In Press: International Journal of Computer Vision

Hypotheses for Image Features, Icons and Textons

Lewis D Griffin[1] & Martin Lillholm[2]

[1]University College London, UK
[2]IT University of Copenhagen, Denmark

We review ideas about the relationship between qualitative description of local image structure and quantitative description based on responses to a family of linear filters. We propose a sequence of three linking hypotheses. The first, the Feature Hypothesis, is that qualitative descriptions arise from a category system on filter-response space. The second, the Icon Hypothesis, is that the partitioning into categories of filter response space is determined by a system of iconic images, one associated with each point of the space. The third, the Texton Hypothesis, is that the correct images to play the role of icons are those that are the most likely explanations of a vector of filter responses. We present results in support of these three hypotheses, including new results on 2-D 1st order structure.

1. Introduction

Research on qualitative image structure is at least 40 years old and in that time many different terms – features, textons, edges, etc. – have been coined and used in different ways by different authors. Our aim in this paper is to attach some specific meanings to these terms and to propose three hypotheses concerning them. In this introduction we first recap the history of the subject, second we summarize a specific model of image filtering – the Scale Space Gaussian Derivative framework, and finally we preview the three hypotheses that are discussed in turn in the body of the paper.

1.1 Qualitative description of local image structure

Although the array of raw pixel values contains the totality of information in an image, an oft-used first step in computer vision is to make a quantitative description of the local image structure at each image location [1]. The simplest way to achieve this is to extract a small (say 7×7) patch at each location, but the preferred way is to measure image structure by convolution with a family of localised filters [2]. The vector of responses to filters centred at a particular image point is a quantitative description of the image in the locality of that point – where the size of the filters determines the extent of the locality. In terms of information content the vector of measurements is very similar to the small patch, but the information has been repackaged into a

form that is more convenient for subsequent analysis [3]. This strategy of local characterization by a family of filters is also that adopted by biological visual systems [4].

Much can be done with these quantitative descriptions of local structure. They can be used for efficient image coding [5], for the calculation of similarity measures to drive image registration processes [6, 7], or for figure/ground segmentation [8, 9]. However many researchers have felt that it would be useful, and should be possible, to derive **qualitative** descriptions of local structure on the basis of the **quantitative** [10]. The intuition here is that a strategic discarding of information at this early stage, achieved by moving from a high-dimensional continuous-valued description to a lower-dimensional combinatorical description, will make higher-level processing (e.g. recognition) more computationally feasible [11]. To some, this engineering intuition is supported by informal observations that natural images are in the main composed from a limited vocabulary of structures (figure 1) and thus are particularly well-suited to qualitative description. Although these informal observations could simply stem from the use of qualitative description by the human visual system, they remain suggestive; since what is good for biological vision is very probably good for machine vision. However, we note that, at this time, no definitive evidence exists for qualitative description of local structure by biological vision: we are unaware of any psychophysical support, and the physiological data, although strongly suggestive [12-16], is open to alternative interpretations.

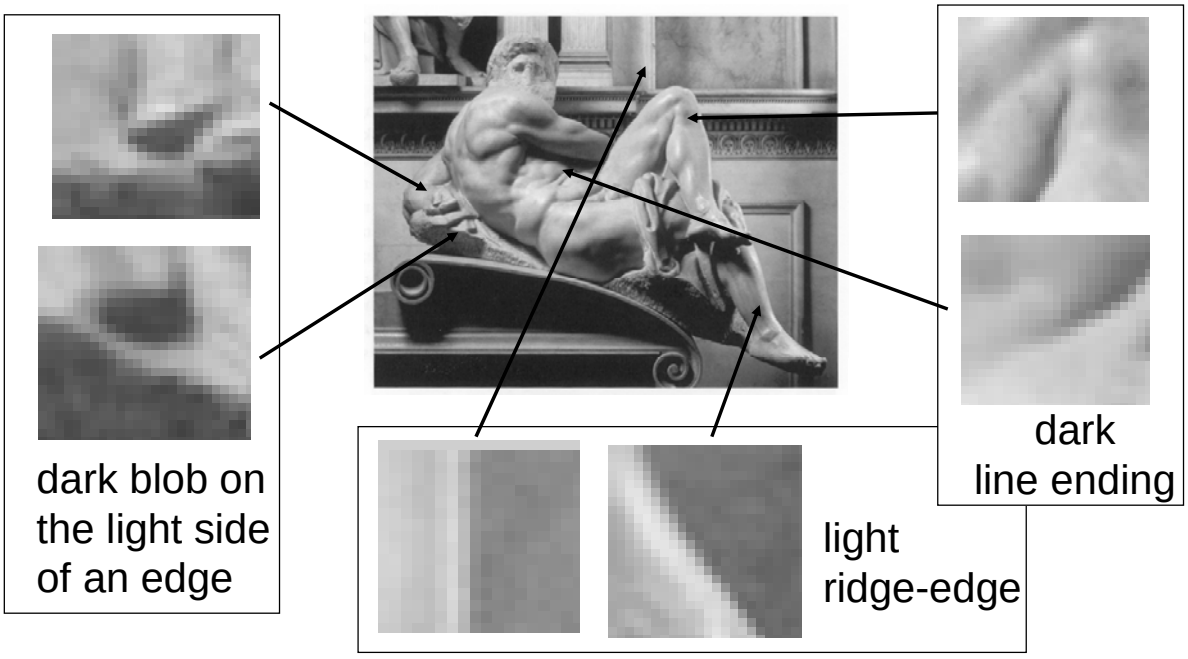

Figure 1 – Much of natural images seems to be composed of the same qualitative structures repeated again and again. One can informally attempt to group and name these repeated structures (for example as shown) but the process becomes very difficult to keep track of once sufficient patches are considered.

Although the hard evidence for qualitative description of local structure in biological vision is scant, the idea did originate in physiology. Barlow, studying the frog retina [17], found neurons that responded only to a specific type of visual stimulation – a small side-to-side moving dark blob – and named them bug detectors. Importantly, the action of these cells could not be understood as a simple linear filtering as they also showed impressive invariance to behaviourally unimportant changes such as the overall luminance level. Reviewing this finding later [18], Barlow analogically described the function of such neurons as signalling the presence of a specific 'letter' in the sensory inputs (later neurons signalling 'words'). This reduction of the input to one of a finite 'alphabet' of possibilities is exactly what we mean by qualitative description.

This approach was taken up in Computer Vision by Marr [19] and elaborated into a scheme to describe local image structure in terms of the feature categories: edge and bar. For the detection of which Marr gave explicit schemes based on the measurements of linear filters [20]. He also explained the link between the quantitative and qualitative descriptions by using the only extant theory of local analysis – differential geometry. Using this, he was able to explain that the loci of zero-outputs of centre-surround filters deserved to be described as 'edges', as such filters could be interpreted as approximately computing the laplacean of a blurred version of the image, and points of zero laplacean, it can be argued, are inevitably close to points of locally maximal gradient; and such points correspond to an informal understanding of 'edge'. Since this reasoning is important, let us labour the point but giving the reasoning in the other direction. The informal idea of 'edge' is captured to a degree by the exact concept: points of maximal gradient. Points of maximal gradient are approximately the same as the exact concept: zero-crossings of the laplacean. Zero-crossings of the laplacean are approximately the same as zero-crossings of the outputs of Marr's centre-surround filters.

The mathematics that links image filtering to differential geometry was greatly expanded by Koenderink in his Scale Space theory of early vision [2, 3, 20-23]. This showed how any required blurred derivative could be precisely computed using the appropriate linear filter, and so a suitable ensemble of such filters could be understood as computing the initial terms of a local Taylor expansion of the blurred image. This elegant analysis makes a theory of qualitative description seem within reach and some successes were had (figure 2). For example, Marr's laplacean scheme was improved so that the detected edges are exactly on points of highest gradient rather than simply near them [24]. To do this one uses filters that are derivatives- rather than differences-of-Gaussians, and one computes not the zero-crossings of $L_{xx} + L_{yy} = 0$ but of $L_x^2 L_{xx} + 2L_x L_y L_{xy} + L_y^2 L_{yy}$ (where L is the image luminance, and subscripts denote differentiation) which can be achieved by combining the outputs of five linear filters.

Promising though the linkage from filters to differential geometry is, as an approach to qualitative description it seems to run out of steam. Four overlapping reasons can be given. (i) differential geometry has no general theory of qualitative aspects of Taylor series. (ii) what qualitative local description there is in differential geometry is limited to the identification of special points (e.g. critical) [25-27] and curves (e.g. parabolic) [28], not to description of general points. (iii) we have no general recipe for producing a combination of derivatives that will signal the presence of a given qualitative structure (e.g. a dark blob by a light bar). (iv) the interpretation of filters as computing the derivatives of the blurred intensity means that in this

approach one is adopting the counter-intuitive strategy of trying to analyze qualitative structure after it has almost blurred away (e.g. in designing a Y-junction detector one would need to characterize the differential structure of a blurred Y-junction).

Figure 2 - Edges, defined as the locus of points where the gradient magnitude is locally maximal in the direction given by the gradient vector, computed on a medical image. The features in this scheme are sparse.

Recently, two developments have renewed the search for methods of qualitative description of local image structure. The first is the observation that in parallel to the view of filters as computing derivatives of the blurred luminance is a second view of them as computing an orthogonal decomposition of a locality in an image [29-32]. Second is the idea that the linkage between filter-response space and qualitative structure is not via special low-dimensional manifolds (e.g. $\nabla^2 L = 0$) in filter response space, but via a partitioning[1] of filter response space

[1] A partitioning of a set *X*, is a collection of subsets of *X* which are pair-wise disjoint and whose union is *X*.

into non-overlapping sub-regions of co-dimension zero[2] – each sub-region corresponding to a category of qualitative structure [11, 33-38]. Because the sub-regions of filter-response space have co-dimension zero, they will give rise to image regions (rather than curves or points) that have a particular feature label. For example, in figure 3 a five-dimensional filter response space has been partitioned into five-dimensional sub-regions. So pixels of the image that receive that label 'edge' form a region. As it happens, these 'edge' regions tend to be long and narrow, but they are formed of pixels as compared to (say) zero-crossings of the laplacean which are curves lying between pixels.

