

SHAPE FROM X

One image:

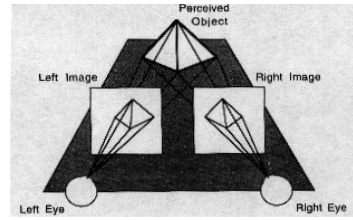
- Texture
- Shading

Two images or more:

- **Stereo**
- Contours
- Motion



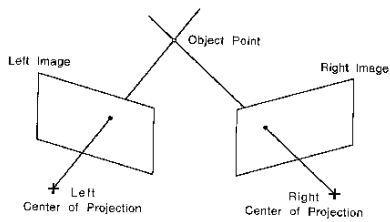
GEOMETRIC STEREO



Depth from two or more images:

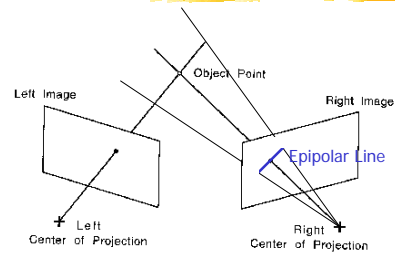
- Geometry of image pairs
- Establishing correspondences

TRIANGULATION



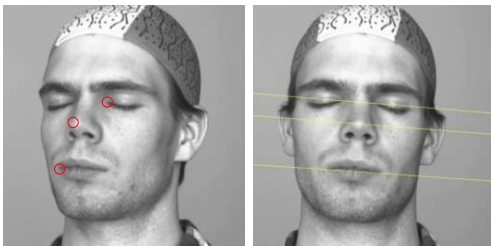
Geometric Stereo: Depth from two images

EPIPOLAR LINE



Line on which the corresponding point must lie.

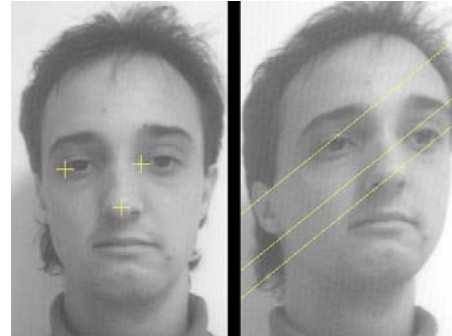
EPIPOLAR LINES



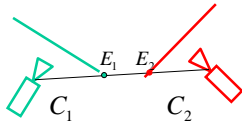
Three points shown as red crosses.

Corresponding epipolar lines.

EPIPOLAR LINES



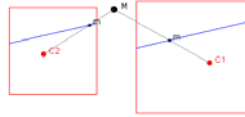
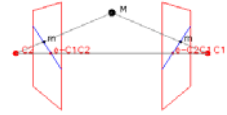
EPIPOLE



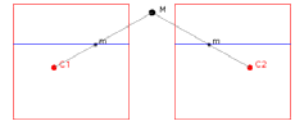
Point at which **all** epipolar lines intersect:
 Located at intersection of line between optical centers and image plane.

EPIPOLAR GEOMETRY

In general:

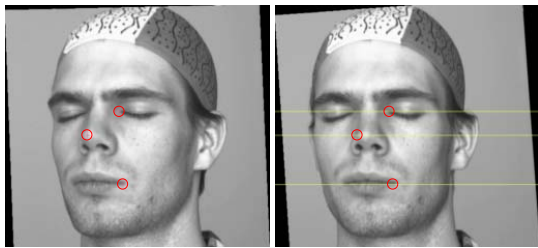


Parallel image planes



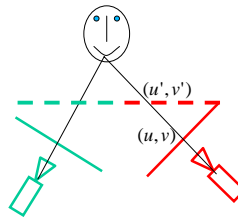
Horizontal baseline

RECTIFICATION



Parallel epipolar lines

RECTIFICATION



$$\begin{bmatrix} U' \\ V' \\ W' \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & 1 \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}$$

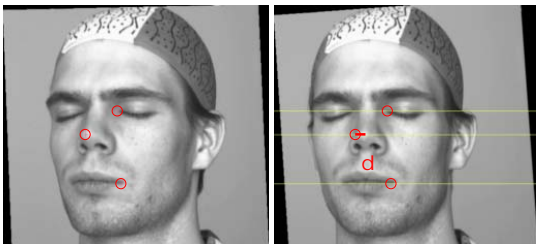
$$u' = U'/W'$$

$$v' = V'/W'$$

Reprojection into parallel virtual image planes:

- Linear operation in projective coordinates
- Real-time implementation possible

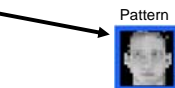
DISPARITY



Horizontal shift along epipolar line, inversely proportional to distance.

