

# Concept, Content and the Convict

Mika 'Lumi' Tuomola, Teemu Korpilahti  
Jaakko Pesonen, Abhigyan Singh  
University of Art and Design Helsinki (Taik)  
Hämeentie 135 C, 00560  
Helsinki, Finland

mika.tuomola@taik.fi, teemu@tmukoo.net,  
jaakko.pesonen@kolumbus.fi,  
abhigyan.singh@gmail.com

Robert Villa, P. Punitha  
Yue Feng, Joemon M Jose  
University of Glasgow  
Sir Alwyn Williams Building  
Glasgow, Scotland

villar@dcs.gla.ac.uk, punitha@dcs.gla.ac.uk,  
yuefeng@dcs.gla.ac.uk, jj@dcs.gla.ac.uk

## ABSTRACT

This paper describes the concepts behind and implementation of the multimedia art work *Alan01 / AlanOnline*, which wakes up the 1952 criminally convicted Alan Turing as a piece of code within the art work - thus fulfilling Turing's own vision of preserving human consciousness in a computer. The work's context is described within the development of associative storytelling structures built up by interactive user feedback via an image and video retrieval system. The input to the retrieval system is generated by *Alan01 / AlanOnline* via their respective sketch interfaces, the output of the retrieval system being fed back to *Alan01 / AlanOnline* for further processing and presentation to the user within the context of the overall artistic experience. This paper, in addition to presenting the productions and image retrieval system, also presents the installation and online production user reception and some of the issues and observations made during the development of the systems.

## Categories and Subject Descriptors

J.5 [Arts and Humanities]: Arts, fine and performing; H.3.3 [Information Search and Retrieval]: Search process

## General Terms

Design, Human Factors

## Keywords

Alan Turing, Art installation, Online production, Image retrieval

## 1. INTRODUCTION: THE CONVICT

"The difference between man and machine is not how they operate but how they are treated." [30]

Alan Turing (1912-1954), a World War II code-breaker considered one of the fathers of modern computing, made a significant and provocative contribution to the debate regarding artificial intelligence: whether it will ever be possible to say that a machine is conscious and can think. As a person he was "an ordinary English homosexual atheist mathematician" [13]. For his wartime achievements he was awarded the OBE, Officer of The Most Excellent Order of the British Empire. In 1952, Turing was convicted of "acts of gross indecency" after admitting to a sexual relationship with a man in Manchester. He was placed on probation and required to undergo hormone therapy to achieve temporary chemical castration. The therapy caused Turing's body to develop female forms, and the conviction resulted in his

security clearance being revoked. He died after eating an apple laced with cyanide in 1954. The death was ruled a suicide.

Not surprisingly, several pieces of drama have been inspired by Turing's life, including *Breaking the Code* by Hugh Whitmore [41] and *Turing* by Miko Jaakkola [14]. The internationally awarded *Breaking the Code* starred Derek Jacobi as Alan Turing in London's West End and New York's Broadway theatres at the end of 80s, while Jacobi's performance was immortalised by the 1996 BBC production of the play [40]. Jaakkola's *Turing* was originally performed by the Helsinki based Circus Maximus in 2000. It had a rebirth recently, when Opera Skaala turned the script into the critically acclaimed multimedia opera *Turing Machine* [35].

Both plays depict Alan Turing as a man who wished to beat death by coding human consciousness. Such an interpretation is supported by the significant biographical facts of Turing's life. At the age of 18 he lost Christopher, his best friend and love – and due to that event, also his faith in religion and life beyond death. Man was needed to invent the eternal life to save loved ones.

The theme is carried on in *Alan01 / AlanOnline*, in which engagers will be able to meet a fictional Alan in dialogue, interactively, as if his consciousness had indeed been carried on by a machine's code. This time the story is experienced by an associational structure suitable to computers, and in describing behaviours of human consciousness: "...the narrative is formed by a series of moments which are linked by common elements and do not rely on chronology or episodic relationships to produce their meaning or effect" [26]. The structural choice also reflects the ACM Multimedia 2009 art exhibition theme of "disophrenia" and the fragmentation and timelessness of consciousness that is transferred into digital representation [2]. Due to the work's interactive nature, the "negotiated narrative" [32] between the authored system and its users, spatial installation design and relatively large database of available media material, it also discusses with Janet Murray's notions of narrative in digital environments as procedural, participatory, spatial and encyclopaedic [21].

The research production carries on the continuing research on associational and new forms of interaction structure and narrative. In *Alan 01 / AlanOnline*, we are particularly interested in looking into user/engager and associational system interaction logics: how will interaction modes and tempo regulate and be influenced by story tempo and mode? Previous production-based research has included, for instance, *Accidental Lovers* [1], which discovered a

clear correspondence between user interaction and story tempo, though free user interaction available at all times [37]. In addition, we were interested in using the image retrieval to convert non-textual input into textual concepts and symbols. As a concept this is fascinating and opens up endless opportunities in the context of interactive art, for instance as a tool for creating sophisticated installation and artwork interfaces.

In this paper we present two art works, one an online interface, the other a physical installation, which use image and media retrieval as the method of driving the interaction between users and the installations, and the associational narrative possibilities enabled by the systems. In *AlanOnline*, users can also interact by drawing and sketching, generating a visual stimulus, which is fed into the image retrieval system. The results generated by the image retrieval system then result in a cascade of other actions in the works themselves, allowing users to interact with the multimedia installations in a visual manner. Physical interaction in *Alan01* and the drawing interface in *AlanOnline*, body movement produced media representations of *mind*, carry on the “disophrenia” theme further.

The paper is structured as follows: in the next section the *Alan01* / *AlanOnline* cross media artwork concept is presented, which is then followed by a description of both of these productions. The image retrieval system used in these productions is then described, along with a detailed description will be given of the retrieval techniques used and an initial evaluation of these techniques. This is followed by a short description of the productions’ audience reception, conclusions and future work.

## 2. THE CONCEPT: Alan01 / AlanOnline

Media and image retrieval (see chapter 3) play central roles in the multimedia art productions that consist of two individual parts: *Alan01* – a physical installation, and *AlanOnline* – the installation’s online counterpart.<sup>1</sup> Both of these systems engage interactive audience in dialogue with a fictional Alan, as if Turing’s consciousness had been coded into a machine at the time of his death. The user/engager can “talk” to *Alan01* and *AlanOnline* by a system of symbols, which we imagine to have been relevant to Alan Turing’s life. In *Alan01* the symbols are physically arranged into sequences by the installation user, while in *AlanOnline*, the physical interaction is simulated by a drawing interface for the symbols’ retrieval. The *AlanOnline* system starts with a way of allowing the user to draw an image, which is then passed to the image retrieval system, which returns one or more matching symbols from a hand-built collection. The search result or results then trigger an associational story in each of the productions, which will be described in sections 2.1 and 2.2.

