OMNIDIRECTIONAL VIEWS SELECTION FOR SCENE REPRESENTATION

Ivana Tosic and Pascal Frossard

Ecole Polytechnique Fédérale de Lausanne (EPFL)
Signal Processing Institute
CH-1015 Lausanne
{ivana.tosic, pascal.frossard}@epfl.ch

ABSTRACT

This paper proposes a new method for the selection of sets of omnidirectional views, which contribute together to the efficient representation of a 3d scene. When the 3d surface is modelled as a function on a unit sphere, the view selection problem is mostly governed by the accuracy of the 3d surface reconstruction from non-uniformly sampled datasets. A novel method is proposed for the reconstruction of signals on the sphere from scattered data, using a generalization of the Spherical Fourier Transform. With that reconstruction strategy, an algorithm is then proposed to select the best subset of n views, from a predefined set of viewpoints, in order to minimize the overall reconstruction error. Starting from initial viewpoints determined by the frequency distribution of the 3d scene, the algorithm iteratively refines the selection of each of the viewpoints, in order to maximize the quality of the representation. Experiments show that the algorithm converges towards a minimal distortion, and demonstrate that the selection of omnidirectional views is consistent with the frequency characteristics of the 3d scene.

Index Terms— image based rendering, camera positioning, non-uniform sampling, omnidirectional vision, 3D scene representation

1. INTRODUCTION

The main objective of Image based rendering (IBR) is to describe a 3-dimensional scene from a set of reference images, taken from discrete viewpoints in the space. It then allows to generate new views of the scene from arbitrary positions using the description provided by reference images. The increasing popularity of such approaches for the coding of 3d scenes is certainly due to high complexity reduction, compared to the classical 3d objects coding methods. Indeed, the 3d information is obtained from normal cameras, the coding step essentially reduces to image coding, and IBR does not require special 3d rendering hardware at the user side.

IBR methods generally use the plenoptic function [2] to describe the light field that characterizes a 3d scene. Each image is a set of samples from the plenoptic function, and the performance of an IBR scheme therefore greatly depends on the choice of viewpoints for the reference images, or equivalently the camera positions. Images have to be chosen such that all parts of the scene are visible, with sufficient sampling density for high quality reconstruction. Moreover, the scene should be described with equal accuracy in all areas, in order to avoid over-sampling and under-sampling effects in the reconstruction of arbitrary views. In general, the number of images directly drives the size of the representation, and the bandwidth requirements for transmission.

Most of the previous work on camera positioning targeted the 3d scene modelling problem, where viewpoints were chosen in order to maximize the visibility of 3d features [12], or similarly, to minimize occlusions [7]. However, since these methods are primarily designed to ensure the complete coverage of the scene, they do not consider the sampling accuracy of the scanned areas, which is certainly an important factor for IBR tasks.

We propose here a framework where a 3d scene is captured by multiple omnidirectional cameras, whose output images are appropriately mapped into spherical images that represent the light in its natural radial form [5]. Processing the light field directly in the spherical domain allows to avoid discrepancies due to Euclidean approximations in common IBR schemes. Moreover, omnidirectional cameras offer generally a much wider field of view, which reduces the number of necessary images for reconstruction of the scene. We consider that 3d objects can be represented as continuous functions in the 3d space, in particular as function belonging to the space of square integrable functions on a unit sphere. This approach is motivated by recent works on representation and compression of 3d objects as functions on the sphere [1, 3, 6]. By modelling a 3d scene with a continuous function we can achieve a consistent reconstruction of arbitrary views without over-sampling nor under-sampling. The view selection problem can then be reduced to a non-uniform sampling problem on the sphere. From initial set of n viewpoints, an iterative algorithm is finally proposed to selectively alter the choice of viewpoints, such that the overall distortion in the scene reconstruction is minimized.

The paper is organized as follows. In Section 2 we describe the framework for scene representation with omindirectional views. A new method for interpolation from non-uniform samples on the sphere is proposed in Section 3. Section 4 presents the view selection algorithm, and experimental results are given in Section 5. Section 6 concludes the paper.

2. FRAMEWORK

We consider a framework where a 3d object is observed by multiple omnidirectional cameras, which represent the light field under a spherical form. The proposed approach for view selection is based on sampling of the 3d object surface, which is modelled as a contin-

This work has been supported by the Swiss National Science Foundation under grants 20001-107970/1 and PP002-68737.

