

МОДЕЛИРОВАНИЕ РАБОТЫ ПРОСТРАНСТВЕННО-ЧАСТОТНЫХ ФИЛЬТРОВ ПРИ ВОСПРИЯТИИ СЛОЖНЫХ ДИНАМИЧЕСКИХ СЦЕН

© 2014 г. Е. В. Логунова^{*}; С. В. Пронин^{**}; Ю. Е. Шелепин^{***}, доктор мед. наук

^{*}Санкт-Петербургский государственный университет, Санкт-Петербург

^{**}Институт физиологии им. И.П. Павлова РАН, Санкт-Петербург

^{***}Санкт-Петербургский национальный исследовательский университет информационных технологий механики и оптики, Санкт-Петербург

E-mail: yshelepin@yandex.ru

В работе исследован процесс восприятия динамических изображений, подвергнутых обработке фильтрами пространственных частот, имитирующими характеристики рецептивных полей нейронов первичной зрительной коры. Была использована методика, позволяющая дать количественную оценку восприятия испытуемыми эмоционального состояния лиц людей на предъявляемых изображениях. Было показано, что в процессе восприятия изображений лиц, помимо вертикальных и горизонтальных составляющих пространственно-частотного спектра, существенную роль играют диагональные составляющие. Несмотря на то, что чувствительность зрительной системы к диагональным составляющим меньше, чем к вертикальным и горизонтальным, содержащаяся в них информация позволяет выделять индивидуальные особенности и эмоциональное состояние лица человека.

Ключевые слова: динамические изображения, зрение, пространственно-частотная фильтрация, ориентационная избирательность, распознавание.

Коды OCIS: 330.7310, 330.6110.

Поступила в редакцию 05.08.2014.

Введение

При построении образа сложной сцены зрительная система человека анализирует физические параметры воспринимаемого сигнала и выбирает из них значимые для наблюдателя свойства. Согласно развиваемым уже более полувека моделям обработки информации в зрительной системе [1–10], воспринимаемое изображение сцены проходит несколько следующих этапов фильтрации:

1) Этап первичной фильтрации, цель которого – анализ физических параметров сигнала. К первичной фильтрации относят согласованную работу оптики глаза и рецептивных полей нейронов сетчатки, подкорковых ядер и зрительной коры. Работа рецептивных полей нейронов от сетчатки до нейронов зрительной коры обычно рассматривают как многоканальную фильтрацию.

2) Этап согласованной фильтрации – осуществление сравнения данного изображения с имеющимся “динамическим шаблоном”, хранящимся в памяти.

3) Этап принятия решения. Только этот этап обычно является осознанным. Он зависит от контекста, воспринимаемой сцены, эмоционального состояния и мотивации человека.

Очевидно, что характеристики используемых зрительным анализатором фильтров должны соответствовать задаче выделения наиболее значимых для наблюдателя свойств объектов. Известно, что для распознавания большого класса объектов достаточно только вертикальных и горизонтальных составляющих пространственно-частотного спектра изображений. В работах [11, 12] это свойство зрительной системы было использовано для сжатия видеосигналов. На использовании только вертикальных и горизонтальных составляющих основан широко применяемый в области технического зрения метод детектирования объектов Виолы–Джонса [13]. Однако известно, что ориентации рецептивных полей первичной зрительной коры человека равномерно покрывают весь диапазон от 0 до 360 градусов. Так как пространственно-частотные характеристики рецептивных полей

отражают статистические характеристики визуальной среды, очевидно, что диагональные составляющие также содержат ценную информацию, позволяющую идентифицировать какие-то важные свойства объектов. Целью данной работы является исследование характера информации, содержащейся в вертикальных, горизонтальных и диагональных пространственно-частотных составляющих изображений.

