Homework 2: Stereo Matching

Prof. Davi Geiger

Due March 7^{th}, 2017

Introduction

If two cameras are parallel to each other (no rotation) and if the translation is simply along "$\hat{X}$", then

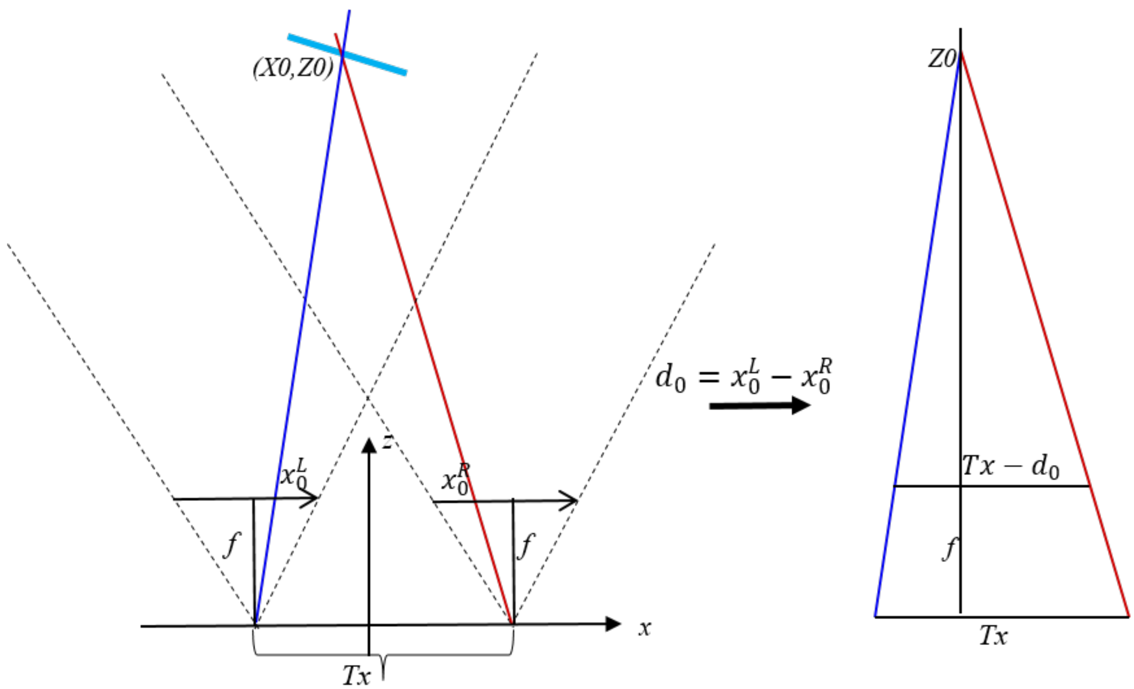

Figure 1: *Projection to parallel cameras. From the triangle on the right we derive* $\frac{Z0}{Z0-f} = \frac{Tx}{Tx-d0} \rightarrow d0 = f\,\frac{Tx}{Z0}$

$$\begin{pmatrix} x^R \\ y^R \end{pmatrix} = \begin{pmatrix} x^L \\ y^L \end{pmatrix} - f \begin{pmatrix} \frac{T_x}{Z^L(x^L, y^L)} \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{cases} Z^R(x^R, y^R) &= Z^L(x^L, y^L) \\ d(x^L, y^L) &= x^L - x^R = f\,\frac{T_x}{Z^L(x^L, y^L)} \end{cases} \tag{1}$$

(see figure 1)

This is also the case for images that have been "rectified". We will be working with one image pair ("the pentagon") from the stereo data set http://vasc.ri.cmu.edu/idb/html/stereo/

A Hypothesis Match

Let us consider a pair of left and right images defined on a grid. The epipolar lines are indexed by $y^L = y^R$, and so we can focus on the x coordinate alone. Let us consider a pixel coordinate x_0^L, integer, in the left image and a window of size δ around it. On the left image (our reference), consider templates of size 5×5 pixels, $\delta = 5$ pixels and investigate templates every 4 pixels.

Let us hypothesize that such a window have a correspondence in the right image with another window centered in a pixel coordinate x_0^R, integer, at the corresponding epipolar line. The window displacement can be measured by the centers difference, we refer to it as $d_0(\vec{x}_0^L) = x_0^L - x_0^R$, integer, i.e.,

$$x_0^L \quad \leftrightarrow \quad x_0^R = x_0^L - d_0(\vec{x}_0^L) \tag{2}$$

where $x_0^R, x_0^L, d_0(\vec{x}_0^L) \in \mathbb{Z}$.

