MAPI – Computer Vision

Active Vision

Active Vision

- Most computer vision systems still use a static camera.
  - Some have a camera mounted on a vehicle or a moving arm, but usually there is little control over the camera’s movements.
- In biological systems, the eye is extremely active.
  - Rapid, frequent **saccades** (fast changes in direction) and **tracking movements** are made.
  - Information pickup is active, not passive. As well as eye movements, body movements are made to change the position of the viewpoint.

Active Vision

- Rapid eye movements recorded during driving (in this case by M.F. Land of BIOLS) provide a good example.

Active Vision

- The term **active vision** is used in at least two senses.
  - “active operation in the world in order to change the images that are being collected in a way which enhances task achievement”
    - This recognises that vision is part of a perception-action cycle.
  - “active autonomous processes (e.g. snakes) which exploit the coherence of images in a sequence in order to efficiently and reliably track aspects of interest over time”
    - This recognises that vision takes place in a temporal continuum.

Active Vision

- Active vision requires new:
  - Hardware
    - active heads, mobile cameras;
  - Software
    - algorithms that integrate information over time;
  - Philosophy
    - the goal of vision is to enable useful interaction with the environment. The goal is not to find the 3-D layout, and is not to make a symbolic description.

Active heads

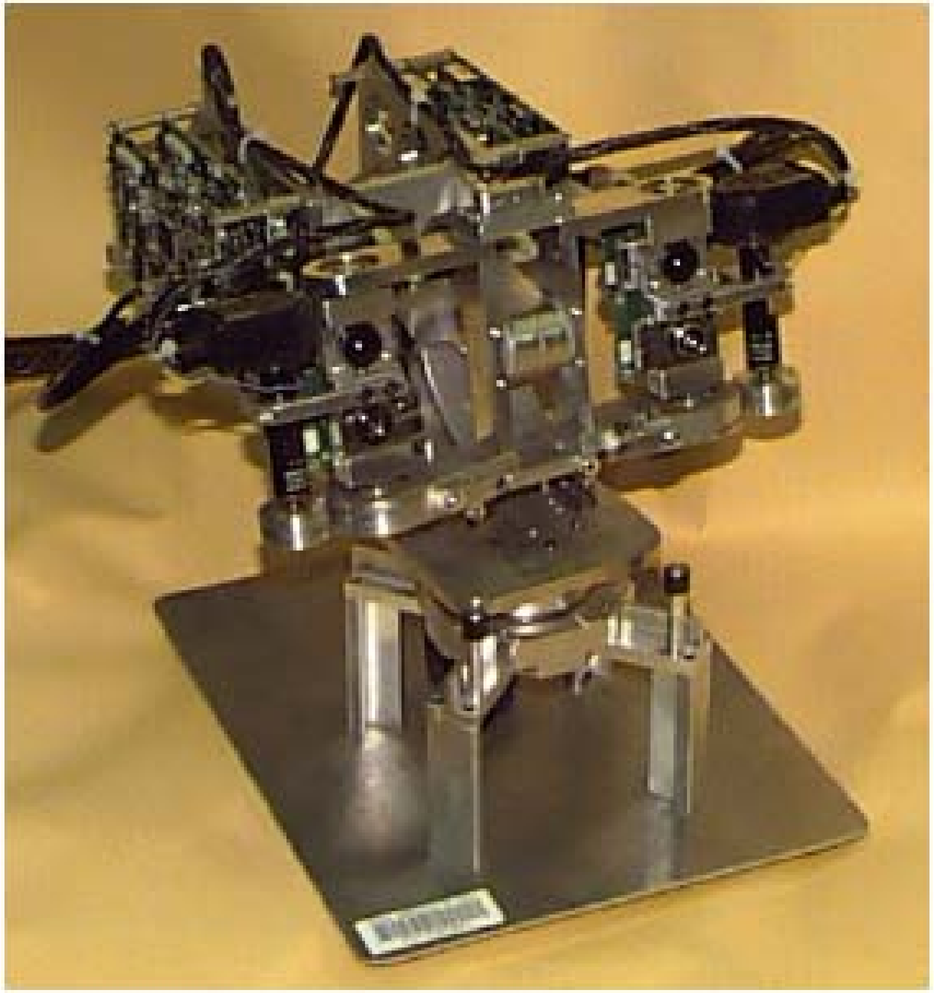
- Recently, there has been much research on developing **active heads**:
  - mountings for cameras that allow very rapid movements for tracking purposes.
  - High angular velocity is necessary for **tracking** whilst high angular acceleration is important for **capture**.
  - Maximum angular velocities of several hundred degrees per second are achieved, with maximum accelerations of thousands of degrees per second
- Such systems are built by research groups, not commercially.
- Motor and controller technology are very important.

