Eliciting Perceptual Ground Truth for Image Segmentation.

Victoria Hodge, John Eakins and Jim Austin.
Department of Computer Science,
University of York
York, UK

Abstract

In this paper, we investigate human visual perception and establish a body of ground truth data elicited from human visual studies. We aim to build on the formative work of Ren, Eakins and Briggs who produced an initial ground truth database. Human subjects were asked to draw and rank their perceptions of the parts of a series of figurative images. These rankings were then used to score the perceptions, identify the preferred human breakdowns and thus allow us to induce perceptual rules for human decomposition of figurative images. The results suggest that the human breakdowns follow well-known perceptual principles in particular the Gestalt laws.

1 Introduction

We hypothesise that perception and thus segmentation varies from person to person and also varies with the domain of application (context). This subjectivity is almost inevitable due to culture, education, expectation, domain of application, mood, age etc. but there must be a core set of commonalities across human judgements that we aim to distil out. There is currently no comprehensive theory of human or computational image and shape segmentation.

Our work forms part of the PROFI (Perceptually-Relevant Retrieval of Figurative Images) project[1]. In PROFI, we aim to develop new techniques for the retrieval of figurative images (i.e. abstract trademarks and logos) from large databases. The techniques will be based on the extraction of perceptually relevant shape features and the matching of these features in the target image against features in the stored images, thereby overcoming many of the limitations of existing methods. This project aims to develop and evaluate new algorithms for:

1. Perceptual segmentation of raw images, and grouping of shape elements.
2. Matching of geometrical patterns representing shape features.
3. Partial matching: fitting part of one shape with part of another.
4. Indexing shape features in huge databases of figurative images.
5. Indexing the relative spatial layout of shape features within these images.

In this paper we focus on task 1.

Existing systems, for example trademark search systems, attempt to match a target against stored images such as those shown in Figs. 1-3 in one of two ways: (a) comparing features generated from the images as a whole, or (b) matching features from individual parts of the images [E01].

[1] PROFI web page: http://www.cs.uu.nl/profi/

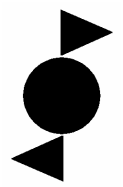


Fig. 1

Fig. 2

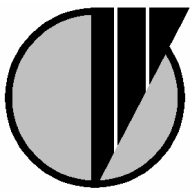


Fig. 3

The principal difficulty in matching by parts is the selection of parts that accurately reflect the image's appearance to a human observer. In Fig. 1 this is reasonably clear (2 triangles and a circle). But in Fig. 2, should the central bars be matched as six individual components, or as two groups of three? And in Fig. 3, should matching be based on a circle and a triangle - neither of which are actually present in the image itself? These are the questions which this current research aims to answer.

For present purposes, therefore, we are primarily interested in clarifying two aspects of human segmentation behaviour: the formation of intermediate-level groupings of image parts; and, the generation of perceived elements not explicitly present in the original image. Our hypothesis is that these will allow us to identify the most salient image elements for matching more accurately than has hitherto been possible.

The seminal paper describing image decomposition for this aspect of the PROFI project is Ren et al. [REB00]. The paper evaluates how human subjects segment trademark images into their perceived constituent parts. The subjects initially break down trademark images into a set of components in as many ways as they see fit. These breakdowns are then fed into the second part of the experiment where subjects rank the breakdowns from part 1 by their perceived likelihood. The paper's main discoveries are that humans partition trademark images into disjoint regions most commonly, then into overlapping or nested regions and partition into separate line segments or groups least commonly. The breakdowns generated are similar to the breakdowns obtained by applying the Gestalt principles [W23], [K63], [K79], [G72] of human perceptual organisation. The authors [REB00] posit that perceptual line grouping, closed-region identification, texture processing, identifying familiar shapes (such as triangles, squares etc.) and uncovering 'hidden' image features (such as figure-ground reversal) are areas requiring further investigation. We aim to augment and complement these results in the current paper and use the results in our development of a computerized image retrieval system.

Dyson & Box [DB97] evaluated how humans sub-divide shapes by providing 3 palettes of symbols. The human subjects selected the border, main shape and any number of other shapes that they perceived to be present in target images. The subjects were permitted to select a single border, single main shape but as many other shapes as they saw fit. These shape descriptions were then fed into the database system and any stored matches retrieved. The conclusion to be distilled from their investigation is that less is more. If the subject described a shape with too many 'other shapes' then in subsequent match tests, too many results will be retrieved. The granularity of human shape descriptions also varies widely, for example, line and triangle versus arrow. People use different terms to describe the same object, for example dot vs. circle, square vs. rectangle. We may conclude that a core set of shapes must be stored in the palette that may not be subdivided or subsumed by other

shapes in the palette; for example, arrow may be subdivided into triangle and line, square is subsumed by rectangle etc. This will prevent ambiguity and prevent over-description.

Mojsilović et al. [MGR02] posit that human vision is a hierarchical process where vision initially detects the edges in an image and breaks the image into primitives (lines, bars, crossings or blobs). These primitives are then grouped by perceptual significance into chains, curves, clusters, regions, or are grouped into built-in geometric elements (circles, squares or ellipses). The primitives are arranged further using clustering, connectivity, symmetry, parallelism, similarity matching and "textureness" to permit figure/ground separation. They accomplish this by firstly performing edge-detection followed by texture segmentation, colour segmentation and foreground/background separation. This divides the image into "meaningful regions". Each region is labelled with its size, position, neighbours, boundary, boundary curvature, texture, elementary shape (boundary, eccentricity, moments and symmetry features), mean colour and colour name. Finally, the labelled regions are analysed and combined. We posit that: "*the power of a system stems from the combination technique and method*".

Biederman et al. [BSBKF99] also propose that human image recognition works on various levels and that an agglomerative technique is used. The most primitive 'basic' level allows images to be named, e.g. chair, elephant, and kettle. The next layer up, 'subordinate' layer, allows, e.g., African elephants to be distinguished from Asian elephants. Their analyses suggest that these two levels employ geon structure descriptions, i.e., the decomposition of the images into components. These geons have qualitative ('non-accidental') properties and relations that allow images to be matched. The authors determine experimentally that qualitative properties have more influence on object matching (whether two images are deemed similar) than quantitative. This agrees with the findings of Ferguson et al. [FAG96] (described later) regarding Gestalt symmetry. Only when there are large differences in quantitative properties are they used. The authors [BSBKF99] go on to posit a hierarchical architecture for image decomposition into geons and their properties and relations. The hierarchy is similar to that of Mojsilović et al. described above. The lowest layer represents edges, the second layer: vertices, axes and blobs (all conjunctions of edges), the third layer: properties of geons, the fourth and fifth layers: relations between geons, the sixth layer: a conjunction of the geon, its properties and its relations to other geons and layer seven: objects within the image (conjunctions of geons). The paper does not describe how to obtain the second layer which is the critical layer; it assumes that the vertices, axes and blobs are provided.

Further support for the hypothesis of human image segmentation stems from Jain & Vailaya [JV98] who propose that humans use semantics during shape matching and that semantically similar images may be visually very different. They posit that an automated method needs to extract salient features from the image and to perceptually group features and elements. Jain & Vailaya's technique struggles to find bull's head shapes as some are line-based and others filled-in. By filling in all shapes to allow generalisation and remove unnecessary detail they improve the technique's recall accuracy. However, they feel that this loses information from within the image (within the holes). Hence, further improvements would result from using image segments for matching rather than a generalised outline. This would necessitate a

robust and accurate segmentation algorithm. This could be further extended to allow local feature matching.

Vecera, Behrmann et al. [VBF01] investigated the role of attention and image parts and posited that their findings unite with theories of object recognition that suggest that objects are decomposed into parts prior to recognition. Baker, Olson & Behrmann [BOB04] investigated the role of attention and perceptual grouping and identified that connectedness - one of the strongest cues for visual grouping - and attention both affect statistical learning. Zemel et al. [ZBMB02] posit that grouping principles, familiarity and task instructions all effect object attention and they provide empirical evidence and citations to support these. Through empirical investigation, they also identified that attention benefits are achieved for newly learned unfamiliar objects. They propose that recent experience determines the perception of occluded shapes. A framework is desired that allows for the rapid formation of novel objects and permits their influence on perceptual organisation. They note that their results tie in more with the Brunswick school of perception which favours the influence of statistically learned rules more than purely stimulus-driven Gestalt principles. However, these two approaches are not dichotomous and even Wertheimer posited that experience modulates perceptual grouping.

In current computational approaches, shapes may be segmented using either the shape's boundary or the shape's interior (fill area) but rarely both compared to the holistic viewpoint used by humans. Humans are posited to decompose shapes using Gestalt principles where symmetry, complexity, structure and deformation are all important along with conceptual (semantic) information. However, humans struggle with orientation and tend to regard similar shapes with differing orientations as more dissimilar then slightly dissimilar shapes with the same orientation. Orientation is less problematic for computational methods than humans.

Previous work on human segmentation analyses includes Hoffman & Richards [HR84] whose work was based on psychophysical observations and the notion that concavities arise when two convex parts are joined. They hence posit the minima rule for image decomposition – divide the surface into parts at loci of negative minima of each principal curvature along its associated family of lines of curvature. They subdivide shapes using only the contours and not the shapes' interiors and the approach does not always produce intuitive results [R93]. Hoffman & Richards identify open questions such as what qualitative and metrical descriptions should be applied to these parts? How are the partitioning contours to be identified for 2-D images? What spatial relations need to be identified?

Other authors have investigated computational methods to mimic human image segmentation. Much research from the computational geometry field has focussed on decomposing polygons into sub-shapes from a palette of shapes (such as triangle, convex, spiral or star-shaped). However, these often do not match the decompositions extracted from human segmentation evaluations.

Work on general image decomposition has built upon the formative work of Hoffman & Richards [HR84] described above and includes Siddiqi & Kimia [SK95] who examined psychophysical and ecological factors and proposed that shapes are segmented using limbs and necks. A limb is "a part-line going through a pair of

negative curvature minima with co-circular boundary tangents on at least one side of the part-line". A neck is "a part-line which is a local minimum of the diameter of an inscribed circle". Singh et al. [SSH99] refute this proposal by providing counter-example images where the proposed breakdown approach would fail for both limbs and necks.

Singh et al. [SSH99] propose a similar technique – short-cut rule - that uses minimum distance and skeletal axes to determine segmentation lines between boundaries where at least one boundary is a concave vertex. They augment their proposal with human experiments on crosses (+) and L-shapes that appear to validate it. It builds on the seminal approach of Hoffman & Richards [HR84] which could identify boundary points for cuts but not the actual cuts. Singh et al. posit that all things being equal, humans prefer to use the shortest cuts to segment shapes. Their approach can also identify cuts that are not necessarily between the local minima points of concave vertices but are in fact, between the most human-oriented cut points.

The shapes in the paper are all very simple with usually only a single cut point or one ambiguous cut point. Rosin [R00] also criticises the technique as it relies solely on boundary information and uses very little global shape information. Singh et al. [SSH99] also use an arbitrary choice of skeletal axis (smoothed local symmetries) with no justification provided. Rosin also provides counter-example images where perceptually relevant cuts need not cross an axis but may on occasions follow an axis and where the most perceptually relevant cut is not the shortest. The approach does not incorporate many Gestalt principles. Singh et al. [SSH99] propose further investigation regarding local symmetry, good continuity (w.r.t. boundary), segmentations that yield fewer segments, and for some shapes: no segmentations and the orientation of the whole shape. This approach seems more generic than Siddiqi & Kimia [SK95] but is only demonstrated on homogeneously shaded shapes. Gestalt principles are intuitively complex and do not operate in isolation. Adding a texture to the shapes, for example, would surely affect where a human perceived the segmentation lines but this is not investigated; all shapes are homogeneously shaded.

Rosin [R00] evaluates various techniques such as Siddiqi & Kimia [SK95], Singh et al. [SSH99] and concludes that the best approach is to use convexity augmented with saliency factors such as good continuity of cuts with boundaries, cut length, size of segmented regions. However, combining these methods is difficult. There also needs to be a stopping criterion that determines when a shape has been segmented sufficiently and also the possibility of generating arc cuts rather than purely straight line cuts.

Tanase & Veltkamp [TV02a, TV02b] propose a segmentation approach using straight-line skeletons. The process comprises two stages: the shape is decomposed into non-overlapping segments using the skeletal bifurcation points. The boundaries of these segments are then simplified and protrusions removed in the second stage. This two-stage process overcomes some of the limitations posited by Rosin [R00]. Removing the protrusions should implement a degree of good continuity. The approach also has an autonomous termination point.

Carlin [C01] includes skeleton features along with geometric moments, Legendre moments, invariant moments, Fourier descriptors, fuzzy and symmetry descriptors

and a mixed feature set in his paper assessing the relative merits of each approach for shape similarity matching. The paper notes that skeleton features perform well on application specific criteria but are not robust to shape deformation.

