Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Directed Evolution of Human Facial Images

Barbora Zahradnikova * **Roman Budjac** * **Peter Schreiber** * **Stefan Rydzi** *

* *Institute of Applied Informatics, Automation and Mechatronics, Faculty of Materials Science and Technology, Slovak University of Technology, J. Bottu 25, 91701 Trnava, Slovakia (e-mail: barbora.zahradnikova@ stuba.sk).*

Abstract In a criminal investigation, situations are quite often where no evidence but a witness is available. In these cases, facial composite is a tool applied in search for a perpetrator of a crime. Facial composites are images of faces drafted by a forensic technician requiring a precise description provided by a witness. Despites deploying the computational technique into the process, the naming rates remain very low (not reaching 5%). In this paper, we present a system developed according to the latest research on facial perception. Incorporating the interactive evolutionary algorithm, whole face images are automatically generated based on the witness's selection. The system is presented in the latest tested version, providing information on the selected EA as well as several image manipulation algorithms inclusive of method for age progression and hair manipulation. Besides describing applied methods and in-office testing, testing in simulated field conditions is part of the paper.

Keywords: facial composite system, evolutionary algorithm, face perception, forensic system

1. INTRODUCTION

It is an undeniable fact that being a police artist requires not only good artistic qualities but also patience, psychological skills and nowadays also deep computational competence. In contrast to other criminology domains where the deployed software solutions achieve very good results, the identification rate when using facial composites is very low.

Facial composites are images of faces used in search for a perpetrator of crime. Originally, they were manually drawn by a police artist according to a witness's description. Mechanical systems were first innovations in process. They were based either on foils laid one upon another in order to create a facial image or on cutting photographs of various faces into strips so that their features could be rearranged to form new faces. The success rates of these systems however were rather low.

Globalization and society informatization penetrated also into the field of composite construction. Since the 1970s, several composite computer systems were developed. The major graphics related problems of the mechanical systems seemed to be overcome by providing users with a large offer of features and additional addons enabling manipulating the image. However, when these systems were applied on constructing a composite from memory, very low naming rates were achieved. The testing results for the feature-based systems are summed up in Table 1.

In order to improve the recall capabilities of witnesses, the psychologists needed to focus on the human face perception. According to the latest research, humans perceive other faces as wholes. The free description provided by a witness is usually followed by a complex questioning in order to provide more details to the forensic technician. Forcing the concentration to be oriented on the individual facial features often leads to memory overshadowing, i.e. replacing the fade original memory with inaccurate details seen during the composition process resulting into misinformation effect (Pitchford et al. (2017)). Based on the conducted studies, the psychologists deduced that the process of decomposing a face into a set of features is

Table 1. Traditional Systems: Comparison

System	Method	Retention interval	Naming rate	Source
Photo-fit	Component based mechanical system for photofits	3-4 hrs	6.2%	Frowd et al. (2005b)
Identi-Kit 2000	Component based software for sketches	2 days	2.9%	Frowd et al. (2007)
E-FIT	Component based software for photofits	<4 hrs	19.0%	Frowd et al. (2005b)
		2 days	<2.0%	Frowd et al. (2005a)
FACES	Component based software for sketches	2 days	3.2%	Frowd et al. (2005a)
PROfit	Component based software for photofits	2 days	5.0%	Frowd et al. (2007, 2011)

not compatible with the way humans perceive faces and is responsible for the low identification rates of identikits.

The proposed approach enables automatic generation of facial images and their presenting to a witness, resembling the process of line-up identification. The user is only supposed to choose a face most resembling the perpetrator. The program is based on a genetic algorithm reaching a satisfying likeness within a few steps.

2. PRELIMINARY RESEARCH

The currently presented research stems from and further elaborates the hypothesis formulated and partly implemented into a system by Kovac et al.(2013). Though the partial success of the previous work, it was still unclear whether the proposed hypothesis could be successful in full implementation. Based on the weaknesses of other currently used systems mentioned above, we defined three main improvements which were to be confirmed using a simple program meeting the minimal requirements:

- The program shall be user friendly
- Selected optimization strategy (EA) may lead to at least one usable solution
- The time consumption shall be minimised

The first version of the program did not have a remarkable graphics, however, it served its purpose as it provided us with several important findings. The initial system validation confirmed the hypothesis that the composite processing time can be decreased and convergence towards the target can be ensured using EA. The results implied that the full functionality could be ensured only with the system based on a suitable database. The size of the database, file type of images and most important of all, the order of individual elements were identified as the primary database requirements. According to expectations, testing the system revealed that the ease of use was considered as the biggest advancement, when compared to the traditional methods.

