Seventh International Conference on Control, Automation, [
Robotics And Vision (ICARCV'02), Dec 2002, Singapore

15 seconds of fame — an interactive, computer-vision based art installation

Franc Solina **Peter Peer** **Borut Batagelj** **Samo Juvan**

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

Abstract

"15 seconds of fame[1]" is an interactive art installation which elevates the face of a randomly selected gallery visitor 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 installation consists of a computer with a flat-panel monitor, a digital camera and proprietary software that can detect human faces in images and graphically transform them.

In this paper we present the technical background of the installation, in particular, how computer vision techniques were applied in this art installation.

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 [3, 12].

The Computer vision laboratory at 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 [1, 2, 11]. The installation described in this paper was envisioned by Franc Solina in 1996 when more reliable computer vision methods capable of detecting human faces in images started to appear and implemented in 2002 with the help of his graduate students.

The idea for 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 painting 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 politics and the arts (i.e. Mao-Tse Toung, Marilyn Monroe, Jackie Kennedy, etc.). Some of these images are true icons of the 20th century.

The installation "15 seconds of fame" intends to make instant celebrities out of common people by putting their portraits on the museum wall. Instead of providing 15 minutes of fame as prophesied by Warhol we decided to shorten this time to 15 seconds to make the installation more dynamic. This time frame 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, the installation tries to convey that fame tends to be not only short-lived, but also random.

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

The visible part of the "15 seconds of fame" in-

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

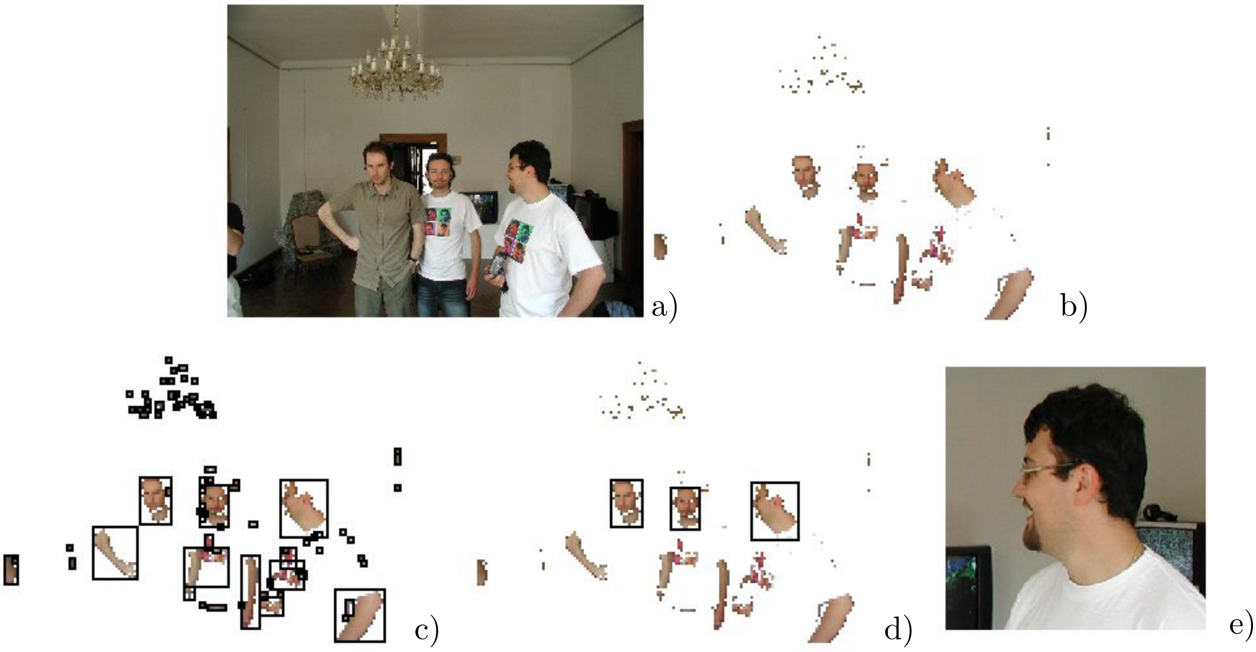

Figure 3: Steps in finding faces: a) downsize the resolution of the original image, b) eliminate all pixels that can not represent a face, c) segment all regions containing face-like pixels, d) eliminate regions which can not represent a face, e) select one of the faces and crop it from the original image.

