Disambiguating Complex Visual Information: Towards Communication of Personal Views of a Scene

Marc Pomplun[a], Helge Ritter[a], Boris Velichkovsky[a,b]

[a] Department of Neuroinformatics, University of Bielefeld, Germany
[b] Unit of Applied Cognitive Research, Dresden University of Technology, Germany
email: impomplu@techfak.uni-bielefeld.de

Abstract. Two experiments on perception and eye-movement scanning of a set of 6 overtly ambiguous pictures are reported. In the first experiment it was shown that specific perceptual interpretations of an ambiguous picture usually correlate with parameters of the gaze-position distributions. In the second experiment these distributions were used for an image-processing of initial pictures in such a way that in regions which attracted less fixations the brightness of all elements was lowered. The pre-processed pictures were then shown to a group of 150 naïve subjects for an identification. The results of this experiment demonstrated that in 4 out of 6 pictures it was possible to influence perception of other persons in the predicted way, i.e. to shift spontaneous reports of naïve subjects in the direction of interpretations that accompanied gaze-position data used for the pre-processing of initial pictures. Possible reasons for a failure of such a communication of personal views in two cases are also discussed.

1 Introduction

Pictures and scenes are notoriously ambiguous. Culture, experience, attention, functional state and dozens of other factors determine that two persons may have completely different subjective perception of one and the same physical situation. For any educated psychologist this is a well-established basic fact which certainly deserves investigation but cannot be changed. There is a long tradition of illustrating this multistable and idiosyncratic character of individual perceptive consciousness with the help of ambiguous figures both in history of art (Gombrich, 1969; Chapman, 1987) and in experimental psychology where perception of ambiguous pictures became the goal of countless studies (Vicholkovska, 1906; Boring, 1942; Velichkovsky, Luria & Zinchenko, 1973; Cooper, 1994; Rock, Hall & Davis, 1994). In other disciplines like e.g. computer science pictures and scenes are processed and transformed but usually from a physicalist point of view, although active vision approach and neural computation paradigm can be regarded as signs of change in the tradition (see Lee & Bajcsy, 1992; Ritter, Martinetz & Schulten, 1992). In this paper we are going to demonstrate that an interdisciplinary convergence of these two lines of research is possible and welcomed on practical reasons: The transformation of physical pictures from the perspective of their perception by an active observer can support an unambiguous communication of the subjective views to other persons.

In order to approach this problem experimentally we used the most reliable (albeit certainly not ideal, see Zinchenko & Vergiles, 1972) objective index of visual perceptual activity, namely the data about eye movements of an observer. Since Yarbus' (1967) and other earlier investigation it is generally accepted that gaze position data are a fairly sensitive index of individual preferences and task attitudes. Our interest in eye movements was based on pragmatic considerations and had no direct relation to hypotheses about their possible casual role in detection, recognition and identification of visual information (e.g. Noton & Stark, 1971). In recent years there were several studies demonstrating the importance of clustering gaze position data (Nodine, Kungel, Toto & Krupinsky, 1992; Pillalamari, Barnette & Birkmire, 1993) for explication of subject's knowledge and strategies. However, these studies did not change the situation of a passive registration of eye movements in principle. The gaze-contingent change of local characteristics of visual displays in dependence on parameters of eye movements remained one of the paradigms of investigation that was predominantly used in the field of reading research. (see e.g. Rayner, Well, Polatsek & Bertera, 1982).

Our intention was to make a further step: The use of the information about gaze position in order to process the picture/scene for reconstruction of its outlook as it could be available to the active observer who produced the eye-movements. In other words, we want to approach if not the famous question of a philosopher "What is it like to be a bat?" then at least "What is it like to be Mrs/Mr. Smith's (visual) perceptual homunculus?". The answer on this last question can be of practical importance as many forms of non-verbal expertise, for instance, in interpretation of medical images (Norman, Coblentz, Brooks & Babcook, 1992), are still hardly available to an objective analysis and public communication. Fluctuations in perception of classic ambiguous pictures seem to provide a suitable experimental model for the study, because variants of their perceptual interpretation are well known and, in addition, eye movements have been investigated in numerous previous studies. In particular, these studies have demonstrated that the eye movement parameters can be specific to the different subjective interpretations of the ambiguous figures (e.g. Ellis & Stark, 1978; Gale & Findlay, 1983).

Recently Garcia-Perez (1992) proposed that eye movements during perception of ambiguous figures, such as the Necker cube and the Boring figure, can lead to a kind of spatial frequency filtration favoring the interpretation which corresponds to the location of gaze in a corresponding "focal area" of the figure (for other similar proposals, see Kawabata & Mori, 1992; Tsal & Kolbet, 1985). The method developed for the present investigation can be helpful in empirically proving hypotheses of this kind. The study was based on the use of an advanced imaging eye-tracker as well as on our previous work on eye-movement mediated communication of attention in cooperative problem solving (Velichkovsky, 1995).

2 Experiment 1: Eye-movement characteristics in perception of ambiguous pictures

This experiment was performed for collection of eye-movement data and evaluation of their specificity to different perceptual interpretations of ambiguous pictures.

2.1 Method

2.1.1 Apparatus

The system used in our experiments (Stampe, 1993) is an example of non-invasive imaging eye-trackers. It is based on the use of ISCAN RK-416PC pupil-tracking boards and two video cameras as inputs of information about the position of the head within the environment and the position of the pupil within the head. Fast calibration that remains stable over the whole period of study and does not suffer from accidental blinks (which are detected and described as such), free head with permitted deviation from the straight-ahead position up to 15^o, and finally a practically unrestricted field of view (80^o in the horizontal dimension and 60^o in the vertical) as well as the possibility to run experiments under normal illumination conditions made this eye-tracker to a perfect device for basic and applied studies. The average absolute precision of the gaze-position measurement with the help of the eye-tracker lies within the range of $0.6 \ldots 0.8^o$. By using a new calibration interface based on parametrized artificial neural networks, we improved the precision of measurement by up to 0.4^o. This made it possible to recruit even subjects wearing spectacles (see Pomplun, Velichkovsky & Ritter, 1994).

2.1.2 Subjects

6 subjects naïve about the purpose of the study participated in Experiment 1. They were students and co-workers of the computer science department at University of Bielefeld. 4 of them had a normal and 2 a corrected to normal vision.

2.1.3 Material and procedure

As the stimuli for this experiment we chose 6 pictures. These pictures are shown in Fig. 1 left. Two of them – the Necker cube and the Boring figure – are classical examples of ambiguous figures with a long history of investigation (Boring, 1942; Garcia-Perez, 1992, among many others). Two others were fragments of "Earth" by Giuseppe Arcimboldo and Maurits Cornelis Escher's "Circlelimit IV". Though popular as examples of perceptual bi-stability, these pictures were not used earlier in connection with eye-movement studies, as far as we know. Another picture was a fragment of Albrecht Duerer's "View of the Val d'Arco". Strictly speaking, this picture can hardly be considered as ambiguous, because the alternatives interpretations are unequal: Almost all observers first see a landscape with the castle and discover only after long delay the possibility to see the rock on the left side as the profile of a human face. It was included nevertheless, because of the realism of this situation (which made this picture especially interesting for a transfer of the method to such domains, as medical imaging). The last picture was the product of our-own morphing of a woman's and a man's faces.

We also prepared two unambiguous versions of each picture. They are shown in Fig. 1 in the middle and in the right row.

All stimuli were presented on a high-resolution 17" colour monitor (ViewSonic 7) with a screen resolution of 640×480 pixels. The distance of observation was 60 cm. The pictures were about 480 pixels high and from 330 (Boring figure) to 620 (Duerer painting) pixels wide. At the beginning of every session two unambiguous and one ambiguous version of the Necker cube were shown. They were used to introduce the task. After the explana-

Figure 1: The original pictures (left column) and their unambiguous variants for interpretations A and B (middle and right column, respectively)

tion all pictures were shown in such a way that two unambiguous versions of a picture always preceded the corresponding ambiguous version. The presentation time of every unambiguous version was 20 sec; the time of presentation of an ambiguous picture was 60 sec. Intervals between the variants of the same picture were 10 sec, intervals between different classes of pictures - 60 sec. This time of 60 sec also included the time of re-calibrating of the eye-tracker which took less than 10 sec. The order of presentation of all pictures was counterbalanced across the subjects.

The subjects received the special task of manual reporting their perception while viewing the ambiguous version of each picture, and therefore had to put their hand on a two-button computer mouse. The task was to push the button on the left as soon as they saw a certain interpretation of the picture – interpretation A – and keep it pressed as long as this state of perception lasted. When seeing the interpretation B they had to push the button on the right. The experimenter told the subjects about the corresponding buttons for each interpretation shortly before the next ambiguous picture was shown.

After the experiment we divided the fixations which were recorded during the presentation of the ambiguous pictures according to the button response data in two sets, namely the fixations belonging to the interpretations A and B, respectively. With additional data obtained during the presentation of the unambiguous variants, there were four different fixation sets derived from every thematic class of pictures: The sets A and B from the two unambiguous variants and the sets A' and B' which were obtained after dividing the pool of fixation data from observation of the ambiguous picture.

2.2 Results

All subjects were able to differentiate the perceptual states of all the pictures without apparent difficulties. The transitions from one perceptual state to another as manifested in manual responses of the subjects were almost instantaneous, i.e. with temporal gaps or overlaps of less than 500 msec, in about 90% of cases. The intermediate state "not one/not other" extended over less than 5% of the observation time of ambiguous pictures. The duration of perceiving a constant interpretation was found between 3 and 13 seconds varying significantly between subjects, but not between pictures. In the following analysis we used as a reference point the moment of a button pushing signifying the transmission into the corresponding perceptual state.

In the few cases with extensive history of previous investigation with the help of eye-movement recording, our results partially replicated previous data (Ellis & Stark, 1978; Gale & Findlay, 1983). Thus, the perception of the Necker cube was mostly connected with the saccades along its main diagonal. The change of a perceptual state correlated with a shift of the fixations to another "core area" of the picture. However, we could not confirm the previous suggestion that these phenomenal changes coincide with longer, so-called "organizational fixations" or other parameters of individual fixations or saccades (cf. Ellis & Stark, 1978). Corresponding data for fixation length as well as average size of pupil are shown in Fig. 2 and Fig. 3, respectively. The same lack of correspondence between parameters of separate fixations and the instants of phenomenal changes was typical also in the case of all other pictures. In order to evaluate stability and possible specificity of eye-movement patterns to perceptual interpretation of pictures in a more objective way a measure of similarity s between two fixation sets was used, which yields similarity values in the interval $[0, 1]$. This is described in detail in appendix A.