The retrieval is made from a limited set of approximately 50 symbolic images, the graphical presentation and selection of which having been made while bearing in mind the nature of the interface. The symbols are similar to what a user might draw in a short time of five to ten seconds. Another requirement for the selection of the symbols is that they are connected to the context of Alan Turing’s life. A couple of examples of the images: bird, hand, fish and heart are shown in Figure 1.

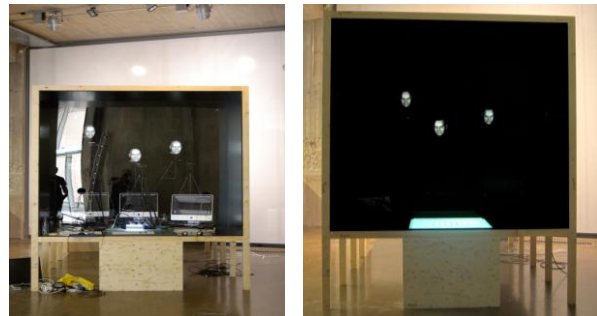
The idea of the drawing interface is therefore to enable a non-textual input to an art piece, which can still be translated to symbols and their textual meanings. From there on, the associational narrative structure script of the art piece can start to function.



**Figure 1: Examples of the symbolic images shared by the installation and its online version**

### 2.1 The Alan01 Installation

The physical installation (Figure 2) is composed of a black darkened room in which is placed a horizontal touch screen, a series of three monitors for displaying videos and user feedback, plus three three-dimensional heads, which function as video projection surfaces.



**Figure 2: Series of monitors and three-dimensional heads displaying videos (left), glowing display and touch screen (right)**

The touch screen provides the interface for interacting with engagers (the glowing horizontal display in Figure 2); its surface is sandblasted to make it suitable for video projection, images being projected via a mirror upwards from beneath the glass. Under the table there is also a video camera that captures any movement that happens above the cube. When a user presses her finger against the glass, the finger reflects the bright white light from the video projector. The software of the installation’s main computer detects these bright blobs on the surface and projects black “paint” blobs on the corresponding spots. This experience of drawing/choosing symbol sequences on the glass is somewhat similar to finger painting. Within the experimental drawing interface that remains to be used in public in future (for the first show of the installation we used choosing the symbols by finger or an object), the interface remains untouched for a couple of seconds, the system displays a timer icon on the screen, which functions as feedback for the user, hinting that unless they resume drawing, the system will move on. Once the timer finishes, the user’s image is saved locally on the computer. This image is then used as input to a media retrieval system, which will return a result symbol, which is then presented to the user. The aim of this setup is to create a sophisticated input device without utilizing traditional buttons or keyboard.

A simple XML file serves as the database and structure for the content of the installation. In the XML file each symbol, as

<sup>1</sup> The productions’ website: <http://mlab.taik.fi/alanonline/>

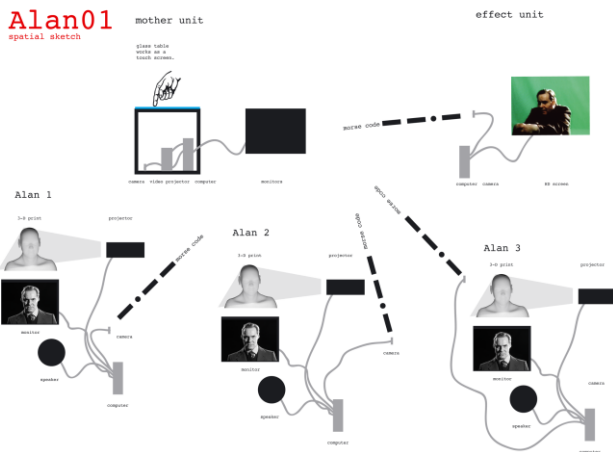
returned by the search, has a set of associative words, the number of which varies from one to five. The associative word is selected randomly from these potential matches and displayed on the surface with the resulting image. Connected to each word there is a light signalling code, structured as either short or long flashes, similar to Morse code, but not mapped to letters like standard Morse code. The main unit sends this code sequence by blinking the touch surface, the flashing code then being detected by other parts of the installation.

Figure 3 illustrates the principle of communication between the installation units. The use of this traditional communication method is a historical reference to the work of Alan Turing as a code breaker during World War II and can also be seen as a reference to how this tradition is still amidst us. Morse code and sending messages with directional spotlights are still used by modern armies as a close range communication method, due to the difficulty of intercepting such signals. From the point of view of the user's experience, it makes the flow of data visible.



**Figure 3: Light signal communication between installation units**

The installation has three receiving units that monitor the main touch screen with web cameras. These receiving units time the duration of the light sequences on the screen and the duration between them (see Figure 4 for a conceptual sketch of the installation). Each receiving unit has an identical XML file that serves as the code key for translating the light message, the length of which varies from one to seven flashes, back to the corresponding textual meanings.



**Figure 4: Conceptual sketch for the Alan01 installation**

Various different media assets are also associated with each code in this XML file. The receiving unit reacts to the message it received by showing the related content through its three output channels: video material of an actor portraying Alan Turing mixed with a 3D animation made with a model of the same actor's head are shown on a screen, as well as projected onto a physical miniature statue that is printed from the same 3D model. The emotional effect of the video projection bears resemblance to

Tony Oursler's projections on balloon-heads, for example, in *System for Dramatic Feedback* [23], while the physicality of the 3D statue might remind the user of the uncanny robotic puppetry of Ken Feingold's *Head* [6]. However, the combination of "virtual" projection onto a statue created from the same 3D model is quite unique.

Dialog sentences are projected onto a surface and played as audio that has been produced with text-to-speech synthesis. In the XML file each machine has its own designated assets, so no two units will react in the same way. The video images are emotional reactions to the symbols and words, while the dialog is spoken by a monotonous machine voice that reveals aspects of the personal history and personality of Alan Turing.

### 2.1.1 Morse code and light signalling

The concept for the light signalling communication between computers was initially developed during a Pure Data workshop held by M. Koray Tahiroglu in Media Lab Helsinki during September 2008. The idea was to create a system where two computers could have a conversation with each other, in a way where the flow of data is visible to the viewers. The possibility of allowing errors where one machine misunderstands the other or where a viewer intervenes would then make the conversation between the two machines interesting and unpredictable.

