

УДК 004.932.001.57

Кязимов Т.Г.¹, Махмудова Ш.Д.²

Институт Информационных Технологий НАНА, Баку, Азербайджан

¹tofig@mail.ru, ²shafagat_57@mail.ru

ОБ УЛУЧШЕНИИ АЛГОРИТМА РАСПОЗНАВАНИЯ В ИДЕНТИФИКАЦИОННОЙ СИСТЕМЕ НА ОСНОВЕ ФОТОПОРТРЕТА

Рассматриваются вопросы по определению доверительных интервалов геометрических характеристик человеческого лица для идентификации человека на основе фотопортрета. Подчеркнута важность данной задачи при определении точности идентификации. Для определения доверительных интервалов геометрических характеристик используется известный метод Стьюдента. Решение данной задачи приводит к повышению эффективности идентификационной системы.

Ключевые слова: биометрические технологии, изображение, геометрические характеристики, доверительный интервал, распознавание, метод Стьюдента.

Введение

Биометрика – это технология, идентифицирующая личность при помощи субъективных физиологических параметров человека (отпечатки пальцев, сетчатка глаза и т.д.). Биометрическая технология была разработана на основе биометрической науки, учитывающей уникальные характеристики отдельного человека. Биометрические технологии широко используются при идентификации личности.

Биометрические характеристики делятся на две основные группы:

- 1) физиологические биометрические характеристики;
- 2) биометрические характеристики, связанные с поведением.

Некоторые биометрические характеристики можно использовать для удостоверения личности или для проверки информации о личности, так как эти характеристики уникальны для конкретного человека [1].

К биометрическим характеристикам относятся:

- линии пальцев;
- форма руки;
- линии сетчатки глаза;
- параметры звука;
- черты лица;
- термограмма лица (схема расположения кровенос сосудов на лице);
- форма подписи.

Биометрическая идентификация используется для защиты от несанкционированного доступа в здания, банкоматы, компьютеры и т.д. Этот подход значительно надежнее, чем применение паролей, ПИН-кодов, смарт-карт, жетонов и т.д.

Применение биометрических технологий разнообразно: доступ к рабочим местам и сетевым ресурсам, защита информации, обеспечение доступа к определенным ресурсам и вопросы безопасности. Ведение электронного бизнеса и электронных правительственных дел возможно только после соблюдения определенных процедур по идентификации личности. Биометрические технологии используются в области безопасности банковских обращений, инвестирования и других финансовых перемещений, а также в розничной торговле, охране правопорядка, вопросах охраны здоровья, сфере социальных услуг. Обращая внимание на стадии развития биометрических технологий, можно заметить, что распознавание человеческого лица по фотопортрету сегодня одна из актуальнейших тем.

Увеличение количества изданий и научных исследований в последние десятилетия в этой сфере показывает актуальность данной проблемы:

- в первую очередь это связано с развитием компьютерных технологий;
- существование множества коммерческих структур и рост социальных проблем увеличивают потребность в биометрических технологиях. Так как автоматическая идентификация личности тесно связана с вышеуказанными проблемами;
- идентификацию личности на основе изображения можно применить для удостоверений личности (паспорта, водительских прав), а также для обеспечения безопасности информации от случайного или преднамеренного доступа лиц, не имеющих на это права, для контроля криминальных происшествий, в банковском секторе и т.д.

Это делается, во-первых, для противодействия международной организованной преступности и терроризму, борьбы с незаконной миграцией, подделкой документов. Кроме того, биометрическая идентификация позволяет упростить саму процедуру паспортного контроля.

Постановка задачи

В настоящее время имеется значительное количество работ, посвященных исследованиям распознавания людей по фотопортретам, а также некоторые рекомендации разработчиков систем идентификации личности по фотографиям. При этом под термином «фотопортрет» подразумевается цифровое изображение лица человека в фас без элементов одежды, украшений, солнечных очков и т.д., которые могут закрывать или искажать части его лица.