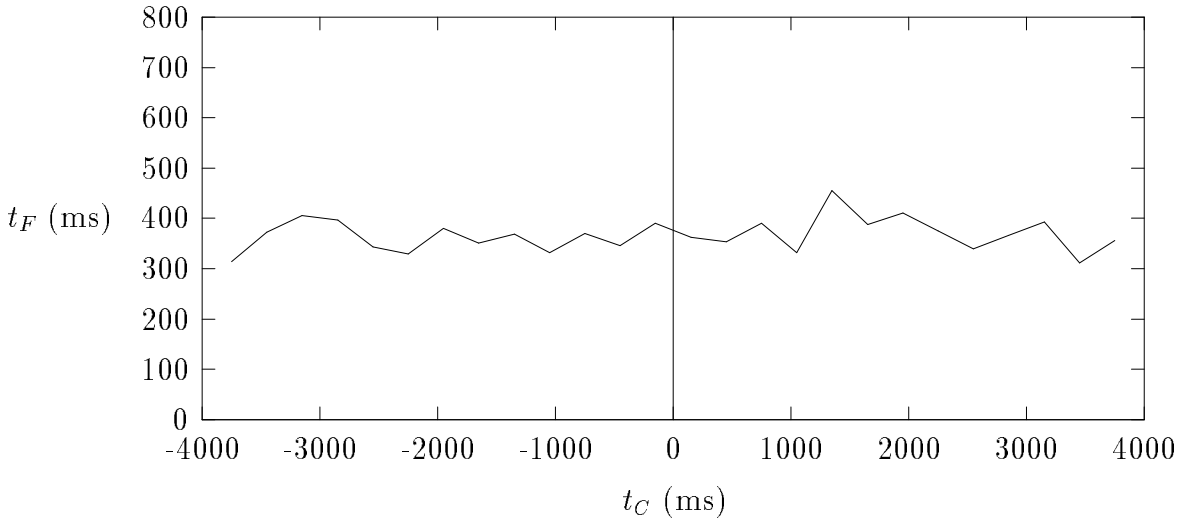


Figure 2: Average duration t_F of fixations as a function of the time t_C relative to changes of interpretations

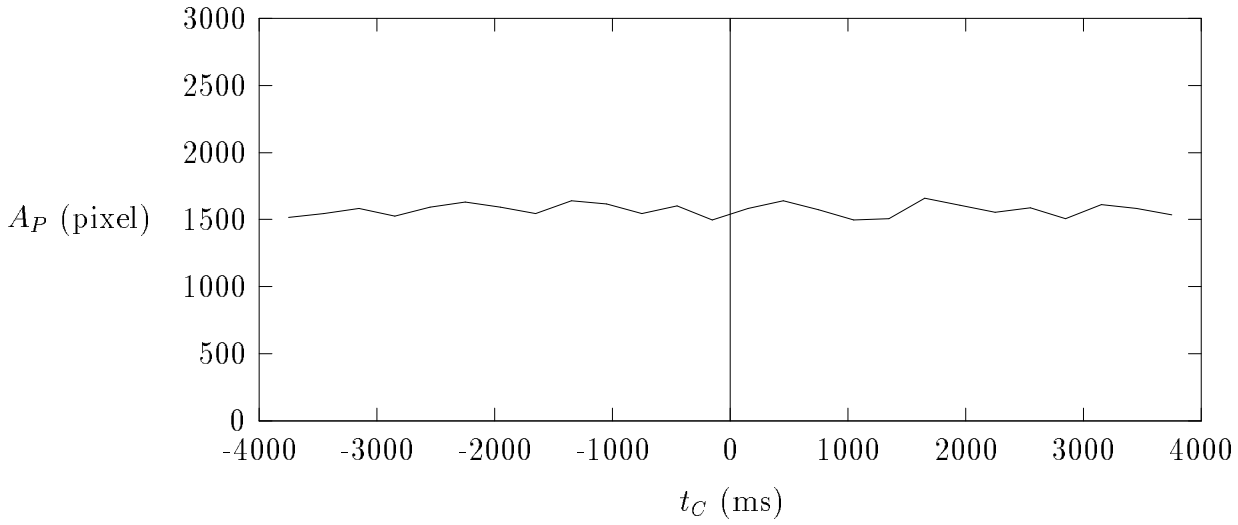


Figure 3: Average pupil size A_P as a function of t_C (measured as the number of pixels in the digital picture of the eye camera)

Comparison	Necker Cube	Duerer	Boring	Escher	Faces	Arcimboldo
A vs. B	64.3	45.4	76.1	15.2	72.2	40.5
A' vs. B'	77.5	27.7	48.5	28.6	34.9	67.5
A vs. A'	61.4	78.6	87.7	72.0	78.2	66.1
B vs. B'	37.5	93.5	88.1	88.2	64.8	83.0
A vs. B'	59.7	39.8	51.4	24.3	48.2	29.9
B vs. A'	48.8	35.6	71.3	15.8	50.3	62.9

Table 1: Similarities of fixation sets in %

The degree of similarity of fixation data for both variants of perception of the same picture was computed and compared between themselves and with the corresponding parameters for every unambiguous variant of the pictures. The computation was performed for individual and group data. The similarity of individual data for identical pictures was in the interval between 76% and 95%. The results of the comparison of group data (i.e. cumulated fixation data across subjects) are presented in Table 1.

For almost all pictures the computed similarity coefficients demonstrate systematic changes which become more prominent when visualized as gray values in matrices shown in Fig. 4. Each 4×4 matrix shows the subset of similarity coefficients pertaining to the 4×4 pairings of the four variants A, B, A', B' of each picture (these matrices are symmetric since the order in a pair is irrelevant, and the main diagonal represents the pairings of each pattern with itself, which is not relevant for our discussion). The brightness of each matrix element increases with the similarity coefficient of the corresponding comparison.

If there was no significant difference between the "statically" (A, B) and "dynamically" (A', B') derived fixation patterns, but between fixation patterns for different interpretations, the comparisons A vs. A' and B vs. B' should demonstrate higher similarity than all others. Obviously, in this case the matrices would present *checker-board patterns*. And in fact, the checker-board patterns can be easily seen in every box, with the sole exception of the box with data of the Necker cube. They are exactly those which can be expected on the basis of hypothesis about specificity of fixation distributions to the type of phenomenal interpretation of a picture.

The effect of higher similarities of A vs. A' and B vs. B' on the whole data set can be visualized also by a cumulative plot (Fig. 5). Here, the similarity coefficients for each class of comparison are presented in ascending order.

The computing of similarity coefficients finally allowed to approach the classic problem about objective indices of the phenomenal changes and the temporal relationships between a change of phenomenal state and the moment of manual report. For all the subjects and all the pictures, the minimal values of the similarity coefficient for fixation patterns A' and B' during a specific perceptual interpretation of an ambiguous picture were achieved if one takes into account a certain "response time" of approximately 900–1000 ms. In order to investigate the subject's response time, we changed the way of deriving the fixation

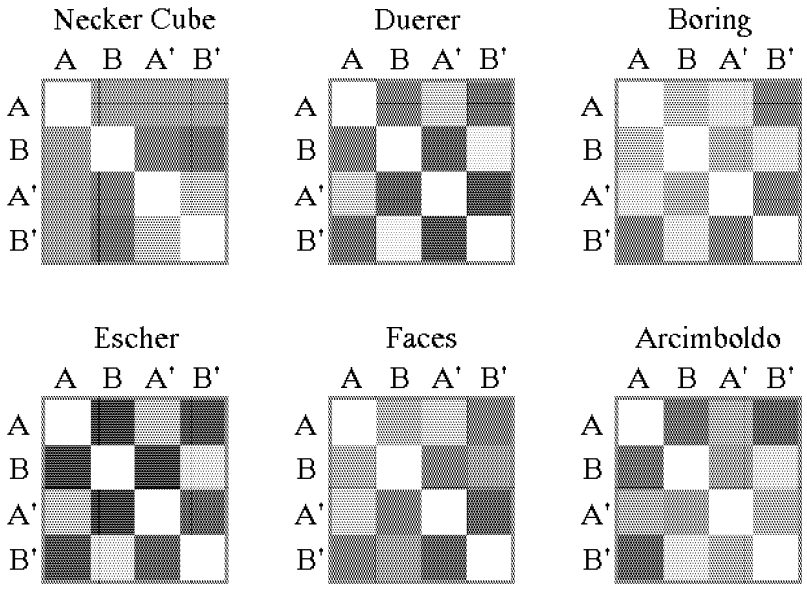


Figure 4: The similarity matrices of the six pictures

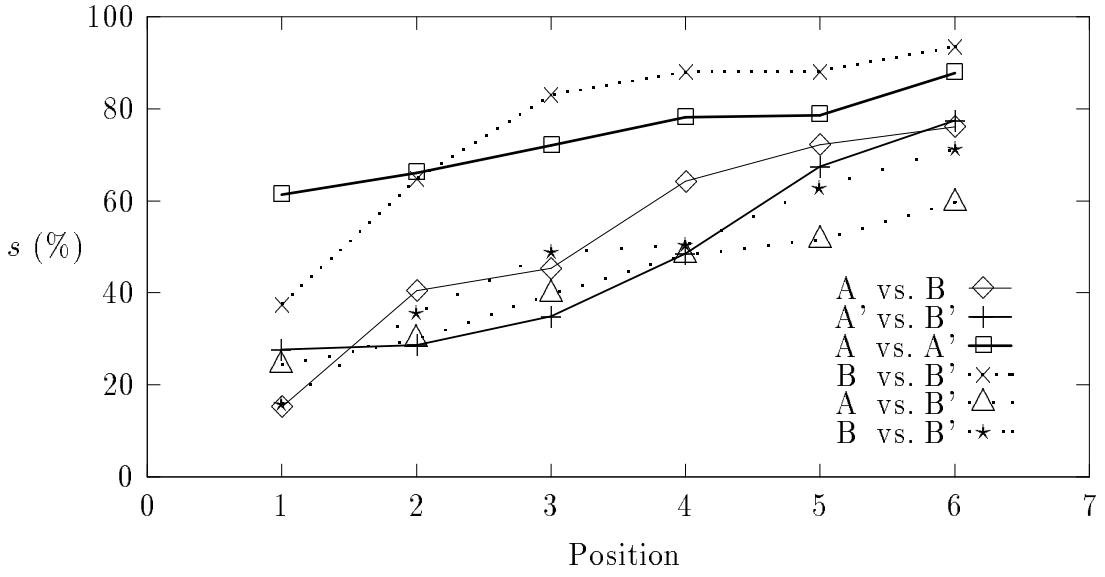


Figure 5: Cumulative plot of similarity values s for the six classes of comparisons

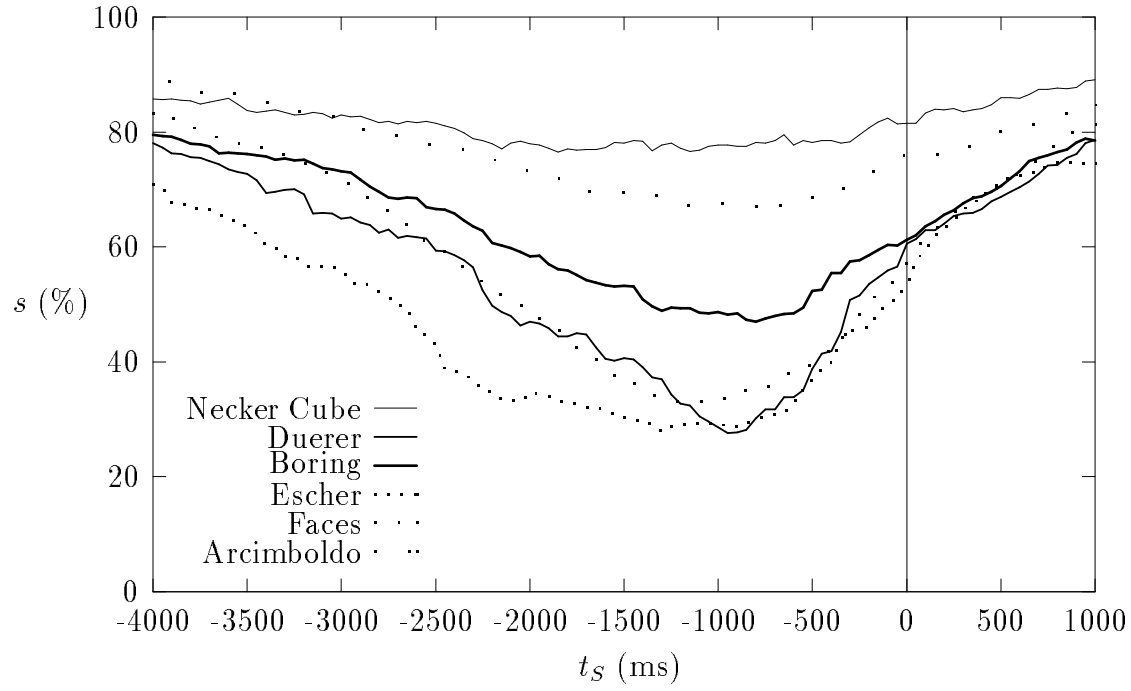


Figure 6: Similarity s of A' and B' for individual pictures and different "time shifts" t_S

sets A' and B' from showing the ambiguous picture. We added a constant "time shift" to every manual report of all subjects, pretending the reports happenend earlier (negative time shift) or later (positive time shift) than registrated. Now the similarity of the fixation sets A' and B' for individual pictures was computed using different time shifts.