In the introduction we noted that in current computational approaches, shapes may be segmented using either the shape's boundary or the shape's interior (fill area) but rarely both compared to the holistic viewpoint used by humans. [LC02] aim to bridge this gap by unifying skeletons and edge detection approaches. The system uses very simple shape primitives and integrates edge detection and skeleton extraction to match trademarks. Initially, it segments the image into regions using the pixel connectivity. For each region, the system then either performs edge detection or performs thinning. The authors posit that: "it is advantageous to use different methods under different situations". They note that for a solid region where the shape conveys much visual information, edge detection is preferable to thinning as it extracts the contour of the region. However, for a region containing curves, thinning is preferable as it extracts the skeleton and "produces a better representation". The system determines whether edge detection or thinning is preferable for a particular region by examining the distribution of the distances between each pixel of the skeleton and the nearest pixel of the contour. If the distance from the skeleton to the nearest contour pixel is small and if this distance remains relatively constant for different skeleton points then the system performs thinning to extract the skeleton. If there is a large variation in the distances from skeleton pixels to the nearest contour pixels, then the system performs edge detection.

Once the system has calculated the skeleton or contour, the system traces the strokes by following the pixel connectivity and extracts features from each stroke which the system uses to classify each stroke as either: line, circle or polygon by assigning a confidence measure (between 0 and 1) for each type. Trademarks are matched by calculating the correspondence between strokes using the spatial order and feature distances. The similarity between trademarks is thus a sum of the stroke matches and the spatial relation similarity between them minus the cost of unmatched strokes.

Humans are posited to decompose images along Gestalt principles. There has been widespread investigation including human experimentation of individual Gestalt principles [W23], [K63], [K79] & [G72]. However, most authors have investigated one principle in isolation. For example, Desolneux et al. [DMM04] have theoretically investigated a wide range of Gestalt principles and derived formulae for many using the Helmholtz principle – *a geometrically meaningful event is an event that, according to probability estimates, should not happen by chance, which therefore implies it is deliberate and meaningful*. They also note that multiplicity suggests that a Gestalt principle can only be active in an image if its application would not create a huge number of arrangements (segmentations). However, they [DMM04] posit that the main challenge remaining is to combine several partial Gestalts (arrangements using one Gestalt principle) and arrive at the point where Gestaltists stopped, namely: *identifying collaborations involving multiple partial Gestalts and resolving any conflicts between the collaborating Gestalts*. This is a deep problem related to neuro-physiological binding. Even further, rules governing the bottom-up construction of principles may be found.

The evaluations we have found are also mainly focussed on recognition and detection experiments rather than how human shape decomposition is affected by the principle. The following focuses on figure/ground separation, symmetry and texture as these were outlined above and in the PROFI proposal as areas requiring further investigation within human segmentation experiments. We have also identified singularity as an area for further investigation although we have not found any specific papers relating to this area in the literature beyond the formative work of Goldmeier [G72]. Goldmeier noted that the singular values (such as symmetry, parallelism, horizontality, perpendicularity, recti-linearity or other regularities) which are most strongly realised have the most effect on similarity. He also posited that the similarity of two images depends on the agreement of their singular phenomena. With respect to singularities, he posited that

- Two spatial directions are most pertinent: vertical and horizontal
- Among the distinguishable features of parts of an image are those which are determined to some degree by the orientation to the vertical/horizontal axes. Some of these features are so important that the language has words for them, e.g., base, top etc.
- These two principal spatial directions are not equivalent. Vertical separates phenomenally equivalent domains, (the two sides), whereas horizontal separates phenomenally non-equivalent domains, (up. down, top-bottom).
- Many figures when viewed as wholes, have preferred, distinguished or singular positions.

Driver & Baylis's [DB95] empirical analyses led them to conclude that figure/ground assignment results in a description of the figural part of an image (as distinct from the background) as a set of convex components. The background is never sub-divided this way, which explains why subjects are able to distinguish the two relatively easily in most cases. The authors refute counter-findings by suggesting that the subjects were primed as to what they were looking for after a small number of trials. They go on to suggest that figure/ground assignment is determined agglomeratively, i.e., by image segmentation factors. However, where this leads to an ambiguity, it is resolved top-down by the strategic allocation of attention. Paradoxically, visual attention is directed not at the dividing edges between the image components but at the entire figure.

Ferguson et al. [FAG96] has evaluated human symmetry classification. Subjects were asked to classify shapes as symmetric/asymmetric. The authors noted that humans classified shapes with both concavity and number of vertices differences more easily than just number of vertices differences and number of vertices differences more easily than concavity. This qualitative versus quantitative preference agrees with the findings of Biederman et al. [BSBKF99] (*regarding pairwise matching of images, e.g., when matching pairs of images of goblets or pairs of bottle images*). Symmetric figures were classified more accurately than asymmetric figure throughout the experiments. We know from this that humans can perceive symmetry and what forms of asymmetry are most significant. We know from previous work that vertical symmetry is more perceptually relevant than horizontal or oblique symmetry. We could therefore extend this experiment by investigating the segmentation of images when components within the image are symmetric and when the same components are made asymmetric focussing on the vertical plane (qualitatively asymmetric such as concave or different numbers of vertices).

Palmer [P85] systematically investigates symmetry using: squares and diamonds; + and x shapes; and diagonal and horizontal/vertical configurations of these shapes in conjunction with textures and bounding boxes (rectangular frames). Human subjects have most difficulty perceiving shapes when the symmetries of the shapes and their configurations or boundary frames are inconsistent, e.g. squares in diagonal arrangements, diamonds in horizontal/vertical arrangements, diamond frame around square or square frame around diamond. Textures that are inconsistent increase reaction times most noticeably when the texture stripes are widest. The orientation of the target, the orientation of the visible context and the gravitational orientation of the environment all affect symmetry perception. Again, we hypothesise that other perceptual factors could interact. The authors note that factors such as the relative contrast or spatial frequency need investigation and their effects quantifying.

Payne et al. [PHS00, PS01] note that texture is more important than colour for human classification and that texture is easy to recognise but hard to define. The main perceptually relevant factors of human texture recognition identified by researchers are: repetitiveness, coarseness, directionality, complexity and contrast. IBM's QBIC [QBIC] image retrieval system performs texture-based retrieval by calculating features of coarseness, contrast and directionality on grey-scale images (colour images are converted to grey-scale). Images may then be compared using vector distance calculations using weighted Euclidean distance in this 3-D space. A similar approach used in the Photobook [Photo] system is the Wold texture model where textures are represented by repetitiveness, directionality and randomness. Textures may then be compared using a single or linear combination of distance metrics such as Euclidean, Mahalanobis etc.

Human texture investigations have generally focused on identifying patches of contrasting texture within textured backgrounds. We have not found any experiments that investigate the effects of textures on shape segmentation. We know from Weigle's [WELTEH00] experiments with jittered dashes that humans can identify textures best that differ in orientation by more than 15 degrees. Humans also recognise textures well when the background is vertical or horizontal but paradoxically humans do not perform well when the target (superimposed on the background) is vertical or horizontal. Nothdurft has investigated the effects of texture form and texture spacing. Desolneux et al. [DMM04] have investigated the perceived visibility of noisy squares on noisy backgrounds. We can use these factors within our segmentation experiments where texture is present or we can replace homogeneous shading with textures obeying the above rules to investigate the affect of texture on human segmentation. In many of these documented experiments there are no control conditions. For example, in Desolneux et al. [DMM04], subjects are asked if they can see a square which obviously focuses their attention. They do not state whether there are any images with no squares as control samples.

The papers cited above agree that humans decompose images into segments. The authors generally accept that this decomposition is performed in line with the Gestalt principles and semantics (which closely relates to Gestalt principles such as familiarity and goodness) although other statistical factors such as experience, mood or culture may influence this. However, these Gestalt principles and statistical factors are not counter-intuitive and may work in tandem. The papers differ as to the final

image units produced from their computational decomposition. Some authors posit geometrically defined parts such as Hoffman & Richard's [HR84] convex parts, Siddiqi & Kimia's [SK95] necks and limbs through to Singh et al.'s [SSH99] and Tanase & Veltkamp's [TV02a, TV02b] skeletons whereas others posit specific shapes from a set of shape primitives such as Dyson & Box's [DB97] palette or Biederman et al.'s [BSBFK99] geons. Some authors have gone on to propose overall architectures for image matching systems which are hierarchical and fit the decomposition approach but fuse their image units agglomeratively. The architectures start from the most primitive elements such as edges and incrementally build increasingly more complex image units from the units below in the hierarchy.

Our approach needs to accommodate human variance. Image perception and hence decomposition varies from human to human. We need to produce a consensus approach that fits most cases and scenarios but not necessarily all instance. That is, we must produce the best compromise system. We need to incorporate the findings from research into the individual Gestalt principles and merge this with findings from our decomposition analyses and previous decomposition analyses such as the fact that qualitative differences overshadow quantitative differences when humans match images or that humans partition into disjoint regions primarily, overlapping regions secondly and separate line groups least often. We need to use our analyses to find the most promising decomposition or decompositions for a broad range of images using Gestalt principles to drive the process for work package 3 and train the system to handle these. We need to carefully analyse our representations a priori. What qualitative and quantitative units should describe image parts? What qualitative and quantitative relations between parts should be used? Should we provide a set of image primitives or familiar shapes that all images are constructed from? We need to ensure that our technique will not produce too many decompositions for a particular image as multiplicity implies that a Gestalt factor can only be active if it does not produce too many decompositions. The decompositions we optimise should also be meaningful and not produce chance decompositions. We should ultimately look to combine Gestalt factors within a principled methodology and permit interactions and conflict resolution.

In the remainder of this report we detail the development and implementation of the experimental methodology and provide some analysis. In the appendices we provide the results of the human analysis experiments as a set of images each with a list of the preferred breakdowns and the preference score for each breakdown.

2 Methodology

The experimental methodology was developed in conjunction with the Psychology Department at the University of York, UK who advised on methodology, ethical considerations and best practice and also provided general advice and guidance.

The central premise for the investigations in this paper is to identify how humans decompose images, the degree of commonality across a range of human subjects and to provide a set of ground truth images. These ground truth images may be further analysed to elicit statistics and preference scores regarding the decomposition preferences of humans: i.e., which decomposition is generally preferred for each image, a ranked order of decompositions for each image, how many potential decompositions there should be for each image. We aim to investigate symmetry,

texture, singularities and also to some extent the effect of figure/ground phenomena. We aim to use the results from our experimental analyses to drive the formation of an integrated computational system that mimics human segmentation. We need to ensure that our resultant computerised technique will not produce too many decompositions for a particular image. The decompositions we optimise should also be meaningful and not produce chance decompositions.

We performed an initial pilot study to allow us to select useful images and to revise and improve the experimental methodology.

A set of trademark and other figurative images was presented to University of York staff, students and their relatives and friends. Each subject received a printed booklet containing 17 pages: a front sheet and 16 pages with 2 images per page in 2 columns giving 32 images in total in each booklet. The subjects also received a copy of the experiment instructions. The subjects were requested to draw (using pen or pencil) their perceived decompositions of each image in turn on to the booklet and to rank each decomposition (1^{st}, $2^{nd,}$ 3^{rd} etc.) according to the order in which they perceived that decomposition. All completed booklets were anonymized and labelled with a subject ID number. All subjects who completed the experiment were entered into a prize draw where the prizes were a £200, £50 and 5 x £10 shopping vouchers. The statistics of the subjects from experiment 1 and experiment 2 are:

- Age range 14 – 70
- Gender mixed
- Nationality mixed international

2.1 Images

Each image was 4.5 cm high although the size of the drawn image varied slightly according to the amount of white space surrounding it. All images were monochrome TIFF images.

2.1.1 Methodology

There were three sets of 32 images. Each set contained some images present in the other sets to act as controls and thus to verify that the subjects in each group are statistically similar. The trademarks were in pairs (14 pairs in each set, p_1 .. p_{14}) along with 4 other images (i1 .. i4). The unpaired images are supplementary control images (i1, i2) and buffer images (i3, i4) in case the subjects do not complete the exercise. The paired images were ordered p_1^1, p_2^1, p_3^1, … p_{14}^1, i1, i2, p_1^2, p_2^2, p_3^2, ... p_{14}^2, i3, i4. The subjects received the first image of a pair and then later, a second paired image: the same image but altered according to symmetry, texture or singularity principles. We note that it is extremely difficult to isolate Gestalt principles within the trademark images. For example, altering an image along symmetrical lines will inevitably alter other Gestalt properties such as familiarity, continuity or perhaps grouping. We attempted to provide as wide a variety of symmetry, texture or singularity alterations as possible. These 3 sets of images were further divided into forward and backward sets giving 6 sets in total (A-Forward, A-Reverse, B-Forward, B-Reverse, C-Forward and C-Reverse). The forward and reverse sets have the order of the images reversed to prevent order bias where the order of image presentation affects the perception:

- Forward - **p_1^1, p_2^1, p_3^1, … p_{14}^1, i1, i2, p_1^2, p_2^2, p_3^2, ... p_{14}^2, i3, i4** and then
- Reverse - **p_{14}^2, p_{13}^2, p_{12}^2, ... p_1^2, i1, i2, p_{14}^1, p_{13}^1, p_{12}^1, ... p_1^1, i3, i4**.

If all subjects receive p_1^1 before p_1^2 then this may influence their perception of p_1^2.

The images are listed in Appendix B and the sets are:
A-Forward: images 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32.
A-Reverse: images 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1, 31, 32.
B-Forward: images 33, 2, 34, 35, 36, 11, 37, 1, 38, 39, 40, 41, 42, 43, 15, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60.
B-Reverse: images 58, 57, 56, 55, 54, 53, 52, 51, 50, 49, 48, 47, 46, 45, 44, 15, 43, 42, 41, 40, 39, 38, 1, 37, 11, 36, 35, 34, 2, 33, 59, 60.
C-Forward: images 61, 62, 33, 63, 2, 36, 64, 65, 66, 10, 67, 8, 68, 11, 15, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 41, 79, 80, 81, 82, 83, 84.
C-Reverse: images 82, 81, 80, 79, 41, 78, 77, 76, 75, 74, 73, 72, 71, 70, 69, 15, 11, 68, 8, 67, 10, 66, 65, 64, 36, 2, 63, 33, 62, 61, 83, 84.

