# Generative Portrait Sketching

Patrick TRESSET and Frédéric FOL LEYMARIE*

*Digital Studios, Goldsmiths College, University of London, U.K.*

**Abstract**

We consider the problem of generating art with a computer system, as it relates to the understanding of a drawing style. Our system in its present development is able to sketch faces automatically, starting from a picture, typically a photographic snapshot of a scene with humans. Once a digital image is considered, the system automatically finds where some of the faces (or face-like patterns) are and isolate these. Each face image pattern is then used to produce a stylised portrait. The style which the system aims at using is derived from the one the first author has developed over the years in his artistic practice. The goal is for the final system to *mimic the process* developed by the artist, rather than aiming at results exactly reproducing his way of drawing a portrait. Nevertheless, the produced sketches are in the style of the artist. Various steps in the process of producing a sketch by the artist were isolated, including: image segmentation, filtering, shape selection and depiction, filling and shading. The implementation of these steps relies on an understanding of human visual perception, of the artist's work process, and of advances made in computer vision.

## 1. Introduction

The first author, the artist whose style we will study and re-interpret via a computerised system, can vividly remember drawing his first faces. They were copies of sketches by Leonardo da Vinci. Since this initiation he has been fascinated by faces and has drawn a multitude of these. As a child he rapidly noticed the fascination a beautifully, or expressively drawn face exercises on an audience. Over his years of practice his technique has evolved to become detached, dexterous and spontaneous. The faces are now sketched in one or very few complex continuous movements (Figure 1).

After some research and experiments the artist came to the conclusion that it was possible to design a system capable of automatically drawing faces in his style. Tresset had learnt to program an old ink plotter and had used the available simple functions to start producing various experimental drawings. It seemed possible to refine this approach by better understanding the possibilities of the algorithmic modification of images, and linking this approach with a breakdown of the artist's way of working toward a final piece. To further explore this research possibility Tresset joined forces with Leymarie at the Digital Studios of Goldsmiths College in 2004.

But why create a system able to do drawings an artist can do in a few seconds? On one hand, we are interested in this problem because the public has a tendency to recognise the artist in the work produced and this tends to bias the artwork appreciation. We value art as a universal language of the sublime, of beauty, of the struggle of human intellect to understand

*Published in the "Proceedings of the 11th International Conference on Virtual Systems and MultiMedia" (VSMM'05), Hal Twaites, ed., ARCHAEOLINGUA, publ., pp. 739–748, Oct. 2005, Ghent, Belgium.

Figure 1: Examples of face sketches by Patrick Tresset; inks on paper; 2002.

and represent the world. We do not think it is important for art to become an expression of the ego. The artist should sit behind the work rather than in front of it. By creating programs generating art the artist is moved a step away. One could argue that we do not want to take responsibility for the artwork or that we do not want to be associated with the beauty we create. Our answer is that our goal is to simply *reveal* beauty.

On the other hand, we wish to better understand how beauty is perceived, how variations of style influence the perceptual outcome. Our approach in this research is to proceed by studying the process of creation of a particular artist, namely Patrick Tresset. This shall represent a first step in a larger study where other styles and influences can also be studied in a similar fashion. The hope is that such an understanding, if applicable to a multitude of artists, of styles, shall give us a deeper knowledge of the perceptual processes involved in both creating and perceiving art.

We report in this paper on our progress toward developing a fully automatic computerised system able to produce portraits in the style of Patrick Tresset. In the following section we review the literature on previous related attempts at creating such systems. Then, in section 3, we describe Tresset's approach to creating face sketches and how this process can be broken down in a sequence of steps. Each such step can then be represented and studied in isolation. This leads us to describe, in section 4, a system where each step can be understood as an algorithm acting on the result of the previous step, in a sequence mimicking the artist's hand.