Роль этой информации можно исследовать, наблюдая восприятие людьми изображений каких-либо значимых объектов (лиц, животных, пейзажей), подвергнутых цифровой фильтрации с целью выделения требуемого диапазона ориентаций пространственно-частотных составляющих. Для этого, прежде всего, необходимо выбрать способ цифровой фильтрации. Известно, что профили весовых функций рецептивных полей корковых нейронов хорошо описываются синусоидой, промодулированной гауссианом, т.е. элементом Габора [5, 7]. Поэтому, целесообразно осуществлять фильтрацию путем свертки изображений с элементом Габора, близким по характеристикам к весовым функциям рецептивных полей. Кроме того, необходима методика, позволяющая хотя бы качественно оценить объем информации, воспринятой наблюдателем из различных видов отфильтрованных изображений.

Методика исследования

На первом этапе были выделены видеофрагменты с длительностью 3 с. Сюжеты видеофрагментов выбирались в соответствии со следующими задачами:

- 1) “выделить на сложном фоне фигуру”,
- 2) “выделить фигуру и узнать знакомое лицо”,
- 3) “определить эмоциональное состояние и настроение человека”.

Для первой задачи были использованы видеофрагменты, где основное действие объектов заключается в передвижении их на сложном заднем плане (фоне). Для второй задачи были отобраны видеофрагменты с перемещением персонажей с заднего плана на передний. Для третьей задачи использовались кадры с крупным изображением лица человека на фоне сложной сцены.

В ходе подготовки стимулов видеофрагменты были разложены на кадры. Затем каждый кадр был подвергнут цифровой фильтрации пу-

тем свертки с элементом Габора. Элементы Габора рассчитывались по следующей формуле:

$$z(x,y) = \sin(2\pi x f) \exp\left(-\frac{(x^2 + y^2)}{2\sigma^2}\right)$$

где x, y – координаты пикселов изображения, f – пространственная частота в периодах на ширину изображения, σ^2 – дисперсия гауссоиды. Значение дисперсии гауссоиды было выбрано равным 35, пространственной частоты – 40. Элементы Габора имели либо вертикальную или горизонтальную ориентацию, либо две диагональные. В последнем случае отфильтрованное изображение представляло из себя сумму двух изображений, полученных при помощи фильтрации элементами Габора с ориентациями 45 и 135°. После фильтрации из полученных кадров формировался видеофрагмент, используемый в экспериментах.

Для того, чтобы как-то оценить способность испытуемых извлекать значимую информацию из синтезированных видеофрагментов, была использована следующая методика. Испытуемым на короткое время (300 мс и 1000 мс) предъявляли видеофрагменты, содержащие изображения эмоциональных человеческих лиц. После предъявления фрагментов им предлагалось выбрать наиболее подходящую эмоцию из предложенных: “Усталость”, “Надежда”, “Спокойствие”, “Беспокойство”, “Интерес”, “Радость”, “Печаль”. В результате анализа определялось, какую эмоцию наиболее часто выбирает испытуемый. От общего количества выборов определялось процентное соотношение к наиболее часто выбираемым эмоциям. По этим соотношениям можно было оценить, насколько точно испытуемым удавалось определить эмоциональное состояние персонажа из видеофрагмента. Кроме того, предлагалось оценить степень проявления эмоции по шкалам, имеющих два полюса: грустная – радостная, выражает интерес – не выражает интерес, усталая – бодрая. На первом этапе оценки испытуемый выбирает полюс, например, грустное или радостное выражение лица. После чего на шкале выставлялась соответствующая оценка (использовали шестибалльную шкалу).

Результаты и их обсуждение

Известно, что при решении задачи “узнавания” лица на простом фоне зрительная система использует горизонтальные составляющие,

которые в воспринимаемом образе формируют брови, глаза и губы. В своих исследованиях С. Дакин и Р. Ватт отмечали, что эти визуальные ключи могут быть упрощены до чередующихся темных и светлых полосок различной ширины, названных ими “штрих-код лица” [15]. При этом, несмотря на различные манипуляции со стимулами (изменение размера, освещения, ракурса) “штрих-код” остается неизменным для конкретного лица и позволяет узнать воспринимаемый образ.

Для иллюстрации данной ситуации, когда лицо и фигура человека является главным объектом восприятия, мы использовали фрагменты из фильма А. Тарковского “Зеркало”, для которых и была применена цифровая фильтрация с использованием элемента Габора в двух ориентациях – вертикальной и горизонтальной.

