PLENOPTIC BASED SUPER-RESOLUTION FOR OMNIDIRECTIONAL IMAGE SEQUENCES

Luigi Bagnato[†]*, Yannick Boursier*[*]*, Pascal Frossard*[†]*, Pierre Vandergheynst*[†]

[†] Ecole Polytechnique Fédérale de Lausanne (EPFL), Signal Processing Laboratory, Lausanne, Switzerland

[*] Aix-Marseille Université, CPPM, CNRS/IN2P3, Marseille, France

ABSTRACT

This paper addresses the reconstruction of high resolution omnidirectional images from a low resolution video acquired by an omnidirectional camera moving in a static scene. In order to exploit the additional information provided by the side images in the video sequence, the ego-motion of the camera must be accurately estimated in a first step. The reconstruction can then be modeled as a plenoptic sampling problem that has to encompass the change of viewpoint between each position of the omnidirectional sensor and the specific discretization of the real scene observed from each position. We formulate this problem as an ill-posed inverse problem that incorporates a regularization term based on a Total Variation (TV) prior. A graph variational formulation is used in order to ease the representation of omnidirectional data and to adapt the discretization of differential operators to the omnidirectional geometry. Experimental results on synthetic images demonstrate the relevance of this approach and its superiority compared to standard super-resolution using a single image.

Index Terms— Omnidirectional, Graph, Total Variation, Super-Resolution, Plenoptic

1. INTRODUCTION

Super-resolution usually describes the problem of reconstructing high quality images from multiple images of lower resolutions, that are typically taken at different instants in time or from slightly different viewpoints. Several efficient solutions of the super-resolution problem have been proposed for images from perspective cameras [1, 2, 3]. In these cases, the related ill-posed inverse problem is formulated as a minimization regularized problem in which the regularization term is based on a Total Variation (TV) prior or a Tikhonov prior (using a classical ℓ_2 norm). This a priori information significantly improves the performance of reconstruction and increases the numerical stability of the solutions.

With the advent of new imaging systems, it becomes crucial to deal with the super-resolution issue for these specific images. Among them, omnidirectional cameras have the advantage to present a wider field of view than perspective cameras, but this often comes at a price of a lower resolution, and especially a lower angular resolution, with most of the common sensors. Super-resolution techniques can improve the quality of the image sequences, by exploiting the correlation between successive images. However, the problem of super-resolution for omnidirectional sensors has not been widely studied.

Algorithms used in super-resolution of perspective images have been applied to omnidirectional images [4], but without exploiting their true omnidirectional geometry. More recently, the specific characteristics of omnidirectional images have been considered in [5], where improved quality is obtained by an iterative projection solution with multiple images captured by pre-defined rotation of the camera around its main axis. The problem of joint registration and super-resolution for omnidirectional images captured with arbitrary rotation is addressed in a spherical framework in [6].

We address here the problem of super-resolution from a sequence of spherical low-resolution images acquired by a single omnidirectional camera moving in a static scene. Since the camera motion and the structure of the scene are unknown, we estimate them as described in [7]. We then use the plenoptic geometry of the scene to perform a registration step between successive frames of the video sequence, and we exploit all the visual information for the generation of a high resolution spherical image. One of the main advantages of the proposed framework is the flexibility, since the reconstruction of a high resolution image is not constrained to a specific position in space. We use a graph-based representation to cope with the irregular sampling of the 2-Sphere, naturally induced by the geometry of the omnidirectional sensors, and we formulate the high resolution image reconstruction as an inverse problem based on total variation regularization. Experimental results show that the proposed algorithm is able to increase the quality of omnidirectional images, both in terms of PSNR and visual quality.

