① # COMPUTER VISION

AI - complete problem ↳ requires: (1) Find robust representations of world
↳ signal to symbol
(2) Maintaining + updating them (with Machine Learning
(3) Interfacing with attention, goals and plans.

Challenging as:

↳ ① Inverse optics ∴ 2D → 3D world. Correspondence problem.
↳ ② Inverse graphics ⇒ need to deal with 3D world but only 2D image with surfaces occluded, shading, etc.
↳ ③ Congnitive Penetrance Paradox : challenging to solve problems that are simple for humans. Since cannot reverse engineer - process of seeing things includes a highly complex model that is tough to ~~unpiece~~ ⇒ eg Facial Recognition
↳ ④ Few tasks can be done bottom-up (data driven). Need top-down (prior knowledge) + model-driven.
↳ ⑤ Signal data is often terrible
↳ ⑥ Goals ~~means~~ of problem is often not $\boxed{\text{well posed}}$
↳ Solution (1) exists, (2) unique (3) depends continuously on the data.

↳ ⑦ Pose-invariance is often a large problem
↳ ⑧ We have to be able to deal with objects that haven't been seen before. (wide variety)

Pixel Arrays, CCD/CMOS sensors, image coding

spatial resolution determined by CCD density and lens properties
↳ pixel size limited by photon flux into small areas
↳ per $\boxed{\text{pixel}}$

↳ CMOS / CCD cameras contain independent sensors, converting incident photons (focused by lenses) into charge proportional to light energy.
↳ charge is coupled to allow voltage to be read out easier

RGB-D sensors capture colour and depth information

filter layer
sensor arrays

↳ Luminance Resolution is the number of distinguishable gray levels = number of bits per pixel
∴ colour arises from three CCD arrays (three types of sensors)

↳ S Video = Luma and Chroma channels. ↳ Composite video uses high-frequency chrominance burst

↳ Framegrabber (strobed sampling block) contains high-speed ADC to discrete video into frames.
↳ video frames stored in 3 byte arrays (each different colour planes)

$\boxed{\text{IMAGE FORMATS}}$ see slide 25:
↳ generally 8 bits / col / pixel = 24 bits / pixel
↳ need to revert format to use image

NYQUIST SAMPLING THEOREM: highest spatial frequency component of information in an image = ½ sampling density of pixel array

(2) Pixel array with ~~640,000~~ columns can represent spatial frequency components of image structure no higher than 320 cycles/image. If image frames sampled at 30 fps, max temporal frequency component of information within moving sequence is 15 Hertz.
↳ can use NIR to do iris mapping ⇒ with pixel variance and mean; imaging the ratio.

BIOLOGICAL VISUAL MECHANISMS

Neurones: richly interconnected cells (analogue + digital) with non-linear, adaptive features. Consist of enclosing membrane ⇒ voltage difference between inside and outside.
↳ lipid bilayer (100 pF) - pores that are differentially selective to ions. Cross the neural membrane through protein pores

(31)

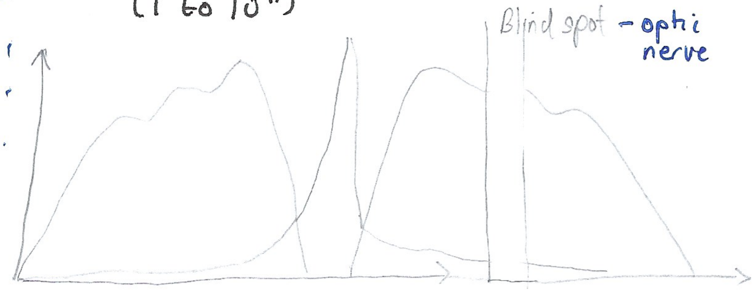
Photo-chemical isomerisation
11-cis-retinal + hν → all-trans-retinal
↓
Carbon double bond flips from cis to trans + causes pore to close to Na^+ ions.

