

УДК 004.932

## ОБНАРУЖЕНИЕ ЛИНИЙ ЭЛЕКТРОПЕРЕДАЧ НА ИЗОБРАЖЕНИИ С ИСПОЛЬЗОВАНИЕМ МНОГОАГЕНТНОГО ПОДХОДА

*Алпатов Б.А., д.т.н., профессор кафедры автоматики и информационных технологий в управлении (АИТУ)*

*Рязанского государственного радиотехнического университета (РГРТУ), e-mail: aitu@rsreu.ru;*

*Бабаян П.В., к.т.н., заведующий кафедрой АИТУ, e-mail: aitu@rsreu.ru;*

*Шубин Н.Ю., к.т.н., доцент кафедры АИТУ, e-mail: aitu@rsreu.ru.*

### POWER LINE DETECTION ON IMAGES USING MULTI-AGENT APPROACH

*Alpatov B.A., Babayan P.V., Shubin N.J.*

*The problem of power line detection on the images is described in this paper. The algorithm based on Radon transform and multi-agent approach is suggested. The results of experimental research of the proposed algorithm are presented.*

**Key words:** power line detection, wire detection, Radon transform, multi-agent system, pattern recognition.

**Ключевые слова:** обнаружение линий электропередач, обнаружение проводов, преобразование Радона, многоагентные системы, распознавание образов.

#### Введение

Беспилотные летательные аппараты (БПЛА) находят всё более широкие возможности применения в различных областях человеческой деятельности. Отсутствие на борту экипажа приходится компенсировать высокой степенью автоматизации управления, либо удалённым управлением. Однако, несмотря на очевидное несовершенство современных бортовых систем управления по сравнению с человеком, даже оператор не всегда способен правильно и быстро оценить обстановку и принять решение о манёвре. Возможность столкновения БПЛА с элементами ландшафта, различными постройками или другими воздушными объектами сильно ограничивает области применения БПЛА как в гражданской, так и в военной сфере. Особую опасность в условиях города для БПЛА представляют линии электропередач. Обнаружение проводов и оценка их параметров с целью информирования об их присутствии либо оператора, либо автопилота является актуальной задачей. Предлагаемый в данной работе способ обнаружения линий на изображении может быть полезен и в ряде других областей применения теории обработки изображений, например, в аэрофотосъемке, картографировании и медицине.

**Цель работы** – разработка алгоритма обнаружения и оценки параметров протяжённых проводов, наблюдаемых с подвижного носителя, изображения которых могут быть подвержены значительному изменению форм и размеров с течением времени.

Задача обнаружения линий электропередач (ЛЭП) на изображении неразрывно связана с более общей задачей обнаружения прямолинейных границ на изображении. Анализ литературы показывает, что применение преобразований Радона (ПР) [1] и Хафа (ПХ) [2] является

*Рассмотрена задача обнаружения линий электропередач на изображении. Предложен алгоритм обработки изображения, решающий поставленную задачу в два этапа: обнаружение отдельных фрагментов линий электропередач с помощью преобразования Радона и объединение этих фрагментов в непрерывные кривые с помощью многоагентного подхода. Представлены результаты экспериментальных исследований разработанного алгоритма.*

одним из распространённых способов обнаружения прямолинейных границ. В то же время при решении задачи обнаружения ЛЭП на изображении эти преобразования используются практически повсеместно [3, 4].

Обычно применение ПР или ПХ подразумевает обработку исходного изображения каким-либо детектором границ. Такое изображение подвергается ПР или ПХ, результат которого содержит локальные максимумы, предположительно соответствующие какой-либо прямой на исходном изображении.

Однако применение в задаче обнаружения ЛЭП распространённых методов обнаружения прямолинейных границ с использованием ПР или ПХ мало эффективно по следующим причинам:

– каждый провод ЛЭП содержит две прямолинейные границы, каждая из которых обнаруживается обычными методами как самостоятельная граница;

– в городской местности, где чаще всего можно наблюдать ЛЭП, достаточно много прямолинейных границ от объектов искусственного происхождения (например, зданий), не имеющих к ЛЭП никакого отношения, вследствие чего возникает проблема отличия простых прямолинейных границ, от границ, принадлежащих ЛЭП;

– на фоне объектов естественного происхождения или просто объектов сложной неоднородной текстуры затруднительно детектировать прямолинейные границы обычными способами, использующими информацию о скорости перепада яркости, но практически игнорирующими направление этого перепада при применении детектора границ;

– в случае наблюдения ЛЭП с близкого расстояния

или на их высоте и при определённых ракурсах некоторые провода не будут прямолинейными, вместо этого они будут примерно соответствовать уравнениям цепной линии [5] с поправкой на растяжимость и неабсолютную гибкость проводов, а также эффект перспективы наблюдения.

В [6] предлагается отличать провода от обычных границ по признаку направления градиента. Показано, что использование информации о направлении градиента позволяет повысить точность обнаружения прямых в условиях шума или сложного неоднородного фона. Таким образом, использование векторного поля градиента для обнаружения ЛЭП на изображении представляется обоснованным. Основываясь на этом принципе, в [7] вводится так называемое интегральное векторное преобразование Радона (ИВПР), вычисляемое на основе обычного:

$$S(s, \alpha) = \cos \alpha \cdot \Phi \left[ \frac{\partial f(x, y)}{\partial x} \right] + \sin \alpha \cdot \Phi \left[ \frac{\partial f(x, y)}{\partial y} \right];$$

$$T(s, \alpha) = \int_{\rho=s_{\min}}^s S(\rho, \alpha) d\rho, \quad s \in [s_{\min}, s_{\max}]. \quad (1)$$

где  $S(s, \alpha)$  – результат ВПР;  $(s, \alpha)$  – параметры прямой;  $\Phi[f]$  – ПР;  $f(x, y)$  – исходное изображение.

С помощью ИВПР можно обнаружить прямые линии проводов, в том числе на неоднородном фоне, а также оценить их толщину в пикселях. В отличие от ПР данное преобразование может иметь как положительные, так и отрицательные значения. В то же время ИВПР обладает и привычным для ПР свойствами – его локальные экстремумы (минимумы и максимумы) косвенно могут указывать на наличие прямолинейной границы. Таким образом, ИВПР представляется наиболее подходящим преобразованием для использования при решении задачи обнаружения ЛЭП.

В соответствии с разработанным на основе ИВПР алгоритмом обнаружения ЛЭП на изображении (рис. 1) выполняются следующие действия:

1) обрабатываемое изображение разбивается на квадратные блоки фиксированного размера, при этом каждый соседний блок сдвинут относительно текущего на половину ширины блоков;

2) фрагменты изображения, принадлежащие каждому из блоков, подвергаются ИВПР, результаты которых подвергаются процедуре поиска локальных экстремумов;

3) полученные списки локальных экстремумов для каждого из блоков подвергаются отбраковке сначала в пределах одного блока (внутриблоковой фильтрации), а затем между блоками (межблоковой фильтрации).

Остановимся подробнее на каждом из пунктов. Исходными данными для ИВПР является векторное поле анализируемого изображения. Так как этому преобразованию подвергаются части изображения, то расчёт градиентного поля выполняется предварительно для всего изображения, а затем для каждого отдельного блока выделяется свой фрагмент этого поля, на основе которого рассчитывается ИВПР.

При обычном разбиении области изображения на квадратные блоки, при котором отсутствует наложение,

возможен случай, когда провод попадает на место стыка двух рядов блоков. В таком случае разные границы одного кабеля попадут на разные блоки, что не позволит распознать этот объект как ЛЭП. Также при слишком малой длине попавшего в блок фрагмента кабеля (по отношению к размерам этого блока) локальный экстремум окажется слабо выражен и может быть проигнорирован. Для компенсации подобных нежелательных эффектов блоковой обработки, в соответствии с предлагаемым алгоритмом, соседние блоки сдвинуты друг от друга не на размер блока, а на его половину. Таким образом, практически любая часть кабеля, так или иначе, попадет в достаточном для обнаружения объёме в какой-либо блок.

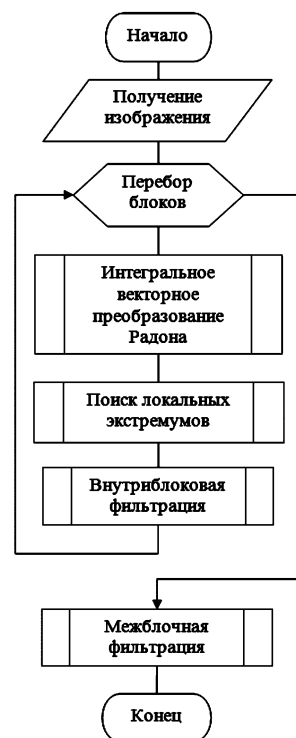


Рис. 1. Блок-схема алгоритма обнаружения линий электропередач

На следующем после разбиения шаге выполняется ИВПР от каждого из блоков. Изображение, содержащее результат этого преобразования, подвергается размытию для уменьшения числа ложных обнаружений. Результат последней операции подвергается поиску локальных экстремумов. Каждый найденный экстремум может соответствовать кабелю на исходном изображении, либо является результатом ложного обнаружения. Для снижения числа ложных тревог на следующем этапе применяется внутриблоковая фильтрация: строится гистограмма контрастности найденных локальных экстремумов, по которой эвристическим критерием выбирается порог фильтрации. В списке найденных локальных максимумов каждого отдельно взятого блока остаются лишь те из них, чья контрастность не ниже пороговой для этого блока.

Экспериментальные исследования показали, что применение к обнаруженным фрагментам проводов пороговой фильтрации по степени выраженности экстремума (контрасту между проводом и фоном, умноженному на длину фрагмента) не достаточно для достижения

приемлемой точности. Более детальный анализ выявил, что распределение плотностей вероятностей контрастности экстремумов для ложных и правильных обнаружений слабо отличимы. Из этого следует, что проблему полностью не решит даже оптимальная пороговая фильтрация.

Ложно обнаруженные сегменты проводов отличает произвольность их положения и ориентации, в то время как сегменты, принадлежащие проводам, с высокой вероятностью имеют своё продолжение в соседних блоках по обеим сторонам в виде других сегментов со схожей ориентацией и расположенных так, что вместе они образуют цепь сегментов с малой кривизной. Подобные цепи в дальнейшем будем называть группами. Для выделения подобных групп применяется **межблочная фильтрация**.

Так как перебор всех возможных соединений различных сегментов друг с другом в поиске наиболее длинной и гладкой цепи является неэффективным, то в данной работе для построения группы сегментов применяется многоагентный подход. В соответствии с одним из вариантов этого подхода [8] каждый сегмент должен быть представлен агентом потребности, а каждая допустимая связь между двумя сегментами – агентом возможности. Связь оценивается как допустимая, если:

- 1) соответствующие сегменты принадлежат разным блокам;
- 2) сдвиг одного блока относительно другого не превышает ширины блока по каждой из осей;
- 3) два сегмента имеют схожую ориентацию и с достаточной точностью могут быть аппроксимированы прямой.

Последний пункт выполняется при условии:

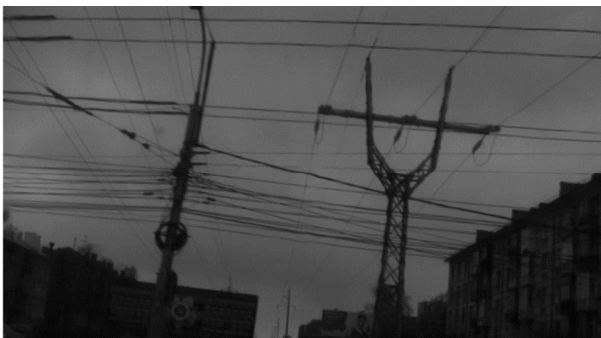
$$\begin{cases} \min(|\alpha_1 - \alpha_2|, |\alpha_1 + \alpha_2 - 180^\circ|) < \Delta_\alpha, \\ |\rho_2 - \rho_1 + \Delta x \cos(\alpha_1) + \Delta y \sin(\alpha_1)| < \Delta_\rho, \\ |\rho_2 - \rho_1 + \Delta x \cos(\alpha_2) + \Delta y \sin(\alpha_2)| < \Delta_\rho, \end{cases} \quad (2)$$

где  $\alpha_1$  и  $\alpha_2$  – углы поворота обоих сегментов;  $\Delta_\alpha$  – порог несовпадения углов  $\alpha_1$  и  $\alpha_2$ ;  $\rho_1$  и  $\rho_2$  – сдвиги обоих сегментов относительно центров их блоков;  $\Delta x$  и  $\Delta y$  – координаты сдвига второго блока относительно первого в пикселях;  $\Delta_\rho$  – порог несовпадения сдвига сегментов. Величины  $\Delta_\alpha$  и  $\Delta_\rho$  задаются заранее.

