# IMAGE UNDERSTANDING
# AND
# AESTHETIC EVALUATION

**Zbigniew Les**

**This book is based on the Master of Science thesis**

**The University of Melbourne, 2004**



This book was designed and produced by QJF Press

P.O. Box 654, Toorak**, VIC 3142**

**www.qjfpl.org**



**Printed in Australia**

**QJF PRESS • MELBOURNE**

## CONTENTS

## Chapter 1 Introduction

Image understanding refers to a computational approach to image interpretation. It stresses functional aspects of the information conveyed by an image, and can be expressed as follows: when an object (in the image) is discovered an action should be undertaken. Pictures which are created for artistic proposes cannot be regarded within this functional framework. However, the possibility exists of applying a similar image-understanding approach to this class of pictures in order to find a methodology that will allow for aesthetic evaluation. This image-understanding approach has been used in the current thesis to develop a method of aesthetic evaluation of a class of pictures (abstract paintings). This method has been applied to a particular class of abstract paintings, the Mondrian class, in order to find composition rules of works of art. To evaluate pictures from the Mondrian class a neural networks method is used.

The method of aesthetic evaluation can be seen as an attempt to develop a hybrid connectionist computational model of human aesthetic evaluation. However, in the present research only a computational aspect of this method is presented. This is a first attempt which uses a computational approach based on image understanding method to examine aesthetic evaluation of works of art.

## Chapter 2. Literature Review

The literature review is presented in three main sections: image understanding, aesthetic evaluation, and image processing methods. Because works of art (paintings) are described by use of such features as shape and texture, and finding a set of descriptors is one of the main tasks, some image processing methods used to analyse shape and texture are included.

### 2.1. Image Understanding

Works of art possess a complex structure. To elaborate a method of aesthetic evaluation which can capture this complexity and also make it possible to create a computational system, a framework taken from a similar field of research is needed.

One of the fields of research which tries to solve the problem of image interpretation is image understanding. The term image understanding has a range of meanings, but in general, image understanding refers to a computational, information processing approach to image interpretation. The term *image understanding* denotes an interdisciplinary research area which includes signal processing, statistical and syntactic pattern recognition, artificial intelligence, and psychology. Image understanding refers to knowledge-based interpretations of visual scenes that transform pictorial inputs into commonly understood descriptions or symbols (see for example: Binford, 1982; McKeown, Harvey, McDermott, 1985; Matsuyama & Shang-Shouq Hwang 1990, Ullman & Richards, 1989). A two-dimensional image is a mediator between an image-understanding system and real world objects in the process of recognition. In simple cases, pattern recognition techniques are used to classify an input into one of several categories (Duda & Hart, 1973; Schalkoff, 1989). Such pattern recognition systems can be seen as the simplest image understanding systems. In contrast to pattern recognition, in which all processing is done based on attributes of objects, image understanding involves reasoning about the structure of the scene. Most natural scenes are composed of objects of various kinds, and to understand a scene we need knowledge about relations between objects, as well as knowledge about their intrinsic properties. The term *object recognition* is often used for systems which recognise specific objects by dichotomising the world into the target object and the background (Wang & Srihari, 1989; Grimson, 1990). However, image understanding considers a scene as a heterogeneous structure composed of mutually related objects of different kinds.

Computer vision is also used to refer to a similar research area (Shirai, 1987; Overington, 1992), but while computer vision emphasises the computational aspects of visual information processing, such as measurement of three-dimensional shape information by visual sensors, image understanding stresses knowledge representation and reasoning methods for scene interpretation. Another field of research which stresses modelling of the human visual system, called computational vision can also be treated as a field of image understanding research. Computational vision is a multidisciplinary and synergetic approach whose main task is to explain the processes of the human visual system and build artificial visual systems (Welcher, 1990). The flow of the information from retina toward the striate cortex is modelled as a flow of information from sensor toward an object recognition system. In this approach invariants play a key role in finding an appropriate set of image transformations. The invariant, which was an important concept in Gibson's theory (Gibson, 1973) allows us to recognise objects despite image variability. In this approach, object recognition systems are created based on transformation techniques which are taken from a study of the human visual system (Wechsler & Zimmerman, 1988). A different definition of computational vision, which is very similar to the definition of image understanding, can be found in Poggio, Tore & Koch (1990). According to Poggio et al. computational vision denotes a new field in artificial intelligence, centred on theoretical studies of visual information processing. Its two main goals are to develop image understanding systems which automatically construct scene descriptions from image input data, and to understand human vision.