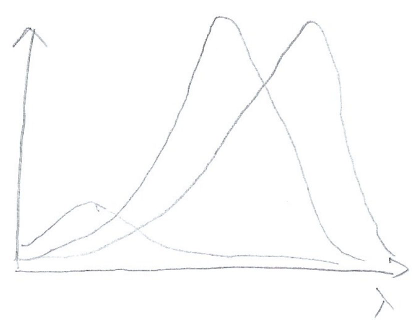
As more positive ions flow into the neurone, voltage becomes positive on the inside reducing membrane's resistance allowing more to enter. This breakdown in resistance constitutes the nerve impulse. (after refractory period (2ms) it's active again). Impulses propogate down axons at $100\,ms^{-1}$. Summations of current flows into neurone from other neurones at synapses ⇒ triggering impulse.

Neural activity is asynchronous - 300 Hz. 2/3 rds of brain receives visual input.
↳ 30 different visual areas (specifically Primary Visual Cortex + Occipital Lobe)

Retina: 1mm thick, 120 mn light-sensitive photoreceptors → 6mn cones + rest are rods.
↳ part of the brain!
120mn photoreceptors
1mn output channels
↳ can reliably see individual photons (1 to 10")
400 - 700 nm
→ axons of Ganglion cells → image processing at first synapse + temporal processing at second synapse
→ specialised red, blue, green
good for night vision
mostly near fovea - central 20°

As Na^+ ions actively pumped (dark current), increased resistance causes increase trans-membrane voltage.





P(photon capture)
$\lambda = c/\nu$

↳ distal neurones are analogue devices. Photoreceptors respond to absorption by hyperpolarisation

Rodes and cones distributed in hexagonal lattices with varying relative densities
↳ imperfect, ~~not~~ → incoherent, not crystalline

Retina network: multi-layered - 3 nuclear layer + 2 plexiform layers
↳ synaptic interconnections
↳ photoreceptors at rear ; 2 directions of signal flow → bipolar cells → ganglion cells
↳ ① Longitudinal (photoreceptors)
↳ ② horizontal + amacrine cells, outer/inner plexiform

3

↳ Therefore, both convergent and divergent signals.
↳ Centre-surround comparisons implemented by bipolar cells
↳ Temporal differentiation by amacrine cells, for motion detection
↳ Separate channels for sustained vs transient image information by different classes of ganglion cells
↳ Right and left visual fields project to different brain hemispheres
↳ at optic chiasm ⇒ crosses over to project only to contralateral brain hemisphere

↳ Projects only to the same brain hemisphere
↳ share information with Corpus Callosum

↳ Projections then go to Lateral Geniculate Nucleus (LGN) - in thalamus
↳ model builds here ⇒ neurones receive input primarily from one eye with left and right eye alternating.
↳ Ocular Dominance Columns have cycle of 1mm.

Orientation Selectivity : new tuning variable
↳ neurones in orientation columns respond to image structures in a preferred range of orientations ⇒ arises from alignment of isotropic subunits in LGN.

↳ Constructed into hypercolumns

Spatial Image Encoding
↳ 5 main DOF in spatial structure: ⇒ position (x,y), orientation, receptive field size, phase.
↳ inferred from boundaries between excitatory and inhibitory regions - bipartite and tripartite.
↳ Receptive field profiles well described by 2D Gabor Wavelets.

MATHEMATICAL IMAGE OPERATIONS

Image processing is built with 2D convolutions of an image with small kernel arrays.
↳ eg. Edge Detection, Filtering, Feature Extraction.
CONVOLUTION ⇔ FILTERING ⇔ FOURIER OPERATION
↳ convolutions in the Fourier domain are much faster
↳ multiplication (given FFT)

Image is superposition of many 2D Fourier components : $f(x,y) = \exp(i\pi(\mu x + \nu y))$
↳ 2D spatial frequency $= \sqrt{\mu^2+\nu^2}$, orientation $= \arctan(\nu/\mu)$
Adding conjugate pair is real valued wave

Convolution Theorem

$f(x,y)$ has FFT $F(\mu,\nu)$, $g(x,y)$ has FFT $G(\mu,\nu)$

$$f(x,y) * g(x,y) = h(x,y) = \int_\alpha \int_\beta f(\alpha,\beta)\, g(x-\alpha, y-\beta)\, d\beta\, d\alpha$$

$$H(\mu,\nu) = F(\mu,\nu)\, G(\mu,\nu)$$

$h(x,y)$ is normally subjected to non-linear operations of various kinds of analysis, segmentation, pattern recognition and object classification.

