

# Lip Reading: Preparing Feature Vectors

Stanislav Soldatov

Graphics & Media Laboratory\*

Department of Computational Mathematics and Cybernetics

Moscow State University, Moscow, Russia

stass@graphics.cs.msu.su

## Abstract

In human-computer interface development speech recognition plays important role. Speech recognition quality can be seriously improved by employing visual information. The subject of this paper is lip reading task, visual front-end task for speech recognition.

For lip reading it is reasonable to use methods, similar to methods, used in speech recognition area, like hidden markov models. This paper studies adaptation of lip feature data (lip contours, etc) for use in HMM-based recognition.

Proposed algorithm consists of several stages, including principal component analysis and vector quantization.

## Аннотация

В распознавании речи, необходимом для развития и совершенствования интерфейсов между человеком и компьютером, все более значительную роль приобретает визуальная система ввода информации, основанная на задаче чтения по губам.

Для задачи чтения по губам естественно использовать методы, аналогичные методам решения задачи распознавания речи; здесь рассматривается применение скрытых Марковских моделей. В данной статье изучается преобразование характерных точек - контуров губ - в вид, пригодный для использования в скрытых Марковских моделях.

Предлагаемый алгоритм состоит из нескольких частей, включая анализ методом главных компонент и векторную квантизацию.

**Keywords:** *lipreading, speech recognition.*

## 1. ВВЕДЕНИЕ

Распознавание речи, одного из важнейших способов человеческой коммуникации, является значительной частью задачи совершенствования интерфейсов между человеком и компьютером. Существующие системы распознавания речи, основанные на звуке, недостаточно надежны, так как их точность резко ухудшается при искажении звуковой информации шумами различной природы. Кроме того, при наличии нескольких говорящих система распознавания речи сталкивается с проблемой идентификации говорящего в данный момент. Поэтому целесообразно одновременное

использование видеоинформации, то есть чтения по губам, что также является важной частью человеческого восприятия речи. Эксперименты, описанные в статье Бреглера и Кёнига [1], показывают, что аудиовизуальная система ввода информации более надежна, чем просто аудио система.

В существующих решениях задачи распознавания речи из аудио-потока очень часто применяются скрытые марковские модели [2]. Поэтому их также естественно использовать в задаче чтения по губам. Это также дает возможность интегрировать аудио и визуальную части системы распознавания речи, улучшая характеристики всей системы распознавания [3].

Для распознавания речи с помощью скрытых марковских моделей могут быть выбраны различные лингвистические единицы. Простейших единиц (например, фонем) в языке меньше, но они более изменчивы в языке и поэтому их распознавание менее надежно. При этом сложных единиц, например, слов, слишком много и для их распознавания требуются слишком большие словари. Поэтому целесообразно строить систему распознавания на основе сочетаний двух или трех простых единиц, которые позволяют увеличить надежность распознавания без сильного увеличения объема требуемых вычислений и данных.

Состояния губ, соответствующие фонемам устной речи, называется виземами. Последовательности из двух визем называются дифонами. В данной статье рассматривается распознавание на основе дифонов.

Для распознавания движений губ применяются не только системы, основанные на СММ. В статье [1] характеристики для распознавания получены анализом изображений губ методом главных компонент. В работе Метью и др. [7] предпринята попытка объединения данных распознавания контуров и масштабирования. В статье Харвей и др. [6] информация о характеристиках губ получается из измерений масштабного пространства.

## 2. ПРЕДСТАВЛЕНИЕ КОНТУРОВ ГУБ В ВИДЕ ВЕКТОРОВ ПРИЗНАКОВ

### 2.1 Исходные данные

Первым этапом решения задачи чтения по губам является выделение контуров губ; это производится с помощью алгоритма выделения контуров губ на цветном изображении, основанного на алгоритме радиального расширения, подробнее описанного в статье [4].

---

\* <http://graphics.cs.msu.su/>



функция принадлежности  
классу кожи

оценочная функция

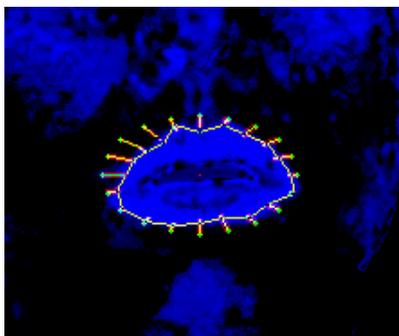
Для работы алгоритма выделения контуров губ необходима подготовка изображений. Изображения переводятся в цветное пространство (r/g,b/g), что позволяет избежать влияния освещения. На основе нескольких изображений с выделенными областями кожи лица и губ строятся цветовые классы. Описание цветового класса представляет собой параметры эллипса, внутри которого находится большинство точек класса на двумерной гистограмме в пространстве цветов (r/g,b/g). На основе определенных цветовых классов строится оценочная функция; функция принадлежности цветовому классу кожи учитывается с обратным знаком.

