FACE PHOTO RETRIEVAL BASED ON SKETCHES

G.A. Kukharev [1, 2], N.L. Shchegoleva [2]
[1] *West Pomeranian University of Technology, Szczecin, Poland,*
[2] *Saint Petersburg Electrotechnical University "LETI", St. Petersburg, Russia*

Abstract

The paper deals with the problem of the automatic retrieval of face photos using sketch drawings based on the witness description. We propose new methods for the generation of a sketch population from the initial one to improve the performance of sketch-based photo image retrieval systems. The method based on the computation of an average sketch from the generated population has been applied to increase the index of similarity in sketch-photo pairs. It is shown that such sketches are more similar to the original photographic images and their use leads to good results. Results of the experiments on CUHK Face Sketch and CUHK Face Sketch FERET databases and open access databases of photo-sketches pairs are discussed.

Keywords: Photo-Sketch Retrieval, Population of Sketches.

Citation: Kukharev GA, Shchegoleva NL. Computer Optics 2016; 40(5): 729-739. DOI: 10.18287/2412-6179-2016-40-5-729-739.

Introduction to sketch synthesis

Almost 20 years after the first papers on the subject [1, 2] were published, there is on-going interest in the community to automatic matching of the subjective portrait resulting from a sketch based on evidence from crime witnesses with a suspect's authentic photo. The input information is stored in the testimony of witnesses and their description of the suspect.

A subjective portrait can be either a drawing or a composite portrait. A drawn portrait is a line or half-tone drawing of the whole face made by an artist or a criminal expert following a description the witness has made of the suspect. A composite portrait is a face image including separate primitives (e.g. the eyebrows, the eyes, the nose, and the mouth) as well as some additional elements such as headgears, spectacles, earrings, bows, clips, etc. The primitives can be either drawings (prepared beforehand) or photographed parts of a face. In both cases this is a drawn-composed or a photo-composed portrait.

A French criminologist P. Chabot offered the method of creating a composite photo based on a verbal description and an identikit (or a "photorobot") in the middle of the 20th century. English research papers use the term sketch [5 – 9] rather than "photorobot". Sketches can take the following forms: a Viewed Sketch, drawn by an artist based on a photo or directly from her/his face; Forensic Sketch, drawn by a forensic artist following a description by an eyewitness. The term Viewed Sketch also means a computer drawing generated automatically basing on a digital photo image. On the other hand, a computer drawing (Viewed Sketch) re-worked by an artist is called an Artist Sketch. If a library of facial features is used for a sketch, the resulting sketch is called a "Composite Sketch". Moreover, if a "Composite Sketch" is based on a verbal description of a suspect's face, then it is known as "Composite Forensic Sketch".

At present, all sketches are generated by special computer programs. The most well-known among them are "IdentiKit", "PhotoFit", and E-FIT and "Mac-a-Mug", "FACES" and "IdentKit2000" [3, 4, 6 – 10]. The basic idea underlying these programs is a "mechanical collection" (collage) of a face area from individual features (primitives) from a library of primitives by an operator.

Although there are now ample libraries of primitives and a perfect technique of "gluing" primitives in computer programs, as well as some highly developed interface, the resulting sketch is highly dependent on the expertise of the specialist working with the program and the subjectivism of the verbal description from the witness. For example, in Fig. 1, the original face photo and Viewed Sketches are practically matching. The result depends on how a Viewed Sketch has been created. However, Composite Sketches (identikits) are less similar to the original photo even though they are made from that very photo, and they are not similar to each other either. This is due to different characteristics of the software used for identikit synthesis.

Original photo   Viewed Sketch   Composite Sketches
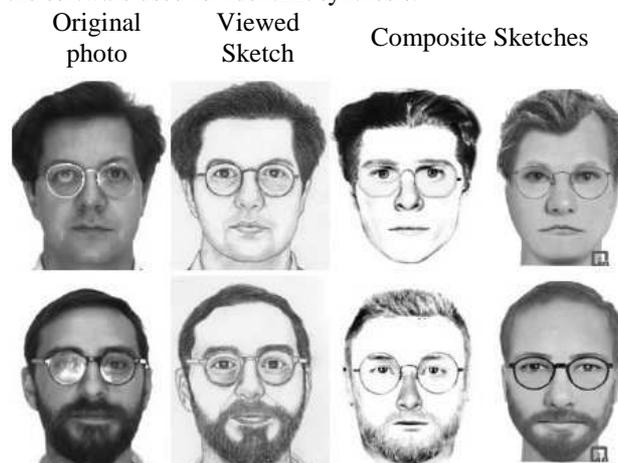
Fig. 1. Original photo, Viewed Sketches and two variants of Composite Sketches [7]

This "dissimilarity" is typical for any method of photo-collage composition. The dissimilarity increases if a photo-collage is not created from an original photo but only from a witness' verbal description. This situation is aggravated if the verbal description made by a witness is based on their recollection some days after the contact with the suspect (or the criminal), when the memory only partially retains the original primitives of the suspect's face.

This "similarity flaw" has fueled a growing interest to developing some improved techniques and systems for generating sketches and has resulted in creating some facial composite systems based on evolutionary algorithms (EA) and interactive strategies. In this case a sketch is not constructed from individual facial features, but instead, is selected from a database of sketches, taking into account the phenotype (facial features resulting from a verbal description) and treating it as a unique entity. All changes in the face are made by EA and corrected interactively by a witness. The simplest examples of a phenotype could be the head shape and/or individual facial features, as well as racial, gender and age face features presented in the form of a verbal description.

The first system using EA and interactive strategy was "E-FIT-V – Eigen FIT version V" system developed by Christopher Solomon [11]. Charlie Frowd was the author of the second system "Evo-FIT – Evolutionary Facial Imaging Technique for Creating Composites" [12]. Both systems use the representation of a face image based on the shape model (Active Shape Model – ASM) and the appearance model (Active Appearance Model – AAM). ASM determines the contour of the entire face area and the AAM of its texture. These model parameters use less than 50 features in the eigenspace based on PCA (Principal Component Analysis) and Karhunen–Loeve transformation. These model parameters are varied in the framework of evolutionary algorithms (EA) using the cloning procedure and random mutation of 50 original features.

After the first attempts of sketch synthesis the "Population 1" of several faces was generated by those systems. Population 1 corresponds to both the phenotype of a query face and a random variation of its parameters [13]. Population 1 represents different faces that have a similar hairstyle but noticeably different shapes of the head and basic facial features.

Then a witness interactively selects a result from Population 1 that best matches the original verbal description or some its individual features. From this result, accepted as the current sketch model, the system generates a new population (Population 2) by changing the parameters according to EA. Thus, Population 2 does represent the genotype of a face of the same person with small variations of primitives.

Finally, from the population 2 a face is selected, which matches best the original verbal description or individual features (not taken into consideration earlier). This process continues until a witness validates the result.

