Content-Based Projections for Panoramic Images and Videos

Leonardo K. Sacht, Paulo Cezar Carvalho, Luiz Velho
(leo-ks, pcezar, lvelho) @ impa.br

Introduction

Panoramic images, i.e., images of wide fields of view, allows us to better represent scenes since they do not have the field of view limitation presented by common cameras.
We represent a scene as a sphere with color information, which we call *viewing sphere*. Using the longitude/latitude representation

$$\begin{array}{rrcl} \mathbf{r}: & [-\pi,\pi]\times[-\frac{\pi}{2},\frac{\pi}{2}] & \to & \mathbb{S}^2 \\ & (\lambda,\phi) & \mapsto & (\cos(\lambda)\cos(\phi),\sin(\lambda)\cos(\phi),\sin(\phi)) \end{array}$$

we associate this sphere to images called *equirectangular images* (the input images in the result section are equirectangular images).
With these notations, we formulate the panoramic image problem as that of finding a good projection

$$\begin{array}{rrcl} \mathbf{u}: & S\subseteq\mathbb{S}^2 & \to & \mathbb{R}^2 \\ & (\lambda,\phi) & \mapsto & (u(\lambda,\phi),v(\lambda,\phi)) \end{array}.$$

Optimizing Content-Preserving Projections

Previous approaches ([2] and [3]) show that the main difficulty of the panoramic image problem is to obtain a result where straight lines in the scene appear straight and shapes of objects are preserved.
Carroll et al. [1] model these properties in terms of $\mathbf{u}$ and propose an optimization solution that minimizes all these distortions.
An energy E_c that measures how distorted shapes are is obtained by discretizing the *Cauchy-Riemann equations*

$$\frac{\partial u}{\partial \phi} = -\frac{1}{\cos(\phi)}\frac{\partial v}{\partial \lambda} \text{ and } \frac{\partial v}{\partial \phi} = \frac{1}{\cos(\phi)}\frac{\partial u}{\partial \lambda}.$$

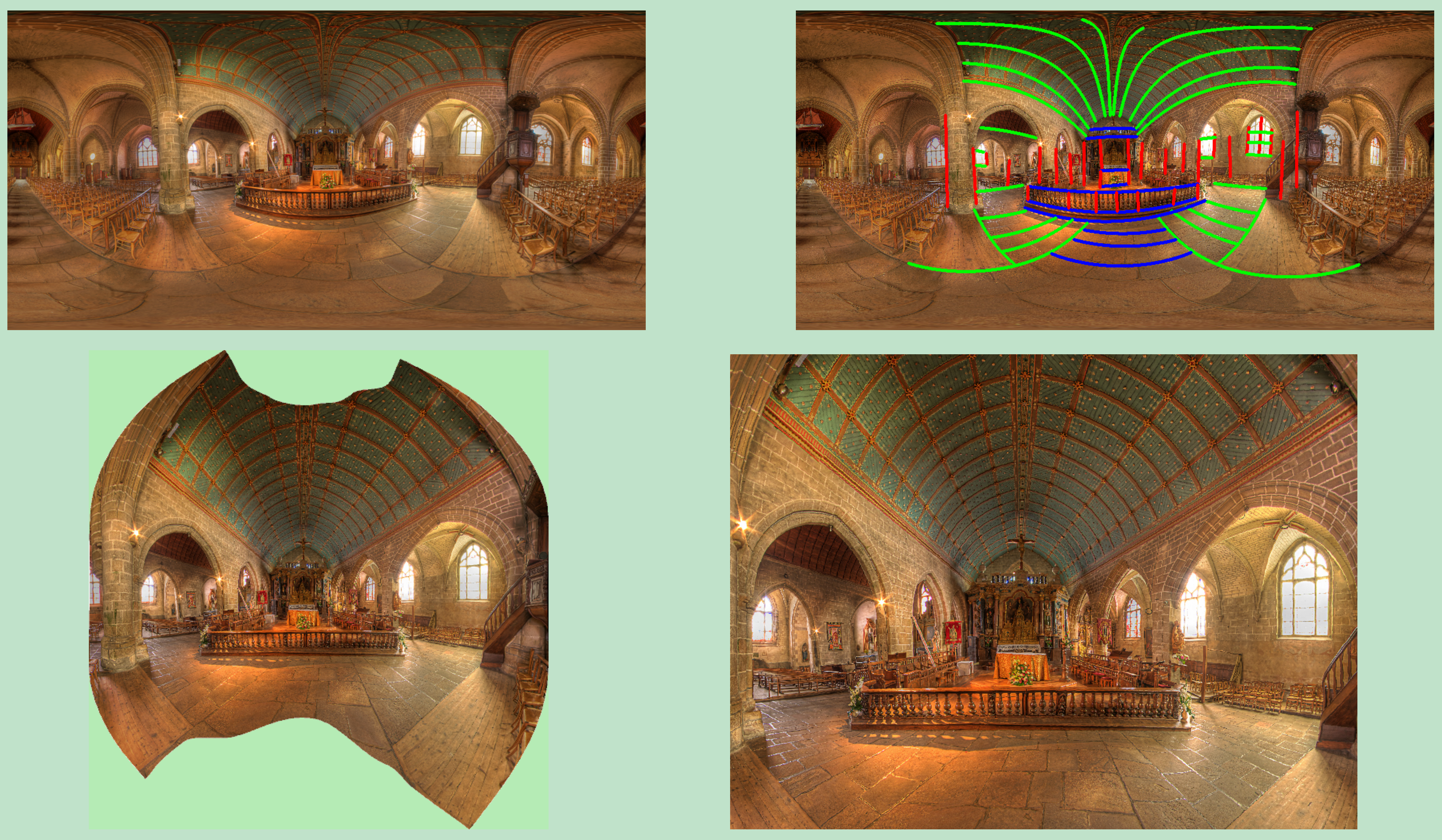
Energies of the form E_l are constructed from the information given by the user about the localization of the lines in the scene (see the images with marked lines in the result section). An additional energy E_s is obtained to measure the variation of the projection.
The final discretized projection is obtained by minimizing

