visual communication

ARTICLE

A taxonomy of the image: on the classification of content for image retrieval

BRYAN BURFORD, PAM BRIGGS AND JOHN P. EAKINS
Institute for Image Data Research, University of Northumbria

ABSTRACT

Image database (IDB) systems are at present often designed to test technology and the efficacy of retrieval algorithms, rather than being oriented towards delivering functionality to users. Research is necessary to design interfaces geared towards human usage of images. The starting point of this research needs to be consideration at a fundamental, user-centred level of how people perceive and interpret images. This article considers literature from many disciplines to describe a taxonomy of image content, from direct sensory elements to high-level abstractions. The nine categories derived will later be validated and used to direct the design of visual query interfaces for IDB systems.

KEY WORDS

content classification • iconography • image data • image database (IDB) • visual perception • visual query formulation

INTRODUCTION

This article argues that previous classification systems are inadequate for describing qualitative types of content required for query expression by users of image databases (IDBs), and presents a taxonomy that will allow user descriptions to be mapped onto interface design solutions. The article begins with a description of some general issues in image data, followed by a summary of previous attempts to classify content. The taxonomy is then presented, with justification and support from a wide-ranging literature.

Images are powerful tools in any means of communication, formal or informal; however, until recently, systems for searching and retrieving image documents were limited to expert users. This situation is changing with the advent of the internet and new broadband forms of communication. IDBs

Vol 2(2): 123–161 [1470-3572(200306)2:2; 123–161;032591]

are found not only in centralized picture archives, but are also now accessible from the desktops of relatively novice users, who have no specific training in the use of particular coding schemes.

For these users, it is vital that some form of visual query interface exists which permits a meaningful search of an IDB without assuming any specialist knowledge. An effective implementation of such an interface has yet to be successfully developed. Current state-of-the-art developments in image search and retrieval systems are imaginative and technologically sophisticated but have not involved a user-centred approach – with the result that the systems do not map properly onto users' needs or abilities. The authors of this article are currently engaged in a project to establish a user-centred interface to an IDB. The starting point naturally lies with the end user: how do people construct and express visual queries? In order to answer these questions, we first need to know something about the nature and range of image content in order to progress to the next stage – an understanding of those aspects of image content required by different users.

This article provides a taxonomy of image content as extracted by the viewer of an image. This taxonomy has been developed by mining a very broad literature on the nature and content of images – with articles drawn from art history, perceptual and cognitive psychology, computer science and vision research. This broad base is important if the taxonomy is to be both comprehensive and concise, allowing any form of description given by a viewer to be effectively categorized without too much unwieldy detail. The intention is that a concise taxonomy can then be mapped onto interface solutions, thus allowing the user-centred development of a visual query interface.

The article is intended for an interdisciplinary readership and specialists within any of the areas may be surprised therefore at some exceptions and inclusions in the literature cited. The aim has been to use a diverse literature, which may not provide depth in any single area, but which can link disparate areas in order to provide an integral, multi-disciplinary account of image perception.

Multiplicity of content

An important first issue is that any one image has varied content, which may be available either consecutively or concurrently to the same or to different viewers. These multiple 'ways of seeing' (Berger, 1972) have been discussed by several writers over the years.

It is worth noting here the contrast with textual data. While textual data can have a multiplicity of content and meaning, in terms of the discrete elements of a query, the visual and linguistic content are homologous. The fundamental building blocks of text databases are ASCII character strings representing words that have a direct semantic interpretation.

In contrast, the pixel values making up digital images have no inherent significance. Considerable processing of the image is necessary even to infer the presence of a simple shape like a circle, let alone a complex object such as a tree. Direct comparison of image bitmaps can tell us only one thing about a given pair of images – whether they are identical or not. Nothing can be deduced about their similarity in terms of the objects they contain, or scenes they represent.

IDBs have been called 'perceptual databases' (Santini and Jain, 1996) in order to emphasize their distinction from traditional, symbolic databases. While text may have polymorphous interpretations when it comes to archiving and retrieval, text mappings are still more constrained and straightforward than for image data.

Firschein and Fischler (1971, 1972) argue that the multiplicity of image content is linked to the *purpose* of an image description. These purposes include: evocation of an image (painting a brief mental picture); classification of an image; retrieval of an image; full reconstruction of an image; and a description in response to specific queries *about* the image. The fact that description of an image varies according to the context in which that description is required is an important recurring theme in this article – a theme that is echoed by several writers in the art history field, including Malinas (1991), Gracia (1994), Heehs (1995) and Rollins (1999).

Malinas (1991) discussed the different vocabularies that may be used to describe the function of a picture: it can denote, refer to, depict, and allow the truth of statements to be ascertained. These vocabularies reflect the different functions described by Firschein and Fischler (1971, 1972). They also indicate different ways in which an image relates to the physical world.

Malinas (1991) distinguished between resemblance and representation, with each taking precedence in different contexts (see also Rollins, 1999). Using the terms of Ornager (1997), resemblance is basically straightforward symbolic content (what a picture is 'of'), while representation is related to expressive content (what a picture is 'about'), though it does contain some higher order symbolic content (this is more true of Rollins's interpretation). For example, a picture of a man will resemble the model who sat for it. The same picture may represent Jesus (symbolic content) and, by extension, represent a story from the Bible, the Christian Church, etc. For some pictures, neither resemblance nor representation are involved. For example, impossible objects are not depicted in a literal sense as they can have no existence outside the picture. Malinas (1991) also mentions the importance of titles in adding 'broad content' to a picture.

The co-existence of more than one type of content within an image is illustrated by the discussions of Gracia (1994) and Heehs (1995). Gracia discusses the role of an audience in the presentation of any text or image, arguing that the role of an audience must be fulfilled for any meaning to be interpreted. This can be fulfilled by the artist himself or herself, but is fundamentally different to that of creator. Acknowledged within this role, however, is the fact that different audiences may bring different things to the same image. This is a tenet of an area of critical theory known as 'reception theory' (e.g. Iser, 1978; Jauss, 1982). The argument is that the meaning of a text (encompassing images, films or even computer interfaces [Persson et al., 2000]) is constructed by the reader/viewer/user in 'negotiation' with the text.

This point is clarified in Heehs's (1995) discussion of narratives. Using the example of Poussin's painting *The Arcadian Shepherds* (see Figure 1), he describes a number of different (and mutually contradictory) narrative interpretations derived by different art historians. Panofsky (1970[1955]) had used the same painting as a point for discussion, concentrating merely on different interpretations of the engraved words at the centre of the painting: 'Et in Arcadia ego'; in this way, subjective inferences, perhaps defined by ideological or political position, radically changed the translation of the phrase, and thus the interpretation of the picture. See also Mannheim (e.g. 1997) for discussion of the multiplicity of meaning that images may contain.

The systematic application of multiple levels of content is demonstrated in Grund (1993), who describes the process of indexing within the ICONCLASS system – originally a multi-volume system for indexing images, now computerized. This required many different types of content to be considered, working through (mostly) hierarchical levels of indexing classes, from over-arching abstract subjects (such as Religion and Magic, Nature, Literature) through to specific objects or actions. These

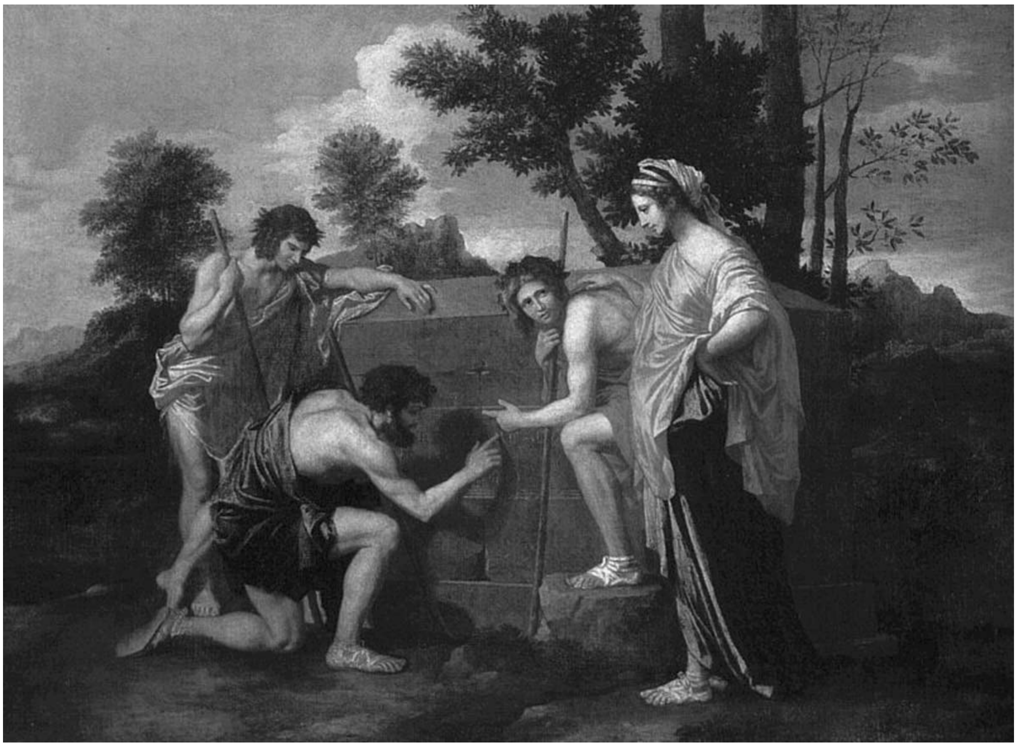

Figure 1 *The Arcadian Shepherds* (Nicolas Poussin 1656, Musée du Louvre, Paris).

classifications encompass both symbolic and expressive content and, while the system is hierarchical, the hierarchy does not follow these types (i.e. symbolic categories can contain expressive categories, and vice versa). The point is that any image can simultaneously be described in terms of a multiplicity of content – thus no user-centred visual query system should assume a strictly systematic or hierarchical determination of content types.

Previous classification schemes

A number of content classification schemes exist in the literature, although none is sufficiently comprehensive for purposes of the current work. Previous studies have identified broad, qualitative differences between types of content, dating from Panofsky (1970[1955]) to Enser (1995), Ornager (1997) and Eakins (1998). These schemes have been devised with a number of purposes in mind from the development of theory in art history, to a means of considering issues involved in the encoding and indexing of images within databases. The current work builds upon these, although our acknowledged purpose is to describe image content from the viewer's perspective.

Earlier approaches provide some useful basic distinctions between content types, with perhaps the simplest being made by Ornager (1997, after Garnier, 1984, and Shatford Layne, 1994): a binary distinction between symbolic and expressive content. This is best considered simply as a distinction between what an image is *of* (symbolic content) and what an image is *about* (expressive content). This crude distinction was illustrated in the previous section, and forms a useful starting point for determining categorization.

Different binary categorizations, not quite orthogonal to those of Ornager (1997), are presented by Enser (1995) and Rasmussen (1997). Enser distinguishes between visual and linguistic content, while Rasmussen distinguishes between content-based and concept-based indexing. Linguistic or conceptual content is content that can be expressed verbally; visual content, or simply content in Rasmussen's terms, is content that cannot be expressed verbally. Both of Ornager's categories are essentially linguistic. Enser does not go into any great detail about what may constitute instances of this content. His discussion is more concerned with identifying the issues involved in mapping visual codes to linguistic queries, and vice versa, rather than describing the details of what would comprise those codes and queries. Thus, although his work is relevant, it also fails to provide a firm basis for a user-centred taxonomy of image content.

Enser cites Panofsky's (1970[1955]) hugely influential distinction between three qualitatively different ways of reading content in art – pre-iconographic content, iconographic content and iconological content. To consider the example of Leonardo da Vinci's *Mona Lisa* (Figure 2), pre-iconographic content refers to the objects portrayed within the image in generic terms (e.g. woman, landscape). Iconographic content, on the other hand, refers to the specific instances of those objects, relying on specific knowledge of those objects (e.g. the model for the painting). Iconological content concerns the more abstract associations of an image, and not necessarily the actual content (for example, the political circumstances of the Renaissance).

Panofsky's work has been seminal in the art history field, and provides much of the inspiration for the current work. However, his approach is too entrenched in art history to be directly generalizable, and the categories do not provide sufficient discrimination for an applied taxonomy.

Eakins (1998) also drew on Panofsky to develop a three-level system generalizable to any image data. This system is the immediate precursor of the current work, in that each level represents a further degree of abstraction from the raw visual image.

