

УДК 621. 391. 037. 372

## ПОИСК ЛЮДЕЙ ПО ФОТОРОБОТАМ: СОСТОЯНИЕ ПРОБЛЕМЫ И ТЕХНОЛОГИИ

Г.А. Кухарев<sup>a</sup>, Ю.Н. Матвеев<sup>b,c</sup>, Н.Л. Щеголева<sup>d</sup>

<sup>a</sup>Западно-Поморский технологический университет в Щецине, Щецин, 70-310, Польша

<sup>b</sup> Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

<sup>c</sup> ООО «Центр речевых технологий», Санкт-Петербург, 196084, Российская Федерация, matveev@mail.ifmo.ru, matveev@speechpro.com

<sup>d</sup> Санкт-Петербургский государственный электротехнический университет (ЛЭТИ), Санкт-Петербург, 197376, Российская Федерация

**Аннотация.** Обсуждается проблема поиска людей по фотороботам, составленным по словесному портрету. Приводится обзор состояния этой проблемы от исходных понятий и используемой терминологии до современных технологий создания фотороботов, реальных сценариев и результатов поиска. Представлены история развития систем формирования композиционных портретов (фотороботов и скетчей) и идеи, реализованные в этих системах. Обсуждается задача автоматического сравнения фотороботов с оригинальными фотографиями, вскрываются причины недостижимости устойчивого поиска фотопортретов-оригиналов по фотороботам в реальных сценариях. Формулируются требования к базам фотороботов в дополнение к существующим бенчмарковым базам изображений лиц, а также способы реализации таких баз. В рамках этих способов обсуждаются методы генерации популяции фотороботов из исходного фоторобота для повышения результативности поиска по нему фотопортрета-оригинала. Представлен метод повышения индекса подобия в паре фоторобот/фотопортрет, основанный на вычислении среднего фоторобота из сформированной популяции. Показано, что такие фотороботы более подобны портретам-оригиналам и их использование в обсуждаемой проблеме поиска может привести к высоким результатам. При этом сформированные фотороботы отвечают требованиям правдивого сценария, поскольку учитывают возможность неполной информации в словесных портретах. Обсуждаются результаты применения этих методов для баз CUHK Face Sketch database и CUHK Face Sketch FERET database, а также опубликованных в открытой печати фотороботов и соответствующих им фотографий.

**Ключевые слова:** изображения лиц, фоторобот, скетч.

**Благодарности.** Данное исследование проводится при частичной финансовой поддержке Правительства Российской Федерации (грант № 074-U01).

## PEOPLE RETRIEVAL BY MEANS OF COMPOSITE PICTURES: PROBLEM STATE-OF-THE-ART AND TECHNOLOGIES

G.A. Kukharev<sup>a</sup>, Yu.N. Matveev<sup>b,c</sup>, N.L. Shchegoleva<sup>d</sup>

<sup>a</sup>West Pomeranian University of Technology, Szczecin, 70-310, Poland

<sup>b</sup>ITMO University, Saint Petersburg, 197101, Russian Federation

<sup>c</sup>STC Ltd., Saint Petersburg, 196084, Russian Federation, matveev@mail.ifmo.ru, matveev@speechpro.com

<sup>d</sup>Saint Petersburg Electrotechnical University (LETI), Saint Petersburg, 197376, Russian Federation

**Abstract.** We discuss the problem of people retrieval by means of composite pictures constructed according to descriptive portrait. An overview of the problem state-of-the-art is provided beginning from the basic concepts and terminology to a modern technology for composite picture creation, real-world scenarios and search results. The development history of systems for forming composite portraits (photo robots and sketches) and the ideas implemented in these systems are provided. The problem of automatic comparison of composite pictures with the original ones is discussed, and the reasons for unattainability of stable retrieval of originals by a composite picture in real-world scenarios are revealed.

Requirements to composite pictures databases in addition to the existing benchmark databases of facial images and also methods for implementation of such databases are formulated. Approaches for generation of sketches population from an initial one that increase effectiveness of identikit-based photo image retrieval systems are proposed. The method of similarity index increasing in the couple identikit-photograph based on computation of an average identikit from the created population is provided. It is shown that such composite pictures are more similar to original portraits and their use in the discussed search problem can lead to good results. Thus the created identikits meet the requirements of the truthful scenario as take into account the possibility of incomplete information in descriptions. Results of experiments on CUHK Face Sketch and CUHK Face Sketch FERET databases and also open access identikits and corresponding photos are discussed.

**Keywords:** face images, composite picture, sketch.

**Acknowledgements.** The work is partially financially supported by the Government of the Russian Federation (grant 074-U01).

### Введение в проблему синтеза фотороботов

В течение почти двадцати лет (начиная с работ [1, 2]) не утихает интерес к проблеме автоматического сравнения субъективного портрета, составленного по показаниям свидетелей некоторого криминального события, и оригинальными фотопортретами подозреваемых. Исходной при этом является информация, содержащаяся в показаниях свидетелей, и записанная с их слов (что определяется как словесный портрет подозреваемого человека). По форме исполнения субъективный портрет может быть представлен как рисованный портрет и как композиционный портрет. Рисованный портрет – это штриховой или полутоновый рисунок всей области лица, выполненный художником или криминалистом по словес-

ному портрету [3]. Композиционный портрет представляет собой изображение лица, составленное из его отдельных примитивов (например, бровей, глаз, носа, губ), а также сопутствующих элементов. Последние включают головные уборы, очки, сережки, банты, заколки и т.д. [3]. В свою очередь примитивы могут быть как рисованными (заранее подготовленными), так и состоять из фрагментов фотоизображений лиц. В этих случаях получается рисовано-композиционный или фотокомпозиционный портрет. Метод изготовления фотокомпозиционных портретов из фрагментов фотографий по словесному портрету был предложен в середине прошлого века французским криминалистом П. Шабо [3–6]. Эти портреты П. Шабо назвал фотороботами, и с тех пор все субъективные портреты, независимо от техники их создания и формы представления, стали называть фотороботами.

В зарубежной научно-технической литературе используется другая терминология для фотороботов, основу которой составляет слово скетч (Sketch [7–11]), что в переводе с английского означает «эскиз или набросок». При этом используются следующие основные формы таких фотороботов: Viewed Sketch; Artist Sketch; Composite Sketch; Composite Forensic Sketch.

Viewed Sketch – рисунок, выполненный художником по фотографии или непосредственно по лицу человека, которого видит художник. Часто под формой Viewed Sketch понимают также компьютерный рисунок, автоматически полученный из исходного цифрового изображения лица. Примеры Viewed Sketch-рисунков, выполненных художником по фотографии, показаны на рис. 1 в колонке «б». Дополнительно доработанный художником компьютерный рисунок (Viewed Sketch) определяется как Artist Sketch.

Forensic Sketch – рисунок, выполненный художником-криминалистом по словесному портрету со слов свидетеля. Если для составления скетча используется библиотека примитивов лица, то результат определяется как «Composite Sketch». Примеры Composite Sketch показаны на рис. 1 в колонке «в». Если при этом Composite Sketch составлен криминалистом по словесному портрету, то он определяется как Composite Forensic Sketch.

В настоящее время все фотороботы выполняются с помощью специальных компьютерных программ, позволяющих составлять композиционные портреты по примитивам, выбираемым из встроенной библиотеки этих программ. Первыми и наиболее простыми программами были «IdentiKit», «PhotoFit» и EFIT, более совершенными являются программы «Mac-a-Mug», «Портрет», «Облик» и «FACES», «IdentiKit 2000» [3–6, 8–12]. Основная идея, реализованная в этих программах, – механическая сборка (коллаж) оператором программы области лица из отдельных примитивов, содержащихся в базе. При этом выбранные в общий фоторобот примитивы могут принадлежать формально разным по генотипу людям. Однако поскольку склеивание примитивов и их подгонка по границам и текстуре области лица выполняется методами компьютерной графики, то это создает эффект хорошо выполненного фотопортрета.

Несмотря на представительные базы примитивов лиц и совершенную технику их склеивания в компьютерных программах в общий портрет, а также развитый интерфейс этих программ, качество получаемых фотороботов существенно зависит от опыта специалиста, обслуживающего саму программу, и субъективизма человека, составившего словесный портрет. Например, на рис. 1 видно, что исходное фото и Viewed Sketch практически точно совпадают между собой, что определено способом получения Viewed Sketch. Однако Composite Sketches (фотороботы) менее подобны исходному фото (хотя выполнены именно по нему), а также различаются и между собой, что обусловлено разными характеристиками использованных программ синтеза фотороботов.