In the system the sending and receiving units share a common code key that translates predefined words to sequences of light flashes (similar to Morse code) that vary in length from one to seven flashes. Each of the flashes is translated to be either long or short, and every burst of light within one message is followed by a short moment of darkness. Each complete message is followed by a dark period, the duration of which is at least twice the length of the short darkness. In this way the system defines when a message is completed. Adjusting the duration of light and dark periods affects the susceptibility to interference. The longer the durations are, the less likely the system is to misunderstand the message.

Within the context of the installation, a user may manipulate the system by blocking a code sequence. The importance of a user's body for communication thus becomes articulated within the installation environment. The system can, for instance, misinterpret a long sequence such as "--...--" for a shorter one like "--...--", if the user decides to block the receiving camera half way through the message. Since the code key is built in the fashion of a pyramid, shorter messages always have some meaning coded into them. Thus it is more likely that short messages are received without error and that longer messages are likely to be mistaken for shorter ones.

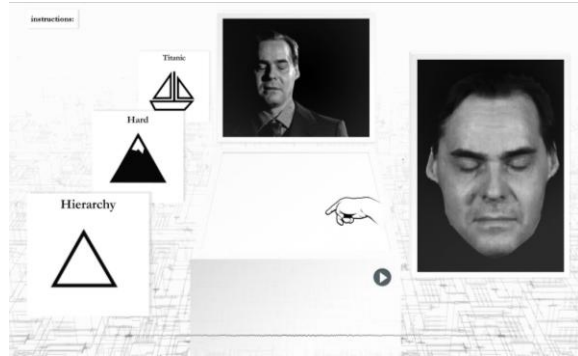
## 2.2 AlanOnline

The online counterpart of the physical installation shares many of its media assets with the installation. The image retrieval system is used in the same fashion and the logic of the connections of the media content in the XML file is similar. The media files are also the same, with the exception that the content is compressed and modified to be suitable for a web browser interface. The different characteristics of physical media installations and non-material artwork, which only exist as software, are an important factor when a physical installation and online version are designed around the same theme, using shared assets. In the case of *AlanOnline*, the browser version is not an attempt to copy the physical installation, but instead to create a similar yet individual

art piece that offers another interface to the same content, using the media elements and providing interaction and output in a way that is most suitable for its presentation media and technology. This is to ensure that the design choices of the online piece are valid, so that it can exist as well as if no physical installation was ever created. Lev Manovich [19] describes the nature of the underlying question in the following way: "Synthetic computer-generated imagery is not an inferior representation of our reality, but a realistic representation of a different reality." Figure 5 shows the present version of *AlanOnline*.

In *AlanOnline* the user faces a white landscape that on closer examination reveals itself as a massive circuit diagram that disappears to a distant imaginary horizon. In the centre of this pseudo space is the main interface – a white slightly translucent cube. Where in the installation there is the main unit and its three subunits, here there is only one machine performing all the tasks, the user being figuratively placed inside the machine.

The user's tool of interaction is a black and white representation of a human hand, with its index finger extended. The top of the white cube is the canvas of the online version, which the user can draw on in much the same fashion as in the physical installation. Once the drawing is finished the Flash application sends the image back to the server, where it is used for the image retrieval that is made based on the same symbol set as in the installation.



**Figure 5: A screen shot of the AlanOnline production**

The result image is uploaded and the result's media content is retrieved. The code that was used in the installation as a visible means of communication, is now revealed to the user as an audible series of telegraph style "dit" and "dah" sounds, which in addition to the historical reference, offer the user feedback that the system is processing data, and actions will soon follow.

Once the associative word has been chosen, it is connected to the image. The dialog sentence is presented as text and played back using the synthesized voice. The voice is played using virtual surround sound so that the source of the sound seems to be behind the user, which enhances the users feeling of being in the centre of the machine. As an additional visual connection between the voice and the imaginary computer a sound wave spectrum reacts to the voice. Following the sentence, instances of the connected emotional video responses of the actor and 3D animations are then played in different parts of the virtual space.

### 3. IMAGE RETRIEVAL SYSTEM

The backend retrieval system is responsible for the retrieval of the images used in *AlanOnline* and is planned to be used in future versions of *Alan01*. It is based on a heavily developed version of

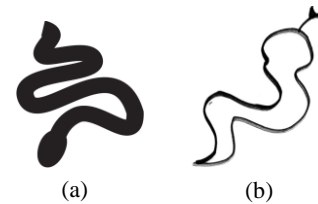
the system described in [38], much expanded with the ability to search images, videos, and external web search services. While essentially, its development has been altered to reflect the differing and challenging requirements of the *Alan01* and *AlanOnline* productions.

The first point to be made is that the visual retrieval system is only required to generate a single result: a single result is used by *Alan01* as the initialising point for the sequence of other actions; in the *AlanOnline* production, it is possible to display multiple results, to provide multiple responses to the input, although this is not considered here. This requirement calls for high precision results, with the added requirement that the system must always generate a result – even if there is nothing in the collection which is similar to the input query, the *closest* should always, ideally, be selected.

Secondly, the matching required by the productions is largely visual in nature – each query input, whether generated by the touch screen in *Alan01* or the online drawing canvas of *AlanOnline*, is only required to be visually similar to the result image in some manner. But this visual similarity should be *understandable*, i.e. users should be able to intuitively see why a result was produced for a query, or be able to determine a property of the search result which matches the input query. If the system cannot return results which cannot be interpreted as similar, the risk is that those interacting with the productions will be less likely to engage with them.

As mentioned in Section 2, for the specific needs of the installation a limited set of roughly 50 images were created (Figure 1). The retrieval system is expected to retrieve the most semantically similar image from this small collection of 50 images, but for an arbitrary hand drawing generated by *Alan01* / *AlanOnline* via their respective sketch interfaces.

Since the collections we are using in our case do not belong to a specific domain, it is difficult to predict if the boundary based or region based algorithms will perform better. Additionally, since we have a very small collection of shapes and the queries are expected to be generated by *Alan01* and *AlanOnline* through an interface where a user can draw any arbitrary image, the similarity retrieval of an object from the collection becomes a big challenge. For instance, the image in the database for the object snake shown in Figure 6(a) is very different from the user drawn snake shown in Figure 6(b) and yet the system should ideally retrieve the snake shown in Figure 6(a).