2. SUPER-RESOLUTION FRAMEWORK

In the following we will work with a sequence of low resolution omnidirectional images l_i, $i \in \{1, 2, \cdots, N\}$, acquired at different spatial positions $\mathbf{x}_i$ in a static scene. Since we assume that the scene does not change over time, we interpret the low resolution image sequence as a spatial sampling of the 5-D *plenoptic function* [8] $\mathcal{L}(\mathbf{x}, \omega)$ describing the light ray intensity at the 3D position $\mathbf{x}$ when looking in direction ω.We choose the 2-sphere S^2 as spatial domain to process the omnidirectional images, for at least two good reasons:

1. It comes natural if we use the plenoptic function: if we fix the spatial position $\mathbf{x} = \mathbf{x}_i$, then the function $\mathcal{L}(\mathbf{x}_i, \omega)$ is defined on the 2-Sphere.
2. For a single effective viewpoint camera there is a one-to-one mapping of the image plane onto a sphere (see [9]).

For a catadioptric camera, one of the most common omnidirectional vision systems, the projection of the catadrioptric plane onto a sphere is given via inverse stereographic projection [10]:

$$\theta = 2\tan^{-1}(2r),\ \phi = \phi$$

where $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi[$ are respectively the zenith angle and the azimuthal angle in a spherical coordinate system, and r is the distance in the image plane from the principal point (we assume for simplicity the focal to be one). For each pixel in the image plane

we can then associate the direction $\omega = (\theta, \phi) \in S^2$ of the corresponding light ray leaving the camera center.

To simplify the notation, but without loss of generality, we will assume in the following that all images l_i have the same angular sampling of the plenoptic function, such that to all of them correspond the same set of directions $\Omega_o = \{\omega_k : k = 1, 2, \cdots, M_o\}$.We can write then: $l_i = \mathcal{L}(\mathbf{x}_i, \Omega_o)$ for $i \in \{1, 2, \cdots, N\}$.

2.1. Light Rays Registration

We want to relate now, by simple geometrical considerations, a directions ω_k at the viewpoint $\mathbf{x}_j$ with the corresponding direction $\tilde{\omega}_k$ at a generic point in space $\mathbf{x}$, as depicted in Figure 1. We fix a common system of coordinates. Let us assume now that from $\mathbf{x}_j$, looking in direction ω_k, we observe a point $\mathbf{p} = D_j \omega_k$ in the scene, where D_j is the distance of the point $\mathbf{p}$ from $\mathbf{x}_j$. Please note that, with an abuse of notation, we use ω_k to represent both the 3D unit vector and the direction on S^2. If $\mathbf{p}$ is visible from position $\mathbf{x}$, since the translation vector between the two viewpoints is simply given by $\mathbf{t}_j = \mathbf{x}_j - \mathbf{x}$, the following relationship holds:

$$\tilde{\omega}_k = \frac{\mathbf{p} - \mathbf{t}_j}{\|\mathbf{p} - \mathbf{t}_j\|}. \tag{1}$$

If the light conditions change slowly over time we can use the brightness consistency equation to relate different samples of the plenoptic function:

$$\mathcal{L}(\mathbf{x}, \tilde{\omega}_k) = \mathcal{L}(\mathbf{x}_j, \omega_k). \tag{2}$$

It should be clear from Eq. (2) that, if we know $\mathbf{p}$ and $\mathbf{t}_j$, we can obtain a new plenoptic sample $\mathcal{L}(\mathbf{x}, \Omega_l)$ on the set of directions

$$\Omega_l = \bigcup_j \{\tilde{\omega}_k\}^j,$$

where we use $\{\tilde{\omega}_k\}^j$ to represent all the new directions $\tilde{\omega}_k$ at position $\mathbf{x}$ obtained from position $\mathbf{x}_j$ and the set of directions Ω_o using Eq. (1).