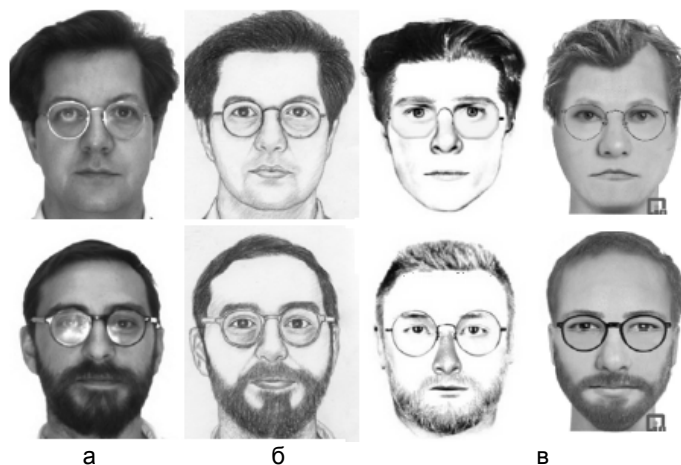


Рис. 1. Исходные изображения (фотопортреты) (а), Viewed Sketches (б) и два варианта Composite Sketches [9] (в)

Можно также отметить, что этот факт «неподобия» наблюдается практически при любых способах механической сборки фотороботов по отдельным примитивам. И мера этого неподобия растет, если фо-

тороботы формируются не по исходному изображению, а непосредственно по показаниям свидетелей или составленных ими словесных портретов. Ситуация эта еще более усугубляется, если исходный словесный портрет составляется свидетелем по памяти через несколько дней после контакта с подозреваемым (или преступником), когда в памяти свидетеля сохранился уже не полный состав исходных примитивов лица подозреваемого.

Эти факты «неподобия» вызвали новый интерес к созданию более совершенных методов и компьютерных систем построения фотороботов, в результате чего появились системы композиции лиц на основе эволюционных алгоритмов (ЭА) и интерактивных стратегий. В этом случае исходные лица для фотороботов не составляются по отдельным примитивам, а выбираются из базы систем с учетом фенотипа (или особенностей лица, вытекающих из словесного портрета) и рассматриваются как единое целое, а все изменения лиц осуществляются в рамках ЭА и корректируются свидетелем в интерактивном режиме. Наиболее простыми исходными данными по фенотипу могут быть, например, форма головы и (или) особенности отдельных примитивов лица, а также, например, национальные особенности лиц, пол и возраст человека, представленные в словесном портрете.

Первой из систем на основе ЭА и интерактивных стратегий была представлена система К. Соломона «EFITV-V – Eigen FIT version V», а затем и система Чарльза Фровда «EvoFIT – Evolutionary Facial Imaging Technique for Creating Composites» [13, 14]. Обе системы используют представление изображений лиц на основе модели формы (Active Shape Model – ASM) и модели внешнего вида (Active Appearance Model – AAM). При этом ASM определяет контур всего лица, а AAM – его текстуру. Параметры этих моделей представлены 50-ю признаками в собственном подпространстве, основанном на методах анализа главных компонент (Principal Component Analysis – PCA) и преобразовании Карунена–Лоэва (KLT – Karhunen-Loeve Transform). И именно эти параметры моделей изменяются в рамках ЭА на основе процедур клонирования и случайной мутации 50-ти исходных признаков.

В этих системах уже на первой пробе синтеза фоторобота генерируется «популяция» из нескольких лиц, отвечающих как индивидуальности искомого лица (фенотипу), так и случайно измененным его параметрам. Пример такой исходной популяции для системы «EFITV-V» из работы [15] показан на рис. 2, а. Далее свидетелем в интерактивном режиме выбирается тот результат из исходной популяции изображений лиц, который наиболее точно соответствует описанию исходного словесного портрета или отдельным примитивам словесного портрета (например, на рис. 2, а, выбранный результат отмечен кругом). Опираясь на этот результат как на текущую модель фоторобота, система генерирует новую популяцию фотороботов по измененным на основе ЭА параметрам текущей модели. Полученный при этом результат – новая популяция лиц – показан на рис. 2, б (пример также взят из работы [15]).

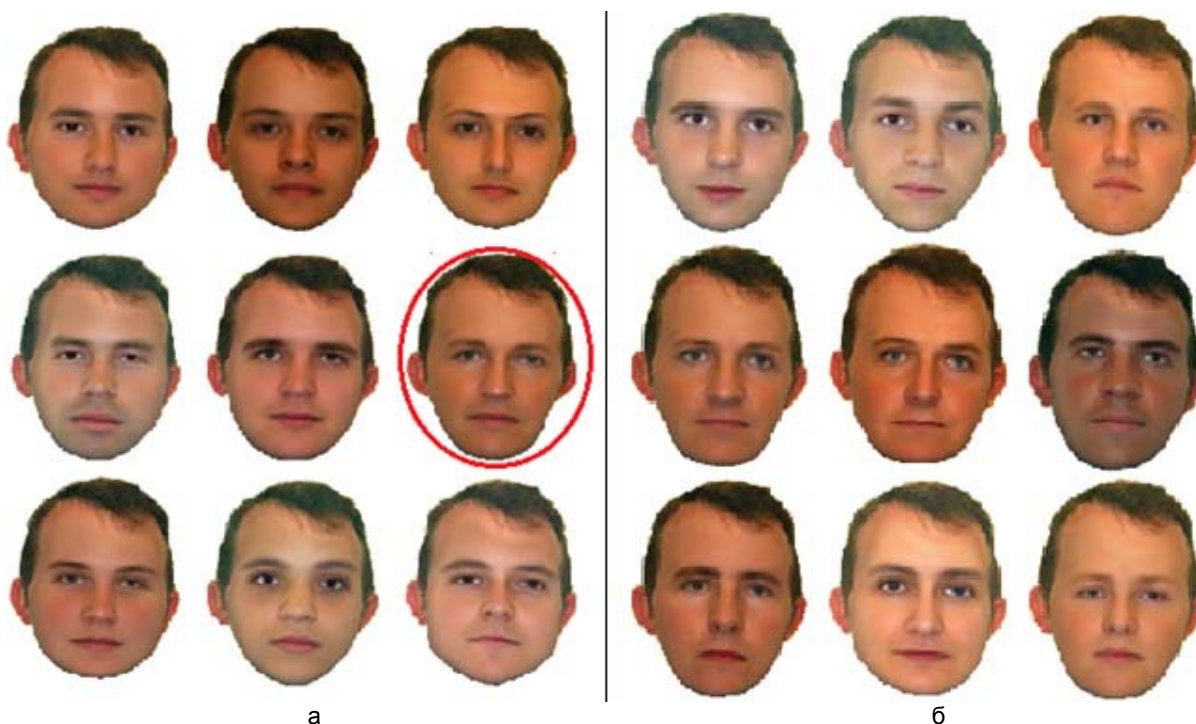


Рис. 2. Примеры популяции из 9 лиц в системе «EFITV-V» [13]: популяция нескольких изображений лиц, отвечающих как индивидуальности искомого лица (фенотипу), так и случайно измененным его параметрам, где выбранная свидетелем модель фоторобота отмечена кругом (а); новая популяция изображений лиц, сгенерированная на основе выбранной свидетелем модели фоторобота (б)

Если сравнить обе популяции изображений лиц, представленных на рис. 2, то можно отметить два факта. Факт первый: популяция на рис. 2, а, представляет разные лица, имеющие подобие только по форме волос на голове, но заметные различия по форме головы и базовым примитивам лица. Факт второй: популяция рис. 2, б, действительно представляет генотип лица одного и того же человека с небольшими вариациями его основных особенностей (примитивов). Наконец, теперь уже из этой популяции (рис. 2, б) выбирается то лицо, которое более точно соответствует описаниям исходного словесного портрета или отдельным (ранее не учтенным) его примитивам. Процесс продолжается далее, до полного утверждения (или согласования) свидетелем полученного результата.

Подход, реализованный в системах «EFIT-V» и «EvoFIT», основан на идеях ЭА и интерфейса человек–компьютер, поскольку промежуточные решения ЭА корректируются самим автором исходного словесного портрета. Окончательное решение также определяется автором, что приводит к быстрому и более точному получению окончательного фоторобота. Именно поэтому эти идеи были определены их авторами как стратегия интерактивной эволюции в создании фотореалистичных композиционных лиц (фотороботов). В ряде выполненных экспериментов было показано, что получаемые в рамках данной стратегии фотороботы известных людей настолько близки своим оригиналам, что относительно легко были идентифицированы экспертами (психологами и криминалистами) [15, 16].