**Figure 6: Examples of collection and query images**

Almost all methods proposed so far for shape retrieval, have a fixed model dataset, containing images which may be occluded, distorted or otherwise transformed. The free style drawings in the collection open up a new research question in the field of shape representation and retrieval. Hence, we have used a variety of boundary based and region based features to investigate the

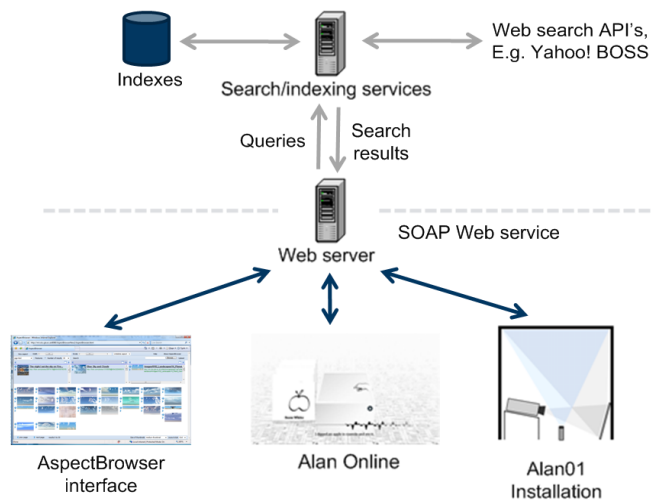
suitability of these feature for free style hand drawn shape recognition and retrieval as explained in later sections.

### 3.1 Architecture

The system architecture is shown in Figure 7, and provides a backend retrieval system which deals with the indexing and retrieval of images and videos, a SOAP web service interface which is used by *Alan01* and *AlanOnline* to carry out content-based searches automatically, and a separate web interface which can be used by the artistic developers to experiment and search their collections of images. This latter search interface is called the “AspectBrowser”, an earlier iteration of this interface having been evaluated and described in [36].

The SOAP web service provided by the backend retrieval system exposes a number of different services, including:

- Search: a search can be carried out based on one or more image examples and/or text (when available). Which features to use in the search, the collection to search, and the number of results required can also be changed
- Return the list of available collections: this returns the list of collections which can be searched
- Get available features for a collection: each collection may be indexed using different types of features, this method allows a client to discover the types of feature which can be used to search in a given collection



**Figure 7: Retrieval system architecture**

Image indexing and retrieval of the *Alan01* and *AlanOnline* data set is achieved via the use of various low-level visual features such as colour, edge, texture, etc, each image being represented via these low level features, and retrieval being carried out via image example. For use in *Alan01* and *AlanOnline*, a number of shape features (also known as *shape descriptors*) were investigated, each of which will be described in the following sections.

## 3.2 Shape Descriptors

### 3.2.1 Edge Histogram

The spatial distribution of edges in an image is a very useful descriptor for similarity search and retrieval [18]. To compute an edge histogram descriptor, the image is divided into 16 blocks and

then the local-edge distribution for each block is represented by a histogram. In total, 80 histogram bins are required to represent each edge histogram of all blocks.

Since considering the block-edge histogram alone may not be sufficient for image matching, global-edge descriptors are also implemented in addition to local block edge descriptors. Additionally, edge distribution information for the whole image, horizontal and vertical semi-global-edge distributions, as well as local edge distributions are also used to improve the matching performance. The global-edge histogram and semi-global-edge histograms are estimated from the local 80 bins. The global-edge histogram is calculated by accumulating the five types of edge distributions for all blocks. The semi-global-edge histograms are estimated from the grouped blocks, which are grouped in the following ways: grouping of four vertical blocks, grouping of four horizontal blocks and grouping of four neighbour blocks. In this case, 13 different segments are created. The corresponding edge histograms for each segment are then calculated using the local-edge histograms. After combining the local, the semi-global and the global histograms, a new histogram with 150 bins is constructed for similarity matching.

To calculate the similarity between two images in the edge domain, the following distance measure using two edge histograms A and B is adopted:

$$EdgeSim(A, B) = \sum_{i=1}^{80} |h_A(i) - h_B(i)| + 5 * \sum_{i=1}^5 |h_A^g(i) - h_B^g(i)| + \sum_{i=0}^{65} |h_A^s(i) - h_B^s(i)|$$

where  $h_A$  and  $h_B$  are the normalized histogram bin values of image A and B, and where  $h_A^g$  and  $h_B^g$  are the normalized histogram bin values for the global-edge histograms of image A and image B, respectively.

### 3.2.2 Contour Shape

Object shape features provide a powerful clue to identity [4]; the contour shape extractor used in the system is implemented in two steps: (i) locate the objects in the image and extract its outer contour, followed by (ii) extract the contour features.

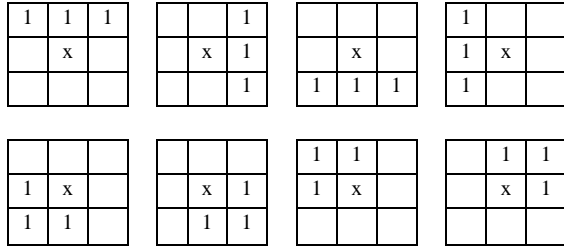
Given an image, it is first transformed to obtain a monochrome image where object and background are represented in contrasting colours. An edge extractor is first applied to obtain the edge information of the object, followed by a morphological open-close operation in order to smooth the computed contours and connect any breakage in the contours.

Although there are many edge detection algorithms available, we make use of very simple convolution in the spatial domain to obtain the boundaries of the object. The structuring elements (also called masks or filters) shown in Figure 8 are used to obtain inner and outer boundaries of objects in an image.

From all such generated region contours, the largest closed contour is selected and used to represent the object, since it is usually the outer boundary of the image object, thus preserves object shape. An example of the largest extracted contour of an apple image is illustrated in Figure 9. This is with the assumption that there is only one object in a image but with many regions within itself.

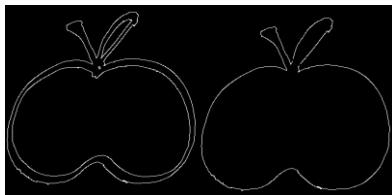
Given the extracted contour, we follow the contour in a clockwise manner and keep track of the direction as we go from one contour pixel to the next, represented using a chain code [16]. Given a

contour pixel, the next contour pixel is a pixel from its 8-connected neighbours. A unique number, from 0 to 7, is used to represent each direction. Looping through all the contour pixels, any contour can be represented by as a numerical array for similarity matching.



