

TV-REGULARIZED GENERATION OF PLANAR IMAGES FROM OMNICAMS

Y. Boursier¹, L. Jacques^{1,2}, D. Raboud¹, P. Frossard¹, M.J. Fadili³, P. Vanderghenst¹

¹Institute of Electrical Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

²Communications and Remote Sensing Laboratory, Université catholique de Louvain (UCL), B-1348 Louvain-la-Neuve, Belgium

³GREYC CNRS-ENSICAEN-Université de Caen, 14050 Caen France

ABSTRACT

This paper addresses the problem of mapping images between different vision sensors. Such a mapping could be modeled as a sampling problem that has to encompass the change of geometry between the two sensors and the specific discretization of the real scene observed by the two different imaging systems. We formulate the problem in a general framework that can be cast as a minimization regularized problem with a linear operator, that applies to any image geometry. We then focus on the particular problem of the generation of planar images from omnidirectional images, in any viewing direction and for any size and resolution. In this regularized approach, the fidelity term is expressed in the original omniscam geometry and the regularization is based on Total Variation (TV) solved here with proximal methods. Experimental results demonstrate the superiority of this approach with respect to alternative schemes based on linear interpolation or TV inpainting.

Index Terms— Sampling, Inverse Problem, Superresolution, Omnidirectional camera, Total Variation Optimization.

1. INTRODUCTION

With the advent of new imaging systems, it becomes important to define appropriate mappings between images from different sensing devices for proper processing and rendering of the light information. In particular, a variety of omnidirectional imaging systems have emerged recently. They offer better models in terms of accuracy and efficiency in the representation of scenes. Interestingly, recent works have provided image processing methods adapted to the particular geometry of these omnidirectional sensors [1, 2, 3]. Most applications require however high quality rendering of planar images in well chosen directions in the scene. It therefore becomes crucial to correctly aggregate the information provided by both omnidirectional and planar sensors. We propose in this paper a novel framework for mapping full or partial images between sensors characterized by different geometries and different intrinsic parameters.

Y. B. is a Postdoctoral Researcher funded by the APIDIS European Project. L. J. is a Postdoctoral Researcher of the Belgian National Science Foundation (F.R.S.-FNRS).

2. FRAMEWORK

In this paper, we denote by \mathbf{x}_c the hypothetical continuous “vision” that a given camera c (planar or omnidirectional) perceived from a 3-D scene. It is a function of the spatial continuous coordinates $\vec{x} = (x, y) \in \mathbb{R}^2$ in the image plane, i.e. it provides the *illumination* (color or gray scale) $\mathbf{x}_c(\vec{x}) \in \mathbb{R}$ on this point. Our basic assumption is that \mathbf{x}_c belongs to some subspace $\mathcal{X} \subset L^2(\mathbb{R}^2)$. For instance, \mathcal{X} can be the set of bandlimited elements of $L^2(\mathbb{R}^2)$ (for some given 2-D cut-off frequencies), or the set of piecewise continuous functions, i.e. $\mathcal{C}^0(\mathbb{R}^2)$.

The actual image \mathbf{x} provided by any camera is a discretization of the continuous view \mathbf{x}_c restricted to the camera Field of View (FOV) on a grid of N points. Mathematically, a sampling operator $\mathbf{E} : \mathcal{X}_{\mathcal{F}} \rightarrow \mathbb{R}^N$ is applied on \mathbf{x}_c , i.e. $\mathbf{x} = \mathbf{E} \mathbf{x}_c$, where $\mathcal{X}_{\mathcal{F}}$ is the restriction of \mathcal{X} to the FOV $\mathcal{F} \subset \mathbb{R}^2$. This sampling depends on the characteristics of the camera (type and parameters) and the N sampling points are lying on grid $\mathcal{G} = \{\vec{x}_i = (x_i, y_i) : 1 \leq i \leq N\}$.