The approach adopted in "E-FIT-V" and "EvoFIT" systems implements the ideas of evolutionary algorithms and a human-computer interaction where the author of an original verbal description can correct the results of evolutionary algorithms. The final decision is also made by this author, resulting in a quickly designed and a more similar final sketch. That is why the developers called their ideas the strategy of evolutionary creation of photorealistic composite faces or sketches. A series of experiments has shown that sketches made using this strategy are close to their originals and easily identified by experts (psychologists and criminologists) [13, 14].

To summarize, the original photo and the synthesized sketch are made similar by the following factors of interactive evolution strategy:
– discarding the "assembly of sketches" mechanism using separate primitives;
– selection of a holistic face image as an initial sketch taking into account the phenotype data;
– using parametric models of a face image in the eigenspace of features;
– variation of model parameters of sketches based on EA;
– selection of the best solution based on the interaction with a witness.

This "dissimilarity" is typical for any method of photo-collage composition. The dissimilarity increases if a photo-collage is not created from an original photo but only from a witness' verbal description. This situation is aggravated if the verbal description made by a witness is based on their recollection some days after the contact with the suspect (or the criminal), when the memory only partially retains the original primitives of the suspect's face.

This "similarity flaw" has fuelled a growing interest to developing some improved techniques and systems for generating sketches and has resulted in creating some facial composite systems based on evolutionary algorithms (EA) and interactive strategies. In this case a sketch is not constructed from individual facial features, but instead, is selected from a database of sketches, taking into account the phenotype (facial features resulting from a verbal description) and treating it as a unique entity. All changes in the face are made by EA and corrected interactively by a witness. The simplest examples of a phenotype could be the head shape and/or individual facial features, as well as racial, gender and age face features presented in the form of a verbal description.

The first system using EA and interactive strategy was "E-FIT-V – Eigen FIT version V" system developed by Christopher Solomon [11]. Charlie Frowd was the author of the second system "Evo-FIT – Evolutionary Facial Imaging Technique for Creating Composites" [12]. Both systems use the representation of a face image based on the shape model (Active Shape Model – ASM) and the appearance model (Active Appearance Model – AAM). ASM determines the contour of the entire face area and the AAM of its texture. These model parameters use less than 50 features in the eigenspace based on PCA (Principal Component Analysis) and Karhunen–Loeve transformation. These model parameters are varied in the framework of evolutionary algorithms (EA) using the cloning procedure and random mutation of 50 original features.

After the first attempts of sketch synthesis the "Population 1" of several faces was generated by those systems. Population 1 corresponds to both the phenotype of a query face and a random variation of its parameters [13]. Population 1 represents different faces that have a similar hairstyle but noticeably different shapes of the head and basic facial features.

Then a witness interactively selects a result from Population 1 that best matches the original verbal de-

scription or some its individual features. From this result, accepted as the current sketch model, the system generates a new population (Population 2) by changing the parameters according to EA. Thus, Population 2 does represent the genotype of a face of the same person with small variations of primitives.

Finally, from the population 2 a face is selected, which matches best the original verbal description or individual features (not taken into consideration earlier). This process continues until a witness validates the result.

The approach adopted in "E-FIT-V" and "EvoFIT" systems implements the ideas of evolutionary algorithms and a human-computer interaction where the author of an original verbal description can correct the results of evolutionary algorithms. The final decision is also made by this author, resulting in a quickly designed and a more similar final sketch. That is why the developers called their ideas the strategy of evolutionary creation of photorealistic composite faces or sketches. A series of experiments has shown that sketches made using this strategy are close to their originals and easily identified by experts (psychologists and criminologists) [13, 14].

To summarize, the original photo and the synthesized sketch are made similar by the following factors of interactive evolution strategy:
− discarding the "assembly of sketches" mechanism using separate primitives;
− selection of a holistic face image as an initial sketch taking into account the phenotype data;
− using parametric models of a face image in the eigenspace of features;
− variation of model parameters of sketches based on EA;
− selection of the best solution based on the interaction with a witness.

2. Problems of photo-sketch matching

Interaction with a witness is crucial in "EFIT-V" and "EvoFIT" systems as it is the witness who makes the final decision about the similarity of a sketch with a subjective description. Sketches generated in this framework in accordance with the interactive evolution strategy do not simplify the problem of automatic matching sketches with original photo images.

The analysis presented [8, 9] shows that stable recognition of Composite Forensic Sketches and Composite Sketches with a corresponding photo from special criminalistics databases is currently not achievable in practice.

This happens due to the following three basic reasons:
1) The low quality of sketches results from verbal descriptions;
2) The methods used to match a pair photo-sketch have some drawbacks;
3) There is a lack of databases with photos and sketches required for this task.

These reasons induce the developing databases of sketches, expanding the existing benchmark face databases [15, 16], improving the methods of photo–sketch matching, and modeling the task of retrieving photos based on given sketches [6 – 9, 17 – 21].

As a result the first database of sketches has been created, including the most popular CUHK Face Sketch database (CUFS) and CUHK Face Sketch FERET database (CUFSF), containing photos and corresponding sketches AR data set and XM2VTS data set [13, 14]. Besides, in [17 – 21] new ideas concerning automatic sketch synthesis from a face photo we, as well as the methods of matching were proposed. In majority of cases these developments were made using the CUHK and CUFSF databases and results are contain sketches in the form of "Viewed Sketch".

Meanwhile the construction of sketches (identikits) is only the first step in solving a more general problem - the problem of automatic (without human participation) face photo retrieval based on given sketches. This problem arises, for example, when searching for the original face photo in a large database by a given sketch or finding correspondences between the faces of people in a surveillance system video, as well as in solving problems of mutual photo-sketch recognition.

In the presentation of results of mutual photo-sketch recognition on CUFS and CUFSF databases, both the methods of processing and parameters of training and test samples were not precisely specified. Therefore unambiguous determination of the model of conducted experiments is impossible. The assessment of obtained results is considerably difficult, as well as verification of the model of conducted experiments in the framework of meta-analysis. The state-of-art and problems mentioned above (approaches, solutions, results and their analysis) are presented in [22, 23].

3. What sketch databases do we need today?

The CUHK database contains sketches generated automatically from original photos and corrected by artists; there are 188 photo-sketch pairs. The CUFSF database contains sketches that are made by an artist from original photos from the FERET database. These sketches retain main facial features and singular facial attributes but have some artefacts (elements of caricature or exaggeration). In fact, both databases contain sketches defined above as "Artist Sketches".

Different methods of making sketches in CUFS and CUFSF databases result in the effect when sketches are recognized using simple methods [22, 23] with high performance (close to 100 %), while others are also recognized with the same simple methods but require precise matching of size and orientation of face areas in the plane XY. Unfortunately in [16] only cropped sketches of extremely poor quality (on resolution, size and texture) are accessible. So we are not able to conduct representative practical research on these databases. The examples of matching cropped faces from the FERERT database with corresponding sketches from the CUFSF database [16] are shown in Fig. 2.

Comparing cropped sketches from the CUFSF database [16] with sketches from Fig. 3 additional retouch of facial fragments and facial features (forehead, nose, mouth, etc.) can be seen.

