Stereo Matching by Neural Networks

Houman Rastgar

Research Centre for Integrated Microsystems
Electrical and Computer Engineering
University of Windsor

Supervisors: Dr. Sid-Ahmed

May 6, 2005

OUTLINE

- Research Objective
- Background – Image Formation
- Introduction to 3D Vision
- Stereo Vision
- Camera Calibration
- Correspondence
  - Area based approaches
  - Feature-based approaches
  - Comparison
- Correspondence – Matching Constraints
- Artificial Neural Network approach
  - Feature Extraction
  - Hopfield Neural Network
  - Occlusion
- Future Work

Research Objective

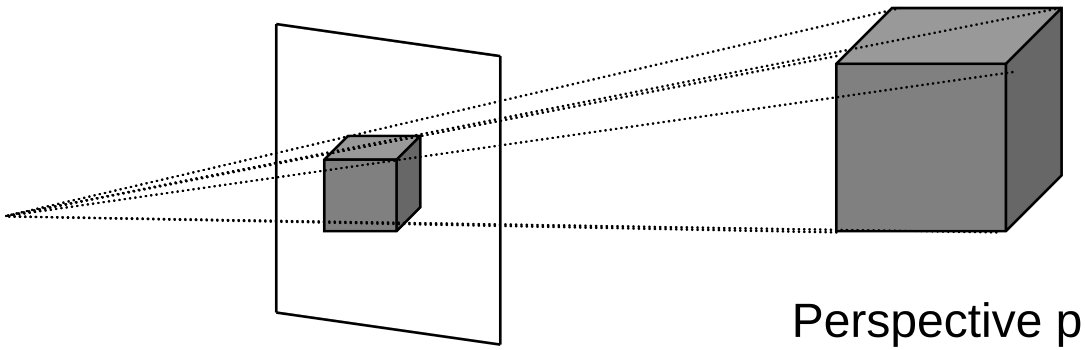
- Develop a Vision System for a robotics close-range position sensing system
- In other words, develop a system that will enable the robot to "see" and interact with its environment
- The algorithm takes images of the environment as input, the output is 3D location of objects in view plus orientation -> give commands to robot to grab object
- Need a way to extract 3D information form images -> stereo vision

Background- Image Formation

- **Image:** is a 2D projection of a 3D scene.
  Mapping from 3D to 2D, i.e., some information is getting lost

- **Computer vision problem:** recover (some or all of) that information.
  The lost dimension 2D → 3D

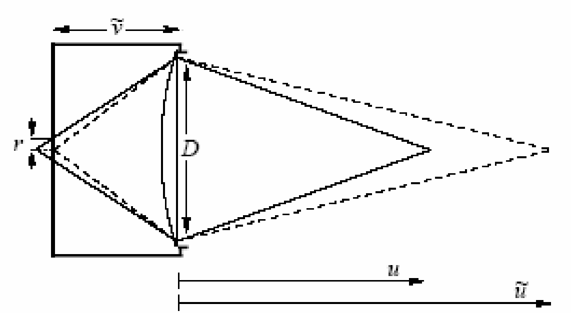
- **3D Vision Goals:**
  - **Reconstruction**: recover a model of the 3D scene from 2D images in order to accomplish close range position sensing



Perspective projection

3D Vision

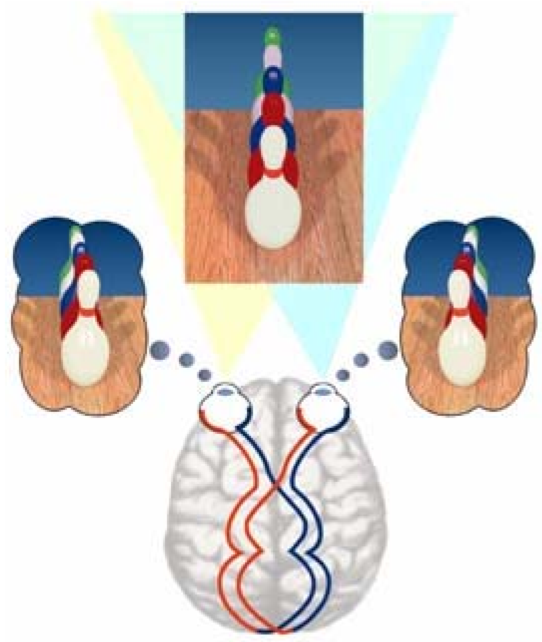
- There are different methods for recovering 3D info. From images:
  - 1- Passive Vision
    - Shape from de-focus
    - Stereo vision
  - 2- Active vision:
    - Laser
- Laser: In active vision, some type of energy such as a laser is emitted into the environment with the reflected light detected by sensors.
- Shape from de-focus: By adjusting the camera’s focus, we can determine when a point is in focus and hence determine it’s depth.



Stereo Vision

- The fundamental basis for stereo is the fact that a single three-dimensional physical location projects to a unique pair of image locations in two observing cameras.

- Given two camera images, if it is possible to locate the image locations that correspond to the same physical point in space, then it is possible to determine its three-dimensional location.



Why Stereo?