The adjoint operator of \mathbf{E} , noted \mathbf{E}^* , performs the reconstruction of \mathbf{x}_c from its discrete representation \mathbf{x} . For instance, \mathbf{E}^* could be any interpolation method relying on some a priori information available on \mathbf{x}_c . In other words,

$$[\mathbf{E}^* \mathbf{x}](\vec{x}) = \sum_i \mathbf{x}(\vec{x}_i) \phi_{\vec{x}_i}(\vec{x}), \quad (1)$$

for some interpolating functions $\{\phi_{\vec{x}_i} \in \mathcal{X}\}$. For instance, if \mathcal{X} is the set of bandlimited functions, and if the grid \mathcal{G} is regular, then $\phi_{\vec{x}_i}(\vec{x})$ is the translation on \vec{x}_i of a mother function $\phi(\vec{x}) = \phi(x)\phi(y)$, where $\phi(x)$ is *sinc* function. In this paper, we select a Gaussian kernel ϕ for reasons that will become clear in the sequel.

Let us now consider two (discrete) images \mathbf{x}_1 and \mathbf{x}_2 provided by two different cameras, which can be pan-tilt-zoom (PTZ), catadioptric, fisheye lens camera or any other imaging system, with two different resolutions N_1 and N_2 and FOVs \mathcal{F}_1 and \mathcal{F}_2 . The operator \mathbf{A} allows to deduce the \mathbf{x}_2 image from the \mathbf{x}_1 image such that

$$\mathbf{x}_2 = \mathbf{A} \mathbf{x}_1. \quad (2)$$

The operator \mathbf{A} is in fact a linear operator of $\mathbb{R}^{N_2 \times N_1}$

since it can be expressed as the combination of three successive operators defined above in the following way:

$$\mathbf{A} = \mathbf{E}_2 \mathbf{\Pi} \mathbf{E}_1^*, \quad (3)$$

where $\mathbf{\Pi} : \mathcal{X}_{\mathcal{F}_1} \rightarrow \mathcal{X}_{\mathcal{F}_2}$ expresses the change of geometry between the two sensors. This operators make our approach flexible and well suited to the processing of images from any geometry to any other: catadioptric, fisheyed, planar or any distorted images as soon as the distortion rules are known.

In Section 3, we tackle the specific problem of generating planar images from fisheye ones. We apply this model to an optimization problem in the context of superresolution, where $N_1 > N_2$ typically, and with a prior of sparsity on the gradients of images which naturally leads to TV-regularized formulation. Simulation results are then given in Section 4 and compared with those provided by linear interpolation and inpainting alternative schemes.

3. RECONSTRUCTION OF PLANAR IMAGES FROM FISHEYE DATA

We focus now on the problem of the generation of perspective planar images in any direction covered by a fisheye sensor. In our case, the original data comes from a perfect fisheye lens $\mathcal{S}_{\text{Fisheye}}$. Figure 1 shows an example of a fisheye image (APIDIS basket ball dataset¹).

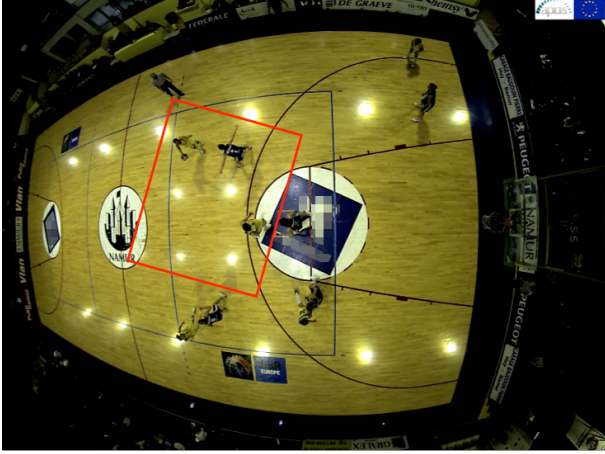


Fig. 1: Fisheye image. Courtesy APIDIS European Project

The generation of planar images has for inputs either the conventional parameters of a virtual pan-tilt-zoom camera \mathcal{S}_{PTZ} (line-of-sight, angle-of-view, resolution) or a region of interest that can be defined by spherical coordinates. Those parameters could be provided for instance in response to a tracking or event detection algorithm. They fully determine the virtual camera lying at the position of the omniscam which will produce the desired planar image.