A different class of systems, which can be treated as a subclass of image understanding systems, are two-dimensional object recognition systems (Reiss, 1992, Wechsler & Zimmerman, 1988). A two-dimensional image can be treated as a transformation of the 3-D real world scene into the 2-D picture domain. A work of art (painting) can be seen as a 2-D picture, and reference to a 3-D world scene is only present in a certain class of works of art: representational paintings. Two-dimensional recognition systems for machine vision and robotics applications are usually model driven in that recognition involves matching the input image with a set of predefined models of parts. Models are usually based on geometric properties describing shape characteristics. The 2-D image representations are based on global features (perimeter, area, moments of inertia), local features (corners, lines, or curve segments) and relational features (distance and adjacency). Matching is implemented via statistical pattern recognition or relational graph matching (Reiss, 1992).

### 2.1.1. Neural networks

The neural network can be treated as part of an image understanding approach (pattern recognition) and is suitable for recognition or classification of objects where picture representation is given in the form of feature vectors.

The formal artificial neuron was introduced by McCulloch & Pitts in 1943 (McCulloch, 1965) when they demonstrated that everything that can be expressed or computed by man can be expressed and computed by a set of binary units. Retrieval of a stored pattern is viewed as the convergence of a dynamical system towards an attractor. This notion of memories as attractors creates the possibility of retrieving a complete pattern starting with a noisy or incomplete pattern.( Kosko, 1987).

An important aspect of neural network systems is that they “learn” by being trained on a set of examples. There are many types of neural networks and many different methods used to train them (see for example: Zurada 1992). One fundamental way of distinguishing between various types of neural networks is by the method used to train the network, and there are two main categories of training methods: supervised and unsupervised learning.

In unsupervised learning, the training set consists only of a given input vector. During the learning stage there is no comparison done with predetermined desired target (output) vector. A neural network must discover for itself any existing patterns or regularities (Kohonen, 1990).

In supervised learning, a training pair consists of an input vector and a desired target vector. One of the popular supervised learning algorithms is a back-propagation algorithm which is used to recognise and classify the objects. Many learning algorithms have produced good results in pattern recognition area (see, for example: Wasserman & Schwartz, 1988, Zurada 1990 ).

Some aspects of the image understanding approach will be used in the present work as a framework to generate a method for aesthetic evaluation of pictures.

As was described above, image understanding is an interdisciplinary approach to analysis of images. Works of art (paintings) possess a complicated structure and to analyse a work of art the knowledge from many research areas is needed. This “knowledge” can be represented as different levels of description. Levels of description mirror the complexity of the picture representation, and stress different aspects of the picture “interpretation”. For example, the perceptual level is concerned with perceptual phenomena which are involved in the process of perception of the picture. The semantic level is concerned with meaning of the picture elements and their symbolic representation. However, not all levels of description are needed to obtain a picture description for a given picture. For example, Mondrian's pictures and Titian's pictures can be described based on different levels of description. In the case of Mondrain's work such levels as perceptual or semantic do not have a big influence on the description of the picture, whereas to describe Titian's work these levels can not be ignored.

In the image understanding approach a raw image at the pixel level is usually analysed to produce primitive image features such as points, lines and regions, and after segmentation a symbolic representation is created from various 2-D image features, such as shape, texture and contour. Aesthetic evaluation of a picture (work of art) is often given in terms of its visual attributes such as colour, line, shape or texture (Arnheim, 1974). We can use image processing methods to obtain a set of descriptors for each attribute (line, shape, colour and texture) in order to use them to achieve aesthetic evaluation of the picture.

In image understanding, we can distinguish two main levels of processing: low-level processing which involves pre-processing and segmentation, and high-level which involves recognition of the object. As was mentioned above, pattern recognition methods are part of an image understanding approach in which all processing is done based on attributes of objects. In the work of art such as abstract painting the picture elements (objects) can be described using an attributes (descriptors) of the objects and the pattern recognition method (neural network) can be used for an aesthetic evaluation of the picture.

