

**Крыловецкий Александр Абрамович** – Воронежский государственный университет; e-mail: aakryl@sc.vsu.ru; 394006, г. Воронеж, Университетская площадь, 1; тел.: +74732304639; кафедра цифровых технологий; зам. декана факультета компьютерных наук; к.ф.-м.н.; доцент кафедры цифровых технологий.

**Черников Игорь Сергеевич** – e-mail: ichernikov@yandex.ru; тел.: +79081446258; кафедра цифровых технологий; аспирант.

**Krylovetsky Alexander Abramovich** – Voronezh State University; e-mail: aakryl@sc.vsu.ru; 1, Universitetskaya ploshad, Voronezh, 394006, Russia; phone: +74732304639; the department of digital technologies; vice dean of computer science faculty; cand. of phis.-math. sc.; associate professor.

**Chernikov Igor Sergeevich** – e-mail: ichernikov@yandex.ru; phone: +79081446258; the department of digital technologies; postgraduate student.

УДК 004.93

**А.Н. Гнеушев**

### **ОЦЕНИВАНИЕ ПАРАМЕТРОВ ЛИЦА ПУТЕМ ОПТИМИЗАЦИИ МОДЕЛИ ИЗОБРАЖЕНИЯ В ПРОСТРАНСТВЕ БАЗИСНЫХ ФУНКЦИЙ**

*Рассматривается задача автоматической локализации характерных точек лица на изображении. На основе исследований [1–4] развивается подход построения текстурной модели объекта путем аппроксимации обучающего множества изображений с помощью оптимизации системы базисных функций. Каждое изображение элемента лица из обучающей выборки описывается весовым вектором в полученном базисе и координатами характерных точек. Оценивание параметров изображения объекта производится с помощью минимизации невязки весовых векторов модели и проекции изображения в базисном пространстве по параметрам преобразования модельной системы координат.*

*Текстурная модель; оптимизация базиса; оценка положения лица; слежение; выделение характерных точек лица; Gabor Wavelet Network.*

**A.N. Gneushev**

### **FACIAL FEATURES ESTIMATION BY IMAGE MODEL OPTIMIZATION IN SPACE OF THE BASIC FUNCTIONS**

*Consideration was given to a problem of the facial features localization in an image. Under investigation [1–4] the approach of the textural model construction by the learning image set approximation by optimization of the basic function system has been improved. Each facial feature's texture of the learning set is described by weight vector in the obtained basis and the set of the characteristic point coordinates. The parameters of the object image under consideration are estimated by minimizing the residual of the model weight vectors and projection of the image in the basic space for parameters of the model coordinate system transformation.*

*Textural pattern, decomposition optimization; Gabor basic functions; face localization; face tracking; face features extraction; Gabor Wavelet Network.*

**Введение.** В связи с повышенным интересом к проблеме распознавания и автоматического анализа видеоизображений объектов со сложной структурой, таких как лицо человека, большое значение имеют задачи обнаружения объекта, отдельных его элементов (глаза, нос, губы), определения их границ. Для решения этих задач необходимо построить текстурную модель изображений целевых объектов, с помощью которой можно локализовать соответствующую область. Текстурную модель изображений с неоднородной структурой целесообразно строить на основе аппроксимации целых областей некоторым семейством базисных функций, выбранных таким образом, чтобы наиболее эффективно представлять текстуру целе-

вых объектов. Так как изображения объекта зависят от его масштаба, ракурса и условий освещения сцены, текстурная модель должна описывать некоторый класс изображений инвариантно к этим факторам. Будем допускать, что изменение масштаба, положения и ракурса объекта можно моделировать с помощью преобразования системы координат изображения.