FINDING A PATTERN IN AN IMAGE

Straightforward Approach



Move pattern everywhere and compare with image.

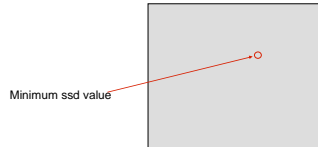
But how?

SUM OF SQUARE DIFFERENCES

Subtract pattern and image pixel by pixel and add squares:

$$ssd(u, v) = \sum_{(x, y) \in N} [I(u+x, v+y) - P(x, y)]^2$$

If identical $ssd=0$, otherwise $ssd > 0$
 → Look for minimum of ssd with respect to u and v .



CORRELATION

$$ssd(u, v) = \sum_{(x, y) \in N} [I(u+x, v+y) - P(x, y)]^2$$

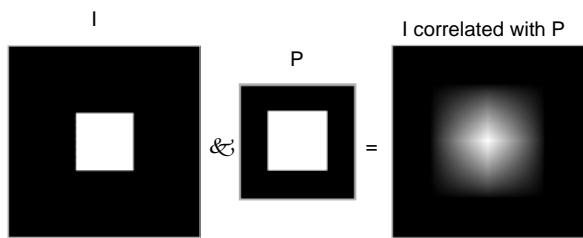
$$= \sum_{(x, y) \in N} I(u+x, v+y)^2 + \sum_{(x, y) \in N} P(x, y)^2 - 2 \sum_{(x, y) \in N} I(u+x, v+y)P(x, y)$$

↑ Sum of squares of the window (positive term)
 ↑ Sum of squares of the pattern (CONSTANT term)
 ↑ Correlation

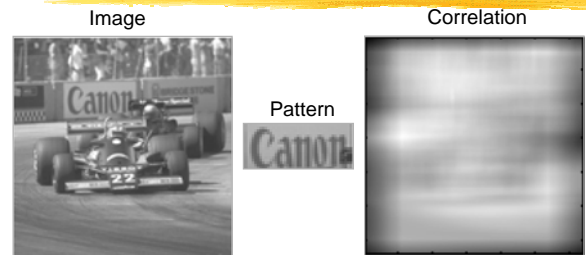
$ssd(u, v)$ is minimized when correlation is largest

→ Correlation measures similarity

SIMPLE EXAMPLE



NOT SO SIMPLE EXAMPLE



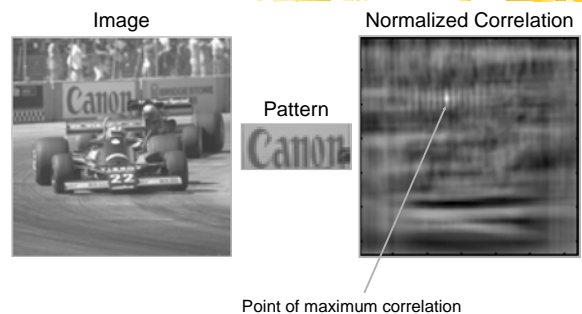
- Correlation value depends on the local gray levels of the pattern and image window
- Need to normalize

NORMALIZED CROSS CORRELATION

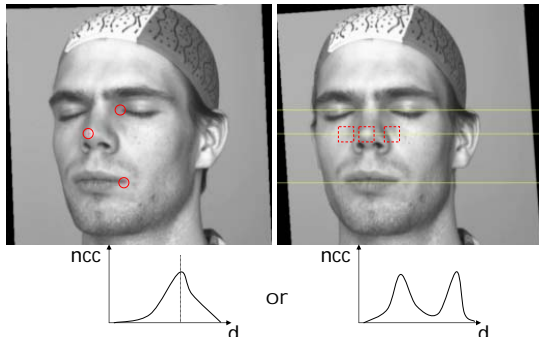
$$ncc(u, v) = \frac{\sum_{(x, y) \in N} [I(u+x, v+y) - \bar{I}][P(x, y) - \bar{P}]}{\sqrt{\sum_{(x, y) \in N} [I(u+x, v+y) - \bar{I}]^2 \sum_{(x, y) \in N} [P(x, y) - \bar{P}]^2}}$$