## 2.2. Aesthetic evaluation

Image understanding approaches can be seen as obtaining knowledge about reality based on the information extracted from the picture. The question of whether art can provide knowledge of, or insight into, reality is as old as philosophy itself. Plato argued in "The Republic" that art has the power to represent only the appearances of reality. According to this theory, a painter reproduces (imitates) a subject on canvas. The counter-position, that art can yield insight into the real, is commonly held by modern philosophers, artists, and critics. Many critics, in fact, allege that art offers a special, non-discursive, and intuitive knowledge of reality that science and philosophy cannot achieve (Langer, 1967, Kuspit, 1984)

However, while one aspect the work of art can be regarded as an object which can give insight into reality, there is another aspect in which a work of art is subjected to aesthetic evaluation. There is little agreement on the issue of aesthetic evaluation and many philosophers have tried to find a solution to this problem (see Tatarkiewicz (1974) for discussion).

A radical position on the issue of aesthetic evaluation is that aesthetic evaluation is simply an expression of preference, and thus cannot be considered either true or false. A contrary notation suggests a work of art possesses unique properties which allow it to be distinguished from an ordinary object. This process involves aesthetic evaluation based on unique properties and a specific knowledge which is needed to interpret content of work of art.

Aesthetic evaluation can be seen also as the interpretation of works of art in terms of their “form” or “structure” and as such is often distinguished from the interpretation of work in terms of their “content” or “association” ( Golaszewska, 1989; Stiny, 1975 ). For the picture “form” corresponds more or less to how a work of art is “composed”. The composition of the picture treats a work of art as a coherent whole that can be divided into several parts, each bearing a relation to the whole. The evaluation of a work of art in terms of its form considers the properties of this whole, its parts, and their relations. These properties in a painting include colour, shape, texture and rule of composition of those elements on the picture’s plane.

In the present study only one class of works of art, abstract painting, in which an evaluation can be based on their “form” has been chosen. It is assumed that a work of art (an abstract painting) possesses unique properties which can be measured, and based on these measurements can be classified (evaluated) as a work of art or not a work of art.

### 2.2.1. Aesthetic Measurements

With the assumption that aesthetic evaluation is based on unique properties (relationships) of a picture, some of the examples of the measurements which have been reported or can be used for aesthetic evaluation will be presented.

One attempt to establish an aesthetic measure was proposed by Birkhoff (1932).

Birkhof's aesthetic measure defines the evaluative criterion M=O/C, where M is an aesthetic measure, O is order, and C is complexity. He applied this measure to several classes of objects by defining formulae for measuring the order and complexity of the elements of each of the classes. Birkof's measure has some disadvantages such as arbitrariness of the measurement of the order and the complexity (see Stiny, 1975). Although it is also impossible for a single formula to be universally acceptable, Birkof's measure provides an attempt at a measure for aesthetic evaluation.

The fractal dimension is often used as a description of an image (Lumb, Low & Di Leo, 1994; Vehel & Mignot ,1994; Nyikos, Balazs & Schiller, 1994) and Nyikos, Balazs & Schiller, used a box counting method for counting a fractal dimension to analyse drawings of some of the great artists.

As Nyikos, Balazs & Schiller showed, works of art can be characterised by computing a fractal dimension. However, to apply this method to a picture (painting) we need to find a method which allows us to represent a picture as a set of curves. To obtain this "representation" we could use edge detection or contour following method.

The proportions can be seen as unique properties of a picture which make it possible to distinguish a work of art from a picture that is not a work of art. The proportion based on ratios derived from the musical scale influenced many artists from ancient Greek to the Renaissance (Gombrich 1982). These proportions were also used by modern artists as a part of composition rules or as a way of organising the canvas. For example, the proportions based on $\sqrt{2}$ or the golden section $\Phi=(1+\sqrt{5})/2$ facilitate the repetition of ratios that fit together to form a whole in an aesthetic way. One of the proportions (ratios) which was often used as an element of a simple composition rule is the golden section. For example, Gris used the golden section in constructing the partial grid that dominates certain of his composition (Adelman & Compton, 1980).