¹<http://www.apidis.org/Dataset/>

3.1. The Forward Model

The discrete fisheye image (our source of information) is noted \mathbf{x}_2 , and the planar image (the one we wish to reconstruct) is noted \mathbf{x}_1 .

In this study, we consider that both source and target images are sampled on regular grids, i.e. \mathbf{x}_1 and \mathbf{x}_2 are defined on the grids $\mathcal{G}_1 = \{(i\delta, j\delta) : i \in \mathbb{N}_p, j \in \mathbb{N}_q\}$ and $\mathcal{G}_2 = \{(i\nu, j\nu) : i \in \mathbb{N}_m, j \in \mathbb{N}_n\}$ respectively with $\mathbb{N}_k = \{0, \dots, k-1\}$. The resolutions are thus $N_1 = pq$ and $N_2 = mn$ with $N_1 > N_2$ with pixel widths $\delta, \nu > 0$.

Therefore, without loss of generality, we can assume that \mathbf{x}_1 (\mathbf{x}_2) is linked to the continuous function $\mathbf{x}_{c_1} \in \mathcal{X}_{\mathcal{F}_1}$ (resp. $\mathbf{x}_{c_2} \in \mathcal{X}_{\mathcal{F}_2}$) of FOV $\mathcal{F}_1 = [0, (p-1)\delta] \times [0, (q-1)\delta]$ (resp. $\mathcal{F}_2 = [0, (m-1)\nu] \times [0, (n-1)\nu]$).

Let us define the 'backprojection' operator \mathbf{A} such that

$$\mathbf{x}_2 = \mathbf{A}\mathbf{x}_1 \quad \text{with} \quad \mathbf{A} = \mathbf{E}_2 \mathbf{\Pi} \mathbf{E}_1^*. \quad (4)$$

This backprojection is composed of three operators that are described as follows. First, \mathbf{E}_1^* is a reconstruction operator such that

$$\mathbf{x}_{c_1} = \mathbf{E}_1^* \mathbf{x}_1.$$

In this study, we assume that the functional space \mathcal{X} is well represented by discrete translation of Gaussians. Therefore, \mathbf{E}_1^* reads

$$\mathbf{x}_{c_1}(\vec{x}) = \sum_{i=1}^p \sum_{j=1}^q \mathbf{x}_1(i, j) \phi(x-i) \phi(y-j).$$

with $\phi(x) = \frac{1}{\sqrt{2\sigma}} \exp -\frac{x^2}{2\sigma^2}$ and with σ set to the pixel size. Other combination of spaces \mathcal{X} and kernels ϕ may be investigated to perform interpolation and ongoing studies concern functions with a compact support, in order to significantly reduce the computation time.

Second, $\mathbf{\Pi}$ drives the change of geometry between the spaces of $\mathcal{S}_{\text{Fisheye}}$ and \mathcal{S}_{PTZ} , knowing that they are observing the same 3-D world. Therefore, this operator applies a change of variables π , i.e.

$$\mathbf{x}_{c_2} = \mathbf{\Pi} \mathbf{x}_{c_1} = \mathbf{x}_{c_1} \circ \pi,$$

defined by the following three transformations: (i) an inverse central projection centered on the line of sight that remaps the planar image onto the sphere of directions, (ii) a particular spherical projection function of the fisheye camera parameters that maps the spherical domain onto the fisheye geometry.

Since we can reasonably assume the required FOV of the planar image amounts to an Angle of View (AOV) smaller than 180° , the final change of variables $\pi : \mathcal{X}_{\mathcal{F}_1} \subset \mathbb{R}^2 \rightarrow \mathcal{X}_{\mathcal{F}_2} \subset \mathbb{R}^2$ is invertible. Separating its effect on the two coordinates of $\mathcal{X}_{\mathcal{F}_1}$, we have $\pi(x, y) = (\pi_x(x, y), \pi_y(x, y))$, and $\pi^{-1}(x', y') = (\pi_x^{-1}(x', y'), \pi_y^{-1}(x', y'))$.