*Fig. 2. Examples of matching cropped face images
and sketches [16]*

These operations were made "manually", i.e. not in an automatic mode. Besides, texture in face areas is "smoothed" by low-pass filtration. And finally, basic anthropometric parameters (eyes' line, inter-eye distance, etc.) were aligned. It is possible that in order to study the methods of photo-to-sketch matching, such alignment is necessary. But in real applications, for instance, in criminal events and scenario where the task of a suspect retrieval based on a given sketch is actual, the condition of alignment is practically unattainable because we do not know in advance how the original photo of a suspect looks like. Thus parameters of the face photo are unknown, and we do not know if it corresponds to the given verbal description (and to the generated sketch based on it).

In Fig. 3 there are photo-sketch pairs (Composite Forensic Sketches). The lack of high similarity between photos and sketches is obvious, in contrary to similarity observed, e.g., in corresponding pairs in Fig. 1 or in Fig. 2.

*Fig. 3. Photo and sketches based on verbal description
of witnesses. http://abclocal.go.com/story?section=news
/national_world&id=7044287*

All four pairs have a noticeable difference in size (height and breadth) of facial areas, in size and location of facial primitives, as well as distortions in symmetry of facial fragments, and different directions of a look. The matching of such sketches with original photos usually taken frontally and with standard normalization is rather difficult or impossible in practice.

This raises a new challenge: how to match sketches and original photos? There is an analogy with the strategies that are implemented in "EFIT-V" and "EvoFIT" systems. Every original sketch should be modified a few times and presented with new parameters concerning geometry of the face area (size, symmetry, and shift) in order to generate "new population" of such sketches. These modifications imitate the generation of $K > 1$ sketches corresponding to "group of K witnesses". Thus for such a generated population it is possible to solve the task of matching sketches with original images. For matching an average sketch of the population or each sketch from the population majority mechanisms or "mixtures of experts" are used. And this is the only case when it is possible to obtain a good matching result!

This is the fact that became the starting point in [20, 21] to generate a new set (population) of sketches for any given original sketch. The mechanism of generation of this specific sketch populations is rather simple that was demonstrated and confirmed by the results obtained for CUFS and CUFSF databases presented in [22, 23].

Similar results can be achieved using the approach for generating "population of caricatures" presented in [24]. However, solutions presented in [24] are based on exact shape (AMS) and appearance (AAM) models of face images, linear transformations between two images (based on PCA and KLT) and different variants of rearrangement and magnification of compact facial fragments or facial primitives.

Taking this into account the approach presented in [24] is worse than approaches in [22, 23] because of complexity of algorithms used for the generation of requested populations. The high accuracy of a face appearance provided by the approach [24] is not sufficient in the task of generating a sketch population as it is unknown if an original photo corresponds to a given sketch.

4. Algorithm for a sketch population generation

The proposed algorithm for generating a "population of sketches" is following. We form $K > 2$ new sketches from a given sketch by geometrical changes in facial areas, for instance, as it is determined by biometric standards. The matrix S of size $M \times N$ represents the original sketch image in a grayscale format. We assume that the facial area occupies no less than 80 % of the whole image. For every $k = 1, 2, ..., K$ we define three parameters p_1, p_2 and p_3 using a generator of random numbers and mapping this values into the interval $\pm d$ so that:

$$p_i = sign(R_n^{(i)}) fix(dR_u^{(i)}), \text{ for } I = 1, 2, 3, \qquad (1)$$

where p_i – parameter; d – maximal value of parameter, and $d \geq 2$; $R_n^{(i)}$ and $R_u^{(i)}$ – normally and uniformly distributed random numbers; $sign(R_n^{(i)})$ – sign of the number.

Parameters p_i generated according to formula (1) have the following meanings: p_1 corresponds to the length of a face in the original image and thus the location of eyes' line; p_2 relates to changes in the breadth of a face; p_3 relates to the changes in location of a symmetry line in the facial area. Parameter d can be related, for example, with the number of pixels that correspond to the changes (increase or decrease) in the central part of a face or location of an eyes' line or a face symmetry line.

Examples of generated sketches from input sketch are show in Fig. 4, where a sketch from the Population 1 has visible changes in width and height of the face area and a sketch from the Population 2 has visible changes in all parameters of the original model with additionally blurred edges of facial primitives and textures. Main parameters of a face model are as follows: W – face width; H – face height; S – distance to the line of symmetry from a contour of a face oval; h – distance to the eyes line; d – distance between centers of eyes.
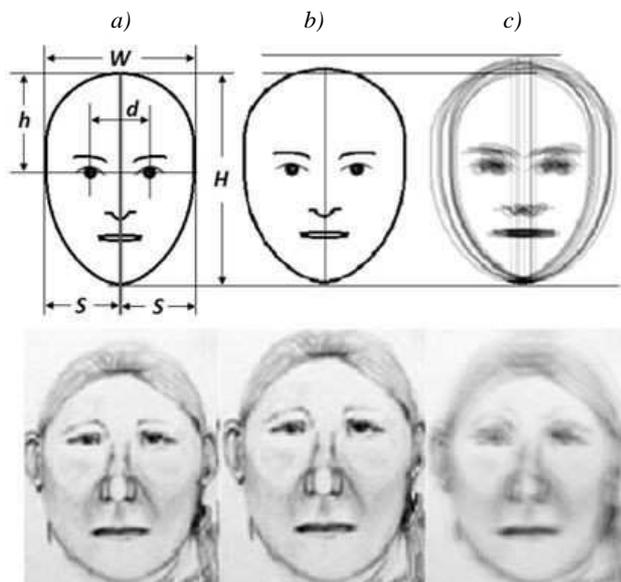
Fig. 4. An idea of sketch generation: an input sketch (a) and sketches from Populations 1 (b) and Populations 2 (c)

The algorithm of geometrical changes in facial area consists of three steps; where in each step one operation of changing in geometry is executed.

Step 1. If $p_1 > 0$, then remove first $(p_1 - 1)$ rows from matrix S. If $p_1 < 0$, then extend matrix S by adding first $(p_1 - 1)$ rows above S. This operation can be described as:

$$S^{(1)}(\text{var} \times N) = \begin{cases} S(p_1 : M, :), & \text{if } p_1 > 0 \\ [S(1 : abs(p_1)); \quad S], & \text{if } p_1 < 0 \end{cases}, \quad (2)$$

where $S^{(1)}(var \times N)$ is matrix with removed or added rows, and indicated by a parameter var.

Resulting matrix $S^{(1)}$ in (2) is longer if condition $p_1 > 0$ is satisfied, or is shorter if condition $p_1 < 0$ is satisfied. Next we rescale $S^{(1)}$ restoring its original size $M \times N$ so that:

$$S^{(1)}(\text{var} \times N) \rightarrow S^{(1)}(M \times N). \quad (3)$$

Here the length of a face in resulting matrix $S^{(1)}$ in (3) increases, if the condition $p_1 > 0$ is met or is shortened if condition $p_1 < 0$ is met. Consequently, eyes' and mouth lines are shifted up or down. The changes we are interested in a facial area in a sketch are within the range of values $\pm d$, in respect to the length of an original face.

