

УДК 004.93+004.85

Сверточное разреженное представление изображений для анализа статических и динамических образов

Князев Б. А.^{1,*}, Черненький В. М.¹

*bknyazev@bmsturu

¹МГТУ им. Н.Э. Баумана, Москва, Россия

В данной работе с целью повышения эффективности классификации статических и динамических образов изучены сверточные и корреляционные свойства изображений, в результате чего разработаны модель и алгоритм представления изображений. Показано, что выходные данные алгоритма схожи с результатами разреженной декомпозиции методом минимизации энергетической функции. Представлены и проанализированы количественные показатели полученных представлений, в том числе данные по производительности и точности разработанной модели при классификации рукописных цифр базы MNIST. Проведено сравнительное тестирование с методом на основе фильтров Габора и локального оператора, усовершенствованным в данной работе. В качестве классификатора применяется метод опорных векторов. Представлены качественные данные экспериментов по исследованию свойств модели при работе с изображениями и их последовательностями, продемонстрированы достоинства и недостатки модели с точки зрения специфики поставленной цели.

Ключевые слова: свертка, фильтры, Габор, параметрическое представление, разреженное представление, метод опорных векторов, рукописные цифр

Введение

Задачей метода разреженного представления (*sparse coding*) сигналов, сформулированная в [1], является минимизация функции $E(\mathbf{x}, \mathbf{z}, \mathbf{W}) = \|\mathbf{x} - \mathbf{Wz}\|_2^2 + \alpha \|\mathbf{z}\|_1$ (иногда называемой энергетической) по отношению к \mathbf{W} , где $\mathbf{x} \in \mathbb{R}^{n_x}$ – входные данные¹ (изображение, видео, аудио), $\mathbf{Wz} = \tilde{\mathbf{x}}$ – реконструированные данные, получаемые проекцией декодирующей матрицы $\mathbf{W} \in \mathbb{R}^{n_x \times n_z}$ на вектор $\mathbf{z} \in \mathbb{R}^{n_z}$; $\|\mathbf{x}\|_2$, $\|\mathbf{z}\|_1$ – нормы ℓ -2 и ℓ -1 соответственно, α – коэффициент регуляризации. Результатом оптимизации такой

¹ Изображение $\mathbf{X} \in \mathbb{R}^{M \times N}$ часто необходимо рассматривать не как матрицу $M \times N$, а как вектор-столбец $\mathbf{x} \in \mathbb{R}^n$, $n = MN$, в котором каждая позиция соответствует индексу пикселя, и является независимой переменной. В данной работе будем придерживаться следующей нотации, если не указано иное: строчные или прописные латинские символы курсивом (например, x , M) обозначают скаляр; строчные латинские символы жирным ($\mathbf{x} = [x_1, x_2, \dots, x_n]^T$) – вектор-столбец; прописные латинские символы жирным (\mathbf{I}) – матрицу скалярных или других значений; \mathbf{XY} – произведение матриц; $\mathbf{X} \circ \mathbf{Y}$ – поэлементное произведение матриц.

функции, например, методом градиентного спуска, является матрица \mathbf{W} с максимально-возможным количеством нулевых значений благодаря компоненту $\alpha\|\mathbf{z}\|_1$ (*sparsity constraint*). Столбцы в \mathbf{W} представляют собой полосовые фильтры (ориентированные и локализованные в пространстве в случае изображений или в пространстве и времени в случае видео [2]), схожие с фильтрами Габора (Морле) [3,4] (рис. 1,а). Ранее на основе данной функции были получены одни из самых высоких результатов в задачах классификации как статических [5-8] изображений, так и их последовательностей (видео) [2] и аудио сигналов [8]. Основные недостатки метода: избыточность получаемых векторов, низкая скорость оптимизации и необходимость разбиения сигнала на области некоторым способом. Поэтому были предложены несколько разновидностей данного метода, среди которых отметим аппроксимирующее (*predictive sparse decomposition*) [6, с. 16; 9] и сверточное (*convolutional sparse coding*) [6, с. 69; 10] разреженные представления. Данные методы успешно конкурируют со сверточными нейронными сетями (CNN) [11,12] (рис. 1,б), их расширениями [13,14] а также ограниченной машиной Больцмана (*RBM*) [15] – другим генеративным методом (рис. 1,в). Некоторые другие интересные работы (например, [16,17]) сложно оценить объективно, так как не было найдено экспериментальных данных по их использованию на выборках открытого доступа.

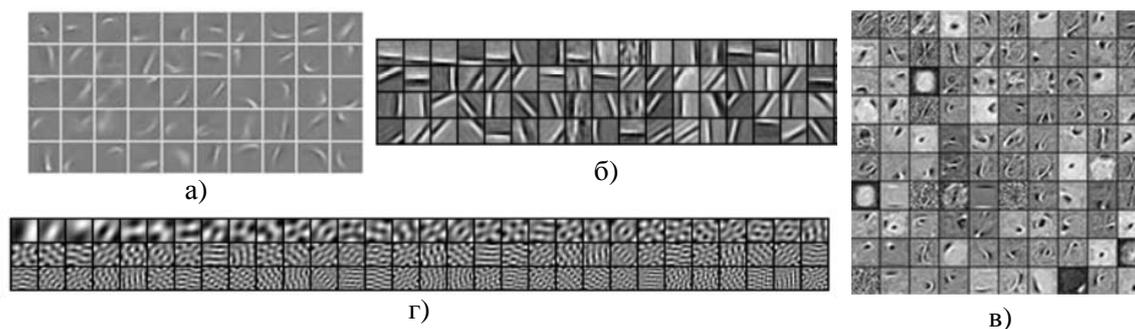


Рис. 1. Паттерны, извлекаемые из изображений или являющиеся результатом обучения: а – «инверсные» фильтры разреженного кодирования [5]; б – некоторые фильтры сверточной сети [14]; в – фильтры машины Больцмана²; г – главные компоненты (PCA) [7]. Нулевым значениям соответствуют серые пиксели, то есть 128 из 255.

Примером неразреженного представления (*dense coding*) является матрица, получаемая из всей выборки вычитанием среднего и проекцией результата на собственные вектора ковариационной матрицы (метод главных компонент – PCA) [18, с. 561; 19] (рис. 1,г). Особенность такого представления в том, что линейные (поворот, масштабирование, смещение) и нелинейные (деформации) изменения исходного изображения приводят к равномерным изменениям практически во всех его размерностях (рис. 2), так как размерности обладают статистическим смыслом (ℓ_2 -норма собственных векторов = 1), но не обладают физическим. Данная особенность относится и к другим генеративным методам, включая разреженное представление: они стремятся к минимальной среднеквадратичной ошибке реконструкции, тогда как большинство изменений

² <http://www.deeplearning.net/tutorial/rbm.html>

изображений приводят к нелинейным искажениям в среднеквадратичном смысле. Данные особенности можно считать недостатками, если целью является классификация образов инвариантная к их искажениям. Ограничением PCA также является вычислительные затраты на поиск ковариационной матрицы для больших изображений и выборок (другие недостатки описаны в [8, с. 4]).

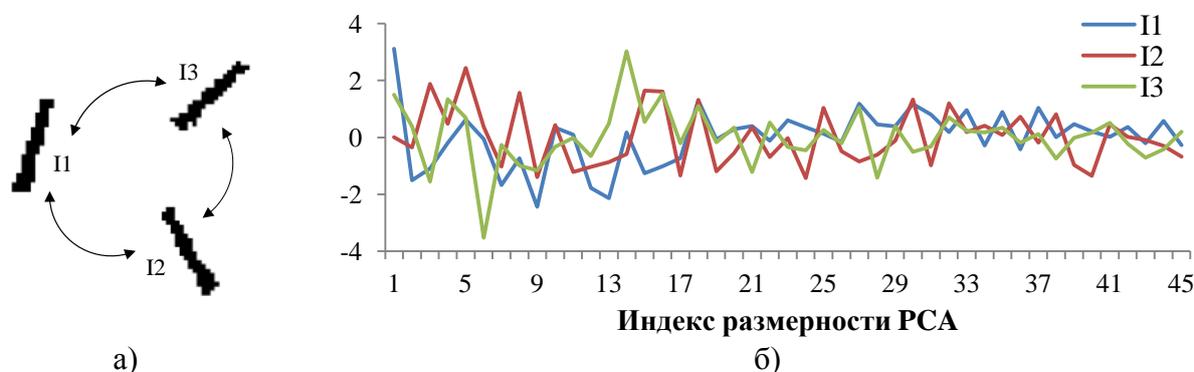


Рис. 2. Демонстрация особенности методов представления изображения на основе минимизации среднеквадратичной ошибки реконструкции на примере метода главных компонент: а – поворот образа; б – значения в 45 размерностях PCA (по оси у – нормированное значение пикселей). В идеальном случае должны меняться значения только некоторых размерностей, в данном случае отвечающих за ориентацию образа.

Изображения на рис. 1,а-в можно интерпретировать как определенные закономерности (паттерны, особенности, фильтры – *patterns, features, filters*), повторяющиеся в изображениях. Во многих работах (например, [1,14]) было отмечено, что эти паттерны напоминают фильтры Габора. Заметим также, что условие разреженности аналогично условию оптимальности разрешающей способности функции Габора (см. следующий раздел). Данные наблюдения во многом послужили мотивацией к исследованию сверточных и корреляционных свойств изображений в текущей работе.

Одним из недостатков большинства методов анализа изображений, в том числе разреженного представления, является разбиение изображения на области (окна поиска – *patches, ROIs*) и использование скользящего окна [20]. Общепринятая методика разбиения отсутствует, а полный перебор всех возможных областей даже для изображений 28×28 пикселей (из [11]) может представлять задачу, нерешаемую за требуемое время. Более того, большинство получаемых откликов могут не являться признаковыми для класса или быть сильно коррелированы между собой. Особенности визуальных образов более естественно представлять произвольными формами, а не прямоугольными областями. Поиск произвольных паттернов возможен в частотной области, что и предлагается в данной работе.

Во многих отмеченных выше трудах, так или иначе, получены фильтры, подобные Габору. Не смотря на это, работ, в которых бы использовалось параметрическое описание таких фильтров, найдено не было за исключением, быть может, [6, с. 37; 9], где некоторые параметры определяются подбором (*fitting*) наиболее похожего фильтра Габора, что неэффективно на практике. Параметрическое описание, предлагаемое в данной работе,

требует задания всего 5-8 параметров (в зависимости от цели), поэтому выигрыш в размерности может достигать двух и более порядков, учитывая, что в предыдущих работах размеры фильтров варьируются от 5×5 до 28×28 и более пикселей.

В данной работе исследуются сверточные и корреляционные свойства изображений с целью более эффективного извлечения паттернов, подобных Габору. Предлагается новая модель представления изображения, включая алгоритм преобразования, на основе описания паттернов с помощью параметров аналитической формы соответствующей функции Габора. Предоставляются результаты количественного анализа полученных паттернов. Экспериментально, на примере изображений рукописных цифр, показано, что получаемый на выходе алгоритма вектор значений может использоваться для классификации образов. В данной работе, как и во многих работах по разреженному представлению, будем работать с изображениями базы MNIST [11], так как: база общедоступна, что упрощает проверку корректности полученных результатов; база представляет собой достаточно большой и разнообразный исследовательский материал с общим количеством 7×10^4 экземпляров (6×10^4 тренировочных и 1×10^4 тестируемых). Помимо этого, на примере обработки по предлагаемому алгоритму видеоизображения из базы МГТУ им. Н.Э. Баумана качественно показано, что также могут быть решены актуальные практические задачи в области автоматизированного анализа последовательностей изображений, часто возникающие в робототехнике [21,22], медицине [23] и безопасности [24]. Одной из более конкретных целей работы является разработка модели описания изображения для последующего анализа их последовательностей по методике, представленной ранее в [25].

1. Параметры функции Габора

Прежде чем перейти к описанию разработанной модели, кратко рассмотрим наиболее обобщенное аналитическое определение *фильтра Габора* (более подробно в [3,4]), описание которого потребуется далее. Формально, фильтр Габора является комбинацией двух *независимых* функций (рис. 3,а):

$$\mathbf{G}(x, y, x_0, y_0, a, b, u_0, v_0) = g(x, y, x_0, y_0, a, b) \xi(x, y, x_\varphi, y_\varphi, u_0, v_0), \quad (1)$$

где $g(x, y, x_0, y_0, a, b) = e^{-\pi[(x-x_0)^2 a^2 + (y-y_0)^2 b^2]}$ – функция Гаусса, $\xi(x, y, x_\varphi, y_\varphi, u_0, v_0) = e^{-2\pi i[u_0(x-x_\varphi) + v_0(y-y_\varphi)]}$ – комплексная гармоническая функция. В частотной области данная комплексная экспонента переносит (модулирует) функцию g на частоту $\pm \omega_0$ относительно 0-ой частоты (рис. 3,б).

Первый множитель в (1) определяет положение фильтра на плоскости (координаты x_0, y_0) и ширину фильтра в пространственной (коэффициенты $a = \frac{1}{\sqrt{2\pi}\sigma_x}$ и $b = \frac{1}{\sqrt{2\pi}\sigma_y}$) и частотной (см. ниже) областях, где σ_x, σ_y – стандартные отклонения функции Гаусса по осям x и y соответственно, которые иногда называют масштабом фильтра Габора; $\gamma = b/a = \sigma_x/\sigma_y$ – эллиптичность. Также неявным параметром является угол поворота β системы координат, при этом неважно пространственной или частотной систем, так как в

силу свойств преобразования Фурье (ПФ) поворот в пространстве на угол β приводит к повороту Фурье-образа (результата ПФ) на такой же угол.