Finally, \mathbf{E}_2 is a sampling operator that evaluates the values of the function \mathbf{x}_{c_2} on the fisheye grid such that

$$\mathbf{x}_2 = \mathbf{E}_2 \mathbf{x}_{c_2}.$$

As the image \mathbf{x}_2 is defined on the regular grid \mathcal{G}_2 , assuming $\nu = 1$, we obtain $\mathbf{x}_2(i, j) = \mathbf{x}_{c_2}(i, j)$ for $0 \leq i \leq m - 1$ and $0 \leq j \leq n - 1$.

Given these notations, the main operator \mathbf{A} acts on the planar image \mathbf{x}_1 as $\mathbf{x}_2(i') = (\mathbf{A} \mathbf{x}_1)_{i'}$ with

$$(\mathbf{A} \mathbf{x}_1)_{i'} = \sum_{i=1}^p \sum_{j=1}^q \mathbf{x}_1(\mathbf{i}) \phi(\pi_{\mathbf{x}}^{-1}(\mathbf{i}') - i) \phi(\pi_{\mathbf{y}}^{-1}(\mathbf{i}') - j),$$

with $\mathbf{i} = (i, j)$ and $\mathbf{i}' = (i', j')$. Each entry $\mathbf{A}_{i'i}$ of the matrix $\mathbf{A} \in \mathbb{R}^{N_2 \times N_1}$ is thus given by $\phi(\pi_{\mathbf{x}}^{-1}(\mathbf{i}') - i) \phi(\pi_{\mathbf{y}}^{-1}(\mathbf{i}') - j)$.

3.2. Inverse Problem and Image Reconstruction

Reconstructing the perspective image \mathbf{x}_1 of a virtual planar camera from the observed fisheye image $\mathbf{x}_2 = \mathbf{A} \mathbf{x}_1$ when $N_1 > N_2$ is a typical inverse problem. We regularized this ill-posed reconstruction by assuming that the initial image \mathbf{x}_1 has a small Total Variation (TV) norm. In other words, we solve the Direct Regularized Problem (DRP)

$$\arg \min_{\mathbf{x}_1} \frac{1}{2} \|\mathbf{x}_2 - \mathbf{A} \mathbf{x}_1\|_2^2 + \lambda \|\mathbf{x}_1\|_{TV}, \quad (\text{DRP})$$

where the regularization parameter λ controls the weight of the *penalty* term, i.e. the TV (pseudo) norm of the solution, compared to the first term measuring the *fidelity* of the solution.

The fidelity term is expressed in the space of the fish-eye plane. At each iteration of the minimization algorithm (described in Section 4.1), the current estimate of \mathbf{x}_1 is then compared to the original fisheye image.

The TV penalty, which measures roughly the ℓ_1 norm of the gradient of the image, promotes smooth areas separated by sharp C^2 curve transitions [4], which is a reasonable model for our images. The numerical implementation of $\|\cdot\|_{TV}$ that we selected is the finite differences method [5].

In **DRP**, the penalty term has to smooth enough the reconstructed function in order to eliminate the artefacts due to the change of geometry and to the deviation between the actual functional space of \mathbf{x}_{c_1} and \mathbf{x}_{c_2} and the selected one \mathcal{X} . For instance, images are not truly bandlimited signals since they are composed of sharp object edges, and they cannot be considered as exactly piecewise smooth functions since they may contain textures.

Another significant artefact may consist in a ringing effect observed in regions of \mathbf{x}_1 that correspond to regions of \mathbf{x}_2 where the change of geometry triggers off strong discrepancies between the regular grid in the planar geometry and the grid resulting of the projection of the regular grid from the fisheye geometry to the planar one.

4. EXPERIMENTAL RESULTS

4.1. Proximal methods

The **DRP** minimization is a convex optimization program of shape

$$\arg \min_{\mathbf{x}_1} f(\mathbf{x}_1) + g(\mathbf{x}_1),$$

where f is convex and differentiable, and g is convex and has a β -Lipschitz gradient. In our case, $f(\cdot) = \frac{1}{2} \|\mathbf{x}_2 - \mathbf{A} \cdot\|_2^2$, and $g(\cdot) = \|\cdot\|_1$.