## 2. Background

### 2.1. Systems producing portraits automatically

In the recent years, a few interesting projects have been reported which aim at producing portraits, often from a photographic input. Giving aesthetically pleasing results is the *Floccugraph* by Golan Levin [17]. Starting from a photograph, the darker areas of (what is assumed to be) a face are used to define a probability field for generating random tugging or repulsion forces on a series of hair-like filaments: around darker areas the turbulences on a filament are more active. One of the aesthetic qualities of this project is in the use that is made of a gray background with black and white filaments. The images produced remind us of certain traditional 16th century drawings such as da Vinci's drapery studies. One of the limitation of this system is that it does not take into account if it is drawing a portrait or not: it would draw a shoe exactly in the same manner. Another recent project giving interesting results is the "Woman Portrait Generator" by Celestio Soddu [15] inspired by Picasso's paintings but giving results in the form of 3D sculptures. This system consists of two main steps. First, a set of algorithms represents how each feature (eye, nose, face lines, and so on) is generated in the style of a Picasso (of a given period). This constitutes the generative notebook, a set of procedures reacting in a random fashion towards an artificial environment (*i.e.*, what has been generated so far); the combination of these procedures can be used to generate complex structures. Second, the construction of the artwork is then organised through an evolutionary process, letting the artwork evolve to generate multiple representations of the same idea. This step is based on a genetic algorithm to produce an endless series of portraits each different, but still recognisable as a "Picasso." Soddu has further used the same genetic principles to generate towns, building and so on.

In some scientific, computing or commercial based systems, the programs take into account that they are drawing a face and nothing else. Different strategies are then deployed.

The system described by Liang *et al.* generates *caricatures* from a photographic input in two stages [11]: *(i)* in a training phase and *(ii)* at run-time. During the training phase the system "learns" how the artist exaggerates certain features. For this an artist has been asked to make caricatures from a set of photographs; then, on each element of each pair photo/caricature a set of chosen feature points is labelled. The system then analyses the correlation between each pair using a least square algorithm, after which it identifies a set of exaggeration prototypes, each set depending on the type of faces caricatured. At run-time the system extracts the features from the image, turns the face into one of the prototypes, and then applies the corresponding distortion. This system is restricted to frontal representations, yet produces caricatures resembling the model. As with any system with a training phase it is not versatile and the training set can become cumbersome. The system is further constrained by allowing only for a simple line drawing to be produced, and, as a result, it would hardly be able to reproduce any kind of rhythm, flow in the line arrangements. However, the way the training phase is implemented is interesting as it can be useful to study how distortions are applied.

The system presented by Hong Cheng *et al.* generates portrait drawings in the *Manga* style [2]. The system again is designed in two stages: *(i)* training and *(ii)* run-time. During the training stage, a set of Manga drawings is created by an artist from a set of photos. Then, an ensemble of prototype drawn features is created corresponding to certain features found in these real images, *i.e.*: certain types of lips, noses, eyes, eyebrows, and so on. At run-time global information about the face is collected in a "map," such as the position and scale of the different features. Then, each feature is classified as belonging to one of the prototypes. The corresponding drawing is then warped to the scale of the original features. Each feature is finally positioned according to the global map. This system is not that powerful at producing Manga drawings *resembling* the model as a limited number of types of features is used when drawing faces.

The system presented by Pei-Ying Chiang *et al.* generates *Manga* style *caricatures* from photographic input [3]. The system localises 119 nodes categorised in eight groups corresponding to the MPEG-4 codification.[1] Then, from the various facial features extracted from the photographic input, the face is warped using the found nodes. This system, even if simple in conception and implementation, gives results of surprisingly good likeness and quality. But the output is closer to a painting than a drawing. It is not likely to give interesting results

[1] MPEG-4 is an international standard which permits the integration of the production, distribution and access of multimedia content; *e.g.*, refer to http://www.mpeg4.net/

with a pure line drawing.

All the systems above differ fundamentally from the one we are implementing as they are only imitating the drawing style of an artist or draughtsman, *i.e.*, the end result, rather than the processes involved when drawing. Still, some of the introduced strategies, such as the use of templates created during training time, are of interest to us. Below, we describe two recent systems which attempt to study the creative process itself rather than just the end result.