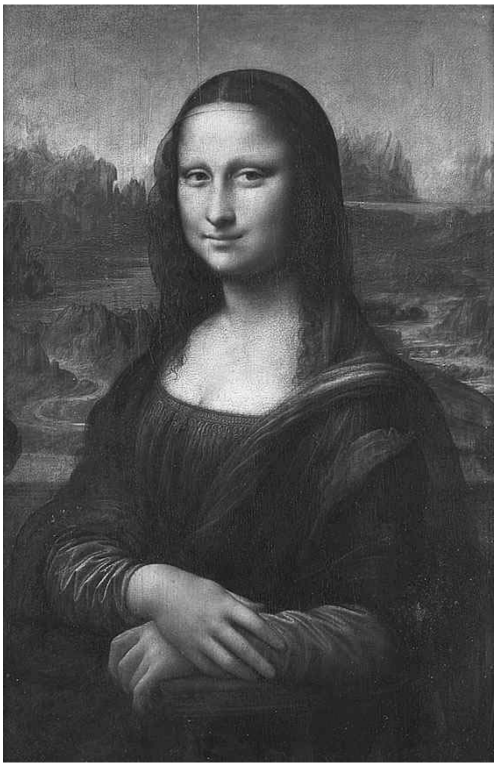

Figure 2 The *Mona Lisa* (Leonardo da Vinci 1506, Musée du Louvre, Paris).

Level 1 in this system refers to primitive features of an image: the shapes, colours and textures that comprise it at a purely visual level. Considering Figure 2, this may include horizontal lines, circles, rectangles and other polygons, and areas of light and dark. Level 2 refers to the logical or derived elements that those primitives make up – a woman, a chair, a landscape. This level contains both the generic labels (woman), and the specific labels of those things (*Mona Lisa*). Level 3 is the abstract or conceptual content that may be attached to an image, which may be associative (e.g. the Renaissance, patronage) or affective (calm, enigmatic).

Eakins, however, classifies queries in terms of the *technical* problems each level will raise. While suitable for consideration of an IDB per se, this system is too coarse for considering the breadth of queries people may present.

Jorgensen (1995) presents 12 classes of image attributes which provide a finer and more comprehensive system, but her classification does not appear to be *systematic* in relation to the user's extraction of meaning. Some levels appear to be overspecified ('literal object', 'location', 'people', all being distinct categories), while others appear to be underspecified ('abstract concepts', 'people-related attributes'). The taxonomy presented in the next section aims to be systematic and concise with respect to the user's relationship with visual information.

A proposed system of classification

Even a brief look at the literature as presented in the previous section supports the intuitive expectation that there are many different elements, at many different levels, which can be perceived in an image. These elements may be described in terms of simple retinal and cortical processes which detect basic features (such as edges and colour) through more complex processes (such as the extraction of shapes and boundaries from those features and the naming of those features) to high-level cognitive processes (such as naming the objects which those features represent – both as general classes and specific instances – placing those objects in a historical context, and associating them with abstract political and emotional terms).

The earlier analyses described previously go some way to classifying image content but are not sufficiently comprehensive for the purpose of developing interface solutions. While the classifications offered by Eakins (1998), Enser (1995), Panofsky (1970[1955]), Ornager (1997), etc. are indeed comprehensive and concise, they are not detailed enough to provide enough resolution or discrimination for the current purpose, i.e. the user-centred development of novel interfaces for IDBs.

It is the starting principle of the current research that in order to design an interface which will allow a user to express a query in a naturalistic way, it is necessary to understand the different ways in which people perceive images, and thus the different forms that query may take. The earlier approaches do not discriminate between content type with sufficient specificity to allow for the variations in user perception and description to be accounted for.

The reason for this is the essential difference between the current work and that which precedes it: the current work is fundamentally *user-centred*, meaning its consideration of images and image content aims to be from the point of view of the user of the image. Previous work has been fundamentally *system-centred*, concerned with technical implications of image content, for encoding, indexing and retrieval (this is equally true of the theoretical system of Panofksy, 1970[1955], as of the technologically oriented system of Eakins, 1998).

From such a viewpoint, the distinction between, say, the significance of a picture (for example, the *Mona Lisa*) as a historical document, and the emotive qualities of that picture, is slight. Both are abstractions for which there are few if any easily identifiable visual cues, and in terms of indexing and retrieval both are equally problematic (these would both be Level 3 in Eakins's 1998 classification). For a user, though, the importance and implications of those different types of image content may be very different indeed.

It is not a given that those subjective differences will lend themselves to different expression in terms of querying a database (that is for the research to uncover), but it is potentially the case. At a lower level, too, there are differences in user perception of an image which may not be significant to the system which is storing and retrieving the image (e.g. light and dark regions may represent a simple two-dimensional pattern, or light and shadow on a three-dimensional object – see Moore and Cavanagh, 1998, for a study relevant to this).

The remainder of this article considers the process of visual perception and cognition to describe nine categories of types of information that may be associated with an image. These include the distinctions between visual and linguistic, 'of' and 'about', description and association, which underpin the systems already described, but elaborate on them to identify other significant distinctions. The categories identified are summarized in Table 1. They draw on literature specialized to each level, encompassing some

Table 1 Classes of image content described in this article

Category	*Definition*
Perceptual primitives	The content extracted by low-level perceptual systems. In a strict sense this is unlikely (even impossible) to be reported. In practical terms, though, colour and some textural descriptions which do not rely on a higher level may be categorized here.
Geometric primitives	Simple two- and three-dimensional non-representational forms, such as line, arc, square, circle, etc.
Visual extension	Visual meaning which requires some inference. Most typical of these will be detection of depth, from shadow, occlusion, perspective, etc.
Semantic units	Names, both general and specific. Most descriptions will have some naming content, though it may be subsumed in higher levels.
Contextual abstraction	Associations or interpretations which depend on environmental knowledge. Such abstractions are presumed to be universal.
Cultural abstraction	Associations which rely on specific cultural knowledge. This may be the viewers' own culture (or subculture), or simply one of which they are aware.
Technical abstraction	Associations which rely on detailed specialist knowledge and vocabulary. Again this may be through direct experience of an area, or second-hand knowledge.
Emotional abstraction	Emotional and affective associations. These may be generalizable, but will be filtered by the viewers' own experiences.
Metadata	Information which describes the image, but is not actual image content, such as image format, size, aspect ratio, etc.

debate in the cognitive component of the early stages of meaning, but attempting overall to produce an internally coherent taxonomy.

The first eight of these categories may be seen as a rough hierarchy of levels, based on different degrees of abstraction and, as such, a development of Eakins's (1998) classification. These levels are roughly analogous to a sequential extraction of meaning, although such an interpretation may be overly simplistic in terms of actual cognitive processes. The taxonomy is intended in the first place as an analytical model, but a taxonomy that will enable the practical consideration of IDB queries.

The ninth category, metadata, has a different relationship to the image, which will be discussed in the metadata section.

TYPES OF CONTENT

Perceptual primitives

The lowest level of visual information is that extracted at the lowest level of perceptual processing. These qualities, termed here *perceptual primitives*, include: luminance level, luminance contrast, colour (hue, intensity and saturation) and contrast polarity. These are features that are extracted at a retinal or early cortical level. Much of the literature dealing with this level of feature is at the neurobiological level, dealing with how qualities and local features are extracted at a cellular level (see Bruce et al., 1996, ch. 3, for a general introduction).

Perceived luminance level is a direct consequence of the firing rate of the primary retinal cells, the rods and cones. These cells are the photoreceptors that respond directly to illumination. The different cell types respond at different thresholds, with rods firing in response to higher levels of illumination than cones. These cells also provide primary information about colour – different rods fire for different frequencies of illuminating light.

Local contrasts of luminance and colour are also extracted at a retinal level, by an intermediate level of cells that aggregate outputs from the photoreceptors and indicate the presence of edges, 'terminators' (contrasts that end in the field of activation of the aggregating cell), and 'blobs' (contrasts wholly contained within the cell's field). These cells may also indicate changes in illumination which may indicate an ambient change, or movement in the visual field.

Low-level elements are also implicated in more abstract levels. Heitger et al. (1992) discussed the role of terminator, or 'end-stop' detectors, arguing that they do not function as detectors per se, but as occlusion indicators. If this is true, this implicitly places their role more directly in higher levels of vision, or conversely implies that the higher levels of form extraction facilitated by occlusion are initiated at a much lower level. Malinas (1991), in his discussion of image content, depiction and representation, acknowledges that the complex objects which comprise the image are composed of the interactions of low-level attributes, with the fundamental primitive unit of all visual cognition being a point of colour.

Retinal signals are passed up the optic nerve to the visual or striate cortex, where they are aggregated further and instances of form and motion are extracted. It is perhaps important to note here that despite the easy assumption, much of what occurs in the *visual* system is not associated with *sight*. Goodale and Humphrey (1998) point out that the pupillary reflex, circadian synchronization to the local light/dark cycle and the visual control of posture are all examples of visual processes that have no association with our subjective experience of sight. Of course, when discussing images, we are wholly concerned with sight but its location in a broader system should not be ignored.

On a related point, Heller (1989) discusses the picture perception of blind people feeling raised outlines, compared to the perception of sighted people performing the same task, finding some similarities between the two groups. This raises the interesting point that while vision is not wholly about sight, perception of a 'visual' scene is not necessarily wholly about vision. The nature of vision is a long-standing philosophical issue, which as Goodale and Humphrey (1998) point out, dates back to Aristotle. However, this low-level epistemology is again not directly relevant to the current work.

At a more abstract level of theorizing, consideration of the elements extracted by the low-level retinal and cortical processes has been dominated, even after 20 years, by the work of Marr. Marr's *Vision* (1982) is still regarded by many as the definitive approach to visual perception. The computational approach he presented describes how low-level sensory elements are accumulated into visual units and thus form a sensible visual scene. Marr's theory however is not necessarily falsifiable, and so should not be over-relied upon. The component processes he describes though are largely observable, and it is the *integration* that he provides that is his main legacy.

Figure 3 provides an approximation of an image described at this level, showing boundaries detected by a low-level algorithm. An approximation is all that is possible as the level described is by definition pre-attentional and pre-cognitive. As such, it is unlikely to be useful for application to query formation or expression, but it is important as the informational underlay of meaning.

This lowest level of perception was termed by Marr the 'raw primal sketch', and is basically composed of the luminance contrasts detected by the intermediate retinal cells. In real terms, it seems that the lowest order of retinal extraction does in fact contain more information than that (i.e. it includes information on colour and motion). However, the pre-cognitive nature of the processes means that they cannot be *directly* reported. Perceptual primitives, however, do form an important class of content which may be indirectly reported, mediated by higher levels – illustrating how the overall process of visual perception and cognition is not simply one-way, bottom-up. That said, in strict terms of abstraction from raw sensory input, the next level, where form begins to be extracted, is the first truly cognitive level.

Geometric primitives

The perceptual primitives extracted from the visual scene by the various retina cells combine to form what is effectively the lowest *meaningful* structural level of visual content, i.e. shape and form, or geometric primitives. Discussion of these features in the literature is again sometimes at the neural level, as features are detected by complex and specialized structures in the visual cortex (often involving interaction between cell clusters on different layers and in different regions of the cortex). Grossberg et al. (1997), for example, discuss how different levels of cells feed back into each other to interpolate scales between the gross resolutions of photoreceptors and of higher level cells. Generally, though, consideration of shape occurs at a higher, more generalizable level.

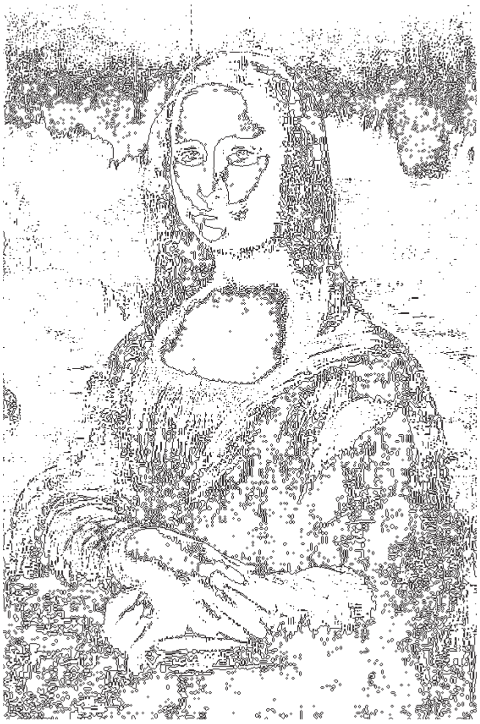

Figure 3 An approximation of the perceptual primitives of an image.