- Uses basic cameras which are inexpensive compared with laser and optical scanners.
- Affordability means vision systems could be constructed using personal computers.
- Active vision drawbacks: slow scanning speed and higher cost.
- Passive sensors have the advantage over laser and radar sensors: the possibility of acquiring data in a noninvasive way and so not altering the environment.
- Biggest advantage over active: interference among sensors of the same type.
- Shape from defocus suffers from low resolution.

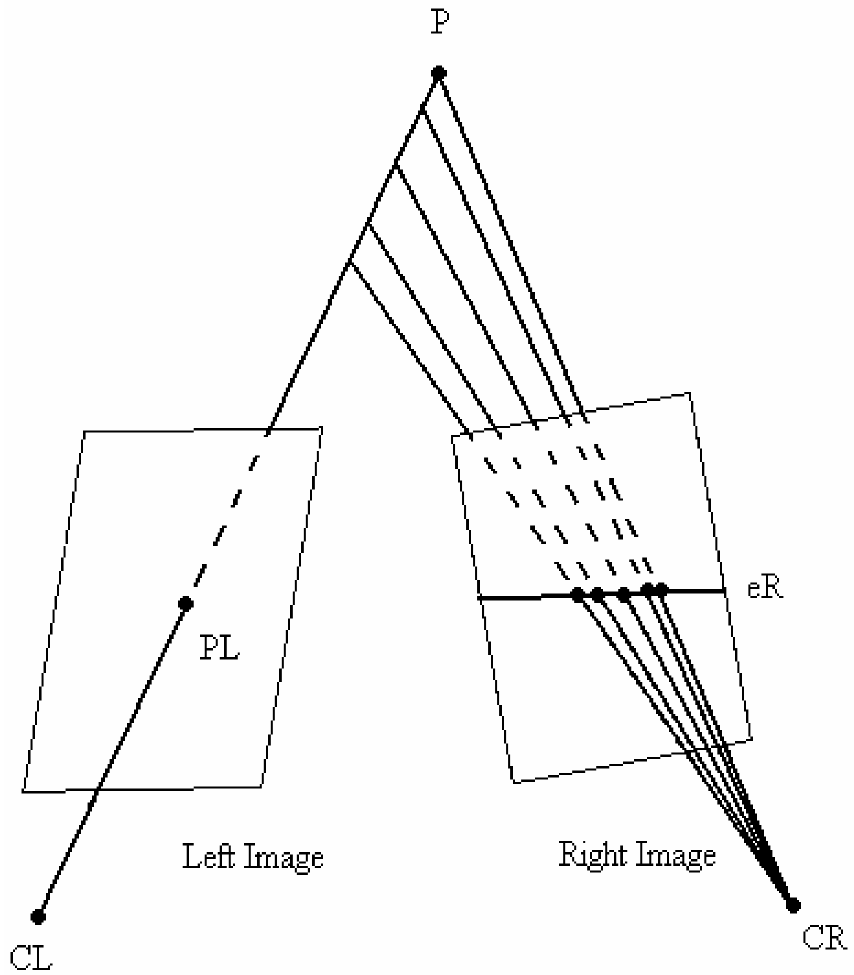
Stereo Vision

- **Steps taken for 3D information using stereo vision systems:**

  - **Calibration**: find geometrical relationship between the 3D space and the cameras.

  - **Correspondence**: Identify the image point that represents the same scene point in the other image.

  - **Reconstruction**: Calculate the depth of the selected location based on the location difference of the corresponding points and camera position.