Если какой-то из сегментов не находит для себя ни одной допустимой связи на начальном этапе, то агент потребности для него не создаётся.

Основная идея многоагентного подхода с агентами потребности и возможности заключается в том, что каждый агент потребности ищет такой агент возможности, связь с которым обеспечила бы ему наиболее выгодное состояние. В данной работе предлагаются следующие правила по построению сети возможности-потребности (ПВ-сети):

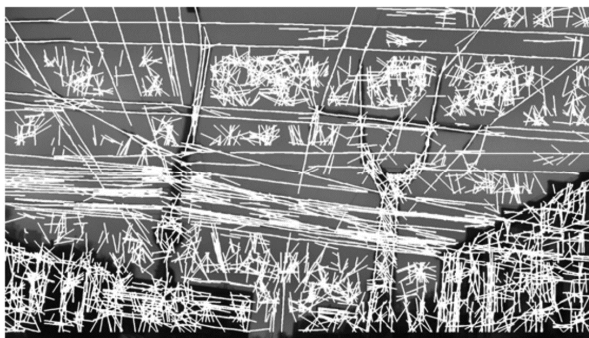
- 1) выгода для некоторой группы связанных в цепь сегментов определяется одинаково для каждого из её членов;
- 2) эта выгода определяется критерием  $k$ , который увеличивается с ростом размера группы  $n$  (число входящих в группу сегментов) и уменьшается с ростом  $\sigma$  – СКО расстояний концов отрезков до аппроксимирующей их кривой (таким образом, находится некий баланс между размером группы и правильностью её формы);



а)



б)



в)



г)

Рис. 2. Примеры результатов основных этапов работы алгоритма: а) – исходное изображение, б) – найденные с помощью ИВПР сегменты, в) – результат внутриблоковой фильтрации, г) – результат межблоковой фильтрации

3) каждый агент возможности обозначает возможную связь двух агентов потребностей в не зависимости от того, представляют ли они отдельные сегменты, или же объединяются две группы в ещё большую группу;

4) каждый агент потребности принимает решение об объединении или не объединении с другим на основе максимизации критерия (т.е. он выступает за объединение с таким агентом, в группе с которым он достигает максимального значения критерия);

5) если два агента потребности выступили за объединение друг с другом посредством голосования за соответствующий агент возможности, оба агента потребности и агент возможности исчезают, и появляется новый агент потребности, включающий в себя новую объединённую группу с соответствующим критерием;

6) объединение в новый агент потребностей необратимо и сопровождается исключением ряда агентов возможности, связанных с участниками объединения, т.к. эти возможности теперь нереализуемы;

7) если на каком-то этапе агент потребности оказывается без единой возможности, то он считается прошедшим фильтрацию, если размер его группы превышает некоторое пороговое значение (в противном случае он исключается).

Критерий  $k$  определяется по выражению:

$$k = \begin{cases} k_0, & \text{если } n = 1; \\ k_1 n - \frac{k_2 \sqrt{2}}{a} \sigma, & \text{иначе,} \end{cases} \quad (3)$$

где  $a$  – размер стороны блока в пикселях;  $k_0$ ,  $k_1$  и  $k_2$  – коэффициенты, подбираемые эмпирически.

Процесс выстраивания соединений является итеративным. В самом начале (на первой итерации) каждый агент потребности представляет один единственный сегмент, стремящийся стать частью группы с максимальным значением критерия. Вступающие во взаимодействие по указанным правилам агенты потребности и возможности в итоге выстраивают законченную ПВ-сеть, содержащую найденные цепи сегментов, принад-

лежащие проводам на исходном изображении.

Пример результатов выполнения каждого из этапов алгоритма приведён на рис. 2.

Для условий наблюдения с низким отношением сигнал/шум для повышения точности работы алгоритма после окончания обработки текущего кадра также применяется временная фильтрация, учитывающая результаты работы алгоритма на предыдущих кадрах.

### Экспериментальные исследования

Для оценки работоспособности алгоритма была использована видеопоследовательность формата Full HD (1920x1080 пикселей), содержащая 7774 кадра, на 4521 из которых ЛЭП присутствует, а на остальных 3253 – отсутствует. Оценка отношения сигнал/шум проводилась на основе наугад взятого типичного кадра из этой видеопоследовательности. На нём присутствовали как алгоритма ЛЭП, так и стволы деревьев, которые, являясь частью фона, могут оказаться основной причиной ложных обнаружений для исследуемого.

Для оценки отношения сигнал/шум в расчёт берутся модули градиентов изображений провода, ствола дерева и всего изображения. Фрагменты изображений, соответствующие первым двум пунктам этого ряда, отмечены на рис. 3. Оценка этих параметров составляет: для провода – 23,09 (3623 пикселя); для ствола дерева – 44,87 (1102 пикселя); для всего изображения – 11,84 (129600 пикселей). Таким образом, отношение средних модулей градиента для провода и ствола дерева составляет 0,5146, а для провода и всего изображения – 1,95, что в обоих случаях говорит о достаточно низком отношении сигнал/шум для данной видеопоследовательности.

По итогам проведённых экспериментальных исследований на данной видеопоследовательности (рис. 4) получены следующие результаты:

– частота правильного обнаружения составляет 89,96 %;

– частота ложного обнаружения составляет 0,09 %.



Рис. 3. Случайный кадр видеопоследовательности, взятый для оценки отношения сигнал/шум, с отмеченными на нём фрагментами, соответствующими изображению провода (чёрная рамка) и ствола дерева (белая рамка)





Рис. 4. Результат обработки одного из кадров алгоритмом обнаружения ЛЭП (тёмными линиями обозначены обнаруженные провода)

### Заключение

Разработанный алгоритм позволяет с высокой точностью обнаруживать ЛЭП на видеопоследовательности. Однако ценность данной работы представляется не столько её настоящими результатами, сколько перспективами дальнейшего развития, такими как: учёт модели более сложной формой провода, чем прямая (при вычислении критерия  $k$ ) введение возможности разрушения созданных групп агентов в случае, если одна из составных частей видит большую выгоду в присоединении к другой группе. Многоагентный подход, используемый в данной работе, может быть использован и в задачах распознавания более сложных образов. В этом случае геометрические примитивы должны быть способны собираться не только в кривые, но и в более сложные геометрические объекты, для чего необходимо усложнить правила взаимодействия агентов.

Исследования выполнены при поддержке гранта Президента РФ для государственной поддержки молодых российских учёных МК-5331.2016.9.

### Литература

1. Toft P.A., «The Radon Transform: Theory and Implementation, PhD Thesis», Technical University of Denmark, 1996.
2. Волегов Д.Б., Гусев В.В., Юрин Д.В. «Обнаружение прямых линий на изображениях на основе преобразования Хартли. Быстрое преобразование Хафа» // Труды

конференции Графикон 2006, 16-я международная конференция по компьютерной графике и ее приложениям, Россия, Новосибирск, Академгородок, 1-5 июля 2006. – С. 182-191.

3. Candamo J., Kasturi R., Goldgof D. and S. Sarkar, «Detection of Thin Lines Using Low Quality Video from Low Altitude Aircraft in Urban Settings», IEEE Transactions on aerospace and electronic systems, vol. 45, No. 3 July 2009.

4. Li, Zhengrong, Liu, Yuee, Walker, Rodney A., Hayward, Ross F., & Zhang, Jinglan (2009) Towards automatic power line detection for a UAV surveillance system using pulse coupled neural filter and an improved Hough transform. Machine Vision and Applications, 21(5), pp. 677-686.

5. Меркин Д.П. Введение в механику гибкой нити. – М.: Наука, 1980. – 240 с.

6. Sanders-Reed J.N., Yelton D.J., Witt C.C., Galetti R.R. «Passive Obstacle Detection System (PODS) for Wire Detection», Proc of the SPIE, 7328, April, 2009.

7. Алпатов Б.А., Бабаян П.В., Шубин Н.Ю. Преобразование Радона в задачах обнаружения линий электропередач // Техническое зрение в системах управления-2016: Сб. тезисов науч.-техн. конф. – М., 2016. – С. 64-65. [Электронный ресурс] <http://tvcs2016.technicalvision.ru/file/tezis.pdf> (дата обращения: 20.06.2016).

8. Виттих В.А., Скобелев П.О. Мультиагентные модели взаимодействия для построения сетей потребностей и возможностей в открытых системах // Автоматика и Телемеханика. – 2003. – №1. – С. 162-169.

## КОМПЛЕКС АЛГОРИТМОВ ВЫДЕЛЕНИЯ И ПРОСЛЕЖИВАНИЯ ДВИЖУЩИХСЯ ОБЪЕКТОВ ДЛЯ БОРТОВОЙ СИСТЕМЫ ТЕХНИЧЕСКОГО ЗРЕНИЯ

*Фельдман А.Б., к.т.н., доцент кафедры автоматики и информационных технологий в управлении (АИТУ) Рязанского государственного радиотехнического университета (РГРТУ), e-mail: aitu@rsreu.ru;  
Ерохин Д.Ю., магистрант, инженер кафедры АИТУ, e-mail: erokhin.d.y@gmail.com.*

### THE COMPLEX OF ALGORITHMS FOR DETECTING AND TRACKING MOVING OBJECTS FOR ONBOARD VISION SYSTEM

*Feldman A.B., Erokhin D.Y.*

*The article deals with the complex of algorithms to detection and tracking of moving objects observed from aerial platform. The complex consist in three main component: the algorithm of estimation of geometric transforms, the algorithm of objects extraction, the algorithm of objects tracking. The orientation of the complex for use in on-board systems allowed to achieve an acceptable level of computational complexity. Experiments confirm that the proposed complex has a good accuracy. The results can be using in development of on-board technical vision systems, including systems for small and unmanned aircrafts.*

**Key words:** object detection, tracking, geometric transformations, technical vision system, the Fourier transform.

**Ключевые слова:** выделение объектов, слежение, геометрические преобразования, система технического зрения, преобразование Фурье.

#### Введение

Анализ тенденций последнего времени показывает, что существенно возрос интерес к бортовым системам технического зрения. К таким системам можно отнести системы автоматического обнаружения и сопровождения объектов, системы определения местоположения летательного аппарата по снимкам местности, системы улучшенного и синтезированного видения.

В данной статье рассматривается проблема выделения и прослеживания движущихся объектов в видеопоследовательности, полученной с помощью бортового видеодатчика. Можно выделить типовые особенности рассматриваемой проблемы:

1. Малый контраст объектов интереса по отношению к фону.
2. Неоднородность оптической среды и различные атмосферные явления.
3. Пересечение траекторий прослеживаемых объектов.
4. Заслонение объектов участками фона.
5. Быстрое изменение ракурса сцены.