Active heads

The device on the right was built as part of the *COG* project at MIT.

See http://www.ai.mit.edu/people/scaz/3heads.html

It has two pairs of cameras, with independent *pan* (sideways) movements but coupled *tilt* (up and down) movements.



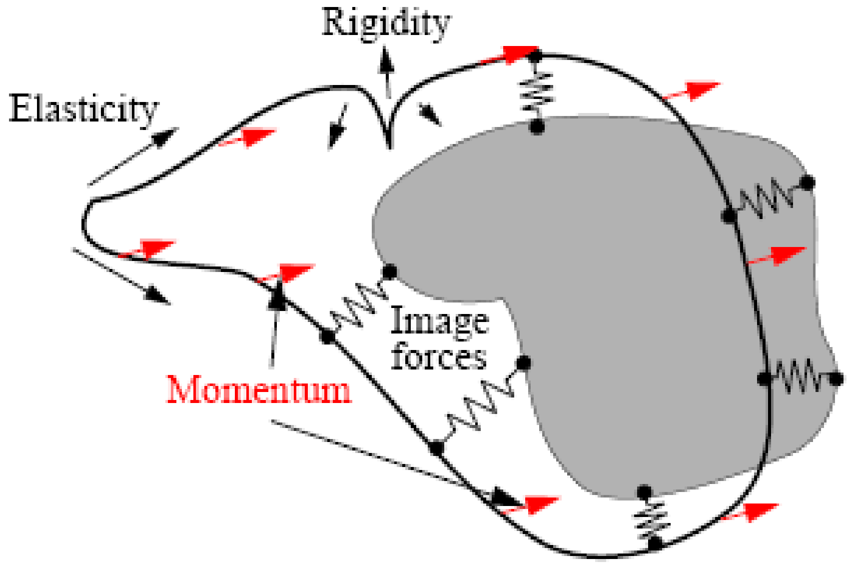
Tracking in Active Vision

- Current tracking methods concentrate on **active contours, deformable templates, dynamic models** and probabilistic predictors such as **Kalman filters**.
- These methods are related to optic flow methods, but make more use of the **history** of the motion — information is integrated over time.

- **Active contours**
  - An approach which combines high-level feature tracking with a model of the motion is the **snake**. This is a parametrised curve in the image (often a spline curve). Snakes have particular advantages for **nonrigid motion.**

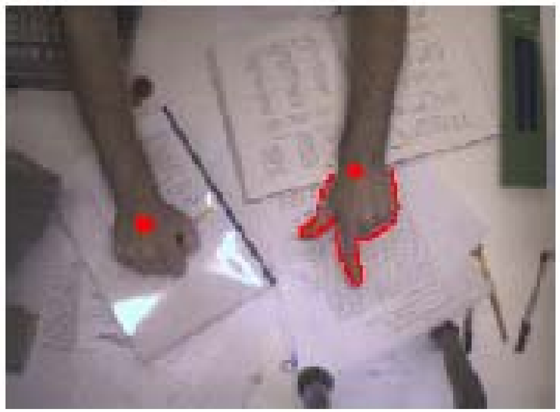
Tracking in Active Vision

- **Simulated dynamics** of a physical system are used to give the contour **momentum**, smoothness, elasticity, and other properties.
  - The equations of motion of the contour combine forces from these intrinsic properties with forces from the image to make the contour track objects through frames.