В дальнейшем осуществляется поиск контура с помощью оценочной функции. Первый шаг алгоритма находит приблизительное положение центра области губ на изображении. Второй шаг находит эллипс, описывающий область губ. Третий шаг находит уточненный контур, с помощью модифицированного алгоритма радиального расширения.

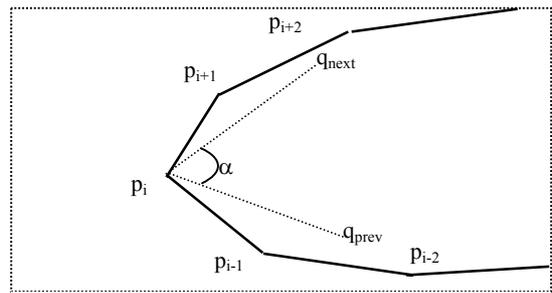
Схема алгоритма удобна для применения к видеопоследовательностям. В этом случае вместо первого шага для поиска положения центра области губ можно воспользоваться положением центра области губ на предыдущем кадре видеопоследовательности.

Процедура выделения контуров губ определяет эллипс, описывающий область рта и набор координат точек. Контур выделяется на наборе изображений, отражающем большинство возможных состояний губ.

Для распознавания движений губ необходимо выделить вектора признаков из полученных данных.



Пример исходных данных



Поиск углов

Процедура поиска контура губ находит  $n$  точек, пронумерованных от  $p_1$  до  $p_n$  по часовой стрелке.

Используемые координаты точек нормализуются: средняя точка эллипса считается началом координат, ось  $x$  направлена по направлению большего радиуса эллипса, большой радиус эллипса считается единицей.

Кроме координат точек, в процессе выделения контуров губ находятся параметры эллипса, описывающего область губ на исходном изображении. Параметры эллипса позволяют сделать выводы о таких общих параметрах области рта, как открыт рот или закрыт.

Нумерация контура начинается с места пересечения контура губ левым большим радиусом эллипса.

## 2.2 Поиск углов

Среди полученных точек необходимо определить правый и левый угол. Несмотря на нумерацию точек, это не всегда точки  $p_1$  и  $p_{n/2}$ . Правым углом считается точка, находящаяся в правой половине контура (между  $p_{n/4}$  и  $p_{3n/4}$ ), у которой угол  $\alpha$  является наименьшим. Угол  $\alpha$  - это угол между средними  $q_{next}$  и  $q_{prev}$ .

Здесь  $q_{next} = (p_{i+1} + \dots + p_{i+k})/k$ ,  $q_{prev} = (p_{i-1} + \dots + p_{i-k})/k$ ,  $k = n/5$ .

Аналогичное правило используется для левого угла.

## 2.3 Векторы Признаков

Следующим шагом после нахождения углов является преобразование набора исходных данных в набор векторов признаков.

В качестве нескольких первых элементов в векторе признаков используются признаки, полученные отдельно от координат - отношение высоты эллипса области губ к его ширине.

Дальнейшие элементы вектора признаков - это координаты левого и правого угла контура, координаты верхней и нижней точек контура, координаты остальных точек контура.

Рассмотрим варианты анализа полученных данных методом главных компонент.

Выделение базиса методом главных компонент позволяет найти основные направления, по которым изменяются вектора признаков. Это дает возможность значительно понизить размерность векторов признаков.

Метод главных компонент применяется к набору векторов признаков, полученных из набора данных, отражающих большинство возможных состояний губ.

### 3. КВАНТОВАНИЕ ВЕКТОРОВ ПРИЗНАКОВ

Каждому вектору признаков необходимо поставить в соответствие символ скрытой марковской модели. Для этого в данной статье используется метод векторной квантизации. С помощью этого метода пространство векторов признаков разбивается на кластеры, по принципу близости к центрам кластеров - кодовым словам. Набор кодовых слов называется кодовой книгой.