На рис. 1 показаны исходное изображение кадра фильма и изображения, полученные в ре-

зультате обработки информации, имитирующей работу ориентационно-избирательных рецептивных полей затылочной коры, настроенных на выделение горизонтальных и вертикальных составляющих в пространственно-частотном спектре наблюдаемого изображения [14, 15]. Горизонтальные составляющие эффективно выделяют лицо человека – это показали С. Дакин и Р. Ватт [15] при цифровой обработке статичных изображений лиц.

В ходе проведенного нами исследования выявлено, что горизонтальные составляющие хорошо выделяют не только лицо, но и объекты других семантических категорий, находящихся на переднем плане. На изображении, полученном в результате фильтрации кадров фильма “Берегись автомобиля”, после выделения горизонтальных пространственно-частотных составляющих (рис. 2б) отчетливо видны машина и мотоцикл, в то время как на изображе-

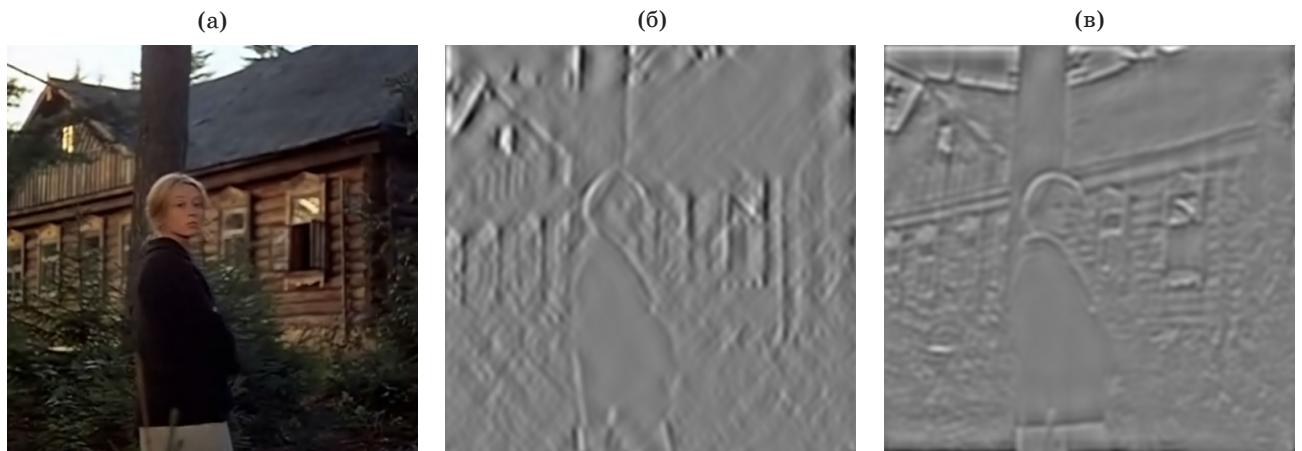


Рис. 1. Исходные кадры фильма (а) и примеры вертикальной (б) и горизонтальной (в) вейвлетной фильтрации.

