Disparities are encoded using a scale factor 4 for gray levels 1 .. 255, while gray level 0 means "unknown disparity". Therefore, the encoded disparity range is 0.25 .. 63.75 pixels. We will use the notation that $D_{\text{min}} = -63$ pixels and $D_{\text{max}} = 1$ pixels. Since no disparity is 0, use 0 disparity for the occlusion pixels.

**FILTERS: COMBINING THEM**

So far we used a feature and feature distance for stereo that is the following combination of two features (and their distances)

$$C_e[l,w+D] = 0.1 \left( \sqrt{W_{\theta = 0, s = 3}(e, l, r(l, w))} + \sqrt{W_{\theta = \pi/2, s = 3}(e, l, r(l, w))} \right)$$

where the feature distance $W_{\theta,s}(e, l, r(l, w))$ is defined in terms of the features $I^L$ and $I^R$ as follows

$$W_{\theta,s}(e, l, r(l, w)) = \min \left( \left( I^L(e, l, \theta, s) - I^R(e, r, \theta, s) \right)^2, \left( I^L(e, l, \theta + \pi, s) - I^R(e, r, \theta + \pi, s) \right)^2 \right)$$

Why did we use the same coefficients for both, $\theta = 0$ and $\theta = \pi/2$ comparisons? What if we use another feature, say intensity grey level comparison, such as $W_i(e, l, r(l, w)) = (I^L(e, l) - I^R(e, r))^2$, how should it be combined with the measures above? We could also use color features, a feature for each channel R, G, B or the average intensity I and ratios such as $r = 255*(R/I)$ and $g = 255*(G/I)$ (blue channels are more noisy), where $I = R+G+B$. How to combine them?

Let us study the case we have n features and feature distances, let us call the distance $F_i(e, l, w); i = 1, ..., n$. A feature distance could be for example $F_i(e, l, w) = W_{\theta,s}(e, l, r(l, w))$. The smaller the value of a feature distance at $(e,l)$ for a disparity $w$, the more likely the disparity is $w$ (the more similar are the image features at left and right image). Let us work with five feature distances

$$F_1 = \sqrt{W_i(e, l, r(l, w))} = AB(S\left(\frac{1}{3} I^L(e, l) - \frac{1}{3} I^R(e, r)\right),$$
\[ F_2 = \sqrt{W_{\theta=0,s=3}(e, l, r(l, w))}, \]
\[ F_3 = \sqrt{W_{\theta=\pi/2,s=3}(e, l, r(l, w))}, \]
\[ F_4 = \sqrt{W_r(e, l, r(l, w))} = 255 * \text{ABS} \left( \frac{R^l(e, l)}{t^l(e, l)} - \frac{R^R(e, r)}{t^R(e, r)} \right), \]
\[ F_5 = \sqrt{W_r(e, l, r(l, w))} = 255 * \text{ABS} \left( \frac{G^l(e, l)}{t^l(e, l)} - \frac{G^R(e, r)}{t^R(e, r)} \right). \]

where \( t^l(e, l) = R^l(e, l) + G^l(e, l) + B^l(e, l) \) and \( t^R(e, r) = R^R(e, r) + G^R(e, r) + B^R(e, r) \)

So one possible definition of a probability of a feature at a given location to have a match at a disparity \( w \) is given by

\[ P_{i,e,l}(w) = \frac{e^{-F_i(e,l,w)}}{\sum_{w=-D_{\text{min}}}^{D_{\text{max}}} e^{-F_i(e,l,w)}} \]

where \( i = 1,2,3,4,5 \). There is an ambiguity as we can chose a weight \( \alpha \) to scale the feature distance without affecting the ranking of the probabilities. So we may refer to each feature distance as being the feature similarity comparison multiplied by some scale factor, we call it \( \alpha \). In the previous homework we used \( \alpha=0.1 \). This factor could vary locally to be \( \alpha_i(e,l,w) \) for each feature distance \( i \).