Основная сложность метода состоит в построении кодовой книги векторов. Размер кодовой книги определяется количеством состояний губ в исходных данных. Кодовая книга известного размера  $k$  строится алгоритмом  $K$  средних [5].

На первом шаге алгоритма случайным образом выбираются  $k$  векторов, считающихся кодовыми словами (центрами кластеров). На следующем шаге каждый входной вектор приписывается к тому кластеру, чье кодовое слово находится на наименьшем расстоянии от него. На третьем шаге кодовые слова каждого кластера пересчитываются. Каждое кодовое слово делается равным среднему арифметическому среди всех векторов кластера. Второй и третий шаги повторяются до тех пор, пока изменения кодовых слов не станут достаточно малы.

Этот алгоритм медленный, но применение анализа главных компонент перед квантованием позволяет понизить размерность и, тем самым, значительно ускорить процесс построения кодовой книги.

Новые исходные данные перед использованием в процессе распознавания квантуются: каждому вектору ставится в соответствие ближайший вектор из кодовой книги, и в дальнейшем вместо вектора в качестве символа скрытой марковской модели используется его индекс в кодовой книге.

### 4. ОБУЧЕНИЕ СИСТЕМЫ СКРЫТЫХ МАРКОВСКИХ МОДЕЛЕЙ

Распознавание по изображению не может работать на уровне визем, так как виземы для различных фонем достаточно близки. При этом распознавание на основе последовательностей визем - дифонов, трифонов - гораздо более надежно.

Для распознавания используется система эргодических скрытых марковских моделей. Каждому дифону соответствует своя СММ. СММ инициализируются равными вероятностями для символов и переходов между состояниями.

Обучение системы СММ производится с помощью последовательности квантованных векторов признаков. Исходные данные вручную разбиваются по обучаемым дифонам, после чего соответствующая СММ обновляется по алгоритму Баума-Велша [2]. Результирующая СММ выдает максимальные значения вероятности на последовательностях, близких к набору для обучения своего дифона.

### 5. РЕЗУЛЬТАТЫ

В результате работы построен эффективный алгоритм построения векторов признаков губ для задачи распознавания речи.

Алгоритм позволяет преобразовать данные контуров губ в наборы признаков, пригодных для распознавания. Алгоритм обладает свойствами надежности и устойчивости и легко интегрируется с системой распознавания речи на основе скрытых марковских моделей.

Эта разработка позволяет повысить точность системы распознавания речи за счет использования дополнительной визуальной информации. Это, в частности, способствует усовершенствованию интерфейса между человеком и компьютером.

### 6. ЛИТЕРАТУРА

- [1] C. Bregler, Y.Konig. "Eigenlips" for robust speech recognition. // Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 1994.
- [2] Рабинер Л.Р. *Скрытые марковские модели и их применение в избранных приложениях при распознавании речи: обзор.* // ТИИЭР, т.77, № 2, февраль 1989, сс.86-120.
- [3] N. Oliver, A. Pentland, F. Berard. *LAFTER: Lips and Face Real Time Tracker.* // IEEE Proceedings of the Conference on Computer Vision and Pattern Recognition '97, pp.123-129.
- [4] Солдатов С.А. *Чтение по губам: распознавание контуров губ.* // Международная научная конференция "Интеллектуализация обработки информации-2002". Тезисы докладов, с.147.
- [5] A. Linde, R. Gray. *An algorithm for vector quantization design.* // IEEE Transactions on Communicatinos COM-28. 1980.
- [6] R. Harvey, I. Matthews, J. A. Bangham, S. Cox. *Lip reading from scale-space measurements.* // Proceedings of the Conference on Computer Vision and Pattern Recognition, 1997.
- [7] I. Matthews, T. Cootes, S. Cox, R. Harvey, J. A. Bangham. *Lipreading Using Shape, Shading and Scale.* // School of Information Systems, University of East Anglia, Norwich.
- [8] G. Potamianos, H.P. Graf, E. Cosatto. *An Image Transform Approach for HMM Based Automatic Lipreading.* // Proceedings of the International Conference on Image Processing, Chicago, vol. III, 1998, pp. 173-177.

### Об авторе

Stanislav Soldatov is a PhD student at Moscow State University, Department of Computational Mathematics and Cybernetics. His contact email is [stass@graphics.cs.msu.su](mailto:stass@graphics.cs.msu.su)