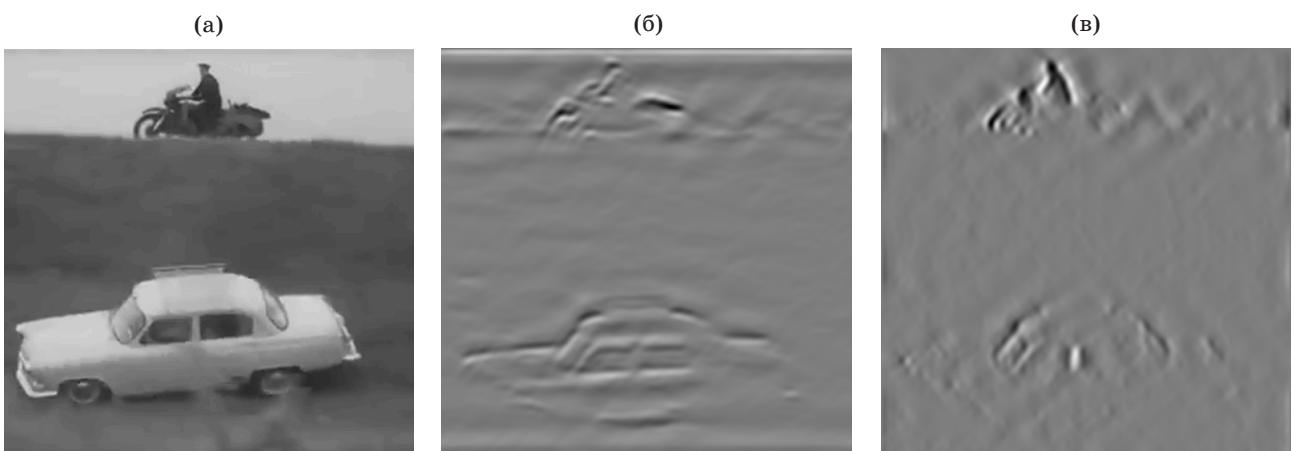


Рис. 2. Исходные кадры фильма (а) и примеры горизонтальной (б) и вертикальной (в) вейвлетной фильтрации.

нии, основой которого являются вертикальные составляющие (рис. 2в), вертикально сидящий мотоциклист выделяется лучше.

Однако, во многих случаях горизонтальных составляющих оказывается недостаточно. Это можно показать, в частности, на задаче выделения фигуры человека на сложном фоне в динамически сменяющемся потоке стимулов. Проиллюстрируем этот случай на примере следующего фрагмента фильма “Зеркало” (рис. 3).

Анализ кадров видеофрагмента показал, что в этом сюжете оптимальной для выделения идущего человека оказывается вертикальная фильтрация. Она наиболее эффективно определяют размер и положение вертикального объекта, что является первоочередной задачей обнаружения человека в динамичном восприятии сложных сцен. Именно вертикальные состав-

ляющие определяют очертания всей фигуры идущего человека.

Обнаружение и категоризация движущегося на сложном фоне объекта являются одной из непростых задач. На периферии поля зрения движение объекта обнаруживается еще до отчетливого восприятия его формы и категоризации. После этого за счет поворота глаз объект попадает в фoveальную часть сетчатки, где и происходит детальный анализ его свойств [16]. При анализе кадров, где основной задачей является выделение фигуры человека, зрительная система в первую очередь выстраивает шаблон для движущегося объекта преимущественно за счет вертикальных составляющих, что и позволяет при восприятии соотнести наблюдаемый объект с определенной семантической категорией.