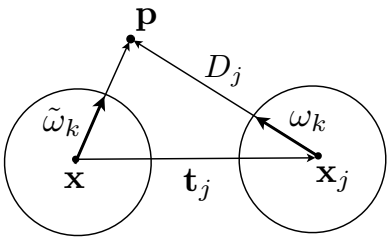


Fig. 1. Light ray geometry between position $\mathbf{x}$ and $\mathbf{x}_j$

2.2. Discrete Embedding of the Sphere in a Graph

Before we further proceed to the formulation of a variational inverse problem, we briefly discuss how we embed the spherical discrete set of data into a graph structure. The main advantage of the graph-based representation is that we can directly define stable differential operators on a discrete domain. Although the 2-sphere is a simple manifold with constant curvature and a simple topology, a naive discretization can cause severe numerical instabilities (see for example [11]). On top of that we have extreme flexibility on the choice of the discrete grid points, so we are not limited to regular grids. A weighted undirected graph $\Gamma = (V, E, w)$ consists of a set of vertices V, a set of vertices pairs $E \subseteq V \times V$, and a weight function $w : E \mapsto \mathbb{R}$ satisfying $w(u, v) > 0$ and $w(u, v) = w(v, u)$, $\forall (u, v) \in E$. Each direction ω can be interpreted as a point on a spherical imaging surface, so we use vertices to represent directions ω in space (i.e. points on the unitary sphere), while edges define connections between directions. The topology of the spherical surface is finally obtained through the definition of weights $w(u, v)$ as a decreasing function of the geodesic distance $g(u, v)$ between the vertices u and v: $w(u, v) = e^{-g^2/\sigma}$. Following Zhou et al [12], we define the gradient and divergence over Γ as :

$$(\nabla^w f)(u, v) = \sqrt{\frac{w(u, v)}{d(u)}} f(u) - \sqrt{\frac{w(u, v)}{d(v)}} f(v) \tag{3}$$

and

$$(div^w F)(u) = \sum_{u \sim v} \sqrt{\frac{w(u, v)}{d(v)}} \left(F(v, u) - F(u, v)\right), \tag{4}$$

where $u \sim v$ stands for all vertices v connected to u and $d : V \mapsto \mathbb{R}$ is the degree function defined as $d(v) = \sum_{u \sim v} w(u, v)$. We also define the local isotropic variation of F at vertex (pixel) v by:

$$\|\nabla_v^w F\| = \sqrt{\sum_{u \sim v} \left[\left(\nabla^w F\right)(u, v)\right]^2}. \tag{5}$$

