Chapter 20. Progress in Computer Vision

The Quest for Artificial Intelligence, Nilsson, N. J., 2009.

Lecture Notes on Artificial Intelligence

Summarized by Heo, Min-Oh and Lee, Sangwoo

Biointelligence Laboratory
School of Computer Science and Engineering
Seoul National Univertisy

http://bi.snu.ac.kr

Contents

Overview of Chapter 20

- Use more information towards object recognition
  - Adding shading information
    - Horn's shape recognizer
  - Adding depth information
    - David Marr's 2½-D Sketch
  - Trial to untangle to recover important 3-D information about the scene
- Finding objects in scenes
  - Reasoning about scenes
    - Knowledge sources in scenes compete and cooperate until a consistent explanation of the scene emerges by consensus
  - Using templates
  - Model-based vision
    - David Marr's 3-D models of objects

Chapter 20. Progress in Computer Vision

20.1 Beyond Line-Finding

Beyond Line-Finding

- Object recognition with more information
  - Adding shading information
    - Horn's shape recognizer 
  - Adding depth information
    - David Marr's 2½-D Sketch 

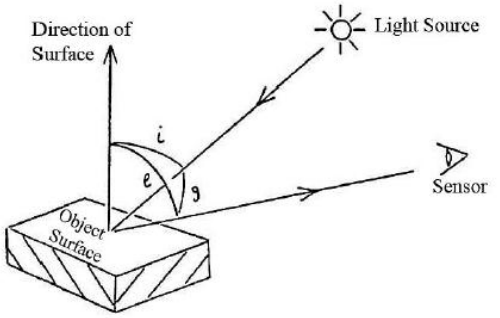


Figure 20.2: Light incident on and reflected by a small piece of a surface.

- David Marr's theory of vision
  - 2-D: Primal Sketch (p.180)
  - 2½-D Sketch
  - 3-D: hierarchy of generalized cones (p.335)

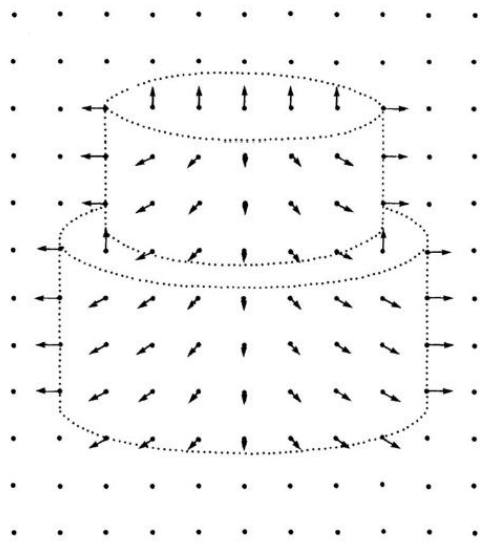


Figure 20.3: A 2½-D sketch.

Beyond Line-Finding

Jay Martin Tenenbaum (left)
Harry Barrow (right)

- Main idea of Intrinsic Images
  - Intensity value of an image resulted from a tangled combination of factors
  - Trial to untangle to recover important 3-D information about the scene
- Intrinsic Images
  - Imaginary images which represent each factors (Fig. 20.5)

