherical image $f_0$, are used to form the prediction $\widetilde{G_0}$ of $G_0$. The prediction error is finally denoted $E_0 = G_0 - \widetilde{G_0}$.

**Algorithm 1** Multiresolution local motion estimation

```
$l = L-1$. $MV_L^i = [0,0], \forall i$, $\delta_\theta^0 = \frac{\pi}{2B_0}$, $\delta_\varphi^0 = \frac{2\pi}{2B_0}$, $B_0$=full resolution
repeat
    $\delta_\theta^l = 2^l \delta_\theta^0$. $\delta_\varphi^l = 2^l \delta_\varphi^0$;
    Divide $G_l$ into $I$ uniform blocks of size $M^l \delta_\theta^l \times N^l \delta_\varphi^l$;
    $i = 0$;
    repeat
        $(p_i, q_i) \leftarrow$ position of $g_l^i$;
        $MV_l^i \leftarrow$ up-sample $MV_{l+1}^i$;
        $\Omega \leftarrow \{(p,q)\}$ such that
        $p \in [p_i + MV_l^i(1) - \frac{S^l \delta_\theta^l}{2} + 1, p_i + MV_l^i(1) + \frac{S^l \delta_\theta^l}{2}]$ and
        $q \in [q_i + MV_l^i(2) - \frac{S^l \delta_\varphi^l}{2} + 1, q_i + MV_l^i(2) + \frac{S^l \delta_\varphi^l}{2}]\}$;
        $f_l^i = arg\,min_\Omega MSE(g_l^i, f_l^i)$;
        $(s_i, t_i) \leftarrow$ position of $f_l^i$;
        $MV_l^i \leftarrow [p_i + s_i, q_i + t_i]$;
        $i \leftarrow i + 1$;
    until $i > I$
    $l \leftarrow l - 1$;
until $l < 0$
```

## 5. EXPERIMENTAL RESULTS

This section presents the results of the local motion estimation algorithm proposed above. Figure 4 shows one spherical image at the second finest resolution level. Figures 5 and 6 show the original spherical images of a static scene captured from two different viewpoints. These images represent real spherical images, but they are shown here as planar images in the $(\theta, \varphi)$ plane, to provide visibility of all image features. Figure 7 represents the prediction $\widetilde{G_0}$ of the second frame, with the local motion estimation algorithm, and Figure 8 shows the corresponding prediction error $E_0$, that has been inverted to highlight the distribution of the residual error (a white pixel corresponds to no error). The number of decomposition levels is $L = 5$. The size of the blocks has been set to $4 \times 4$. The size of the search window can vary from one resolution level to another. We have chosen the window size for the lowest level to be $32 \times 32$ and for all higher levels $8 \times 8$. This way, the proposed algorithm can capture big motions with low search complexity. It can be seen that the motion estimation is quite efficient, since the predicted image provides a very good approximation of $G_0$. Also, the prediction error is almost exclusively located along high frequency components, as expected from the high-pass characteristics of motion estimation.

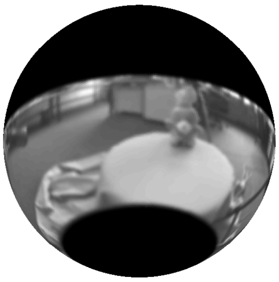

Figure 4: Spherical image displayed on the sphere, level l=1

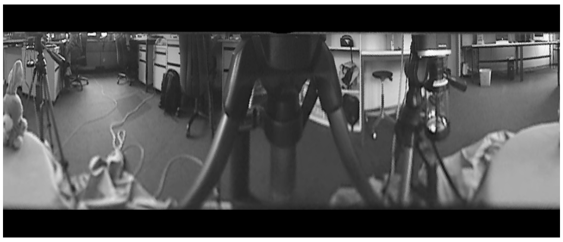

Figure 5: First original spherical image, $F_0$.

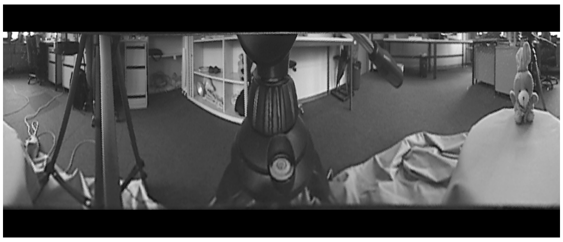

Figure 6: Second original spherical image, $G_0$.

Figure 9 represents the motion field that corresponds to the $3^d$ level of resolution. It can be seen that the motion field is mostly consistent with the spherical image information. For example, motion vectors are very small in uniform and static areas like the table (on the right-hand side of the predicted image). As expected from a local motion estimation algorithm driven by MSE criteria, the motion vectors do not however necessarily follow semantic objects, but rather pair areas with similar luminance information. This behavior can be encountered for large motions where the change in lightning

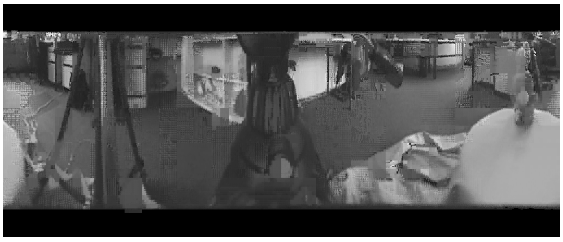

Figure 7: Motion predicted image, $\widetilde{G_0}$.