Люди значительно отличаются друг от друга по размерам и расположению таких лицевых элементов, как глаза, брови, нос, уши, рот и т.д. Поэтому первое решение проблемы автоматической идентификации человека по изображению лица основывалось на выборе и сравнении некоторых антропометрических особенностей лица. Эта методика годами использовалась в экспериментальной криминалистике. Данный способ был особенно эффективен в случае, когда у человека не было иной фотографии, кроме той, что в паспорте (контроль паспорта) [2].

Работа [3] посвящена распознаванию человека на основе фотопортрета. Для распознавания лица на основе фотопортрета авторы данной работы разработали алгоритм вычисления геометрических характеристик человеческого лица на основе выбранных 19 антропометрических точек лица и расстояний между этими точками. Эти точки должны быть прочными к таким изменениям, как ракурс, свет, мимика, косметика, возраст и т.д. Одним из основных методов распознавания является определение геометрических характеристик лица. Суть метода состоит в том, что на лице отмечаются ключевые точки. Затем определяются расстояния (геометрические характеристики) между соответствующими ключевыми точками. Ключевые точки на лице человека могут находиться в различных местах. Например, на краю глаза, губе, мочке уха, носу и т.д.

В работе [4] был разработан алгоритм для автоматического добавления в базу данных вычисленных геометрических характеристик, а также поиск лица человека и идентификация на основе этих характеристик. Описана созданная система автоматизированной биометрической идентификации (САБИ) «Распознавание». Первоначально для САБИ «Распознавание» была организована база данных. В нее были включены фотографии 102 человек. Кроме того, в базу данных для каждого лица были включены индивидуальные данные (имя, фамилия, отчество, день рождения, цвет глаз, рост). Кроме этой информации, для каждой личности в базе сохраняется 18 вычисленных признаков, определяющих геометрическую характеристику лица [3].

Систему «Распознавание» можно применить для удостоверения личности (паспорта, водительских прав), а также для обеспечения безопасности информации от случайного или преднамеренного доступа лиц, не имеющих на это права, для контроля криминальных происшествий, в банковском секторе и т.д.

Определение точности идентификации имеет важное значение для распознавания лиц на основе их фотопортрета. Определение доверительного интервала геометрических характеристик является основным фактором для распознавания лиц на основе их фотопортрета.

В этой статье рассматриваются вопросы по определению доверительных интервалов геометрических характеристик для распознавания лиц на основе их фотопортрета.

Доверительный интервал – интервал, который строится по данным выборочного исследования для оценивания параметра генеральной совокупности. Преимущества испытаний, которые проводились по определению доверительного интервала, нижеследующие:

- 1) незначительное влияние внешних факторов на основной процесс во время испытаний;
- 2) более быстрое и точное испытание;
- 3) проведение нужного количества наблюдений в испытаниях [5].

Для определения доверительного интервала были использованы геометрические характеристики, вычисленные с помощью САБИ «Распознавание» [4]. В первую очередь были проведены опыты над геометрическими характеристиками 102 человек из базы данных. Эти геометрические характеристики, присущие 102 людям, делятся на 18 кластеров по одинаковым признакам.

При работе САБИ «Распознавание» могут быть два случая:

- значение геометрических характеристик может попасть в найденный для них доверительный интервал. В этом случае система продолжает свою работу и переходит к следующему этапу;

- некоторые значения геометрических характеристик могут не попасть в найденный для них доверительный интервал. В этом случае система сообщает об этом пользователю, для определения геометрических характеристик заново вычисляются соответствующие расстояния и процесс продолжается.

В общем для вычисления доверительных интервалов было использовано $102 \cdot 18$ количество значений геометрических характеристик [4]. Кластеры по геометрическим характеристикам обозначены как $Ns_1, Ns_2, \dots, Ns_{18}$.

Для примера была показана таблица 1 со значениями геометрических характеристик 102 человек одного кластера (Ns_1).

Количество таких таблиц 18.

Для определения доверительного интервала используется метод Стьюдента [5].

