**Содержание**

1 Сведения об объекте исследования 3

2 Цель работы 4

3 Выбор направления исследования 5

3.1 Краткая характеристика и оценка состояния рассматриваемой проблемы 5

3.1.1 Распознавание контуров губ в видеопотоке 6

3.1.2 Применение скрытых Марковских моделей 9

3.2 Актуальность работы 13

3.3 Обоснование выбранного метода направления исследования 13

5. Ожидаемые результаты 16

Список использованных источников 17

1 Сведения об объекте исследования

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. Одним из способов получения входных данных для данной задачи является поиск характерных точек, в частности распознавание контуров губ. Объектом данного исследования являются методы компьютерной обработки и анализа цифрового изображения, а предметом можно считать цифровое изображение, содержащее губы человека, и методы их локализации. От точности решения задачи локализации во многом зависит точность решения задач анализа изображения области губ более высокого уровня. Разработки, рассмотренные в данной работе, являются важным этапом на пути к усовершенствованию интерфейса между человеком и компьютером.

2 Цель работы

Цель написания работы – анализ существующих подходов к решению задачи локализации области губ человека на изображении. А так же ознакомиться с применением всевозможных методов распознавания.

3 Выбор направления исследования

3.1 Краткая характеристика и оценка состояния рассматриваемой проблемы

Технологии автоматического обнаружения и распознавания лица используются в ряде современных систем компьютерного зрения: биометрическая идентификация, человеко-машинный интерфейс, зрение роботов, компьютерная анимация, видеоконференции. Основное отличие данных приложений между собой – это целевые классы, которые являются объектами распознавания. Целевыми классами в задачи распознавания лица могут являться: лицо, лицо с элементами перекрытий, морда животного, лицо человека, живое лицо человека, мимика лица, черты лица, пол, раса, личность человека. Выбор одного из таких целевых классов определяет специфику алгоритма распознавания, остальные классы являются второстепенными и скорее играют роль признаков при распознавании целевого класса.

Основным способом извлечения данных о речи из видеоинформации является анализ движений губ, или чтение по губам. Главным источником получения входных данных для данной задачи является поиск характерных точек, в частности распознавание контуров губ. В большей части исследований, посвященных задаче распознавания речи, поиск контуров губ осуществляется с помощью активных контурных моделей.

Существует также другой возможный способ получения информации о губах, заключающийся в выделении особенностей области рта без поиска характерных точек.

В данной работе рассматривается задача распознавания контуров губ в цветном видеопотоке и локализация с помощью скрытых Марковских моделей.

3.1.1 Распознавание контуров губ в видеопотоке

В данной задаче распознавания контуров губ в видеопотоке предметом анализа являются цветные изображения при условии фронтального расположения лица. Кроме того, рассматривается видеопоток низкого разрешения, характерный для бытовых видеокамер. Такая постановка задачи накладывает ограничения на скорость работы алгоритма, распознавание контуров должно оставлять достаточно времени для распознавания речи. Важными факторами для разрешения этой задачи являются использование цвета в качестве главного источника информации, устойчивость архитектуры, а также применение быстрого алгоритма поиска контуров.

Для распознавания используется цветовая информация. Выделение по цвету человеческой кожи и губ достаточно устойчиво, их цветовые характеристики практически не зависят от освещения. Поэтому цветовое пространство, в котором будет осуществляться поиск, не должно учитывать освещение. Этому условию удовлетворяет цветовое пространство (r/g,b/g), которое используется при построении цветовых классов.

Для улучшения качества распознавания выделяются два цветовых класса - кожа и губы. В используемом двухмерном цветовом пространстве на основе выделенных на изображениях областей строятся двухмерные гистограммы (рис. 1). Предполагая нормальное распределение цветов, можно сократить описание цветовых классов с двухмерной гистограммы до пяти параметров. Функция принадлежности цветовому классу имеет значения в диапазоне [0,1].

На основе определенных цветовых классов строится оценочная функция; функция принадлежности цветовому классу кожи учитывается с обратным знаком. Несмотря на возможность использования обобщенных цветовых классов, для получения лучших результатов используется предварительная подстройка под пользователя.

