Computer vision

Develop computational models and algorithms that can be used for solving visual tasks and for interacting with the world.

Applications:
- robot vision (autonomous vehicles, manufacture, hazardous environments)
- automatic inspection
- satellite images
- analysis and visualization of medical data
- teleoperation, data compression
- management of huge image databases (storage, retrieval)

Approaches

- Mathematical modeling of
  - image formation (geometry, illumination)
  - information contents
- Experimental evaluation
  - development and evaluation of algorithms
  - experimental setup
- Relations to biological vision (human, mammals)
  - neuroscience
  - psychophysics

What is meant by active process

Active:
The vision system has the ability to control its sensory parameters (e.g. view direction, tracking, focus).
Images are not given beforehand.
Several problems can be simplified by such control.

Process:
Vision is not performed in isolation (related to task/behaviors).
No "final solution". Perception is a result of hypothesis generation/verification.
Attention is essential to keep down the complexity.
Image analysis

- **Purpose:** Generate a useful description of the image
- **Examples:**
  - Image classification
  - Environmental control, cartography
  - Character recognition, fingerprint analysis
  - Quality control

Components in a computer vision system

*< underdetermined 2D \(\rightarrow\) 3D problem >*

Main assumptions:

- The world we observe is constructed from coherent matter.
- We can therefore perceive it as constructed from smooth surfaces separated by discontinuities (of simple nature).

In human vision, this way of perceiving the world can be said to precede understanding.

The importance of discontinuities

Under general assumptions, a *discontinuity in image brightness* may correspond to a discontinuity in

- depth
- surface orientation
- surface structure
- illumination

Subjective contours

The image physically consist of 3 circles with pie shaped wedges removed and 3 angles formed from straight lines. When you look at the images, you will likely see one triangle on top of another. The top triangle typically appears lighter, although the background is physically uniform. The triangular forms and the apparent lightness of the top triangle in comparison with the bottom one are subjective. Note that the illusion occurs with both colored and black/white stimuli.

Depth illusion

Biological Vision vs. Computer Vision

Human vision: Existence proof
- $10^5$-$10^6$ years of evolution
- adequate
- not perfect (illusions, ambiguities, inconsistencies)

Computer Vision: Automated visual perception
- over 2-3 decades
- limited success

Try to learn/get inspiration from biological vision - methods of analysis
- neurophysiology
- psychophysics

Basic question:
What information necessary and how to determine it?
What is Perception?

Visual perception is not:

- "seeing the world as it is"
- "transferring a picture to the brain"

Coren, Ward and Enns:
“Sensory stimulation provides the data for our hypotheses about the nature of the external world and it is these hypotheses that form our perceptions of the world.”

Should machines see as humans?

- Emulation of human vision still not possible
  - limited (and speculative) knowledge
  - human cognition not understood
- Human vision not perfect
  - illusions, ambiguities
  - inconsistencies
  - hallucinations, overlook
- Computational constraints on digital computers different from biological tissue
  - parallelism
  - complexity/speed
  - learning
  - human vision restricted by evolution

Fraser’s spiral

Radiometry: measurement of optical radiation, electromagnetic radiation between $3 \times 10^{11}$ and $3 \times 10^{16}$ Hz. This corresponds to wavelengths between $10$ and $10^6$ nm, which included regions called the ultraviolet, the visible and the infrared.
Photometry

Photometry: measurement of light, electromagnetic radiation detectable by the human eye. It is restricted to the wavelengths from about 360 to 830 nm.

Photometry is just like radiometry except that everything is weighted by the spectral response of the eye.

Image formation

Image formation is primarily a physical process that captures scene illumination through a lens system and relates the measured energy to a signal.