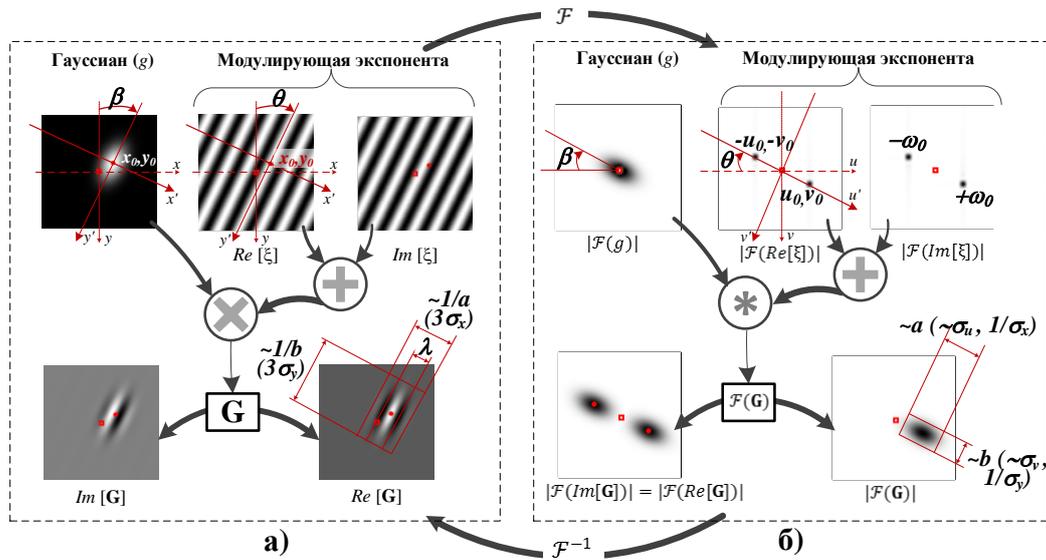


Рис. 3. Формирование фильтра Габора: а – пространственная область, размер изображения 128×128 , $x_0 = 15$, $y_0 = -10$, $\lambda = 6,67\pi$, $\theta = \beta = \pi/7$, $\sigma_x = 2\pi$, $\sigma_y = 3,33\pi$ ($\gamma = 0,6$), $\varphi = 0$; б – частотная область, в целях иллюстрации цвета инвертированы и изменены параметры: размер 28×28 , $\lambda = 1,33\pi$, $\sigma_x = 0,4\pi$, $\sigma_y = 0,67\pi$.

Второй множитель в (1) определяет положение фильтра в частотной области заданием частоты $\omega_0 = \sqrt{u_0^2 + v_0^2}$ или длины волны $\lambda = 1/\omega_0$ модулирующей комплексной экспоненты ξ . Координаты x_φ, y_φ совместно с u_0, v_0 также определяют фазу φ данной экспоненты, то есть сдвиг ее максимума относительно нулевых координат. Если положить $x_\varphi = x_0, y_\varphi = y_0$, то функцию ξ можно переписать в следующем виде: $\xi(x, y, x_0, y_0, u_0, v_0, \varphi) = e^{-2\pi i[u_0(x-x_0) + v_0(y-y_0) + \varphi]}$.

Фурье-образ фильтра, определенного в (1), также является комбинацией двух аналогичных функций [3]:

$$\mathcal{F}(\mathbf{G}) = \widehat{\mathbf{G}}(u, v) = e^{-\pi \left[\frac{(u-u_0)^2}{a^2} + \frac{(v-v_0)^2}{b^2} \right]} e^{-2\pi i[x_0(u-u_0) + y_0(v-v_0)]}, \quad (2)$$

где $a = \sqrt{2\pi}\sigma_u$ и $b = \sqrt{2\pi}\sigma_v$ определяют ширину фильтра в частотной области по осям u и v соответственно (рис. 3,б), σ_u, σ_v – стандартные отклонения функции Гаусса.

Ключевым параметром фильтра Габора является его *ориентация*, в общем случае определяемая углом поворота первого (β) и второго (θ) множителей в (1). Для того чтобы менять угол β или θ , но при этом сохранять координаты x_0, y_0 в изначальной координатной системе изображения, используют матрицу поворота на плоскости $\mathbf{R}(\alpha)$, например, по часовой стрелке:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \mathbf{R}(\alpha) \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix} = \begin{bmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{bmatrix} \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix},$$

где α – угол β или θ . Аналогично для матрицы поворота против часовой стрелки $\mathbf{R}(-\alpha)$, при использовании которой фильтр будет «вращаться» против часовой стрелки с увеличением угла. Тогда функция Гаусса, повернутая **на угол β** , записывается как:

$$g(x, y, x_0, y_0, a, b, \beta) = e^{(-\pi[x'^2 a^2 + y'^2 b^2])}, \quad (3)$$

где $x' = (x - x_0) \cos \beta + (y - y_0) \sin \beta$ и $y' = -(x - x_0) \sin \beta + (y - y_0) \cos \beta$, ось y' совпадает с ориентацией β .

Так как $u_0 = \omega_0 \cos \theta_f$ и $v_0 = \omega_0 \sin \theta_f$ (см. рис. 3,б), где $\theta_f = \theta$ в силу отмеченного выше свойства преобразования Фурье, то выражение $u_0(x - x_0) + v_0(y - y_0)$ в функции ξ можно переписать как $\omega_0(x - x_0) \cos \theta + \omega_0(y - y_0) \sin \theta$. Тогда, модулирующая функция, повернутая **на угол θ** , записывается как:

$$\xi(x, y, x_0, y_0, \lambda, \varphi, \theta) = e^{(-2\pi i/\lambda \cdot [x' + \varphi])}, \quad (4)$$

где $x' = (x - x_0) \cos \theta + (y - y_0) \sin \theta$, при этом заметим, что $\tan(\theta) = v_0/u_0$, а ось x' совпадает с модулирующей осью. Таким образом, строго говоря, фильтр Габора уникально определяется *восемью* свободными параметрами $(x_0, y_0, \lambda, \varphi, \theta, a, b, \beta)$, как и показано в [3].

Формулы (3) и (4) использовались для генерации фильтров на рис. 3, на котором показаны значения всех восьми параметров.

Некоторые распространенные ограничения на свободные параметры. На практике некоторые свободные параметры фиксируют, мотивируя нейрофизиологическими ограничениями, полученными в ходе исследований клеток зрительной коры млекопитающих, а также вычислительной нагрузкой в случае «неограничения» вариаций фильтров. Так, в [3] аргументируется, что угол поворота функции Гаусса β в (3) в большой степени коррелирует с углом ориентации второго множителя θ в (4) (рис. 3,а). Примеры фильтров, у которых данные углы различаются, будут представлены далее (рис. 5,а). Коэффициент $\gamma = b/a$, определяющий эллиптичность функции Гаусса (3), также меняется в относительно небольшом диапазоне ($0,2 < \gamma < 0,9$ [4]) и иногда берется равным 0,5-0,6. Более того, при $\gamma \approx 1$ угол β практически не влияет на конечный вид фильтра Габора. Отношение σ_x/λ определяет количество всплесков и выбирается таким, чтобы было 2-5 основных всплесков. Так как фильтр в основном используется для свертки с изображением, то координаты его центра также не влияют на результат. Поэтому в большинстве работ параметры θ и λ (или σ_x) являются основными для формирования фильтров. Например, ограничиваются 5-9 значениями σ_x и 8 или 16 ориентациями θ , как в [23,26]. В [27] использовали 68 трехмерных фильтров (37 ориентаций одного масштаба и 31– другого).

Подчеркнем, что комбинация выражений (3) в (4) с приведенными выше допущениями, хотя и является наиболее распространенной формой, не является оригинальной и поэтому приводит к заведомо ограниченному набору фильтров. Более того, часто используется только вещественная часть функции (4), тогда как именно комплексная форма обладает *оптимальными* свойствами [28, с. 616].

Оптимальность фильтра. Оптимальный фильтр определяется как фильтр, обладающий «наилучшей» разрешающей способностью, т.е. наименьшей площадью, как в

пространственной, так и в частотной области. Разрешающая способность ограничена принципом неопределенности, который в двумерном виде записывается как [3, с. 3]:

$$s_{min} = (\Delta x)(\Delta y)(\Delta u)(\Delta v) \geq 1/16\pi^2, \quad (5)$$

где $\Delta x, \Delta y$ – эффективная ширина и длина фильтра в пространственной области (рис. 3,а); $\Delta u, \Delta v$ – эффективная ширина и длина фильтра в частотной области (рис. 3,б), которые пропорциональны среднеквадратичным отклонениям функций $\mathbf{G}(x, y)$ и $\widehat{\mathbf{G}}(u, v)$ соответственно. Комплексная форма (то есть (3), (4)) является оптимальным фильтром, так как значение $s_{min} = \frac{\sigma_x \sigma_y}{2} \frac{\sigma_u \sigma_v}{2}$ теоретически не превышает $1/16\pi^2$ (подробнее в [3]) независимо от параметров, но на практике может превышать ввиду дискретизации. Вещественная или мнимая части по отдельности не являются оптимальными, так как их Фурье-образ имеет зеркальную (отраженную от 0-ой частоты) составляющую с координатами $(-u_0, -v_0)$ или $(+u_0, +v_0)$ в зависимости от знака перед 2π в (4).

Принцип неопределенности в данном контексте означает, что невозможно одновременно (на одном изображении) идеально точно описать частотные и пространственные составляющие изображения. Например, чем меньше Δx (σ_x) и Δy (σ_y) фильтра Габора, тем точнее отклик на фильтр будет отражать пространственную информацию исходного изображения, то есть отклик будет близок к исходному изображению, но отражать относительно немного частотных свойств. Аналогичные ограничения наблюдаются в частотной области.

Заметим, что ограничение (5) связано с разреженностью, достигаемой в матрице \mathbf{W} энергетической функции введением ограничения $\alpha|\mathbf{z}|_1$. Однако, в случае (5) разреженность достигается не только в пространственной, но и в частотной области, что позволяет в наиболее компактном виде представить пространственные и частотные свойства изображения, – цель модели представления в данной работе.

При разработке модели, представленной ниже, важно было также учитывать, что параметры в (3) и (4) можно модулировать (делать зависимыми от x, y) не только по амплитуде как в (1), но и по частоте или фазе (например, фильтры с линейной частотной модуляцией или чирплеты – *chirplets*). Если в (4) вместо x' использовать $[|x'| + k/2 \cdot x'^2]$, где k – коэффициент увеличения частоты, то частота всплесков будет нарастать с удалением от центра фильтра. В частотной области это приводит к появлению дополнительных пиков, то есть «растяжению» спектра (Фурье-образа), – эффект, часто наблюдаемый при извлечении паттернов из изображения (рис. 4,г).

2. Модель представления изображения

2.1 Генерация откликов без аналитической формы

При использовании фильтра Габора в качестве ядра свертки с изображениями – классический подход, сходный с вейвлет-преобразованием [4,23,26,27], – фильтр обладает недостатками, затронутыми выше и отмеченными, например, в [29, с. 38]:

- отсутствие однозначной методики выбора *восьми* свободных параметров;

– требование вычислительных ресурсов как для генерирования самого фильтра в соответствии с его аналитической формой, так и операций свертки и хранения полученных откликов, пропорциональных количеству и размеру используемых фильтров.

Фильтр Габора можно получить без использования аналитической формы, тем самым частично решая первую проблему. Как было отмечено выше, методом оптимизации энергетической функции могут быть получены паттерны, подобные фильтру Габора. Другой способ основан на схожести фильтра с производными функциями Гаусса (3-4 порядка, см. в [28]), причем порядок производной равен количеству пересечений функции нуля, то есть на 1 меньше количества всплесков (рис. 4,д). Отличия заключаются в конечности количества всплесков производной Гаусса и форме огибающей, что несущественно при дискретных вычислениях. Однако сама функция Гаусса требует аналитической формы.

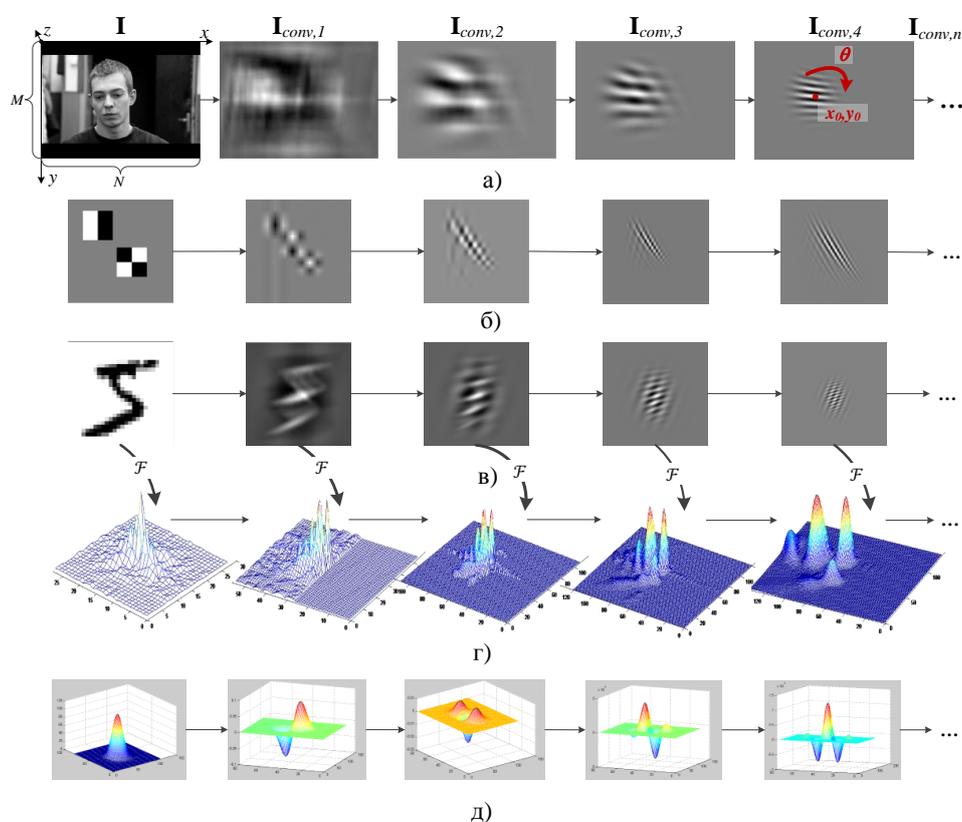


Рис. 4. Применение оператора свертки n -го порядка k : а – изображению лица; б – комбинации функций Хаара, таких как в [20]; в – изображению цифры [11]; г – абсолютные значения Фурье-образов (каждый пик или экстремум соответствует определенному фильтру Габора); д – производные n -го порядка функции Гаусса, схожие с функцией Габора. На рисунках а-г размеры $I_{conv,j} \approx$ размеру I , так как после каждой j -ой итерации происходит уменьшения размера в ≈ 2 раза.