Необходимо отметить, что анализ изображений должен выполняться в реальном масштабе времени, но вместе с тем ресурсы бортовой аппаратуры, как правило, весьма ограничены. Еще одной проблемой, связанной с созданием бортовых систем технического зрения, является отсутствие априорной информации об объектах интереса, характере фонов и условиях окружающей среды. Ввиду большого разнообразия возможных объектов, фонов и условий наблюдения алгоритмы анализа изображений, предназначенные для бортовых систем технического зрения, должны обладать должной уни-

*Рассматривается комплекс алгоритмов для выделения и прослеживания движущихся объектов при наблюдении с борта летательного аппарата. Комплекс алгоритмов включает в себя: алгоритм оценки и компенсация геометрических преобразований изображений, алгоритм выделения движущихся объектов, алгоритм прослеживания объектов и прогнозирования их местоположения. Учет ориентации комплекса на применение в бортовых системах позволил добиться приемлемого уровня вычислительной сложности составляющих его алгоритмов. Экспериментальные исследования, проведенные с использованием натуральных видеосюжетов, подтверждают, что предлагаемый комплекс дает возможность решать задачи выделения и прослеживания движущихся объектов с удовлетворительной точностью. Результаты работы могут быть востребованы при создании бортовых систем технического зрения летательных аппаратов, в том числе относящихся к малой и беспилотной авиации.*

версальностью и обеспечивать работу при минимуме априорной информации.

Задача автоматического обнаружения объектов является одной из ключевых для большинства приложений технического зрения. На данный момент предложено множество методов ее решения. Между тем, с точки зрения применения в бортовых системах, существующие методы обладают как определенными достоинствами, так и некоторыми недостатками.

В работе [1] предложен метод автоматического выделения объектов с использованием гистограмм направленных градиентов [2]. Метод демонстрирует хорошие результаты, однако требует задания формы объекта и связан с большим объемом вычислений

Известны методы выделения объектов на основе информации об их цвете [3, 4]. Однако, устанавливать на борту цветной видеодатчик не всегда целесообразно в виду увеличения нагрузки на каналы передачи и аппаратуру обработки изображений.

Существуют подходы к выделению объектов с использованием машинного обучения [5]. Для применения подобных подходов необходимо сформировать обучающую выборку, что не всегда легко сделать в силу большого разнообразия объектов интереса и условий наблюдения.

Большое число работ посвящено развитию методов выделения объектов с помощью модели на основе смеси гауссовских распределений [6, 7]. Между тем, в большинстве публикаций рассматривается использование неподвижной видеокамеры. Существуют и работы, ориентированные на применение в условиях движения видеодатчика [8].

Несмотря на большое многообразие известных подходов к выделению объектов на видеоизображениях, можно заметить, что они либо не всегда обладают приемлемой вычислительной сложностью, либо слишком требовательны к наличию априорной информации об объектах интереса, либо не демонстрируют должного качества в условиях быстрого перемещения и поворотов видеодатчика.

В настоящей работе предложен комплекс алгоритмов, предназначенных для решения задач выделения и прослеживания движущихся объектов. Алгоритмы не требуют предварительного задания информации об объектах интереса и обладают приемлемой вычислительной сложностью. Комплекс включает следующие основные элементы: алгоритм оценки и компенсации геометрических преобразований изображений, алгоритм выделения движущихся объектов, алгоритм прослеживания объектов и прогнозирования их местоположения. Названные элементы комплекса рассматриваются ниже.

### Алгоритм оценки и компенсации геометрических преобразований

Наиболее распространенной практикой является установка видеодатчика на гиростабилизированной платформе, размещаемой на борту летательного аппарата. С помощью платформы датчик может вращаться относительно носителя. Кроме того, необходимо учитывать собственное движение носителя системы технического зрения. Таким образом, между соседними кадрами имеет место геометрическое рассогласование.

Оценка и компенсация геометрических преобразований поступающих изображений является одним из ключевых вопросов, связанным с решением задачи выделения и прослеживания объектов. Решение этого вопроса существенным образом зависит от выбора математической модели. Между тем, можно заметить, что в задачах выделения и прослеживания движущихся объектов, как правило, используются датчики с узким полем зрения. Также учтем, что расстояние от центра видеодатчика до наблюдаемой сцены существенно больше чем ее глубина и больше того расстояния, на которое за время формирования очередного кадра успевает переместиться летательный аппарат – носитель системы технического зрения. В этих условиях допустимым и оправданным с точки зрения экономии вычислительных ресурсов является использование евклидовой модели геометрических преобразований, которая предполагает

наличие составляющих поворота и сдвига по координатным осям. Иллюстрация процесса формирования изображений показана на рис. 1.

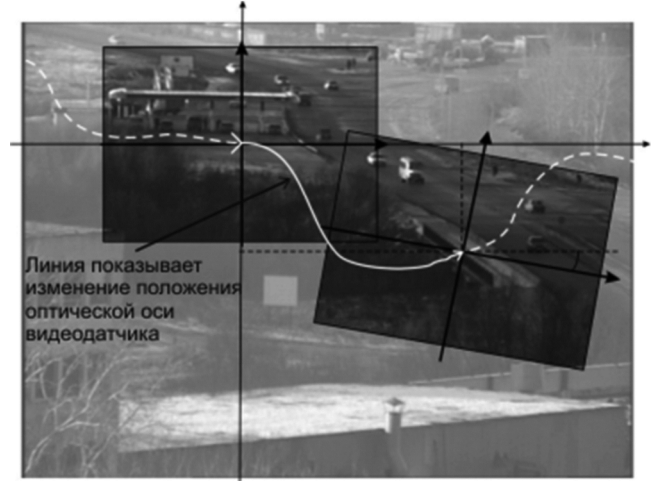


Рис. 1. Иллюстрация процесса формирования изображений при евклидовой модели геометрических преобразований

Для оценки сдвигов может быть использован хорошо зарекомендовавший себя на практике алгоритм фазовой корреляции [9]. Однако, данный алгоритм может применяться и для оценки поворотов, если перейти к представлению амплитудного спектра изображений в полярной системе координат. При повороте изображения происходит такой же поворот его амплитудного спектра [10], сдвиг в данном случае не оказывает влияния. В этом случае можно отдельно оценить поворот, который будет пропорционален смещению логарифма амплитудного спектра в полярной системе. Иллюстрация этого эффекта показана на рис. 2.

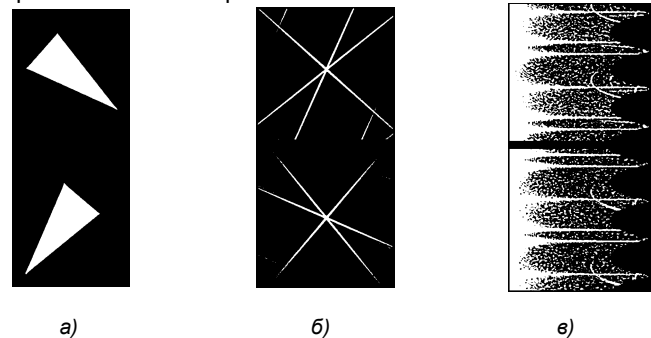


Рис. 2. Исходные изображения (а); амплитудные спектры изображений в декартовой системе координат (б); амплитудные спектры изображений в полярной системе координат (в)

Таким образом, компенсация евклидовых преобразований происходит в два этапа: на первом этапе выполняется оценка и компенсация поворота, на втором этапе – оценка и компенсация сдвига. Укрупненная блок-схема алгоритма представлена на рис. 3.

Отметим преимущества данного подхода. Алгоритм позволяет оценивать сдвиг и поворот с субпиксельной точностью. Если сравнивать данный алгоритм оценки с корреляционно экстремальными алгоритмами [11], осуществляемыми в пространственной области, то можно наблюдать существенное сокращение требуемого объема вычислений. Помимо корреляционных известны, например, подходы к оценке геометрических преобразо-

ваний на основе выбора и сопоставления особых точек [12, 13]. Однако на практике такие алгоритмы позволяют работать только с изображениями хорошего качества, чего редко удается добиться в условиях бортовой видеосъемки, когда наблюдение ведется с большого расстояния в присутствии вибраций носителя и прочих мешающих факторов.

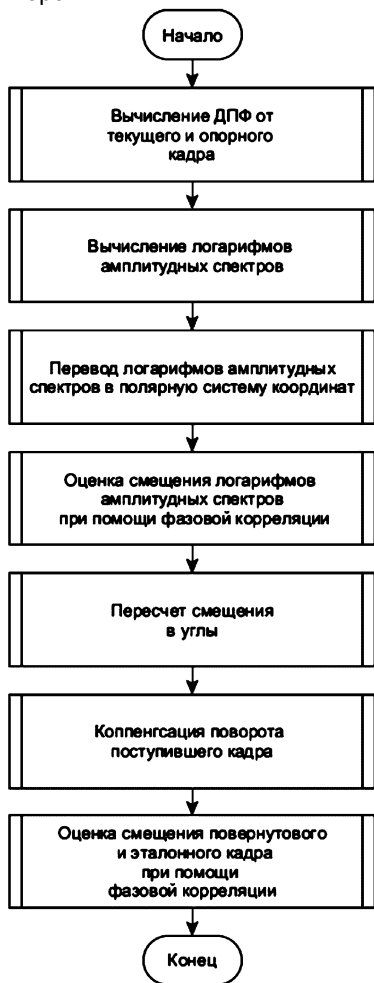


Рис. 3. Блок-схема алгоритма оценки геометрических преобразований

Рассмотренный алгоритм на основе фазовой корреляции позволяет оценивать вектор параметров геометрических преобразований с достаточно высокой точностью, что позволяет в дальнейшем получать оценку фона и решать задачу выделения движущихся объектов. Для дополнительного сокращения вычислительных затрат можно отказаться от обработки всего кадра целиком и ограничиться некоторой небольшой его областью, например, использовать окно размером 256 на 256 пикселей. Следует отметить, что дальнейшее снижение размеров окна не является целесообразным, так как приводит к заметному снижению точности оценки параметров преобразований. Таким образом, рекомендуется использовать квадратную область кадра с размером, кратным степени двойки, что связано с особенностями реализации быстрого преобразования Фурье.

#### Алгоритм выделения движущихся объектов

Задачей алгоритма выделения сегментов является классификация точек наблюдаемого изображения по их

принадлежности к одному из классов – к объектам интереса или фону. Результатом является бинарное изображение, где единичное значение яркости соответствует объекту, а нулевое – фону.

Для простоты рассмотрим случай с одним объектом. В качестве модели наблюдаемого изображения примем аппликативную модель [10]:

$$l(i, j, n) = r(i, j, n)h(i, j, n) + (1 - r(i, j, n))g(i, j, n) + \zeta(i, j, n) \quad (1)$$

где  $i, j$  – координаты пикселя,  $n$  – номер кадра,  $l(i, j, n)$  – наблюдаемое изображение,  $g(i, j, n)$  – изображение фона,  $h(i, j, n)$  – изображение объекта,  $r(i, j, n)$  – бинарное изображение, единичные значения которого соответствуют объекту,  $\zeta(i, j, n)$  – нормальный гауссовский шум с среднеквадратическим отклонением (СКО)  $\sigma(i, j, n)$ . Изображение объекта априори неизвестно. В этом случае будем считать, что яркости точек объекта имеют равномерное распределение. Таким образом, необходимо по наблюдаемому изображению оценить  $r(i, j, n)$ . Обозначим оценку  $r(i, j, n)$  символом  $\hat{r}(i, j, n)$ .

