Чтение по губам: распознавание контуров губ.

Солдатов Станислав Александрович,

Лаборатория компьютерной графики, ВМК МГУ,

e-mail: stass@graphics.cs.msu.su , centrist@inbox.ru

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. Одним из способов получения входных данных для данной задачи является поиск характерных точек, в частности распознавание контуров губ. В данной статье рассматривается задача распознавания контуров губ в цветном видеопотоке. Использование цветовой информации увеличивает надежность алгоритма, основанного на архитектуре "увеличивающейся области внимания". Данный алгоритм дает приемлемые результаты при меньшем числе итераций, чем алгоритмы на основе активных контурных моделей.

Введение

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, описанными в статье Бреглера и Кёнига [5], показывают, что аудиовизуальная система ввода информации более надежна, чем просто аудио система.

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 бóльшей части исследований, посвященных задаче распознавания речи, (напр., в статьях Бреглера и Кёнига [5], Левина и др. [6]), 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. Такие методы рассматриваются в статье Харвей и др. [7]. В работе Метью и др. [8] предпринята попытка объединения данных распознавания контуров и масштабирования.

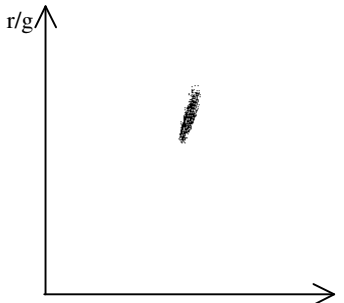
В задачах распознавания особую роль играет устойчивость алгоритма. В данной работе используется архитектура "увеличивающейся области внимания" (Incremental Focus of Attention, IFA), описанная в статье Тоямы и Хагера [1].

В данной работе цветовая информация в видеопотоке используется для поиска областей человеческой кожи. В работе Зарита [9] сделан обзор методов определения человеческой кожи на изображении по цвету.

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.

Распознавание по цвету

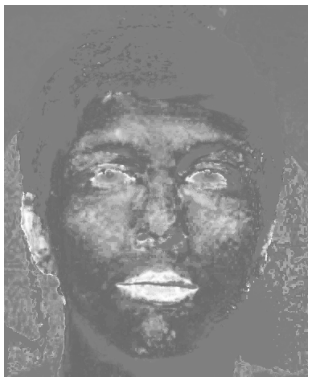
Для распознавания используется цветовая информация. Выделение по цвету человеческой кожи и губ достаточно устойчиво, их цветовые характеристики практически не зависят от освещения [2]. Поэтому цветовое пространство, в котором будет осуществляться поиск, не должно учитывать освещение. Этому условию удовлетворяет цветовое пространство (r/g,b/g), которое используется при построении цветовых классов.



двухмерная гистограмма, губы

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. 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].

функция принадлежности классу кожи


оценочная функция

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.

Увеличивающаяся область внимания

Устойчивость алгоритма обеспечивает архитектура IFA [1], 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.

Поиск контура

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 (x_i,y_i) интерпретируется как количество попаданий случайных величин X и Y в диапазон $x_i \le X < x_i + 1, y_i \le Y < y_i + 1$. 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 x_m,y_m , угол поворота α и средние квадратические отклонения d_x,d_y однозначно определяют эллипс с центром в (x_m,y_m) , повернутый относительно оси на угол α и с радиусами d_x и d_y

На третьем этапе определяются характеристики формы губ, нужных для последующего распознавания. Контур уточняется с помощью алгоритма радиального расширения [3]. 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 x_i определяются как

$$F_i = F_i^{out} + F_i^{in} + F_i^{\text{int}}$$

$$F_i^{out} = k_{out}$$

$F_i^{in} = -k_{in}$, если в точке x_i оценочная функция меньше порогового значения

$$F_i^{\text{int}} = -k_{\text{int}} \frac{(2x_i - x_{i-1} - x_{i+1})v_i}{|2x_i - x_{i-1} - x_{i+1}|}$$

Где k - коэффициенты, v_i - направление перемещения точки x_i

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.

Результаты

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.

Данная разработка, являясь одним из способов анализа речевой информации, способствует повышению точности систем распознавания речи. Это важный этап на пути к усовершенствованию интерфейса между человеком и компьютером.

Литература:

[1] K. Toyama, G. Hager. "Incremental Focus of Attention for Robust Vision-Based Tracking". // The International Journal of Computer Vision, Nov. 1999, 35(1), pp. 45-63.

[2] M. Fleck, D. Forsyth, C. Bregler. "Finding Naked People". // European Conf. on Computer Vision, 1996, Vol II, pp. 592-602.

[3] K. Toyama. "Look, Ma No Hands!" Hands-Free Cursor Control with Real-Time 3D Face Tracking". // http://www.research.microsoft.com/PUIWorkshop/Papers/Toyama.pdf

[4] T.F. Cootes, A. Hill, C.J. Taylor, J. Haslam. "The use of active shape models for locating structures in medical images". // Proceedings of the 13th International Conference on Information Processing in Medical Imaging, June 1993.

[5] C. Bregler, Y.Konig. "Eigenlips" for robust speech recognition", Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 1994.

[6] M.Lievin, P.Delmas, P.Y. Coulon, F. Luthon and V. Fristot. "Automatic Lip Tracking: Bayesian Segmentation and Active Contours in a Cooperative Scheme", Signal and Image Laboratory, Grenoble National Polytechnic Institute.

[7] 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

[8] 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

[9] B.H. Zarit. "Skin Detection In Video Images", University of Illinois at Chicago, 1999