Step 2. Next, if $p_2 > 0$, we remove first $(p_2 - 1)$ columns from matrix $S^{(1)}$. If $p_2 < 0$, then we remove last $(p_2 - 1)$ from matrix $S^{(1)}$. These operations can be written as follows:

$$S^{(2)}(M \times \text{var}) = \begin{cases} S^{(1)}(:, p_2 : N), & \text{if } p_2 > 0 \\ S^{(1)}(:, 1 : N - abs(p_2)), & \text{if } p_2 < 0 \end{cases} \quad (4)$$

Here the resulting matrix $S^{(2)}$ in (4) have $abs(p_2 - 1)$ columns less irrespective to the value p_2.

Next we perform rescaling of matrix $S^{(2)}$ restoring its original size of $M \times N$:

$$S^{(2)}(M \times \text{var}) \rightarrow S^{(2)}(M \times N), \quad (5)$$

that inevitably leads to increased breadth of a facial area in the original image and cyclic shift of a facial area to the left or right. If to neglect change in length of a facial area according to (3), the breadth increase according to (5) is determined by a value close to the value of d.

Step 3. In this step we carry out a cyclic shift of matrix $S^{(2)}$ to the left by $(p_3 - 1)$ columns, if $p_3 > 0$, or to the right, if $p_3 < 0$. These operations can be written as follows:

$$S^{(3)}(M \times N) =$$
$$= \begin{cases} [S^{(2)}(:, p_3 + 1 : N) \quad S^{(2)}(:, 1 : p_3)], & \text{if } p_2 > 0 \\ [S^{(2)}(:, 1 : N - abs(p_3) + 1 : N) \\ \quad S^{(2)}(:, 1 : N - abs(p_3))], & \text{if } p_2 < 0 \end{cases}, \quad (6)$$

resulting in cyclic shifts of the whole sketch image, violating the symmetry in respect to the line of central symmetry of a face.

Now we rewrite the result (6) in a new form:

$$S^{(k)} = S^{(3)}, \quad (7)$$

where the resulting matrix (7) represents a new k-th sketch of Populations 1 (fig. 4b).

Next we write down the result $S^{(k)}$ in the memory and go back to determine parameters p_1, p_2 and p_3, and further to steps $1 - 3$ until we form a new sketch $S^{(k+1)}$ and so on.

The following questions may arise: how is it possible to evaluate the similarity measure between an input image and corresponding sketch? How the visible similarity (visible to a human) should correspond to some formal index? And what is more important to evaluate: similarity visible by a human observer of an input image and its sketch, or some formal index of similarity? These questions are particularly important in the chains "witness→verbal face description→sketch" and "sketch→photo retrieved based on this sketch".

5. A method to increase a similarity index in photo-sketch pairs

In [22, 23] it is presented that if facial sketches are created directly from original face photos (Viewed Sketch or Artist Sketch), then they are recognized by simple systems with 100 % performance. But if sketches are created basing on verbal descriptions, the recognition performance decreases significantly. This situation takes place in a composite sketch based on a verbal description or in an automatically generated sketch based on inexact (or incomplete) input data. When we try to increase the subjective similarity of a generated sketch with the original one, we decrease the formal similarity measure (Structural SIMilarity index - SSIM) [26, 27]. SSIM index allows to evaluate the degree of similarity between two images depending on the following factors: changes in luminance and contrast; loss of correlation. The performance of

sketch recognition depends on the SSIM index: the higher the SSIM index, the higher the recognition result. Thus our next task is to process sketches in such a way that they retain subjective similarity significantly gaining in the SSIM index at the same time. To do this we propose to generate a new sketch Population 2 (fig. 4c): the input is coming consistently with K sketches $S^{(k)}$ represented by data matrices of size $M \times N$, at the output we have K new sketches $S^{(k)}$ obtained as:

$$\tilde{S}^{(k)} = \left(\sum_{j=1}^{k} S^{(j)} \right) / k \text{, for } k = 1, 2, ..., K, \quad (8)$$

where each matrix $\tilde{S}^{(k)}$ represents the resulting sketch being the average of previous k sketches.

6. Analysis of results

In this section we demonstrate that sketches from Population 2 have a similarity index with an original photo greater than the ones from Population 1. Fig. 5 demonstrates: a – a Viewed Sketch from [22]; b – an original photo from the CUHK Student Sketch Database; c – an Artist Sketch from the CUHK database. In the bottom left part values of the SSIM index between the original photo and the Viewed Sketch from Population 1 (lower curve $P1$) and Population 2 (upper curve $P2$) are shown. In the bottom right values of the SSIM index between the original photo and the Artist Sketch from Population 1 (lower curve $P1$) and Population 2 (upper curve $P2$) are shown. Curves show the values of the SSIM index for nine sketches from mentioned populations. The horizontal red lines mark threshold values 0.62 and 0.495 of the similarity index between the original photo and sketches a) and c).
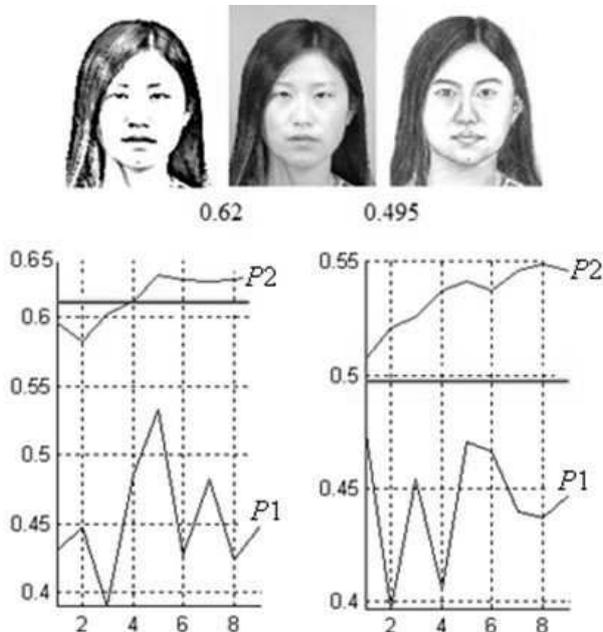
Fig. 5. Original data and corresponding values of the ISSIM

As it can be seen from the results presented the values of the SSIM index for sketches from Population 2 exceed thresholds in both cases. So we claim that sketches from Population 2 are more similar to original photos and consequently the quality of sketches has become higher [25]. Generated sketches are also useful in real-life scenarios

with inaccurate or incomplete information about original photos and their parameters.