Подводя итог настоящему разделу, отметим, что эффект высокого уровня подобия между фотопортретом-оригиналом и синтезированным фотороботом определяется следующими характеристиками стратегии интерактивной эволюции:

1. отказом от механизма «сборки фотороботов» по отдельным примитивам;
2. выбором в качестве исходного фоторобота целостного изображения лица с учетом данных по фенотипу;
3. использованием параметрических моделей изображения лица в собственном пространстве признаков;
4. изменением параметров модели фоторобота на основе ЭА;
5. выбором наилучших решений в рамках интерактивного общения со свидетелем.

### **Проблемы автоматического сравнения фотороботов с оригинальными фотографиями**

Возвращаясь к исходной проблеме, отметим, что построение фотороботов является только первым шагом в решении более общей задачи – задачи автоматического (т.е. без участия человека) поиска соответствий между фотороботом и оригинальным изображением лица. Эта задача появляется, например, при поиске оригинального изображения лица в большой базе данных по заданному фотороботу или при поиске соответствий между фотороботами и лицами людей в системах видеонаблюдений, а также при решении задач взаимного распознавания оригинального фото и скетча.

Как было показано выше, интерактив со свидетелем является чрезвычайно важным (можно сказать, даже принципиальным) в системах «EFIT-V» и «EvoFIT», поскольку только свидетель принимает окончательное решение о подобии фоторобота субъективному портрету. Однако даже фотороботы, построенные в рамках стратегии интерактивной эволюции, не решают проблемы автоматического сравнения фотороботов с оригинальными фотографиями.

Как показали результаты поиска, представленные в работах [10, 11], стабильное распознавание Composite Sketches и Composite Forensic Scetches и устойчивый поиск соответствующих им фотопортретов в специальных криминалистических базах являются на сегодня практически недостижимыми.

Этот печальный факт является следствием трех основных причин:

1. низким качеством скетчей, отображающих реальный словесный портрет;
2. несовершенством методов взаимного распознавания пары скетч-фотопортрет;
3. отсутствием необходимых для этих случаев баз фотопортретов и соответствующих им фотороботов.

Первая причина определяется технологией перевода имеющегося словесного портрета в соответствующее изображение. И здесь мы сталкиваемся с субъективизмом свидетелей, как правило, случайных людей, не связанных с криминалистикой, и особенностями их внимания и описания подозреваемого, сложенным с участием свидетеля в процессе перевода словесного портрета в скетч и высокой ответственностью свидетеля за ошибки. Очевидно, что лучшее решение здесь можно получить при использовании систем «EFIT-V» и «EvoFIT», хотя это и не всегда возможно на практике. В этом случае либо отсутствует возможность использования подобного класса систем, либо отсутствует возможность интерактивного общения со свидетелем, либо, наконец, интерактивное общение проходит так поздно (через несколько дней), что свидетель утрачивает (забывает) часть важной информации о подозреваемом.

Вторая причина связана с незрелостью методологии (идей, методов, подходов и практических решений) сравнения изображений скетч-фотопортрет, если скетчи выходят за рамки классов Viewed Sketch и Artist Sketch, а также отсутствием опыта такого сравнения ввиду недоступности информации по объектам сравнения (например, личностей криминального характера).

Третья причина является следствием неприспособленности «старых баз фотопортретов» преступников к современным компьютерным технологиям обработки изображений в лицевой биометрии. Кроме того, практически отсутствуют доступные специальные оценочные (бенчмарковые) базы данных, в которых каждый класс содержит оригинальный фотопортрет и различные варианты (или вариации) скетчей для него. Следует отметить, что из перечисленных причин третья является основополагающей: именно она не позволяет развивать методологию сравнения изображений скетч–фотопортрет, поскольку отсутствует объект исследований, что ограничивает или сдерживает выполнение задач сравнения на практике.

Именно поэтому исследователи обратились к задачам создания баз скетчей в дополнение к известным бенчмарковым базам изображений лиц [17, 18], к разработке методов сравнения скетчей с соответствующими фотопортретами и к моделированию задач поиска фотопортретов по заданным скетчам [6–9, 17–21]. В результате этого появились первые базы скетчей, наиболее популярные среди которых – CUHK Face Sketch database (CUFS) и CUHK Face Sketch FERET database (CUFSF), содержащие фото и полученные по ним скетчи классов Viewed Sketch и Artist Sketch [15, 16]. Кроме того, в работах [19–23] предложены новые идеи автоматического построения скетчей из исходных фотопортретов людей и методы их сравнения. При этом исследования выполнены на базах CUHK и CUFSF, а результатом являлись также скетчи класса Viewed Sketch. Если же говорить о методах сравнения фото и скетчей, то можно отметить следующее: использование достаточно сложных методов обработки в приложении к скетчам из баз CUHK и CUFSF не является оправданным. Возможно, большую роль здесь сыграла «мода» на перечисленные ниже методы, а не обоснованность их выбора. Среди них: скрытые марковские модели (HMM – Hidden Markov Models), методы на основе CITE (coupled information-theoretic encoding) и CITP (coupled information-theoretic projection), методы SIFT (Shift Invariant Feature Transform) и LBP (Local Binary Pattern), гистограммы BPH (Binary Pattern Histogram) и др., использованные в работах [8, 9, 17–23]. Заметим также, что при представлении результатов взаимного распознавания фото↔скетч в рамках баз CUFS и CUFSF в этих работах недостаточно ясно изложены не только методы обработки, но и параметры обучающих и тестовых выборок изображений, что не позволяет однозначно представить модель выполненных экспериментов. А это существенно затрудняет оценку представленных результатов, не позволяет проверить реализованную модель выполненных экспериментов и использовать полученные результаты в рамках мета-анализа.

Состояние дел по этим проблемам (подходы, решения, результаты и их анализ) было представлено в работах [24, 25] и в настоящей работе дальше не анализируется.

### Какие базы скетчей нужны сегодня?

База CUHK содержит скетчи, сгенерированные автоматически из исходных фото и дорисованные художниками, и включает 188 пар фото-скетч. База CUFSF содержит скетчи, передающие основные особенности портретов людей из базы FERET и особенности их лиц, но с элементами изменений (небольшого преувеличения с элементами карикатуризма), внесенными художником. Обе эти базы фактически составлены из скетчей, которые были определены выше как Artist Sketch.

Различие способов получения скетчей базы CUFS и базы CUFSF привело к тому, что первые из них распознаются простыми методами [24, 25] с большей эффективностью (почти до 100%), в то время как вторые также распознаются простыми методами, но только при очень точном согласовании размеров и ориентации области лица в плоскости XY. К сожалению, на сайте [18] доступны только выделенные изображения (cropped sketches) скетчей крайне низкого качества (по разрешению, размерам и текстуре), что практически не позволяет проводить с ними репрезентативные исследования.

Пример согласования выделенных областей лиц базы FERET и соответствующих им скетчей базы CUFSF [18] приведен на рис. 3.

В сравнении с выделенными скетчами (cropped sketches) из базы [18], на скетчах рис. 3, б, можно увидеть дополнительную ретушь областей лиц и элементов на ней (лба, носа, рта...). И, как видно, эти операции выполнены «вручную», т.е. не в автоматическом режиме обработки изображений. Кроме того, текстура на области лиц «выглажена» низкочастотной фильтрацией. И, наконец, выполнено согласование основных антропометрических параметров (линии глаз, расстояния между центрами глаз и т.д.).

Возможно, что для введения в задачи исследования методов сравнения скетчей и фото и их взаимного распознавания такое выравнивание необходимо. Однако, если исходить из реальных задач и, например, криминальных событий и сценариев (где актуальной становится задача поиска подозреваемого по заданному скетчу), то «условие выравнивания лиц по антропометрике» практически недостижимо. И это связано с тем, что заранее неизвестно, как выглядит оригинальное фото подозреваемого, а, следовательно, неизвестны и параметры лица на фото, и в какой степени будет оно соответствовать имеющемуся словесному портрету (и сгенерированному по нему скетчу).

Обратимся к рис. 4, где показаны пары фото/фоторобот (Composite Forensic Sketches), и отметим, что между фото и фотороботом нет такого высокого подобия, которое наблюдается, например, в соответствующих парах на рис. 1 или рис. 3.