2.1.2 Experiment 1

Experiment 1 involved the first 28 subjects who were thus effectively a test group to allow fine-tuning although their results were used in the final analysis. 25 subjects were presented with a booklet of one set of 32 images (1 from the 6 sets described above) and 3 subjects were presented with 3 booklets (3 sets) (84 images in total when repetitions are excluded) and allowed to draw their perceptions unsupervised with their initial perception first and any other perceptions in the order that they perceived them. The results from this study were used in the final analysis but were also used to improve and fine-tune the experimental instructions. We note that some subjects (13 of the 28 who completed a single booklet) only drew one decomposition per image. As a result, we revised the instructions of the subsequent experiment 2 as we felt some of them may have misunderstood the instructions. However, we note from feedback from the subjects, that not everyone is able to see more than one breakdown per image so not all of these 13 subjects had necessarily misunderstood the instructions.

2.1.3 Experiment 2

The final analysis involved 25 staff and students drawn from across the University. They were invited to a series of four 1-hour sessions spread across 27th June 2005 starting at noon with the final session at 3 pm. The sessions were supervised. Each subject received one printed booklet of 32 images (1 from the 6 sets described above) and was invited to draw their perceptions of each image, drawing their initial perception first and any other perceptions in the order that they perceived them. The subjects from experiments 1 and 2 were entered into a prize draw to win shopping vouchers.

2.2 Overview

Of the 6 sets of images: 10 people analysed set A-Forward, 11 people analysed set A-Reverse, 9 people analysed set B-Forward, 8 people analysed set B-Reverse, 11 people analysed set C-Forward and 9 people analysed set C-Reverse.

The first stage of analysing the images was to collate the breakdowns drawn by the subjects and to note the rank. Each image had a list of the breakdowns perceived.

Each breakdown had a list of the ID of the subjects who perceived that breakdown and the rank they awarded it (1^{st}, 2^{nd} etc.). For each image, if two subjects had drawn identical or extremely similar breakdowns then the breakdowns were marked as the same and the subjects' IDs and the rank they awarded the breakdown added to the list for that specific breakdown. Otherwise, the breakdowns were marked as two separate breakdowns and the subjects' IDs and ranks added to the respective breakdowns' lists. The output from this analysis is a listing of all breakdowns for each image in turn along with a listing of all subjects who drew that breakdown and the rank that each subject gave it (1^{st}, 2^{nd}, 3^{rd} etc.)

2.3 Preference Scoring Mechanism

Ren et al. [REB00] used a slightly different experimental methodology compared to us. We aggregated their two-stage process into a single stage: Ren et al. used volunteers to elicit the breakdowns in stage 1 and then used a second set of volunteers to rank the breakdowns in stage 2. We conflated this into a single stage as we had difficulty recruiting volunteers at University of York due to expectations of payment which is the norm at the University and no funds were available within the budget. This conflation was in full agreement with the recommendations from the psychology advisors. This also required a slightly different scoring mechanism compared to that used by Ren et al..

For the vast majority of the images (74 of the 84), the subjects who drew that image drew 1, 2 or 3 breakdowns each so we used this number of breakdowns to devise our scoring mechanism. 10 images had a maximum number of breakdowns of 4 or 5; 2 subjects drew most of these 4 or 5 breakdowns per image with another 3 subjects drawing 4 breakdowns per image once each. Therefore, for all images we scored 3, 2, 1, ½ and ¼ for ranks 1 to 5 respectively.

For each breakdown the scores were totalled and divided by the total of the scores across all breakdowns for that image. This gives the preference score for each breakdown of each image.

The listing is given in Appendix B where, for a selection of the 84 images (those images discussed in section 4), we list each breakdown drawn by two or more subjects coupled with the breakdown's preference score. The images numbers are to allow the authors to cross-reference the images and the images are not listed in numerical order but rather arranged so that image pairs are listed together. The breakdowns are drawn in the order we analysed them and are not sorted in any way. Note that the breakdown numbers are again to allow the authors to cross-reference the breakdowns and are not significant although they do provide some notion of the number of breakdowns seen by one subject only (i.e., the omitted numbers from the list are the singular breakdowns). The breakdowns seen by only one person are not listed as there were simply too many. Where the individual components are difficult to distinguish, we have added red crosses to the diagram to allow the individual components or groups of components to be identified.

3 Results

Results from the analysis of the perceptions derived from the various sets of subjects indicate that the number of breakdowns drawn by the subjects varies quite widely

from image to image as shown in Figure 1. If the number of human breakdowns is large then the search space required for any computerised shape decomposition system will be large to allow an identical decomposition to be created by the computerised system. The search space will also be large for a computerised system matching components from one image against components in other stored images due to the large potential search space.

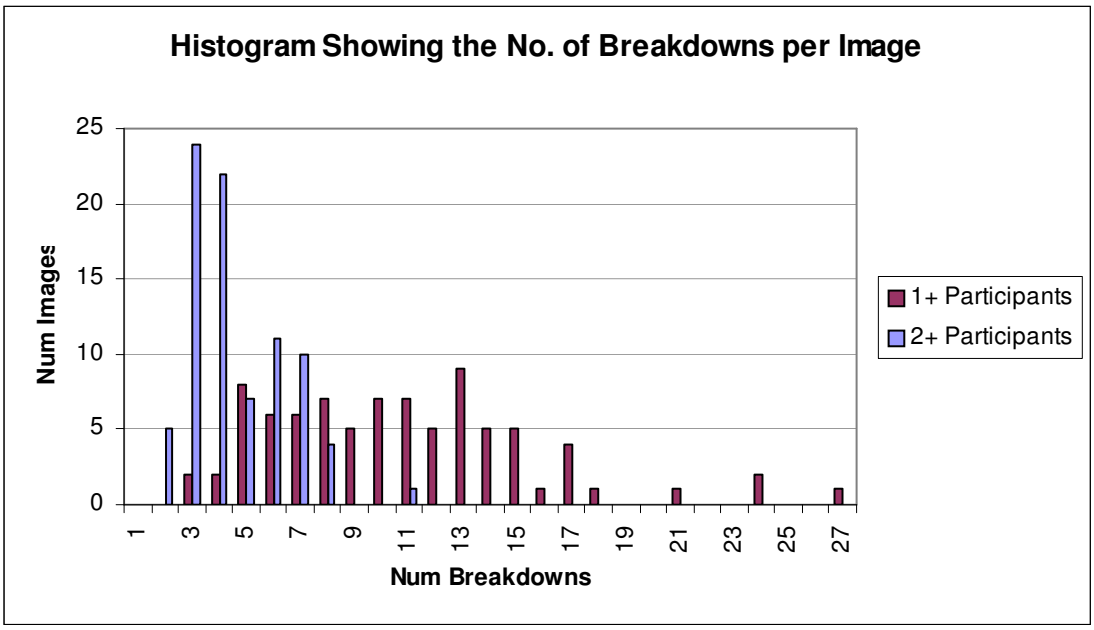


Figure 1. Graph showing the distribution of the number of breakdowns (seen by at least 1 subject and seen by at least 2 subjects) for each of the 84 images.

Another factor that we would expect to affect the number of breakdowns is the number of degrees of freedom available within the image. Images 1, 2, 7, 23, 33, 36 & 72 all produced at least 17 breakdowns seen by at least one subject and each of these images has a large number of potential components and a large number of possible arrangements of components. The search space for a computerized decomposition system or image component matching system processing these images would be large.

The graph in Figure 1 shows that the number of breakdowns seen by 2 or more subjects is much more closely grouped than the number of breakdowns perceived by 1 or more subjects with between 2 and 11 breakdowns perceived by 2 or more people. The mode value is 3 and only one image had more than 8 breakdowns perceived (image 25).

Ren at al. [REB00] had between 1 and 4 breakdowns for each image in their analyses. We found the unrestricted breakdown generation that we allowed the subjects coupled with consolidating Ren et al.'s two-stage process into a single stage allowed more scope for subject variation.

4 Analyses

While the limited number of results makes it impossible to perform any detailed quantitative analysis, qualitative analysis of individual results yields a number of insights which we expect to prove useful in subsequent phases of the project. In the following, we analyse the core set of breakdowns for each image seen by 2 or more subjects. We note that there are often multiple Gestalt differences between the images and their analogues as it is almost impossible to alter one Gestalt rule without affecting others. During our analyses, we try to focus on the main Gestalt change in each image though we acknowledge that this is a subjective process.

From analysing the subjects' drawings, we noted that the subjects may be focussed purely on eliciting the component breakdowns of each image. We feel they may concentrate on the individual components and do not always see the "larger picture". For example, where 6 triangles are arranged in a hexagonal shape many subjects drew 6 triangles but not the overall hexagonal shape. We took this hypothesis into consideration when analysing the breakdowns drawn by the subjects and we also feel that this should be taken into consideration when using the component breakdowns. Hence, the breakdowns should be used purely for eliciting components and the larger picture should be noted with regard to the overall shape arrangements.

4.1 Singularity

Changing the orientation of image components changes the perception. This is particularly true for textures where altering the angle of the texture can change the figure/ground perception (see also the discussion below regarding figure/ground for an example). Also, familiar image components such as human figures or aircraft are less often perceived when distorted or not in their natural orientation although the reduction may only be slight.

Images 8, 24 & 80 - These images were selected to study the effects of orientation on grouping of otherwise identical bars. The most common interpretation of all three images was of three groups of bars, as would be expected from the Gestalt rule of proximity. However, changing the orientation of the central element made a slight difference to the results. Where the central bar is vertical, the propensity for three separate groups is reduced and the preference of grouping the bar with the similar oriented group is increased. Paradoxically, for the horizontal bar, the tendency to three separate groups is higher and the grouping of the central horizontal bar with the horizontal group is reduced. When the central bar is diagonal, the tendency is for three separate groups but with separated components more often perceived than for the horizontal or vertical central bars

Images 38, 54, 63 & 73 - An illustration of the effects of changes in singularity comes from these two images. In image 38, by far the most popular interpretation is of a white shape (resembling an aircraft?) on a black background. By contrast, in image 54, where the white figure looks more like a cross, this interpretation, though still the most popular, receives much less support. Familiar images and shapes are less often perceived when distorted or when not in their natural orientation.

Images 61 & 70 - In these images the stylised human figure is the expected orientation in images 61 but is oriented upside down while holding the flag in image

70. Although the recognition of the human is reduced from 61 to 70, the reduction is only slight.

4.2 Familiarity

When elements of an image are gradually removed/reorganized so as to destroy familiarity of the image then the human breakdowns change to be based on individual components rather than the entire image and tend to proximity-based grouping.

Images 2, 18, 46 & 74 - These images were selected to study the effect of the gradual removal of elements of familiarity (the image could be interpreted as a human figure), symmetry and good continuation on image grouping. In fact, the most popular of the eight interpretations listed for image 2 was for a complete breakdown into elementary components, with no intermediate grouping at all. Some of the less popular interpretations showed grouping of the two U-shaped components (through good continuation?), but this did not seem to be a major effect. When the upper four components were tilted (image 18), the trend to complete decomposition was even stronger, though there was weak evidence of an intermediate grouping formed by the four tilted components. When the components were scrambled to remove any effect of symmetry or good continuation (image 46), a variety of groupings was observed, mostly based on proximity. When the upper circle (effectively the head in the human figure interpretation) is changed to an outline rather than a solid fill then the perception is very similar to image 2. However, in one breakdown (D11) the outline circle is perceived as a hole in the paper indicating a figure/ground variation.

Images 12 & 28 - These images show that displacement of one image element (laterally by about 20% of the image diameter) can have a marked effect on perception. Image 12 was seen by most subjects either as four separate segments of a circle or as a circle crossed by three horizontal bars, image 28 either as four separate segments or three grouped and one ungrouped segment. The interpretation of three white bars on a black background was severely weakened; suggesting that perception of additional shapes through figure-ground reversal may require a regular-shaped and familiar background such as a circle or a square.

Images 68 & 81 - These images depict a familiar chef's head with an asymmetrical variant in image 68 and a symmetrical variant in image 81. The tendency to subdivide into hat, face and bow tie is higher in the symmetrical and more familiar variant (image 81) than the less familiar image 68 where the face is less well recognised and more fragmented.

4.3 Symmetry

When symmetry is removed from an image, the human decompositions tend to individual components or image halves. This is particularly true for illusory contours and images where axial symmetry is removed.

Images 4 & 20 - These images were selected to show the effects of vertical displacement of part of an image. In the modified image 20, the frequency with which the two large bars are perceived as a single group is markedly reduced when compared to image 4. This can be explained through the destruction of symmetry and good continuation.

Images 36, 49 & 75 - This set of images compares the results of linear and angular changes in structure on a line-based image. The most popular interpretation of image 36 is of a series of overlapping unbranched line elements showing evidence of good continuation, though interpretations based on identification of letters of the alphabet can also be perceived. Altering the angles of the previously horizontal lines to about 30º (preserving symmetry but reducing instances of good continuation) reverses the relative importance of these two types of interpretation (image 49). Vertical displacement of the right-hand half of the image by about 15% (image 75), by contrast, leads to a different interpretation (dominated by branched lines) predominating. It is of interest that preservation of symmetry while reducing instances of good continuation can result in markedly different partitioning.