Для определения доверительного интервала, используемого в идентификации человека на основе фотопортрета с помощью метода Стьюдента, прежде всего должно быть исследовано подчинение закону нормального распределения отбора, состоящего из геометрических характеристик.

Определение нормального распределения имеет большое значение по разным причинам. В большинстве случаев оно считается самой лучшей сходимостью к функции. Статистическое распределение многих природных явлений считается нормальным. Например, определение веса или объема товаров, измерение роста людей, проходящих медицинский осмотр, и т.д.

Таблица 1

Значения геометрических характеристик 102 человек одного кластера

Ns1					
№		№		№	
1	2,7	40	2,13333	79	1,8125
2	2,14286	41	1,88235	80	1,75
3	2,0625	42	1,875	81	2
4	2,21429	43	2,38462	82	2,07143
5	2,28571	44	2,45455	83	1,94118
6	2,92857	45	2,05882	84	2,30769
7	2,4	46	2,53846	85	2,25
8	2,66667	47	2,30769	86	2,53846
9	2	48	1,63158	87	2,17647
10	2,15385	49	2,10526	88	2,30769
11	2,69231	50	2,05882	89	2,30769
12	3	51	1,76471	90	2,125
13	2,05882	52	2	91	2,46667
14	2,42857	53	2,45455	92	2
15	2,88889	54	2,33333	93	2,23529
16	2,54545	55	3,36364	94	2
17	2,61538	56	2,06667	95	2,5
18	2,17647	57	2,2	96	2,1875
19	3,44444	58	2,16667	97	1,77778
20	1,94444	59	2,61538	98	2,21429
21	2,35714	60	2,13333	99	2,42857
22	2,26667	61	2,4375	100	2,4375
23	2,41667	62	1,8	101	2,26667
24	2,28571	63	2,76923	102	2,15
25	2,33333	64	2,15385		
26	3	65	2,21053		
27	2	66	3,3		
28	1,94444	67	2,625		
29	2,07143	68	2,05263		
30	2,0625	69	1,86667		
31	1,8125	70	2,13333		
32	2,23529	71	2,13333		
33	2,84615	72	2,5625		
34	2,66667	73	1,75		
35	2,91667	74	2,2		
36	2,11111	75	2		
37	2,06667	76	2,5		
38	2,64286	77	1,85714		
39	2,33333	78	1,94444		
	$\bar{X} = 2,2823$	$\Sigma(x_i - \bar{X}) = 0,2356$	$\Sigma(x_i - \bar{X})^2 = 12,7397$		

Нормальное распределение определяется по двум основным параметрам: средние и стандартные ошибки [6].

Среди значений геометрической характеристики была создана конечная совокупность выборов из 17 случайных опций значений. Среднее значение вычисляется на основе конечной совокупности выборов. В результате проведенных вычислений среднее значение равнялось $\mu = 2,45787$.

Выбор распределения был представлен в конечном выборе на основе двух значений и на их основе было вычислено среднее значение за каждое значение выбора конечного значения. Частота средних значений и их сумма вычислены и показаны в таблице 2.

Таблица 2

Частота средних значений и их сумма

Частота f	$f\bar{x}$
$\sum f = 135$	$\sum f\bar{x} = 331,66$

Математическое ожидание на основе распределения выбора было рассчитано по следующей формуле:

$$E(\bar{x}) = \frac{\sum f\bar{x}}{\sum f} = \frac{331,662}{135} = 2,46$$

Учитывая, что среднее значение выбора данных $\mu = 2,45787 \approx 2,46$, то

$$E(\bar{x}) = \mu = 2,46.$$

Стандартная ошибка для основной нормально распределенной совокупности была определена по следующей формуле [6].

$$SE_{(\bar{x})} = \sqrt{\frac{(N-n)\sigma^2}{(N-1)n}}, \tag{1}$$

где σ^2 показывает основные дисперсии.