Dyson and Box (1997) describe an approach that takes the idea of geometric shapes as reductive components of an image, and apply it as a way of describing and classifying an image for retrieval. They distinguish between ‘mainshape’ and ‘border’ (amongst other things). Mainshape is the dominant shape in an image, while border is the shape that contains all components in an image. ‘Othershapes’ may also be included in an image. This approach, mainly suited for simple, high-contrast images, is similar to that described algorithmically by Burt and Adelson (1983), and has been applied to trademark images by researchers at the Institute for Image Data Research, in the Artisan system (e.g. Eakins et al., 2001). This principle of ‘scaling resolution’ allows individual boundaries within an image to be extracted at one resolution, and the bounding/border shape to be extracted at another.

Figure 4 shows dominant shapes identifiable in the *Mona Lisa*. This level, of lines and shapes, coupled with texture retained from the lower level is roughly what comprises Marr’s (1982) ‘full primal sketch’. The features of this level contain no additional information to that present in the image: they do not involve *meaning*.

Geometric primitives, however, do not necessarily have to be two-dimensional, nor simple. One group of theories sees visual perception as stemming from the construction of objects from generalized primitives in three-dimensional space. Perhaps the most well known of these is

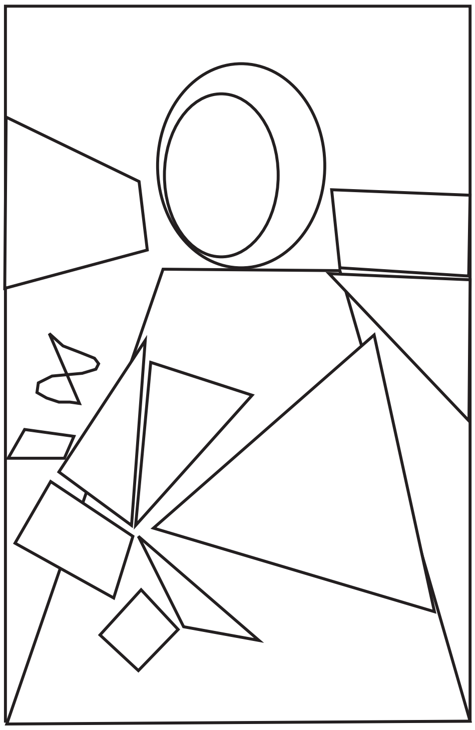

Figure 4 Geometric primitives in the *Mona Lisa*.

Biederman's (1987) 'recognition by components', although it drew influence from Marr and Nishihara's (1978) 'generalized cones', among others.

Recognition by components considers objects as composed of primitives, derived from generalized cones, called geons. Any object, according to this theory, can be described by a set of geons, their sizes, and the relationships between them. The geons are defined by differences in fundamental properties: straightness, symmetry and parallelism. Geons, while existing in three dimensions, can be described in two dimensions, and can be two-dimensional themselves, though this may raise questions of the extent to which the affine projection of volumes reflects the volume itself.

Recognition by components also provides support for elements of Gestalt theory. Biederman (1987) demonstrates that 'non-recoverable representations', which contain occluded (or omitted) elements which cannot be inferred using Gestalt rules such as good continuation (usually vertices), render objects much harder to recognize than presentations which occlude or remove elements that can be so inferred. Although questioned more recently by Koch and Abbey (1999), this has implications for an understanding of viewpoint–invariant object recognition.

Discussion of viewpoint–invariant recognition requires some conception of what a viewpoint varies *from*. The idea of a 'canonical view' (Neisser, 1967) describes a viewpoint from which essential elements of an object are visible, and extraneous elements not visible – for example, the canonical view of a cat would be side on, with four legs, tail and shape of head clearly visible. A canonical view need not be an actual view – it may be more stylized than anything that would actually be seen from any natural viewpoint.

Biederman and Gerhardstein (1993) addressed the question of viewpoint invariance, demonstrating that the structural descriptions allowed by geons are robust enough to be interpreted from novel viewpoints – the implication from the earlier finding is that this will be true as long as vital vertices are not occluded (for example, an object viewed obliquely is easily recognized; while viewed from directly above, it is not, see Biederman, 1987: 144). Biederman and Gerhardstein (1993) argue that a geon structural description should be unique for any given object, and that any viewpoint of the same object should have the same description.

Viewpoint independence based on structural descriptions requires some rotational mapping, adding to the cognitive overhead associated with these approaches. Takano (1989), however, distinguished between orientation-free and orientation-bound features (e.g. the relative length of components and the angles between them, and the cross-sectional and longitudinal-section shape of the components, respectively). Liter (1998) compared the effects of these two types of feature on object recognition. He found, similar to Takano's argument, that multiple-viewpoint recognition is much better when attention is focused on orientation-free features that do not require such mapping. However, the nature of viewpoint dependence is still a moot question (Tarr and Bölthoff, 1998).

Part-based approaches are highly indirect, in that the composite representations are highly complex, and may be contrasted with the highly direct perception approach of Gibson (see the 'Visual extension' section of this article). These approaches are sometimes called structural description theories, in contrast with view-based theories that treat the whole visual field as the raw material for cognition (see Tarr and Bölthoff, 1998, and Hummel, 2000, for discussions of the relative merits of structural and view-based approaches).

One question about these approaches is that they do not always demonstrate how an image is segmented into these component primitives in the first place. Some approaches have considered local 'energy minima' (first or second differential of local curves), but these are not always reliable. Kimia et al. (1992, described in Siddiqi et al., 1996, amongst others) considered the segmentation problem and proposed the 'shape triangle', which relied on a number of elements to explain how the same pattern of minima can be described as parts, bends or protrusions (Figure 5). These are distinguished

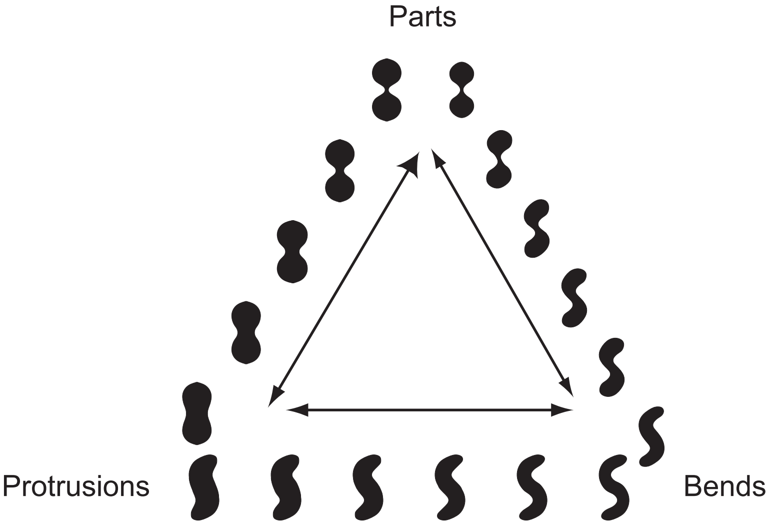

Figure 5 The 'shape triangle' (from Kimia et al., 1992)

in terms of their ‘shock-based descriptions’. This description, in terms of different types of shock, or curve singularity is presented as a more robust and informative method of description than other part-based methods (Siddiqi et al., 2001).

In whatever way the primitives at this level are derived – as formal geometric shapes or arbitrary boundaries, two- or three-dimensional – their common defining feature is that they contain no *meaning* beyond that shape. Meaning beyond the purely visual is introduced in the next category.

Visual extension

There are purely visual features that do contain meaning beyond the simple perceptual pattern, but not *semantic* content. They are distinct from the previous categories that represent nothing more than their visual pattern, but from the following categories in that all meaning is *available from* the visual stimuli. These paradoxical elements contain information on the immediate environment, such as depth, orientation and motion (actual or apparent). It is likely that these features are extracted by further complex cellular interactions in the visual cortex, but the majority of the literature in this area has been at a qualitatively different level, concerned much more with the cognitive and behavioural implications of this level of perception, rather than the processes underlying it.

Features at this level include indications of: depth (which may be due to a gradient in the resolution of texture, parallax or stereopsis, or occlusion); apparent motion (in a still image indicated by blur or the physical attitude of people or objects); and shape, again indicated by occlusion, shadow or surface contours. Response to these features develops in infancy (see, e.g., Slater, 1998), but whether this is due to the maturing of innate systems, or to learning, is a matter of debate.

These features (and others like them) have been studied in depth individually over the decades. They form what Marr (1982) described as the $2^1/_2$-d sketch. However, this is perhaps an appropriate point at which to introduce two highly significant (though aging) theories of visual perception: Gestalt theory, and Gibson’s ecological theory. Both of these theories operate at a much higher level than considered in the previous categories, and are potentially more useful for application to the expression of visual queries.

The approach of the Gestalt school (e.g. Koffka, 1935) is of course typically summarized by the statement that an image is ‘greater than the sum of its parts’. The implication is that the generation of meaning is not simply an additive process, but must be considered holistically, comprised of multiple interactions. While wholly unconcerned with how the processes may occur at a neural level, the Gestaltists identified a number of rules with regard to the extraction of information from an image, at least some of which are now understood at a computational level.

The central theme of these rules is organization – the visual system extracts (or imposes?) order from a visual scene, grouping objects by proximity, similarity, or 'common fate'. It identifies continuous lines, and closed boundaries, and will prefer stability and symmetry over uncertainty or asymmetry in a figure. Perceptions of orientation, depth and motion are derived from such rules.

Gibson's ecological theory of vision (developed over many years, finally presented as a whole in Gibson, 1979) has in some ways much in common with the Gestalt approach. Gibson paid little or no attention to the questions of how vision operated at a biological level, but was concerned with a holistic, phenomenological view of what is seen, and how it is seen (deliberately) ignoring the lower levels of form extraction.

Gibson's fundamental argument (or at least one part of it) is that the 'visual world' is not composed of abstract geometrical constructs – of points, lines and planes – but of objects: fibres and surfaces. His most radical argument was that we do not 'see' the third dimension – depth – by means of computation or inference. We do not see it at all, he argues; what we see are 'invariants of structure'. For example, when we view a rectangular table from one end, we consistently see a rectangular table, we do not see a trapezoid (which is the shape actually cast on the retina). In fact, the shape cast on the retina will rarely be a true rectangle, as we will rarely see a table directly from above (and even then with some distortion), but we will always 'see' a rectangle.

This point has significance for the viewing of two-dimensional pictures. Viewers comparing a picture of a table taken from a low angle, and of the trapezoid matching that shape, may not see much similarity, at least in an initial viewing. They will see a real object, a solid rectangular table, and an abstract geometric shape. As the fidelity of the picture of a table decreases, perhaps to a black and white photograph, to a shaded drawing, to a line drawing, the similarity may increase, but Gibson would take this as further evidence that what is seen in the real world is texture, rather than lines and angles.

An interesting visual extension operating in pictures is the effect of surface attraction – that is, the apparent contours of an object presented in silhouette change depending on the orientation of a surface implied by other objects in the scene. With a full scene, this effect is possibly redundant, as other elements of the image will provide the details anyway, but it does demonstrate the contingencies by which some elements are interpreted. A similar reflection on the minimal information required to discern form comes from Moore and Cavanagh's (1998) discussion of the extraction of form from shadow. In a series of studies presenting participants with black and white patterns of varying complexity matching the patterns of shadows on three-dimensional objects, they found that delineation of volumetric primitives (such as geons) was insufficient to interpret depth without some higher-level model of the object represented. It may be that such explicit top-down influences are not necessary for higher fidelity greyscale or colour images where further contextual cues are available. The point is though that depth is not perceived through light and shadow, but rather light and shadow appear to be interpreted through an understanding of depth (cf. Williams and Hanson, 1996).

Gibson (1979) describes a nomenclature for the description of scene elements to distinguish surface geometry, which applies in the visual world people actually see and interact with, from the abstract geometry used to describe forms theoretically. In this, a plane becomes a surface, for example, and the intersection of two surfaces becomes an edge rather than a line. Scenes are defined in relation to the *ground*, literally the surface of the earth in the scene viewed. Objects are defined in relation to the ground and to other objects, being attached or detached, and in relation to the viewer, showing enclosing surfaces. Other objects are described in real-world terms: sticks, sheets, fissures.

