

УДК 004.93
05.11.00

Разработка и исследование алгоритмов локализации контрольных точек с использованием модели активной формы

07, июль 2012

Прояев П.О., Жиденко М.В.

*Студенты,
кафедра «Биомедицинские технические системы»*

*Научный руководитель: Самородов А.В.,
к.т.н., доцент, доцент кафедры «Биомедицинские технические системы»*

МГТУ им. Н.Э. Баумана
prozelek@mail.ru
zhidenk@rambler.ru

Введение

Общение людей между собой было бы не таким полноценным без зрительного восприятия собеседника. Оценка таких факторов, как положение или наклон головы, направление взгляда, мимика лица, позволяет определять душевное состояние человека, его настроение, спрогнозировать его намерения и замыслы. Для людей оценка этих параметров не вызывает принципиальных затруднений, что нельзя сказать об автоматических системах. Одним из наиболее удачных алгоритмов автоматического определения особенностей лица является разработанные в 1995 г. Тимом Кутесом и Крисом Тейлором статистические модели формы объектов (ASM, active shape models, модели активной формы) [1]. Возможные области практического применения алгоритма - детекторы настроения, лжи, основанные на автоматическом анализе мимики. Несмотря на потенциально широкие возможности алгоритмов ASM, качество их работы существенно зависит от параметров изображений, ракурса и других факторов. Целью данной работы является рассмотрение, реализация и исследование алгоритма ASM с использованием нового метода выделения локальных текстурных признаков.

Алгоритм локализации контрольных точек лица на основе ASM

Алгоритм основан на получении статистической модели лица по наиболее информативным его элементам (брови, глаза, нос, уши, рот и овал лица), При этом извлекаются локальные признаки информативных точек, называемых

контрольными точками (КТ), которые затем используются для поиска наиболее схожих точек на новом изображении.

Построение статистических моделей формы объекта по набору обучающих изображений было впервые предложено в работе [2]. Пусть дана обучающая выборка изображений с нанесенными контрольными точками, в качестве которых используются точки вдоль границ элементов лица. Форма лица кодируется вектором, составленным из координат контрольных точек. Форму лица удобно представить в виде суммы “средней” по обучающей выборке формы, набора линейных приращений, определяющих нежесткие (не сохраняющие расстояние между точками) преобразования средней формы и набора параметров, задающих жесткие преобразования формы – поворот, масштабирование и параллельный перенос. С целью поиска направлений приращений, соответствующих нежестким преобразованиям, к векторам формы обучающей выборки применяется процедура выравнивания, позволяющая исключить жесткие преобразования модели формы, содержащиеся в обучающей выборке. Направления линейных приращений средней формы, соответствующие нежестким преобразованиям, определяются как главные компоненты матрицы ковариаций выравненных векторов формы обучающей выборки. Подробное описание алгоритма построения модели формы приведено в [2, 3].

На этапе обучения извлекается информация о свойствах текстуры из изображений обучающей выборки. Для этого в окрестности i -ой КТ модели для j -го изображения обучающей выборки извлекается текстурный профиль, представляющий собой вектор $g_{i,j}$, состоящий из n_p значений разности интенсивностей изображения в точках, расположенных вдоль нормали к границе модели [5].

В процессе поиска предполагаемых перемещений точек в окрестности каждой КТ модели из тестового изображения извлекается профиль g (по нормали к границе модели) длиной $l > n_p$. Далее на этом профиле выбирается точка d_{best} такая, что профиль тестового изображения с центром в этой точке наиболее схож с эталонным профилем [5]. В качестве меры схожести используется расстояние Махаланобиса:

$$f_{prof}(d) = (h(d) - \bar{g})^T S_g^{-1} (h(d) - \bar{g}),$$

где $h(d)$ – подмножество профиля g длины n_p с центром в точке d ,

S_g – матрица ковариаций профилей обучающей выборки.

Точка d_{best} выбирается в качестве предполагаемого нового положения соответствующей КТ. Совокупность координат точек $d_{best,i}$ задает приближение модели формы s^k на k -й итерации [6].

Описанный процесс повторяется итерационно до тех пор, пока не будет выполнено условие

$$\|s^{k+1} - s^k\| < \varepsilon \quad (1)$$

или не будет превышено максимально допустимое число итераций [6]. Процедуры обучения и поиска проводятся по пирамиде соответствующего изображения произвольной глубины.