В данной работе предлагается *рекурсивный оператор свертки n -го порядка* произвольного изображения $I \in \mathbb{R}^{M \times N}$:

$$I_{conv,n} = Conv_n(Conv_{n-1}[\dots Conv_2(Conv_1(I_{conv,0}))]), \quad (6)$$

где $Conv_j(I_{conv,j-1}) = Conv(I_{conv,j-1}, I_{conv,j-1}) = I_{conv,j}$ – результат свертки³ изображения с собой, $j \in \{1, \dots, n\}$, $I_{conv,0} = I$ (рис. 4,а-г). Теоретически вместо операции свертки можно использовать автокорреляционную функцию, позволяющую получить визуально схожий результат, но, как будет аргументировано в п. 0 и п. 0, ее использование нецелесообразно. Также, для сравнения, приведем операцию производной n -ого порядка:

$$I_{diff,n} = Diff_n(I_{diff,n-1}), \quad (7)$$

где значение каждого пикселя вычисляется как $i_{diff,n}(x, y) = i_{diff,n}(x, y + 1) - i_{diff,n}(x, y)$, $x \in \{1, N\}$, $y \in \{1, M\}$ (рис. 4,д).

Результатом применения оператора (6) к некоторому статистически нормализованному (см. в [18, с. 567]) изображению, например, лица или одного или комбинации вейвлетов Хаара, является изображение с периодической структурой, которое будем называть паттерном или *откликом*⁴. В зависимости от исходного изображения и порядка оператора можно наблюдать как отклики, схожие с одним из фильтров Габора, так и более сложные структуры. При этом, так как свертка функции Габора с собой также является функцией Габора, то последующие итерации до определенного n ($n = 6-10$) обычно приводят к более четкой структуре. При дальнейшем увеличении n результат начинает напоминать дельта-функцию (дельта-функции используются, например, в [18 с. 155]).

Таким образом, операторы (6) и (7) выражают взаимосвязь произвольных изображений, вейвлетов Хаара, функций Гаусса и Габора. Также можно говорить еще об одном относительно простом, но в то же время явном экспериментальном подтверждении присутствия паттернов, подобных Габору, в структуре изображений.

2.2 Параметрическое описание «идеального» отклика

Одним из вариантов определения параметров отклика **G** является выбор наиболее похожего фильтра Габора, полученного аналитически по формулам (3) (4). Так, в [6, с. 37; 9] позиция, ориентация и частота откликов определяется с помощью перебора (*fitting*) фильтров, что не подходит для цели данной работы по двум причинам. Во-первых, для этого необходимо предварительно сконструировать набор фильтров, руководствуясь некоторыми априорными знаниями о параметрах. Во-вторых, если говорить об анализе видеоизображений, то объем требуемых вычислительных ресурсов для данной операции перебора приведет к соотношению порядка 100/1 (и более) длительности анализа к длительности видео. Поэтому вместо этого предлагается определять параметры отклика **G** по его характеристикам в пространственной и частотной областях.

³ В данной работе, ввиду ограничения объема, описания базовых операций, таких как свертки, преобразования Фурье, ковариационной матрицы и других, не приводятся и могут быть найдены в соответствующей литературе.

⁴ Результатом применения данного оператора к статистически *ненормализованному* изображению, у которого все значения положительны (например, в диапазоне [0,1] или [0,255]), является паттерн, схожий с функцией Гаусса.

В соответствии с рис. 3, выражениями (1)-(4) и экспериментальными наблюдениями параметры $(x_0, y_0, \lambda, \varphi, \theta, a, b, \beta)$ предлагается вычислять следующим образом.

Параметры модулирующего множителя $\xi: x_0, y_0, \lambda, \varphi, \theta$. Первые два параметра, центр (x_0, y_0) , соответствуют позиции максимума абсолютного значения (модуль комплексного числа) отклика в пространственной области: $\mathbf{G}(x_0, y_0) = \max|\mathbf{G}(x, y)|$. Длина волны λ обратно пропорциональная пространственной частоте ($\lambda = 1/\omega_0$), определяется по спектру как $\omega_0 = \sqrt{u_0^2 + v_0^2}$, где u_0, v_0 – координаты Фурье-образа фильтра $\widehat{\mathbf{G}}(u, v) = \mathcal{F}(\mathbf{G}(x, y))$ (см. (2)), соответствующие позиции частоты с максимальной абсолютной амплитудой: $\widehat{\mathbf{G}}(u_0, v_0) = \max|\widehat{\mathbf{G}}(u, v)|$. Фаза φ определяется из фазы Фурье-образа в точке (u_0, v_0) , то есть как фаза основной гармоники: $\varphi = \arg(\widehat{\mathbf{G}}(u_0, v_0))$, либо в пространственной области как $\varphi = \arg(\mathbf{G}(x_0, y_0))$, что соответствует фазе аналитического определения (4). Ориентация модулирующей составляющей θ определяется из ориентации основной гармоники: $\theta = \text{atan}(v_0/u_0)$, $\theta \in [-\pi/2, \pi/2]$ в силу симметричности фильтра.

Параметры функции Гаусса $g: \sigma_x(a), \sigma_y(b), \beta$. Помимо того, что функция Гаусса является симметричной, существуют неопределенности при выборе значений $\sigma_x, \sigma_y, \beta$, если g рассматривать независимо от модулирующего множителя ξ . Заметим, что параметры σ_x, σ_y можно менять местами в зависимости от того, какую ось подразумевать под x и y (рис. 3,а). Для разрешения данной неоднозначности будем руководствоваться принципом, что $\gamma = b/a < 1$ или $\sigma_y > \sigma_x$, в результате чего также получаем однозначный выбор угла β . Данное допущение, хотя и распространенное, не отражает истинные параметры фильтра. На рис. 5 многие из фильтров (в основном, в нижнем ряду) имеют $\gamma > 1$, если под ось y подразумевать ориентацию модулирующей составляющей. В данной работе также испытывались другие стратегии преодоления обозначенной проблемы, позволяющие в некоторых случаях более точно оценивать γ и β , однако это приводило к резким колебаниям значений в случаях, когда $\gamma \approx 1$ или когда $|\beta - \theta| \approx \pi/4$, и как следствие, к ухудшению точности классификации статических образов (п. 0) и наглядности анализа последовательностей изображений (п. 0).

Параметры $\sigma_x, \sigma_y, \beta$ могут быть определены как в пространственной, так и в частотной области. Во втором случае, например, предлагается сначала находить координаты всех значений более некоторого порога τ_σ , т.е. $(\mathbf{u}, \mathbf{v}): \widehat{\mathbf{G}}(u, v) > \tau_\sigma \widehat{\mathbf{G}}(u_0, v_0)$. Затем вычисляются 2 собственных вектора (\mathbf{e}_1 и \mathbf{e}_2) и соответствующие им собственные значения (d_1 и d_2) ковариационной матрицы $\mathbf{X} = \text{cov}(\mathbf{u}, \mathbf{v})$, откуда с учетом допущения $\sigma_y > \sigma_x$ получаем: $\sigma_y = \max(d_1, d_2)$, $\sigma_x = \min(d_1, d_2)$. Угол β определялся как угол между \mathbf{e}_1 и \mathbf{e}_2 . Зная σ_x, σ_y можно определить значения σ_u, σ_v и наоборот. В данной работе экспериментально установлено, что для наиболее корректного определения эффективной ширины в пространственной (σ_x, σ_y) и частотной (σ_u, σ_v) областях и эффективной

площади s_{min} порог $\tau_\sigma \approx e^{-2}$, а для определения β оптимальным был меньший порог, равный 0,01.

2.3 Алгоритм преобразования произвольного изображения

Входными данными алгоритма преобразования f является изображение \mathbf{I} размером $M \times N$ в оттенках серого, максимальное количество откликов, которых необходимо вернуть на выходе ($|\mathcal{H}|_{max}$), а также максимальный порядок оператора свертки (n_{max}). Параметры $|\mathcal{H}|_{max}$ и n_{max} можно интерпретировать как ограничение ресурсов.

Выходными данными алгоритма f является вектор значений $\mathbf{I}_{\mathcal{H}} \in \mathbb{R}^{8 \times |\mathcal{H}|_{max}}$, определяемый как:

$$\mathbf{I}_{\mathcal{H}} = f(\mathbf{I}) = [\mathbf{h}_1^T, \dots, \mathbf{h}_i^T, \dots, \mathbf{h}_{|\mathcal{H}|_{max}}^T], \quad (8)$$

где $\mathbf{h}_i^T = [x_{i,0}, y_{i,0}, \lambda_i, \varphi_i, \theta_i, \sigma_{i,x}, \sigma_{i,y}, \beta_i]^T$. Ниже кратко приводятся шаги разработанного алгоритма, в основе которого лежит метод извлечения откликов, сходных фильтру Габору, из изображения (п. 0), а также методы оценки их параметров (п. 0).

Входные параметры: $\mathbf{I} \in \mathbb{R}^{M \times N}$, $n = 1, n_{max}$, $|\mathcal{H}|_{max}$.

Выходные данные: вектор значений $\mathbf{I}_{\mathcal{H}}$.

1. Стандартизация \mathbf{I}' , так, чтобы оно имело нулевое среднее значение и единичную дисперсию, где \mathbf{I}' – изображение \mathbf{I} , дополненное нулевыми значениями до размеров $(2M - 1) \times (2N - 1)$.
 2. Вычисление Фурье-образа $\mathcal{F}(\mathbf{I}')$.
 3. В случае если \mathbf{I}' является вещественным, вычисление преобразования Гильберта: $\mathbf{I}^H = \mathcal{F}^{-1}(-i \text{sign}(\omega) \cdot \mathcal{F}(\mathbf{I}'))$, где $\mathbf{I}^H \in \mathbb{C}^{M \times N}$, $\text{sign}(\omega) = \{1, -1, 0\}$ для $\omega: \{> 0, < 0, = 0\}$ соответственно, $i^2 = -1$.
 4. Вычисление оператора (6) для текущего значения n как квадрат $\mathcal{F}(\mathbf{I}^H)^2 = \mathcal{F}(\mathbf{I}^H) \circ \mathcal{F}(\mathbf{I}^H)$, тогда $\mathbf{I}_{conv,n} = \mathcal{F}^{-1}[\mathcal{F}(\mathbf{I}^H)^2]$.
 5. Понижение размерности $\mathbf{I}_{conv,n}$.
 6. Поиск локальных экстремумов в $\mathbf{I}_{conv,n}$ и в $|\mathcal{F}(\mathbf{I}_{conv,n})| \rightarrow \text{max_values}$.
 7. **Если** количество экстремумов > 1 , то для каждого i -го экстремума в max_values (рис. 4, г):
 - а. Вычисляем $\hat{\mathbf{I}}_i = \mathcal{F}(\mathbf{I}_{conv,n}) \circ g(u, v)$, так чтобы текущий экстремум был глобальным, где $g(u, v)$ – функция Гаусса, локализованная в окрестностях экстремума;
 - б. Вычисляем $\mathbf{G}_i = \mathcal{F}^{-1}(\hat{\mathbf{I}}_i)$ – i -ый паттерн, сходный с фильтром Габора.
 - **Иначе:** $i = 1$, $\mathbf{G}_i = \mathbf{I}_{conv,n}$.
 - с. Определяем параметры \mathbf{G}_i в соответствии с пунктом 0 \rightarrow
 $\mathbf{h}_i^T = (x_0, y_0, \lambda, \varphi, \theta, \sigma_x, \sigma_y, \beta)$, $\mathbf{I}_{\mathcal{H}} = \mathbf{I}_{\mathcal{H}} \cup \mathbf{h}_i^T$;
 - д. если $|\mathbf{I}_{\mathcal{H}}| \geq |\mathcal{H}|_{max}$, ($|\mathbf{I}_{\mathcal{H}}|$ – общее количество откликов), **выход** $\rightarrow \mathbf{I}_{\mathcal{H}}$.
8. $n++$. **Если** $n \leq n_{max}$:
 - а. $\mathbf{I} = \mathbf{I}_{conv,n}$, возвращаемся на шаг 2.
 - **Иначе:** **выход** $\rightarrow \mathbf{I}_{\mathcal{H}}$.

В экспериментах, проводимых в данной работе, использовался коэффициент понижения размерности 1,8 (шаг 5), так как при значениях $> 1,8$ наблюдалась потеря информативных экстремумов, как в частотной, так и в пространственной областях. В целях оптимизации скорости работы параметр n_{max} был ограничен значением 7, а

$|\mathcal{H}|_{max} = 32$. Также, после наблюдений, **шаги 5-7** пропускались для $n = 1$; были установлены отдельные ограничения $|\mathcal{H}|_{max}$ на каждом порядке свертки значениями 7, 7, 6, 5, 4, 3 для $n = 2-7$ соответственно, добавлено ограничение максимальной эффективной площади $\max_{\mathcal{H}}(s_{min}) = 0,02$.

