

РАСПОЗНАВАНИЕ РУКОПИСНЫХ СИМВОЛОВ НА ОСНОВЕ АНАЛИЗА ДЕСКРИПТОРОВ ФУНКЦИЙ ДЛИНЫ ХОРДЫ

С. А. Запрягаев, А. И. Сорокин

Воронежский государственный университет

Поступила в редакцию 10.10.2009

Аннотация. В работе рассматривается подход к распознаванию рукописных символов, основанный на получении признаков объектов. Описаны способы построения инвариантных дескрипторов функций длины хорды и эллиптических дескрипторов Фурье. Также в работе приведены некоторые подходы к уменьшению размерности пространства признаков на основе метода главных компонент и методов линейного дискриминантного анализа. Сформулированы алгоритмы обучения и распознавания объектов.

Ключевые слова: Распознавание, признаки, рукописные, символы, уменьшение размерности.

Annotation. In the article a feature-based approach to handwritten character recognition is presented. The methods of building invariant chord-length descriptors and elliptic Fourier descriptors are discussed. The article describes some approaches to the reduction of feature space dimension by means of Principal Component Analysis and Linear Discriminant Analysis. The algorithms of learning and object recognition are presented.

Keywords: recognition, handwritten character.

ВВЕДЕНИЕ

Задачи, связанные с распознаванием объектов на плоском изображении, имеют высокую практическую значимость. Одним из приложений распознавания объектов является распознавание текстов, в том числе рукописных. В данной работе рассматривается подход к распознаванию рукописных символов, основанный на получении признаков объектов путем построения инвариантных дескрипторов функций длины хорды [3,4]. В качестве исходных данных для построения инвариантных дескрипторов используется внешний контур объекта (символа), представляющий собой периодическую функцию, описывающую форму объекта

$$f(t) = \{x(t), y(t)\} \subset R^2$$

или, в дискретном случае, набор координат

$$f_i = (x_i, y_i) \in R^2.$$

Описание объектов посредством инвариантных дескрипторов функций длины хорды позволяет получить относительно большое по сравнению, например, с эллиптическими дескрипторами Фурье, количество признаков каждого объекта, что приводит к необходимости отбора значимых признаков, например, ме-

тодом главных компонент или линейным дискриминантным анализом.

1. ИНВАРИАНТНЫЕ ДЕСКРИПТОРЫ ФУРЬЕ

Понятие дескрипторов Фурье было введено Косгриффом (Cosgriff) в 1960 году применительно к использованию преобразований Фурье в задачах описания геометрических фигур. Основная идея данного метода — представление контура (замкнутой кривой) конечным набором коэффициентов разложения Фурье. Дескрипторы Фурье широко используются в задачах распознавания образов, при обработке медицинских, картографических и т.п.

1.1. ДЕСКРИПТОРЫ ФУРЬЕ

Рассмотрим периодическую функцию действительной переменной $f(t)$ с периодом T . Для этой функции справедливо следующее представление в виде разложения в тригонометрический ряд Фурье:

$$f(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos(k\omega t) + b_k \sin(k\omega t)), \quad (1)$$

где $\omega = \frac{2\pi}{T}$ — фундаментальная частота.

$$a_k = \frac{2}{T} \int_0^T f(t) \cos(k\omega t) dt, \quad (2)$$

$$b_k = \frac{2}{T} \int_0^T f(t) \sin(k\omega t) dt.$$

Коэффициенты, участвующие в разложении (1) и описывающие (descript) функцию $f(t)$, называют дескрипторами Фурье функции $f(t)$.

1.2. СВОЙСТВА ДЕСКРИПТОРОВ ФУРЬЕ ФУНКЦИИ ДЕЙСТВИТЕЛЬНОЙ ПЕРЕМЕННОЙ

Покажем, как изменяются дескрипторы с изменением описываемой ими функции.

1. Рассмотрим функцию $f'(t) = f(t + \alpha)$, полученную из $f(t)$ сдвигом на константу α . Дескрипторы Фурье для этой функции будут заданы согласно (2) следующим образом:

$$\begin{aligned} a'_k &= \frac{2}{T} \int_0^T f'(t) \cos(k\omega t) dt = \\ &= \frac{2}{T} \int_0^T f(t + \alpha) \cos(k\omega t) dt = \\ &= \frac{2}{T} \int_\alpha^{T+\alpha} f(s) \cos(k\omega s - k\omega\alpha) ds = \\ &= \frac{2}{T} \int_\alpha^{T+\alpha} f(s) [\cos(k\omega s) \cos(k\omega\alpha) + \\ &\quad + \sin(k\omega s) \sin(k\omega\alpha)] ds = \\ &= a_k \cos(k\omega\alpha) + b_k \sin(k\omega\alpha). \end{aligned}$$

Аналогично,

$$\begin{aligned} b'_k &= \frac{2}{T} \int_0^T f'(t) \sin(k\omega t) dt = \\ &= \frac{2}{T} \int_0^T f(t + \alpha) \sin(k\omega t) dt = \\ &= \frac{2}{T} \int_\alpha^{T+\alpha} f(s) \sin(k\omega s - k\omega\alpha) ds = \\ &= \frac{2}{T} \int_\alpha^{T+\alpha} f(s) [\sin(k\omega s) \cos(k\omega\alpha) - \\ &\quad - \cos(k\omega s) \sin(k\omega\alpha)] ds = \\ &= b_k \cos(k\omega\alpha) - a_k \sin(k\omega\alpha). \end{aligned}$$