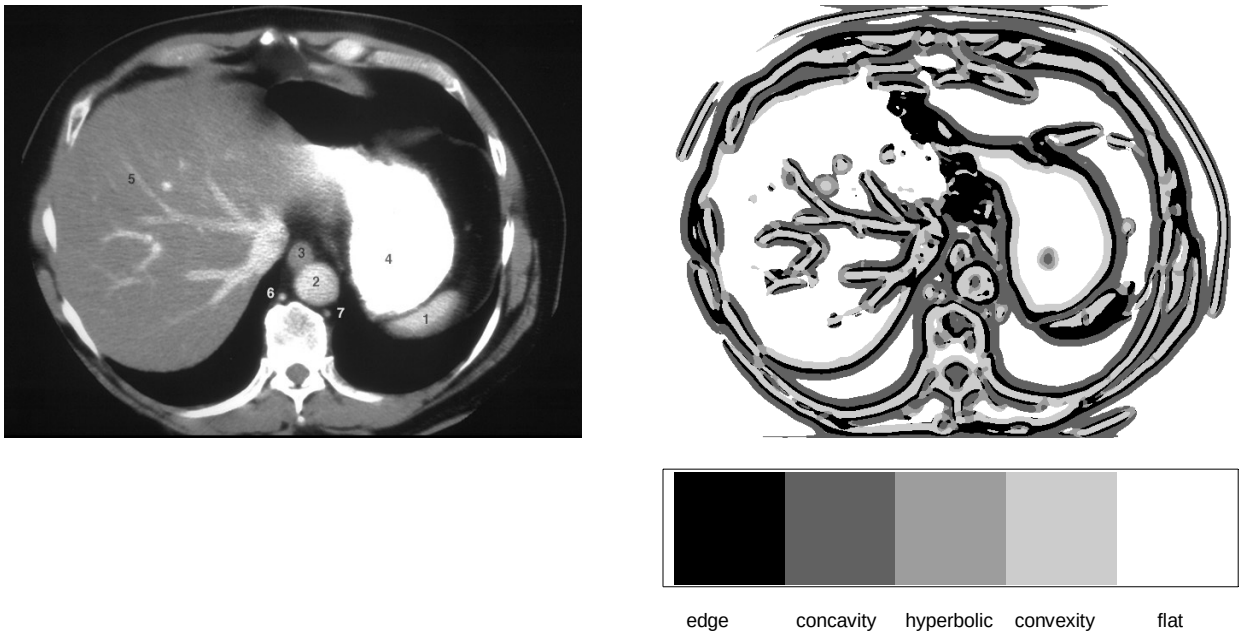

Figure 3 – An example of the dense feature labelling (right) that results when using a particular scheme of qualitative description based on the partitioning of a particular filter response space.

1.2. The Gaussian Derivative Local Jet

The hypotheses on qualitative description that we will present are not specific to a particular family of linear filters, but for concreteness we will use the derivative of Gaussian (DtG) filter family developed particularly by the Scale Space school [39-42]. Not only do DtGs of 1st to 4th order provide a well-fitting model of the receptive fields of simple cells in mammalian V1 [43-45], but they have many formal properties to recommend their use in computer vision [3, 21-23, 46, 47]. We will refer to the measurements given by a set of co-localised DtGs up to some order as a local jet, and the space of possible local jets as the jet space. Our uses of 'jet' are similar but

[2] That is to say that the regions of the partition have the same dimensionality as the space of which they are a portion.

not identical to its uses in the differential geometry literature [48] which is concerned with infinitesimal, rather than blurred, derivatives.

As we noted earlier, there are two distinct ways to interpret the application of DtG filters. First is that a local jet is interpretable as the initial terms of the Taylor series of the image blurred to the same degree as the scale of the DtG filters [2]. Second is that a local jet is interpretable as the initial terms of a Hermite Transform; the Hermite Transform being a local analogue of the Fourier Transform [30, 31, 49, 50], where the locality is specified using a Gaussian aperture (or weighting) function.

A 1-D Gaussian kernel can be written as $G_\sigma(x) = \left(2\pi\sigma^2\right)^{-\frac{1}{2}} e^{-\frac{x^2}{2\sigma^2}}$; and, because of separability, a 2-D kernel can be written as $G_\sigma(x,y) = G_\sigma(x)G_\sigma(y)$ [27]. In both cases the parameter σ specifies the width or scale of the kernel. DtGs are spatial derivatives of the 2-D Gaussian kernel. Differentiating with respect to different co-ordinate systems leads to different families of DtGs [22, 23]. For instance, using a Cartesian system gives the family (shown up to 4th order) in figure 4 (top middle). An alternative 'wavetrain' family made of filters, each of which is a derivatives in only a single direction, can also be formed (fig. 4 bottom middle). These two families are equivalent in the sense that any of the nth order filters can be created by affine combination of the corresponding nth order filters from the other family. Figure 4 also illustrates that DtG filters can alternatively be produced by multiplying an Gaussian aperture of scale $\sigma\sqrt{2}$ by Hermite-Weber functions i.e. $G_{\sigma\sqrt{2}}H_n = G_\sigma^{(n)}$. This is the interpretation one uses when treating application of DtGs as being computation of terms of a Hermite Transform: the aperture isolates a local patch of the image, which is then probed by the Hermite-Weber functions, just as sinusoids are used to probe the entire unwindowed domain when computing the Fourier Transform [50].

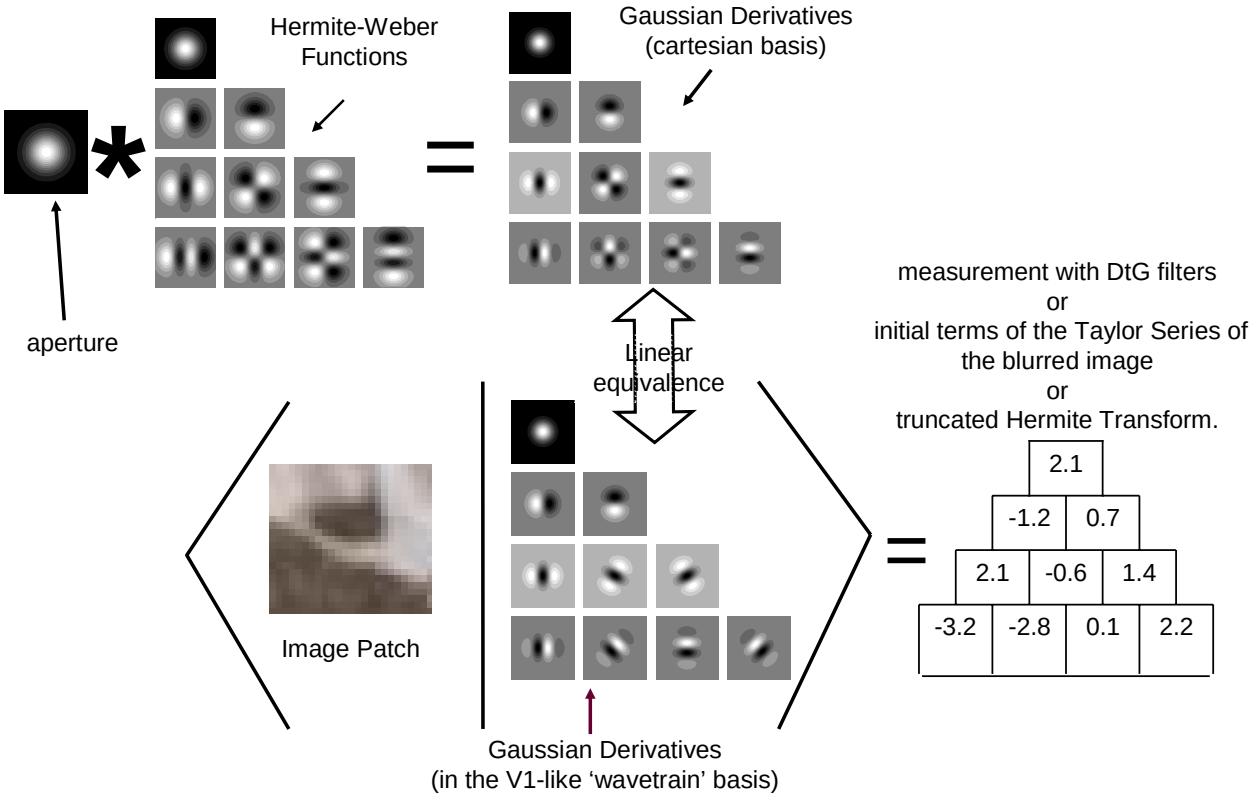

Figure 4 – Middle Column: two alternative families of Gaussian Derivative (DtG) filters. Bottom row: how measurement with a DtG family gives rise to the local jet. Top row: equivalence between measurement with DtGs and computation of the Hermite Transform.

1.3. The Hypotheses

We propose three hypotheses about qualitative description of local structure that we believe capture some of the recent literature in this area. In phrasing them we have decided on particular definitions of 'Feature', 'Icon' and 'Texton' that match reasonably those in general use, though inevitably there will be mis-matches with some authors. One of our intentions here is to establish some terminology that will assist future debate in this area.

Discussion of the three hypotheses will form the body of the paper, but we will preview them at this point. First is the *Feature Hypothesis*, which is that there exists a system of qualitative description of local structure that is (i) based on a partitioning of jet space, and (ii) is useful for subsequent stages of processing. Second is the *Icon Hypothesis*, which is that (i) for each point of jet space there exists some iconic image patch that measures to that jet, and (ii) the equivalence relation of qualitative similarity of icons is what determines the partitioning of the Feature Hypothesis. Third is the *Texton Hypothesis* which is that the icons of the Icon Hypothesis are the image patches that, of those that measure to the jet, are the most common in natural images.

2. Features

Application of DtG filters allows expression of local image structure as a point in jet space (figure 4, bottom right). Several authors [33-36, 51] have recently suggested that non-overlapping regions of jet space correspond to features (though the term texton is sometimes used). We make the idea explicit with the following statement:

> Feature Hypothesis: There exists a system of qualitative description of local structure, useful for later stages of processing, which is based on a partitioning of jet space. In this system, two image patches get the same qualitative description (i.e. feature label) if and only if their local jets are in the same region of jet space.

Typically the justification for the 'features = jet space regions' hypothesis is in terms of clustering of similar filter responses [33, 52]. Such a clustering approach is inviting if one looks at the jet space histogram of an extended texture, as one will often see a multi-modal structure. The multiple peaks of the histogram corresponding to elements of the texture that are repeated many times in similar form. However, if one looks at a broader range of images than a single texture then the multi-modal structure of the jet space histogram disappears. For example, figure 5 shows the jet space histogram of the 1-D 1^{st}-2^{nd} order jet of profiles taken from the van Hateren [53] database of natural images. There is absolutely no hint of clusters in this histogram and yet some degree of qualitative description should be possible even in this low-order, low-dimensional example. The key point that we wish to make is that the idea of features corresponding to regions of jet space is coherent without any consideration of the frequency of occurrence of particular image structures. The distinctive consequence of the feature hypothesis is that images analyzed in this framework will be densely labelled with feature types (figure 3) rather than having sparse 1- or 0-dimensional features (figure 2).