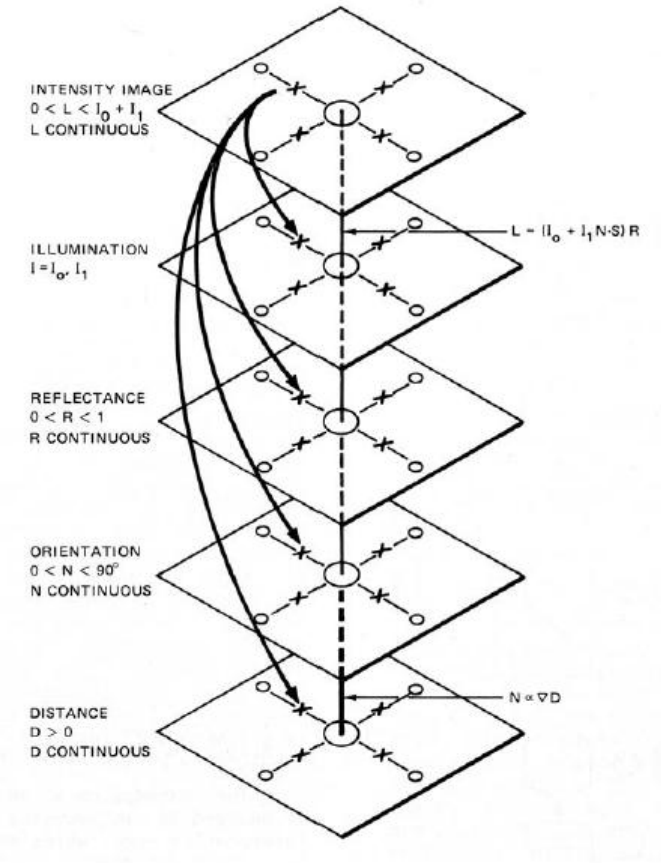


Figure 20.5: Intrinsic images.

Chapter 20. Progress in Computer Vision

20.2 Finding Objects in Scenes

Finding objects in scenes

- Reasoning about scenes
  - MSYS system 
    - Knowledge sources in scenes compete and cooperate until a consistent explanation of the scene emerges by consensus
    - Trying to find the most likely overall set of region interpretation