The proposed method of increasing the quality of sketches is suitable to be applied to sketches used in crime detection practice. We employ the sketch from paper [8] which is interesting because it was recognized with the rank 72, i.e. corresponds to the original photo at the position 72 in the ranked list of retrieval results. Fig. 6 presents a sketch and a corresponding original photo also taken from [8] and corresponding values of the SSIM index. The first row gives: a sketch, a modification of a sketch generated in Population 1 (numbers above it are parameters p_1, p_2 and p_3); a sketch generated in Population 2 for the value $k = 10$. The second row gives: an original image, a modification of a photo generated in Population 1 (numbers above it are values of parameters p_1, p_2 and p_3); a photo generated in Population 2 for value $k = 10$.
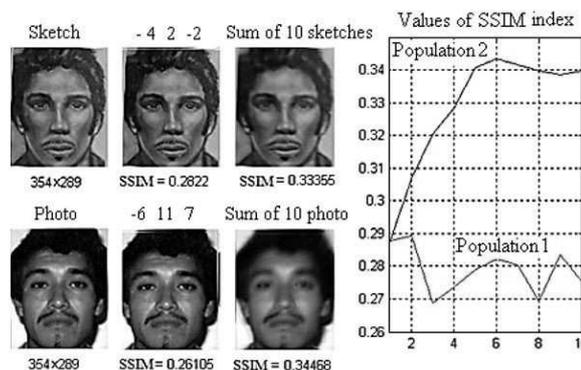
Fig. 6. Sketch, original photo and corresponding values of the SSIM index

Presented results show that the SSIM index for sketches generated in Population 2 is higher than the SSIM index for Population 1.

These examples (Fig. 6 and Fig. 7) demonstrate that the presented method of generation of sketches possess a universal trait since it can be used independently (for sketches and original photos) and for any available databases of photos.

Summing up the results of the experiment the following is necessary to note:

a) The modification of the original data (identikits/sketches) and their representation in the form of Population 1 emulate the acquisition of new data from a group of K witnesses. This effect can be seen as more objective representation of sketches for the original data available. Having these assumptions it is possible to solve the problem of comparing the new data with the original photo rather effectively even within Population 1;

b) Thus, the comparison may be performed with an average (for the whole population) sketch or with each sketch from a population on the basis of majoritarian mechanisms or, for example, on the basis of mixtures of experts [28];

c) The transformation of results of Population 1 to results of Population 2 improves the similarity in photo-sketch pairs. The marked effect in conjunction with the mechanisms mentioned in b) creates new conditions for more effective comparison of sketches with original photos;

Photo   Sketch   SSIM index

Shift = 8 -2 2   Sum of 10 Faces

SSIM = 0.25074   SSIM = 0.32794
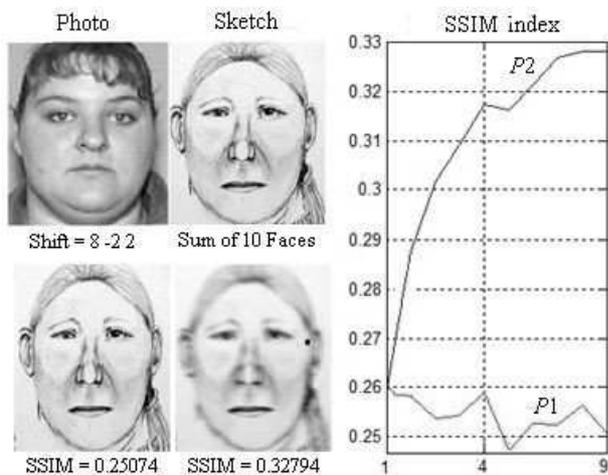
Fig. 7. Original photo, composite forensic sketch and corresponding values of the SSIM index

d) For the similarity assessment in original photo-sketch pairs it is possible to use the SSIM index as it estimates the correlation and texture of local areas in the original data [26, 27];

e) As shown in [8, 9], Multiscale Local Binary Patterns (MLBP) provides the most useful result for the practice. In face recognition applications MLBP represents a genotype as a set of its modifications i.e. some kind of a population! Therefore, the proposed approach corresponds to modern trends of face photo retrieval using sketches.

7. Experiments

The aim of experiments presented below is to confirm the hypothesis that the ISSIM of pairs original photo-sketch from Population 2 can be efficiently used as a search criterion in face photo retrieval based on sketches even with the use of simple methods. In this section we show the results of experiments with sketches from Population 2 that are generated from sketches taken from CUFS and CUFSF databases.

Experiments on the CUFS database

Based on $K = 100$ photo-sketch pairs from the CUFS database we generated new subsets of sketches of Populations 1 and 2 with $\sigma \leq 3$. Using both original and generated data we conducted three experiments with the same subset of original photos (used as references) but with different subsets of sketches: original sketches from the test part of the CUFS database; corresponding sketches from Population 1 and generated from them sketches of Population 2.

The structure of the FaRetSys for the first experiment is presented in Fig. 8*a*. The main parts of the system are CUHK/CUFS databases that are represented by a block of original photos (BOP) and a block of original sketches (BOS); a feature extraction unit (FEU) and a comparator (CMP).

The extended structure of the FaRetSys for the second and third experiments is presented in Fig. 8*b*. The following blocks are added: a generator of populations of sketches (SPG); a database that stores Populations 1 and 2 (P1 и P2) of sketches generated from the original sketch.

a) Simple FaRetSys *b*) Extended FaRetSys
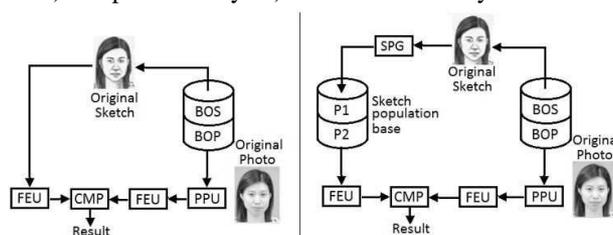
Fig. 8. Structures of the face photo retrieval system used in experiments

The aim of each experiment is to search an original photo $P(k)$, $1 < k \leq K$ in the CUFS database by a given sketch (Query Sketch $S(j)$ where $1 < j \leq J$) and evaluate this search accuracy numerically/qualitatively. The results of the experiment 1 serve as a "reference database" and are used further to compare with the results for sketches from Populations 1 and 2.

Parameters and models of retrieval systems in experiment 1

The dimensions of original photos and sketches are 250×200 pixels. The preprocessing (PP) step consists of transforming images to GRAY scale selecting the area of interest within a facial image (size of 200×180) and blurring the selected area using a "blur-filter" with a window of size 5×5 or 7×7.

A feature extraction procedure is based on the two-dimensional discrete cosine transform (2D DCT) of original images (photos and sketches). This procedure can be performed in two ways. The first way is to represent images (photos or sketches) using values starting from the upper-left corner of a spectral matrix of order d. The total number of spectral components in this case is $d \times d$. In the second way we also use components starting from the upper-left corner of a spectral matrix of order d but they are selected with a simplified "zigzag" method [29]. In this case a facial area is described by $d(d+1)/2$ spectral components. In both cases the parameter d changes ranging from 10 to 50.