- Between -1 and 1
- Invariant to linear transforms
- Independent of the average gray levels of the pattern and the image window

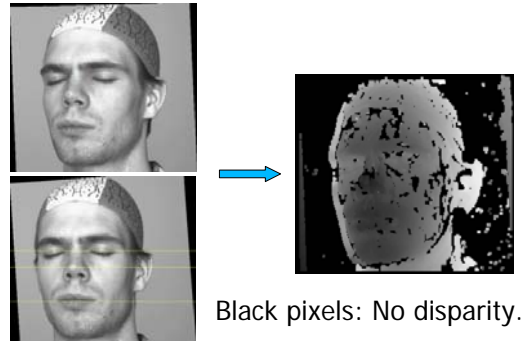
NORMALIZED EXAMPLE



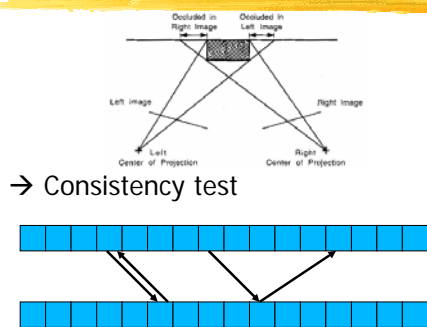
SEARCHING ALONG EPIPOLAR LINE



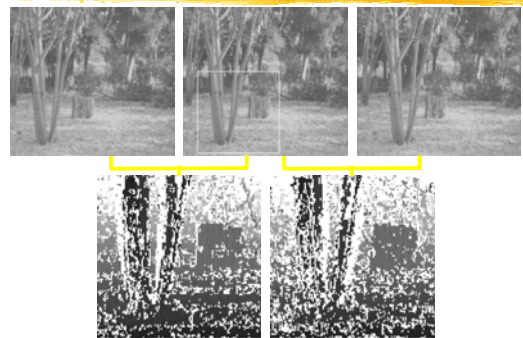
DISPARITY MAP



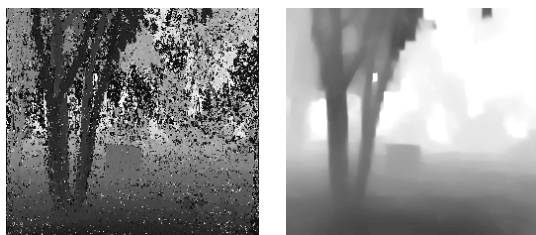
OCCLUSSIONS



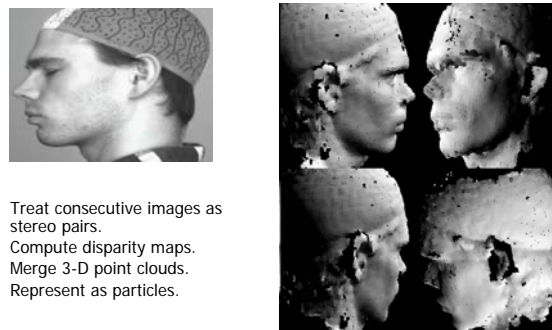
GROUND LEVEL STEREO



COMBINING DISPARITY MAPS

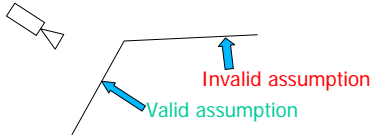


SHAPE FROM VIDEO



FRONTO-PARALLEL ASSUMPTION

The disparity is assumed to be the same in the whole correlation window, which is equivalent to assuming constant depth.



→ Ok when the surface faces the camera but breaks down otherwise.

