

УДК 534.78

СОРТИРОВКА АВТОКОРРЕЛЯЦИОННЫХ ПОРТРЕТОВ ДЛЯ УСКОРЕНИЯ РАСПОЗНАВАНИЯ РЕЧЕВЫХ КОМАНД

© 2014 А.П. Ерофеев, В.Р. Крашенинников

Ульяновский государственный технический университет

Поступила в редакцию 19.10.2014

Исследуется способ ускорения работы алгоритма распознавания речевого сигнала, использующий этап предварительного исключения из рассмотрения заведомо несовпадающих автокорреляционных портретов речевых команд. На этапе предобработки для каждого портрета создается уменьшенная копия, которая используется для быстрого принятия решения о схожести эталонной команды с произнесенной. В случае положительного решения, производится сопоставление полноразмерных изображений. Применение алгоритма позволило сократить максимальное время распознавания речевой команды вдвое при сохранении эффективности распознавания.

Ключевые слова: *распознавание, речевая команда, интенсивный шум, автокорреляционный портрет, эталон, сортировка*

В настоящее время большое внимание уделяется разработке речевых информационно-управляющих систем, в которых запрос на выдачу информации или управление исполнительными механизмами осуществляется по голосовым командам оператора, например, пилота самолёта. Основной проблемой, возникающей при разработке этих систем, является распознавание речевых команд (РК) оператора, которое особенно затруднено наличием сильных акустических помех, например, от двигателя летательного аппарата. В то время как имеется значительное количество промышленных систем, обеспечивающих вероятность правильного распознавания, приближающуюся к 100%, их эффективность существенно снижается в условиях сильных посторонних акустических шумов.

Одним из методов, успешно применяемых в условиях сильных акустических помех, является идентификация РК, преобразованных в особые изображения – автокорреляционные портреты (АКП). Этот метод заключается в обнаружении во входном потоке РК [1], построения для них АКП и последующего сопоставления с АКП эталонных РК, предварительно записанных и сохраненных в памяти [2]. При этом АКП представляет собой двумерное изображение, строки которого есть выборочные значения нормированной автокорреляционной функции отрезка

оцифрованного акустического сигнала. Это преобразование позволяет применить методы обработки изображений [3-5] для распознавания речевых команд.

Эффективность распознавания РК по их АКП значительно снижают ошибки определения границ РК, для уменьшения их влияния применяется варьирование оцененных границ [6]. Большое значение имеет также состав словаря (библиотеки) распознаваемых команд и набор эталонов, оптимизация которых рассмотрена в [7]. Основной же мешающий фактор это вариативность произношения РК. Исследование вариативности произношения и способы её учёта в алгоритмах распознавания рассмотрены в [8], в частности, предложено использование кросскорреляционных портретов вместо АКП [6, 9, 10]. Распознавание улучшается также за счёт различной предобработки АКП [11]. При использовании речевого управления системой, работающей в реальном времени (транспортным средством, производственным механизмом), важна также и скорость отклика на произнесенную РК. Значительные задержки в процессе распознавания недопустимы.

При сравнении АКП распознаваемой РК с эталонными АКП используется метод, основанный на динамическом программировании, когда для каждой строки АКП распознаваемой РК подбирается отстоящая на минимальное расстояние от неё строка АКП эталонной команды по некоторой метрике. В то время как построение АКП эталонных команд выполняется лишь один раз при запуске программы, сопоставление АКП распознаваемой РК приходится выполнять с

Ерофеев Александр Павлович, аспирант. E-mail: a.p.erofeev@gmail.com

Крашенинников Виктор Ростиславович, доктор технических наук, профессор, заведующий кафедрой «Прикладная математика и информатика». E-mail: kvr@ulstu.ru

каждым из АКП эталонных РК, что и составляет основное время процесса распознавания.

С точки зрения ускорения работы алгоритма целесообразно каким-то образом заранее уменьшить количество возможных эталонов, с которыми сравнивается каждая поступающая команда. Один из способов – разделение всех команд на несколько немногочисленных групп, состоящих из функционально близких команд определённого режима работы системы (например, взлёт, манёвр или посадка). При этом распознавание текущей РК производится только в её функциональной группе, что достигается указанием текущего режима работы системы. В то же время внутри одной функциональной группы могут оказаться команды, АКП которых будут отличаться весьма существенно. Они будут различимы даже при сопоставлении уменьшенных копий. При этом временные затраты на сопоставление будут меньше пропорционально уменьшению размера АКП распознаваемой и эталонных команд. После того, как произведен отсев заведомо неподходящих эталонов, производится сопоставление с полноразмерными АКП эталонных изображений.

Для уменьшения размеров АКП в настоящей работе был использован алгоритм ближайшего соседа, реализованный в библиотеке OpenCV [12]. Алгоритм заключается в выборе из оригинального изображения ближайшего пикселя, соответствующего искомому в уменьшенном изображении. При проведении эксперимента оригинальные АКП размером

32x150 пикселей уменьшались до 24x90 пикселей. Уменьшенные АКП сохранялись в памяти. При поступлении на вход очередной РК создается АКП и его уменьшенный вариант. Пусть при этом размер группы эталонов равен N_g . Производится сопоставление с набором АКП эталонных РК, результаты сортируются по возрастанию расстояния между портретами и выбираются $N_s = \max(0, 3N_g, 4)$ АКП с минимальным расстоянием до АКП распознаваемой РК. После этого производится сопоставление полноразмерных АКП выбранных РК алгоритмом.