Устойчивость алгоритма обеспечивает архитектура IFA, при которой алгоритм разделен на несколько этапов. Данные, полученные на предыдущем этапе, уточняются на следующем, и, таким образом, неуспешный поиск на текущем этапе означает возврат к предыдущему. В соответствии с архитектурой IFA решение задачи распознавания контуров разбито на три этапа. Первый этап - приблизительный поиск области губ на изображении, второй - ограничение контура губ эллипсом, третий - уточнение контура алгоритмом, основанном на радиальном расширении.

r/g



b/g

Рисунок 1. Двухмерная гистограмма, губы

Для распознавания важно, чтобы характеристики губ, полученные в результате, были инвариантны относительно освещения и положения лица на изображении. Первый этап - это предварительный поиск положения губ на изображении. На этом этапе происходит поиск начального приближения. Предполагая приблизительный размер области губ, на изображении осуществляется поиск области соответствующей площади. На этом этапе вычисляются приблизительные координаты центра области губ.

Предполагая работу в реальном времени, задачу поиска положения можно упростить, так как она может проводиться не во всех кадрах. Для работы в каждом кадре используются в качестве начального значения результаты поиска положения на предыдущем кадре.

На втором этапе область губ ограничивается эллипсом с помощью статистических методов. Рассматриваются значения оценочной функции в области, найденной на предыдущем этапе. Рассматриваются точки, в которых оценочная функция выше порогового значения. Значение f оценочной функции в точке изображения (xi,yi) интерпретируется как количество попаданий случайных величин X и Y в диапазон ≤ < +1, ≤ < +1 *i i i i x X x y Y.* Для случайных величин X и Y вычисляются математические ожидания и матрица ковариации. Оси координат совмещаются с направлениями, соответствующими собственным векторам матрицы. Для случайных величин X’ и Y’, соответствующих новой системе координат, считается среднее квадратическое отклонение. Математические ожидания xm,ym, угол поворота α и средние квадратические отклонения dx,dy однозначно определяют эллипс с центром в (xm,ym), повернутый относительно оси на угол α и с радиусами dx и dy.

На третьем этапе определяются характеристики формы губ, нужных для последующего распознавания. Контур уточняется с помощью алгоритма радиального расширения. В качестве начального приближения для контура используются точки на эллипсе, полученном на предыдущем этапе. Точки на эллипсе берутся в соответствии с точками спецификации MPEG4. Точки перемещаются по радиусам эллипса в зависимости от действия сил. Силы для точки *xi* определяются как

*Fi Fiout Fiin iint,*

*Fiout = kout,*

*iint =-kin,*

*iint =-kint*

Где *k* - коэффициенты, *vi* - направление перемещения точки *xi*.

За счет наложения более жестких ограничений на форму контура данный алгоритм позволяет производить более точное и быстрое распознавание, чем алгоритмы на основе активных контурных моделей. Количество итераций также сокращается за счет получения хорошего начального приближения на втором этапе.

В результате работы построен эффективный алгоритм поиска контуров губ для задачи распознавания речи. В отличие от алгоритмов на основе активных контурных моделей, этот алгоритм дает приемлемые результаты при меньшем числе итераций (4-8). Алгоритм позволяет надежно и быстро искать контуры губ в видеопотоке, освобождая больше времени для задачи чтения по губам.

3.1.2 Применение скрытых Марковских моделей

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.

Для работы алгоритма выделения контуров губ необходима подготовка изображений. Изображения переводятся в цветовое пространство (r/g,b/g), что позволяет избежать влияния освещения. На основе нескольких изображений с выделенными областями кожи лица и губ стоятся 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.

Процедура поиска контура губ находит n точек, пронумерованных от *p1* до *pn* 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.

Затем выполняем поиск углов (рис. 2). Среди полученных точек необходимо определить правый и левый угол. Несмотря на нумерацию точек, это не всегда точки *p1* и *pn/2*. Правым углом считается точка, находящаяся в правой половине контура (между *pn/4* и *p3n/4*), у которой угол α является наименьшим. Угол α - это угол между средними *qnext* и *qprev.* Здесь *qnext=* (*pi+1+…+ pi+k)/k,* *qprev=(pi-1+…+ pi-k)/k, k=n/5.* Аналогичное правило используется для левого угла.

Рисунок 2 - Поиск углов

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.

Каждому вектору признаков необходимо поставить в соответствие символ скрытой Марковской модели. Для этого используем 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.

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.

