Color-based face detection in the "15 seconds of fame" art installation*

Franc Solina, Peter Peer, Borut Batagelj,
Samo Juvan, Jure Kovač

franc.solina@fri.uni-lj.si
University of Ljubljana,
Faculty of Computer and Information Science
Tržaška 25, SI-1000 Ljubljana, Slovenia

27th November 2002

Abstract

"15 seconds of fame[1]" is an interactive art installation which elevates the face of a randomly selected observer for 15 seconds into a "work of art". The installation was inspired by Andy Warhol's statement that "*In the future everybody will be world famous for fifteen minutes*" as well as by the pop-art style of his works.

The critical part of the installation is detection of faces which enables the cropping of faces from the wide-angle image of people standing in front of the installation. In the paper we describe color-based face detection in detail and many problems that had to be solved to make it reliable in different illumination conditions and scenes.

1 Introduction

Technology influenced artistic production throughout history. In the later part of the 20th century the fast advent of computer and information technology in particular left a strong impression in the art world [5, 20].

*Submitted to *Computer Vision and Image Understanding*, Special issue on Colour for Image Indexing and Retrieval

[1]This work was supported by the Ministry of Education, Science and Sports, Republic of Slovenia (Program Computer Vision 1539-506)

Figure 1: LCD computer monitor dressed up like a precious painting. The round opening above the picture is for the digital camera lens.

The Computer Vision Laboratory at the University of Ljubljana collaborates with the Video and New Media Department at the Academy of Fine Arts in Ljubljana by supporting artistic creativity using Internet, teleoperation and web cameras since 1995 [2, 4, 18]. The installation described in this paper was envisioned by Franc Solina already in 1996 and finally fully implemented in 2002 with the help of his graduate students. The "15 seconds of fame" installation was publicly exhibited at the 8th International festival of computer arts, 28 May-1 June 2002 in Maribor, at interINFOS'02, 21-25 October 2002 in Ljubljana and in the Finžgar gallery, 14-30 November 2002 in Ljubljana [1].

The installation "15 seconds of fame" was inspired by Andy Warhol's celebrated statement that "*In the future everybody will be world famous for fifteen minutes*" and his photography derived paintings of famous people. Warhol took faces and persons from mass media, banal in their newspaper everydayness, and transformed them into paintings by performing some kind of color surgery on them. By separating the face from the background or selecting distinct facial features (i.e. mouths, eyes, hair) he obtained image areas which he highlighted or covered with paint. Warhol portrayed in this fashion celebrities from arts and politics (i.e. Mao-Tse Toung, Marilyn Monroe, Jackie Kennedy, etc.). Some of these images are true icons of the 20th century [7].

The installation "15 seconds of fame" intends to make instant celebrities out of common people by putting their portraits on the museum wall (Fig. 1). Instead of providing 15 minutes of fame as prophesied by Warhol we decided

Figure 2: A group of people in front of the installation.

to shorten this time to 15 seconds to make the installation more dynamic. This period also limits the time necessary for computer processing of each picture. Since the individual portraits, which are presented by the installation, are selected by chance out of many faces of people who are standing in front of the installation in that moment, the installation tries to convey that fame tends to be not only short-lived, but also random. However, people can prolong their fame by ordering their portrait over the Internet, printing it on paper, framing it and putting it on the wall.

1.1 How the installation works

The visible part of the "15 seconds of fame" installation consists of a LCD computer monitor which is framed like a precious picture and hangs on the wall. The digital camera is hidden behind the frame above the computer monitor so that only a round opening for the lens is visible (Fig. 1). Pictures of gallery visitors which are standing in front of the installation (Fig. 2) are taken by the digital camera using a wide-angle lens setting (Fig. 3a). The camera is connected with the hidden computer via USB connection, so that the camera can be remotely controlled by the computer. Each digital photo is analyzed by the computer to detect faces in it (Fig. 3b, c, and d show steps in the face detection). The software then randomly selects one of the faces and crops it from the larger picture (Fig. 4 top left). This processing performs the same function as a photographer who would take from that point a portrait of one of the visitors using a telephoto lens.

The selected portrait is then transformed using randomly selected color filters to automatically generate a Warhol inspired pop-art portrait (Fig. 4).

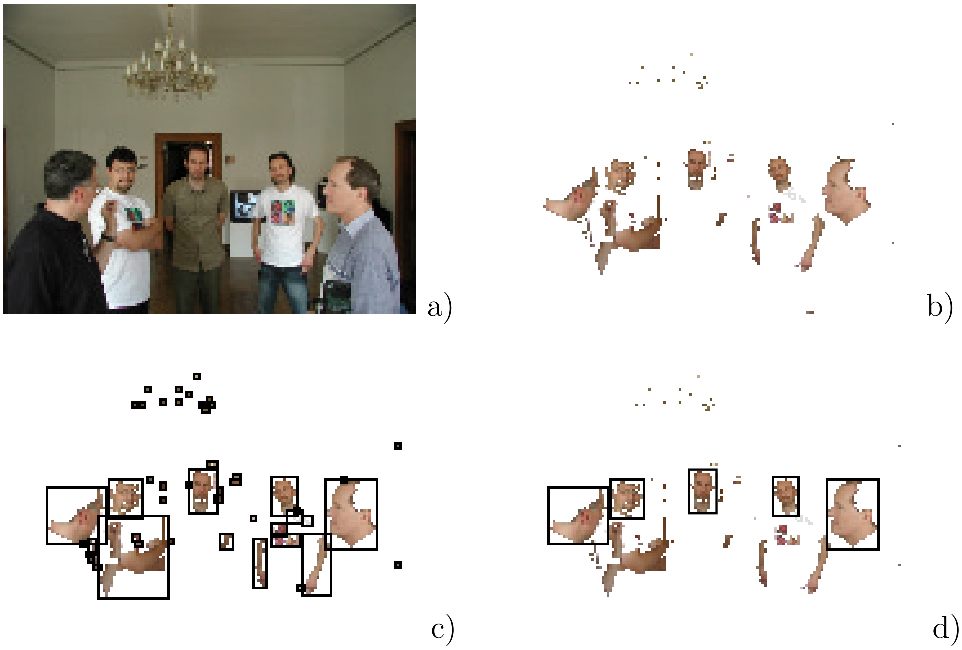

Figure 3: Stages in the process of finding faces in an image: a) downsize the resolution 2048×1536 of the original image to 160×120 pixels, b) eliminate all pixels that can not represent a face, c) segment all the regions containing face-like pixels using region growing, d) eliminate regions which can not represent a face using heuristic rules.

The resulting portrait is then displayed for 15 seconds on the picture/monitor together with a unique ID number. In the mean time, the processing of the next portrait is taking place, so that after fifteen seconds another pop-art portrait can be displayed. In this fashion every fifteen seconds a new picture is taken, a new portrait selected and processed, so that it can be displayed in the next 15 second period.

