Lecture 1
Introduction to computer vision
1. Introduction to computer vision

- History
- Perception versus measurement
- Simple vision system
- Taxonomy of computer vision tasks
Exciting times for computer vision

Robotics

Medical applications

Gaming

Driving

Mobile devices

Accessibility
To see

“What does it mean, to see? The plain man's answer (and Aristotle's, too). would be, to know what is where by looking.”

To discover from images what is present in the world, where things are, what actions are taking place, to predict and anticipate events in the world.
The Summer Vision Project

Author: Papert, Seymour A.

Citable URI: http://hdl.handle.net/1721.1/6125
Date Issued: 1966-07-01

Abstract:
The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which allow individuals to work independently and yet participate in the construction of a system complex enough to be real landmark in the development of "pattern recognition". The basic structure is fixed for the first phase of work extending to some point in July. Everyone is invited to contribute to the discussion of the second phase. Sussman is coordinator of "Vision Project" meetings and should be consulted by anyone who wishes to participate. The primary goal of the project is to construct a system of programs which will divide a videsector picture into regions such as likely objects, likely background areas and chaos. We shall call this part of its operation FIGURE-GROUND analysis. It will be impossible to do this without considerable analysis of shape and surface properties, so FIGURE-GROUND analysis is really inseparable in practice from the second goal which is REGION DESCRIPTION. The final goal is OBJECT IDENTIFICATION which will actually name objects by matching them with a vocabulary of known objects.
Just a few years ago...
"Mask RCNN", He et al. 2017
<table>
<thead>
<tr>
<th>Question</th>
<th>Strategy</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>What color is the vase?</td>
<td>classify<a href="attend%5Bvase%5D">color</a></td>
<td>green</td>
</tr>
<tr>
<td>Is the bus full of passengers?</td>
<td>measure[is](combine[and](attend[bus], attend[full]))</td>
<td>yes</td>
</tr>
<tr>
<td>Is there a red shape above a circle?</td>
<td>measure[is](combine[and](attend[red], re-attend<a href="attend%5Bcircle%5D">above</a>))</td>
<td>no</td>
</tr>
</tbody>
</table>
#edges2cats [Chris Hesse]

[“pix2pix”, Isola et al. 2017]
Number of submitted (blue) and accepted (green) papers in CVPR by year.

Source: CVPR 2019, Derek Hoiem
https://medium.com/reconstruct-inc/the-golden-age-of-computer-vision-338da3e471d1
Why is vision hard?
To see: perception vs. measurement
To see: perception vs. measurement
To see: perception vs. measurement

What the machine gets

The camera is a measurement device, not a vision system
The camera is a measurement device, not a vision system.
To see: perception vs. measurement

Depth processing is automatic, and we can not shut it down...

by Roger Shepard ("Turning the Tables")
To see: perception vs. measurement

Depth processing is automatic, and we can not shut it down...

by Roger Shepard ("Turning the Tables")
To see: perception vs. measurement

(c) 2006 Walt Anthony
To see: perception vs. measurement
The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".
Problem set 1
The “one week” vision project

The goal of the first problem set is to solve vision
A Simple Visual System

• A simple world
• A simple image formation model
• A simple goal
A Simple World

The problem of machine recognition of pictorial data has long been a challenging goal, but has seldom been attempted with anything more complex than alphabetic characters. Many people have felt that research on character recognition would be a first step, leading the way to a more general pattern recognition system. However, the multitudinous attempts at character recognition, including my own, have not led very far. The reason, I feel, is that the study of abstract, two-dimensional forms leads us away from, not toward, the techniques necessary for the recognition of three-dimensional objects. The per-

Complete Convex Polygons. The polygon selection procedure would select the numbered polygons as complete and convex. The number indicates the probable number of sides. A polygon is incomplete if one of its points is a collinear joint of another polygon.
A Simple World
A simple image formation model

Simple world rules:
• Surfaces can be horizontal or vertical.
• Objects will be resting on a white horizontal ground plane
A simple image formation model

Perspective projection

Parallel (orthographic) projection
A simple image formation model

(right-handed reference system)
A simple image formation model

Image and projection of the world coordinate axes into the image plane

World coordinates

$\begin{align*}
  x &= X + x_0 \\
  y &= \cos(\theta) Y - \sin(\theta) Z + y_0
\end{align*}$
A simple goal

To recover the 3D structure of the world

We want to recover $X(x,y)$, $Y(x,y)$, $Z(x,y)$ using as input $I(x,y)$
Why is this hard?

