Homework 2: Stereo Matching
Prof. Davi Geiger

Due March 21\textsuperscript{th}, 2017

Introduction

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

\[
\begin{align*}
Z_0 - f &= d_0 = x_0^R - x_0^L \\
0 &= \frac{Z_0 - f}{Tx - d_0} \\
&= f \frac{Tx}{Z_0 - f} \\
\end{align*}
\]

(see figure 1)

\[
\begin{align*}
\begin{pmatrix} x^R \\ y^R \end{pmatrix} &= \begin{pmatrix} x^L \\ y^L \end{pmatrix} - f \begin{pmatrix} \frac{Tx}{Z_0(x^L, y^L)} \\ 0 \end{pmatrix} \\
\end{align*}
\]

\[
\begin{align*}
\begin{pmatrix} Z^R(x^R, y^R) \\ d(x^L, y^L) \end{pmatrix} &= \begin{pmatrix} Z^L(x^L, y^L) \\ x^L - x^R = f \frac{Tx}{Z^L(x^L, y^L)} \end{pmatrix}
\end{align*}
\]

(1)

Figure 1: Projection to parallel cameras. From the triangle on the right we derive \(\frac{Z_0}{Z_0 - f} = \frac{Tx}{Tx - d_0} \rightarrow d_0 = f \frac{Tx}{Z_0} \)
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/

For a larger data set, and with higher resolution and ground truth disparity, see http://vision.middlebury.edu/stereo/data/

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 \( x^L_0 \), integer, in the left image and a window of size \( \delta \) around it. On the left image (our reference), consider templates of size \( 5 \times 5 \) 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 \( x^R_0 \), integer, at the corresponding epipolar line. The window displacement can be measured by the centers difference, we refer to it as \( d_0(x^L_0) = x^L_0 - x^R_0 \), integer, i.e.,

\[
x^L_0 \leftrightarrow x^R_0 = x^L_0 - d_0(x^L_0)
\] (2)

where \( x^R_0, x^L_0, d_0(x^L_0) \in \mathbb{Z} \).

Problem 1: Wavelet Transform for Matching Hypothesis

Let us consider the Morlet Wavelets

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

where \( \vec{e}_\theta^T = (\cos \theta \quad \sin \theta) \). 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

\[
\mathcal{W}_{\sigma,\theta}[I](\vec{u}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(\vec{x}) \psi_{\theta,\sigma}(\vec{u} - \vec{x}) \, d^2\vec{x}
\] (4)
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 $x^L = x^L_0 + \delta x^L$ and $x^R = x^R_0 + \delta x^R$. Thus, $\delta x^L = x^L - x^L_0$ and $\delta x^R = x^R - x^R_0$. The disparity $d_0$ is an unknown we want to recover.

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

\[
\text{We}[I](\vec{u}) = \max_{\theta} \left| W_{\sigma=2,\theta}^{\text{imaginary}} [I](\vec{u}) - W_{\sigma=2,\theta}^{\text{real}} [I](\vec{u}) \right|
\]  

(5)

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

**Problem 2:** Finding the Disparity I

The matching hypothesis for the window in the left image, centered in $x^L_0$ and size $\delta = 5$ pixels, can then be written as

\[
\text{We}[I^L](x^L, y^L_0) \approx \text{We}[I^R](x^R, y^R_0) \quad x^L \in [x^L_0 - \delta, x^L_0 + \delta] \\
\approx \text{We}[I^R](x^L - d_0(x^L_0), y^L_0)
\]  

(6)

The pixel error
\begin{equation}
\epsilon^1_{x_0}(d_0) = \sum_{x^L = x_0^L - \delta}^{x_0^L + \delta} \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
\end{equation}

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.

![Graph](image.png)

**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 errors 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 $y^L_0$, as $30 < y^L_0 < H-30$, where $H$ is the height of the image. The choice of $\Delta = 30$ is to stay away from the image boundaries. For each $y^L_0$ (epipolar line) vary $x^L_0$, as $30 < x^L_0 < W-30$, where $W$ is the width of the image, and for each $(x^L_0, y^L_0)$ search for a disparity $-5 < d_0(x^L_0, y^L_0) < 15$ to find the one that minimizes $\epsilon_{x^L_0}(d_0)$. Show the image of $d_0(x^L_0, y^L_0)$ (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).