Fig. 6 shows the average similarity function of the fixation patterns A' and B' as a function of the time shift used for the separation of the fixations. The Necker cube values demonstrate no significant dependence from the underlying time shift, but the other pictures indicate a more or less distinct "U"-shape. If one takes as a temporal reference point the moment located about 900 ms before the subject's manual report, then the similarity functions reach their absolute minimum.

The empirical data on the fixation patterns A' and B' that corresponded to the different perceptual interpretations of the same pictures were further used in Experiment 2 of the study.

3 Experiment 2: Visualization and transfer of subjective views of the pictures

The aim of this experiment was to attempt an objective reconstruction of different subjective views of the ambiguous pictures on the basis of the eye-movement data collected in Experiment 1.

3.1 Method

3.1.1 Subjects

150 subjects naïve about the purpose participated in this study All of them were students of natural sciences and mathematics at University of Bielefeld.

3.1.2 Material and procedure

In order to process pictures in a gaze-dependent way, one should decide what the form of the visibility function connected with such fixations is. Three lines of research can be of relevance for the answer of this question: "useful field of view" and "useful resolution" studies (Ball, Beard, Roenker, Miller & Griggs, 1988; Mackworth, 1976; Shioiri & Ikeda, 1989), investigation of asymmetry in dynamic distribution of attention in dependence on the direction of eye movements in reading (Rayner, Well, Polatsek & Bertera, 1982) and experiments with images stabilized on the retina that demonstrate a kind of dissociation between anatomical and "functional" fovea (Zinchenko & Vergiles, 1972). Unfortunately, it is impossible to use these data directly, because all of them were obtained under rather specific conditions. Therefore we assumed the relatively restricted and conservative hypothesis that the average efficient field of view coincides with the idealized anatomical fovea (Hood & Finkelstein, 1986). According to this working hypothesis the visibility function is a two-dimensional Gaussian distribution with the center at the registered fixation point and the standard deviation of one degree of visual angle.

Our further hypothesis was that the visibility functions of individual fixations can be collapsed without taking into account their temporal order. The aim of our image processing is to emphasize the regions of a picture which received the highest attention from the subject. There are many methods to achieve this, for example:

- lowering of brightness,
- enhancing of brightness,
- reduction of contrast,
- reduction of optical resolution in the "valleys" of attentional landscapes, i.e. outside of the highs of the gaze-position clusters.

More information about our image processing can be found in appendix B. In this experiment we only used the method "lowering of brightness". The processing was based on the corresponding fixation sets A' and B', respectively, derived from showing the ambiguous pictures in Experiment 1. The resulting pictures are presented in Fig. 7.

These 12 pictures were used together with 6 originals as material in Experiment 2. Subjects were individually presented with counterbalanced subsets of 6 pictures which included 2 originals and 4 processed pictures – representing all different thematic classes of pictures used in current study one and only one time. Subjects were asked to describe the content of the pictures. The descriptions were then subjected to a blind forced choice evaluation, so that a consistent "first sight interpretation" for every stimulus and every subject was agreed between three experts.

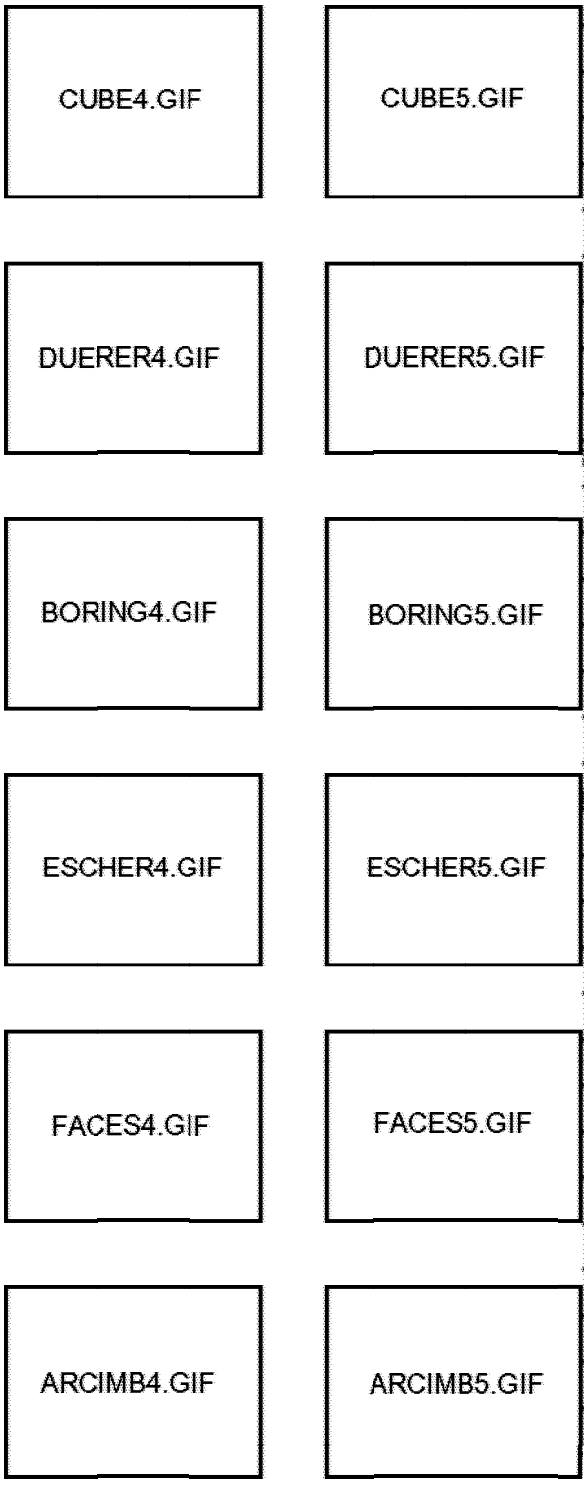


Figure 7: The "highlighted" pictures for each interpretation

	Necker Cube	Duerer	Boring	Escher	Faces	Arcimboldo
$I_A(O)$	39	49	23	41	11	38
$I_A(A)$	37	50	15	44	39	50
$I_A(B)$	38	23	7	11	4	8
$I_B(O)$	11	1	27	9	39	12
$I_B(A)$	13	0	35	6	11	0
$I_B(B)$	12	27	43	39	46	42

Table 2: Results of Experiment 2

3.2 Results

The results of this experiment are summarized in Table 2. The value of $I_A(O)$, e.g., tells us how many of the subjects came to interpretation A when the original picture was shown to them. Due to the fact that the decision always was either for A or B, the equation $I_A(x) + I_B(x) = 50$ is true for each of the 18 presented pictures. Statistical analysis of the data was performed with the help of a one-sided four fields χ^2 test (Lienert, 1973). The analysis demonstrates that the processing of initial pictures in terms of distribution of fixations had a significant and predicted influence on their further perception, although this influence was not documented in all cases. In particular, both variants of processing (i.e. in the direction of the interpretations A or B) had no influence on the perception of the Necker cube ($u_A = 0.58$, $p > 0.05$; $u_B = 0.23$, $p > 0.05$). In the case of the Boring figure the effect was only significant for the enhancing of the interpretation "old woman" ($u_B = 3.49$, $p < 0,001$). Paradoxically, the processing towards perception of the young woman rather diminished the frequency of this interpretation ($u_A = 1.65$). In all other cases, the effect of transfer of perceptual experience was fairly strong. When the base-line frequency was initially shifted towards one of the interpretations an appropriate processing either made the hidden version obvious or, at least in a tendency, additionally enhanced the dominating version of perception: Duerer's painting ($u_A = 1.43$, $0.05 < p < 0.10$; $u_B = 5.51$, $p < 0.001$), Escher ($u_A = 1.11$, $0.05 < p < 0.10$; $u_B = 5.80$, $p < 0.001$). For the remaining two pictures the results were even more homogeneous: Faces ($u_A = 5.61$, $p < 0.001$; $u_B = 1.96$, $p < 0.05$) and Arcimboldo ($u_A = 3.87$, $p < 0.001$; $u_B = 5.81$, $p < 0.001$).

4 Discussion

The present study brought about some old as well as some new results. In the line with earlier work we were able to testify in Experiment 1 that in the case of several pictures allowing more than one interpretation there are specific "focal areas" whose fixation correlates with specific perceptual interpretations (Gale & Findlay, 1983). In addition, we have demonstrated that a general change in distribution of gaze position patterns, as evalu-

ated with the help of a new wholistic measure of similarity, usually preceded the manual response about the change of phenomenal perception by a time of about 900 ms. For experiments on perceptual identification (see e.g. Posner, 1978) this is a reasonably long reaction time to suppose that the manual report indeed is a reaction on the phenomenal changes. The result in general corresponds to the introspective observation that phenomenal changes, while being expected, often slightly astonished observers. The phenomenal changes themselves, of course, can coincide, precede or perhaps follow the changes in eye movements. In contrast with one previous report, characteristics of separate saccades or fixations (as well as fluctuations of the pupil size) were insufficient for a differentiation of alternative perceptual interpretations of pictures from our set (cf. Ellis & Stark, 1978).

In Experiment 2 we attempted to use the data about eye fixation patterns of ambiguous pictures for the visualization of actual perception. This processing was done in such a way that the regions of the pictures which attracted the gaze fixations during specific interpretations were highlighted. In the present study this processing was based on the simplest assumption about the form and size of visibility function whereby we equated the parameters of the "functional fovea" with the idealized anatomical fovea, i.e. the Gaussian function with the standard deviation of one degree of visual angle. Despite this oversimplification the experiment was basically successful: In four out of six pictures we found a clear transfer effect of such a processing on the perceptual interpretation of naïve subjects. All the pictures that demonstrated this transfer were relatively complex, colorful stimuli with several levels of contrast.

From these data it seems to follow that both line drawings in our set – the Necker cube and the Boring figure – have a special status. Although exactly these figures were considered earlier from the perspective of their dependence on the eye movement based filtering of spatial frequencies (Garcia-Perez, 1992), the rather similar transformation used in the present study did not lead to the expected control of phenomenal experience. What are possible reasons for such a failure?

In the case of the Necker cube, for instance, there seems to be a built-in deceit: The very shift of the focus of attention to the "focal area" of an alternative interpretation creates a higher probability of reversal in the opposite direction. Indeed, in the middle of both "focal areas" one finds a vertex which has to be perceived as a component of the back(ground) plane of the corresponding 3D-interpretation. The fixation of the vertex can however provide it with a figure status and therefore provoke the reversal of the whole configuration. In the Boring figure there was an unexpected shift of initially more or less evenly distributed probabilities of both interpretations to one of them as a result of the image processing. The shading-out of the surrounding information obviously limits the possibility to see a young woman. This is the perceptual interpretation which is mostly conveyed by global information about the posture of the body as a whole and not so much by details like eye or mouth. An additional reason for the relative failure of our procedure in the case of black-and-white line-drawings may lay in the fact that the introduced modulation of brightness was too weak to be integrated into the main graphical elements of such pictures.