Camera Calibration

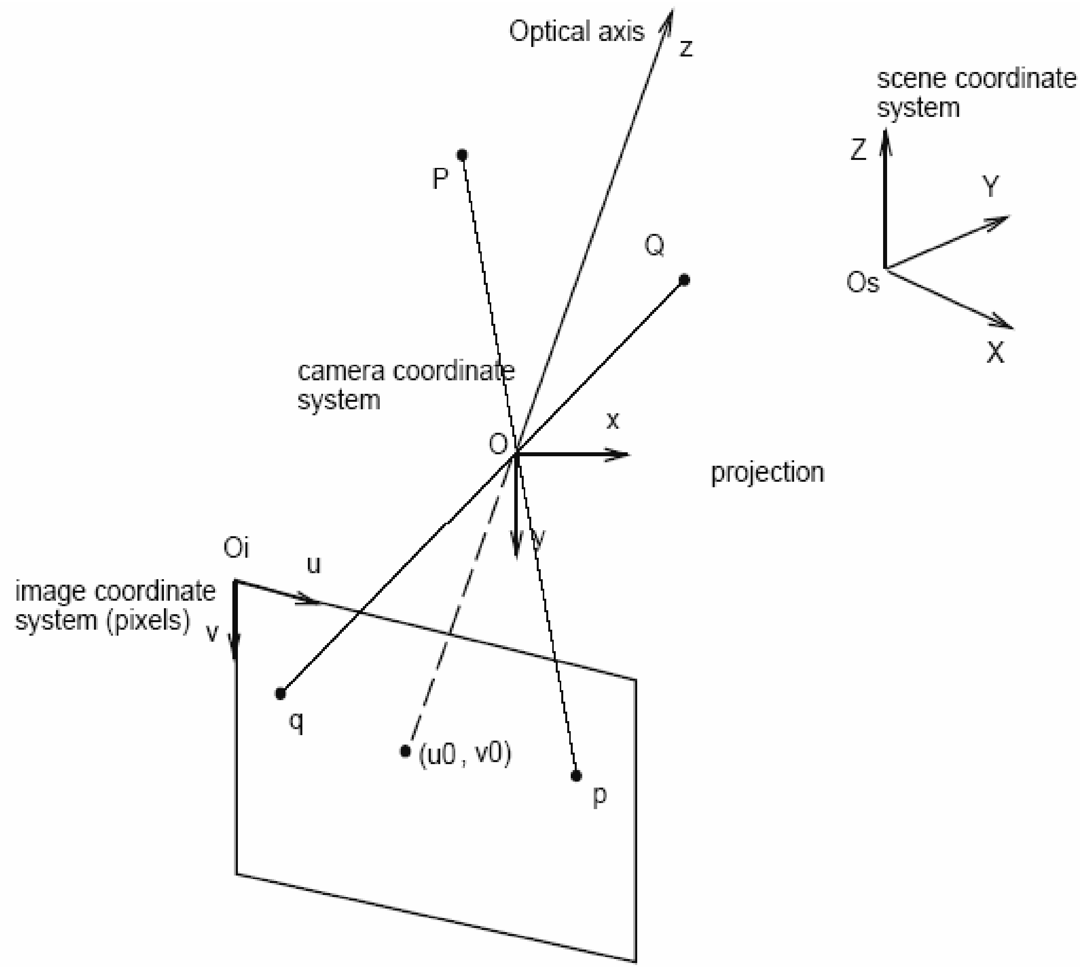
- Projection Matrix

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} \beta_1 & \beta_2 & \beta_3 & \beta_4 \\ \beta_5 & \beta_6 & \beta_7 & \beta_8 \\ \beta_9 & \beta_{10} & \beta_{11} & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

1- A 3D to 3D transformation: rigid camera displacement

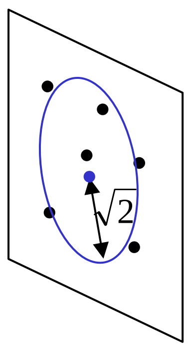
2- A 3D to 2D transformation (perspective projection)

3- A 2D to 2D transformation: normalized camera parameters



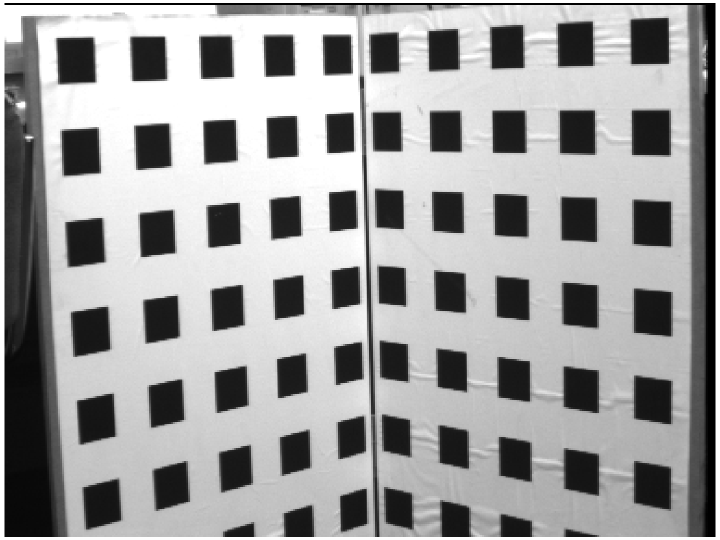
Calibration

- Camera calibration is the process of estimating the extrinsic and intrinsic parameters of a camera *or* finding the projection matrix.
- The steps involved in calibrating a camera:
  - Taking images of the calibration target at known world coordinates.
  - Using the image locations of the interest points on the calibration target and the known world coordinates, calculate projection matrix.
- There are several methods to solve this numerical problems, the most popular is the SVD.
- It is important to normalize the data points for stability purposes.

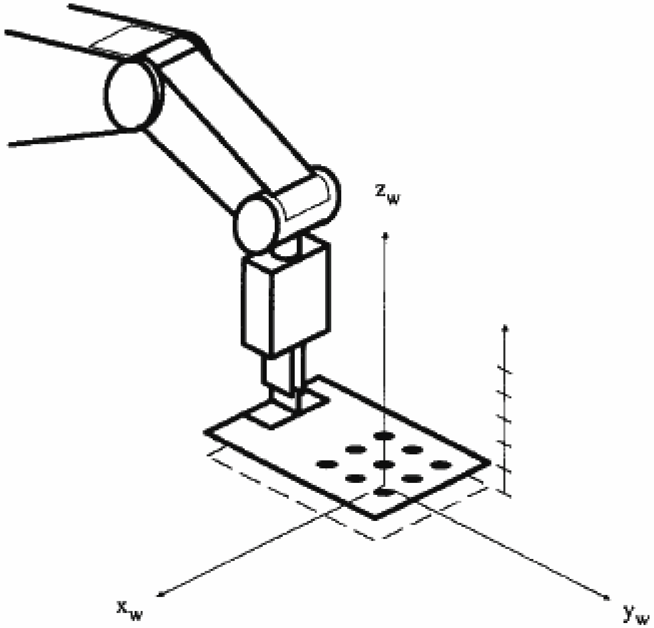


Calibration examples

- Canny edge detection
- Straight line fitting to the detected edges
- Intersecting the lines to obtain the images corners