Для оценки эффективности распознавания речевых команд при использовании описанного алгоритма ускорения была проведена серия экспериментов, в которых использовались 112 записей РК авиационной тематики, произнесенных на фоне сильных акустических помех. Полученные записи оцифрованы с частотой 44 кГц. Значения констант, используемых для выбора АКП, были выбраны на отложенной выборке РК размером 112 команд, записанных в тех же акустических условиях, что и контрольная. Константы подбирались так, чтобы минимизировать общее время распознавания при сохранении его точности. Результаты, полученные в ходе эксперимента, приведены в табл. 1. Использование предложенного алгоритма позволило повысить скорость распознавания речевых команд при использовании АКП размером 32x150 пикселей на 40-45%.

Таблица 1. Время распознавания речевых команд

Способ организации распознавания	Максимальное время распознавания РК, сек	Среднее время распознавания РК, сек
без сопоставления уменьшенных копий	0,072	0,037
с сопоставлением уменьшенных копий	0,043	0,022

СПИСОК ЛИТЕРАТУРЫ:

1. Крашенинников, В.Р. Обнаружение речевой активности на фоне помех по изменению квазипериода в двухканальной схеме регистрации звукового сигнала / В.Р. Крашенинников, А.П. Ерофеев, В.К. Капырин // Известия Самарского научного центра РАН. 2012. Вып. 4(3). С. 820-822.
2. Крашенинников, В.Р. Распознавание речевых команд на фоне интенсивных помех с помощью автокорреляционных портретов / В.Р. Крашенинников, А.И. Армер, Н.А. Крашенинникова, А.В. Хвостов // Научно-технические технологии. 2007. № 9. С. 65-76.
3. Васильев, К.К. Статистический анализ многомерных изображений / К.К. Васильев, В.Р. Крашенинников. – Ульяновск: УлГТУ, 2007. 172 с.
4. Krasheninnikov, V.R. Estimating Parameters of Interframe Geometric Transformation of an Image Sequence by the Fixed Point Method / V.R. Krasheninnikov, M.A. Potapov // Pattern Recognition and Image Analysis. 2010. Vol. 20, No. 3. P. 316-323.
5. Krasheninnikov, V.R. Estimation of Parameters of Geometric Transformation of Images by Fixed Point Method / V.R. Krasheninnikov, M.A. Potapov // Pattern Recognition and Image Analysis. 2012. Vol. 22, No. 2. P. 303-317.

6. Крашенинников, В.Р. Вариация границ речевых команд для улучшения распознавания речевых команд по их кросскорреляционным портретам / В.Р. Крашенинников, Е.Ю. Лебедева, В.К. Капырин // Известия Самарского научного центра Российской академии наук. 2013. Т. 15, №4(4). С. 928-930.
7. Krasheninnikov, V.R. Optimization of Dictionary and Model Library for Recognition of Speech Commands Based on Cross-Correlation Portraits / V.R. Krasheninnikov, N.A. Krasheninnikova, V.V. Kuznetsov, E.Yu. Lebedeva // Pattern Recognition and Image Analysis. 2013. Vol. 23, No. 1. P. 80-86.
8. Крашенинников, В.Р. Модели изменчивости речевых команд / В.Р. Крашенинников, А.И. Армер // Научные технологии. 2007. № 9. С. 56-64.
9. Krasheninnikov, V.R. Cross-Correlation Portraits of Voice Signals in the Problem Recognizing Voice Commands According to Patterns / V.R. Krasheninnikov, A.I. Armer, V.V. Kuznetsov, E.Yu. Lebedeva // Pattern Recognition and Image Analysis. 2011. Vol. 21, No. 2. P. 192-194.
10. Лебедева, Е.Ю. Алгоритм распознавания речевых команд в шумах по кросскорреляционным портретам с использованием Фурье-преобразования / Е.Ю. Лебедева, А.И. Армер, А.П. Ерофеев // Радиотехника. Математическое моделирование инфокоммуникационных систем. 2011. Вып. 175. С. 41-44.
11. Ерофеев, А.П. Предобработка автокорреляционных портретов для улучшения показателей распознавания речевых команд // Междисциплинарные исследования в области математического моделирования и информатики. Мат-лы 3-й науч.-практ. конф. – Ульяновск: SIMJET, 2014. С. 189-191.
12. Geometric Image Transformations: [Электронный ресурс]. 2014. - http://docs.opencv.org/modules/imgproc/doc/geometric_transformations.html

SORTING OF AUTOCORRELATION PORTRAITS TO ACCELERATE SPEECH COMMANDS RECOGNITION

© 2015 A.P. Erofeev, V.R. Krasheninnikov

Ulyanovsk State Technical University

We investigate a method to accelerate the recognition algorithm of the speech signal using the phase prior exclusion from consideration deliberately mismatched autocorrelation portraits of voice commands. At the stage of pre-processing for each portrait created a small copy of which is used for a quick decision on the similarity with the reference team delivered. In the case of a positive decision, a comparison of the full-size image is performed. Application of the algorithm has reduced the maximum time command recognition twice, while maintaining the efficiency of recognition.

Key words: speech commands recognition, noise, autocorrelation portrait, standard, pre-sorting

Alexander Erofeev, Post-graduate Student. E-mail: a.p.erofeev@gmail.com
Viktor Krasheninnikov, Doctor of Technical Sciences, Professor, Head of the Department "Applied Mathematics and Computing Science".
E-mail: kvr@ulstu.ru