The preliminary system showed that holistic approach is a step towards the way humans perceive other faces. However, it also revealed several weak points. Above all, the generated composites did not achieve the expected likeness. Aditionally, the processing time was too long (cca 10 minutes per generation). And eventually, automated merging facial features into a whole face based on EA using the proposed strategy was evaluated as a difficult and time consuming task since manual ordering of the individual features had been required. Therefore, leaning on preliminary results, applying a suitable model of face and choosing an appropriate evolutionary strategy were defined as the main areas for the future research.

3. FACINATOR I

The system developed within the latest years is marked by both improved graphics and faster operating evolutionary algorithm. Not only the system is more user friendly but first of all the facial representation is much more realistic. This was possible due to incorporating the Active Shape Model which is a mathematical model of face based on the presumption that all human faces are statistically similar and therefore, from a database of faces, new faces may be derived.

3.1 Facial Model

The method is based on Principal Component Analysis, which is a technique utilised e.g. for face recognition or image compression. PCA is a statistical tool enabling reducing the dimensionality of data space to the smaller dimensionality of feature space by extracting the major axis of variation in a data set (Kim et al. (2002)).

Preprocessing data Prior to applying the PCA, all images in the data set are marked by control points representing the facial features. These landmarks often referred to as shape, stand for the external representation of a face. In this testing model, there were 46 control points selected mirroring the location of the basic facial features, i.e. facial contour, mouth, eyes, nose. Before the PCA can be applied, data representing the shape and the texture (centered image describing only greylevel intensity) need to be separated and centered to the mean.

Texture, a shapeless image is obtained by piecewise affine warping.

The n X– and Y– coordinates (mirroring the position of facial features) are marked and converted to a 1D vector $\mathbf{x} = (x_1, \dots, x_n)^T$ by concatenating the rows into a long vector. Based on the M vectors (corresponding to image shapes), the mean shape is obtained as:

$$\bar{x} = \frac{1}{M}\sum_{i=1}^{M} x_i \tag{1}$$

The original images (also represented as 1D vectors) are warped to the mean shape in order to obtain the textures (grayscale information) $\mathbf{g} = [g_1, g_2, \dots, g_n]$.

Shape Model Applying PCA, separately on shapes and on textures, eigenvectors and eigenvalues are extracted from the images and the images are represented in a form:

$$\mathbf{x} = \bar{\mathbf{x}} + \phi_s \mathbf{b}_s \tag{2}$$

$$\mathbf{g} = \bar{\mathbf{g}} + \phi_g \mathbf{b}_g \tag{3}$$

where ϕ represents orthogonal modes of variations (eigenvectors corresponding to the largest eigenvalues) and $\mathbf{b}$ codes the parameters of the shape/texture. Using this model, shape and texture information may be used as separate input for the evolution of new faces.

Applying the genetic algorithm, each input data set (texture and shape eigenvectors) is to be optimised independently. The newly obtained shape was merged with the novel texture utilising a method of 2-D warp as proposed by Kroon (2010).

Warping is a method used to distort an image according to new coordinates. Image warping usually stands for a geometric transformation defined by a set of control points in the source and destination image. For warping images to new landmarks, we used piecewise affine transformation. Affine transformation is a linear mapping method which preserves points, straight lines and planes. The transformation is applied locally, on the triangles defined by landmark points:

• Firstly, several triangular regions are defined for the original and mean shape according to the control points.
• Then, coordinates of the three vertices of each triangle in base shape and original shape are counted.
• Matrix A is calculated from the coordinates according to Durrleman (2010).
• The coordinates of individual pixels are calculated and rounded.
• Pixel coordinates are found in the warped image.

3.2 Genetic algorithm

The purpose of developing the system automatically generating facial composites is to find an image approximating the likeness of a sought person. When constructing a facial composite, a witness (a user) is supposed to evaluate the likeness of the composite. Interactive evolutionary algorithms provide a high level of usability in such cases. We selected and tested mainly three evolutionary strategies. The algorithms were evaluated with regard to their convergence ability as well as time consumption. Operator usability and likeness of the constructed composites were evaluated later.

