Solution notes

Computer Vision 2005 – Paper 9 Question 11 (JGD)

(*a*) [*This question relates to image analysis and the properties of filters.*]

(*i*) In 2D Fourier terms, this is a *bandpass filter.* It responds only to a certain band of spatial frequencies lying between some lower and upper limits. [2 marks]

(*ii*) This filter does not discriminate between different orientations of image structure, e.g. edges at different angles. It is undiscriminating for angle because both the Laplacian and the Gaussian parts of the definition of $f(x, y)$ are isotropic. [2 marks]

(*iii*) The spatial frequency bandwidth of $f(x, y)$ is approximately 1.3 octaves. [1 mark]

(*iv*) By concatenating a filtering operation (such as convolution with a Gaussian of a certain scale σ) and a differentiating operation (such as taking its Laplacian), one can construct an edge detection operator that only finds the edges in an image at a certain coarseness or fineness: hence "at a certain scale of analysis." The scale parameter σ of the blurring Gaussian, as in the definition given for $f(x, y)$, determines this scale of analysis. A "scale-space fingerprint" is a mapping of the edges detected as a function of changing σ from fine (small) to coarse (large). The larger σ, the fewer edges are detected, and they are only the most pronounced and longest ones. [3 marks]

(*b*) [*This question relates to image formation, and to the inference of object surface properties from image properties.*]

The physical photonic properties of a surface determine how it scatters light, e.g. over a broad range of angles (a Lambertian surface) or only over a narrow range of angles obeying Snell's Law that angle of emission equals angle of incidence between an illuminating ray and the local surface normal (a specular, or mirror-like, surface). Any surface can be described as lying somewhere on a continuum between these two extremes. These different surface behaviours make it difficult to interpret image data in terms of surface shape, since the scattering angles for reflected light depend on unknown factors. In the case of face images, the relative wetness or oilyness of the skin at a given moment can transform the face from a Lambertian surface to a specular surface, thus confounding "shape-from-shading" methods for infering the facial structure. [3 marks]

(*c*) The `.jpeg` image format is designed for compression and consists of Huffman-coded coefficients computed by Discrete Cosine Transforms on local patches of the image. Because pixels themselves are not directly available, one cannot access the luminance or chrominance data as a function of position in the image until the `.jpeg` data is reformatted as (e.g.) `.bmp` data. [1 mark]

(*d*) [*This question relates to biologically-inspired strategies for vision, and in particular to hypothesis-generation-and-testing as the core basis of vision.*]

When visual data leaves the retina down the million fibres of either optic nerve and reaches its first synapse at the thalamus (or LGN, lateral geniculate nucleus), it is met there by a much larger flood of feedback signals coming back down from the visual cortex. This feedback projection is estimated to contain as many as ten times more fibres than the afferent fibres bringing data up from the retina. One interpretation of this puzzling observation is that vision works by a kind of hypothesis-generation and testing process, in which graphical models are constructed in the brain about the external world and the objects that populate it (and these "graphics" are really what one sees); and the graphics are shaped, constrained, and updated by the 2D image data coming from the retina. Hence we see not image data, but 3D models constructed to be consistent with such data. This is the theory of "vision as [inverse] graphics." [4 marks]

(*e*) A problem is defined (by Hadamard) as "well-posed" if all of these three conditions apply: (1) a solution exists; (2) the solution is unique; and (3) the solution depends continuously on the data. In general, the problem of infering object properties from image properties violates one or more of these three conditions. The problems only become soluble by adding ancillary assumptions, or other data (such as past knowledge learned from experience or from other modalities). The task of infering the spectral reflectances of object surfaces from image colours (the wavelengths of light received at the sensor) is soluble only if one knows the wavelength composition of the illuminant, since that distribution is multiplied by the spectral reflectance function of the object surface. The problem of colour inference becomes well-posed and almost instantly soluble if the illuminant is known; but in the absence of that information or of strong assumptions, it is ill-posed. [4 marks]

[*This question relates to the formally ill-posed character of many problems in computer vision, specifically the inference of object properties from image properties.*]