Из представленных преобразований видно, что сумма $a_k^2 + b_k^2$ устойчива к сдвигу функции ($f'(t) = f(t + \alpha)$), т.е.

$$a_k'^2 + b_k'^2 = a_k^2 + b_k^2. \quad (3)$$

Данный факт может быть использован при построении инвариантных дескрипторов. Действительно,

$$\begin{aligned} a_k'^2 + b_k'^2 &= (a_k \cos(k\omega\alpha) + b_k \sin(k\omega\alpha))^2 + \\ &\quad + (b_k \cos(k\omega\alpha) - a_k \sin(k\omega\alpha))^2 = \\ &= a_k^2 \cos^2(k\omega\alpha) + 2a_k b_k \cos(k\omega\alpha) \sin(k\omega\alpha) + \\ &\quad + b_k^2 \sin^2(k\omega\alpha) + b_k^2 \cos^2(k\omega\alpha) - \\ &\quad - 2b_k a_k \cos(k\omega\alpha) \sin(k\omega\alpha) + a_k^2 \sin^2(k\omega\alpha) = \\ &= a_k^2 + b_k^2. \end{aligned}$$

2. Для функции $f'(t) = sf(t)$, где s — некоторое ненулевое число соответствующие дескрипторы a'_k и b'_k равны:

$$a'_k = sa_k, \quad b'_k = sb_k.$$

3. Найдём дескрипторы a'_k и b'_k функции $f'(t) = f(t) + f_0$:

$$\begin{aligned} a'_k &= \frac{2}{T} \int_0^T f'(t) \cos(k\omega t) dt = \\ &= \frac{2}{T} \int_0^T [f(t) \cos(k\omega t) + f_0 \cos(k\omega t)] dt = \\ &= a_k + \frac{2}{T} f_0 (\sin(k\omega T) - \sin(0)) = a_k, \\ a'_0 &= \frac{2}{T} \int_0^T f'(t) \cos(k\omega t) dt = \\ &= \frac{2}{T} \int_0^T [f(t) + f_0] dt = a_k + 2f_0, \\ a'_0 &= a_0 + 2f_0, \quad a'_k = a_k \quad k > 1 \quad b'_k = b_k. \end{aligned}$$

Таким образом, дескрипторы функций $f'(t)$ и $f(t)$ совпадают, за исключением a'_0 и a_0 .

1.3. ПРЕДСТАВЛЕНИЕ ДИСКРЕТИЗИРОВАННОЙ ФУНКЦИИ ДЕЙСТВИТЕЛЬНОЙ ПЕРЕМЕННОЙ ДЕСКРИПТОРАМИ ФУРЬЕ

Пусть периодическая функция $f(t)$ с периодом T задана m дискретными значениями:

$$f_j = f(j\tau), \quad j = 1..m, \quad \tau = \frac{T}{m}.$$

Каждое слагаемое вида $a_k \cos(k\omega t) + b_k \sin(k\omega t)$ в представлении (1) — периодическая тригонометрическая функция переменной t с частотой $\frac{k}{T}$.

Как следует из теоремы Найквиста—Шеннона [1] в случае m дискретных значений максимально достижимая частота восстановленной по этим значениям функции — $\frac{m}{2T}$. Найдём

номер слагаемого вида $a_k \cos(k\omega t) + b_k \sin(k\omega t)$, обеспечивающего максимально достижимую частоту:

$$\frac{k}{T} = \frac{m}{2T}, \quad k = \frac{m}{2}.$$

Таким образом, в случае задания $f(t)$ m дискретными значениями имеет смысл учитывать для описания функции не более $m/2$ дескрипторов.

Запишем представление приближения $f(t)$.

$$f(t) \approx \frac{a_0}{2} + \sum_{k=1}^{m/2} (a_k \cos(k\omega t) + b_k \sin(k\omega t)), \quad (4) \quad \text{где}$$

где (по методу трапеций)

$$a_k = \frac{2}{m} \sum_{j=1}^m f_j \cos(k\omega j\tau),$$

$$b_k = \frac{2}{m} \sum_{j=1}^m f_j \sin(k\omega j\tau), \quad \tau = \frac{T}{m}.$$

1.4. ЭЛЛИПТИЧЕСКИЕ ДЕСКРИПТОРЫ ФУРЬЕ, ПРЕДСТАВЛЕНИЕ ЗАМКНУТОЙ ПЛОСКОЙ КРИВОЙ

Покажем, как можно использовать разложение (4) для описания некоторой плоской кривой. Кривая $f(t)$ на плоскости xOy может быть задана параметрически при помощи двух уравнений [2], описывающих её координаты: $x = x(t)$, $y = y(t)$, $f(t) = \{x(t), y(t)\} \subset R^2$, где t – переменный параметр.

Выберем некоторую точку O , лежащую на непрерывной кривой и одно из двух возможных направлений, тогда каждой точке A кривой может быть поставлено в соответствие некоторое число s , равное по модулю длине дуги OA , взятой со знаком «+» или «-» в зависимости от направления. Если кривая описывается при помощи переменного параметра s , то такую параметризацию называют естественной (рис. 1).