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

stallation consists of a flat-panel computer monitor which is framed like a precious picture and hangs on the wall. The digital camera is hidden behind the frame and placed 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. 3 a). 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. 3 b, 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. 3 e). This processing performs the same function as a photographer who would take 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. The resulting portrait is then displayed for 15 seconds on the monitor. 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 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. Examples of such transformed portraits can be seen in Figure 4.

The outline of the rest of the paper is as follows: in Section 2 we explain the procedure for detecting human faces in images, graphical transformations to achieve pop-art effects are described in Section 3, Section 4 describes how the resulting portraits are displayed and, finally, Section 5 concludes the article.

Figure 4: Pop-art portraits of authors of this article produced by the described art installation (F. Solina, P. Peer, B. Batagelj and S. Juvan from top left in clock-wise direction).

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 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 [6]):

- *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. Though this approach can be easily extended to multiple views, it is unable to work well under different image conditions because the image structure of the facial features vary too much to be robustly detected by the feature detectors.

- In *texture-based approach* faces are detected by examining the spatial distribution of the gray-level information in the subimage (using Space Gray Level Dependency (SGLD) matrices). This is again not easily extensible to multiple viewpoints.

- *Neural network approach* detects faces by subsampling different regions of the image to a standard-sized subimage and then passing it through a neural network filter. In general, the algorithm performs very well for frontal-parallel faces, but performance deteriorates when extended to different views of the face. It is still not possible to extend the algorithm to detect faces in profile views.

- *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. In the first step pairs of stereo images containing frontal views are sampled from the input video sequence. Then 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.

2.1 Our original face detection method

We developed a face detection method which consists of two distinct parts: making face hypotheses and verification of these face hypotheses [9]. This face detection method tries to join two theories: it is based on detection of shape features, i.e. eye pairs (bottom-up feature-based approach), and skin color (color-based approach). The method assumes certain constraints and is therefore not applicable universally. Given the constraints, it is effective though for applications where fast execution is required.

The two basic limitations of the method that originate from the constraints are:

- the input image must have a high enough resolution (the face must be big enough) and

- it is sensitive to the skin complexion (fair skin is assumed).

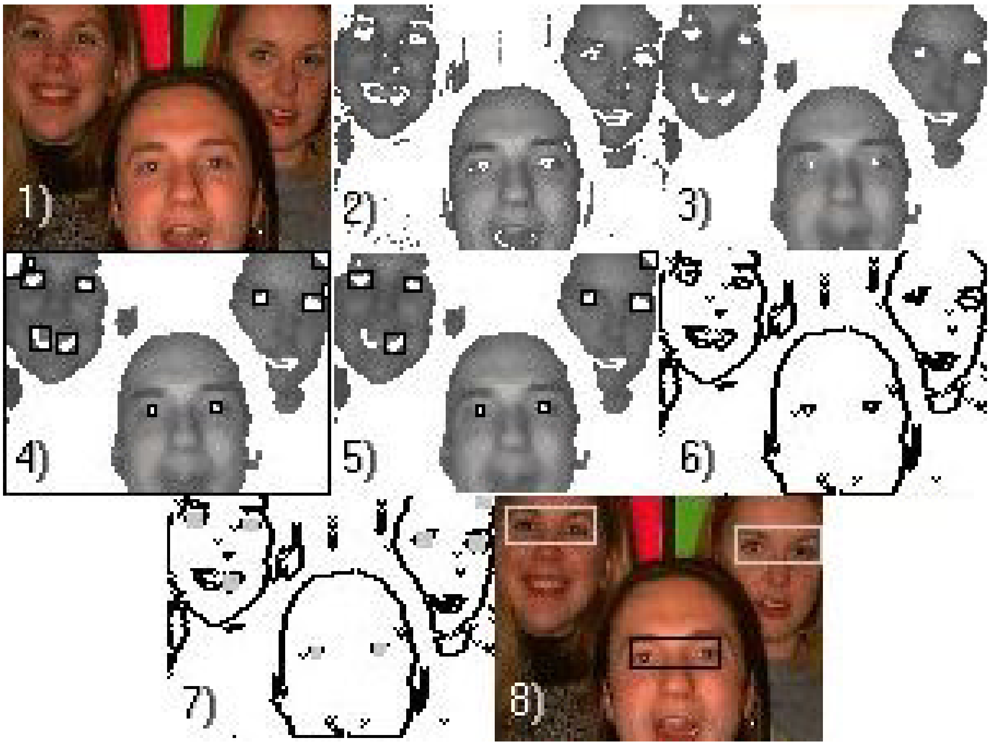

Figure 5: Basic steps of our face detection method described in [9]: 1) input image, 2) eliminated all non-skin colors, 3) image filtered with median filter, 4) segmented white regions, 5) eliminated insignificant white regions, 6) traced edges, 7) best possible circles within regions of interest, 8) output image. An additional feature of the method is also evident from the figure: because of greater distance of the faces in the background, the algorithm ascribe to the latter a lower probability based on the fact, that they have lower resolution and worse illumination then the face in the foreground.