Images 39 & 53 - This image pairing demonstrates the effects of symmetry. In image 39 the hexagon is split in half horizontally but in image 53 the hexagon is split into 1/3 & 2/3. The lack of symmetry affects the decompositions markedly. There is no analogue in the breakdowns from image 53 that matches the favoured breakdown of image 39. The breakdowns of image 39 are generally more regular.

There are exceptions where the removal of symmetry has little effect on the decompositions particularly for images that trace the outlines of shapes.

Images 7 & 23 - Here, two line-based images - one symmetric, the other modified to remove axial symmetry - are compared. A wide variety of segmentations can be observed for both images - though there is interestingly no evidence that removal of symmetry significantly affects segmentation in this case. The results are in contrast to those for images 4 & 20.

4.4 Continuity

Reducing the continuity alters the human perceptions with a tendency to proximity grouping and decomposition into individual components.

Images 1, 17 & 52 - These three images were selected to study the effects of small alterations in image structure on hidden contour perception. All consisted of six black circles on which the corners of a white cube were superimposed. In image 1 all were correctly oriented, while in image 17 three were rotated, and in image 52 two were rotated. The results for image 1 showed that by far the most common interpretation of the image was indeed six circles plus the "hidden" cube, as one would expect from the Gestalt rule of good continuation. Results for the other two images, on the other hand, were much more equivocal, suggesting that only a small perturbation of the image is needed to inhibit the perception of illusory contours.

Images 9 & 25 - These two images show the effect of removing corners from a line-based image. Image 9 generated 7 different perceptions common to two or more observers, including examples of both region- and line-based segmentations as defined by Ren et al. [REB00]. Image 25, in which internal corners had been removed, appeared to generate less consensus - 11 different interpretations were recorded.

Images 67 & 79 - Images 67 and 79 investigate the seminal Necker cube phenomenon. Image 67 produces the expected perception of a cube. However, if we

terminate the ends of the components, we interrupt the good continuity and perhaps familiarity and there is a tendency to decompose image 79 into individual components or groups of diagonally aligned components.

When continuity is reduced in conjunction with symmetry removal then the decomposition differs from when continuity alone is removed. An asymmetric image promotes the perception of good continuity whereas a symmetric variant of the image promotes proximity grouping.

Images 43 & 58 - From the Gestalt principles, humans are posited to favour good continuation and grouping of similar objects. Images 43 and 58 examine these principles coupled with symmetry. The asymmetric variant (image 58) produces good continuity where the dots are effectively joined as a line. In contrast, the symmetric variant (image 43) elicits grouped decompositions.

4.5 Figure/ground

If the components of an image are tilted or inverted then the figure/ground perception changes. If the components are textured with stripes then the figure/ground perception changes from the untextured image and if the texture is strengthened with a darker texture then the figure/ground perception changes even more. A uniform background enhances the perception of figure/ground reversal whereas familiarity of image components reduces the figure/ground reversal.

Images 5, 21, 35 & 48 - This set of images illustrates the effect of variations in background on illusory contour formation. In all cases, the illusory boundary between striped and black areas as clearly recognized, though observers were divided on whether image 5 should be perceived as a black overlay on a striped background, a striped overlay on a black background, or two disjoint areas, one striped and one solid black. Interestingly, the modifications to the image (tilting and - more markedly - inversion) all caused fewer observers to perceive a striped background. The reasons for this are not immediately obvious, though it provides a useful reminder that the direction from which an image is viewed can significantly affect its perception.

Images 6 & 22 - This pair of images helps to illustrate the conditions under which image components can be generated through figure-ground reversal. In image 6, the most common interpretation is the obvious one of four distinct triangles. In image 22, where they are shaded to suggest a continuous background, interpretations suggesting an element of figure-ground reversal are more prominent.

Images 11, 27, 51 & 82 - This set of images was also selected to observe the effects of changing background on the generation of perceived image components through figure-ground reversal. Again, with unshaded image elements (image 11), the most common interpretation is solely of unmodified image components. Adding a striped texture to the three image elements and leaving their contours implicit (image 27) strengthened the perception of figure-ground reversal to some extent; adding a darker texture with explicit contours (images 51 & 82) strengthened this perception still more. It should be noted that even in these cases there was still significant support for the original partitioning into the three explicitly-drawn components.

4.6 Texture/shading

When the texture is altered the perception changes. Texture change particularly affects the perception of figure/ground and proximity grouping. However, changes in shading are overridden by changes in continuity or symmetry, component shape and component positioning.

Images 13 & 29 - This pair of images also illustrates the effects of shading on perception. The shapes, comprising interlocking "canoe" shapes differing only in their shading which is symmetric in 13 and asymmetric in figure 29, showed some differences in the way they were partitioned. The decomposition into 4 “U” shapes is less favoured for the asymmetric variant and this asymmetric variant also produces more decompositions compared to the symmetric figure.

Images 33, 45 & 72 - These images also compare the effects of changes in image structure and shading on perception. As observed elsewhere, replacing solid black areas in image 33 with stripes (image 45) appears to have only minor effects, while changing structure (in this case inverting the right-hand half of the image to remove symmetry and reduce instance of good continuation) leads to interpretations where the image is regarded as two separate halves.

Images 37 & 51 - The observation that images consisting of overlapping circles are partitioned in a similar way whether or not they are filled by shading reinforces the principle that differences in shading are of only minor importance in partitioning. Although we note that the subjects generally draw the bounding-box for image 37 but not for the shaded image 51.

5 Conclusion & Future Work

Our results concur with previous investigations such as [REB00] in that image decomposition appears to follow a set of perceptual principles analogous to the Gestalt laws. The experiments and analyses show that these Gestalt laws interact and possibly conflict as noted by [DMM04]. The experiments also indicate that there are a core set of decompositions for each image perceived by 2 or more people along with a set of decompositions seen only by individuals.

We have identified some possibilities for additional work that would generate useful data. The experimental analyses detailed in this paper are very human-oriented. Humans generate all the breakdowns with no recourse as to whether they are feasible for a computer system to generate. Therefore, after we have used the data from these analyses to develop and refine our computational system, we could use the resultant system to generate a set of breakdowns for further images. We can then present these sets of breakdowns, for each image in turn, to human subjects who can rank them *1 to n* where *n* is the number of images in the set. This will allow us to fine-tune the computational system further using tangible computer-generated breakdowns.

6 References

[BOB04] Baker, C., Olson, C. and Behrmann, M.
Role of attention and perceptual grouping in visual statistical learning.
Psychological Science, 15(7): 460-466, 2004.

[BSBFK99] Biederman, I., Subramaniam, S., Bar, M., Kalocsai, P, & Fiser, J. Subordinate-Level Object Classification Re-examined. Psychological Research, 62:131-153, 1999.

[C01] Carlin, M.
Measuring the performance of shape similarity retrieval methods. Computer Vision and Image Understanding, 84(1): 44-61, October 2001, special issue on empirical evaluation of computer vision algorithms, ISSN: 1077-3142.

[DMM04] Desolneux, A., Moisan, L., and Morel, J.-M.
A theory of digital image analysis. 2004. Book in preparation

[DB95] Driver, J. and Baylis G.
One-sided edge assignment in vision: 2. Part decomposition, shape discrimination, and attention to objects. Current Directions in Psychological Science, 4: 201-206, 1995.

[DB97] Dyson, M., and Box, H.
Retrieving symbols from a database by their graphic characteristics: are users consistent? Journal of Visual Languages and Computing 8(1): 85-107, 1997.

[E01] Eakins, J.P.
Trademark image retrieval. In M. Lew (Ed.), *Principles of Visual Information Retrieval* (Ch 13). Springer-Verlag, Berlin, (2001).

[FAG96] Ferguson, R. W., Aminoff, A., and Gentner, D.
Modelling qualitative differences in symmetry judgments, Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society. Hillsdale, NJ: Lawrence Erlbaum Associates, 1996.

[G72] Goldmeier, E.
Similarity in Visually Perceived Forms *[1936]*, in Psychological Issues VIII(1), ed. Herbert J. Schlesinger, International Universities Press, 1972.

[HR84] Hoffman, D.D. and Richards, W.A.
Parts of recognition. Cognition, 18:65-96, 1984

[JV98] Jain, A. K. and Vailaya, A..
Shape-based Retrieval: A Case Study with Trademark Image Databases, Pattern Recognition, 21(9): 1369-1390, 1998.

[LC02] Leung, W.H. and Chen, T..
"Trademark retrieval using contour-skeleton stroke classification", *IEEE Intl. Conf. on Multimedia and Expo.* (ICME 2002), Lausanne, Switzerland, August 2002.

[MGR02] Mojsilovic, A., Gomes, J. and Rogowitz, B.

"ISee: Perceptual features for image library navigation", Proc. 2002 SPIE Human Vision and Electronic Imaging, San Jose, January 2002.

[P85] Palmer, S.
The role of symmetry in shape perception. Acta Psychol (Amst). May;59(1):67-90, 1985. PMID: 4024984

[PHS00] Payne, J., Hepplewhite, L and Stoneham, T. J.
Applying perceptually-based metrics to textural image retrieval methods. Proc SPIE Electronic Imaging, 3959:423-433, Jan 2000.

[PS01] Payne, J. and Stoneham, T. J.
Can Texture and Image Content Retrieval Methods Match Human Perception. ISIMP, May 2001

[Photo] Photobook
http://vismod.www.media.mit.edu/vismod/demos/photobook/

[QBIC] QBIC
http://wwwqbic.almaden.ibm.com/

[REB00] Ren, M., Eakins, J. P. and Briggs, P.
Human perception of trademark images: implications for retrieval system design. Journal of Electronic Imaging, 9 (4):564-575, October 2000.

[R93] Rom, H.
Part Decomposition and Shape Description. PhD Thesis, University of Southern California, December 1993. IRIS Technical Report 93-319

[R00] Rosin, P. L.
Shape partitioning by convexity. IEEE Transactions on Systems, Man, and Cybernetics, Part A 30(2): 202-210 (2000).

[SK95] Siddiqi, K. and Kimia, B. B.
Parts of visual form: Computational aspects. Pattern Analysis And Machine Intelligence, 17(3):239-251, March 1995

[SSH99] Singh, M., Seyranian, G. & Hoffman D.D.
Parsing silhouettes: The short-cut rule. Perception & Psychophysics, 61:636-660, 1999.

[TV02a] Tanase, M and Veltkamp, R.C.
Polygon decomposition based on the straight line skeleton. Symposium on Computational Geometry 2003: 58-67 2002.

[TV02b] Tanase, M and Veltkamp, R.C.
Polygon Decomposition Based on the Straight Line Skeleton. Theoretical Foundations of Computer Vision 2002: 247-267.

[VBF01] Vecera, S.P., Behrmann, M., Filapek, J.C.

Attending to the parts of a single object: part-based selection limitations. Perception and Psychophysics, 63(2):308-321, 2001.

[WELTEH00] Weigle, C., Emigh, W., Liu, G., Taylor, R., Enns, J., and Healey, C. Oriented Texture Slivers: A Technique for Local Value Estimation of Multiple Scalar Fields. In Proceedings Graphics Interface 2000 (Montreal, Canada), pp. 163-170, 2000.

[ZBMB02] Zemel, R., Behrmann, M., Mozer, M. C., and Bavelier, D. Experience-Dependent Perceptual Grouping and Object-Based Attention. Journal of Experimental Psychology: Human Perception and Performance, 28(1):202-217, 2002.

Appendix A - Experiment Documentation

Instructions - Investigation of Image Perception

This investigation aims to understand how people see and interpret images and the shapes of their component parts. The investigation supports a programme of research on Perceptually-Relevant Image Retrieval at the Department of Computer Science, University of York.

During the investigation, you will be presented with a series of images. Your task is to **DRAW** the shapes of the component parts and the shapes of natural groups of component that **YOU** perceive in each image presented. *Two examples are given overleaf.* This process is subjective and as such there are no right or wrong answers; all answers are correct.

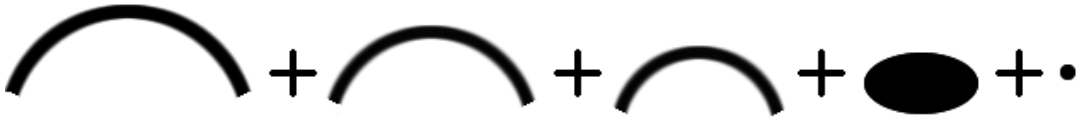
You should attempt to draw the arrangement of shapes that make your **initial** perception **first**. If you can then draw **ANY other shape arrangements** that you perceive **in the order** that you perceive them (2^{nd}, 3^{rd}, 4^{th} etc.)

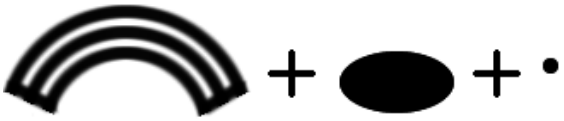
Thank you.

