LOW BIT-RATE COMPRESSION OF OMNIDIRECTIONAL IMAGES

Ivana Tošić and Pascal Frossard

Signal Processing Laboratory (LTS4)
Ecole Polytechnique Fédérale de Lausanne (EPFL)
{ivana.tosic, pascal.frossard}@epfl.ch

ABSTRACT

Omnidirectional images represent a special type of images that are captured by vision sensors with a 360-degree field of view. This work targets the compression of such images by taking into account their particular geometry. We first map omnidirectional images to spherical ones and then perform sparse image decomposition over a dictionary of geometric atoms on the 2D sphere. A coder based on Matching Pursuit and adaptive quantization is finally proposed for efficient compression of omnidirectional images. The experiments demonstrate that the proposed system outperforms JPEG2000 coding of unfolded images. Since most omnidirectional sensors can be parametrized with a spherical camera model, the proposed method is generic with respect to different sensor constructions.

Index Terms— omnidirectional images, compression, sparse

1. INTRODUCTION

Although omnidirectional cameras have found wide range of applications lately, there has been little work done on compression of the acquired images. The projective geometry of omnidirectional cameras, such as catadioptric or fish-eye cameras, has a specific nature since lines in the 3D space project into curves on the omnidirectional image. This leads to a clear conclusion that applying standard 2D planar image compression techniques to omnidirectional images is suboptimal. To perform proper compression of such images, we have to design spatial correlation reduction methods for their specific projective space. For catadioptric systems for example, the projective surface is the mirror manifold. However, designing compression methods for each mirror type results in a specific codec for each camera type. Fortunately, Geyer et al. [1] have proposed a unifying framework where omnidirectional images obtained from catadioptric systems with different mirrors can be uniquely mapped to spherical images. This allows us to represent the light field directly in the spherical domain and design a coding method on the sphere which can be used for different types of mirrors. Two main approaches for representing signals on the sphere have dominated in the past: the Spherical harmonics transform [2]; and the spherical wavelet transform SWT [3]. Spherical harmonics have global spatial support, which makes them unsuitable for local analysis on the sphere. SWT has spatially localized and mutli-scale spherical basis functions, but it fails to represent multi-dimensional singularities like contours in omnidirectional images. Another multi-resolution representation of spherical signals is the Laplacian Pyramid on the sphere [4], which is not as efficient for compression as wavelets due to its redundant nature.

This work has been supported by the Swiss National Science Foundation under grant 200020-120063.

To the best of our knowledge, the only work that considered the geometry of catadioptric systems for omnidirectional image and video compression has been presented by Bauermann et al. [5]. They proposed a preprocessing step that maps each image in an omnidirectional video into a panoramic image and then compresses that video using the standard H.264 coder. However, panoramic mapping cannot fully exploit the correlation statistics of the radial light field captured by an omnidirectional camera. This is because the sampling density of light rays is non-uniform on a panoramic cylinder, while the pixel distribution on the cylinder is uniform.

We propose to represent a spherical image as a series of oriented and anisotropically refined functions taken from a redundant dictionary of atoms on the sphere. These atoms are edge-like functions living on the 2D sphere, which can take arbitrary positions, shapes and orientations. Since the number of atoms in the redundant dictionary is usually much higher than the dimension of the signal, there is a high probability that a given signal is well approximated with few atoms. Such signal representations are called sparse, as the selected atoms are sparse in the dictionary. Due to compact signal representation, redundant expansions can give good compression performance. Hence, we exploit the proposed geometric dictionary on the sphere to design a novel omnidirectional image coder, based on the Matching Pursuit (MP) algorithm. An encoder is proposed that compresses the stream of atoms by adaptive coefficient quantization. Experimental results demonstrate that the new coder outperforms in rate-distortion the standard JPEG2000 coder on unfolded images for 2dB in average, at low rates. It also significantly outperforms the SPIHT-encoded Laplacian Pyramid on the sphere.