The basic idea of the algorithm 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 complexion information.

The method was developed over a training set of different pictures. The goal of the method was to reach maximum classification accuracy over the images and meet the following demands and constraints:

- real-time operation on a standard personal computer,
- plain background,
- uniform ambient illumination,
- faces of fair-complexion which must be present in the image in their entirety (frontal position), and
- faces turned away from the frontal view for at most 30 degrees.

The basic principle of the method described in [9] is illustrated in Figure 5.

The method requires some thresholds, which play a crucial role for proper processing. They are set quite loosely (tolerantly), but they become effective in a sequence. All thresholds were defined experimentally using the training set. The method was tested over an independent testing set of two public image-databases (M2VTS [8], PICS [10]) with good results [9].

2.2 The simplified face detection method

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.

In order to evaluate a given face candidate as a face or as a non-face 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 down to 160×120 pixels. We implemented this resolution reduction in a pyramid manner. The system searches for face candidates in a lower resolution to speed up the process, but the selected face is later cropped from the original resolution for final processing.

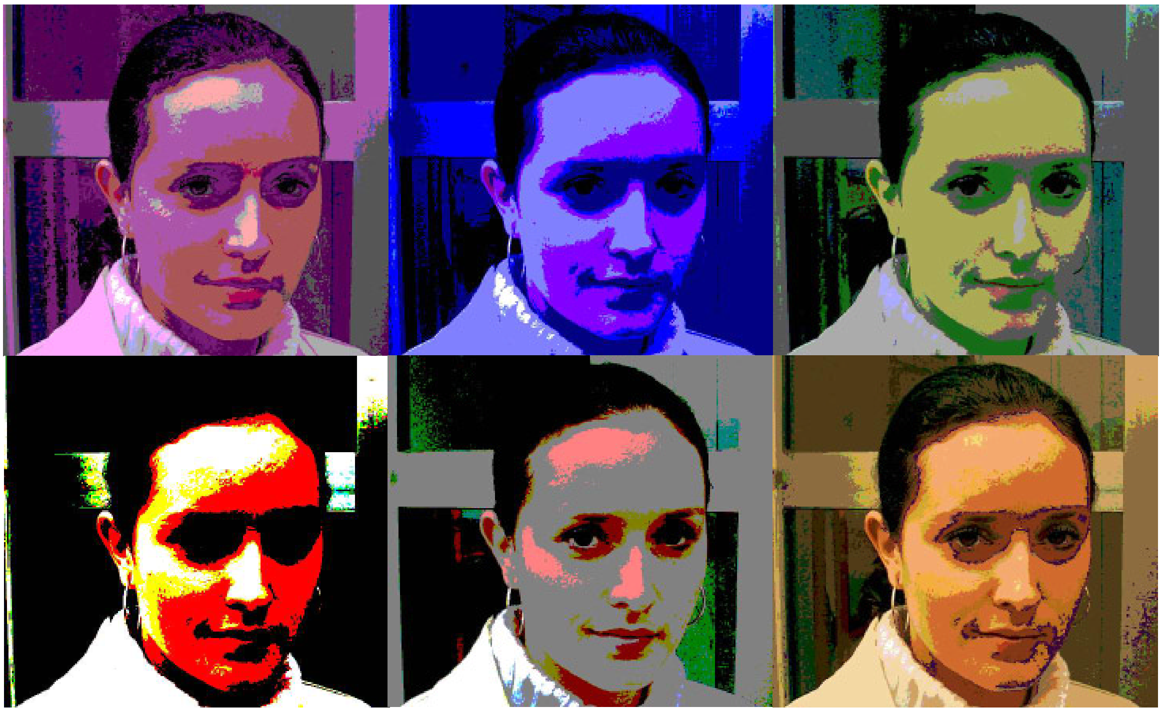

Figure 6: Some graphical transformations of a selected face in pop-art manner.

The system then eliminates from the picture all pixels that do not correspond to skin color. After the color segmentation is done, 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. The 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 pass some other heuristic tests (such as the percentage of skin pixels within the box around the candidate face region).

Initially, we had problems with bare arms which the system recognized as faces. But since we are checking the percentage of skin-like pixels within the region, the results were much better, although still not perfect. However, this is not too annoying—now and then it still happens that someone's arm becomes famous for 15 seconds.