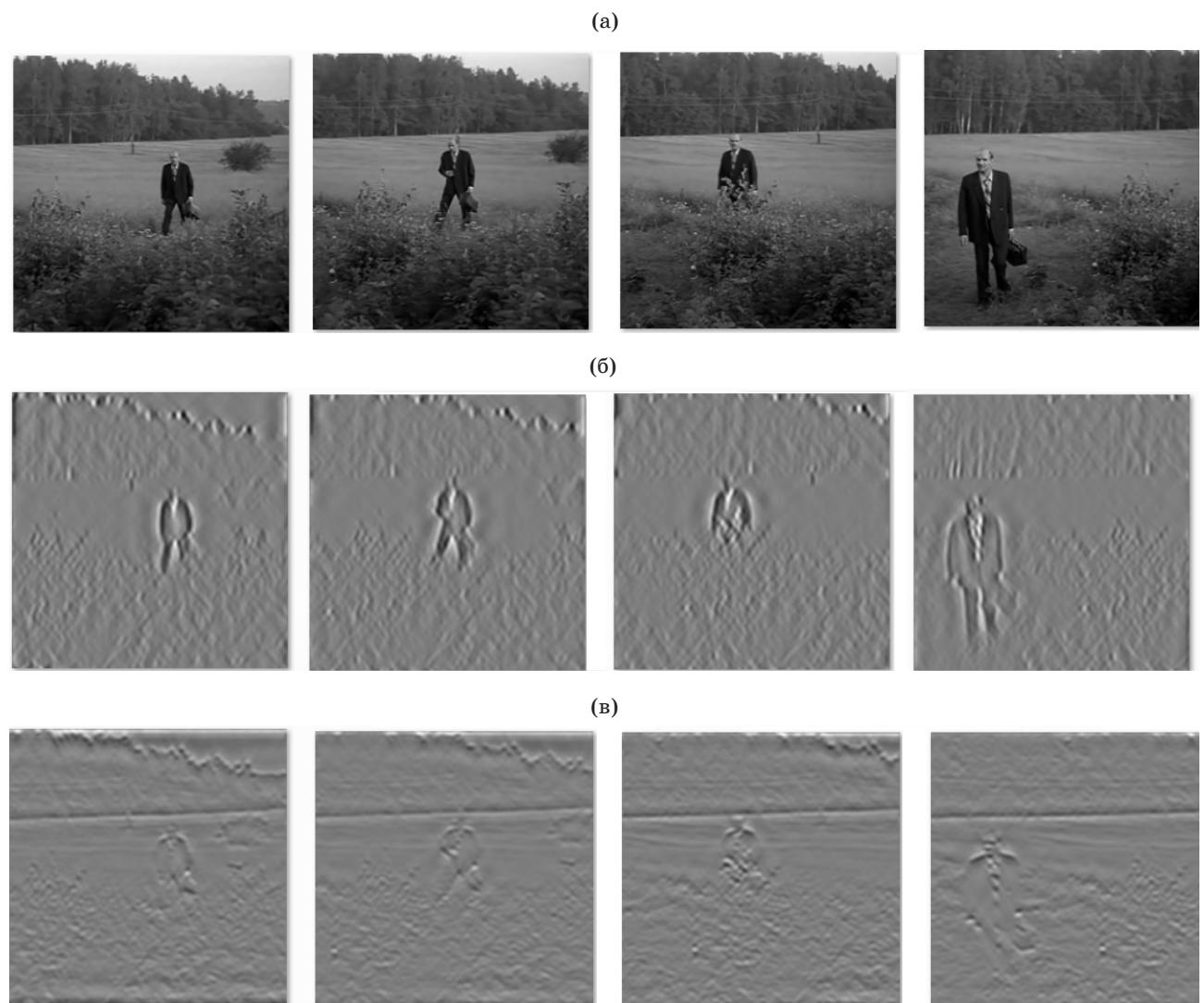


Рис. 3. Пример вейвлетной фильтрации при анализе задачи “выделение фигуры из фона”. Стока (а) – исходные кадры, (б) – вертикальная вейвлет-фильтрация, (в) – горизонтальная вейвлет-фильтрация.

Рассмотрим более сложную ситуацию, когда в динамично сменяющихся сценах требуется одновременное выполнение двух задач: “выделить из фона человека” и “узнать знакомое лицо” (рис. 4).

Изображения, полученные в результате фильтрации, показывают, что чем ближе фигуры детей, тем эффективнее становятся горизонтальные составляющие и ослабевает значение вертикальных. Так, на первых кадрах за счет вертикальных составляющих зритель выделяет фигуры детей на сложном фоне. На последующих кадрах с помощью горизонтальных составляющих определяет лицо конкретного индивидуума. Конструкция “зрительного мозга”, используя вертикальные составляющие, строит шаблон, который позволяет соотнести движущийся объект с категорией “человек”, после чего с помощью горизонтальных составляющих строится “штрих-код

лица”, с помощью которого и происходит идентификация человека [14, 15].

В реальной жизни, когда происходит перевод взора и фовеа на лицо, описание фона на периферии поля зрения становится нечетким, но размытости мы не ощущаем, т.к. на периферии поля зрения нет высокочастотных каналов. Высокочастотные каналы в фовеа детализируют выражение глаз и мимику рта, то есть выделяют области, содержащие паттерны невербальной коммуникации. Когда объектом анализа в зрительной системе наблюдателя является лицо человека, для выявления паттернов невербальной коммуникации, проявляющихся в минимальных изменениях мимики, уже недостаточно информации, содержащейся в горизонтальных и вертикальных пространственно-частотных составляющих изображений. Предположим, что