В случае естественно параметризованной замкнутой кривой функции $x(t)$ и $y(t)$ – пе-

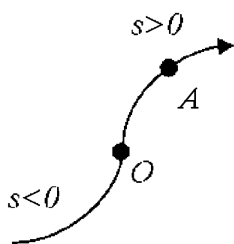


Рис. 1. Естественная параметризация

риодические, с периодом T , равным длине кривой и для них справедливо представление (1):

$$x(t) = \frac{a_{x0}}{2} + \sum_{k=1}^{\infty} (a_{xk} \cos(k\omega t) + b_{xk} \sin(k\omega t)),$$

$$y(t) = \frac{a_{y0}}{2} + \sum_{k=1}^{\infty} (a_{yk} \cos(k\omega t) + b_{yk} \sin(k\omega t)), \quad (5)$$

$$a_{xk} = \frac{2}{T} \int_0^T x(t) \cos(k\omega t) dt,$$

$$b_{xk} = \frac{2}{T} \int_0^T x(t) \sin(k\omega t) dt,$$

$$a_{yk} = \frac{2}{T} \int_0^T y(t) \cos(k\omega t) dt,$$

$$b_{yk} = \frac{2}{T} \int_0^T y(t) \sin(k\omega t) dt.$$

1.5. ИНВАРИАНТНЫЕ ЭЛЛИПТИЧЕСКИЕ ДЕСКРИПТОРЫ ФУРЬЕ

Перепишем представление (5) для функции $f(t)$ в виде:

$$f(t) = \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \frac{1}{2} \begin{bmatrix} a_{x0} \\ a_{y0} \end{bmatrix} + \sum_{k=1}^{\infty} \begin{bmatrix} a_{xk} & b_{xk} \\ a_{yk} & b_{yk} \end{bmatrix} \begin{bmatrix} \cos(k\omega t) \\ \sin(k\omega t) \end{bmatrix}.$$

Пусть $f'(t)$ получена из $f(t)$ поворотом на угол ρ вокруг точки $\left(\frac{a_{x0}}{2}, \frac{a_{y0}}{2}\right)$, тогда

$$f'(t) = \begin{bmatrix} x'(t) \\ y'(t) \end{bmatrix} = \frac{1}{2} \begin{bmatrix} a_{x0} \\ a_{y0} \end{bmatrix} + \begin{bmatrix} \cos \rho & \sin \rho \\ -\sin \rho & \cos \rho \end{bmatrix} \sum_{k=1}^{\infty} \begin{bmatrix} a_{xk} & b_{xk} \\ a_{yk} & b_{yk} \end{bmatrix} \begin{bmatrix} \cos(k\omega t) \\ \sin(k\omega t) \end{bmatrix}.$$

Вычислим соответствующие дескрипторы a'_{xk} , a'_{yk} , b'_{xk} , b'_{yk} :

$$a'_{xk} = a_{xk} \cos \rho + a_{yk} \sin \rho,$$

$$b'_{xk} = b_{xk} \cos \rho + b_{yk} \sin \rho,$$

$$a'_{yk} = a_{xk} (-\sin \rho) + a_{yk} \cos \rho,$$

$$b'_{yk} = b_{xk} (-\sin \rho) + b_{yk} \cos \rho.$$

Заметим, что

$$a'_{xk}{}^2 + a'_{yk}{}^2 = a_{xk}^2 + a_{yk}^2 \text{ и } b'_{xk}{}^2 + b'_{yk}{}^2 = b_{xk}^2 + b_{yk}^2 \quad (6)$$

Таким образом, суммы дескрипторов (6) инвариантны относительно поворота исходной функции в плоскости вокруг ее «центра тяжести» $\left(\frac{a_{x0}}{2}, \frac{a_{y0}}{2}\right)$. Тем не менее, эти суммы зависят от выбора начальной точки отсчёта при параметризации. Используя (3) и (6) можно построить сумму, не зависящую ни от поворота кривой в плоскости, ни от выбора точки отсчёта.

Пусть заданы плоские кривые:

$$f'(t) = f(t + \alpha),$$

$$f''(t) = \frac{1}{2} \begin{bmatrix} a_{x0} \\ a_{y0} \end{bmatrix} + \begin{bmatrix} \cos \rho & \sin \rho \\ -\sin \rho & \cos \rho \end{bmatrix} \sum_{k=1}^{\infty} \begin{bmatrix} a'_{xk} & b'_{xk} \\ a'_{yk} & b'_{yk} \end{bmatrix} \begin{bmatrix} \cos(k\omega t) \\ \sin(k\omega t) \end{bmatrix}.$$

Обозначим соответствующие им дескрипторы тем же количеством штрихов. Рассмотрим сумму:

$$\begin{aligned} & (a''_{xk} + a''_{yk}) + (b''_{xk} + b''_{yk}) \stackrel{(6)}{=} \\ & \stackrel{(6)}{=} (a'_{xk}{}^2 + a'_{yk}{}^2) + (b'_{xk}{}^2 + b'_{yk}{}^2) \\ & \stackrel{(3)}{=} (a_{xk}^2 + a_{yk}^2) + (a_{xk}^2 + b_{xk}^2) + (a_{yk}^2 + b_{yk}^2) \\ & \stackrel{(3)}{=} (a_{xk}^2 + b_{xk}^2) + (a_{yk}^2 + b_{yk}^2). \end{aligned}$$

Сумма

$$a_{xk}^2 + a_{yk}^2 + b_{xk}^2 + b_{yk}^2 \quad (7)$$

не зависит ни от выбора начальной точки параметризации на кривой $f(t)$, ни от поворота вокруг $\left(\frac{a_{x0}}{2}, \frac{a_{y0}}{2}\right)$. При $k > 0$ на сумму (7) не влияет сдвиг кривой $f(t)$ на плоскости, т.е. изменение положение «центра тяжести» $\left(\frac{a_{x0}}{2}, \frac{a_{y0}}{2}\right)$. Тем не менее, эта сумма не устойчива к масштабированию. Рассмотрим функцию $f'''(t)$, полученную из $f''(t)$ растяжением в s раз.