This study is only a first attempt of elicitation of perceptual experience on the basis of eye movement data. Several additional problems should be solved before the outlined approach could become a more reliable method. First of all, the shape of the visibility function has to be considered anew with the possibility that it can vary depending on objective and subjective factors. The second in the list is the problem of temporal char-

acteristics of processing – to what extent can the temporal order information be ignored in such studies (cf. Hacisalihzade, Stark & Allen, 1992) and what is the possible window size of accumulation of fixations to the "attention landscapes"? Another problem to be solved is an adjustment of our processing algorithms to the spatial frequency characteristics of pictures and to the corresponding perceptual attitudes of observers, e.g. as it would be necessary in the case of the Boring figure (for an investigation of related issues, see Caelli, 1988). A combination of our approach with methods of visual scene parsing and depth planes analysis from computer vision research (Ballard & Brown, 1982) could also be fruitful. Finally, one should of course be aware that not every fixation is "filled with attention", so states of "empty gaze" have to be differentiated. It seems that this problem could be solved on the basis of an analysis of eye movements themselves. However, the possible key to the solution may be situated in a slightly different domain, namely in the domain of micro eye movements (Gippenreiter & Romanov, 1972).

Non-verbal visual expertise plays an important role in everyday life, technology and medicine (see e.g. Norman, Coblentz, Brooks & Babcook, 1992; Velichkovsky, Pomplun & Rieser, 1995). The demonstrated fact that it is possible to convey to other persons a specific perceptual interpretation made by other people even in the case of relatively complex pictures gives grounds to believe in an applied significance of the gaze-dependent processing approach. Being well aware of shortcomings of the present study, we believe that future methods which like ours unite traditional perceptual research with contemporary image processing possibilities will support a more subject-oriented phase in the development of information and communication technologies. This will in turn open the way to communication of not only declarative knowledge but also practical expertise.

Acknowledgements. We wish to thank Larry Stark, Vladimir Zinchenko, and – last not least – Richard Gregory for discussing the results of this study and for encouraging us to present them to PERCEPTION. Two anonymous reviewers helped us in improving the final version of the text. Thomas Clermont and Peter Munsche participated in supervision and running the experiments. Our special thanks are due to Eyal Reingold and Dave Stampe for the development of the eye-tracker used in our experiments. This study was supported by a grant from the German Science Foundation (DFG SFB 360/B4).

Appendix A: A similarity measure for the comparison of distributions of fixations

In order to derive a similarity measure for two fixation patterns which depends on the holistic distribution of attention and not on separate eye movements, we first subdivided the monitor screen into $n_x \cdot n_y$ squares. Then the sum s_n of total duration of fixations $\in F_0$, which were located in the square n, had to be calculated for each square n, $n = 1 \ldots (n_x \cdot n_y)$. We obtained distribution vector $\vec{v}_0$ consisting of the values s_n and therefore having $n_x \cdot n_y$ dimensions. Our hope was that this vector contained sufficient information about how much attention or at least dwell time was spent in each of the squares.

To compute the similarity of two fixation sets F_1 and F_2, first the distribution vectors $\vec{v}_1$ and $\vec{v}_2$, respectively, had to be determined in the way described above. Then the cosine of the angle α between these two vectors was calculated according to the following simple equation:

$$\cos\alpha = \frac{\vec{v}_1 \cdot \vec{v}_2}{|\vec{v}_1| \cdot |\vec{v}_2|}$$

The value of $\cos\alpha$ was taken as the similarity measure. In fact, it has several important features. It yields similarity values in the interval $[0, 1]$, since both $\vec{v}_1$ and $\vec{v}_2$ have only nonnegative components. It does not take into account the *number* of fixations, but only their *distribution* over the screen. In addition, it can be easily weighted or corrected for duration of fixations.

Nevertheless, this measure still has an unpleasant property: Its values depend on the size and position of the squares on the picture. This dependency on position could be nearly removed by calculating similarity coefficients for different x- and y-offsets of the whole square grid and by taking the average similarity as the result. We used 10 equidistant x-offsets, which were chosen in order to allow a maximum global shift of the length of one square's side. These x-offsets were combined with 10 analogous y-offsets, so there have to be computed 100 "elementary" similarities on the whole to derive the position-invariant measure.

And how can we avoid the dependency on the square's *size*? Fig. 8 illustrates the functional relationship between square size (or "granularity") and calculated similarity coefficients for different fixation patterns. As an example the data for the Boring and the Arcimboldo picture are displayed. The Boring picture causes differences between the fixation sets A' and B' on a *small scale*, the Arcimboldo picture on a *large scale*. This fact will be discussed later in the text; at this point these sets should be considered as "technical" examples.

Obviously all similarity values generally increase with the underlying square's size. This fact can easily be explained by two extreme cases: If we used a square size of only *one* pixel, the similarity would be very low, because only very few fixations would be located in corresponding squares. On the other hand, if we used squares as large as the screen, the similarity value would always be 100%, because all fixations would lay in the same (single) square. Fig. 8 demonstrates another two important facts: First, the order of similarities remains invariant for the four fixation patterns with respect to granularity, at least for the investigated range from 5 to 300 pixels. This confirms the stability of our measure. Second, the maximum difference between similarities A vs. A' and A' vs. B' varies with the pictures. The Boring picture causes a maximum difference at a granularity of about 25 pixels (small scale), the Arcimboldo picture at about 60 pixels (large scale). To derive

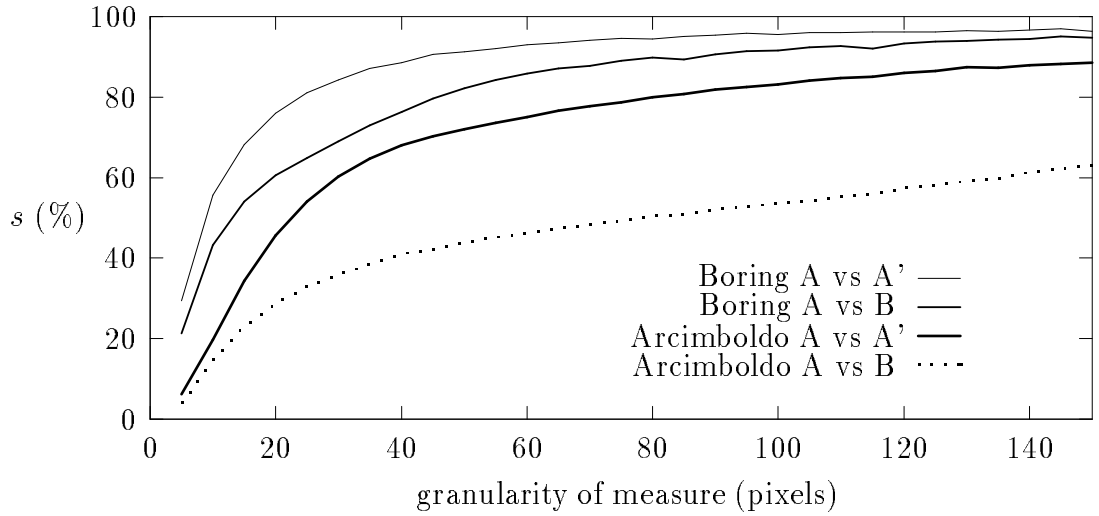


Figure 8: Similarity of specific fixation sets as a function of the granularity used for the similarity measure

a "fair" measure we decided to use the average similarity coefficient for comparison on 25, 40, and 64 pixel granularity, which is a geometric series in the relevant range. The use of smaller squares is not sensible, since the human foveal vision has a certain extent and, in addition, the eye-tracker accuracy itself is limited (see Pomplun, Velichkovsky & Ritter, 1994). Larger squares are not capable to improve the measure, because no further important information can be found on the scale of the presented pictures. Our definite similarity measure now uses 300 elementary distribution vector comparisons. It has all desired properties and its stability was proved in various tests.

Figure 9: The original picture

Appendix B: Methods of gaze-dependent image processing

The important precondition for the gaze-contingent image processing is a continuous "attention function" $a(x, y)$ which is defined all over the picture and is built on the basis of the recorded fixations. In order to find a suitable function we define a two-dimensional Gaussian distribution centered at the current fixation point, where the standard deviation is one degree of visual angle. Then we simply sum up these Gaussian distributions for all recorded fixations weighted for their durations to receive the desired function $a(x, y)$.

In order to illustrate the procedure, let us consider a test picture (Fig. 9). Its accumulative "attentional landscape" is shown in Fig. 10. This form of representation is derived from empirical fixations, and the peaks of this function corresponding to the eyes and the mouth in the woman's picture are significant.

The gaze-contingent processing can be realized in several different ways, depending on the chosen type of image processing function $f_P : A \times O \rightarrow P$ which combines the attentional landscape A and the original picture O to the resulting picture P. The effect of four different functions is illustrated in Fig. 11 to 14, where the gaze-contingent processing of a prototype picture was coupled with (a) lowering of brightness, (b) enhancing of brightness, (c) reduction of contrast, or (d) reduction of optical resolution in regions with lower values of attentional landscape. The last of these possibilities was already considered as a prospective method of disambiguation of ambiguous pictures, however, without considering eye movements (see Shiori & Ikeda, 1989).

Many different combinations of these procedures are easily realizable either beetween themselves or with different modes of processing. For the present study we used the

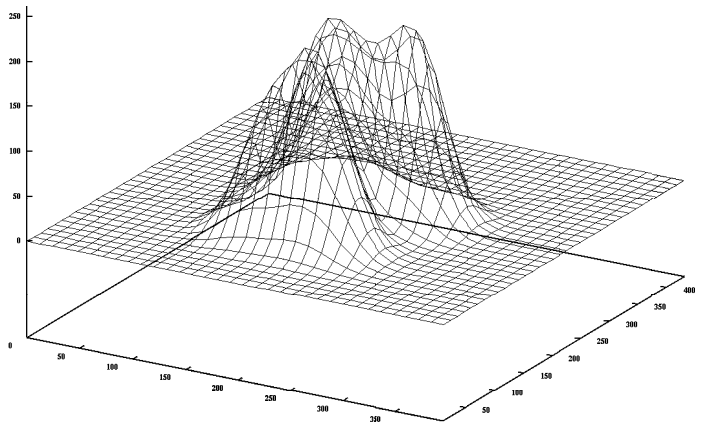


Figure 10: "Attentional landscape" distributed over the monitor screen as obtained from a subject watching the picture shown in Fig. 9

first of the described procedures: The brightness outside of attended regions was reduced according to the following transformational equation (1), which is applied on every pixel (x,y) of the picture:

$$\vec{p}_{xy} = \Theta_{xy} \cdot \vec{o}_{xy} \tag{1}$$

Here, $\vec{p}_{xy}$ and $\vec{o}_{xy}$ are the RGB-vectors (i.e. the *red, green*, and *blue* component of a colour) of pixel (x,y) in the processed and the original picture, respectively. The transformation factors Θ_{xy} can be calculated by equation (2):

$$\Theta_{xy} = m + (1 - m)\frac{a(x,y)}{a_{max}} \quad , \tag{2}$$

where $a(x,y)$ is the value of "attentional landscape" for pixel (x,y), a_{max} is the maximal value in the whole picture, and m is a constant which determines the minimum brightness remaining in the processed picture. If m, e.g., is set to 0.1, the regions of the picture with attention value 0 will keep 10% of their initial brightness, if $m = 1$ the picture will not change at all. In this experiment we always set $m = 0.1$.