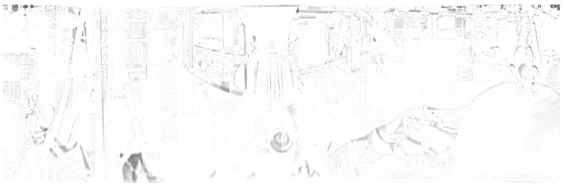

Figure 8: Motion prediction error, $E_0$.

conditions can induce discrepancies. On the other side, the obtained motion field precisely depicts smaller movements.

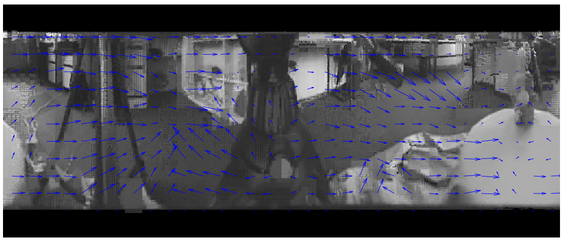

Figure 9: Motion field at resolution level 3.

Figure 10 presents the evolution of the residual energy relative to the original image energy, as a function of the size of the solid angle, and the size of the search window. It can be seen that a larger search window at the coarsest resolution level generally improves the quality of the motion estimation. Moreover, smaller block size provides a better prediction, since details can be better approximated. In a coding perspective however, a trade-off needs to be found between the accuracy of the motion estimation, and the coding cost, which generally increases with the number of motion vectors.

## 6. CONCLUSIONS

In this paper a local motion estimation algorithm has been presented, that captures the correlation between omnidirectional images taken from arbitrary viewpoints. A multiresolution approach has been proposed to improve the motion filed accuracy, while limiting the computational complexity of the motion estimation scheme. The local motion estimation algorithm has been shown to be quite efficient since the residual error is kept very small and mostly located around edges or high frequency components in the predicted image. The proposed scheme can certainly represent an important building block in a rate-distortion efficient encoder for distributed omnidirectional images.

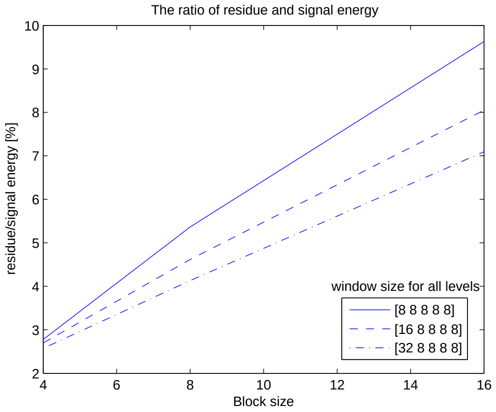

Figure 10: Relative energy of the prediction error, for different block and search window sizes ($L = 5$).

## REFERENCES

[1] C. Geyer and K. Daniilidis, “Catadioptric projective geometry,” *International Journal of Computer Vision*, vol. 45, no. 3, pp. 223 – 243, December 2001.

[2] C. Geyer and K. Daniilidis, “Paracatadioptric camera calibration,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, May 2002.

[3] L. S. A. Makadia and K. Daniilidis, “Rotation estimation from spherical images,” in *Proceedings of the 17th International Conference on Pattern Recognition*, vol. 3, 2004, pp. 590–593.

[4] A. Makadia and K. Daniilidis, “Direct 3d-rotation estimation from spherical images via a generalized shift theorem,” in *Proceedings of the IEEE Computer Vision Pattern Recognition*, vol. 2, 2003, pp. 217–224.

[5] G. Conklin and S. Hemami, “Multi-resolution motion estimation,” in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1997.

[6] E.H. Adelson and J.R. Bergen, *Computational Models of Visual Processing*. M. Landy and J.A. Movshon, eds., MIT Press, Cambridge, 2001, pp. 3 – 20.

[7] S. Mallat, *A Wavelet Tour of Signal Processing*. Academic Press, 1998.

[8] P. J. Burt and E. H. Adelson, “The laplacian pyramid as a compact image code,” *IEEE Transactions on Communications*, vol. COM-31,4, pp. 532–540, 1983. [Online]. Available: citeseer.ist.psu.edu/burt83laplacian.html

[9] D. Healy Jr. and D. Rockmore and P. Kostelec and S. Moore, “Ffts for the 2-sphere - improvements and variations,” *Journal of Fourier Analysis and Applications*, vol. 9, no. 3, pp. 341 – 385, 2003.

[10] I. Bogdanova, P. Vandergheynst, J.-P. Antoine, L. Jacques and M. Morvidone, “Discrete wavelet frames on the sphere,” in *In proceedings of EUSIPCO 2004, Vienne, Austria*, EUSIPCO. EUSIPCO, September 2004.