Для значения основных геометрических характеристик:

$$\sigma^2 = 0,3551, \quad N = 102, \quad n = 2.$$

Если мера основной совокупности больше, чем выбор совокупности ($n/N = 0,0196 \leq 0,05$) тогда

$$\sqrt{\frac{(N-n)}{(N-1)}} = 0,995 \approx 1,$$

и стандартная ошибка равна:

$$SE_{(\bar{x})} = \sqrt{\frac{\sigma^2}{n}} = 0,2511.$$

Как видно из расчетов, нормальное распределение значений геометрических характеристик было доказано (рис.1).

Доверительный интервал имеет вид:

$$\bar{X} - t_{\beta} m_{\bar{X}} \leq \tilde{X} \leq \bar{X} + t_{\beta} m_{\bar{X}}, \tag{2}$$

где \bar{X} – среднее значение генеральной совокупности; $m_{\bar{X}}$ – ошибка среднего арифметического, вычисляемая:

$$m_{\bar{X}} = \pm \frac{\sigma}{\sqrt{k}}, \quad (3)$$

$$\sigma = \pm \sqrt{\frac{\sum (x_i - \bar{X})^2}{k-1}}, \quad (4)$$

где t_{β} – число Стьюдента, выбранное из таблицы Стьюдента; $x_i (i = \overline{1, k})$ – показывает значение генеральной совокупности.

Используя формулу (2) для кластера Ns1, получим:

$$k = 102,$$

$$\sigma = 0,3551,$$

$$\beta = 95\%, \quad t_{\beta} = 1,98,$$

$$m_{\bar{X}} \approx \pm 0,0353,$$

$$2,21232796 \leq \bar{X} \leq 2,3522703.$$

Оценки t_{β} были взяты из таблицы Стьюдента на основе оценок k и β .

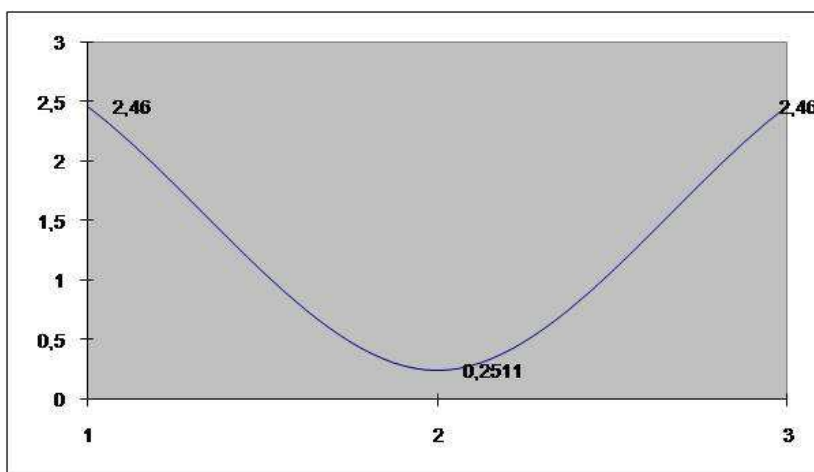


Рис. 1. График для нормальных распределений значений геометрических характеристик

По этим правилам были вычислены оценки доверительного интервала по остальным кластерам и указаны в таблице 3. На рис.2 указан график доверительных интервалов геометрических характеристик человеческого лица для идентификации человека на основе фотопортрета.

Оценки доверительного интервала

Ns1	Ns2	Ns3
2,21 ≤ 2,28 ≤ 2,35	0,05 ≤ 0,42 ≤ 0,79	5 ≤ 5,76 ≤ 6,51
Ns4	Ns5	Ns6
0,18 ≤ 0,22 ≤ 0,62	0,9 ≤ 1,13 ≤ 1,35	0,49 ≤ 0,79 ≤ 1,08
Ns4	Ns5	Ns6
0,18 ≤ 0,22 ≤ 0,62	0,9 ≤ 1,13 ≤ 1,35	0,49 ≤ 0,79 ≤ 1,08
Ns7	Ns8	Ns9
1,94 ≤ 2,03 ≤ 2,11	0,25 ≤ 0,59 ≤ 0,92	1,21 ≤ 1,23 ≤ 1,25
Ns10	Ns11	Ns12
0,33 ≤ 0,665 ≤ 0,98	1,6061 ≤ 1,6198 ≤ 1,6334	3,215 ≤ 3,2434 ≤ 3,271
Ns13	Ns14	Ns15
0,2544 ≤ 0,29 ≤ 0,33	0,01 ≤ 0,293437745 ≤ 0,74	5 ≤ 5,76 ≤ 6,51
Ns16	Ns17	Ns18
0,18 ≤ 0,22 ≤ 0,62	0,33 ≤ 0,66 ≤ 0,98	1,6061 ≤ 1,6198 ≤ 1,6334