Problem 1: Wavelet Transform for Matching Hypothesis

Let us consider the Morlet Wavelets

$$\psi_{\theta,\sigma}(\vec{x}) \;=\; \frac{C_1}{\sigma}\left(e^{i\,\frac{\pi}{2}\vec{e}_\theta\cdot\frac{\vec{x}}{\sigma}} - C_2\right) e^{-\frac{|\vec{x}|^2}{2\sigma^2}} \tag{3}$$

where $\hat{e}_\theta^T = \begin{pmatrix}\cos\theta & \sin\theta\end{pmatrix}$. Since we will work with $\sigma = 2$, we can consider filters of size $13x13$ pixels.

The Morlet wavelet transform of an image is given by the convolution

$$\mathcal{W}_{\sigma,\theta}[I](\vec{u}) \;=\; \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(\vec{x})\,\psi_{\theta,\sigma}\,(\vec{u}-\vec{x})\;d^2\vec{x} \tag{4}$$

Let us consider the "Wavelet-edge transformation" for $\sigma = 2$, i.e., we construct for every pixel $\vec{u}$

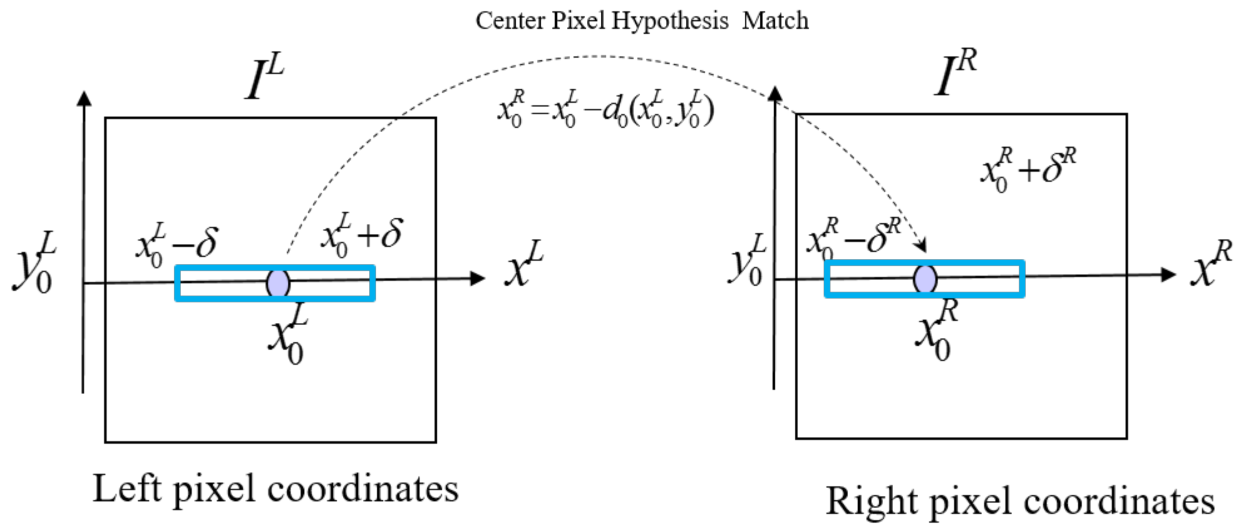

Figure 2: *We show a left and right images, with the blue segment representing the matching line segments, with sizes* $[-\delta, \delta]$ *and* $[-\delta^R, \delta^R]$*, respectively. Points along the segments are represented by* $x^L = x_0^L + \delta x^L$ *and* $x^R = x_0^R + \delta x^R$*. Thus,* $\delta x^L = x^L - x_0^L$ *and* $\delta x^R = x^R - x_0^R$*. The disparity* d_0 *is an unknown we want to recover.*

$$\mathcal{W}e[I](\vec{u}) \quad = \quad \max_{\theta} \left| \mathcal{W}^{imaginary}_{\sigma=2,\theta}[I](\vec{u}) - \mathcal{W}^{real}_{\sigma=2,\theta}[I](\vec{u}) \right| \tag{5}$$

Problem 1: For the pentagon stereo pairs. Show the left and right "Wavelet-edge Transformation" for each image (2 images).