Sinha & Adelson 93
Why is this hard?
Why is this hard?

Figure 1. (a) A line drawing provides information only about the $x$, $y$ coordinates of points lying along the object contours. (b) The human visual system is usually able to reconstruct an object in three dimensions given only a single 2D projection. (c) Any planar line-drawing is geometrically consistent with infinitely many 3D structures.

Sinha & Adelson 93
A simple visual system
The input image
Edges

Occlusion

Change of Surface orientation

Contact edge

Horizontal 3D edge

Vertical 3D edge

Shadow boundary
Finding edges in the image

Image gradient:

$$\nabla I = \left( \frac{\partial I}{\partial x}, \frac{\partial I}{\partial y} \right)$$

Approximation image derivative:

$$\frac{\partial I}{\partial x} \approx I(x, y) - I(x - 1, y)$$

Edge strength

$$E(x, y) = |\nabla I(x, y)|$$

Edge orientation:

$$\theta(x, y) = \angle \nabla I = \arctan \frac{\partial I/\partial y}{\partial I/\partial x}$$

Edge normal:

$$n = \frac{\nabla I}{|\nabla I|}$$
Finding edges in the image

\[ \nabla I = \left( \frac{\partial I}{\partial x}, \frac{\partial I}{\partial y} \right) \]

\[ n = \frac{\nabla I}{|\nabla I|} \]

\( I(x,y) \) and \( E(x,y) \) and \( n(x,y) \)
Edge classification

• Figure/ground segmentation
  – Using the fact that objects have color

• Occlusion edges
  – Occlusion edges are owned by the foreground

• Contact edges
From edges to surface constraints

$Y(x,y)$

$Z(x,y)$

$X(x,y)$

?
From edges to surface constraints

• **Ground**

\[ Y(x,y) = 0 \] if \((x,y)\) belongs to a ground pixel

• **Contact edge**

\[ Y(x,y) = 0 \] if \((x,y)\) belongs to foreground and is a contact edge

• **What happens inside the objects?**

... now things get a bit more complicated.
Generic view assumption: the observer should not assume that he has a special position in the world… The most generic interpretation is to see a vertical line as a vertical line in 3D.

Freeman, 93
Non-accidental properties

Perceptual Organization and Visual Recognition

D. Lowe, 1985

Principle of Non-Accidentalness: Critical information is unlikely to be a consequence of an accident of viewpoint.

Three Space Inference from Image Features

<table>
<thead>
<tr>
<th>2-D Relation</th>
<th>3-D Inference</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Collinearity of points or lines</td>
<td>Collinearity in 3-Space</td>
<td></td>
</tr>
<tr>
<td>2. Curvilinearity of points of arcs</td>
<td>Curvilinearity in 3-Space</td>
<td></td>
</tr>
<tr>
<td>3. Symmetry (Skew Symmetry?)</td>
<td>Symmetry in 3-Space</td>
<td></td>
</tr>
<tr>
<td>4. Parallel Curves (Over Small Visual Angles)</td>
<td>Curves are parallel in 3-Space</td>
<td></td>
</tr>
<tr>
<td>5. Vertices—two or more terminations at a common point</td>
<td>Curves terminate at a common point in 3-Space</td>
<td></td>
</tr>
</tbody>
</table>

Figure 4. Five nonaccidental relations. (From Figure 5.2, Perceptual organization and visual recognition [p. 77] by David Lowe. Unpublished doctoral dissertation, Stanford University. Adapted by permission.)
Non-accidental properties
in the simple world

Using $E(x,y)$
Using $\theta(x,y)$
From edges to surface constraints

How can we relate the information in the pixels with 3D surfaces in the world?

• **Vertical edges**

$$x = X + x_0$$

$$y = \cos(\theta) Y - \sin(\theta) Z + y_0$$

Given the image, what can we say about $X$, $Y$ and $Z$ in the pixels that belong to a vertical edge?

$$Z = \text{constant along the edge}$$

$$\frac{\partial Y}{\partial y} = \frac{1}{\cos(\theta)}$$
From edges to surface constraints

- **Horizontal edges**

Given the image, what can we say about $X$, $Y$ and $Z$ in the pixels that belong to an horizontal 3D edge?

$$Y = \text{constant along the edge}$$

$$\frac{\partial Y}{\partial t} = 0$$

Where $t$ is the vector parallel to the edge

$$t = (-n_y, n_x)$$

$$\frac{\partial Y}{\partial t} = -n_y \frac{\partial Y}{\partial x} + n_x \frac{\partial Y}{\partial y}$$
From edges to surface constraints

• What happens where there are no edges?