However, in order to develop a suitable algorithm, several other factors needed to be taken into account:

(1) Human usability - Human abilities to observe, differentiate and recognise unfamiliar human faces need to be reflected. The most limiting factor of the interactive human evaluation is the exhaustion of a "witness" which is also necessary to be taken into consideration when proposing the population size; similarly to lineup identifications, the optimal number of seen faces is usually lower.
(2) Convergence towards a suitable solution - The proposed strategy must converge towards the desired solution and must be able to approach to a target resembling the searched face within a reasonable time period.
(3) Variability of population - Although the individual faces shall become more and more similar in each generation, the variability of the population shall be maintained until the optimal solution can be chosen by the user. As the population size is to be modest, the variability of the individuals needs to be ensured using the genetic operators. The mutation and crossover rate need to be set higher.

Based on this pre-testing, IDE (Interactive Differential Evolution) was chosen as the strategy best fullfilling the requirements of the system. More information on the tested algorithms and test procedures might be found in Zahradnikova et al. (2017).The selected strategy is based on simplified basic genetic algorithm, completely eliminating the operator of crossover. In each generation 12 individuals are composed. The zero generation is generated randomly, however the model is chosen based on the user input provided on the application start. In the next iterations, the indivudals represent muted copies of the best individual from the previous generation. With regard to ensuring the population variability, the mutation rate needed to be increased to 0.2. The termination condition for the algorithm is the user's decission on the final likeness. The schematic diagram is depicted in Figure 1

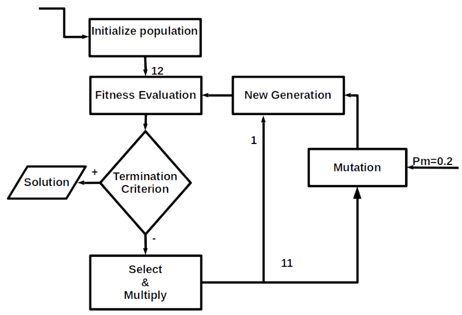

Figure 1. IDE:Interactive differential evolution

Based on the available research papers (Frowd et al. (2011)), the external parts of the face were not part of the face evolution. However, as several users had expressed an interest in adding a hairstyle, additional program was developed enabling basic external feature manipulation. Lateron, the hairstyle selection was added as a basic feature to the programm with the possibility to be turned off.

3.3 Hairstyle - Merging external and internal features

External features were gained based on the images from the YALE database of faces (Georghiades et al. (2001)) containing both male and female hairstyles of different length and shape.

At first sight, selection and application of the appropriate hairstyle seemed to be an easy task to implement. However, several issues arose during enabling the feature: Firstly the sizes of hairstyle images did not match the developed image dimensions. Secondly, the lightning conditions of various images were too different. Thirdly, the transitions were very visible with the contrast on the edges.

In order to obtain a working solution displaying a whole face image, several steps were necessary:

(1) The images with internal facial parts had to be resized to match the hairstyle images (260x360).
(2) The exposure and contrast of the images were matched based on their histograms (Figure 2). Balancing based on the maximal, minimal, average or median pixel intensity did not lead to satisfying results.

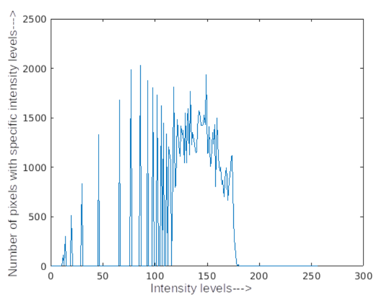

Figure 2. Example of histogram of a selected hairstyle image (the histogram is calculated for each hairstyle)

(3) The binary mask (Figure 3) was prepared in order to fit both images internal regions.
(4) The edges of the mask were blurred applying averaging and gaussian toi refine the transitions (Figure 3).

Figure 3. Example of a binary mask before and after blurring the edges

(5) The image of internal parts was multiplied by the mask, the hairstyle image was multiplied by the mask image complement in order to mutually fit (Figure 4).

Figure 4. Merging hairstyle and internals into a whole face

(6) The images were concatanated (Figure 4).

3.4 Age progression

Utilising the ASM for creating novel faces usually reaches satisfying results depicting almost authentic human faces. However, when faces of older people are to be generated, the structure of faces remains smooth and due to the averaging of the faces, only minimal wrinkles typical for older appearance are present. Therefore, in order to capture faces of all age categories, we applied an additional technique based on filtering.