If several people are standing in front of the installation then the software tries to select each time a different person. Even if there is just a single person present in front of the installation, the randomly selected graphic effects assure that the displayed portraits are never the same. If the system does not detect any face, the last detected face is being displayed, but with a different graphic effect in each 15 seconds period. Some examples of such transformations of a single portrait can be seen in Fig. 4. Since portraits displayed by the installation are saved in a database, they can be ordered by sending e-mail to `15sec@lrv.fri.uni-lj.si` with the corresponding portrait ID number in the subject field.

The outline of the rest of the paper is as follows: in Section 2 we explain

Figure 4: Some color transformations in pop-art manner of the face selected in Fig. 3.

the process of detecting human faces in images, Section 3 is on eliminating the influence of non-standard illumination in images, graphical transformations to achieve pop-art effects are described in Section 4, Section 5 describes how the resulting portraits are displayed and how e-mail ordering of portraits works, and finally, Section 6 concludes the article.

2 Finding faces in color images

Automatic face detection is like most other automatic object-detection methods difficult, especially if sample variations are significant. Large sample variations in face detection arise due to a large variety of individual face appearances and due to differences in illumination. Note that any face recognition must be preceded by face detection.

There are a few distinct approaches to face detection (for a detailed survey see [13]). *Top-down model-based approach* assumes a different face model at different coarse-to-fine scales. For efficiency, the image is searched at the coarsest scale first. Once a match is found, the image is searched at the next finer scale until the finest scale is reached. In general, only one model is assumed in each scale (usually in the frontal-parallel view) and thus it is difficult to extend this approach to multiple views (faces seen from profile).

Bottom-up feature-based approach searches the image for a set of facial features and groups them into face candidates based on their geometric relationship. This approach is difficult to extend to multiple views because the image structure of the facial features varies too much.

In *texture-based approach* faces are detected by examining the spatial distribution of the gray-level information in the subimage. This is again not easily extensible to multiple viewpoints.

Neural network approach detects faces by subsampling different image regions to a standard-sized subimage and then passing it through a neural network filter. The algorithm performs well for frontal-parallel faces, but performance deteriorates when extended to different views of the face.

Color-based approach labels each pixel according to its similarity to skin color and subsequently labels each subregion as a face if it contains a large blob of skin color pixels. It can cope with different viewpoint of faces but it is sensitive to skin color and face shape.

Motion-based approach uses image subtraction to extract the moving foreground from the static background. The face is then located by examining the silhouette or the color of the difference image. This approach will not work well when there are a lot of moving objects in the image.

In *depth-based approach* primary facial features are localized on the basis of facial depth information. On a pair of stereo images containing frontal face views point correspondences over a large disparity range are determined using a multiresolution hierarchical matching algorithm. Finally, the facial features are located based on depth information.

We decided to use in the installation a color-based approach of face detection that we developed earlier [16]. In next Subsection 2.1 we describe our original method and in Subsection 2.2 the simplifications of this method that we had to make for the installation are revealed.

2.1 Our original face detection method

We developed a face detection method which consists of two distinct parts: making of face hypotheses and verification of these face hypotheses [16]. This method combines two approaches: color-based and feature-based approach.

The basic idea of the method is as follows: find in the image all skin-like regions, which contain possible candidates for an eye, then on the basis of geometric face characteristics try to join two eye candidates into an eye pair and finally, confirm or refuse the face candidate using skin color information. The main steps of the method described in [16] is illustrated in Fig. 5.

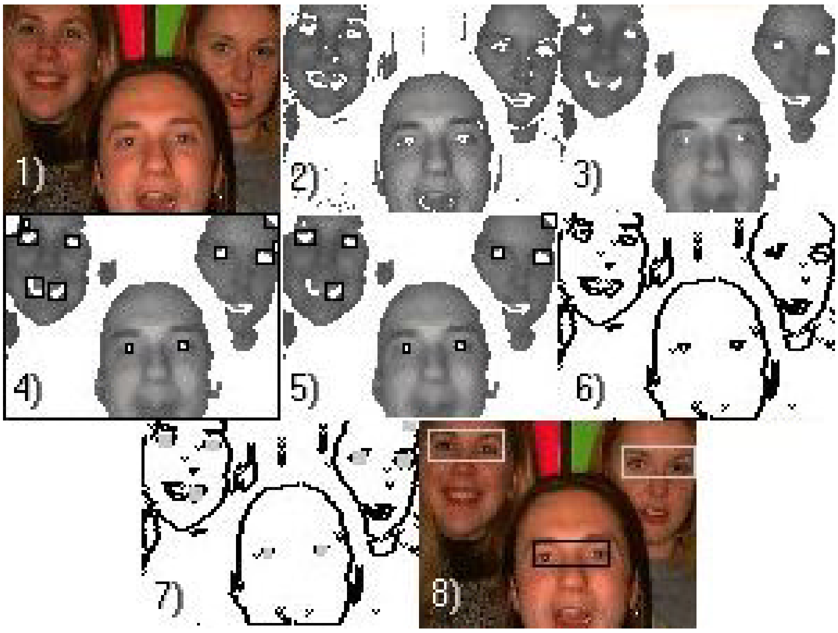

Figure 5: Basic steps of our face detection method [16]: 1) input image, 2) eliminated all non-skin colors, 3) image filtered with median filter, 4) segmented white regions, 5) eliminated insignificant white regions, 6) traced edges, 7) circles within candidate face regions indicating possible eye positions, 8) output image with confirmed faces.

The goal of the method was to reach maximum classification accuracy over the learning and testing sets of images under the following constraints: near real-time operation on a standard personal computer, plain background, uniform ambient natural or studio illumination, faces of fair-complexion, which must be entirely present in the image, and faces turned away from the frontal view for at most 30 degrees.

During the processing the method requires some thresholds, which are set quite tolerantly, but they become effective in a sequence. All thresholds were defined experimentally. The method was developed using different training sets of pictures [3] and tested over an independent testing set of two public image-databases [15, 17] with good results [16].

Two basic limitations of the method are that the face must be big enough (measured in pixels) and that it is sensitive to the skin complexion.

Since the original face detection algorithm is computationally demanding, we decided to develop a simpler version for integration in the "15 seconds of fame" installation.

2.2 The simplified face detection method

In order to confirm a given face candidate in our installation, we modified the original face detection method as follows. When the color picture is downloaded from the digital camera (we normally use resolution 2048×1536 pixels) to the computer, the system reads the picture and first decreases the resolution in a pyramid manner down to 160×120 pixels. The system searches for face candidates in the lower resolution to speed up the process, but the selected face is later cropped from the original resolution for final processing. The system then eliminates from the picture all pixels that do not correspond to skin color. After this color segmentation the system applies a region growth algorithm, which segments all face-like pixels into candidate face regions.