В качестве альтернативного алгоритма (ASM+) был рассмотрен и реализован алгоритм выделения локальных текстурных признаков в окрестности КТ с помощью фильтра Гаусса. Фильтр представляет собой набор из 16-ти т.н. ядер, каждое из которых чувствительно только к одному из 16-ти направлений изменения текстуры и находится как разность между двумя функциями Гаусса [7]:

$$G_{pqf_k}^- = \frac{1}{2\pi\sigma^2} * \exp\left(-\frac{(p-q*\cos(f_k))^2 + (p-\sin(f_k))^2}{2\sigma^2}\right),$$

$$G_{pqf_k}^+ = \frac{1}{2\pi\sigma^2} * \exp\left(-\frac{(p+q*\cos(f_k))^2 + (p+\sin(f_k))^2}{2\sigma^2}\right),$$

где σ - СКО,

f_k – угол поворота, $f_k = k * 22.5^\circ$, $k=0..15$, $p \in [-3; 3]$, $q \in [-3; 3]$.

Далее происходит вычисление свертки по всей области интереса по следующей формуле [7]:

$$CE_{ijf_k} = \sum_{p,q} C_{i-p,j-q} * (G_{pqf_k}^- - G_{pqf_k}^+),$$

где C_{ij} представляет собой набор цветовых параметров изображения.

В оригинальном источнике [7] рассматриваются параметры «тон», «насыщенность» и «яркость», однако, как показали проведенные исследования данного алгоритма, достаточно использовать один серого с 256 градациями. Максимальный отклик среди 16-ти значений CE_{ijf_k} определяет максимальное изменение локальной границы вокруг данной точки.

Вектор текстурных признаков пикселя изображения состоит из 49 элементов. На рисунке 1 показано, в каком виде рассматривается окрестность точки ($\Psi = 22.5^\circ$). Каждый узел сетки – это результат свертки окрестности данного узла с описанным выше фильтром. Расстояния R_0 , R_1 и R_2 выбираются исходя из практических соображений.

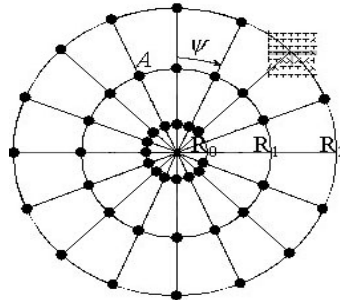


Рис 1. Вычисление текстурных признаков контрольной точки

Затем происходит сравнение вектора текстурных признаков текущей точки с шаблонным с вычислением меры K [7]:

$$K = \sum_{i=1}^{49} th\left(\frac{1}{16} \sum_{j=1}^{16} Sg(A_{ij}^b; A_{ij}^{rw})\right),$$

где A_{ij} является значением свертки для j -ой ориентации в конкретной области для i -ой точки входного окна,

b – шаблонный вектор текстурных признаков,

rw – текущий вектор текстурных признаков,

$$Sg(a,b) = \begin{cases} 1 & \text{if } (a > 0) \wedge (b > 0), \\ 1 & \text{if } (a = 0) \wedge (b = 0), \\ 0 & \text{otherwise.} \end{cases}$$

$$th(x) = \begin{cases} x & \text{if } x > threshold, \\ 0 & \text{otherwise.} \end{cases}$$

Пороговое значение (threshold) подобрано эмпирически и принято равным 0,2. Чем выше полученное значение меры K , тем больше схожесть окрестности рассматриваемой точки с шаблонной, полученной при обучении.

Реализация алгоритма ASM

Алгоритм ASM был реализован на языке C++ в среде Microsoft Visual Studio 2010 с применением пакета OpenCV 2.2.0. Обобщенная структурная схема алгоритма приведена на рис 2. Используемая схема расстановки КТ на изображениях лица показана на рис. 3.