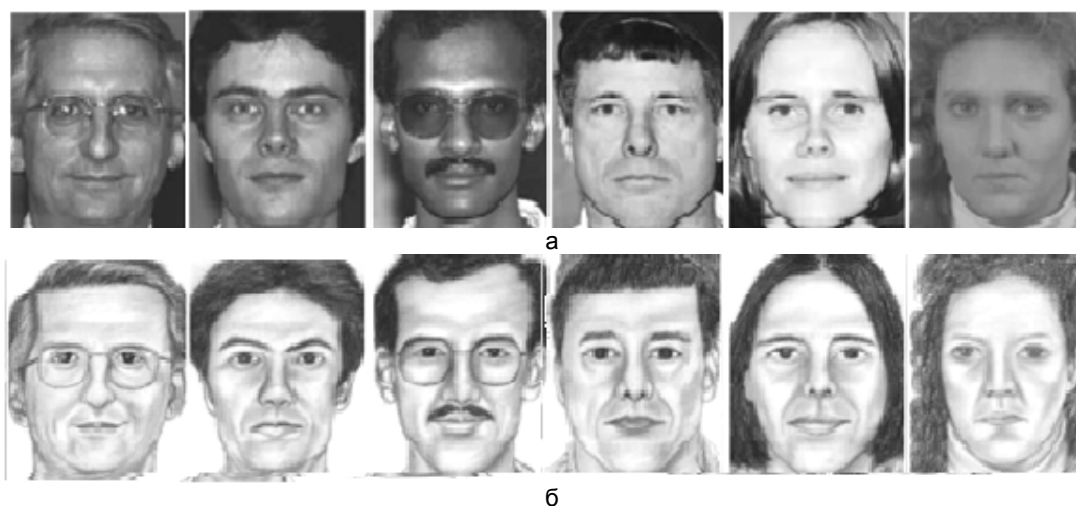


Рис. 3. Примеры согласования областей лиц на фото (а) и скетчах [18] (б)

На рис. 4 на всех четырех парах фото/скетч заметно различие размеров (высоты и ширины) областей лиц, размеров и положения примитивов на лицах, а также нарушение симметрии областей лиц и различие в ориентации взгляда. Естественно, что сравнение таких скетчей с фотооригиналами, ориентированными, как правило, анфас и нормализованными по стандарту, будет весьма затруднительным или практически невозможным.

При этом возникает новый вопрос: как же тогда сравнивать между собой скетчи и фотооригиналы? И здесь напрашивается аналогия с теми стратегиями, которые реализованы в системах «EFIT-V» и «EvoFIT». Ответ получаем такой: каждый исходный скетч должен быть несколько раз модифицирован и представлен с новыми параметрами по геометрии областей лиц (размеру, симметрии, сдвигу) с целью создания «новой популяции» таких скетчей. Данная модификация как бы симитировала получение  $K > 1$  скетчей от «группы из  $K$  свидетелей». И, уже в рамках этой популяции, можно решить задачу сравнения скетчей с фотооригиналом. Собственно само сравнение может выполняться со средним (для всей популяции) скетчем или с каждым скетчем из полученной популяции на основе мажоритарных механизмов или, например, на основе смеси экспертов (Mixtures of experts). И только в таком случае можно надеяться на хороший результат сравнения!

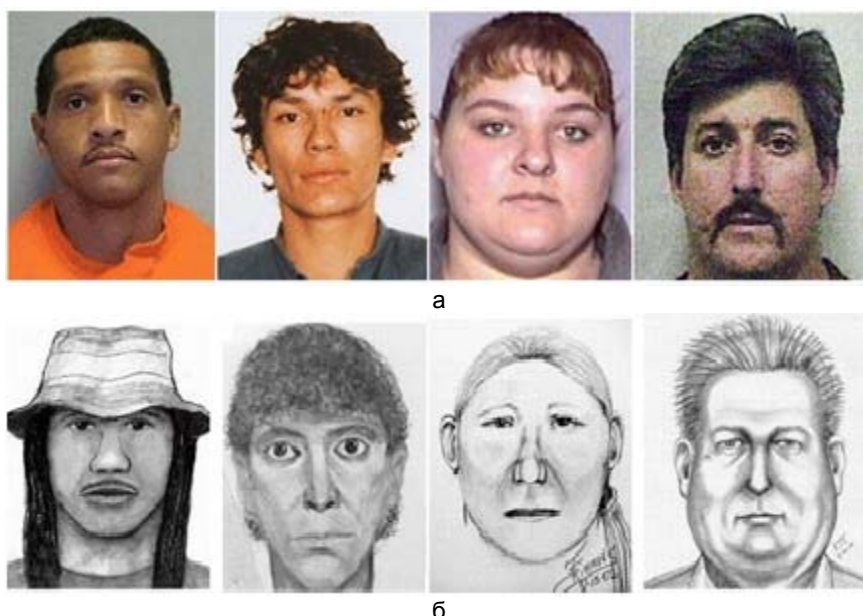


Рис. 4. Фотографии (а) и соответствующие им скетчи, составленные по описаниям свидетелей (данные с сайта <http://prikol.i.ua/view/474038>) (б)

Именно эти факты послужили отправной точкой для предложенных в работах [22, 23] подходов по генерации нового состава (популяции) скетчей из любого заданного исходного скетча. Причем механизм получения этих своеобразных популяций скетчей реализуется достаточно просто, что показано и о чем свидетельствуют результаты экспериментов на примере баз CUF5 и CUF5F, представленные в работах

[24, 25]. Справедливости ради отметим, что подобные же результаты можно было бы получить и на основе подходов для генерации «популяции карикатур», изложенных в работе [26]. Однако решения в [26] основаны на точных моделях форм (ASM) и моделях внешнего вида (AAM) изображений лиц, линейных преобразованиях между двумя изображениями (основанных на PCA и KLT) и различных вариантах переконфигурации и «преувеличения» компактных областей лиц или примитивов лиц. С учетом этого отметим, что подход [26] проигрывает подходам [24, 25] по сложности алгоритмов генерации требуемых популяций лиц. А высокая точность моделей внешнего вида, достигаемая в подходе [26], совершенно не обязательна в задаче генерации популяции скетчей, поскольку, по определению, мы не знаем того фотопортрета, к которому принадлежит заданный нам скетч.

В следующем разделе, в отличие и в дополнение к [24, 25], приводится полное описание процесса генерации «популяции скетчей» и исследуются их характеристики на примере Composite Forensic Sketches. Хотя вместо скетчей можно генерировать и популяции новых фотопортретов из некоторого заданного фотопортрета (например, из портрета mug-shot, часто используемого в криминалистике [10, 11]).

#### Алгоритм генерации «популяции скетчей»

Блок-схема алгоритма генерации «популяции скетчей» приведена на рис. 5. По этому алгоритму из исходного скетча (поданного на вход 1) формируются  $K > 2$  новых скетчей с небольшими геометрическими изменениями области лица. На вход 2 подаются параметры изменений скетчей.

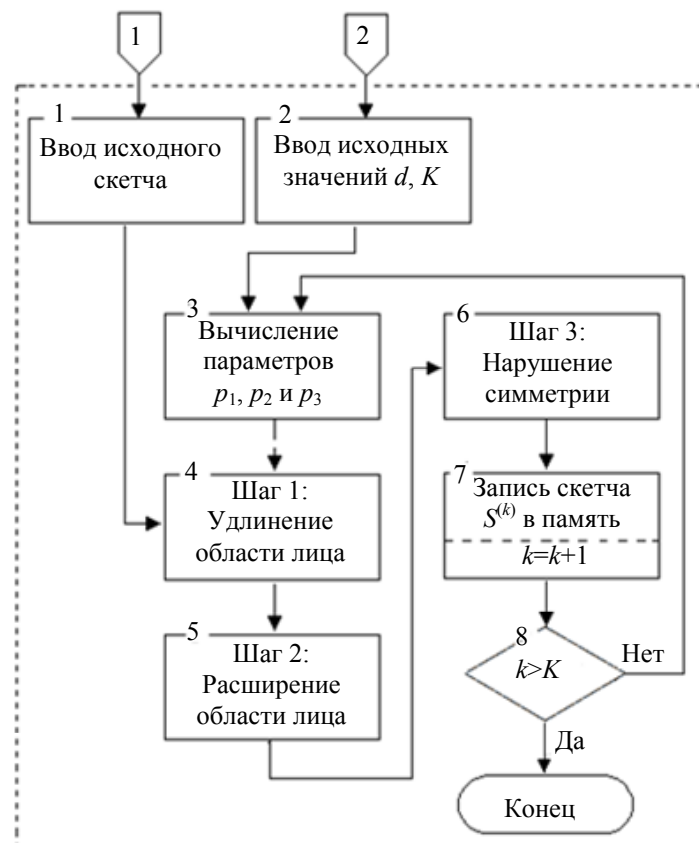


Рис. 5. Алгоритм генерации популяции скетчей, где 1 и 2 – входы для данных и параметров

Пусть задана матрица  $S$  размером  $M \times N$ , представляющая исходное изображение скетча в формате GRAY. При этом примем, что область лица на скетче занимает не менее 80% всего исходного изображения, что определено, например, биометрическими стандартами.