↳ For explicit convolution algorithm, see (60)

Actually, $O(2(\log_2(n)+1))$

Differentiation Theorem: Computing derivatives of $f(x,y)$, ~~$F(\mu,v)$~~ is equivalent to multiplying its 2DFT, $F(\mu,v)$ by the corresponding spatial frequency coordinate $(x\, i)$ raised to the power equal to order of differentiation

$$\left(\frac{\partial}{\partial x}\right)^m\left(\frac{\partial}{\partial y}\right)^n f(x,y) \xRightarrow{2DFT} (i\mu)^m (iv)^n F(\mu,v)$$

Notably, for Laplacian:

$$\nabla^2 f(x,y) \equiv \left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right) f(x,y) \xRightarrow{2DFT} -(\mu^2+v^2)F(\mu,v)$$

$\boxed{\text{EDGE DETECTION}}$

Why? - demarcate boundaries + occlusions. Helps solve stereo correspondence problem

↳ DISCONTINUITIES = INFORMATION

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

can be discretized by finite differences

↳ convolution with FINITE DIFFERENCE KERNEL

$[-1, 1]$

↳ can concatenate to get higher derivatives + 2D kernel

↳ Effectively high pass filters

Grad direction $\theta = \arctan\left(\frac{\partial f}{\partial y} \Big/ \frac{\partial f}{\partial x}\right)$

$$\|\vec{\nabla} f\| = \sqrt{\left(\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2\right)}$$

apply threshold on this

$$\begin{bmatrix} -1 & 2 & -1 \\ -1 & 2 & -1 \\ -1 & 2 & -1 \end{bmatrix}$$

But this only works in a specific orientation

Integrating vertically, second derivative horizontally

Isotropic Operator - Laplacian: has no preferred orientation - approximation is: $\begin{bmatrix} -1 & -2 & -1 \\ -2 & 12 & -2 \\ -1 & -2 & -1 \end{bmatrix}$

Laplacian of Gaussian works well:

$$\nabla^2 G_\sigma(x,y) = \left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right) G_\sigma(x,y) = \frac{x^2+y^2-2\sigma^2}{2\pi\sigma^6}\exp(-(x^2+y^2)/2\sigma^2)$$

↳ smoothing parameter (σ)

↳ to extract good information, we define a scale analysis:
- → multi-scale family of filters
- → Laplacian pyramid

extracts image structure in octave bands of spatial frequency

In 2D fourier domain, ∇^2 multiplies by paraboloid (μ^2+v^2)

↳ Blurring Laplacian by a Gaussian limits the high-frequency components:

(5) Scale parameter σ determines where the high-frequency cut-off occurs. Zero-crossings correspond to edge locations. Bandwidth of $\nabla^2 G_\sigma(x,y)$ filter is 1.3 octaves.

↳ Logan's Theorem shows this doesn't satisfy one-octave constraint.

↳ Doesn't matter what order Laplacian and Gaussian are applied

CANNY EDGE OPERATOR Removes spurious edges that are detected.

1. Smooth image with Gaussian filter to reduce noise
2. Compute $\vec{\nabla} I(x,y)$ over image
3. Non-max suppression to remove spurious edges
4. Double threshold to local gradient magnitude: strong, weak, suppressed
5. Impose connectivity constraint: edges that are weak + not connected to strong are eliminated

WAVELETS & ACTIVE CONTOURS

Gabor Wavelets:- proposed as model for receptive field profiles of neurones in visual cortex. Wavelets optimal for extracting orientation, position and modulation of image structure. Achieves theoretical lowerbound over variables.