Taking into account this factor, we should define the probabilities as

\[ P_{i,e,l}(w) = \frac{e^{-\alpha_i(e,l,w) F_i(e,l,w)}}{\sum_{w=-D_{\text{min}}}^{D_{\text{max}}} e^{-\alpha_i(e,l,w) F_i(e,l,w)}} = \frac{e^{-\alpha_i(e,l,w) F_i(e,l,w)}}{Z_w} \]

How to estimate these parameters \( \alpha_i(e,l,w) \)? We do have the true disparity for these images and so we do have the correspondence between pixels. We can evaluate the value of \( F_i(e,l,w) \) for each pixel of the image and its correspondence.

1. Construct a histogram for each \( F_i(e,l,w) \), i.e., for each value \( E_i = F_i(e,l,w) \) define an interval \( \Delta \) and count the number of matched pixels falling in this interval. Normalize the histogram by dividing the count by the number of pixels considered. Plot the histogram. This histogram represents the probability \( P(E_i = F_i(e,l,w)) = \frac{e^{-\alpha_i(e,l,w) E_i}}{Z_E} \), so we need to extract the parameter \( \alpha_i(e,l,w) \) from it.

2. One can see that

\[ \log P_i(E_i = F_i(e,l,w)) = -\alpha_i(e,l,w) E_i - \log Z_E \]

and so

\[ \alpha_i(E_i = F_i(e, l, w)) = -\frac{\partial \log P_i(E_i = F_i(e,l,w))}{\partial E_i} = -\frac{1}{P_i(E_i)} \frac{\partial P_i(E_i)}{\partial E_i} \]

Thus, we can take these derivatives of the histogram to obtain the local parameters. Note that these parameters vary with \( E_i = F_i(e,l,w) \) so plot \( \alpha_i(E_i) \) per feature \( i \) and report the average value \( \bar{\alpha}_i \) per image, i.e.,
\[
\bar{\alpha}_i = \max_{E_i} \sum_{E_i=0}^{\max E_i} \alpha_i(E_i) \cdot P_i(E_i)
\]

3. Disparity Changes. Create the histogram of disparity changes. For each two consecutive epipolar-line-pixels on all horizontal lines (epipolar lines), compute the magnitude of the disparity change and add to one count for its bin on the histogram. So the x-axis of the histogram represent the possible values of the disparity change, i.e., it should vary from 0 to \(D_{\text{max}} - (D_{\text{min}}) = D_{\text{max}} + D_{\text{min}}\). Normalize the histogram by dividing the bin counts by the total number of counts (total number of consecutive pixels considered). This histogram generates

\[
Pr(E = |w_{e,l} - w_{e,l-1}|) = \frac{e^{-\lambda(e,l,w) E}}{Z_E}
\]

so we need to extract the parameter \(\lambda(e,l,w)\) from it.

4. One can see that

\[
\log Pr(E = |w_{e,l} - w_{e,l-1}|) = -\lambda(e,l,w) E_i - \log Z_E
\]

and so

\[
\lambda(E = |w_{e,l} - w_{e,l-1}|) = -\frac{\partial \log Pr(E = |w_{e,l} - w_{e,l-1}|)}{\partial E_i} = -\frac{1}{Pr(E)} \frac{\partial Pr(E)}{\partial E}
\]

Thus, we can take these derivatives of the histogram to obtain the local parameters. Note that the parameter depends on \(E_i = |w_{e,l} - w_{e,l-1}|\), so plot \(\lambda_i(E_i)\) and report the average value \(\bar{\lambda}_i\) per image

\[
\bar{\lambda} = \sum_{E=0}^{\max E} \lambda(E) \cdot Pr(E)
\]

5. Use these parameters learned from these images and apply to the Tsukuba image. Note that in the case of the \(\lambda_i(E_i)\) we only use \(\lambda\) since the dynamic programing would have to be modified to allow for local variations. However, for \(\alpha_i(E_i)\) we can use a local value that depends on the feature matching error \(E_i = f_i(e,l,w)\).