### 2.2. Systems simulating an artist's vision for artistic purposes

The system described by Matthew Brand starts with the picture of an animal and finishes with a three dimensional sculpture [1]. The singularity of this system is that it tries to mimic an artist's way of forming a mental representation of reality. The system is designed to work in four steps:

1. Analyse the image and identify salient parts.
2. Build a schematic mental image that represents the salient regions.
3. Make an initial design by projecting this representation into a new medium.
4. Refine the result to reinforce resemblances to the original image and impose some formal aesthetic qualities.

Different traditional computer vision techniques are deployed to achieve this. It has to be noted that as for the system we are implementing, Brand's system is also a type of *expert system*. The part of the system of greatest interest to us is designed to extract information from bodies. During the first stage the system tries to identify salient parts of interest for the "artificial artist" such as the anatomical structure of the subject; for this a combination of gradient extraction, boundary eliminations, mid-pass filtering, skeletonization and other traditional computer vision algorithms are applied in an expert manner to extract meaningful information from the image such as: muscles position, main joints, limbs and so on.[2] This system is conceptually and practically original, using sophisticated ideas.

The approach described by Doug de Carlo and Anthony Santella to stylise photographs based on *visual attention* [4, 13], is also very interesting both conceptually and in its results. In their approach the system starts from a photograph. A human user then interacts with the system to identify the meaningful content of the image. This is modelled by having the user look at the image for a short period of time, during which a map is built with attention data gathered from an *eye-tracking* device. The map represents which elements of the image carries important information. This map is then used to create a simplified picture which reflects the selection of zones of greater impact on visual attention. The images produced are in our view a very convincing stylisation of the perceived reality. They can be seen as a good simulation of the type of images produced by a trained illustrator. Another interesting feature of the system is that an untrained user can produce work of high quality. The system could be seen as a kind of *cyber-artist-hand*.

The above represent recent approaches which attempt at capturing a style of rendering a picture. We will now go into the details of our own approach, which first requires to understand how the artist proceeds when drawing a portrait.

## 3. Tresset's Style of Drawing Portraits

There are many methods of drawing perceived reality. Still the generic initial stage follows a common thread. The artist begins by observing the scene or the subject, be it an object, a

[2] The rest of the system builds an articulated 3D geometric mobile from the extracted information.

figure, a landscape, and so on, to identify a salient structure. The draughtsman then has to establish the structuring lines or surfaces, and position these on the medium. This combination of observation and structure depiction has to be repeated until enough structure is established, using the lines already drawn as elements of comparison. Then the drawing process proper starts. The presence of structure allows the drawing to be executed in a spontaneous yet precise manner. Spontaneity is important if one wants to achieve a good drawing as it brings naturalness, flow and rhythm. This part depends a lot on the style of drawing to be achieved. Shading follows, where shadows are represented by some patterns generally integrating one, two or all of the following qualities: intensity, a gradient orientation, and gradient speed. Some artists might start the shading before the drawing proper, which results in a technique closer to painting, often used with charcoal, chalk, or pastels, giving the possibility to "sculpt with light" on a 2D canvas.

With experience it becomes possible to do without the first step — which still is present as an inner visualisation of the mind. The style of the drawing is largely dependent on the distortions occurring during the first stage (whether laid out explicitly on the canvas or in one's mind projection), on the quality of the lines used for the drawing, on the ways of translating the curves, and on the patterns used for shading.