Made of wavelet family codebook:

position in image

$$f(x,y) = \exp\left(-\left[(x-x_0)^2/\alpha^2 + (y-y_0)^2/\beta^2\right]\right)\exp\left(-\left[u_0(x-x_0) + v_0(y-y_0)\right]\right)$$

effective width and length

modulation of spatial frequency

$\omega_0 = \sqrt{(u_0^2 + v_0^2)}$

$\theta_0 = \arctan(v_0/u_0)$

Given a generic 2D Gabor wavelet, $\psi(x,y)$ can generate daughter wavelets:

$$F(u,v) = \exp\left(-\left[(u-u_0)^2\alpha^2 + (v-v_0)^2\beta^2\right]\right) \times \exp\left(i\left[x_0(u-u_0) + y_0(v-v_0)\right]\right)$$

$$\psi_{mpq\theta}(x,y) = 2^{-2m}\psi(x',y')$$

→ to incorporate dilations in size by 2^{-2m}, translations (p,q) and rotations θ

$$x' = 2^{-m}[x\cos\theta + y\sin\theta] - p$$

$$y' = 2^{-m}[-x\sin\theta + y\cos\theta] - q$$

Quadrature Wavelets: used for automatic localisation of facial features

↳ most features can easily be captured using a handful of wavelets.

↳ taking the modulus of facial image after convolving with complex 2D wavelets find features easily.

It is possible to find only circular and parabolic boundary shapes by computing derivatives of contour integrals

$$g(x,y) = \int_\alpha \int_\beta e^{-((x-\alpha)^2 + (y-\beta)^2)/\sigma^2} \cos(\omega(x-\alpha)) I(\alpha,\beta)\, d\beta\, d\alpha$$

$$h(x,y) = \int_\alpha \int_\beta e^{-((x-\alpha)^2 + (y-\beta)^2)/\sigma^2} \sin(\omega(x-\alpha)) I(\alpha,\beta)\, d\beta\, d\alpha$$

$$A^2(x,y) = g^2(x,y) + h^2(x,y)$$

(6)

Hough Transform: find curves whose parameters wrt to increasing radious of contour integras

↳ Hough Transform is a voting scheme to find instances of shapes within certain class of objects.

↳ accumulator space groups edge evidence - parameters of curve.
↳ gradient magnitudes.
↳ output of Canny operator.

↳ For each edge pixel, increment all the compatible accumulator cells. Accumulator cell for which greatest edge evidence found.

Active Contours: deformable shape models (snakes) (spline) - by energy minimisation
↳ pull it towards object contours

Can also split or merge contours as well.

↳ Changes shape under competing forces: (1) Image Forces
(2) Internal Forces - resist excessive deformations

↳ $\boxed{\text{External Energy}}$ - reflects how poorly snake is fitting a contour

↳ $\boxed{\text{Internal Energy}}$ - reflects how much snake is bent or stretched

↳ Sum of energies minimised by: (1) Gradient descent, (2) Simulated annealing (3) PDEs
↳ BUT: numerical instability + stuck in local minima

Scale-Invariant Feature Transform (SIFT)

(1) Object recognition with $\boxed{\text{geometric invariance}}$
↳ photometric invariance

does not deal with non rigid deformations

try to estimate a homography by identifying keypoints that correspond in different images & find transformation

(2) Matching corresponding parts of different images or objects

(3) 3D Scene Understanding + Action Recognition

↳ Feature detectors with orientation index and scale index

find orientation by edge detectors

↳ Eg. extrema, Gaussian image pyramid and resampling

bins of orientation histogram normalized relative to dominant grad direction

↳ for each region, orientation histogram constructed from gradient directions

↳ MATCHING PROCESS → matches sought across wide range of scales + positions, 30° orientation bin sizes.

↳ compare relative configurations of groups of minutiae

↳ Best candidate match determined as nearest neighbors in extracted keypoints
↳ Best bin first priority queue

↳ use Hough Transform voting.

PARALLEL FUNCTIONAL STREAMS

Multiple parallel functional streams in brain for specific visual subdomain: (1) form, (2) colour, etc.

dorsal & ventral hierarchies

Also, conscious and unconscious vision

dorsal stream → Spatial information of visual information

Primary Visual input

ventral stream → Higher level processing of object form

But lots of reciprocal pairwise connections between separate areas.