Gibson (1979) devotes the final section of his book to picture perception, which he sees as – while obviously related – not necessarily a simple subset of ambient or ambulatory visual perception. He defines a picture as composed of 'optical invariants': 'The depiction ... captures and displays [the optical invariants] in an optic array, where they are more or less the same as they would be in the case of direct perception' (p. 262).

These optical invariants are the informational components of vision – which, for Gibson, are mainly resolvable to texture information: discontinuities and gradients. It is these components that are recorded and perceived in a picture: '[A picture] is not a substitute for going back and looking again. What it records, registers or consolidates is information, not sense data' (p. 281). The distinction is perhaps moot, philosophically speaking, but arguments of identity and representation are beyond the scope of the current work.

Gibson also notes the paradox of a picture being a scene (i.e. a complex informational array) and a surface (a flat, informationally neutral panel) at the same time. It is assumed that when people are oriented towards picture content, the surface quality is not an issue, though that may not be the case (particularly when images of different media are compared – does a scanned or digitized photograph have more or less of a surface quality than a digitally originated scene?). This point is contained in Fidel's (1997) distinction between the image when used as an object (information in itself), and when used as data (a source of information).

Other studies have found that affine shape (the projected shape or silhouette) is adequate for object recognition within certain parameters. This is related to the part-based approaches discussed previously.

The 'directness' of visual perception has formed the basis of an ongoing debate in philosophical circles. In these debates, Gibson is often held up as the epitome of direct perception, whereby we perceive things such as depth in perspective, etc. because these are ecologically 'true' patterns. This directness leads to such counterintuitive arguments as 'depth is not perceived.'

In opposition, Goodman (1984) argued that phenomena such as perspective are merely conventions specific to Western cultures, and that Gibson's invocation of them as representing the reality of human perception is fallacious. Support for this view is the observation that perspective in painting is usually only considered in terms of depth, not height or breadth. However, this point has itself been rebutted. For example, Boynton (1993) argues that while Western art has made perspective conventional, it *does* reflect actual visual patterns, and cannot be rejected as wholly conventional:

> Just as the rules of cooking ... reflect and perpetuate the laws of nutrition, so the rules of artificial perspective can be thought of as reflecting and perpetuating the laws of ecological optics. Perspective may best be conceived as a conventional system that guides the way that ecological laws are used in Western art. (p. 63)

An issue that drives much of the debate is what constitutes the basic object of vision, particularly the status of the retinal array. For much of the history of vision research, it has been portrayed as an image, on which further perceptual and cognitive processes operate. One of Gibson's (1950) tenets was that such a view, while intuitive on the basis of the optics of the eye, was essentially fallacious – the retinal pattern is an integral as well as a primary part of the visual process. As Gibson said: 'It should not be thought of as an image, and even less, as a literal picture. It is an event composed not of light, but of nerve-cell discharges' (p. 50). Cognition does not act *on the* retina, the retina is *part of* the cognitive process.

The view of the retina as an image persists for some though. Wetherick (1999) proposed a behaviourist approach (after Taylor, 1962, and in response to Sharrock and Coulter, 1998) in which the visual scene as represented on the retina is the stimulus, perception the response. Sharrock and Coulter (1999) dismiss this, stating that Wetherick was 'presenting a ... solution to the Gibsonian problem which we had attempted not to solve but to *dissolve*' (abstract, p. 557).

Despite their complexity, none of the features or elements of vision mentioned in this section involve any semantic, or specifically linguistic, input to describe or perceive. That is not to say that they *cannot* have this dimension, simply that they do not *require* it. The types of content discussed to this point, and the following categories of content, fall either side of Enser's (1995) distinction between visual and linguistic content.

Semantic units

'Semantic' in this context refers to meaning that is not derived from purely visual information. A square has no meaning beyond that it is a square, a building does. Note that 'semantic' is used rather than 'linguistic' because the units or features at this level do not necessarily require *naming*, but they do require a knowledge of meaning and (if appropriate) function. Semantic units may be general or specific, referring to categories of object, or to specific instances.

There is a vast literature that could be invoked here – addressing the coding and organization of memory, and its access for different purposes – but it is not intended to pursue every theoretical avenue. Instead, a brief summary is given. There are basically two elements to this: (a) the organization of concepts, and (b) the retrieval of names.

Underlying the nature of object naming is the theoretical literature of concept acquisition and concept organization. The basics of this follow, but the focus for the current work is how names are associated with concepts, and how they are retrieved. In this article, we are interested in the types of content within an image, which may be used to construct a query. Therefore, we are concerned primarily with how names are associated with visual stimuli, rather than the theoretical structure of the concepts those names refer to. It is also likely that non-verbal descriptions such as sketches are reliant on the same process – object drawing is also driven conceptually.

Approaches to questions of concept acquisition and organization are rooted in the work of Rosch (e.g. Rosch et al., 1976; Rosch and Lloyd, 1978). In this approach, concepts are arranged hierarchically, with the 'basic level' of a concept being that which is 'at the most inclusive level at which there are attributes common to all or most members of the category' (Rosch and Lloyd, 1978: 31). Examples of the basic level are 'table', 'chair'. Superordinate categories ('furniture') may share only a small proportion of those attributes and are over-inclusive, while subordinate categories ('kitchen chair') contain more specific attributes, and are less inclusive. Hoffmann and Kaempf (1985) found that primary, basic-level concepts are preferred for naming, though Tanaka and Taylor (1991) found that expertise raises reporting of subordinate category judgements to the level of basic level categories.

Instances within categories are defined by prototypes, 'the clearest cases of category membership defined operationally by people's judgements of goodness of membership in the category' (Rosch and Lloyd, 1978: 36). There is not a single prototype for any given category or subcategory, rather it is a subjective indicator of representativeness, judged by the 'cue validity' of visible (or available) attributes. In visual terms, the canonical view discussed previously (Neisser, 1967) approaches a prototype, though there are important differences (for example, a canonical view can be defined fairly precisely, while a prototype can be fluid and vary with context).

Landau et al. (1998a) distinguish between instinctive and theoretical similarity. Generally, instinctive similarity is that which is determined by perception, thus including shape, texture and colour, while theoretical similarity might include function. For example, a picture of a rake might have instinctive similarity with a comb, but theoretical similarity with a hoe.

Landau and Jackendoff (1993), on the other hand, found, as might be expected, that shape is not important for place recognition or naming.

Cox (1986) similarly distinguished between intellectual vs visual realism – demonstrated in the way that children will draw an object in its canonical view, or even more abstractly, rather than how it is actually visually perceived. For example, a table is always drawn with four legs visible, a house with a pitched roof, and the sky as a blue band at the top of the page. Bremner and Moore (1984) argue that children do produce realistic drawings of unknown, unlabelled objects they have not handled, suggesting that this is not a failure of visual cognition of eye–hand mapping, but a result of deeper cognitive processing.

There is evidence though that different heuristics for generalization apply for artefacts compared to natural objects, where texture or composition may be more important than shape. Function for natural objects is not generally an appropriate, or realistic, attribute – what is the function of a tree? Humphrey et al. (1994) found that surface properties (such as colour, texture) aid natural object naming, but not of manufactured objects. An exception is musical instruments, which show the same pattern as natural objects (Dixon et al., 2000). It is possible that, as natural objects cannot readily be discriminated on the basis of function, other attributes may be primary.

The distinction of function and shape has also been studied by D'Arcais and Schreuder (1987) and a body of work by Landau (e.g. Landau et al., 1998a, 1998b; Landau and Leyton, 1999; Landau and Shipley, 2001). D'Arcais and Schreuder (1987) distinguish between perceptual elements (P-elements, essentially visual) and functional elements (F-elements) – cf. Ornager's (1997) of/about distinction. Landau's work has also considered the relationship of each group of elements with age (after Gentner, 1978 – see also Furmanski and Engel, 2000; Merriman et al., 1993). The evidence is not entirely conclusive, but it appears that the influence of function increases with age, although for some classes of object, shape returns to dominance.

For known objects, the process of naming is more interactive, with a principle of mutual exclusivity constraining potential categorizations on the basis of both shape *and* function (e.g. a comb would not be categorized as a rake). However, in Landau et al.'s (1998a, 1998b) studies, generalization to a new form of a known object was either on the basis of shape ('it's a paper comb'), or not at all ('it's a piece of paper'). Function does not generalize as it does for novel objects. The early role of semantic qualities is demonstrated by Mitchell et al. (1996) who found that when presented with visual stimuli, children found it easier to choose a 'silly' or inappropriately coloured object than an appropriately coloured one. Price and Humphreys (1989) found that incongruent colour disrupts naming of similarly shaped objects, and disrupts classification of dissimilar objects.

The approach described by recognition by components (Biederman, 1987) has also been used to consider how objects are named. Biederman and

Cooper (1991) found that priming effects for partial figures is only for the exact figure, not the complements of the original stimuli. Frazier and Hoyer (1992) found that the effect of the amount of fragmentation increased with age, while Brown and Koch (1993) found that identification of partial and fragmented objects is better when contours are left open (presumably because completion is aided). Interestingly, providing an occluding object can hinder identification, though only on a small scale (presumably it interferes with continuation of line).

Davidoff and De Bleser (1994) found a naming deficit in brain-damaged patients only for picture naming – not for visual or tactile inspection of actual objects. It may be that the role of shape may be impaired, but functional access is not. As Humphrey et al. (1999) suggest, the process is complex and not strictly linear. They suggest both visual and semantic information constrain naming/concept identification bi-directionally although in naming latency reaction time tasks, there is evidence that visual priming facilitates object naming (Bar, 2000; Davidoff and Ostergaard, 1988; Landau et al., 1998b; Ostergaard and Davidoff, 1985). Naming and category judgement are not the same though – naming is in one sense a refinement of category, but is not necessarily sequential.

Meyer and Van der Meulen (2000) and Meyer et al. (1998) have identified the fact that access to a name is completed before gaze shifts, suggesting that name selection is a discrete phase. Levelt et al. (1999) support this finding, in that when multiple options are presented, a single phonological alternative is activated after selection, rather than multiple alternatives being activated before selection.

As mentioned earlier, Gibson (1979) sees form and function as essentially integrated – meaning is integral to viewing the visual array. This is exemplified in the concept of 'affordances', the functions that visual objects suggest to us. Analysing cognitive development, Gibson says: 'The affordance of an object is what the infant begins by noticing. The meaning is observed before the substance and surface, colour and form, are seen as such' (p. 134). It is essentially how a person is *able to* interact with the environment, and refers to all elements, not just complex artefacts. The *ground*, in Gibson's terms, affords weight bearing, as long as it fulfils parameters of extent and rigidity. Gibson distinguishes his definition from similar earlier concepts developed by the Gestalt theorists, in that while these theorists (e.g. Koffka, 1935) referred to the 'valence' as a component of the viewer's expectation of the object, Gibson's definition of affordance is an invariant aspect of an object, and remains part of the viewer's awareness even if it is not particularly salient at the time (for example, a post-box retains the affordance of posting something, even when the viewer has nothing to post).

Panofsky (1970[1955]) also discusses the centrality of perceived function to the way an image, or indeed any object, is seen. However, he takes more of the Gestaltist approach, identifying function with the demands and expectations of the viewer:

> When a man looks at a tree from the point of view of a carpenter, he will associate it with the various uses to which he might put the wood; and when he looks at it from the point of view of an ornithologist he will associate it with the birds that might nest in it. (p. 34)

This highlights a potential failing of Gibson's (1979) view – what happens when there is more than one affordance for a given object? Gibson may reply that they are all equally valid, and there is no reason why one needs to have priority over another, but it is a fact that people perceive one available function or role over another, in a given context.

While Panofsky's (1970[1955]) discussion is for the most part restricted to art, he argues in common with Gibson that we do not see line and form but objects, and that these objects are intrinsically meaningful. His lowest level of consideration, pre-iconographical content, fits quite neatly into this level of general semantics: 'The world of pure forms thus recognized as carriers of primary or natural meanings may be called the world of artistic motifs' (p. 54). These 'primary meanings' are the identification of forms as human beings, animals, etc.

To some extent, Panofsky's next level, iconography, may also be considered to fall within this category of semantic units. However, it requires a specific level of knowledge and abstraction which makes it qualitatively different, and is discussed in the next section.