**Figure 8: The structuring elements used for finding boundaries of an object in the image**

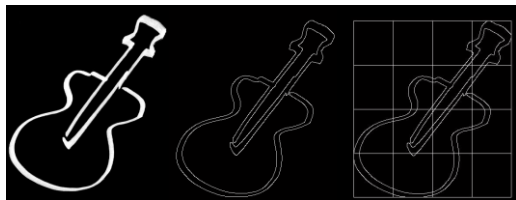
Using chain codes is efficient because of the constraints on their construction. Only a starting point is represented by its location; the other points on the shape curve are represented by successive displacements from grid to grid along the curve. Since the chain code is invariant under boundary rotation, it makes the similarity matching between two contours relatively easy.



**Figure 9: The extracted contour of apple**

### 3.2.3 Object Signature

Due to the fact that the shape features for irregular, free style drawings can be very complicated, we have also studied the pixel distribution of the object in an image. However, depending on the pressure an artist applies on the instrument used to draw the picture, the concentration of pixels in regions may vary. It is more appropriate therefore to thin the boundaries of the object to obtain a single pixel thickness shape. This helps to generate a fair matching irrespective of the thickness due to different pressures. However, employing any thinning algorithm will distort the shape of the object, hence we extract the boundaries of the object as explained in the previous section.

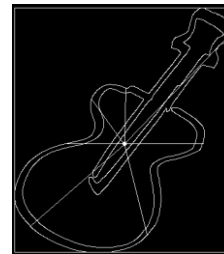


**Figure 10: The boundaries of guitar and the minimum bounding rectangle**

Once the boundaries of the object are extracted, a minimum bounding rectangle is used to fit the object. A minimum bounding rectangle (MBR) is the smallest rectangle that can completely contain an object in an image. The object along within the MBR can be normalized, i.e., scaled down or up, to obtain a standard sized image for easier matching. The minimum bounding

rectangle is then partitioned into a number of blocks, as shown in Figure 10. The ratio of object pixels and the background pixels are computed for each block and are recorded as a feature descriptor.

The boundary pixels of the object are used to generate another feature descriptor for the image, a “Signature” for the object. Using every object pixel on the object boundary, the centre of the object is computed. With this centre, the object is scanned in a counter clockwise direction from 0 degree to 360 degrees with a certain interval. The distance of the pixel at, say,  $d$  degrees, from the object centre is computed and recorded as another feature descriptor. Figure 11, shows an instance during the generation of the centroid profile for a guitar object, showing the direction angles at intervals of 45 degrees.



**Figure 11: Centroid profile showing the direction angles at intervals of 45 degrees, used in the Object Signature feature**

## 3.3 Fusion

Each of the image features described in Section 3.2 generate a separate result list for a given query image. Since a query may be relevant to multiple different collection images in multiple different ways, we require a method of fusing the search results generated by the different features into a single result list. After fusion of result lists, we can remove the top ranked image as the final result for use in *Alan01* and *AlanOnline*.

In order to achieve this, we have implemented and tested six different fusion methods:

- Reciprocal rank, which is a simple summing of the reciprocal of the rank of each result [22]. For an image  $i$  and rank list  $j$ , the reciprocal rank is defined as:

$$r(image_i) = \frac{1}{\sum_j 1/position(image_{ij})}$$

Where  $position(image_{ij})$  is the rank position of image  $i$  in list  $j$ .

- The Borda count method, which gives a number of points to each image at a given rank, and then sums the points to determine the final ranks. We use the system described in [22], where if there are  $n$  possible image results across all ranked lists being merged, an image at rank  $i$  will be given  $n-i$  points.
- The Condorcet method takes account of the relative positions of the ranked images, building up a matrix of which image results come above, below, or are tied in rank with other images. Again, we use the system described in [22].
- A voting method, which counts the number of ranked lists an image result is part of, and then ranks primarily on this number [29].
- The minimum, maximum and sum of the similarities generated by the different features for each result image. In this scheme,

the similarity scores for an image result across all ranked lists are used as input, where a new score is generated by taking the minimum, maximum or the sum of the scores [7].

The first four of these methods are based solely on the ranks of the results, the final three instead utilise the similarity scores generated by the similarity matching function used in retrieval.

### 3.4 Evaluation

In order to evaluate which of the combinations of features and fusion techniques produced the best results, a small evaluation was carried out which simulated a series of known item searches. Two image collections were used, the first being the collection of target images, the other a set of hand-drawn query images.

As a first step, a set of relevance judgements was manually created, where each query image was matched to its ideal target image (an example is given in Figure 12). In addition to an ideal “target”, we also defined for each query zero or more alternative images which were deemed to be acceptable results for the query, but not ideal (Figure 13). This latter list was defined to enable us to consider the matching between query and target(s) as a fuzzy mapping, in order to model that which was thought to be “acceptable” in the implementation of *Alan01* and *AlanOnline*, where the output generated need not always be exact. Indeed, the collection may not necessarily contain any images similar to the query.



Hand drawn query image Target image from

Figure 12: Example query image and associated target image

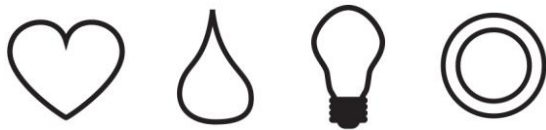


Figure 13: Alternative matches for the query in Figure 12

Once the manual relevance judgements were constructed, each query image was matched to 47 collection images, for each combination of feature, fusion technique, and ten different rank list depths. In addition to the three features defined in Section 3.2 which consisted of Edge Histogram (Eh), Contour Shape (Cs), and Signature (S), two other colour based features were used for comparison purposes: Colour Histogram (Ch) and Colour Layout (Cl). Both of these features were as defined by the MPEG-7 standard [4]. While these features were not expected to perform as well as the others, they are commonly used in a number of other retrieval situations, and provide an interesting comparison to the others used.

The fusion techniques were as described in Section 3.3, consisting of the reciprocal rank (Rrank), voting (Vote), Borda, Condorcet, Min, Max, and Sum methods. Since each of these fusion methods operate on ranked lists, we also altered the depth of the ranked lists fused to between one and ten, in order to investigate the impact of fusion depth. By increasing or decreasing the depth, we introduce more or fewer image results for fusion. In the situation

where only a single feature is used to perform the retrieval, the fusion is of course not required.