Примеры извлекаемых откликов \mathbf{G}_i показаны на рис. 5, на котором заметно, что по данному алгоритму извлекаются отклики, соответствующие основным паттернам изображения. Так как алгоритм представляет собой поиск пиков в частотной области (рис. 4,г), то извлекаемые паттерны соответствуют произвольным регионом в пространственной области. Методы разреженного кодирования позволяют извлечь похожие паттерны, но, как было отмечено ранее, они требуют разбиения изображения на регионы некоторым способом (например, 13×13 пикселей в [7]), а также длительное время для схождения энергетической функции. Вообще говоря, используя метод разреженного представления только в пространственной области, паттерны, идентичные представленным на рис. 5,а, могут быть найдены только полным перебором всех возможных областей изображения, что невыполнимо на практике.

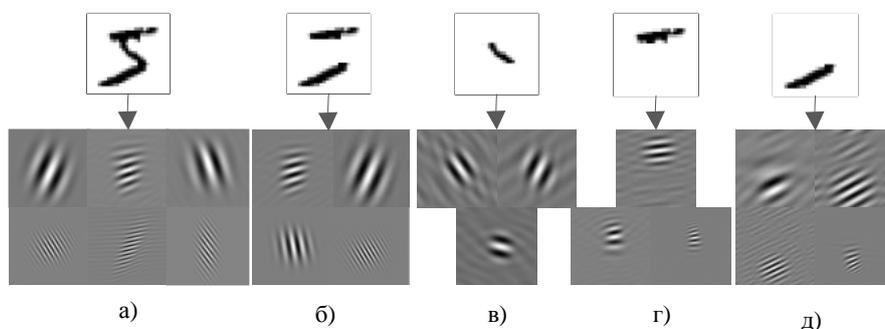


Рис. 5. Демонстрация результатов работы алгоритма: а, сверху – изображение из базы MNIST [11], снизу – некоторые отклики \mathbf{G}_i , полученные по разработанному алгоритму; б-д, сверху – сегментированные *вручную* части (паттерны) изображения; снизу – некоторые отклики, полученные по тому же алгоритму для данных частей.

Преобразования Гильберта на **шаге 3** необходимо для удаления зеркальной составляющей Фурье-образа $\mathcal{F}(\mathbf{I}')$ и уменьшения вычислительной нагрузки при поиске экстремумов, и как показали эксперименты, увеличения точности классификации. Наиболее ресурсоемким является **шаг 7.а**, на котором необходимо несколько раз генерировать функцию Гаусса пока текущий экстремум не станет единственным. При этом необходимо использовать максимально широкую функцию, чтобы сохранить структуру отклика.

3. Экспериментальная часть

Для оценки эффективности разработанной модели представления и алгоритма преобразования проводилось три эксперимента:

1. Исследование точности определения параметров отклика (**шаг 7.с** алгоритма).
2. Исследование диапазона значений параметров откликов и их корреляционных свойств с целью выявления наиболее информативных параметров.
3. Решение задачи классификации образов с целью оценки репрезентативности разработанной модели представления.

3.1 Точность определения параметров

Для данного эксперимента было сгенерировано по $\sim 3 \cdot 10^4$ фильтров Габора в соответствии с аналитической формой (3) и (4) для каждого из трех размеров фильтров (табл. 1). Для каждого из фильтров были вычислены параметры по разработанному алгоритму (п. 0), после чего найдены средние ошибки и стандартные отклонения между параметрами, заданными аналитически и вычисленными по алгоритму. Помимо восьми параметров, сравнивалась эффективная площадь s_{min} с эталонной (см. (5)), количество экстремумов спектра $k_{chirp} = N_{max_values}$ ($k_{chirp} = 1$ для идеального Габора как в частотной, так и в пространственной областях), а также $\gamma = \sigma_x / \sigma_y$. Также, по вычисленным параметрам были сгенерированы фильтры и вычислены ошибки между вещественными (ε_{real}), мнимыми (ε_{imag}) и абсолютными (ε_{abs}) составляющими изображений фильтров.

Таблица 1. Средние (m) и среднеквадратичные (σ) ошибки определения параметров фильтров*

Параметр	16×16 пикселей		32×32 пикселей		64×64 пикселей	
	m	σ	m	σ	m	σ
x_0	-0,36	2,22	0	0	0	0
y_0	-0,37	2,21	0	0	0	0
λ	2,44	5,62	0,04	0,62	-0,01	1,32
φ , рад	0,29	2,61	0	0	0	0
θ , рад	0,06	0,22	0,00	0,04	-0,01	0,05
σ_x	-1,92	1,03	-1,35	1,42	-1,02	1,77
σ_y	-1,48	1,19	-1,10	1,33	-1,00	1,74
β , рад	-0,002	0,27	-0,01	0,17	-0,01	0,15
s_{min}	0,0015	0,0027	0,0051	0,0033	0,0051	0,0046
γ	-0,16	0,51	-0,18	0,40	-0,13	0,34
k_{chirp}	-0,56	0,61	0	0	0	0
ε_{real}	1,54	0,85	3,56	2,24	7,07	5,62
ε_{imag}	1,55	0,91	3,67	2,55	7,31	6,10
ε_{abs}	1,73	0,80	4,18	2,61	7,45	7,38
t , мс	5,99	1,13	7,13	1,43	8,78	1,52
Σ	12,47	19,05	14,10	11,39	24,02	24,46

*Интенсивность зеленой/красной заливки возрастает с увеличением/уменьшением значений за исключением последних двух рядов; Σ – суммарное значение без учета t .

Оба множителя, функция Гаусса и модулирующая экспонента, идеальных откликов *симметричные*, что затруднило объективное вычисление ошибок для параметров $\theta, \sigma_x, \sigma_y, \beta, \gamma$. Например, не смотря на, казалось бы, верное значение θ , найденное алгоритмом, вторым вариантом всегда является значение $\pm \pi$, что могло не соответствовать ориентации, заданной аналитически и вызывать резкое возрастание ошибки. Аналогично, для остальных указанных параметров. Ввиду того, что данные ошибки не являются принципиальными, в табл. 1 представлены результаты без учета периодичности углов θ, β и реверса параметров $\sigma_x, \sigma_y, \gamma$.

По результатам эксперимента выявлено, что в целом для размеров 32×32 и 64×64 пикселей все параметры, за исключением параметров σ_x, σ_y и связанной с ней эффективной площади s_{min} , вычисляются точнее, чем для случая 16×16. Генерация фильтров больших размеров (>64×64 пикселей), а также определение их параметров,

требует значительных вычислительных ресурсов (>18 мс на фильтр для 128×128). Поэтому проводились только ограниченные исследования для таких фильтров, которые позволяют предположить, что с дальнейшим увеличением размера точность остается прежней или возрастает, так как фактически увеличивается частота дискретизации сигнала. Так как для ошибок ε_{real} , ε_{imag} и ε_{abs} усреднение проводилось только по количеству фильтров, то учитывая увеличение общего количества пикселей в 4 раза во втором и третьем экспериментах, можно оценить, что средняя ошибка уменьшается, в то время как среднеквадратичная – немного возрастает. Ошибки всех параметров увеличиваются, в основном, в тех случаях, когда фильтр начинает напоминать дельта-функцию в пространственной или в частотной области. Данные случаи ведут к резкому уменьшению (вплоть до 0) или увеличению (вплоть до 0,05) значения s_{min} , поэтому частично были отсеяны, однако было сложно учесть все подобные случаи, учитывая количество фильтров. Оптимальным размером отклика можно считать 32×32 пикселей, при котором достигается высокая точность разработанного алгоритма и незначительно возрастает время его работы.

3.2 Диапазон значений и корреляционные свойства откликов

В данном эксперименте нас интересует, какие паттерны присутствуют в изображениях, для которых выполняется автоматизированный анализ, например, с целью классификации образов. Для данного эксперимента, а также для решения задачи в следующем пункте, была обработана выборка изображений рукописных цифр MNIST [11]. По алгоритму, представленному в п. 0, было извлечено $|\mathcal{H}|_{max} = 23$ отклика из каждого изображения, что для всей тренировочной выборки соответствует чуть менее $23 \times 60 \times 10^3$ откликам из-за того, что из некоторых экземпляров (обычно простых, таких как 1 и 7) возвращается менее 23 откликов (табл. 2, рис. 6). Принимая во внимание средние значения и диапазон, можно сделать вывод о том, что угловые параметры φ, θ, β имеют наибольший разброс значений. Однако их величина несколько завышена из-за проблемы периодичности. Так, например, отклики с ориентациями $\theta = -1,55$ рад и $\theta = 1,55$ рад при прочих равных параметрах практически идентичны.

Обобщая результаты, сгруппированные в табл. 1 и табл. 2, можно показать, что разработанная модель является разреженной формой представления изображений. Во-первых, значения s_{min} (табл. 2) показывают, что отклики \mathbf{G}_i , возвращаемые алгоритмом, являются разреженными как в пространственной, так и в частотной области, что согласуется с принципом (5). Отклонения от эталонного значения ($1/16\pi^2 = 0,0063$) не превосходят отклонений, вычисленных для аналитически сгенерированных функций (табл. 1). Действительно, $s_{min} < 0,0146$ и $s_{min} < 0,0150$ для >99% откликов в данном и предыдущем (для случая 32×32) экспериментах соответственно. Во-вторых, среднее количество возвращаемых откликов (суммарное для всех порядков n) около 25, в то время как общее количество откликов, извлеченных из MNIST⁵, более 10^6 . Многие из них сильно коррелированы между собой и в зависимости от критерия идентичности отклика можно оставить 10^3 - 10^5 откликов, в результате чего получаем разреженность >2,5%, которой можно управлять ограничениями s_{min} и $|\mathcal{H}|_{max}$ аналогично коэффициенту α в энергетической функции.

⁵ Первые 10^4 откликов доступны на bmstu.ru/ps/~bknyazev/fileman/ls/2014/DigitRecognition/MyModel.

Таблица 2. Статистика параметров откликов*

Парам.	m	σ	Мин.	Макс.
x_0	2,52	7,45	-76,00	91,00
y_0	2,75	10,57	-91,00	91,00
λ	6,72	3,92	1,54	24,26
σ_x	6,14	1,79	0,98	16,32
σ_y	9,38	2,14	3,05	32,63
γ	0,66	0,13	0,11	1,00
φ , рад	0,02	1,81	-3,14	3,14
k_{space}	1,60	1,65	1,00	127,00
θ , рад	0,08	0,85	-1,57	1,57
β , рад	0,08	0,52	-1,57	1,57
s_{min}	0,0083	0,0021	0,0020	0,0200
n	3,80	1,44	2,00	7,00
n_2	5,92	1,46	1,00	7,00
n_3	6,99	0,12	3,00	7,00
n_4	5,69	0,71	1,00	6,00
n_5	3,44	1,15	0,00	5,00
n_6	2,10	0,91	0,00	4,00
n_7	1,32	0,63	0,00	3,00
t , сек	1,07	0,36	0,32	4,66

* k_{space} – количество экстремумов в пространственной области, n – порядок свертки, n_i – количество паттернов, возвращаемых после оператора (б) i -го порядка (были вычислены с ограничением $|H|_{max} = 32$),
 t – время обработки изображения.

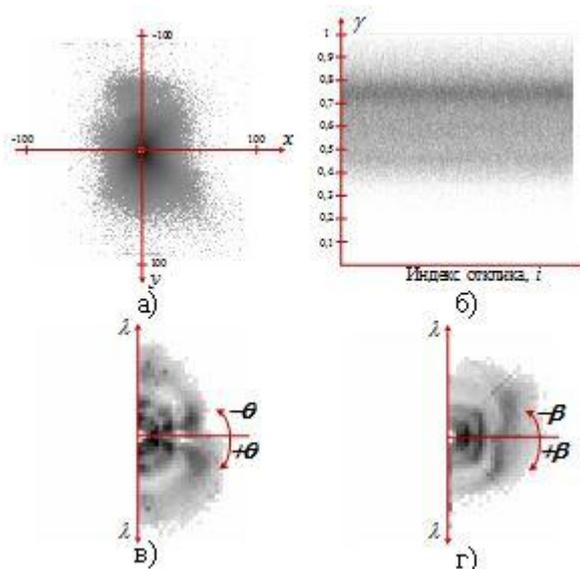


Рис. 6. Распределение значений откликов, полученных из изображений MNIST, в логарифмической шкале: а – x_0, y_0 ; б – γ ; в, г – θ и β в полярных координатах совместно с длиной волны λ .

В дополнении к этому, представленные количественные результаты в табл. 3 также согласуются с работами [3,4], которые ссылаются на исследования клеток зрительной

коры млекопитающих. Действительно, наблюдается корреляция параметров λ и σ_x ($r = 0,22$), углов θ и β ($r = -0,26$), а значения $\gamma > 0,27$ в $>99\%$ случаев. Однако в нашем эксперименте корреляция последних оказалась отрицательной, а сами значения несколько заниженными опять же ввиду проблемы периодичности. В табл. 3 отсутствует φ , так как фаза не имеет корреляции $>0,1$ ни с одним из параметров. С ростом порядка n оператора (6) значение λ и эффективная площадь s_{min} уменьшается ($r = -0,47$ и $r = -0,44$ соответственно), что означает, что всплески в частотной области отдаляются от нулевой частоты и становятся более вытянутыми (рис. 4,г). Положительная корреляция между \widehat{G} и λ ($r = 0,25$) свидетельствует об уменьшении значений экстремумов с отдалением от 0-й частоты.