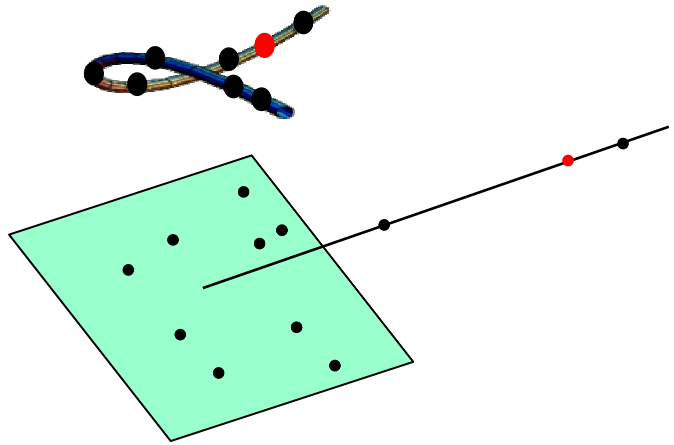


- Move the target to three different heights
- Threshold the image
- Border follow the resulting image
- Find centroid

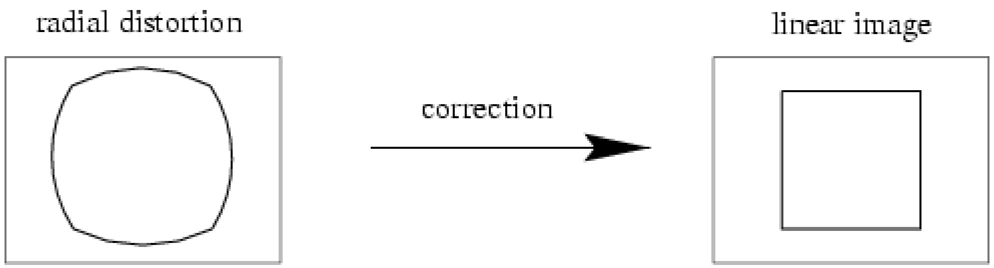


Calibration

- Degenerative configurations:
  - Camera and points on a twisted cubic
  - Points lie on a plane or on a single line containing the camera centre



- Lens Distortion
  - Real lenses suffer from a number of aberrations
  - Correct image coordinates to those that would been obtained under the linearity assumption



$$p = \begin{pmatrix} 1/\lambda & 0 & 0 \\ 0 & 1/\lambda & 0 \\ 0 & 0 & 1 \end{pmatrix} MP$$

$$\lambda = 1 + \kappa_1 \hat{r}^2 + \kappa_2 \hat{r}^4 + \ldots$$

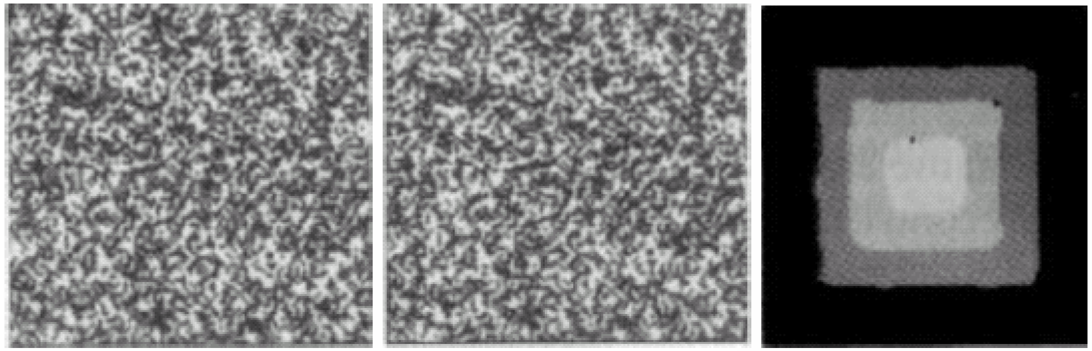
Correspondence

- Correspondence needs to be established between points corresponding to the same scene point.
- The disparity is used to extract the 3D information.
- Disparity is the difference in location of corresponding features seen by the left and right cameras.
- Parallel cameras make matching easier: not the case in this project.
- Epipolar line computation becomes necessary, but the advantage of greater overlap of the images.
- Challenges of correspondence:
    - ambiguity (low-contrast regions)
    - missing data (occlusions)
    - intensity error (quantization, sensor error)
    - position error (camera calibration)