7

Structure From Texture

supports figure/ground segmentation by dipole statistics

Texture is helpful to identify shapes, both 2D and 3D
↳ defined by statistical correlations across the image
Variations in the texture reveal 3D shape, slant, distance etc.

Quasi-periodicity detected best by Fourier-related methods - can estimate especially using Gabor wavelets. Energy within the periodicities with modulus of Gabor wavelets coefficients reveal texture energy variation
↳ Phase Analysis for person identification is particularly powerful

NB. Resolving textural spectra with location information limited by Heisenberg's Uncertainty Principle + optimized by Gabor wavelets.

Colour Information

$R(\lambda) = I(\lambda)\ O(\lambda)$

wavelength mixture received by camera at corresponding point

wavelength composition of the illuminant

↳ spectral reflectance of the object.

Colour assignments are a matter of calibration

RETINEX

$\boxed{M}$

1. Find max (r_{max}, g_{max}, b_{max}) across all pixels
2. Assume scene contains objects that reflect all red, blue, green etc.
3. $M = I(\lambda)$
4. Hence $(r, g, b) \rightarrow (r/r_{max}, g/g_{max}, b/b_{max})$
   ↳ discounted the illuminant.

Can also be done in local areas rather than just global

Stereo Vision

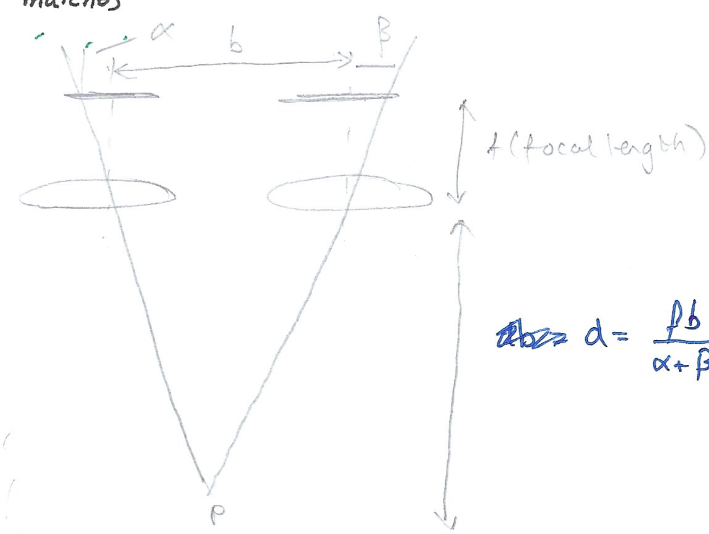
2 eyes with base of separation having stereoscopic disparity, depending on 3D geometry and camera properties. But requires solving correspondence problem.

Parallax: if objects project onto different parts of the images.
↳ disparity ∝ distance of object in front or behind point of fixation

Base of triangulation: ~~increased~~ distance between the two cameras

permutation-matching space is greatly attenuated, terminating with single-pixel precision matches

multi-scale image pyramid ↔ steers search by course-to-fine strategy to maximise efficiency



$d = \frac{fb}{\alpha + \beta}$

Optical Flow

Apparent motion in a scene due to relative motion between observer and the scene (and camera)
↳ ego-motion

Motion estimation requires the solving of the correspondence problem.

Create velocity vector field for image
↳ may be necessary to assign more than one velocity vector to any local image region ↳ motion transparency

Need to detect a coherent overall motion pattern across options
↳ need to disambiguate object motion from contour motion

8

Intensity Gradient Models

$$-\frac{\partial I(x,y,t)}{\partial t} = \vec{v} \cdot \vec{\nabla} I(x,y,t)$$

Dynamic Zero-Crossing Models

Dynamic Zero-Crossing Models: measure image velocity finding edges and contours.

$$-\frac{\partial}{\partial t}\left[\nabla^2 G_\sigma(x,y) * I(x,y,t)\right]$$

Time-derivative of Laplacian of Gaussian-convolved image

In vicinity of Laplacian zero-crossing. Amplitude is estimate of speed, sign of quantity determines direction of motion relative to normal to contours

Fourier Methods

generally looking for correlated signals across time.

$$F(\omega_x, \omega_y, \omega_t) = \int_x \int_y \int_t I(x,y,t) \exp(-i(\omega_x x + \omega_y y + \omega_t t))\, dt\, dy\, dx$$

optical flow also used for localisation through SLAM and LIDAR

Local spatio-temporal spectrum collapses onto 2D inclined plane.