Таблица 3. Корреляционная матрица параметров откликов, а также классов (l) и времени (t)*

	x_0	y_0	λ	σ_x	σ_y	γ	\widehat{G}	k_{space}	θ	β	s_{min}	n	t	l
x_0		0,07	-0,13	0,09	0,09	0,03	0,08	-0,04	0,05	0,03	-0,05	0,20	0,02	0,03
y_0	0,07		-0,10	0,11	0,10	0,03	0,12	-0,05	0,05	-0,03	-0,11	0,19	-0,02	0,01
λ	-0,13	-0,10		0,22	0,09	0,20	0,25	0,10	-0,06	-0,03	0,04	-0,47	0,08	0,04
σ_x	0,09	0,11	0,22		0,69	0,62	0,21	-0,07	-0,06	0,01	-0,24	0,41	0,05	0,05
σ_y	0,09	0,10	0,09	0,69		-0,11	0,08	-0,06	-0,05	-0,02	-0,18	0,39	0,02	-0,11
γ	0,03	0,03	0,20	0,62	-0,11		0,17	-0,04	-0,03	0,04	-0,11	0,13	0,05	0,18
\widehat{G}	0,08	0,12	0,25	0,21	0,08	0,17		-0,06	0,05	-0,16	-0,29	0,23	0,00	-0,01
k_{space}	-0,04	-0,05	0,10	-0,07	-0,06	-0,04	-0,06		0,00	0,09	0,44	-0,25	0,09	0,01
θ	0,05	0,05	-0,06	-0,06	-0,05	-0,03	0,05	0,00		-0,26	0,02	0,03	-0,05	-0,04
β	0,03	-0,03	-0,03	0,01	-0,02	0,04	-0,16	0,09	-0,26		0,12	-0,12	-0,03	0,02
s_{min}	-0,05	-0,11	0,04	-0,24	-0,18	-0,11	-0,29	0,44	0,02	0,12		-0,44	0,14	0,09
n	0,20	0,19	-0,47	0,41	0,39	0,13	0,23	-0,25	0,03	-0,12	-0,44		0,03	-0,01
t	0,02	-0,02	0,08	0,05	0,02	0,05	0,00	0,09	-0,05	-0,03	0,14	0,03		-0,07
l	0,03	0,01	0,04	0,05	-0,11	0,18	-0,01	0,01	-0,04	0,02	0,09	-0,01	-0,07	

* \widehat{G} – значение $|\widehat{G}(u, v)|$ в i -ой точке экстремума.

В целом по полученным количественным оценкам в данном эксперименте можно говорить о согласованности результатов алгоритма с имеющимися знаниями о фильтрах Габора. Полученные результаты также могут использоваться для конструирования более подходящих фильтров для эффективного решения конкретных задач, например, задачи классификации образов.

3.3 Решение задачи классификации

Задачами данного эксперимента являются 1) оценить возможности использования алгоритма, предложенного в п. 0, для классификации изображений; 2) получить/улучшить результаты, используя predetermined фильтры Габора и 3) метод их комбинаций с локальным оператором минимума и максимума [7]; 4) сравнить результаты с данными предыдущих работ. В качестве тренировочной и тестовой выборки использовались

изображения рукописных цифр MNIST аналогично предыдущему эксперименту, однако могут быть использованы и другие изображения.

В качестве классификатор (или метода машинного обучения) использовалась машина опорных векторов (SVM) [30], и кросс-платформенная библиотека `libsvm` [31], которая ее реализует. Изначально, метод SVM являлся бинарным классификатором, но во многих работах были предложены и в `libsvm` реализованы расширения метода по принципу сравнения каждого с каждым (*one-against-one*, *one-vs-one*) и каждого со всеми остальными (*one-vs-all*, *one-vs-rest*) [18, с. 338]. В данной работе использовался вариант по умолчанию в `libsvm` – сравнение каждого с каждым. Для определения оптимальных параметров метода SVM (C , γ), а также количества главных компонент метода PCA (N_{PCA}) и количества откликов ($|\mathcal{H}|_{max}$) метода, разработанного в данной работе, проводилась процедура перекрестного тестирования (или кросс-валидация, *cross-validation*) [18, с. 32], как и в других работах [7,11,13,27]. Для этого использовались первые 10^4 тренировочных экземпляров, так как для большего количества процедура требовали вычислительных ресурсов, не всегда соизмеримых с улучшением точности. Напомним, что при перекрестном тестировании выборка разбивается на n_{fold} частей (в данной работе $n_{fold} = 5$). Тестируемая выборка задействуется только для вычисления ошибки классификации, а для настраивания модели не используется.

Линейный и нелинейный метод опорных векторов. В первой части эксперимента проверялось преимущество нелинейной функции ядра метода SVM по сравнению с линейной. В качестве обучаемого вектора брались все $N = 784$ пикселя изображения. Дополнительная обработка изображений, за исключением статистической нормализации (такой, как в **шаге 1** алгоритма в п. 0), не осуществлялась⁶. Для линейной функции ошибка классификации ε составила 7,44%. Ошибка 1,41%, полученная для радиальной базисной функции (*RBF*), соответствует аналогичным работам [7,11], не смотря на независимую от других работ процедуру выбора оптимальных параметров (C , γ) (рис. 7,а).

С применением других нелинейных функций ядра (полиномиальной, сигмоидальной), встроенных в библиотеку `libsvm`, проводились только ограниченные эксперименты. Во-первых, они не показали преимуществ по сравнению с *RBF*, во вторых, задачей эксперимента является исследование модели представления, а не выбор наилучшей функции ядра.

⁶ При использовании других методов классификации, в частности метода k ближайших соседей, бинаризация изображений (порог 0,2) могла существенно (для базы MNIST это 1-2%) увеличить точность классификации.

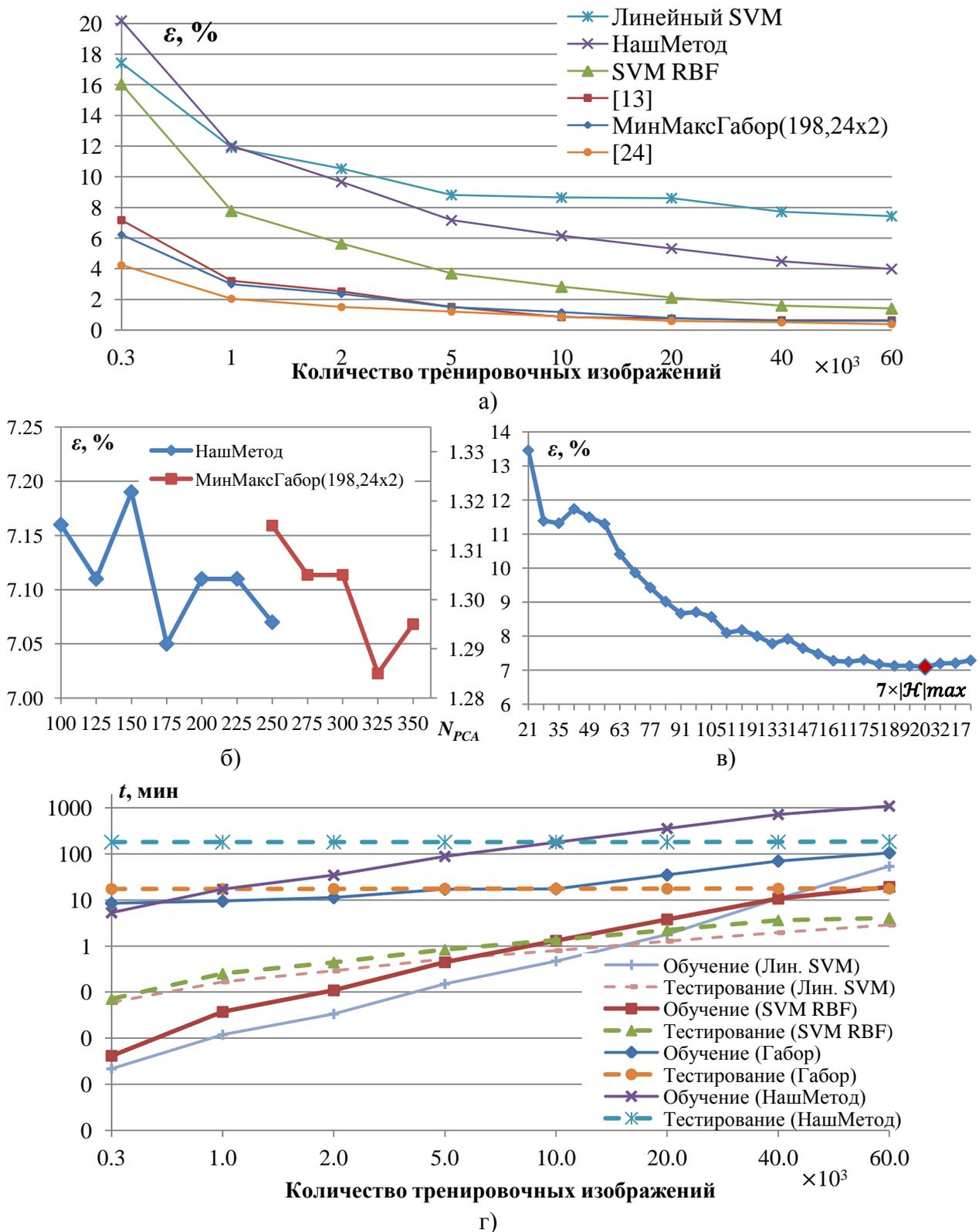


Рис. 7. Экспериментальные данные, полученные в работе: а – зависимости ошибок классификации (ε) методов в зависимости от размера тренировочной выборки (тестируемая фиксирована – 10^4 экземпляров); б – зависимость ε от количества размерностей PCA (N_{PCA}); в – зависимость ε от размера вектора $\mathbf{I}_{\mathcal{H}}$, равного $7 \times |\mathcal{H}|_{max}$; г – суммарное время (t), потраченное на обработку *тренировочной* выборки и *обучение* и на обработку *тестируемой* выборки и *тестирование* в зависимости от размера тренировочной выборки.

Фильтры Габора, компонентный анализ и метод опорных векторов. В данной части тестирования вначале исследовались возможности метода обработки изображений предопределенным набором фильтров Габора, как в [4,26,23,27]. При использовании классического набора из 40 фильтров (5 масштабов и 8 ориентаций) получаем вектор из $N = 40 \times 784 = 31360$ значений. Оптимизация функции опорных векторов для данных такой размерности и выборки из 60×10^3 экземпляров на практике трудноосуществима (по крайней мере, используя вычислительные ресурсы персонального компьютера) ввиду нелинейной сложности существующих алгоритмов оптимизации данной функции. Поэтому часто применяют метод главных компонент, который позволяет снизить размерность данных до $N_{PCA} \in \{1, N\}$, не ухудшая или даже увеличивая точность (например, при применении PCA с $N_{PCA} = 45$ к необработанным изображениям точность могла быть улучшена на 0,01-0,05%). В результате за счет применения PCA, а также предварительного снижения размерности данных с помощью удаления высокочастотной части изображений, были проведены эксперименты с дополнительно сконструированными фильтрами (сначала 64, затем 76). Обычно использует половинное деление спектра (как в диадном вейвлет преобразовании), но было замечено, что при таком делении теряются важные паттерны, поэтому был выбран коэффициент 1,8. С использованием эффективного алгоритма (встроенного в MATLAB) вычисления первых наиболее значимых собственных векторов была вычислена матрица $\mathbf{Y} \in \mathbb{R}^{N \times N_{PCA}}$ главных компонент на основе 10^4 экземпляров. Не смотря на то, что увеличение количества фильтров приводило к уменьшению ошибки (1,15% для 64; 1,11% для 76), дальнейшее увеличение данным методом не было возможным ввиду ограничений памяти при вычислении ковариационной матрицы $\mathbf{X} \in \mathbb{R}^{N \times N}$, а также собственных векторов в \mathbf{X} : сложность может достигать $O(N^3)$. Тем не менее, логично было предположить, что увеличение количества фильтров до некоторого предела могло бы положительно сказаться на результатах тестирования, что послужило мотивацией следующего эксперимента.

На этом этапе также было установлено преимущество использования комплексной формы фильтра Габора по сравнению с вещественной. Так, используя первые 10^4 экземпляров для обучения, разница в ошибки на тестируемой выборке составляла 1,2% (3,56% против 4,76%).

Фильтры Габора, локальный оператор минимума-максимума, компонентный анализ и метод опорных векторов. Отличие данного метода от предыдущего заключается в применении оператора F^{minmax} , вычисляющего минимум и максимум в некоторой области изображения [7]. Так как за счет такого оператора происходит понижение размерности на порядок, то возможно увеличение количества фильтров. При этом по нашим наблюдениям использование дополнительных фильтров лучше сказывается на точности, чем сохранение размерности откликов на фильтры. В [7] с использованием 160 фильтров Габора и F^{minmax} для областей размером 9×9 пикселей (9 областей для изображений цифр размером 28×28) была достигнута ошибка всего 0,71%, а

с использованием 169 паттернов, полученных методом разреженного представления, и аналогичного оператора F^{minmax} – 0,59%.

Так как детали 160 фильтров в [7] неизвестны, то необходимо было сконструировать набор фильтров Габора. Для этого экспериментальным путем были выбраны 3 масштаба σ_x , 4 масштаба σ_y , 7 ориентаций $\theta = \beta$ и 2 отношения σ_x/λ , $\varphi = 0$, $x_0 = y_0 = 0$, т.е. всего 168 фильтров, в результате чего была получена ошибка 0,77%, что несколько хуже, чем в [7]. При этом оператор F^{minmax} вычислялся для 16 областей размером 7×7 пикселей, иначе ошибка была больше. Далее удалось понизить ошибку до 0,72%, добавив 12 *составных фильтров*, то есть состоящих из суммы двух фильтров Габора (в этом случае комбинация параметров x_0 , y_0 двух фильтров влияет на результат). Стоит заметить, что в данных экспериментах оптимизация параметров C , γ не проводилась, и метод PCA не применялся, так как далее была получена меньшая ошибка. Сначала выбирался способ, который позволял достичь лучших результатов с использованием параметров по умолчанию, а уже затем проводилась оптимизация, которая обычно уменьшала ошибку еще на 0,03-0,05%.

