

УДК 004.93

МНОГОМОДАЛЬНАЯ ИДЕНТИФИКАЦИЯ ЛИЧНОСТИ ПО ФОРМЕ ЛАДОНИ И ГОЛОСУ

© Бакина И.Г., Местецкий Л.М.

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМ. М.В. ЛОМОНОСОВА
ФАКУЛЬТЕТ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И КИБЕРНЕТИКИ
ЛЕНИНСКИЕ ГОРЫ, 1, Г. МОСКВА, 119234, РОССИЯ

E-MAIL: irina_msu@mail.ru, L.Mest@ru.net

Abstract. The paper demonstrates the method of improving the quality of the one-mode method for personal identification by a palm shape on the base of combining it with a less reliable recognition method (by uttered word). Theoretical and experimental estimates proving the efficiency of the proposed method are given.

ВВЕДЕНИЕ

На данный момент существует достаточно много одномодальных методов биометрической идентификации человека: анализ радужной оболочки глаз, дактилоскопия, распознавание по форме ладони, подписи, геометрии лица, голосу человека и т.д. Однако, в любом из методов идентификации есть ошибки, и задача повышения качества распознавания остается актуальной. Один из возможных вариантов решения этой проблемы — комплексный анализ по нескольким биометрическим признакам.

В данной работе рассматривается метод двумодальной идентификации личности, основанный на распознавании человека по форме ладони и голосу этого человека. Их интеграция рассматривается с точки зрения комбинирования классификаторов, когда каждый из методов обучается отдельно по своей модальности. По терминологии [3] под модальностью понимается набор признаков, относящийся к одной группе измерений (например, параметры голоса или параметры формы ладони). Распознавание по форме ладони обладает относительно высоким, но недостаточным уровнем правильной идентификации. Однако, как показывают теоретические оценки и проведенные эксперименты, в совокупности с менее надежным методом распознавания, таковым является распознавание по голосу, качество идентификации может быть улучшено.

1. РАСПОЗНАВАНИЕ ЧЕЛОВЕКА ПО ФОРМЕ ЛАДОНИ

Существует несколько способов описания изображений ладоней, на основе которых строятся различные критерии их сравнения. Один из таких критериев рассматривается в работе [2]. Он базируется на так называемом циркулярном разложении ладони (Рис.1а), описывающем ширину пальцев. Другой критерий, основанный на сравнении формы гибких объектов (Рис.1б), предложен в [3]. Здесь измеряется площадь симметрической разности объектов при их наложении друг на друга и "шевелении", при котором обеспечивается максимальное совпадение формы. Использование гранично-скелетного представления формы объектов позволяет построить еще

несколько количественных мер для сравнения формы ладоней. В работе [4] рассмотрены критерии, основанные на измерении длины пальцев, ширины ладони, ширины пальца вдоль его оси, а также график искривления самой этой оси. Ширина пальца описывается в виде зависимости радиуса окружности, вписанной в силуэт пальца, от расстояния между ее центром и кончиком пальца. Искривление пальца представляется в виде функции, описывающей отклонение серединной оси пальца от прямой линии.



Рис. 1. (а) Растровый образ, его скелет и циркулярное разложение; (б) Представление ладони в виде гибкого объекта и сравнение таких объектов.

В данной работе за основу был взят метод распознавания человека по форме ладони, изложенный в [9]. Он включает в себя следующие основные шаги:

1. Построение гранично-скелетного представления тестового изображения ладони.
2. Сравнение полученного представления с эталонными образцами из имеющейся базы (сравнение происходит по нескольким критериям).
3. Определение Парето-оптимальных эталонных образцов ладоней ("ближайших").
4. Классификация среди "ближайших".

Для интеграции данного метода с распознаванием человека по голосу выполнение всех четырех шагов не требуется, достаточно определить лишь множество "ближайших", не проводя дальнейшую классификацию. Фактически, рассматривается неполный вариант метода, выходом которого является список персон, чьи изображения ладони попали в число "ближайших", и число таких попаданий.

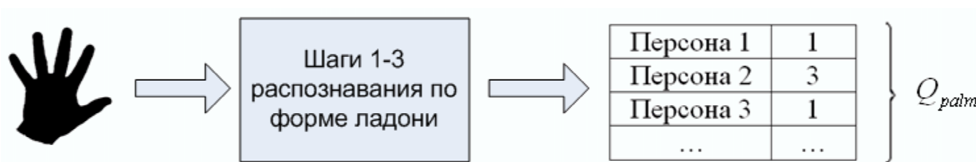


Рис. 2. Схема работы неполного метода распознавания человека по форме ладони.