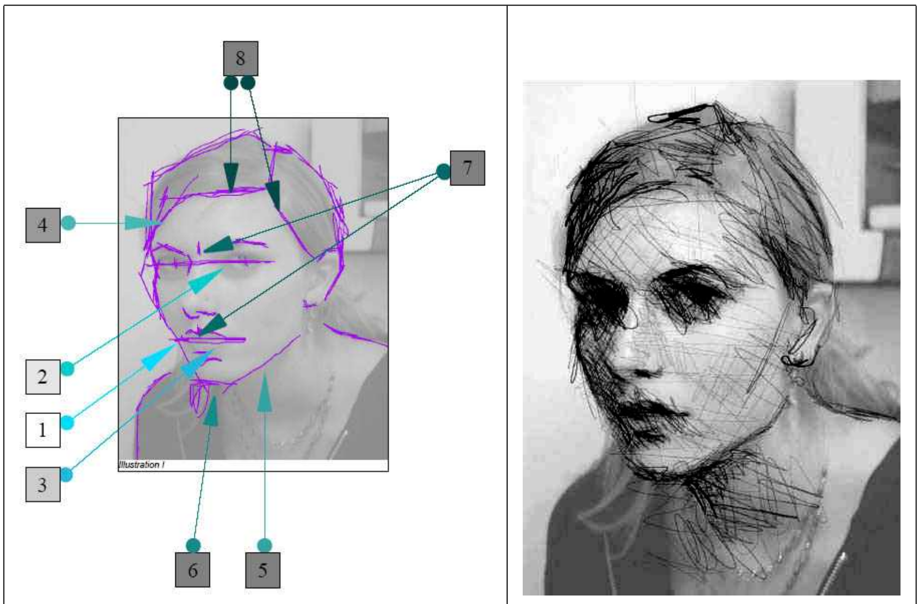
In this work, the modelled drawing process is not the most precise, but the *most spontaneous* technique Tresset uses, which gives the most dramatic effects. The aesthetic effect of the drawing, its "mood," is due in large part to its apparent imprecision. Yet, to control the "mood" and achieve a desired effect, the drawing must have a certain level of accuracy. Also, because of the execution speed, there is a coherence in the arrangement of the lines creating a flow effect (Figure 1).

### 3.1. Finding salient structures: An example

The process of finding salient structures is an acquired skill. It starts by measuring proportions or angles with a pencil or brush held at arm's length. After some time of training the same process can be performed without the need for the held pencil. Tresset performs a sweeping of lines, scanning his field of vision in different directions. The speed of the scan along a line decreases with the number of salient regions it hits. If the scan slows down under a certain speed, a record is made of a salient structuring line. When the process is done mentally a series of distortions and errors are more likely to happen, thus influencing very strongly the style of the resulting drawing.

When drawing a face we are looking for specific structuring lines as described below (Figure 2.(a)):

1. Position lines indicating the direction of edges of the face only if the slope is significant.
2. Position the pair of lines going through the middle of the eyes. The slope is important. Place the middle of the pupil.
3. Position the line going through the middle of the mouth; it is (nearly) parallel to the line for the eyes.
4. Position the edge of the forehead, using the intersection with 1 and 2 (above).
5. Position the line of the jaw, its main orientation.
6. Position the bottom of the chin.
7. Position marks for the middle of the forehead and middle of the mouth.
8. Position other structuring lines.
9. Roughly draw the mouth, positioning marks for the bottom of the nose, eyebrows directions, ears.




(a) Example of structure evaluation; numbers correspond to list in §3.1

(b) Shading process example.

Figure 2: An example of the drawing process.

The above outlined process of drawing structuring lines is performed by a constant feedback mechanism, where the artist visually evaluates the (partial, on-going) drawing, locates intersections and compares the structure with his/her mental memory of what needs to be drawn first. Note that the human visual system proves not very good at generating absolute values *(i.e.*, accurate positions), but is quite good at relative ones, for slopes, lengths, and the general topological layout of relations among structures.

The shadows are then placed iteratively, first by very roughly locating shadows around the eye sockets, then under the nose, under the mouth, and under the chin (Figure 2.(b)). The shadows structuring the volume are then placed. More shadows are added, and regions thus emphasised get darker and more precise. Some contours are placed such as for the eyebrows, the sides of the face, the nostrils, the mouth shape, the details of the eyes, and such. Lines and shadows are positioned on the basis of observation and experience.

## 4. An Automatic Computerised "Portraitist:" AIKON

In its present implementation, AIKON (Automatic IKONic drawing) is a relatively simple system originally written in the "python" programming language (www.python.org), and now being ported to "java" (www.java.sun.com) for greater flexibility and availability. AIKON's outputs (Figure 3) evoke Tresset's scribbled face sketches (Figure 1).