Table 1 shows the ten feature, fusion and rank list depth combinations which identified the greatest number of correct target images (left), or the greatest number of alternative target images (right). Combinations are coded as a list of features, the fusion technique, followed by the ranked list size. Note that it is only if the top result of the ranked list matches either the target image or one of the alternative targets will a correct result be counted – if the correct result appears at rank position two or below, the combination will be counted as a fail. While absolute, this reflects the needs of the *Alan01* and *AlanOnline* systems. Additionally it should be noted that when searching for an alternative target image, the ideal target is also considered as a correct response to the query.

Table 1: The percentage of target images and alternative images (including target) identified by the top ten features, fusion and result list combinations

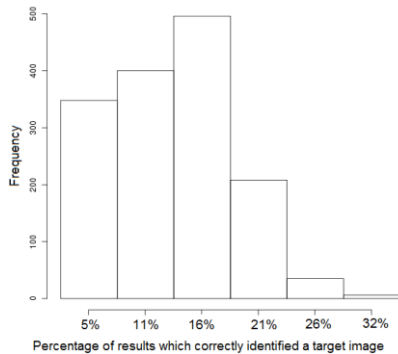
% of ideal targets		% of alternative targets	
Combination	% Corr	Combination	% Corr
Eh/Cs/S-Vote-4	32	Cl/Cs/S-Vote-10	47
Cs/S-Vote-7	32	Cl/S-Vote-9	42
Cs/S-Vote-4	32	Cl/Eh/Cs/S-Vote-10	42
Cs/S-Vote-3	32	Cl/Cs/S-Vote-9	42
Cl/Eh/Cs/S-Vote-10	32	Cl/Ch/Cs/S-Vote-9	42
Ch/Cs/S-Vote-3	32	Cl/Ch/Cs/S-Vote-10	42
Eh/S-Rrank-9	26	Cl/S-Vote-8	37
Eh/S-Rank-10	26	Cl/S-Vote-10	37
Eh/Cs/S-Vote-3	26	Cl/Eh/Cs/S-Vote -9	37
Eh/Cs/S-Min-10	26	Eh/S-Vote-4	32

It can be seen in Table 1 that the best techniques for identifying the target image are correct for 32% of the queries; when also including the alternative targets, this increases to over 40% of queries correct. The best combinations for both situations are, however, different. For detecting a single ideal target image, the top 6 combinations include the Contour Shape and Signature features at various sizes of rank list. When also allowing for the alternative target images, the Colour Layout is present in 9 out of the top 10 best combinations, while a combination of Colour Layout and Signature performs as well as other combinations which include Contour Shape.

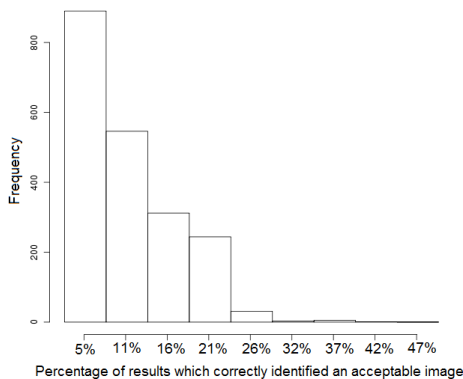
Looking at the fusion techniques, the voting method dominates the best performing combinations in both retrieval situations. Rank list size does vary, however, with there being a trend for smaller ranked lists when retrieval aims to only return the single ideal target image. The best performing combination which is common to both situations combines Colour Layout, Edge Histogram, Contour Shape, and Signature together with voting fusion and a ranked list size of 10.

To expand on Table 1, Figures 14 and 15 show the distribution of the different retrieval combinations for different performance levels on the x-axis. Figure 14 shows that there are a significant number of possible combinations (over 200) which can return the correct image 21% of the time. On Figure 15, to retrieve any of the images classed as acceptable, it can be seen that the

distribution is skewed to the right – only very few of the combinations performed better than 26% correct, although there are again very many which were able to perform at a level of 21% correct results.



**Figure 14: Number of retrieval combinations for different performance levels (retrieval of single target image)**



**Figure 15: Number of retrieval combinations for different performance levels (retrieval of any acceptable target image)**

### 3.5 Previous work

The problem of recognising arbitrary shaped images irrespective of various geometric transformations, such as, orientation, scaling, translation and shearing effects has been tackled in the fields of Robotics, object recognition and computer vision, among others. Object recognition with shape features deals with finding a match between certain features obtained from the shape of a query submitted with that of the different instances of the model objects in the database.

A number of approaches have been proposed for 2D object recognition based on shape features that can be categorised into two groups: boundary based methods or region based methods [10, 24, 15, 16] depending on the feature extraction mechanism.

The boundary based methods extract global/local features from the outer boundaries of an object. The simplest way of representing a shape is via a chain code representation, where the direction of the neighbour pixel to a reference pixel is recorded while a boundary is traversed in a clockwise or a counter clockwise direction. Shape boundaries, also known as contours, are represented, by many researchers, using statistical moment variants [17, 31, 34, 43]. Zahn and Roskies, [42], Persoon and Fu, [27], and Wallace and Wintz [39] used Fourier descriptors to describe object shape boundaries. Although Hough transforms

have a very high computational cost, they are well accepted for better recognition of lines and curves. Ballard [3] and Pao et al. [25] proposed methods based on a generalised Hough transform to represent shapes by a set of tangent lines. Some researchers found that besides curve and lines in a shape some dominant points are more useful for shape representation. Shape boundaries were represented on the basis of convex and concave curvature points by [5, 8]. Based on the behaviour of the local region pixels, an incremental circle transform was proposed by Han et al. [12] to take care of any effect due to image orientation. Gary and Mehrotra [9] represented shapes by local structural features.

Despite the fact that boundaries are very helpful to preserve the shape of an object, they are very sensitive to noise – the presence of additional pixels or the absence of a few pixels on the boundary completely varies the representative feature of the object. Most methods which work on boundaries are sensitive to starting points and also to the directions in which the boundary is traversed. They generally work in only limited setups. In addition to these disadvantages, there are many applications where the boundary features are not relevant and hence demand the use of region characteristics, i.e., the characteristic of an object within the boundary of the object.

Statistical moments being applicable to all areas in general, were also used to compute region features by a few researchers [33, 31]. Other, features such as those of the MPEG7 framework was used in [28]. A few simpler region features can be found in [10]. Some researchers have also worked on combining boundary features and region features for effective retrieval and wider applicability of the methods. Among these, methods combining moments were proposed by Mehtre et al. [20]. To avoid any discrepancy in the crisp value representation mechanisms, shapes were represented by symbolic features in [11].