$$f'''(t) = \frac{1}{2} \begin{bmatrix} a_{x0} \\ a_{y0} \end{bmatrix} + s \sum_{k=1}^{\infty} \begin{bmatrix} a''_{xk} & b''_{xk} \\ a''_{yk} & b''_{yk} \end{bmatrix} \begin{bmatrix} \cos(k\omega t) \\ \sin(k\omega t) \end{bmatrix}.$$

В этом случае,

$$\begin{aligned} & a'''_{xk}{}^2 + a'''_{yk}{}^2 + b'''_{xk}{}^2 + b'''_{yk}{}^2 = \\ & = s(a''_{xk}{}^2 + a''_{yk}{}^2 + b''_{xk}{}^2 + b''_{yk}{}^2), \end{aligned}$$

в частности,

$$\begin{aligned} & a'''_{x1}{}^2 + a'''_{y1}{}^2 + b'''_{x1}{}^2 + b'''_{y1}{}^2 = \\ & = s(a''_{x1}{}^2 + a''_{y1}{}^2 + b''_{x1}{}^2 + b''_{y1}{}^2). \end{aligned}$$

Используя последние два уравнения, получим инвариантные дескрипторы, устойчивые к сдвигу, выбору начальной точки параметризации, масштабу и повороту в любых комбинациях.

$$d_k = \frac{a_{xk}^2 + a_{yk}^2 + b_{xk}^2 + b_{yk}^2}{a_{x1}^2 + a_{y1}^2 + b_{x1}^2 + b_{y1}^2}, \quad k \geq 2 \quad (8)$$

Таким образом, получим вектор инвариантных дескрипторов объекта:

$$D = \{d_2, \dots, d_{m/2}\} \subset R^{m/2-1}.$$

2. ПРИМЕНЕНИЕ ФУНКЦИИ ДЛИНЫ ХОРДЫ ДЛЯ ПОЛУЧЕНИЯ ХАРАКТЕРНЫХ ПРИЗНАКОВ ОБЪЕКТА

2.1. ПОНЯТИЕ О ФУНКЦИИ ДЛИНЫ ХОРДЫ

Бин Ванг и Чаоджян Син в своих работах [3, 4] предложили использовать для описания замкнутых плоских кривых функцию длины хорды (chord-length function). Суть метода состоит в следующем. Пусть $f(t) = \{x(t), y(t)\} \subset R^2$ — естественно параметризованная замкнутая плоская кривая, T — ее периметр (а следовательно, и период функций $x(t), y(t)$), точка $O = (x(0), y(0))$ — точка отсчета в естественной параметризации. Функция длины хорды определяется как:

$$L_k^j \{f\}(t) = \left\| f\left(t + j \frac{T}{k}\right) - f(t) \right\|, \quad j = 1 \dots k-1,$$

где k — некоторая целочисленная константа, а $\|\cdot\|$ — евклидова норма.

Функции $L_k^j \{f\}(t)$ являются периодически, с периодом T .

Для того чтобы пояснить смысл функции длины хорды, рассмотрим случай для $k=4$. зафиксируем некоторый параметр $t_0 \in [0, T]$. Разобьем замкнутую кривую, определяемую

$f(t)$, точками $s_j = f\left(t_0 + j \frac{T}{k}\right)$, $j = 0 \dots k-1$ на 4

дуги равной длины. Проведем 3 хорды из точки s_0 в каждую из точек s_1, s_2, s_3 . Длины хорд $\overline{s_0s_1}$,

$\overline{s_0s_2}, \overline{s_0s_3}$ будут совпадать со значениями функций длин хорд в t_0 . Например, $L_4^1 \{f\}(t_0) = |s_0s_1|$,

$L_4^3 \{f\}(t_0) = |s_0s_3|$.

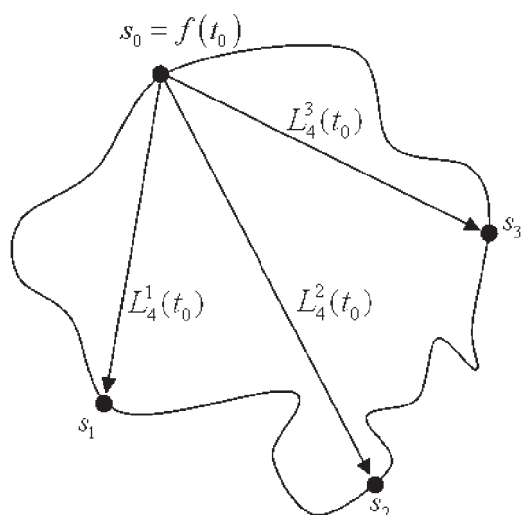


Рис. 2. Функция длины хорды.

Заметим, что

$$L_k^j \{f\}(t) = L_k^{k-j} \{f\}(t + j \frac{T}{k}).$$

$$L_k^{k-j} \{f\}(t + j \frac{T}{k}) =$$

$$= \left\| f \left(t + j \frac{T}{k} + (k-j) \frac{T}{k} \right) - f \left(t + j \frac{T}{k} \right) \right\| =$$

$$= \left\| f(t+T) - f \left(t + j \frac{T}{k} \right) \right\| =$$

$$= \left\| f \left(t + j \frac{T}{k} \right) - f(t) \right\| = L_k^j \{f\}(t).$$

Таким образом, периодические функции $L_k^j \{f\}(t)$ и $L_k^{k-j} \{f\}$ будут отличаться друг от друга только сдвигом на $j \frac{T}{k}$. Следовательно, для описания плоской кривой достаточно использовать только $k/2$ функций $L_k^j \{f\}(t)$.