Example 1

Given this image:

Your perceptions of the shapes of the parts and natural groups of components may be:



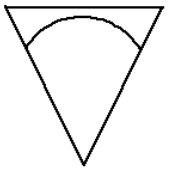


OR

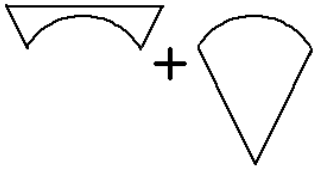
some other perception of components.

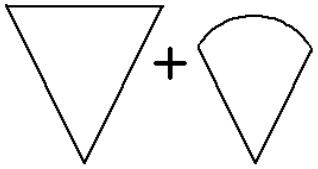
Example 2

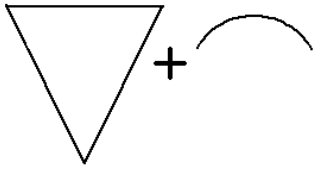
Given this image:

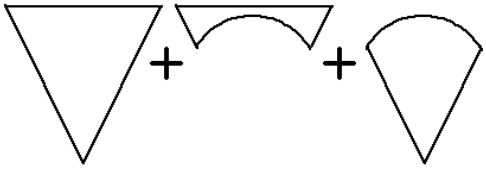


Your perceptions of the shapes of the parts may be:





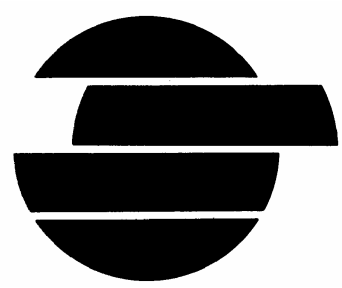




OR

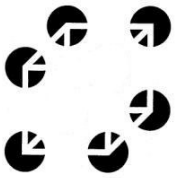
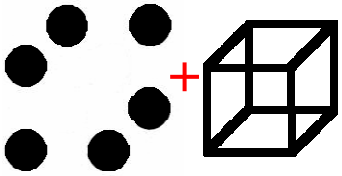
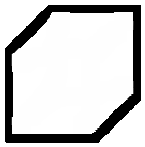
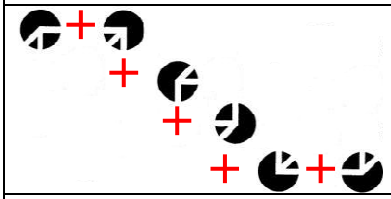
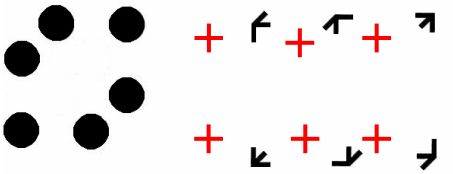
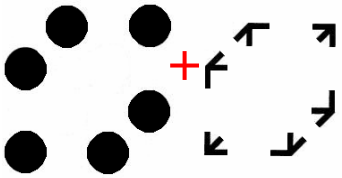
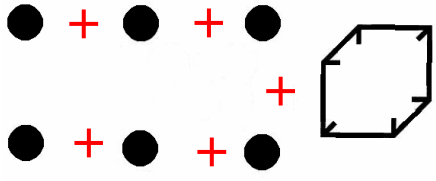
some other perception of components.

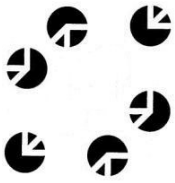
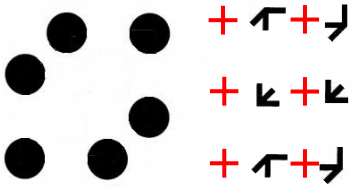
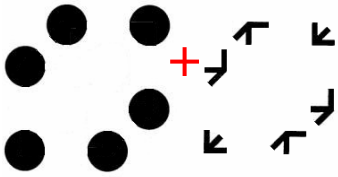
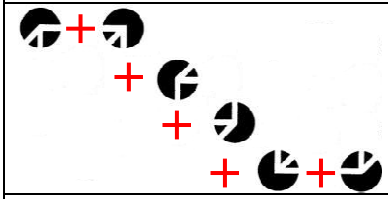
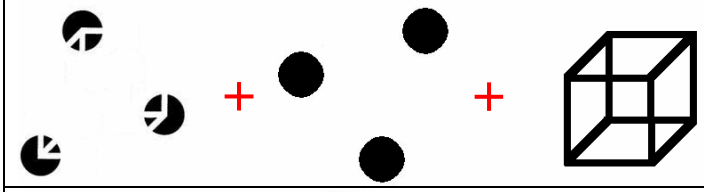
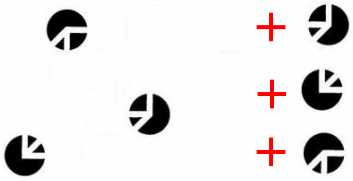
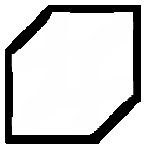
An Example Page From The Booklet Given to Subjects

Given Image	Given Image
	
Your Perception(s) (in order) of the Component Parts.	Your Perception(s) (in order) of the Component Parts.

Appendix B – Experiment Results

Page 27 **Table Listing A Selection of the 84 Images With Their Respective Breakdowns Seen By 2 Or More Subjects.**

	Image 1	
	Decomposition	Score
	D1	0.448
	D6	0.072
	D3	0.064
	D5	0.064
	D2	0.056
	D12	0.048

	Image 17	
	Decomposition	Score
	D1	0.192
	D3	0.141
	D2	0.128
	D6	0.115
	D11	0.077
	D5	0.064
	D7	0.051

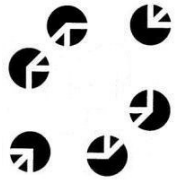
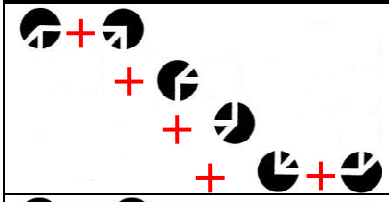
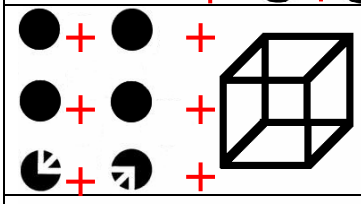
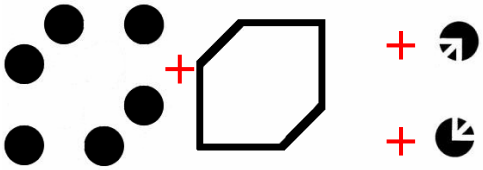
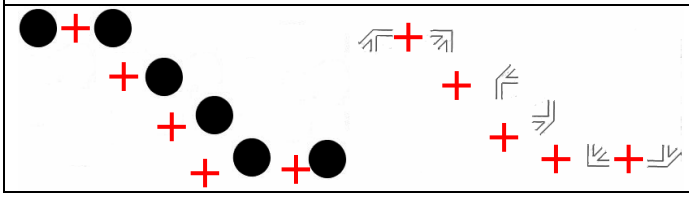
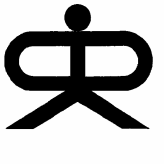
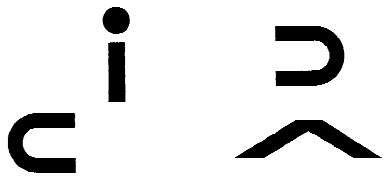
	Image 52	
	Decomposition	Score
	D3	0.259
	D2	0.155
	D5	0.103
	D1	0.086

	Image 2	
	Decomposition	Score
	D1	0.228
	D4	0.161
	D11	0.14
	D2	0.124
	D10	0.088
	D5	0.073
	D12	0.073
	D6	0.01

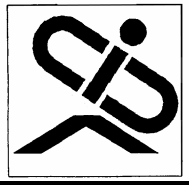
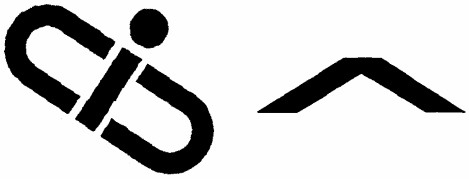
	Image 18	
	Decomposition	Score
	D1	0.405
	D4	0.214
	D5	0.107
	D2	0.095

	Image 46	
	Decomposition	Score
	D4	0.145
	D8	0.145
	D9	0.145
	D3	0.129
	D6	0.129
	D1	0.097
	D5	0.097

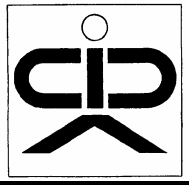
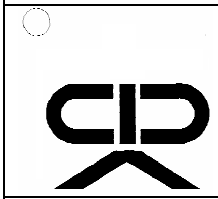
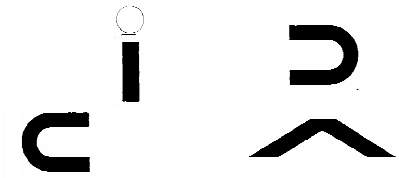
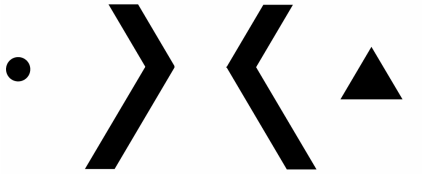
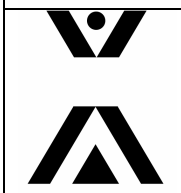
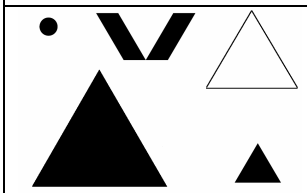
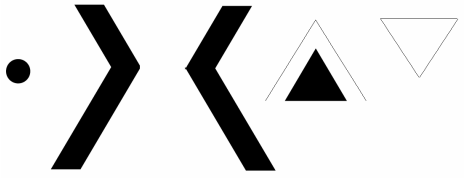
	Image 74	
	Decomposition	Score
	D5	0.293
	D3	0.122
	D4	0.110
	D7	0.098
	D2	0.085
	D10	0.077
	D11	0.061
	D6	0.045

	Image 4	
	Decomposition	Score
	D4	0.333
	D2	0.311
	D5	0.067
	D11	0.056
	D7	0.044
	D3	0.022

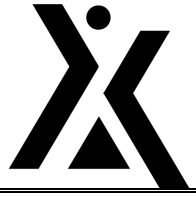
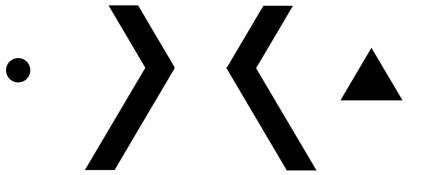
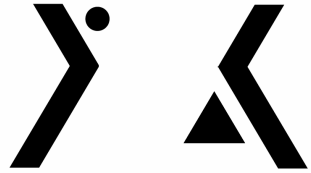
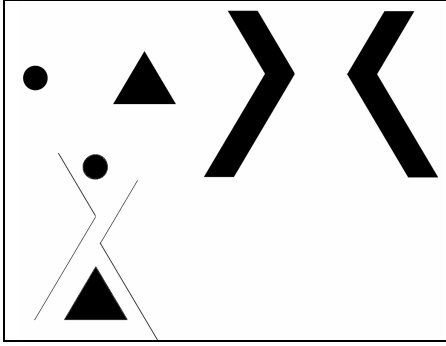
	Image 20	
	Decomposition	Score
	D1	0.564
	D5	0.128
	D2	0.09
	D6	0.077

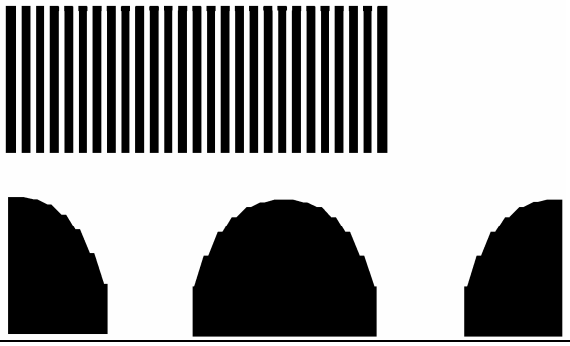
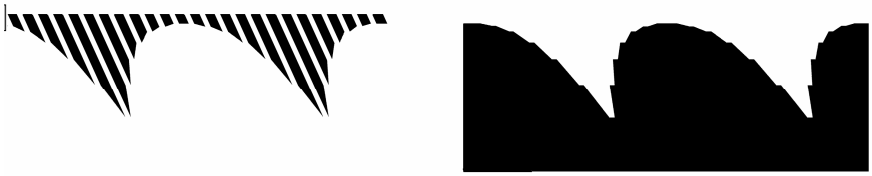
	Image 5	
	Decomposition	Score
	D3	0.355
	D2	0.289
	D1	0.276
	D7	0.053

	Image 21	
	Decomposition	Score
	D1	0.403
	D2	0.338
	D3	0.182

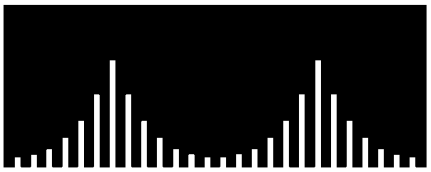
	Image 35	
	Decomposition	Score
	D1	0.436
	D2	0.418
	D3	0.091