**2. Оптимизация текстурной модели.** Пусть  $f(\mathbf{x})$  – функция изображения, где  $\mathbf{x} = (x, y)^T$  – координаты точки. Будем ее рассматривать как функцию из пространства  $L^2(\mathfrak{R}^2)$  и искать ее среднеквадратичное представление в виде конечного ряда, в векторной форме:

$$\hat{f}(\mathbf{x}) = \Psi^T \mathbf{w}, \quad (1)$$

где  $\Psi = (\psi_{\mathbf{n}_1}, \dots, \psi_{\mathbf{n}_N})^T$ ,  $\mathbf{w} = (w_1, \dots, w_N)^T$ ,  $\psi_{\mathbf{n}_i}(\mathbf{x})$  – базисные функции в пространстве  $L^2(\mathfrak{R}^2)$  с вектором параметров  $\mathbf{n}_i \in \Omega_{\mathbf{n}}$ ,  $\Omega_{\mathbf{n}}$  – множество допустимых векторов  $\mathbf{n}_i$ ,  $w_i$  – веса разложения по базисным функциям,  $N$  – число базисных функций,  $\hat{f}$  – представление изображения  $f(\mathbf{x})$ . При заданном векторе  $\Psi$  вектор весов  $\mathbf{w}$  находится путем проецирования изображения на систему базисных функций:

$$\mathbf{w} = \mathbf{G}_{\Psi}^{-1}(\Psi, f), \quad (2)$$

где  $\mathbf{G}_{\Psi} = (\langle \psi_{\mathbf{n}_i}, \psi_{\mathbf{n}_j} \rangle)$  – матрица Грама,  $\langle \cdot, \cdot \rangle$  – операция скалярного произведения относительно  $\mathbf{X}$ ,  $(\Psi, f) = (\langle \psi_{\mathbf{n}_1}, f \rangle, \dots, \langle \psi_{\mathbf{n}_N}, f \rangle)^T$ .

Параметры базисных функций  $\mathbf{n}_i$  находятся путем минимизации среднеквадратичного отклонения обучающей выборки изображений целевого объекта и разложения (1):

$$\mathcal{E}^2 = \min_{\mathbf{n}_i \in \Omega_{\mathbf{n}}, w_i^k} \left( \sum_{k=1}^K \|f_k - \Psi^T \mathbf{w}^k\|_2^2 \right), \quad (3)$$

где  $f_k$  – изображение с номером  $k$  из обучающего множества,  $K$  – общее количество изображений во множестве,  $\mathbf{w}^k = (w_1^k, \dots, w_N^k)^T$  – вектор весов разложения изображения  $f_k$  по базису,  $\mathcal{E}$  – среднеквадратичная ошибка.

В разложении (1) используется только конечное число базисных функций. Таким образом, одному представлению (2) (проекции) соответствует несколько исходных изображений, близких по норме в  $L^2(\mathfrak{R}^2)$ , и, следовательно, ряд (1) будет представлять некоторый класс изображений.

*Определение 1.* Классом изображений  $C_{\varepsilon}$  относительно нормы в  $L^2(\mathfrak{R}^2)$  определим такое множество изображений  $\{f_k\}$ , что для  $\forall i, j \exists \varepsilon$ :  $\|f_i - f_j\| \leq \varepsilon$ .

*Определение 2.* Пусть на множестве рассматриваемых изображений определены система координат  $M_{cs}$  и ее непрерывное взаимнооднозначное преобразование  $T_a(\mathbf{x}): \mathfrak{R}^2 \rightarrow \mathfrak{R}^2$  с вектором параметров  $\mathbf{a}$ . Классом изображений  $C_{\varepsilon, T}$  с преобразованием  $T_a$  относительно нормы в  $L^2(\mathfrak{R}^2)$  определим такое множество изображений  $\{f_k\}$ , что для  $\forall i, j \exists \mathbf{a}(i, j)$ ,  $\varepsilon: \|f_i(\mathbf{x}) - f_j(T_a(\mathbf{x}))\| \leq \varepsilon$ .