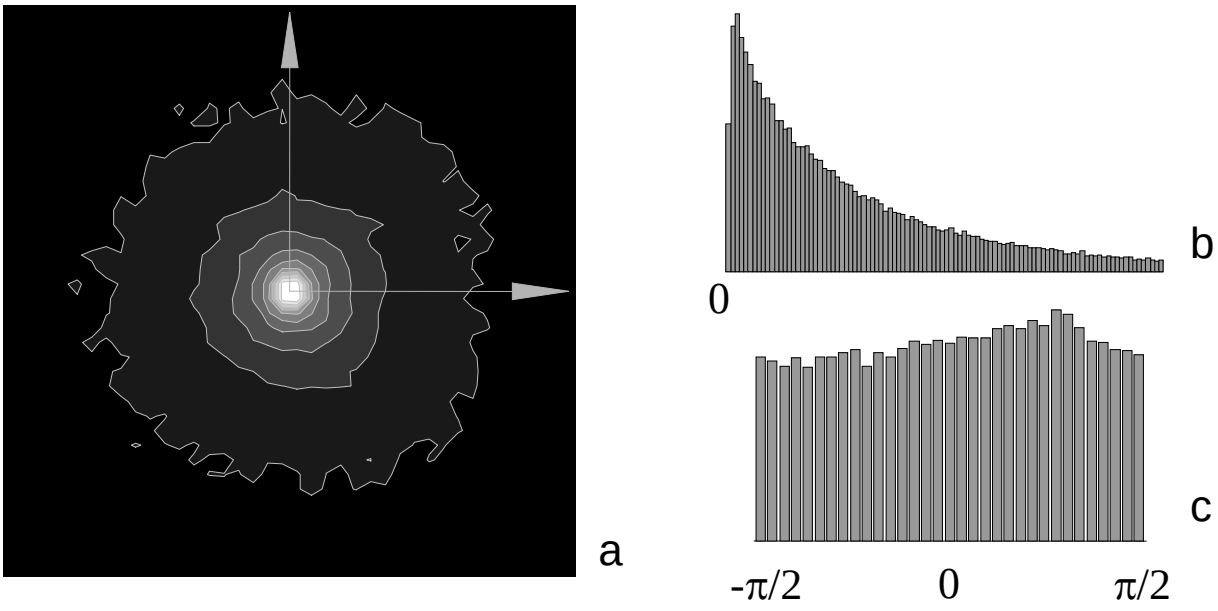

Figure 5. The density plot at left shows the jet space histogram for the 1-D, 2nd order jet. The horizontal axis is the 1st order filter response. The vertical axis is σ times the 2nd order filter response. The 0th order response is not plotted. The density has been square root transformed to improve visibility. On the right are plotted the polar marginal histograms; at top for the radial variable and at bottom for the phase. Note the absence of multiple clusters.

There is an interesting precedent for the partitioning of jet space into categories that may provide insights. The precedent occurs not in spatial vision but in colour vision, and has it force because of a similarity in formal structure between these two modalities. The similarity is between the measurements performed by the human visual system of the distribution of spectral energy across the visible spectrum, and the measurements performed on the distribution of luminance across a portion of the visual field [54]. Thus the analogy is between a 1-D and a 2-D system.

Colour vision in humans is subserved by three classes of light-sensitive retinal cells known as cones [55]. The three cone classes have different spectral sensitivities (figure 6, top-left). Ignoring various complications, one can assume that at each retinal location one of each class of cone is present, and so the distribution of spectral energy in the light incident on the location is transduced into three positive real numbers, where each number is the inner product of the sensitivity function and the spectral energy distribution of the incident light. Although this is similar to the way that linear filters probe the distribution of intensity at a location of an image, the analogy seems limited as the uni-modal cone sensitivity functions seem very different in form to the oscillating DtG functions. However, if one linearly combines the cone sensitivities one can produce results that looks very like 1-D 0th, 1st and 2nd order DtGs [54, 56]. Since the action of the cone functions is linear, linearly combining the sensitivities is equivalent to linearly combining the outputs of the raw cone functions. Hence, barring signal vs. noise issues, the linearly combined system has identical sensitivity to the original system. The similarity between the cone colour system and a 1-D 0th-2nd order DtG system is even greater if one is also allowed to warp the wavelength axis (figure 6, bottom-left). The warping that maximizes the similarity is

fairly mild, and since the visual system has little or no access to any metric on the wavelength axis one is not distorting the sensitivity of the cone system, this is simply a non-standard alternative picture of it. Thus the trichromatic colour system can be regarded as performing something equivalent to a 1-D, 2nd order Hermite Transform of the spectral energy distribution, just as a family of co-localized V1 simple cells can be regarded as performing a 2-D, 4th order Hermite Transform of the retinal luminance distribution.

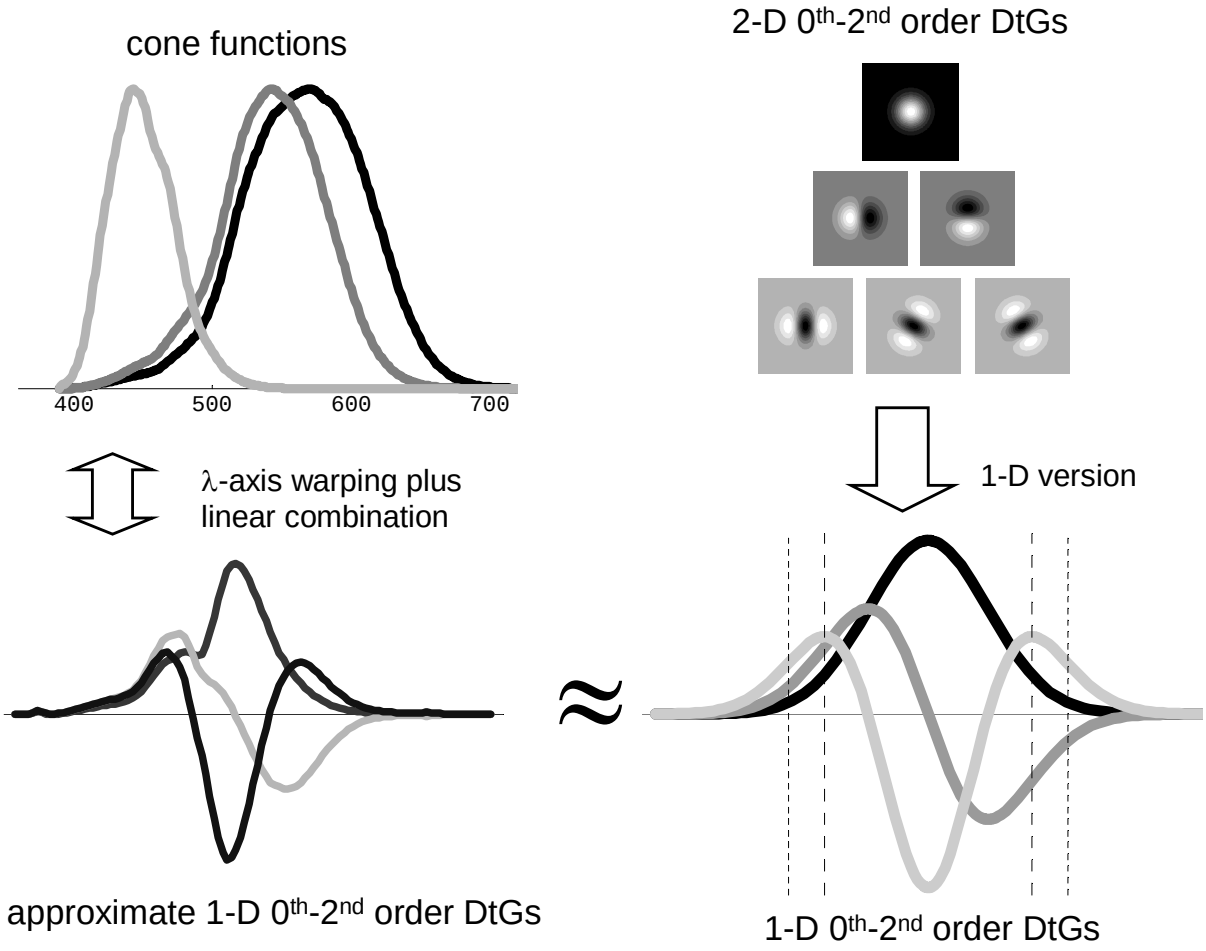

Figure 6 – Illustrates the analogy between spatial and colour vision via the Hermite Transform. The cone functions (top-left) can be transformed into a form (bottom-left) that approximates the 0^{th}-2^{nd} order 1-D DtGs (bottom right). This family of three DtGs are the 1-D equivalent of the six filters making up the 0^{th}-2^{nd} order 2-D DtG family (top-right).

It is not simply that there is an analogy between spatial and colour vision via the Hermite Transform that interests us here though. What interests us is that human colour vision, in addition to computing a quantitative description of the spectral energy distribution as described, also computes a qualitative description, and this description is based on a partitioning of the cone response space into non-overlapping regions of qualitatively similar colours. The words (black, grey, white, red, orange, yellow, green, blue, purple, pink & brown) referencing these colour

categories are known as the Basic Colour Terms [57]. Terms with very similar referents are found in all sufficiently developed languages, and as such they are a universal endpoint or at least bottleneck in the cultural evolution of colour language. Less developed languages typically have fewer than eleven terms, but it is claimed [58] (and disputed [59] and claimed again [60]) that these are always a subset of the full eleven. This trend towards universality has suggesting to some [43], though not all [44], that they have a pre-cognitive universal basis.

The fact of the basic colour terms sets a precedent for the useful partitioning of Hermite Transform spaces in visual processing and thus makes the Feature Hypothesis less out-of-the-blue. Also, we can consider the types of explanation that have been put forward for why the Basic Colours are the ones they are, rather than some others, as inspiration for the types of explanation that could underlie a system of image features. Note that because of the many differences between the domains of colour and spatial vision we do not expect anything stronger than suggestions for types of explanation. Explanations have been advanced for the Basic Colours in terms of the domains of neurophysiology, language, ecological optics and visual ecology. We review these below.

Neurophysiological explanations tie the Basic Colours to the form of the cone spectral-sensitivity functions [55, 61], to some later processing stage such as opponent channels [62-66] or to dedicated neural mechanisms [67]. Neurophysiology may also limit what categories are possible, for example we may lack cognitive structures capable of representing disconnected or non-convex regions of colour space [68].

Explanations based on shared language appeal to factors such as: limitations of some language-acquiring brain module [69], effects of the process of achieving consensus on semantics [67], and advantages of agreement despite inter-individual variations in colour vision [70].

Explanations from ecological optics [71] consider how common optical processes affect colour, and what invariants exist despite these processes. Examples are: colour categories being shaped so that neither shadowing nor highlights alter the hue of reflected light; and categories being shaped so that they achieve reasonable stability despite variations in illuminant [72].

In explanations from visual ecology, the colour statistics of the environment are taken into account. For example, categories could correspond to clusters of naturally-occurring colours [73]. Another possibility is that categories are particular effective for certain types of interaction with the world, such as search, identification, recognition, discrimination or classification [74]. Effectiveness for classification ties in well to a recent suggestion about psychological categories in general: good systems of categories are those that effectively support induction [75]. In the context of colour, the argument would go like this. A ‘green’ category is useful as it allows inferences like the following: the majority of ‘green’ things that I have seen have been plants, therefore this ‘green’ thing is probably a plant. If instead of a ‘green’ category one had ‘turquoise’ and ‘glaucous’ categories then this useful inference would no longer be so easily made.

Of these various types of explanation only the language-based ones cannot easily be translated into something plausible for image features. For example, an equivalent for features of

explanation of the basic colours from neurophysiology would be that image features are a 'natural' partitioning of the jet space derivable simply through consideration of the Hermite Transform. An example of an equivalent of explanation from ecological optics would be that features are such that they are stable with respect to changes in viewpoint or illumination geometry. While a feature-equivalent of explanation from visual ecology would be that features are such as to maximize our ability to recognize objects, people, or places on the basis of them.

To end this section we note that while it is tempting to coin the term 'Basic Image Features' we think this would be unhelpful given the continuing contention and hard-argued debates that surround the Basic Colours.

3. Icons

Our next hypothesis – the Icon Hypothesis – concerns what might underlie the categories of the Feature Hypothesis in the previous section. The hypothesis originates with Koenderink [76-78] but the following is our own statement of his suggestion:

> Icon Hypothesis: i) for each possible local jet (i.e. point of jet space) there exists some iconic image that has the specified local jet at the origin, and (ii) the equivalence relation of qualitative similarity of icons is what determines the partitioning of the Feature Hypothesis.