The final step of finding a face candidate is selecting one of the candidates (randomly, but with help of some heuristics, like giving bigger priority to bigger regions and regions that are higher in the picture), mapping the coordinates of the region in the picture with lower resolution to the picture with the original resolution and cropping the face candidate out of the original picture. Fig. 3 illustrates the described process.

3 Pop-art color transformations

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 [13]. His techniques of transforming a photography into a painting could be easily described with a set of formal rules, like the ones used in shape grammars [7, 5].

In this project we try to achieve somewhat similar effects with much simpler means in a purely automatic way. The system does not search for any features in the image, just processes the input image with selected filters.

The selected portrait is cropped from the original digital image as a square area and resized to the resolution of 400×400 pixels. The system then randomly selects one out of 34 predetermined filters and applies it to the portrait. 17 filters consist of different combinations of three well known filters: posterize, color balance and hue-saturation balance. One of the properties of these filters is that they drastically reduce the number of different colors in the image. 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: first

number of images*	actual number of faces	detected face regions	false positives	false negatives
806	1972	1458	272	768

Table 1: Face detection statistics for the first public showing of the installation (*only images containing faces were taken into account).

the system selects one pixel from the already filtered image and remembers the pixel's value. Next, the system randomly selects one color from the RGB color space and, finally, the system replaces all pixels within the image that have the same value as the selected pixel with the new selected color. In this way we achieve millions of different filtering effects, so that the filtered portraits almost never look completely alike.

Six different randomly obtained filtering effects on the same input portrait can be seen in Fig. 6.

4 Display of results

The installation displays the selected portrait in 15 second intervals. 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 or 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 flat-panel display is a counter counting the seconds from 15 to 0, reminding the currently "famous" visitor that his fame is fading away.

At the end let us summarize the sequence of processing steps in the "15 seconds of fame" installation:

1. A color picture of 2048×1536 pixels is captured with the digital camera and downloaded to the computer.
2. The resolution of the original image is downsized to 160×120 pixels.
3. Based on color, all pixels that cannot represent a face are eliminated and replaced with white color.
4. The color image is converted into grayscale image.
5. Using a region growth algorithm the face-like pixels are segmented into candidate face regions.
6. Candidate face regions that do not meet predetermined criteria are eliminated.
7. One among the remaining face regions is randomly selected, cropped from the original, and resized to the resolution of 400×400 pixels.
8. A pop-art filter is randomly selected and applied to the resized cropped image.
9. The resulting pop-art portrait is shown on the framed LCD display for 15 seconds.

5 Conclusions

The first public showing of the "15 seconds of fame" installation was in Maribor, at the 8th International festival of computer arts, 28 May–1 June 2002 (Figs. 7, 8). The installation was very well accepted by the audience, but nevertheless we have plans for some improvements.

The installation was active for five consecutive days and the face detection method turned out to be quite stable despite the ever changing scene and illumination (Table 1). On the technical side we plan to make the face detection algorithm more robust but still fast enough. False negatives were due mostly to the small size of the face regions, because heads were turned away too much from the frontal position, and when faces were placed too close together. To prevent false positive face detection we should devise additional constraints (for example, check the compactness of skin-like regions or use the distribution of edge points within the face candidate regions). The most pressing problem is to make the face detection less dependent on the type of illumination (natural light, various types of artificial light). Based on these results we believe that our face detection method is reliable and effective for applications which require fast execution.

There are also possible improvements on the conceptual side of this art installation. We could try, for example, to determine the facial expression of each selected portrait to classify it as happy, angry or sad [4] and influence the color space for the graphic effects according to the mood of the person.

We are also considering if the digital copies of the pop-art portraits should be made available to the public. A strict interpretation the "15 seconds of

Figure 7: The installation "15 seconds of fame" in the gallery "Stari Rotovž" in Maribor, Slovenia, during the 8th International festival of computer arts, 28 May–1 June 2002

Figure 8: A gallery visitor and her 15-seconds pop-art portrait.

fame" installation philosophy requires that the portraits are not saved. To make the portraits publicly available we can store all images in a database and display along the currently famous portrait on the computer monitor also its identification code. By e-mailing this code in the subject field to our server, the server could return to the sender the requested picture. By printing, framing and displaying the portrait, people could stay famous just a little bit longer.

References

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

[2] 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.

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

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

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

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

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

[8] M2VTS Face Database: *http://www.tele.ucl.ac.be/M2VTS/*.

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

[10] PICS Image Database: *http://pics.psych.stir.ac.uk/*.

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

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

[13] 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.