$$E = w_c^2 E_c + w_s^2 E_s + w_l^2 \sum_{l\in L} E_l.$$

Minimization is done by solving a sparse, symmetric linear system.

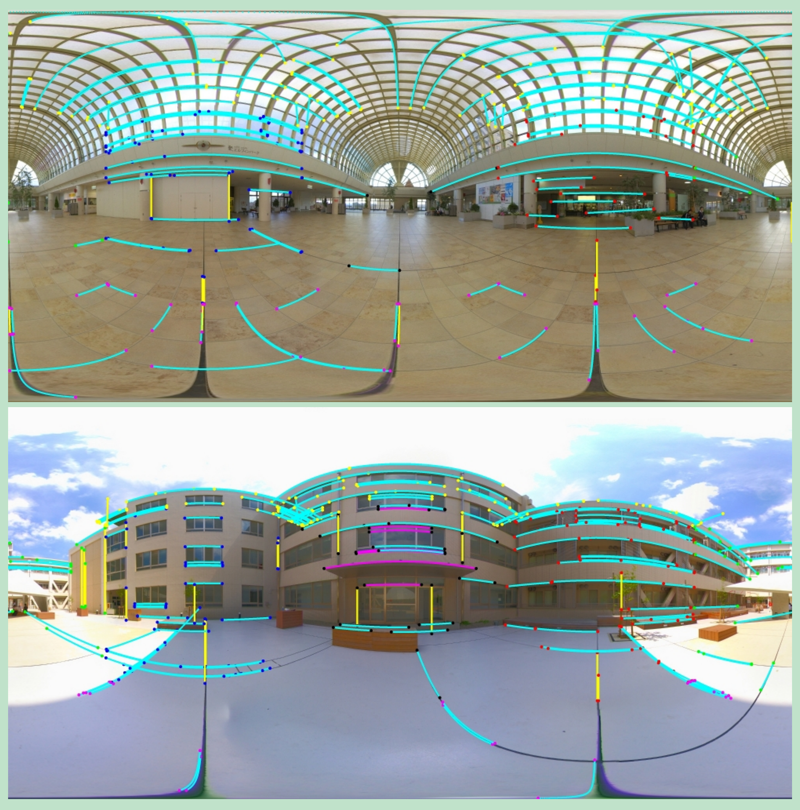
Results

We show below two results obtained using our implementation of the method explained in the previous section.