После установления геометрического соответствия между кадрами может быть выполнена оценка фонового изображения. Предположим, что к настоящему моменту получена достаточно хорошая оценка фона  $\hat{g}(i, j, n)$ . В соответствии с критерием Неймана-Пирсона должны выполняться следующие условия:

$$P(\hat{r}(i, j, n) = 1 / r(i, j, n) = 0) = p_- \quad (2)$$

$$P(\hat{r}(i, j, n) = 1 / r(i, j, n) = 1) = p_+ \rightarrow \max.$$

где  $p_-$  – вероятность принятия решения о присутствии объекта в точке  $(i, j)$ , где объект отсутствует;  $p_+$  – вероятность принятия решения о присутствии объекта в точке  $(i, j)$ , где он на самом деле присутствует. Так как решение о принадлежности точки объекту или фону принимается независимо для каждой точки, то целесообразно будет опустить в дальнейшем индексы  $(i, j, n)$ . Пусть яркость объекта равномерно распределена в диапазоне от  $c_{\min}$  до  $c_{\max}$ . Найдем отношение правдоподобия:

$$\Lambda(l) = \frac{p(l/r=1)}{p(l/r=0)} = \frac{1/(c_{\max} - c_{\min})}{p_{\xi}(l - \hat{g})} \quad (3)$$

где  $p_{\xi}(l - \hat{g})$  – плотность вероятности шума. Тогда оптимальную оценку бинарной маски объекта можно найти из выражения:

$$\hat{r} = \begin{cases} 1, & \text{при } \Lambda(l) > \Lambda_0 \\ 0, & \text{иначе,} \end{cases} \quad (4)$$

где  $\Lambda_0$  – пороговое отношение правдоподобия, которое определяется соотношением:

$$\int_{\Lambda(l) > \Lambda_0} p(l/r=0) dl = p_- \quad (5)$$

После некоторых упрощений правило выделения точек объекта может быть представлено в виде:

$$\hat{r}(i, j, n) = \begin{cases} 1, & \text{при } |l(i, j, n) - \hat{g}(i, j, n)| > k\hat{\sigma}(i, j, n), \\ 0, & \text{иначе,} \end{cases} \quad (6)$$

где  $k$  – пороговый коэффициент, который определяется допустимым уровнем ложных тревог,  $\hat{\sigma}(i, j, n)$  – оценка СКО шума. На практике пороговый коэффициент  $k$  может выбираться эмпирически из диапазона от 2.7 до 3.2.

Оценки фона и дисперсии шума в точках, не закрытых объектами (т.е. в точках, где  $\hat{r}(i, j, n) = 0$ ) формируются с помощью рекурсивных фильтров:

$$\hat{g}(i, j, n+1) = \beta \hat{g}(i, j, n) + (1-\beta)l(i, j, n), \quad (7)$$

$$\hat{\sigma}^2(i, j, n+1) = \beta \hat{\sigma}^2(i, j, n) + (1-\beta)(l(i, j, n) - \hat{g}(i, j, n))^2. \quad (8)$$

Здесь  $\beta$  – коэффициент, определяющий эффективную память экспоненциального фильтра ( $0 < \beta < 1$ ). Обычно выбирают значение  $\beta$  в диапазоне от 0,85 до 0,97. В точках, где  $\hat{r}(i, j, n) = 1$ , оценки фона и дисперсии не обновляются.

Приведенный алгоритм выделения объектов может быть улучшен. Как показано в работе [10], в связи с наличием геометрических преобразований сцены, наибольшие ошибки оценивания яркостей точек фона будут наблюдаться на границах контрастных объектов. В таком случае можно снизить влияние этих ошибок, если при выделении объектов учитывать интенсивность яркостных перепадов. С этой целью в зависимости от величины градиента  $|\nabla(i, j, n)|$  изменяется пороговое значение в правиле (6). При этом правило выделения точек объектов принимает вид:

$$\hat{r}(i, j, n) = \begin{cases} 1, & \text{при } |l(i, j, n) - \hat{g}(i, j, n)| > k (\hat{\sigma}(i, j, n) + |\nabla(i, j, n)|), \\ 0, & \text{иначе.} \end{cases} \quad (9)$$

Использование относительно простых моделей приводит к тому, что неизбежными являются ошибки выделения точек, принадлежащих объектам. Такие ошибки обычно проявляются в виде отдельных точек на бинарном изображении. Ошибки другого рода связаны с тем, что не удается выделить часть точек, принадлежащих объектам интереса. В этом случае один объект может быть представлен двумя и более сегментами на одном изображении. Данный эффект является нежелательным, так как приводит к идентификации отдельных сегментов в качестве разных объектов. С целью улучшения результатов выделения объектов бинарное изображение обрабатывается вначале медианным фильтром, а затем к нему применяются морфологические операции [14]. Операция морфологического замыкания используется для объединения близко расположенных сегментов, а также позволяет заполнить разрывы. Операция морфологического размыкания, напротив, позволяет избавиться от небольших сегментов, которые остались после медианной фильтрации.

После бинаризации необходимо получить список сегментов текущего кадра, для этого используется алгоритм разметки и параметризации, представленный в [10]. В результате получаем список сегментов, которые в дальнейшем необходимо сопоставить с прослеживаемыми объектами. Одновременно с разметкой осуществляется фильтрация, которая позволяет отбросить сегменты по размеру. Из дальнейшего рассмотрения удаляются те сегменты, которые по своему размеру заведомо не могут относиться к объектам интереса.

Пример, иллюстрирующий работу алгоритма оценки фона и выделения объектов, представлен на рис. 4.

Ограничением рассмотренного подхода является то,

что для решения задачи выделения объектов необходимо сформировать достоверную оценку яркостей точек фона. Таким образом, зона, в которой могут быть выделены объекты, ограничена и зависит от быстроты геометрических преобразований наблюдаемой сцены. Эксперименты показывают, что с использованием экспоненциального фильтра (7), (8) для получения достаточно достоверной оценки яркости точки фона необходимо учесть примерно 25-30 измерений.

### Алгоритм прослеживания объектов

Для слежения за подвижными объектами необходимо устанавливать соответствие между списком отслеживаемых объектов и списком сегментов, выделенных в текущем кадре. Для решения изложенной задачи в работе используется так называемый траекторный граф [15].

Для построения траекторного графа необходимо выбрать количественную меру соответствия между отслеживаемыми объектами и выделенными сегментами. В качестве такой меры будем использовать евклидово расстояние между центром выделенного сегмента и прогнозируемым центром объекта, вычисленным с помощью фильтра Калмана. В результате получаем полный двудольный граф, где вершины первого множества соответствуют сегментам, вершины второго множества соответствуют объектам, а в качестве весов ребер выступает расстояние между соответствующей парой вершин. Далее необходимо из полного графа получить траекторный граф. Назовем компоненту связности элементарной, если в ней присутствует не более трех вершин.

Процедура построения траекторного графа состоит из нескольких основных шагов:

1. Найти в графе связные компоненты.
2. Во всех неэлементарных компонентах удалить ребро с максимальным весом.
3. Если в графе остались неэлементарные компоненты, то перейти к п. 1, иначе траекторный граф построен.

В результате выполнения изложенных выше пунктов получаем траекторный граф, пример которого представлен на рис. 4.

В результате построения траекторного графа могут образоваться пять типов элементарных компонент (рис. 5):

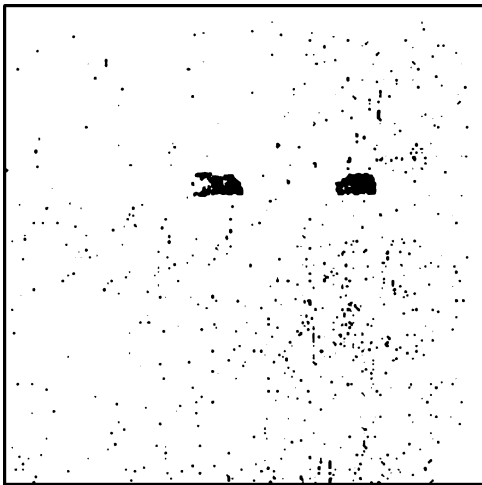
1. В компоненте связности одна вершина, которая соответствует объекту. В этом случае на текущем кадре данный объект не обнаружен (рис. 5, компонента G1).
2. В компоненте связности одна вершина, которая соответствует сегменту. В этом случае на текущем кадре обнаружен новый объект (рис. 5, компонента G2).
3. В компоненте две вершины первая соответствует объекту, а вторая соответствует сегменту. В этом случае на текущем кадре обнаружен прослеживаемый объект (рис. 5, компоненты G4, G6).
4. В компоненте три вершины, две из которых соответствуют объектам, а одна соответствует сегменту. В этом случае произошло объединение объектов (рис. 5, компонента G5).
5. В компоненте три вершины, две из которых соответствуют сегментам, а одна соответствует объекту. В этом случае произошло разделение объектов (рис. 5, компонента G3).



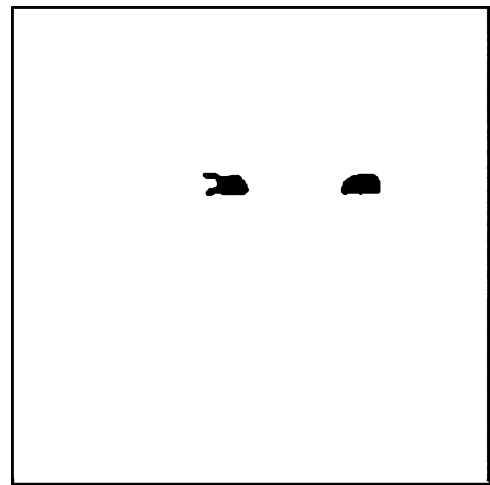
а)



б)



в)



г)

Рис. 4. Текущий кадр (а); оценка фона (б); результат выделения сегментов (в); результат морфологической обработки (г)

Если объект не был обнаружен в текущем кадре, его положение прогнозируется с помощью фильтра Калмана. При длительном пропадании объекта он удаляется из списка отслеживаемых.

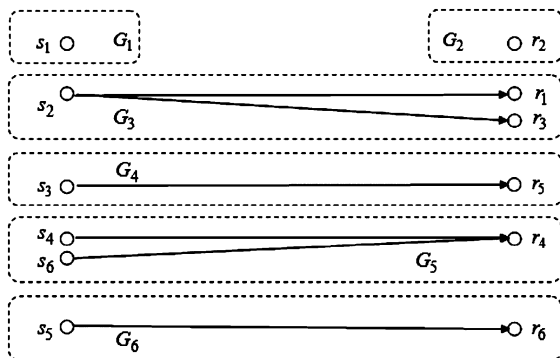


Рис. 5. Траекторный граф

### Экспериментальные исследования

В ходе экспериментальных исследований была выполнена оценка эффективности предложенного комплекса алгоритмов.

Для оценки точности алгоритма оценки и компенсации геометрических преобразований изображений были синтезированы две видеопоследовательности. На каж-

дом кадре присутствует случайный сдвиг в пределах от -5 до 5 пикселей, а также поворот на  $\pm 5$  градусов. В табл. 1 показано, как влияет размер окна, в котором осуществляется оценка параметров преобразований, на точность работы алгоритма. В результате эксперимента были вычислены СКО ошибок определения смещения по осям  $Ox - \sigma_x$  и  $Oy - \sigma_y$ , а также СКО ошибки определения поворота -  $\sigma_\varphi$ .

Таблица 1. Влияние размера окна на точность определения параметров геометрических преобразований

Размер окна	Параметр	Видеопоследовательность	
		№1	№2
128 на 128	$\sigma_x$ , пикс.	0,12	0,11
	$\sigma_y$ , пикс.	0,13	0,1
	$\sigma_\varphi$ , град.	0,07	0,06
256 на 256	$\sigma_x$ , пикс.	0,09	0,094
	$\sigma_y$ , пикс.	0,1	0,95
	$\sigma_\varphi$ , град.	0,06	0,05
400 на 400	$\sigma_x$ , пикс.	0,08	0,08
	$\sigma_y$ , пикс.	0,84	0,07
	$\sigma_\varphi$ , град.	0,06	0,06



Как видно из табл. 1, окна размером 256 на 256 достаточно для определения вектора евклидовых преобразований с хорошей точностью. Дальнейшее увеличение размеров окна почти не оказывает влияния на точность, однако помогает избежать ситуаций, когда в действительности происходит движение видеодатчика, но изображение остается неизменным, что может иметь место в случае регулярной структуры сцены.