Each candidate face region must pass some simple tests to qualify as a true face region. Since the confirmation of hypothesis in the simplified method relies entirely on simple heuristic tests and not on structural information (i.e. eye pairs), we had to make them more stringent. Initially, for example, we had problems with bare arms which the system recognized as faces. The face candidate region must be large enough (based on the assumption about the minimal size of the face in the original picture), have rough face proportions (width/height ratio of the rectangular box containing the candidate face region), and have a high enough percentage of skin pixels within this box. The results are much better, although still not perfect. However, this is not too annoying—now and then it still happens that someone's arm or palm becomes famous for 15 seconds. The side benefit of the simplified method is that faces seen from the profile can also be detected.

The final step of finding a face for the portrait is selecting one of the faces (randomly, but with help of some heuristics, like giving bigger priority to bigger regions and regions that are higher in the picture), mapping the face coordinates from the lower resolution picture to the original resolution picture and cropping the face out of the original picture. Fig. 3 illustrates the described process. The selected portrait is cropped from the original digital image as a square area and resized to the resolution of 400×400 pixels.

The databases that we used for developing and testing of our original face detection method contained pictures taken in daylight or under white-balanced studio illumination. When the performance of the face detection algorithm during the first public exhibition was not fully satisfactory, it was mainly due to difficult illumination conditions (Tab. 1). Once we realized that we will not be able to completely control the illumination in different exhibition spaces, we decided to improve our face detection results by eliminating the influence of non-standard illumination.

All faces	Detected	TP	FP	FN	TP/Det.	FN/All
1972	1458	1168	272	768	81.34	38.95

Table 1: Face detection statistics on a subset of images taken at the first public showing of the installation. They were taken under close to standard illumination and no preprocessing was applied before. Shown are the number of all faces, the number of all detections (Detected), the number of detected faces as true positives (TP), number of false detections as false positives (FP) and number of faces missed as false negatives (FN). TP/Det shows the percentage of true positives out of all detections and FN/All shows the percentage of false negatives out of all faces in the image set. The first percentage is for the installation extremely important, while the second one is merely informative, since we have consciously eliminated faces that were too small for further processing, but were included in the number of all faces!

3 Eliminating the influence of non-standard illumination

The purpose of studying methods for eliminating the influence of non-standard illumination in our project is to improve the results of our face detection algorithm. Non-standard illumination are by definition those that are more or less different from daylight illumination (CIE standard D65) [21]. We find such illumination almost anywhere in enclosed spaces with artificial illumination, where the installation could potentially be exhibited. There are two main groups of methods for addressing this problem: color compensation methods and color constancy methods.

3.1 Color compensation methods

Methods in this group have low time complexity (order of $O(n^2)$) and they do not need a preliminary learning step. This means that they are easy and straightforward to implement. Their effectiveness is relatively high on sets of images with some input constraints. Illumination should be relatively close to standard illumination. The input image is transformed in the way that the colors in the image are leveled in respect to some statistical quantity.

3.1.1 Grey World

Grey World (GW) [11] algorithm presents simple and fast method for color compensation on images which are defined in RGB color space [10]. It is based on the presumption that the average surface color on the image is acromatic. This means that the average color, which is reflected from the surfaces, corresponds to the color of the illumination. To execute the algorithm we have to calculate the averages for each channel R, G and B for the whole image. Averages are then transformed with the help of a linear scale factor to values that correspond to the mean gray value of one standard illuminant. The corresponding scale factor for each channel R, G and B is calculated as

$$S_C = \frac{C_{std}}{C_{ave}}, \tag{1}$$

where C is one of the channels R, G and B, std stands for standard gray value and ave stands for average value of the channel in the whole image. We change the input image so that for each pixel we multiply the pixel's R, G and B values with corresponding scale factor (Eq. 1).

There are several possible values that can represent the mean gray value. One example of such a value is $\frac{1}{2}RGB_{std}$, which represents one half of ideal gray color under standard illumination (CIE D65 [21]). On the other hand, we can choose a similar constant value, which is reflected by the previously calculated average or by some test results.

The method is very effective if we have to correct the illumination in images that were captured in conditions close to standard illumination. Though the method is powerless in case of extreme illumination conditions, e.g. lights in a discotheque. The root of the problem can be found in the already mentioned presumption that the color of the illumination is equal to the average color of the image. Such a presumption is of course naive for realistic color reconstruction.

3.1.2 Modified Grey World

Modified GW [9] method is very similar to basic GW algorithm with the difference that the scale factors are calculated in a slightly different way. The average of each channel is calculated by counting each color only once, regardless of how many times the color appears in the image. By doing so, we eliminate the influence of colors represented by a large number of pixels in the image on the average color. The method is effective on images, which do not contain a lot of different colors. In the basic GW method prevailing colors can have big influence on the average color, resulting in lowering the influence of other colors in the image. The modified GW ensures that all the colors in the image are equally important in calculations.

For calculating the average we need information about all the colors in the image. This information can be gathered with the help of color histograms. This brings additional processing in the algorithm, making it more computationally demanding than the basic GW method.

3.1.3 White-Patch Retinex

Retinex [11] method is like the Modified GW method just a special version of the basic GW method. The difference lies again in the method of calculating the scaling factors. In case of Retinex, instead of the average color we use the maximal value of each channel in the whole image (compare with Eq. 1)

$$S_C = \frac{C_{std}}{C_{\max}}. \tag{2}$$

The Retinex method is above all suitable for dark images. In intensively illuminated scenes the maximal value of each channel is close to the saturation value of 255. Necessary changes in such cases do not take place. The complexity of the Retinex method is the same as the complexity of the basic GW method.

3.2 Color constancy methods

Methods belonging to this group differ from the color compensation methods above all in the need to integrate a preliminary learning step. They need the knowledge about illumination properties and properties of the capturing devices, e.g. digital cameras. The input image is then transformed in such a way that it reflects the state, which is independent of the illumination. Thus a stable presentation of colors under different illuminations is achieved. Generally speaking the methods consist of two distinct steps: scene illumination detection and standard illumination reconstruction. In the first step, the algorithm determines with the help of preliminary knowledge which illumination out of the set of known illuminations is present in the image. In the second step, it applies the necessary transformations to reconstruct the standard (or other wanted) illumination.

3.2.1 Color by Correlation

In the Color by Correlation method [9] the preliminary knowledge about the illumination properties is represented with a set of colors, which can appear under specific illumination, i.e. colors that are visible under specific illuminant.

Processing is done in 2D color space [10] and it enables us to reconstruct the color of illumination up to the precision of a multiplicative constant.

With this estimation we are able to transform colors within the input image into an illumination independent state. After this illumination independent state is known (image illuminant is detected), it is possible to reconstruct the image under other possible illuminations, e.g. standard illumination.

For 2D color space YUV space was chosen, because its usage is widely spread and it is similar to the TSL color space in the sense of chroma and brightness separation. TSL color space proved itself very well regarding face detection [19]. U and V coordinates represent chroma and Y represents brightness caused by illumination.