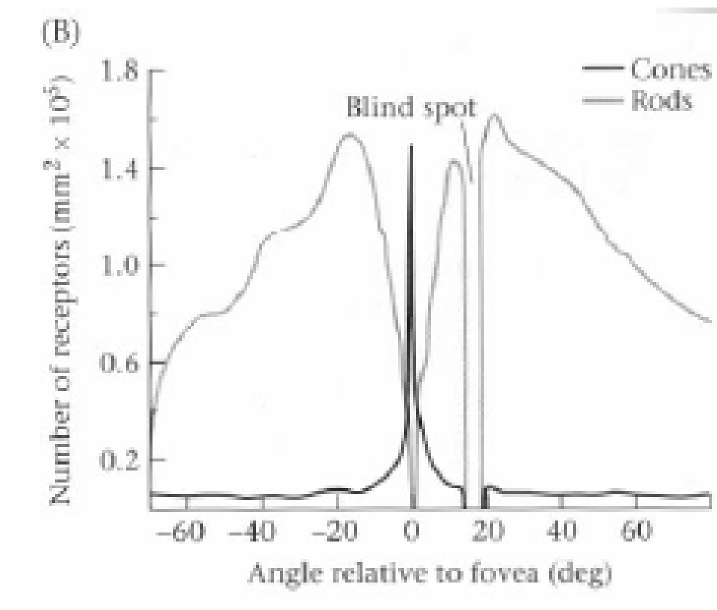
Tracking in Active Vision

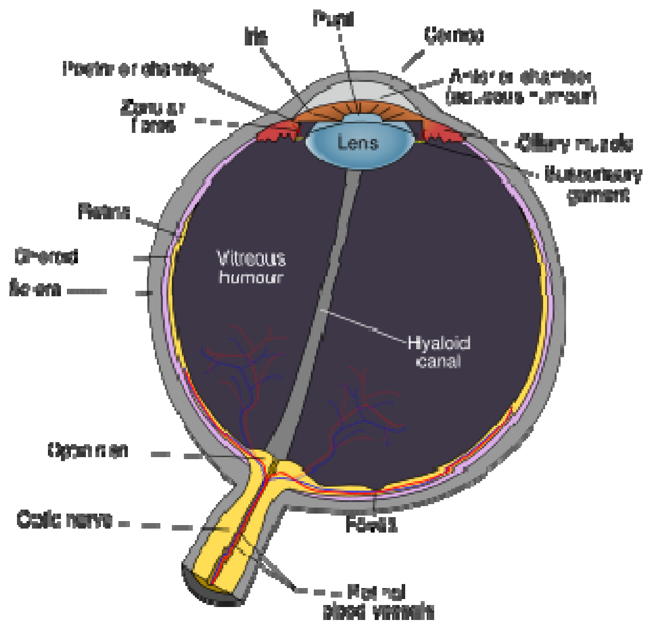
- **Tracking using anticipation**
  - An key to successful tracking is to focus computational resources on important parts of the image.
  - Using *anticipation* allows this to occur. Knowledge of the properties of the tracked object — geometrical and dynamic — is central to this.



Active Vision and foveal vision

- The distribution of receptor cells (particularly cones) on the human retina is extremely non-uniform.
  - The greatest concentration is at the **fovea.**
  - This contrasts with the uniform representation usually used in computer vision.