2.3. Variational Problem Formulation and Solution

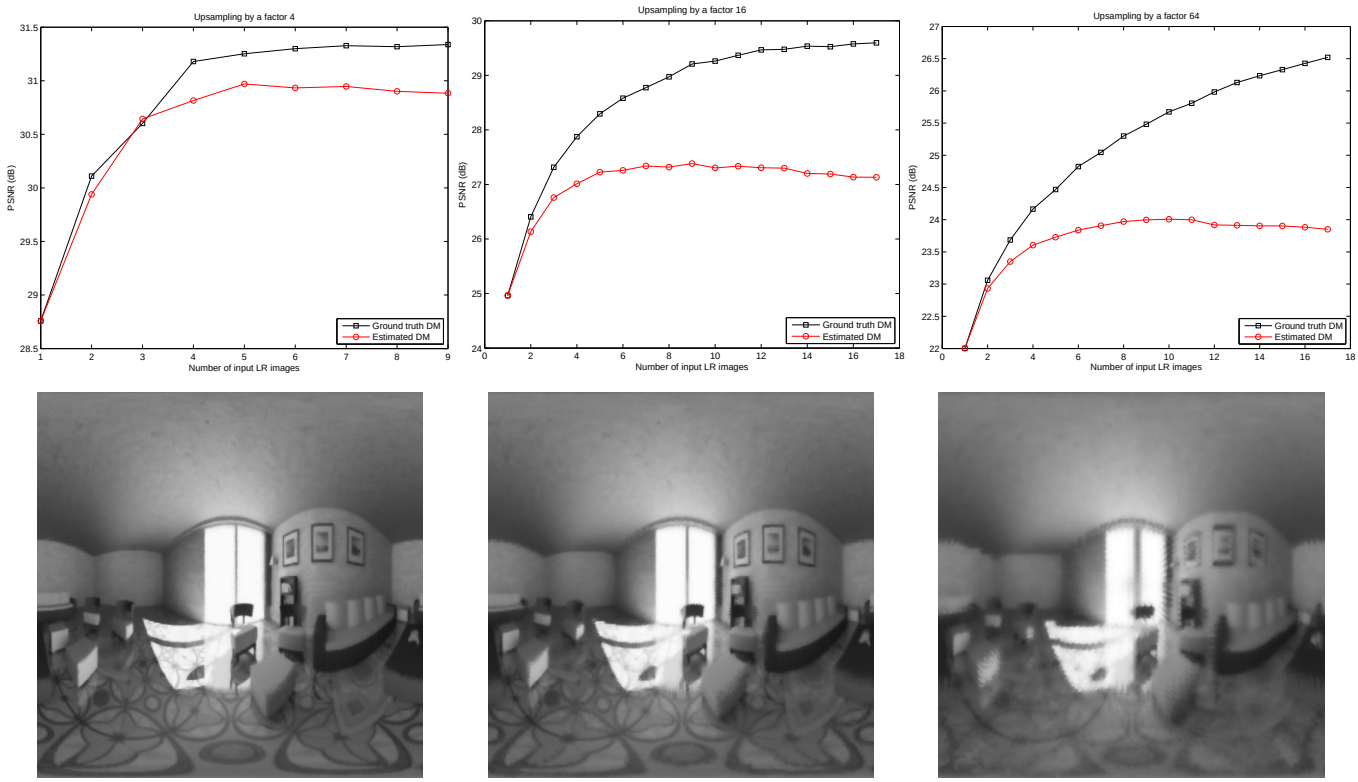
We assume from now on that we know the camera motion parameters, i.e., translation and rotation for each pair of images l_i and l_j, as well as a dense depth map estimate $D_i(\omega_k)$ for each spatial position $\mathbf{x}_i$. We obtain such estimates in a pre-processing step using the same structure from motion algorithm described in [7]. For each couple of successive frames l_i and l_j the camera ego-motion and the depth map $D_i(\omega_k)$ are jointly estimated in variational framework. Using the ego-motion parameters and the depth information, we perform an image registration step where we obtain, for a generic position $\mathbf{x}$ in space, the full set of directions Ω_l as described in Section 2.1. We want to stress that, thanks to the framework flexibility, $\mathbf{x}$ does not have necessarily to coincide with one of the camera positions. We set $b = \mathcal{L}(\mathbf{x}, \Omega_l) \in \mathbb{R}^m$ where m is the cardinality of Ω_l. The super-resolution problem then consists in estimating from b a sample of the plenoptic function at a given position $\mathbf{x}$ and for a given set of directions $\Omega_h = \{\omega_h : h = 1, 2, ...M_h\}$, where typically $M_h \gg M_o$. Let us call f the full set of plenoptic function values on the set of direction $\Omega = \Omega_l \cup \Omega_h$, i.e. $f = \mathcal{L}(\mathbf{x}, \Omega) \in \mathbb{R}^n$ when n is the cardinality of Ω. We formulate then the following variational inverse problem where we look for a regularized solution f^* s.t.

$$f^* = \underset{f}{\operatorname{argmin}} \|b - \Phi f\|^2 + \lambda \sum_v \|\nabla_v^w f\|, \tag{6}$$

where Φ is a matrix defined as:

$$\Phi = \mathbb{I}_{\Omega_l} \in \mathbb{R}^{m \times n} \tag{7}$$

i.e., it is obtained from the identity matrix $\mathbb{I} \in \mathbb{R}^{n \times n}$ by keeping only the rows corresponding to directions in Ω_l. In other words Φ restricts f to the available data. Since the second energy term in Eq. (6) defines the total variation norm of f on the graph, the minimization problem defined in Eq. (6) can be seen as a TV inpainting scheme on graphs. It is out of the scope of this paper to tackle convex optimization algorithm on graphs, so we refer to the solution proposed in [13, 11].



(a) Using 256x256 (b) Using 128x128 (c) Using 64x64

Fig. 2. PSNR plots for 3 video sequences as function of the number of images used in the reconstruction. The black curve is obtained using ground truth registration. In red we show the curve obtained from estimated registration. On the bottom there are the corresponding HR image reconstructions.