Для оценки точности выделения подвижных объектов была экспериментальным путем построена операционная характеристика. Для этого использованы три натуральных видеосюжета с межкадровыми сдвигами и поворотами. Для каждого 25-го кадра этих сюжетов были вручную размечены бинарные маски, соответствующие наблюдаемым объектам. Среднее отношение сигнал/шум на тестовых видеосюжетах рассчитывалось по формуле:

$$SNR = \frac{1}{K} \sum_{k=1}^K \frac{|\bar{g}^k - \bar{h}^k|}{\sigma_{\xi}^k}, \quad (10)$$

где  $\bar{h}^k$  – средняя яркость точек  $k$ -го объекта,  $k = \overline{1, K}$ ,  $\bar{g}^k$  – средняя яркость фона в окрестности  $k$ -го объекта,  $\sigma_{\xi}^k$  – СКО шума, вычисленное для кадра, в котором наблюдался  $k$ -й объект. Экспериментальная операционная характеристика, полученная при среднем отношении сигнал/шум 3, представлена на рис. 6.

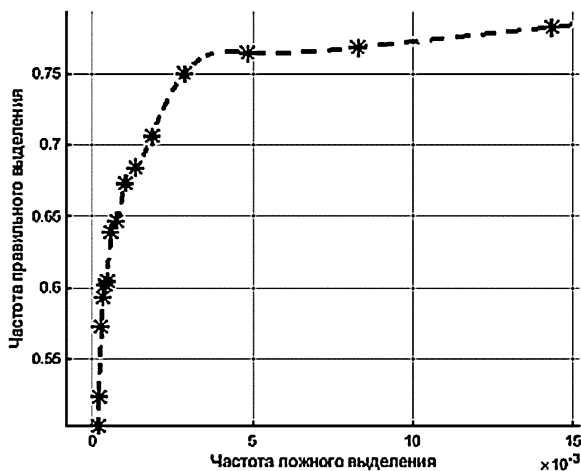


Рис. 6 Операционная характеристика алгоритма выделения точек объекта

Оценка точности определения координат подвижных объектов была проведена с использованием 3-х натуральных видеосюжетов длительностью 100 кадров каждый. Сюжеты получены в видимом диапазоне, содержат городской ландшафт и элементы пересеченной местности. На первом и втором сюжете присутствуют межкадровые сдвиги и поворот, на третьем видеосюжете камера является неподвижной. Для каждого кадра вручную были введены истинные положения объектов на изображении. Средний размер объектов составляет 30x20 пикселей. В ходе эксперимента были определены значения СКО ошибки оценивания координат объекта по оси  $Ox$  –  $\delta_x$  и по оси  $Oy$  –  $\delta_y$ . Результаты эксперимента приведены в табл. 2. Можно заметить, что на сюжете со статическим фоном результаты лишь немного лучше, чем на сюжетах, полученных движущейся камерой. Это

говорит о хорошей работе алгоритма компенсации геометрических преобразований.

Таблица 2. СКО ошибок определения координат объектов

СКО ошибки	Сюжет 1	Сюжет 2	Сюжет 3	Среднее
$\delta_x$	1,82	1,97	1,54	1,78
$\delta_y$	1,87	2,35	1,63	1,95

Для проверки работоспособности алгоритма прослеживания подвижных объектов была использована тестовая видеопоследовательность с сайта <http://www.yale.edu/perception/Brian/demos/MOT-Basics.html>.

На данной видеопоследовательности имеется множество объектов интереса, которые перемещаются по пересекающимся траекториям. Также были проведены испытания на реальных видеосюжетах. Испытания показали, что алгоритм прослеживания на основе траекторного графа успешно справляется со случаями заслонения объектов интереса. На рис. 7 представлена работа алгоритма прослеживания на реальном видеосюжете.



а)

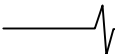


б)

Рис. 7. Объекты до (а) и после (б) пересечения траекторий

### Заключение

В рамках данной работы была рассмотрена задача обнаружения и прослеживания подвижных объектов при



сдвиге и повороте видеодатчика. Представленный комплекс алгоритмов был реализован на языке C++. Обработка кадров размером 640 на 480 пикселей при использовании ЭВМ на базе IntelCore 2 DuoT7100 осуществляется в реальном масштабе времени с частотой 25 кадров в секунду. Полученные результаты экспериментов свидетельствуют о возможном применении данного комплекса алгоритмов в составе бортовой системы технического зрения.

Работа выполнена на средства гранта Президента РФ для поддержки молодых российских ученых – кандидатов наук МК-3732.2015.10.

### Литература

1. Lu W.L., Little J.J. Simultaneous tracking and action recognition using the pca-hog descriptor // The 3rd Canadian Conference on Computer and Robot Vision (CRV'06). – IEEE, 2006. – С. 6.
2. Dalal N., Triggs B. Histograms of oriented gradients for human detection //2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). – IEEE, 2005. – Т. 1. – С. 886-893.
3. Lefevre S. et al. A new way to use hidden Markov models for object tracking in video sequences //Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on. – IEEE, 2003. – Т. 3. – С. III-117-20 vol. 2.
4. Gorry B. et al. Using mean-shift tracking algorithms for real-time tracking of moving images on an autonomous vehicle testbed platform //Proceedings of World Academy of Science, Engineering and Technology. – 2007. – Т. 25. – №. 11. – С. 1307-6884.
5. Stauffer C., Grimson W.E.L. Adaptive background mixture models for real-time tracking //Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on. – IEEE, 1999. – Т. 2.
6. Stauffer C., Grimson W.E.L. Adaptive background mixture models for real-time tracking //Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on. – IEEE, 1999. – Т. 2.
7. Bouwmans T., El Baf F., Vachon B. Background modeling using mixture of gaussians for foreground detection-a survey //Recent Patents on Computer Science. – 2008. – Т. 1. – №. 3. – С. 219-237.
8. Moo Yi K. et al. Detection of moving objects with non-stationary cameras in 5.8 ms: Bringing motion detection to your mobile device //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. – 2013. – С. 27-34.
9. Reddy B.S., Chatterji B.N. An FFT-based technique for translation, rotation, and scale-invariant image registration //IEEE transactions on image processing. – 1996. – Т. 5. – №. 8. – С. 1266-1271.
10. Алпатов Б.А., Бабаян П.В., Балашов О.Е., Степашкин А.И. Обработка изображений и управление в системах автоматического сопровождения объектов. – Рязань, 2011. – 235 с.
11. Баклицкий В.К., Бочкарев А.М., Мусьяков М.П. Методы фильтрации сигналов в корреляционно-экстремальных системах навигации //М.: Радио и связь. – 1986. – 216 с.
12. Lowe D.G. Object recognition from local scale-invariant features //Computer vision, 1999. The proceedings of the seventh IEEE international conference on. – IEEE, 1999. – Т. 2. – С. 1150-1157.
13. Bay H. et al. Speeded-up robust features (SURF) // Computer vision and image understanding. – 2008. – Т. 110. – №. 3. – С. 346-359.
14. Гонсалес Р., Вудс Р. Цифровая обработка изображений. – М.: Техносфера, 2006. – 1072 с.
15. Alpatov B., Babayan P. Multiple object tracking based on the partition of the bipartite graph //SPIE Security+ Defence. – International Society for Optics and Photonics, 2011. – С. 81860B-81860B-6.

УДК 004.932

## СРАВНЕНИЕ ТРЕХ КРИТЕРИЕВ РАБОТОСПОСОБНОСТИ МНОГОЭТАЛОННОГО АЛГОРИТМА ОЦЕНКИ КООРДИНАТ ОБЪЕКТОВ

*Корепанов С.Е., ведущий инженер каф. Автоматики и информационных технологий в управлении (АИТУ)*

*Рязанского государственного радиотехнического университета (РГРТУ), e-mail: aitu@rsreu.ru;*

*Смирнов С.А., к.т.н., доцент каф. АИТУ РГРТУ, e-mail: aitu@rsreu.ru;*

*Стротов В.В., к.т.н., доцент каф. АИТУ РГРТУ, e-mail: aitu@rsreu.ru.*

## THE COMPARISON OF THREE PERFORMANCE INDICATORS FOR THE OBJECT POSITION ESTIMATION ALGORITHM

*Korepanov S.E., Smirnov S.A., Strotov V.V.*

*In this work three performance indicators for the object coordinate estimation algorithm based on multiple template matching are considered. They include: criterion based on temporal reference area image variability, criterion based on criteria function shape analysis in the nearest neighborhood of its extremum and the criterion based on image gradient analysis. The results of the comparative experimental examinations are given. The natural video is used for the experimental examination.*

**Key words:** the object coordinate estimation algorithm, multiple template matching, SAD criterion, vision systems, performance indicators.

**Ключевые слова:** многоэталонный алгоритм оценки координат объектов, разностная критериальная функция, системы обработки изображений, критерии работоспособности.

### Введение

В настоящее время системы обработки изображений получили широкое распространение. Они успешно используются при решении задач наземной и воздушной навигации, при разработке мобильных роботов, в охранных системах видеонаблюдения, в сфере специальных и военных технологий.

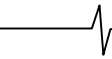
Системы обработки изображений могут быть установлены как на стационарную платформу, так и на борт мобильного носителя. В таких системах могут использоваться датчики, регистрирующие изображения в различных спектральных диапазонах. Наибольшее распространение получили видеодатчики, формирующие изображения в видимом и инфракрасном диапазонах.

Основным классом задач, решаемых в таких системах, является обнаружение и измерение координат объектов интереса. Зачастую, при работе системы обработки изображений необходимо оценить координаты объектов от момента, когда размеры объекта близки к точечным и до момента, когда объект заслоняет поле зрения видеодатчика. Подобное изменение размеров объекта проявляется в случае приближения или удаления друг от друга объекта наблюдения и видеодатчика. Для получения координат в таких условиях необходимо использовать алгоритм, ориентированный на адаптацию к изменению размеров объекта. К сожалению, на сегодняшний момент среди существующих алгоритмов нет такого, который обладал бы удовлетворительным качеством в таких условиях наблюдения. Добиться приемлемого качества оценки координат объекта можно за счет использования комплексного подхода, заключаю-

*Рассмотрены три критерия работоспособности многоэталонного алгоритма оценки координат объектов: критерий на основе анализа временной изменчивости опорного участка, критерий на основе анализа поведения разностной критериальной функции в окрестности минимума и критерий на основе анализа градиентного изображения. Представлены результаты сравнительных экспериментальных исследований, выполненные с использованием натуральных видеосюжетов.*

щегося в переключении между алгоритмами, каждый из которых будет ориентирован на определенный этап слежения за объектом. В работе [1] осуществляется переключение между алгоритмом на основе пространственно-временной фильтрации, ориентированным на оценку координат объектов малых и средних размеров, и многоэталонным алгоритмом, предназначенным для оценки координат объектов больших размеров. В работе предлагается осуществлять переключение между алгоритмами, исходя из текущих размеров наблюдаемого объекта. Однако, при использовании такого принципа переключения возможны ситуации, когда многоэталонный алгоритм на начальных кадрах будет выдавать координаты объекта с повышенной ошибкой. Этот факт объясняется тем, что на наблюдаемом изображении нет необходимого количества опорных участков с достаточным контрастом. В качестве альтернативного принципа переключения между алгоритмами можно рассматривать переключение на основании критериев работоспособности [2]. В таком случае, эффективность комплексного подхода в значительной степени будет зависеть от правильности выбора критерия работоспособности.

В данной работе проведено сравнение трёх критериев работоспособности: критерия на основе анализа временной изменчивости опорного участка, критерия на основе анализа поведения разностной критериальной функции в окрестности минимума и критерия на



основе анализа градиентного изображения. Сравнение будет производиться с точки зрения ошибок принятия решений и сложности вычислительной реализации. Перед тем как привести подробное описание критериев работоспособности, приведем описание многоэталонного алгоритма оценки координат.

### Многоэталонный алгоритм оценки координат объектов

Для измерения координат крупноразмерных объектов в работе [1] используется многоэталонный алгоритм. В основе алгоритма лежит принцип, построенный на выборе нескольких опорных участков на изображении и определении их положения от кадра к кадру. Для слежения за опорными участками используется алгоритм оценки положения на основе сопоставления с эталоном. Будем называть его разностным алгоритмом, т.к. критериальная функция представляет собой модуль разности эталонного и наблюдаемого изображений [3]. Многоэталонный алгоритм состоит из следующих этапов.