Из определения 2 следует, что  $\forall f_k \in C_{\varepsilon, T} \exists \mathbf{a}(k): \{f_k(T_a(\mathbf{x}))\} \subseteq C_\varepsilon$ . Изображения класса  $C_{\varepsilon, T}$  являются эквивалентными относительно преобразования  $T_a$ , и, следовательно, их геометрическая структура может характеризоваться одним множеством точек в некоторой выделенной системе координат. Введем модельную систему координат  $M_{cs}$ , в которой определим множеством характерных точек  $\mathbf{P}_M = \{\mathbf{x}_m\}$ . Текстурно-геометрическую модель для изображений некоторого объекта из класса  $C_{\varepsilon, T}$  представим множеством  $M_O = \{\Psi, \mathbf{w}^k, \mathbf{P}_M, T\}_{k=0, \dots, K}$ . Модель  $M_O$  строится с помощью обучающей процедуры, в которой для каждого исходного изображения  $f_k^O \in C_{\varepsilon, T}$  эксперт задает множество характерных точек  $\mathbf{P}_O^k = \{\mathbf{x}_o^k\}$ . Далее формируется обучающая выборка  $\{f_k\}_{k=1, \dots, K} \subset C_\varepsilon$  путем преобразования:  $\{f_k^O(\mathbf{x})\} \rightarrow \{f_k(T_{a(k)}(\mathbf{x}))\}$ , где  $T_{a(k)}: \mathbf{P}_O^k \rightarrow \mathbf{P}_M$ . Обучение модели состоит в итеративном решении задачи (3) методом доверительных областей [3, 5].

Пусть имеется обученная модель и изображение  $f(\mathbf{x})$ , содержащее целевой объект. Оценивание параметров изображения объекта заключается в получении множества характерных точек  $\mathbf{P}_O = \{\mathbf{x}_o\}$  в системе координат  $f(\mathbf{x})$ . Для этого находятся такие параметры преобразования  $T_a: \mathbf{P}_M \rightarrow \mathbf{P}_O$ , при которых минимизируется невязка проекции изображения  $f(T_a(\mathbf{x}))$  на систему базисных функций и модельных образов  $\mathbf{w}^k$ :

$$\min_{\substack{\mathbf{a} \in \Omega_a, \\ k=1..K}} \left\| \mathbf{w}^k - \mathbf{G}_\Psi^{-1}(\Psi(\mathbf{x}), f(T_a(\mathbf{x}))) \right\|_\Psi^2, \quad (4)$$

где  $\Omega_a$  – множество допустимых векторов  $\mathbf{a}$ , задающее ограничения на преобразование;  $\mathbf{w}^k$  – весовой вектор образа  $k$  в модели; норма определена как  $\|\tilde{\mathbf{w}} - \tilde{\mathbf{v}}\|_\Psi^2 = \|\mathbf{w}/\|\mathbf{w}\|_\Psi - \mathbf{v}/\|\mathbf{v}\|_\Psi\|_\Psi^2$ ,  $\|\mathbf{w}\|_\Psi \neq 0$ ,  $\|\mathbf{v}\|_\Psi \neq 0$ , где  $\|\mathbf{w} - \mathbf{v}\|_\Psi^2 = (\mathbf{w} - \mathbf{v})^T \mathbf{G}_\Psi (\mathbf{w} - \mathbf{v})$ ,  $\mathbf{w}$  и  $\mathbf{v}$  – векторы в базисном пространстве [1].

Для инициализации процесса оптимизации (4) задается начальный вектор  $\mathbf{a}$  на основе приближенной оценки характерных точек  $\mathbf{P}_O$  на этапе грубой локализации объекта. Оптимизация методом доверительных областей [5] состоит в решении следующей квадратичной подзадачи на каждой итерации  $i$ :

$$\min_{\mathbf{d}_i \in \Omega} \mathbf{g}_i^T \mathbf{d}_i + 1/2 \mathbf{d}_i^T \mathbf{A}_i \mathbf{d}_i, \quad \|\mathbf{D} \mathbf{d}_i\|_2 \leq \Delta_i,$$