Для каждого значения  $k = 1, 2, \dots, K$  сформируем три параметра  $p_1, p_2$  и  $p_3$  с использованием генератора случайных чисел и отмасштабируем их значения в диапазоне  $\pm d$  так, что

$$p_i = \text{sign}(R_n^{(i)}) \cdot \text{fix}(d R_u^{(i)}), \text{ для } i = 1, 2, 3, \quad (1)$$

где  $p_i$  – параметр;  $d$  – выбранное значение изменения параметров, причем  $1 < d \leq \sigma$ ;  $\sigma$  – максимальное значение границ изменения параметров;  $R_n^{(i)}$  и  $R_u^{(i)}$  – нормально и равномерно распределенные случайные числа;  $\text{sign}(R_n^{(i)})$  – знак числа.

При этом параметры  $p_1$  и  $p_2$  связаны с изменением длины и ширины лица на исходном изображении, что приводит к изменению положения линии глаз и расстоянию между центрами глаз. Параметр  $p_3$

связан с изменением положения линии симметрии на области лица. А параметр  $\sigma$  может быть, например, связан с числом пикселей, на которые изменяется (увеличивается или уменьшается) центральная область лица, положение линии глаз и расстояния между центрами глаз или оси симметрии лица. Далее алгоритм геометрических изменений области лица складывается из трех шагов, на каждом из которых выполняется одна операция изменения геометрии области лица.

Шаг 1. Если  $p_1 > 0$ , то удаляем первые  $(p_1-1)$  строк из матрицы  $S$ . Если  $p_1 < 0$ , то удлиняем матрицу  $S$ , дописывая ее первые  $(p_1-1)$  строк сверху матрицы  $S$ . Указанные действия можно записать следующим образом:

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

где  $S^{(1)}(var \times N)$  – матрица с уменьшенным или увеличенным числом строк, что определено параметром «var».

Далее выполняем перемасштабирование матрицы  $S^{(1)}$  до ее первоначального размера  $M \times N$  так, что  $S^{(1)}(var \times N) \rightarrow S^{(1)}(M \times N)$ . (3)

При этом длина лица в матрице-результате  $S^{(1)}$  в (3) увеличивается, если выполнялось условие  $p_1 > 0$ , или уменьшается, если выполнялось условие  $p_1 < 0$ . При этом линия глаз и линия рта смещаются вверх или вниз. Заметим, что интересующее нас изменение области лица скетча будет находиться в пределах текущих значений  $d$  от исходной длины лица.

Шаг 2. Далее, если  $p_2 > 0$ , то удаляем первые  $(p_2-1)$  столбцы из матрицы  $S^{(1)}$ . Если  $p_2 < 0$ , то удаляем последние  $(p_2-1)$  столбцы матрицы  $S^{(1)}$ . Указанные действия можно записать следующим образом:

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

При этом матрица-результат  $S^{(2)}$  в формуле (4) становится на  $abs(p_2-1)$  столбцов меньше при любом значении  $p_2$ .

Далее опять выполним перемасштабирование матрицы  $S^{(2)}$  до первоначального размера  $M \times N$ :  $S^{(2)}(M \times var) \rightarrow S^{(2)}(M \times N)$ , (5)

что неминуемо приведет к расширению области лица в поле изображения и к асимметрии области лица. При этом без учета изменения длины области лица в (3) расширение области лица по (5) определится величиной, близкой к значению  $d$ .

Шаг 3. Выполним циклический сдвиг матрицы  $S^{(2)}$  влево на  $(p_3-1)$  столбцов, если  $p_3 > 0$ , или вправо, если  $p_3 < 0$ . Этот шаг запишем следующим образом:

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

что приводит к циклическому сдвигу всего изображения скетча, нарушающему его симметрию относительно центральной линии симметрии лица.

Теперь перепишем формулу (6) в новой форме:

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

где матрица-результат (7) представляет собой новый скетч.

Запишем результат  $S^{(k)}$  в память и перейдем снова к формированию параметров  $p_1$ ,  $p_2$  и  $p_3$  и далее к шагам 1–3, вплоть до формирования нового скетча  $S^{(k+1)}$  в соответствии с (1)–(7) и т.д.

Теперь ответим на следующие вопросы: как и чем оценить меру подобия между исходным фото и соответствующим ему скетчем? Должно ли их внешнее сходство (видимое человеку) соответствовать некоторому формальному показателю? И что нам важнее – видимое человеком внешнее сходство фото и скетча или формальная оценка этого сходства? Эти вопросы особенно актуальны в связках «свидетель→словесный портрет→скетч» и «скетч→найденное по нему фото». Некоторые ответы на эти вопросы можно найти в работах [24–28]. Ниже дополним эти ответы новым анализом результатов и обзором по ним.

### Метод повышения индекса подобия в паре скетч-фото

В работах [24, 25] было показано, что, если скетчи получаются непосредственно из исходных фотопортретов (Viewed Sketch или Artist Sketch), то они распознаются простыми системами со 100% результатом. Однако, как только эти скетчи пройдут этап локальных изменений области лица, результативность их распознавания существенно снижается. Отмеченные факты вполне объяснимы, поскольку мы изменили локальную структуру скетчей, добиваясь приближения их к ситуации, отвечающей правдивому сценарию. Подобные ситуации возникают при композиции скетчей по словесному портрету или при автоматическом получении скетчей для неточно заданных (или неполных) параметров «исходных данных». Если



же рассуждать на уровне «субъективного и формального подобия», в рамках полученной популяции скетчей мы сохранили их субъективное внешнее сходство с исходным скетчем, но существенно уменьшили значение формального параметра подобия – индекса структурного подобия (Structural SIMilarity Index – SSIM [27, 28]). Индекс SSIM (ISSIM) позволяет оценить степень подобия (искажения) двух изображений как комбинацию трех факторов: яркостных изменений, изменений контраста и потери корреляции между изображениями. При этом, поскольку результативность распознавания скетчей линейно связана со значением ISSIM, нашей следующей задачей является проведение нового (дополнительного) этапа обработки, сохраняющего субъективное внешнее сходство скетчей, но существенно увеличивающее значение индекса ISSIM.

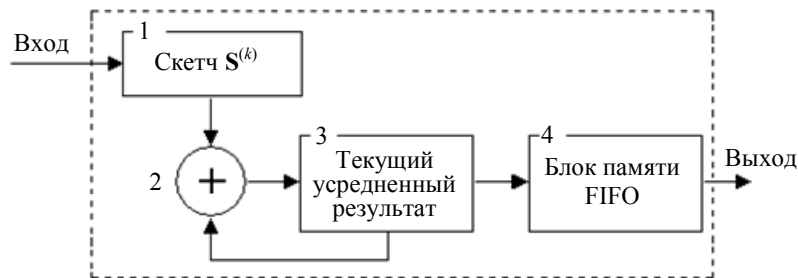


Рис. 6. Блок-схема алгоритма создания новой популяции скетчей: блок 1 – ввод исходного скетча; блок 2 – кумулятивное суммирование всех поступающих скетчей; блок 3 – вычисление текущего среднего скетча; блок 4 – запись результата в память

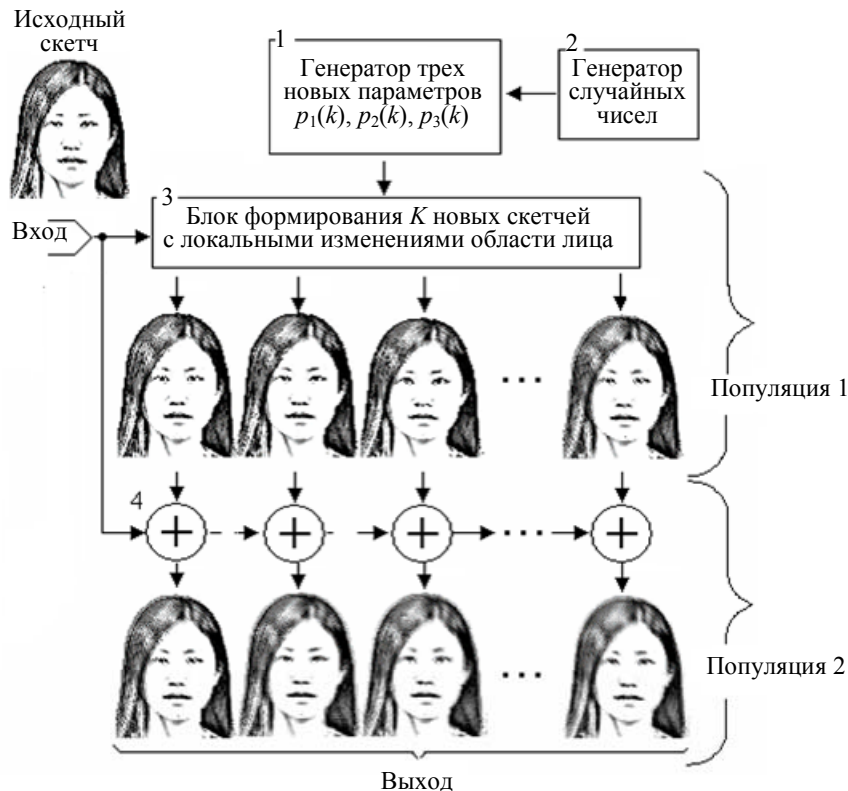


Рис. 7. Структура метода формирования популяции скетчей: блок 1 – генератор параметров  $p_1(k)$ ,  $p_2(k)$ ,  $p_3(k)$ ; блок 2 – генератор случайных чисел; блок 3 – формирование  $K$  новых скетчей с локальными изменениями области лица; блок 4 – суммирование скетчей