(1) A high pass averaging filter is applied on a face I_{old} (mirroring the age to be achieved) and the filter F is substracted from the image obtaining a wrinkle map (see relation 4). The process is depicted in Figure 5

$$\mathbf{M} = \mathbf{I}_{old} - \mathbf{F} \tag{4}$$

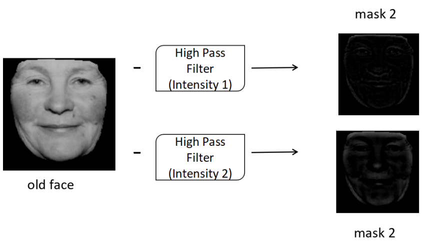

Figure 5. Process of obtaining a mask for manipulating the age of a subject

(2) In a similar manner, the upper layer of the younger image is filtered away with the aim of smoothening the base.
(3) The obtained mask is warped to fit the base image.
(4) The mask and the base face are concatenated resulting in an image with original features but enhance texture. Examples of the approach are depicted in Figure 6.

Figure 6. Images obtained applying the mask with two different structure intensities

4. TESTING

During the development of the system, its individual features were regularly verified in order to assure that the software fully meets the requirements.

4.1 Tuning the GA

As already mentioned three different genetic algorithms were proposed to be tested with the system. The parameter setting described in previous section, however, was carefully selected after running a series of tests with different parameter settings.

The time consumption and convergence rate were utilised as main criteria for tuning the parameters. As expected, the smaller the population was, the higher the mutation rate needed to be set.

4.2 Evaluation in accordance with Gold Standard Protocol

In order to enable objective evaluation of systems, Frowd et al. (2005b) proposed a “gold standard” protocol for testing. The protocol is designed to evaluate the system performance under simulated real conditions: Two groups of participants are included in two different tasks: constructing a facial composite and recognising an individual depicted in the composite. To begin with, the participants-witnesses are given 1 minute to look at the photograph of an unfamiliar target. After a certain time (retention interval), they should attempt to describe the target face. During this process, they are questioned by an interviewer (system operator). The cognitive interview is used to help a witness recall as many details about a face as possible. Afterwards, the cued recall - prompted description of each individual detail of the face - should be invoked by the interviewer. The second set of participants (familiar with the depicted person) are then asked to attempt to identify the composites.

Several studies (Frowd et al. (2005a,b); Carlson et al. (2012)) indicated quite promising results after testing the systems under laboratory conditions - constructing a composite with a target image present or from memory with up to 3-4 hours delay after seeing the target - reaching around a 20% naming rate. Evaluation of the systems following the “gold standard” protocol and including more realistic retention interval of one to two days, however, led to a drop down in naming rates. In different evaluations, the tested systems did not reach 5%(Frowd et al. (2005a, 2007)).

4.3 Preparation phase

Photographs of 6 people were prepared and distributed among 11 students. The software was prepared for both feature-based and holisitic approach. Additionally, system assessment sheets including questions concerning the ease of use, intuitiveness, time consumption and user friendliness of the system were drawn up. Finally, the ranking lists were composed to enable evaluation of the final facial composite images. After signing the conformity papers, the students received one photograph and they were asked to confirm they were not familiar with the depicted person.

Afterwards, they were given 1 minute to look at the photograph.

4.4 Composition phase

After two days, the same students were subject to a holistic cognitive interview. They were asked to think about the face in general and to try to recall the context the face was depicted at. They were recommended to focus on the character or personality of the person. After approximately 1 minute, the students were questioned for more detailed description concerning the overall appearance of the offender. They were asked to rate the person on the following characteristics: intelligence, masculinity, distinctiveness, friendliness, healthiness, attractivity, arrogance, aggressivity, dominance.

Followingly, they were asked to create a facial composite using either a feature based software or Facinator (the system screenshot is depicted in Figure 8). Facial composites were to be constructed either in photo-quality or as a sketch. During the whole process they were assisted by a person skilled with the used software tools.

