INFORMATION IN A CULTURE-ALERT SOCIETY
Can Photogrammetry measure the Visual Perception of Works of Art ?
Invited Keynote paper

Petros Patias

President of ISPRS Commission V
President of CIPA
The Aristotle University of Thessaloniki, Greece
Department of Cadastre, Photogrammetry and Cartography, Faculty of Surveying Engineering
patias@topo.auth.gr

KEY WORDS: Heritage, Art, Entropy, Digital filters, Eye-gazing

ABSTRACT:

The intention of this presentation is to propose a new point-of-view, from which a quest for innovative applications of Photogrammetry in assessment of cultural heritage can be initiated. As such, the main point rests on the presentation of new ideas, rather than on completely resolving the involved technical issues. Wölfflin's principles in for comparative examination of pieces of art and architecture, belonging to different art periods are evaluated with photogrammetric means. The "clarity" of figures is assessed with edge-enhancement and image segmentation. The perceived "depth" is assessed through virtual emboss of the images and creation of "virtual" stereo pairs. The image information is assessed through entropy calculations. The existence of "focal points" is assessed through interest-point operators and the development of an eye-gaze tracking system. Examples from Renaissance and Baroque artists are given, to prove the feasibility of the approaches.

1. PREAMBLE

The intention of this presentation is to propose a new point-of-view, from which a quest for innovative applications of Photogrammetry in assessment of cultural heritage can be initiated. As such, the main point rests on the presentation of new ideas, rather than on completely resolving the involved technical issues.

2. MOTIVATION

From the dawn of its conception, Photogrammetry based its relative advantage to other surveying techniques on the novel and attractive idea of "*Measuring what cannot be touched*". Starting with recording the optical spectrum, the above principle was re-phrased to "*Measuring what cannot be touched but can be seen*". For many years, all sorts of techniques for recording, measuring, modeling and reconstructing 2D and 3D objects, which can be brought under the status of the above principle, have been successfully devised and exercised.

In later years, the metaphor has been broadened to "*Measuring what cannot be touched but can be visually sensed*", meaning not necessarily the optical part of the spectrum, but including also other parts like x-rays, thermal, microwaves, infrared, ultraviolet, etc.; even ultrasounds.

To what extent emotional impact relies on visual perception, it is not a new issue. From the early days of rock painting to the contemporary media and communication industry, images are tools of a central role in triggering and driving emotions. The question, however, on the degree to which the images are actually visually perceived the intended way, has up to now, only been an issue of psychological approaches and behavioral science analysis, heavily based on subjective procedures and non-measurable indices.

Therefore, if Photogrammetry can contribute to the above research in a rational and scientific (meaning exact, measurable and repetitive) way, then we may be in the beginning of a new much broader metaphor "*Measuring the intangible*".

In line to the above thoughts, this presentation will try to build up a concept and the respective photogrammetric procedures for measuring the visual perception impact of a painting and describing it in concrete, exact and measurable terms. This should be perceived only as a demonstration example contributing to a much broader discussion.

3. THE CONCEPT

When the famous art historian Heinrich Wölfflin (Böckelman , 1938, Gantner, 1949, Schmitz, 1993) formulated his five antithetical principles, namely :

1. *linear vs. pictorial*
2. *surficial vs. anaglyph*
3. *closed vs. open form*
4. *multiplicity vs. unity*
5. *absolute vs. relative clarity*

for comparative examination of pieces of art and architecture, belonging to different art periods, he set the basis for a rational and objective assessment. However, probably he had never imagined that his principles could also be evaluated in a similarly rational manner, using objective procedures.

For the economy of this presentation we are going to limit our discussion on how Photogrammetry can be involved in applying these principles in comparative examination of two periods of art: Renaissance and Baroque. And we are going to demonstrate it by using examples of paintings. The following Table (Table 1) summarizes the differences in the two periods in terms of Wölfflin's principles, as well as the photogrammetric tools we intend to use.

4. THE TECHNICAL APPROACH

4.1 Clarity of borders

Proposition 1: *The clarity of the borders, silhouettes and figures can be measured by studying the edge-enhanced image. Besides the image can be segmented and the number, the size as well as the degree of definition of the obtained clusters can be used as indicators.*

For all the following examples, the Corel's Photo-Paint and Trace software has been used.

Table 1: Wölfflin's principles as applied to Renaissance & Baroque Art

Princ	**Renaissance** *(1420-1600)*	**Baroque** *(1600-1790)*	**Photogrammetric Tools**
1	• Tangible silhouettes • Undoubted outlines • Clear geometrical figures	• Complicated silhouettes • Disputable outlines • Sense of motion	• Study of contours • Image segmentation
2	• Planar surfaces • Perspective views	• Sense of depth • Intense contrast	• Depth studies
3	• Geometrical rigidity	• Non-geometric	• Study of contours
4	• Self-inclusive forms • Figure isolation • Diffused focus	• Absolute unity • Concentration of action • Focal points	• Study of contours • Image segmentation • Interest-points studies • Eye motion studies
5	• Extension to whole surface • Visible details • As much information as the viewer can handle	• No clarity • No borders • Less information	• Study of contours • Image segmentation • Histogram studies • Entropy studies

In Fig. 1, two typical images have been enhanced by high-pass filter and subsequently the edges have been traced, with the same parameters for both images. The clarity of figures is obvious in Renaissance as compared to Baroque. However this is not "measurable".

What is, though, "measurable" is the number of clusters when the images are segmented. In Fig. 2, the images have been segmented in a similar and consistent manner. The number of clusters, above the min. size, in Renaissance is bigger and also the clusters are smaller in size than those in Baroque.

A similar analysis to more paintings revealed the same trend as it is shown in Table 2 and Fig. 3.

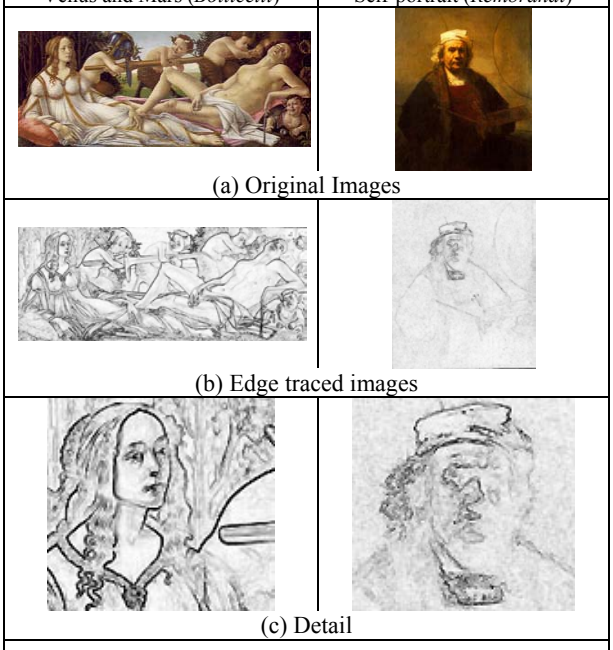


Figure 1: Edge traced images. In both images the same "sensitivity value" (Level=50 for Corel Paint) has been used.