After being presented an image, or after collecting an image from the environment, *e.g.*, via the Internet or with a camera, AIKON first locates regions likely to be representative of human skin. Then, it eliminates regions which are unlikely to be faces, and extracts the best remaining face-like region. Follows a shading process, obtained by segmenting the selected region in different gray-levels. Finally, the system controls a plotter to perform scribbling movements of varying density corresponding to the extracted gray-level maps (Figure 3). Such drawings are presently being realised on an old-fashion pen-plotter (OCE 8500), allowing the use of high quality traditional drawing paper and China ink. The results are real

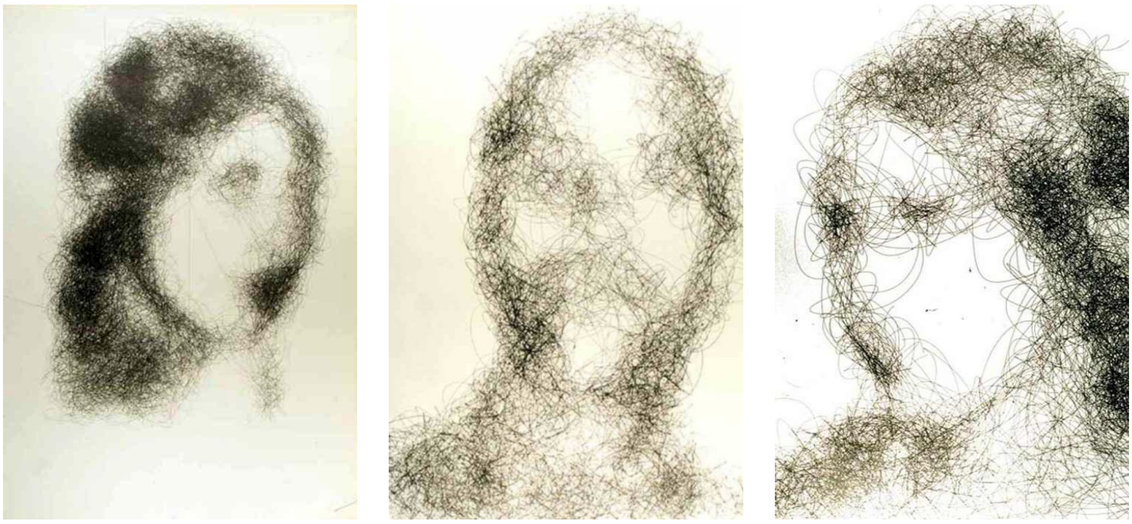


Figure 3: Examples of plotted drawings on the *OCE 8500 pen plotter, A1* format, on Arches paper, 2004.

drawings not prints.

The plotter is driven in real-time by the software; furthermore, the instructions to the plotter are not kept in record. In this way each drawing can be kept as a *unique* piece. The size of the drawings produced ranges between A6 to A0. The system's simplicity does not hamper the sketches from being aesthetically pleasing to the eye, and artistically interesting; to some degree these sketches can even be unexpected in result as they do not precisely reproduce Tresset's touch (compare Figure 3 with Figure 1).

### 4.1. Face localisation

There has recently been a lot of research in the field of automatic face detection mainly due to the growing interest in surveillance and identification for security purposes. There are some well defined strategies for facial localisations in colour or grayscale space. One of the most efficient and practical method is to first segment out the skin-coloured pixels and then eliminate regions which cannot count as "faces" [5].

Different approaches are available to label a pixel as "skin" or "non-skin," for example as studied in the survey by Vezhnevets *et al.* [16]. The results for skin segmentation do not sufficiently differ to prefer one method over another. But a combination of different methods is likely to diminish the risk for false negative while it may result in slightly more false positives to be eliminated further on. From experience, the "CIE L*a*b" colour space seems to give less false negatives and is less sensitive to ambient light than other colour spaces tested.[3]

Prior to skin segmentation it is useful to apply a colour constancy algorithm, to perform a colour correction compensating for the effect of the illuminant light colour. Three different methods are commonly used for colour constancy purpose:[4] *(i)* The *grey world* method assumes that the average of all colours in a picture is gray. By computing the average of the pixels in the image it is possible to deduct what is the illuminant color. *(ii)* A common method favored by photographers using a digital software package consists of performing an

[3]The "CIE L*a*b" is a colour space calculated in a way to be more consistent with human perception. For details refer for example to the website of Charles Poynton: http://www.poynton.com/ColorFAQ.html .