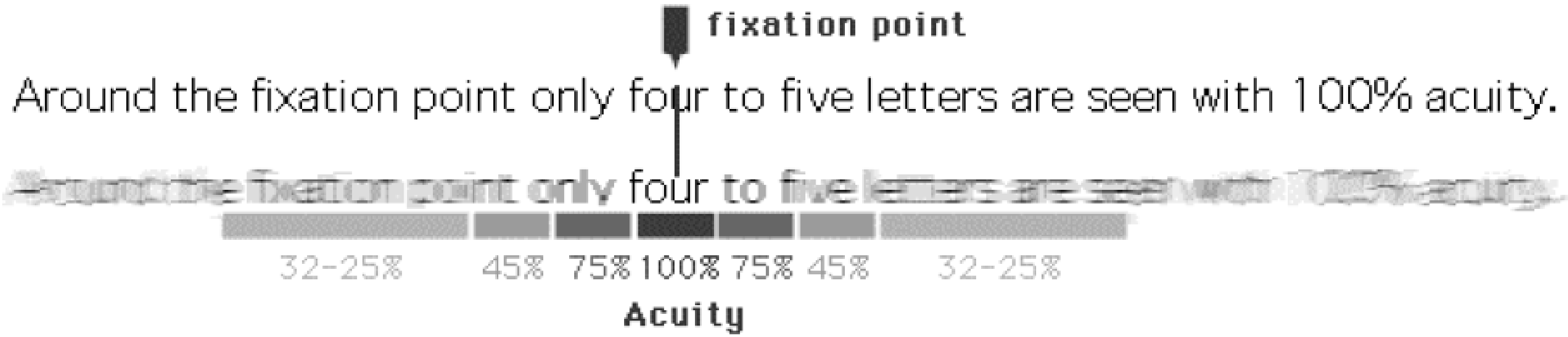




Active Vision and foveal vision

- we believe that we perceive a scene completely and sharply, only a small subpart is sharp (the one on the fovea).

- The complete perceived sharp image is created by *saccading* over the scene.



Active Vision and foveal vision

- Two parameters are of high importance in visual sensor design: **resolution** and **field of view**:
  - resolution determines the scale of the smallest details that can be detected in the images;
  - field of view determines how far from the central view point objects can be detected.

- Computer vision systems usually control efficiency by controlling image size, either by
  - reducing resolution or the field of view:
    - image resolution is reduced uniformly according to some desired criteria;
    - field of view is reduced by defining windows of interest around objects, where further processing is preformed.

Active Vision and foveal vision

- These strategies have been applied successfully in some structured environments, where good models for object sizes and motions exist.
- However, they are too rigid to be applied in more unstructured situations.
  - In the first case, resolution may not be enough to detect objects whose scale changes along time.
  - In the second case, moving objects easily move away from the windows of interest.

Active Vision and foveal vision

- *Foveation* deals with methods to represent images efficiently
  - preserving both the field of view
  - maximum resolution,
    - at the expense of reducing resolution at some parts of the image.
- With such a representation, the sensory strategy can allocate high resolution areas to objects of interest as they are detected in a wide field of view.

Active Vision and foveal vision

- Active vision has greatly increased the interest in foveal vision.
  - Active heads can be used to **fixate** objects of interest ensure that their image remains in the densely sampled region. Meanwhile the periphery can continue to pick up important events that might signal a new region of interest.

Active Vision and foveal vision

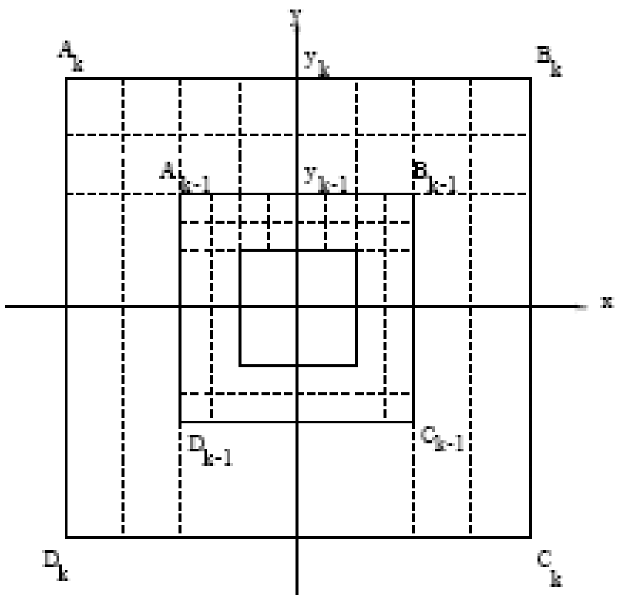
- Foveal vision offers
  - concentration of computational resources on important parts of the scene
  - task-based visual processes
  - good algorithms for template matching and optic flow estimation.

Active Vision and foveal vision

- Some systems uses two cameras for each eye, one to give a broad view and one to give a detailed central ('foveal') view.