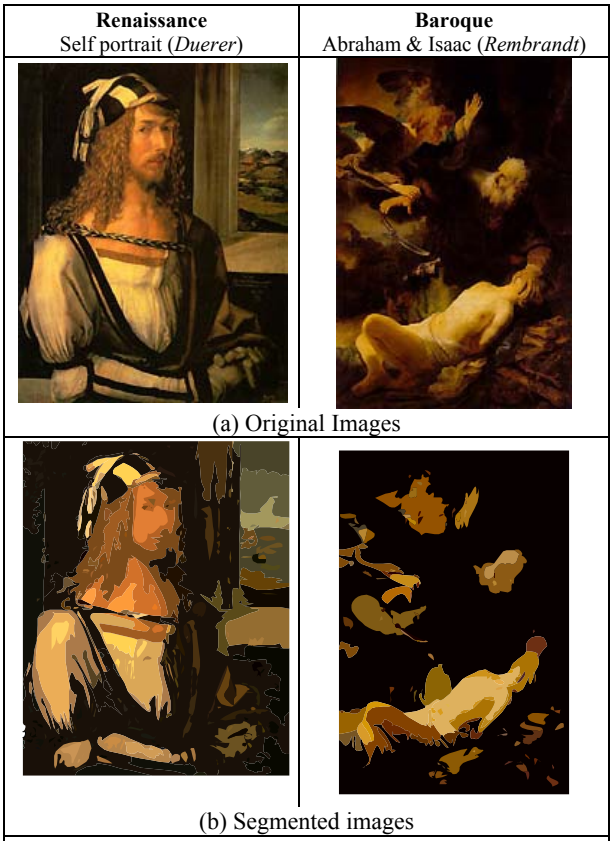


Figure 2: Segmented images. In both images the same "segmentation parameters" (Complexity=5, Node_reduction=50, Object_size=50, Smooth=Y for Corel Trace) have been used.

Table 2: Image segmentation in Renaissance & Baroque Art

Period	Artist	Title	Number of segments	Segments per pixel ($x10^{-4}$)
Renaiss.	Botticelli	The Birth of Venus	537	6.2
	Botticelli	Venus and Mars	267	6.0
	Giorgione	Adoration of the Magi	233	5.0
	Raphael	The school of Athens	507	11.6
	Michelan.	The Creation of Adam	432	6.8
Baroque	Rebrandt	Abraham and Isaac	128	1.6
	Rebrandt	The Holy Family	82	0.9
	Rebrandt	Bathsheba at Her Bath	121	1.4
	Rebrandt	Return of prodigal son	86	1.0
	Rebrandt	Self-portrait	173	2.1

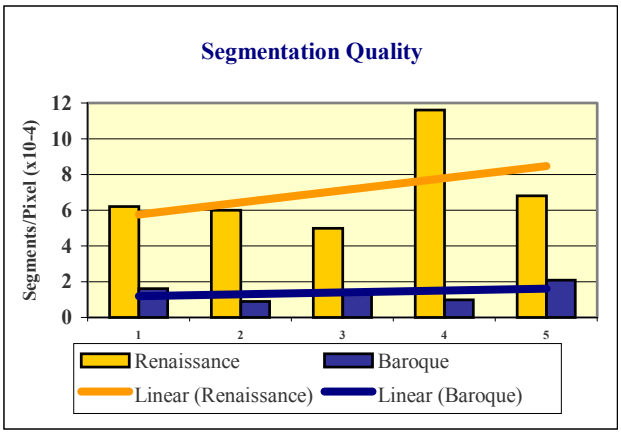


Figure 3: Number and size of clusters in Renaissance and Baroque art.

4.2 The perceived "depth"

Proposition 2: *The perceived "depth" can be measured by virtually embossing the image.*

Virtual emboss is achieved by creating shades (caused by a positioned light source) according to color differences. This means that the higher the contrast the wider the shades, corresponding thus to more intense anaglyph, in a similar way as it is perceived by the viewer.

Such a "virtual" depth corresponds to a perceived depth, and as expected in Renaissance is much lower than it is in Baroque. This is quite obvious in Fig. 4, where two typical examples have been processed with the same embossing parameters.

This, however, analysis is only of qualitative nature, and more quantitative indicators should be sought. An interesting approach is the one illustrated in Fig. 5.

Since there is no actual 3rd dimension, we need to emulate it. One way to do so is to virtually illuminate the image by two light sources from different directions and create two different embossed images. This introduces "virtual parallax" which in turn gives an impression of "depth".

In Fig. 5, artificial light sources have been placed alternatively at 0 ^{o} and 180^{o} azimuth creating thus "pseudo" relief. The resulted images have been subsequently colored in Red and Blue so as to give the anaglyph view.

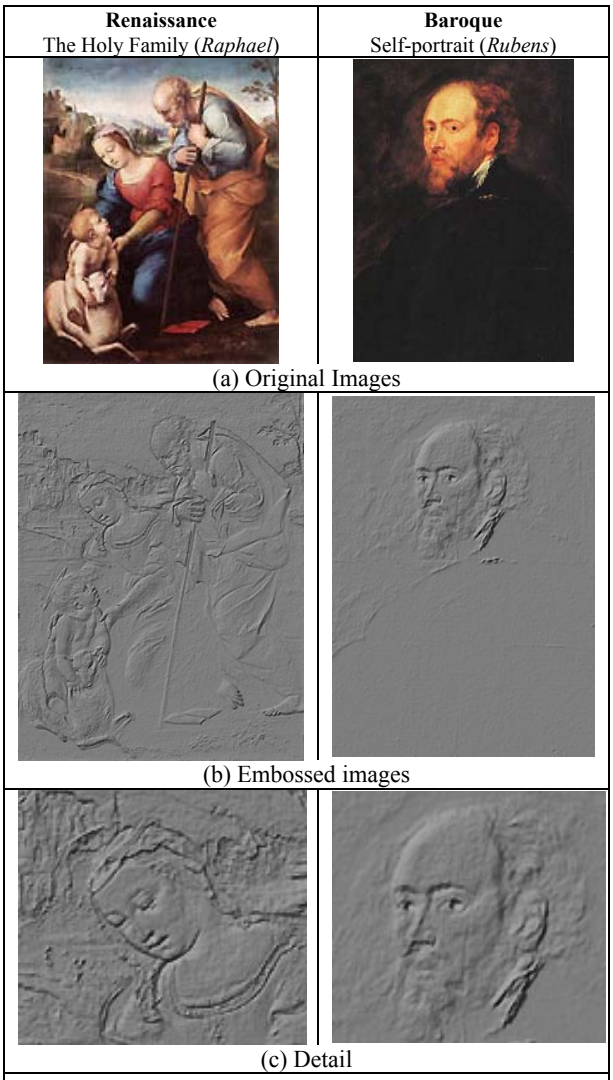


Figure 4: Embossed images. In both images the same "embossing parameters" (Detail=100, Depth=5, Smoothness=10, Light direction=315^{o} for Corel Paint) have been used.

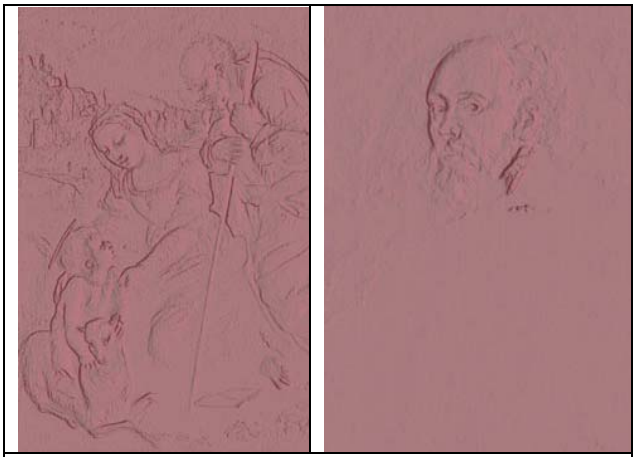


Figure 5: 3D "pseudo" anaglyph images.