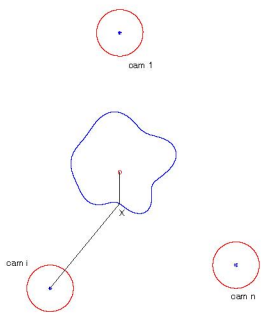

Fig. 1. A 2d slice of a synthetic scene captured with n cameras (only 3 are displayed for simplicity).

uous function on the sphere. It is described in the form $r = f(\theta, \varphi)$, where (r, θ, φ) are the spherical coordinates of a point on the object surface, when the center of the coordinate system coincides with the center of the object. This parametrization is only valid for star-shape models, i.e., models where each point on the surface corresponds to only one direction in the spherical coordinate system. However, a parametrization into a finite number of spheres is feasible for more complex models, with an appropriate tessellation of the surface.

Figure 1 shows a 2d slice of a synthetic scene, recorded by n omnidirectional cameras. Suppose that camera i captures the distance of the point X to the center of the sphere, i.e., the radial coordinate of this point in the coordinate system of camera i. We will denote this system the camera i frame (F_i), while the coordinate system of the object will be called the world frame (F_0). Cartesian coordinates of a point in different frames are related via the rigid body transformation [13] in the following way:

$$X_i = R_i \cdot X_0 + T_i, \tag{1}$$

where X_i and X_0 are the coordinates of the point X in F_i and F_0 respectively, and R_i and T_i are the rotation and translation matrices of a frame F_i with reference to the frame F_0. The world frame is uniquely defined as (r, θ, φ) coordinates in a spherical coordinate system.

We then consider a set of viewpoints P that are distributed around the object. In other words, we define a discrete sampling of the world frame F_0. In order to obtain a discrete set of values for θ and φ coordinates, we have chosen to use the HEALPix [9] tessellation of the sphere, which results in sphere partitions of equal area and therefore gives the same importance to all possible viewpoints (placed in centers of the partitions). In HEALPix, a base mesh consists of 12 parts of the sphere, and finer meshes are obtained by successive dyadic partitioning of each part in the base mesh. The number of possible viewpoints is given with $N_{dir} = 12 N_{side}^2$, where $N_{side} = 2^{l_r}$ and l_r is the resolution level. We finally choose n_r arbitrary values for the distance coordinate r, so the total number of viewpoints in our system is given by $N_{pos} = 12\, n_r\, N_{side}^2$.

The view selection problem consists in finding in P, the subset of n viewpoints, $P_n \subset P$, such that the overall distortion in the reconstruction of the scene is minimized. The Mean Square Error (MSE) is used to assess the quality of the reconstruction, defined on the sphere as :

$$\begin{aligned} MSE(f, \tilde{f}) &= \|f - \tilde{f}\|^2 = \langle f - \tilde{f}, f - \tilde{f} \rangle & (2) \\ &= \frac{1}{4\pi} \int_\theta \int_\varphi (f(\theta, \varphi) - \tilde{f}(\theta, \varphi))^2 sin\theta d\theta d\varphi, & (3) \end{aligned}$$

where f is the signal on the sphere and $\tilde{f}$ is its reconstruction.

3. RECONSTRUCTION FROM A SET OF OMNIDIRECTIONAL IMAGES

By capturing depth information from multiple arbitrarily positioned omnidirectional cameras, we obtain non-uniformly distributed samples on the sphere. Signal reconstruction from non-uniform samples for 1d and 2d signals on the plane, has been studied intensively in literature, where both theoretical and practical issues have been considered [10]. However, interpolation from scattered data on the sphere has been studied by only a few authors [4, 11], who have mostly developed the theoretical frameworks, without considering the practical implementation issues. We propose here a method for object reconstruction from non-uniform samples obtained by multiple cameras, based on the Fast Spherical Fourier Transform - FST [8].
FST decomposes a signal that belongs to the Hilbert space of square-integrable functions on the two-dimensional sphere $L^2(S^2, d\omega)$ into a series of spherical harmonics $Y_l^m(\theta, \varphi)$:

$$f(\theta, \varphi) = \sum_{l \in \mathbb{N}} \sum_{|m| \leq l} \hat{f}(l, m) Y_l^m(\theta, \varphi). \tag{4}$$

The Fourier coefficients $\hat{f}(l, m)$ are given with:

$$\hat{f}(l, m) = \int_{S^2} f \bar{Y}_l^m(\theta, \varphi) d\omega, \tag{5}$$

where $d\omega(\theta, \varphi) = d\cos\theta d\varphi$ is the rotation invariant Lebesgue measure on the sphere.

