

## ИДЕНТИФИКАЦИЯ ДИКТОРА ПО ГОЛОСУ С ПОЗИЦИЙ ТЕОРИИ АКТИВНОГО ВОСПРИЯТИЯ

В.Е. Гай, И.В. Поляков

В статье рассматривается решение задачи идентификации диктора с позиций теории активного восприятия. Приводятся результаты вычислительных экспериментов

Ключевые слова: распознавание диктора, теория активного восприятия, цифровая обработка сигналов

Контроль доступа к различным физическим объектам и информационным ресурсам является актуальной задачей для любой современной организации. Решение этой задачи, обычно, выполняется с помощью системы контроля управления доступом (СКУД). Работа СКУД основана на сравнении идентификационных признаков физического лица или объекта с эталоном. В качестве идентификационных признаков, часто, используется биометрическая информация. Биометрические параметры для проведения идентификации можно разделить на два класса: статические и динамические. К статическим параметрам относятся изображения отпечатков пальцев, радужной оболочки и сетчатки глаза, форма лица, форма ладони, расположение вен на кисти руки и т. д. Эти параметры практически не меняются со временем. Динамические параметры – параметры, изменяющиеся во времени: голос, почерк, клавиатурный почерк, личная подпись и т. п.

В настоящее время, при решении задачи идентификации часто используются записи речи физического лица. В этом случае задачу идентификации можно рассматривать как задачу распознавания образов. Известны различные варианты задач идентификации диктора по голосу: текстозависимая и текстонезависимая, идентификация на открытом множестве пользователей и на закрытом.

Процесс распознавания образов, с позиций системного анализа, можно разделить на три этапа: формирование исходного описания, нахождение системы признаков и построение решающего правила [1].

Известны проблемы, связанные с применением существующих методов распознавания образов [2] при решении поставленной задачи:

1) проблема формирования исходного описания: данная проблема связана с тем, что существующие модели и методы распознавания адаптированы к конкретному классу прикладных задач и требуют априорного знания свойств анализируемых сигналов;

2) проблема формирования системы признаков, связанная с выбором конечного множества признаков, обеспечивающих однозначность решения задачи классификации на этапе распознавания и отвечающая требованиям необходимости и достаточности. Этап выбора системы признаков необходим для сокращения размерности входного описания. Учитывая, что задача сокращения размерности -- оптимизационная задача, то для её решения необходимо использование критерия информативности. Отсутствие модели априорной неопределённости и модели её раскрытия породило большое количество методов в выборе критерия информативности, что, в свою очередь, порождает большое число возможных вариантов признаков;

3) проблема принятия решений в условиях априорной неопределённости. Этап принятия решения заключается в сравнении с имеющимся эталоном признакового

Гай Василий Евгеньевич – НГТУ, канд. техн. наук, доцент, e-mail: iamuser@inbox.ru  
Поляков Игорь Владимирович – НГТУ, магистрант, e-mail: polyakovigor92@gmail.com

описания анализируемого вибросигнала. Предполагается, что эталону соответствует компактное множество точек в системе признаков. Однако помехи, структурные изменения одного и того же представителя класса приводят к перекрытию классов. Поэтому проблема принятия решения замыкается на проблемы формирования системы признаков, позволяющей сформировать эталон, имеющий компактное представление.

Решение указанных проблем может быть выполнено на основе теории активного восприятия [1].

Реализация предварительной обработки изображения, с позиций теории активного восприятия, заключается в выполнении операции интегрирования, реализация процесса нахождения признаков заключается в выполнении операции дифференцирования.

Результатом первого преобразования является множество «визуальных» масс, упорядоченных в координатах исходного пространства  $\|m_{ij}\|$ . Результатом второго является вектор  $\mu = (\mu_0 \equiv m_0, \mu_1, \mu_2, \dots, \mu_{15})$ . Преобразование дифференцирования реализуется с помощью 16 фильтров Уолша системы Хармута. Технология их применения отличается тем, что они применяются к изображению только после выполнения преобразования интегрирования. Пара преобразований, первое из которых – интегральное, второе – дифференциальное, определены  $U$ -преобразованием.