В конечном счете, опытным путем удалось получить наилучшие показатели (ошибку 0,60%, рис. 7,а), уменьшив количество ориентаций до 6 (144 фильтра), увеличив количество составных фильтров до 18 (162 фильтра) и добавив 18 фильтров с фазой $\varphi = -\pi/2$ и 18 с $\varphi = \pi$, т.е. всего 198 фильтров. В данном случае изображение разбивалось на 24 области размером 4×7 пикселей, и применялся метод PCA, результаты выбора оптимального N_{PCA} для которого представлены на рис. 7,б. После выбора N_{PCA} , были найдены оптимальные параметры C , γ методом перекрестного тестирования (см. начало пункта).

Мульти-классовая классификация методом SVM осуществляется на основе сравнения количества положительных значений расстояний от гиперплоскости, разделяющих классы, из возможных $k(k-1)/2 = 45$, где $k = 10$ – количество классов [18, с. 338]. Таким образом, помимо наилучшего выбора, можно было проанализировать второй и последующие кандидаты классов экземпляров. В данном эксперименте из 60 неправильно классифицированных (из всего 10^4 тестируемых экземпляров), только 15 экземпляров не оказались ближе к правильному классу при выборе второго кандидата.

Таким образом, модифицируя метод, предложенный в [7], удалось получить одни из лучших результатов классификации изображений MNIST на сегодняшний день, не смотря на то, что метод является одним из самых простых. Его недостаток в том, что выбор фильтров критично влияет на результат, поэтому они должны быть тщательно сконструированы с учетом ограничений вычислительных ресурсов и специфики конкретной задачи, что затрудняет его распространенное использование на практике.

Алгоритм преобразования, предложенный в данной работе. Для изначального тестирования использовались все восемь параметров (см. п. 0), а также k_{space} , S_{min} и значения экстремумов $\hat{\mathbf{G}}$, так как, не смотря на сильную корреляционную зависимость отдельных параметров (табл. 3), в некоторых случаях именно комбинация параметров

позволяла получить более высокие результаты. В ходе экспериментов было установлено, что наиболее информативными параметрами являются $x_0, y_0, \lambda, \theta, \sigma_x, \sigma_y, \beta$, то есть все параметры выражений (3) и (4) за исключением фазы φ , которая не улучшала результат классификации.

Для машины опорных векторов не имеет значения, в каком порядке следуют значения классифицируемого вектора, так как все значения рассматриваются как независимые переменные. Главное, чтобы этот порядок был одинаковым для всех экземпляров. В случае вектора $\mathbf{I}_{\mathcal{H}}$ (см. (8)) имеем $|\mathcal{H}|_{max}$ наборов \mathbf{h}_i^T из 7 значений. В шаге 7 алгоритма в п. 0 порядок следования экстремумов не задается (подразумевается, что отклики возвращаются в порядке значений $\hat{\mathbf{G}}$ в точках экстремума). Проблема формирования вектора, порядок значений которого не зависит от экземпляров, заключается в том, что порядок максимумов ($\hat{\mathbf{G}}$) может меняться среди изображений одного класса из-за их индивидуальных особенностей. В результате получаем, что в $\mathbf{I}_{\mathcal{H}}$ каждого экземпляра вектора \mathbf{h}_i^T могут следовать в некотором своем порядке, уникальном для конкретного экземпляра, что неприемлемо для SVM и приводит к относительно большим ошибкам. Было замечено, что сортировка векторов \mathbf{h}_i^T по тому или иному параметру может положительно влиять на результат. Начав с сортировки по углу θ , был сделан вывод, что необходим учет различных вариантов сортировок. Это привело к следующему расширенному формату вектора:

$$\mathbf{I}_{\mathcal{H}}^* = \left[\bigcup_{n=2}^{n=7} \bigcup_{k=\theta, \beta, \hat{\mathbf{G}}, \lambda} \text{sort}(\mathbf{I}_{\mathcal{H},n}^*, k) \right] \cup \left[\bigcup_{m=\theta, \lambda, y_0} \text{sort}(\mathbf{I}'_{\mathcal{H}}, m) \right], \quad (9)$$

где $\mathbf{I}_{\mathcal{H},n}^* = [\mathbf{h}_{1,n}^T, \dots, \mathbf{h}_{i,n}^T, \dots, \mathbf{h}_{|\mathcal{H}|_{max},n}^T, \mathbf{h}_{1,n}^{(\theta+\pi)T}, \dots, \mathbf{h}_{i,n}^{(\theta+\pi)T}, \dots, \mathbf{h}_{|\mathcal{H}|_{max},n}^{(\theta+\pi)T}]$, – вектор параметров откликов, возвращаемых оператором (6) порядка n , $\mathbf{h}_i^{(\theta+\pi)T} = [x_{i,0}, y_{i,0}, \lambda_i, \theta_i + \pi, \sigma_{i,x}, \sigma_{i,y}, \beta_i]^T$ – вектор, добавляемый ввиду периодичности ориентации θ отклика, $\text{sort}(\mathbf{I}_{\mathcal{H},n}^*, k)$ – оператор сортировки векторов $\mathbf{I}_{\mathcal{H},n}^*$ по параметру k , $\text{sort}(\mathbf{I}'_{\mathcal{H}}, m)$ – оператор сортировки вектора $\mathbf{I}'_{\mathcal{H}} = \bigcup_{n=2}^{n=7} \mathbf{I}_{\mathcal{H},n}^*$ по параметру m (без учета порядка n). Оптимальное количество откликов $|\mathcal{H}|_{max} = 29$ (рис. 7,в), что соответствует длине вектора $\mathbf{I}_{\mathcal{H}}$, равной $29 \times 7 = 203$ в (8). В (9) длина вектора $\mathbf{I}_{\mathcal{H}}^*$ увеличивается до $29 \times 14 \times 4 + 29 \times 14 \times 3 = 2842$. Для данного вектора был применен метод PCA, результаты выбора оптимального N_{PCA} для которого представлены на рис. 7,б.

Минимальная ошибка, полученная данным методом, составляет 3,99%, что больше всех исследуемых методов на основе SVM за исключением его линейного варианта (7,44%, табл. 4). Однако, в отличие от других методов кривая обучения на рис. 7,а получилась более крутая. В то время как кривые других методов, начиная с $\sim 20 \times 10^3$ тренировочных экземпляров, становятся практически плоские – методы перестают обучаться, в случае нашего метода кривая все еще имеет большой угол, а значит, имеет перспективу уменьшения ошибки с увеличением объема выборки. Тем не менее, тот факт, что обучение происходит медленно, является минусом модели. При использовании автокорреляционной функции вместо (6) ошибка оказалась больше ($\sim 1\%$ при кросс-

валидации), что говорит о преимуществе оператора свертки. Это может быть связано с потерей фазовых составляющих в случае использования автокорреляционной функции.

Таблица 4. Сравнительная таблица полученных результатов для MNIST [11]*

Метод	$T_{обр.}$, мин	$T_{обуч.}$, мин	C/γ	N/N_{PCA}	ϵ , %
Линейный SVM	0	54	4/-/-	784/-	7,44
Данная работа	1078	11	$2^{1,5}/2^{-7}$	2842/175	3,99
RBF SVM [7]	-	-	1291,5/21,5	784/-	1,42
RBF SVM	0	19	$2^{1,25}/2^{-8,5}$	784/-	1,41
Габор _{17×17} ⁶⁴	~57	-	1/(1/N)	21964/300	1,15
Габор _{17×17} ⁷⁶	140	12	1/(1/N)	18496/300	1,11
Габор _{17×17} ⁸⁸	-	-	1/(1/N)	25432/-	недостаточно памяти
Сверточная сеть LeNet-5 [11]	-	-	-/-	-	0,95
МинМаксГабор _{16×2} ¹⁶⁸	-	-	1/(1/N)	5120/-	0,77
МинМаксГабор _{16×2} ¹⁸⁰	-	-	1/(1/N)	5760/-	0,72
МинМаксГабор _{9×2} ¹⁶⁰ [7]	-	-	31,5/21,5	2880/-	0,71
Разреженное кодирование [5]	-	-	-/-	-/-	0,62-0,64
МинМаксГабор _{24×2} ¹⁹⁸	104	1,4	$2^{2,75}/2^{-10}$	9504/325	0,60
Разреженное кодирование [7]	-	-	93/2,5	3042/-	0,59
Другие варианты SVM [11]	-	-	-/-	-/-	0,56-1,1
Разреженные операторы [13]	-	-	-/-	4 ² «диска»/-	0,43
Сверточная нелинейная сеть [14]	-	-	-/-	-/-	0,39

*серым выделены методы, реализованные и тестируемые в данной работе, $T_{обр.}$ – время обработки, $T_{обуч.}$ – время обучения, N – размер вектора признаков, ϵ – ошибка классификации.

В таблице указаны только работы, в которых не использовались дополнительные тренировочные данные. Так, например, в [12] с помощью комбинации 35 сверточных сетей, обученной на выборке MNIST с добавлением искусственно искаженных экземпляров, получена ошибка всего 0,23%, что близко к результатам классификации образов человеком.

4. Обсуждение модели

Производительность. Для обработки изображения размером 28×28 пикселей методом, предлагаемым в данной работе, требуется в среднем ~1 сек (табл. 2), что приводит к общему времени обработки всех изображений MNIST и обучения около суток (рис. 7,в). Для сравнения, например, свертка изображения с 40 фильтрами занимает всего 8-10 мс (с использованием преобразования Фурье), что на 2 порядка быстрее. Тем не менее, время, требуемое для нашего метода, можно считать стандартным для многих современных методов (сверточная сеть в [12], ограниченная машина Больцмана обучается около двух дней⁷ или неделю в [15, с. 8]). Производительность *оптимизированного* метода сверточного разреженного кодирования примерно соответствует нашему методу: в [10]

⁷ Информация с сайта <http://www.cs.toronto.edu/~rsalakhu/DBM.html>.

для изображения размером 50×50 пикселей требуется $\sim 2,5$ сек для сходимости энергетической функции.

Некоторые свойства модели. Разработанная модель перспективна для применения в области анализа последовательностей изображений, так как обладает следующим полезным для этой задачи свойством: изменения образа, включая смещение, вращение вокруг осей, растяжение/сжатие (вдоль или вокруг осей x, y, z соответственно) приводит к характерным изменениям откликов в пространственной и/или частотной областях, включая появление дополнительных частотных и/или амплитудных модуляций отклика. В данной статье это свойство предлагается наблюдать качественно (рис. 8).

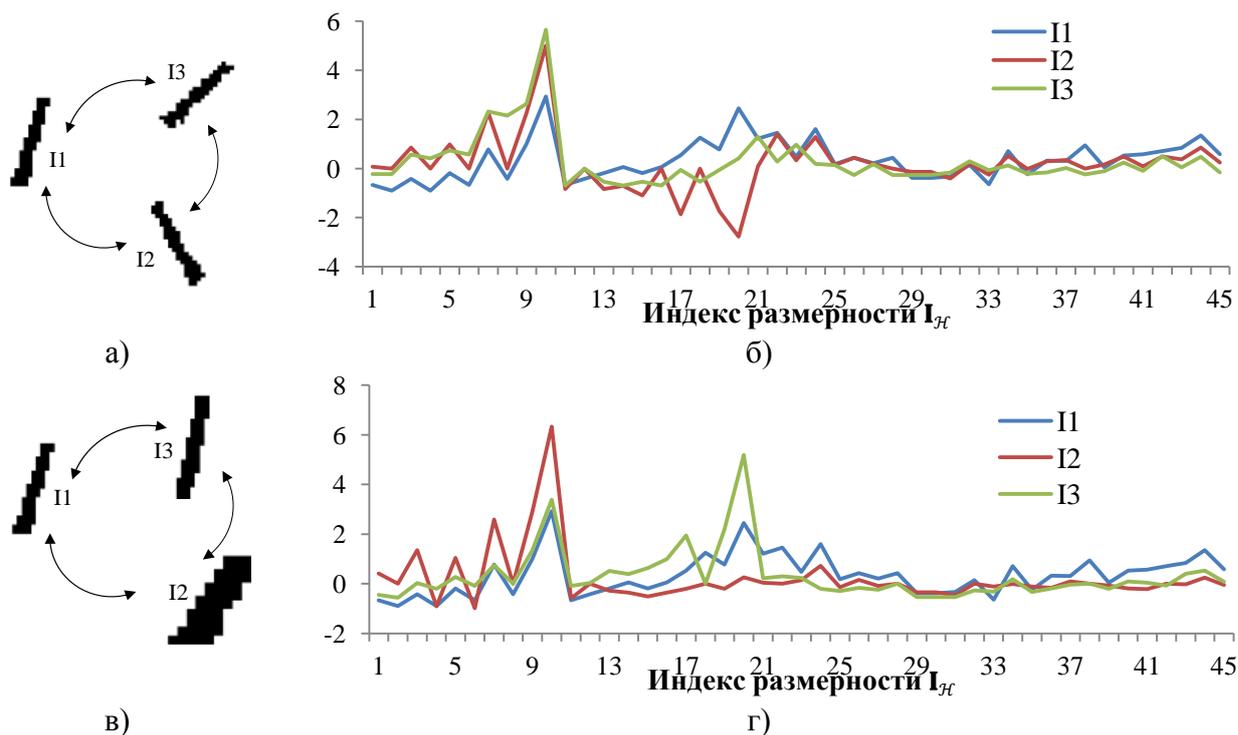


Рис. 8. Демонстрация свойств модели: а – преобразование, идентичное изображенному на рис. 2,а; в – увеличение масштаба образа по осям x и y ; б,в – значения в некоторых размерностях полученного представления.

Всплески значений сконцентрированы, в основном, на тех позициях вектора $I_{ж}$, на которых произошли изменения, а не распределены по всем позициям как в методе РСА (рис. 2). В отличие от генеративных методов на основе минимизации ошибки реконструкции, данный метод представляет размерности, обладающие более явным физическим смыслом.