3. EXPERIMENTAL RESULTS

We test our algorithm with synthetic photorealistic images. The video sequence is obtained from spherical renderings of a 3D model of a living room. Each frame is generated from a different position in space and the spatial distance between successive frames is fixed. We reconstruct a high-resolution (HR) image on an equiangular grid of 512x512 pixels, using three different low-resolution (LR) spherical video sequences each composed of 17 frames. Each video sequence has a different resolution: 64x64, 128x128 and 256x256. We compare the results with respect to a HR ground truth image, and we have also access to the registration ground truth. We use the same set of parameters for all simulations. In particular we fix the regularization parameter in Eq.(6) to $\lambda = 0.05$. This is not an optimal choice in terms of absolute results (i.e. PSNR values), since λ should vary as function of the resolution. However, the scope of the simulations is to test the efficacy of the super-resolution scheme and not the sensitivity to the regularization parameters. In Figure 2 we show the PSNR curve as function of the number of images used in the reconstruction. The black curve is obtained using registration ground truth and represents the ideal system behavior. When we use the estimated registration, the PSNR curve (in red) tends to saturate earlier, but still shows a significant improvement of up to 2dB. The difference with the ideal curve is more evident when we reconstruct the HR image from very low resolution video (e.g., 64x64): this effect is easily explained if we consider that the estimation of the depth map using only two consecutive frames is an error prone operation, and the quality of the estimate degrades at low resolutions. It is interesting to note, however, that the PSNR gap does not correspond to a degradation of the visual image quality as we can clearly observe in Figure 4. Finally a visual inspection of Figure 3 confirms that using an increasing number of frames leads to more pleasant reconstructions.

4. CONCLUSIONS

In this paper we propose a novel approach to super-resolution from omnidirectional image sequences with unknown registration. Omnidirectional camera images are naturally defined on the 2-sphere. Since classical approaches fail to apply to the spherical domain we propose a flexible embedding of the spatial domain on a graph, which handles naturally irregular sampling of the sphere. We first perform a registration step between the spherical low resolution images and a new position in space, based on a dense depth map estimate. Then we use this information to formulate a variational inverse problem based on TV regularization. Tests on synthetic images support the validity of the proposed approach that leads to consistent improvements in the reconstructed high resolution images both in terms of PSNR and visual quality.

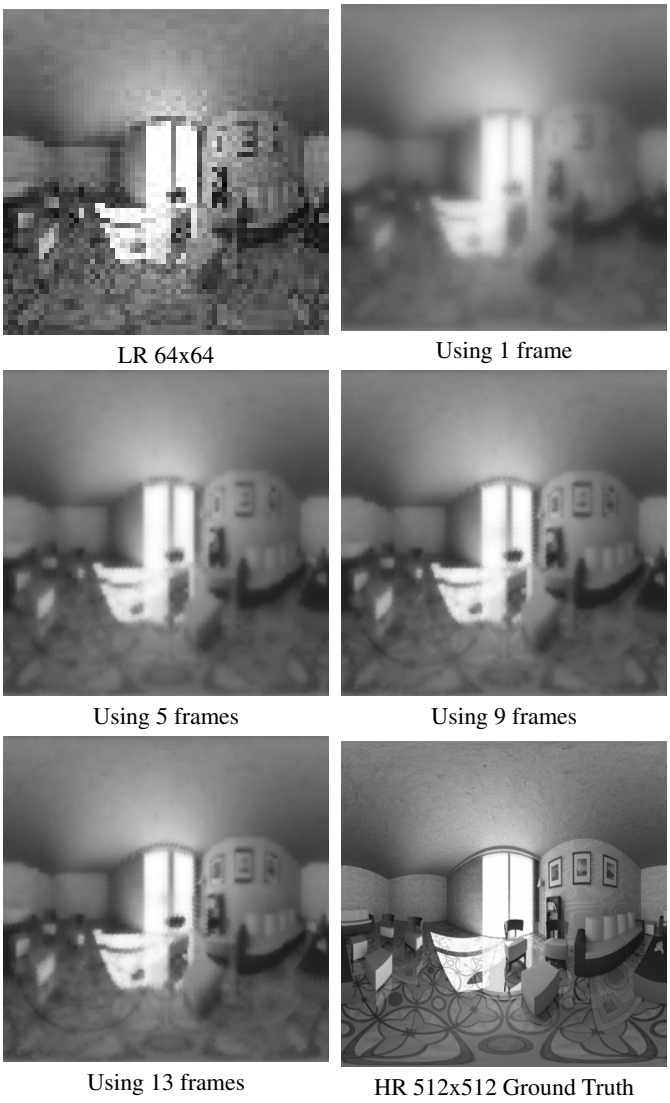


Fig. 3. Visual results using increasing number of frames (1,5,9,13) from the 64x64 video sequence. Top left: the LR image. Bottom right: the HR ground truth image.

