
ТЕОРИЯ И КОНСТРУИРОВАНИЕ

Л.Х. Кашапова, Е.Ю. Латышева, И.Н. Спиридонов

АЛГОРИТМ РАСПОЗНАВАНИЯ ЭМОЦИОНАЛЬНОГО СОСТОЯНИЯ ПО ИЗОБРАЖЕНИЯМ ЛИЦА С ИСПОЛЬЗОВАНИЕМ ДИСКРИМИНАНТНОГО АНАЛИЗА И ФИЛЬТРОВ ГАБОРА

Аннотация

В статье описан алгоритм оценки эмоционального состояния по изображению лица с использованием дискриминантного анализа и фильтров Габора. Для классификации используются 6 базовых эмоций. Обучение и тестирование алгоритма проводились на базе изображений, включающих в себя изображения базы данных JAFFE. Вероятность правильного определения базового эмоционального состояния на тестовой выборке изображений составляет 72,38 %.

В современных условиях человек постоянно подвергается повышенным психологическим нагрузкам. Изменения социальной и физической среды, повышение темпов жизни и ее стоимости, разрушение традиционных семейных структур, социальные и экологические катаклизмы способствуют общему неспецифическому снижению психической резистентности. Актуальными задачами являются контроль психоэмоционального состояния и своевременное распознавание психических расстройств у работников профессий, ответственных за жизнь и здоровье людей. Кроме того, объективная оценка психоэмоционального состояния традиционна для психиатрии, где изучение эмоциональной сферы пациента является одним из основных инструментов врача-психиатра.

Одним из наиболее информативных показателей эмоционального состояния являются мимические движения лица. Существует несколько подходов к оценке эмоционального состояния по изображению лица в зависимости от используемого пространства признаков. Пространство признаков, вычисляемое по изображениям лица, может вклю-

чать в себя геометрические или текстурные признаки [1].

Геометрические признаки представляют собой расстояния, углы и другие особенности, характеризующие особенности расположения и размеров элементов лица. Использование комплекса мимических признаков из 26 расстояний и 7 углов между контрольными точками лица для классификации базовых эмоций позволяет достичь точности распознавания 85 % при правильной расстановке контрольных точек лица [2]. Однако погрешность автоматической расстановки контрольных точек является критическим звеном для анализа видеоизображений [3]. Оператору необходимо контролировать расстановку контрольных точек на каждом кадре видеоизображения и проводить коррекцию в случае необходимости.

В данной статье рассмотрен подход к оценке эмоционального состояния по текстурным признакам изображения лица. Текстурные признаки изображений вычисляются без привязки к контрольным точкам лица. В качестве текстурных признаков рассмотрены вейвлет-признаки. Наиболее

распространенными вейвлетами для анализа изображений лица являются вейвлеты Габора, которые также называются фильтрами Габора [4].

Фильтры Габора

Банк фильтров Габора представляет собой так называемый направленный микроскоп. Вейвлет-преобразование – это свертка исходного изображения с некоторой функцией ϕ -вейвлетом. Двумерный вейвлет (ядро, фильтр) Габора, соответствующий ориентации μ и масштабу ν , записывается как

$$\phi_{\mu, \nu}(x, y) = \frac{\|k_{\mu, \nu}\|^2}{\sigma^2} e^{-\frac{\|k_{\mu, \nu}\|^2 \|z\|^2}{2\sigma^2}} \left[e^{ik_{\mu, \nu} z} - e^{-\frac{\sigma^2}{2}} \right],$$

где $z = (x, y)$ – координаты вектора; $\|\bullet\|$ – норма вектора; вектор $k_{\mu, \nu} = k_\nu e^{i\theta_\mu}$; $k_\mu = k_{\max} / f^\nu$; $\theta_\mu = \pi\mu / 8$; $k_{\max}$ – максимальная частота.

Обычно используют банк фильтров Габора со следующими характеристиками: 5 различных масштабов $\nu \in \{0, \dots, 4\}$, 8 различных направлений $\mu \in \{0, \dots, 7\}$, $\sigma = \pi$, $k_{\max} = \pi / 2$ и $f = \sqrt{2}$ [4].

Изображение $I(x, y)$, подаваемое на вход алгоритма фильтрации, имеет размер $n_r \times n_c$ и является полутонным. Преобразование Габора изображения представляет собой свертку изображения с импульсной характеристикой фильтра Габора:

$$X_{\mu, \nu}(x, y) = I(x, y) * \phi_{\mu, \nu}(x, y),$$

где $*$ – оператор свертки.

