Scalable Motion-Aware Panoramic Videos

Leonardo Sacht* Luiz Velho Diego Nehab Marcelo Cicconet
IMPA, Rio de Janeiro, Brazil

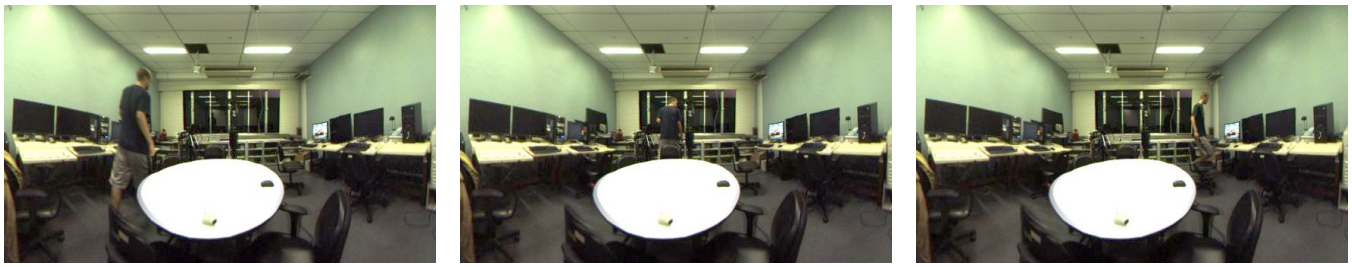

Figure 1: *Our method generates panoramic videos considering both spatial and temporal constraints. Our results comprehend wide fields of view and preserve straight lines and object shapes. Shapes of moving objects are especially preserved.*

Abstract

This work presents a method for obtaining perceptually natural panoramic videos, that is, videos composed of wide-angle frames where straight lines and object shapes are preserved. The shape preservation for moving objects has a special treatment that ensures temporal coherence. Energies describing these properties are obtained for the case of a fixed omni-directional camera, and an optimization procedure is presented. Our optimization works per-frame, which makes the method scalable for arbitrarily long scenes.

Keywords: 2D Morphing and Warping, Panoramic Video.

1 Introduction

Recent developments in computational photography gave rise to affordable omni-directional cameras, which are output video in the equirectangular format. Applications in navigation, immersion, and surveilance are the key driving forces of omni-directional video.

We consider a new use for the spherical videos captured by these cameras: videos with wide-angle frames. We improve on recent developments in the understanding of the distortions involved in wide-angle images [Zorin and Barr 1995; Kopf et al. 2009; Carroll et al. 2009], and study the implications of the introduction of the time dimension.

Our method is divided into three phases: pre-processing, optimization and post-processing. As in the work of [Carroll et al. 2009], our pre-processing interface allows the user to select lines that are to be straight in the final result video, as well as their final orientations. The user can also specify the field-of-view (FOV) to be considered in the input video. While these operations are performed, our system detects the moving objects in the video using background subtraction. Using this input, an optimization phase is performed (see following sections). Finally, the post-processing stage consists of specifying the rectangular area of the optimization results to store in the output video.

*e-mail: leo-ks@impa.br

2 Spatial Constraints

We aim to find an optimal projection of the input video

$$\begin{aligned}\mathbf{U} : [\lambda_o, \lambda_f] \times [\phi_o, \phi_f] \times [0, t_f] &\to \mathbb{R}^3 \\ (\lambda, \phi, t) &\mapsto (U(\lambda, \phi, t), V(\lambda, \phi, t), t)\end{aligned} \tag{1}$$

where $[\lambda_o, \lambda_f] \times [\phi_o, \phi_f] \subseteq [-\pi, \pi] \times [-\frac{\pi}{2}, \frac{\pi}{2}]$ is the FOV specified by the user and $[0, t_f]$ is a time interval. The projection of the points (λ_j, ϕ_i, t_k) obtained by uniformly discretizing the longitude-latitude-time domain are the unknowns in the problem:

$$\mathbf{U}_{ijk} := (U_{ijk}, V_{ijk}, t_k) = \mathbf{U}(\lambda_j, \phi_i, t_k). \tag{2}$$