Abstraction

Some types of information content require non-visual information to decode – these are referred to here as abstracted types. Four are presented separately here, as requiring distinct types of non-visual knowledge, although there are overlaps that will be discussed when relevant. These types also mark the shift, in Ornager's (1997) terms, from discussing what a picture is *of*, to what it is *about*. The four types of abstraction distinguished here are presented in increasing degree of idiosyncrasy, i.e. the extent to which they rely on *individual* experience and knowledge.

Abstraction corresponds, to a greater or lesser extent depending on specifics, to Panofsky's (1970[1955]) definition of *iconography*. He described this with particular reference to art history: '[iconography] presupposes a familiarity with specific themes or concepts transmitted through literary sources or an oral tradition' (p. 61). In other words, it invokes background knowledge. In the following descriptions, the definition of 'literary and oral sources' is generally adequate, but not for emotional abstraction, where 'experience' is the primary source.

It is in these categories that the issues raised by reception theory become relevant. Viewers'/users' biographical, political or ideological background can, and in most cases will, influence their interpretation of images, and accordingly the way they express a query. However, the contexts which may so influence viewing are massively heterogeneous; thus, while the categories described in the next sections are based on the broad roles and contexts a viewer may bring to an image, detailed contexts are not addressed other than as illustration. With regard to the stated aim of enhancing IDB interfaces, catering for too fine distinctions of viewer/user is deemed counterproductive.

Contextual abstraction

Contextual abstraction refers to non-visual information which is presumed to be universal, in that it is derived from knowledge of the environment. A simple example is telling whether an image represents day or night, an inside or outside scene, or different weather conditions.

Contextual abstraction has something in common with the visual extension described earlier (and in a purely Gibsonian view would fundamentally be part of the visual array) but, particularly when concerned with images rather than an ambient array, requires some abstraction of meaning from visual cues. Figure 6 shows similar views of the Empire State building in different weather conditions. Purely visually, there are differences in luminance and colour, but the difference in weather conditions is an abstraction from these differences.

There is a degree to which contextual and cultural abstraction overlap. Consider a scene of sunbathers on a beach: to someone in the UK, it may suggest the summer; to an Australian, the winter. Latitude and climate play a part in both familiarity and associations with different environmental

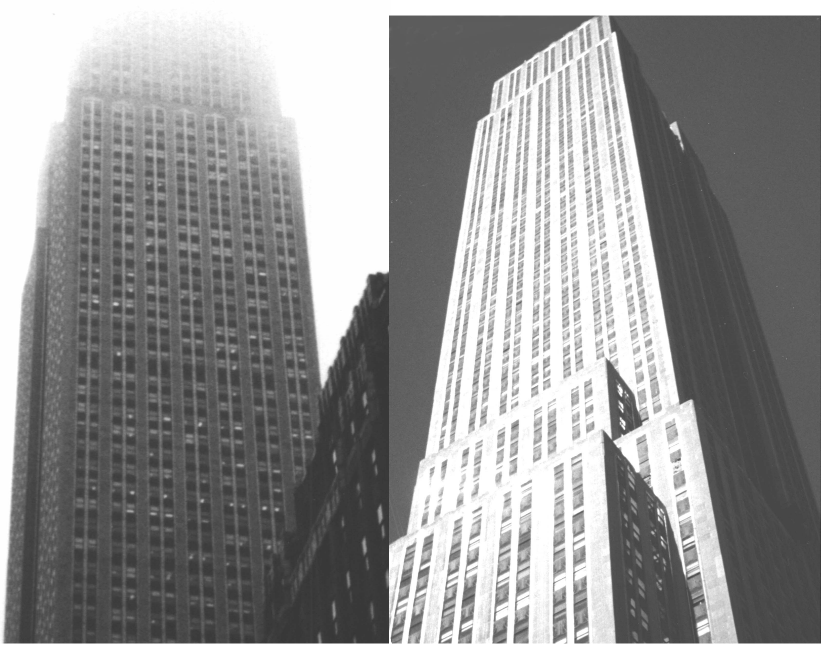

Figure 6 View of the Empire State building in different weather conditions. © B. Burford, 2002.

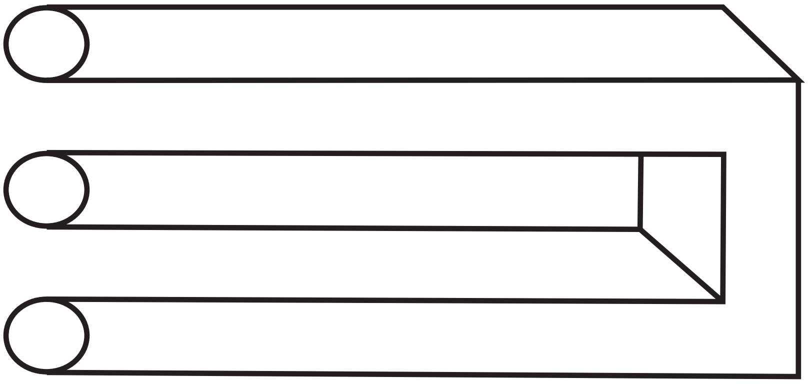

Figure 7 An 'impossible object'.

conditions – consider the associations a snow scene may have to someone living within the Arctic Circle as compared to someone living in the Tropics.

Contextual abstraction also has more esoteric implications. Malinas's (1991) discussion of depiction raises the idea of objects that may be non-depictable and, conversely, images that may represent 'impossible objects' (such as the example in Figure 7). The point is also made by Natsoulas (1999). Such images illustrate a peculiar interaction between visual extension and contextual abstraction. Their form is perceived by visual extension, while the impossibility is determined by contextual abstraction 'out of' the picture into the real world.

This sort of ontological distinction is discussed in a slightly different way by Smith (2000), who highlights the distinction between boundaries which exist because of actual physical discontinuities, and thus exist independent of cognition (abstruse philosophical arguments notwithstanding), and boundaries which exist only because of imposed cognition (for example, country or county borders that are not coastal or do not follow watercourses). These 'fiat objects' may impinge directly on visual perception, albeit transiently – Smith describes the horizon as a 'one-dimensional fiat boundary in the interior of the visual field' (p. 324). Optically, the horizon is 'real' but does not actually exist as a boundary. It could be argued, then, that all images present fiat boundaries as they do not contain the actual boundaries perceived by the viewer (even if they may be depictions of actual objects). For example, the boundary implied by the affine shape of an orange is roughly circular, but the boundary perceived by a viewer is a convex hemisphere.

This discussion has much in common with Natsoulas's (1999) discussion of the 'virtual objects' Gibson (1979) refers to in his ecological theory. Gibson defines a virtual object as an object in a photograph or an inkblot: 'They are not perceived, and yet they are perceived' (p. 283). Natsoulas (1999) concludes that such objects (including images in photographs, images seen in inkblots, shadows and magnified objects) 'either

... was, is, or will be an actual part of the physical world or the item has no existence' (p. 368). While his argument is not entirely clear (the quoted statement is preceded by 'there are no virtual objects that are not also actual objects' (p. 368), the essence is that any object which casts a pattern on the retina is directly perceived and is de facto 'real'. The discussion does though render quite ironic Gibson's (1979) statement that the use of the term 'virtual' is independent of its usage in classical optics because the latter is 'swamped in epistemological confusion' (p. 283).

As a final illustration of the complexities of the ontological arguments about representation, depiction, etc., consider Figure 8, a single horizontal line. This line could *represent* a horizon and *resemble* it, but does not *depict* it. It could, however, resemble and depict a section of railway line through a snowfield, seen from a great height, but may represent winter travel, or it could depict, resemble and represent another straight line. In each of these cases, which would be perceived, the representation or the depiction – and to what extent will the line be perceived as a depiction, or as an image in its own right?

Such questions of similarity and identity are philosophical questions beyond the interests of the current project but they do serve to illustrate the complexities of image content, and that a typology such as the one presented here can only be an approximation.

Figure 8 A horizon, a railway line from the air, or an illustration of another line.

Cultural abstraction

Cultural abstractions are presumed to be fairly generalized in that the 'literary sources and oral tradition' which inform them comprise the general culture of the viewer. Subcultural interests and expertise mean that the boundaries of cultural knowledge cannot be prescribed, but the type of image content perceived and the way in which it is derived are assumed to be generalizable.

All the abstractive levels described in this article are hermeneutic, in principle, in that they are interpretative rather than objectively descriptive. As such, they may all be interpreted as cultural: Hart (1993) provides a discussion of hermeneutics in reference to images, citing Mannheim's (1952[1923]) assertion that all perception is based on cultural meaning (or *weltanschauung*). While literally this means, of course, that all interpretations are culturally defined, it is probably safe to assume – and it certainly aids precision if we do so – that professional codes (for example) supersede cultural ones, where relevant.

Cultural abstractions may refer to many things; however, in the most general cases, they may refer to political or sporting events, or to the historical era that an image represents. Consider Figure 9 that depicts a

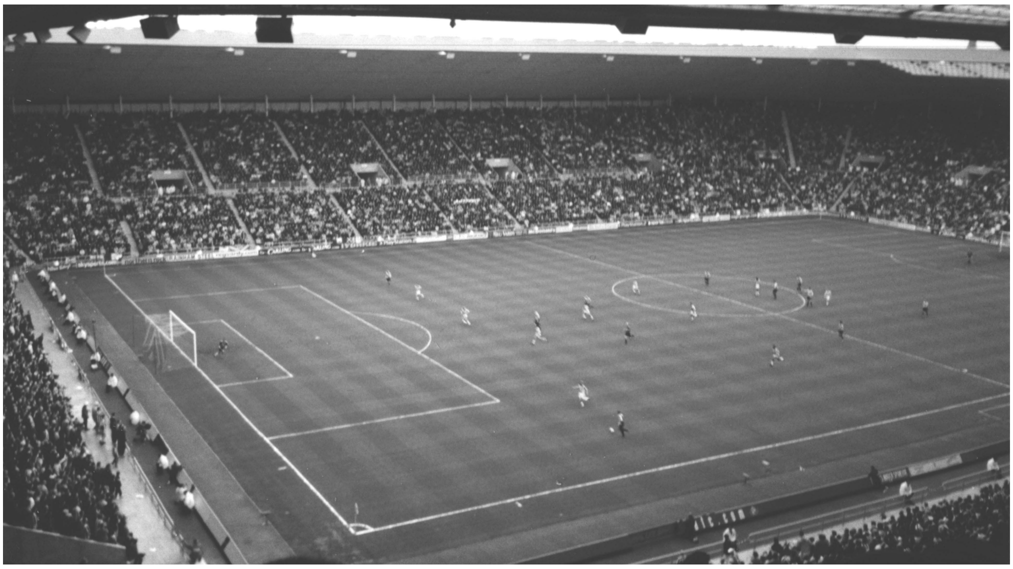

Figure 9 Sunderland vs Newcastle United, 2001. © B. Burford, 2001.

Sunderland vs Newcastle football match. To identify all the details and associations of this picture (teams, dates, the historical rivalry between the teams) would require a depth of knowledge available through cultural immersion and some specific subcultural expertise – merely identifying the picture as football-related requires cultural knowledge. Even this image though does not have the additional cultural weight of the more iconic picture of Bobby Moore holding the World Cup aloft after England's victory in 1966. More generally, interpreting photographs of holidays, religious festivals, etc. requires knowledge of that culture: for example, a Christmas tree is linked to Northern European and North American culture; a Menorah to Hanukah; a knotted handkerchief to English seaside holidays.

Professional/technical abstraction

Professional or technical abstraction is defined here as that information which requires specific technical expertise to interpret or extract. It is the type of abstraction which most directly corresponds to Panofsky's (1970[1955]) definition of iconography, formulated with art historians in mind. It may equally apply to professional image users in other domains though, where meaning can only be discerned through the application of specific, expert knowledge derived through learning and training. As an example, consider the image in Figure 10. To a lay viewer, it is an abstract pattern, but to a clinician it is an informative cell

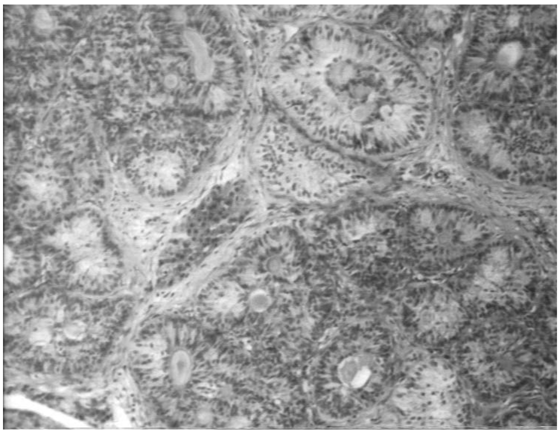

Figure 10 Stained histological micrograph. © Bristol Biomedical Image Archive.

structure. The quote of Panofsky's given in the previous section emphasizes this: the cell structure is the tree to the doctor's carpenter.