The Color by Correlation method consists of three steps. First, we have to code the illumination properties that represent the preliminary knowledge about which colors are possible under specific types of illumination. This knowledge is in the second step correlated to the color information in the actual input image. For each possible illumination we calculate the possibility that this illumination represents the illumination used in the input image. In the third step, we use this result to reconstruct the required illumination.

Correlation matrix $\mathbf{M}$ (Tab. 2) is used to code the preliminary knowledge. First of all, we have to define the range of colors for each possible illumination, which are visible under specific illumination. We can obtain this information by taking one or more pictures of differently colored objects (e.g. Macbeth color checker in Fig. 6) under the same illumination and check which colors are present in the image. With the help of this information we compose the probability distribution. This distribution presents the probability of appearance of each color under specific illumination. Probability distributions for all possible illuminations compose the columns of correlation matrix $\mathbf{M}$ (Tab. 2).

As we compose the correlation matrix $\mathbf{M}$, we compose also the vector $\vec{\mathbf{v}}$ out of the input image by counting pixels of the same color (N_{x_i,y_j}). The vector components are set to 1 or 0 according to Eq. 3

$$\vec{\mathbf{v}}_{x_i,y_j} = \begin{cases} 1 & ; \quad N_{x_i,y_j} > 0 \\ 0 & ; \quad \text{otherwise.} \end{cases} \tag{3}$$

If in vector $\vec{\mathbf{v}}$ there is a 0 for color (x_i, y_j), then this color is not represented in the image.

For each color in the image we calculate the probability P_c. The probability is calculated by counting all pixels of the same color (N_c or N_{x_i,y_j}) and then dividing that number with the number of all pixels in the image (N_{all})

$$P_c = \frac{N_c}{N_{all}}. \tag{4}$$

color	ill_1	ill_2	ill_3	$\ldots$	$\ldots$
x_1, y_1	$p(x_1, y_1\|ill_1)$	$p(x_1, y_1\|ill_2)$	$p(x_1, y_1\|ill_3)$	$\ldots$	$\ldots$
x_1, y_2	$p(x_1, y_2\|ill_1)$	$p(x_1, y_2\|ill_2)$	$p(x_1, y_2\|ill_3)$	$\ldots$	$\ldots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_1, y_n	$p(x_1, y_n\|ill_1)$	$p(x_1, y_n\|ill_2)$	$p(x_1, y_n\|ill_3)$	$\ldots$	$\ldots$
x_2, y_1	$p(x_2, y_1\|ill_1)$	$p(x_2, y_1\|ill_2)$	$p(x_2, y_1\|ill_3)$	$\ldots$	$\ldots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_n, y_n	$p(x_n, y_n\|ill_1)$	$p(x_n, y_n\|ill_2)$	$p(x_n, y_n\|ill_3)$	$\ldots$	$\ldots$

Table 2: Correlation matrix $\mathbf{M}$: The first row and the first column are not part of the matrix, their purpose is merely representational. In the first column colors in 2D color space [10] are given. Matrix columns represent probability distribution of colors for each possible illumination. Each cell in the matrix reveals the probability of color x_i, y_j under illumination ill_k.

The correlated value of two input images tells us how the input images coincide. The bigger is their correlation value the more alike are the input images. For instance; vectors $\vec{\mathbf{a}}$ and $\vec{\mathbf{b}}$ highly coincide if the value $\vec{\mathbf{a}} \circ \vec{\mathbf{b}}$ is big. The correlation value is computed by operation $\circ$ (correlation)

$$\vec{\mathbf{a}} \circ \vec{\mathbf{b}} = \vec{\mathbf{a}}^{\mathrm{T}} \cdot \vec{\mathbf{b}}. \tag{5}$$

Now we have to determine the correlation between each of the possible illuminations and the input image. This can be done by calculating the vector

$$\vec{\mathbf{l}} = \vec{\mathbf{v}}^{\mathrm{T}} \circ \mathbf{M}. \tag{6}$$

The elements of this vector present the correlation value for each of possible illuminations.

In the final step, we have to estimate the color of the actual illumination by finding the illumination, which has the maximal correlation with the input image

$$\begin{aligned} \hat{\mathbf{c}} &= Max(\vec{\mathbf{l}}) \cdot \mathbf{C}_{ill} \\ Max(\vec{\mathbf{x}}) &= \vec{\mathbf{x}}'; \quad \vec{\mathbf{x}}'_i = \begin{cases} 1; & \vec{\mathbf{x}}_i = \max(\vec{\mathbf{x}}) \\ 0; & \text{otherwise,} \end{cases} \end{aligned} \tag{7}$$

where $\hat{\mathbf{c}}$ is a 2-component (U,V) color vector, which represents the estimation of color of illumination, $\mathbf{C}_{ill}$ is a matrix with dimensions $N \times 2$ that stores the estimations of colors of N illuminations, between which we choose from, and function Max returns a vector in which a value 1 is associated only with the component with the maximal value of input vector, while all other components are set to 0. $\vec{\mathbf{l}}$ is expressed in Eq. (6). $\hat{\mathbf{c}}$ actually represents the detected illumination and $\mathbf{C}_{ill}$ represents all possible illuminations!

The correlation method solves the problem of color constancy using the maximal probability approach. Within such a correlation framework we could also implement other algorithms [9], e.g. Gray World, neural networks based approach, Convex Gamut Mapping, Color by Voting etc. All we need to do is to change the correlation matrix in a way that it reflects the presumptions, constraints and connections between colors and illuminations suggested by the mentioned algorithms.

3.2.2 Illumination reconstruction

After the illumination detection based on correlation technique takes place (value $\hat{\mathbf{c}}$ in Eq. 7), we need to reconstruct the image scene under some standard illumination conditions. In order to perform such reconstruction, certain transformations should be applied. In our work this transformations are linear and based on straight line calculations.

To calculate the transformation parameters, we need the information about the spectral power distribution. We can gain this information with the help of the Macbeth color checker (Fig. 6) [8]. We need two images of the Macbeth color checker captured under different illuminations. The first one should be captured under the same illumination as the input image, which we want to reconstruct. The second image should be captured under standard (or other wanted) illumination, which we want to use in the reconstruction process.

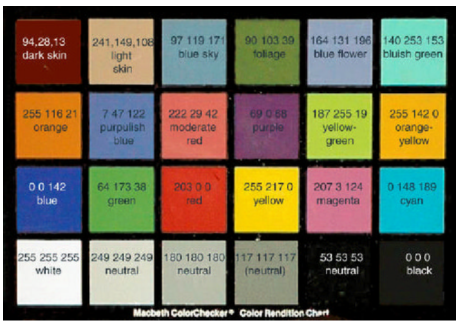

Figure 6: The Macbeth color checker [8].