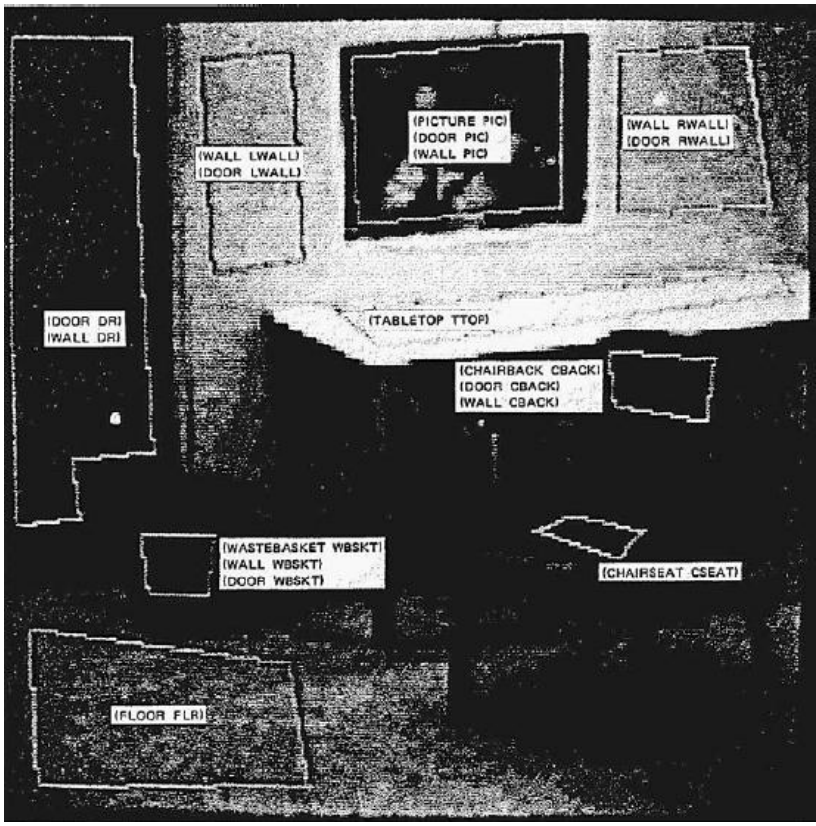


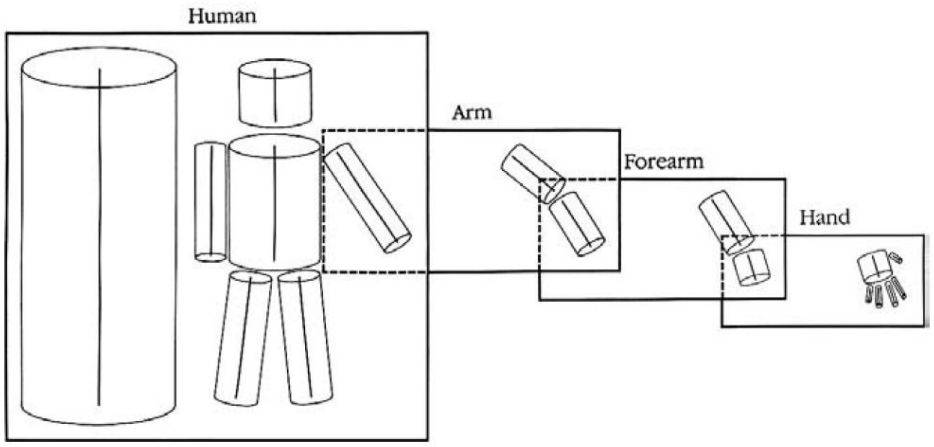
Figure 20.6: An MSYS scene with some regions detected and labeled

Finding objects in scenes

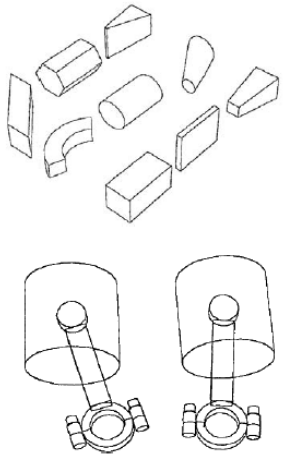
- Using templates
  - Object recognition based on object ‘templates’ that can be matched against images
  - Fischler & Elschlager’s stretchable templates
    - To find objects (faces or particular terrain features) in photos containing target objects
    - Developing dynamic-programming style method

Finding objects in scenes

- Model-based vision
  - Marr's 3-D models of object (in theory of vision)
    - Hierarchies of models (generalized cones)
      - Decomposing into subparts and each subpart is represented as submodels
    - Helping to identify and locate objects in a scene with other 2-D and 2½-D information
  - Brooks – ACRONYM system
    - Used generalized cones to model several different kinds of objects



An example of one of Marr's 3-D model hierarchies



Primitive generalized cones and piston models constructed from generalized cones

Finding objects in scenes

- Using Models (purposive or interactive approach)
  - Marr's position: attempting to use vision to reconstruct entire scenes
  - Y. Aloimonos
    - Claim: the goal of vision is to perceive what is required to guide action
    - The descriptions of space-time that the system needs to derive are not general purpose, but are purposive
  - Neuroscience community
    - "What is vision for? Is a perfect internal recreation of the three-dimensional world really necessary? Biological and computational answers to these questions lead to a conception of vision quite different from pure vision [as advocated by Marr]"

Chapter 20. Progress in Computer Vision

20.3 DARPA's Image Understanding Program

DARPA's Image Understanding Program

- Duration
  - Started in 1976
  - Planned to be a 5-year research
  - Continued until approximately 2001
- Participation
  - MIT, Stanford, University of Rochester, SRI, and Honeywell
  - USC-Hughes Research Laboratories
  - Univ. of Maryland-Westinghouse, Inc.
  - Purdue Univ.-Honeywell, Inc.
  - CMU-Control Data Corporation
- Goal
  - To develop the technology required for automatic and semiautomatic interpretation and analysis of military photographs and related images

