Binocular Stereo Vision

Region-based stereo matching algorithms
Properties of human stereo processing

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 \( n = \text{number of pixels in patch} \)

(2) normalized correlation

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

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) (normalized correlation)
if \((m \times n_{\text{best}})\), 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)
Properties of human stereo processing

Use features for stereo matching whose position and disparity can be measured very precisely.

Stereoacuity is only a few seconds of visual angle

\[
\text{difference in depth} \approx 0.01 \text{ cm at a viewing distance of 30 cm}
\]

Matching features must appear similar in the left and right images.

For example, we can’t fuse a left stereo image with a negative of the right image...
Properties of human stereo processing

Only “fuse” objects within a limited range of depth around the fixation distance. **Vergence** eye movements are needed to fuse objects over larger range of depths.

Properties of human stereo vision

We can only tolerate small amounts of *vertical disparity* at a single eye position. Vertical eye movements are needed to handle large vertical disparities.
Properties of human stereo processing

In the early stages of visual processing, the image is analyzed at multiple spatial scales...

Stereo information at multiple scales can be processed independently

Neural mechanisms for stereo processing

G. Poggio & colleagues: complex cells in area V1 of primate visual cortex are selective for stereo disparity neurons that are selective for a larger disparity range have larger receptive fields

zero disparity: at fixation distance
near: in front of point of fixation
far: behind point of fixation
In summary, some key points...

- Image features used for matching: simple, precise locations, multiple scales, similar between left/right images
- At single fixation position, match features over a limited range of horizontal & vertical disparity
- Eye movements used to match features over larger range of disparity
- Neural mechanisms selective for particular ranges of stereo disparity