This optimization is practically solved by monotone operator splitting and proximal methods [6, 7]. More precisely, as the ℓ_1 -norm is non-differentiable, the Forward-Backward splitting is used [8]. Forward-backward (FB) splitting is essentially a generalization of the classical gradient projection method for constrained convex optimization. It can be written in the compact form

$$x^{(t+1)} = S_{\mu_t} \circ (\text{Id} - \mu_t \nabla f_2)(x^{(t)}), \quad (5)$$

where $0 < \inf_t \mu_t \leq \sup_t \mu_t < 2/\beta$ for the iteration to converge (weakly in general), S_γ is the component-wise soft-thresholding operator with threshold γ , and ∇ the gradient operator.

4.2. Experimental Setup

We evaluate the above planar image generation method with a 600×800 fisheye image² (see Fig. 1). In particular, we reconstruct a 300×272 planar image with an horizontal angle-of-view equal to 35° . The part of the fisheye image exploited is of size 265×208 , so that $\frac{N_1}{N_2} = 1.5$. Results are shown on Figure 2. We compare the performance of the **DRP** algorithm with the two methods described below. In this example, we use color processing by handling the channels independently.

Linear interpolation: The value assigned to each pixel of \mathbf{x}_1 depends on all the pixel values of \mathbf{x}_2 with the rule

$$\mathbf{x}_1(i, j) = \sum_{i'=0}^{m-1} \sum_{j'=0}^{n-1} w_{i', j'} \mathbf{x}_2(i', j').$$

In the Inverse Distance Weighted approach³, the weight function $w_{i', j'}$ is a function of the euclidian distance between a pixel (i, j) of \mathcal{G}_1 and a pixel (i', j') of \mathcal{G}_2 projected onto the planar geometry.

TV Inpainting: This method reconstructs \mathbf{x}_1 by inpainting using the TV flow [9]. We first compute \mathbf{x}_1 on a subset Ω . Each pixel of \mathcal{G}_1 which is the nearest neighbour, in the planar geometry, of a pixel of the grid \mathcal{G}_2 projected on the planar geometry by π^{-1} will belong to Ω . Its value will be equal to its nearest neighbour value. We obtain a partial image \mathbf{x}_1^Ω

²This image was acquired during an acquisition campaign in the context of the european APIDIS project

³The full interpolation scheme is described at http://www.ems-i.com/smshelp/Data_Module/Interpolation/Inverse_Distance_Weighted.htm

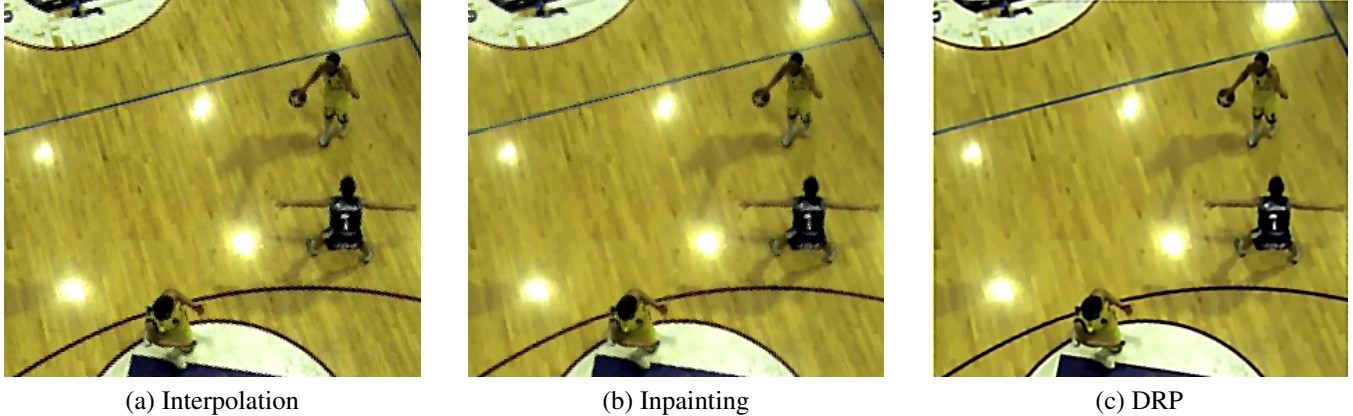


Fig. 2: Results of reconstruction of planar image from fisheye lens image for each method.