The didactic route to icons starts with consideration of metamerism, which is the phenomenon where sensory measurements fail to fully determine the proximal stimulus [79]. The most familiar examples are in colour vision, where the outputs of the three cone classes do not and cannot fully determine the form of the spectral energy density that they measure. This is because the spectral energy density has a huge (if not infinite) number of degrees of freedom, whereas the cone system measures only three. Thus there are an infinite number of spectral energy density functions that are compatible with any triple of cone responses. The same situation occurs in spatial vision, where a local jet fails to uniquely determine the form of the retinal intensity even within the locality that the local jet is the measure of [38]. Pairs of inputs, in colour or spatial vision, that measure to the same values are said to be metameric; and the set of possible inputs that measure to a given vector of values is the metamery class for that vector of values.

The fact of metamerism raises the question of how a visual system should treat a local jet. There are three possibilities [38]. (1) by ignoring metamerism i.e. by using the numbers that define a metamery class as a *symbol* standing for the class but never representing or reasoning about individual class members. (2) by using the class definition as a *code* that defines the class, allows generation of class members, and facilitates testing of membership, while remaining uncommitted as to which class element is the true stimulus (cf. 'multiple visual worlds' [80]). (3) By 'sticking its neck out' [80] and selecting a particular iconic representative of a metamery class and attaching the icon's qualities to the full metamery class. The Icon Hypothesis holds that for spatial vision, strategy (3) is appropriate. The question for spatial vision then becomes: how to select icons so that the qualities that thus accrue to the metamery classes are simple, conservative and representative? In considering icon selection, one can also keep an eye on the

second part of the Icon Hypothesis that states that the 'correct' icons will induce the partitioning of jet space hypothesized by the Feature Hypothesis.

In the papers where Koenderink proposed the Icon Hypothesis the particular icon selection rule that he considered was (more-or-less): that the icon of a metamery class should be the element with the smallest range of intensity values. It can be shown [81] that this rule leads to icons of a very distinct type: they have only two intensity values, and the transition locus between the two values always has the form of an n^{th} order algebraic curve, where n is the order of the local jet specifying the metamery class. What makes this icon selection rule especially attractive is that the second part of the Icon Hypothesis – that an equivalence relation of qualitative similarity of icons should determine a partitioning of jet space – is easily fulfilled. Consider the 2^{nd} order jet for example. The transition loci of the range-minimizing icons always have the form of a 2^{nd} order curve i.e. a conic, and conics naturally partition into classes: ellipse, hyperbolae plus degenerate cases. Similarly for the 3^{rd} order jet one obtains icons with cubic curve transition loci, for which one may use Newton's classification of the cubics [82] to define equivalence classes.

Range-minimization is not the only plausible rule for icon selection though, and others have been looked at [83]. In particular we have looked [11, 37, 38, 84] at other rules based on selecting the metamery class element that minimizes some measure of complexity. The definitions of 'complexity' that we have considered are norms. For an image I, we define its luminance norms as $L^r(I) = \min_{\mu_r \in \mathbb{R}} \left(\int_{\vec{x} \in \mathbb{R}^2} \left| I(\vec{x}) - \mu_r \right|^r \right)^{\frac{1}{r}}$, and its gradient norms as $D^r(I) = \left(\int_{\vec{x} \in \mathbb{R}^2} \left| \nabla I(\vec{x}) \right|^r \right)^{\frac{1}{r}}$. In both cases, infinity norms are defined by taking the limit $r \to \infty$. Using this notation, the norms we have looked at are: the range of an image (L^∞), the variance (L^2), the total variation (D^1) and the 'roughness' (D^2). These four norms were chosen as their minimizers show a degree of resemblance to structures found in natural images (or to put it the other way, patches in natural images tend to be low in these norms when compared to their metameres). This is not true for all norms. For example, the L^1 norm minimizers consist of very 'unnatural' small collections of weighted delta functions. We also note that in the context of colour vision, the smoothest reflectance functions (i.e. D^2 minimizing) that could account for each triple of cone responses have been determined numerically [85, 86].

Summary results of our investigations into the effectiveness of these norms for icon selection are shown in figures 7-9. Figure 7 shows the icons that would be selected for the 1-D, 1^{st} and 2^{nd} order jets. The conclusions of this study were that: (i) while low order L^∞ minimizers are quite representative of natural structure, at higher orders they are distinctly less so, (ii) the standard definition of D^1 in 2-D seems to be an incorrect generalization of the 1-D definition (also argued for elsewhere [87]), (iii) even so, D^1 minimization is the most successful at selecting icons that are natural looking.

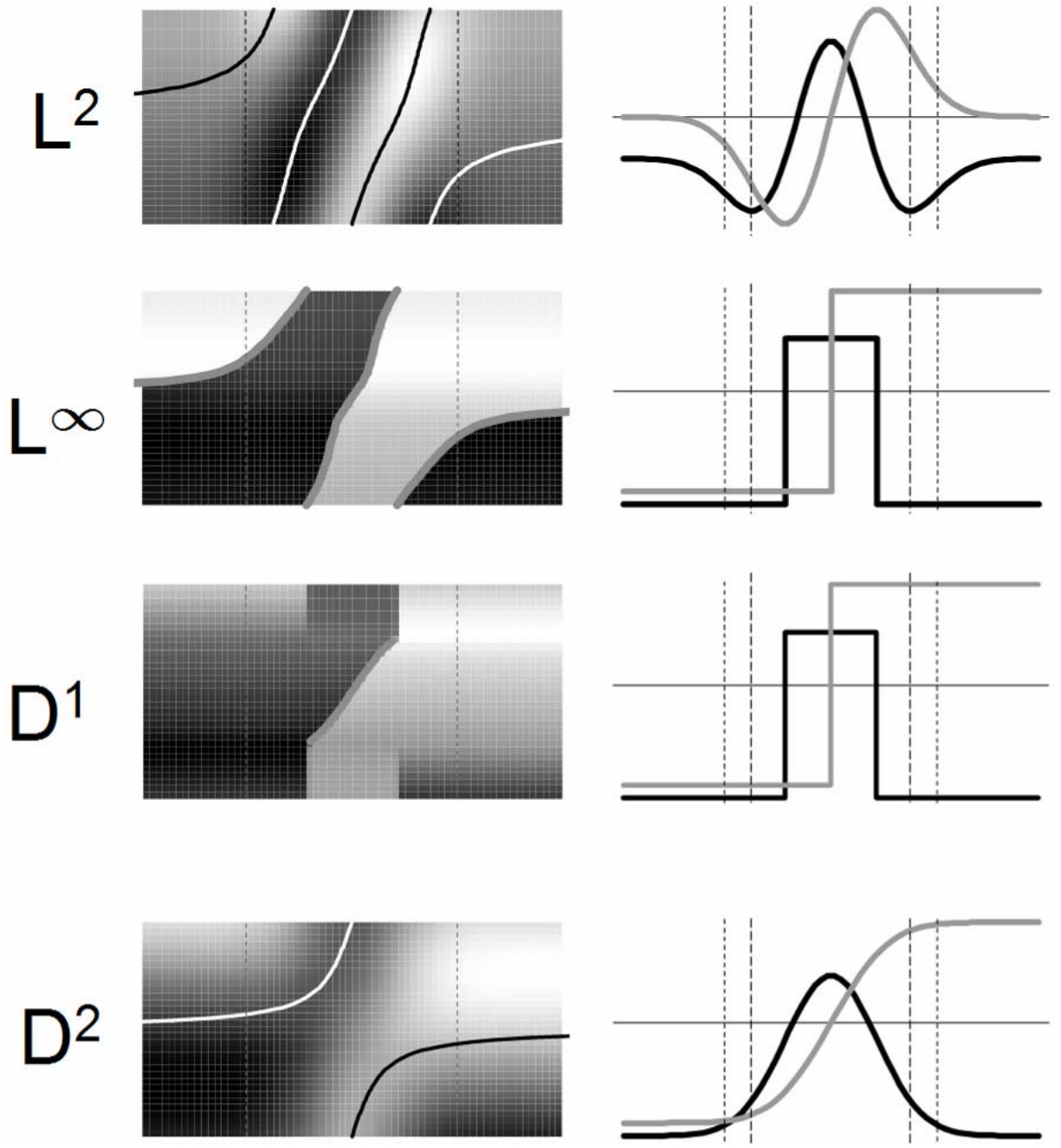

Figure 7 – Each row shows icons selected according to minimization with respect to a different norm (indicated in the leftmost column). The grey curves in the right column are the icons for the 1-D, 1st order jets. The rows of the density plots are the icons for the 1-D, 2nd order jets for different phase ratios between the 1st and 2nd DtG filter responses. The top and bottom rows of each density plot corresponds to large 2nd order DtG response, and small 1st order response (i.e. cosine phase); the central row to small 2nd order response and large first order (i.e. sine phase). To assist visualization, lines have been overlaid on these density plots to show the location of extrema and discontinuities. The right hand column of the figure shows the cosine-phase (black) and the sine-phase (gray) minimizers as regular plots.