It should be noted that now the two images can now be "matched" to each other and the differences in x-coordinates of conjugate points constitute the virtual "x-parallaxes". Average

and most notably variances can thus be calculated. As it is expected the average and standard deviations in x-parallax is respectively mean $_{Px}$ = 0.00 pixels and σ_{Px} = 2.21pixels for Fig. 5a (Renaissance) and mean $_{Px}$ = 0.83 pixels and σ_{Px} = 3.31pixels for Fig. 5b (Baroque).

4.3 Information content

Proposition 3: *The amount of information included in the image and possibly perceived by the viewer can be measured through the image Entropy.*

4.3.1 The notion of Entropy

Entropy, as a measure of information, was first proposed by Shannon (Shannon, 1948). Assume that a random variable x is distributed according to the probability function P(x). If an experiment is carried out, where the output is governed according to P(x), then the information generated by the experiment, for each output, is determined by the entropy of the distribution P(x).

The self-information of x is the quantity I(x), which is inverse to the probability P(x). If the event x always occurs then P(x) = 1, no (new) information can be transferred. The sum of the self-information of all elements is the average information per source output, which is described by the Entropy H (Bansal, 1999).

The value H increases with increase of the number of elements and thus more information is associated with the source. Besides, if the probability of each element is the same, the entropy is maximized and the source provides the greatest possible average information per element. Said another way, entropy measures the uncertainty in the output value generated by the experiment. Shannon derived a mathematical measure of the entropy, H(x), of a probability distribution to be:

$$H(x) = -\sum_{x} P(x)\ln(P(x)) \quad (1)$$

4.3.2 Entropy in the Image context

The quantification of the information contained in an image is quite similar to the information in communication theory as defined above. By modifying eq. (1) and setting the logarithmic base to 2 (this is appropriate in order to yield a unit of bits per pixel) we get (Leung et al, 2001) :

$$H = -\sum_{i=1}^{G} P(i)\ln_2(P(i)) \quad (2)$$

Where G is the number of grey/color levels of the image's histogram and P(i) the normalized frequency of occurrence of each level. The average information content is estimated in units of "bits/pixel".

For a Binary image (2 levels) the maximum entropy H_{max} (ie. when all levels have equal distribution, or in other words the same occurrence frequency) is 1 bits/pixel. This value implies that if any one pixel is extracted, only 1 bit of information can be interpreted (either "black" or "white"). And this holds true regardless of the size of the image. Similarly, the maximum entropy H_{max} for a variable with 256 possible realizations is 8 bits/pixel. That is each grey level has the same probability and we need $\ln_2(256) = 8$ bits to code it. Obviously, for an RGB color image H_{max} = 24 bits/pixel.

It is interesting to compare the entropy H of an image to the maximum entropy H_{max} for the greatest possible number of grey levels that can be contained in the image. This way we can define the relative entropy (Volden et. al., 1994) as :

$$H_{rel} = H / H_{max} \quad (3)$$

(The H_{rel} is commonly presented in percentage of H_{max})

Actually, the entropy of an image essentially measures two factors: the number of grey levels and the shape of their relative frequency distribution. The relative entropy H_{rel} is a measure of the contribution to the entropy due to the shape of the histogram, whereas the maximum entropy H_{max} measures the part due to its length.

4.3.3 Discussion

According to the above analysis, the Entropy H of 12 paintings belonging to Renaissance period and 9 belonging to Baroque has been computed, as shown in Table 3 and Fig. 5 .

Table 3: Information content in Renaissance & Baroque Art

Period	Artist	Title	Entropy *(bits/pixel)*				*Relative Entropy (%)*
			R	G	B	Total	
Renaissance	Botticelli	The Birth of Venus	7.6	7.5	6.4	**21.5**	89
	Botticelli	Venus and Mars	7.7	7.7	7.6	**23.1**	96
	Da Vinci	St. Jerome	7.3	6.8	5.5	**19.7**	82
	Da Vinci	The Annunciation	7.6	7.4	6.9	**21.9**	91
	Duerer	Self-portrait	7.5	7.2	6.2	**20.9**	87
	Giorgione	Adoration of Magi	7.5	6.8	5.6	**19.9**	83
	Giorgione	Venus asleep	7.8	7.8	7.4	**23.0**	96
	Raphael	The Holy Family	7.8	7.7	7.6	**23.1**	96
	Raphael	The school of Athens	7.6	7.3	7.6	**22.8**	95
	Veronese	The Marriage at Cana	7.4	7.3	6.2	**20.9**	87
	Michelan.	The Creation of Adam	7.3	7.5	7.3	**22.0**	92
	Michelan.	The Fall of Man	7.6	7.7	7.7	**22.9**	96
Baroque	Rubens	Battle of the Amazons	7.4	7.1	6.5	**21.0**	88
	Rubens	Self-Portrait	5.5	5.0	4.8	**15.4**	64
	Rebrandt	Abraham and Isaac	7.1	5.5	4.4	**17.0**	71
	Rebrandt	The Holy Family	7.0	6.4	5.4	**18.9**	79
	Rebrandt	Night-watch	7.3	6.5	5.0	**18.7**	78
	Rebrandt	Bathsheba at her Bath	7.2	7.0	5.4	**19.7**	82
	Rebrandt	Man with magn. glass	6.7	6.0	5.1	**17.7**	74
	Rebrandt	Return of prodigal son	7.6	7.6	5.5	**20.7**	86
	Rebrandt	Self-portrait	7.4	6.9	5.3	**19.6**	82

The immediate result is that Renaissance art conveys higher levels of information than Baroque; a result which was intuitively expected, but never actually quantified.

In Fig. 6, examining the total Entropy H, it is obvious that although the linear trends (Renaissance: y = 0.10x + 21.14 , Baroque: y = 0.20x + 17.73) in both cases are almost parallel (slopes 0.1 and 0.2 respectively) the y-offset of the Renaissance

line is almost 4 bits/pixel larger. This means that the amount of information in the analyzed paintings of the Renaissance period is (4/24=) 17% more than those of the Baroque period, and generally, with only few exceptions, require more than 20 bits/pixel (ie. at average 21 to be compared to the maximum of 24 bits/pixel) to be coded. To the contrary, Baroque art in general requires less than 20 bits/pixel (ie. at average 17.7 bits/pixel).