Пример применения банка фильтров Габора представлен на рис. 2, где можно наблюдать выделение локальных признаков лица фильтрами в соответствующем направлении и на определенном масштабе. На рис. 1 представлено изображение, подаваемое на вход алгоритма.



Рис. 1. Пример изображения

После применения банка фильтров Габора для изображений обучающей выборки получаем $N \times O \times S$ матриц признаков, где N – количество изображений обучающей выборки; O – количество ориентаций; S – количество масштабов.

Для того чтобы признаки стали некоррелированными и матрицы признаков имели меньшую размерность, используется метод главных компонент (2DPCA). Применение 2DPCA к строкам матриц признаков уменьшает количество строк в мат-

рицах. Для перехода в новое пространство признаков необходимо получить ковариационную матрицу [4]

$$C = \frac{1}{NOS} \sum_{i, \mu, \nu} (X_{\mu, \nu}^i - \bar{X})(X_{\mu, \nu}^i - \bar{X})^T,$$

где

$$\bar{X} = \frac{1}{NOS} \sum_{i, \mu, \nu} X_{\mu, \nu}^i$$

– матрица среднего по всем матрицам признаков обучающей выборки; $()^T$ – операция транспонирования.

Поскольку матрица C размера $n_r \times n_r$ является симметричной, всегда возможно построить ортонормированный базис, состоящий из собственных векторов данной матрицы. При этом будем использовать собственные векторы, соответствующие $d \leq n_r$ ненулевым собственным значениям:

$$V_d = [\bar{v}_1, \bar{v}_2, \dots, \bar{v}_d].$$

Преобразование главных компонент для каждого изображения будет определяться формулой

$$Y_{\mu, \nu}^i = V_d^T (X_{\mu, \nu}^i - \bar{X}),$$

где $Y_{\mu, \nu}^i$ – матрица признаков размера $d \times n_c$, соответствующая μ -ориентации и ν -масштабу i -го изображения.

Для того чтобы признаки, соответствующие различным классам, были линейно разделимы, применяем после 2DPCA метод прямого линейного дискриминанта Фишера (2DFLD) к матрицам признаков, уменьшая таким образом размерность матриц по столбцам. Критерий Фишера запишем следующим образом [4]:

$$J(\alpha) = \frac{\alpha^T S_B^P \alpha}{\alpha^T S_W^P \alpha},$$

где S_B^P – матрица межклассового разброса признаков; S_W^P – матрица внутриклассового разброса признаков.

Матрицы S_B^P , S_W^P имеют размерность $n_c \times n_c$:

$$S_B^P = \frac{1}{NOS} \sum_{i=1}^K N_i SO(\bar{Y}_i - \bar{Y})^T (\bar{Y}_i - \bar{Y});$$

$$S_W^P = \frac{1}{NOS} \sum_{i, j, \mu, \nu} (Y_{\mu, \nu}^{ij} - \bar{Y}_i)^T (Y_{\mu, \nu}^{ij} - \bar{Y}_i),$$

где

$$\bar{Y}_i = \frac{1}{N_i OS} \sum_{j, \mu, \nu} Y_{\mu, \nu}^{ij}$$

– матрица среднего по матрицам признаков i -го класса;

$$\bar{Y} = \frac{1}{NOS} \sum_{i, j, \mu, \nu} Y_{\mu, \nu}^{ij}$$

– матрица среднего по матрицам признаков всех классов; K – количество классов.

Определим собственные векторы β и собственные значения Σ матрицы S_B^P : $S_B^P \beta = \Sigma \beta$; выбираем g собственных векторов, соответствующих $g \leq n_c$ ненулевым собственным числам. Затем проецируем матрицу S_W^P в подпространство, соответствующее g выбранных собственных векторов: $\beta_g^T S_W^P \beta_g = S_W^B$. Получаем матрицу S_W^B размера $g \times g$, которая может быть диагонализирована: $\gamma^T S_W^B \gamma = \eta$, где γ и η матрицы собственных векторов и собственных значений матрицы S_W^P . Отобрав $q \leq g$ собственных векторов, формируем итоговую матрицу признаков, соответствующую μ -ориентации и ν -масштабу i -го изображения:

$$Z_{\mu,\nu}^i = Y_{\mu,\nu}^i \beta_g \gamma_q.$$

Матрица $Z_{\mu,\nu}^i$ имеет размер $d \times q$ и содержит признаки, максимально разделенные в столбцах по классам. Объединяя матрицы признаков $Z_{\mu,\nu}^i$, относящиеся к одному изображению i , сформируем комплекс признаков i -го изображения.