где  $\mathbf{g}_i = -2\mathbf{J}^T(\mathbf{a}_i)\mathbf{G}_\Psi(\tilde{\mathbf{w}}^k - \tilde{\mathbf{v}}(\mathbf{a}_i))$  – градиент функции невязки для (4);  $\tilde{\mathbf{v}} = \mathbf{v}/\sqrt{\mathbf{v}^T\mathbf{G}_\Psi\mathbf{v}}$  – нормированный весовой вектор,  $\mathbf{v}(\mathbf{a}_i) = \mathbf{G}_\Psi^{-1}(\Psi(\mathbf{x}), f(T_{\mathbf{a}_i}(\mathbf{x})))$  – образ изображения в базисе;  $\mathbf{A}_i = 2\mathbf{J}^T(\mathbf{a}_i)\mathbf{G}_\Psi\mathbf{J}(\mathbf{a}_i)$  – симметричная матрица, аппроксимирующая матрицу Гессе;  $\mathbf{J}(\mathbf{a}) = \partial\tilde{\mathbf{v}}(\mathbf{a})/\partial\mathbf{a}$  – матрица Якоби;  $\mathbf{d}_i$  – решение подзадачи, приращение к текущему вектору параметров,  $\mathbf{a}_{i+1} = \mathbf{a}_i + \mathbf{d}_i$ ;  $\mathbf{D}$  – диагональная матрица, задающая эллипсоидную область;  $\Delta_i$  – размер доверительной области, корректируемый на каждой итерации в зависимости от степени сходимости. Элементы матрицы Якоби для нормированной проекции  $\tilde{\mathbf{v}}$  находятся из выражения:

$$\partial\tilde{\mathbf{v}}(\mathbf{a})/\partial\mathbf{a} = \left( (\mathbf{v}^T\mathbf{G}_\Psi\mathbf{v}) \frac{\partial\mathbf{v}}{\partial\mathbf{a}} - \left( \mathbf{v}^T\mathbf{G}_\Psi \frac{\partial\mathbf{v}}{\partial\mathbf{a}} \right) \mathbf{v} \right) (\mathbf{v}^T\mathbf{G}_\Psi\mathbf{v})^{-\frac{2}{3}},$$

где

$$\frac{\partial\mathbf{v}(\mathbf{a})}{\partial\mathbf{a}} = \mathbf{G}_\Psi^{-1} \left[ \left| \mathbf{J}(T_{\mathbf{a}})(\mathbf{x}) \right| \frac{\partial \left| \mathbf{J}(T_{\mathbf{a}}^{-1})(T_{\mathbf{a}}(\mathbf{x})) \right|}{\partial\mathbf{a}} \Psi(\mathbf{x}) + \frac{\partial\Psi(\mathbf{x})}{\partial\mathbf{x}} \frac{\partial T_{\mathbf{a}}^{-1}(T_{\mathbf{a}}(\mathbf{x}))}{\partial\mathbf{a}}, f(T_{\mathbf{a}}(\mathbf{x})) \right],$$

$\left| \mathbf{J}(T_{\mathbf{a}}^{-1})(\mathbf{x}) \right|$  – модуль якобиана преобразования.

Заметим, что скорость сходимости (4) не зависит от площади целевого объекта на исходном изображении, на нее влияет только точность начального приближения вектора  $\mathbf{a}$ .

**3. Результаты.** Для построения текстурных моделей лица в качестве базисных функций используется семейство нечетных функций Габора, удовлетворяющих условию “допустимости вейвлетов” [1]. Проекция (2) изображения в таком базисе устойчива к неравномерной освещенности сцены, вследствие того, что базисные функции локальны и имеют нулевое среднее. Класс  $C_{e,T}$  изображений областей лица определим относительно аффинного преобразования

$$T_{\mathbf{a}}(\mathbf{x}) = \begin{pmatrix} s_x & s_{xy} \\ 0 & \gamma s_x \end{pmatrix} \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} \left( \mathbf{x} - \begin{pmatrix} c_x \\ c_y \end{pmatrix} \right), \quad s_x > 0, \quad \gamma > 0, \quad \text{с вектором параметров}$$

$\mathbf{a} = (c_x, c_y, s_x, \gamma, s_{xy}, \varphi)^T$ . Для каждого элемента (левого глаза, правого глаза и рта) и всего лица в целом строится своя текстурная модель по обучающей выборке из 100 изображений. Для модели лица площадью 48x50 точек используется 70 базисных функций, для моделей элементов площадью 48x30 точек – по 52 функции. Модели элементов иерархически связаны с моделью всего лица.