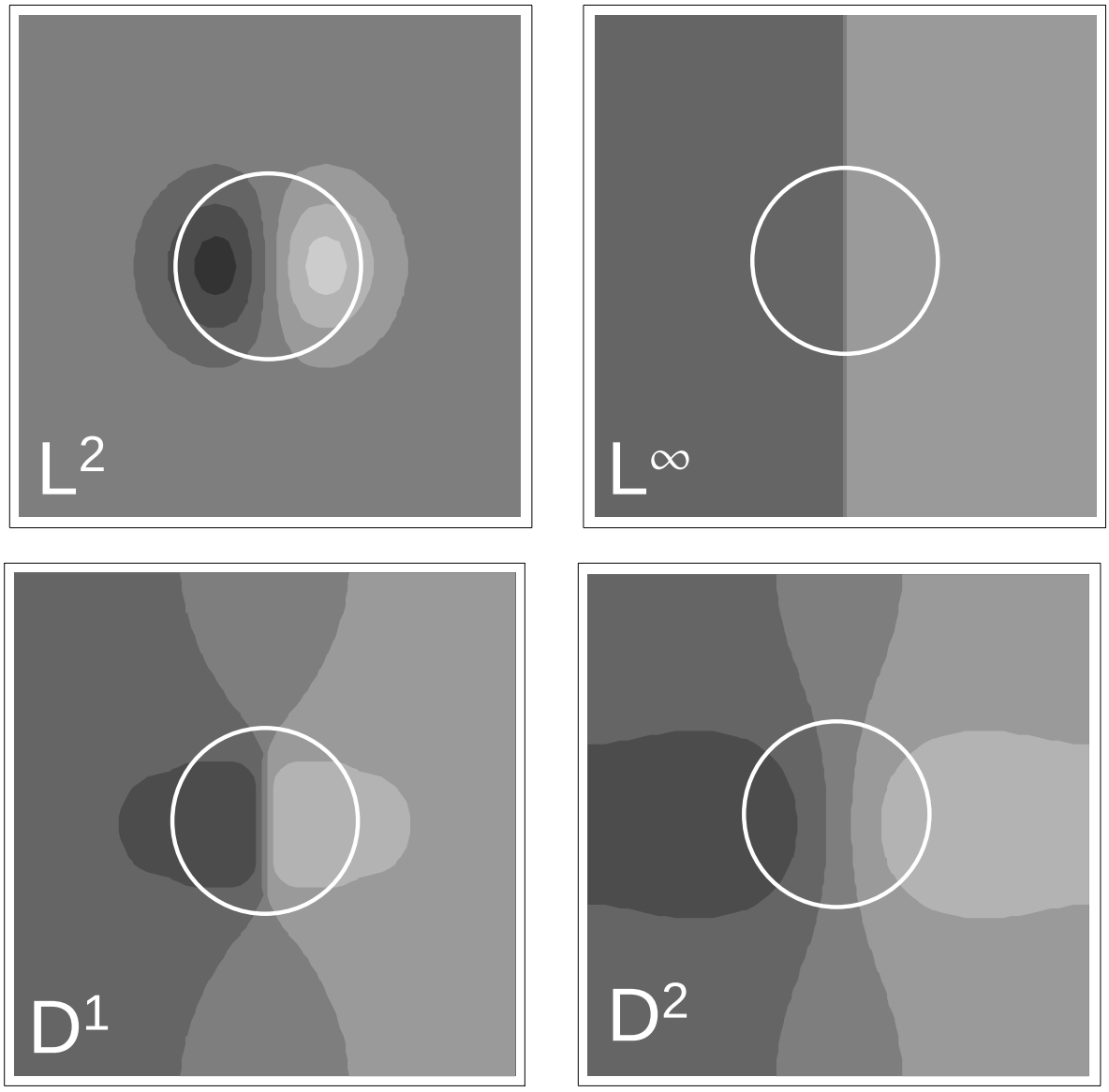

Figure 8 – Shows the icons for the 2-D 1st order jet selected according to minimization with respect to different norms as indicated. The overlaid circle roughly indicates the size of the locality where the local jet is measured.

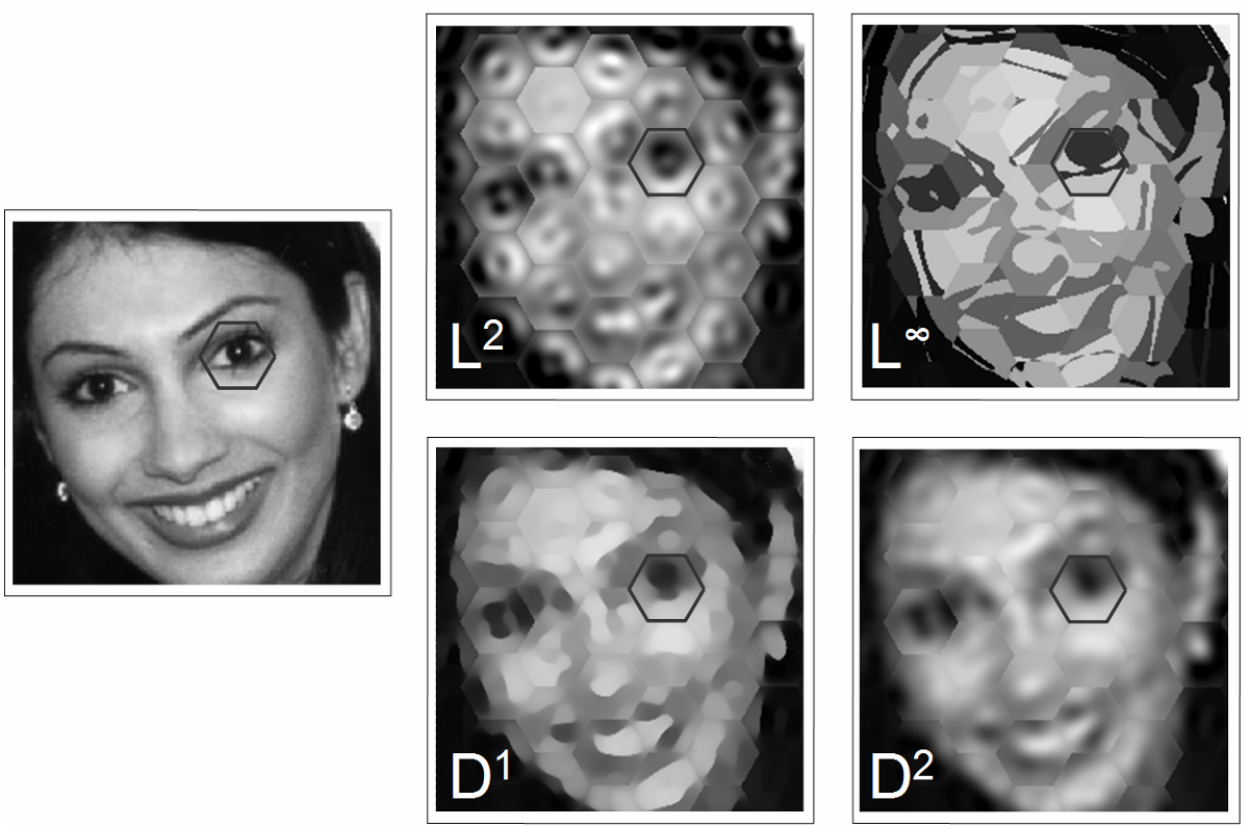

Figure 9 allows comparison of icon selection based on minimizing different norms. The original image is at left. To produce each of the other panels, a hexagonal array of sites was set-up and the 4^{th} order local jet was measured at each site at a scale indicated by the hexagons. Each hexagonal patch was then replaced by an icon with the same 4^{th} order structure. The icon selection rule is indicated by the label at the bottom-left of each panel.

In summary, we have found that suitable icon selection rules do lead to icons that are quite natural looking and so are reasonable representatives of their metamery classes. Also, icon selection rules based on minimizing complexity do lead to simple icons, from which natural partitionings of jet space are easy to define. However the partitioning that results does depend on the icon selection rule used.

4. Textons

To recap, the Feature Hypothesis posits that there is some (unknown) partition of the jet space that is effective for qualitative description and the Icon Hypothesis states that the partition is determined by iconic images, one associated with each point of jet space. The Texton Hypothesis is concerned with what determines those icons.

> Texton Hypothesis: the icons of the Icon Hypothesis are the elements of metamery classes that are the most common in natural images.

The Texton Hypothesis can be seen as a refinement of the icon selection rules discussed in the previous section, but with ‘most probable’ substituted for ‘most simple’ This is a familiar substitution given the linkages between probability and simplicity that have been established through principles such as Minimum Description Length [88]. The Texton Hypothesis also captures some of the flavour of approaches based on the modes of the histogram of local jets [89, 90] and also of approaches that attempt the construction of an optimum code book of image patches [91]. Certainly there is a crucial difference of detail as in these approaches the relevant probabilities are those of the appearance of local jets or patches, whereas in the Texton Hypothesis, the probabilities are those of image patches given the local jet.

We have pursued the consequences of the Texton Hypothesis by computing icons according to the definition. Initially these computations were neither easy nor quick since no good algorithm for high dimensional mode estimation existed. However we have now developed such an algorithm (earlier versions described in [11, 38, 92], final version in [93]) so the computations are now easy (but still slow). To use this mode estimation algorithm, 2-D patches or 1-D profiles are extracted from a suitably prepared large database [53] of natural images. These patches or profiles are then normalized by linear transformation of their intensities. The normalization is such as to factor out two components – an additive pedestal and a contrast magnitude – weakly justified by a hunch that qualitative structure should be invariant to linear transformation of image intensity. Patches, but not profiles are also rotated so that their gradient vector is in a canonical orientation. These steps remove two degrees of freedom from the 1-D jets of the profiles, and three from the 2-D jets of patches. We then select the subset of patches or profiles with a local jet sufficiently close to the local jet that we wish to determine the icon/texton for. These patches or profiles are then treated as points in a high-dimensional space and as such are input to the mode estimation algorithm.

Our first study was of the 1-D, 1^{st} order jet [38]. The 1-D, 1^{st} order jet space has two dimensions (the 0^{th} and 1^{st} order filter responses) so strictly speaking there is a 2-D family of metamery classes each with a separate icon to be discovered. However our normalization step collapses this 2-D family of metamery classes down to a single canonical class, and so there is only a single icon to discover. As shown in figure 10, we found that for natural images the maximum likelihood profile for the 1-D, 1^{st} order jet was a step edge. This contrasts with the maximum likelihood profiles for Gaussian and Brownian noise images which were found to be a Gaussian first derivative and an error function respectively (fig. 10). In the case of the noise images we were also able to prove that these were the correct forms. The finding that the step edge was the maximum likelihood profile was consistent with a hypothesis that we ventured at the time that maximum likelihood forms for natural images were either the L^{∞} or D^{1} minimizers from metamery classes (figure 7).