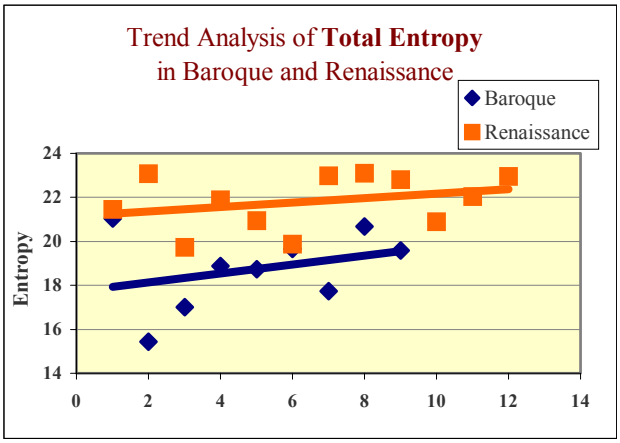


Figure 6: In Renaissance the total entropy is at average almost 20% larger than in Baroque

Fig. 7, is even more revealing. Analyzing the relative entropy H_{rel} and its linear trend (Renaissance y = 0.43x + 88.10, Baroque: y = 0.85x + 73.88) one can draw several conclusions: First, in Renaissance more than 88% of the available colors are used at average, whereas in Baroque only almost 74% at average. Second, the relative contribution of the colors is rather even in Renaissance, whereas in Baroque the contribution is quite uneven with the dark colors to be highly preferred. Fig. 7 shows two rather typical examples of such histograms, where one can notice that there is a considerable discrepancy between the average RGB value (47) and its median (34) showing an uneven distribution. This should be compared to the average (122) and median (123) values of the Renaissance example, which clearly shows an even distribution.

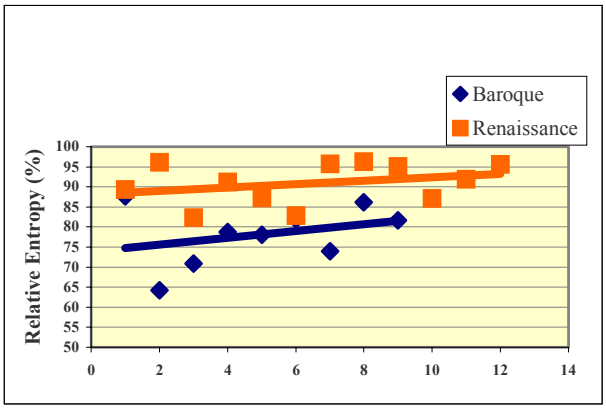


Figure 7: In Renaissance 88% of the available colors are used vs. 74% in Baroque

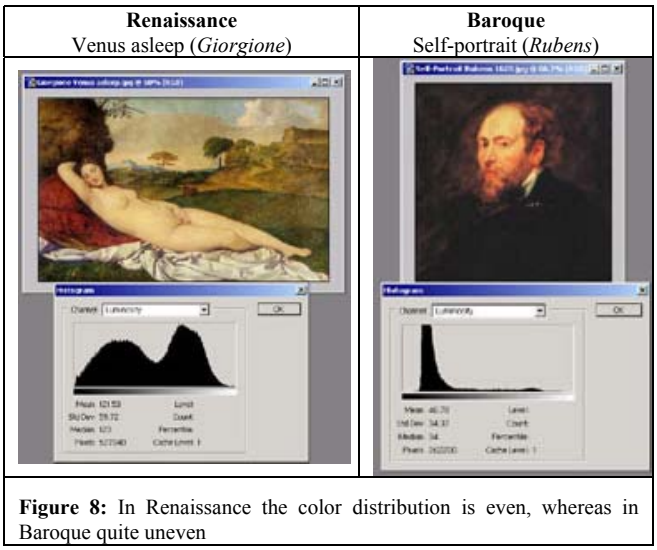


Figure 8: In Renaissance the color distribution is even, whereas in Baroque quite uneven

Similar conclusion one can reach by analyzing the relative contributions of the R, G and B channels to the total entropy (see Fig. 9). In Renaissance the relative contributions are pretty much balanced (average values R:7.6, G:7.4, B:6.8), whereas in Baroque (R:7.0, G 6.4, B:5.3).

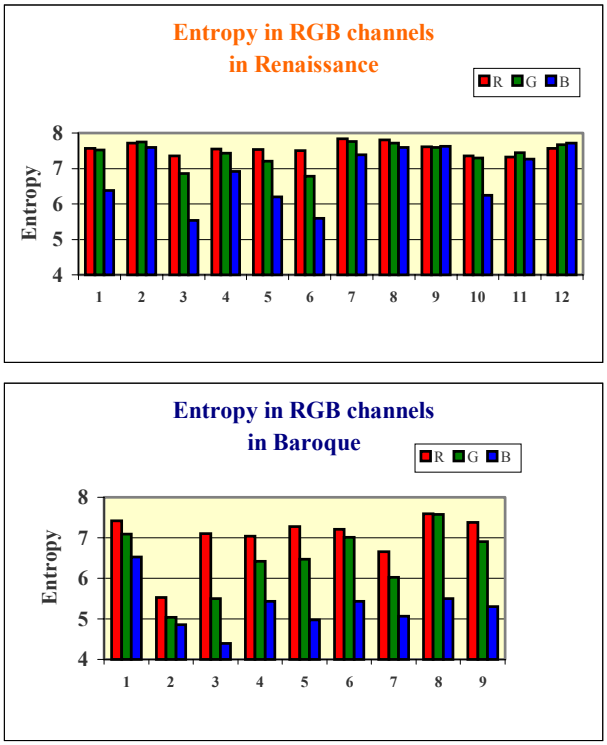


Figure 9: The relative contribution of RGB channels to the total image information is balanced in Renaissance but quite uneven in Baroque

4.4 Existence of "focal points"

Proposition 4: *The existence of specific "focal points" can be visualized and measured by either applying "interest point" operators, or record actual eye-gaze fixations on the picture.*

4.4.1 How computers "look" at pictures ?

Obviously, computer vision lacks any aesthetical appreciation to art. What can be "important" or "interesting" for a computer is only what can be sensed as "different". This notion, which is quite adequate for subsequent technical analysis, is the theoretical basis of "interest-point-operators". In the following we will use the Plessey operator, a typical point detector of this kind (Stylianidis, 2001).

The Plessey feature point detector (Harris and Stephens, 1988) makes use of an average squared gradient matrix M (eq 4-6) whose eigenvalues λ_1, λ_2 are used for point detection.

$$M = \begin{bmatrix} \left(\frac{\partial I}{\partial x}\right)^2 & \left(\frac{\partial I}{\partial x}\right)\left(\frac{\partial I}{\partial y}\right) \\ \left(\frac{\partial I}{\partial x}\right)\left(\frac{\partial I}{\partial y}\right) & \left(\frac{\partial I}{\partial y}\right)^2 \end{bmatrix} \quad (4)$$