1. В начальный момент времени на изображении объекта выбирается заданное число участков  $s_k$ , называемых опорными. Количество опорных участков и их размер определяются исходя из возможностей аппаратной платформы, на которой реализуется рассматриваемый алгоритм.

2. По положению опорных участков на наблюдаемом изображении оценивается местоположение объекта. Координаты центров опорных участков определяются с субпиксельной точностью разностным алгоритмом.

3. На основании полученных оценок местоположения опорных участков находятся оценки параметров геометрических преобразований сдвига, поворота и изменения масштаба наблюдаемого изображения объекта.

4. Определяются новые координаты центра изображения наблюдаемого объекта и его размеры с учетом полученных параметров геометрических преобразований.

5. Непригодные к дальнейшему использованию опорные участки заменяются на новые участки.

Особенность данного алгоритма заключается в том, что опорные участки могут быть выбраны только на изображении объекта. Поэтому для выбора участков используется априорная информация об объекте и фоне изображения. Если объект наблюдается на слабоконтрастном или однородном фоне, то выбор опорных участков может производиться по центральной зоне наблюдаемого изображения по критерию:

$$(\hat{\alpha}_k, \hat{\beta}_k) = \operatorname{argmin}_{\alpha, \beta} \left( \frac{1}{\sum_{(i,j) \in S_k} (\Delta_i I(i, j, n))^2} + \frac{1}{\sum_{(i,j) \in S_k} (\Delta_j I(i, j, n))^2} \right), \quad (1)$$

где запись вида  $\Delta_i I$  означает операцию численного дифференцирования изображения  $I$  по координате  $i$ , а  $S_k$  – множество точек  $k$ -го участка  $S_k$ , расположенного в точке с координатами  $(\alpha, \beta)$  [4].

### Критерии работоспособности

Перейдем к разработке критериев работоспособно-

сти многоэталонного алгоритма. Поскольку для корректной работы данного алгоритма требуется определенное минимальное количество опорных участков, изменения положений которых на кадрах видеопоследовательности вычисляются разностным алгоритмом, то критерий работоспособности многоэталонного алгоритма сводится к проверке критериев работоспособности разностного алгоритма на выбранных опорных участках. Если количество опорных участков, на которых разностный алгоритм работоспособен, не меньше минимального количества опорных участков, достаточных для работы многоэталонного алгоритма, то многоэталонный алгоритм работоспособен, в противном случае – нет.

Критерии работоспособности должны служить индикатором качества оценки координат и при этом не опираться на эталонные данные о местоположении опорных участков. Сущность критерия работоспособности должна быть связана с особенностями многоэталонного алгоритма оценки координат объектов, а именно в способе выбора опорных участков и в сопоставлении опорных участков на основе анализа разностной критериальной функции. Для принятия решения о работоспособности необходимо сравнить полученное значение критерия с пороговым значением, полученным экспериментальным путем по методике подробно описанной в [5].

**Критерий работоспособности №1.** Данный критерий основывается на анализе временной изменчивости опорного участка. Пусть имеется дискретное эталонное изображение опорного участка  $H$ , представляющее собой множество связанных элементов  $h(i, j)$ . Данное эталонное изображение является частью наблюдаемого изображения  $l(i, j)$ ,  $(i, j) \in L$ . Поиск эталонного изображения  $H$  на наблюдаемом в  $n$ -м кадре изображении  $L$  заключается в минимизации разностной критериальной функции вида:

$$F(\alpha, \beta, n) = \sum_{(i,j) \in H} |l(i + \alpha, j + \beta, n) - h(i, j)|, \quad (2)$$

где  $l(i, j, n)$  – наблюдаемое в  $n$ -м кадре изображение,  $\alpha, \beta$  – параметры, характеризующие смещение объекта вдоль осей координат.

При наблюдении за опорным участком его изображение будет изменяться от кадра к кадру. Изменчивость опорного участка во времени можно охарактеризовать минимальным значением критериальной функции, получаемой в очередном кадре. Если показатель временной изменчивости опорного участка примерно соответствует показателю отличия опорного участка и окружающего его фона, то измерение координат вряд ли возможно. В то же время, если измерение координат возможно, т.е. опорный участок и фон не похожи друг на друга, то показатель отличия опорного участка от фона будет больше, чем показатель временной изменчивости опорного участка [2].

Показатель отличия изображения опорного участка от окружающего фона можно охарактеризовать величиной  $\min_{(\alpha, \beta) \in \Gamma} F(\alpha, \beta, n)$ , где  $\Gamma$  – множество координат, соответствующих границе зоны поиска. Показатель отличия эталонного изображения на кадре  $n$  от изображения

опорного участка на кадре  $n+1$  описывается величиной  $\min_{(\alpha, \beta) \in W} F(\alpha, \beta, n+1)$ , т.е. минимальным значением критериальной функции в  $(n+1)$ -м кадре, которое характеризует межкадровую изменчивость изображения опорного участка. Таким образом, если  $\min_{(\alpha, \beta) \in \Gamma} F(\alpha, \beta, n)$  больше

$\min_{(\alpha, \beta) \in W} F(\alpha, \beta, n+1)$ , то можно утверждать, что измерение координат возможно. Следовательно, отношение этих величин можно использовать как признак, оценивающий возможность измерения координат опорного участка в  $(n+1)$ -м кадре:

$$KF_1(n+1) = \frac{\min_{(\alpha, \beta) \in \Gamma} F(\alpha, \beta, n)}{\min_{(\alpha, \beta) \in W} F(\alpha, \beta, n+1)}. \quad (3)$$

**Критерий работоспособности № 2.** Второй критерий основан на анализе поведения критериальной функции вида (2) в окрестности минимума. Минимум критериальной функции  $F$  определяется выражением:

$$(\alpha^*, \beta^*) = \arg \min_{(\alpha, \beta) \in L} \sum_{(i, j) \in H} |l(i + \alpha, j + \beta, n) - h(i, j, n)|. \quad (4)$$

Для расчета критерия работоспособности определяется среднее значение в окрестности точки  $(\alpha^*, \beta^*)$  размером  $(2k+1) \times (2k+1)$  точек:

$$M = \frac{\sum_{\alpha=\alpha^*-k}^{\alpha^*+k} \sum_{\beta=\beta^*-k}^{\beta^*+k} F(\alpha, \beta, n) - F(\alpha^*, \beta^*, n)}{(2 \cdot k + 1)^2 - 1}. \quad (5)$$

Затем определяется число точек лежащих ниже уровня  $T$ , вычисляемого по формуле:

$$T = z \cdot M, \quad (6)$$

где  $z$  – коэффициент, определяемый экспериментальным путем (на практике используется  $z = 1,4$  и  $k = 3$ ).

В том случае, если число точек критериальной функции, лежащих ниже уровня  $T$ , меньше порогового значения, можно говорить об уверенном измерении координат опорного участка.

**Критерий работоспособности № 3.** Критерий основан на анализе градиентного изображения. По данному изображению можно судить о контрасте опорного участка. Критерий работоспособности будем рассматривать как отношение оценки сигнал-шум.

Поскольку выбор положения опорных участков производится в областях с максимальным контрастом по яркости в соответствии с формулой (1), то в качестве сигнала можно рассматривать величину:

$$L_Z = \sqrt{\left( \frac{1}{\sum_{(x, y) \in Z} (Z_{\Delta_x, \phi}(x, y, n))^2} + \frac{1}{\sum_{(x, y) \in Z} (Z_{\Delta_y, \phi}(x, y, n))^2} \right)^{-1}}, \quad (7)$$

где  $Z_{\Delta_x, \phi}(x, y, n)$ ,  $Z_{\Delta_y, \phi}(x, y, n)$  – градиентные изображения зоны поиска, сглаженные усредняющим фильтром. Размер маски фильтра соответствует размеру эталона.

В качестве характеристики, определяющей шум, будем рассматривать его среднеквадратическое отклонение (СКО)  $\sigma_\xi$ . Подход к оценке СКО шума приведен в работе [6]. Использование данной величины позволяет нормировать оценку отношения сигнал-шум. Таким образом, критерий работоспособности можно определить следующим образом:

$$KF_3 = \frac{L_Z}{\sigma_\xi} \quad (8)$$

**Экспериментальные исследования**

Сравнительные исследования рассматриваемых критериев работоспособности проводились с использованием натуральных сюжетов, снятых в видимом диапазоне. Длительность сюжетов варьировалась от 1000 до 1400 кадров, средняя длительность – 1270 кадров. На видеосюжетах объекты были неподвижны. В процессе съемки диафрагма камеры постепенно закрывалась и открывалась вновь, что обеспечивало изменение освещенности сцены. На рис. 1 приведены кадры одного из сюжетов при различной степени открытости диафрагмы.

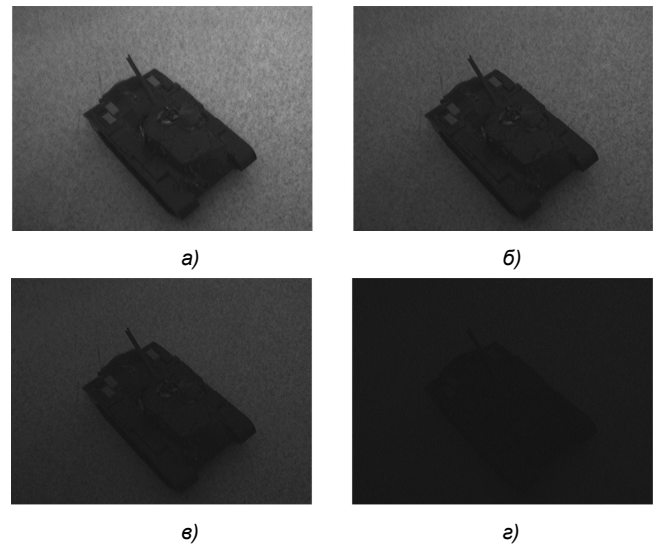
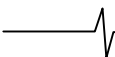


Рис. 1. Примеры кадров из тестового сюжета при различных значениях диафрагмы видеодатчика

Для каждого сюжета выбиралось 8 опорных участков размером  $32 \times 32$  пикселя. Опорные участки выбирались таким образом, чтобы они включали в себя изображения частей объекта, изображения фона, а также участки, содержащие одновременно и фон и части объекта. Положение опорных участков, указанное человеком-экспертом на первом кадре, принималось за эталонное (истинное) положение. На рис. 2 представлен кадр из тестового сюжета, на котором квадратами показаны выбранные опорные участки.

Для каждого видеосюжета местоположение опорных участков в кадре определялось разностным алгоритмом. При возникновении больших ошибок в оценке положения осуществлялся перезапуск алгоритма. При этом по истинным координатам формировалось новое целеуказание, на основе которого выполнялось обновление эталонного изображения. На основании сравнения измеренных и эталонных координат принималось решение о пригодности разностного алгоритма для оценки коор-



динат конкретного опорного участка на каждом кадре видеопоследовательности. Параллельно с этим на каждом кадре для опорных участков вычислялись значения трёх критериев работоспособности. Таким образом, для каждого критерия работоспособности были сформированы два массива значений, соответствующих кадрам на которых разностный алгоритм был «работоспособен» и «не работоспособен».