2. GEOMETRY OF OMNIDIRECTIONAL IMAGES

Various omnidirectional cameras exist on the market today, and almost all of them can be considered as spherical cameras using an appropriate mapping. We will consider a typical parabolic catadioptric sensor, which is realized with a parabolic mirror placed in front of a camera approximating an orthographically projecting lens, as shown in Fig. 1a). In this construction, the ray of light incident with the focus of the parabola is reflected to a ray of light parallel to the parabola's axis. The entire information seen by the catadioptric system can be described with the intensity distribution of the pencil of light rays incident to the focal point of the mirror. Obviously, the most natural representation of this distribution is in the spherical coordinate system. It has been shown in [1] that there is an equivalence between any central catadioptric projection and a composition of two conformal mappings on the sphere. First mapping is a projective representation, where the projective space is a sphere centered at the focal point of the mirror. The second mapping is the stereographic projection from the pole of the sphere to the catadioptric

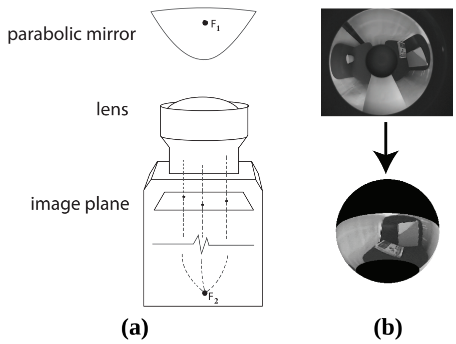


Fig. 1. a) Omnidirectional system with parabolic mirror [1]; b) Mapping from image plane to the spherical image.

plane (image sensor plane). We can thus recover the spherical coordinates of incoming light rays through a simple inverse stereographic projection of the sensor images, as shown in Fig. 1b).

Similar mapping schemes can be derived for different system constructions (with hyperbolic or elliptic mirrors), by employing the inverse stereographic projection from a point specified by the chosen construction [1]. Therefore, designing a compression scheme for spherical images leads to a generic coding method for omnidirectional images obtained with various catadioptric devices.

3. SPARSE APPROXIMATIONS ON THE SPHERE

3.1. Redundant expansions

Given an overcomplete dictionary $\mathcal{D} = \{\phi_k\}$ in a Hilbert space H, every signal $y \in H$ can be represented as a linear combination of atoms ϕ_k. However, since the dictionary is over-complete, there are infinitely many possible linear representations of the image. In order to find a compact image approximation one has to search for a linear expansion that contains a small number of components. In other words, we look for a *sparse* representation of y in $\mathcal{D}$, which is a linear combination of a small number of atoms in $\mathcal{D}$ up to an approximation error η, i.e., $y = \sum_{k\in I} \phi_k c_k + \eta$ where I labels the set of atoms $\{\phi_k\}_{k\in I}$ participating in the representation. Unfortunately, finding the sparsest representation has combinatorial complexity. However, there exist polynomial time algorithms, such as Matching Pursuit (MP), that search for a suboptimal solution. MP is an iterative algorithm, which selects at each iteration the atom that best matches the signal and removes its contribution from the signal to form the residue. It then continues the same procedure on the residue until it becomes sufficiently small.

Redundant expansions have shown interesting approximation properties in the decomposition of signals with multidimensional singularities such as contours in natural images. They provide a lot of freedom in the design of the bases or dictionaries. In particular, it is possible to design structured dictionaries which include rotation or anisotropy in the basis functions. These two properties are keys to the development of efficient algorithms for image approximation.

3.2. Redundant dictionary on the 2-D sphere

We propose to decompose spherical signals as a series of atoms, taken from a redundant dictionary of functions defined on the 2D sphere. Dictionaries are in general constructed as a set of different waveforms, where each waveform is defined by a generating function. Each generating function can serve as a base for building the overcomplete dictionary, simply by changing the function parameters or indexes (e.g., position or scale indexes). While there is a priori no restriction on the construction of the dictionary, the usage of generating functions advantageously leads to structured and parametric dictionaries, whose indexes directly correspond to atom characteristics.

The construction of the dictionary on the 2D sphere is mostly based on the dictionary presented in [6] for 3D surfaces, with some modifications in order to account for statistics of the spherical images. The dictionary design involves the three following steps:

- definition of the generating function(s) on the sphere,
- anisotropic scaling of atoms
- definition of the motion of atoms on the sphere, and rotation around their axis.

Since the signal to be approximated is defined in the space of square-integrable functions on a unit 2-sphere S^2, the atoms have obviously to live in the same space. Let g denote a generating function on the 2D sphere. By combining motion, rotation and scaling, we form an overcomplete set of atoms $\phi \equiv g_\gamma$, where $\gamma = (\tau, \nu, \psi, \alpha, \beta) \in \Gamma$ is the atom index. This index is described by five parameters that respectively represent the position of the atom on the sphere: τ along the zenith angle θ and ν along the azimuth angle φ, its orientation ψ, and the scaling parameters (α, β).