Within a supervised framework of face photo retrieval based on sketches we know the class of a given sketch $S(j)$ a priori. This makes it possible to evaluate numerically the accuracy of retrieval (Retrieval Rate – *RR*). RR is calculated as the ratio of *R* correctly found photos to a total number of *K* search attempts and is expressed in percent. Therefore $RR = (100 \times R)/K$ where *K* is the maximal number of search attempts that is equal to the number of photo-sketch pairs. The classifier is implemented basing on the criterion of minimum distance (CMD) in the L1 metric with rank 1. The task of classification of sketch $S(j)$ is reduced to calculation of all distances $dis(k) = distance\ (P(k),\ S(j)),\ \forall k$ and $\forall\ j \leq K$. Index k corresponds to the minimum value of dis(k) that defines the maximal proximity (or similarity) measure between $S(j)$ and a photo of class k in a database. The result of retrieval is considered correct if $j \equiv k$ and this is checked within a supervised classification.

The FaRetSys models for the first and second ways of selecting spectral components take the following form [29]:

CUFS [100//1 Photo (PP)//1 Sketch]{2DDCT:

$250 \times 200 \rightarrow (d \times d)$}[CMD//L1//rank=1]

and

CUFS [100//1 Photo (PP)//1 Sketch]{2DDCT:

$250 \times 200 \rightarrow (d(d+1)/2)$}[CMD//L1//rank=1],

where CUFS – image database name,
[100//1 Photo (PP))//1 Sketch] – base parameters: the number of classes in the database // number of test images, database images category, PP – preprocessing; {2DDCT: $250 \times 200 \rightarrow (d \times d)$} – features extraction method, the dimension of the input images, the dimension of the feature vector; [CMD//L1//rank=1] – classifier type (minimum distance classifier), distance estimation metric, ranking the result of classification.

The results of these models are shown in Fig. 9 where: a) examples of an original photo before and after preprocessing; b) a query sketch and an original photo as the result of retrieval; c) and d) dynamics of the RR for first and second ways respectively.

Fig. 9 shows that the accuracy of face photo retrieval by sketches grows with the increasing of the parameter d and reaches $RR = 100\%$ when $d \geq 30$. In further experiments we use the second method as it provides much smaller (approximately twice) number of spectral components. The presented results show that the problem of face photo retrieval based on sketches can be solved successfully within rather simple FaRetSys frameworks. On the other hand, these results are observed on the CUFS database where photo-sketch pairs have high values of ISSIM (>0.6). In real-world conditions such high values of ISSIM are not achievable and hence $RR = 100\%$ is also impossible.
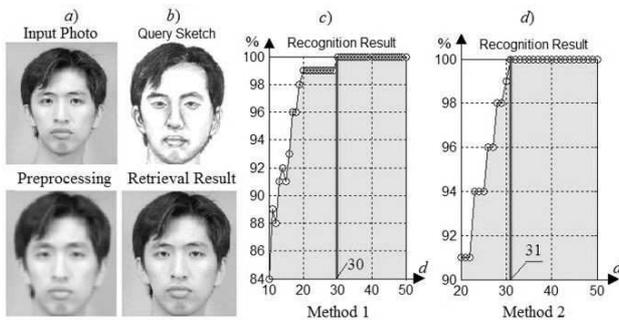
Fig. 9. Initial data and results for the experiment 1

*Retrieval system parameters and models
for experiments 2 and 3*

Population 1 imitates sketches created according to the descriptions of "group of witnesses" and corresponded to the testimony with incomplete information in verbal portraits. In this case the index of similarity of sketches of Population 1 with corresponding original photos is rather low and so the accuracy of face photo retrieval based on these sketches. Characteristics of sketches of Population 1 correspond to real-world conditions.

Population 2 sketches on the contrary have a significantly higher similarity index with original photos. Therefore, their use may become a satisfactory solution

for the problem of face photo retrieval based on these sketches. The experiments 2 and 3 confirm this fact.

So in experiment 2 tests are "median sketches" from Population 1 and in the experiment 3 tests are sketches with number $L = 10$. Taking these into account FaRetSys models for experiments 2 and 3 (and the second method of feature extracting) are as follows:

CUFS [100//1 Photo (PP)//1 Sketch/P1]{2DDCT:

$250 \times 200 \rightarrow (d(d+1)/2)$}[CMD//L1/rank=1]

and

CUFS [100//1 Photo (PP)//1 Sketch/P2]

{2DDCT:$250 \times 200 \rightarrow (d(d+1)/2)$}[CMD//L1/rank=1].

The results of these models are shown in Fig. 10, where: a) examples of sketches of Population 1; b) examples of sketches of Population 2; c) results of face photo retrieval based on these sketches.
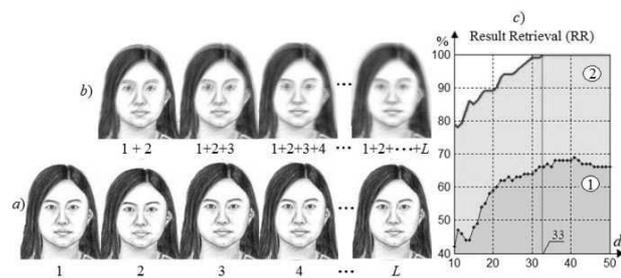
Fig. 10. Initial data and results for experiments 2 and 3

Serial numbers of sketches from Population 1 are marked with 1, 2, … L. Sketches from Population 2 (obtained by averaging of sketches from the Population 1) are marked with expressions $1 + 2$, $1 + 2 + 3$. Digits in circles define boundaries of RR's for sketches from Populations 1 and 2. The parameter d defines the order of a spectral matrix (or the number of spatial spectral components). The line marked "33" defines the minimum value of d where the accuracy of retrieval of 100 original photos by 100 sketches of the Population 2 achieves 100 % ($RR = 100\%$).

At a priori unknown value of d, in each experiment there is a variation task on selecting the "best value" of d. To solve it we have performed 40 attempts to search for $10 \leq d \leq 50$. The number of selected spectral components in this case was in the range from 55 to 1275 (or 1274 if the spectral component (1.1) is excluded).

As expected, results of face photo retrieval based on sketches from Population 1 are not satisfactory. Depending on the parameter σ RR values are in the interval $60 \div 80\%$. At the same time the results of face photo retrieval based on sketches from Population 2 are excellent: $RR = 100\%$ is already achieved at $d = 33$ that is very close to the result shown in Fig. 9d. In comparison with the curve "1" this result is higher on average by 30 %. This means that characteristics of sketches from Population 2 are close to characteristics of original sketches from the CUFS database. It is that what we intended to show and this result is consistent with the results shown in Fig. 5, 6 and 7, matching original photos with sketches from Population 2 using ISSIM criterion.

This high value of RR is mainly the result of the following factors: specific properties of sketches from Popu-

lation 2; using the area of interest in original photos and sketches; blurring the selected face area in an original photo (an averaging filter with a window of size 5×5 or 7×7). The first factor contributed approximately 20 %, while the second and the third ones added approximately 10 % altogether to the observed increase.

Estimating these results we have to note that only within the third experiment (where we used sketches of Population 2) the desirable result has been achieved. Experimentally we have proved that the usage of sketches from Populations 1 and 2 with $\sigma > 7$ gives RR reduced by $5 - 10\%$ on average. It is caused by lack of additional anthropometric alignment of original sketches and photos. Unfortunately, such an alignment, as practice shows, is not always possible due to the unavailability of original photos.