Let now $\mathcal{P}_M$ be the space of polynomials on the sphere, given by:

$$p(\theta, \varphi) = \sum_{l=0}^{N-1} \sum_{|m| \leq l} a(l, m) Y_l^m(\theta, \varphi). \tag{6}$$

An arbitrary sampling problem can then be considered as a discretization of above polynomials on the sphere. In particular, in the case of uniform sampling we have:

$$f(\theta_j, \varphi_k) = p(\pi j / 2N, \pi k / N) \tag{7}$$

The sampling theorem for uniformly distributed samples on the sphere has been established by Driscol and Healy [8]. It states that if a signal on the sphere is bandlimited, i.e., if $\hat{f}(l, m) = 0$ for $l \geq N$, then it can be perfectly recovered from its uniform samples $\theta_j = \pi j / 2N, \varphi = \pi k / N; j, k = 0, ..., 2N - 1$.

Consider next the problem of the reconstruction of an N - bandlimited signal on the unit sphere, from non-uniform samples obtained from n omnidirectional images. The signal is given by n sets of samples $S_i, i = 1, ..., n$, where each set contains q_i samples of the object: $S_i = (r_j^i, \theta_j^i, \varphi_j^i)$, $j = 1, ..., q_i$. Hence, using the formula (6), for each sample $(r_j^i, \theta_j^i, \varphi_j^i)$, $i = 1, ..., n; j = 1, ..., q_i$, we have:

$$r_j^i = f(\theta_j^i, \varphi_j^i) = \sum_{l=0}^{N-1} \sum_{|m| \leq l} a(l, m) Y_l^m(\theta_j^i, \varphi_j^i). \tag{8}$$

The previous relation can be rewritten in a matrix form as:

$$V \cdot \mathbf{a} = \mathbf{r}, \tag{9}$$

where

$$V = \{Y_l^m(\theta_j^i, \varphi_j^i)\}_{q \times N^2}, \tag{10}$$

$$\mathbf{a} = \{a(l,m)\}_{N^2 \times 1}, \tag{11}$$

$$\mathbf{r} = \{r_j^i\}_{q \times 1}, \tag{12}$$

and $q = \sum_{i=1}^{n} q_i$.

By solving this linear system we obtain the values for coefficients $a(l,m)$, which are the approximates of Fourier coefficients for the signal f on the sphere :

$$\hat{f}(\theta, \varphi) \approx \mathbf{a} = V^{-1} \cdot \mathbf{r}. \tag{13}$$

Since V is not a square matrix, finding its inverse is not an easy task. The pseudo-inverse will give a minimal norm solution, resulting in a stable reconstruction, but it is computationally very expensive. Instead of directly solving (9), we propose to multiply each side with $V^* = conj(V^T)$, which results in :

$$V^* \cdot V \cdot \mathbf{a} = V^* \cdot \mathbf{r}, \tag{14}$$

or equivalently :

$$T \cdot \mathbf{a} = \mathbf{R}, \tag{15}$$

with

$$T = V^* \cdot V; \qquad \mathbf{R} = V^* \cdot \mathbf{r}. \tag{16}$$

The matrix T is now a square matrix of size N^2, so that solving the system given in (15) instead of (9) is much less computationally demanding. Another advantage of this transformation is that adding or removing samples does not change the size of the system, as it amounts to a simple addition (resp. subtraction), as given by :

$$(T \pm \Delta T) \cdot \mathbf{a} = \mathbf{R} \pm \Delta \mathbf{R}, \tag{17}$$

where ΔT and $\Delta \mathbf{R}$ are obtained using (16) for set of added (resp. deleted) samples.