DARPA's Image Understanding Program

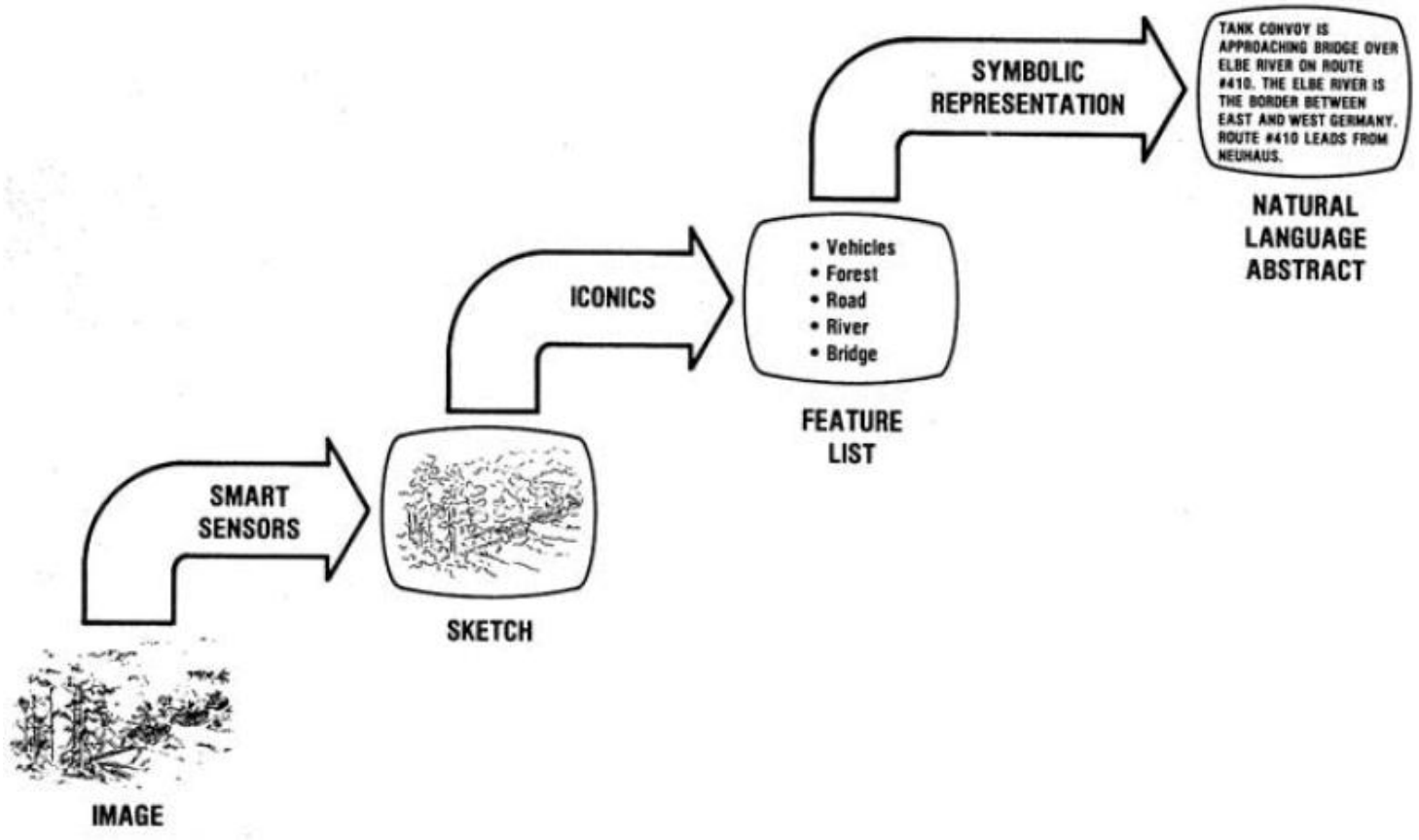


Figure 20.9: An illustration of IU goals.

Chapter 20. Progress in Computer Vision

Appendix

20.1.1 Shape from Shading

Berthold Horn

- Human perceive an appropriately shaded image of a circle as sphere
- Horn's shape recognizer
  - Trial to determine the shape of an object from its shading