The first step in our optimization is to compute an optimal projection for the background. The unknowns U_{ij} and V_{ij} for the background are put together in a solution vector denoted by $\mathbf{x}_{bg}$. We minimize energies related to well known perceptual properties for panoramic images: conformality, smoothness and straight lines. For conformality and smoothness, we use slight modifications of the energies proposed by Carroll et al. [2009]. We denote these energies by $E_c(\mathbf{x}_{bg})$ and $E_s(\mathbf{x}_{bg})$. To preserve straight lines, we use an energy term that expresses how the projection of L deviates from the normal direction to $\mathbf{n}_L$, for all line segments L:

$$E_l(\mathbf{x}_{bg}, \{\mathbf{n}_L\}) = \sum_{L \text{ line}} \sum_{\mathbf{e}_w \in V_L} \left(\mathbf{n}_L^T \cdot (\mathbf{U}_{w+1}^L - \mathbf{U}_w^L)\right)^2. \tag{3}$$

Above, $\mathbf{U}_{w+1}^L$ is a virtual output vertex defined by $\mathbf{e}_{w+1}$, the edge intersected by L immediately after $\mathbf{e}_w$. V_L is the set of edges of the discretized domain intersected by L. The normals associated to lines with no specified orientations are treated as unknowns. In order to make the linear system associated to our final energy nonsingular, we add a term $E_a(\mathbf{x}_{bg})$ that penalizes the mesh corner points from deviating from specified coordinates. The projection for the background is obtained by solving for $\mathbf{x}_{bg}$ in

$$\arg\min_{\mathbf{x}_{bg}} \; E_{bg}\big(\mathbf{x}_{bg}, \{\mathbf{n}_L\}\big), \quad \text{s.t.} \quad \|\mathbf{n}_L\|^2 = 1, \tag{4}$$

where

$$\begin{aligned}E_{bg}\big(\mathbf{x}_{bg}, \{\mathbf{n}_L\}\big) = w_c^2 E_c(\mathbf{x}_{bg}) + w_s^2 E_s(\mathbf{x}_{bg}) \\ + w_l^2 E_l\big(\mathbf{x}_{bg}, \{\mathbf{n}_L\}\big) + w_a^2 E_a(\mathbf{x}_{bg})\end{aligned} \tag{5}$$

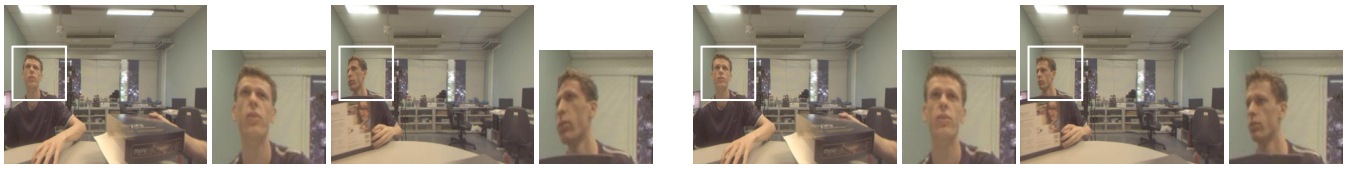

Figure 2: *Left: Two frames of the result video after step 1 and detail of them. The man's head starts with reasonable proportions but gets too distorted when passing near line constraints, revealing a temporal incoherent behavior. Right: The same two frames of the result video after step 3. Now the man's head is better preserved along time. Artificial lines were marked over the backgorund and minimized for both steps to emphasize differences between their results.*

In all our tests, we use weights $w_c = 1$, $w_s = 0.5$, $w_l = 3$, and $w_a = 0.01$. To obtain a solution, we alternate between solving for $\mathbf{x}_{bg}$ with $\{\mathbf{n}_L\}$ fixed, and solving for $\{\mathbf{n}_L\}$ with $\mathbf{x}_{bg}$ fixed. By continuing this process, we obtain a convergent sequence of energy values for $E_{bg}(\mathbf{x}_{bg}, \{\mathbf{n}_L\})$.

3 Temporal Constraints

Applying the background projection for all frames can result in temporal incoherence in the shape of moving objects (see left of Figure 2). These inconsistencies are caused by strong variations of the differential north and east vectors

$$\mathbf{H}(\lambda,\phi,t) = \begin{pmatrix} \frac{\partial U}{\partial \phi}(\lambda,\phi,t) \\ \frac{\partial V}{\partial \phi}(\lambda,\phi,t) \end{pmatrix} \quad \text{and} \quad \mathbf{K}(\lambda,\phi,t) = \frac{1}{\cos(\phi)} \begin{pmatrix} \frac{\partial U}{\partial \lambda} \\ \frac{\partial V}{\partial \lambda} \end{pmatrix},$$

in different areas of the background projection. When a moving object passes over these areas, these variations become more pronounced, leading to unpleasant effects, especially near straight lines.