Используя нечеткое вычисление, можно определить доверительный интервал значения геометрических характеристик [7].

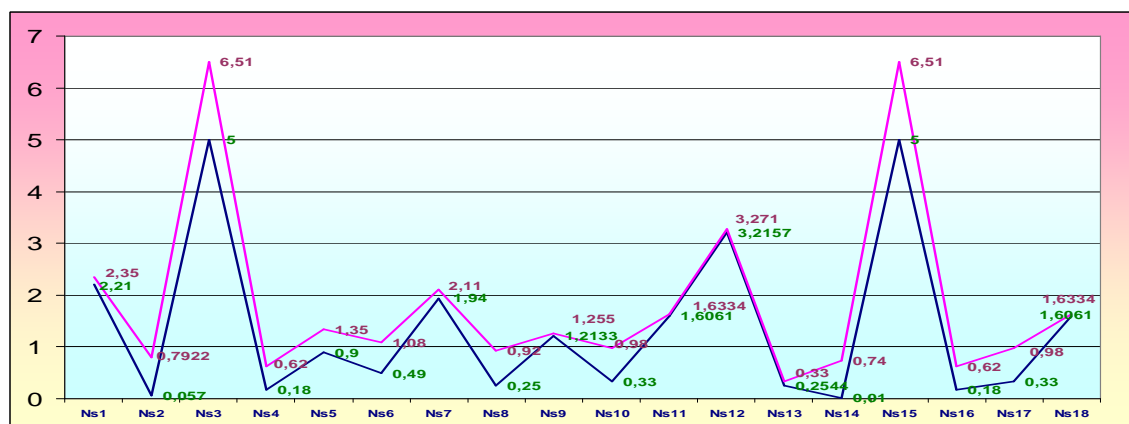


Рис. 2. График доверительных интервалов геометрических характеристик человеческого лица для идентификации человека на основе фотопортрета

Следует отметить, что на основе исследований, проведенных авторами, был установлен реальный интервал значений расстояний между антропометрическими точками человеческого лица. Чтобы найти интервал значений геометрических характеристик в соответствии с теми же расстояниями с помощью нечеткого расчета была использована следующая формула.

$$S_i^* / S_{i+1} \leq P_i^* \leq S_i / S_{i+1}^* \quad i = 1, n-1 . \quad (5)$$

Здесь реальное максимальное значение i – антропометрического расстояния человеческого лица было указано как S_i^* ($i = \overline{1, n}$), а реальная минимальная стоимость как S_i ($i = \overline{1, n}$), значение геометрических характеристик было указано как P_i^* ($i = \overline{1, n}$).

Реальный интервал значений геометрических характеристик, определенный с помощью (5), показан в таблице 4.

Таблица 4

Реальный интервал значений геометрических характеристик

	Minimum	Maksimum
P_1^*	1,42	2,8
P_2^*	0,04	0,8
P_3^*	2,8	6,7
P_4^*	0,1	1
P_5^*	0,352	1,79
P_6^*	0,4	1,29
P_7^*	0,68	2,22
P_8^*	0,1	1,6
P_9^*	1	2,031
P_{10}^*	0,2	3,33
P_{11}^*	0,21	1,73
P_{12}^*	0,33	3,56
P_{13}^*	2,4	0,54
P_{14}^*	0,01	1
P_{15}^*	2,8	6,8
P_{16}^*	0,1	1
P_{17}^*	0,23	1,33
P_{18}^*	0,21	1,8

При сравнении значений P_i^* ($i = \overline{1, n}$) в таблице 4 со значениями P_i ($i = \overline{1, n}$) в таблице 3 будут обеспечены следующие условия:

$$\text{Max} (P_i) \leq \text{Max} (P_i^*) \quad i = \overline{1, n} ,$$

$$\text{Min} (P_i) \geq \text{Min} (P_i^*) \quad i = \overline{1, n} .$$

Из этого следует, что значения P_i не превышают пределы интервала реального значения.