На первом этапе локализации путем оптимизации (4) ищется наилучшее соответствие образов модели всего лица изображению, и находятся параметры аффинного преобразования  $\mathbf{a}_{\text{face}}$ . На втором этапе производится поиск каждого элемента лица путем оптимизации (4) соответствующей модели по параметрам  $\mathbf{a}_{\text{el}}$ . Для первоначальной оценки  $\tilde{\mathbf{a}}_{\text{el}}$  используется вектор  $\tilde{\mathbf{P}}_{\text{el}} = T_{\mathbf{a}}(\mathbf{P}_M, \mathbf{a}_{\text{face}})$ , полученный на первом этапе. Итоговый вектор характерных точек элемента находится из выражения  $\mathbf{P}_{\text{el}} = T_{\mathbf{a}}(\mathbf{P}_M, \mathbf{a}_{\text{el}})$ .

Реализованные алгоритмы тестировались на размеченной базе изображений 1000 лиц с разрешением от 180\*180 до 1200\*1600 точек, снятых в различных условиях освещения и разными камерами. Для начального положения лица, найденного с точностью 15 % от расстояния между глазами, получены следующие средние ошибки локализации точек в процентах от расстояния между зрачками: центры зрачков – 3,13 %, внешние уголки глаз – 4,13 %, внутренние уголки глаз – 3,63 %, центр верхней и нижней губы – 4,83 %, уголки губ – 6,36 %. Для сравнения, в работе [2] для двух баз изображений лиц с разрешениями 320\*243 и 256\*384, с расстоянием около 55 точек между глазами при сравнимом времени оценивания получена суммарная средняя ошибка, равная 10 точкам (18 % от расстояния между глазами).

**Заключение.** Описанные методы развивают и обобщают работы [1-3]: обобщен подход для моделирования более широкого класса изображений лиц [4]; текстовая модель строится на множестве изображений объекта, а не на одном выбранном; модели элементов описывают области, а не окрестности точек лица; оптимизация (4) производится в системе координат модели, а не исходного изображения; используется устойчивый метод доверительных областей и оптимизация по нормированному расстоянию. Результаты экспериментов показывают, что основным недостатком метода является риск нахождения неверного локального минимума (4). Частично проблема решается путем уменьшения количества базисных функций в модели, но при этом увеличивается средняя ошибка. Для уточнения результата можно использовать нелинейное преобразование  $T_a$ . С помощью предложенного подхода можно оценивать также качественные характеристики, такие как раса и пол, так как при оптимизации (4) находится ближайший образ в модели, с которым на этапе обучения можно связать соответствующие характеристики.

#### БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Feris R., Krueger V., Cesar Jr.R.* Efficient real-time face tracking in wavelet subspace // Proc. IEEE ICCV Workshop on Recog., Anal. and Tracking of Faces and Gestures in Real-time Systems. Vancouver, 2001.
2. *Feris R., Gemmell J., Toyama K. et al.* Hierarchical wavelet networks for facial feature localization // Proc. IEEE ICCV Workshop on Recog., Anal. and Tracking of Faces and Gestures in Real-time Systems. Vancouver, 2001.
3. *Гнеушев А.Н.* Построение и оптимизация текстурно-геометрической модели изображения лица в пространстве базисных функций Габора // Изв. АН. Теория и системы управления. – 2007. – № 3. – С. 85–96.
4. *Гнеушев А.Н.* Оптимизация текстурно-геометрической модели изображения для оценивания параметров лица // Автоматика и телемеханика. – 2012. – № 1. – С. 159-168.
5. *Byrd R., Schnabel R.B., Shultz G.A.* A trust region algorithm for nonlinearly constrained optimization // SIAM J. Numer. Anal. – 1987. – № 24. – P. 1152-1170.

Статью рекомендовал к опубликованию д.ф.-м.н. В.И. Цурков.

**Гнеушев Александр Николаевич** – Федеральное государственное бюджетное учреждение науки Вычислительный центр им. А.А. Дородницына Российской академии наук; e-mail: gneushev@ccas.ru; 119333, г. Москва, ул. Вавилова, 40; тел.: 84959307558; отдел сложных систем; научный сотрудник; к.ф.-м.н.

**Gneushev Alexander Nikolayevich** – Institution of Russian Academy of Sciences Dorodnicyn Computing Centre of RAS; e-mail: gneushev@ccas.ru; 40, Vavilov street, Moscow, 119333, Russia; phone: +74959307558; department of complex systems; leading researcher; cand. of phis.-math. sc.