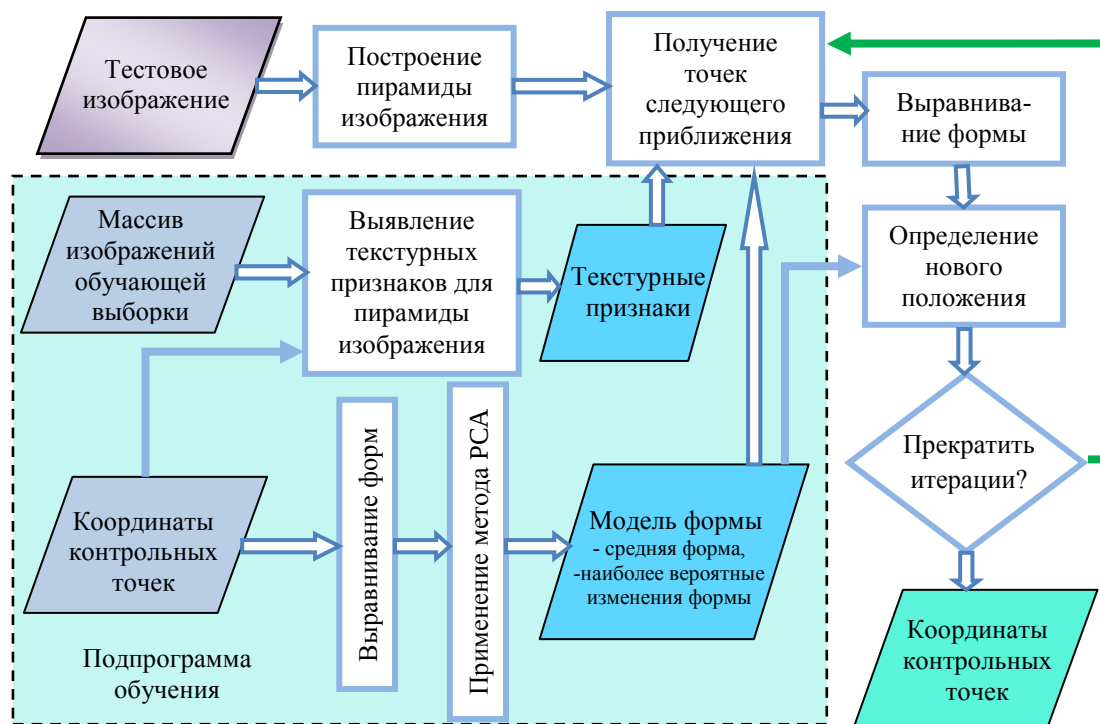


Рис 2. Структурная схема алгоритма ASM

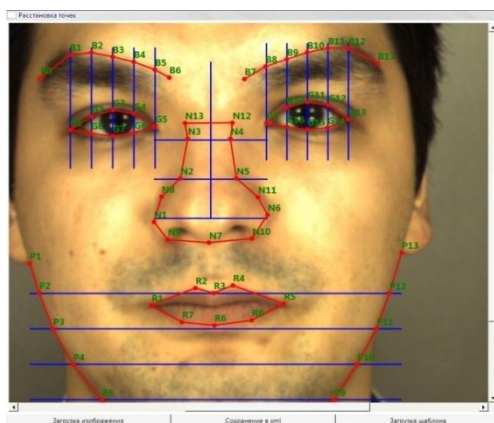


Рис 3. Схема расстановки контрольных точек лица

Пример результатов работы реализованного алгоритма ASM показан на рис 4. Черной линией показано начальное положение формы, белой – найденное. На левом рисунке – верное определение точек, на правом – ошибочное из-за наличия выраженных мимических изменений.

В качестве варьируемых параметров алгоритма выбраны:

- shapeaccuracy – точность построения формы (по этому параметру определяется число главных компонент);
- pyrlevels – глубина построения пирамиды изображения;
- zero_tresh – пороговое значение критерия того, что все модели формы обучающей выборки выравнены
- profilelength – длина текстурного профиля КТ в пикселях;
- searchdepth – длина (в одну сторону) исследуемого сегмента около каждой КТ модели формы для поиска на нем участка, наиболее похожего на текстурный профиль, полученных при обучении;
- maxiterations – число итераций поиска положения модели формы для каждой ступени пирамиды изображения;
- norm_tresh – пороговое значение критерия ε (1) того, что положение модели формы найдено на текущем изображении лица.

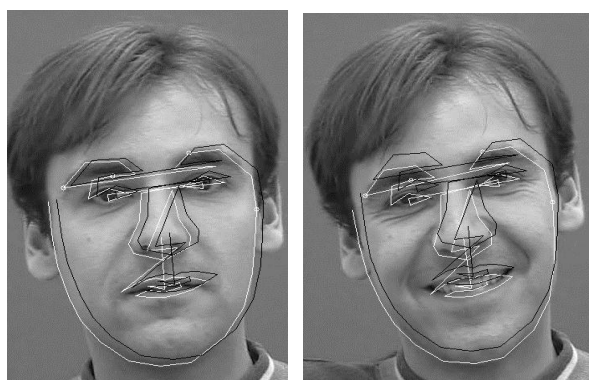


Рис 4. Пример обработки изображений базы Franck

Была проведена оптимизация указанных параметров алгоритма, исходя из критериев наивысшего качества его работы и максимального быстродействия для случая построения модели формы, содержащей 64 КТ. Наивысшее качество построения модели формы достигается при следующих значениях параметров: shapeaccuracy = 0,95; pyrlevels = 6; zero_tresh = 0,001; profilelength = 9; searchdepth =

14; maxiterations = 70; norm_tresh = 0,003. Максимальное быстродействие построения модели формы достигается при: shapeaccuracy = 0,90; pyrlevels = 3; zero_tresh = 0,004; profilelength = 9; searchdepth = 8; maxiterations = 25; norm_tresh = 0,005.

