Homework 2.A: An investigation on the stereo matching criteria

Davi Geiger

Due March 1st, 2016

Introduction

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

\[ r = l - \frac{T_x}{Z_l(e,l)} \quad e = e^R = e^L \quad Z^R(e,r) = Z^L(e,l) \]

\[ d(e,l) = l - r = \frac{T_x}{Z_l(e,l)} \quad (1) \]

where $Z^L_l$ is the depth from the left coordinate view. This is also the case for images that have been "rectified". We will be working with an image pair from the middlebury 2014 stereo data set (http://vision.middlebury.edu/stereo/data/2014/).

We will work with the image pair

http://vision.middlebury.edu/stereo/data/2014/datasets/Piano-perfect/im0.png
http://vision.middlebury.edu/stereo/data/2014/datasets/Piano-perfect/im1.png

and the ground truth disparity

http://vision.middlebury.edu/stereo/data/2014/datasets/Piano-perfect/disp0.pfm
http://vision.middlebury.edu/stereo/data/2014/datasets/Piano-perfect/disp1.pfm

Question 1: Occlusions

The first step is to detect the occlusion regions. The input is the two disparity maps provided in the web site (one is the ground truth for the left image, $d^L_{true}$, and the other is the ground truth for the right image, $d^R_{true}$).

How to find the occlusion regions? It is possible that disparities assigned the value infinity ($\infty$) are the occlusions. Can you check that with the alternative method? An
alternative method is to generate an occlusion map by crosschecking the pair of disparity maps. More precisely, for each epipolar line, \( e = 1, \ldots, N \), and pixel \( l \) along the line, and its correspondent match \( r = l - d^L_{\text{true}}(l) \), check that \( d^L_{\text{true}}(l) = d^R_{\text{true}}(r) \). If \( |d^L_{\text{true}}(l) - d^R_{\text{true}}(r)| < 0.3 \) pixels, then there is no occlusion, i.e., \( O^L_l = 0 \). Else if \( |d^L_{\text{true}}(l) - d^R_{\text{true}}(r)| \geq 0.3 \) pixels, then it is occluded, i.e., \( O^L_l = 1 \). Do you get the same answer as the places where the disparity is assigned \( \infty \) or there are new occlusions found this way?

In this way, you construct an occlusion variable, \( O^L_l \) (left eye) that is 0 or 1.

Create an occlusion variable for the left eye, and show the image on the left eye with pixels \( I(l,e) = (0,0,0) \) ("black") if the pixel is such that \( O^L_l = 1 \).

**Question 2: A First Algorithm for Stereo**

On the left image (our reference), consider templates of size 37 \( \times \) 37 pixels.

![Matching Space](image)

Figure 1: Matching Space. We will fill these nodes with data, i.e., the matching of left and right pixel data. We will fill the array from the left eye coordinate, \( l \). Let us refer to this array as \( M(e,l,w) \). Let us only use the scale \( \sigma = 6 \) data and two directions \( \theta = 0, \frac{\pi}{2} \).

The matching at \( l \) is from the Wavelet responses at different angle directions. Let us consider only two directions \( \theta = 0, \frac{\pi}{2} \). We are using the coordinate system \( (l,w) \). At pixels where there is enough contrast, we use the wavelet response to compare, and at pixels where the wavelet response is very low, for now, we simply ignore. More precisely, we create the array \( M(e,l,w) \) as follows.
Extract from the ground truth left disparity (disp0) the minimum disparity value $D_{\text{min}}$ and maximum disparity value $D_{\text{max}}$.

Initialize the arrays

\[ M(e,l,w) = \infty; \quad \text{stores the matching values} \]
\[ \text{Prev}(e,l) = 0; \quad \text{stores the previous pixel with a “good” matching value} \]
\[ w(e,l) = 0; \quad \text{stores the “best” disparity match} \]

Main Loop

for $e = 7, \ldots, N - 7$

\[ \text{count}(e) = 0; \]
\[ l_{\text{prev}} = 0; \]

for $l = 7, \ldots, N - 7$

Left-image-wavelet response $WI^L(\theta, \sigma = 6, e, l)$ are not computed beyond these limits.

if (O(l) = 0) && (\exists \theta^L \text{ where } |WI^L(\theta^L, \sigma = 6, e, l)| > \text{Tolerance})

TempMinDiff = $\infty$;

for $w = D_{\text{min}}, \ldots, D_{\text{max}}$

\[ r = l - w; \quad \text{ (check } 7 \leq r \leq N - 7) \]

if \( |WI^R(\theta^L, \sigma = 6, e, r)| > \text{Tolerance} \) (note it is the same $\theta^L$)

\[ \text{Diff} = WI^R(\theta^L, \sigma = 6, e, r) - WI^L(\theta^L, \sigma = 6, e, l) \]
\[ M(e,l,w) = |\text{Diff}| \]

if \( |\text{Diff}| < \text{TempMinDiff} \)

\[ w_{\text{best}}(e,l) = w; \]
\[ \text{TempMinDiff} = |\text{Diff}|; \]

end

\[ \text{Prev}(e,l) = l_{\text{prev}}; \]
\[ l_{\text{prev}} = l; \]
\[ \text{count}(e) = \text{count}(e) + 1; \quad \text{number of pixels that are evaluated} \]

end

end

\[ \text{Prev}(e, N - 6) = l_{\text{prev}}; \text{ store the last one at epipolar line } e \]

end
Note that $|Diff| = \sqrt{Diff \times \text{conjugate}(Diff)}$. We can construct a list of all pixels that have passed the test as follows.

*For each epipolar line retrieve the list of all pixels that are evaluated by the Wavelet Measure*

```latex
for e = 7, \ldots, N - 7
    for i = count(e) : 2 steps of -1
        l(e,i-1)=\text{Prev}(e,i);
    end
end
```

**Question 3: Analysis and comparison to ground truth**

**Errors of disparity estimate:** We investigate every pixel where the wavelets transform of the image can offer a solution. They are stored in $l(e,i-1)$. We do have the true disparity value at these pixels, $d_{\text{true}}^L(e,l)$, so we can compare to our estimates (stored at $w_{\text{best}}^L(e,l)$).

3a Create an image-like of $\text{error}(e,l) = |d_{\text{true}}^L(e,l) - w_{\text{best}}^L(e,l)|$, for the pixels in the list $l(e,i)$. Display it so that we can see the errors. Are they large? small ?

3b. Create an image-like of $M(e,l, w_{\text{best}}^L(e,l))$, for the pixels in the list $l(e,i)$. Display it so that we can see the matching quality (how small is the matches). Are they large? small ?

**Plot a histogram**: We build the histogram $H(|Diff(e,l(e,i),d_{\text{true}}^L)|)$, a voting graph over all points in the list $l(e,i)$. It is constructed by having the "x-axis" to be the possible values for the magnitude of the differences $|Diff| = |WIR(\theta^L, \sigma = 6, e,l - d_{\text{true}}^L(e,l)) - WIL(\theta^L, \sigma = 6, e,l)|$ for pixels in the list $l(e,i)$. The "y-axis" is the accumulation of votes for each error.

The trick is to create bins of error range. How to set the bins? it is a bit of an art. "Try your best" and make sure that your program has this parameter (the range of each bin) easy to be changed, so we can discuss in class.

Our goal in building the histogram is to find the threshold that guarantees us to include the ground truth disparity, given our matching criteria.