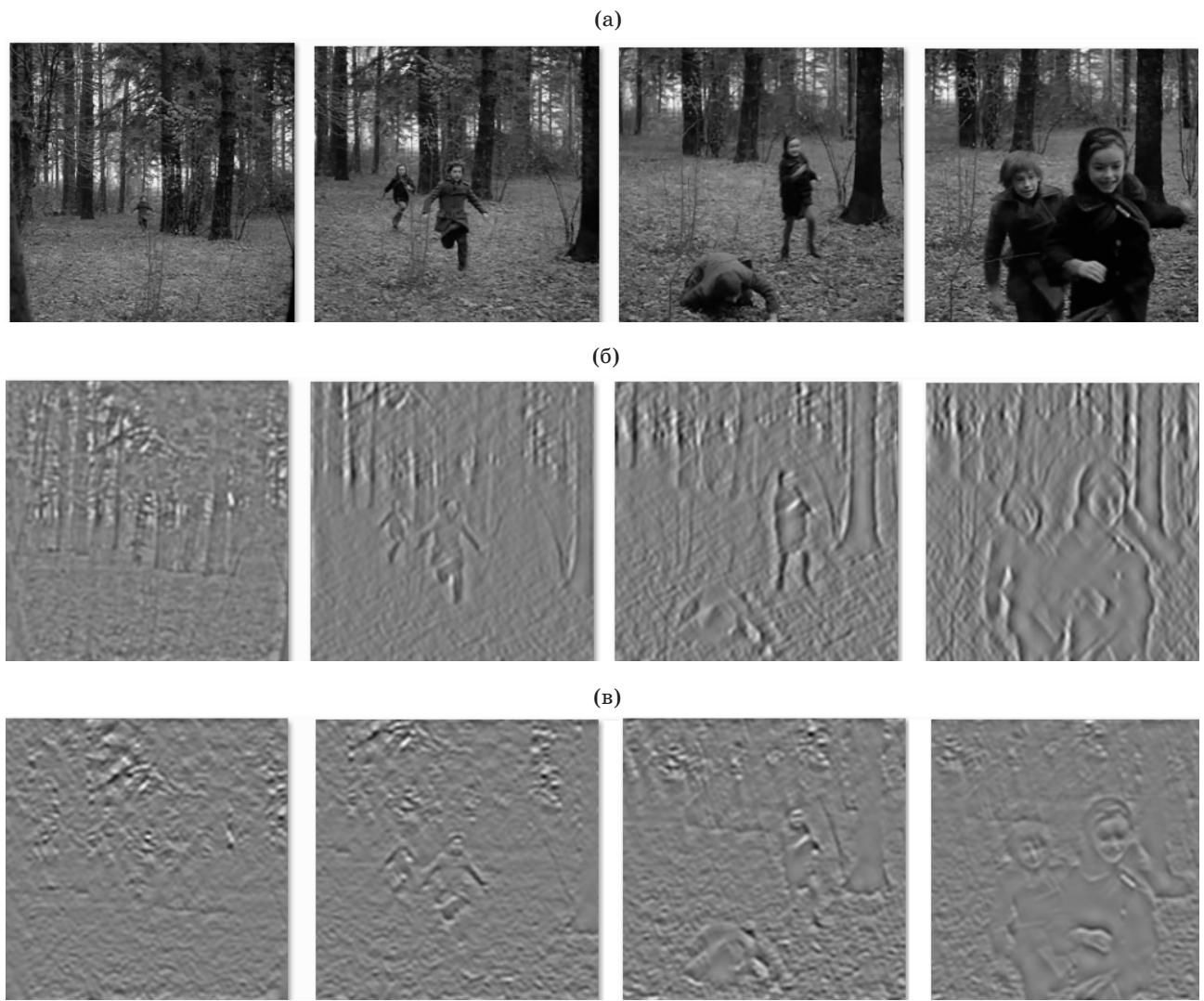


Рис. 4. Пример вейвлетной фильтрации при комбинированении задач “выделение фигуры из фона” и “узнавание лица”. Стока (а) – исходные кадры, (б) – вертикальная вейвлет-фильтрация, (в) – горизонтальная вейвлет-фильтрация.