Image formation in Digital Camera

- Optic Parameters of the Lens
  - lens type, focal length, field of view, angular apertures
- Photometric/Radiometric Properties:
  - type, intensity, and direction of illumination
  - reflectance properties of the viewed surfaces
  - characteristics of the photoreceptors
- Geometric Parameters
  - type of projections
  - position and orientation of the camera in space
  - perspective distortions introduced by the imaging process
- Intrinsic Parameters
  - physical properties of the photosensitive matrix
  - quantization of the intensity scale

Basics

Hardware, Digital Camera

- CCD (Charged Coupled Device): Irradiance to video
- Frame Grabber: Video signal to digitized array.
- Host Computer: Image Processing and Vision

Digital Image Representation

- Numerical Matrix, $E$, with $N$ rows and $M$ columns
- $E(i,j)$ denotes image value at pixel (picture element i,j)
- Encodes the intensity received by the photosensor of the CCD that contributes to that pixel
- $E(i,j)$ in the range of [0,255] (typically)
Pinhole camera - Perspective geometry

The coordinate systems in the world and in the image domain are parallel. The optical axis is \( \perp \) image plane.

The image plane is usually modeled in front of the optical center.

Segmentation

Simplest Image Segmentation is done by thresholding. This requires that an object has a homogenous intensity and a background with a different intensity level. Such an image can be segmented into two regions by simple thresholding:

\[
g(x, y) = \begin{cases} 
1 & \text{if } f(x, y) > T \\
0 & \text{otherwise}
\end{cases}
\]

Example:
### Connected component

- For every $p$, the set of all points $q$ connected to $p$ is said to be a **connected component**.

### Medial axis/ Skeleton

Let $B$ be a set of boundary points. For each point $p$ in the region, find its closest neighbors (using some metric) on the region boundary. If more than one boundary point is the minimum distance from $x$, then $x$ is on the skeleton of the region.

- **Applications**
  - curve representation, character and fingerprint recognition
- **Desirable properties**
  - no creation of holes, no deletion of blobs
  - no end point erosion, no excessive depth erosion
General

Under rather general assumptions about the image formation the world consists of smooth regular surfaces with different reflectance properties.

One can assume that a discontinuity in image brightness corresponds to a discontinuity in (and their combinations):

- depth,
- surface orientation,
- reflectance, or
- illumination.

Thus, find discontinuities in image brightness (edges) and characterize these with respect to the physical phenomena that gave rise to them.

Edge detection applications

- In computer vision:
  - object recognition
  - line drawing analysis
  - motion algorithms
- In image analysis:
  - segmentation
  - enhancement

Note: An edge representation may correspond to substantial data reduction - 2D data is represented by 1D subset.

Moreover: If a perfect line drawing can be computed, subsequent analysis will be simplified.

Sobel operator and noise suppression

Simplified model of the vision system
Visual front end

The first stages of visual processing performed directly on the raw image data.

- What operations to perform?
  - A mathematical theory exists.
  - Under rather general assumptions it states that the natural operators to apply are convolutions with Gaussian kernels and their derivatives.
  - Inspiration from biology.

Questions

- What principles to follow when designing a vision system?
- If the goal is to solve a specific task
  - in principle, any set of algorithms that solves the task is sufficient.
- If the goal is a flexible system capable of solving a variety of visual tasks
  - aim at general design,
  - generic low-level modules which are shared between mid-level processes for solving different visual tasks.

Structural ideas

- Approach vision problem without strong presumptions about tasks.
- Fundamental question: What information should be extracted at earliest stages?
- Axiomatic approach:
  - Impose structural (symmetry) constraints on an uncommitted visual front-end.
- The output from these can be used as the basis for expressing a large number of visual operations:
  - Feature detection,
  - Optical flow, correspondence, tracking,
  - Cues to surface shape.

Basic constraints on a visual front-end

We perceive objects in the world (and hence image features) as meaningful entities only over certain range of scale.

Ex. molecule (nm) | rain drops
leaf (cm) | cloud
tree (m) | atmosphere
forest (km) |

- Representations of the world must take this property into account.
- Physics: Different types of description
  - particle physics ⇒ quantum mechanics ⇒ thermodynamics ⇒ solid mechanics ⇒ astronomy ⇒ relativity theory
- Cartography:
  - building ⇒ city ⇒ county ⇒ country ⇒ world
Basic constraints (cont)

- Physics and cartography
  Qualitatively different types of models at different scales
- Scale dependency is in clear contrast to certain idealized mathematical entities, such as point and line, which are independent of the scale observation.
- When to process real-world data by automatic methods?