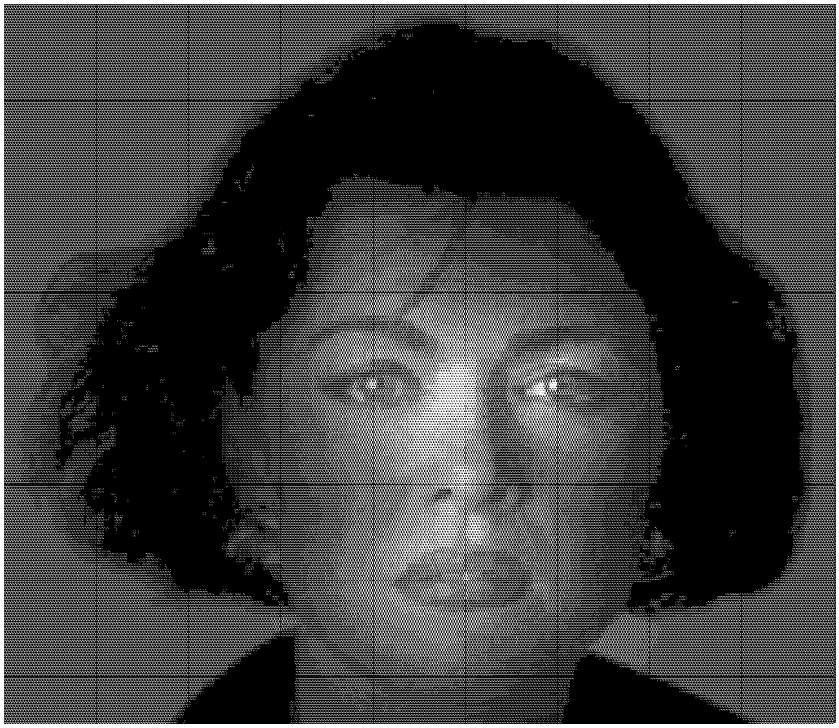


Figure 11: After a partial decrease of brightness one of the face regions seems to be "highlighted" (variant a).

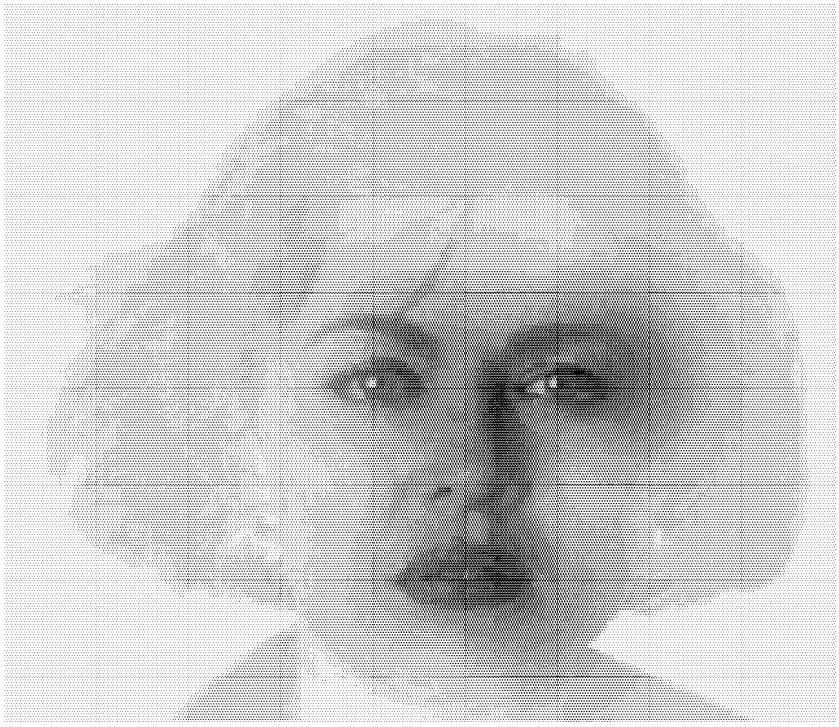


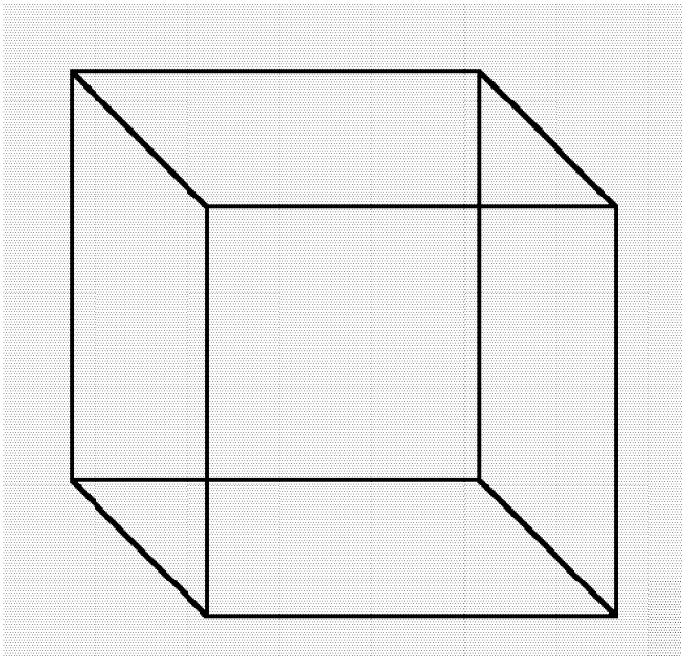
Figure 12: The less inspected areas seem to disappear behind a veil of mist after enhancing brightness (variant b).

Figure 13: The differences in colour decrease in the peripheral regions after reducing contrast (variant c).

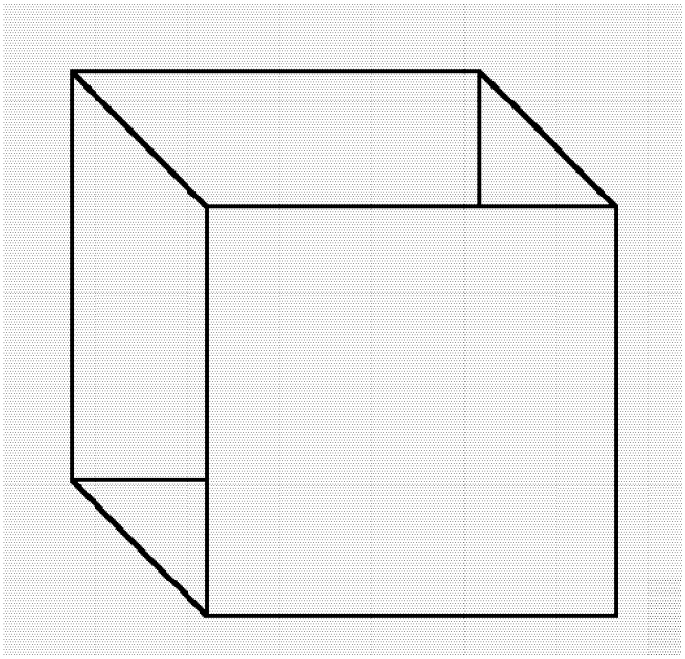
Figure 14: The areas of lower attention are blurred, so the Figure looks like a camera picture focussing the woman's face (variant d).