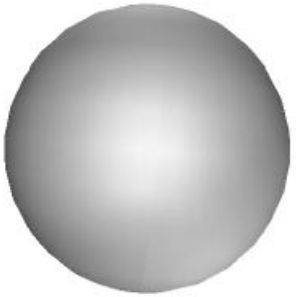


Figure 20.1 (right) a shaded circle

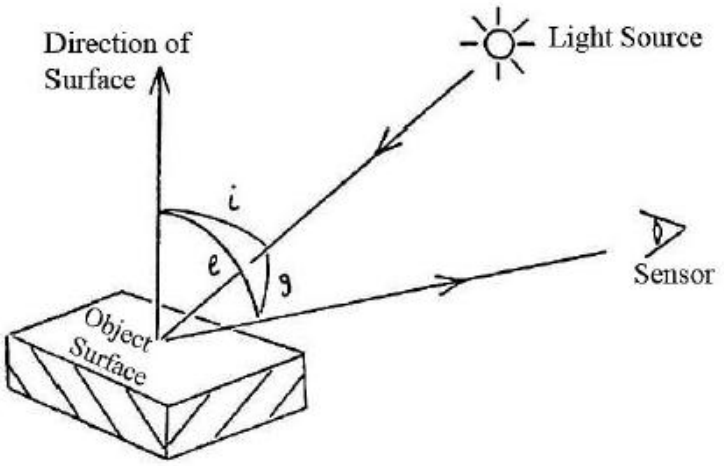


Figure 20.2: Light incident on and reflected by a small piece of a surface.

20.1.2 The 2½-D Sketch

- Viewer infer from image shading and **depth cues** of the scene's 3-D attribute
  - Surface shapes, shapes occluding other shapes, abrupt changes between smooth surfaces, etc.
- David Marr's 2½-D Sketch
  - Include above cues of 3-D attribute
- David Marr's theory of vision
  - 2-D: Primal Sketch (p.180)
  - 2½-D Sketch
  - 3-D: hierarchy of generalized cones (p.335)

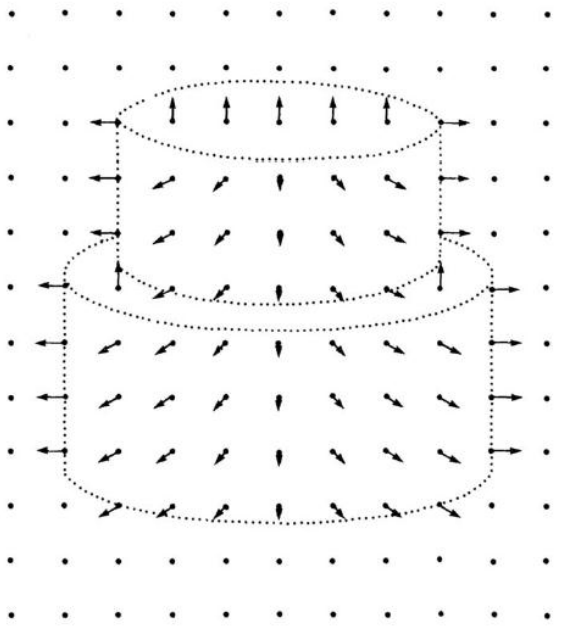


Figure 20.3: A 2½-D sketch.

20.1.3 Intrinsic Images

Jay Martin Tenenbaum (left)
Harry Barrow (right)

- Main idea
  - Intensity value of an image resulted from a tangled combination of factors
  - Trial to untangle to recover important 3-D information about the scene

- Intrinsic Images
  - Imaginary images which represent each factors (Fig. 20.5)

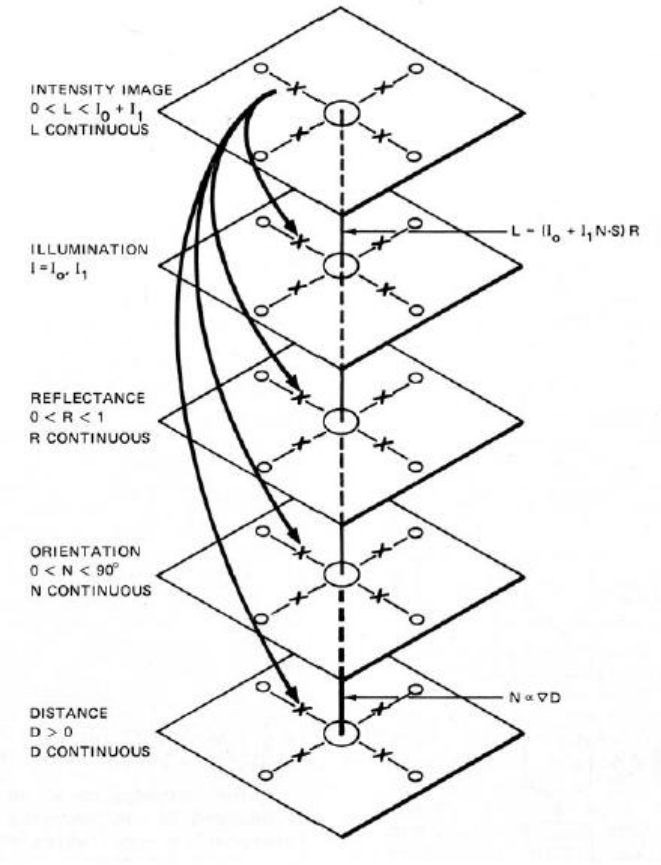


Figure 20.5: Intrinsic images.