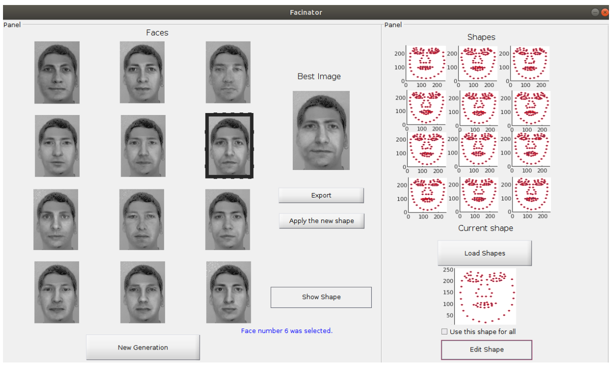

Figure 7. Facinator - system screenshot

After finishing the composition process, the students were given a questionnaire concerning the work with the particular system. The ease of use, intuitiveness, time consumption and user friendliness of the system were to be rated on the 1-5 scale (as in the school). Eventually, they were provided with the original photograph and the corresponding composite and asked to rank the likeness on 1-10 scale, with 10 being very good likeness.

Figure 8. Example of evolving a facial composite

4.5 Recognition phase

People familiar with the persons depicted in the target photographs, mostly the colleagues, were asked to name the person. At the beginning, they were just said they are familiar with the person. If the person was not named, or he/she was named incorrectly, they were given additional cues such as "It is a former colleague, age ca. 30 years".

After finishing the recognition phase, they were also asked to evaluate the composite likeness on the 1-10 scale, without having the original photograph, just by knowing the identity.

4.6 Results

36 facial composites were created using traditional and evolutionary software system within the "field testing". 11 of them were selected and used for the recognition part of the experiment. Using the traditional feature based software, whole face images depicting persons from the neck to the hair were created by the students. No facial composite was named correctly by the second group of people without being provided with a cue. Cued recognitions resulted in naming rate of 2.5%.

System	**Facinator (Photo)**		**Feature based**
Characteristics	internal parts only	whole face	
Number of evaluations	40	24	40
Named correctly (without a cue)	0	2	0
Naming rate (uncued)	0	8.33%	0
Named correctly (with a cue)	2	14	1
Naming rate (cued)	5.00%	58.33%	2.50%
Likeness (1-10 scale)	4.08	7.38	2.29

Table 2. Summarised results of the testing

Based on the research conducted by Frowd et al. (2011), facial composites evolved utilising Facinator did not depict the external features, i.e. only internal part of the face excluding hair, neck or ears was displayed. However, when they were distributed among people to be recognised, none of the 20 asked could identify anyone on the image without a cue. Having a cue, only 5% recognitions were successful. After the external features were added, the uncued recognitions reached 8.33%. Adding external features and providing the asked group with a cue on the sought person resulted in a naming rate reaching 58.33%.

Characteristics	**Facinator Photo**	**Facinator Sketch**	**Feature based**
Ease of Use	2.2	2.9	3
Understandability	2.7	2.9	2.7
Intuitiveness	2.5	2.9	2.9
Time Consumption	1.8	2.1	2.9
User Friendliness	2.8	2.9	3.6
Final Likeness of the composite (1–10 scale)	5.9	4.7	4.2

Table 3. Summarised results of the evaluation

The composite likeness was evaluated based on the memory of the people familiar with the depicted person. Facial composites constructed utilising a feature based system reached 2.3 out of 10 points. Although the recognition rate for the facial composites displaying only internal parts of the face was 0%, when being told the identity, the identifiers could see the similarity with the target person, evaluating the composite likeness in average with 4 points. Adding external features to the facial composites created utilisng Facinator produced an improvement in composite assesment, reaching 7.375 on the 1 - 10 evaluation scale. The complete results are depicted in Table 2.

According to the system use evaluation, the Facinator Photo was rated to be the easiest system to use and to understand. It was also marked as the most intuitive software. The subjective time perception (not measured) was evaluated as lowest among the utilised products. Moreover, it gained the grades 1 and 2 only from more than a half of the users in all the characteristics. Besides, the final composite likeness was assessed with 6 points out of 10, which was almost 2 points higher than the other evaluated systems. The results of the user evaluation are summarised in Table 3.

5. CONCLUSION

Based on the psychological and criminological review in the field, Facinator, a system automatically generating facial composites applying the evolutionary algorithm and utilising the Active Shape Model to model facial images, was developed as the main contribution of the presented research. Within this paper, a detailed description of the applied computational techniques, as well as a deep overview of the testing methods were delivered.

Besides standard verification processes, the system was tuned up based on the experimental procedures. Moreover, it was tested among the students according to the Gold Standard Protocol reaching naming rate of 58.33%. Additionally, it was evaluated with respect to the user friendliness and usability and the evolved facial composites were ranked both by the 'witnesses' and 'recognisers' achieving the best results among the tested systems.