Other examples of such technical imaging situations are radiology, where X-ray and MRI images require the expertise of trained clinicians to interpret beyond the identification of gross anatomical features (limbs, torso, etc.).

Ingwersen (1982, 1996) discusses meaning in terms of classification, and the fact that object recognition can be influenced by context – people will classify an object differently in different contexts, as context primes conceptual structures. This illuminates the phenomenon discussed by Heehs (1995) that different narratives can be imposed on/seen in a picture depending on the viewer's background in approaching it. Professionals approaching an image in a professional context will classify the image and extract content that differs from that extracted by non-professionals.

Firschein and Fischler's (1972) discussion of types of description is also significant here – different purposes will elicit different descriptions. Firschein and Fischler also noted that goal-oriented descriptions are richer and contain more specific information than generalized descriptions. Thus it might be expected that a professional's description *qua* a professional will be more detailed than a general description.

The specifics of professional queries are discussed by Gross and Do (1995: architects), Larsgaard (1996; Larsgaard and Carver, 1995: geographical images), Tagare et al. (1997: medical images) and Hastings (1995: art historians). The discussion of Gross and Do (1995) is particularly interesting as it describes the visual shorthand and analogies that creative professionals can use (for example, using a crude spiral to signify the complex layout of the Guggenheim Museum).

An example of how different backgrounds can confuse image query and retrieval is given by Keister (1994) who describes an instance in which an image was requested which depicted polio sufferers. After much searching, and almost by accident, it was discovered that what was required was a street scene at the turn of the 20th century, in which polio sufferers could be seen. The image, however, was indexed in terms of location and period, with illness not considered a significant part of the scene – only specific historical knowledge would associate endemic polio with that time period. (This example also serves to illustrate an overlap between cultural and professional abstraction: as events move further into the past, what was once common cultural knowledge becomes the domain only of the older population, and ultimately of expert historians.)

Emotional abstraction

While one of the most intuitively obvious, emotional abstraction is perhaps the hardest type of content to specify. It refers to affective or emotional associations or responses people may have to an image. It is distinct from cultural or technical abstraction in that a generalized affective response does not rely on particular, identifiable expertise or experience; although emotional responses and interpretation *are* obviously based on experience, that experience will *generally* be idiosyncratic.

Obviously *sometimes* this content will be universal, for example a picture of a smiling face would be expected to be associated with happiness, regardless of the viewer's cultural background. However, in other instances, there will be cultural influences on the emotional content perceived in an image. For example, a picture of happy English football supporters at the 1966 World Cup may have associations of disappointment for a German viewer. However, just as validly, the positions may be reversed if the English viewer sees the picture as representing a lost time and an unhealthy nostalgia, and the German viewer as a low point long left behind. (A similar image of the 2001 qualifying match might not have these corollaries.) The abstraction is fundamentally idiosyncratic, whatever other points it might touch on.

This category contains much of what Panofsky (1970[1955]) calls iconological content. This consists of 'intrinsic meaning or content ... qualified by one personality and condensed into one work' (p. 55). These 'symbolical' values require 'something more than a familiarity with themes or concepts as transmitted through literary sources ... we need a mental faculty comparable to that of a diagnostician' (p. 61). The content is intuitive and ill defined, and is very much in the eye of the beholder.

This content may be representational or evocative. For example the two crude examples in Figure 11 may both represent happiness or brightness. Figure 11(a) does this representationally, Figure 11(b) evocatively. In Panofsky's terms it may be argued that Figure 11(a) is pre-iconographical, while Figure 11(b) depends on iconological associations.

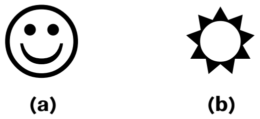

Figure 11 (a) A face icon representing happiness; (b) a sunshine icon evoking happiness.

Of course, all images have the potential for emotional content or association, and mapping this content to visual content is perhaps the biggest challenge for IDB design.

Metadata

The term 'metadata' is often used to refer to classification and coding schemes for particular IDB formats and technical specialities (e.g. ICONCLASS). These classifications are contained within the categories previously described, and are termed metadata because the tags, indexing fields or whatever are attached to the image data (Jorgensen, 1999, provides a review of such indexing systems).

However, there remains a class of non-content information, which can play an important part in describing an image (and so forming part of a query) but which *cannot* be derived from the image itself. It has absolute,

Perceptual primitives
-texture
-hue
Geometric primitives
-rectangular sections
-arc
Visual extension
-shadow
-texture gradient
Semantic units
-woman
-dog rose
Contextual abstraction
-water
-spring/summer
Cultural abstraction
-art gallery
-old-fashioned
Emotional abstraction
-sadness
-loss
Technical abstraction
-pre-Raphaelite art
-Shakespeare

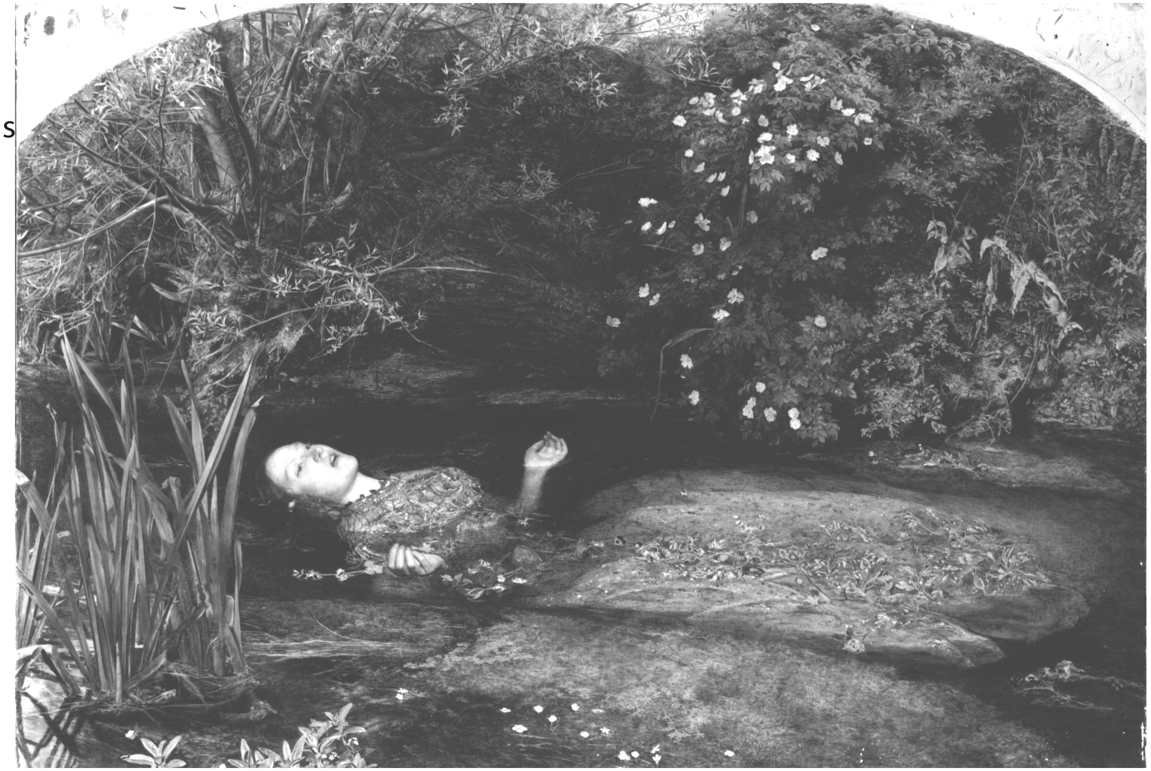

Figure 12 Illustration of content types: painting (*Ophelia* by John Everett Millais). © Tate, London 2002.

given values for each instance of an image and is not open to interpretation or labelling by the viewer.

These 'true' metadata may include details such as the date and time a photo was taken, copyright information, or technical details such as file size, file type, compression ratio, colour depth, etc. Elements such as aspect ratio, framing and composition may also fall into this category as they do not refer directly to image content but to the presentation of the image surface.

Some metadata fields may to some extent correlate with content-based information (e.g. time of day may be indicated by light levels, time of year by a colour histogram, date by contextual abstraction, low colour depth by an obviously restricted palette), but the metadata themselves are discrete and may only be accessed directly. Nevertheless, some content-based queries may be efficiently mapped onto metadata searches.

CONCLUSION

This article has presented a system for the classification of image content consisting of nine categories. These categories are derived from literature in many areas, from visual perception to art history. Figures 12 and 13 apply the taxonomy to example images, illustrating how each type may be instantiated in an image. The instances given are illustrative only; the descriptions given by actual image users may be far more varied and contextually rooted, but it is pointless at this stage to second-guess that context.

To begin with, the two images illustrate different metadata – one being a painting, the other a photograph. Both are reproduced in black and white from colour originals, though in practice they may differ in this respect. There is a clear difference in aspect ratio: an aspect of metadata which may be used in description. Another, more subtle, distinction in terms of metadata is that the image of *Ophelia* is of a distinct, unique original, of definite size, while the photo has no definitive 'original' size (although the scene it depicts obviously has).

Both pictures obviously contain perceptual primitives, though as discussed in that section, actually reporting this level is not *directly* possible. However, some elements are only meaningfully classified as perceptual primitives, as including them in higher categories is misleading (naming an attribute is different to naming an object). So, for pragmatic reasons, colour (including hue, saturation and brightness) as well as brightness of greyscales and texture may be described in these two images.

As reproduced, there is a range of greyscales in both images. In the originals, there are different hues – predominantly green in the case of *Ophelia*, yellow/brown and blue in the photograph of the castle harbour. Both also feature varied textures and, while some of them may be described at the semantic level (water, foliage), some descriptions are meaningful only as analogues to the perceptual level (speckled, fuzzy, rough, smooth). Textures described in this way are not necessarily features of the scene

Perceptual primitives
- -texture
- -contrast

Geometric primitives
- -triangles
- -quadrilaterals

Visual extension
- -shadow
- -texture gradient

Semantic units
- -sand
- -castle

Contextual abstraction
- -sunlit
- -wet
- -hot

Cultural abstraction
- -holidays
- -summer

Emotional abstraction
- -history
- -security

Technical abstraction
- -building age
- -aperture

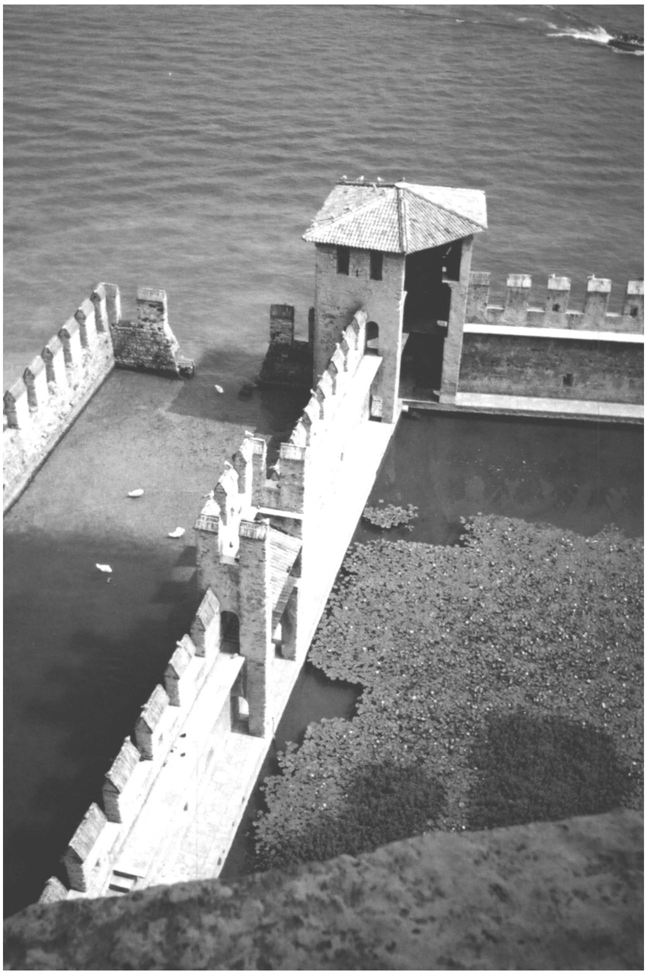

Figure 13 Illustration of content types: photograph (castle harbour in Sirmione, Italy). © B. Burford, 2002.

depicted, but can be qualities of the image surface as well – an out-of-focus or badly reproduced image may be described as 'fuzzy' or 'hazy'.