Однако это не может быть операция, обратная той, что была описана выше, поскольку мы и дальше исходим из предположения «неполного знания параметров исходного фотопортрета» или «полного незнания этих параметров». Это будет этап создания новой популяции скетчей, предложенный в [24, 25], идеи которого представляет рис. 6. Здесь на вход 1 последовательно поступают  $K$  скетчей  $\mathbf{S}^{(k)}$ , представленных матрицами данных размером  $M \times N$ . На выходе также получаем  $K$  новых скетчей  $\tilde{\mathbf{S}}^{(k)}$ , что достигается кумулятивным суммированием всех поступающих скетчей и вычислением текущего среднего скетча  $\tilde{\mathbf{S}}^{(k)}$ :

$$\tilde{\mathbf{S}}^{(k)} = (\sum_{j=1}^k \mathbf{S}^{(j)})/k, \text{ для } k = 1, 2, \dots, K. \quad (8)$$

Операция кумулятивного суммирования (8) реализуется в блоках 2 и 3. Результат записывается в память (на промежуточном этапе это может быть память типа FIFO – блок 4), что показано на рис. 6. На рис. 7 показан фрагмент взаимодействия блоков 3 и 4, где под номером 4 показана линейка сумматоров.

### Анализ результатов

Теперь покажем, что скетчи популяции 2 имеют более высокий индекс подобия с исходным изображением, чем скетчи, полученные в популяции 1.

Рассмотрим рис. 8, на котором показаны исходный фотопортрет (рис. 8, б) и соответствующие ему скетчи из [17]: Viewed Sketch (рис. 8, а) и Artist Sketch (рис. 8, в). На рис. 8, г, показаны значения индекса SSIM между исходным фотопортретом и Viewed Sketch из популяции 1 (нижняя кривая П1) и популяции 2 (верхняя кривая П2). На рис. 8, д, показаны значения индекса SSIM между исходным фотопортретом и Artist Sketch из популяции 1 (нижняя кривая П1) и популяции 2 (верхняя кривая П2). Кривые значений ISSIM приведены для 9 скетчей из этих популяций. Прямые горизонтальные линии отмечают пороги 0,62 и 0,495 как значения индекса подобия между фотопортретом (рис. 8, б) и исходными скетчами (рис. 8, а, в).

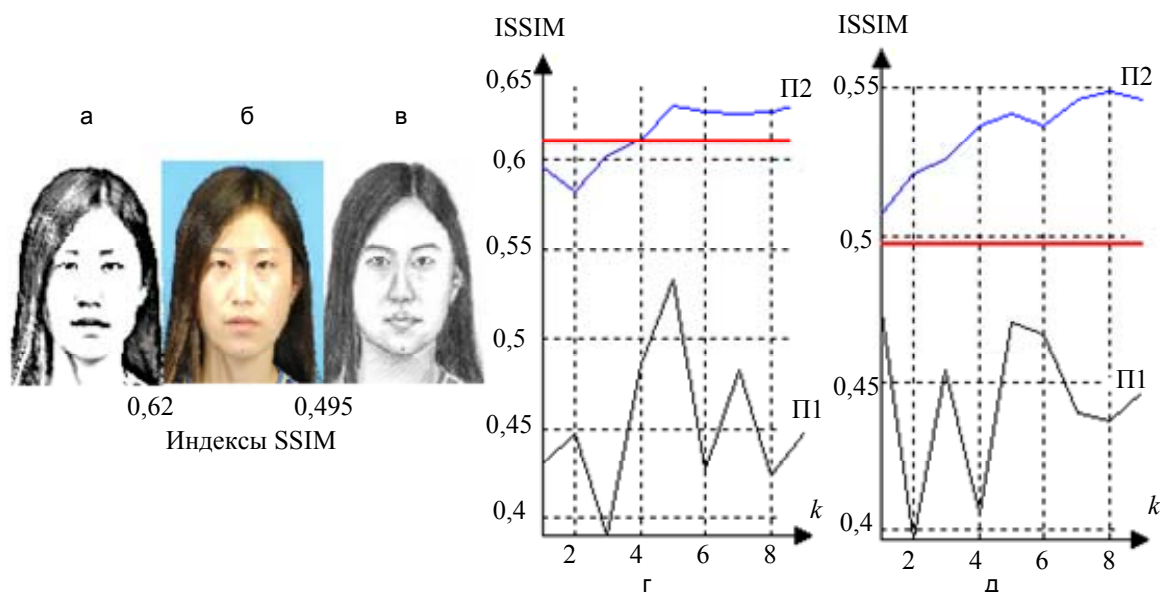


Рис. 8. Изменения индекса SSIM для исходных скетчей и сгенерированных популяций скетчей: Sketch (а); исходный фотопортрет (б); Artist Sketch (в); значения индекса SSIM между исходным фотопортретом и Viewed Sketch из популяции 1 (нижняя кривая П1) и популяции 2 (верхняя кривая П2) (г); значения индекса SSIM между исходным фотопортретом и Artist Sketch из популяции 1 (нижняя кривая П1) и популяции 2 (верхняя кривая П2) (д)

Как видно из приведенных результатов, значения индекса SSIM для скетчей, сформированных в популяции 2, выше порога сравнения в обоих случаях. Таким образом, можно утверждать, что скетчи из популяции 2 ближе к портретам-оригиналам и, следовательно, качество этих скетчей стало выше. При этом сформированные скетчи отвечают требованиям правдивого сценария, поскольку учитывают возможность неполной информации о портретах-оригиналах – неточно заданных или неполных параметрах «исходного фотопортрета».

А теперь покажем, что предлагаемый способ повышения качества скетчей может быть применен также и для фотороботов, применяемых в криминалистической практике. Используем фоторобот из работы [10], который интересен тем, что был распознан при ранге 72, т.е. портрет-оригинал для него находился на 72 месте в последовательности найденных по нему результатов.

На рис. 9 приведены: оригинальное фото (рис. 9, а) и соответствующий ему фоторобот (рис. 9, б) из работы [10], а также значения индексов SSIM между ними в сформированных популяциях. В верхнем ряду показаны: портрет-оригинал, вариант этого портрета, полученного в популяции 1 (цифры над ним – значения параметров  $p_1$ ,  $p_2$  и  $p_3$ ); портрет, полученный в популяции 2 для значения  $k = 10$ . В нижнем ряду показаны: фоторобот, вариант фоторобота, полученного в популяции 1 (цифры над ним – значения параметров  $p_1$ ,  $p_2$  и  $p_3$ ); фоторобот, полученный в популяции 2 для значения  $k = 10$ . Уже из этих результатов видно, что индекс SSIM для фотороботов популяции 2 выше, чем индекс SSIM популяции 1. Этот вывод подтверждается динамикой изменений индекса SSIM, что показано на графиках рис. 10.