Correspondence

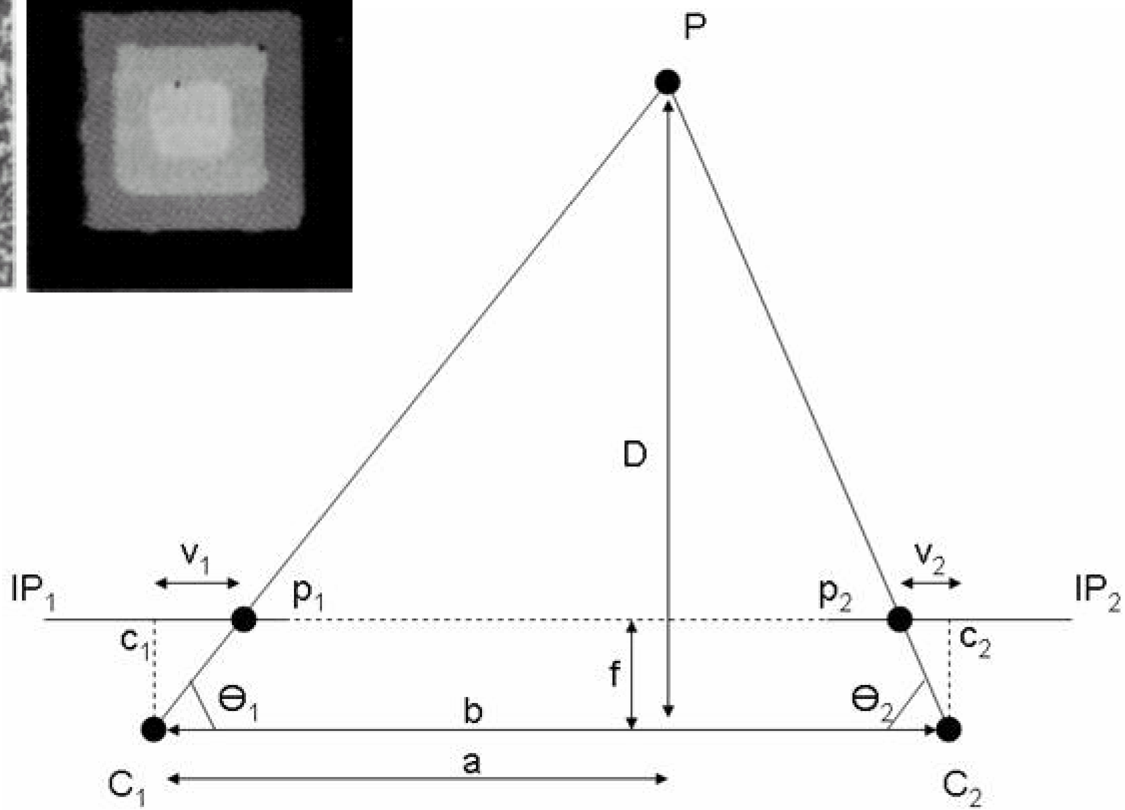
- Relationship between depth and disparity

left image　　right image　　disparity map



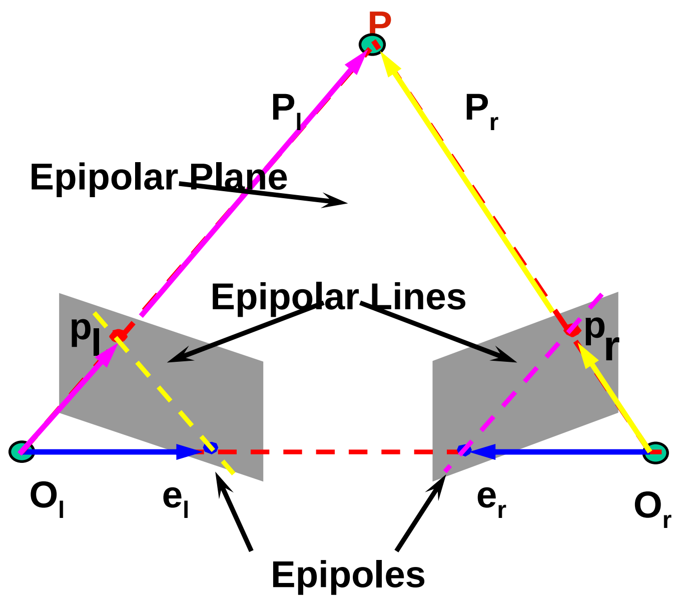
From similar triangles:

$$D = \frac{bf}{v_1 + v_2}$$



Correspondence

- Due to the limited resolution of images, increasing the baseline distance b gives us a more precise estimate of depth.
- Large b -> views will be very different -> difficult to establish correspondence.
- If a non-parallel scheme is used, it is important to determine the Fundamental Matrix during the calibration stage, this will reduce the search from 2D to 1D.
  - Computable from corresponding points
  - Simplifies matching
  - Allows to detect wrong matches
  - Related to calibration