Under the assumption that spherical images are mostly composed of smooth surfaces, and singularities aligned on pieces of great circles, we propose to build the dictionary over two generating functions. First, in order to efficiently capture the singularities, we use a generating function that resembles to a piece of contour on the sphere. Our choice is a spherical function which is a Gaussian function in one direction and its second derivative in the orthogonal direction. Scales α and β additionally perform anisotropic refinement of the generating function in two orthogonal directions, to finally obtain the edge-like atom function:

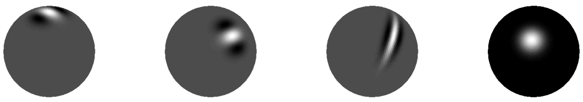
$$\begin{aligned} g_{HF}(\theta,\varphi) &= \frac{1}{K_1}\left(16\alpha^2\tan^2\frac{\theta}{2}\cos^2\varphi - 2\right) \\ &\cdot \exp\left(-4\tan^2\frac{\theta}{2}\left(\alpha^2\cos^2\varphi + \beta^2\sin^2\varphi\right)\right) \quad (1) \end{aligned}$$

where K_1 is a normalization factor. The motivation for the choice of a Gaussian kernel lies in its optimal joint spatial and frequency localization. On the other side, the second derivative in the orthogonal direction is used to filter out the smooth polynomial parts of the signal and capture the signal discontinuities.

Second, in order to represent efficiently the smooth areas in the spherical signals corresponding to low-frequency (LF) components, we propose to use a second generating function for the construction of the dictionary. The second function is a two-dimensional Gaussian function on the sphere, anisotropicaly scaled:

$$g_{LF}(\theta,\varphi) = \frac{1}{K_1}\exp\left(-4\tan^2\frac{\theta}{2}\left(\alpha^2\cos^2\varphi + \beta^2\sin^2\varphi\right)\right). \quad (2)$$

The functions g_{HF} and g_{LF} define atoms that are centered exactly on the North pole. We further form the redundant dictionary by applying geometric transforms to these functions, on the 2D sphere, i.e., we apply different anisotropic scales α and β, and move the generating functions on the sphere. Motion and rotation belong to the group of affine transformations of the unit 2D sphere S^2. They are both realized by one transform composed of three rotations by



(a) **(b)** **(c)** **(d)**

Fig. 2. a-c) Samples of HF atoms; d) LF atom.

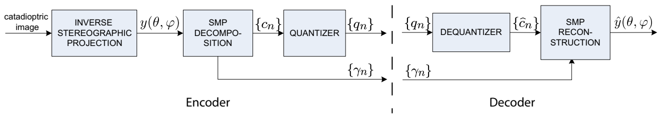


Fig. 3. Omni-SMP coding scheme.

Euler angles (ν, τ, ψ). The generating function is therefore transformed into an atom that can be moved to a particular point (τ,ν) on the sphere, and rotated. By a proper choice of the transformation parameters, one finally obtains an overcomplete dictionary of functions, which is used to represent spherical signals. Sample atoms are illustrated in Fig. 2.

4. OMNIDIRECTIONAL IMAGE CODER

Our objective is to build on the effective approximation properties offered by redundant expansions to obtain compressed versions of omnidirectional images. The block diagram of the proposed coder, called Omni-SMP (omnidirectional image coder based on Spherical Matching Pursuit[1]), is shown in Fig. 3. At the encoder side, an omnidirectional image is first mapped to a spherical image $y(\theta, \varphi)$. Matching Pursuit algorithm is then performed on the spherical image to select a series of atoms from the dictionary described above, with their relative coefficients. Atoms are first sorted along the decreasing magnitude of their coefficients. The coefficients are then uniformly quantized with a decaying quantization range, based on the method proposed by Frossard et al. [7]. This method takes advantage of the property that the energy of MP coefficients is limited by an exponentially decaying upper-bound. The decoder first performs inverse quantization of MP coefficients and then reconstructs the approximated spherical image by linear combination of atoms whose relative weights are given by the MP coefficients.