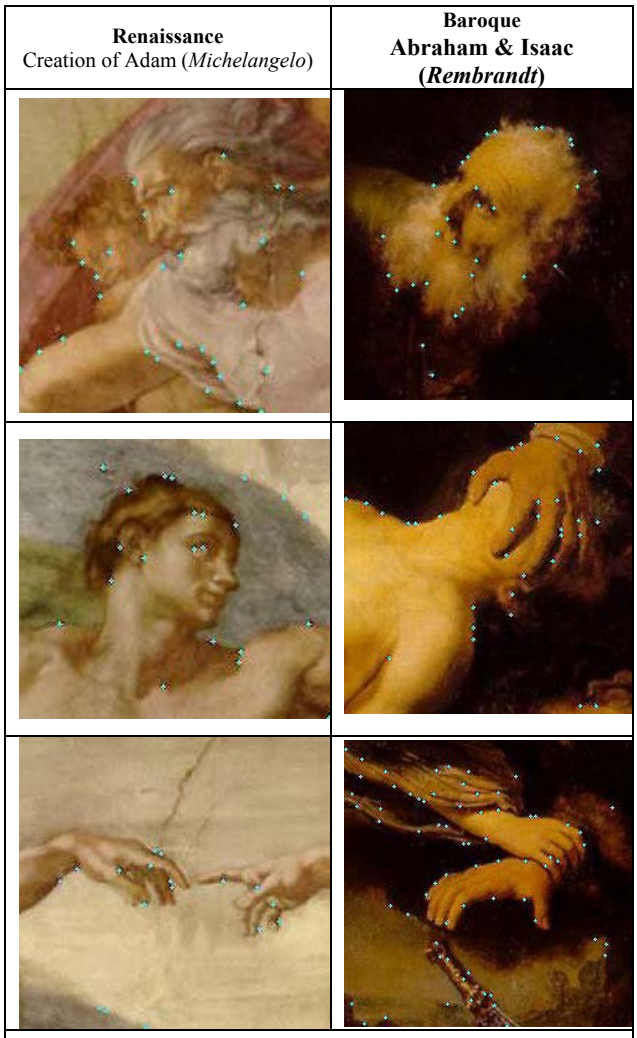


Figure 10: Interest points detected by the Plessey operator

In fact these points will be in regions where matrix M will be of full rank, and thus a function, which effectively measures the rank deficiency of M (response function) is used. This function R is defined as:

$$R = \det(M) - k\,(trace(M))^2 \quad (5)$$

with $\det(M) = \lambda_1\lambda_2$ and $trace(M) = \lambda_1 + \lambda_2$ (6)

and the regularization parameter k is of the order of 0.04 – 0.05 (Deriche, 1987, Schmid, 1996).

In Fig. 10, two art pieces have been processed by the Plessey point detector. The main conclusion is that in Renaissance the detected points are spread around the whole area, due to the facts that there are a lot of details and in many places there is detected changes in colors. In Baroque art, the points are detected along the major differences in colors (edges of sharp color gradients).

The remaining question, however, is whether humans perceive images much like the computer does. Do they spot the same "interest" points? In a similar manner? The only way to discover this is to actually record the way a human sees.

4.4.2 How people look at pictures ?

It is reasonable to assume that during looking, the direction of gaze in indicative of what the observer is interested in – although he might not always attend to what he is looking for.

And to quote Yarbus (Yarbus, 1967), "Eye movement reflect the human though processes; so the observer's thought may be followed to some extent from the records of eye movements. It is easy to determine from these records, which elements attract his eye (and, consequently, his thought), in what order, and how often".

We must confess that the aim is to understand the ways the eye moves and devise a way to record these movements and correlate them to perception and cognition. The relevance and importance of this should be left to behavioural scientists and art critics to judge. Here, we are only interested in the technical part of how this process can be rationalized.

4.4.2.1 The physiology of the eye

The eye-ball has a diameter of ca. 20 mm and is enclosed by three membranes: the *cornea & sclera*, the *choroid*, and the *retina*.

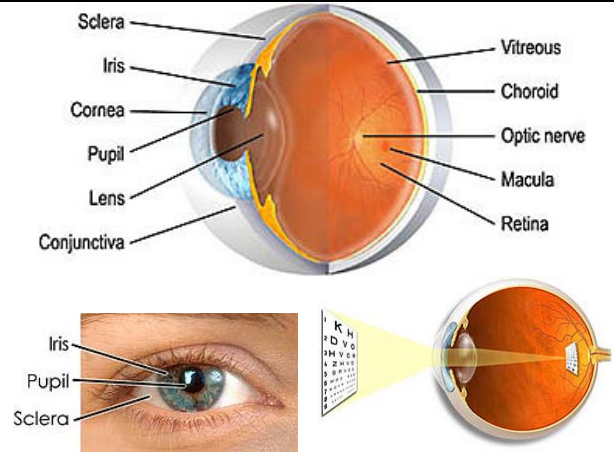


Figure 11: Human eye structure and the formation of the image

The eye lens is made up of fibrous cells and is flexible. It is controlled by three sets of muscles (for horizontal, vertical and rotational movement) and to focus on distant objects it gets flatter (and vice versa).

The image is formed at the back of the eye: the *retina*, which is not homogeneous (Fig. 11). It consists of a large outer ring of *rods* (ca. 130 million per eye), which provides peripheral vision (less acute), and a small central region: the *fovea* densely full of photoreceptor *cones* (ca. 7 million) for highly acute vision. Perception takes place by the relative excitation of light receptors. These receptors transform radiant energy into electrical impulses that are ultimately decoded by the brain.

Thus, detailed observation of a reasonable part of the surrounding world requires moving the eye (and head and body) to successively focus different parts of the reflected light rays on the fovea, thus foveating various regions of the observed scene.

The distance between the center of the lens and the retina (*focal length*) varies from 17 mm to 14 mm (refractive power of lens goes from minimum to maximum).

4.4.2.2 Eye movements and picture viewing

It is beyond the scope of this presentation the detailed presentation of years of research in the subject, and the interested reader is routed in the relevant literature for that (eg. Babcock, et al., 2002). We are only going to present, in brief, the main useful conclusions, which are productive to our aim.

Yarbus (Yarbus, 1967) showed that the perception of a complex scene involves a complicated pattern of *fixations* (when the eye pauses on a particular spatial location) and *saccades* (when the eye makes rapid angular rotations to foveate a new location).

Treated as "low level processes", eyes exhibit seven different types of movements (Glenstrup et al., 1995): Convergence, Rolling, Saccades, Pursuit motion, Nystagmus, Drift and microsaccades, and Physiological nystagmus. In "higher level" gaze patterns, there are three general types of looking (Kahneman, 1973): *Spontaneous* looking, *Task-relevant* looking, and *Orientation-of-thought* looking. From these, the spontaneous and the task-relevant looking are the types of looking we are mainly interested in.

Buswell (Buswell, 1935), experimented on the perceptual and cognitive significance of eye movements, in his pioneer work in behavioral science, where he examined more than 200 viewers. He concluded that observers exhibit two forms of eye movement behavior: (a) viewing sequences characterized by a general survey of the image, where a succession of brief pauses is distributed over the main features (b) observers make long fixations over smaller sub-regions. In general, people are inclined to make quick, global fixations early, transitioning to longer fixations (and smaller saccades) as viewing time increases.

Other useful conclusions are that observers often fixate on the same spatial locations in an image, but not necessarily in the same temporal order. Further, these consistencies revealed that people do not randomly explore pictures. Instead, the eye tends to focus on foreground elements like faces and people, rather than background elements, like clouds and foliage.

Brandt (Brandt, 1945) in his remarkable publication on a general analysis of eye movement patterns of people looking at advertisements, also concluded that, aside the individual differences, there exist similar patterns in viewing behavior. Later on, Yarbus (Yarbus, 1967) proved that eyes are directed to areas in an image that are useful or essential to perception. Other researchers also agree that viewers generally tend to fixate on semantic-rich and informative regions of a scene.

Generally, we can conclude, that the eyes are not attracted by the physical qualities of the items in the scene, but rather on how important the viewer would rate then to be. Thus, eg. when viewing faces, mostly the eyes, the lips and the nose will attract the eyes of the viewer.