Correspondence: Area-based algorithms

- Comparison between brightness patterns in the neighborhood of a pixel

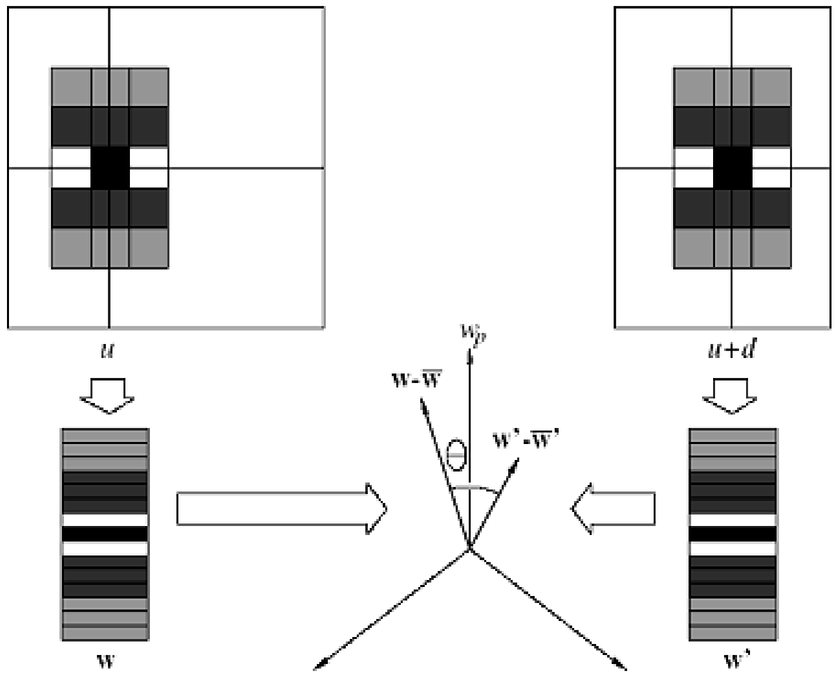
$$NCC = \frac{\sum_{u,v}(I_1(u,v)-\bar{I}_1)(I_2(u+d,v)-\bar{I}_2)}{\sqrt{\sum_{u,v}(I_1(u,v)-\bar{I}_1)^2(I_2(u+d,v)-\bar{I}_2)^2}}$$

- Use NCC, SSD, NSSD, SAD,…

- Drawbacks:

  1- Use intensity values directly thus sensitive to distortion.

  2- Occluding boundaries confuse the matching process: erroneous depth values.



Correspondence: Feature-based algorithms

- Primitives that are to be matched should correspond to physical items that have identifiable physical properties.
- Various problems with area-based techniques.
- Symbolic features derived from intensity value are used instead of intensity values themselves for matching.
- Edge points or edge segments (intensity and direction) are the most commonly used features.
- Faster than area based algorithms since only the attributes of features are compared instead of intensity values.
- The system is more insensitive towards changes in contrast and ambient lighting.

Area-based versus Feature-based

Feature-based

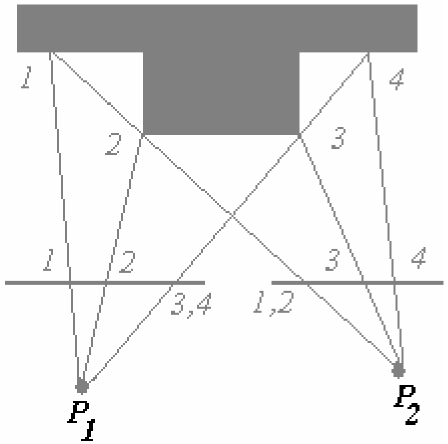
- Sparse maps
- Ideal for feature-rich images
- More efficient and robust against image variation
- Needs preprocessing: feature extraction
- Need for dense maps and improvement in efficient area matching: decline in this area

Area-based

- Dense disparity map
- Ideal for highly textured environments
- Easy to implement
- Computation of correlation is very expensive
- Perform poorly in occluded areas

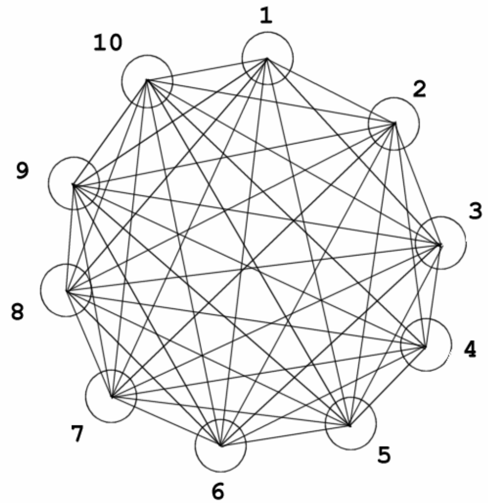
Constrains

- Epipolar line: horizontal in case of parallel cameras, must be computed in nonparallel geometry
- Regional disparity continuity constraint: smooth surfaces thus smooth disparities on surfaces
- Figural continuity constraint: contours project as continuous curves in both images
- Uniqueness of a match
- Preserved ordering of matches along horizontal scanlines
- Disparity limit



