INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION

INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
34100 TRIESTE (ITALY) - P.O.B. 586 - MIRAMARE - STRADA COSTIERA 11 - TELEPHONE: 2240-1
CABLE: CENTRATOM - TELEX 460392 - I

SMR/302-20

COLLEGE ON NEUROPHYSICS:
"DEVELOPMENT AND ORGANIZATION OF THE BRAIN"
7 November - 2 December 1988

"Interaction of Modules in Depth Perception"

Heinrich H. BULTHOFF
Center for Biological Information Processing
M.I.T.
Cambridge, MA
USA

Please note: These are preliminary notes intended for internal distribution only.

Interaction of Modules in Depth Perception

Heinrich H. Bülthoff

Center for Biological Information Processing

MIT
Cambridge
USA

Collaboration

- **Hanspeter Mallot, Mainz Univ., FRG**
- **Andrew Blake, Oxford University, UK**
- **Dan Kersten, Brown University, USA**
- **Tomaso Poggio, MIT, USA**

Overview

■Integration:

- in Machine Vision
- in Human Vision

■Integration: Depth Perception

- Shape from Stereo
- Shape from Shading
- Shape from Texture
- Shape from Specularities

Overview

Study Integration

■ in Machine Vision

- with MIT Vision Machine Project
- Eye-Head System
- Connection Machine

■in Human Vision

- Computer Graphics Psychophysics
- Quantitative Depth Perception

Vision Machine

■ Goal

- to use information from several cues simultaneously
- to help refine the initial estimation of surface discontinuities , which are typically noisy and sparse

■ Strategy

- Couple different cues to the image data (in particular intensity edges)
- Cues (the continuous fields) and their discontinuities are coupled to each other through the discontinuities in the physical properties of the scene

Vision Machine

■ a two Camera Eye/Head input device:

the camera platform can pan and tilt, and the pointing direction of each of two cameras can be controlled with 3 degrees of freedom. Lenses with controllable zoom, focus and aperture permit active exploration of the scene.

■ a Connection Machine:

Single Instruction Multiple Data (SIMD)

a retinotopic mapping into the CM memory

Vision Machine

■Vision Algorithms

map quite naturally onto the CM

a retinotopic mapping into the CM memory

map also onto

- VLSI architectures

  analog VLSI

  mixed analog digital VLSI

  specialized digital processors,

  eg., specialized CM

Observable Discontinuities

Stereo and Motion share a common framework

■Assumption:

surfaces locally vary slowly in depth and motion

■Algorithm:

find locally consistent values in depth and motion

Interestingly, the process computing depth and motion, directly yields cues for object boundaries, since the smooth variation assumption fails here (see motion talk).

Parallel Algorithms -- CM

Vision algorithms from early, middle and high-level vision place different requirements on a fine-grained parallel machine.

- ■Early Vision:
  - such as edge detection, stereo motion
  - compute locally in the spatial domain
- ■Middle Vision:
  - such as line following, visual routines
  - has regular, but context dependent comm.
- ■High-Level Vision:
  - such as recognition
  - uses arbitrary communication, which is in the CM provided by the Router

Parallel Algorithms -- VM

A Vision Machine has many components: their communication requirements are critical in their implementation on fine-grained machine.

- ■Early Vision:
  - modules translate directly into separate layers of the CM memory
- ■Discontinuities
  - are found from local evidence
- ■Linked
  - into larger structures by non-local communication (Router)

Physics of Image Events

Intensity edges can arise from a variety of physical causes

- **Occlusion -- Depth Discontinuity**
- **Shadow -- change in Illumination**
- **Specularity -- surface property, surface orientation and illumination**
- **Albedo -- change in reflectance properties**
- **Orientation-- change in suface orientation**

Future Directions

- **Estimating parameters from examples**
- **"Learning" to label surface discontinuities**
- **Coupling "Occlusion Rues" to MRF lattice**
- **VLSI implementations**
- **Physiology of Integration**
- **Psychophysics of Integration**

Physiology of Integration

in collaboration with
Peter Schiller and Nicos Logothetis
MIT
Brain and Cognitive Science Department

■behavior (monkey psychophysics)

■recordings from V1, V4 and MT

■looking for "authentic depth cells"

■lesions

Integration Psychophysics

in collaboration with
Peter Schiller and Nicos Logothetis
MIT
Brain and Cognitive Science Department

■monocular cues to help stereopsis
(W. Richards, 1976)

■interpolating surfaces in depth
(T. Collet, 1985; D. Buckley
and J. Frisby, 1988)

■Integration of Depth Modules
(H. Bülthoff and H. Mallot, 1987; 1988)

■Integration along discontinuities
(H. Bülthoff, in prep.)