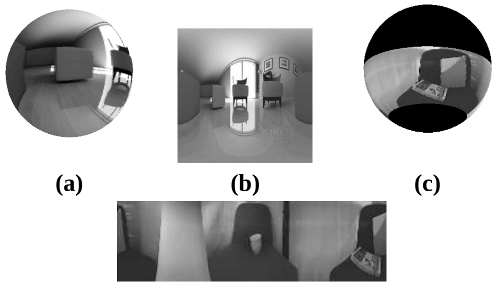
5. EXPERIMENTAL RESULTS

5.1. Implementation

In the dictionary presented in Section 3.2, the atom indexes obviously take discrete values. We first use the equiangular spherical grid to drive the values of the position parameters, τ and ν; both parameters are uniformly distributed on the interval $[0, \pi]$, and $[-\pi, \pi)$, respectively, with a resolution that is identical to the input image. The rotation parameter ψ is uniformly sampled on the interval $[0, \pi]$, with 16 orientations. This choice is mostly due to the use of the fast correlation on the SO(3) group within the Matching Pursuit algorithm. Finally, scaling parameters are distributed in a logarithmic manner, from 1 to half of the resolution of the input signal, with a granularity of one third of octave. To reduce the dictionary size and hence the complexity, we limit the scale β to be $\beta \leq \alpha$. This constraint gives a dictionary with atoms elongated along edges, which efficiently represent image structures. For low-frequency atoms, the maximal value for the scaling parameters is chosen to be $1/16$ of the signal resolution. Motion and rotation parameters are discretized in the same way as for HF atoms.

5.2. Compression performance evaluation

The compression performance of the proposed Omni-SMP coder has been evaluated on synthetic spherical and natural omnidirectional images. The original spherical image of the synthetic scene Room, rendered by the ray tracer engine[2], is shown in Fig. 4(a). Its unfolded version is also shown in Fig. 4(b) in order to display all image features. Natural Lab image has been captured by a paracatadioptric camera with a mirror from the Remote Reality Corporation[3], which has 360 degrees view in the azimuth angle and the zenith view which ranges from 35 to 92.5 degrees. It has been mapped to a spherical image in Fig. 4(c), whose unfolded version is shown in Fig. 4(d). Both test spherical images have resolution 256x256 pixels on the equiangular spherical grid.



(a) **(b)** **(c)**

(d)

Fig. 4. Original images, resolution 256x256: (a) spherical image Room; (b) unfolded image (a); (c) omnidirectional image Lab mapped to a spherical image; (d) unfolded image (c).

The rate-distortion (RD) performance of Omni-SMP has been evaluated, reporting the number of bits per pixel of the image needed to achieve the image quality measured by the PSNR (Peak Signal to Noise Ratio). PSNR is a logarithmic measure of the mean square error, evaluated as the mean of the spherical l_2 norm of the error. For the natural Lab image PSNR has been evaluated only on the non-black (informative) part of the sphere. Fig. 5 and Fig. 6 present the RD performance of the Omni-SMP coder for Room and Lab images, respectively. In the same figures, we plot the RD performance of the JPEG2000 coder on unfolded images, and the performance of a multiresolutional method that employs the Spherical Laplacian pyramid (SLP) [4], followed by the SPIHT coding of LP coefficients. Since there are no spherical wavelet-based methods adapted to the compression of omnidirectional images, but only to shape compression, we compare to JPEG2000, which is a wavelet based coder for planar images and currently state-of-the-art method in image coding. Performing compression on unfolded spherical images using the planar image coder JPEG2000 is very similar to projecting images to panoramic images and then applying JPEG2000. To have a

[1]Spherical Matching Pursuit refers to the Matching Pursuit of signals on the sphere, using a dictionary of spherical atoms.

[2]http://yafray.org/

[3]http://www.remotereality.com/

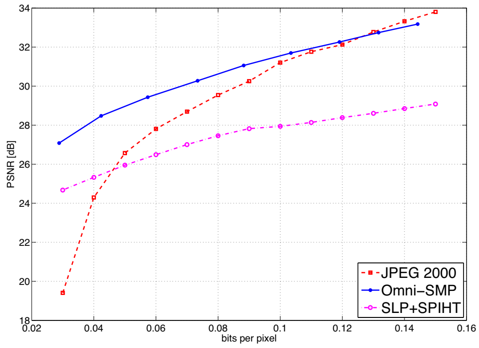


Fig. 5. Rate-distortion performance for the Room image.