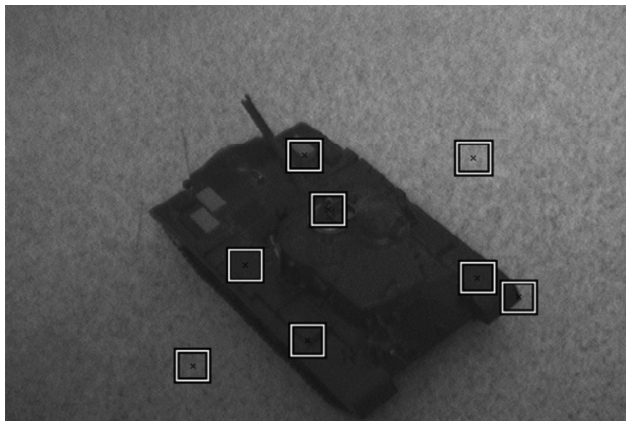


Рис. 2. Опорные участки

Согласно методике, приведенной в [5], для оценки работоспособности разностного алгоритма осуществляется переход от значений критерия работоспособности к показателю качества, принимающим значения «работает»/«не работает», путём сравнения критериев работоспособности с пороговыми значениями. Процедура нахождения пороговых значений по полученным массивам критериев работоспособности подробно описана там же. При определении пороговых значений возможны ошибки двух видов:

- ошибка первого рода – ошибочное принятие решения, что алгоритм неработоспособен,
- ошибка второго рода – ошибочное принятие решения, что алгоритм работоспособен.

Результаты исследований критериев приведены в табл. 1.

Таблица 1. Пороговые значения признаков и ошибки классификации

Критерий работоспособности	Пороговое значение	Ошибка первого рода, %	Ошибка второго рода, %
№1	1,3	0	8
№2	152	5	47
№3	0,9	0	4

Вычислительная сложность расчета критерия работоспособности № 1 и критерия работоспособности № 2 в большей степени определяется затратами на нахождение минимума критериальной функции вида (2). Если пренебречь несколькими операциями вычитания и деления, то из формул (3) и (5) следует, что вычислительная сложность двух критериев отличается затратами на расчет показателя отличия изображения опорного участка от окружающего фона  $\min_{(\alpha, \beta) \in \Gamma} F(\alpha, \beta, n-1)$ , кото-

рая рассчитывается при работе разностного алгоритма на  $n-1$  кадре. Таким образом, можно придти к выводу,

что затраты на вычисление признаков примерно одинаковы. При использовании этих двух критериев для оценки работоспособности многоэталонного алгоритма вычислять минимум критериальной функции вида (2) необходимо для каждого опорного участка. Следовательно, вычислительная сложность критерия работоспособности многоэталонного алгоритма будет расти с увеличением количества опорных участков.

Вычислительные затраты критерия работоспособности №3 главным образом определяются расчетом величины  $L_z$  в формуле (7). Учитывая, что данная величина вычисляется в многоэталонном алгоритме при выборе опорных участков по формуле (1), то вычислительной сложностью данного критерия можно пренебречь.

## Заключение

Результаты сравнительного исследования эффективности критериев с точки зрения ошибок принятия решений позволяют сделать выводы, что использование критерия работоспособности на основе анализа градиентного изображения (критерия № 3) целесообразно для анализа работоспособности многоэталонного алгоритма оценки координат объектов. Невысокая вычислительная сложность позволяет использовать данный критерий для повышения эффективности перехода на многоэталонный алгоритм для оценки координат объектов больших размеров.

Исследования выполнены при использовании Стипендии Президента РФ СП-2598.2016.5.

## Литература

1. Стротов В.В., Корепанов С.Е. Слежение за объектом, размеры которого значительно изменяются со временем // Вестник Рязанского государственного радиотехнического университета. – Рязань, 2012. – № 39. – С. 9-14.
2. Бабаян П.В., Смирнов С.А. Слежение за объектом на основе алгоритма сопоставления с эталоном при одновременном наблюдении в видимом и инфракрасном диапазонах // Цифровая обработка сигналов.– №4. – 2010. – С.18-21.
3. Алпатов Б.А., Селяев А.А., Степашкин А.И. Цифровая обработка изображений в задаче отслеживания движущегося объекта // Изв. вузов. Сер. Приборостроение. – 1985. – №2. – С. 39-43.
4. Стротов В.В. Выбор опорных участков в многоэталонном алгоритме определения параметров геометрических преобразований изображений // Вестник Рязанского государственного радиотехнического университета. – Рязань, 2009. – №2(28). – С.93–96.
5. Бабаян П.В., Смирнов С.А. Оценивание ожидаемой работоспособности алгоритмов измерения координат объектов // ГрафиКон' 2010: 20-я Международная конференция по компьютерной графике и зрению. – Санкт-Петербург. – 2010. – С. 329-330.
6. Алпатов Б.А., Бабаян П.В., Балашов О.Е., Степашкин А.И. Методы автоматического обнаружения и сопровождения объектов. Обработка изображений и управление. – М.: Радиотехника, 2008. – 176 с.



УДК 004.932.4

## МАТИРОВАНИЕ ВИДЕОПОСЛЕДОВАТЕЛЬНОСТЕЙ С ИСПОЛЬЗОВАНИЕМ ВОССТАНОВЛЕННОГО ФОНА

*Ерофеев М.В., аспирант Московского государственного университета им. М.В. Ломоносова, факультет вычислительной математики и кибернетики, e-mail: merofeev@graphics.cs.msu.ru; Ватолин Д.С., к.ф.-м.н., с.н.с. Московского государственного университета им. М.В. Ломоносова, e-mail: dmitriy@graphics.cs.msu.ru.*

### VIDEO MATTING WITH AID OF RECONSTRUCTED BACKGROUND

*Erofeev M.V., Vatolin D.S.*

*Formally, matting is a problem of image decomposition into foreground image, background image and foreground transparency map. This problem is extremely important for such video and image editing problems as: background replacement, applying transform to background or foreground only, stereoscopic image generation. In this paper we propose video matting method based on Learning Based Matting method. We describe modification to base method which enabled us to use reconstructed background sequence as additional input. We also propose iterative method for spatio-temporal transparency map smoothing. Finally, we show that proposed approach outperforms 11 image and video matting methods in objective comparison.*

**Key words:** matting, video processing, background reconstruction, background inpainting, matting Laplacian.

**Ключевые слова:** матирование, обработка видео, восстановление фона, матирующий Лапласиан.

#### Введение

Одной из ключевых задач, возникающих при редактировании изображений и видеопоследовательностей, является построение карты прозрачности (матирования) объекта переднего плана для последующей замены фона или элементов фона, или изменения положения объекта относительно фона.

Формальная постановка задачи матирования имеет следующий вид: разделить данное изображение  $I$  на изображение объекта переднего плана  $F$ , изображение заднего плана  $B$  и карту прозрачности переднего плана  $\alpha$  таким образом, чтобы было верно следующее уравнение:

$$I = \alpha F + (1 - \alpha) B \quad (1)$$

Задача матирования видео является обобщением задачи матирования изображений на видеопоследовательности. Такое обобщение порождает дополнительные требования к алгоритмам матирования видео:

Результирующая карта прозрачности должна быть стабильна во времени, ввиду высокой чувствительности зрительной системы человека к различиям между соседними кадрами.

Алгоритм должен иметь адекватную объемам видеоданных вычислительную сложность.

В данной работе предлагается алгоритм матирования видеопоследовательностей, в основу которого положен алгоритм матирования изображений на основе обучения (Learning Based Matting [1]). Отличительной особенностью предложенного алгоритма является использование изображений фона, содержащих фрагменты, скрытые объектом переднего плана на исходной

*Задача матирования состоит в разделении данного изображения или видео на передний план, задний план и карту прозрачности переднего плана. Построение такого разделения важно для решения многих задач редактирования и обработки видео, таких как: замена фона, применение спецэффекта только к фону либо только к переднему плану, построение стереоизображения. Предлагается метод матирования видеопоследовательностей, основанный на методе матирования изображений Learning Based Matting. Описывается способ модификации базового метода, позволивший использовать информацию о фрагментах фона, скрытых объектом переднего плана, в качестве дополнительных данных; а также способ итеративной фильтрации последовательности карт прозрачности. Приводятся результаты сравнения предложенного алгоритма с 11 аналогами, иллюстрирующие преимущества предложенного подхода.*

видеопоследовательности, в качестве дополнительных входных данных. Для построения таких изображений фона применяется алгоритм, описанный в [2], выполняющий восстановление фрагментов фона, скрытых на данном кадре, путем копирования этих же фрагментов из других кадров видеопоследовательности, где они являлись видимыми. Предлагаемый подход также выполняют два последовательных прохода по видеопоследовательности с целью обеспечения большей стабильности во времени результата в сравнении с результатом независимого применения алгоритма [1] к кадрам видеопоследовательности. Подробное описание алгоритма содержится в разделе «Предлагаемый алгоритм».

Предложенный алгоритм был протестирован на наборе из 10 видеопоследовательностей с известными эталонными картами прозрачности (данный набор был подготовлен в рамках работы [3] и сравнен с 11 другими методами матирования изображений и видео). Описание проведенной экспериментальной оценки содержится далее в разделе: «Экспериментальная оценка».

---

## Обзор области исследований

Исследования в области матирования видео, в сравнении с областью матирования изображений, затруднены большим объемом входных данных, требуемых для работы алгоритмов матирования видео, и необходимостью обеспечения стабильности во времени результирующей карты прозрачности. Ввиду этого исследования в области матирования видео всегда отставали от прогресса в области матирования изображений. Несмотря на это, на сегодняшний день существуют 12 оригинальных методов матирования видео, а также объективная методика сравнения алгоритмов матирования видео, описанная в [3].

Большинство методов матирования видео основаны на одном из существующих методов матирования изображений. При этом основное внимание их авторы уделяли повышению временной стабильности результирующей последовательности карт прозрачности и уменьшению объема дополнительных данных, требуемых от пользователя.

Наиболее широко используемым в литературе способом сокращения объема дополнительных данных является распространение тернарных масок (карт, задающих три области: гарантированный передний план, гарантированный задний план и неизвестная область, для которой требуется решить задачу матирования) ключевых кадров на все кадры видеопоследовательности. Данный метод позволяет пользователю предоставить тернарные маски лишь для части кадров, затем алгоритм вычисляет вектора оптического потока между кадрами исходной видеопоследовательности и использует их для переноса масок на остальные кадры. Такой метод применяется в алгоритмах матирования видео [4, 5, 6, 7]. Схожим образом задачу сокращения объема ввода пользователя решает алгоритм [8], требуя в качестве дополнительных входных данных лишь карту прозрачности первого плана. Следует также отметить алгоритм [9], который, аналогично алгоритму сегментации изображений Grab Cut [10], лежащему в его основе, не требует никаких дополнительных данных от пользователя. Однако получаемые с его помощью карты прозрачности имеют крайне низкое качество.

Для обеспечения стабильности во времени результирующей последовательности карт прозрачности существует множество разнообразных подходов. Первый из них был описан в работе [5], ее авторы внедрили в оптимизируемую алгоритмом функцию стоимости слабое, моделирующее совместное распределение градиента исходной видеопоследовательности и последовательности карт прозрачности.

Большая группа существующих на сегодня алгоритмов матирования выполняет поиск итоговой карты прозрачности как минимума квадратичной функции стоимости, аналогичной описанной в [11]. Работы [12, 13, 14] обеспечивают стабильность результата во времени, внедряя в эту функцию слабое, требующее постоянства карты прозрачности соседних кадров вдоль векторов движения. В работах [6, 8] пространственный матирующий лапласиан с окнами  $3 \times 3$ , описанный в [11], за-

менен пространственно-временным матирующим лапласианом с окнами  $3 \times 3 \times 3$ . Эта идея была развита в работе [15], заменившей матирующий лапласиан с окнами  $3 \times 3 \times 3$  нелокальным матирующим лапласианом – лапласианом, задающим связи между пикселями, не обязательно находящимися по соседству. В работе [16] пространственно-временной лапласиан был заменен быстрым методом фильтрации Guided Filter [17], позволившим уменьшить время обработки одного кадра.

Другой подход к обеспечению стабильности карт прозрачности во времени состоит в создании моделей переднего плана и фона для групп кадров или всего видео вместо моделей, построенных для каждого кадра независимо (как в алгоритмах матирования изображений). Примерами этого подхода являются: метод [12], переносящий параметры классификаторов между соседними кадрами; построение тернарной маски с использованием модели цвета переднего плана текущего и предыдущего кадра в работе [8]; пространственно-временное сэмплирование образцов пикселей, примененное в [13]; использование глобального словаря цветов переднего плана и фона в методе [18].