Both images feature quite strong, though implicit, geometric content. The overall structure of *Ophelia* may be seen as two quadrilaterals above a triangle, although further detail may include the curved edges of these. At a more detailed level, areas of the picture have circular and triangular elements (e.g. the flowers and leaves). The photograph is largely defined by the irregular quadrilaterals on the left, right and top of the picture, although the triangular aspect of the wall seen at the bottom right corner is also dominant. Other polygons can be found in the walls and curves in the water. The use of geometric primitives in description is not limited to regular shapes, but includes rough approximations (indeed for most natural scenes there will be very few completely regular forms).

Visual extension generally refers to any of a number of cues for depth, and the observation of that depth. Thus, noting that the plant stem at the bottom of the painting is in the foreground is an instance of visual extension, while there are obvious cues in the photograph that it has been taken from a height. Both images contain examples of texture gradients, while the shadows cast in both pictures, particularly in the photograph, are also indicative of depth. The photograph actually provides a non-linear cue from texture gradient as its depth of field leaves the foreground out of focus relative to the rest of the image (usually, closer objects show finer detail).

Semantic units may be specific or general. 'Wall' is a semantic unit at a very general level, as is 'plant', 'woman' or 'river'. 'Ophelia' as a proper noun refers to the image content at a semantic level, as well as to the painting at a metadata level. Finer levels of detail may describe the individual plants identifiable in the painting, or the stone in the walls visible in the photograph, although these may be more appropriately considered as technical abstractions (obviously there is overlap here – types are not always mutually exclusive even for particular instances of content).

Contextual abstraction is also possible in both images. Evening is implied by the general illumination of the scene in *Ophelia*, and the season by the foliage. The photograph is clearly taken in late morning to early afternoon from the height and angle of the shadows. Water is indicated by reflection and refraction in both pictures. Dampness around the water may also be inferred by contextual abstraction, as might heat in the photograph, and the scent of undergrowth in the painting. These are non-visual associations made on the basis of experience of similar environments and are equally valid potential abstractions as visual or abstract ones.

Cultural abstraction relies on background information derived from a common culture. In the case of *Ophelia*, this could refer to art galleries, or the historical association of the style of painting (e.g. it looks old-fashioned, Victorian). For the photograph of the harbour, associations might be holidays, medieval history, etc.

Some of the elements of *Ophelia*, however, may fall between cultural and technical abstraction. For some people, recognition of the picture, the scene it depicts, associations with Shakespeare, etc. will be a part of a cultural background. For others, these consequent associations may not be available and might be taken to require specific expertise. Broadly though, instances of subcultural knowledge remain within the category of cultural abstraction.

However, even without knowledge of the story depicted by *Ophelia*, it might be expected that the painting would have the emotional abstraction of sadness and loss, although an alternative reading might be of relaxation. Likewise, the castle harbour might suggest solitude and security, although the cultural abstraction of 'holiday' may have happy associations. Both images may also have associations of the endurance of nature; for example, the presence of greenery against human elements.

Technical abstraction for both images may refer to both content and non-content elements. For the painting, the most obvious form would be biographical information about the artist and the pre-Raphaelite movement of which he was a part, with content-related abstraction, possibly in terms of technique and materials. For the photograph, a similar distinction of content and non-content abstraction is possible – perhaps identifying the building period on the one hand, and elements such as film stock, aperture, etc. on the other. Of course, both images, as digital reproductions, are open to technical abstraction addressing whether they have been compressed, by how much, printing process, etc.

These illustrations show that categories within the taxonomy can sometimes overlap in that elements may not be uniquely classifiable without inferring the viewer's thinking, and surmising *why* they have used particular terms. It is hard to see how such polymorphism (such as between semantic, cultural and technical abstraction) may be avoided. These are fundamental qualities of image content and image perception though, and do not indicate problems peculiar to this taxonomy. In practice, it may be that the appropriate category can be inferred from context and the individual providing a description. For a theoretical taxonomy, such data are not available.

The aim of this article – to derive a comprehensive, yet concise taxonomy for the categorization of image content – has certainly been fulfilled. However, this taxonomy is not intended to be just for theoretical discussion but for *application* to professional image users and, ultimately, the design of interfaces for IDBs. It is likely that considering genuine descriptions of images will highlight any weaknesses or gaps in the taxonomy (including the problem of polymorphism referred to earlier).

For this reason, work in progress aims to validate the taxonomy with actual image descriptions before incorporating the taxonomy into a user-centred design process. The first stage of this work has involved interviewing image users, taking their descriptions of images, and seeing how completely their descriptions can be coded using the terms of the taxonomy. This will (a) indicate whether any additional categories are necessary; and (b) show which, if any, category is preferred for description.

It is assumed that different categories of content will lend themselves to different interface solutions. Within a database query, dominant categories are likely to be used more, and so should be catered for accordingly. Some categories may not be used at all, and so may be safely ignored within an interface, while some queries will require dynamic combinations of categories as well as interface elements. Empirical work will pursue this question as part of an iterative design process.

We have succeeded in theoretically codifying the ways in which people perceive images. It remains to be seen whether that theory is robust enough for application.

REFERENCES

Bar, M. (2000) 'Conscious and Nonconscious Processing of Visual Object Identity', in Y. Rossetti and A. Revonsuo (eds) *Beyond Dissociation:*

Interaction between Dissociated Implicit and Explicit Processing. Advances in Consciousness Research. Amsterdam: John Benjamins.

Berger, J. (1972) *Ways of Seeing*. Harmondsworth: Penguin.

Biederman, I. (1987) 'Recognition by Components: A Theory of Human Image Understanding', *Psychological Review* 94: 115–47.

Biederman, I. and Cooper, E.E. (1991) 'Priming Contour-Deleted Images – Evidence for Intermediate Representations in Visual Object Recognition', *Cognitive Psychology* 23(3): 393–419.

Biederman, I. and Gerhardstein, P.C. (1993) 'Recognising Depth-Rotated Objects: Evidence and Conditions for Three-Dimensional Viewpoint Invariance', *Journal of Experimental Psychology: Human Perception and Performance* 19(6): 1162–82.

Boynton, D.M. (1993) 'Relativism in Gibson's Theory of Picture Perception', *Journal of Mind and Behavior* 14(1): 51–70.

Bremner, J.G. and Moore, S. (1984) 'Prior Visual Inspection and Object Naming – 2 Factors that Enhance Hidden Feature Inclusion in Young Children's Drawings', *British Journal of Developmental Psychology* 2: 371–6.

Brown, J.M. and Koch, C. (1993) 'Influences of Closure, Occlusion, and Size on the Perception of Fragmented Pictures', *Perception & Psychophysics* 53(4): 436–42.

Bruce, V., Green, P.R. and Georgeson, M.A. (1996) *Visual Perception: Physiology, Psychology & Ecology*, 3rd edn. Hove: Psychology Press.

Burt, P. and Adelson, E.H. (1983) 'The Laplacian Pyramid as a Compact Image Code', *IEEE Transactions on Communication* COM-31: 532–40.

Cox, M.V. (1986) 'Cubes Are Difficult Things to Draw', *British Journal of Developmental Psychology* 4: 341–5.

D'Arcais, G.B.F. and Schreuder, R. (1987) 'Semantic Activation during Object Naming', *Psychological Research-Psychologische Forschung* 49(2–3): 153–9.

Davidoff, J. and De Bleser, R. (1994) 'Impaired Picture-Recognition with Preserved Object Naming and Reading', *Brain and Cognition* 24(1): 1–23.

Davidoff, J.B. and Ostergaard, A.L. (1988) 'The Role of Color in Categorical Judgements', *Quarterly Journal of Experimental Psychology Section A-Human Experimental Psychology* 40(3): 533–44.

Dixon, M.J., Piskopos, M. and Schweizer, T.A. (2000) 'Musical Instrument Naming Impairments: The Crucial Exception to the Living/Nonliving Dichotomy in Category-Specific Agnosia', *Brain and Cognition* 43(1–3): 158–64.

Dyson, M.C. and Box, H. (1997) 'Retrieving Symbols from a Database by their Graphic Characteristics: Are Users Consistent?', *Journal of Visual Languages and Computing* 8(1): 85–107.

Eakins, J.P. (1998) *Techniques for Image Retrieval* (Library and Information Briefings 85, October). London: British Library and South Bank University.

Eakins, J.P., Edwards, J.D., Riley, J. and Rosin, P.L. (2001) 'A Comparison of the Effectiveness of Alternative Feature Sets in Shape Retrieval of Multi-component Images', in M.M. Yeung, C.-S. Li and R.W. Lienhart (eds) *Storage and Retrieval for Media Databases 2001*, Proceedings of Society of Photo-optical Instrumentaion Engineers (SPIE) 4315: 196–207.

Enser, P.G.B. (1995) 'Pictorial Information Retrieval', *Journal of Documentation* 51: 126–70.

Fidel, R. (1997) 'The Image Retrieval Task: Implications for the Design and Evaluation of Image Databases', *New Review of Hypermedia and Multimedia* 3: 181–99.

Firschein, O. and Fischler, M.A. (1971) 'Describing and Abstracting Pictorial Structure', *Pattern Recognition* 3: 421–43.

Firschein, O. and Fischler, M.A. (1972) 'A Study in Descriptive Representation of Pictorial Data', *Pattern Recognition* 4: 361–77.

Frazier, L. and Hoyer, W.J. (1992) 'Object Recognition by Component Features – Are There Age Differences?', *Experimental Aging Research* 18(1–2): 9–14.

Furmanski, C.S. and Engel, S.A. (2000) 'Perceptual Learning in Object Recognition: Object Specificity and Size Invariance', *Vision Research* 40(5): 473–84.

Garnier, F. (1984) *Thesaurus iconographique: Systeme descriptif des representations*. Paris: Leopard d'or.

Gentner, D. (1978) 'On Relational Meaning: The Acquisition of Verb Meaning', *Child Development* 49(4): 988–98.

Gibson, J.J. (1950) *The Perception of the Visual World*. Boston, MA: Houghton Mifflin.

Gibson, J.J. (1979) *The Ecological Approach to Visual Perception*. Boston, MA: Houghton Mifflin.

Goodale, M.A. and Humphrey, G.K. (1998) 'The Objects of Action and Perception', *Cognition* 67: 181–207.

Goodman, N. (1984) *Of Mind and Other Matters.* Cambridge, MA: Harvard University Press.

Gracia, J.J.E. (1994) 'Can There Be Text without Audiences? The Identity and Function of Audiences', *Review of Metaphysics* 47: 711–34.

Gross, M.D. and Do, E. Yi-Leun (1995) 'Diagram Query and Image Retrieval in Design', in *International Conference on Image Processing Proceedings*, Vol. 0–8186–7310–9/95, pp. 308–11. New York: Institute of Electrical and Electronics Engineers (IEEE).

Grossberg, S., Mingolla, E. and Ross, W.D. (1997) 'Visual Brain and Visual Perception: How Does the Cortex Do Perceptual Grouping?', *Neuroscience* 20: 106–11.

Grund, A. (1993) 'ICONCLASS: On Subject Analysis of Iconographic Representations of Works of Art', *Knowledge Organization* 20(1): 20–9.

Hart, J. (1993) 'Erwin Panofksy and Karl Mannheim: A Dialogue on Interpretation', *Critical Inquiry* 19: 534–66.

Hastings, S.K. (1995) 'Query Categories in a Study of Intellectual Access to Digitized Art Images', *ASIS Annual Meeting Proceedings*, 9–12 October, pp. 3–8. Silver Springs, MD: American Society for Information Science (ASIS).

Heehs, P. (1995) 'Narrative Painting and Narrative about Paintings: Poussin among the Philosophers', *Narrative* 3(3): 211–31.