Line Detection

We developed a method to detect straight lines in equirectangular images.
The viewing sphere is projected to six perspective images. These images are processed and the Hough transform is applied on each of them. The detected lines are then mapped back to the equirectangular image.
We show below two results. In the first one, 161 line segments were detected and, in the second one, 225 segments were detected.



Face Detection

We implemented a method to detect faces in equirectangular images.
The viewing sphere is projected to its corresponding Mercator image. This image is processed and a face detector is applied on it. The detected faces are then mapped back to the equirectangular image.
We show below two results of this method.

Details about both detection methods can be found in Sacht et al. [4].

Panoramic Videos

In this section we consider the problem of obtaining panoramic videos, i.e., videos where each frame is a panoramic image.

We separate the problem in 3 cases:
- Case 1: Stationary VP, stationary FOV, moving objects;
- Case 2: Stationary VP, moving FOV, stationary objects;
- Case 3: Moving VP.

We formulate perceptual properties that are desirable in panoramic videos:
- Each frame must be a good panoramic image;
- Moving objects must be less distorted;
- Temporal coherence for the scene;
- Temporal coherence for the objects.

The information of a scene through time is represented by the *temporal viewing sphere*, which is the image of the following function:

$$\begin{array}{rcl} \mathbf{R}: [-\pi,\pi]\times[-\frac{\pi}{2},\frac{\pi}{2}]\times[0,t_0] & \to & \mathbb{R}^4 \\ (\lambda,\phi,t) & \mapsto & (\mathbf{r}(\lambda,\phi),t) \end{array}$$

Temporal viewing spheres are denoted by $\mathbb{TS}^2$ and represented by *equirectangular videos*. With these notations we state the panoramic video problem as that of finding a projection

$$\begin{array}{rrcl} \mathbf{U}: & S\subseteq\mathbb{TS}^2 & \to & \mathbb{R}^3 \\ & (\lambda,\phi,t) & \mapsto & (U(\lambda,\phi,t),V(\lambda,\phi,t)) \end{array}.$$

We now consider solutions for case 1. A simple solution that presents temporal incoherence for the objects is to generate a video where each frame is an optimizing image obtained by the method studied in last sections.
To correct these problems, we present temporal coherence equations for the moving objects

$$\frac{\partial U}{\partial \phi}(\lambda^{ob}_{t_1,t_2},\phi^{ob}_{t_1,t_2},t_2) - \frac{\partial U}{\partial \phi}(\lambda,\phi,t_1) = 0$$
$$\frac{\partial V}{\partial \phi}(\lambda^{ob}_{t_1,t_2},\phi^{ob}_{t_1,t_2},t_2) - \frac{\partial V}{\partial \phi}(\lambda,\phi,t_1) = 0$$

and for the scene

$$\frac{\partial U}{\partial \phi}(\lambda,\phi,t_2) - \frac{\partial U}{\partial \phi}(\lambda,\phi,t_1) = 0$$
$$\frac{\partial V}{\partial \phi}(\lambda,\phi,t_2) - \frac{\partial V}{\partial \phi}(\lambda,\phi,t_1) = 0 .$$

We show below two frames of a video produced by our method

References

[1] R. Carroll, M. Agrawala, A. Agarwala. Optimizing content-preserving projections for wide-angle images. In *Siggraph 2009*.

[2] D. Zorin, A. Barr. Correction of geometric perceptual distortions in pictures. In *Siggraph 95*.

[3] L. Zelnik-Manor, G. Peters, P. Perona. Squaring the circle in panoramas. In *ICCV' 2005*.

[4] L. K. Sacht, P.C. Carvalho, L. Velho, M. Gattass. Face and straight line detection in equirectangular images. In *WVC 2010*.