Finally, it must be noted that the *scanpath* (the loci of the *fixations*) is determined by the composition of the scene (and, of course the individual viewer). This is hardly a surprise: for centuries artists have known that they could, to some extent, control the way the viewers would view their paintings, by exploiting composition.

Interestingly enough, studies on eye movements with emphasis in map reading and cognition has appeared for long in the cartographic society (see eg. Castner & Eastman, 1984, 1985), but, to writer's knowledge, never in photogrammetric literature.

4.4.2.3 Eye–gaze tracking

There are two major components to gaze direction: the orientation of the subject's head, and the orientation of the subject's eyes within their sockets. In this section we are dealing with the second one. In the next section an approach towards the solution of the first component is presented.

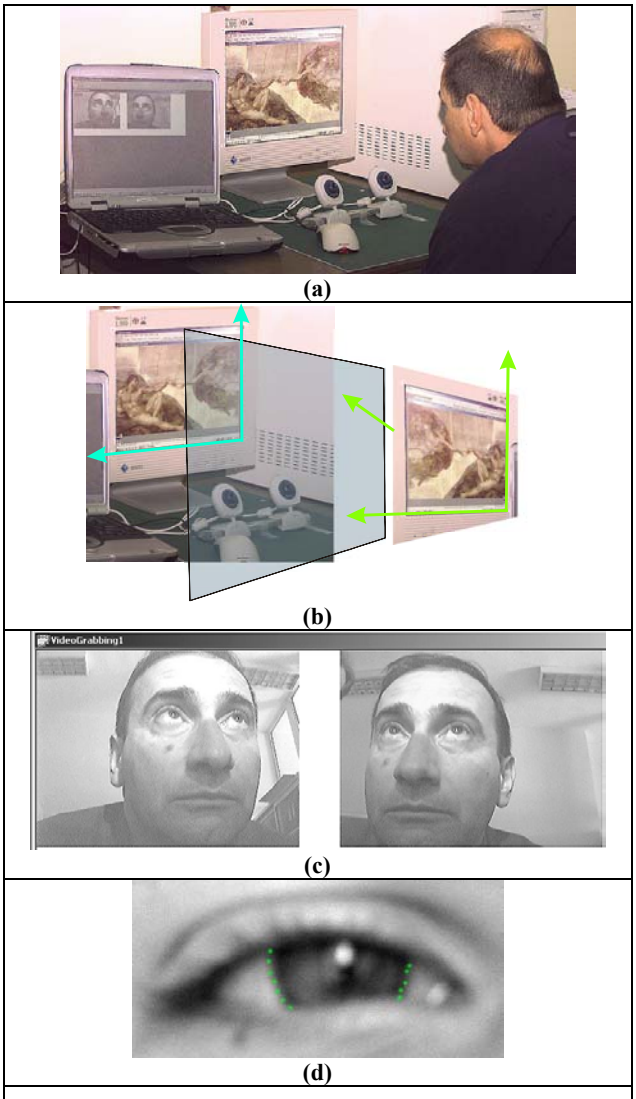


Figure 12: The proof-of-idea prototype system

Concerning the eye gaze determination, in the related vast literature, three classes of techniques can be recognized (eg. Young, et al, 1975; Glenstrup, et al., 1995): (a) techniques based on reflected light (ie. Limbus tracking, Pupil tracking, Corneal & Pupil reflection relationship, Artificial Neural Networks, Purkinjie image tracking), (b) techniques based on Electric Skin Potential, and (c) techniques based on Contact Lenses. From these, the techniques based on reflected light seem to be the most popular, and the interested reader will have no problem to find a long reference list (eg. Chenyz et al., 2000; Corno et al, 2002; Ebisawa, 1998; Kim, 1999; Kim et al, 1999; Ko et al, 1999, Liu, 1998; Rao et al, 1997; Shih et al, 2000; Smith et al, 2000; Talmi et al, 1999; Varchmin et al, 1998).

The prototype (Tsioukas, 2003), we are going to describe here, is based on mere photogrammetric techniques and possesses all the advantages of the techniques based on reflected light (portability, low-end equipment etc.), without the disadvantage of using infrared reflected light, which is usually interfered with other light sources and ambient light.

At this stage, the prototype system is rather primitive (Fig. 12a), since its aim is only to prove the feasibility of using photogrammetric techniques in eye tracking. The research is in progress, and this is only a preliminary report.

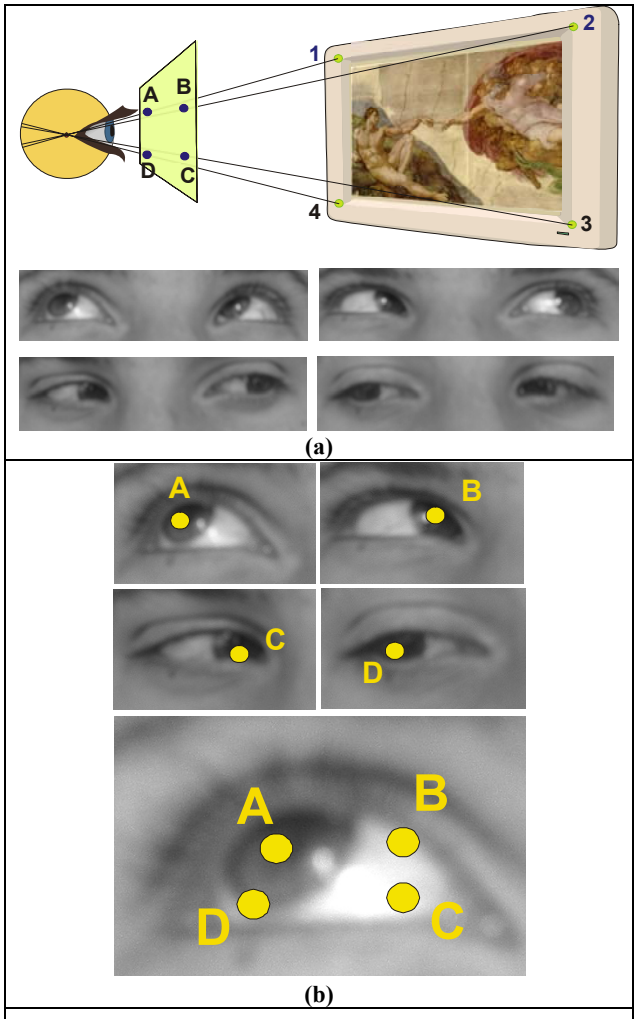


Figure 13: Determination of the gaze line

The technical approach is rather simple: Two low-end web cameras (c = 7mm, sensor: 3.2mm x 2.4mm) are set in front of a viewer and are controlled by a Notebook-PC. The viewer is shown a painting on a larger screen, and his fixations on the screen are recorded. Since we are only interested in the spontaneous and task-relevant looking patterns (not the saccades), currently the recording is done with a low frequency of 1Hz (ie. once a second).

The determination of the interior and exterior orientation of the two cameras is performed once, using portable calibration fields and e.o. devices. In our first attempts we have used a mirror in front of the screen (Fig. 12b), using thus a "virtual' calibration field of mirrored (fixed on the monitor) control points. Once the orientation of the cameras is known, the 3D position of the eyes' pupils in space can be computed by simple resection (Fig. 12c).