Find motion by applying filters to image sequence observing centre frequencies are co-planar, in this 3-space. Azimuth and elevation correspond to direction and speed of motion.

1. Have $I(x,y,t)$ and $F(\omega_x, \omega_y, \omega_t)$. Detecting $\bar{v} = (v_x, v_y)$
2. $I(x,y,t) = I(x - v_x t_0, y - v_y t_0, t - t_0)$
3. $F(\omega_x, \omega_y, \omega_t) = \exp(-i(\omega_x v_x t_0 + \omega_y v_y t_0 + \omega_t t_0))\, F(\omega_x, \omega_y, \omega_t)$
4. (3) only true if $F() = 0$ where exp factor ~~=0~~ $\neq 1$ — speed $= \sqrt{(v_x^2 + v_y^2)}$
5. $\therefore F(\dots) \neq 0$ only on 3D plane $\omega_x v_x + \omega_y v_y + \omega_t = 0$ — azimuth = direction = $\arctan\left(\frac{v_y}{v_x}\right)$

SURFACE AND REFLECTANCE MAPS

Albedo: fraction of illuminant re-emitted from a surface in all directions

↳ Light reflectance is ~~also~~ dependent on albedo and geometric factors based on angle

① LAMBERTIAN SURFACES: Amount of light (diffuse / matte) reflected dependent on angle of incidence (Lambert's Law) not on angle of emission — $\phi(i,e,g) = \cos(i)$

② SPECULAR SURFACES: Snell's law, perfect reflections — $\phi(i,e,g) = 1$ when $i = e$ ($g = i + e$) else $= 0$

Most surfaces on continuum between Lambertian and specular.

③ LUNAR SURFACES — reflection depending on ratio of cosines of angle of incidence and angle of emission. $\Rightarrow \phi(i,e,g) = \frac{\cos(i)}{\cos(e)}$

↳ Why looks spherical

↳ Faces: $\phi(i,e,g) = \frac{1}{2}(s(n+1)(2\cos i \cos e - \cos g))^n + (1-s)\cos i$

specular: n reflects sharpness

~~specular~~ Lambertian

Reflectance Maps: $\phi(i,e,g)$ fraction of incident light reflected per unit solid angle in direction of camera. Otherwise flux/steradian

i: angle of illuminant

e: angle of light re-emitted

g: angle between emitted ray and illuminant

Shape-from-shading

Shape-from-shading: Requires disambiguation of

- ↳(1) Illumination geometry
- ↳(2) Reflectance properties of surface (and variations)
- ↳(3) Geometry of surface
- ↳(4) Rotations of surface
- ↳(5) Variations in surface albedo

all not well known so that the problem is well-posed.

(9)

SHAPE REPRESENTATION & CODON SHAPE GRAMMARS

Curvature map : $\Theta(s) = \lim_{\Delta s \to 0} \frac{1}{r(s)}$ where local radius of curvature defined as limiting radius of circle that best fits contour at position s.

↳ Curvature sign depends on if circle is inside or outside the figure

↳ Concavities linked with minima

↳ convexities linked with maxima

can result from active contours

Properties of curvature-map descriptions :

↳ (1) Position - independent

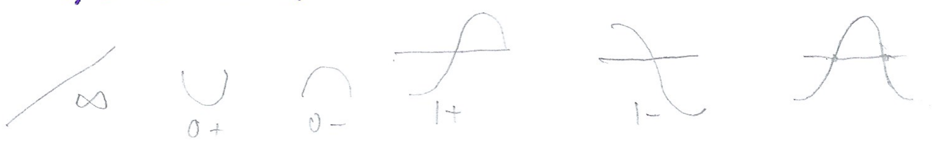
↳ (2) Orientation - independent

↳ (3) Perimeter traversed in opposite direction by changing sign of s.