Experiments on FERET and CUFSF databases

To perform experiments we have built and used our own database that contains 220 photo-sketch pairs from the CUFSF database harmonized by facial geometric parameters using reference points defined in the CUFSF database. Each image in a pair represents a selected facial area of 160×128 pixels in gray scale. The alignment was performed automatically using the express method presented in [19]. It did not allow in some cases to achieve precise alignment of position, size and orientation of selected facial areas between photos and sketches. No additional operations to enhance the quality of original photos from the FERET database were done (!).Therefore, the average value of ISSIM in photo-sketch pairs does not exceed 0.3. First of all it speaks about poor quality of original photos especially in terms of dynamics of their brightness, glare and orientation. However, neither the first nor the second contradicts our initial assumptions about the low similarity of photos and sketches in real-world situations and scenarios.

Further, from this initial database of sketches new sketches of Population 1 with the parameter $\sigma \leq 5$ and from them sketches of Population 2 have been generated. Face photo retrieval based on sketches from Population 2 has been done. We used the following model of FaRetSys:

{FERET+CUFSF} [NUM//1 Photo//L Sketches] {2DDCT:$160 \times 128 \rightarrow (d(d+1)/2)$}[CMD//L1//rank=1:10], where $L = 9$ – number of sketches of Population 2 in each class, NUM $= 100$ and 220; $d \geq 21, 22$.

The screenshot of results of an original photo retrieval based on a given sketch is shown in Fig. 11. Face photo retrieval based on sketches is executed within a supervised classification where we calculate a rank of correct classification and depending on the sketch number k in the Population 2 ($k = 1, 2, ..., 10$). The results are summarized in Table 1.
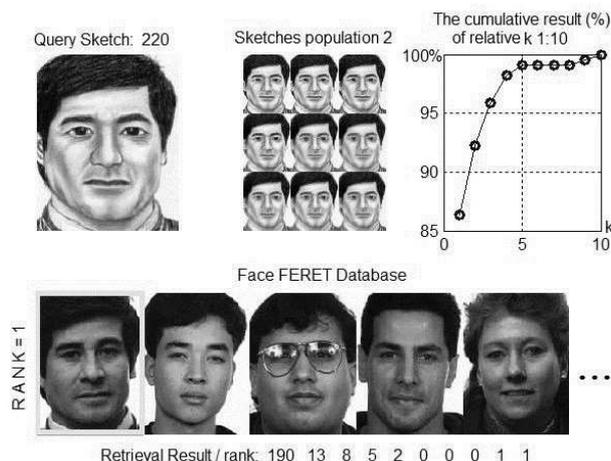
Fig. 11. Screenshot of resultы of an original photo retrieval based on a given sketch

Evaluating the results we note that the accuracy of matching photos from the FERET database with sketches from the CUFSF database (which were transformed into sketches of Population 2) is equal to 86.4 % for $k = 1$. This result is consistent, for example, with the case of recognition of photos «fa» from the FERET database based on photos «fb» from the same database. It also indicates a relatively high accuracy of recognition. The recognition result for $k = 5$ and $k = 7$ is more than 99 % for NUM $= 100$ and NUM $= 220$ respectively.

Table 1. Results of an original photo retrieval based on a given sketch

Population of sketches /NUM	Form of the result	Sketch number k in the Population 2 (for $d = 21$)									
		1	2	3	4	5	6	7	8	9	10
2 / 100	**correct classification**	91	6	2	1	0	0	0	0	0	0
	$100 \times R / K = RR$ %	91	6	2	1	0	0	0	0	0	0
Recognition result		-	97	99	100	100	100	100	100	100	100
2 / 220	**correct classification**	190	13	8	5	2	0	0	0	1	1
	$100 \times R / K = RR$ %	86,4	5.9	3.6	2,27	0,9	0	0	0	0,45	0,45
Recognition result		-	92,3	95.9	98,2	99,1	99,1	99,1	99,1	99,5	100

8. Analysis of experimental results

The high accuracy of results of face photo retrieval based on sketches received during the experiments on the CUFS database are related mainly to high quality of sketches and photos in this database.

The low accuracy of results received during the experiments on FERET and CUFSF databases are related to the low quality of sketches and photos in these databases. However, in both cases it has been shown that use of sketches from Population 2 improves the accuracy of retrieval results. It is obvious that when face photo retrieval is based on Forensic Sketches we cannot expect the same results [5, 6].Taking this and our previous experience into account we proposed to compare Forensic Sketches not with original photos but with original sketches from these databases (Viewed Sketches or Artist Sketches).

To improve the accuracy of retrieval we offered three solutions. In the first one original sketches are trans-

formed to populations and then within these populations a sketch that is similar to a given sketch (Forensic Sketch) is defined. The class of the sketch found in a population is a priori uniquely corresponds to the class of the original photo. In the second solution a Forensic Sketch is transformed to a population of sketches, and all original sketches from the initial database are compared to sketches from a new population "Forensic Sketches". The correspondence between classes is defined the same way as it was done in the first solution. The third solution includes the generation of population of sketches using both all original sketches and Forensic Sketches. The next step is matching of sketches from the two populations. These ideas are to be investigated in further studies.

Conclusion

The paper gives an overview of tasks related to the problem of face photo retrieval using sketches and systemizes the experience and results accumulated over the past two decades on this issue. The primary concepts, terminology used, ideas and modern technologies for construction of sketches are presented, and also the difficulties and the reasons of failures that arise in real-world search scenarios are shown. The history of developing systems for *constructing* facial composites (identikits and sketches) and the ideas realized in these systems are provided. The analysis of automatic matching tasks of sketches with original photo images has been made and the reasons of low performance of search in real-world scenarios are brought to light.

We have formulated the additional requirements for the existing databases of sketches and also methods of creating of such databases. The methods for generating a population of sketches from the initial sketches to improve the performance of searching of sketch-based photo image retrieval systems are discussed. A method to increase the similarity index in photo-sketch pairs based on computation of an average sketch from the generated population is provided. It is shown that such sketches are more similar to original photos and their use in the discussed matching problem may lead to good results. But for all that, the created sketches meet the requirements of the truthful scenario as they allow the possibility of incomplete information in verbal descriptions.

The results of experiments on CUHK Face Sketch and CUHK Face Sketch FERET databases and also open access databases of sketches and corresponding photos are discussed. These examples show that the proposed method of generation of sketches possesses the characteristic of universality as it can be independently used for sketches and identikits and any other correspondent available databases. Based on the results of the survey we conclude that the methods of matching sketches with the appropriate photos should be based on the approaches focused on specific scenarios. Therefore, further studies should be linked with the analysis of different scenarios taken from real situations. It is necessary to seek and create new options for the synthesis of sketches and methods for their recognition.

References

[1] Uhl R., Lobo N. da Vitoria. A framework for recognizing a facial image from a police sketch, Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, CA, USA, Jun. 18-20, 1996, pp. 586–593.

[2] Konen W. Comparing Facial Line Drawings with Gray-Level Images: A Case Study on PHANTOMAS, Proc. International Conference on Artificial Neural Networks, Bochum, Germany, Jul. 16-19. 1996. pp. 727–734.