[4]For details, refer for example to: http://www.cs.sfu.ca/∼colour/research/colour-constancy.html .


Figure 4: Face localisation examples: (a) Source image, Armelle, 2005 (natural light). (b and c) Skin colour regions; colour corrected: (b) gray world, (c) normalisation. (d) Detected face enclosed by bounding box. (e) Source image, Anaelle, 2005 (natural light). (f) Detected face in bounding box. (g) Source image, Maeliss, 2003 (flash light). (h) Detected face in bounding box.

*histogram stretching* of the red, green and blue components of the image. *(iii)* The *white patch* method assumes the presence of bright white pixels in the picture, representative of the ambient light.

Once candidate skin regions have been segmented, we can use the following criteria to further rule-out regions not likely to be part of a face [5]: (a) wrong ratio of height/width, (b) wrong ratio of perimeter/area, (c) region without a mouth, eyes, (d) wrong triangle mouth–eyes. Examples of skin and face detection are given in Figure 4.

### 4.2. Shadows

There are several possible ways to implement the shading process. We make use of $K$-means clustering [8] to quantize and segment the colour face region into different grey level patches and create a set of binary maps used during the drawing process (Figure 5.(a,b,c)).[5] Two different manners of treating the shading are then implemented: *(i)* shading along a medial axis, and *(ii)* random shading.

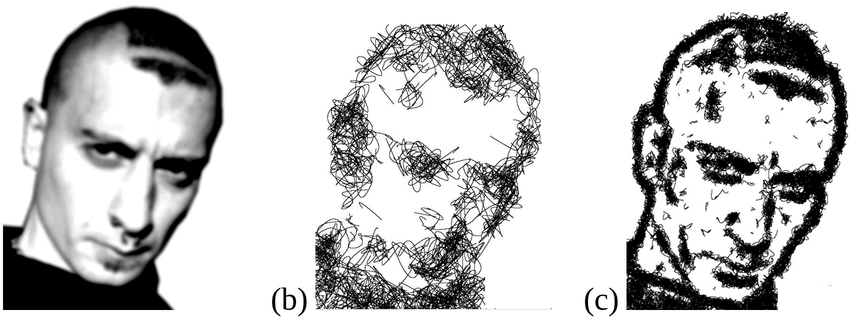
To perform shading along a *medial axis* [10], the outline contour of each region is first smoothed using the "curvature morphology" technique [9, 7] to reduce the number of spurious short axes.[6] Then the path of each axis is traversed and recorded (Figure 5.(d)). During the shading stage, along each axis a point $O$ is moved at a certain speed $s$; the width $w$ at the position $O$ and an orientation angle $\alpha$ are also measured; a sinusoidal is traced with an amplitude $w:2$ and slanted at an angle $\alpha$; the speed $s$ varies in relation to the gradient speed

[5] $K$-Means clustering is an iterative automatic method to generate a specific number ($K$) of disjoint groups of pixels or clusters.

[6] The medial axis of a region or an outline is a skeletal graph located in the middle of the region bounded by the outline [10].

(a) (b) (c) (d) (e) (f)

Figure 5: Example of *medial-axis based shading.* (a) Segmented face region, Maeliss, 2003. (b) K-Mean clustering after blurring. (c) Binary map representing two gray levels. (d) Medial axis map of binary map in (c). (e) Example of shading executed using (c) and (d). (f) Example of the shading process for four segmented levels.



(a) (b) (c)

Figure 6: Examples of *random shading*. (a) Source image, preprocessed manually, Guilio, 2002. (b) Example of random shading extracted from (a). (c) Another variant of random shading for a different set of parameters.

in the original image (Figure 5.(e,f)). This previous step can then be repeated with varying angles $\alpha$ depending on the desired effect.

To perform *random shading*, we first randomly select an initial location $L$ on a binary map, and an additional $n$ points $P$; $L$ and $P$ are then used to calculate Bezier curves [14] whose control points are also randomly chosen in circular areas of varying diameters. Each control point is calculated to allow for a smooth transition between the succeeding curves, which create together a scribbling effect of length of factor $n$. This process can be iterated as a function of the number of black pixels in the map (Figure 6).