Assumption of planar faces:

\[
\begin{align*}
\frac{\partial^2 Y}{\partial x^2} &= 0 \\
\frac{\partial^2 Y}{\partial y^2} &= 0 \\
\frac{\partial^2 Y}{\partial y \partial x} &= 0
\end{align*}
\]

Information has to be propagated from the edges
A simple inference scheme

All the constraints are linear

\[ Y(x,y) = 0 \]

if \((x,y)\) belongs to a ground pixel

\[ \frac{\partial Y}{\partial y} = \frac{1}{\cos(\theta)} \]

if \((x,y)\) belongs to a vertical edge

\[ \frac{\partial Y}{\partial t} = 0 \]

if \((x,y)\) belongs to an horizontal edge

\[ \frac{\partial^2 Y}{\partial x^2} = 0 \]

\[ \frac{\partial^2 Y}{\partial y^2} = 0 \]

\[ \frac{\partial^2 Y}{\partial y \partial x} = 0 \]

if \((x,y)\) is not on an edge

A similar set of constraints could be derived for \(Z\)
Discrete approximation

We can transform every differential constrain into a discrete linear constraint on $Y(x,y)$

$$\frac{dY}{dx} \approx Y(x,y) - Y(x-1,y)$$

A slightly better approximation
(it is symmetric, and it averages horizontal derivatives over 3 vertical locations)
Discrete approximation

Transform the “image” \( Y(x,y) \) into a column vector:

\[
\begin{align*}
0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{align*}
\]

\[
\frac{dY}{dx} \approx Y(x,y) - Y(x-1,y) = Y(2,2) - Y(1,2) = 1
\]

\( x=2, \ y=2 \)
A simple inference scheme

\[ AY = b \]

Constraint weights

\[ Y = (A^T A)^{-1} A^T b \]

\[ \frac{\partial Y}{\partial y} = \frac{1}{\cos(\theta)} \]
Results

- Edge strength
- 3D orientation
- Contact edges
- Depth discontinuities
Changing view point

New view points:
Impossible steps
Impossible steps
Tasks: what machines are good at

Seeing small and precise details

Unwarped Iris image

Iris code

http://www.ind.homeoffice.gov.uk/managingborders/technology/iris/
Tasks: what machines are good at

Doing many times the same thing without losing attention
Tasks: what machines are good at
Tasks: what machines are good at

Image stitching
Tasks: what machines are not so good at

Great at tasks that require exact computations, memory and exploration

Not so good on common sense reasoning, like what is needed for going from an arbitrary set of visual instructions to behavior.
Tasks: what machines are not so good at

Visual intelligence, reasoning
Tasks: what humans care about
Tasks: what humans care about

Verification: is this a building?

Recognition: which building is this?
Tasks: what humans care about

Image classification: list all the objects present in the image

- Building
- Grass
- People
- Trees
- Sky
- Columns
- …
Tasks: what humans care about

Scene categorization

- Outdoor
- Campus
- Garden
- Clear sky
- Spring
- Group picture
- …
Tasks: what humans care about

Semantic segmentation:
Assign labels to all the pixels in the image

Related tasks:
• Semantic segmentation
• Object categorization
Tasks: what humans care about

Detection: Locate all the people in this image
Tasks: what humans care about

Recognition: who is this person?
Tasks: what humans care about

Rough 3D layout, depth ordering
Tasks: what humans care about

Making new images
Tasks: what humans care about

Adding missing content

Input image

Colorized output
Tasks: what humans care about

Predicting future events

What is going to happen?
Tasks: generic formulation

\[ \mathbf{x} \xrightarrow{f} \mathbf{y} \]

- \( \mathbf{x} \): Image / Sequence
- \( f(x) \): Function
- \( \mathbf{y} \): Labels
  - Image/sequence
1. Introduction to computer vision

- History
- Perception versus measurement
- Simple vision system
- Taxonomy of computer vision tasks