4. VIEW SELECTION ALGORITHM

Now that a method has been defined to reconstruct an object from scattered data, we can address the problem of selecting the subset of images or viewpoints P_n that minimizes the overall distortion after reconstruction. The view selection algorithm first selects the initial viewpoints based on the frequency distribution of the 3d scene. Depth estimation is performed from all images captured from each viewpoint in P, in order to obtain a set of samples $S = \bigcup_{i=1}^{N_{pos}} S_i$, where S_i is a set of depths seen from i^{th} viewpoint. The scene f is then approximated into $\tilde{f}_{all}$ from S, using the FST-based method (see eq. 15). Substituting $\tilde{f}_{all}$ in the spherical harmonic differential equation, the distribution of frequency $w = \sqrt{l^2 + m^2}$ of the 3d object can be estimated. An *importance factor* IF_i is finally computed for each view in P, as

$$IF_i = \frac{\sum_{j=1}^{q_i} \tilde{w}(\theta_j^i, \varphi_j^i)}{q_i},$$

where $\tilde{w}(\theta_j^i, \varphi_j^i)$ is an estimate of the frequency w at a sample $(\theta_j^i, \varphi_j^i)$ originating from a view i. The n viewpoints with highest importance factor then form an initial set of images, $P_{n,0} = \{V_{1,0}, ..., V_{n,0}\}$. Initial positioning obtained by using this frequency estimation approach highly increases the probability that the view selection algorithm converges towards the global minimum of the error.

The view selection algorithm then iteratively refines each of the initial viewpoints, as described in Algorithm 1. It basically proceeds in three steps:

- *Step 1:* The initial viewpoints $P_{n,0} = \{V_{1,0}, ..., V_{n,0}\}$ are chosen as the positions with the highest importance factors.
- *Step 2:* The first viewpoint is moved to other possible positions, while keeping all viewpoints fixed. In each case, the object is reconstructed using the FST-based method from depth maps seen only from n cameras, for each of these possible positions of camera 1, and creates $\tilde{f}_{1,0}^m; m = 1, ..., t$. The distortion is calculated for each one of these approximations, with respect to $\tilde{f}_{all}$. The minimal value for MSE will give the best position for camera 1 at this iteration, and $V_{1,0}$ is assigned the value of the best viewpoint for camera 1. The same procedure is repeated for all the viewpoints, by altering one of the positions, while all the other stay unchanged.
- *Step 3:* After finding the best position for camera n, the algorithm goes back to the camera 1 and repeats the Step 2 with a new set of camera positions $P_{n,1}$, obtained from a previous iteration. The search is continued until it reaches the stable solution (P_n), i.e. when the change of position of any camera does not lead to reduction in MSE, i.e., $P_{n,i} = P_{n,i-1}$.

Algorithm 1 View Selection

Input: $P, S, \tilde{f}_{all}, P_{n,0} = \{V_{1,0}, ..., V_{n,0}\}$
$i = 0; j = 0$
repeat
  $S_i = \bigcup_{k=1}^{n} S_{k,i}$, where $S_{k,i}$ is a set of surface samples seen from view $V_{k,i}$
  $P_{n,i}^0 = P_{n,i}$
  repeat
    $j = j + 1$
    $PV_{j,i} = P \setminus (P_{n,i}^{j-1} \setminus V_{j,i}) = \{V_{j,i}^1, ..., V_{j,i}^t\}$
    $PS_{j,i}^m = S_{j,i}^m \cup \{\bigcup_{k=1;k\neq j}^{n} S_{k,i}\}$ for $m = 1, ..., t$
    reconstruct $\tilde{f}_{j,i}^m$ from $PS_{j,i}^m$ using the FST method for $m = 1, ..., t$
    $L_m = MSE(\tilde{f}_{j,i}^m, \tilde{f}_{all})$ for $m = 1, ..., t$
    $V_{j,i} = \{V_{j,i}^M | L_M = min_{m=1,...,t} L_m\}$
    $P_{n,i}^j = \{V_{1,i}, ..., V_{n,i}\}$
  until $j = n$
  $i = i + 1;$
  $P_{n,i} = P_{n,i-1}^n$
until $P_{n,i} = P_{n,i-1}$

5. EXPERIMENTAL RESULTS

The performance of the proposed method is evaluated on an arbitrarily shaped synthetic object, shown on Figure 2. The number of possible viewpoints used in simulations is set to $Npos = 144$, with

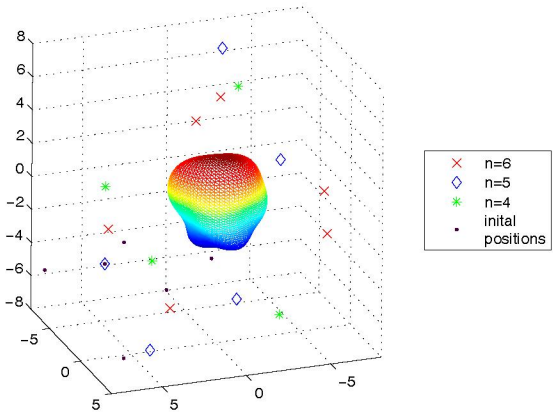

Fig. 2. View selection algorithm: results for 4, 5 and 6 cameras, with initial viewpoints set.