Appendix C: The picture series

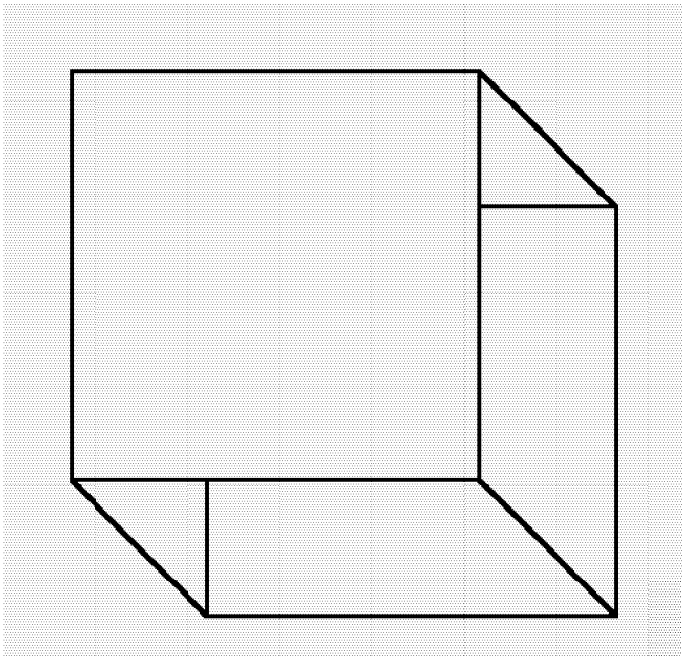
C.1 The Necker cube series



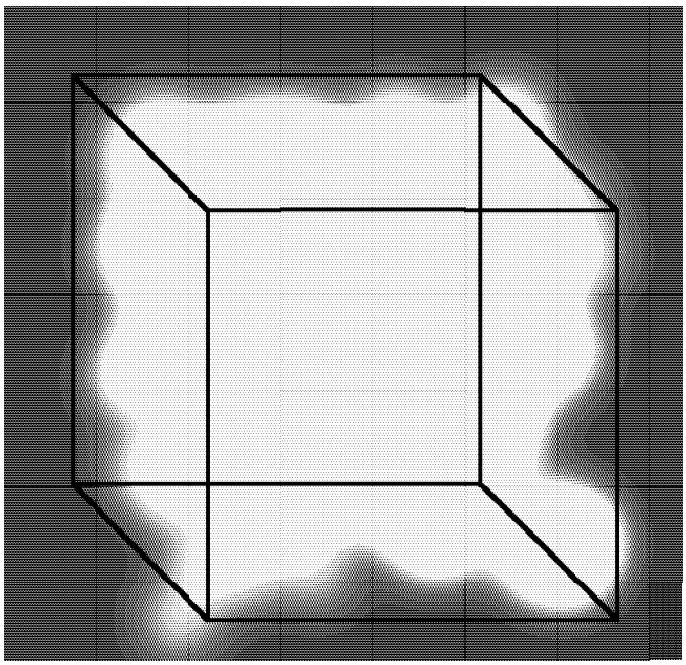
CUBE1.GIF



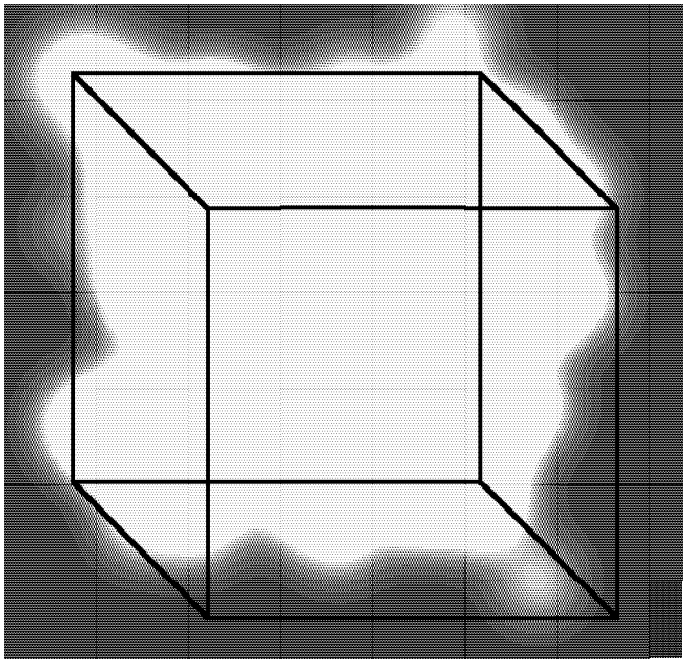
CUBE2.GIF



CUBE3.GIF

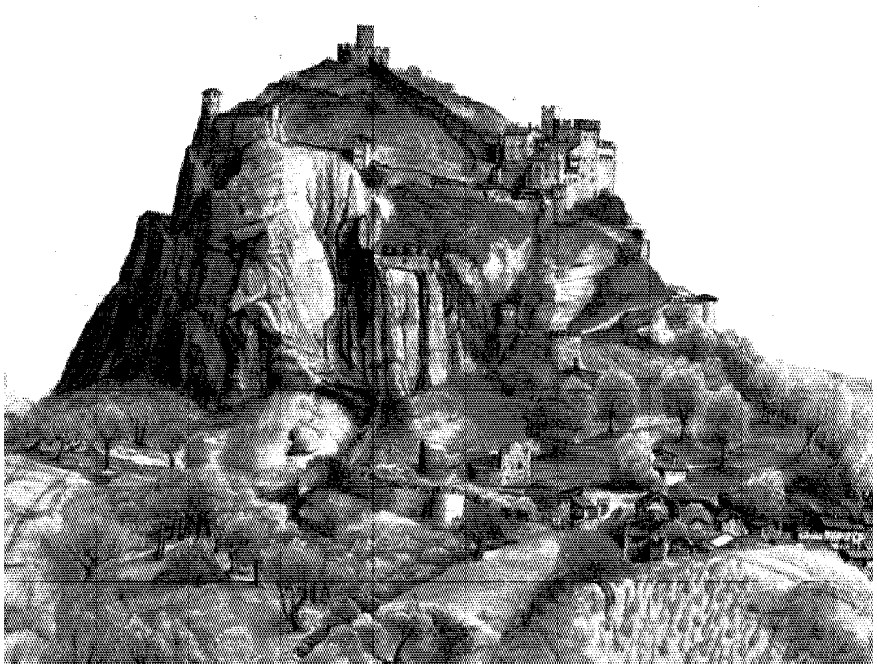


CUBE4.GIF

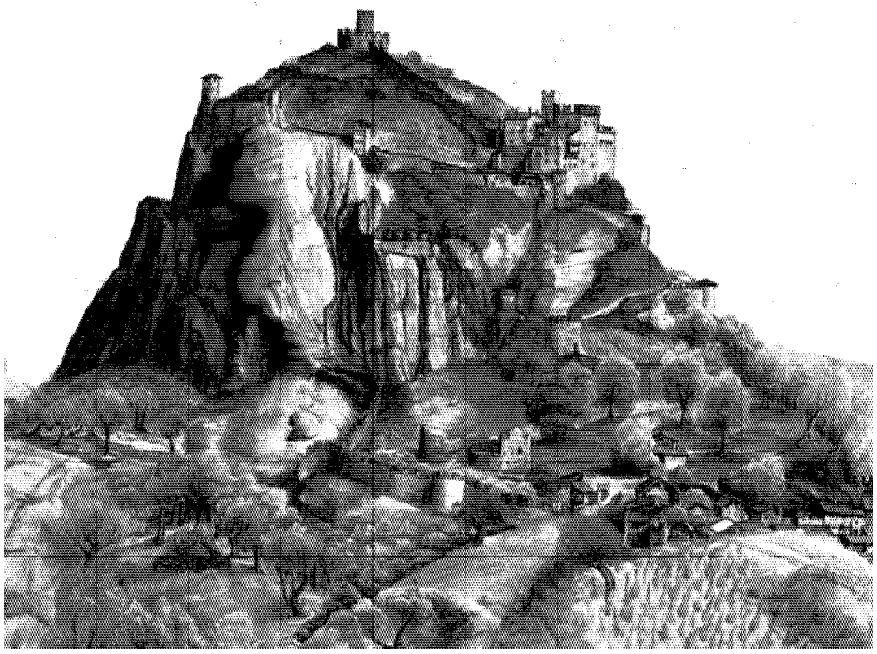


CUBE5.GIF

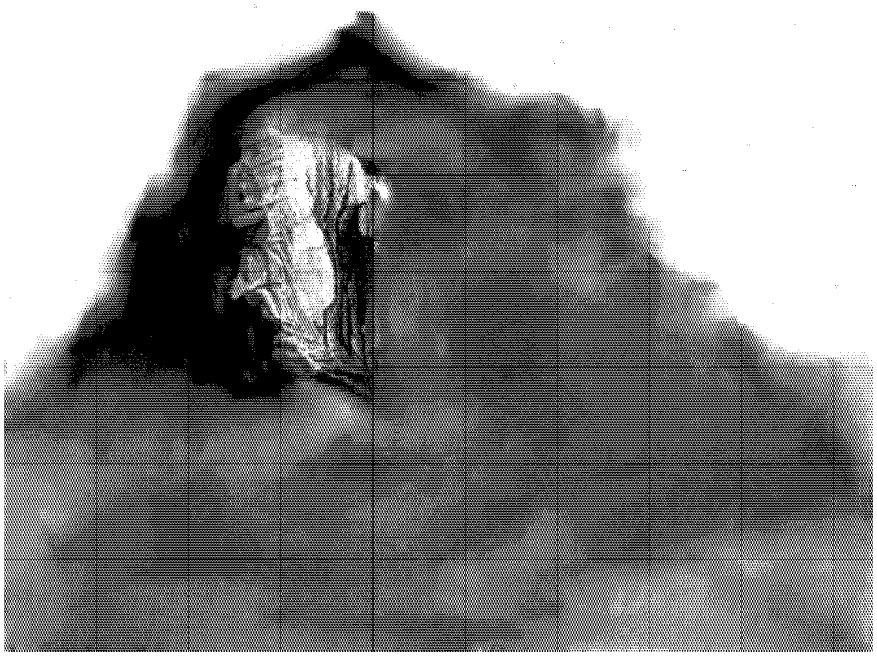
C.2 The Duerer series



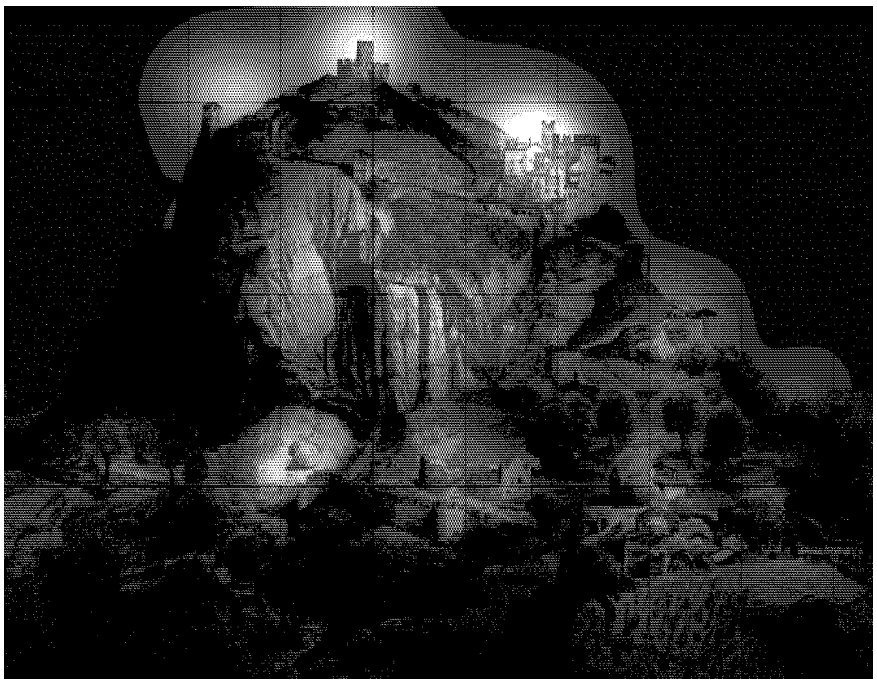
DUERER1.GIF



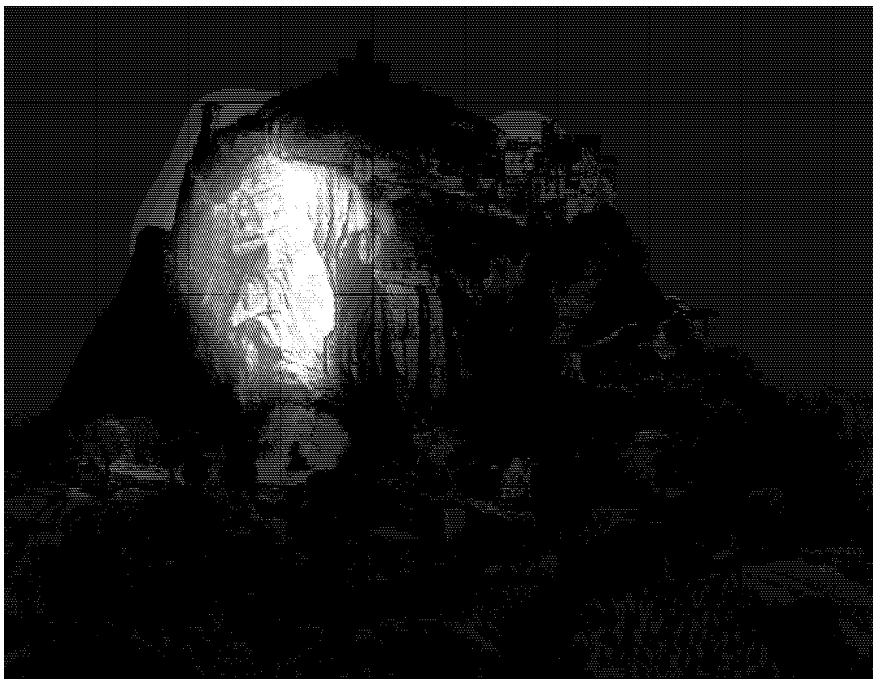
DUERER2.GIF



DUERER3.GIF



DUERER4.GIF



DUERER5.GIF

C.3 The Boring series

BORING1.GIF

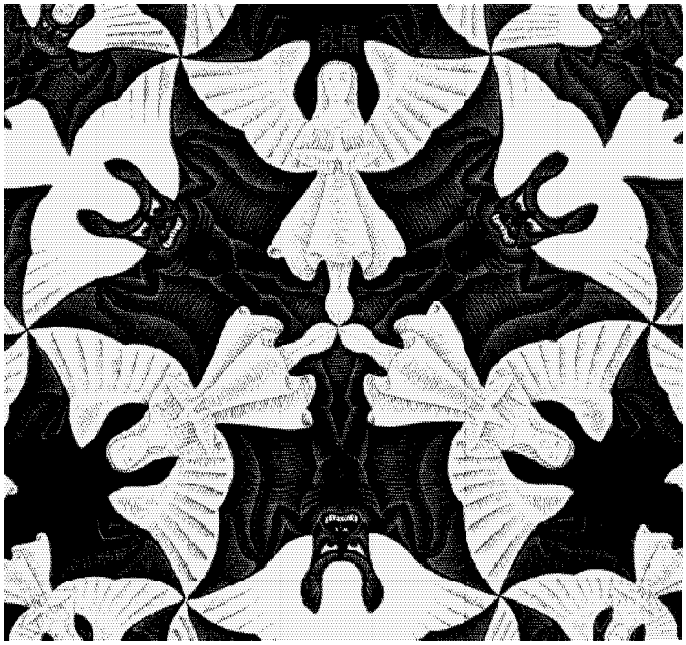
BORING2.GIF

BORING3.GIF

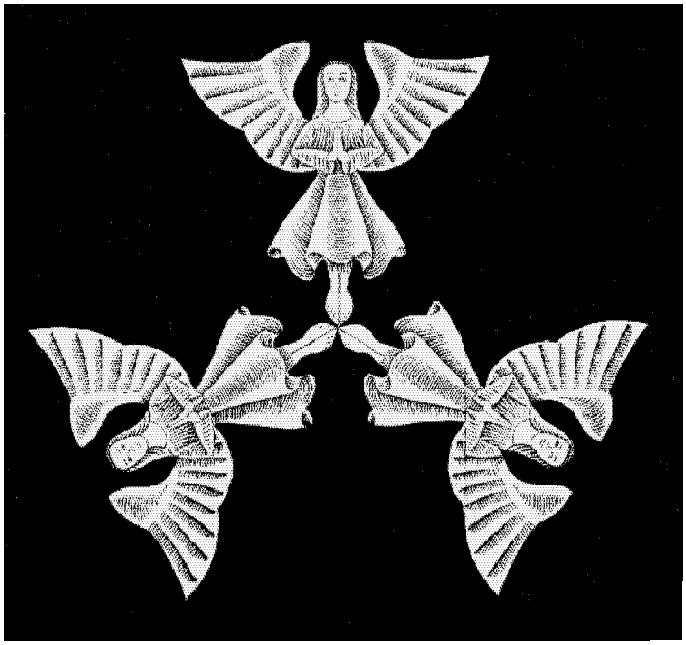
BORING4.GIF

BORING5.GIF

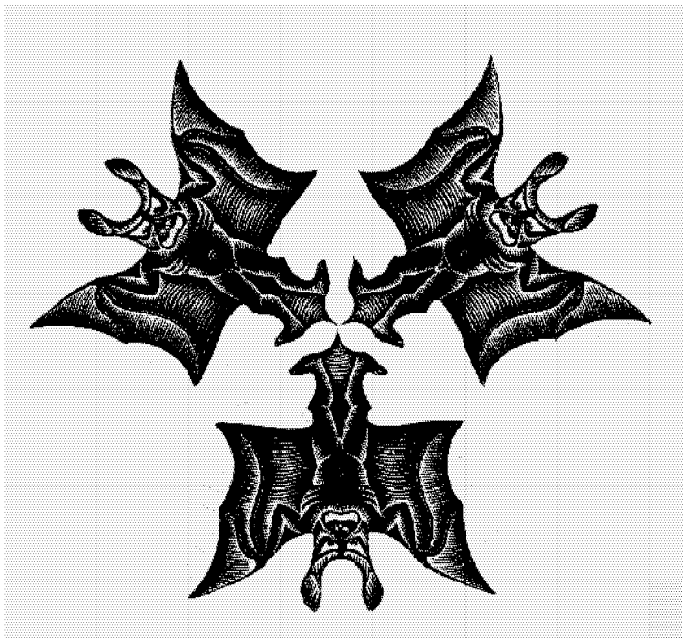
C.4 The Escher series



ESCHER1.GIF

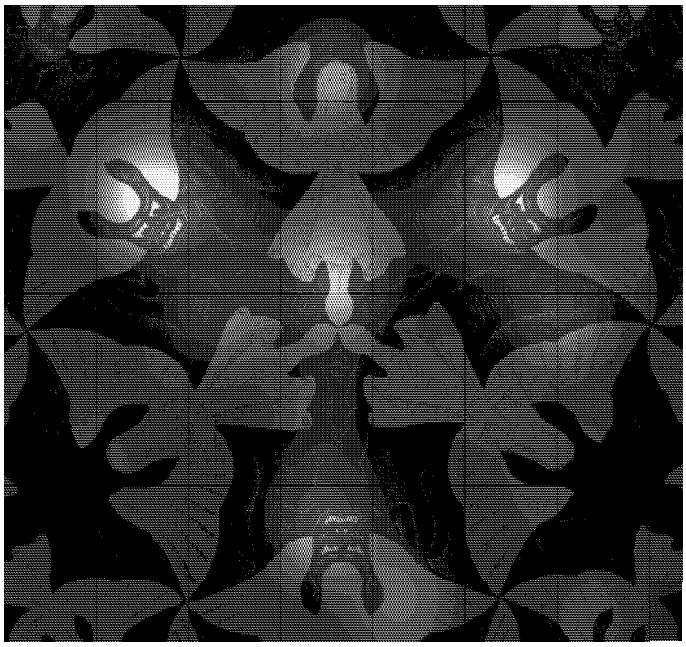


ESCHER2.GIF



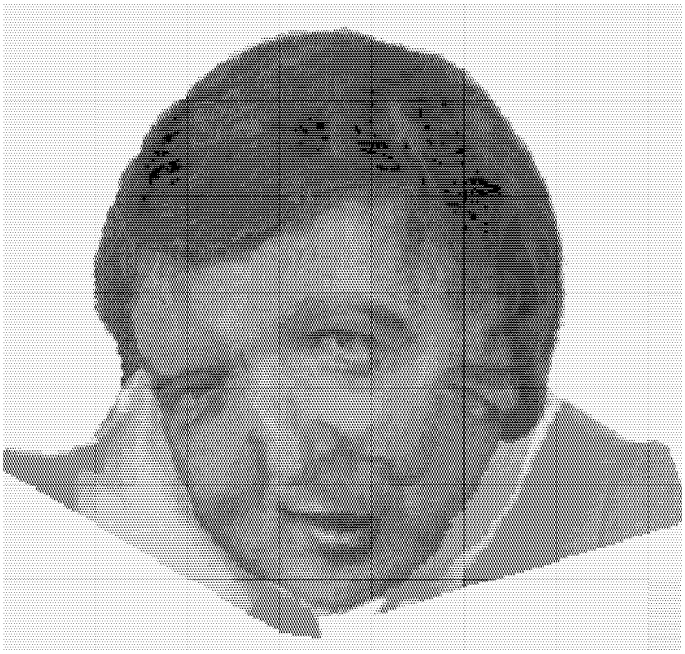
ESCHER3.GIF

ESCHER4.GIF



ESCHER5.GIF

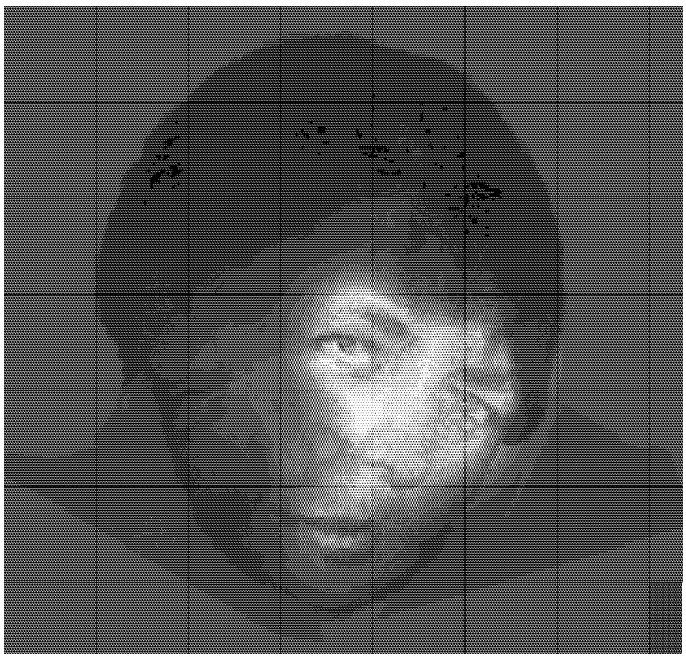
C.5 The faces series



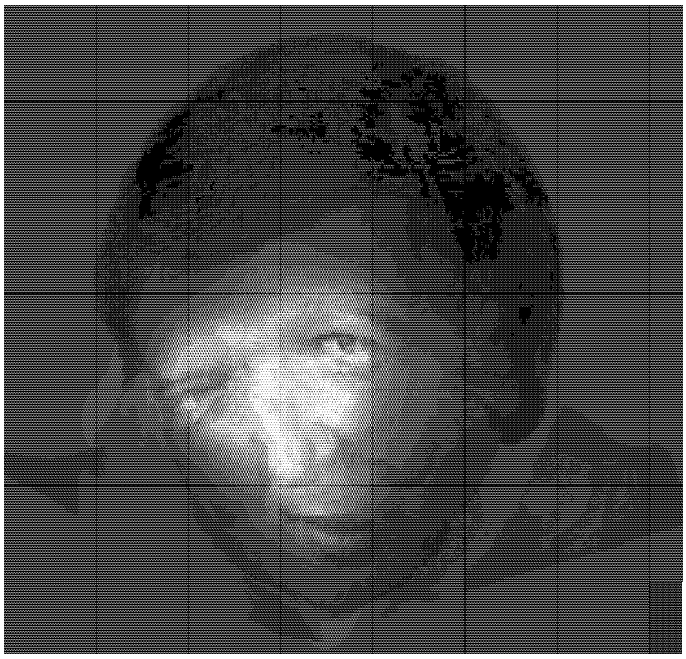
FACES1.GIF

FACES2.GIF

FACES3.GIF

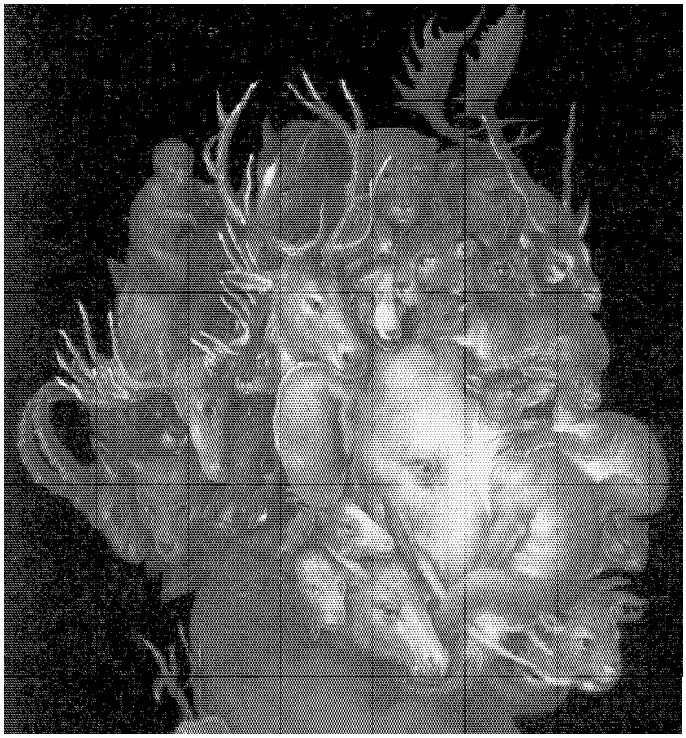


FACES4.GIF

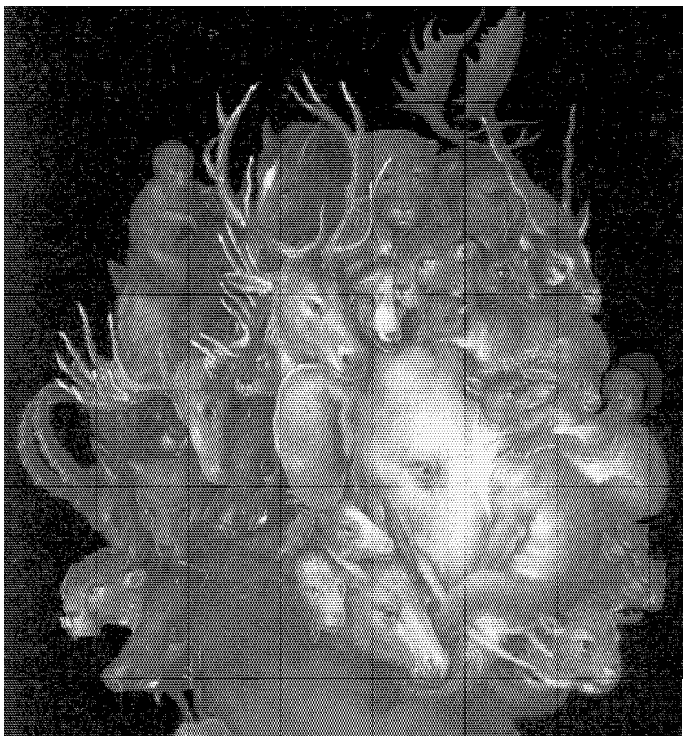


FACES5.GIF

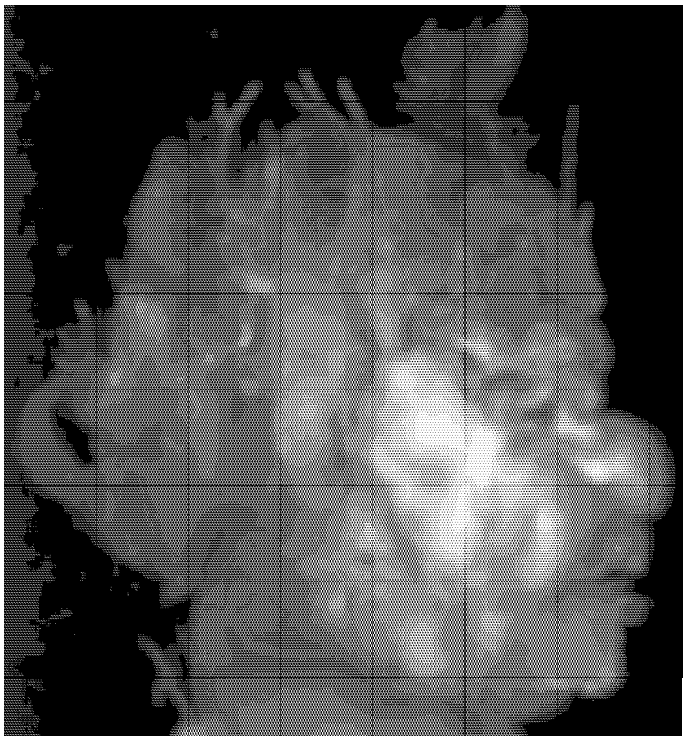
C.6 The Arcimboldo series



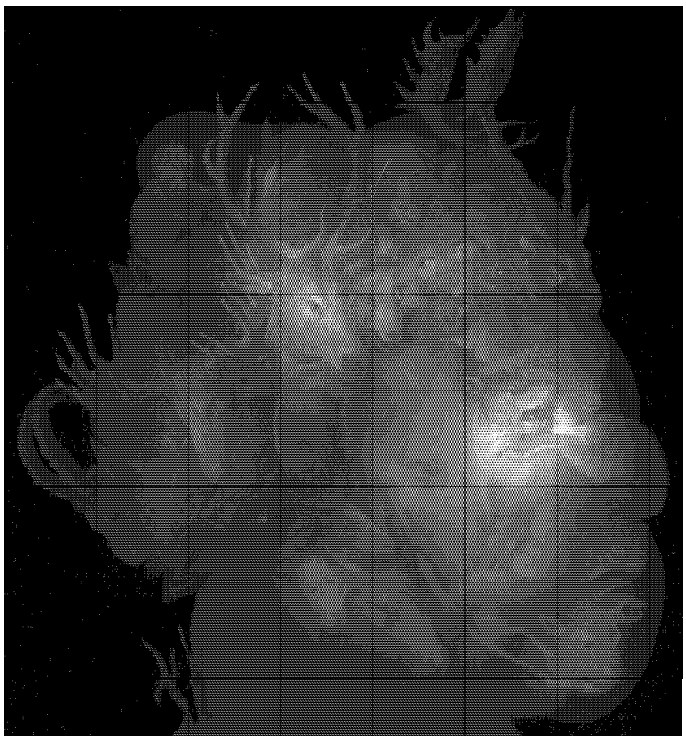
ARCIMB1.GIF



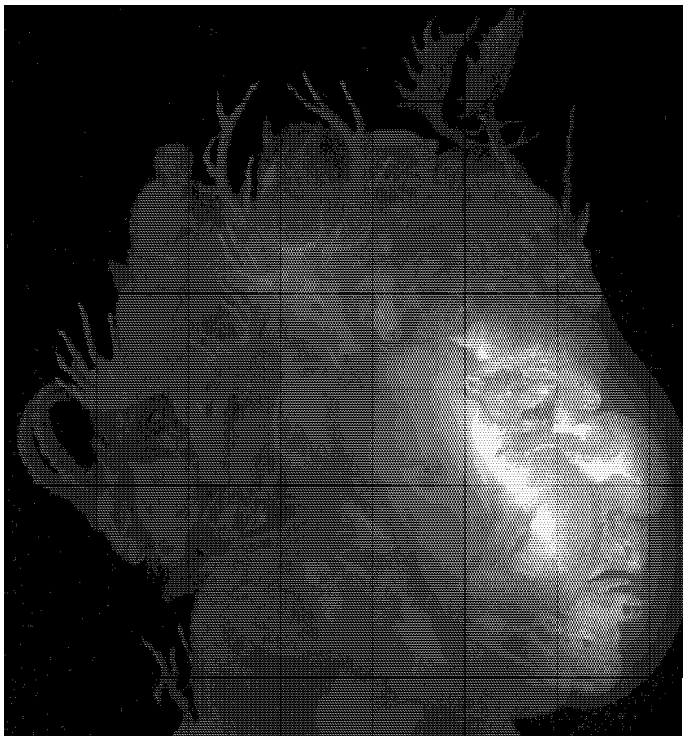
ARCIMB2.GIF



ARCIMB3.GIF



ARCIMB4.GIF



ARCIMB5.GIF

References

Ball K K, Beard B L, Roenker D L, Miller R L, Griggs D S, 1988 "Age and visual search: Expanding the useful field of view" *Journal of the Optical Society of America* A, **5** 2210–2219

Ballard D, Brown C, 1982 *Computer vision* (Engliwood-Cliffs: Prentice Hall)

Boring E G, 1942 *Sensation and perception in the history of experimental psychology* (New York: Irvington)

Caelli T M, 1988 "An adaptive computational model for texture segmentation " *IEEE Transactions on Systems, Man, and Cybernetics* **18** 9–17

Chapman J (Ed), 1987 *The Arcimboldo Effect: Transformation of the face from the sixteenth to twentieth century* (Milano: Fratelli Fabri Editori)

Cooper L, 1994 "Mental representation of visual objects and events" in *International perspectives on psychological science: The state of the art* Eds G d'Ydewalle, P Eelen & P Bertelson (Hove, UK/Hillsdale, NJ: Lawrence Erlbaum Associates)