### 4.3. Drawing

Presently, AIKON draws sketches on different outputs including old fashioned pen plotters. There are a few motivating factors encouraging us to use pen plotters. They have been and are still used by a group of pioneering computer artists called “Algorists” [12]. Pen plotters have the advantage of really drawing, on paper, using various means: it is the combination of the paper and pen movements that creates the drawing, making it possible to use traditional inks, watercolours, as well as various paper types. The results have far more materiality and great flexibility in the choices (inks, papers, etc.) given to the artist. Another interesting possibility to be explored would be to create etchings, giving the potential of a classical art’s medium.

**5. Conclusions**

We have presented a progress report toward developing a sophisticated automatic computerised system, called AIKON, able to produce portraits in the style of the first author. We have shown an outline and preliminary results of the first implemented prototype. We are in the process of developing a more mature and ambitious system which will depend on a refined systemic approach to model the artist's creative process, as well as include more diverse and sophisticated computer vision techniques. The systemic representation shall depend on the use of *saliency maps* (§2), inspired in particular by the work of Laurent Itti on the simulation of visual attention [6], and the work of Matthew Brand [1].

# References


[1] Matthew Brand. Artificial artist, March 1997. http://alumni.media.mit.edu/∼brand/artist.html.

[2] Hong Chen et al. Example-based composite sketching of human portraits. In *Proceedings of the 3rd International ACM Symposium on Non-Photorealistic Animation and Rendering*, pages 95–153, New York, USA, 2004.

[3] Pei-Ying Chiang, Wen-Hung Liao, and Tsai-Yen Li. Automatic caricature generation by analyzing facial features. In *Proceedings of Asian Conference on Computer Vision (ACCV)*, Jeju Island, Korea, January 2004.

[4] Doug DeCarlo and Anthony Santella. Stylization and abstraction of photographs. In *Proceedings of SIGGRAPH*, pages 769–776, San Antonio, Texas, USA, 2002.

[5] Rein-Lien Hsu, Mohamed Abdel-Mottaleb, and Anil K. Jain. Face detection in color images. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24:696–706, May 2002.

[6] Laurent Itti. *Models of Bottom-Up and Top-Down Visual Attention*. PhD thesis, California Institute of technology, Passadena, CA, USA, January 2000.

[7] Andrei C. Jalba, Michael H.F. Wilkinson, and Jos B.T.M. Roerdink. Automatic segmentation of diatom images for classification. *Microscopy Research and Technique,*, 65:72–85, 2004.

[8] T. Kanungo et al. An efficient $k$-means clustering algorithm: Analysis and implementation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):881–892, July 2002.

[9] F. Leymarie and M. D. Levine. Curvature morphology. Technical Report TR-CIM-88-26, CIM, McGill University, Montreal, Canada, December 1988.

[10] F. Leymarie and M. D. Levine. Simulating the grassfire transform using an active contour model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(1):56–75, January 1992.

[11] Lin Liang et al. Example-based caricature generation with exageration, 2002.

[12] Verostko Roman. The algorists, historical notes, 2004. http://www.verostko.com/algorist.html.

[13] Anthony Santella and Doug DeCarlo. Visual interest and NPR: An evaluation and manifesto. In *Proceedings of the 3rd International Symposium on Non-Photorealistic Animation and Rendering (NPAR)*, pages 71–78, Annecy, France, 2004.

[14] Peter Shirley. *Fundamentals of Computer Graphics*. A.K. Peters, 2002.

[15] Celestino Soddu. Argenia, art's idea as generative code. In Annie S. Wu, editor, *Proceedings of the Genetic and Evolutionary Computation Conference*, 2000.

[16] Vladimir Vezhnevets, Vassili Sazonov, and Alla Andreeva. A survey on pixel-based skin color detection techniques. In *Proceedings of GraphiCon*, pages 85–92, Moscow, Russian Federation, September 2003.

[17] Carlo Zanni. Interview: Golan Levin. *Magazine Electronique du Centre International d'Art Contemporain de Montréal (CIAC)*, (19), Summer 2004. http://www.flong.com/resume/short.html.