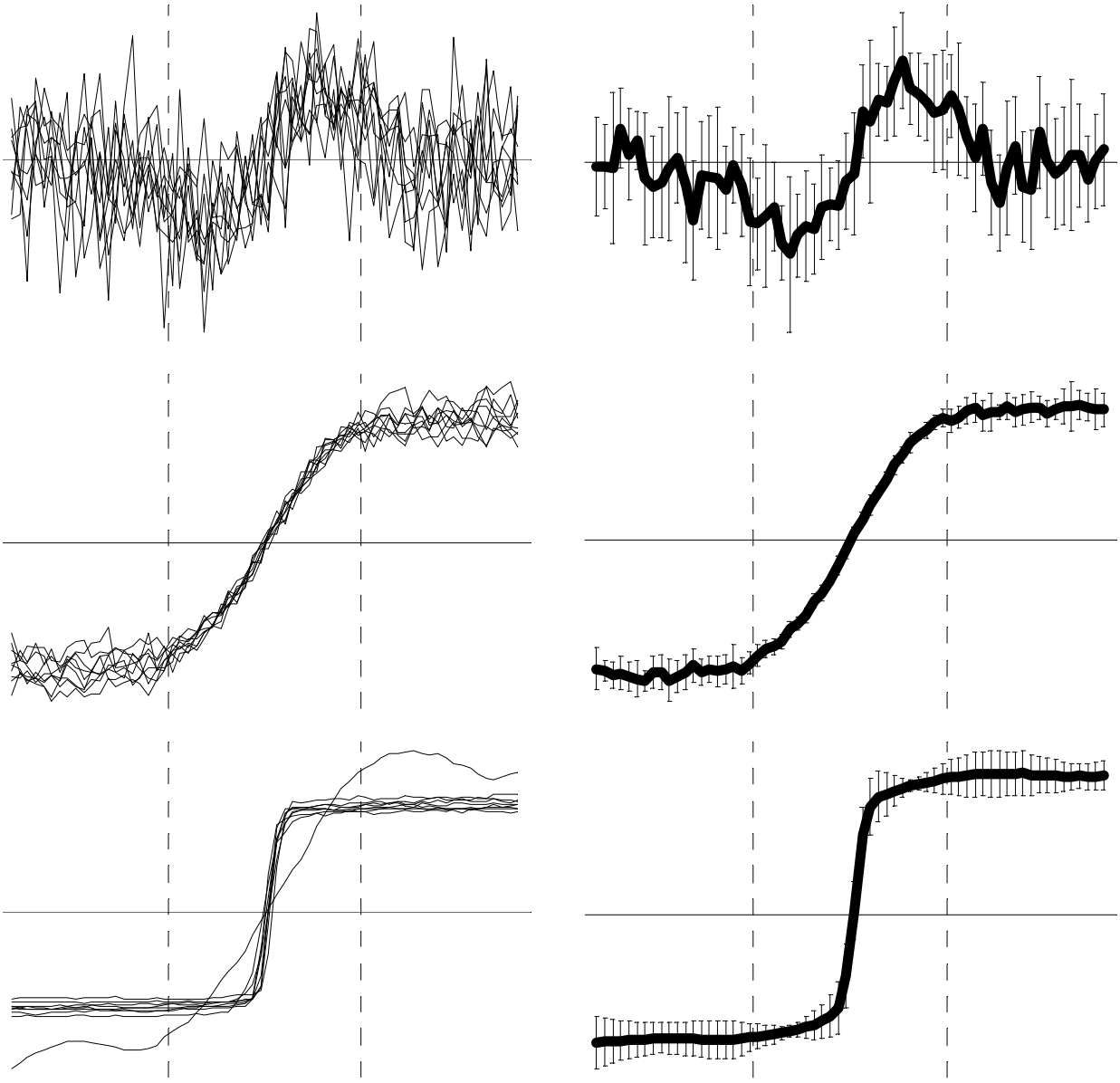

Figure 10 – Maximum likelihood (modal) profiles for different classes of image – Gaussian noise (top), Brownian noise (middle), and natural images (bottom). Profiles were normalized before mode estimation so that their 1st order jets were equal. The jets were measured in the centre of the profiles using DtG filters of a scale indicated by the dashed lines. Each plot in the left hand column shows nine separate mode estimates each based on 5×10^6 profiles. The plots at the right show the average mode estimate and one-sd-of-scatter error bars. Note that one of the nine maximum likelihood natural image profiles is of slope rather than step edge form. This suggests, not surprisingly, that these low order metamery classes are multi-modal.

Recently, we have managed to compute maximum likelihood patches (rather than profiles). Our first results from this are shown in fig 11 which shows the maximum likelihood patches from

natural images when they have been normalized so that they agree in their 1st order local jet. Not surprisingly, given the previous 1-D study, the maximum likelihood forms are straight step edges. Whereas the 1-D study results were consistent with the maximum likelihood forms being identical to the L^∞ or D^1 minimizing forms, the 2-D results are only consistent with the L^∞ minimizing form (see figure 8).

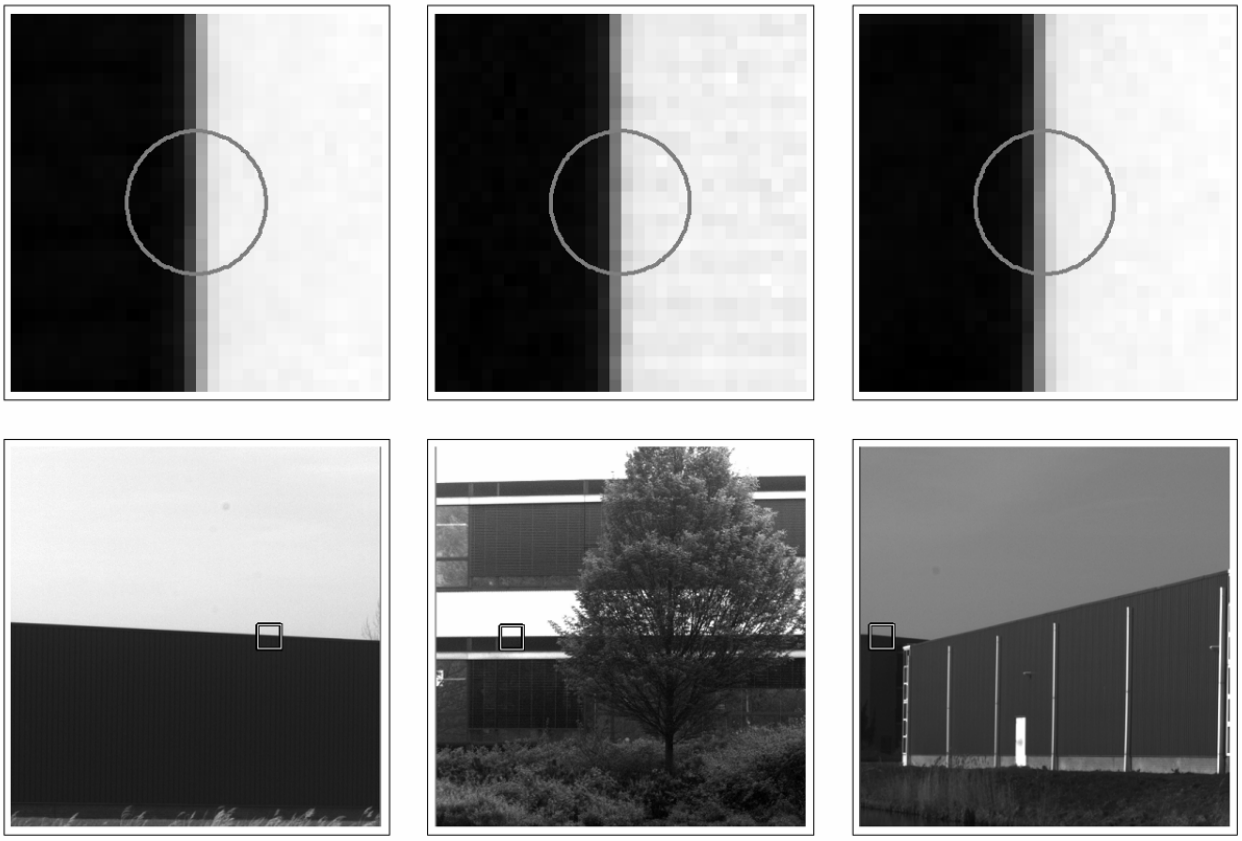

Figure 11 – Top row: Estimates of the maximum likelihood natural image patch for natural images. Each estimate was based on 350,000 patches. The bottom row shows the context of the patches that were chosen as the mode estimates.

The final mode estimation study we will report is of the 1-D, 2nd order jet [11, 84]. This jet has three degrees of freedom, two of which are absorbed by the normalization process leaving a single degree of freedom which we index using a phase variable running from $-\pi/2$ to $\pi/2$. The phase variable measures the ratio between the first and second derivatives in the jet. Jets with phases near the extremes of the range (i.e. cosine phase) are dominated by second order structure, those with phase near zero (i.e. sine phase) by first order structure. We divided the phase range into 33 bands and calculated separate mode estimates for each of these bands. The results are shown in figure 12 (top-left). As the figure shows, for sine-phase the maximum likelihood form is a step edge; this was a likely but not inevitable finding given the 1-D 1st order result. For cosine-phases the maximum likelihood forms are roughly a step-edged bar or pass. The maximum likelihood forms are similar, but definitely different from the L^∞ and D^1 norm

minimizing forms (figure 7), so the hypothesis that the maximum likelihood forms would coincide with the minimizers of one of the norms we had been considering was rejected at this stage. This does not prove that there is not some other norm, as yet unspecified, whose minimizers coincide with the maximum likelihood forms, but this seems unlikely.

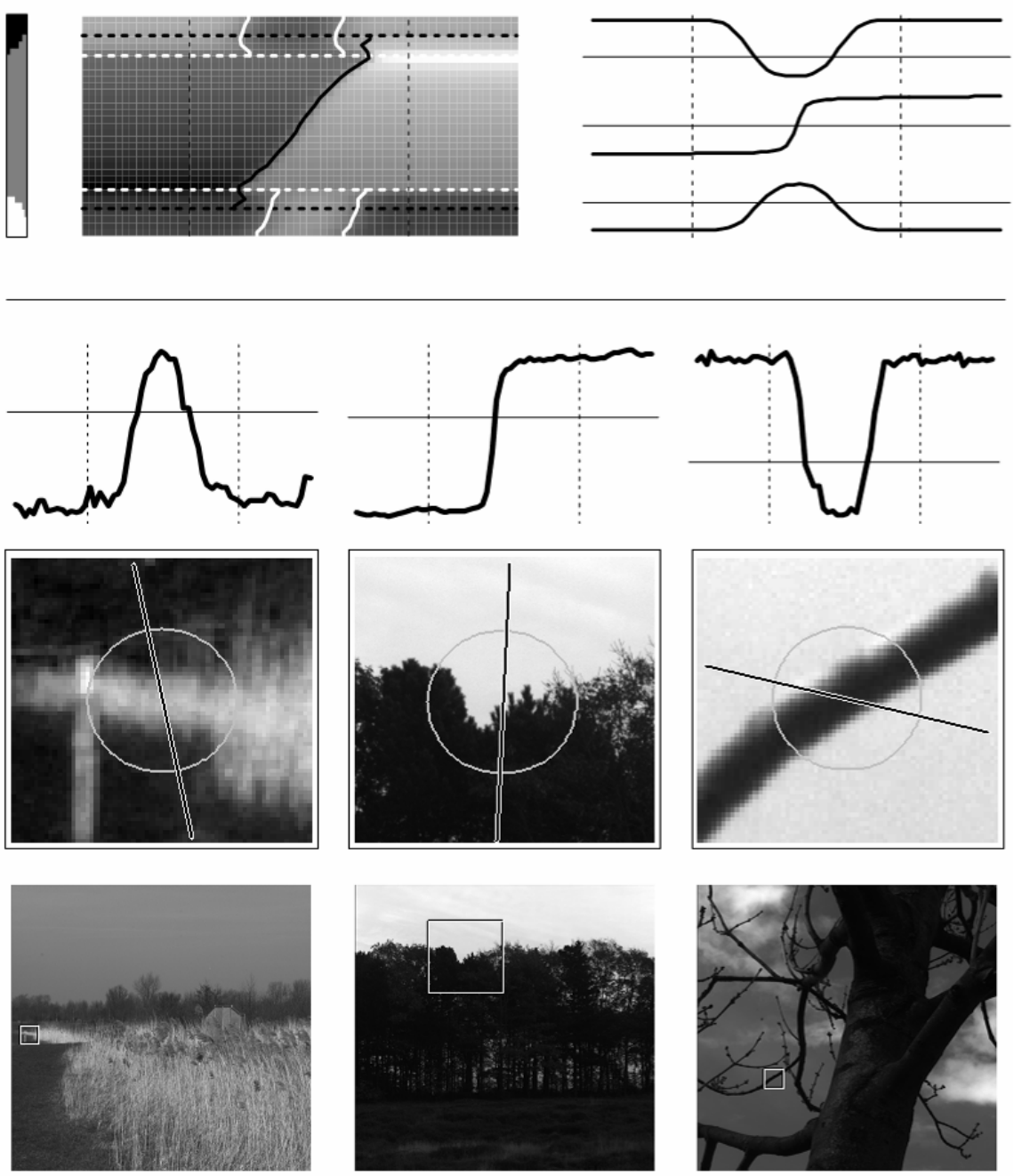

Figure 12 – Results from a study of the maximum likelihood profiles in natural images, when the profiles are conditioned on the 2nd order local jet. The modal profiles are shown in the top-left panel, where each row is the mode estimate for a different phase (i.e. ratio of 2nd order to 1st order derivative). The overlaid lines on the panel show a model that has an excellent fit to the data. The model consists of three template profile forms (shown at the right) that are shifted and scaled differently for each phase. For most phases the mode estimate corresponds to one of the template forms, but there are narrow bands where the modal profile is a mixture of two templates. The vertical bar to the left of the plot shows the amount of each template that is present at each phase. Examples of the three template forms occurring in natural images are shown in the lower 2/3rds of the figure.