The next problem is the tracking of the pupils through the image sequence. This is done by processing the hough-transformed images and computing the "mostly voted" circle, while setting conditions for its center and its size (close to that in previous image in the sequence). Next, from the edge points, a best-fitting circle and its center is computed for better accuracy, although simple average of the "black" pixels of the iris may work as well (Fig. 12d).

The final two steps are (a) the determination of the *gaze line* (the line determined by the center of the eyeball and the center of the pupil) at each instant, and (b) the determination of the *Point of fixation* (the intersection of the gaze line of one eye with the surface of the object glanced).

From all the imaginable approaches, we chose to solve both the above problems in one step. The restrictions we impose are (a) that the head does not move during the monitoring, and (b) that the iris/pupil moves on a plane, not the actual sphere of the eyeball.

The steps we follow are described next:

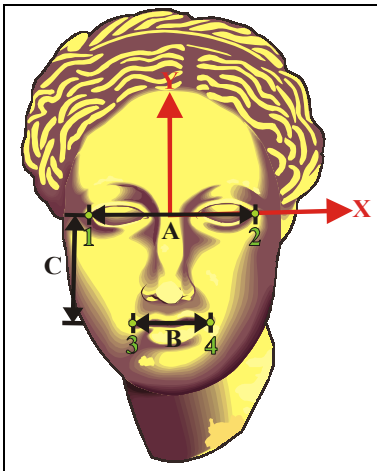
- The viewer is instructed, in the beginning, to focus on the four points (1, 2, 3, 4) marked on the monitor (Fig. 13a).
- During these fixations the 3D coordinates of the pupil is computed (A, B, C, D) and the system is thus "calibrated" to the user (Fig. 13b).
- The plane ABCD is rectified to plane 1234, once, and the rectification parameters are then used to transform the pupil coordinates to viewing screen coordinates.

4.4.2.4 Recovering the pose of the characters

When dealing with artist's desire to draw viewer's attention to specific part of the canvas, we should not underestimate another factor as well: "Humans have little difficulty sensing where another person is looking, often using this information to re-deploy their own visual attention. Even pre-Renaissance artists were aware of this, using the gaze of the characters within the painting to draw the viewer's eye to some significant part of the canvas." (Gee & Cipolla, 1994).

In the very interesting work of Gee & Cipolla, the problem of defining the face direction of the characters in a painting is tackled. Here, we are following a completely photogrammetric approach: (a) Four control points are defined as in Fig. 14a. The used anthropometrical dimensions are the same used by Gee & Cipolla and are in accordance to dimensions used by Michelangelo; (b) The "photo"-coordinates of these points are also measured; (c) the "virtual camera" (ie. the perspective used by the artist) rotations are computed by single-photo-resection.

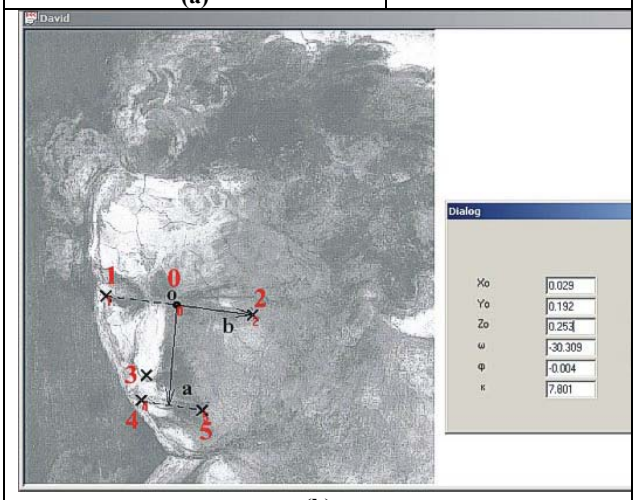
To compare our rather simple technique with the more complicated of Gee & Cipolla, we used their example of prophet Daniel of Capela Sistina (Fig. 14b). Our algorithm results to a horizontal tilt of 30° vs. 27° of theirs.



Pt	X	Y
1	-A/2	0
2	A/2	0
3	-(A-B)/2	-C
4	(A-B)/2	-C

Assumption :
A = 10 units
B = 4 units
C = 10 units

(a)



(b)

Figure 14: Anthropometrical dimensions and the example of Daniel

The procedure is further applied to different characters of Michelangelo (Fig. 15) and their face gaze direction has been estimated (Table 4), which is pretty much in accordance to our intuition.

Table 4: Face gaze direction of different characters

Character	ω (°)	φ (°)	κ (°)
Madonna	53.1	22.9	-33.7
Joseph	3.5	8.9	8.3
Angel	45.5	12.9	-13.1
God	35.7	0.5	22.9

5. CONCLUSIONS

The intention of this, rather unusual presentation is to propose a new point-of-view, from which innovative applications of Photogrammetry can be initiated.

Of course, the focal point here is applications in cultural heritage, but it is obvious, we hope, for the reader that the procedures suggested have much broader applicability than only in heritage assessment. For example, the estimation of the face gaze direction is necessary if one wishes to automate the eye-gazing estimation. The latter, on the other hand, can be applied in a myriad of cases.

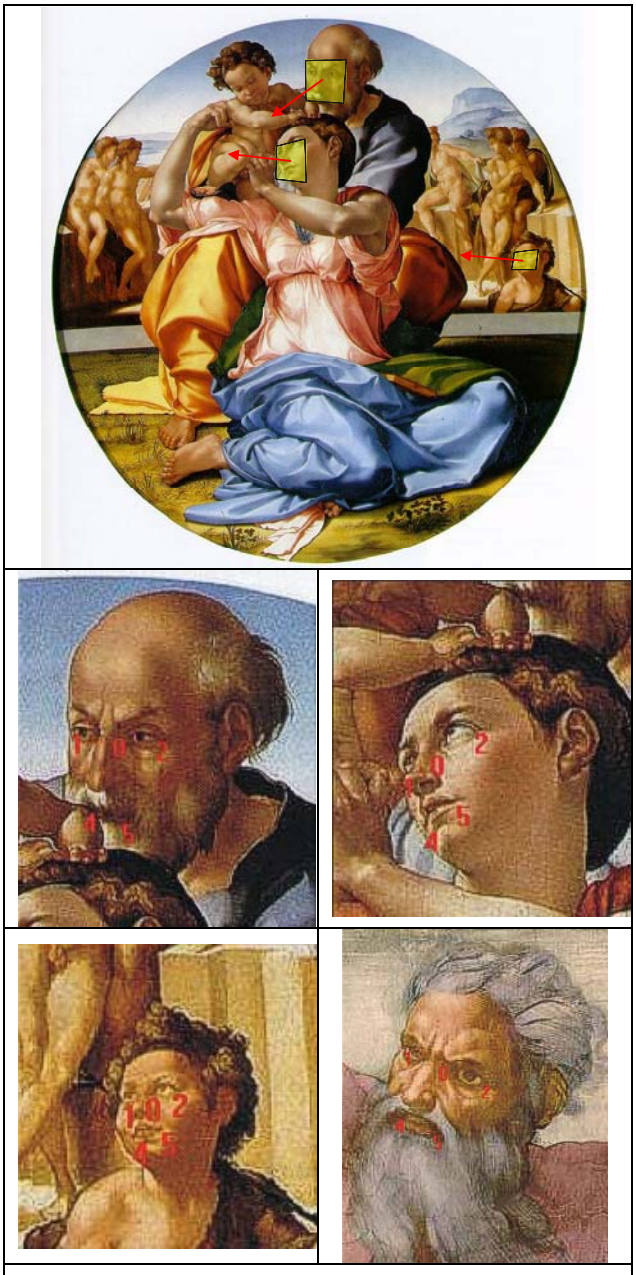


Figure 15: Face gaze direction in Michelangelo characters

The following list is non-exhaustive at all :

- Development of assistant devices for severely disabled people;
- Development of systems for monitoring drivers' alertness;
- Eye-controlled camera focusing devices;
- Development of human-computer eye-gaze control interfaces;
- Eye-gaze (Interest and Emotion Sensitive - IES) media (Recreative viewing, Commercials, Infromation browsing systems, video games);
- Development of interactive, eye-gaze based, multi-resolution displays.