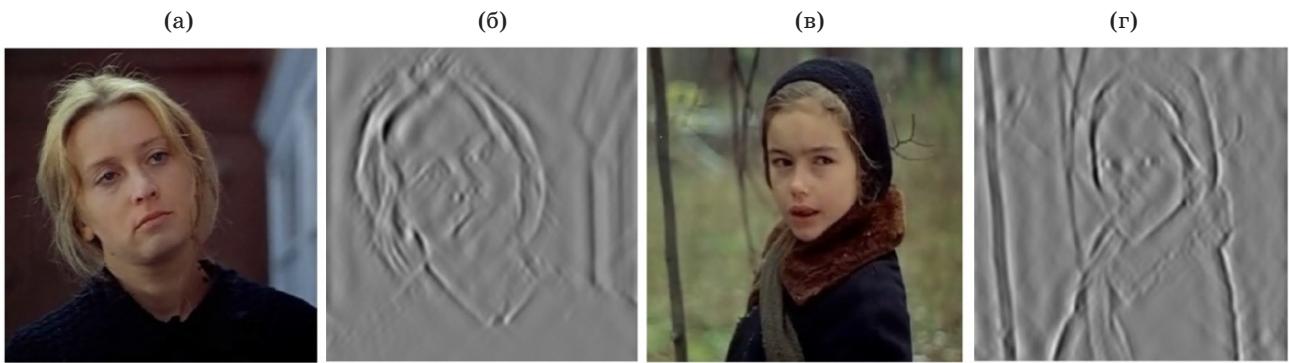


Рис. 5. Пример диагональной вейвлетной фильтрации. а, в – исходные кадры, б, г – после вейвлетной фильтрации.

она содержится в диагональных составляющих. Чтобы выяснить роль информации, содержащейся в диагональных составляющих изображений, нами был поставлен эксперимент в соответствии с подходом, изложенным в разделе “Методика исследования”. Для этой цели нами были использованы видеосюжеты “Девочка” и “Мать” (рис. 5).

Особенность выбранных кадров заключается в том, что задний фон в них размыт, а фигура и лицо персонажей остаются четкими. Это облегчает восприятие и выводит на первый план паттерны невербальной коммуникации.

В проведенном эксперименте данные видеосюжеты предъявлялись испытуемым четыре раза с интервалами 1–1,5 недели. Первые два раза предъявлялись отфильтрованные изображения сцен, третий и четвертый раз – исходные изображения сцен. Время первого и третьего предъявления стимулов – 300 мс. Время второго и четвертого предъявления стимулов – 1000 мс. Всего было опрошено 15 респондентов. После предъявления стимулов респондентам предлагали выбрать наиболее подходящую эмоцию из следующих: “Усталость”, “Надежда”, “Спокойствие”, “Беспокойство”, “Интерес”, “Радость”, “Печаль”. Как показали результаты эксперимента, диагональная фильтрация лица матери подчеркнула такую эмоциональную составляющую, как “Усталость” (отметили 80% опрошенных респондентов, на исходном изображении эмоциональное выражение лица оценивалось у большинства респондентов, как “Печаль”), а лица дочери – “Надежду” и “Интерес” (отметили 75%, в то время как на исходном изображении данную эмоцию выбрали только 27% респондентов).

Таким образом, незначительные изменения в деталях, такие как усиление за счет диагональной фильтрации мелких морщин в угол-

ках рта и глаз могут существенно изменить восприятие эмоциональных аспектов лиц, запечатленных на изображениях. Подготовка изображения для выделения незначительной, практически скрытой информации происходит неосознанно в рецептивных полях затылочной коры в результате процессов, напоминающих пространственно-частотную фильтрацию с использованием вейвлетов, подобных элементам Габора. Эту фильтрацию можно рассматривать как особую маску, “грим”, подчеркивающий мимические особенности при восприятии лиц, которые являются опорными сигналами для механизмов принятия решения на более высоких уровнях зрительной системы [17].