**Problem 3: Dynamic Programming**

Say for each epipolar we have the error function $\epsilon(x^L, x^R)$, i.e., for each $x^L$ it gives us the error of matching $x^L - x^R = d(x^L)$. Now we devise the following dynamic programming (DP) algorithm.

Say we ask the following subproblem at match point $(x^L, x^R)$: "What is the optimal cost for the solution to pass through $(x^L, x^R)$?" Such a path must come from either of these three previous match points $\{(x^L-1, x^R), (x^L, x^R-1), (x^L-1, x^R-1)\}$ (see figure 6). We assume such the subproblems at these previous matches have already been
solved. The cost of transition from them to \((x^L, x^R)\) is defined as follows

1. From \((x^L - 1, x^R) \rightarrow (x^L, x^R)\), \(\text{CostTransition}[(x^L - 1, x^R); (x^L, x^R)] = \text{Occ. Fixed Cost}\).

2. From \((x^L, x^R - 1) \rightarrow (x^L, x^R)\), \(\text{CostTransition}[(x^L, x^R - 1); (x^L, x^R)] = \text{Occ. Fixed cost}\).

3. From \((x^L - 1, x^R - 1) \rightarrow (x^L, x^R)\), \(\text{CostTransition}[(x^L - 1, x^R - 1); (x^L, x^R)] = \epsilon(x^L, x^R). \text{The Cost to Match}\).

Thus, the solution to the subproblem is given by

\[
\begin{align*}
\text{BestCost}[x^L, x^R] &= \min\{\text{Occ} + \text{BestCost}[x^L - 1, x^R], \text{Occ} + \text{BestCost}[x^L, x^R - 1], \\
&\quad\epsilon_{x^L-1}(x^L - x^R) + \text{BestCost}[x^L - 1, x^R - 1]\}
\end{align*}
\]

\[
\begin{align*}
\text{BestPrev}[x^L, x^R] &= \arg \min\{\text{Occ} + \text{BestCost}[x^L - 1, x^R], \text{Occ} + \text{BestCost}[x^L, x^R - 1], \\
&\quad\epsilon_{x^L-1}(x^L - x^R) + \text{BestCost}[x^L - 1, x^R - 1]\}
\end{align*}
\]

(9)

For each epipolar line, we can then loop for \(x^L = \text{start}, \ldots, \text{end}\) and for each \(x^L\) visit all \(x^R\) within the disparity range \(d \in [D_{\text{min}}, D_{\text{max}}]\), i.e., where \(D_{\text{min}} \leq x^L - x^R \leq D_{\text{max}} \rightarrow x^L - D_{\text{max}} < x^R < x^L - D_{\text{min}}\).
The Pseudo Code looks like

**Forward Algorithm:**

For each epipolar line $y^L$

For $x^L = x^L_{\text{start}}, \ldots, x^L_{\text{end}}$

For $x^R = x^L - D_{\text{max}}, \ldots, x^L - D_{\text{min}}$

$$\text{BestCost}[x^L, x^R] = \min\{\text{Occ} + \text{BestCost}[x^L - 1, x^R], \text{Occ} + \text{BestCost}[x^L, x^R - 1],$$

$$\epsilon_{x^L-1}(x^L - x^R) + \text{BestCost}[x^L - 1, x^R - 1]\}$$

$$\text{BestPrev}[x^L, x^R] = \arg\min\{[x^L, x^R - 1], [x^L - 1, x^R], [x^L - 1, x^R - 1]\}$$

end

end

**Backtrack Algorithm:**

For each epipolar line $y^L$

Find best $x^R$ at $x^L_{\text{end}}$, i.e.,

$$x_{\text{best}} = \arg\min_{x_R} \text{BestCost}[x^L_{\text{end}}, x^R]$$
Find best previous \([x^L, x^R]_i^* = \text{BestPrev}[x^L_{\text{end}}, x^R_{\text{best}}]\).

Iterate: at stage \(i\) one has best solution \([x^L, x^R]_i^*\) and we find
\([x^L, x^R]_{i-1}^* = \text{BestPrev}[x^L, x^R]_i^*\) until we reach \(x^L_{\text{start}}\).

For each pair \([x^L, x^R]_i^*\) recover the disparity \(d_i(x^L, y^L) = x^L - x^R\).

Show the disparity image for the Pentagon stereo pair: range of disparity \(D_{\text{min}} = -5\) and \(D_{\text{max}} = 15\).

Bonus, show the disparity image for the Piano stereo pair: range of disparity \(D_{\text{min}} = 35\) and \(D_{\text{max}} = 210\).