Для удобства назовем полученный список персон Q_{palm} . Заметим, что в случае, когда метод отклоняет предъявленное изображение ладони, $Q_{palm} = \emptyset$.

В качестве метода распознавания по форме ладони мог быть рассмотрен любой из перечисленных выше. Единственное условие — метод должен работать с базой эталонных изображений ладоней и на выходе выдавать список, составленный из наиболее похожих изображений. Выбор метода [9] обусловлен тем, что он позволяет анализировать сразу несколько критериев, и потому множество "ближайших" изображений может быть определено более корректно.

Эксперименты по распознаванию ладоней с использованием этих критериев показали, что они довольно хорошо позволяют определить круг наиболее похожих образцов из базы эталонов. Вместе с тем, ошибки классификации остаются достаточно большими. Особенно это сказывается на ошибках ложного узнавания (*False Acceptance Rate*) в случае, когда объем базы эталонов достаточно велик.

2. РАСПОЗНАВАНИЕ ЧЕЛОВЕКА ПО ГОЛОСУ

Распознавание человека по голосу рассматривается в данной статье с точки зрения распознавания человека по произнесенному парольному слову. Каждая персона характеризуется парольным словом, при этом парольные слова разных людей могут совпадать. Например, в качестве парольного слова может использоваться фамилия человека. Имеется база эталонных записей парольных слов: аудиозапись парольного слова каждого человека хранится в отдельном файле в формате ".wav".

В данной работе не ставится задача распознавания человека по произносимой речи. Запись парольного слова представляет собой достаточно короткий звуковой сигнал, поэтому методы, основанные на построении модели диктора (например, метод квантования векторов [5, 6]), здесь неприменимы.

Сравнение двух записей парольных слов осуществляется на основе метода динамического выравнивания [7]. С этой целью дискретный звуковой сигнал разбивается на перекрывающиеся фреймы длины $N_t = 512$ измерений с перекрытием в 100 измерений. Дальнейшей обработке подвергаются лишь фреймы, амплитуда сигнала в которых больше заданного порога. К каждому из полученных фреймов применяется оконная функция Хэмминга, затем фрейм описывается вектором признаков $\vec{f} = (f_1, f_2, \dots, f_{12})^T$, составленным из первых 12 кепстральных коэффициентов (0-ой коэффициент исключается из рассмотрения, так как он содержит информацию только об энергии сигнала). Кепстральные коэффициенты задаются следующим образом: если известен дискретный звуковой сигнал $s(t)$, а N_t — число временных отсчетов в записи, то n -ый кепстральный коэффициент может быть вычислен по формуле:

$$f_n = \text{Re} \left[\frac{1}{N_t} \sum_{l=0}^{N_t-1} \left\{ \ln \left(\sum_{t=0}^{N_t-1} \left| s(t) \exp(-i \frac{2\pi l t}{N_t}) \right| + 1 \right) \right\} \exp(i \frac{2\pi l n}{N_t}) \right], n = \overline{0, N_t}$$

Расстояние между фреймами F_1 и F_2 , описываемых векторами-признаками $\vec{f}^1 = (f_1^1, f_2^1, \dots, f_{12}^1)^T$ и $\vec{f}^2 = (f_1^2, f_2^2, \dots, f_{12}^2)^T$ соответственно, определяется по формуле:

$$D(F_1, F_2) = \sqrt{\sum_{i=1}^{12} (f_i^1 - f_i^2)^2}$$

Как уже упоминалось, в качестве меры близости двух звуковых сигналов рассматривается расстояние, полученное по методу динамического выравнивания.

В работе также рассматривалось признаковое описание звукового сигнала на основе спектральных коэффициентов, однако для данного метода они оказались недостаточно информативными. Их использование приводило к значительному увеличению ошибок распознавания по сравнению с случаем, когда в качестве признакового описания рассматривались кепстральные коэффициенты.

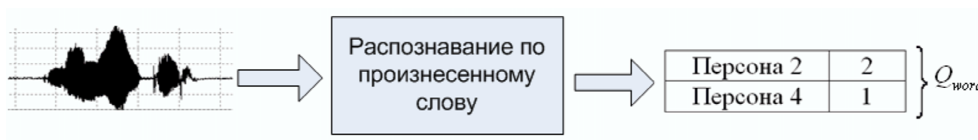


Рис. 3. Схема работы метода распознавания человека по произнесенному слову ($k = 3$).