Segmentation

- Goal: Divide a given image into regions which are homogeneous in a certain respect.
- Common additional requirement: Union of neighboring regions should not be homogeneous in the same respect.
- Thus, minimize within variance and maximize between variance.

Segmentation by pixel classification - Overview

- Pixel classification
  - thresholding based on histograms
  - compensation for non-uniform illumination
  - variations: small objects, two thresholds, image smoothing, adaptive threshold
- Application: junction classification
- Feature classification in higher dimensions
  - automatic cluster detection
  - nearest centroid classification
  - nearest neighbor classification
- Watershed segmentation

Examples of classifications

- Object / Background
- Edge / non-edge / corner
- Category X
- Texture type
- ...

Pattern Recognition ..... 

- Sensors give measurements, which should be converted to features (can be pure measurements, e.g., pixels!)
- Ideally, a feature value is identical for all samples in one class

- However:
  - Measurement, discretisation noise
  - Variation between samples
  - Poor features

Statistical Pattern Recognition Learning

Sets of Object Examples: A or B or C or

Set of Objects to be Classified

Object of Known Class

Minimize Error
How to select the categories and tree?
How to estimate the distributions of features for each class?

Example

Problems......

Pattern recognition

Clustering: find natural groups of samples in unlabelled data
Density estimation: make a statistical model of the data
Classification: find functions separating the classes
Regression: fit lines or other functions to data (not in this course)
What defines an object?
- Sometimes very subjective
- Has been well studied

Object recognition may be based on searching for distinctive features that define an object.
Let us consider the letter Z: two horizontal and one diagonal line.

Let us play “Spot the Z”.
How to approximate a curve given a set of pixels?
Stereo

- Use the difference between two (or more) views of the same scene
  - more direct cues to the three-dimensional structure (e.g. position and surface shape)
- reduced ambiguity in 3D to 2D mapping
- If corresponding points can be identified in the left and right images, then, in principle, depth can be computed by triangulation
- Inference of surface shape less critically dependent on specific assumptions

Simple test of your stereo vision

Hold a pen in each hand, at about a relaxed arms length, place the ends toward each other separated about 10cm

- Try to bring the pencils together:
  1. with one eye shut
  2. using both eyes

- About 5-10% of humans are either stereo deficient or have no stereo perception.

Stereopsis in biological vision

- Fusion: The (disparate) images from the left and right eye are merged into a single unified percept
  - Only objects at approximately the same distance as the fixation target will be fused. Other objects give rise to “double vision” (diplopia).
- The word “stereo” comes from the Greek word “stereos” which means firm or solid.

Methods for establishing correspondence

- feature based methods
  - edges
  - corners
  - blobs
- intensity based
  - correlation
  - differential
  - phase based
  - variational
Methods for establishing correspondence

- Several methods are hierarchical (coarse to fine)
  - compute initial estimate at coarse scales and improve these estimates (iteratively) at finer scales
- Sometimes, intensity based methods are guided by an initial feature based matching method
- Combination with the epipolar constraint and/or geometric constraints

Motion

- Observe the world over time \( \Rightarrow \) one more dimension added to the sensory input
- Moving observer \( \Rightarrow \) richer cues to the structure of the world (generalization of geometric stereo)
- Dynamic stereo
  - motion descriptors essential to capture how variations occur in the world
Roles of motion in biological vision

- Avoid obstacles when moving
- Find other objects approaching/leaving (hunt/escape)
- Find direction of heading (navigation and stabilization)
- Cues to depth and shape (compare with stereo)
- Figure-ground segmentation / object segmentation (similar velocities and/or depths)
- Controlling eye movements (saccades and smooth pursuit)

A motion blind patient (Munich 1983)

- Lesion in area 16 ⇒ inability to see visual motion
- Difficulty to pour tea or coffee, because fluid appears to be frozen
- Cloud not stop at right moment, because she could not perceive the movement in the cup when the fluid rose

A motion blind patient (Munich 1983), Cont.

- Difficulty to cross roads
  - “When I am looking at the car first, it seems far away. But then when I want to cross the road, suddenly it is very near.”
- Notable: Defect selective to visual motion perception (e.g., the patient could recognize cars)
  - Moreover, motion perception based on tactile and auditory input was unaffected.