Теория активного восприятия не ограничивается только формированием спектрального представления сигнала. В состав теории входит раздел «Алгебра групп», посвящённый анализу зависимостей между спектральными коэффициентами разложения. Обнаруженные зависимости допускают своё использование на этапах принятия решения и понимания анализируемого сигнала. Пусть каждому фильтру  $F_i \in \{F_i\} \equiv F$  соответствует координатно-определённый бинарный оператор  $V_i \in \{V_i\} \equiv V$ ; тогда компоненте  $\mu_i \neq 0$  вектора  $\mu$  допустимо поставить в

соответствие оператор  $V_i$  либо  $\bar{V}_i$  в зависимости от знака компоненты. В результате вектору  $\mu$  ставится в соответствие подмножество операторов из  $\{V_i\}$ , имеющих аналогичную фильтрам конструкцию, но разное значение элементов матрицы ( $+1 \leftrightarrow 1$ ;  $-1 \leftrightarrow 0$ ). Задавая на множестве  $\{V_i\}$  операции теоретико-множественного умножения и сложения, имеем алгебру описания изображения в двумерных булевых функциях.

На множестве операторов формируется алгебра групп анализируемого сигнала:

1) семейство алгебраических структур (названных полными группами)  $\{P_{ni}\}$  вида  $P_n = \{V_i, V_j, V_k\}$  мощности 35, где каждая группа изоморфна на сложный нейрон, а их подмножество мощности 32 – на кристаллографические группы;

2) семейство алгебраических структур (названных замкнутыми группами)  $\{P_{si}\}$  вида  $P_s = \{V_i, V_j, V_k, V_r\}$  мощности 105, где каждая группа изоморфна на сверхсложный нейрон и образована из пары определённым образом связанных полных групп.

С помощью замкнутых и полных групп выполняется спектрально-корреляционный анализ. Полные группы позволяют выявить корреляционные связи между операторами. Замкнутые группы позволяют выявить корреляционные связи между полными группами. Если множество операторов – алфавит, то множества групп – более сложные грамматические описания наблюдаемого сигнала: полная группа – слово, замкнутая группа – словосочетание.

С учётом алгебры групп, предлагается следующий подход к вычислению признаков описания сигнала  $f$ . Обработка сигнала ведётся во временной области.

1) отсчёты анализируемого сигнала разбиваются на множество сегментов  $S = \{s_k\}$ ,  $k = \overline{1, N}$ , где  $N$  – число сегментов в сигнале  $f$ ;

2) к каждому сегменту  $s_k$  применяется  $U$ -преобразование ( $U$ -преобразование является базовым в теории активного восприятия), в результате, формируется спектральное представление каждого



сегмента:  $u_k = U[s_k]$ ,  $u = \{u_k\}$ , где  $U$  – оператор вычисления  $U$ -преобразования;

4) по вычисленному спектральному представлению  $u_k$  сегмента  $s_k$  определяется описание сигнала с помощью одной из структур, входящих в алгебру групп (например, с помощью полных или замкнутых групп):

$p = P[u]$ , где  $P$  – оператор вычисления групп по спектральному представлению сигнала,  $p = \{p_k\}$ ,  $p$  – описание сигнала, выполненное с помощью алгебры групп;

5) вычисляется гистограмма частоты появления пар групп:  $h = H[p]$ , где  $H$  – оператор формирования гистограммы замкнутых групп, которая и является признаковым описанием сигнала  $f$ . В результате признаковым описанием сигнала является гистограмма  $h$ . Предлагается использовать одномерные, двумерные и трёхмерные гистограммы. После вычисления гистограммы возможна нормализация

системы признаков по амплитуде к отрезку  $[0; 1]$ .

Тестирование предложенного подхода к решению задачи идентификации диктора по голосу, будет выполнено на самостоятельно собранной базе данных звуковых записей, включающей 50 дикторов (25 мужчин и 25 женщин). Для каждого диктора в базе данных хранится 10 файлов длительностью 4 секунды каждый. Глубина кодирования – 16 бит, частота дискретизации – 44100 Гц.

Вычислительный эксперимент выполнялся на основе метода перекрёстной проверки (массив данных разбивался на 10 элементов). Решение задачи идентификации диктора по голосу выполняется на закрытом множестве дикторов. Рассматривается текстонезависимая идентификация диктора. В табл. 1 приведены известные результаты решения задачи идентификации диктора, в табл. 2 приведены результаты, полученные на основе предложенного подхода.