Анализ последовательностей. В [23] показаны временные диаграммы, используемые для классификации эмоциональных состояний по видеоизображению. С помощью разработанной модели получены аналогичные графики изменений параметров $(x_0, y_0, \lambda, \sigma_x, s_{min}, \gamma, \theta, \beta)$ на рис. 9,а и параметров $(x_0, y_0, \sigma_x, \gamma, \theta)$ на рис. 9,б, вычисленные для видеоизображения с человеком, выполняющим жесты элементов лица и тела в четкой последовательности. Предварительный анализ показал, что изменения параметров носят

закономерный характер и согласуются с соответствующими движениями элементов лица или тела. Для извлечения поведенческих паттернов конкретных элементов лица и тела требуется тщательный анализ получаемых последовательностей с привлечением специальных методов обработки временных данных, исследованию которых будут посвящены последующие работы.

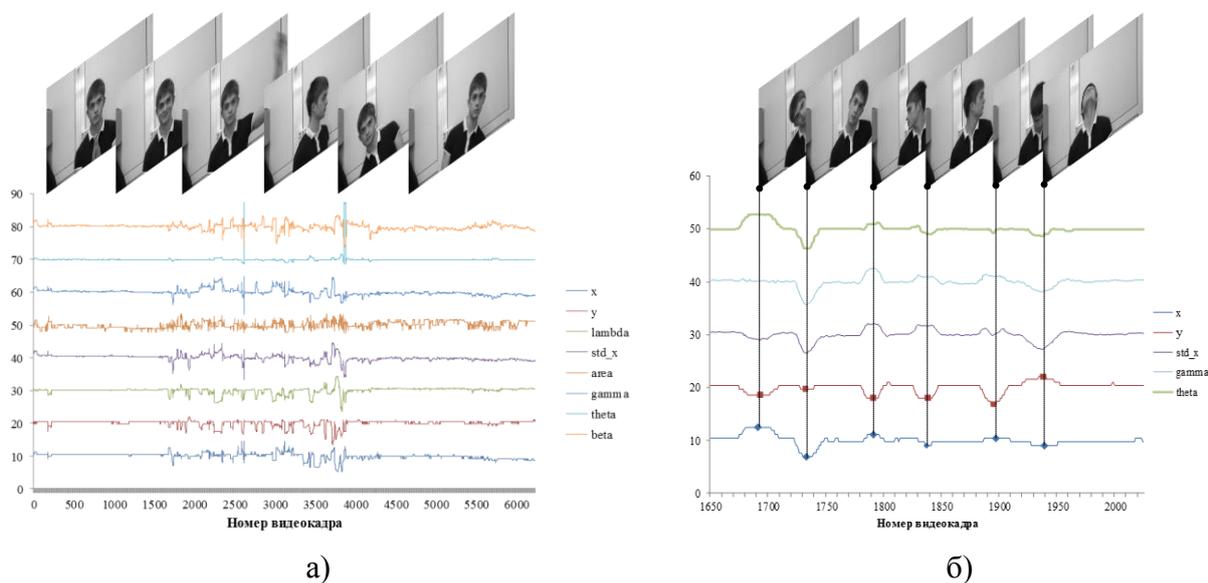


Рис. 9. Анализа последовательностей изображений: а – временная диаграмма поведения человека (1-6240 кадры видеозображения) в некоторых размерностях разработанного представления; б – временная диаграмма на интервале 1650-2025 кадры, на которых последовательно выполняются 6 движений головы. Значения на диаграммах нормализованы.

Возможность реконструкции. В случае возможности реконструкции изображения \mathbf{I} из его представления $\mathbf{I}_{\mathcal{H}}$ можно говорить об отсутствии потерь в таком представлении. Недостатком рекурсивной свертки образа с собой (оператор (6)), применяемой для вычисления $\mathbf{I}_{\mathcal{H}}$, является то, что однозначное восстановление исходного образа нетривиально. Чтобы это понять, необходимо перейти к комплексному представлению Фурье-образа. Свертка образа \mathbf{I} с собой в пространственной области означает перемножение его Фурье-образа с собой в частотной области: $\mathcal{F}(\mathbf{I} * \mathbf{I}) = \mathcal{F}(\mathbf{I}) \circ \mathcal{F}(\mathbf{I}) = \mathcal{F}(\mathbf{I})^2$. Фурье-образ, являясь комплексным представлением, может быть выражен как $\mathcal{F}(\mathbf{I}) = \mathbf{A}e^{i\varphi}$. Тогда, квадрат Фурье-образа $\mathcal{F}(\mathbf{I})^2 = \mathbf{A}^2e^{i2\varphi}$. (При использовании автокорреляционной функции (\star) в (6) реконструкция принципиально не осуществима, так как $\mathcal{F}(\mathbf{I} \star \mathbf{I}) = \mathbf{A}e^{i\varphi} \circ \mathbf{A}e^{-i\varphi} = \mathbf{A}^2$.) Восстановление исходного сигнала, казалось бы, тривиально, $\mathbf{I} = \mathcal{F}^{-1}(\sqrt{\mathcal{F}(\mathbf{I})^2})$. Однако, заметим, что $e^{i2\varphi/2}$, присутствующий в данной формуле, не дает однозначного результата, так как $e^{i2\pi k} = e^{i2\pi}$, где k – произвольное целое число. Например, предположим, что исходное значение фазы равно $\varphi \in \{0, 2\pi\}$. Для произвольного φ всегда можно найти $\varphi^* \in \{0, 2\pi\}$ и k такие, что $\varphi = \varphi^* + \pi k$, то есть противоположную на комплексной плоскости фазу. При умножении исходной фазы на два (2φ) имеем также, что $2\varphi = 2\varphi^* + 2\pi k$, или, в силу периодичности фазы с периодом 2π , $2\varphi = 2\varphi^*$. При делении 2φ на два (восстановлении) получаем соответственно два

варианта: исходную фазу φ и противоположную φ^* . Тривиального способа определить, какая из двух фаз изначально присутствовала в образе \mathbf{I} (изображении), не найдено. Между тем выбор φ^* вместо φ критически (в терминах среднеквадратичной разницы) влияет на результат реконструкции. В случае изображения размером $M \times N$ имеем соответственно $2^{M \times N}$ вариантов реконструкций, только один из которых идентичен исходному \mathbf{I} , что является недостатком модели. Одним из вариантов устранения данного недостатка является хранение информации о знаке фазы для каждой точки отклика, что приведет к дополнительному хранению $2^{M \times N}$ бит. Реконструкция также может быть осуществлена итеративно. Для каждого значения Фурье-образа необходимо выбирать ту фазу, которая ведет к меньшей ошибке реконструкции $\varepsilon = \frac{1}{\sqrt{MN}} \|\mathbf{I} - \mathbf{I}_{\text{рек}}\|_2$. Недостаток данного метода в том, что необходимо знать оригинальное изображение \mathbf{I} . Проблеме реконструкции изображения из формируемого разработанным алгоритмом представления будут посвящены последующие работы.

Заключение

В данной работе предложена, реализована и протестирована модель разреженного представления статического изображения. Разработка данной модель мотивирована возрастающей востребованностью универсальных методов, способных максимально эффективно решать узкопоставленные задачи, не имея априорных знаний о самой задаче, – одна из составляющих автономного искусственного интеллекта. Разработанная модель может быть расширена для случаев последовательностей изображений либо путем модификации предложенного в работе оператора свертки n -го порядка, либо путем анализа динамики параметров статического представления.

Разработан алгоритм, выходными данными которого являются пространственно-частотные паттерны (отклики), явно присутствующие в изображениях, что делает данный метод аналогичным методу разреженного представления. В отличие от других работ, извлекаемые паттерны соответствуют *произвольным* частям изображений. Другим отличием является то, что вместо попиксельного представления, характерного для предыдущих методов, предлагается параметрическое описание паттернов, что позволяет более компактно формировать описание изображений независимо от их размеров. Показано, что вычисляемые параметры репрезентативны, так как позволяют решать задачу классификации образов с приемлемой точностью (ошибка 3,99% в совокупности с методом PCA и SVM для базы MNIST). Относительно большая ошибка в первую очередь обусловлена сложностью подготовки вектора для метода опорных векторов, который требует одинаковый порядок значений для всех экземпляров выборки. Поэтому актуальным является исследование альтернативных методов обучения и классификации.

Помимо этого, осуществлен вклад в метод комбинации фильтров Габора в совокупности с локальным оператором минимума-максимума, а именно, сконструированы более подходящие фильтры и подобраны оптимальные параметры модели. В результате

экспериментально показано, что данный метод является эффективным (ошибка 0,60% для базы MNIST) и может конкурировать со сверточными сетями и другими передовыми разработками.

В работе также отмечены другие особенности, свойства, достоинства и недостатки разработанной модели и алгоритма. Анализируемое на рис. 9 видеоизображение и исходные коды разработанных и используемых в данной работе алгоритмов, могут быть получены по запросу к авторам, некоторые материалы также доступны на <http://www.bmstu.ru/ps/~bknyazev/fileman/Is/2014/DigitRecognition>.

Статья подготовлена в рамках выполнения государственного задания № 2014/104.

Список литературы

1. Olshausen B., Field D. Emergence of simple-cell receptive field properties by learning a sparse code for natural images // Nature. 1996. No. 381 (6583). P. 607-609. DOI: [10.1038/381607a0](https://doi.org/10.1038/381607a0)
2. Baccouche M., Mamalet F., Wolf C., Garcia C., Baskurt A. Spatio-Temporal Convolutional Sparse Auto-Encoder for Sequence Classification // In: Proc. British Machine Vision Conference. University of Surrey, Guildford, United Kingdom, 2012, Vol. 18, no. 5-6.
3. Daugman J.G. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters // Journal of the Optical Society of America A. 1985. Vol. 2, no. 7. P. 1160-1169.
4. Petkov N. Biologically motivated computationally intensive approaches to image pattern recognition // Future Generation Computer Systems. 1995. Vol. 11, iss. 4-5. P. 451-465. DOI: [10.1016/0167-739X\(95\)00015-K](https://doi.org/10.1016/0167-739X(95)00015-K)
5. Ranzato M., Fu Jie Huang, Boureau Y.-L., LeCun Y. Unsupervised Learning of Invariant Feature Hierarchies with Applications to Object Recognition // IEEE Conference on Computer Vision and Pattern Recognition (CVPR'07). IEEE, 2007. P. 1-8. DOI: [10.1109/CVPR.2007.383157](https://doi.org/10.1109/CVPR.2007.383157)
6. Kavukcuoglu K. Learning Feature Hierarchies for Object Recognition. Ph.D. Dissertation. New York University, New York, NY, USA, 2010.
7. Labusch K., Barth E., Martinetz T. Simple Method for High-Performance Digit Recognition Based on Sparse Coding // IEEE Transactions on Neural Networks. 2008. Vol. 19, no. 11. P. 1985-1989. DOI: [10.1109/TNN.2008.2005830](https://doi.org/10.1109/TNN.2008.2005830)
8. Raina R., Battle A., Lee H., Packer B., Y. Ng A. Self-taught learning: transfer learning from unlabeled data // In: Proceedings of the 24th International Conference on Machine Learning (ICML '07). ACM, New York, NY, USA, 2007. P. 759-766. DOI: [10.1145/1273496.1273592](https://doi.org/10.1145/1273496.1273592)
9. Gregor K., LeCun Y. Emergence of Complex-Like Cells in a Temporal Product Network with Local Receptive Fields // arXiv.org, 2010. Art. no. [arXiv:1006.0448](https://arxiv.org/abs/1006.0448)

10. Bristow H., Eriksson A., Lucey S. Fast Convolutional Sparse Coding // 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR '13). IEEE Computer Society, Washington, DC, USA, 2013. P. 391-398. DOI: [10.1109/CVPR.2013.57](https://doi.org/10.1109/CVPR.2013.57)
11. LeCun Y., Bottou L., Bengio Y., Haffner P. Gradient-based learning applied to document recognition // Proceedings of the IEEE. 1998. Vol. 86, no. 11. P. 2278-2324. DOI: [10.1109/5.726791](https://doi.org/10.1109/5.726791)
12. Cireşan D., Meier U., Schmidhuber J. Multi-column Deep Neural Networks for Image Classification // 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'12). IEEE, 2012. P. 3642-3649. DOI: [10.1109/CVPR.2012.6248110](https://doi.org/10.1109/CVPR.2012.6248110)
13. Bruna J., Mallat S. Invariant Scattering Convolution Networks // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2013. Vol. 35, iss. 8. P. 1872-1886. DOI: [10.1109/TPAMI.2012.230](https://doi.org/10.1109/TPAMI.2012.230)
14. Mairal J., Koniusz P., Harchaoui Z., Schmid C. Convolutional Kernel Network // arXiv.org, 2014. Art. no. [arXiv:1406.3332](https://arxiv.org/abs/1406.3332)
15. Hinton G.E., Osindero S., Teh Y.-W. A Fast Learning Algorithm for Deep Belief Nets // Neural Computation. 2006. Vol. 18, no. 7. P. 1527-1554. DOI: [10.1162/neco.2006.18.7.152](https://doi.org/10.1162/neco.2006.18.7.152)
16. Трубаков А.О. Методы и алгоритмы многомерного моделирования пространства характеристик изображений: дис. ... канд. техн. наук. Брянск, 2011. 214 с.
17. Сафронов К.В. Иерархический итерационный метод распознавания объектов на основе анализа многомерных данных: дис. ... канд. техн. наук. Уфа, 2008. 164 с.
18. Bishop C.M. Pattern Recognition and Machine Learning. Berlin: Springer, 2006. 738 p. (Ser. Information Science and Statistics).
19. Конспект лекции «Уменьшение размерности описания данных: метод главных компонент» по курсу «Математические основы теории прогнозирования» // MachineLearning.ru : информационно-аналитический ресурс по машинному обучению, распознаванию образов и интеллектуальному анализу данных, 2011. Режим доступа: http://www.machinelearning.ru/wiki/images/a/a4/MOTP11_5.pdf (дата обращения 02.10.2014).
20. Viola P., Jones M. Rapid object detection using a boosted cascade of simple features // 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'01). Vol. 1. IEEE, 2001. P. 511-518. DOI: [10.1109/CVPR.2001.990517](https://doi.org/10.1109/CVPR.2001.990517)
21. Нехина А.А., Князев Б.А., Кашапова Л.Х., Спиридонов И.Н. Использование онтологической модели знаний и программных средств сенсора Kinect описания позирования человека // Биомедицинская радиоэлектроника. 2012. № 12. С. 54-60.
22. Князев Б.А., Нехина А.А. Исследование и разработка мультиагентного аппаратно-программного комплекса распознавания позы человека // Инженерный вестник. 2013. № 7. С. 523-538. Режим доступа: <http://engbul.bmstu.ru/doc/598836.html> (дата обращения 01.10.2014).

