

Сведения об авторах

- Татьяна Анатольевна Маркина** — аспирант; Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кафедра вычислительной техники; E-mail: tmark812@mail.ru
- Андрей Юрьевич Щеглов** — д-р техн. наук, профессор; Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кафедра вычислительной техники; E-mail: info@npp-itb.spb.ru

Рекомендована кафедрой
вычислительной техники

Поступила в редакцию
23.12.13 г.

УДК 004.627

А. А. ТРОПЧЕНКО

МЕТОДЫ ПОВЫШЕНИЯ РОБАСТНОСТИ РАСПОЗНАВАНИЯ В МУЛЬТИМОДАЛЬНЫХ БИОМЕТРИЧЕСКИХ СИСТЕМАХ

Рассмотрены методы повышения робастности (устойчивости) распознавания в мультимодальных биометрических системах идентификации личности.

Ключевые слова: распознавание личности, мультимодальные системы распознавания, объединение признаков, робастность.

Для аутентификации и идентификации человека применяется биометрика — область знаний, использующая индивидуальные биологические особенности. Биометрические данные включают множество признаков (модальностей): отпечаток пальца, изображение лица, речь, геометрия кисти руки и ушной раковины, сетчатка глаза, подпись, динамика нажатия клавиш, походка, физиологические сигналы (электрокардиограмма) и т.д. У использования каждого признака есть свои преимущества и ограничения с точки зрения точности, устойчивости и удобства работы. Например, использование сетчатки обеспечивает высокую точность и устойчивость распознавания, но требует дорогого оборудования и существенных затрат времени.

Идентификация в динамике гораздо более сложная задача, особенно когда велико число зарегистрированных в системе пользователей. Динамические идентификационные системы на основе анализа аудиосигналов достигают высокой производительности, когда высоко отношение сигнал—шум (SNR) распознаваемого отрезка (текста). Однако устойчивость быстро ухудшается, когда SNR набора тестов уменьшается [1]. Для исследования устойчивости (робастности) распознавания личности при различных уровнях искажений был проведен ряд экспериментов.

На рис. 1 продемонстрировано последовательное снижение качества тестового изображения для аудиовизуальной базы данных XM2VTS, уровень вносимых искажений определяет параметр QF (Quality Factor).



QF=50



QF=14



QF=2

Рис. 1

База данных XM2VTS содержит видеоролики, на которых 295 человек произносят тестовое предложение (“Joe took fathers green shoe bench out”) 4 раза с интервалами в месяц. Это предложение считается фонетически сбалансированным для английского языка.

Результаты экспериментов по распознаванию речи при различных уровнях зашумленности сигнала приведены на рис. 2 (P — вероятность правильного распознавания). Аудиосигнал был предварительно обработан для увеличения мощности в области более высоких частот с использованием фильтра $H(z) = 1/(1 - 0,97z^{-1})$. Затем сигнал был разделен на участки (фреймы) длиной 20 мс с перекрытием в 10 мс, что обеспечивает частоту аудиокадров 100 Гц. Далее из каждого фрейма извлекаются мел-частотные кепстральные коэффициенты (MFCC).

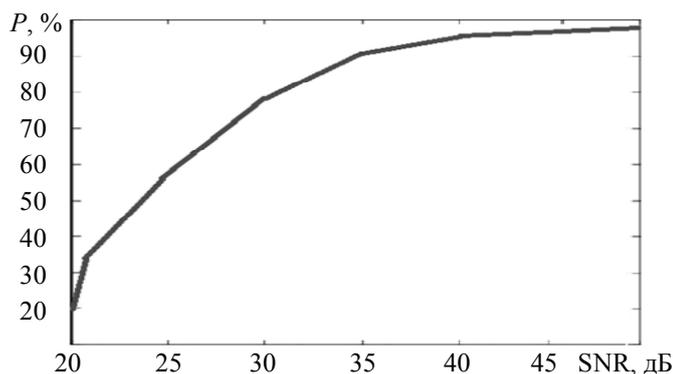


Рис. 2

Вероятность правильного распознавания 97,6 % была достигнута при SNR = 48 дБ. При снижении SNR до 21 дБ вероятность уменьшалась до 37 %. Для исследования робастности метода распознавания визуальных признаков использовались скрытые марковские модели [2]. На рис. 3 приведена зависимость точности распознавания от уровня искажений для лицевой модальности, показавшей значительно более высокую устойчивость к искажениям (при QF = 2 точность составила 48 %).