	Image 48	
	Decomposition	Score
	D2	0.527
	D1	0.418

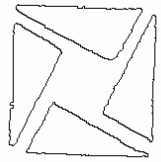
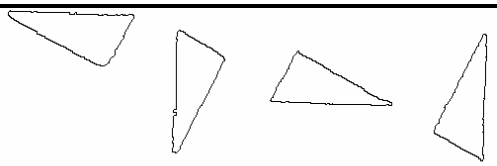
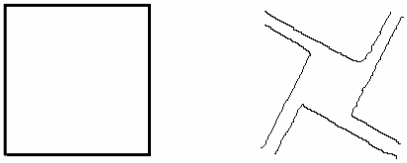
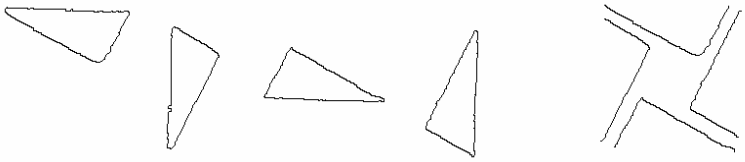
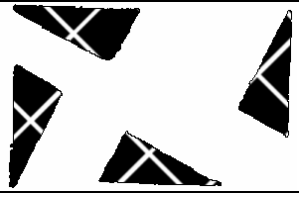
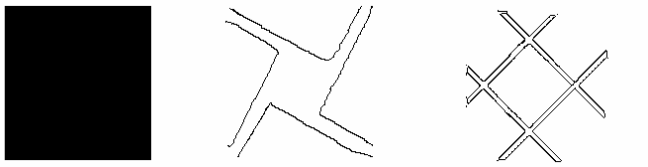
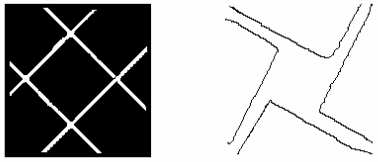
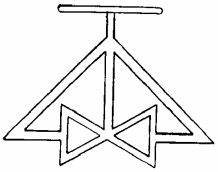
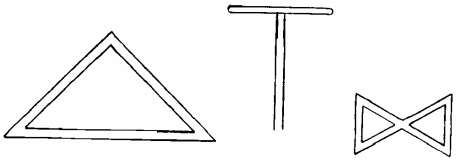
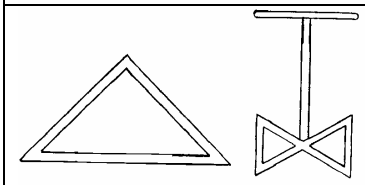
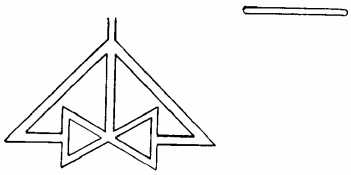
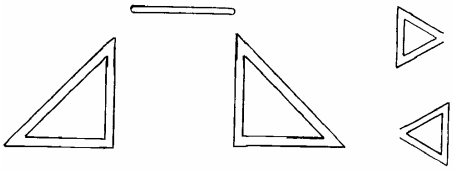
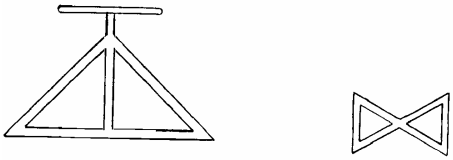
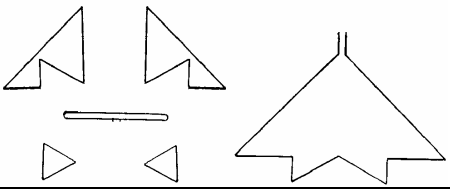
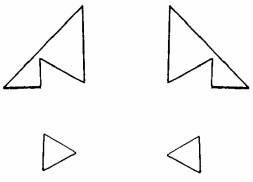
	Image 6	
	Decomposition	Score
	D1	0.563
	D2	0.213
	D3	0.113

	Image 22	
	Decomposition	Score
	D1	0.291
	D4	0.291
	D2	0.241

	Image 7	
	Decomposition	Score
	D2	0.237
	D5	0.142
	D3	0.083
	D13	0.071
	D6	0.059
	D15	0.059
	D4	0.030

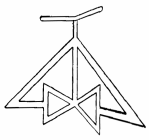
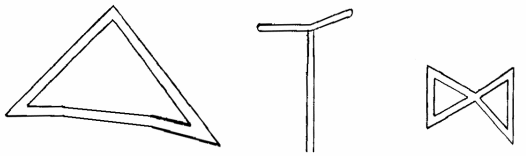
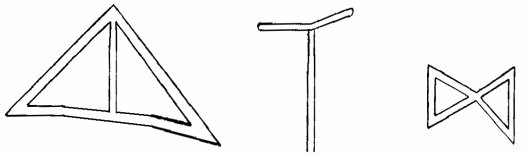
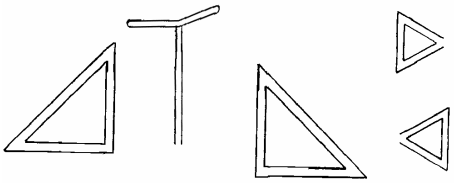
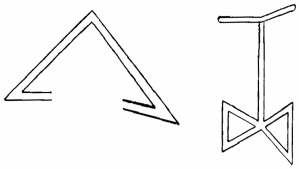
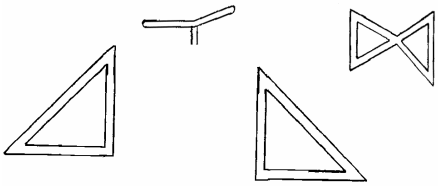
	Image 23	
	Decomposition	Score
	D6	0.205
	D4	0.123
	D1	0.082
	D2	0.068
	D16	0.068

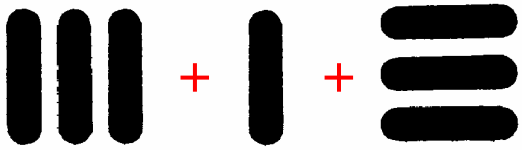
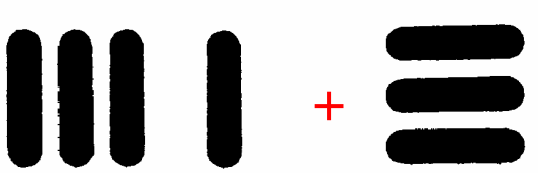
	Image 8	
	Decomposition	Score
	D1	0.555
	D2	0.212
	D3	0.168

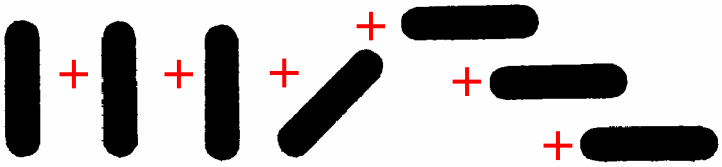
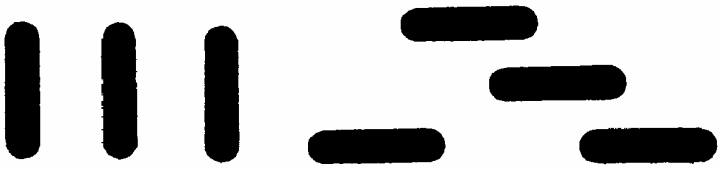
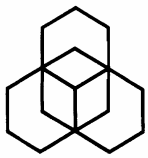
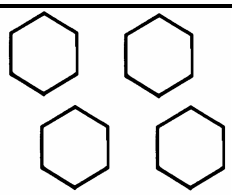
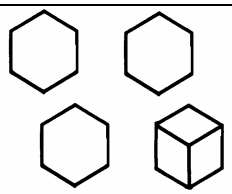
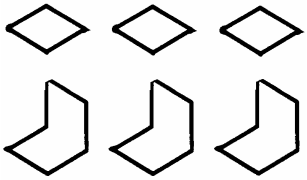
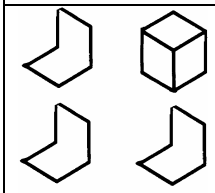
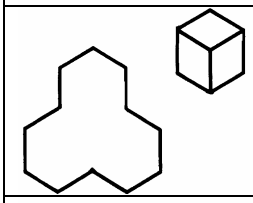
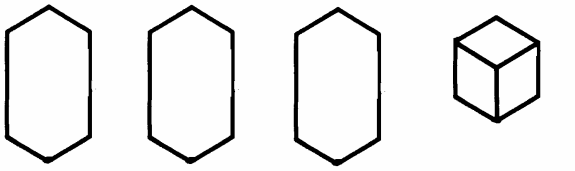
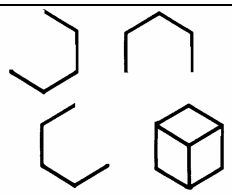
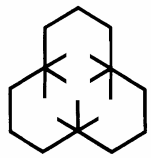
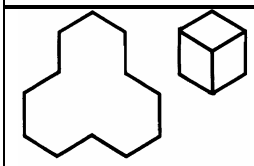
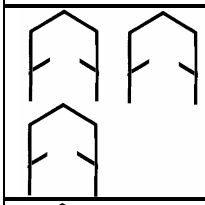
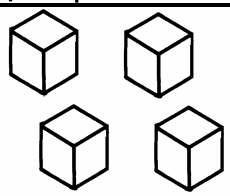
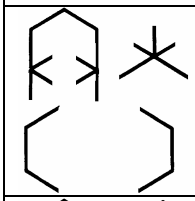
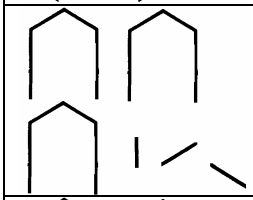
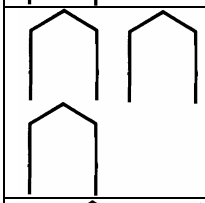
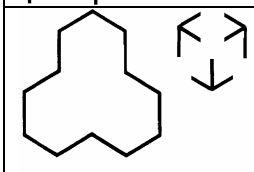
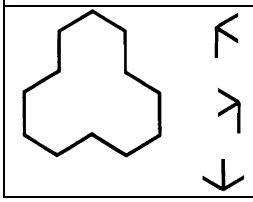
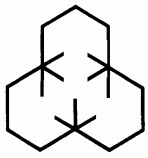
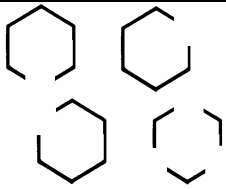
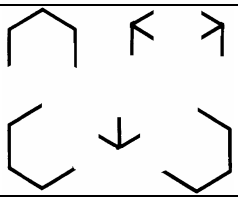
	Image 24	
	Decomposition	Score
	D1	0.592
	D2	0.268
	D3	0.07

	Image 80	
	Decomposition	Score
	D2	0.641
	D4	0.154
	D3	0.064

	Image 9	
	Decomposition	Score
	D1	0.297
	D5	0.154
	D3	0.110
	D4	0.088
	D8	0.066
	D7	0.055
	D2	0.044

	Image 25	
	Decomposition	Score
	D3	0.125
	D12	0.113
	D1	0.075
	D6	0.075
	D8	0.075
	D11	0.075
	D4	0.063
	D9	0.063

	D13	0.063
	D7	0.05
	D2	0.038

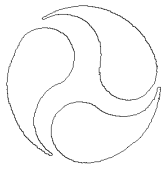
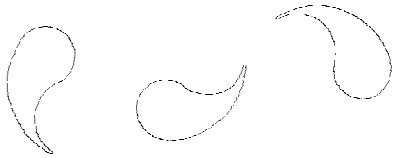
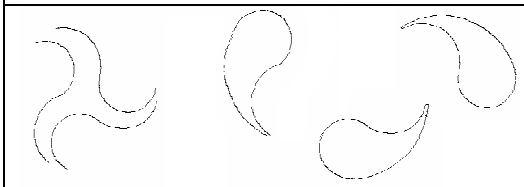
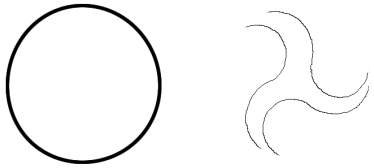
	Image 11	
	Decomposition	Score
	D1	0.573
	D5	0.213
	D2	0.191

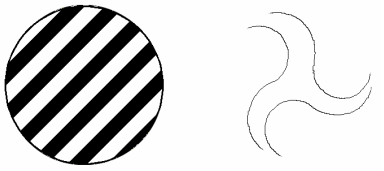
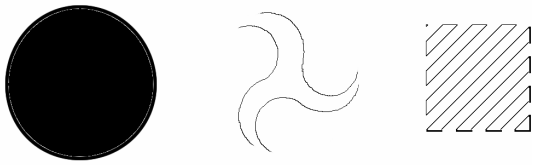
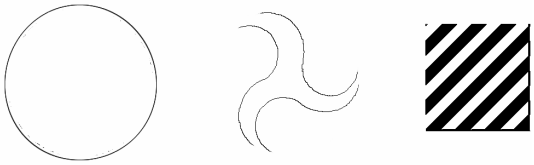
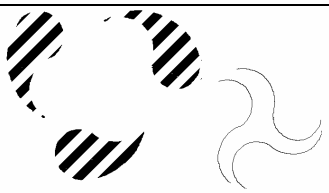
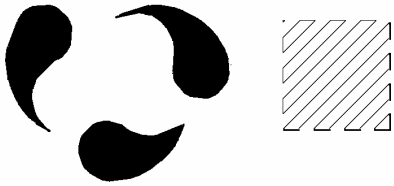
	Image 27	
	Decomposition	Score
	D3	0.25
	D4	0.202
	D2	0.19
	D7	0.107
	D9	0.071
	D1	0.06