Общая схема работы метода заключается в следующем. При произнесении парольного слова звуковой сигнал сравнивается со всеми эталонными образцами из имеющейся базы, после чего они упорядочиваются по возрастанию меры близости с тестовым образцом. Из полученного списка отбираются k наиболее похожих эталонных записей, и строится множество персон Q_{word} , аналогичное Q_{palm} . В отличие от распознавания по форме ладони множество Q_{word} не может быть пустым (так как отбор похожих персон происходит всегда).

Эксперименты показывают, что при сравнении тестового образца с базой эталонов, вероятность того, что персона попадет в начало упорядоченного списка, весьма высока. Это позволяет рассчитывать на низкий уровень ошибок ложного отказа (*False Rejection Rate*). Однако в случае, когда персона не представлена в базе, для нее все равно найдется достаточно похожий эталон. Это происходит потому, что две записи парольного слова одного человека зачастую оказываются менее похожими, чем записи парольных слов двух разных людей (с точки зрения меры близости, рассмотренной выше). Поэтому не удастся уверенно классифицировать ситуацию, когда тестовый образец принадлежит "чужой" персоне. Это приводит к очень высокому уровню ошибок ложного узнавания.

3. КОМБИНИРОВАНИЕ КЛАССИФИКАТОРОВ

Целью комбинирования классификаторов является преодоление указанных недостатков каждого из описанных методов.

Для распознавания человек предъявляет системе свою ладонь и произносит парольное слово. Отдельно по каждому из этих биометрических признаков происходит распознавание, и формируются два множества Q_{palm} и Q_{word} , описанные выше. Далее ищется пересечение этих множеств $Q = Q_{palm} \cap Q_{word}$, при этом возможны следующие варианты:



Рис. 4. Комбинирование классификаторов: (a) $|Q| = 1$; (b) $|Q| > 1$.

1. $Q_{palm} = \emptyset$, и в этом случае человек признается как "чужой".
2. Ситуация $|Q| = 1$ представлена на Рис.4а. Человек идентифицируется как персона, попавшая в пересечение.
3. В случае $|Q| > 1$ считается общее число попаданий в группу "ближайших" по ладонке и по голосу одновременно. Например, для ситуации на Рис.4б для Персоны 2 эта величина равна $n_1 = 3 + 2 = 5$, а для Персоны 3 - $n_2 = 1 + 1 = 2$. Так как $n_1 > n_2$, то человек идентифицируется как Персона 2. В случае, когда для нескольких персон эти величины совпадают, система признает человека как "своего", но идентифицировать отказывается.

Обозначим через FRR_{palm} и FAR_{palm} соответственно *False Rejection Rate* и *False Acceptance Rate* для неполного метода распознавания человека по форме ладони, а через FRR_{word} и FAR_{word} - по произнесенному слову. Пусть u_o — "свой" человек, а u_s — "чужой". Для рассмотренных методов распознавания эти величины определяются следующим образом:

1. $FRR_{palm} = P(\{u_o \notin Q_{palm}\})$;
2. $FAR_{palm} = P(\{Q_{palm} \neq \emptyset \mid u_s\})$;
3. $FRR_{word} = P(\{u_o \notin Q_{word}\})$;
4. $FAR_{word} = P(\{Q_{word} \neq \emptyset \mid u_s\}) = 1$, так как множество всегда непустое.

Оценим FAR и FRR для рассматриваемой двумодальной системы. Как отмечалось выше, человек определяется "чужим" в случае, когда он не попал в пересечение списков $Q = Q_{palm} \cap Q_{word}$ персон для обоих методов. Поэтому:

$$\begin{aligned} FRR &= P(\{u_o \notin Q\}) = P(\{u_o \notin Q_{palm} \cap Q_{word}\}) = \\ &= 1 - P(\{u_o \in Q_{palm}\}) \cdot P(\{u_o \in Q_{word}\}) = \\ &= FRR_{palm} + FRR_{word} - FRR_{palm} \cdot FRR_{word} \quad (1) \end{aligned}$$

"Чужой" признается "своим" в случае, когда множество оказывается непустым, оценка для FAR имеет следующий вид:

$$FAR = FAR_{palm} \cdot \frac{k}{n} \quad (2)$$

В этой формуле k — число отбираемых эталонных записей парольных слов, а n — число персон в базе. Здесь предполагается, что вероятности попадания парольных слов в k "ближайших" равны, определяются лишь количеством парольных слов в списке и не зависят от самих слов.

