Solution notes

Computer Vision 2005 – Paper 8 Question 11 (JGD)

(*a*) [*This question relates to face recognition, and to visual learning.*]

(*i*) Any given presenting face is represented in terms of factors precomputed from the database population of example faces, by a process called Principal Components Analysis (see answer to (ii) below). The database represents "knowledge" about what human faces are like, and what some of their main forms of variability are. Thus the population database should have the same kinds of diversity as the images that will be presented for recognition. [3 marks]

(*ii*) A statistical procedure called Principal Components Analysis finds the major forms of variation among the database face images. Using linear algebraic methods (diagonalizing a covariance matrix to find its eigenvectors and their corresponding eigenvalues), a set of "eigenfaces" are precomputed (these are the eigenvectors). They have the mathematical property that the greatest amount of variability in the database is spanned by the smallest number of basis vectors, up to any desired degree of approximation, when these eigenvectors are used as the representation. Any given face is represented simply as a linear combination of these eigenfaces – usually about 20 to 40 – and that short sequence of numbers (the eigenvalues) is the distinguishing code for a particular face. [4 marks]

(*iii*) A strength of the method is that the representation for a given face is very compact, and so searches can be performed at great speed. A second strength is that the basis vectors (eigenfaces) are orthogonal, ordered by importance, and they capture the greatest amount of variability in the smallest number of terms. The weaknesses of the method are that: (1) the representation is image-based, i.e. two-dimensional appearance based, and so it captures no invariances for pose angle or perspective angle; (2) similarly it has no invariance for changes in facial expression; and (3) it is very sensitive to changes in illumination; – so much so that usually the 2nd or 3rd eigenface is just an illumination factor (e.g. above versus below), and so if a person is enrolled under one type of illumination he tends not to be recognized under another. The same problem occurs if size normalization is imperfect. [4 marks]

(*iv*) The algorithm is efficient because the Principal Components Analysis of the database is precomputed off-line; any given presenting face then only needs to be projected onto each of the precomputed eigenfaces (a simple series of inner product operations). The algorithm can learn over time as more faces are encountered, simply by continuing the process of PCA as the database grows. However, its lack of fundamental invariances is its fatal flaw. In trials of the method, error rates of 43% to 50% have been found either when there are large changes in illumination geometry, or for images taken more than one year apart. [3 marks]

(*b*) The three main terms in Bayes Rules have the following meanings:

$P(C_k|x)$ is the *posterior probability* of object class C_k, given the observation of data set x. This is the outcome of the Bayesian inference calculation: it assigns a probability to each of the possible classification hypotheses.

$P(x|C_k)$ is the *class-conditional likelihood* that the data set x would have been observed, if the object belonged to class C_k. This is one of the ways in which Bayesian inference relies on prior expert knowledge about the problem domain.

$P(C_k)$ is the *prior,* or the *unconditional likelihood* of class C_k. In the absence of data this is the plausibility of the hypothesis C_k, and it is another way in which Bayesian inference incorporates prior expert knowledge. When Bayesian inference is applied in an iterative way for learning over time as more data arrives, the last calculated posterior $P(C_k|x)$ is used as the new prior $P(C_k)$.

[3 marks]

[*This question relates to formal inference methods in pattern classification.*]

(*c*) The *reflectance map* relates intensities in an image to the surface orientations of objects, in terms of angles to the illuminant and to the observer. It specifies the fraction of incident light reflected per unit surface area, per unit solid angle, in the direction of the camera. The three variables on which it depends are: i, the angle of the illuminant, relative to the surface normal; e, the angle of a ray of light re-emitted from the surface; and g, the angle between the emitted ray and the illuminant. [3 marks]

[*This question relates to interpreting shape and surface geometry from shading information.*]