Заключение

Для определения доверительного интервала геометрических характеристик для распознавания лиц на основе их фотопортрета использовался метод Стьюдента.

Была доказана необходимость нормального распределения геометрических характеристик для определения доверительного интервала геометрических характеристик,

используемых для идентификации человеческого лица на основе фотопортрета, с помощью метода Стьюдента.

Определение и использование доверительного интервала ускоряют процесс распознавания лиц на основе их фотопортрета и способствуют более эффективной работе системы идентификации.

Литература

1. Болл Р.М., Коннел Дж.Х., Панканти Ш., Ратха Н.К., Сеньор Э.У. Руководство по биометрии. Москва, Техносфера, 2007, 368 с.
2. Самаль Д.И., Старовойтов В.В. Подходы и методы распознавания людей по фотопортретам // Минск, 1998, 54 с. (Препринт / Ин-т технической кибернетики НАН Беларуси, №8).
3. Кязимов Т.Г., Махмудова Ш.Дж. Система компьютерного распознавания людей по фотопортретам // Информационные технологии, Москва, 2009, №1, с.13–16.
4. Кязимов Т.Г., Махмудова Ш.Дж. Автоматизированная система по распознаванию человека на основе идентификационных геометрических характеристик изображения лица // Телекоммуникации, Москва, 2008, №11, с. 22–25.
5. Орлов А.И. Математика случая: Вероятность и статистика – основные факты: Учебное пособие. Москва, МЗ-Пресс, 2004, 110 с.
6. Эддоус М., Стэнфилд Р. Методы принятия решений. М.: Аудит, ЮНИТИ, 1997, 590 с.
7. Мацеевский С.В. Нечеткие множества. Учебное пособие. Калининград: Изд-во КГУ, 2004, 176 с.

UOT 004.932.001.57

Kazimov Tofiq H.¹, Mahmudova Şafəqət C.²

AMEA İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan

¹tofig@mail.ru, ²shafagat_57@mail.ru

Fotoşəkil əsasında identifikasiya sistemində tanınma alqoritminin təkmilləşdirilməsi haqqında

Fotoşəkil əsasında şəxsin identifikasiyası üçün insan sifətinin həndəsi xarakteristikalarının inam intervallarının müəyyən edilməsi məsələsinə baxılmışdır. İdentifikasiyanın dəqiqliyinin müəyyən edilməsinin əhəmiyyəti qeyd olunmuşdur. Həndəsi xarakteristikaların inam intervallarının müəyyən edilməsi üçün Student metodundan istifadə olunmuşdur. Bu məsələnin həlli identifikasiya sisteminin effektivliyinin artmasına gətirib çıxarır.

Açar sözlər: *biometrik texnologiyalar, təsvir, həndəsi xarakteristikalar, inam intervalı, tanınma, Student metodu.*

Tofiq H. Kazimov¹, Shafagat J. Mahmudova²

Institute of Information Technology, National Academy of Sciences of Azerbaijan

¹tofig@mail.ru, ²shafagat_57@mail.ru

About improving of algorithm of recognition in identification system on the basis of the photo-portrait

Problems of estimating the confidence intervals of geometrical characteristics of human face for people identification based on photo portraits are considered. The importance of a given problem is emphasized at the identification precision determination. To define the confidence intervals of geometrical characteristics, the well-known Student method is used. The solution of the present problem results in the identification system's efficiency increase.

Key words: *biometrical technologies, image, geometrical characteristics, confidence interval, recognition, Student method.*