Набор функций длины хорды $L_k^j \{f\}$ позволяет получить содержательные характеристики объекта, захватывая более «мелкие» детали при малых значениях параметра j и более общие признаки объекта при j близком к $k/2$.

2.2. ИНВАРИАНТНЫЕ ДЕСКРИПТОРЫ ФУНКЦИИ ДЛИНЫ ХОРДЫ

Перейдём к дискретному случаю описания кривой с помощью функций длины хорды. Пусть естественно параметризованная замкнутая кривая $f(t)$ периметра T , задана m равностоящими (в смысле длины дуги кривой) точками:

$$f_i = f(t_i), \quad t_i = i \frac{T}{m}, \quad i = 0..m-1.$$

Для некоторого параметра k построим $k/2$ функций длины хорды:

$$L_k^j \{f\}(t), \quad j = 1..k/2.$$

В дискретном случае каждая из функций $L_k^j \{f\}(t)$ будет определяться набором m дискретных значений:

$$L_k^j \{f\}(t_i) = \left\| f \left(t_i + j \frac{T}{k} \right) - f(t_i) \right\|, \quad (9)$$

$$i = 0..m-1.$$

Применим к $L_k^j \{f\}(t)$, заданной m дискретными значениями (9) преобразование (4):

$$L_k^j \{f\}(t) = \frac{a_0}{2} + \sum_{s=1}^{m/2} (a_s \cos(s\omega t) + b_s \sin(s\omega t)),$$

где

$$a_s = \frac{2}{m} \sum_{l=1}^m L_k^j \{f\}(t_l) \cos(s\omega t_l),$$

$$b_s = \frac{2}{m} \sum_{l=1}^m L_k^j \{f\}(t_l) \sin(s\omega t_l),$$

$$\tau = \frac{T}{m}.$$

Функции $L_k^j \{f\}(t)$ по определению устойчивы к повороту исходной кривой $f(t)$, а также изменению ее положения, тем не менее, они зависят от выбора начальной точки отсчета $(x(0), y(0))$ и растяжения/сжатия исходной функции $f(t)$.

Согласно (3) сумма $a_s^2 + b_s^2$ инвариантна к сдвигу аргумента функции $L_k^j \{f\}(t)$. Заметим, что сдвиг в параметризации кривой $f'(t) = f(t + \alpha)$ приведет к такому же сдвигу параметризации функции длины хорды и $L_k^j \{f'(t)\} = L_k^j \{f\}(t + \alpha)$.

Аналогично, в случае сжатия/растяжения кривой $f'(t) = \gamma f(t)$: $L_k^j \{f'(t)\} = \gamma L_k^j \{f\}(t)$. Соответствующие дескрипторы Фурье: $a'_s = \gamma a_s$, $b'_s = \gamma b_s$.

Разделим каждую из сумм $a_s^2 + b_s^2$ на $a_0^2 + b_0^2$, получим $m/2$ дескрипторов вида:

$$g^{(s)} \{L_k^j\} = \frac{a_s^2 + b_s^2}{a_0^2 + b_0^2}, \quad s = 1..m/2.$$

Каждый из дескрипторов g_s устойчив к сдвигу, смещению, повороту и масштабированию исходной кривой $f(t)$.

Проводя описанную выше процедуру при заданном параметре k для функции $f(t)$, мы

получим $k/2$ функций длины хорды $L_k^j \{f\}(t)$, $j = 1..k/2$, для каждой из которых определены $m/2$ инвариантных дескрипторов Фурье

$$G_k^j = \{g^{(1)}\{L_k^j\}, \dots, g^{(m/2)}\{L_k^j\}\} \subset R^{m/2}. \quad (11)$$

3. РАСПОЗНАВАНИЕ ОБЪЕКТА ПО ПРИЗНАКАМ

Рассмотренные выше методы получения инвариантных дескрипторов были с успехом применены для распознавания символов рукописного текста. Рассмотрим подходы, использованные для обучения системы распознавания рукописного текста, а также методы представления и хранения образцов данных.

3.1. ВЫДЕЛЕНИЕ ЗНАЧИМЫХ ПРИЗНАКОВ МЕТОДОМ ГЛАВНЫХ КОМПОНЕНТ

Оценим количество параметров, получаемое при построении инвариантных дескрипторов функции длины хорды. Для определенности рассмотрим замкнутую границу объекта, представленную $m = 50$ равностоящими точками:

$$f_i = (x_i, y_i), \quad i = 0..m-1, \quad m = 50.$$

Зафиксируем значение параметра k , используемого при построении набора функций длины хорды. Полагая $k = 16$, получим 8 функций длины хорды вида (9):

$$L_k^j \{f\}, \quad j = 1..8.$$

Таким образом, инвариантные дескрипторы вида (10)

$$G_k^j = \{g^{(1)}\{L_k^j\}, \dots, g^{(25)}\{L_k^j\}\} \subset R^{25}, \quad j = 1..8$$

будут представлены $m/2 \times k/2$ скалярными значениями, т.е. в случае $m = 50$ и $k = 16$ получим 200 действительных чисел – инвариантных признаков объекта.