Результаты экспериментальных исследований

Экспериментальные исследования реализованного алгоритма ASM и ASM+ проводились на базе размеченных изображений Franck (5 000 изображений). Для тестирования был выбран вариант алгоритма с оптимумом быстродействия. Результаты сравнивались с работой аналогичных алгоритмов, реализованных в современных пакетах VOSM (Visionopen) [8] и ASMLib [9]. Показателем точности считалось среднее расстояние в пикселях между истинным положением контрольной точки и найденным для данной области. На рис. 4 приведены примеры изображений базы Franck, а в таблице – результаты тестирования алгоритмов. Быстродействие алгоритмов ASM и VOSM составляет порядка 0,5 с., а ASMLib – 0,3 с. ASM+ обрабатывает одно изображение базы Franck за 1,1 с.

Таблица. Результаты тестирования алгоритмов

Область	ASM	ASM+	VOSM	ASM Lib
Все лицо	10,3	8,1	8,8	6,9
Рот	8,6	7,9	10,0	7,5
Глаза	8,5	5,3	4,6	4,8
Брови	8,3	4,9	5,0	4,8

Обсуждение результатов и выводы

Наилучший результат, как по быстродействию, так и по точности показала работа алгоритма пакета ASMLib. Точность ASM удовлетворительна, но несколько ниже, чем у других алгоритмов, ASM+ показывает точность, превосходящую алгоритм ASM и сопоставимую с библиотекой ASMLib, но из-за нетривиального способа вычисления текстурных признаков он работает значительно медленнее. Исследование параметров алгоритма ASM показало важность оптимизации их значений, особенно глубины построения пирамиды и числа итераций поиска. Пирамиду большой глубины следует строить для изображений высокого разрешения, при этом точность и быстродействие находятся в обратной зависимости.

Дальнейшая работа может быть направлена как на оптимизацию реализованных алгоритмов, особенно процедуры построения вектора текстурных признаков, так и на существенную их модернизацию. При попытке построения модели формы на лице, сильно искаженном мимикой или с ракурсом, отличающимся от фронтального, качество работы исследованных алгоритмов существенно снижается. Необходимо создать модель формы, обеспечивающую последовательное приближенное определения координат КТ при значительных отклонениях ракурса лица от фронтального и выраженных мимических изменениях с уточнением координат КТ на последних этапах работы алгоритма.

Список литературы

1. Cootes T.F., Taylor C.J., Cooper D.H., Graham J. Active shape models - their training and application. 1995. C. 38 – 59.
2. Taylor C. J., Cooper D. H., Graham G. Training models of shape from sets of examples. In Proc. British Machine Vision Conference. 1992. C. 9 – 18.
3. Cootes T. F., Wheeler G.V., Walker K.N., Taylor C. J. View-based active appearance models. Image and Vision Computing, 20. 2002. C. 657 – 664.
4. Cootes T. F., Taylor C. J., Cooper D., Graham J. Active shape models - their training and application. Computer Vision and Image Understanding, 61(1). Jan. 1995. C. 38 – 59.
5. Stan Z. Li, Anil K. J. Handbook of face recognition. Springer Science. 2005. C. 39 – 63.
6. Hill A., Cootes T. F., Taylor C. J. Active shape models and the shape approximation problem. In Proc. British Machine Vision Conference, BMVA Press, UK, Sept 1995. C 157 – 166.
7. Gao X.W., Anishenko S., Shaposhnikov D. High-precision Detection of Facial Landmarks to Estimate Head Motions Based on Vision Models. Journal of Computer Science, Vol.3, No.7. 2007. C. 528 – 532.
8. <http://www.visionopen.com/downloads/vosm/>
9. <http://code.google.com/p/asmlib-opencv/>
10. Romdhani S., Gong S., Psarrou A. A multi-view non-linear active shape model using kernel PCA. In T. Pridmore and D. Elliman. editors, Proc. British Machine Vision Conference, Vol. 2. 1999. C 483 – 492.