To avoid these problems, we restrict the projection to be smoother over moving objects. For all points (λ_j, ϕ_i, t_k) detected as belonging to a moving object, if $(\lambda_{j+1}, \phi_i, t_k)$ and $(\lambda_j, \phi_{i+1}, t_k)$ also belong to a moving object, we enforce

$$\begin{cases} \mathbf{H}(\lambda_j,\phi_i,t_k) = \mathbf{H}(\lambda_{j+1},\phi_i,t_k) \\ \mathbf{K}(\lambda_j,\phi_i,t_k) = \mathbf{K}(\lambda_{j+1},\phi_i,t_k) \end{cases} \tag{6}$$

and

$$\begin{cases} \mathbf{H}(\lambda_j,\phi_i,t_k) = \mathbf{H}(\lambda_j,\phi_{i+1},t_k) \\ \mathbf{K}(\lambda_j,\phi_i,t_k) = \mathbf{K}(\lambda_j,\phi_{i+1},t_k). \end{cases} \tag{7}$$

These requirements alone could still lead to temporal incoherences, since no information about the neighboring frames is being considered. For example, abrupt changes in the scene such as objects suddenly coming in and out of the frame could lead to changes in the resulting video. To consider the information coming from neighboring frames, we impose (6) and (7) to points that belong to objects in adjacent past and future frames (between 16 frames before and after the current frame), and ponder these constraints by Gaussian weights centered at the current frame.

Denoting by $\mathbf{x}_k$ the vector of unknowns U_{ijk} and V_{ijk} and discretizing (6) and (7) leads to an energy that captures distortion in moving objects, which we denote by $E_{sh}(\mathbf{x}_k)$. Step 2 of our optimization consists of minimizing this energy and restricting the points that do not belong to moving objects to be projected as in the solution $\mathbf{x}_{bg}$ of step 1 . To do so, we minimize the energy

$$E_{ob}(\mathbf{x}_k) = \gamma^2 E_{sh}(\mathbf{x}_k) + \|\mathbf{x}_k - \mathbf{x}_{bg}\|^2, \tag{8}$$

where γ is a parameter that controls shape distortion for moving objects. We set $\gamma = 2.5$. The solution $\mathbf{x}_k^{ob}$ of this step may not be satisfactory yet, since straight lines near moving objects may now appear bent due to the extra constraints. To alleviate this problem, the third and last step of our optimization minimizes the energy

$$E_{final}(\mathbf{x}_k) = E_{bg}(\mathbf{x}_k, \{\mathbf{n}_L\}) + \gamma^2 \|\mathbf{x}_k - \mathbf{x}_k^{ob}\|^2. \tag{9}$$

Here we choose the set of normals $\{\mathbf{n}_L\}$ for lines with no specified orientation as the last one obtained in step 1. One result of step 3 is shown in Figure 2 (right).

Figure 3: *A one-minute long result of our method.*

4 Results

We use a mesh with resolution around 70×70 vertices for each frame of each result. Step 1 of our minimization takes a few seconds and is performed only once. Step 2 takes about 0.05 seconds and step 3 takes about 0.005 per frame.

A result of our method is shown in Figure 1. This video comprehends a field of view of 170 degrees longitude by 110 degrees latitude. It has 170 frames. Other result that comprehends 135×100 degrees is shown in Figure 3. This result highlights the scalability of our method. It has 800 frames and is one minute long. These and other results are available in the accompanying video.

Acknowledgements

We acknowledge the grant of the first author provided by CNPq.

References

CARROLL, R., AGRAWAL, M., and AGARWALA, A. 2009. Optimizing content-preserving projections for wide-angle images. *ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH 2009)*, 28(3):43.

KOPF, J., LISCHINSKI, D., DEUSSEN, O., COHEN-OR, D., and COHEN, M. F. 2009. Locally adapted projections to reduce panorama distortions. *CGF (Proc. of EGSR'09)*, 28(4):1083–1089.

ZORIN, D. and BARR, A. H. 1995. Correction of geometric perceptual distortions in pictures. In *Proceedings of ACM SIGGRAPH 1995*, pages 257–264.