We have found a simple model that describes well the maximum likelihood natural images profiles for the 2nd order jet and their variation with phase (i.e. the data of figure 12, top-left). The model consists of three template profiles (figure 12 top-right) – roughly an edge, a light bar and a dark bar – that by spatial shifting, intensity scaling and adding can be made to fit the ML profile at each phase. For phases near cosine-phase, one or other of the unmixed bar templates fits the data. For phases around sine-phase the unmixed edge template fits the data. For only six of our 33 phase bins is it necessary to use a bar-edge mixture. These six are arranged in two groups of three and separate the pure edges from the pure bars. The relative weighting of the bar and edges in these transition zones follow a monotonic pattern from pure edge to pure bar (figure 12 top, far-left). The plot of figure 12 shows the model profiles overlaid with lines that indicate how the bars and edges are shifted for different phases.

In figure 12 we also show actual natural image profiles that have our three template forms and how they arose in the images from which they were extracted. The example having the light bar form arose from a foreshortened view of a region of the ground plane containing lighter vegetation than the surround. The dark bar arose from the simpler situation of a branch silhouetted against the sky. The edge example is another silhouette, but this time of a tree canopy against the sky.

The model that we have fitted to the maximum likelihood profiles for the 2nd order jet is fully compatible with the Texton and Icon Hypotheses in that icons determined as maximum likelihood profiles induce a partition of the jet space. The partition is into three classes separated by two fuzzy intermediate bands. The Icon Hypothesis posits that a partition structure will be induced by qualitative similarity of icons. This is the case here, but the nature of the qualitative similarity is surprisingly simple: equality modulo shifting and affine scaling. This is even simpler than was the case for the norm-minimizing profiles we previously considered (figure 7). It appears that our use of the term Texton for our final hypothesis is particularly apt, since on the strength of these results natural image structure appears to be dominated by a small vocabulary of particular image structures.

4. Discussion

We will conclude with discussion of two issues. First, alternative positions that can be held if the hypotheses are not accepted. Second, the evidence presented for the hypotheses.

Considering alternative positions to those of the hypotheses, we consider the Feature Hypothesis first. The strongest rejection of the hypothesis is to call into question the necessity and utility of qualitative description. One could for example opt to keep local jets intact and uncategorized, and instead focus on developing a sophisticated metric for jet space so that distances in the space were a good reflection of the average dissimilarity between patches having those jets. A weaker rejection of the Feature Hypothesis is to accept the utility of qualitative description, but to prefer a fuzzy approach in which descriptions are vectors of fuzzy membership-of-feature-category scores. The weakest rejection of the Feature Hypothesis is to embrace unfuzzy categorical

description but to prefer a sparse-features approach where features only occur at isolated points and curves, with the remaining majority of the image being unlabelled [94]. The issue that we have with the metric- or fuzzy-type rejections of the Feature Hypothesis is that the reduction of information necessary to facilitate later processing without combinatorial explosion has simply been postponed. Against the view that feature labelling should include sparsely distributed curves and points as well as regions, we note that in our experience of practical image processing, the potential that curves and points have to 'lie between pixels' leads to much more complicated algorithms than is the case if only regions need to be dealt with.

Next, consider the Feature Hypothesis being accepted but the Icon Hypothesis rejected. Such a stance accepts that there is some partitioning of jet space to be found, but holds that it is determined in some way other than by icons. There are several possible plausible alternatives, as is clear from the analogy we drew with the basic colour categories of colour vision. Most of the types of explanation advanced for the basic colours can be translated into something plausible for feature categories of spatial vision, but the most readily convincing would be those concerning ecological optics (feature categories are such as to be as stable as possible with respect to changes in viewpoint or illumination geometry) and visual ecology (features are such as to maximize our ability to recognize objects, people, or places on the basis of them). We stress though that we regard the colour/spatial vision analogy as too weak to lead to any expectation that the explanations in the different domains should be the same. The single biggest difference (in the light of the analogy) between the domains is that the stimulus that the filters probe is more accessible in spatial vision than in colour – think for instance how informative are the changes to local jets caused by small translations of the retinal image, but for colour there is no mechanism that causes shifts along the wavelength axis.

Lastly, consider the Feature and Icon Hypotheses being accepted but the Texton Hypothesis rejected. This position would hold that there is some family of icons to be found that induce an effective partitioning of jet space, but that the icons were not the maximum likelihood members of metamery classes and were instead determined by some other rule. As we have discussed in section 3, norm-minimization is a plausible rule for icon selection but we have been unable to identify a norm which gives results which look to be worth further testing. Moreover one can question what rationale there could be behind a norm-minimization approach that was not phrased in terms of maximizing probability relative to some prior; and so given that, why not use maximum likelihood directly (which is equivalent to letting the database of natural images specify the prior). A radically different way that icons could be determined would be for them to arise from some specific strategy of cartoon-ification of the world. Perhaps some cartoon-ification scheme could be more conservative and so more robust than maximum likelihood [83].

Of the three hypotheses, we regard the Icon Hypothesis as the shakiest. Some of the ways in which a partitioning of jet space could arise without a basis in a system of icons are plausible.

Finally, we consider the evidence that has been advanced in favour of the combined Feature, Icon and Texton Hypotheses in this and our previous papers on the subject [11, 37, 38, 81, 84, 92]. We have determined the maximum likelihood icons for the 1^{st} order jet in 1-D and 2-D, and the 2^{nd} order jet in 1-D. The results for the 2-D 1^{st} order have not been presented previously, and

the results for 1-D 1st order are improvement of those previously published since they were computed using the latest version of our algorithm. In all cases the icons found were simple and were natural-seeming in the sense of corresponding to very familiar types of image structure with easily understood physical causes. Equally crucial is their determination of a partitioning of jet space. So far we have only presented results on this for 1-D jets, and results for 2-D jets will be necessary before we can assess whether the partitioning is 'useful' as required by the Feature Hypothesis. However, setting the need for further testing to one side, the 1-D, 2nd order jet space partitioning that we have determined is as good as could be expected. Firstly it was very readily determined by the icons, there was no need to argue for some subtle definition of qualitative identity of icons. Secondly, it agrees well with what one would guess as the answer. Thirdly, the surprising result that the icons were simply translates of some template forms, rather than the more complicated variations seen with norm-minimizers, supports the idea that the distribution of local structures actually seen in natural image structure is sparse, which in its turn supports the idea that underpins the Texton Hypothesis: qualitative description of local image structure derives from the inference of likely real-world appearances from low-dimensional quantitative measurements.

In conclusion, we have made careful statements of a sequence of three increasingly specific hypotheses concerned with how qualitative description of local image structure can best be performed. We have reviewed work that we have done on testing these hypotheses. We would characterize these results as supportive, rather than confirming, of the hypotheses.

References

1. ter Haar Romeny, B.M., *Front-end vision and multi-scale image analysis*. 2003: Kluwer.
2. Koenderink, J.J., *The Structure of Images.* Biological Cybernetics, 1984. **50**(5): p. 363-370.
3. Koenderink, J.J. and A.J. van Doorn, *Representation of Local Geometry in the Visual-System.* Biological Cybernetics, 1987. **55**(6): p. 367-375.
4. Hubel, D.H. and T.N. Wiesel, *Receptive fields and functional architecture of monkey striate cortex.* Journal of Physiology, 1968. **195**: p. 215-243.
5. Lawson, S. and J. Zhu, *Image compression using wavelets and JPEG2000: a tutorial.* Electronics & Communication Engineering Journal, 2002. **14**(3): p. 112-121.
6. Cen, F., et al., *Robust registration of 3-D ultrasound images based on gabor filter and mean-shift method*, in *Computer Vision and Mathematical Methods in Medical and Biomedical Image Analysis*. 2004. p. 304-316.
7. Zhilkin, P. and M.E. Alexander, *3D image registration using a fast noniterative algorithm.* Magnetic Resonance Imaging, 2000. **18**(9): p. 1143-1150.
8. Martin, D.R., C.C. Fowlkes, and J. Malik, *Learning to detect natural image boundaries using local brightness, color, and texture cues.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004. **26**(5): p. 530-549.
9. Heiler, M. and C. Schnorr, *Natural image statistics for natural image segmentation.* International Journal of Computer Vision, 2005. **63**(1): p. 5-19.
10. Thom, R., *Structural stability and morphogenesis*. 1972, Reading MA: W. A. Benjamin, Inc.

11. Griffin, L.D. and M. Lillholm, *Image features and the 1-D, 2nd order gaussian derivative jet*, in *Proc. Scale Space 2005*. 2005, Springer. p. 26-37.
12. Logothetis, N.K., J. Pauls, and T. Poggio, *Shape Representation in the Inferior Temporal Cortex of Monkeys.* Current Biology, 1995. **5**(5): p. 552-563.
13. Nakamura, K., et al., *Visual Response Properties of Single Neurons in the Temporal Pole of Behaving Monkeys.* Journal of Neurophysiology, 1994. **71**(3): p. 1206-1221.
14. Sigala, N. and N.K. Logothetis, *Visual categorization shapes feature selectivity in the primate temporal cortex.* Nature, 2002. **415**(6869): p. 318-320.
15. Vogels, R., et al., *Inferior temporal neurons show greater sensitivity to nonaccidental than to metric shape differences.* Journal of Cognitive Neuroscience, 2001. **13**(4): p. 444-453.
16. Wilson, M. and B.A. Debauche, *Inferotemporal Cortex and Categorical Perception of Visual- Stimuli by Monkeys.* Neuropsychologia, 1981. **19**(1): p. 29-41.
17. Barlow, H.B., *Summation and inhibition in the frog's retina.* Journal of Physiology (London), 1953. **119**: p. 69-88.
18. Barlow, H.B., *Single units and sensation: a neuron doctrine for perceptual psychology?* Perception, 1972. **1**: p. 371-394.
19. Marr, D., *Vision*. 1982, New York: W H Freeman & co.
20. Marr, D. and E. Hildreth, *Theory of edge detection.* Proceedings of the Royal Society Series B, 1980. **20**: p. 187-217.
21. Koenderink, J.J., *Operational Significance of Receptive-Field Assemblies.* Biological Cybernetics, 1988. **58**(3): p. 163-171.
22. Koenderink, J.J. and A.J. van Doorn, *Receptive-Field Families.* Biological Cybernetics, 1990. **63**(4): p. 291-297.
23. Koenderink, J.J. and A.J. van Doorn, *Generic Neighborhood Operators.* Ieee Transactions on Pattern Analysis and Machine Intelligence, 1992. **14**(6): p. 597-605.
24. ter Haar Romeny, B.M. and L.M.J. Florack, *Higher-order differential structure of images.* Image and Vision Computing, 1994. **12**(6): p. 317-325.
25. Griffin, L.D., *Critical Points in Affine Scale Space*, in *Gaussian Scale-Space Theory*, S. Sporring, et al., Editors. 1997. p. 165-180.
26. Griffin, L.D., *Descriptions of Image Structure*. 1995, London: PhD thesis, University of London.
27. Griffin, L.D. and A.C.F. Colchester, *Superficial and Deep-Structure in Linear Diffusion Scale-Space - Isophotes, Critical-Points and Separatrices.* Image and Vision Computing, 1995. **13**(7): p. 543-557.
28. Koenderink, J.J. and A.J. Van Doorn, *The structure of relief*, in *Advances in Imaging and Electron Physics, Vol 103*. 1998. p. 65-150.
29. Debnath, L., *On Hermite Transforms.* Mathematicki Vesnik, 1964. **1**(16): p. 285-292.
30. Makram-Ebeid, S. and B. Mory, *Scale-space image analysis based on hermite polynomials theory*, in *Proc. Conf. on Scale Space Methods in Computer Vision*, L.D. Griffin and M. Lillholm, Editors. 2003, Springer. p. 57-71.
31. Martens, J.B., *Local orientation analysis in images by means of the Hermite transform.* IEEE Transactions on Image Processing, 1997. **6**(8): p. 1103-1116.
32. Rivero-Moreno, C.J. and S. Bres, *Conditions of similarity between hermite and gabor filters as models of the human visual system*, in *Computer Analysis of Images and*