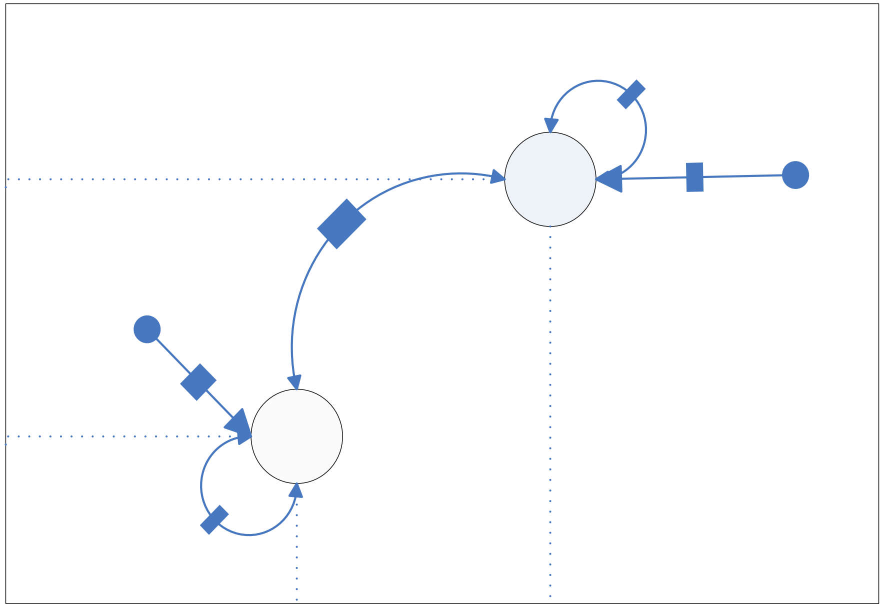
ANN Approach

- NN used to implement cognitive mechanisms: fit for vision
- NN is an energy surface, min energy = solution to many optimization problems
- Matching problem can be formulated as minimization of a cost function, where all the constraints can explicitly be included
- 2D Hopfield Networks
  are good candidates, used for
  pattern association and optimization



ANN Approach: Hopfield NN

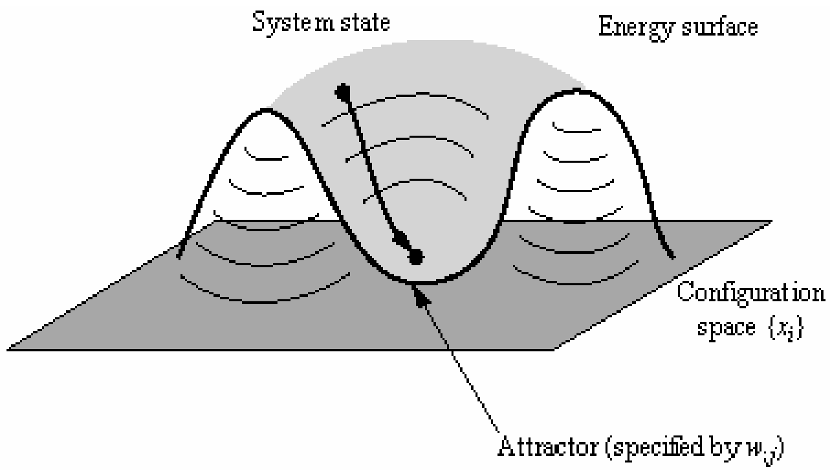
- Matching =an optimization where an energy function representing the constraints will be minimized using HNN
- HNN, different from multilayer scheme
- 2D array of N_r x N_l
- No self feedback, T_{ijkl}=0
- Symmetric, T_{ij}=T_{kl}
- Binary network
- Random updating



$$E = (-\frac{1}{2})\sum_{i=1}^{N_l}\sum_{k=1}^{N_r}\sum_{j=1}^{N_l}\sum_{l=1}^{N_r} T_{ikjl}V_{ik}V_{jl} - \sum_{i=1}^{N_l}\sum_{k=1}^{N_r} I_{ik}V_{ik}$$

ANN Approach

- The stereo constraints are the starting point for every stereo system
- Point of using the constraints: narrow the search, match selection, false match detection
- Design an energy function: associate to every constraint a term that decreases when approaching a match
- Every neuron n_{ik} represents a matching possibility between the respective elements



ANN Approach: Feature extraction

- Choosing the right primitives: important
- Features should be:
  - General: represent majority of the useful info in a picture
  - Matchable: should be easy to match
  - Available: a convenient method for extraction should exist
- Use feature points: local maxima of the directional variance minima
- The Moravec operator is used to extract these points:

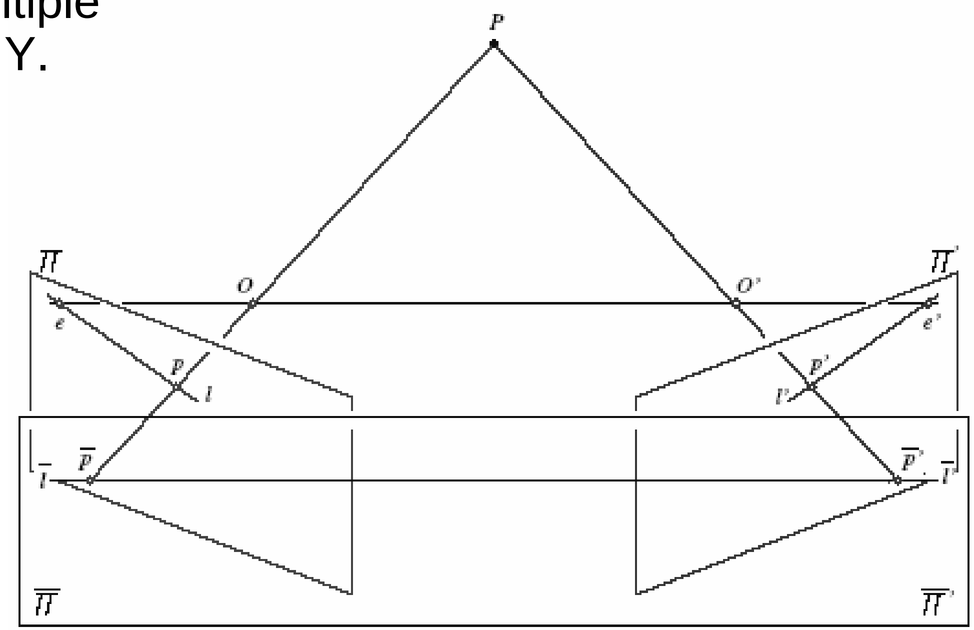
$$V(i,j) = \min V_\theta(i,j) \quad with\ \theta = \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}, \pi$$