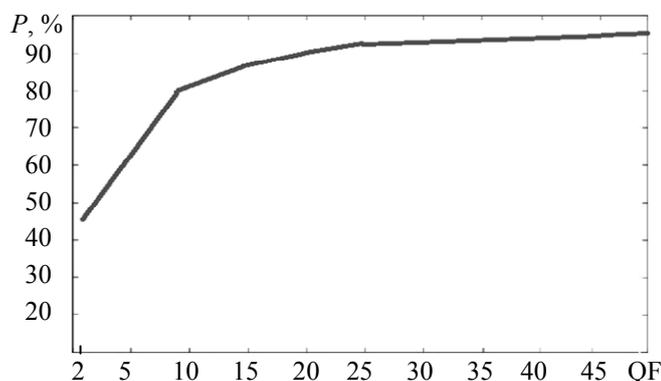


Рис. 3

Для объединения (фузирования) признаков различных модальностей необходимо в каждый момент времени располагать набором признаков от каждой модальности. Такие наборы должны быть выравнены по размеру, для этого использование разных модальностей, например звуковых и визуальных, должно быть синхронизировано с использованием интерполированных кадров до этапа слияния данных, что позволит использовать корреляции между ними [5]. Для оценки корреляции рассматривались задержки между последовательностями аудио- и видео-признаков в диапазоне 0—100 мс (с шагом 10 мс), причем точка с минимальным значением ошибки оказалась в значении 40 мс. Следует отметить, что в той же точке был достигнут максимум функций корреляций между аудио- и видеопризнаками [4].

Для дальнейшего повышения вероятности точного распознавания были проведены эксперименты по объединению модальностей. На рис. 4 приведены зависимости вероятности точного распознавания объединенной аудиовизуальной модальности (1) от уровней искажений SNR для звука (2) и QF для видео (3).

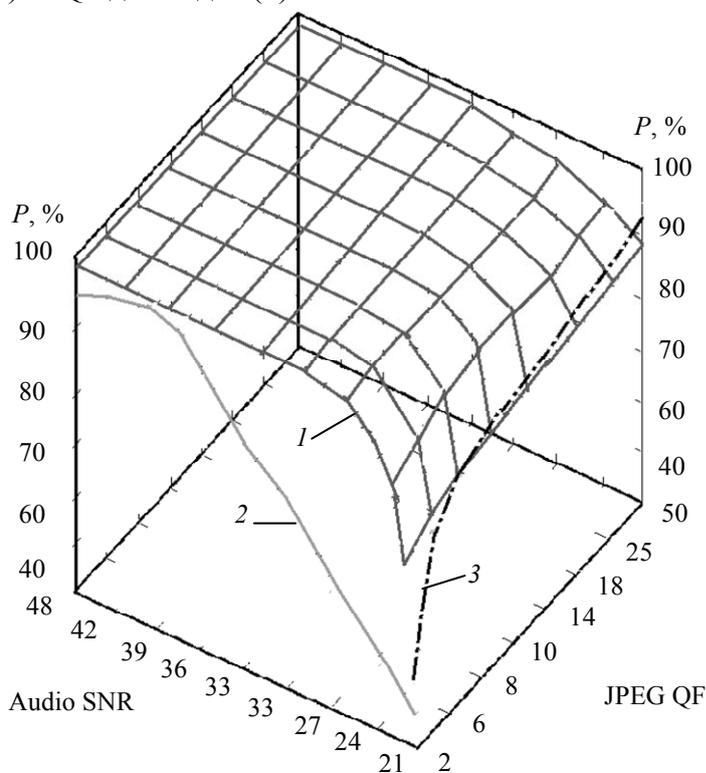


Рис. 4

В то время как использование аудиомодальности при максимальном уровне шумов обеспечило $P = 37\%$, а визуальной — 48% , использование объединенной аудиовизуальной модальности обеспечило $P = 70\%$ (см. таблицу).