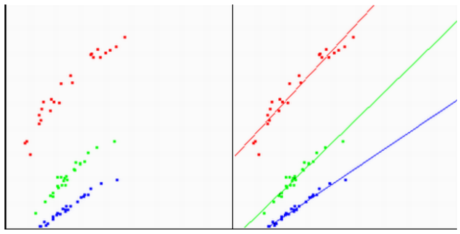

Figure 7: Graph of Macbeth's color averages for standard and unknown illumination: On the left side are shown the points of average values of each color box for all three color channels R, G and B. On the right side the linear approximation is given. x axis denotes the standard illumination averages and y axis represents the unknown illumination averages.

The Macbeth color checker is a device used in professional photography, where it serves as a reference for determining lighting conditions. The checker consist of 24 boxes that are of different colors. The chosen colors represent natural objects like human skin, plants, sky etc. Its purpose is mainly to help in the process of recognition and reconstruction of listed objects under different illuminations.

We obtain the transformation parameters by calculating the average values of all color channels R, G and B in each color box of Macbeth's checker. We do that for both checkers, i.e. for the one captured under standard illumination and for the one captured under unknown illumination. These values can be presented on a graph where we have standard illumination averages on the x axis and unknown illumination averages on the y axis (Fig. 7). We can observe that the points that belong to the same color channel indicate a linear shape, which can be well approximated with a straight line using the min-square method. The results of the min-square method reveal the parameters k and n of the straight line model $y = kx + n$ for each color channel (Fig. 7). This straight line model is then used to transform the colors from one illumination into the colors of the second (target) illumination. In this way, the transformation is simple and fast.

We can put these transformations for all possible illuminations also directly into the matrix $\mathbf{C}_{ill}$ (Eq. 7), if we want that vector $\hat{\mathbf{c}}$ carries the information about the needed transformation.

3.3 Comparison of methods

3.3.1 Color compensation methods

In order to determine the influence of these algorithms on our face detection results, some experiments were performed on the set of images gathered in

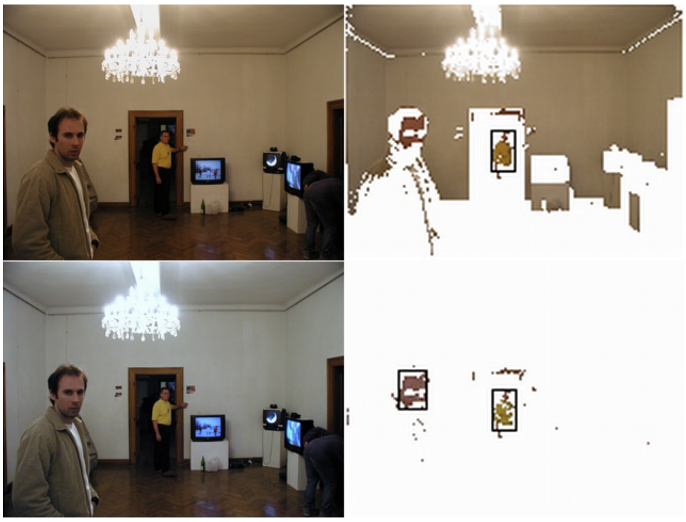

Figure 8: GW performance: the upper part of the figure shows how face detector failed to detect faces due to confusion caused by surrounding colors. This was improved by GW preprocessing as seen in the lower part of the picture. The image was chosen from the incandescent subset of images.

our lab and at the first public showing of the installation. The testing set is composed of 160 images taken under four different types of illumination conditions. One subset of images (40 images) was taken under standard daylight, in the second subset (40 images) objects were illuminated by incandescent lamps assembled into a chandelier, the third subset (40 images) was taken under camera flash light conditions, and the last subset of images (40 images) was taken under neon light illumination conditions. After that, one of the color compensation methods was applied and finally, face detection algorithm was applied to original and preprocessed images.

Results gathered in Tab. 3 show perceivable improvement in face detection on images taken under different than standard illumination conditions when one of the compensation algorithms was previously applied (see Figs. 8 and 9). Grey World algorithm performed especially well since for the flashlight, incandescent and neon light conditions a considerable increase in TP/Det percentage can be noticed. Whenever another algorithm performed better than Grey World, this had a significant influence on the FN/All percentage. Which means that the advantage has been gained not by increasing of true positives but by decreasing of false positives and as a consequence false negatives increased. We can easily observe this behavior in the Modified Grey World algorithm results. Either way, if the number of true positives is not too small, this has a very positive effect on our application, since from the

Method	**None**	**GW**	**MGW**	**RET**	**None**	**GW**	**MGW**	**RET**
Illuminant	**standard**				**incandescent**			
All Faces	109				95			
Detected	75	70	65	76	45	57	42	43
TP	68	65	60	68	28	45	31	29
FP	7	5	5	8	17	12	11	14
FN	40	44	48	40	67	50	64	66
TP/Det	90,66	92,85	92,31	89,47	62,22	78,95	73,81	67,44
FN/All	36,70	40,37	44,04	36,70	70,53	52,63	67,37	69,47
Illuminant	**flashlight**				**neon**			
All Faces	112				78			
Detected	55	47	43	39	63	64	29	60
TP	38	39	36	32	50	54	26	49
FP	17	8	7	7	13	10	3	11
FN	74	73	76	79	28	24	52	29
TP/Det	69,09	82,98	83,72	82,05	77,77	84,37	89,66	81,66
FN/All	66,07	65,18	67,86	70,53	35,90	30,77	66,66	37,18

Table 3: Color compensation results show the number of all detections (Detected), the number of detected faces as true positives (TP), number of false detections as false positives (FP) and number of faces missed as false negatives (FN) on four subsets of images which represent different illumination conditions (standard, incandescent, flashlight and neon), previously preprocessed by Grey World (GW), Modified GW (MGW), White Patch Retinex (RET) or no preprocession at all (None). Row All Faces shows the number of faces in particular subset of images. TP/Det shows the percentage of true positives out of all detections and FN/All shows the percentage of false negatives out of all faces in the subset. For the installation the first percentage is extremely important, while the second one is merely informative, since we have consciously eliminated faces that were too small for further processing but were included in the number of all faces!

Figure 9: Retinex performance: In the upper part of the figure we see two false detections, while on the lower part this problem was overcome by the Retinex algorithm. The image was taken from the flashlight subset.

application point of view it is crucial not to show false detections too often. A measure againts displaying false positive face detections is also built in the selection mechanism, which in principle selects faces at random, but still gives a higher priority to those faces which are bigger and higher in the input image.

All preprocessing techniques showed little or no improvements at standard illumination conditions. This was somehow expected since the original face detection algorithm was developed under presumption that standard illumination is present in the image.

All these results are dependent on our skin detection technique used in face detection algorithm, where skin color is detected which works in the 3D color space (RGB). Skin detection in 2D color space (YUV) might improve this results considerably as it is less brightness dependent than detection in the 3D color space (RGB) [10].

The results also show that the performance of these techniques depends very much on the type of the illumination. Therefore a considerable amount of precaution should be taken in decisions about the usage of these techniques. On the other hand all of these algorithms are very effective from the time complexity point of view and as such they enable the possibility of performing a simple initialization test when the scene is first placed under certain illumination. In this way we can determine which algorithm would produce the best results under certain type of illumination.