К полученному комплексу признаков i -го изображения применим простейший классификатор с вычислением расстояния между двумя матрицами признаков, где признаки линейно разделены в столбцах. Матрицы признаков, соответствующие двум различным изображениям, с фиксированной ориентацией и масштабом

$$Z_{\mu,\nu}^n = [Z_{\mu,\nu,1}^n, Z_{\mu,\nu,2}^n, \dots, Z_{\mu,\nu,q}^n]$$

и

$$Z_{\mu,\nu}^m = [Z_{\mu,\nu,1}^m, Z_{\mu,\nu,2}^m, \dots, Z_{\mu,\nu,q}^m];$$

расстояние между двумя комплексами признаков, относящихся к двум различным исходным изображениям

$$dis(Z^n, Z^m) = \sum_{\mu,\nu,k} \|Z_{\mu,\nu,k}^n - Z_{\mu,\nu,k}^m\|_2,$$

где

$$\|Z_{\mu,\nu,k}^n - Z_{\mu,\nu,k}^m\|_2$$

– евклидово расстояние между двумя соответствующими вектор-столбцами матриц признаков $Z_{\mu,\nu,k}^n$ и $Z_{\mu,\nu,k}^m$. Вычисляя расстояние между комплексом признаков тестируемого изображения Z^{test} и N -числом комплекса признаков обучающей выборки Z^w , определяем принадлежность тестируемого изображения классу $K - Z^w \in K$:

$$dis(Z^{test}, Z^w) = \min_j dis(Z^{test}, Z^j).$$

Сбор и подготовка базы изображений для обучения и тестирования разрабатываемого алгоритма

Для формирования обучающей и тестируемой выборок использовали базу JAFFE, состоящую из 213 изображений 10 японских женщин, показывающих 7 эмоций (классов):

- 0 (neutral/покой) – 30 изображений;
- 1 (sad/печаль) – 31 изображение;
- 2 (happy/счастье) – 31 изображение;
- 3 (fear/страх) – 32 изображения;
- 4 (disgust/отвращение) – 29 изображений;
- 5 (anger/гнев) – 30 изображений;
- 6 (surprise/удивление) – 30 изображений.

Имеется от 2 до 4 изображений женщин каждого класса. Размер каждого изображения 256 x 256 пикселей, изображения представлены в формате TIFF. Изображения, подаваемые на вход алгоритма фильтрации банком фильтров Габора, подвергаются дополнительной обработке: на каждом изображении выделяется и вырезается область лица,

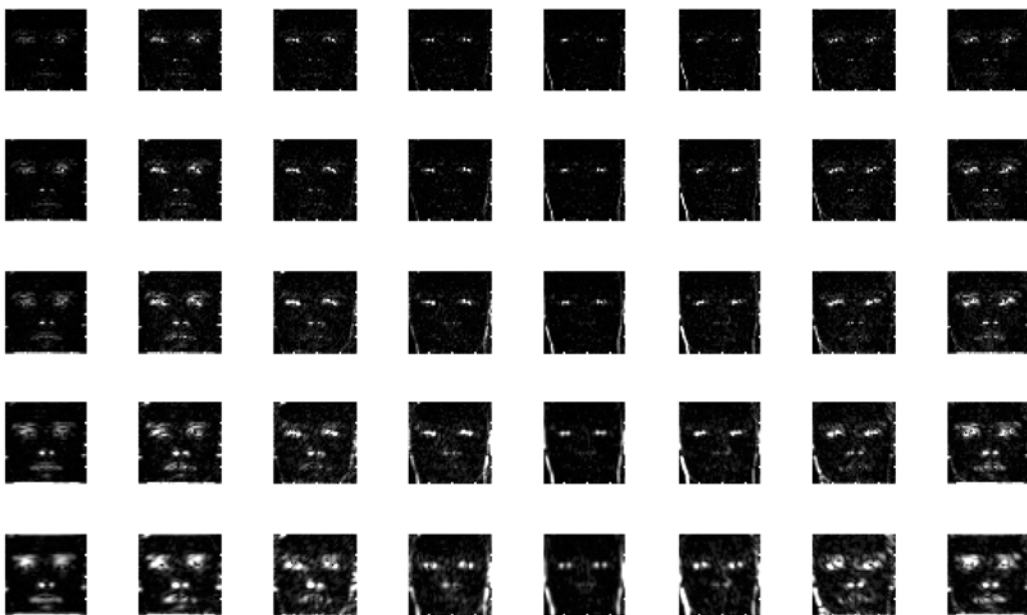


Рис. 2. Амплитудная характеристика свернутого изображения

изображение конвертируется в полутоновый формат. В эксперименте обучающая выборка представляла собой 7 классов по 15 изображений в каждом, набранных произвольным образом из базы. Остальные 105 изображений составили тестируемую выборку.