Problem2: Finding the Disparity I

The matching hypothesis for the window in the left image, centered in x_0^L and size $\delta = 5$ pixels, can then be written as

$$\begin{aligned} \mathcal{W}e[I^L](x^L, y_0) &\approx \mathcal{W}e[I^R](x^R, y_0^L) \qquad x^L \in [x_0^L - \delta, x_0^L + \delta] \\ &\approx \mathcal{W}e[I^R](x^L - d_0(\vec{x}_0^L), y_0^L) \end{aligned} \tag{6}$$

The pixel error $\left| \mathcal{W}e[I^L](x^L, y_0) - \mathcal{W}e[I^R](x^L - d_0(x_0^L), y_0) \right|^2$ values small Wavelet responses, so we consider instead the following error

$$\epsilon_{\vec{x}_0^L}(d_0) = \sum_{x^L = x_0^L - \delta}^{x_0^L + \delta} \frac{\mathcal{W}e[I^L](x^L, y_0) + \epsilon}{\mathcal{W}e[I^R](x^L - d_0(x_0^L), y_0) + \epsilon} + \frac{\mathcal{W}e[I^R](x^L - d_0(x_0^L), y_0) + \epsilon}{\mathcal{W}e[I^L](x^L, y_0) + \epsilon} \tag{7}$$

which values relative differences and it is minimized at the optimal choice of $d_0(x_0^L)$. The parameter $\epsilon = 0.001$ avoids divisions by zero.

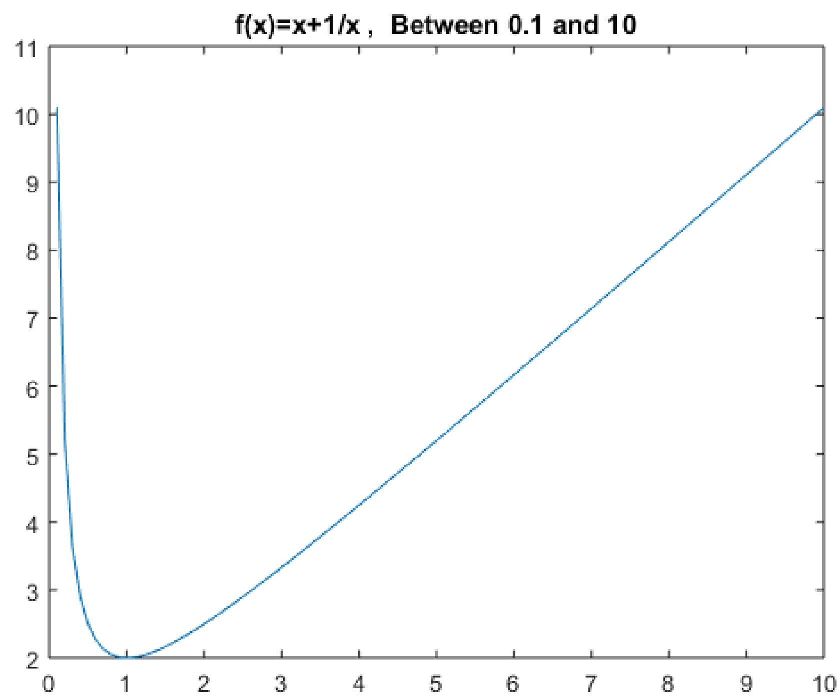

Figure 3: *Plot of the pixel error function,* $f(x) = x + \frac{1}{x}$*, where* $x(d_0) = \frac{\mathcal{W}e[I^L](x^L, y_0) + \epsilon}{\mathcal{W}e[I^R](x^L - d_0(x_0^L), y_0) + \epsilon}$.

Problem 2: Compute the best $d_0(x_0^L)$, according to the error above, for each pixel (x_0^L, y_0^L) in the image. Assume that $-5 < d_0(x_0^L, y_0^L) < 15$, so the disparity range is of 20 pixels (from -5 to 15).

More precisely, vary y_0^L, as $30 < y_0^L <$ H-30, where H is the height of the image. The choice of $\Delta = 30$ is to stay away from the image boundaries. For each y_0^L (epipolar line) vary x_0^L, as $30 < x_0^L <$ W-30, where W is the width of the image, and for each (x_0^L, y_0^L)

search for a disparity $-5 < d_0(x_0^L, y_0^L) < 15$ to find the one that minimizes $\epsilon_{\vec{x}_0^L}(d_0)$. Show the image of $d_0(x_0^L, y_0^L)$ (for each pixel (x^L, y^L) show the disparity d_0 as an image for the stereo the pair of images (1 image of the disparity).