Ellis S R, Stark L, 1978 "Eye movements during the viewing of Necker cubes" *Perception*, **7** 575–581

Gale A G, Findlay J M, 1983 "Eye movement patterns in viewing ambiguous figures" in *Eye movements and psychological functions: International views* Eds R Groner, C Menz, D F Fisher & R A Monty (Hillsdale, NJ: Lawrence Erlbaum Associates)

Garcia-Perez M A, 1992 "Eye movements and perceptual multistability" in *The role of eye movements in perceptual processes* Eds E Chekaluk & K R Llowellyn (Amsterdam: North Holland)

Gippenreiter Yu B, Romanov V Ya, 1972 "A method of investigation of the internal form of visual activity" in R MacLeod and H L Pick, Jr, 1974 *Perception: Essays in honor of James J Gibson* (Ithaca: Cornell University Press)

Gombrich E H, 1969 *Art and illusion: A study in the psychology of pictorial representation* 2nd ed (Princeton, NJ: Princeton University Press)

Hacisalihzade S S, Stark L W, Allen J S, 1992 "Visual perception and sequences of eye movement fixations" *IEEE Transactions on Systems, Man, and Cybernetics* **22** 474–481

Hood D C, Finkelstein M A, 1986 "Sensitivity to light" in *Handbook of perception and human performance, Vol 1: Sensory processes and perception* Eds K R Boff, L Kaufman, J P Thomas (New York: John Wiley and Sons)

Lee S W, Bajcsy R, 1992 "Detection of specularity using color and multiple views" *Image and Vision Computing* **10** 643–653

Lienert G, 1973 *Verteilungsfreie Methoden in der Biostatistik* 2 Auflage, Bd 1 (Meisenheim/Glan: Verlag Anton Hain)

Mackworth N H, 1976 "Stimulus density limits the useful field of view" in *Eye movements and psychological processes* Eds R A Monty, J W Senders (New York: John Wiley and Sons)

Nodine C F, Kungel H L, Toto L C, Krupinsky E A, 1992 "Recording and analysing eye-position data using a microcomputer workstation" *Behavioral Research Methods, Instruments, and Computers* **24** 475–485

Norman G R, Coblentz C L, Brooks L R, Babcook C J, 1992 "Expertise in visual diagnostics: A review of the literature" *Academic Medicine Rime Supplement* **67** 78–83