23. Hamm J., Kohler C.G., Gur R.C., Verma R. Automated facial action coding system for dynamic analysis of facial expressions in neuropsychiatric disorders // *Journal of Neuroscience Methods*. 2011. Vol. 200, no. 2. P. 237-256.
24. Князев Б.А., Гапанюк Ю.Е. Распознавание аномального поведения человека по его эмоциональному состоянию и уровню напряженности с использованием экспертных правил // *Инженерный вестник*. 2013. № 3. С. 509-524. Режим доступа: <http://engbul.bmstu.ru/doc/568250.html> (дата обращения 01.10.2014).
25. Князев Б.А., Черненький В.М. Методика и модель кластеризации паттернов двигательной активности лица как преобразований метаграфов // *Вестник МГТУ им. Н.Э. Баумана. Сер. Приборостроение*. 2014. № 4. С. 34-54.
26. Кашапова Л.Х., Латышева Е.Ю., Спиридонов И.Н. Алгоритм распознавания эмоционального состояния по изображениям лица с использованием дискриминантного анализа и фильтров Габора // *Медицинская техника*. 2012. № 3. С. 1-4.
27. Solmaz B., Assari S. M., Shah M. Classifying Web Videos using a Global Video Descriptor // *Machine Vision and Applications (MVA)*. 2013. Vol. 24, no. 7. P. 1473-1485.
28. Lindeberg T. A computational theory of visual receptive fields // *Biological Cybernetics*. 2013. Vol.107, iss. 6. P. 589-635. DOI: [10.1007/s00422-013-0569-z](https://doi.org/10.1007/s00422-013-0569-z)
29. Самаль Д. М. Алгоритмы идентификации человека по фотопортрету на основе геометрических преобразований: дис. ... канд. техн. наук. Минск, 2002. 167 с.
30. Cortes C., Vapnik V. Support-Vector Networks // *Machine Learning*. 1995. Vol. 20, no. 3. P. 273-297. DOI: [10.1007/BF00994018](https://doi.org/10.1007/BF00994018)
31. Chang C.-C., Lin C.-J. LIBSVM: A library for support vector machines // *ACM Transactions on Intelligent Systems and Technology*. 2011. Vol. 2, iss. 3. Article no. 27. DOI: [10.1145/1961189.1961199](https://doi.org/10.1145/1961189.1961199)

Convolutional Sparse Coding for Static and Dynamic Images Analysis

B.A. Knyazev^{1,*}, V.M. Chernenkiy¹

[*bknyazev@bmsturu](mailto:bknyazev@bmsturu)

¹Bauman Moscow State Technical University, Moscow, Russia

Keywords: convolution, filters, Gabor, parametric representation, sparse coding, support vector machine, handwritten digits

The objective of this work is to improve performance of static and dynamic objects recognition. For this purpose a new image representation model and a transformation algorithm are proposed. It is examined and illustrated that limitations of previous methods make it difficult to achieve this objective. Static images, specifically handwritten digits of the widely used MNIST dataset, are the primary focus of this work. Nevertheless, preliminary qualitative results of image sequences analysis based on the suggested model are presented.

A general analytical form of the Gabor function, often employed to generate filters, is described and discussed. In this research, this description is required for computing parameters of responses returned by our algorithm. The recursive convolution operator is introduced, which allows extracting free shape features of visual objects. The developed parametric representation model is compared with sparse coding based on energy function minimization.

In the experimental part of this work, errors of estimating the parameters of responses are determined. Also, parameters statistics and their correlation coefficients for more than 106 responses extracted from the MNIST dataset are calculated. It is demonstrated that these data correspond well with previous research studies on Gabor filters as well as with works on visual cortex primary cells of mammals, in which similar responses were observed. A comparative test of the developed model with three other approaches is conducted; speed and accuracy scores of handwritten digits classification are presented. A support vector machine with a linear or radial basic function is used for classification of images and their representations while principal component analysis is used in some cases to prepare data beforehand. High accuracy is not attained due to the specific difficulties of combining our model with a support vector machine (a 3.99% error rate). However, another method is improved, which is based on Gabor filters and the local minimum-maximum operator. Compound filters are designed and optimal parameters are experimentally found leading to high accuracy (a 0.60% error rate). Performance of this method might be further increased by constructing new filters with parameters based on the statistics obtained by our algorithm.

References

1. Olshausen B., Field D. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 1996, no. 381 (6583), pp. 607-609. DOI: [10.1038/381607a0](https://doi.org/10.1038/381607a0)
2. Baccouche M., Mamalet F., Wolf C., Garcia C., Baskurt A. Spatio-Temporal Convolutional Sparse Auto-Encoder for Sequence Classification. In: *Proc. British Machine Vision Conference*. University of Surrey, Guildford, United Kingdom, 2012, Vol. 18, no. 5-6.
3. Daugman J.G. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America A*, 1985, vol. 2, no. 7, pp. 1160-1169.
4. Petkov N. Biologically motivated computationally intensive approaches to image pattern recognition. *Future Generation Computer Systems*, 1995, vol. 11, iss. 4-5, pp. 451-465. DOI: [10.1016/0167-739X\(95\)00015-K](https://doi.org/10.1016/0167-739X(95)00015-K)
5. Ranzato M., Fu Jie Huang, Boureau Y.-L., LeCun Y. Unsupervised Learning of Invariant Feature Hierarchies with Applications to Object Recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'07)*. IEEE, 2007, pp. 1-8. DOI: [10.1109/CVPR.2007.383157](https://doi.org/10.1109/CVPR.2007.383157)
6. Kavukcuoglu K. *Learning Feature Hierarchies for Object Recognition. Ph.D. Dissertation*. New York University, New York, NY, USA, 2010.
7. Labusch K., Barth E., Martinetz T. Simple Method for High-Performance Digit Recognition Based on Sparse Coding. *IEEE Transactions on Neural Networks*, 2008, vol. 19, no. 11, pp. 1985-1989. DOI: [10.1109/TNN.2008.2005830](https://doi.org/10.1109/TNN.2008.2005830)
8. Raina R., Battle A., Lee H., Packer B., Y. Ng A. Self-taught learning: transfer learning from unlabeled data. In: *Proceedings of the 24th International Conference on Machine Learning (ICML '07)*. ACM, New York, NY, USA, 2007, pp. 759-766. DOI: [10.1145/1273496.1273592](https://doi.org/10.1145/1273496.1273592)
9. Gregor K., LeCun Y. *Emergence of Complex-Like Cells in a Temporal Product Network with Local Receptive Fields*. arXiv.org, 2010, art. no. [arXiv:1006.0448](https://arxiv.org/abs/1006.0448)
10. Bristow H., Eriksson A., Lucey S. Fast Convolutional Sparse Coding. *2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR '13)*. IEEE Computer Society, Washington, DC, USA, 2013, pp. 391-398. DOI: [10.1109/CVPR.2013.57](https://doi.org/10.1109/CVPR.2013.57)

11. LeCun Y., Bottou L., Bengio Y., Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998, vol. 86, no. 11, pp. 2278-2324. DOI: [10.1109/5.726791](https://doi.org/10.1109/5.726791)
12. Cireşan D., Meier U., Schmidhuber J. Multi-column Deep Neural Networks for Image Classification. *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'12)*. IEEE, 2012, pp. 3642-3649. DOI: [10.1109/CVPR.2012.6248110](https://doi.org/10.1109/CVPR.2012.6248110)
13. Bruna J., Mallat S. Invariant Scattering Convolution Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, vol. 35, iss. 8, pp. 1872-1886. DOI: [10.1109/TPAMI.2012.230](https://doi.org/10.1109/TPAMI.2012.230)
14. Mairal J., Koniusz P., Harchaoui Z., Schmid C. *Convolutional Kernel Network*. arXiv.org, 2014, art. no. [arXiv:1406.3332](https://arxiv.org/abs/1406.3332)
15. Hinton G.E., Osindero S., Teh Y.-W. A Fast Learning Algorithm for Deep Belief Nets. *Neural Computation*, 2006, vol. 18, no. 7, pp. 1527-1554. DOI: [10.1162/neco.2006.18.7.152](https://doi.org/10.1162/neco.2006.18.7.152)
16. Trubakov A.O. *Metody i algoritmy mnogomernogo modelirovaniia prostranstva kharakteristik izobrazhenii. Kand. dis.* [Methods and algorithms for multi-dimensional modeling of the space image characteristics. Cand. diss.]. Briansk, 2011. 214 s. (in Russian).
17. Safronov K.V. *Ierarkhicheskii iteratsionnyi metod raspoznavaniia ob"ektov na osnove analiza mnogomernykh dannykh. Kand. dis.* [Hierarchical iterative method for object recognition based on the analysis of multidimensional data. Cand. diss.]. Ufa, 2008. 164 c. (in Russian).
18. Bishop C.M. *Pattern Recognition and Machine Learning*. Berlin, Springer, 2006. 738 p. (Ser. *Information Science and Statistics*).
19. Konspekt lektsii “Umen'shenie razmernosti opisaniia dannykh: metod glavnykh component” po kursu “Matematicheskie osnovy teorii prognozirovaniia” [Lecture notes “Reducing the dimension of data description: The principal components method” on the course “Mathematical foundations of the theory of prediction”]. MachineLearning.ru : information and analytical resource on machine learning, pattern recognition and data mining, 2011. Available at: http://www.machinelearning.ru/wiki/images/a/a4/MOTP11_5.pdf , accessed 02.10.2014. (in Russian).
20. Viola P., Jones M. Rapid object detection using a boosted cascade of simple features. *2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'01). Vol. 1*. IEEE, 2001, pp. 511-518. DOI: [10.1109/CVPR.2001.990517](https://doi.org/10.1109/CVPR.2001.990517)

21. Nekhina A.A., Kniazev B.A., Kashapova L.Kh., Spiridonov I.N. Applying an ontology approach and Kinect sdk to human posture description. *Biomeditsinskaia radioelektronika = Biomedical Radioelectronics*, 2012, no. 12, pp. 54-60. (in Russian).
22. Kniazev B.A., Nekhina A.A. Research and development of multi-agent hardware-software complex of recognition of human body posture. *Inzhenernyi vestnik MGTU im. N.E. Baumana = Engineering Herald of the Bauman MSTU*, 2013, no. 7, pp. 523-538. Available at: <http://engbul.bmstu.ru/doc/598836.html> , accessed 01.10.2014. (in Russian).
23. Hamm J., Kohler C.G., Gur R.C., Verma R. Automated facial action coding system for dynamic analysis of facial expressions in neuropsychiatric disorders. *Journal of Neuroscience Methods*, 2011, vol. 200, no. 2, pp. 237-256.
24. Kniazev B.A., Gapaniuk Iu.E. Recognition of abnormal human behavior through his emotional state and the level of tension using expert rules. *Inzhenernyi vestnik MGTU im. N.E. Baumana = Engineering Herald of the Bauman MSTU*, 2013, no. 3, pp. 509-524. Available at: <http://engbul.bmstu.ru/doc/568250.html> , accessed 01.10.2014. (in Russian).
25. Kniazev B.A., Chernen'kii V.M. Method and Model for Clustering Facial Activity Patterns using Metagraph Transformations. *Vestnik MGTU. Ser. Priborostroenie = Herald of the Bauman MSTU. Ser. Instrument Engineering*, 2014, no. 4, pp. 34-54. (in Russian).
26. Kashapova L.Kh., Latysheva E.Iu., Spiridonov I.N. Discriminant Analysis of Two-Dimensional Gabor Features for Facial Expression Recognition. *Medit'sinskaia tekhnika*, 2012, no. 3, pp. 1-4. (English translation: *Biomedical Engineering*, 2012, vol. 46, no. 3, pp. 89-92. DOI: [10.1007/s10527-012-9274-9](https://doi.org/10.1007/s10527-012-9274-9)).
27. Solmaz B., Assari S. M., Shah M. Classifying Web Videos using a Global Video Descriptor. *Machine Vision and Applications (MVA)*, 2013, vol. 24, no. 7, pp. 1473-1485.
28. Lindeberg T. A computational theory of visual receptive fields. *Biological Cybernetics*, 2013, vol.107, iss. 6, pp. 589-635. DOI: [10.1007/s00422-013-0569-z](https://doi.org/10.1007/s00422-013-0569-z)
29. Samal' D.M. *Algoritmy identifikatsii cheloveka po fotoportretu na osnove geometricheskikh preobrazovanii. Kand. dis.* [Algorithms of face recognition based on geometric transformations. Cand. diss.]. Minsk, 2002. 167 p. (in Russian).
30. Cortes C., Vapnik V. Support-Vector Networks. *Machine Learning*, 1995, vol. 20, no. 3, pp. 273-297. DOI: [10.1007/BF00994018](https://doi.org/10.1007/BF00994018)
31. Chang C.-C., Lin C.-J. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2011, vol. 2, iss. 3, article no. 27. DOI: [10.1145/1961189.1961199](https://doi.org/10.1145/1961189.1961199)