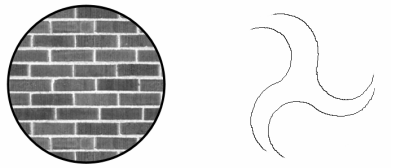
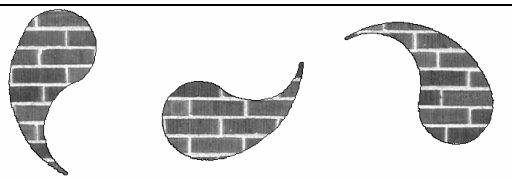
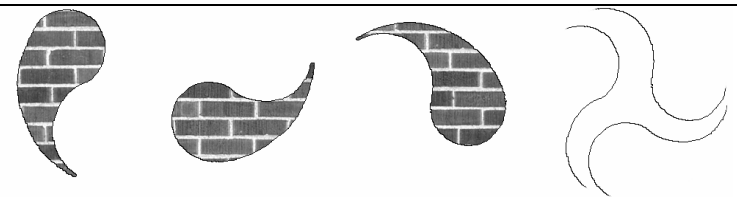
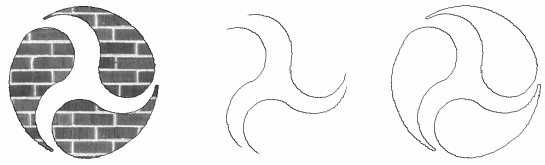
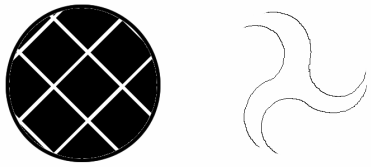
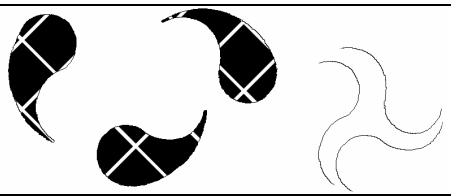
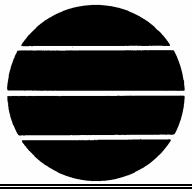
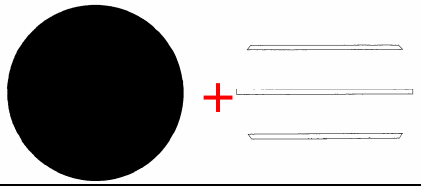
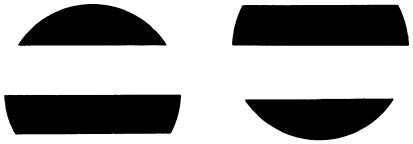
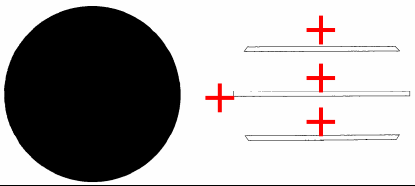
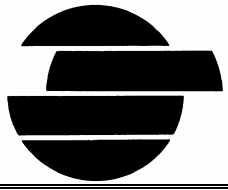
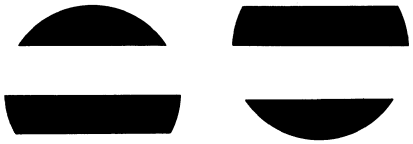
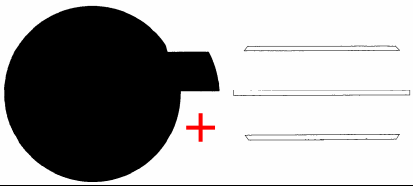
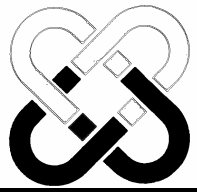
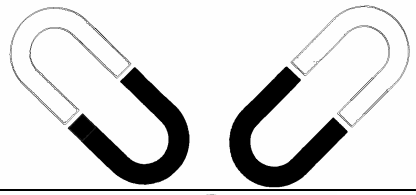
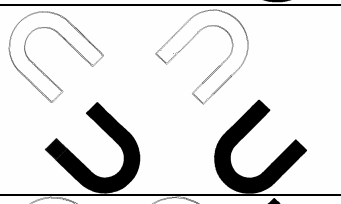
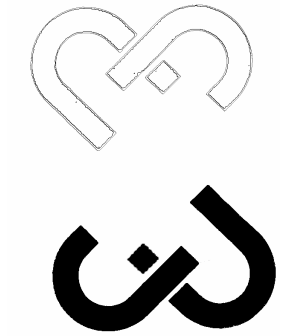
	Image 50	
	Decomposition	Score
	D1	0.424
	D2	0.322
	D3	0.102
	D4	0.102

	Image 82	
	Decomposition	Score
	D1	0.466
	D3	0.247
	D4	0.096

	Image 12	
	Decomposition	Score
	D1	0.451
	D2	0.407
	D3	0.055

	Image 28	
	Decomposition	Score
	D1	0.481
	D2	0.346
	D4	0.123

	Image 13	
	Decomposition	Score
	D1	0.289
	D2	0.289
	D3	0.178
	D7	0.122

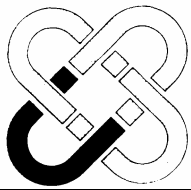
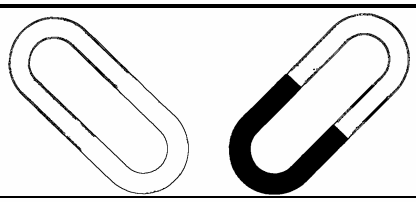
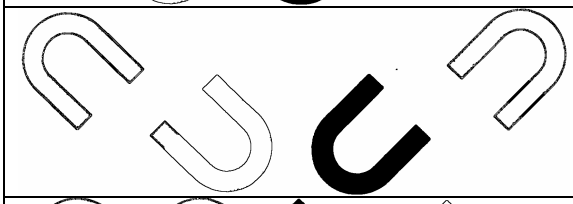
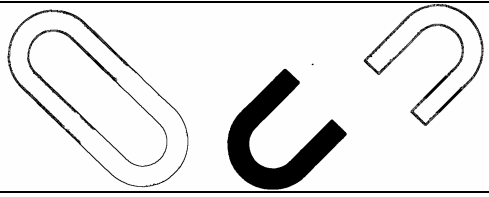
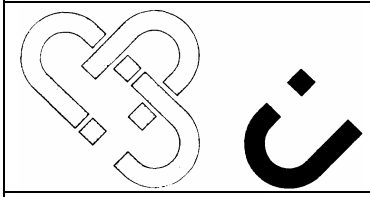
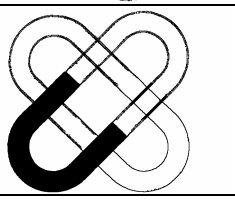
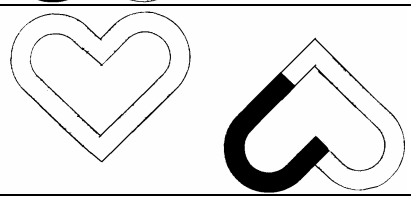
	Image 29	
	Decomposition	Score
	D1	0.29
	D2	0.183
	D3	0.14
	D6	0.129
	D4	0.054
	D5	0.043
	D10	0.022

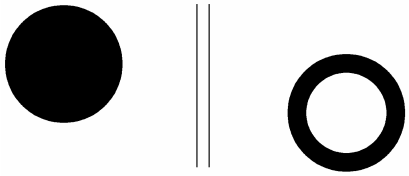
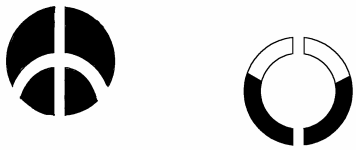
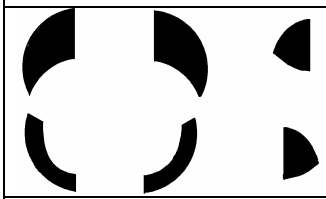
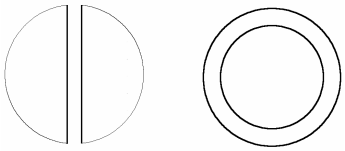
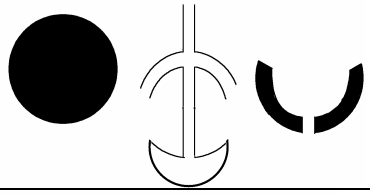
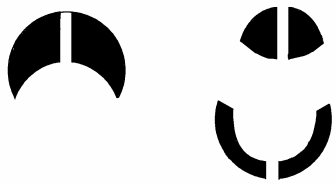
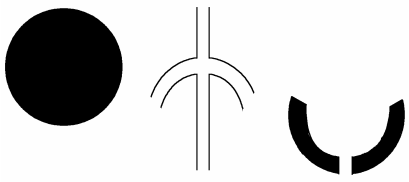
	Image 33	
	Decomposition	Score
	D1	0.169
	D10	0.105
	D16	0.105
	D12	0.064
	D3	0.056
	D15	0.056
	D19	0.040

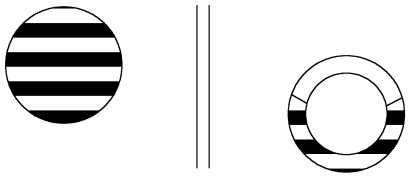
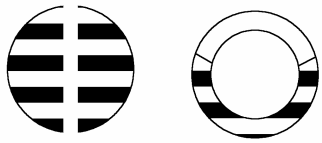
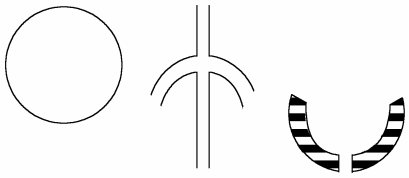
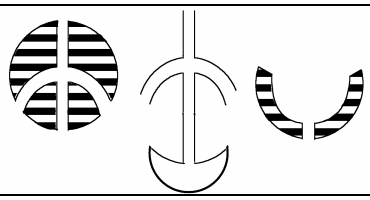
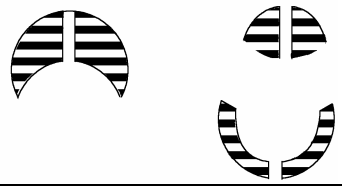
	Image 45	
	Decomposition	Score
	D3	0.316
	D4	0.105
	D1	0.088
	D2	0.088
	D9	0.088

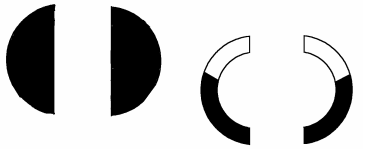
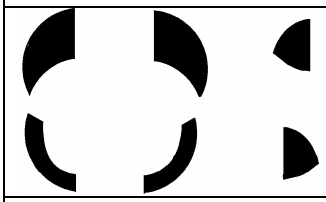
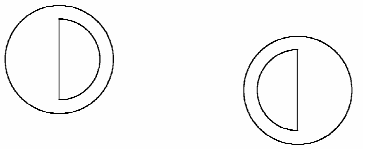
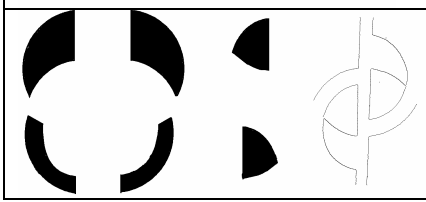
	Image 72	
	Decomposition	Score
	D5	0.224
	D15	0.099
	D6	0.087
	D1	0.062

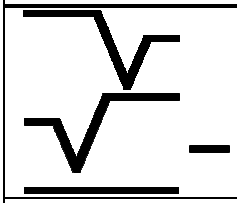
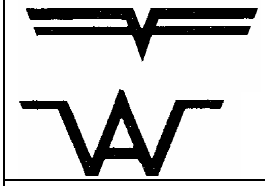
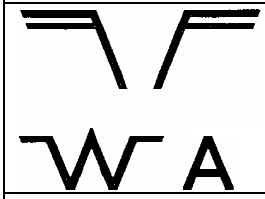
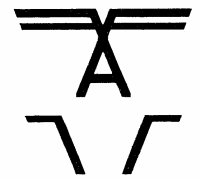
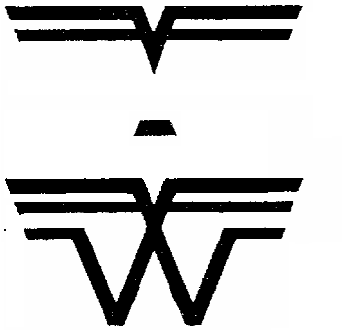
	Image 36	
	Decomposition	Score
	D2	0.195
	D7	0.140
	D6	0.074
	D3	0.056
	D5	0.037
	D15	0.037
	D17	0.047