Табл. 1

Точность работы известных подходов

| Источ-ник | Система признаков  | Классификатор               | Число дикторов | Точность идентификации (в %) |
|-----------|--|-----------------------------|----------------|------------------------------|
| [3]       | Модель гауссовой смеси, мел-частотные кепстральные коэффициенты (12 элементов)                   | классификатор Байеса        | 16             | 94.5                         |
| [4]       | 64-компонентная модель гауссовой смеси, мел-частотные кепстральные коэффициенты (12 элементов)   | классификатор Байеса        | 50             | 87                           |
| [5]       | 1024-компонентная модель гауссовой смеси, мел-частотные кепстральные коэффициенты (13 элементов) | Метод опорных векторов, 1-1 | 64             | 92                           |
| [6]       | Модель гауссовой смеси, коэффициенты перцептивного линейного предсказания (39 элементов)         | Метод опорных векторов, 1-1 | 47             | 92                           |

Табл. 2

Точность работы предлагаемых подходов

| Система признаков   | Классификатор                | Точность идентификации (в %) |
|---|------------------------------|------------------------------|
| Гистограмма замкнутых групп (840-мерный вектор), без нормализации | Метод $K$ -ближайших соседей | 68                           |

|  |                             |    |
|--|-----------------------------|----|
| Двумерная гистограмма полных групп на операции умножения (матрица $140 \times 140$ ), без нормализации | Метод опорных векторов, 1-1 | 84 |
| Двумерная гистограмма полных групп на операции умножения (матрица $140 \times 140$ ), без нормализации | Метод опорных векторов, 1-N | 95 |
| Гистограмма замкнутых групп (840-мерный вектор), с нормализацией                                       | Метод K-ближайших соседей   | 68 |
| Двумерная гистограмма полных групп на операции сложения (матрица $140 \times 140$ ), с нормализацией   | Метод опорных векторов, 1-1 | 83 |
| Двумерная гистограмма полных групп на операции сложения (матрица $140 \times 140$ ), с нормализацией   | Метод опорных векторов, 1-N | 94 |

Результаты вычислительных экспериментов подтверждают эффективность предложенного подхода к решению задачи идентификации диктора.

Результаты исследований могут найти применение при построении модулей СКУД для обработки речевых сообщений, в системах текстозависимой и текстонезависимой идентификации, в системах верификации дикторов.

#### Литература

1. Perez-Meana H. (ed.). *Advances in Audio and Speech Signal Processing: Technologies and Applications: Technologies and Applications*. – Igi Global, 2007., Chapter XIII, page 374

2. В. А. Утробин Физические интерпретации элементов алгебры изображения // *Успехи физических наук*, Т. 174, № 10, 2004, С. 1089–1104.

3. Reynolds Douglas A, Rose Richard C. Robust text-independent speaker identification

using Gaussian mixture speaker models // *Speech and Audio Processing*, IEEE Transactions on. — 1995. Vol. 3, no. 1. — Pp. 72–83.

4. Reynolds Douglas A. Automatic speaker recognition using Gaussian mixture speaker models // *The Lincoln Laboratory Journal / Citeseer*. — 1995.

5. Apsingekar Vijendra Raj, De Leon Phillip L. Support vector machine based speaker identification systems using GMM parameters // *Signals, Systems and Computers*, 2009 Conference Record of the Forty-Third Asilomar Conference on / IEEE. — 2009. — Pp. 1766–1769.

6. Boujelbene Siwar Zribi, Mezghanni Dorra Ben Ayed, Ellouze Nouredine. Robust Text Independent Speaker Identification Using Hybrid GMM-SVM System. // *JDCTA*. — 2009. — Vol. 3, no. 2. — Pp. 103–110.

ФГБОУ ВО «Нижегородский государственный технический университет им. Р. Е. Алексеева»  
Nizhny Novgorod State Technical University named after R. E. Alekseev

## IDENTIFICATION OF SPEAKER'S VOICE IN THE VIEW OF THE ACTIVE PERCEPTION THEORY

V.E. Gay, I.V. Polyakov

The article discusses the problem of speaker's identification from the standpoint of the active perception theory. The article presents the results of computational experiments

Key words: speaker recognition, theory of active perception