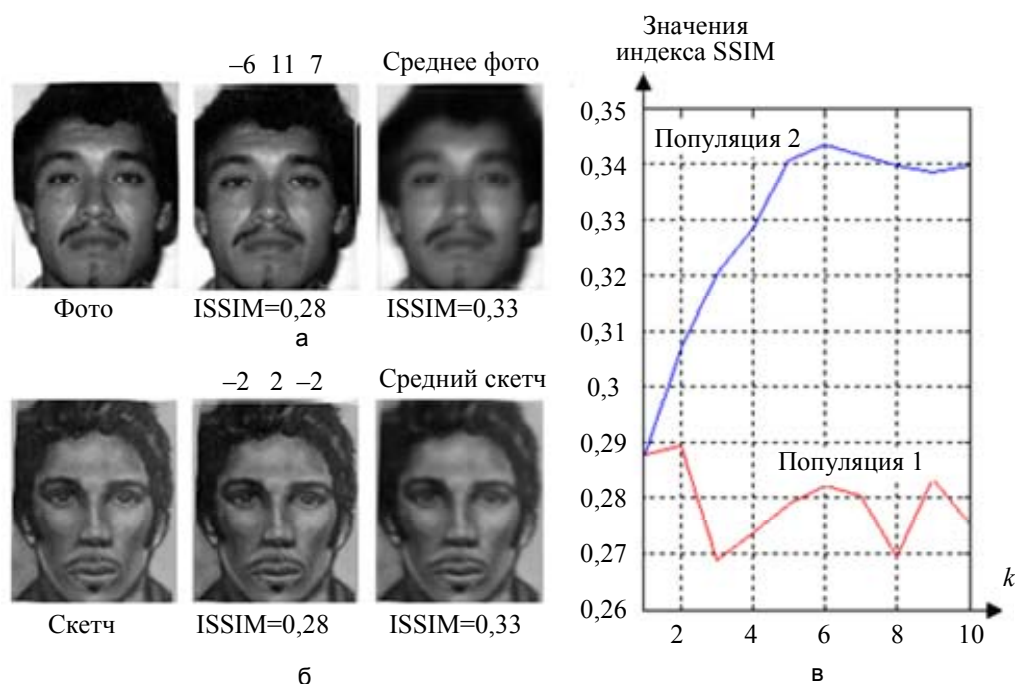


Рис. 9. Портрет-оригинал (фото) (а); фоторобот (скетч) (б) и значения индекса SSIM для них (в)

На рис. 10 показаны варианты Composite Forensic Sketches (см. исходные скетчи на рис. 4) для популяции 1 и 2, а на рис. 11 – значения индекса SSIM для них. Здесь также видно, что индекс SSIM для скетчей популяции 2 выше, чем индекс SSIM популяции 1.

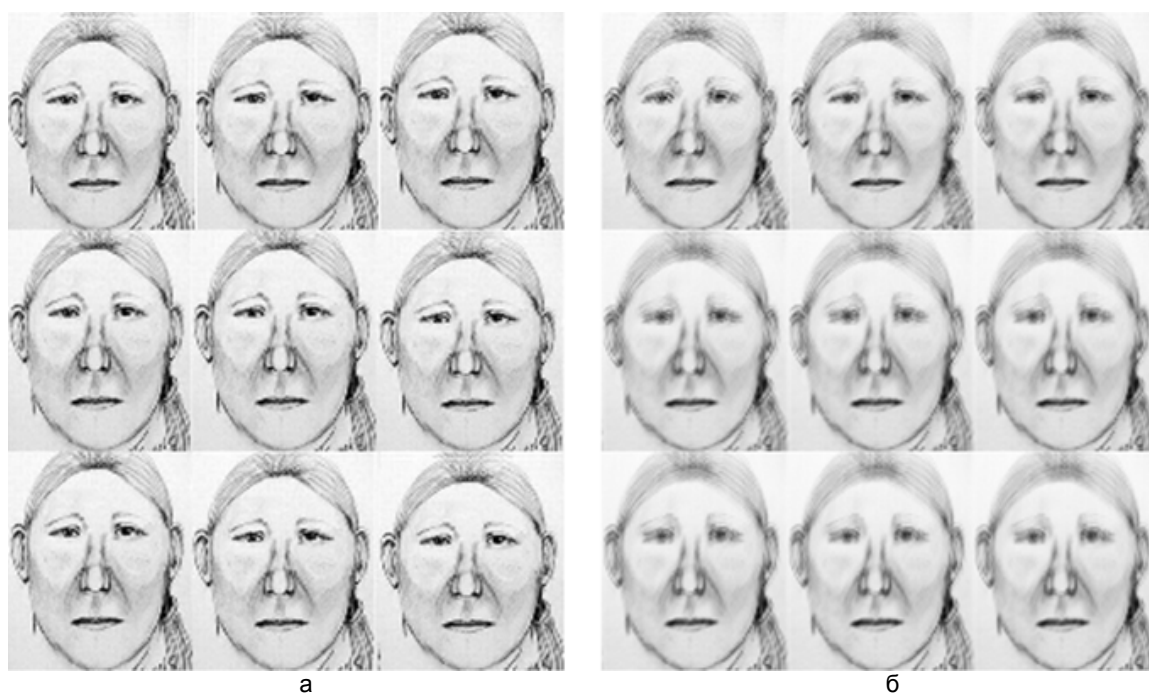


Рис. 10. Варианты Composite Forensic Sketches для популяций: 1 (а); 2 (б)

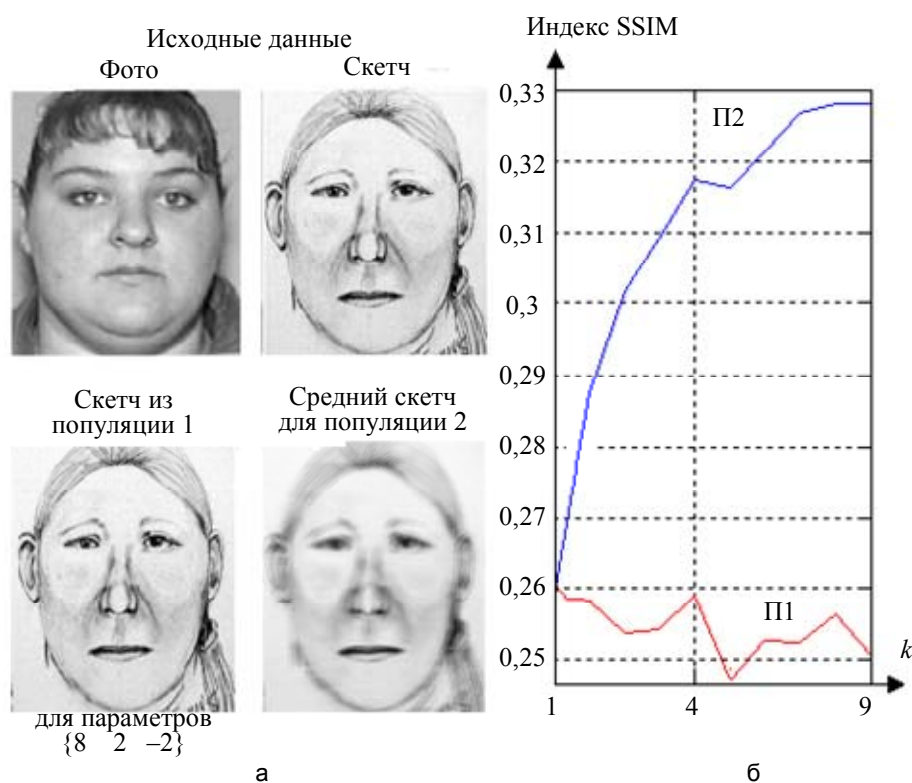


Рис. 11. Характеристики фотооригинала и Composite Forensic Sketch (а) и значения индекса SSIM для популяций 1 и 2 (б)

Подводя итог, отметим следующее.

1. Модификация исходных данных (фотороботов/скетчей) и их представление в форме популяции 1 имитирует получение новых данных от группы из  $K$  свидетелей. Этот эффект можно воспринимать как создание более объективного представления для фоторобота по имеющимся исходным данным. В таких предпосылках, уже в рамках популяции 1, можно достаточно эффективно решать задачу сравнения новых данных с фотооригиналом.
2. При этом само сравнение может выполняться со средним (для всей популяции) фотороботом или с каждым фотороботом из популяции на основе мажоритарных механизмов или, например, на основе смеси экспертов (Mixtures of experts [29]).
3. Модификация результата популяции 1 в результате популяции 2 улучшает подобие между парами фоторобот–оригинальный фотопортрет. С учетом отмеченного эффекта и в совокупности с механизмами, перечисленными в пункте 2, это создает новые условия для еще более эффективного сравнения фотороботов с фотопортретами-оригиналами.
4. Для оценки подобия между парами фоторобот–оригинальный фотопортрет можно использовать индекс SSIM, поскольку он оценивает корреляцию и текстуру локальных областей между исходными данными [27, 28].

### Заключение

В настоящей работе выполнен обзор задач, связанных с проблемой поиска людей по фотороботам, а также систематизирован опыт и результаты, накопленные за последние два десятилетия по этой проблеме. Представлены исходные понятия, используемая терминология, идеи и современные технологии создания фотороботов, а также показаны трудности и причины неудач, возникающих в реальных сценариях поиска. Представлена история развития систем формирования композиционных портретов (фотороботов и скетчей) и идеи, реализованные в этих системах. Выполнен анализ задач автоматического сравнения фотороботов с оригинальными фотографиями, вскрыты причины недостижимости устойчивого поиска фотопортретов-оригиналов по фотороботам в реальных сценариях.