20.2.1 Reasoning about Scenes

- MSYS System (Barrow & Tenenbaum)
  - Reasoning about scenes
    - Knowledge sources in scenes compete and cooperate until a consistent explanation of the scene emerges by consensus
    - Trying to find the most likely overall set of region interpretation

MSYS System

- MSYS's reasoning example

1. Regions PIC, WBSKT, and CBACK cannot be WALL or DOOR, because their brightnesses are much less than that along the top edge of the image vertically above them, which violates [knowledge about the brightness of walls and doors].
   Consequently, region PIC must be the PICTURE, WBSKT must be WASTEBASKET, and CBACK must be CHAIRBACK.

2. Region LWALL and RWALL must then be WALL, since they are adjacent to region PIC, and DOOR cannot be adjacent to PICTURE.

3. Region DR cannot be WALL because all regions labeled WALL are required to have the same brightness. Therefore, region DR must be DOOR.

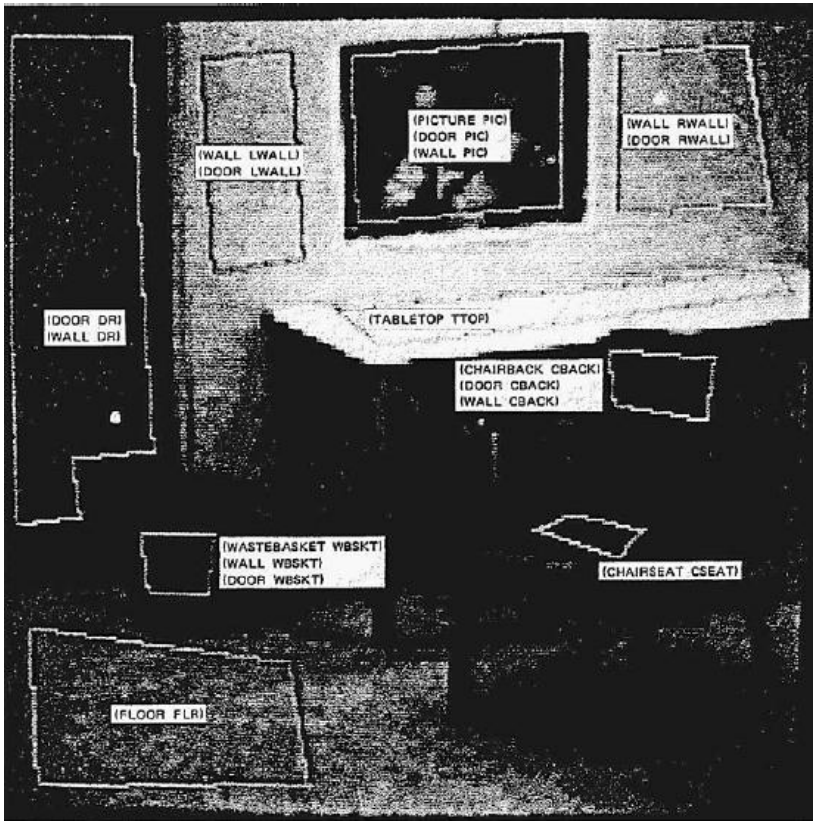


Figure 20.6: An MSYS scene with some regions detected and labeled

20.2.2 Using Templates and Models

- Using Templates
  - Object recognition based on object ‘templates’ that can be matched against images
  - Fischler & Elschlager’s work
    - Stretchable templates
      - To find objects (faces or particular terrain features) in photos containing target objects
      - Developing dynamic-programming style method
    - Main point
      - Specifying an allowable range of spatial relations which the ‘primitive parts’ must satisfy for the object to be present
      - Determining whether or not some of the parts occur depends on whether or not the whole object occurs and vice versa

