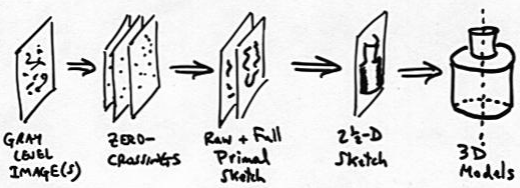
3

MARR'S THEORY of VISION



- IMAGE: ARRAY of INTENSITIES, WHERE INTENSITY AT A PIXEL IS A FUNCTION of: (i) GEOMETRY, (ii) Surface Characteristics, (iii) Illumination, (iv) Viewpoint, (v) Sensor Characteristics
- ZERO-CROSSINGS: "EDGE FINDING" — i.e. Points of intensity discontinuity [resulting from many different physical causes:

Illumination change
Depth change
Surface Orientation change
Surface Reflectance change

4

- PRIMAL SKETCH: A REP. THAT Makes Explicit significant image events:
  Intensity Changes
  Spatial Arrangements
  that represent physical events
- 2½-D SKETCH: Makes explicit the Orientation and "rough" Depth of visible surfaces
- 3D MODEL: MODULAR, HIERARCHICAL Rep. of Shapes and their Spatial Organization. Object-centered Coordinates.

PRIMAL SKETCH

- REPRESENTATION THAT MAKES EXPLICIT SIGNIFICANT IMAGE EVENTS — i.e. INTENSITY CHANGES AND SPATIAL ARRANGEMENTS

- MAIN IDEAS:

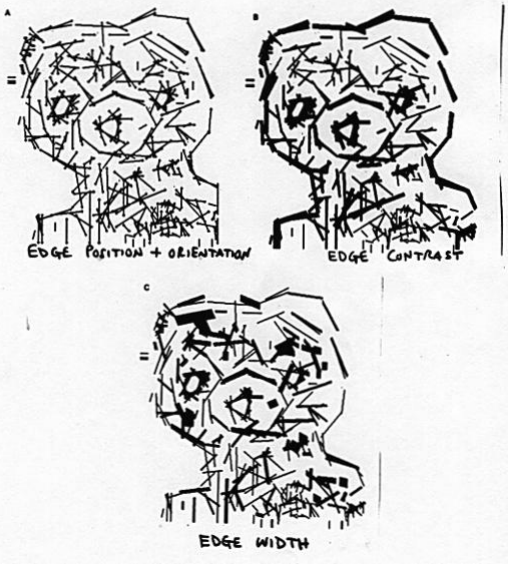
1. Tokens are of same kinds, but varying scales
   - Tokens represent physical events.
   - Such events occur at multiple scales.

   E.g. - Zebra — Stripes — Hair — Flea —...
   ...

   Primitives are:
   - Edge
   - Bar (Line)
   - Blob (Region)
   - Termination (Endpoint)

   Attributes are:
   Orientation, Contrast, Length, Position, Width
   ...

PRIMAL SKETCH



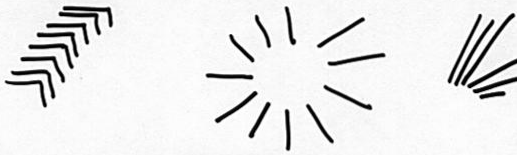
2. PRIMAL SKETCH is built in stages:

* ZERO-CROSSINGS DETECTION AT MULTIPLE SCALES
* Create RAW PRIMAL SKETCH BY "PERCEPTUAL GROUPING RULES" TO LINK ZERO-CROSSINGS
* Create FULL PRIMAL SKETCH
  Add tokens that represent the spatial organization of Raw P.S.
  E.g. Virtual Lines
  Densities / Clusters
  Distances between items
  Boundaries
  texture, shadow, occlusion,
  Repeat recursively.

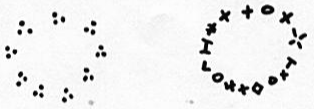
3. TRADE-off between Accuracy and Usefulness — Primal Sketch is used to recover underlying visible surfaces.
⇒ Don't need great accuracy.