Noton D, Stark L W, 1971 "Scanpaths in eye movements during pattern perception" *Science* **171** 308–311

Pillalamari R S, Barnette B D, Birkmire D, 1993 "Cluster: A program for the identification of eye-fixation-cluster characteristics" *Behavioral Research Methods, Instruments, and Computers* **25** 9–15

Pomplun M, Velichkovsky B M, Ritter H, 1994 "An artificial neural network for high precision eye movement tracking" in *Lecture notes in artificial intelligence: AI-94 Proceedings* Eds B Nebel & L Dreschler-Fischer (Berlin: Springer Verlag)

Posner M, 1978 *Chronometric exploration of mind* (Hillsdale, NJ: Lawrence Erlbaum Associates)

Rayner K, Well A D, Polatsek A, Bertera J H, 1982 "The availability of useful information to the right of fixation in reading" *Perception and Psychophysics* **31** 537–550

Ritter H, Martinetz T, Schulten K, 1992 *Neural computation and self-organizing maps* (Reading, MA: Addison-Wesley)

Rock I, Hall S, Davis J, 1994 "Why do ambiguous figures reverse?" *Acta Psychologica* **87** 33–59

Stampe D M, 1993 "Heuristic filtering and reliable calibration methods for video-based pupil-tracking systems" *Behavioral Research Methods, Instruments, and Computers* **25** 137–142

Stampe D M, Reingold E, 1993 "Eye movement as a response modality in psychological research" in *Proceedings of the Seventh European Conference on Eye Movements, Durham* University of Durham, 31st of August – 3rd of September

Velichkovsky B M, 1995 "Communicating attention: Gaze-position transfer in cooperative problem solving" *Pragmatics and Cognition* **3**(2), 199–222.

Velichkovsky B M, Luria A R, Zinchenko V P, 1973 *Psychology of perception* (Moscow: Moscow University Press [in Russian])

Velichkovsky B M, Pomplun M, Rieser J, 1995 in press "Attention and Communication: Eye-Movement-Based Research Paradigms" *Visual Attention and Cognition* Eds W H Zangemeister, H S Stiehl & C Freksa (Amsterdam: Elsevier Science Publishers)

Vicholkovska A, 1906 "Illusion of reversible perspective" *Psychological Review* **13** 276–290

Yarbus A, 1967 *Eye movements and vision* (New York: Plenum Press)

Zinchenko V P, Vergiles N Yu, 1972 *Formation of visual image: Studies of stabilized retinal images* (New York: Plenum Press)