WINDOW SIZE

Small windows:

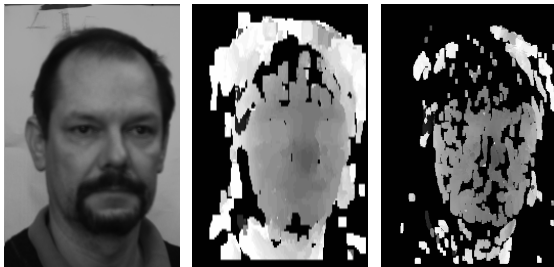
- Good precision
- Sensitive to noise

Large windows:

- Diminished precision
- Increased robustness to noise

→ Same kind of trade-off as in the case of edge-detection.

WINDOW SIZE



14x14

7x7

SCALE-SPACE REVISITED



Gaussian pyramid

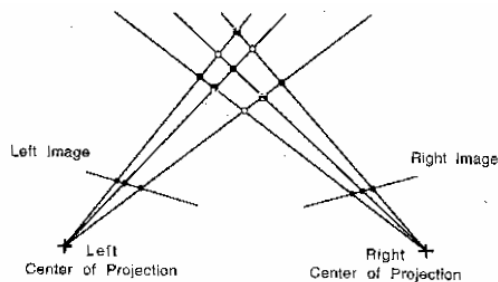
Difference of Gaussians

• Using a small window on a reduced image is equivalent to using a large one on the original image.

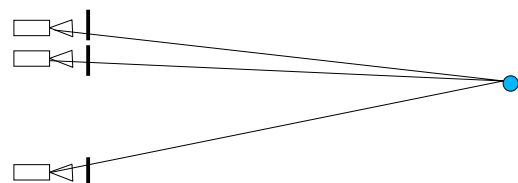
• Using difference of Gaussian images is an effective way of achieving normalization.

→ It becomes natural to use results obtained using low resolution images to guide the search at higher resolution.

UNCERTAINTY



SHORT vs LONG BASELINE



Short baseline:

- Good matches
- Few occlusions
- Poor precision

Long baseline:

- Harder to match
- More occlusions
- Better precision

MARS ROVER



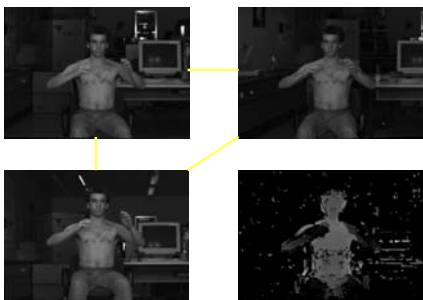
There are four cameras!

VIDEO-BASED MOTION CAPTURE



Fitting an articulated body model to stereo data.

TRINOCULAR STEREO



MULTI-CAMERA CONFIGURATIONS

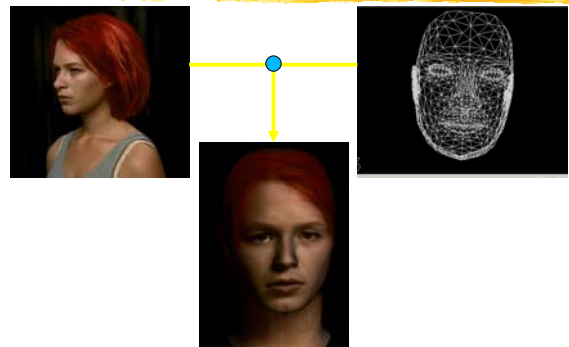
- ● ● 3 cameras give both robustness and precision
- ● ● ● 4 cameras give additional redundancy
- ● ● 3 cameras in a T arrangement allow the system to see vertical lines.

