Computer Vision
Professor: James L. Crowley

M2R GVR Mid-term Exam   November 2014
Duration: 3 hours

Test conditions: All documents and reference materials are authorized. You may NOT communicate with anyone other than the exam Proctor or the course professor (or his representative). You must answer all questions in INK on the official exam paper. You may use scratch paper to prepare your answer, but your scratch paper will not be graded. You may respond in English or French (or both), but you MUST write legibly. Illegible text will not be graded. Use mathematics as well as English and/or French to communicate.

1) (2 points) What is a Lateral Geniculate Nucleus (LGN)? What function does it provide for the human visual system?

ANS
The optical nerves leave the retina and are joined at Optic Chiasm.
Nerves then branch off to the Lateral Geniculate Nucleus (LGN) and the Superior Colliculus. The LGN acts as a kind of central switching system and relay for the optical nerve. Nerves branch out from the LGN to provide "retinal maps" to the different visual cortexes as well as the "Superior Colliculus".
Surprisingly, 80% of the excitation of the LGN comes from the visual cortex! The LGN seems to act as a filter for visual attention.

2) (2 points) What is the Superior Colliculus? What function does it provide for the human visual system?

ANS
The Superior Colliculus is a Feed-Forward (predictive) control system for binocular fixation. The Superior Colliculus is composed of 7 layers receiving stimulus from the frontal cortex, the lateral and dorsal cortexes, the auditory cortex and the retina. The output layer directly drives the optical muscles to achieve vergence, version and cyclotorsion (rotation around the optical axis).

3) (16 points) A ratio of histograms can be used with a vector of image derivatives, $\vec{X} = (P_x, P_y, P_{xx}, P_{yy}, P_{xy})$ to build a pixel level Bayesian detector for patterns. In this problem, we will use this approach to build a detector for human faces in gray-scale images. Assume that you are given a training set of M images, $P_m(x,y)$, with face regions ($P(x,y) \in face$) indicated in each image, where each image is of size 1024 x 1024 pixels, and each face is of (approximately) of size 64x64 pixels.

a) (2 points) Give the full formulae for the sampled Gaussian derivatives for each of the filters $(G_x, G_y, G_{xx}, G_{yy}, G_{xy})$.
Ans:

$$G_x(i,j,\sigma) = -\frac{i}{\sigma^2}G(i,j,\sigma)$$

$$G_y(i,j,\sigma) = -\frac{j}{\sigma^2}G(i,j,\sigma)$$

$$G_{xx}(i,j,\sigma) = \frac{i^2 - \sigma^2}{\sigma^4} G(i,j,\sigma)$$

$$G_{xy}(i,j,\sigma) = \frac{xy}{\sigma^4} G(i,j,\sigma)$$

$$G_{yy}(i,j,\sigma) = \frac{y^2 - \sigma^2}{\sigma^4} G(i,j,\sigma)$$

Where :

$$G(i,j,\sigma) = \frac{1}{B} W_N(i,j) \cdot e^{-\frac{(i^2+j^2)}{2\sigma^2}}$$

with

$$w_N(i,j) = \begin{cases} 1 & \text{for } -R \le i \le R \text{ and } -R \le j \le R \\ 0 & \text{otherwise} \end{cases}$$

Typically: for R should be $\ge 3\sigma$

The normalization factor B is

$$B = \sum_{x=-R}^{R} \sum_{y=-R}^{R} e^{-\frac{(i^2+j^2)}{2\sigma^2}} \approx 2\pi\sigma$$

b) Give the formula for the 2-D convolution $P_x(x,y) = P * G_x(x,y)$.

$$P_x(x,y) = \sum_{i=-R}^{R} \sum_{j=-R}^{R} P(x-i,y-j)G_x(i,j)$$

c) Explain how to use the training data to obtain the histogram $h(\vec{X})$ for estimating the probability of a vector $p(\vec{X})$.

Ans:

Allocate a 5 Dimensional table $h(\vec{X})$ where each dimension has N cells. h() is of size N^5.

$\forall_m \forall_{i,j}$: Compute $\vec{X}_m(i,j)$ and quantize the values of each component of $\vec{X}$ to N values;

$$h(\vec{X}_m(i,j)) = h(\vec{X}_m(i,j)) + 1$$

The table $h(\vec{X})$ then represents the frequency of occurrence for each possible derivative vector $\vec{X}$.   Given M images of 2^{20} pixels, then the probability is .

$$P(\vec{X}) = \frac{1}{M \cdot 2^{20}} h(\vec{X})$$

d) Explain how to use the training data to train the histogram $h_{face}(\vec{X})$ for calculating the probability of a vector given that the pixel is part of the face, $p(\vec{X} \mid P(x,y) \in face)$.