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. Результирующая СММ выдает максимальные значения вероятности на последовательностях, близких к набору для обучения своего дифона.

В результате работы строится 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.

3.2 Актуальность работы

Развитие компьютерной техники ведет к усовершенствованию интерфейсов между человеком и компьютером. Один из важнейших способов человеческой коммуникации - речь, поэтому надежный ввод речевой информации является важным направлением усовершенствования человеко-машинных интерфейсов. Даже у самых надежных систем распознавания речи, основанных только на звуке, точность резко ухудшается при искажении звуковой информации шумами различной природы. При наличии нескольких говорящих система распознавания речи сталкивается с проблемой идентификации говорящего в данный момент. Поэтому целесообразно использование также видеоинформации, что также является важной частью человеческого восприятия речи. Эксперименты ученых, данной области, показывают, что аудиовизуальная система ввода информации более надежна, чем просто аудио система, а значит, актуальна в наше время.

3.3 Обоснование выбранного метода направления исследования

Входе проведенного анализа существующих подходов к решению задачи локализации области губ человека на изображении, были выявлены достоинства и недостатки методов. В методе распознавание контуров губ в видеопотоке, за счет наложения более жестких ограничений на форму контура, алгоритм позволяет производить более точное и быстрое распознавание. Количество итераций также сокращается за счет получения хорошего начального приближения на втором этапе и в отличие от алгоритмов на основе активных контурных моделей, этот алгоритм дает приемлемые результаты при меньшем числе итераций. Рассматривая применение метода скрытых Марковских моделей можно выделить некоторые преимущества и недостатки. Такой метод позволяет работать непосредственно со сжатыми изображениями, такими как JPEG и MPEG, в которые на сегодняшний день являются распространёнными форматами хранения изображений и видео, что является большим преимуществом. А так же одним из полезных свойств СММ является способность сегментировать распознаваемое изображение. Метод скрытых Марковских моделей позволяет учесть локальные деформации и взаимное расположение участков изображений. Но в отличие от оптических потоков и других методов сопоставления деформациями, псевдодвумерная модель учитывает характер деформаций, а то какими именно могут быть возможные деформации, псевдодвумерные СММ усваивают в процессе обучения. Еще одним недостатком является то, что скрытые Марковские модели не обладают различающей способностью. Т.е. алгоритм обучения только максимизирует отклик каждой модели на свои классы, но не минимизирует отклик на другие классы, и не выделяются ключевые признаки, отличающие один класс от другого.

Данная область исследования является актуальной в наше время и не может не заинтересовать, потому как методы локализации недостаточно изучены и требуют дальнейших разработок.

4. Этапы выполнения НИРС

Таблица 1 – Этапы разработки

|  |  |  |
| --- | --- | --- |
| № | Этапы разработки | Недели |
| 1. | Сведения об объекте исследования | 1-3 |
| 2. | Цель работы | 3-5 |
| 3. | Выбор направления исследования | 5-8 |
| 4. | Защита НИРС | 10 |

5. Ожидаемые результаты

В дальнейших работах планируется устранение недостатков метода скрытых Марковских моделей, а именно улучшения способов начального представления изображения и алгоритмов тренировки, что позволило бы повысить точность распознавания при меньшем объеме информации. А так же эта разработка должна повысить точность системы распознавания речи за счет использования дополнительной визуальной информации. Это, в частности, способствует усовершенствованию интерфейса между человеком и компьютером.

Список использованных источников

1. Michael J. Jones, James M. Rehg, Statistical Color Models with Application to Skin Detection. *In CVPR*, 1999
2. B.D. Zarit, B.J. Super, and F.K.H. Quek, Comparison of five color models in skin pixel classification. In *Proceedings of the International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, pages 58-63, Kerkyra, Greece, September 1999.
3. Вежневец В. П. Локализация человеческого лица на цветном растровом изображении // Труды конференции «Математические методы распознавания образов» (ММРО - 10). – 2001
4. Самаль Д.И., Старовойтов В.В. Выбор признаков для распознавания на основе статистических данных // Цифровая обработка изображений. - Минск: ИТК, 1999. - С. 105-114.
5. Самаль Д.И., Старовойтов В.В. Методика автоматизированного распознавания людей по фотопортретам // Цифровая обработка изображений. - Минск: ИТК, 1999. - С. 81-85.