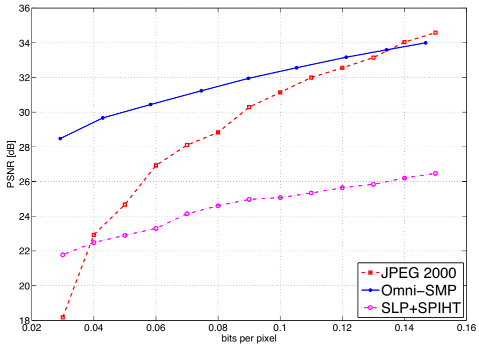


Fig. 6. Rate-distortion performance for the Lab image.

fair comparison, the mean square error for JPEG2000 is evaluated on the sphere, as for the Omni-SMP. We can see that the proposed Omni-SMP coder outperforms JPEG2000 for up to 6 dB at low rates. However, Omni-SMP is not efficient at high rates since the designed dictionary is not optimized to approximate texture information dominant at high rates, but rather structural image information. Due to the redundancy of the Laplacian Pyramid, the RD performance of this method is much worse with respect to Omni-SMP and JPEG2000.

Finally, we observe the visual image quality of the proposed Omni-SMP coder and compare it to JPEG2000. Fig. 7 (a) and (b) show respectively the decoded images using Omni-SMP and JPEG2000 coder, at the same bit rate 0.057bpp. We can see that Omni-SMP decoded image is more visually pleasing, with sharper edges and smoother flat regions. Moreover, we can see how the atoms in the image decomposition fit to 3D lines that are projected to curvatures on the sphere. On the other side, JPEG2000 introduces artifacts that degrade the structure of the image. At higher rates, decoded images with Omni-SMP and JPEG2000 become closer in PSNR value, but the coding artifacts are less annoying for Omni-SMP, as shown in Fig. 7 (c) and (d) depicting the decoded Room images at bit rate 0.088bpp. Similar observations can be made for the Lab image (see Fig. 8).

6. CONCLUSIONS

We have presented a new compression method for omnidirectional images, based on spherical mapping. As many omnidirectional images can be represented in the spherical domain, the new method

a) Omni-SMP 0.057 bpp 29.43 dB | b) JPEG2000 0.057 bpp 27.56 dB | c) Omni-SMP 0.088 bpp 31.06 dB | d) JPEG2000 0.088 bpp 30.25 dB

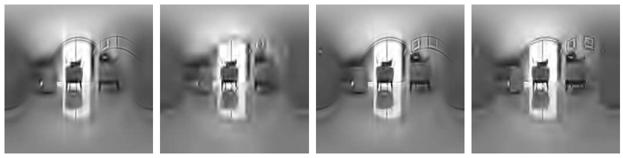


Fig. 7. Decoded Room images

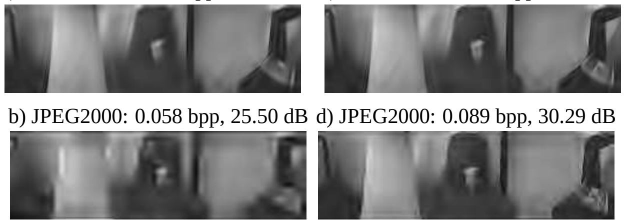


Fig. 8. Decoded Lab images

is quite generic and can be used for different camera constructions. The spherical images are decomposed over a redundant dictionary of multi-dimensional atoms on the 2D sphere, which efficiently approximate the geometry- specific curved discontinuities in the image. The proposed encoder employs the Matching Pursuit algorithm with adaptive quantization of coefficients. The new method outperforms state-of-the-art JPEG 2000 coder on unfolded (panoramic) images, at low bit rates where proper coding of the image geometry is typically more important than texture coding.

7. 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] J. R. Driscoll and D. M. Healy, "Computing fourier transform and convolutions on the 2-sphere," *Advances in Applied Mathematics*, vol. 15, no. 2, pp. 202–250, June 1994.

[3] J. Antoine and P. Vandergheynst, "Wavelets on the 2-sphere : a group theoretical approach," *Applied and Computational Harmonic Analysis*, vol. 7, no. 3, pp. 262–291, November 1999.

[4] I. Tošić, I. Bogdanova, P. Frossard and P. Vandergheynst, "Multiresolution Motion Estimation for Omnidirectional Images," in *Proceedings of the European Signal Processing Conference*, September 2005.

[5] I. Bauermann, M. Mielke, and E. Steinbach, "H. 264 based coding of omnidirectional video," in *Proceedings of International Conference on Computer Vision and Graphics*, September 2004.

[6] I. Tošić, P. Frossard and P. Vandergheynst, "Progressive Coding of 3-D Objects Based on Overcomplete Decompositions," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 11, pp. 1338–1349, November 2006.

[7] P. Frossard, P. Vandergheynst, R.M. Figueras i Ventura and M. Kunt, "A posteriori quantization of progressive matching pursuit streams," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 525–535, February 2004.