4. ЭКСПЕРИМЕНТЫ

Рассматривалась группа людей из 24 человек. Для каждого из них 12 раз производились снимок ладони и запись парольного слова. Полученные пары вида "изображение – запись" разделялись на две части: эталоны и контроль. К эталонам были отнесены 7 из сформированных пар, оставшиеся 5 — к контролю.

При оценивании FRR рассматривались все эталонные и контрольные пары: каждая контрольная пара подавалась на вход программе один раз, после этого подсчитывалось число пар, ошибочно отнесенных к классу "чужой".

При оценивании величины FAR также рассматривались все эталонные и контрольные пары. Однако сравнение предъявляемой контрольной пары происходило лишь с эталонными парами, не относящимися к тому же человеку. В эксперименте подсчитывалось число пар, ошибочно пропущенных системой.

Результаты эксперимента представлены в *Таблице 4*. В последних двух колонках содержатся теоретические и практические оценки FRR и FAR . Теоретические оценки получены по формулам (1) и (2).

Таблица 1. Результаты эксперимента.

	Распознавание по ладонке	Распознавание по слову	Теоретическая оценка	Экспериментальная оценка
FRR	2.5%	5.8%	8.2%	8.33%
FAR	30%	100%	3.75%	3.33%

Из таблицы видно, что использование менее надежного метода распознавания позволяет существенно снизить FAR , правда, за счет некоторого увеличения FRR .

ЗАКЛЮЧЕНИЕ

Проведенные эксперименты и полученные теоретические оценки показывают, что предложенный в работе метод комбинирования классификаторов действительно позволяет повысить надежность распознавания. Прослеживается значительное снижение величины FAR , что согласуется с полученными теоретическими оценками. Наблюдаемое при этом увеличение FRR оказывается не настолько весомым.

Конечно, полученные оценки FAR и FRR все еще остаются значительными. Это объясняется тем, что исходные одномодальные методы распознавания, по форме ладони и произнесенному слову, изначально обладают высоким уровнем ошибок ложного узнавания и ложного отказа. В данной работе большее внимание уделялось изучению не самих методов распознавания, а проверке гипотезы о том, что предложенный подход к их комбинированию приводит к улучшению качества идентификации. В дальнейшем планируется обратиться к анализу указанных методов и поработать над увеличением надежности каждого из них в отдельности.

Работа выполнена при поддержке РФФИ, гранты 08-01-00670 и 08-07-00305-а.

СПИСОК ЛИТЕРАТУРЫ

1. *Mestetskiy L.M.* Shape comparison of flexible objects similarity of palm silhouettes // Proceedings of the 2nd International conference on computer vision theory and applications (VISAPP 2007), Volume IFP/IA, Barcelona, Spain, 2007, P. 390-393.
2. *Mestetskiy L.M., Semenov A.B.* Palm shape comparison based on fat curves // Proceedings of 7th International conference on Pattern recognition and image analysis: new information technologies, St.Petersburg, 2004, P. 788-791.
3. *Местецкий Л.М.* Сравнение изображений гибких объектов на основе нормализации // Труды 17 международной конференции ГРАФИКОН-2007, Москва, ВМК МГУ, 2007, С. 203-210.
4. *Mekhedov I.S., Mestetskiy L.M.* Construction of a classifier for person biometric identification using a palm shape // Proceedings of the Ninth International conference on Pattern recognition and information processing (PRIP'2007), Volume I, Minsk, Belarus, May 22-24, 2007, P. 290-294.
5. *Evgeny Karpov* Real-Time Speaker Identification // University of Joensuu, Department of Computer Science, Master's Thesis.
6. *Tomi Kinnunen, Ismo Karkkainen and Pasi Franti* Is Speech Data Clustered? — Statistical Analysis of Cepstral Features // In EUROSPEECH-2001, 2627-2630.
7. *Sergios Theodoridis, Konstantinos Koutroumbas* Pattern Recognition // second edition, Elsevier 2003.
8. *Татарчук А.И., Елисеев А.П., Моттль В.В.* Комбинирование классификаторов и потенциальных функций в многомодальном распознавании образов // Доклады Всероссийской конференции ММРО-13, 2007, С. 220-222.

Статья поступила в редакцию 25.04.2008