## 4. RECEPTION OF THE PRODUCTION

The *Alan01* installation was placed on display from the 4<sup>th</sup> of June, 2009. The first version of the online counterpart was also published at the same time. The first production versions’ audience reception was investigated 16-22 June 2009. In total 26 users, 11 males (42%) and 15 females (58%) participated in the evaluation by filling out the survey form. Out of these 18 (69%) were interviewed.

Overall experience of *Alan01* Installation received excellent feedback with 95% positive score. Quality of overall production was found to be very high. Amongst the individual attributes, Quality of Head Animations, Graphic Design and Video clips received excellent response with 96%, 86% and 82% positive scores respectively. In the interviews these were found to be most appreciated and engaging features of the installation. *AlanOnline*’s overall experience received high scores (76% positive) and was found easy to use (43% positive and 29% average). Amongst individual attributes, Graphic Design was most appreciated (53% positive). Users identified relationship between drawing and resulting symbols as weak (47% average and 28% negative). Experience due to meaningfulness of user’s action received neutral scores for both *Alan01* Installation (60% average) and *AlanOnline* (47% average). Whereas recognition of relationship between user’s action and interface response received equally distributed results for both *Alan01* Installation (43% positive and 30% negative) and *AlanOnline* (51% positive and 33% negative).

Three independent academia and industry expert reviews (by a design researcher, museum visitor generated content researcher and interactive media writer & director) were generally praising, though narrative and spatial coherence between the two pieces was wished to be made more substantial in future.

## 5. CONCLUSIONS: THE CONTENT

The *Alan01* and *AlanOnline* concept and production present a very demanding challenge to an image retrieval system, where a visual retrieval is required which is consistent with the behaviour of the human visual perceptive system, and which can present high precision results for input queries. The level of accuracy of the results has to be high to ensure that the retrieval within the production doesn't just become a technical gimmick. Instead it should be an integral part of the artwork.

In the present production one of the central issues has become the *predictability* of the retrieval results, an aspect which is not typically considered as important in the field of image retrieval. In an interface that uses this technology, the user is tempted to start to test the system or even play against it. Seeing the results which the system has delivered previously affects the imagery that a user starts to draw henceforth. If a user tries to replicate the images he/she has seen in previous results, the following results need to be consistent in order to avoid the feeling of randomness in the system, and to ensure that the communicated illusion of the installation is not broken.

When the resulting image set is limited and preselected, the level of graphic detail also needs to be relative to the system's ability to recognize details. In the case of the experimental art productions, an added factor is limiting the number of search results. In a conventional image retrieval system, if the user is presented with the top ten retrieval results, it is typically considered sufficient if a significant ratio of those results is relevant – the fact that the highest ranked result isn't relevant doesn't render the whole result list unusable, which is the case here.

From the point of view of the retrieval, the implementation of an image retrieval system tailored to the needs of the *Alan01 / AlanOnline* production has been a very challenging endeavour. While the image collection to be searched is small, the reality has been that creating a content based image retrieval system which is acceptable for production use has been extremely difficult, and is likely to continue to be difficult. Problems include the lack of training data, and the difficulty in the judgment of relevance within the context of the work. While initially the aim was to concentrate purely on visual similarity, in practice this is difficult: we naturally think in terms of the semantics of the image. I.e. the knowledge that an image is an "apple" or "computer" can override particular visual similarities which may be present between two images. Working on the problems raised by the *Alan01* and *AlanOnline* productions has resulted in a rethinking of the needs and roles of the content-based image retrieval.

The ability to convert non-textual input into textual concepts or symbols is fascinating and opens up endless opportunities in the context of interactive art. Image retrieval technology shows great promise as a tool for creating sophisticated installation and artwork interfaces, potentially allowing the creation of rich user interfaces, but ones which can still be used immediately by a visitor, whether adult or child.

The first phase tests and audience reception of the *Alan01 / AlanOnline* delivery systems' ability to produce engaging mini-narratives is promising. Quality of overall production was found to be very high. With the logic and tempo of the associational and procedural narrative structure script working, future concentration is on the rhetoric of physical space, user body movement, and the use of moving image and sound to support the logic. The final success of the production's future will be defined on how we continue to tie up the physical space and online rhetoric with the system logic. The independent interactive media writer & director expert reviewer (see chapter 4) commented on the production: *"This ambitious installation addresses a complex topic, experimenting boldly with dramaturgy as well as content and form. It is a genuine work of visual and computer-enhanced dramatic art, reaching high aesthetic standards, with a confident sense of integrity and through-composition, both in subject and medium."*

## 6. ACKNOWLEDGMENTS

This research was supported by the European Commission contract FP6-027122-SALERO. *Alan01 / AlanOnline* is produced by Crucible Studio, Media Lab, University of Art and Design Helsinki (producer: Tea Stolt).

## 7. REFERENCES

- [1] Accidental Lovers. 2006. Dir. Mika Lumi Tuomola. Finnish Broadcasting Company YLE Channel 1 & Crucible Studio, Media Lab, University of Art and Design, Helsinki.
- [2] ACM Multimedia 2009 Interactive Art Program Call for Exhibition Entries and Papers  
<http://www.acmmm09.org/IAP.aspx>
- [3] Ballard, D.H. 1981. Generalizing the Hough transform to detect arbitrary shapes. *Pattern Recognition* 13, 111-122.
- [4] Bober, M. 2001 MPEG-7 visual shape descriptors, In *Circuits and Systems for Video Technology*, IEEE Transactions on , vol.11, no.6, pp.716-719, Jun 2001
- [5] Dinesh R and Guru D. S., 2004, Recognition of Partially Occluded Objects Using B-tree Index Structure: An Efficient and Robust Approach, *ICVGIP*, 246-251.
- [6] Feingold, Ken. 1999. Head. <http://www.kenfeingold.com/> [the site last visited April 27, 2009]
- [7] Fox, E. A. and Shaw, J. A. 1994. Combination of Multiple Searches. In D. Harman (Ed.) *The Second Text Retrieval Conference (TREC-2)*, Gaithersburg, MD, USA
- [8] Fu, A.M.N., Yan, H., 1997. Effective classification of planar shapes based on curve segment properties. *Pattern Recognition Lett.* 18, 55-61.
- [9] Gary J E., and Mehrotra R., 1993, Similar shape retrieval using a structural feature index, *Information Systems*, Vol. 18, No. 7, 523-537.
- [10] Gonzalez R.C and Woods R. E, 2002, *Digital Image Processing*, Prentice Hall, ISBN 0201180758.
- [11] Guru D. S., Nagendraswamy H. S., 2007, Symbolic representation of two dimensional shapes, *Pattern Recognition Letters*, Vol 28, 144-155.