Построим одномерный вектор, содержащий признаки объекта, последовательно перечислив инвариантные дескрипторы Фурье для каждой из 8 функций длины хорды:

$$x = \left(g^{(1)}\{L_k^1\}, \dots, g^{(m/2)}\{L_k^1\}, g^{(1)} \times \{L_k^2\}, \dots, g^{(m/2)}\{L_k^2\}, \dots, g^{(m/2)}\{L_k^j\} \right)^T. \quad (12)$$

Будем использовать представление (12) для задания признаков объектов, используемых для обучения системы, а также объектов, проверяемых на принадлежность некоторому классу, т.е. распознаваемых объектов. Пусть представление (12) получено для каждого объекта из

набора объектов, предназначенных для обучения системы. Для успешного решения задачи распознавания объектов, необходим некоторый анализ обучающего множества образцов, связанный с выделением наиболее значимых признаков, что в свою очередь может привести к более компактному хранению обучающих данных посредством уменьшения размерности.

Часто применяемым методом выделения значимых признаков и сжатия объемов хранимой информации является метод главных компонент [5]. Сущность метода главных компонент заключается в поиске линейных подпространств меньшей размерности, в ортогональной проекции на которые среднеквадратичное расстояние между точками максимально.

Приведем формальное описание метода главных компонент. Будем называть X матрицей данных

$$X = [x_1 \quad x_2 \quad \dots \quad x_n]^T \in R^{n \times p}. \quad (13)$$

Строки матрицы X – результаты отдельных опытов (в нашем случае – векторы признаков объектов), а столбцы соответствуют некоторым величинам, определяющим признаки исследуемого объекта.

Для применения метода главных компонент необходимо перейти от (13) к матрице \hat{X} , содержащей центрированные данные:

$$\hat{X} = [x_1 - \bar{x} \quad x_2 - \bar{x} \quad \dots \quad x_n - \bar{x}]^T,$$

где $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$.

Без ограничения общности будем считать, что матрица X содержит центрированные данные.

Метод главных компонент [5] определяет ортогональное линейное преобразование системы координат. Применение метода главных компонент к набору векторов $x_i \in R^p$ позволяет перейти к базису пространства R^p , такому что: первая главная компонента (первый вектор базиса) соответствует направлению, вдоль которого дисперсия векторов исходного набора максимальна. Направление второй главной компоненты (второго вектора базиса) выбирается таким образом, чтобы дисперсия исходных векторов вдоль него была максимальной при условии ортогональности первому вектору базиса. Аналогично определяются остальные векторы базиса.

Поиск главных компонент приводит к задаче диагонализации ковариационной матрицы

$$C = \frac{1}{n-1} X^T X. \quad (14)$$

Векторы главных компонент представляют собой ортонормированный набор собственных векторов e_1, \dots, e_p ковариационной матрицы (14), расположенных в порядке убывания собственных значений $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$. Найдём представление исходного набора векторов $x_i \in R^p$ в этом базисе:

$$x_i = \bar{x} + \sum_{j=1}^p g_{ij} e_j, \quad (15)$$

где $g_{ij} = (x_i - \bar{x}, e_j)$ – j -я координата вектора x_i в новом базисе.

Учитывая, что разложение по первым главным компонентам содержит основную полезную информацию, наряду с (15) используется приближённое представление:

$$x_i \approx \bar{x} + \sum_{j=1}^q g_{ij} e_j, \quad q < p. \quad (16)$$

Заметим, что представление (16) является проекцией вектора $x_i \in R^p$ на линейное подпространство меньшей размерности. Таким образом, (16) описывает процедуру «сжатия» данных.

В [5] приведён критерий информативности метода главных компонент, характеризующий относительную долю дисперсии, вносимой первыми q компонентами:

$$I(q) = \frac{\lambda_1 + \lambda_2 + \dots + \lambda_q}{\lambda_1 + \lambda_2 + \dots + \lambda_q + \dots + \lambda_p}. \quad (17)$$

Выбор количества компонентов, которые можно убрать из рассмотрения для уменьшения размерности данных без особого ущерба, представляет собой нетривиальную задачу. На практике использовалось количество компонент, обеспечивавшее $I(q) \approx 0,9$.

Удобным и численно устойчивым способом получения главных компонент является применение сингулярного разложения матрицы данных $X \in R^{n \times p}$ [6]:

$$X = USV^T,$$

где U, V – ортонормированные матрицы размера $n \times n$ и $p \times p$,

а S – диагональная матрица размером $n \times p$, состоящая из неотрицательных вещественных элементов, упорядоченных по убыванию, т.е. $s_{i,i} \geq s_{j,j}$, если $i < j$.

Столбцы $v_1 \dots v_p$ матрицы V представляют собой собственные векторы матрицы $X^T X$, а

отвечающие им собственные значения соответственно равны $\lambda_i = (s_{i,i})^2$, $i = 1 \dots p$. Таким образом, векторы $v_1 \dots v_p$ являются искомыми главными компонентами. Запишем уравнение для получения проекции вектора пространства R^p на подпространство меньшей размерности. Будем использовать только первые q главных компонент, сформируем матрицу проектора на подпространство меньшей размерности, составив столбцы матрицы из векторов $v_1 \dots v_q$, получим:

$$P = [v_1 \dots v_q] \in R^{p \times q}, \quad (18)$$

$$y = P^T x,$$

где $y \in R^q$ – проекция вектора $x \in R^p$ на линейное подпространство, натянутое на вектора $v_1 \dots v_q$.