Active Vision and foveal vision

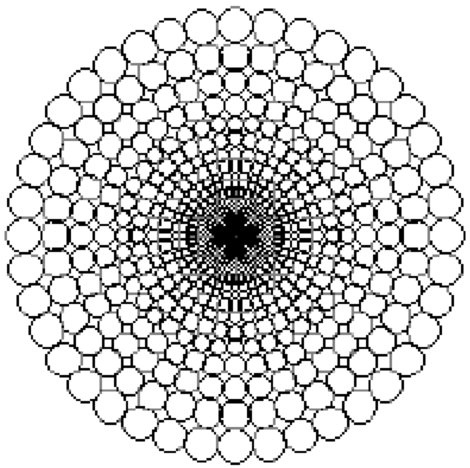
- It is more elegant to use a nonuniform distribution of receptors in a single camera.
  - *superpixel* methods



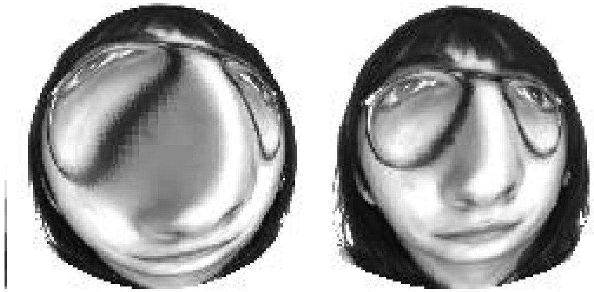
Face recognition using foveal vision
Silviu Minut *et al*

Active Vision and foveal vision

- It is more elegant to use a nonuniform distribution of receptors in a single camera.
  - One common receptor layout for foveal vision is the **log-polar sampled image** (LSI).
    - Positions are represented by **ring** and **wedge** rather than by row and column.



Face recognition using foveal vision
Silviu Minut *et al*



Active Vision and foveal vision

- In the LSI, rotation and expansion are very simple transformations.
- Translation is complex, so active vision is needed to fixate and track points of interest.
- LSI cameras have been implemented in hardware, though much work uses software simulations.

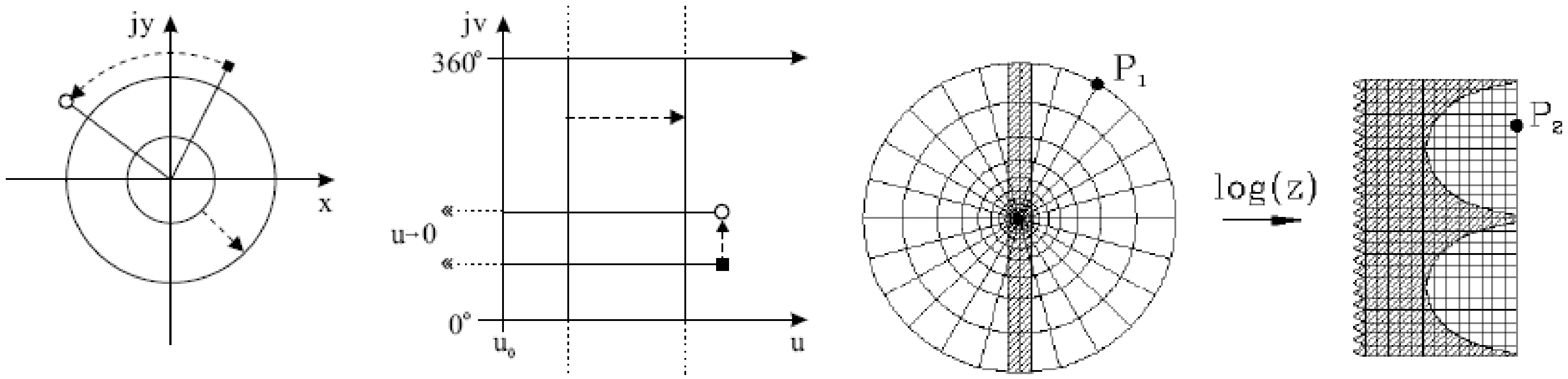
Active Vision and foveal vision

- **The Logpolar Transformation**
- Is based on the complex logarithmic function $\log(z)$, which is used to approximate the retino-cortical mapping of primates.
- Let us consider the complex retinal and cortical planes, represented by the variables $z = x + jy$ and $w = u + jv$, respectively, where j is the complex imaginary unit:
- $w = \log(z) = \log(|z|) + j \arg(z) = \log(\rho) + j\theta$