3.3.2 Color constancy methods: Illumination detection

Tests for the correlation method were performed in order to determine the best merit (function) for illumination detection. The testing set contained 156 images with none, one or multiple faces, taken under 7 various illumination conditions.

The *white* subset contains images taken without any light except for the flash light of the camera. This set represents the standard illumination conditions as it approximates conditions of standard daylight most closely. The color of other illumination conditions can be recognized from the name. Subset red_1 and $blue_1$ are taken under very extreme lighting conditions of particular color, while *red*, *blue*, *green* and *yellow* subsets represents more mild illumination conditions. Illuminations conditions were created with one regular incandescent 40 Watt lamp and one 40 Watt red lamp. Other colors were applied by creating a simple filter out of plastic glasses, which were put over an incandescent lamp.

Correlation merits represent the numbers stored in the correlation matrix and consequently they represent a base for determining the amount of correlation among two image illuminations. P stands for probability of color under particular illumination, $\log(P)$ is the logarithm of that probability, t/f (true/false) merit shows only whether certain color is possible under the illumination in question.

Results (Tab. 4) show that the best illumination detection is achieved by t/f merit, i.e. the smallest number of wrong decisions were made by this merit. We have also to take into consideration the fact that most false detections appeared at extreme illuminations and are due to the lack of color information. This means that one or two color channels at these illuminations had extreme values in almost every image in the subset. Also the majority of these false illumination detections were detected as similar illuminations, e.g. red_1 was detected as *red*.

Good detection results were also achieved by the probability merit, if we don't need detection of standard or close to standard illumination. The reason for these false detections lies in probability merit which is biased in favor of colors that are found in other than standard (*white*) illuminations. For instance, blue color will have a very high probability under blue illumination since blue is the predominating color in such illumination, while under standard illumination blue will have a relatively small probability as there will be many other colors present beside it. Under standard illumination more colors are possible and that is why their probabilities are relatively small. As such they are too small to beat other illuminations. Logarithm of probability and t/f merit reduce this bias, but are facing some other types of

Light	N_{Pics}	$\log(P)$	P	t/f
blue	26	1	0	0
$blue_1$	20	8	0	6
green	18	0	0	0
red	29	10	8	1
red_1	20	20	1	20
yellow	23	0	0	0
white	20	0	20	0
$\sum$	156	29	29	27

Table 4: Performance of correlation method: Illumination detection based on comparison of different correlation merits (probability, logarithm of probability and true/false merit). The table represents numbers of false illumination detections in the selected subset of pictures, except for N_{Pics} column, which stands for the number of pictures in a subset.

false detections. These false detections mostly occur at extreme illumination conditions ($red1$, $blue1$) and are often mistaken for the similar illuminations ($red1$ is labeled *red*), but almost never two illuminations of totally different types are mixed up.

3.3.3 Color constancy methods: Illumination reconstruction

Testing was performed basically on the same set of images as described in section 3.3.2. Subsets red_1 and $blue_1$ were removed since their illumination is to extreme to be successfully reconstructed, also some images which contain no faces were eliminated from this experiment.

First, illumination detection on images was performed with correlation technique, then the images were reconstructed under approximation of standard illumination (subset *white*) and finally face detection algorithm was applied on images (Figs. 10 and 11).

The results are summarized in Tab. 5. Results of subset *white* are shown as a comparable reference to other results. Columns with no method previously applied (None) can contain zero detections. This can occur if face detection algorithm finds no skin color in an image, which can often be the case in extreme illumination conditions.

Results in Tab. 5 show the positive effect of color correction on images with non-standard illumination conditions. If we do not apply preprocessing, the face detector finds some faces only on images from the yellow subset. After the preprocessing step almost all faces were recovered under this illu-

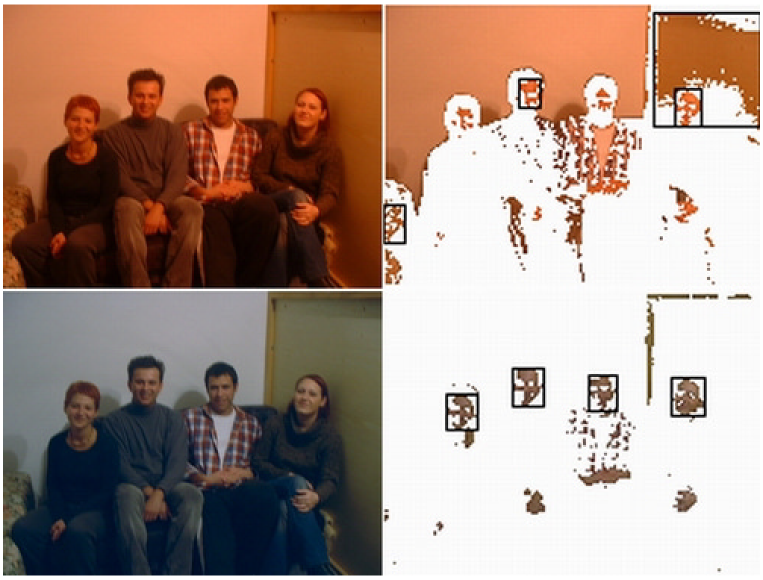

Figure 10: Correlation performance on image from the yellow subset: detection results on image without preprocessing (top) and after illumination reconstruction (bottom).

mination, while under all other illuminations the number of recovered faces was not that high, but the difference with the results gained without preprocessing is enormous.

In the blue subset we see a very large number of false detections caused by the mixture of incandescent and blue light. This mixture was necessary for enhancement of other than blue color channels (R and G) since blue filter was very strong. If only blue channel is present, we have as much information as in achromatic pictures. Incandescent light caused an interesting effect of appearing all shadows slightly yellow in reconstructed pictures and as a consequence many false faces were found.

Nevertheless, with the exception of blue illumination these results are quite comparable with the results in Section 3.3.1, where face detection was tested under close to standard illumination conditions.