Вернёмся к рассмотренному выше примеру, когда каждый объект обучающего набора представляет собой вектор из пространства R^{200} . Пусть совокупность обучающих образцов задана набором векторов признаков вида (12): x_1, \dots, x_n , будем считать, что данные этого набора – центрированы. На практике количество образцов n может быть достаточно велико, будем считать что $n > 300$. Хранение «несжатого» набора данных было бы равносильно хранению порядка 60000 скалярных значений. Пусть долю дисперсии (17) $I(q) > 0,9$ доставляют первые 50 главных компонент, в этом случае «сжатый» набор данных потребовал бы хранения 50×200 скалярных значений, соответствующих координатам главных компонент и 300×50 координат проекций на линейное подпространство, натянутое на 50 главных компонент, всего 25000 скалярных значений, т.е. двукратная экономия. Если считать, что количество главных компонент не увеличится при увеличении объёма обучающего набора до $n = 1000$ образцов, то количество хранимых скалярных значений «несжатых» и «сжатых» данных составило бы соответственно: 200 000 и 60 000, т.е. объём «сжатых» данных был бы уменьшен более чем в 3 раза по сравнению с первоначальным объёмом.

3.2. РАСПОЗНАВАНИЕ ОБЪЕКТА МЕТОДОМ ГЛАВНЫХ КОМПОНЕНТ

Опишем алгоритмы обучения и распознавания объекта по характерным признакам с использованием главных компонент. Пусть обучающий множество объектов задано набором векторов-признаков $x_i \in R^p$, ($i = 1..n$), пред-

ставляющих собой инвариантные эллиптические дескрипторы Фурье (8) или дескрипторы функции длины хорды (12).

Сформулируем алгоритм обучения.

1. Найти среднее данных $\bar{x} \in R^p$:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

2. Построить центрированную матрицу данных:

$$X = [x_1 - \bar{x} \quad x_2 - \bar{x} \quad \dots \quad x_n - \bar{x}]^T \in R^{n \times p}.$$

3. Получить сингулярное (SVD) разложение X :

$$X = USV^T.$$

4. Учитывая, что диагональные элементы матрицы S представляют собой корни собственных значений матрицы ковариации, т.е. $s_{i,i} = \sqrt{\lambda_i}$, $\lambda_i \in \sigma(X^T X)$, найти такое число главных компонент q , при котором

$$I(q) = \frac{\lambda_1 + \lambda_2 + \dots + \lambda_q}{\lambda_1 + \lambda_2 + \dots + \lambda_q + \dots + \lambda_p} > \alpha,$$

5. где $\alpha \in [0, 1]$ — некоторый заранее заданный порог «качества» приближения.

6. Составить матрицу проектора P из первых q столбцов матрицы V :

$$P = [v_1 \dots v_q] \in R^{p \times q}.$$

7. Получить «сжатое» представление для центрированной матрицы данных:

$$Y = XP,$$

где $X \in R^{n \times p}$, $Y \in R^{n \times q}$.

8. Сохранить среднее данных \bar{x} и «сжатую» матрицу образцов Y .

Сформулируем алгоритм распознавания объекта.

1. Построить для классифицируемого объекта дескриптор — вектор признаков $d \in R^p$.

2. Центрировать вектор признаков, используя среднее исходных данных:

$$\hat{d} = d - \bar{x}.$$

3. Получить проекцию $\hat{d} \in R^p$ на линейное подпространство с базисом $v_1 \dots v_q$:

$$u = P^T \hat{d}, \quad u \in R^q.$$

4. Найти вектор-строку матрицы Y , ближайшую к u . Номер строки матрицы Y будет номером образца, ближайшего к распознаваемому объекту.

3.3. УМЕНЬШЕНИЕ РАЗМЕРНОСТИ ДАННЫХ И РАСПОЗНАВАНИЕ ОБЪЕКТОВ МЕТОДОМ ЛИНЕЙНОГО ДИСКРИМИНАНТНОГО АНАЛИЗА

Метода главных компонент позволяет выделить наиболее значимые признаки объектов и понизить размерность пространства признаков, тем не менее, основным его недостатком является игнорирование информации о классах, к которым принадлежат обучающие образцы.

Линейный дискриминантный анализ предлагает иной подход к получению проекций данных по сравнению с методом главных компонент. Линейный дискриминантный анализ позволяет уменьшить расстояние между проекциями векторов признаков, принадлежащему к одному классу и, наоборот, увеличить расстояние между векторами, относящимися к разным классам.

Рассмотрим некоторые способы получения данного преобразования [7]. Будем считать, что обучающие данные сгруппированы по нескольким классам, т.е. векторы признаков $x_1^1, \dots, x_{n_1}^1 \in R^p$ описывают обучающие образцы первого класса, $x_1^2, \dots, x_{n_2}^2 \in R^p$ — второго и т.п. Всего распознаваемых классов — K , а количество образцов

$$N = \sum_{i=1, K} n_i.$$

Построим матрицы ковариации внутри классов и между классами.