(ENDPOINTS)

EVIDENCE THAT TERMINATORS ARE USEFUL:

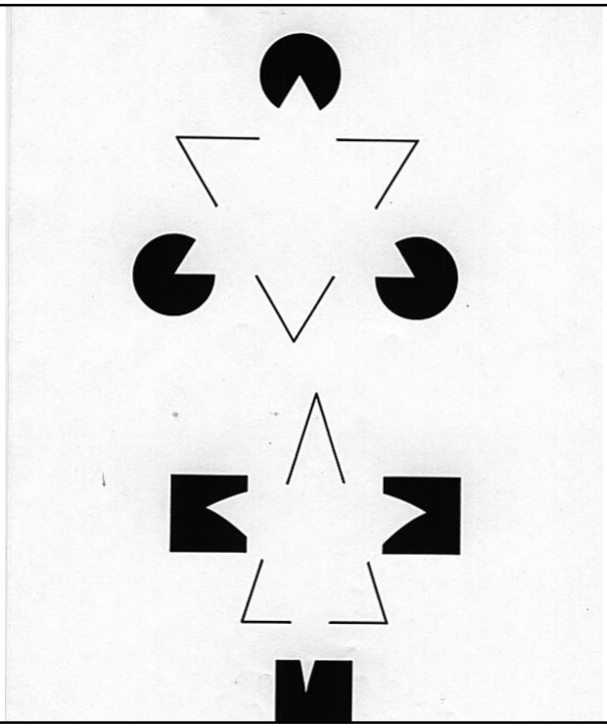


EVIDENCE of HIGHER-LEVEL TOKENS:



EVIDENCE FOR VIRTUAL LINES

< GLASS PATTERNS >



$2\frac{1}{2}$D Sketch

- **Viewer-Centered Representation** of visible surfaces which makes explicit the surface orientation and rough depth at each point.

- Observations & Properties:

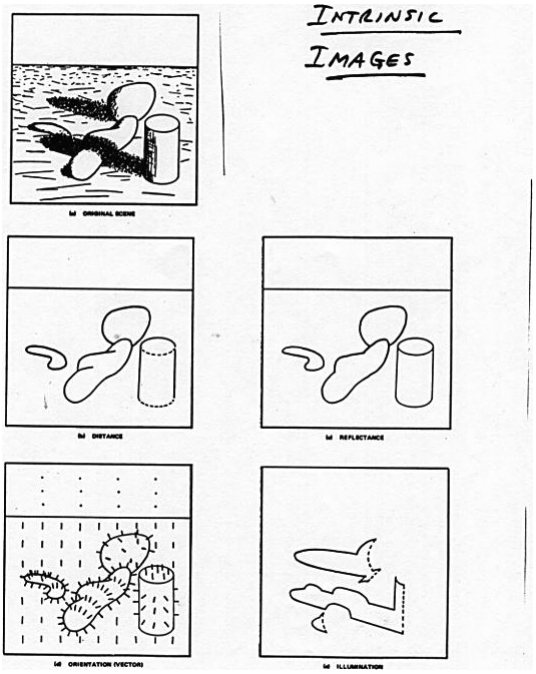
  * Stereopsis w/ random-dot stereograms
    ⇒ depth map memory buffer

  * Easy to "see" depth in complex scene, but hard to remember it
    ⇒ not merely matching obj. models