References and Selected Bibliography

Babcock, J.S, M. Lipps, and J.B. Pelz, 2002. How people look at pictures before, during and after scene capture: Buswell revisited, *Human Vision and Electronic Imaging VII, Rogowitz & Pappas (Ed), Proceedings of SPIE,* Vol. 4662, pp. 34-47.

Bansal, R., 1999. Information Theoretic Integrated Segmentation and Registration of Dual 2D Portal Images and 3D CT Images, PhD Dissertation, Yale University.

Böckelman Walter, 1938. *Die grundbegriffe der Kunstbetrachtung bei Wölfflin und Dvorak*, Dresden.

Brandt, H.F., 1945. *The Psychology of seeing*, Philosophical Library, New York.

Buswell, G. T., 1935. *How People Look at Pictures: A Study of The Psychology of Perception in Art*, The University of Chicago Press, Chicago.

Castner, H. and R. Eastman, 1984. Eye-movement parameters and perceived map complexity - I, *The American Cartographer*, Vol. 11, No. 2, pp. 107-117.

Castner, H. and R. Eastman, 1985. Eye-movement parameters and perceived map complexity - II, *The American Cartographer*, Vol. 12, No. 1, pp. 29-40.

Chenyz Yong-Sheng, Chan-Hung Suyz, Jiun-Hung Chenz, Chu-Song Cheny, Yi-Ping Hungyz_, and Chiou-Shann Fuhz, 2000. Video-based Realtime Eye Tracking Technique for Autostereoscopic Displays, *Proc. of the 5th Conference on Artificial Intelligence and Applications*, Taipei, Taiwan, Nov. 2000, pp. 188–193.

Corno F., L. Farinetti and I. Signorile, 2002. A Cost-Effective Solution for Eye-Gaze Assistive Technology, *ICME2002: IEEE International Conference on Multimedia and Expo, Lausanne, Switzerland.*

Deriche, R., 1987. Optimal edge detection using recursive filtering. In: *Proc. 1st Int. Conf. on Computer Vision*, London, pp. 501-505.

Ebisawa, Y., 1998. Improved video-based eye-gaze detection method, *IEEE Trans. On Instrumentation and Measurement*, Vol. 47, No. 4, August, pp. 948-955.

Gantner Joseph, 1949. *Schönheit und Grenzen der klassischen Form: Burckhardt, Groce, Wölfflin. Drei Vorträge,* Wien.

Gee, A.H. and R. Cipolla, 1994. Determining the gaze of faces in images, Report No. CUED/F-INFENG/TR 174, University of Cambridge, Department of Engineering.

Glenstrup, A. and T. Engell-Nielsen, 1995. Eye Controlled Media: Present and Future State, Thesis, University of Copenhagen, Institute of Computer Science.

Harris, C. and M. Stephens, 1988. A combined corner and edge detector. In: *4th Alvey Vision Conference*, pp. 147-151.

Kahneman, D., 1973. *Attention and Effort*, Prentice-Hall Inc., Englewood Cliffs, New York.

Kim K.-N., 1999. Contributions to Vision-Based Eye-Gaze Tracking, MSc. Thesis, Kwangju Institute of Science and Technology.

Kim Kyung-Nam and R. S. Ramakrishna, 1999. Vision-Based Eye-Gaze Tracking for Human Computer Interface, *IEEE International Conf. on Systems, Man, and Cybernetics*, Tokyo, Japan, Oct.1999.

Ko Jong-Gook, K-N. Kim and R.S. Ramakrishna, 1999. Facial Feature Tracking for Eye-Head Controlled Human Computer Interface, *IEEE, TENCON'99*, Cheju, Korea, Sept. 1999.

Leung, L., B. King, V. Vohora, 2001. Comparison of image data fusion techniques using entropy and INI, *Proc. of 22nd Asian Conference on Remote Sensing*, Nov. 5-7, Singapore.

Liu, Jin, 1998. Determination of the Point of Fixation in a Head-Fixed Coordinate System, *Proc. of 14th International Conference on Pattern Recognition*, 16-20 August, 1998, Brisbane, Australia.

Rao, R., G. Zelinsky, M. Hayborne, D. Ballard, 1997, Eye Movements in Visual Cognition: A Computational Study, Technical Report 97.1, Department of Computer Science, University of Rochester

Schmid, C., 1996. Appariements d'images par invariants locaux de niveaux de gris (in french), *Thèse de Doctorat*, INP Grenoble.

Schmitz Norbert, 1993. *Kunst und Wissenschaft im Zeichen der Moderne: Exemplarische Studien zum Verhältnis von klassischer Avantgarde und zeitgenössischer Kunstgeschichte in Deutschland: Holzel, Wölfflin, Kandinsky, Dvorak,* Wuppertal.

Shannon, C.E., 1948. A mathematical Theory of Communication, *The Bell System Technical Journal*, Vol. 27, July, pp. 379-423, 623-656.

Shih, Sheng-Wen, and Jin Liu, 2000. A calibration-free gaze tracking technique, *IEEE Trans. on Circuits and Systems for Video Technology*, No. 4, Vol. 10, June 2000, pp.518-529.

Smith P., M. Shah, and N. da Vitoria Lobo, 2000. Monitoring Head/Eye Motion for Driver Alertness with One Camera. *Fifteenth IEEE International Conference on Pattern Recognition*, September 3-8, 2000.Barcelona, Spain.

Stylianidis, E., 2001. *Digital Photogrammetry and edge technologies in mapping and documentation of monuments and sites*, PhD Dissertation, AUT.

Talmi, K. and J. Liu, 1999. Eye gaze tracking for visually controlled interactive stereoscopic displays, *Signal Processing: Image Communication*, Vol. 14, pp. 799-810.

Tsioukas, V., 2003, *Personal Communication*.

Varchmin A.C., R. Rae and H. Ritter, 1998. Image based Recognition of Gaze Directions using Adaptive Methods, In: *Gesture and Sign Language in Human-Computer Interaction,* vol. LNAI 1371 of *Lecture Notes in Artificial Intelligence*, Ed. I. Wachsmuth et. al., Springer-Verlag, Berlin, Heidelberg, pages 245-257.

Volden, E., G. Giraudon, M. Berthod, 1994. Modeling Image Redundancy, INRIA Rapport de Recherche No 2440, December.

Yarbus, A.L., 1967. *Eye movements during perception of complex objects*, In: Eye movements and vision, Chap. VII, pp. 171-196, Plenum Press, New York.

Young, L. and D. Sheena, 1975. *Methods & Designs: Survey of eye movement recording methods*, In: Behaviour Research Methods & Instruments, Vol. 7 (5), pp. 397-429.