Table 1. **MSE for reconstruction from n best viewpoints, with n=4, 5 and 6.**

n	MSE_{min}	i
4	8.8556e-13	16
5	1.8161e-13	16
6	9.8843e-15	22

$Nside = 2$ and $n_r = 3$ ($r_1 = 7, r_2 = 10, r_3 = 13$). The viewpoints selected by the view selection algorithm are represented for the cases of $n = 4, 5$ and 6 cameras. The initial set of viewpoints chosen from the frequency distribution is also represented for $n = 6$ (for $n = 4$ and $n = 5$ the initial set of viewpoints is a subset of the initial set for $n = 6$). The resulting selection of viewpoints is consistent with the object shape, namely camera positions are chosen so that the object is sampled more densely in the areas with higher frequency content. Moreover, it is interesting to note that most of the viewpoints are placed on the sphere with the smallest radius, which means that the influence of view resolution on positioning is higher than the visibility of the object surface, when the number of viewpoints is large enough.

Table 1 presents the results of camera positioning for $n = 4, 5$ and 6 cameras respectively, in terms of MSE of the reconstruction. It also represents the number of iterations of the algorithm, i, which are required to reach convergence. We see that the MSE decreases with the number of viewpoints, as expected, and that the convergence is quite fast, even for increasing number of cameras in the system.

6. CONCLUSIONS

This paper introduces a view selection algorithm for 3d scene representation with a set of n omnidirectional cameras. 3d objects are modelled as functions on the unit sphere, and a method for the reconstruction of 3d objects from non-uniformly spaced samples is presented. The view selection algorithm first initializes the viewpoints based on the frequency distribution of the object. It then iteratively refines each of the camera positions, until the distortion of the reconstruction is minimized. Experimental results show that the proposed algorithm selects viewpoints, in such a way that parts of the scene with higher frequencies are sampled more densely; it is exactly the intuition one would start with in dealing with the positioning problem. The proposed framework and view selection algorithm represent a promising alternative in the representation and coding of 3d scenes, based on spherical image-based rendering approaches.

7. REFERENCES

[1] A. Khodakovsky, P. Schröder and W. Sweldens. Progressive geometry compression. *Siggraph '00 Conference Proceedings*, July 2000.

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

[3] H. Hoppe and E. Praun. Shape compression using spherical geometry images. *Symposium on Multiresolution in Geometric Modeling. Cambridge*, September 2003.

[4] S. Hubbert and T.M. Morton. l_p-error estimates for radial basis function interpolation on the sphere. *Journal of Approximation Theory*, (129):58–77, 2004.

[5] I. Tosic, I. Bogdanova, P. Frossard and P. Vandergheynst. Multiresolution Motion Estimation for Omnidirectional Images. In *Proceedings of EUSIPCO*, September 2005.

[6] I. Tosic, P. Frossard and P. Vandergheynst. Progressive coding of 3d objects based on overcomplete decompositions. EPFL-ITS Technical Report TR-ITS-2005.026, 1015 Ecublens, October 2005.

[7] J. Maver, R. Bajcsy. Occlusions as a guide for planning the next view. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(5):417–433, May 1993.

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

[9] K.M. Górski, E. Hivon, A.J. Banday, B.D. Wandelt, F.K. Hansen, M. Reinecke and M. Bartelmann. HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. *The Astrophysical Journal*, 622(2), 2005.

[10] F. Marvasti. *Nonuniform sampling: Theory and Practice*, chapter 6, pages 284–290. Kluwer Academic/Plenum Publishers, 2000.

[11] F. Narcowich and J.D. Ward. Scattered data interpolation on spheres: error estimates and locally supported basis functions. *SIAM Journal on Mathematical Analysis*, 33(6):1393–1410, 2002.

[12] P. K. Allen, M. K. Reed and I. Stamos. View Planning for Site Modeling. *in Proc. DARPA Image Understanding Workshop, Monterey*, pages 1181–1192, November 1998.

[13] Y. Ma, S. Soatto, J. Košecká and S.S. Sastry. *An Invitation to 3-D Vision: From Images to Geometric Models*, chapter 2, pages 19–34. Springer, 2004.