Сформулированы требования к базам фотороботов в дополнение к существующим базам изображений лиц, а также способы реализации таких баз. Как один из возможных вариантов рассмотрены методы генерации популяции фотороботов из исходного фоторобота для повышения результативности поиска по нему фотопортрета-оригинала. Представлен метод повышения индекса подобия в паре фоторобот–фотопортрет (оригинал), основанный на вычислении среднего фоторобота из сформированной популяции. Показано, что сформированные таким образом фотороботы более подобны портретам-оригиналам, и

их использование в обсуждаемой проблеме поиска может привести к высоким результатам. При этом сформированные фотороботы отвечают требованиям правдивого сценария, поскольку учитывают возможность неполной информации в словесных портретах. Обсуждаются результаты применения этих методов для двух популярных баз фото-скетчи, а также опубликованных в открытой печати фотороботов и соответствующих им фотопортретов. Эти примеры демонстрируют, что выявленный способ генерации скетчей обладает характеристикой универсальности за счет возможности независимого его использования (для скетчей и фотороботов) и любых доступных баз исходных данных.

По результатам выполненного обзора можно прогнозировать, что методы сравнения скетчей с соответствующими фотопортретами должны быть основаны на подходах, ориентированных на конкретные сценарии. Исходя из этого, дальнейшие исследования необходимо связывать с анализом различных сценариев, взятых из реальных ситуаций. Именно для этих случаев необходимо искать и создавать новые варианты синтеза скетчей и методы их распознавания.

### References

1. Uhl R.J. Jr., Lobo N.V. A framework for recognizing a facial image from a police sketch. *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. San Francisco, USA, 1996, pp. 586–593. doi: 10.1109/CVPR.1996.517132
2. Konen W. Comparing facial line drawings with gray-level images: a case study on PHANTOMAS. *Lecture Notes in Computer Science*, 1996, vol. 1112 LNCS, pp. 727–734. doi: 10.1007/3-540-61510-5\_123
3. *Kriminalistika* [Criminalistics]. 3<sup>rd</sup> ed. Ed. N.P. Yablokov. Moscow, Yurist Publ., 2005, 781 p.
4. *Azbuka Kriminalistiki* [ABC of Criminology]. Available at: <http://www.expert.aaanet.ru/arhiv/identif.htm> (accessed 11.08.2014).
5. *Identi-Kit Solutions*. Available: <http://www.identikit.net> (accessed 11.08.2014).
6. *IQ Biometrics. Faces Software*. Available at: <http://www.iqbiometrix.com> (accessed 11.08.2014).
7. Yuen P.C., Man C.H. Human face image searching system using sketch. *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans*, 2007, vol. 37, no. 4, pp. 493–504. doi: 10.1109/TSMCA.2007.897588
8. Tang X., Wang X. Face photo-sketch synthesis and recognition. *Proc. 9<sup>th</sup> IEEE International Conference on Computer Vision*. Nice, France, 2003, vol. 1, pp. 687–694.
9. Wang X., Tang X. Face photo-sketch synthesis and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, vol. 31, no. 11, pp. 1955–1967. doi: 10.1109/TPAMI.2008.222
10. Klare B., Li Z, Jain A.K. Matching forensic sketches to mug shot photos. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, vol. 33, no. 3, pp. 639–646. doi: 10.1109/TPAMI.2010.180
11. Han H., Klare B.F., Bonnen K., Jain A.K. Matching composite sketches to face photos: a component-based approach. *IEEE Transactions on Information Forensics and Security*, 2013, vol. 8, no. 3, pp. 191–204. doi: 10.1109/TIFS.2012.2228856
12. Davies G.M., Valentine T. Facial composites: forensic utility and psychological research. In *Handbook of Eyewitness Psychology*. Mahwah, LEA, 2007, vol. 2, pp. 59–83.
13. Gibson S.J, Solomon C.J, Pallares-Bejarano A. Synthesis of photographic quality facial composites using evolutionary algorithms. *Proc. British Machine Vision Conference*. Norwich, UK, 2003, pp. 221–230.
14. 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, vol. 1, pp. 1–21.
15. 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. 10<sup>th</sup> Annual Conference on Genetic and Evolutionary Computation*. Atlanta, USA, 2008, pp. 1485–1490.
16. Frowd C.D., Pitchford M., Skelton F., Petkovic A., Prosser C., Coates B. Catching even more offenders with EvoFIT facial composites. *Proc. 3<sup>rd</sup> International Conference on Emerging Security Technologies, EST-2012*. Lisbon, Portugal, 2012, pp. 20–26. doi: 10.1109/EST.2012.26
17. *CUHK Face Sketch Database*. Available at: <http://mmlab.ie.cuhk.edu.hk/facesketch.html> (accessed 11.08.2014).
18. *CUHK Face Sketch FERET Database*. Available at: <http://mmlab.ie.cuhk.edu.hk/cufsf/> (accessed 11.08.2014).
19. Zhang W., Wang X., Tang X. Coupled information-theoretic encoding for face photo-sketch recognition. *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Colorado Springs, USA, 2011, pp. 513–520. doi: 10.1109/CVPR.2011.5995324
20. Li X., Cao X. A simple framework for face photo-sketch synthesis. *Mathematical Problems in Engineering*, 2012, vol. 2012, art. 910719. doi: 10.1155/2012/910719
21. Kiani Galoogahi H., Sim T. Face photo retrieval by sketch example. *Proc. 20<sup>th</sup> ACM International Conference Multimedia*. Nara, Japan, 2012, pp. 949–952. doi: 10.1145/2393347.2396354



22. Sharma A., Jacobs D.W. Bypassing synthesis: PLS for face recognition with pose, low-resolution and sketch. *Proc. 24<sup>th</sup> IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Colorado Springs, USA, 2011, pp. 593–600. doi: 10.1109/CVPR.2011.5995350
23. Chang L., Zhou M., Han Y., Deng X. Face sketch synthesis via sparse representation. *Proc. 20<sup>th</sup> International Conference on Pattern Recognition*. Istanbul, Turkey, 2010, pp. 2146–2149. doi: 10.1109/ICPR.2010.526
24. Kukharev G.A., Buda K., Shchegoleva N.L. Methods of face photo-sketch comparison. *Pattern Recognition and Image Analysis*, 2014, vol. 24, no. 1, pp. 102–113. doi: 10.1134/S1054661814010076
25. Kukharev G., Buda K., Shchegoleva N. Sketch generation from photo to create test databases. *Przegląd Elektrotechniczny (Electrical Review)*, 2014, vol. 90, no. 2, pp. 97–100. doi: 10.12915/pe.2014.02.26
26. Yu H., Zhang J.J. Mean value coordinates-based caricature and expression synthesis. *Signal, Image and Video Processing*, 2013, vol. 7, no. 5, pp. 899–910. doi: 10.1007/s11760-011-0279-8
27. Wang Z., Bovik A.C. A universal image quality index. *IEEE Signal Processing Letters*, 2002, vol. 9, no. 3, pp. 81–84. doi: 10.1109/97.995823
28. Wang Z., Bovik A.C., Sheikh H.R., Simoncelli E.P. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 2004, vol. 13, no. 4, pp. 600–612. doi: 10.1109/TIP.2003.819861
29. Masoudnia S., Ebrahimpour R. Mixture of experts: a literature survey. *Artificial Intelligence Review*, 2014, vol. 42, no. 2, pp. 275–293. doi: 10.1007/s10462-012-9338-y

<b>Кухарев Георгий Александрович</b>	– доктор технических наук, профессор, профессор, Западно-Поморский технологический университет в Щецине, Щецин, 70-310, Польша, gkukharev@wi.zut.edu.pl
<b>Матвеев Юрий Николаевич</b>	– доктор технических наук, профессор, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация; главный научный сотрудник, ООО «Центр речевых технологий», Санкт-Петербург, 196084, Российская Федерация, matveev@mail.ifmo.ru, matveev@speechpro.com
<b>Щеголева Надежда Львовна</b>	– кандидат технических наук, доцент, Санкт-Петербургский государственный электротехнический университет (ЛЭТИ), Санкт-Петербург, 197376, Россия, NLSchegoleva@etu.ru
<b>Georgy A. Kukharev</b>	– D.Sc., Full Professor, Westpomeranian University of Technology, Szczecin, 70-310, Poland, gkukharev@wi.zut.edu.pl
<b>Yuri N. Matveev</b>	– D.Sc., Professor, ITMO University, Saint Petersburg, 197101, Russian Federation; Chief scientist, STC Ltd., Saint Petersburg, 196084, Russian Federation, matveev@mail.ifmo.ru, matveev@speechpro.com
<b>Nadezhda L. Shchegoleva</b>	– PhD, Associate professor, Saint Petersburg Electrotechnical University (LETI), Saint Petersburg, 197376, Russian Federation, NLSchegoleva@etu.ru

Принято к печати 25.09.14  
Accepted 25.09.14