- [12] Han, D., Bien, Z., You, B.J., 1994. A theory of generalized incremental circle transform and its application for recognition of two-dimensional objects. *Pattern Recognition Lett.* 15, 769–780.
- [13] Hodges, Andre. 1983. *Alan Turing: The Enigma*. Simon and Schuster, New York.
- [14] Jaakkola, Miko; Lehtonen, Jussi. 2000. *Turing*. Circus Maximus theatre, Helsinki.
- [15] Khotanzod, A. and Y.H. Hong. 1990. Rotation invariant image recognition using features selected via a systematic method. *Pattern Recognition* 23, 1089-1101.
- [16] Kim, H.K., Kim, J.D., 2000a. Region-based shape descriptor invariant to rotation, scale and translation. *Signal Process.: Image Comm.* 16, 87–93.
- [17] Kim, W.Y., Kim, Y.S., 2000b. A region-based shape descriptor using Zernike moments. *Signal Process.: Image Comm.* 16, 95–102.
- [18] Manjunath, B. S., Salembier, P., Sikora, T. 2002 *Introduction to MPEG-7*, WILEY, ISBN 0 471 48678 7
- [19] Manovich, L. 2001. *The Language of New Media*. Massachusetts Institute of Technology
- [20] Mehtre B. M., Kanakanahalli M. S., and Lee W. F., 1997, Shape measures for content based image retrieval: A comparison, *Information Processing and Management*, Vol. 33, No.3, 319-337.
- [21] Murray, Janet H. 1997. *Hamlet on the Holodeck: The Future of Narrative in Cyberspace*. New York, London: The Free Press.
- [22] Nuray, R., and Can, F. 2006. Automatic ranking of information retrieval systems using data fusion. In *Information Processing and Management*, 42, 595-614
- [23] Oursler, T. 1994. *System for Dramatic Feedback*. <http://tonyoursler.com/> [the site last visited April 27, 2009].
- [24] Ozugur T., Denizhan Y., and Panayirci E., 1997, Feature extraction in shape recognition using segmentation of the boundary curve, *Pattern Recognition Letters*, Vol. 18, 1049-1056.
- [25] Pao, D.C.W., Li, H.F., and Jayakumar, R., 1992. Shapes recognition using the straight line Hough transform: Theory and generalization. *IEEE Trans. Pattern Anal. Machine Intell.* 14 (11), 1076–1089.
- [26] Parker, Philip. 2002. *The Art and Science of Screenwriting*. Intellect, London
- [27] Persoon, E. and K-S. Fu. 1977. Shape discrimination using Fourier descriptors. *IEEE Trans. Syst. Man Cybernet.* 7, 170-179.
- [28] Prasad , B. G., Biswas , K. K., Gupta, S. K. 2004, Region-based image retrieval using integrated color, shape, and location index, *Computer Vision and Image Understanding*, v.94 n.1-3, p.193-233
- [29] Punitha, P., Urruty, T., Feng, Y., Halvey, M., Goyal, A., Hannah, D., Klampanos, I., Stathopoulos, V., Villa, R., Jose, J. 2008. Glasgow University at TRECVID 2008, TRECVID Workshop at NIST in Gaithersburg, MD.
- [30] Saarinen, L. 2008. *Turing Enigma*. Script for the experimental production *Turing Enigma* (Crucible Studio / Media Lab, University of Art and Design Helsinki 2008) at <http://fullhouse.uiah.fi/turingenigma/> [the site last visited April 10, 2009].
- [31] Taubin, G., Cooper, D.B., 1991. Recognition and positioning of rigid objects using algebraic moment invariants. In: *SPIE Conf. on Geometric Methods in Computer Vision*, vol. 1570, pp. 175–186.
- [32] Tuomola, Mika. 2002. Drama in the digital domain: Commedia dell’Arte, characterisation, collaboration and computers. In: Beardon, Colin & Malmborg, Lone (editors). *Digital Creativity, A Reader - Innovations in Art and Design*. Swets & Zeitlinger. p. 217.
- [33] Teh, C.-H. and R.T. Chin. 1986. On digital approximation of moment invariants. *Computer Vision, Graphics, and Image Processing* 33,318-326.
- [34] Teh, C.H., Chin, R.T., 1988. On image analysis by the methods of moments. *IEEE Trans. Pattern Anal. Machine Intell.* 10 (4), 496–513.
- [35] *Turing Machine (Opera)*. 2008. Dir. Janne Lehmusvuo, libretto (based on Jaakkola’s play *Turing*) by Taina Seitovirta, composition by Eepi Ursin & Visa Oscar. Opera Skaala & Crucible Studio, Media Lab, University of Art and Design, Helsinki.
- [36] Urban, J., Hilaire, X., Hopfgartner, F., Villa, R., Jose, M., Chantamunee, S., Gotoh, Y. 2006. Glasgow University at TRECVID 2006. *Proc. TRECVID 2006 - Text RETrieval Conference TRECVID Workshop*. Gaithersburg, Maryland
- [37] Ursu, M. F., Thomas, M., Kegel, I. C., Williams, D., Tuomola, M. L., Lindstedt, I., Wright, T., Leuridijk, A., Zsombori, V., Sussner, J., Myrestam, U., Hall, N., 2008. *Interactive TV Narratives: Opportunities, Progress and Challenges*. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 4, 4 (Oct 2008)
- [38] Villa, R., Gildea, N., and Jose, J. 2008. *FacetBrowser: a user interface for complex search tasks*. *ACM Multimedia 2008*: 489-498
- [39] Wallace, T.P. and Wintz, P. A. 1980. An efficient three-dimensional aircraft recognition algorithm using normalised Fourier descriptors. *Computer Graphics and Image Processing* 13,99-126.
- [40] Wise, H (director) 1996. *Breaking the Code*. British Broadcasting Company (BBC), UK.
- [41] Whitmore, Hugh. 1986. *Breaking the Code*. Theatre Royal, Bath & London. Also Neil Simon Theatre 1987, New York.
- [42] Zahn, C.T. and R.Z. Roskies. 1972. Fourier descriptors for plane closed curves. *IEEE Trans. Comput.* 21, 269-281.
- [43] Zhu, Y., De Silva, L.C., Ko, C.C., 2002. Using moment invariants and HMM in facial expression recognition. *Pattern Recognition Lett.* 23, 83–91.