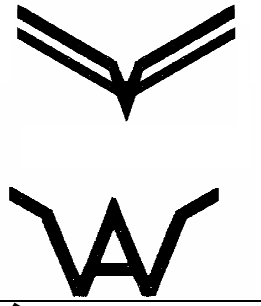
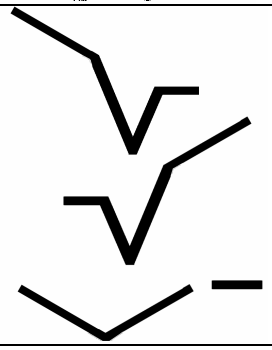
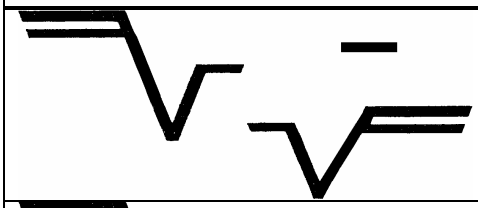
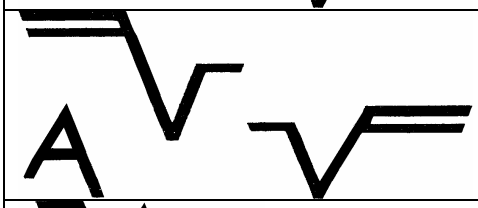
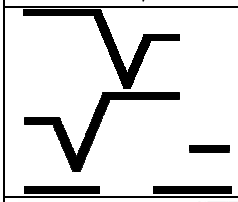
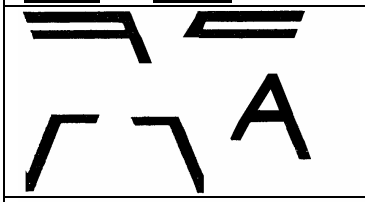
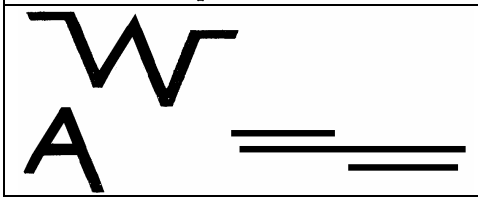
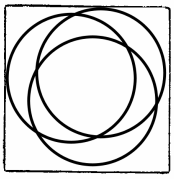
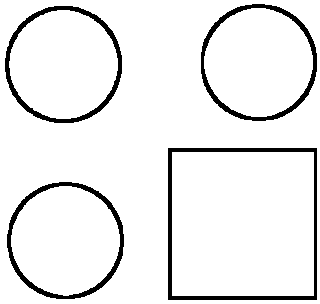
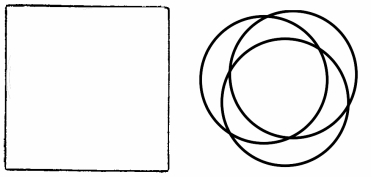
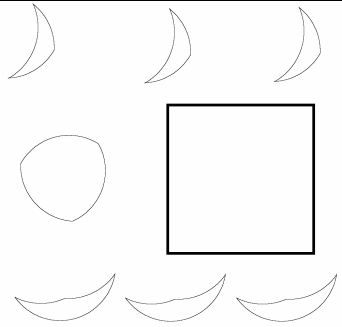
	Image 49	
	Decomposition	Score
	D3	0.273
	D1	0.109
	D4	0.109

	Image 75	
	Decomposition	Score
	D1	0.353
	D3	0.109
	D2	0.068
	D5	0.068
	D12	0.061
	D6	0.054

	Image 37	
	Decomposition	Score
	D1	0.636
	D4	0.109
	D2	0.091

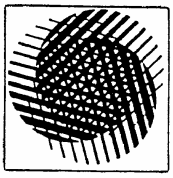
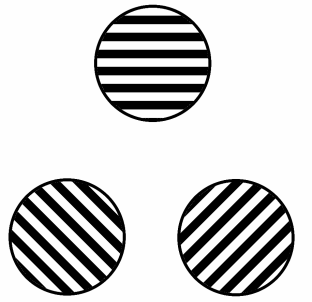
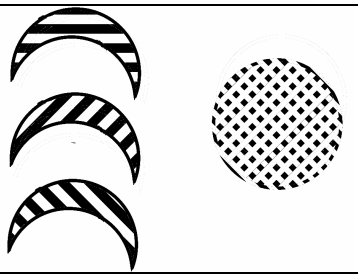
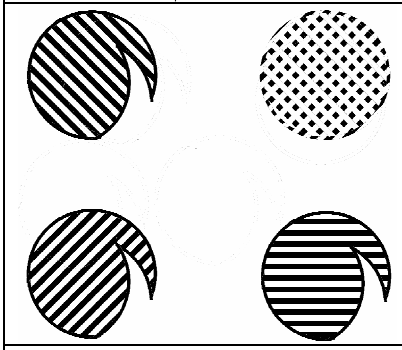
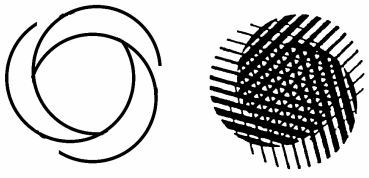
	Image 51	
	Decomposition	Score
	D1	0.529
	D3	0.118
	D5	0.118
	D6	0.118

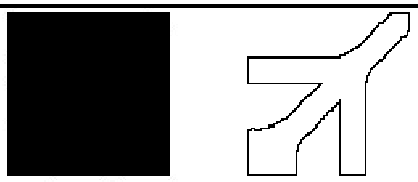
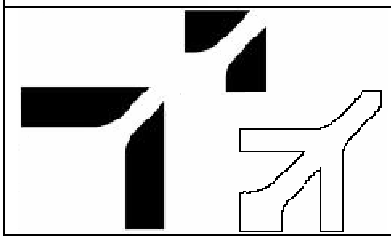
	Image 38	
	Decomposition	Score
	D1	0.567
	D4	0.15
	D2	0.1
	D5	0.1

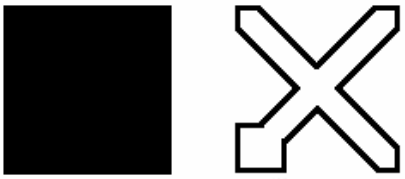
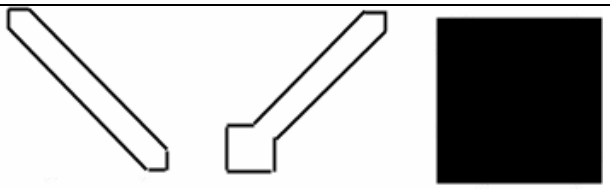
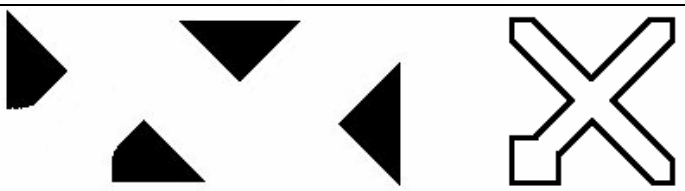
	Image 54	
	Decomposition	Score
	D1	0.242
	D8	0.113
	D2	0.097
	D4	0.097
	D5	0.097
	D6	0.081

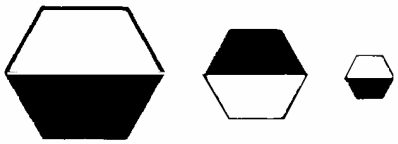
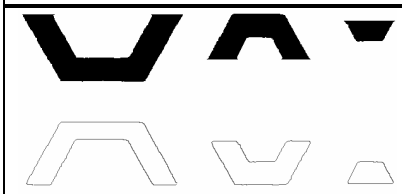
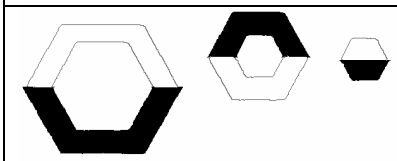
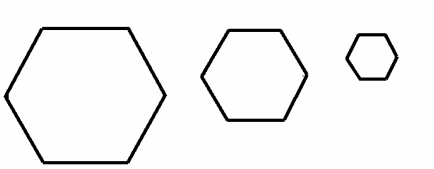
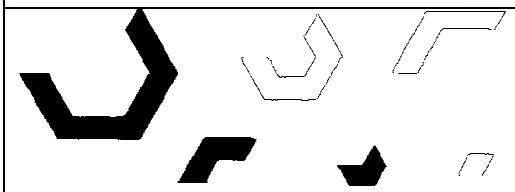
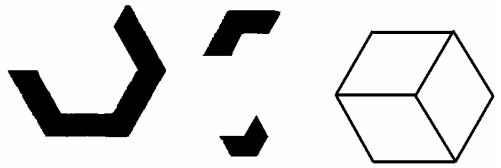
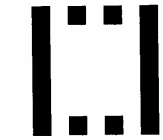
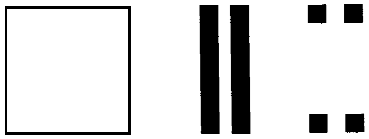
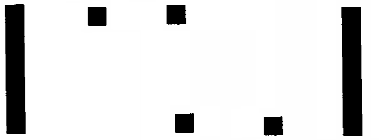
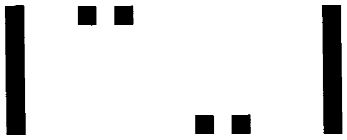
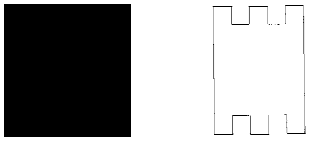
	Image 39	
	Decomposition	Score
	D3	0.278
	D6	0.167
	D2	0.093
	D4	0.093

	Image 53	
	Decomposition	Score
	D2	0.23
	D6	0.098
	D7	0.098
	D1	0.082

	Image 43	
	Decomposition	Score
	D1	0.433
	D2	0.15
	D3	0.133
	D4	0.1

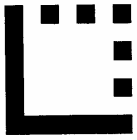
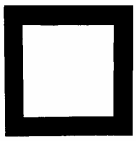
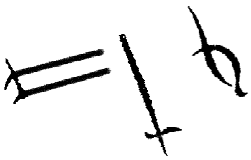
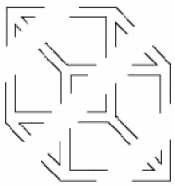
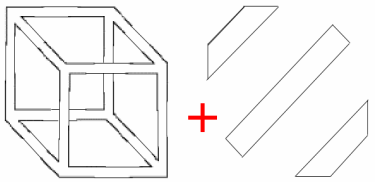
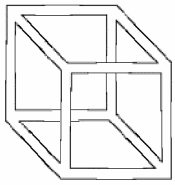
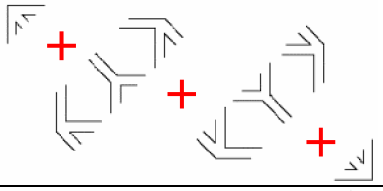
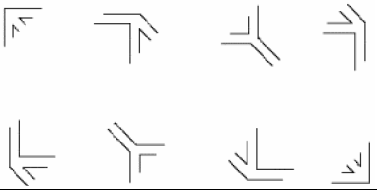
	Image 58	
	Decomposition	Score
	D1	0.656
	D3	0.115

	Image 61	
	Decomposition	Score
	D3	0.288
	D5	0.113
	D6	0.1
	D1	0.075
	D4	0.075
	D10	0.05
	D12	0.041

	D11	0.019

	Image 70	
	Decomposition	Score
	D8	0.227
	D1	0.12
	D3	0.12
	D11	0.093
	D5	0.08
	D4	0.067
	D10	0.067

	Image 67	
	Decomposition	Score
	D2	0.309
	D1	0.176
	D3	0.162
	D5	0.103

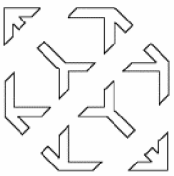
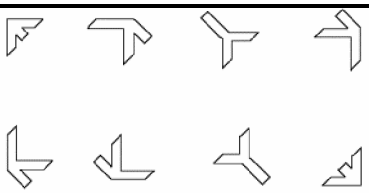
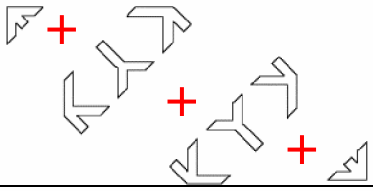
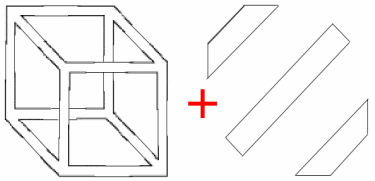
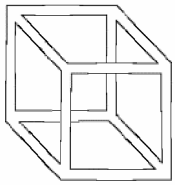
	Image 79	
	Decomposition	Score
	D1	0.466
	D4	0.247
	D2	0.068
	D3	0.055

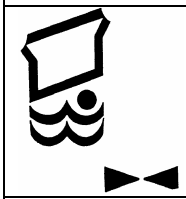
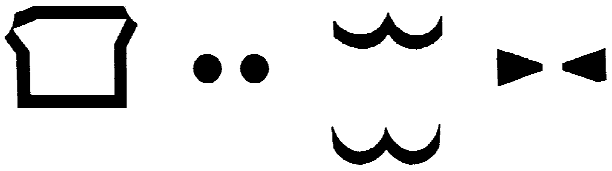
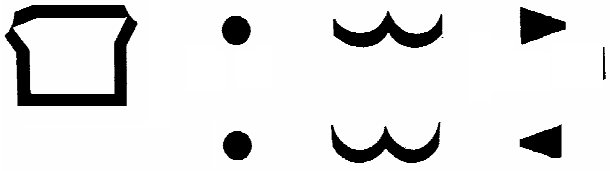
	Image 68	
	Decomposition	Score
	D1	0.293
	D4	0.237
	D5	0.12
	D3	0.08
	D9	0.08

	Image 81	
	Decomposition	Score
	D3	0.356
	D4	0.151
	D7	0.123
	D9	0.041