Вероятность правильного распознавания аудиовизуальной модальности

QF	$P, \%$, при SNR, дБ									
	48	45	42	39	36	33	30	27	24	21
50	99,2	99,2	99,2	99,2	99,2	99,2	98,4	96,4	93,2	87,3
25	99,2	99,2	99,2	99,2	99,2	99,2	98,4	96,4	93,0	87,1
18	99,2	99,2	99,2	99,2	99,2	99,2	98,4	96,4	91,6	87,0
14	99,2	99,2	99,2	99,2	99,2	99,2	98,4	96,0	91,2	86,9
10	99,2	99,2	99,2	99,2	99,2	99,2	98,0	95,0	90,8	86,0
8	99,2	99,2	99,2	99,2	99,2	99,2	97,6	94,0	90,2	83,2
6	99,0	99,0	99,0	99,0	99,0	99,0	97,2	92,9	86,9	78,1
4	99,0	99,0	99,0	99,0	99,0	98,6	96,4	92,0	82,7	75,9
3	99,0	99,0	99,0	98,8	98,4	97,6	95,6	91,2	81,3	71,2
2	99,0	99,0	99,0	98,2	98,0	97,0	95,2	91,0	80,5	70,5

Результаты проведенных на двух аудиовизуальных базах данных экспериментов показали высокую устойчивость (робастность) мультимодальных биометрических систем идентификации личности [5, 6]. Однако требуется дальнейшее исследование влияния алгоритмов и уровня фузирования различных биометрических модальностей на робастность распознавания [7].

СПИСОК ЛИТЕРАТУРЫ

1. Городецкий В. И., Серебряков С. В. Методы и алгоритмы коллективного распознавания: обзор // Тр. СПИИРАН. 2006. Т. 1, вып. 3. С. 139—171.

2. Синицын И. Н., Новиков С. О., Ушмаев О. С. Развитие технологий интеграции биометрической информации // Системы и средства информатики. 2004. Вып. 14. С. 5—36.
3. Тропченко А. А., Тропченко А. Ю. Нейросетевые методы идентификации человека по изображению лица // Изв. вузов. Приборостроение. 2012. Т. 55, № 10. С. 31—36.
4. Ушмаев О. С. Методы мультибиометрической идентификации. М.: Изд-во ИПИ РАН, 2009. 114 с.
5. Dass S. C., Nandakumar K., Jain A. K. A principled approach to score level fusion in multimodal biometric systems // Audio- and Video-Based Biometric Person Authentication. 2005. P. 1049—1058.
6. Karam W., Bredin H., Greige H., Chollet G., Mokbel C. Talking-face identity verification, audiovisual forgery, and robustness issues // EURASIP J. Adv. Signal Process. 2009. Vol. 4. P. 1—15.
7. Ross A., Govindarajan R. Feature level fusion using hand and face biometrics // Proc. of the SPIE Conf. on Biometric Technology for Human Identification. Orlando, USA, 2005. P. 196—204.

Сведения об авторе

Андрей Александрович Тропченко — канд. техн. наук, доцент; Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, кафедра вычислительной техники;
E-mail: zayka_98rus@mail.ru

Рекомендована кафедрой
вычислительной техники

Поступила в редакцию
23.12.13 г.

УДК 004.912

И. В. КАЛИНИН, С. В. КЛИМЕНКОВ, А. Е. ХАРИТОНОВА, Е. А. ЦОПА

ПРЕОБРАЗОВАНИЕ ЕСТЕСТВЕННОГО ЯЗЫКА В ФОРМАТ RDF С ПОМОЩЬЮ СЕМАНТИЧЕСКИХ АНАЛИЗАТОРОВ ТЕКСТОВОЙ ИНФОРМАЦИИ

Решена задача автоматического преобразования естественного языка (русского) в формат RDF средствами семантического анализа. Приведен алгоритм работы программных модулей, созданных для решения задачи.

Ключевые слова: текст, естественный язык, RDF, семантический анализ, тезаурус, AOT, Jena.

Введение. Консорциумом Всемирной паутины для машиночитаемого представления данных, в особенности — метаданных, была разработана модель RDF (Resource Description Framework) [1]. Одним из направлений развития сети Интернет является реализация механизмов машинной обработки информации [2].

В основе этих механизмов лежит работа с метаданными, однозначно идентифицирующими характеристики и содержание ресурсов Интернета. Обработка метаданных должна прийти на смену используемому в настоящий момент текстовому анализу документов [3].

Формирование RDF-описаний ресурсов обычно осуществляется вручную — авторами. Во многих случаях такой подход неэффективен: в частности, при необходимости формирования большого объема метаданных или при формировании содержимого ресурсов пользователями (интернет-энциклопедии, социальные сети). Таким образом, возникает потребность в автоматизации процесса формирования RDF-метаданных.

В настоящей статье рассмотрено решение задачи автоматизации преобразования текста на естественном языке в формат RDF с помощью технологий семантического анализа. Приведен пример разработки алгоритма анализа текстов технической документации на русском языке.