↳ (4) Scaling property : $\Theta(s) \to K\Theta(s)$ to scale an object.

Codon Grammar: used to create a taxonomy of closed shapes

Number of zero crossings as well as before or after maximum



Therefore, can generate 3 codon pairs, 5 codon triples, 9 codon quads

Therefore, object recognition and classification as follows:

↳ Active contours to fit deformable

↳ Extract codon string from $\Theta(s)$ by traversing outline

↳ Use codon string as index to lexicon

↳ Object then classified by shape with lots of invariance

Description of 3D shape

Superquadratics : represent objects as union and/or intersections of generalised superquadratic closed surfaces, loci of points in (x, y, z) space :

$\boxed{\text{Spheres: } A = B = C}$ $Ax^{\alpha} + By^{\beta} + Cz^{\gamma} = R$

Rotations produce cross terms in (xy, xz, yz). Parameters define object dimensions.

VISION AS MODEL BUILDING

bottom-up path

analysis + recognition

Induction

Signal Features → Compare with model hypotheses and estimate likelihood — Hermeneutical Cycle — Generate new model hypotheses + derive expressions → Symbolic Features

top-down path

synthesis + verification

Deduction

↳ human vision not veridical - illusions expected

↳ can learn from neurological traumas (aphasias and agnosias)

↳ slide 149

(10)

BAYESIAN INFERENCE

Impossible to perform computer vision tasks in a bottom up fashion. Can use bayesian method to use priors

$$p(H|D) = \frac{p(D|H)p(H)}{p(D)}$$

↳ ① Some events more probable than others
↳ ② Matter doesn't disappear
↳ ③ Objects rarely change surface colour
↳ ④ Uniform texturing much more likely
↳ ⑤ Rigid rotation more likely then boundary deformation

↳ can apply the rule recursively using latest posterior as the new prior.

Statistical Decision Theory: Pattern classification on basis of vector of acquired features.
↳ decide whether feature vector is consistent with a particular class.
↳ in 2-state decision problem, feature vectors arise from overlapping probability distribution.

For OCR, slide 161 onwards

$$\text{detectability (discriminability)} = d' = \frac{|\mu_2 - \mu_1|}{\sqrt{\frac{1}{2}(\sigma_2^2 + \sigma_1^2)}}$$

(≥ 3 is normally considered good)

For each class seperately, measure how likely any sample value x: $P(x|C_k)$

$$P(x) = \sum_k P(x|C_k)P(C_k)$$

$$\text{Posterior } P(C_k|x) = \frac{1}{P(x)} \underbrace{P(x|C_k)}_{\text{class conditional likelihood}} \underbrace{P(C_k)}_{\text{Prior}}$$

Minimise total probability if assign each observation to class with highest posterior
↳ can rewrite minimum by as unused denominator in Bayes' probability rule is independent of C_k

$$P(x|C_k)P(C_k) > P(x|C_j)P(C_j) \quad \forall j \neq k$$

Discriminant Functions: construct $\boxed{\text{set of functions}}$ $y_k(x)$ of data x, one function for each class C_k, st classification decisions made by assigning x to C_k if $y_k(x) > y_j(x) \quad \forall j \neq k$
↳ discriminant functions ⇒ normally posterior prob functions: $P(C_k|x)$ or: $P(x|C_k)P(C_k)$

Discriminative Methods: Learn function $y_k(x) = P(C_k|x)$ that maps features x to class labels C_k.

Generative Methods: Learn likelihood model expressing prob data features x would be observed in instances C_k, which can be used for classification using Bayes' rule. Have predictive power as allow samples from joint distribution $P(x, C_k)$

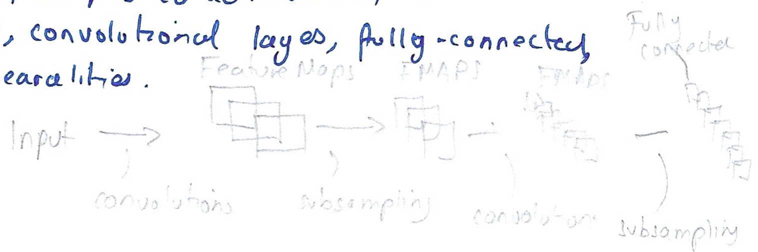
Convolutional Neural Networks: Feed-forward artificial neural networks.