Human vs Machine Vision

Human Vision

- **reliable and fast for *natural* images**
- **slow for *synthetic* images ("Julesz Spiral")**

Machine Vision

- **reliable and fast for *synthetic* images**
- **slow and problematic for *natural* images (eg., with transparency or specularities)**

Solution: Integration of Modules

Single Modules

- **From 2D to 3D with n images:**
- **n = 1: shading, texture, perspective, knowledge**
- **n = 2: binocular stereo**
- **n = 3: structure from motion**

These and other modules have been formalized in terms of computational theory and implemented as single modules in machine vision systems .

Multiple Modules

Human Vision

- Shape-from-X
  - Stereo
  - Shading
  - Texture
  - Specularity
  - Motion

Machine Vision

- Shape-from-X
  - Stereo
  - Shading
  - Texture
  - Specularity
  - Motion

We have to find out how integration works

Integration of Eye and AI

Study of Integration
by
Integration of Human and Machine Vision

- **Human Vision** : Existence Proof for Integration
- **Machine Vision** : Demonstrates Problematic

Computer Graphics Psychophysics

- "Controllable Natural Images"
- **Shading** : Lambertian or Phong
- **Texture** : Texture Mapping or Solid Modelling
- **Stereo** : Computer version of Gregory's Pandora's Box
- **Animation** : Rotate 3D objects or specify camera path through 3D world

Quantitative Depth

Local Depth Measurements

- **Stereo Depth Probe**
  - \+ exact depth map
  - \- interacts with monocular cues

Global Shape Comparisons

- **Variable Shape Probe**
  - \+ works for all cues
  - \- no exact depth map

Edge vs Intensity Stereo

Human Vision

- edge-based stereo
  - Primal Sketch (Marr)
  - Peak Matching (Mayhew & Frisby)
- intensity-based stereo
  - ???

Machine Vision

- edge-based stereo
  - zero-crossing schemes (Marr, Hildreth)
  - Edge Detector (Canny)
- intensity-based stereo
  - Grey-level Correlation

Images without Edges

To study intensity-based stereo we have to make sure that our images do not contain edge information

- **numerical** test: zero-crossings of DOG-filtered images
- **analytical** test: no ZC's in Laplacian of Lambertian sphere, but ZC's for strongly elongated ellipsoids

Images without Edges

- **Luminance Profiles:**

  with inflection points, but ZC's strength 100 times smaller

- **Brightness Profiles**

  compression non-linearity in photoreceptors cancels inflection

Matching Primitives

- **Edges ? no, because**
  - **no ZC's**
- **Peaks ? no, because**
  - **only one broad peak in ellipsoid which is difficult to match**
  - **token experiment (peak substituted by token) shows that surface interpolation is a spatially distributed mechanism**
  - **a single token gives smaller depth than full intensity disparities**

Matching Primitives

- **Intensity** ? ok,
  - but underestimates depth
- **Edges + Intensity** give best results

Intensity-based stereo can be very useful in surface interpolation for sparse depth data from edge-based stereo.

Properties of Modules

- **Edge-based Stereo** :
  - best depth map
  - but data points can be very sparse
- **Intensity-based Stereo:**
  - dense depth map
  - but depth map is too flat
    regularization?
- **Shape-from-Shading** :
  - not quantitative
  - but good for qualitative properties of complex surfaces

Interaction of Modules

■Edge-based Stereo

- vetoes Intensity-based Stereo
- vetoes Shape-from-Shading

■Intensity-based Stereo

- inhibits Shape-from-Shading

■Shape-from-Shading

- enhences Shape-from-Texture

■Shape-from-Texture

- enhences Shape-from-Shading

Shape-from-Shading

■why is the depth map so flat?

- Interaction between binocular and monocular Shading (Inhibition)
- better performance with Shape comparison technique
- depends on knowledge about reflectance function, but how can we know it ?
- shading is convex-concave ambiguous (Crater Illusion)

Single Light Source ?

- **Ramachandran:**
  "brain evolved in a solar system with only one sun"
- **true, but what about**
  - **overcast days (San Diego ?)**
  - **light from all directions in forest**
  - **indoor lighting with multiple lights**
  - **4 out of 10 have multiple perceptions**
  - **works best with non-realistic shading Amiga vs High-Res Graphics?**

Single Light Source ?

- **Crater Illusion**
  - **depends strongly on background**
  - **no effect with black or colored background**
  - **best when carved out of surface**
- **cast shadows only on concave surfaces**

Shape from Specularities

in collaboration with Andrew Blake, Oxford

■Idea : Surface Curvature from Disparate Highlights

■Test : two perceptual experiments

1. Convex-Concave Perception depends on 3D-position of highlight

2AFC experiment for convex/concave

2. perception of glossiness is best for correct 3D-position of highlight

adjustment experiment with movable highlight

Interaction

Shape from Specularities is a good example for the fruitful interaction between Psychophysics and AI

■a solid computational theory of Shape from Specularities (Andrew Blake)

■theory can be tested with Computer Graphics Psychophysics