The research presented within this paper focused on developing and testing an appropriate system able to generate human facial images. The major advantage of an evolutionary system is its efficiency in situations when, due to various factors, witnesses cannot provide a good description. Applying the appropriate holistic questioning and implementing an evolutionary strategy can be helpful even if a witness cannot properly recall the perpetrator of the crime. Although, it is clear that moving towards the psychological assumptions may open up new perspectives in the field, there are still situations which need to be further addressed.

So far, only several papers have been published studying the impact of a viewpoint in facial composite construction. In the area of viewpoint influence, further research needs to be conducted aiming at specifying the appropriate procedures for facial composite construction. Implementing a 3D model might be helpful in this area.

Automated recognition is a field requiring further investigation, as well. A software designed specifically to compare facial composites against a database of mugshots has recently been developed (Mittal et al., 2017; Klum et al., 2014), however, it has never been tested in combination with a holistically evolved facial composite. A solution based on these two state-of-the-art approaches is very promising and would be greatly useful in criminal investigation if accomplished.

ACKNOWLEDGEMENTS

This publication is a result of implementation of the project: " UNIVERSITY SCIENTIFIC PARK: CAMPUS MTF STU - CAMBO" (ITMS: 26220220179) supported by the Research & Development Operational Program funded by the EFRR.

This paper is a result of implementing the project KEGA 0009STU-4/2018: "The innovation of the subject Intelligent Control Methods at the Faculty of Materials Science and Technology of Slovak University of technology" supported by the KEGA.

REFERENCES

Carlson, C.A., Gronlund, S.D., Weatherford, D.R., and Carlson, M.A. (2012). Processing differences between feature-based facial composites and photos of real faces. *Applied Cognitive Psychology*, 26(4), 525–540.

Durrleman, S. (2010). Affine and non-linear image warping based on landmarks.

Frowd, C.D., Carson, D., Ness, H., McQuiston-Surrett, D., Richardson, J., Baldwin, H., and Hancock, P. (2005a). Contemporary composite techniques: The impact of a forensically-relevant target delay. *Legal and Criminological Psychology*, 10(1), 63–81.

Frowd, C.D., Carson, D., Ness, H., Richardson, J., Morrison, L., Mclanaghan, S., and Hancock, P. (2005b). A forensically valid comparison of facial composite systems. *Psychology, Crime & Law*, 11(1), 33–52.

Frowd, C.D., McQuiston-Surrett, D., Anandaciva, S., Ireland, C., and Hancock, P.J. (2007). An evaluation of us systems for facial composite production. *Ergonomics*, 50(12), 1987–1998.

Frowd, C.D., Pitchford, M., Bruce, V., Jackson, S., Hepton, G., Greenall, M., McIntyre, A.H., and Hancock, P.J. (2011). The psychology of face construction: giving evolution a helping hand. *Applied Cognitive Psychology*, 25(2), 195–203.

Georghiades, A.S., Belhumeur, P.N., and Kriegman, D.J. (2001). From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE transactions on pattern analysis and machine intelligence*, 23(6), 643–660.

Kim, K.I., Jung, K., and Kim, H.J. (2002). Face recognition using kernel principal component analysis. *Signal Processing Letters, IEEE*, 9(2), 40–42.

Klum, S.J., Han, H., Klare, B.F., and Jain, A.K. (2014). The facesketchid system: Matching facial composites to mugshots. *IEEE Transactions on Information Forensics and Security*, 9(12), 2248–2263.

Kováč, M. and Schreiber, P. (2013). Identikit creation with the use of genetic algorithms. *Applied Mechanics and Materials*, 248, 491–494.

Kroon, D. (2010). Opensurf (including image warp).

Mittal, P., Jain, A., Goswami, G., Vatsa, M., and Singh, R. (2017). Composite sketch recognition using saliency and attribute feedback. *Information Fusion*, 33, 86–99.

Pitchford, M., Green, D., and Frowd, C.D. (2017). The impact of misleading information on the identifiability of feature-based facial composites. In *2017 Seventh International Conference on Emerging Security Technologies (EST)*, 185–190.

Zahradnikova, B., Sutova, Z., and Schreiber, P. (2017). Interactive evolution of facial composites. *IFAC-PapersOnLine*, 50(1), 11776–11781.