defined on Ω and with missing values on $\mathcal{G}_1 \setminus \Omega$. Typically, if $\frac{N_1}{N_2}$ increases, the size of Ω decreases. The complete image \mathbf{x}_1 is reconstructed by solving the problem expressed as

$$\arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{x}_1^\Omega - \mathbf{x}_1\|_\Omega^2 + \lambda \|\mathbf{x}_1\|_{TV}, \quad (\text{TVI})$$

where the fidelity term is only calculated on the subset Ω whereas the regularization term is calculated on the whole grid \mathcal{G}_1 . The minimization problem is solved by the proximal methods described above.

4.3. Reconstruction of planar images

The experimental results are shown on Figure 2 for the three methods described below. No PSNR is available since no ground truth is available in that case, but the relevance of our approach is evidenced by a clear visual improvement. Unsurprisingly, a better reconstruction is achieved with methods incorporating a TV regularization, which results in sharper edges when comparing with the linear interpolation approach (see the blue area at bottom or the arms and hands of the player with the basket).

It can be seen also that DRP, which includes an appropriate geometry mapping, is superior to the TVI solution. This is mostly evidenced in the sensitive areas : straight lines that are strongly distorted in the original fisheye image, digits of players or edges of players are very sharp and accurate and suffer from less artifacts than with the two other methods.

This gain of quality can be explained by the fact that for DRP, the Π operator provides information, even poor, from the omnidirectional image in these areas, whereas for TVI, there is a lack of information since these pixels have missing values. Overall, the regularization parameter λ , which realizes the tradeoff between the regularization term and the fidelity term, appear to be an important point of the algorithm.

5. CONCLUSIONS

We have addressed the problem of mapping images between different sensors. The proposed framework incorporates the

change of geometry, size and resolution between the images of different sensors (real or virtual). The discretization of the light information is seen as a geometry driven sampling problem. The corresponding inverse reconstruction problem is solved by a TV-regularization approach. Experimental results for the generation of planar images from fisheye image show the benefit of the proposed reconstruction methods compared to interpolation or inpainting strategies.

A study of the influence of the planar camera parameters (direction of view, resolution, FOV) is ongoing. In addition, we are working on improving the performance of the reconstruction algorithms with other kernels ϕ and with improved TV norm computation adapted to band limited functions [10].

6. REFERENCES

- [1] I. Tosic and P. Frossard, "Geometry-Based Distributed Scene Representation With Omnidirectional Vision Sensors," *IEEE T. Im. Proc.*, **17**(7): 1033–1046, 2008.
- [2] Y. Wiaux, J. D. McEwen, P. Vandergheynst, and O. Blanc, "Exact reconstruction with directional wavelets on the sphere," *Mon.Not.R.Astron.Soc.*, **388**:770, 2008.
- [3] C. Geyer and K. Daniilidis, "Catadioptric projective geometry," *Int. J. Comp. Vis.*, **45**(3):223–243, Dec. 2001.
- [4] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal," *Physica D.*, **60**:259–268, 1992.
- [5] A. Chambolle and J. Darbon, "On total variation minimization and surface evolution using parametric maximum flows," Tech. Rep., UCLA CAM Report, 2008.
- [6] P.L. Combettes, "Solving monotone inclusions via compositions of nonexpansive averaged operators," *Optimization*, **53**(5): 475–504, 2004.
- [7] M.J. Fadili and J.-L. Starck, "Monotone operator splitting for fast sparse solutions of inverse problems," *SIAM Journal Im. Sc.*, 2009, submitted.
- [8] P.L. Combettes and V.R. Wajs, "Signal Recovery by Proximal Forward-Backward Splitting," *Mult. Mod. Sim.*, **4**(4):1168, 2006.
- [9] Jianhong Shen, "Inpainting and the fundamental problem of image processing," *SIAM News*, **36**(5):1–4, 2003.
- [10] L. Moisan, "How to discretize the total variation of an image?," in *CIAM07 Minisymposia - Part. Diff. Eq.*, Switzerland, Aug. 2007.