Выводы

Несмотря на относительную простоту пространственно-частотная фильтрация с использованием вейвлетов, имитирующих свойства рецептивных полей нейронов зрительной коры, позволяет успешно осуществить выделение значимых для наблюдателя свойств визуальных объектов. Подобная фильтрация при восприятии сцены позволяет экономично распределять “информационно-вычислительные” ресурсы головного мозга. В зависимости от того, где находится объект в воспринимаемой сцене, зрительная система динамично переключает обработку сигнала с одного участка видимого пространственно-частотного диапазона на другой. Этот процесс, зачастую неосознанный, является важнейшим этапом восприятия для своевременного принятия решений.

Работа финансируется грантом Российского научного фонда (РНФ) № 14-15-00918 “Технологии оптимизации и восстановления когнитивных функций человека виртуальной средой”.

* * *

ЛИТЕРАТУРА

1. *Campbell F., Robson J.* Application of Fourier analyses to the visibility of gratings // *J. Physiol.* 1968. V. 197. P. 551–556.
2. *Watson A.B., Barlow H.B., Robson J.G.* What does the eye see best? // *Nature.* 1983. V. 302. № 5907. P. 419–422.
3. *Шелепин Ю.Е., Колесникова Л.Н., Левкович Ю.И.* Визоконтрастометрия. Л.: Наука, 1985. 105 с.
4. *Watson A.B.* The spatial standard observer: a human vision model for display inspection // *SID Symposium Digest of Technical Papers.* 2006. V. 37. P. 1312–1315.
5. *Weiman C.* Efficient discrete Gabor functions for robot vision // *SPIE Conference on Wavelet Applications.* Orlando, FLA, 1994. April. V. 2242. P. 148–160.
6. *Watson A., Ramirez C.V., Salud E.* Predicting visibility of aircraft // *PLoS ONE.* 2009. V. 4. № 5594. P. 1–16. <http://dx.doi.org/10.1371/journal.pone.0005594>
7. *Филд Д.* Согласованные фильтры, вейвлеты и статистика натуральных сцен // *Оптический журнал.* 1999. Т. 66. № 9. С. 25–36.
8. *Жеребко А.К., Луцив В.Р.* Согласованная фильтрация в естественных и искусственных нейронных сетях // *Оптический журнал.* 1999. Т. 66. № 9. С. 69–72.
9. *Красильников Н.Н., Шелепин Ю.Е., Красильникова О.И.* Применение принципов оптимального наблюдателя при моделировании зрительной системы человека // *Оптический журнал.* 1999. Т. 66. № 9. С. 17–25.
10. *Мирошников М.М.* Согласованная фильтрация при зрительном восприятии и информационное согласование в иконике // *Оптический журнал.* 1999. Т. 66. № 9. Р. 5–8.
11. *Цуккерман И.И.* О согласовании пространственно-частотных фильтров зрительного анализатора со статистикой изображений // *Биофизика.* 1978. Т. XXIII. Вып. 6. С. 1108–1109.
12. *Цуккерман И.И., Шостацкий Н.Н.* Анизотропия пространственно-частотной характеристики зрения // *Физиология человека.* 1978. Т. 4. № 1. С. 17–20.
13. *Viola P., Jones M..* Robust real-time face detection // *Int. J. Computer Vision.* 2004. 57(2). P. 137–154.
14. *Мурыгин К.В.* Оптимизация габоровских вейвлет-преобразований для задачи распознавания человека по изображению лица // *Искусственный интеллект.* 2003. № 4. С. 223–229.
15. *Dakin S.C., Watt R.J.* Biological “bar codes” in human faces // *J. Vision.* 2009. V. 9(4):2. P. 1–10. <http://journalofvision.org/9/4/2/>, doi:10.1167/9.4.2.
16. *Кемпбелл Ф.В., Шелепин Ю.Е.* Возможности фовеолы в различении объектов // *Сенсорные системы.* 1990. Т. 4. № 2. С. 181–185.
17. *Шелепин Ю.Е., Фокин В.А., Меньшикова С.В., Борачук О.В., Коскин С.А., Соколов А.В., Пронин С.В., Харашузов А.К., Васильев П.П., Вахрамеева О.А.* Методы иконики и методы картирования мозга в оценке функционального состояния зрительной системы // *Сенсорные системы.* 2014. Т. 28. № 2. С. 61–75.