↳ Multiple layers of small collections of neurons

↳ Tiling and overlapping of outputs to achieve shift invariance

↳ pooling layers, convolutional layers, fully-connected, point non-linearalities.

↳ Little pre-processing



OCR

CNN designed by Yann LeCun (slide 170)

↳ input is 32×32.

↳ Trained with 100,000+ examples, using supervised back-propogation. Target output +1, rest to −1. Error back propogate to produce feature maps. Neurons have 5×5 kernes, convolved with input

↳ Trained to extract visual feature. Subsequent feature maps achieve size, slant and style invariance. Neurons in final layer identify input as a target.

Output of each neuron at (i,j) applies non-linear activation function f_{act} to sum of its input pixels × weights w_{mn} and bias term:

$$o_{ij} = f_{act}\left(w_0 + \sum_m \sum_n w_{mn} I_{(i-m),(j-n)}\right)$$

FACE DETECTION, RECOGNITION, IDENTIFICATION

Facial detection very challenging: within-class variation > between-class variation

↳ pose, illumination, family, time

↳ Treat as 3D problem or 2D problem.

Viola-Jones Face Detection:

Use cascade of weak classifiers to build a strong detector. 30+ layers

↳ feature detector with 2D Haar Wavelets. - multiplication ~~add~~ not required, therefore quicker

$$h_j(x) = \begin{cases} -p_j & \text{if } f_j < \theta_j \\ p_j & \text{else} \end{cases}, \quad h(x) = \text{sign}\left(\sum_j \alpha_j h_j\right)$$

At intermediate point, face provisionally accepted if $h(x) > 0$. Only those accepted passed on to next layer

AdaBoost: supervised, adapts weights such that early layers have high accept rates and later are more discriminatory

↳ cascade evaluated using sliding window approach

Gabor Wavelets: act as effective compact code

↳ features represented with handful of wavelets

(12)

Eigenfaces

↳ database of pre-normalised for size, position, frontal pose, decomposed into Principal Components as sequence of orthonormal eigenfunctions with descending eigenvalues

↳ Extract 20 most eigenfaces and for presenting photos, project onto eigenfaces and store coefficients

↳ Accurate (90-95%) + fast to use

↳ But pose and illumination

↳ deal with this by brute force, having lots of cameras.

3D Approaches

Need shape model and texture model

↳ laser, LIDAR, stereo cameras, multiple images

Project texture onto shape model

↳ Can then be used by extracting correct pose to do 2D comparisons

FaceNet

2017 Face Recognition Comp tested with non-ideal images, etc

FaceNet: CNN with 22 layers and 140 mn parameters

↳ trained on 200 mn face images => 8mn identities ∴ 2,000 hours training using back-propagation

↳ use Euclidean distance as metric.

↳ Use triplets of images - one pair from same person, minimise loss function L

$$L = \sum_i \left[\| f(x_i^a) - f(x_i^p) \|^2 - \| f(x_i^a) - f(x_i^n) \|^2 \right]$$

↳ Embedding create compact code for each face

↳ Euclidean distance gives decision of same vs different

Affective Computing

Affective Computing: faces used for emotions - lots of brain to interpret others' faces

↳ use MRIs to show brain areas as interpretting different facial expressions

Facial Action Coding System

Facial Action Coding System: taxonomy of facial expressions

↳ 32 Action Units by 7 muscles

↳ 14 Action Descriptors - use message judgement to use AUs and ADs to get meaning

↳ ① Pre-processing - face detection + normalisation

↳ ② Feature Extraction (1) appearance based or using (2) spatio temporal ideas

↳ ③ AU temporal segmentation, classification, intensity estimation.

generative models - infer state from muscular actions

~~deformat~~ discriminator methods fit a deformable model.

Issues

① Small dataset, often not reliable as well

② Manual scoring required.