FACES FROM MONOCULAR SEQUENCES

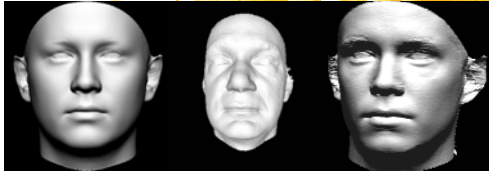


- No calibration data
- Relatively little texture
- Difficult lighting

SIMPLE FACE MODEL



PCA FACE MODEL

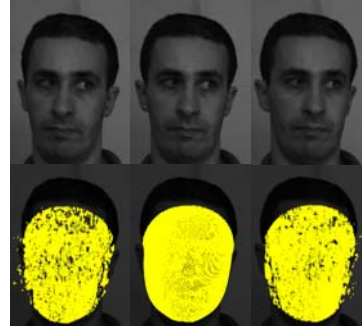


$$S = \bar{S} + \sum_{i=1}^{99} \alpha_i S_i$$

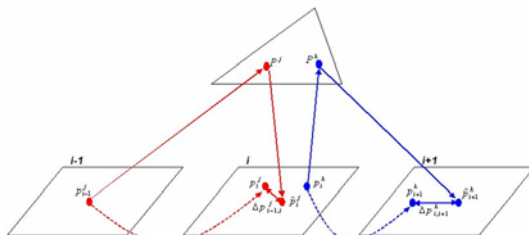
\bar{S} : Average shape
 S_i : Shape vector
 α_i : Shape coefficients

V. Blanz and T. Vetter, "A Morphable Model for the Synthesis of 3-D Faces" in Computer Graphics, SIGGRAPH Proceedings, Los Angeles, CA, August 1999.

CORRESPONDENCES



TRANSFER FUNCTION



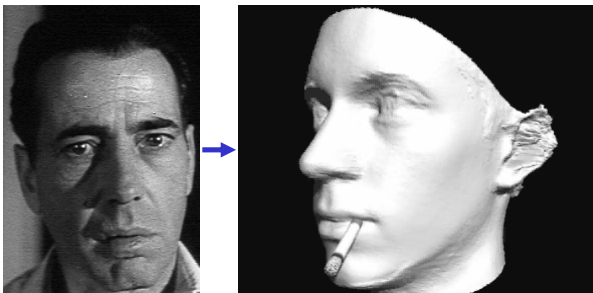
$$F_3(A, C_{i-1}, C_i, C_{i+1}) = \sum_{j \in Q_{i-1}} \|\Delta p_{i-1}^j\|^2 + \sum_{k \in Q_i} \|\Delta p_{i+1}^k\|^2$$

MODEL BASED BUNDLE ADJUSTMENT



→ Median accuracy greater than 0.5mm

MODEL FROM OLD MOVIE



ENERGY MINIMIZATION

Disparity continuous in most places,



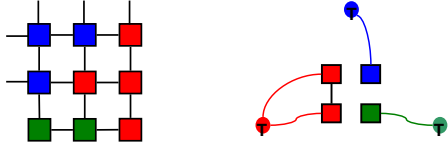
except at depth discontinuities

1. Matching pixels should have similar intensities.
2. Most nearby pixels should have similar disparities

→ Minimize

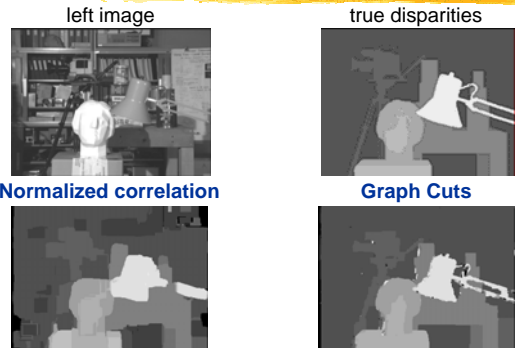
$$\sum [I_1(x+D(x,y), y) - I_2(x,y)]^2 + \lambda \sum [D(x+1,y) - D(x,y)]^2 + \mu \sum [D(x,y+1) - D(x,y)]^2$$

GRAPH CUTS

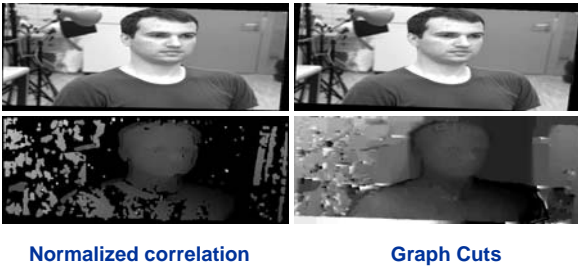


1. Stereo is a labeling problem
 2. Graph cut corresponds to a labeling.
- Assign edge weights cleverly so that the min-weight cut gives the minimum energy!

NCC vs GRAPH CUTS



NCC vs GRAPH CUTS



STRENGTHS AND LIMITATIONS

Strengths:

- Practical method for recovering depth.
- Runs in real-time on ordinary hardware.

Limitations:

- Requires multiple views.
- Only applicable to reasonably textured objects.