3.4 Method selection

The nature of illumination in some galleries can represent a real problem since it normally differs from daylight illumination. In older rooms in particular we usually have chandeliers or similar types of illumination that emit a prevailing yellow light into the room and as a consequence a prevailing yellow color in captured images. This causes the shift of a large part of image color space

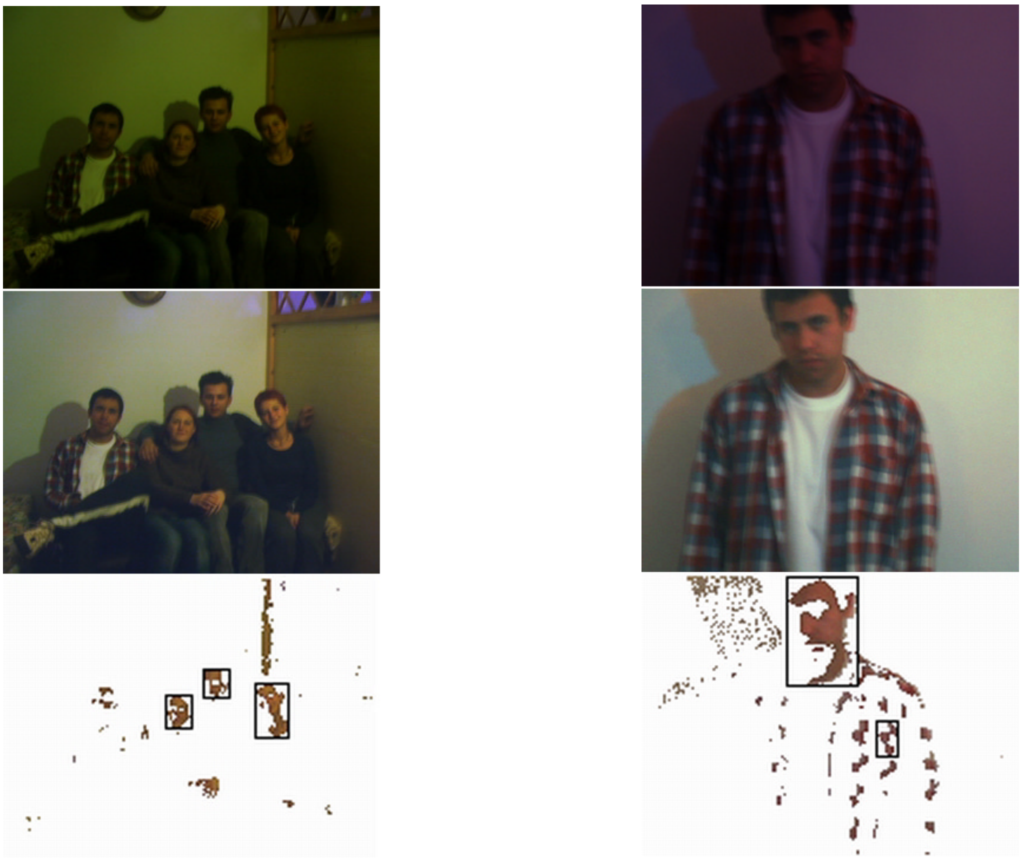

Figure 11: Correlation performance on images from the green (left) and blue (right) subset. On the original image (top) no faces could be detected. After corrected illumination (middle) three faces could be detected.

into the color subspace, which represents skin color. An illustrative example of this property can be observed on the white walls of a room. Normally they are white, but under incandescent lamp illumination they are more bright yellow than white. And since walls can occupy large parts of an image, it can happen that most of the image pixels are recognized as skin-like pixels (see Figs. 8 and 10). This type of illumination can have a serious negative influence on the number of false face detections (false positives).

In case of incandescent lamp illumination we should choose among color compensation methods described in Section 3.1. Based on the results of these algorithms and constraints discussed in Section 3.3.1, we decided to use GW algorithm as it performs best when minor lighting deviations from standard illumination are present. Although, some form of automatic selection is taken into future consideration.

Illuminant	*white*	*yellow*		*green*		*blue*		*red*	
Method	**None**	**None**	**C**	**None**	**C**	**None**	**C**	**None**	**C**
All Faces	38	35		23		42		34	
Detected	43	13	40	0	18	19	80	10	25
TP	35	7	34	0	13	0	29	0	23
FP	8	6	6	0	5	19	51	10	2
FN	2	28	1	23	10	42	13	34	11
TP/Det	81,39	53,85	85,00	0	72,22	0	36,25	0	92,00
FN/All	5,26	80,00	2,86	100,00	43,48	100,00	30,95	100,00	32,35

Table 5: Correlation results show the number of all detections (Detected), number of correct face detections as true positives (TP), number of detections that turned out not to be faces as false positives (FP) and the number of faces missed by detection algorithm as false negatives (FN) for different subsets of images (*white*, *yellow*, *green*, *blue* and *red*), previously processed by correlation algorithm (C) and with no preprocessing at all (None). Row All Faces shows the number of faces in a particular subset of images. TP/Det shows the percentage of true positives out of all detections and FN/All shows the percentage of false negatives out of all faces in the subset. The TP/Det is for the installation extremely important, while FN/All is merely informative for the performance of our face detector. Note that small faces are deliberately eliminated from further processing already by the face detection algorithm.

A totally different story can be observed in discotheques, where illumination emits color light (e.g. blue, green, red etc.) into the room. This shifts all scene colors towards the color of the illumination. Consequently, a lot of skin-like pixels are recognized as non-skin-like pixels and the number of correctly detected faces (true positives) is decreased, since we can not reliably find skin-like pixels.

When deviations from standard illumination are much more noticeable, we must choose a correlation technique with proper illumination reconstruction. When illumination conditions are constant over a large period of time, no illumination detection is necessary. By manually selecting the illumination we eliminate all the risks linked with false illumination detection and assure the best illumination reconstruction.

Eliminating the influence of non-standard illumination before face detection ensures much better results. The whole system is much more flexible and the installation can be exhibited almost anywhere.

4 Pop-art color transformations

Figure 12: Pop-art portraits automatically generated by the installation "15 seconds of fame".

As mentioned in the introduction, Andy Warhol took photographs of people and performed some kind of color surgery on them. In this process he sometimes segmented the face from the background, highlighted the mouth or the eyes, delineated the contours, started the process with the negative instead the positive photography, overlayed the photo with geometric color screens etc. [7]. His techniques of transforming a photography into a painting could be described with a set of formal construction rules, like the ones used in shape grammars [12, 14]. The application of such rules for generation of new pop-art portraits which our installation tries to perform would require automatic segmentation of input images into its constituent perceptual parts: face/background, eyes, mouth, hair etc. These tasks are still too complex to be routinely solved in a few seconds on a large variety of input images. We decided therefore to try to achieve somewhat similar effects with much simpler means. Our system does not search for any features in the image but just processes the input image with selected filters.

We selected filters that achieve effects similar to segmentation. They drastically reduce the number of colors by joining similar looking pixels into uniform looking regions. The system randomly selects one out of 34 predetermined filter combinations and applies it to the portrait. 17 filters consist of different combinations of three well known filters: posterize, color balance and hue-saturation balance. The other half of filters are actually based on the first 17 filters, but with an important additional property, which we call random coloring.

Random coloring works in the following way: the system selects a pixel from the already filtered image and then finds all other pixels in the image of the same color. Then the system randomly selects a new color from the *RGB* color space for these pixels. In this way we achieve millions of different coloring effects and our pop-art portraits almost never look completely the same. Eight different randomly obtained filtering effects on the same input portrait can be seen in Fig. 4. Further six portraits of different people are collected in Fig. 12.