Composition rules (relationships) based on proportion include rules used as a part of composition of the picture or as a more general aesthetic system. For example, the painting by J.Gris “The Watch” is partly constructed by using a set of square triangles with proportion 3:4:5 (Adelman & Compton, 1980). Also the proportions of triangles (rectangles) based on Fibonacci series (1:2:3:5:...) were used by Gris to build a composition. Le Corbusier created the first modern system of proportion, which he called the Modulor. This system developed a linear scale of lengths based on irrational number (Kappraff, 1990 ).

Artists often used regular grids as means of organising the objects within the canvas. This causes the objects to possess an unique properties which can be derived from regular grids. For example, series of works by Van Doesburg were based on grids of 4, 8, or 16 subdivisions vertically and horizontally, together with diagonals (Dabrowski, 1985). These subdivisions introduce a set of constraints to which picture object is subjected and by this form a regular structure based on the set of proportions and relations.

Colours in works of art have also been used in such a way that preserves relations between colour primaries. Unique properties of the picture can be seen as factors which describe relations between colour primaries. These relations were often expressed in the quantitative form as a set of proportions or distance on the chosen scale. For example, the Ostwald’s colour-circle which contains 24 hues was based on the system of four psychological primaries, red, yellow, blue and green (Lynton, 1980). Based on this circle Ostwald established the principle of colour-harmony which depends on the balance of values (the black and white content of each hue), and, among the hues round the circle, on the juxtaposition of those which are found at intervals of 3, 4, 6, 8 or 12. These results were used by the Dutch movement De Stijl to which Mondrian and van Doesburg belonged. Paintings of these artists have been chosen for the purpose of the present research.

There are also many qualitative “measures” which are often used for evaluative description of works of art. These qualitative properties characterize picture elements or composition rules. For example, Arnheim used the weight of visual objects as one of the properties to describe the balance of the composition of a picture (Arnheim, 1974). It is possible to express such properties as the weight in the form of measurements such as area, radius, moments or a set of shape descriptors, and to formulate “a balance of the composition” using these measurements. Also, many of composition rules (relations) are given in qualitative form. For example, “the clear identification of the opposition qualities of shape, with one kind of shape clearly dominant because it appears most often and more importantly”, is often used as a rule of composition (Ellinger, 1980). To apply these kinds of rules their quantitative representation needs to be found. One way is to transform their qualitative features into a quantitative form. “The opposition qualities” of shape can be expressed in the form of shape descriptors, and “kind of shape” can be described in the form of its boundary descriptors or region descriptors. The second is to use artificial intelligence (AI) methods of knowledge representation and try to form a set of rules which can capture this symbolic representation.

As was described above, the unique properties of pictures and the relations between picture elements are part of the description and evaluation of the works of art. It is justifiable to assume that an aesthetic evaluation of pictures can be based on measurement of these properties. Aesthetic evaluation (description) of a picture is also given in terms of its visual attributes like shape or texture (Arnheim, 1974) . Such properties as shape and texture of the picture can be analysed by using image processing methods.

### 2.3. Image processing.

Image processing methods are used in the image understanding approach. In aesthetic evaluation of pictures these methods can be used to find a set of picture descriptors. Selecting the appropriate image understanding method to obtain a set of picture descriptors depends on the class of pictures to be analysed. Edge detection, segmentation and filtering are some of the methods used on many of the pictures. A description of these methods can be found in Jain (1986), Rosenfeld & Kak, (1982).

The visual attributes of picture objects such as shape and texture can be characterised by a set of descriptors obtained by applying image processing methods. A method of aesthetic evaluation of the picture which was established for the purpose of this research allows evaluation of a large class of pictures. Depending on the class selected, the different image processing methods can be applied to find a set of picture descriptors. To find a set of picture descriptors for the classes such as Kandinsky or Mondrian, the methods of shape analysis are needed whereas for the Pollock class the methods of texture analysis are needed. For the description of a generic class (Mondrian, Kandinsky, Pollock) see chapter 4.4.

### 2.3.1. Shape

In art shape is often described in terms of boundary curve or as a region of a given colour. A discussion of this problem can be found in Arnheim (1974). In an “abstract” painting, shape is often described as a circle, ellipse, square or irregular shape (The Tate Gallery, 1980a). To find a set of shape descriptors which can be expressed in quantitative form we need methods of shape analysis.