Ans:

Allocate a second 5 Dimensional table $h_{face}(\vec{X})$ where each dimension has N cells. $h_{face}(\vec{X})$ is of size N^5.

$$\underset{m}{\forall}\ \underset{i,j\in face}{\forall}\ : \text{Compute } \vec{X}_m(i,j) \text{ and quantize the values of each component of } \vec{X} \text{ to N}$$

values.

$$h_{face}(\vec{X}_m(i,j)) = h_{face}(\vec{X}_m(i,j)) + 1$$

The table $h(\vec{X})$ then represents the frequency of occurrence for each possible derivative vector $\vec{X}$. Given M images of 2^{12} pixels, then the probability is .

$$P(\vec{X}\mid face) = \frac{1}{M \cdot 2^{12}} h_{face}(\vec{X})$$

e) (2 points) How large should the histograms be?
Ans:

Choose N such that $8N^5 \leq M \cdot 2^{12}$ i.e. $N < \sqrt[5]{2^9 M}$

f) How can you estimate the a-priori probability that a pixel will be part of a face $p(P(x,y) \in face)$?

Ans:

For the training date : $p(face) = \dfrac{M \cdot 2^{12}}{M \cdot 2^{20}} = \dfrac{1}{2^8}$.
In other words 1 pixel out of 512 is a face pixel.

g) (2 points) Explain how to use the histograms to determine the probability that a pixel represents part of a face given the value of its derivative vector, $p(P(x,y) \in face \mid \vec{X})$.

$$P(face \mid \vec{X}) = \frac{P(\vec{X}\mid face)P(face)}{P(\vec{X})} = \frac{\dfrac{1}{M2^{12}} h_{face}(\vec{X}) \cdot \dfrac{M2^{12}}{M2^{20}}}{\dfrac{1}{M2^{20}} h(\vec{X})} = \frac{h_{face}(\vec{X})}{h(\vec{X})}$$

h) (2 points) How can you determine what value of σ to use?
Ans: This is a very hard question with no easy answer. The best σ depends on the training data. Try all possible values from an exponential scale and use the one that works best.

i) Explain how to compute an estimate of the likelihood that an image window of size WxH contains a face based on the estimated probabilities at each pixel.

ANS:
Given a target probability image $p_{face}(i,j)$ and a ROI of (t,l,b,r):

Sum: $$S = \sum_{i=l}^{r} \sum_{j=t}^{b} p_{face}(i,j)$$

We can estimate the "likelihood" as the average detection probability:

likelihood: $$L = \frac{S}{(b-t)(r-l)}$$

j) Explain how to compute the center of gravity of the face region from the probability of face in the image window of size WxH. The center of gravity should be expressed in image coordinates.

$$x = \mu_i = \frac{1}{S}\sum_{i=l}^{r}\sum_{j=t}^{b} P_{face}(i,j)\cdot i$$

$$y = \mu_j = \frac{1}{S}\sum_{i=l}^{r}\sum_{j=t}^{b} P_{face}(i,j)\cdot j$$

Position is the center of gravity: *(m_i, m_j)*

k) Explain how to compute the second moment of the face region.

$$\sigma_i^2 = \frac{1}{S}\sum_{i=l}^{r}\sum_{j=t}^{b} T(i,j)\cdot (i-\mu_i)^2$$

$$\sigma_j^2 = \frac{1}{S}\sum_{i=l}^{r}\sum_{j=t}^{b} T(i,j)\cdot (j-\mu_j)^2$$

$$\sigma_{ij}^2 = \frac{1}{S}\sum_{i=l}^{r}\sum_{j=t}^{b} T(i,j)\cdot (i-\mu_i)\cdot (j-\mu_j)$$
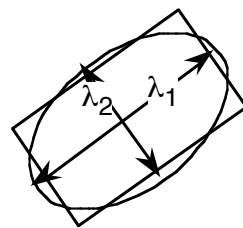
These compose a covariance matrix: $$C = \begin{pmatrix} \sigma_i^2 & \sigma_{ij}^2 \\ \sigma_{ij}^2 & \sigma_j^2 \end{pmatrix}$$

l) Explain how to determine the width, length and orientation of the face region from the second moment.

The principle components (λ_1, λ_2) determine the length and width.
The principle direction determines the orientation of the length.
We can discover these by principle components analysis.

$$RCR^T = \Lambda = \begin{pmatrix} \lambda_1^2 & 0 \\ 0 & \lambda_2^2 \end{pmatrix}$$

where

$$R = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$$

The length to width ratio, λ_1/λ_2, is an invariant for shape.
The angle θ is a "Covariant" for orientation.