5 Results

5.1 Display of portraits

The installation displays the selected portrait in 15 second intervals on the framed monitor. For the display of the final result the system also selects randomly among five possible configurations: in 75% of cases it shows just a single processed portrait, in 25% of cases it subdivides the square display area into four smaller squares each showing a smaller version of the portrait. Each of the four smaller portraits can be processed with a different filter and the right-hand portrait column can be mirrored along the vertical axis. This way of stacking together multiple images also resembles Andy Warhol's way of displaying and amassing of images.

Since gallery visitors often stay in front of this installation for a longer period of time, we integrated a rule that prevents the selection of the same face, or a face at the nearly same location, in two subsequent 15 second intervals.

In the lower left corner of the LCD display is a counter counting the seconds from 15 to 0, reminding the currently "famous" visitor that his fame is fading away.

5.2 E-mail ordering of portraits

When the installation was exhibited publicly for the first time almost everybody who was featured by the installation wanted to get a copy of his or her pop-art portrait. Since we were already storing all images in a database for

Figure 13: Temporary web page showing the e-mail ordered portrait.

later experiments in face detection, we decided to make the finished portraits available to the public.

Therefore, we display now along each portrait a unique ID number in the lower right corner. A visitor who would like to get his portrait must note this ID number and send it in the subject field of an e-mail message addressed to `15sec@lrv.fri.uni-lj.si` up to one month after the end of the exhibition. The system periodically checks the e-mail messages, finds the requested portraits in the database and e-mails them as an attachment back to the sender of the request. In addition, a temporary web page valid for three days is generated, showing the requested portrait (Fig. 13). On this page one can randomly change the pop-art effects and save the new versions. If during an exhibition the installation is not connected to the Internet online, the portraits are sent later. From all requested portraits a special web gallery of famous people is automatically built for each public exhibition on the project's web page [1].

6 Conclusions

The use of computer vision methods in art projects is stimulating and somewhat specific. The concept of the installation "15 seconds of fame" requires from the vision system to find at least one face in the image which is then in turn transformed into a pop-art portrait. Therefore a high percentage of true positive face detections is very important, so that the installation does not display too often other body parts or objects. False negative face detections are on the other hand not a problem if at least one face out of many is found in each input image, since only one face is selected randomly anyway.

Making the installation robust and easy to install and administer was also a challenge. In this paper we described in detail how the problem of non-standard illumination was solved so that the installation can be exhibited under a large variety of illuminations. Since the installation will hopefully be exhibited at several art festivals in the near future, it was also important to make the whole system easy to administer, so that the presence of the authors is not required during the whole exhibition period.

As a side product of the installation a huge database for experimentation with face detection has been assembled. This database which shows mainly groups of people (i.e. Figs. 3a, 8 and 9) is available on the Internet along with our older CVL face database[2] [3].

The installation "15 seconds of fame" was well received by the public (Fig. 14). Even people without any prior information on how the installation works quickly realized that the installation displays portraits of people who are present at the moment. But getting a share of that "fame" proved to be more elusive. People who would step right in front of the installation, trying somehow to force the system to select them for the next 15 second period of fame, would be moreoften disappointed seeing that somebody else way back or on the side was selected instead. The only strategy that worked was to stay around the installation long enough.

There are also several possible improvements on the conceptual side of this art installation. We could try, for example, to determine the facial expression of each selected face to classify it as happy, angry or sad [6] and choose for the pop-art portrait a range of colors that matches that mood. A completely new version of the installation entitled "secret thoughts", which is under development, will paste the texture of the selected face on an animated 3D face model and make the face speak out controversial messages.

[2]The older CVL face database consists of: one face per image, 7 images of the same person in different position and with different face expressions, 114 persons, all captured under studio illumination.

Figure 14: Children having fun in front of the installation.

References

[1] 15 seconds of fame: `http://black.fri.uni-lj.si/15sec`

[2] Art Net Lab: `http://black.fri.uni-lj.si`

[3] CVL face database: `http://www.lrv.fri.uni-lj.si/database.html`

[4] S. Dragan, F. Solina. New information technologies in fine arts. In F. Solina and S. Dragan, editors, *Proc. Conf. New Information Technologies in Fine Arts, Fourth Int. Multi-Conf. Information Society – IS2001*, pp. 441–444, Ljubljana, 2001.

[5] T. Druckrey, Ed. *Ars Electronica: Facing the Future.* MIT Press, Cambridge MA, 2002.

[6] I. Essa, A. Pentland. Coding, analysis, interpretation and recognition of facial expressions. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(7):757–763, 1997.

[7] F. Feldman (Editor), J. Schellman, C. Defendi, J. Schellman (Contributor), A. Warhol. *Andy Warhol Prints: A Catalogue Raisonne 1962–1987.* Distributed Art Publishers, 3rd Rev edition, October 1997.

[8] G. D. Finlayson. Color Constancy in Diagonal Cromathicity Space, *Proceedings of the 5th International Conference on Computer Vision (ICCV'95)*, pp. 218–223, 1995.

[9] G. D. Finlayson, S. D. Hordley, P. M. Hubel. Color by Correlation: A Simple, Unifying Framework for Color Constancy, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1209–1221, 2001.

[10] A. Ford, A. Roberts. Colour Space Conversions, 1998, texttthttp://www.inforamp.net/˜poynton/PDFs/coloureq.pdf

[11] B. Funt, K. Bernard, L. Martin. Is Machine Colour Constancy Good Enough, *Proceedings of the 5th European Conference on Computer Vision (ECCV'98)*, pp. 445–459, 1998.

[12] J. Gros, F. Solina. Describing artworks using shape grammars. *Electrotechnical Review*, 59(5):314–320, 1992.

[13] E. Hjelmas, B. K. Low. Face detection: A survey. *Computer Vision and Image Understanding*, 83:236–274, 2001.

[14] J. L. Kirsch, R. A. Kirsch. A rule system for analysis in the visual arts, *Leonardo*, 21(4):445–452, 1988.

[15] M2VTS Face Database:
`http://www.tele.ucl.ac.be/M2VTS/`

[16] P. Peer, F. Solina. An Automatic Human Face Detection Method, *Proceedings of Computer Vision Winter Workshop*, Ed. N. Brändle, pp. 122–130, Rastenfeld, Austria, 1999.

[17] PICS Image Database:
`http://pics.psych.stir.ac.uk/`

[18] F. Solina, Internet based art installations. *Informatica*, 24(4):459–466, 2000.

[19] J. C. Terrillon, M. N. Shirazi, H. Fukamachi, S. Akamatsu. Comparative Performance of Different Skin Chrominance Models and Chrominance Spaces for the Automatic Detection of Human Faces in Color Images, *Proceedings of 4th IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 54-61, 2000.

[20] S. Wilson, *Information Arts. Intersections of Art, Science, and Technology.* MIT Press, Cambridge MA, 2001.

[21] Joint ISO/CIE Standard ISO 10526:1999/CIE S005/E-1998 CIE Standard Illuminants for Colorimetry, 1999, `http://www.cie.co.at/cie/`