In image understanding, shape analysis provides a means of symbolic description of the image. Three-stage processes are used in shape analysis: shape detection, shape description and representation, and shape classification (Kasturi & Jain, 1991). The last stage, based on shape features, labels a given shape to the known class of shapes. Representation and shape description schemes are usually chosen in such a way that the descriptors are invariant to rotation, translation, and scale change. One of the most popular methods of shape analysis is the Hough transformation. A generalised Hough algorithm given in Ballard (1981), uses edge information to define mapping from the orientation of an edge point to the reference point of shape. This method can be suitable for a class of pictures which possess elements of regular shape and simple structure.

Syntactic shape analysis is often used to analyse a regular contour (Pavlidis & Ali, 1979). This method, which enables shape description as a set of primitives and a set of rules, can be used for aesthetic evaluation of the picture. The rules which are thus obtained can be used to discover regularities among the shapes of a work of art. The syntactic shape analysis method can be applied to analyse regular shapes of the pictures from the Herbin or Mondrian class. The techniques which allow analysis of more general class of shapes (regular or irregular shapes) are methods based on the neural network approach (Pal, Pal & Basu, 1993) or combining pattern recognition approach (Bala & Wechsler, 1993). In neural networks approach the shape is approximated by polygon and angle between two consecutive line segments. Results are taken as features, a sequence of which is modelled by an autoregressive (AR) process. The AR coefficients vector is used as the input for multilayer perceptron network for learning and classification.

The method for regular or irregular shape detection and discrimination based on combining pattern recognition techniques such as a morphological processing with concepts from artificial intelligence and machine learning such as genetic algorithms is described in Bala & Wechsler, (1993). In this approach, a genetic algorithm is used to generate the population of morphological expressions for shape discrimination.

Shape can be also described in terms of its boundary and, in this case, shape description can be reduced to curve description. Measurement of such properties of curve as inflection **(**Monkhtarian & Mackworth,1986).and curvature **(**Rangarajan, Shah & VanBrackle, 1989) can be used for curve description. Also, various types of local features such as the value of maximum distance between a chord and its arc, ratio of arc length to chord length on a curve, can be detected by drawing chords (Philips & Rosenfeld, 1987). The coefficients of the discrete Fourier transformation of boundary points can be also used as the shape descriptors ( Jain 1986, Ghosh & Jain, 1993).

Works of art (abstract paintings) consist of regions (picture elements) of different shapes. For a chosen class of pictures (a generic class, see chapter 4.4) an appropriate method of shape analysis is needed. In the present research, as a generic class, the Mondrian class is choosen. The pictures of the Mondrian class consist of rectangular shapes and a set of picture descriptors are obtained based on picture model (see chapter 4.5). The detailed description obtaining a set of picture descriptors (including shape) for the Mondrian class will be given in chapter 4.7.

### 2.3.2. Texture

Texture can be viewed as a global pattern arising from the repetition, either deterministically, or randomly, of local patterns and can be defined as a repetition of the basic texture elements called texels (Zucker & Terzopoulos, 1980). In painting texture relate to the surface of the picture and is a result of applying different techniques to evoke an artistic effect.

Texture elements are arranged according to placement rules which can be random or regular. For example, Pollock's picture, which was obtained by dripping a liquid paint from a can or spattering it in long arabesques with a stick, possesses a texture which can be described as a “random” placement of texture elements. This class of pictures (Pollock class) can be analysed by applying methods of “random” texture modelling and description such as the Markov Random Field (Cross & Jain, 1993).

Also, some of the picture regions (picture elements) possess a characteristic texture which can be analysed by using methods of texture analysis. There are many methods of texture analysis, such as autocorrelation function method, texture transforms method, congruence matrix or co-occurrence matrix (Tomita & Tsuji 1990, Kasturi & Jain, 1991) which can be used to obtain a set of picture descriptors.

In the present research, as a generic class the Mondrian class is chosen. The pictures of this class do not posses a characteristic texture. However, some of the pictures of this class are created in such a way that the paint is not uniformly spread on the picture plane. It gives an impression of patterns and can be treated as a “specific” texture which could be analysed by applying the method of texture analysis. In this research this “specific” texture will not be analysed. It will be a topic for further research.