$$V(i,j)_{\frac{\pi}{2}} = [f(x,y) - f(x+1,y)]^2$$

ANN: Occlusion

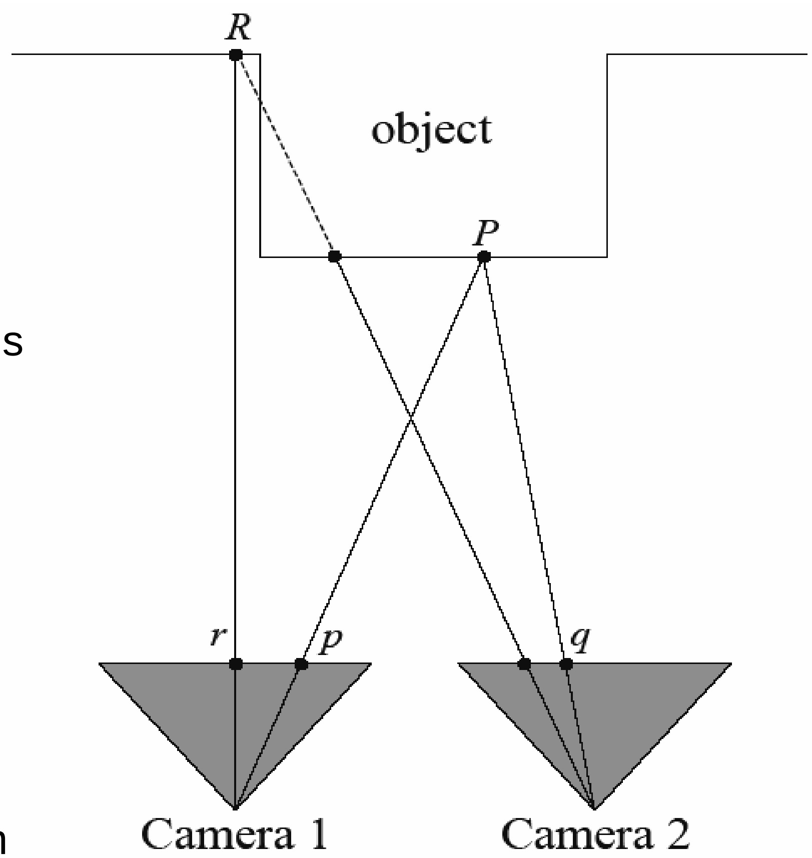
- Carry out epipolar rectification: easier to spot outliers and detect occlusion.
- ANN method: after finding all points, if multiple match, check Y disparity against average Y.
- ANN method: only feature points matched thus lower chance of error.
- Finally, use simple cross checking if too many errors due to occlusion.

Rectified Stereo Pair



Occlusion

- Much of the stereo research in the last decade: detecting and measuring occlusion



1- Detect Occlusion :

ANN

Simplicity, overall good performance,

implemented in many real-time stereo systems

2- Reduce Sensitivity to Occlusion:

Adaptive Window Size for Correlation

3- Model Occlusion Geometry:

Graph Cuts

Integrate knowledge of the occlusion geometry itself into the search process

*Other methods: use more cameras, use active vision in addition to passive sensing

Future Work

- Investigate the effects of using various point features, move to line matching
- Finding the number of optimal feature points that will characterize an object in space
- Incorporate epipolar constraint in the matching cost of the ANN algorithm
- Experiment with alternative calibration techniques and report on their feasibility
- Report on the robustness of the ANN method in presence of measurement noise and compare with other methods (correlation)

References

[1] U.R. Dhond and J.K. Aggarwal, "Structure from Stereo—A Review," IEEE Trans. Systems, Man, and Cybernetics, vol. 19, pp. 1489-1510, 1989.

[2] Y. Boykov, O. Veksler, and R. Zabih, "Fast Approximate Energy Minimization via Graph Cuts," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 23, no. 11, pp. 1222-1239, Nov. 2001.

[3] V. Kolmogorov and R. Zabih, "Computing Visual Correspondence with Occlusions Using Graph Cuts," Proc. Int'l Conf. Computer Vision, 2001.

[4] M. Z. Brown, D. Burschka, and G. D. Hager. Advances in computational stereo. IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(8):993–1008, 2003.

[5] Simon Haykin. *Neural Networks – a Comprehensive Foundation*. Prentice Hall, New Jersey, 2nd edition, 1999. ISBN 0-13-273350-1.

[6] K. Achour, L. Mahiddine. Hopfield Neural Network Based Stereo Matching Algorithm. Journal of Mathematical Imaging and Vision, 16:17-29, 2002

[7] D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms," Int'l J. Computer Vision, vol. 47, no. 1, pp. 7-42, 2002.