Patterns, N. Petkov and M.A. Westenberg, Editors. 2003, Springer-Verlag: Berlin. p. 762-769.
33. Leung, T. and J. Malik, *Representing and recognizing the visual appearance of materials using three-dimensional textons.* International Journal of Computer Vision, 2001. **43**(1): p. 29-44.
34. Liu, X.W. and D.L. Wang, *A spectral histogram model for texton modeling and texture discrimination.* Vision Research, 2002. **42**(23): p. 2617-2634.
35. Varma, M. and A. Zisserman. *Classifying images of materials: achieving viewpoint and illumination independence*. in *ECCV '02*. 2002. Copenhagen: Springer.
36. Zhu, S.-C., et al., *What are textons?* International Journal of Computer Vision, 2005. **62**(1): p. 121-143.
37. Griffin, L.D., *Local image structure, metamerism, norms, and natural image statistics.* Perception, 2002. **31**(3): p. 377-377.
38. Griffin, L.D., M. Lillholm, and M. Nielsen, *Natural image profiles are most likely to be step edges.* Vision Research, 2004. **44**(4): p. 407-421.
39. *Scale Space '99*. in *Scale Space '99*. 1999. Corfu, Greece: Springer.
40. *Scale Space '01*. in *Scale Space '01*. 2001. Vancouver, Canada: Springer.
41. *Scale Space '03*. in *Scale Space '03*. 2003. Isle of Skye, UK.: Springer.
42. *Scale Space '05*. in *Scale Space '05*. 2005. Hofgeismar, Germany: Springer.
43. Young, R.A., *The Gaussian derivative model for spatial vision: I. Retinal mechanisms.* Spatial Vision, 1987. **2**: p. 273-293.
44. Young, R.A., R.M. Lesperance, and W.W. Meyer, *The Gaussian Derivative model for spatial-temporal vision: I. Cortical model.* Spatial Vision, 2001. **14**(3-4): p. 261-319.
45. Young, R.A. and R.M. Lesperance, *The Gaussian Derivative model for spatial-temporal vision: II. Cortical data.* Spatial Vision, 2001. **14**(3-4): p. 321-389.
46. Georgeson, M.A. and T.C.A. Freeman, *Perceived location of bars and edges in one-dimensional images: Computational models and human vision.* Vision Research, 1997. **37**(1): p. 127-142.
47. Florack, L.M.J., et al., *Families of Tuned Scale-Space Kernels*, in *Computer Vision - ECCV '92*. 1992. p. 19-23.
48. Majthay, A., *Foundations of Catastrophe Theory*. 1985, London: Pitman Publishing Ltd.
49. Debnath, L., *Integral Transforms and their Applications*. 1995: CRC Press.
50. Koenderink, J.J. and A.J. van Doorn, *Local Image Operators and Iconic Structure*, in *Algebraic Frames for the Perception-Action Cycle*, G. Sommer and J.J. Koenderink, Editors. 1997, Springer. p. 66-93.
51. Manmatha, R., S. Ravela, and Y. Chitti, *On computing local and global similarity in images*, in *Human Vision and Electronic Imaging III*. 1998. p. 540-551.
52. Varma, M. and A. Zisserman, *A statistical approach to texture classification from single images.* International Journal of Computer Vision, 2005. **62**(1-2): p. 61-81.
53. van Hateren, J.H. and A. van der Schaaf, *Independent component filters of natural images compared with simple cells in primary visual cortex.* Proceedings of the Royal Society of London Series B-Biological Sciences, 1998. **265**(1394): p. 359-366.
54. Koenderink, J.J. and A.J. van Doorn, *Perspectives on color space*, in *Colour Perception: Mind and the Physical World*, R. Mausfield and D. Heyer, Editors. 2003, OUP: Oxford. p. 1-56.

55. Griffin, L.D., *Similarity of Pyschological and Physical Colour Space shown by Symmetry Analysis.* Color Research and Application, 2001. **26**(2): p. 151-157.
56. Geusebroek, J.M., et al., *Color constancy from physical principles.* Pattern Recognition Letters, 2003. **24**(11): p. 1653-1662.
57. Berlin, B. and P. Kay, *Basic Color Terms: their Universality and Evolution.* 1969, Berkeley: University of California Press.
58. Kay, P. and T. Regier, *Resolving the question of color naming universals.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(15): p. 9085-9089.
59. Davidoff, J., I. Davies, and D. Roberson, *Colour categories in a stone-age tribe.* Nature, 1999. **398**(6724): p. 203-204.
60. Kay, P., *Color categories are not arbitrary.* Cross-Cultural Research, 2005. **39**(1): p. 39-55.
61. Buchsbaum, G. and O. Bloch, *Color categories revealed by non-negative matrix factorization of Munsell color spectra.* Vision Research, 2002. **42**: p. 559-563.
62. Kay, P. and C.K. McDaniel, *The linguistic significance of the meanings of the basic color terms.* Language, 1978. **54**: p. 610-646.
63. DeValois, R.L., I. Abramov, and G.H. Jacobs, *Analysis of response patterns of LGN cells.* Journal of the Optical Society of America, 1966. **56**: p. 966-977.
64. Hering, E., *Outlines of a theory of the light sense*. 1920, Harvard: Harvard University Press.
65. Hurvich, L.M. and D. Jameson, *An opponent-process theory of color vision.* Psychological Review, 1957. **64**: p. 384-404.
66. Kay, P. and L. Maffi, *Color appearance and the emergence and evolution of basic color lexicons.* American Anthropologist, 1999. **101**: p. 743-760.
67. Steels, L. and T. Belpaeme, *Coordinating perceptually grounded categories through language. A case study for colour*. Behavioral and Brain Sciences, 2005. **In Press**.
68. Gärdenfors, P., *Conceptual Spaces: the geometry of thought*. 2000, Cambridge MA: MIT Press.
69. Dowman, M., *Modelling the acquisition of colour words*, in *Al 2002: Advances in Artificial Intelligence*. 2002. p. 259-271.
70. Jameson, K.A., *Culture and Cognition: what is universal about color experience?* Cognition and Culture, 2005. **in press**.
71. Gibson, J.J., *The Ecological Approach to Visual Perception*. 1979: Houghton Mifflin.
72. Bimler, D., *Personal Communication*. 2004.
73. Yendrikhovskij, S.N., *Computing color categories from statistics of natural images.* Journal of Imaging Science and Technology, 2001. **45**(5): p. 409-417.
74. Roberson, D., *Color categories are culturally diverse in cognition as well as in language.* Cross-Cultural Research, 2005. **39**(1): p. 56-71.
75. Ellison, T.M., *Induction and inherent similarity*, in *Similarity and Categorization*, U. Hahn and M. Ramscar, Editors. 2001, OUP: Oxford. p. 29-49.
76. Koenderink, J.J., *What is a feature?* Journal of Intelligent Systems, 1993. **3**(1): p. 49-82.
77. Koenderink, J.J. and A.J. van Doorn, *Receptive Field Assembly Specificity*. Journal of Visual Communication and Image Representation, 1992. **3**(1): p. 1-12.

78. Koenderink, J.J. and A.J. van Doorn, *Metamerism in complete sets of image operators*, in *Advances in Image Understanding: A Festschrift for Azriel Rosenfeld*, K.W. Bowyer and N. Ahuja, Editors. 1996, Wiley-IEEE Computer Society Press. p. 113-129.
79. Richards, W., *Quantifying Sensory Channels - Generalizing Colorimetry to Orientation and Texture, Touch, and Tones.* Sensory Processes, 1979. **3**(3): p. 207-229.
80. Koenderink, J.J., *Multiple visual worlds (editorial).* Perception, 2001. **30**: p. 1-7.
81. Tagliati, E. and L.D. Griffin, *Features in Scale Space: Progress on the 2D 2nd Order Jet*, in *LNCS*, M. Kerckhove, Editor. 2001, Springer. p. 51-62.
82. Newton, I., *Enumeratio linearum tertii ordinis*. 1706.
83. van den Boomgaard, R., *Least squares and robust estimation of local image structure*, in *Proc. Scale Space Methods in Computer Vision*, L.D. Griffin and M. Lillholm, Editors. 2003. p. 237-254.
84. Griffin, L.D., *Feature classes for 1-D, 2nd order image structure arise from the maximum likelihood statistics of natural images.* Network-Computation in Neural Systems, 2005. **in press**.
85. van Trigt, C., *Smoothest Reflectance Functions .2. Complete Results.* Journal of the Optical Society of America a-Optics Image Science and Vision, 1990. **7**(12): p. 2208-2222.
86. van Trigt, C., *Smoothest Reflectance Functions .1. Definition and Main Results.* Journal of the Optical Society of America a-Optics Image Science and Vision, 1990. **7**(10): p. 1891-1904.
87. Kimmel, R. and A.M. Bruckstein, *Regularized Laplacian Zero Crossings as Optimal Edge Integrators.* International Journal of Computer Vision, 2003. **53**(3): p. 225-243.
88. Rissanen, J., *Modeling by shortest data description.* Automatica, 1978. **14**: p. 465-471.
89. Pedersen, K.S., *Statistics of Natural Image Geometry*, in *Department of Computer Science*. 2003, University of Copenhagen: Copenhagen.
90. Lee, A.B., K.S. Pedersen, and D. Mumford, *The nonlinear statistics of high-contrast patches in natural images.* International Journal of Computer Vision, 2003. **54**(1-2): p. 83-103.
91. Wu, S.W. and A. Gersho, *Lapped Vector Quantization of Images.* Optical Engineering, 1993. **32**(7): p. 1489-1495.
92. Griffin, L.D. and M. Lillholm. *Mode Estimation by Pessimistic Scale Space Tracking*. in *Scale Space '03*. 2003. Isle of Skye, UK: Springer.
93. Griffin, L.D. and M. Lillholm, *The multiscale mean shift algorithm for mode estimation.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005. **submitted**.
94. Lillholm, M., M. Nielsen, and L.D. Griffin, *Feature-based image analysis.* International Journal of Computer Vision, 2003. **52**(2-3): p. 73-95.