[3] Identi-Kit, Identi-Kit Solutions. Source: ⟨http://www.identikit.net⟩.

[4] FACES 4.0. Source: ⟨http://www.iqbiometrix.com⟩.

[5] Yuen Pong C.A., Man C.H. Human Face Image Searching System using Sketch, Proc. Workshop on Machine Vision Applications (MVA2002), Nara-ken New Public Hall, Nara, Japan, Dec. 11-13. 2002. pp. 500–503.

[6] Tang X., Wang X. Face Photo-Sketch Synthesis and Recognition, Proc. 9th IEEE International Conference on Computer Vision, Nice, France, Oct. 13-16, 2003. Vol. 1. pp. 687–694.

[7] Tang X., Wang X. Face Photo-Sketch Synthesis and Recognition. IEEE Transactions on PAMI 2009; 31(11): 1955–1967.

[8] Klare B.F., Li Z., Jain A.K. Matching Forensic Sketches to Mug Shot Photos. IEEE Transactions on PAMI 2011; 33(3): 639–646.

[9] Hu H., Klare B., Bonnen K., Jain A.K. Matching Composite Sketches to Face Photos: A Component-Based Approach. IEEE Transactions on Information Forensics and Security 2013: 8(3): 191–204.

[10] Davies G.M., Valentine T. Facial composites: forensic utility and psychological research. In Handbook of eyewitness psychology. Memory for people, Mahwah: LEA 2007; 2: 59–83.

[11] Gibson S., Solomon C., Bejarano A. Synthesis of photographic quality facial composites using evolutionary algorithms. Proc. of the British Machine Vision Conference, BMVA Press, University of East Anglia, Norwich, UK, Sep. 9-11. 2003. pp. 221–230.

[12] Frowd C.B., Hancock P. J. B., Carson D. EvoFIT: A Holistic, Evolutionary Facial Imaging Technique for Creating Composites/ ACM Transactions on Applied Psychology 2004; 1(1): 19–39.

[13] George B., Gibson S.J., Maylin M.I.S., Solomon C.J. EFIT-V - Interactive Evolutionary Strategy for the Construction of Photo-Realistic Facial Composites. Proc. of the Conference on Genetic and Evolutionary Computation (GECCO), Atlanta, GA, USA, Jul. 12-16. 2008. pp. 1485–1490.

[14] Frowd C.D., Pitchford M., Skelton F., Petkovic A. Catching Even More Offenders with EvoFIT Facial Composites. Proc. IC EST-2012, Lisbon, Portugal, Sep. 5–7. 2012. pp. 20–26.

[15] Student Sketch Database. Source: ⟨http://mmlab.ie.cuhk.edu.hk/facesketch. html⟩.

[16] Face Sketch FERET Database Source: ⟨http://mmlab.ie.cuhk.edu.hk/cufsf⟩.

[17] Zhang W., Wang X., Tang X. Coupled Information-Theoretic Encoding for Face Photo-Sketch Recognition. Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, RI, USA, Jun. 20-25. 2011. pp. 513–520.

[18] Li X., Cao X. A Simple Framework for Face Photo-Sketch Synthesis. Mathematical Problems in Engineering 2012; 2012: 19 p.

[19] Galoogahi H.K., Sim T. Face Photo Retrieval by Sketch Example. Proc. International Conference ACM Multimedia, Nara, Japan, 29 Oct. –02 Nov., 2012. pp. 949–952.

[20] Sharma A., Jacobs D.W. Bypassing synthesis: PLS for face recognition with pose, low-resolution and sketch. Proc. 24th IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Colorado Springs, CO, USA, Jun. 20-25. 2011. pp. 593–600.

[21] Liang C., Mingquan Z., Yanjun H., Xiaoming D. Face Sketch Synthesis via Sparse Representation. Proc. 20th International Conference on Pattern Recognition (ICPR), Istanbul, Aug. 23-26, 2010. pp. 2146–2149.

[22] Kukharev G.A., Buda K., Shchegoleva N.L. Methods of Face Photo-Sketch Comparison. Pattern Recognition and Image Analysis 2014; 24(1): 102–113.

[23] Kukharev G.A., Buda K., Shchegoleva N.L. Sketch generation from photo to create test databases. Przegląd Elektrotechniczny (Electrical Review) 2014; 9(2): 97–100.

[24] Yu H., Zhang Jian J. Mean value coordinates–based caricature and expression synthesis. Signal, Image and Video Processing (SIViP) 2013; 7(5): 899–910.

[25] Kukharev G.A., Matveev Yu.N., Shchegoleva N.L. Matching of a sketches with an original photos. Proc. XVIII International Conference on Soft Computing and Measurements (SCM). 2015. – P. 157-159.

[26] Wang Z., Bovik A.C. A Universal Image Quality Index. IEEE Signal Processing Letters 2002; 9(3): 81–84.

[27] Wang Z., Bovik A.C., Sheikh H.R., Simoncelli E.P. Image quality assessment: From error measurement to structural similarity. IEEE Transactions on Image Processing 2004; 13(1): 1–14.

[28] Saeed M., ResaE. Mixture of experts: a literature survey. Artificial Intelligence Review 2014; 42(2): 275 –293.

[29] Methods of facial images processing and recognition in biometrics, edited by M.Hitrov, SPb.: Politechnika, 2013, 338 p.

Authors' information

Georgy A. Kukharev Born in Leningrad, Russia. Received Ph.D. degree (1997) from the Fine Mechanics and Optics Institute (Leningrad, Russia) and Doctor of Technical Science degree (1986) from the Institute of Automatics and Computer Facilities (ABT, Riga, Latvia), Full Professor (2006). From 1993, he works at Szczecin University of Technology, Faculty of Computer Science & Information Systems (Poland). From 2003, he works at Saint Petersburg State Electrotechnical University LETI, Department of Computer Software Environment. In 2001–2003 visiting professor of Ecole Centrale de Lyon, Department of Mathematics & Computer Science. Since 2005 visiting professor of Hanoi University of Technology, Department of International Training Programmer. Author of more ten monographs, over 100 scientific papers, and over 44 patents in the areas: computer architecture of signal processing, image processing, and pattern recognition. Current interests: biometrics, including face detection and face recognition, Face/Sketch Retrieval. E-mail: *moevm@mail.ru* .

Nadegda L. Shchegoleva was born in Komsomolsk-on-Amur, Russia. Received Ph.D. degree (2000) in Saint Petersburg Electrotechnical University (LETI). From 2001-2006 she was Senior Researcher in Federal State Unitary Enterprise Central Research Institute "Morphyspribor" (JSC "Concern" OKEANPRIBOR"). From 2007, she is Associate Professor in the Department of Software Engineering and Compute Application (Saint Petersburg Electrotechnical University "LETI", St.-Petersburg, Russia. Co-author of two monographs, 4 inventions, more than 60 scientific articles. Her research interests include biometric identification systems and access control systems, face recognition, synthesis and modeling of the recognition systems.