20.2.2 Using Templates and Models

- Using Models: Model-based vision
  - Marr's 3-D models of object (in theory of vision)
    - Hierarchies of models (generalized cones)
      - Decomposing into subparts
      - Each subpart is represented as submodels
    - Helping to identify and locate objects in a scene with other 2-D and 2½-D information

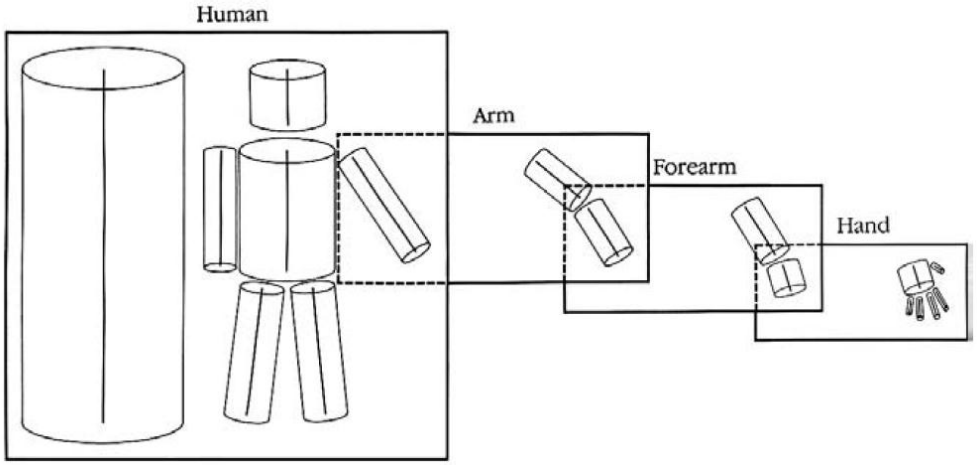


Figure 20.7: An example of one of Marr's 3-D model hierarchies.

20.2.2 Using Templates and Models

- Using Models: Model-based vision
  - Binford
    - Introducing the idea of ‘generalized cylinder’ (also called as generalized cones)
  - Brooks – ACRONYM system
    - Used generalized cones to model several different kinds of objects

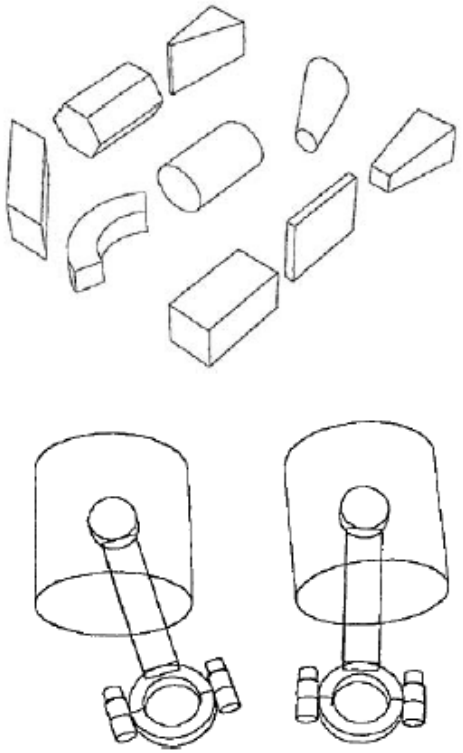


Figure 20.8: Primitive generalized cones and piston models constructed from generalized cones.

20.2.2 Using Templates and Models

- Using Models (purposive or interactive approach)
  - Marr's position: attempting to use vision to reconstruct entire scenes
  - Y. Aloimonos
    - Claim: the goal of vision is to perceive what is required to guide action
    - The descriptions of space-time that the system needs to derive are not general purpose, but are purposive
  - Neuroscience community
    - "What is vision for? Is a perfect internal recreation of the three-dimensional world really necessary? Biological and computational answers to these questions lead to a conception of vision quite different from pure vision [as advocated by Marr]"