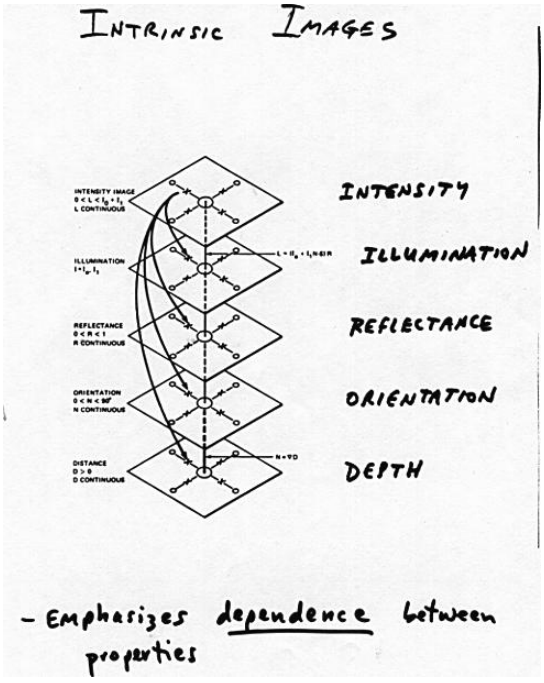
  * Early vision processes extract info from surfaces directly w/o regard to objects
    ⇒ represent surfaces

  * Can perceive shape of unknown objs.
    ⇒ represent orientation

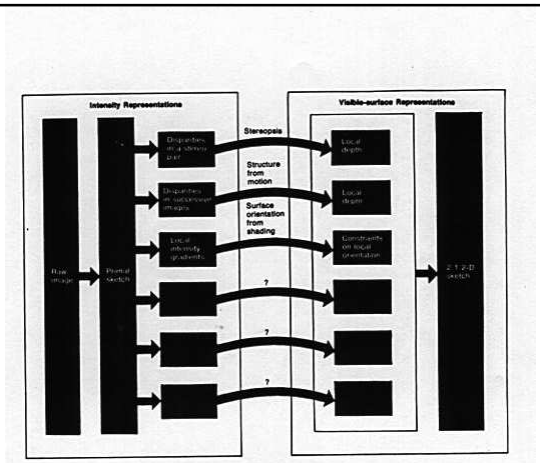
Intrinsic Images



Intrinsic Images



- Emphasizes dependence between properties

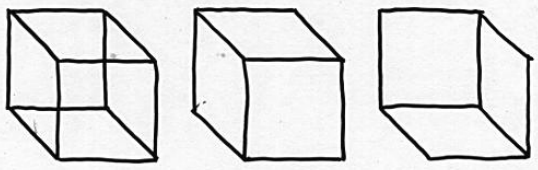


- MODULAR ORGANIZATION
  MANY INDEPENDENT PROCESSES
- LITTLE TOP-DOWN KNOWLEDGE
- USE PHYSICAL CONSTRAINTS, E.G. SMOOTHNESS
- AVOID ITERATIVE, COOPERATIVE ALGORITHMS WHICH ARE TOO SLOW AND HARD TO IMPLEMENT W NEURONS

- STRUCTURE of 2½D SKETCH

CONTAINS:

- CONTOURS of SURFACE DISCONTINUITY
- ROUGH DEPTH
- ORIENTATION
- MECHANISM TO KEEP IT CONSISTENT
  E.G. NECKER ILLUSION:



- RETINOCENTRIC (VIEWER-CENTERED)
- VARIABLE RESOLUTION (DUE to FOVEA)
- REBUILT for EACH EYE MOVEMENT

15

- CONCLUSIONS:
  - * AT LEAST 1 INTERNAL REP. of Depth and Orientation w/ each point
  - * Representation preceeds Decomposition/ Segmentation of Scene into Objects (usually)
  - * Construction driven almost entirely by perceptual processes; only slightly by domain-specific knowledge
  - * Many (independent) processes contribute to it

3D MODELS

- DESCRIPTION of 3D SHAPES AND THEIR SPATIAL ORGANIZATION; SUITABLE FOR RECOGNITION
- DESIGN CRITERIA:
  - * Object-Centered coordinate system. easier for recognition, harder to construct
  - * Volumetric primitives, not surfaces. more info, less sensitive to "noise"
  - * Modular Organization hierarchy of description(s). E.g. Hierarchies of "Generalized Cylinders":
    - Spine (Axis)
    - Cross-Section
    - Spatial Relations