Binocular Stereo Vision

Region-based stereo matching algorithms

Processing stereo images

left

right

- extract features from the left and right images, whose stereo disparity will be measured

- match the left and right image features and measure their disparity in position

  “stereo correspondence problem”

- use stereo disparity to compute depth
Constraints on stereo correspondence

• Uniqueness
  each feature in the left image matches with only one feature in the right (and vice versa…)

• Similarity
  matching features appear “similar” in the two images

• Continuity
  nearby image features have similar disparities

• Epipolar constraint
  simple version: matching features have similar vertical positions (in general, stereo projection is more complex)

Solving the stereo correspondence problem
Measuring goodness of match between patches

(1) sum of absolute differences

\[
\frac{1}{n} \sum_{\text{patch}} | p_{\text{left}} - p_{\text{right}} |
\]

optional: divide by \( n \), number of pixels in patch

(2) normalized correlation

\[
\frac{1}{n} \sum_{\text{patch}} \frac{(p_{\text{left}} - \bar{p}_{\text{left}})(p_{\text{right}} - \bar{p}_{\text{right}})}{\sigma_{p_{\text{left}}} \sigma_{p_{\text{right}}}}
\]

\( \sigma = \) standard deviation of values within patch

Region-based stereo matching algorithm

for each row \( r \)

for each column \( c \)

let \( p_{\text{left}} \) be a square patch centered on \((r, c)\) in the left image

initialize best match score \( m_{\text{best}} \) to \( \infty \)

initialize best disparity \( d_{\text{best}} \)

for each disparity \( d \) from \(-d_{\text{range}}\) to \(+d_{\text{range}}\)

let \( p_{\text{right}} \) be a square patch centered on \((r, c+d)\) in the right image

compute the match score \( m \) between \( p_{\text{left}} \) and \( p_{\text{right}} \)

(sum of absolute differences) \( \text{(normalized correlation)} \)

if \( m < m_{\text{best}} \), assign \( m_{\text{best}} = m \) and \( d_{\text{best}} = d \)

record \( d_{\text{best}} \) in the disparity map at \((r, c)\)

How are the assumptions used??
The real world works against us sometimes...

Example: Region-based stereo matching, using filtered images and sum of absolute differences

(from Carolyn Kim, 2013) (results before improvements)