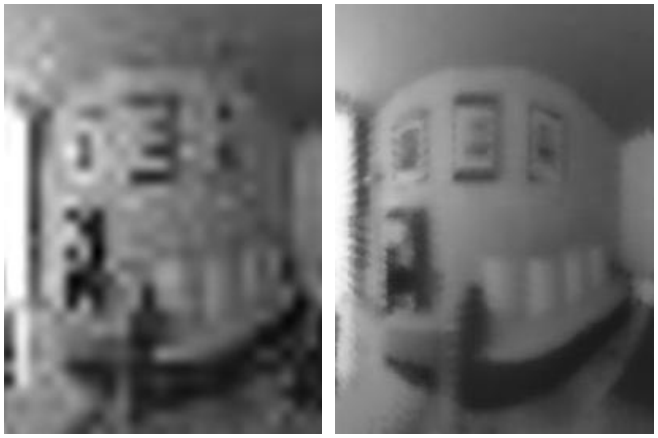


Fig. 4. Zoom on a detail: on the left the LR image (64x64) we use in the reconstruction. On the right the reconstructed HR image (512x512) using 12 images with estimated registration.

5. ACKNOWLEDGMENTS

The authors would like to thank Zafer Arican, Laurent Jacques and Ivana Tosic for the fruitful discussions.

6. REFERENCES

[1] C. A. Segall, R. Molina, and A. K. Katsaggelos, "High-resolution images from low-resolution compressed video," *IEEE Signal Processing Magazine*, vol. 20, pp. 37–48, 2003.

[2] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Advances and challenges in super-resolution," *International Journal of Imaging Systems and Technology*, vol. 14, pp. 47–57, 2004.

[3] A. Marquina and S. J. Osher, "Image super-resolution by tv-regularization and bregman iteration," *J. Sci. Comput.*, vol. 37, no. 3, pp. 367–382, 2008.

[4] H. Nagahara, Y. Yagi, and M. Yachida, "Super-resolution from an omnidirectional image sequence," *Industrial Electronics Society, 26th Annual Conference of the IEEE*, vol. 4, 2000.

[5] Z. Fan and Z. Qi-dan, "Super-resolution image reconstruction for omni-vision based on pocs," *Control and Decision Conference*, pp. 5045–5049, 2009.

[6] Z. Arican and P. Frossard, "l1 regularized super-resolution from unregistered omnidirectional images," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2009.

[7] L. Bagnato, P. Vandergheynst, and P. Frossard, "A Variational Framework for Structure from Motion in Omnidirectional Image Sequences," *Submitted to IEEE Transactions on Image Processing*, 2009.

[8] E. H. Adelson and J. R. Bergen, "The plenoptic function and the elements of early vision," in *Computational Models of Visual Processing*. 1991, pp. 3–20, MIT Press.

[9] S. Baker and S.K. Nayar, "A theory of single-viewpoint catadioptric image formation," *Int J Comput Vis*, vol. 35, no. 2, pp. 175–196, Jan 1999.

[10] C. Geyer and K. Daniilidis, "Catadioptric projective geometry," *Int J Comput Vis*, vol. 45, no. 3, pp. 223–243, Jan 2001.

[11] L. Bagnato, P. Frossard, and P. Vandergheynst, "Optical flow and depth from motion for omnidirectional images using a TV-L1 variational framework on graphs," in *Proceedings of ICIP*. IEEE, 2009.

[12] D. Zhou and B. Scholkopf, "A regularization framework for learning from graph data," *ICML Workshop on Statistical Relational Learning and Its . . .*, Jan 2004.

[13] G. Peyre, S. Bougleux, and L. Cohen, "Non-local regularization of inverse problems," in *Computer Vision-Eccv 2008*, Jan 2008.