Heitger, F., Rosenthaler, L., Vonderheydt, R., Peterhans, E. and Kubler, O. (1992) 'Simulation of Neural Contour Mechanisms – from Simple to End-Stopped Cells', *Vision Research* 32(5): 963–81.

Heller, M.A. (1989) 'Picture and Pattern Perception in the Sighted and the Blind – the Advantage of the Late Blind', *Perception* 18(3): 379–89.

Hoffmann, J. and Kaempf, U. (1985) 'Mechanismen der Objektbenennung–parallele Verarbeitungskaskaden/The Mechanisms of Object Naming: Parallel Processing Cascades', *Sprache und Kognition* 4(4): 217–30.

Hummel, J.E. (2000) 'Where View-Based Theories Break Down: The Role of Structure in Human Shape Perception', in E. Dietrich and A.B. Markham (eds) *Cognitive Dynamics: Conceptual and Representational Change in Humans and Machines*. Mahwah, NJ: Erlbaum.

Humphrey, G.K., Goodale, M.A., Jakobson, L.S. and Servos, P. (1994) 'The Role of Surface Information in Object Recognition: Studies of a Visual Form Agnosic and Normal Subjects', *Perception* 23(12): 1457–81.

Humphrey, G.K., James, T.W., Gati, J.S., Menon, R.S. and Goodale, M. (1999) 'Perception of the McCollough Effect Correlates with Activity in Extrastriate Cortex: A Functional Magnetic Resonance Imaging Study', *Psychological Science* 10(5): 444–8.

Ingwersen, P. (1982) 'Search Procedures in the Library Analysed from the Cognitive Point of View', *Journal of Documentation* 38: 165–91.

Ingwersen, P. (1996) 'Cognitive Perspectives of Information Retrieval Interaction: Elements of a Cognitive IR Theory', *Journal of Documentation* 52(1): 3–50.

Iser, W. (1978) *The Act of Reading*. Baltimore, MD: Johns Hopkins University Press.

Jauss, H.R. (1982) *Toward an Aesthetic of Reception*. Minneapolis: University of Minnesota Press.

Jorgensen, C. (1995) 'Image Attributes: An Investigation', PhD thesis, Syracuse University, New York.

Jorgensen, C. (1999) 'Access to Pictorial Material: A Review of Current Research and Future Prospects', *Computers and the Humanities* 33: 293–318

Keister, L.H. (1994) 'User Types and Queries: Impact on Image Access Systems', in R. Fidel, T. Bellardo Hahn, E.M. Rasmussen and P.J. Smith (eds) *Challenges in Indexing Electronic Text and Images*, pp. 7–22. Medford, NJ: Learned Information, Inc.

Kimia, B.B., Tannenbaum, A.R. and Zucker, S.W. (1992) 'The Shape Triangle: Parts, Protrusions, and Bends', Technical Report TR-92–15, McGill University Research Center for Intelligent Machines.

Koch, C. and Abbey, L. (1999) 'Role of Geons in Object Recognition across Ages', *Perceptual and Motor Skills* 88(3): 983–91.

Koffka, K. (1935) *Principles of Gestalt Psychology*. New York: Harcourt Brace.

Landau, B. and Jackendoff, R. (1993) '"What" and "Where" in Spatial Language and Spatial Cognition', *Behavioral and Brain Sciences* 16(2): 217–65.

Landau, B. and Leyton, M. (1999) 'Perception, Object Kind, and Object Naming', *Spatial Cognition and Computation* 1(1): 1–29.

Landau, B. and Shipley, E. (2001) 'Labelling Patterns and Object Naming', *Developmental Science* 4(1): 109–18.

Landau, B., Smith, L. and Jones, S. (1998a) 'Object Perception and Object Naming in Early Development', *Trends in Cognitive Sciences* 2(1): 19–24.

Landau, B., Smith, L. and Jones, S. (1998b) 'Object Shape, Object Function, and Object Name', *Journal of Memory and Language* 38(1): 1–27.

Larsgaard, M.L. (1996) *Content-Based Searching of Large Image Databases*, http://www.uky.edu/~kiernan/DL/larsgaard.html

Larsgaard, M.L. and Carver, L. (1995) 'Accessing Spatial Data Online – Project Alexandria', *Information Technology and Libraries* 14(2): 93–7.

Levelt, W.J.M., Roelofs, A. and Meyer, A.S. (1999) 'A Theory of Lexical Access in Language Production', *Behavioral and Brain Sciences* 22: 1–38.

Liter, J.C. (1998) 'The Contribution of Qualitative and Quantitative Shape Features to Object Recognition across Change of View', *Memory & Cognition* 26(5): 1056–67.

Malinas, G. (1991) 'A Semantics for Pictures', *Canadian Journal of Philosophy* 21(3): 275–98.

Mannheim, K. (1952[1923]) *Essays on the Sociology of Knowledge*. London: Routledge & Kegan Paul.

Mannheim, K. (1997) *Essays on the Sociology of Culture*. London: Routledge & Kegan Paul.

Marr, D. (1982) *Vision*. San Francisco: WH Freeman.

Marr, D. and Nishihara, H.K. (1978) 'Representation and Recognition of the Spatial Organisation of Three-Dimensional Shapes', *Proceedings of the Royal Society of London Bulletin* 200: 269–94.

Merriman W.E., Scott, P.D. and Marazita, J. (1993) 'An Appearance Function Shift in Children's Object Naming', *Journal of Child Language* 20(1): 101–18.

Meyer, A.S., Sleiderink, A.M. and Levelt, W.J.M. (1998) 'Viewing and Naming Objects: Eye Movements during Noun Phrase Production', *Cognition* 66(2): B25-B33.

Meyer, A.S. and Van der Meulen, F.F. (2000) 'Phonological Priming Effects on Speech Onset Latencies and Viewing Times in Object Naming', *Psychonomic Bulletin and Review* 7(2): 314–19.

Mitchell, P., Davidoff, J. and Brown, C. (1996) 'Young Children's Ability to Process Object Colour: Coloured Pictogens and Verbal Mediation', *British Journal of Developmental Psychology* 14(3): 339–54.

Moore, C. and Cavanagh, P. (1998) 'Recovery of 3D Volume from 2-Tone Images of Novel Objects', *Cognition* 67: 45–71.

Natsoulas, T. (1999) 'Virtual Objects', *Journal of Mind and Behavior* 20(4): 357–78.

Neisser, U. (1967) *Cognitive Psychology*. New York: Appleton-Century-Crofts.

Ornager, S. (1997) 'Image Retrieval: Theoretical Analysis and Empirical User Studies on Accessing Information in Images', *ASIS '97: Proceedings of the 60th ASIS Annual Meeting*, Vol. 34. Silver Springs, MD: American Society for Information Science (ASIS).

Ostergaard, A.L. and Davidoff, J.B. (1985) 'Some Effects of Color on Naming and Recognition of Objects', *Journal of Experimental Psychology: Learning, Memory, and Cognition* 11(3): 579–87.

Panofsky, E. (1970[1955]) *Meaning in the Visual Arts*. London: Penguin.

Persson, P., Höök, K. and Simsarian, K. (2000) 'Human–Computer Interaction versus Reception Studies: Objectives, Methods and Ontologies', paper presented at NorFA Research Seminar on 'Reception: Film, TV, Digital Culture', Stockholm University, 4–7 June. [http://www.sics.se/~perp/AComparisonofHuman.doc 4th March 2002]

Price, C.J. and Humphreys, G.W. (1989) 'The Effects of Surface Detail on Object Categorization and Naming', *Quarterly Journal of Experimental Psychology: Human Experimental Psychology* 41(4-A): 797–827.

Rasmussen, E. (1997) 'Indexing Images', *Annual Review of Information Science and Technology* 32: 169–96.

Rollins, M. (1999) 'Pictorial Representation: When Cognitive Science Meets Aesthetics', *Philosophical Psychology* 12(4): 387–413.

Rosch, E. and Lloyd, B.B. (eds) (1978) *Cognition and Categorization*. London: Erlbaum.

Rosch, E., Mervis, C., Gray, W., Johnson, D. and Boyes-Braem, P. (1976) 'Basic Objects in Natural Categories', *Cognitive Psychology* 8(3): 382–439.

Santini, S. and Jain, R. (1996) 'The Graphical Specification of Similarity Queries', *Journal of Visual Languages and Computing* 7: 403–21.

Sharrock, W. and Coulter, J. (1998) 'On J.J. Gibson: A Response to Our Commentators', *Theory & Psychology* 8(2): 177–81.

Sharrock, W. and Coulter, J. (1999) 'Elucidation vs. Pseudo-Explanation in Psychology: A Response to Professor Wetherick', *Theory & Psychology* 9(4): 557–63.

Shatford Layne, S. (1994) 'Some Issues in the Indexing of Images', *Journal of the American Society for Information Science* 45(8): 583–8.

Siddiqi, K., Kimia, B.B., Tannenbaum, A. and Zucker, S.W. (2001) 'On the Psychophysics of the Shape Triangle', *Vision Research* 41(9): 1153–78.

Siddiqi, K., Tresness, K. and Kimia, B.B. (1996) 'Parts of Visual Form: Psychophysical Aspects', *Perception* 25: 399–424.

Slater, A. (1998) 'Visual Organization and Perceptual Constancies in Early Infancy', in V. Walsh and J. Kulikowski (eds) *Perceptual Constancy: Why Things Look as They Do*. Cambridge: Cambridge University Press.

Smith, B. (2000) ‘Truth and the Visual Field’, in J. Petitot, F.J. Varela, B. Pachoud and J-M. Roy (eds) *Naturalizing Phenomenology: Issues in Contemporary Phenomenology and Cognitive Science*. Stanford, CA: Stanford University Press.

Tagare, H.D., Jaffe, C.C. and Duncan, J. (1997) ‘Medical Image Databases: A Content Based Retrieval Approach’, *Journal of the American Medical Informatics Association* 4(3): 184–98.

Takano, Y. (1989) ‘Perception of Rotated Forms – A Theory of Information Types’, *Cognitive Psychology* 21(1): 1–59.

Tanaka, J.W. and Taylor, M. (1991) ‘Object Categories and Expertise: Is the Basic Level in the Eye of the Beholder?’, *Cognitive Psychology* 23(3): 457–82.

Tarr, M.J. and Bölthoff, H.H. (1998) ‘Image-Based Object Recognition in Man, Monkey and Machine’, *Cognition* 67: 1–20.

Taylor, J.G. (1962) *The Behavioural Basis of Perception*. London: Yale University Press.

Wetherick, N.E. (1999) ‘On What Gibson (and the Rest of Us) Cannot Do Without: Comment on Sharrock and Coulter (1998)’, *Theory & Psychology* 9(4): 551–6.

Williams, L.R. and Hanson, A.R. (1996) ‘Perceptual Completion of Occluded Surfaces’, *Computer Vision and Image Understanding* 64(1): 1–20.

BIOGRAPHICAL NOTES

BRYAN BURFORD is a Research Associate in the Institute for Image Data Research (IIDR). With a background in psychology and ergonomics, since 1993 he has worked on many research projects at the University of Northumbria. These projects have examined a number of theoretical and practical issues in human–computer interaction, including the origins of users’ self-confidence and their trust in technology. He is currently working on the user-centred development of interface specifications for image databases.

Address: Institute for Image Data Research, University of Northumbria, Newcastle upon Tyne, NE1 8ST, UK. [email: b.burford@unn.ac.uk]

PAM BRIGGS is a Research Professor in Psychology at the University of Northumbria and a founder member of the Institute for Image Data Research. Since gaining her PhD in visual word recognition 20 years ago, she has led a number of government- and industry-sponsored research projects on cognitive and perceptual aspects of human–computer communication. She is the author of numerous publications on cognitive science and human computer interaction.

Address: as Bryan Burford. [email: p.briggs@unn.ac.uk]

JOHN P. EAKINS, Professor of Computing and Director of the Institute for Image Data Research at Northumbria University, has over 10 years’ experience of research into storage and retrieval techniques for drawings and images. He is the author of over 20 papers on the subject, including commissioned reviews of the state of the art in content-based image retrieval (CBIR) and trademark retrieval. He has given many invited presentations on the topic of image retrieval, most recently at the third European Summer School in Information Retrieval, held in Varenna, Italy, in September 2000. He has for three years been co-chair of the Challenge of Image Retrieval conferences set up to promote exchange between researchers and practitioners in this field.

Address: as Bryan Burford. [email: john.eakins@unn.ac.uk)