Таблица 1

Результаты тестирования ($p = g = q = 20$, Σ_i – количество изображений i -го класса тестируемой выборки, % – процент распознавания)

	0	1	2	3	4	5	6	Σ_i	%
0	11	–	–	1	–	–	3	15	73,3
1	1	10	–	2	–	–	–	13	76,9
2	3	4	9	–	–	–	–	16	56,3
3	1	2	1	9	2	–	2	17	52,9
4	–	–	–	1	10	2	1	14	71,4
5	3	2	–	–	1	9	–	15	60,0
6	–	–	–	3	1	–	11	15	73,3

Таблица 2

Результаты тестирования ($p = g = q = 10$, Σ_i – количество изображений i -го класса тестируемой выборки, % – процент распознавания)

	0	1	2	3	4	5	6	Σ_i	%
0	11	–	–	–	–	–	4	15	73,3
1	1	10	–	2	–	–	–	13	76,9
2	3	2	11	–	–	–	–	16	68,7
3	1	3	–	9	2	–	2	17	52,9
4	–	–	–	–	12	1	1	14	85,7
5	3	1	–	–	1	10	–	15	66,7
6	–	–	–	2	–	–	12	14	80,0

Таблица 3

Результаты тестирования ($p = g = q = 5$, Σ_i – количество изображений i -го класса тестируемой выборки, % – процент распознавания)

	0	1	2	3	4	5	6	Σ_i	%
0	12	–	–	–	–	–	2	15	80,0
1	1	11	–	1	–	–	–	13	84,6
2	2	2	12	–	–	–	–	16	75,0
3	1	3	–	9	2	–	2	17	52,9
4	–	2	1	–	11	–	–	14	78,6
5	4	1	–	–	1	9	–	15	66,7
6	2	–	–	2	–	–	11	15	80,0

Тестирование разработанного алгоритма: таблицы результатов

Результаты эксперимента приведены в табл. 1-3.

При рассмотрении таблиц результатов можно сделать вывод, что наилучший процент распознавания эмоций печали, счастья и нейтрального состояния достигается при $p = g = q = 5$, эмоции отвращения – при $p = g = q = 10$, эмоций гнева, удивления – при $p = g = q = 5$ и $p = g = q = 10$.

Выводы и результаты

Результаты показывают, что фильтры Габора могут быть использованы для оценки психоэмоционального состояния по изображению лица. Вероятность правильного распознавания на тестовой выборке составила 72,38 %. По сравнению с алгоритмом, основанным на расстановке контрольных точек лица, алгоритм с использованием фильтров Габора не требует контроля и коррекции со стороны оператора.

Список литературы:

1. Зинин А.М. Руководство по портретной экспертизе. Учебное пособие. – М.: Эксмо, 2006. 208 с.
2. Кашапова Л.Х., Крылов Ю.В., Спиридонов И.Н. Разработка алгоритма оценки эмоционального состояния по видеоизображению лица / IX МНТК «Физика и радиоэлектроника в медицине и экологии ФРЭМЭ'2010». Материалы НМШ по биометрическим технологиям для медико-биологических анализов и функциональной диагностики. – Владимир, 2010. С. 83-88.
3. Кашапова Л.Х., Хрулев А.А., Спиридонов И.Н. Использование видеоизображений для автоматизированной оценки мимических изменений / Медико-технические технологии на страже здоровья. Сб. докладов X РНТК, г. Монастир (Тунис), 28 сентября – 5 октября 2008 г. – М.: НИИ РЛ МГТУ им. Н.Э. Баумана, 2008. С. 76-78.
4. Mutelo R.M., Woo W.L., Dlay S.S. Discriminant analysis of the two-dimensional Gabor Features for face recognition // The Institution of Engineering and Technology. 2008. PP. 37-49.

Лилия Ханифовна Кашапова,
ассистент,

Екатерина Юрьевна Латышева,
студентка,

Игорь Николаевич Спиридонов,
д-р техн. наук, профессор,
зав. кафедрой,

кафедра биомедицинских
технических систем (БМТ1),
МГТУ им. Н.Э. Баумана,

г. Москва,

e-mail: kashapova.liliya@gmail.com