Во-первых, найдём средние векторов признаков каждого класса:

$$\mu^i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_j^i.$$

Найдём среднее векторов признаков всех классов:

$$\mu = \frac{1}{N} \sum_{k=1}^K \sum_{i=1}^{n_k} x_i^k.$$

Построим матрицу ковариации внутри класса:

$$S_w = \sum_{k=1}^K \sum_{i=1}^{n_k} (x_i^k - \mu_i)(x_i^k - \mu_i)^T, \quad S_w \in R^{p \times p}.$$

Матрица ковариации между классами имеет вид:

$$S_b = \sum_{k=1}^K (\mu^k - \mu)(\mu^k - \mu)^T, \quad S_b \in R^{p \times p}.$$

Получим матрицу преобразования $W \in R^{p \times q}$, определяющую проекцию в пространство мень-

шей размерности. По аналогии с методом главных компонент будем искать набор ортогональных направлений $w_1 \dots w_q \in R^p$ (столбцов матрицы W), образующих базис в подпространстве меньшей размерности. Аналогично, $w_1 \dots w_q$ упорядочены по убыванию дисперсии каждой из координат проекции.

Вектор w_1 , связанный с наибольшей дисперсией, может быть получен, например, посредством критерия Фишера, заключающегося в максимизации функционала:

$$J(w) = \frac{w^T S_b w}{w^T S_w w}. \quad (19)$$

Функционал (19) показывает, насколько «хорошо» координата проекции, связанная с направлением w , описывает расстояние между классами по сравнению с расстоянием внутри класса на всём множестве обучающих данных.

Решение задачи оптимизации

$$J(w) \rightarrow \max \quad (20)$$

приводит к необходимому условию оптимальности

$$\frac{\partial J}{\partial w} = 0,$$

откуда

$$S_b w = J(w) S_w w.$$

Таким образом, задача поиска оптимального w сводится к нахождению собственного вектора, отвечающего максимальному собственному значению матрицы $S_w^{-1} S_b$, а в качестве набора ортогональных направлений $w_1 \dots w_q \in R^p$ используются q собственных векторов матрицы $S_w^{-1} S_b$, упорядоченных в порядке убывания собственных значений. В [8] приведён подробный обзор численно-стабильных методов получения собственных значений $S_w^{-1} S_b$ без необходимости обращения S_w при помощи QR и SVD разложений.

Алгоритмы обучения и распознавания объектов методом линейного дискриминантного анализа практически эквивалентны алгоритмам на основе метода главных компонент, за исключением способа нахождения проектирующей матрицы W .

4. ПРИМЕР ВЫЧИСЛИТЕЛЬНОГО ЭКСПЕРИМЕНТА

В качестве примера работы алгоритма по распознаванию рукописного текста, представлен рукописный текст, изображенный на рис. 3.

съешь еще этих
мягких французских
булок да выпей чаю

Рис. 3. Входные данные для вычислительного эксперимента

Результат распознавания текста в соответствии с рассмотренными выше алгоритмами следующий:

«съешь еще этих мягких французских булок да выпей чаю»

В данном тексте не распознано только два символа из 44. Статистический результат, накопленный при обработке большего количества информации, демонстрирует среднюю ошибку распознавания рукописного текста порядка 15 %.

ЗАКЛЮЧЕНИЕ

В данной работе рассмотрен подход для распознавания рукописных и печатных символов русского языка. На практике все образцы были представлены набором признаков, полученных построением инвариантных дескрипторов функции длины хорды. Инвариантные эллиптические дескрипторы Фурье (8) при любом выборе числа равноотстоящих точек не обеспечивали лучшего качества распознавания по сравнению с дескрипторами (12). Для уменьшения размерности полученного пространства признаков на практике использовался метод главных компонент.

СПИСОК ЛИТЕРАТУРЫ

1. Nixon M., Aguado A. Feature extraction by shape matching // Feature Extraction and Image Processing. — Oxford; N.Y.: Newnes. — 2002. — P. 247—277
2. Аминов Ю. А. Дифференциальная геометрия и топология кривых. — М.: Наука. Гл. ред. физ.-мат. лит., 1987. — 160 с.
3. Bin Wang, Chaojian Shi S. Matching Using Chord-Length Function // IDEAL. — 2006. — P. 746—753
4. Bin Wang, Chaojian Shi A. Novel Fourier Descriptor for Shape Retrieval // FSKD. — 2006. — P. 822—825
5. Айвазян С. А., Бухштабер В. М., Енюков И. С., Мешалкин Л. Д. Прикладная статистика. Т. 3 Классификация и снижение размерности. — М.: Финансы и статистика, 1989 — С. 334—344.

6. *Lay D.* Linear Algebra and Its Applications, 3rd Updated Edition. — Redwood, Addison-Wesley, 2005 — P. 426—445.

7. *Fukunaga K.* Introduction to Statistical Pattern Recognition. — San Diego: Academic Press Professional, Inc., 1990 — P. 445—466.

8. *Nain N., Gour P., Agarwal N., Talawar R. P., Chandra S.* Face Recognition using PCA and LDA with Singular Value Decomposition (SVD) using 2DLDA // Proceedings of the World Congress on Engineering, Vol. I. — London, U.K, 2008.

Запругаев Сергей Александрович — д. ф.-м. н, проф. каф. цифровых технологий Воронежского государственного университета. Тел. (4732)208-257. E-mail: zsa@main.vsu.ru

Сорокин Андрей Игоревич — аспирант, кафедра цифровых технологий, Воронежский государственный университет. Тел. (4732) 208-257

Zapryagaev S. A. — Doctor of Physics-math. Sciences, Professor of the dept. of digital technologies Voronezh State University. Tel. (4732) 208-257. E-mail: zsa@main.vsu.ru

Sorokin Andrey Igorevich — Post-graduate student of the dept. of digital technologies Voronezh State University. Tel. (4732)208-257