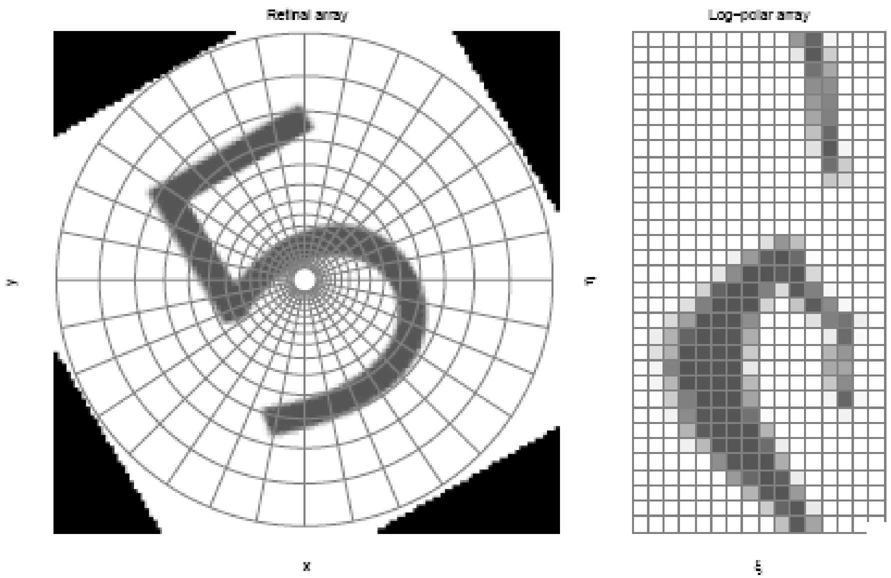
ρ is the input eccentricity and θ is the input angle.

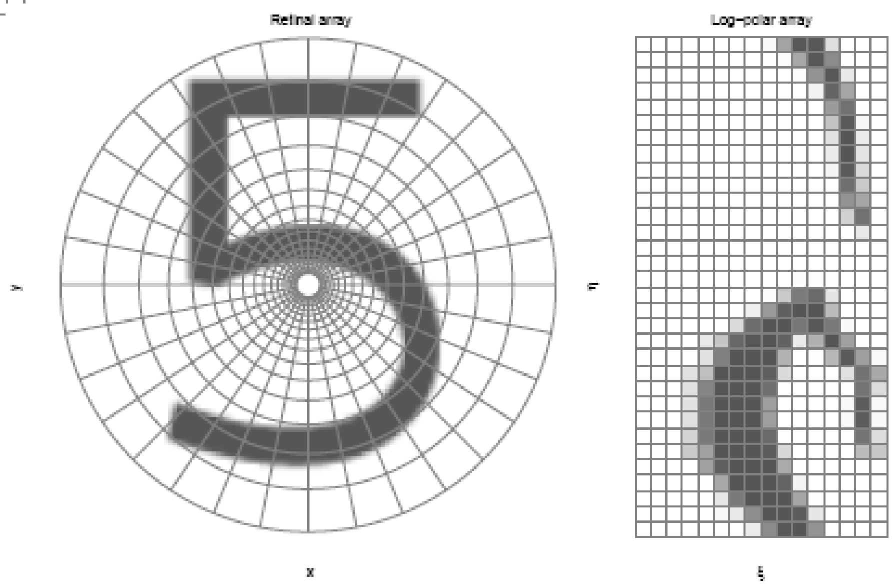
Active Vision and foveal vision

- **The Logpolar Transformation**
- Image rotations and scalings in the center of the retinal plane become simple translations along the *jv* and *ju* axes in the cortical plane, respectively.



Active Vision and foveal vision





Active Vision and foveal vision