Последним из рассматриваемых в данном обзоре методов обеспечения стабильности карты прозрачности во времени является темпоральная фильтрация карт прозрачности. В [7] авторы применяют алгоритм фильтрации на основе дистантного преобразования [19] для обеспечения стабильности во времени последовательности карт прозрачности, полученных с помощью алгоритма матирования изображений [20].

В заключение обзора области следует также упомянуть работу [3], в рамках которой было подготовлено 10 видеопоследовательностей с эталонными картами прозрачности и тернарными масками, предложены методы объективной оценки качества карт прозрачности, а также было проведено объективное сравнение ряда существующих алгоритмов матирования изображений и видео.

### Предлагаемый алгоритм

Как уже отмечалось во введении, отличительной особенностью предлагаемого алгоритма является использование восстановленного фона. Очевидно, что знание цвета пикселей фона  $B$  в неизвестной области тернарной маски уменьшает множество возможных решений уравнения (1). Однако в случае задания изображения можно лишь примерно оценить их цвет (что и делают многие алгоритмы матирования изображений). Ситуация кардинально меняется в случае матирования видеопоследовательности: фрагменты фона, скрытые объектом переднего плана на одном кадре, зачастую видимы на других кадрах этой видеопоследовательности. Этот факт позволяет вычислить цвет пикселей фона неизвестной области путем копирования данных с соседних кадров.

Несмотря на то, что различными авторами ранее предлагались алгоритмы матирования видеопоследовательностей, использующие восстановленный фон, данная работа имеет новизну, благодаря:

Использованию алгоритма восстановления фона, применимого к более широкому классу видеопоследо-

вательностей, чем алгоритмы ранних работ.

Предлагаемому обобщению алгоритма матирования, основанного на матирующем лапласиане, используемому восстановленный фон в качестве дополнительных входных данных (ранее в литературе восстановленный фон использовался только в методах, основанных на сэмпировании (последние методы существенно отличаются от первых)).

Методу вычисления доверия к векторам оптического потока, оптимизированному для задачи матирования видео.

Входными данными предлагаемого в данной работе алгоритма являются: исходная видеопоследовательность, последовательность тернарных масок. Алгоритм выполняет следующие основные шаги:

Применить к исходной видеопоследовательности алгоритм восстановления фрагментов фона, скрытых объектами переднего плана.

Применить к каждому кадру исходной последовательности модифицированную версию алгоритма матирования изображений на основе обучения, использующую восстановленный фон в качестве дополнительных входных данных.

Применить итеративный алгоритм пространственно-временного сглаживания последовательности карт прозрачности.

Следует отдельно отметить использование итеративного подхода к сглаживанию карт прозрачности. Формально это может быть записано следующим образом: пусть  $\alpha_t^0$  – кадр  $t$  последовательности карт прозрачности, вычисленной изначально;  $S(\cdot)$  – оператор сглаживания,  $\alpha_t^i$  – кадр  $t$  результата фильтрации на итерации  $i$ , тогда

$$\alpha_t^i = S(\alpha_{t-1}^{i-1}, \alpha_t^{i-1}, \alpha_{t+1}^{i-1}). \quad (2)$$

Доминирующим подходом в литературе на сегодняшний день является метод последовательной фильтрации, результат которой ( $\alpha_t^i$ ) может быть записан как

$$\alpha_t^1 = \alpha_t^0 \quad (3)$$

$$\alpha_t^i = S(\alpha_t^{i-1} - 1, \alpha_t^0, \alpha_t^0 + 1). \quad (4)$$

Согласно проведенным экспериментам, последовательная фильтрация склонна накапливать ошибку, то есть расстояние между вычисленной картой прозрачности и эталоном увеличивается с номером кадра. Итеративный подход лишен данного недостатка, а также допускает параллельное вычисление карт прозрачности нескольких кадров. Далее мы рассмотрим основные шаги алгоритма более детально.

### Восстановление фрагментов фона

Для восстановления фрагментов фона, скрытых передним планом, был использован алгоритм, описанный в [2]. Входными данными алгоритма являются видеопоследовательность и набор масок для каждого ее кадра, отмечающих области, подлежащие восстановлению. Результатом работы алгоритма является видеопоследовательность, совпадающая с исходной в областях, не отмеченных на масках, и содержащая фрагменты вос-

становленного фона в остальных областях. Дополнительными выходными данными алгоритма является набор масок областей, восстановление которых не было выполнено. Принцип работы алгоритма состоит в экстраполяции поля векторов движения в область восстановления и последующем копировании фрагментов фона других кадров согласно экстраполированным векторам. В данной работе в качестве маски восстановления используется объединение областей тернарной маски, отвечающих переднему плану и неизвестной области.

### Базовый метод матирования изображений

Для лучшего понимания дальнейшего описания предлагаемого алгоритма матирования видео рассмотрим краткое описание алгоритма матирования изображений [1], лежащего в его основе. Входными данными алгоритма являются изображение в цветовом пространстве RGB и тернарная маска. Алгоритм выполняет обход изображения с использованием скользящего окна  $3 \times 3$  пиксела. Далее для каждого окна  $\omega$ , содержащего хотя бы один неизвестный пиксел, выразим прозрачность каждого его пиксела  $\alpha^i$  как линейную комбинацию  $w^i$  прозрачностей его соседей  $\alpha_\omega$ :

$$\alpha^i = w^i \alpha_\omega. \quad (5)$$

Для получения  $w^i$  воспользуемся приемом, схожим с линейной регрессией. Пусть, зная цвет пиксела  $x^i = [R; G; B; 1]$  и всех его соседей  $X_\omega$  (каждая строка матрицы состоит из цвета пиксела в системе RGB и единицы), требуется предсказать его прозрачность  $\alpha^i$ . Дополнительно предположим, что известны прозрачности всех его соседей в текущем окне  $\alpha_\omega$ , тогда:

$$\alpha^i = x^{iT} \beta, \quad (6)$$

где  $\beta$  – столбец из четырех элементов, таких, что:

$$\beta = \arg \min_{\beta} \| X_\omega \beta - \alpha_\omega \|^2 + \varepsilon \| \beta \|^2, \quad (7)$$

где  $\varepsilon = 10^{-6}$  – параметр регуляризации. Последнее выражение содержит задачу минимизации квадратичного многочлена с положительно определенным гессианом, а значит, может быть сведено к:

$$\beta = (X_\omega^T X_\omega + \varepsilon I)^{-1} X_\omega^T \alpha_\omega. \quad (8)$$

Подставляя правую часть (8) в уравнение (6) и сопоставляя с уравнением (5), получим:

$$w^i = x^{iT} (X_\omega^T X_\omega + \varepsilon I)^{-1} X_\omega^T. \quad (9)$$

Вектор-строка  $w$  вычисляется для каждого пиксела в каждом окне, содержащем хотя бы один пиксел неизвестной области. Для удобства записи дополним полученные вектора нулями, чтобы было верно:

$$\alpha^i = w^i \alpha, \quad (10)$$

где  $\alpha$  – коэффициенты прозрачности всех пикселов изображения. Матрица  $W$  образуется из дополненных нулями векторов. Теперь задача вычисления карты прозрачности сводится к следующей задаче оптимизации:

$$\alpha = \arg \min_{\alpha} \| \alpha - W \alpha \|^2 + (\alpha - T)^T D (\alpha - T), \quad (11)$$

где  $T$  – тернарная маска,  $D$  – диагональная матрица, у которой элементы главной диагонали, отвечающие пикселам неизвестной области, равны нулю, а осталь-

ные –  $10^6$ . Решение последней задачи может быть сведено к решению разреженной системы линейных уравнений. Ее решение и есть результат работы базового метода матирования изображений.

### Использование восстановленного фона

Рассмотрим, как базовый метод может быть модифицирован для использования восстановленного фона в качестве дополнительных данных. Важно, чтобы данная модификация была выполнена с учетом следующих фактов:

- результаты работы метода восстановления фона могут содержать неверно восстановленные фрагменты.
- восстановленный фон зачастую смещен на несколько пикселей из-за накопления ошибки восстановления.

Для тех окон, которые не содержат восстановленного фона, базовый метод применяется без модификаций. Для каждого окна  $\omega$ , содержащего хотя бы один пиксел восстановленного фона, дополним матрицу  $X_\omega$  (см. уравнение (7)) цветами пикселей восстановленного фона. Использование соответствий окон фона окнам изображений вместо прямых соответствий между пикселями обеспечивает устойчивость метода к сдвигам фона до 3 пикселей. Для обеспечения устойчивости метода к ошибкам восстановления фона модифицируем уравнение (7), взвесив образцы пикселей в матрице  $X_\omega$  на доверие к ним:

$$\beta = \arg \min_{\beta} \|C(X_\omega \beta - \alpha_\omega)\|^2 + \varepsilon \|\beta\|^2, \quad (12)$$

где  $C$  – диагональная матрица, элементы главной диагонали которой определяют доверие к соответствующим строкам матрицы  $X_\omega$ . В предлагаемом методе доверие ко всем образцам фона устанавливается равным 0,08 и 1 для остальных пикселей. Задача (12), аналогично предыдущему разделу, может быть переписана в виде:

$$\beta = (X_\omega^T C X_\omega + \varepsilon I)^{-1} X_\omega^T C \alpha_\omega, \quad (13)$$

тогда (аналогично уравнению (9)):

$$w' = x'^T (X_\omega^T C X_\omega + \varepsilon I)^{-1} X_\omega^T C. \quad (14)$$

Так как прозрачность пикселей восстановленного фона равна нулю, в уравнении (5) подставим нули вместо соответствующих им элементов вектора  $\alpha_\omega$ , эта операция эквивалентна удалению из вектора  $w'$  элементов, отвечающих пикселям восстановленного фона. Аналогично базовому методу получаем задачу вида (11), которая сводится к решению системы линейных уравнений. Таким образом, полученный метод расширяет каждое окно базового метода образцами пикселей восстановленного фона, прозрачность которых заведомо равна нулю, и учитывает их с низким доверием для обеспечения устойчивости к неточностям восстановленного фона.

### Обеспечение стабильности результата во времени

Как уже упоминалось ранее, предложенный метод выполняет итеративную фильтрацию последовательно-сти карт прозрачности по схеме, описанной в уравнении (2). Здесь и далее запись  $v_k^n$  обозначает поле векторов

движения от кадра  $n$  к кадру  $k$ . Рассмотрим, какие операции выполняются при фильтрации кадра  $t$ :

1. Вычисление полей векторов движения  $v_{t-1}^t, v_{t+1}^t, v_{t-1}^{t-1}$  и  $v_{t+1}^{t-1}$  с помощью алгоритма, описанного в [21] (алгоритм [21] вычисляет поле векторов движения  $v_k^n$ , выполняя поиск блоков кадра  $k$ , наиболее похожих на блоки кадра  $n$ ; смещения между блоками кадра  $n$  и найденными блоками кадра  $k$  образуют искомое поле векторов).

2. Вычисление карт доверия к полям векторов движения  $K_{t-1}^t$  и  $K_{t+1}^t$ .

3. Применение базового метода матирования, учитывающего восстановленный фон, с дополнительными требованиями стабильности результата во времени.

Для вычисления доверия к векторам движения, полученным на первом шаге, используются две вспомогательные величины: **FBC** и  $E$ . **FBC** – несоответствие между полями векторов движения для прямого и обратного проходов. Значение **FBC** в точке с координатой  $x$  может быть выражено формулой:

$$\mathbf{FBC}_k^n(x) = \|v_k^n(x) - v_n^k(x + v_k^n(x))\|_2^2. \quad (15)$$

$E_k^n(x) - L_1$  расстояние между блоком 8x8 пикселей с центром в точке  $x$  кадра  $n$  исходной видеопоследовательности и аналогичным блоком кадра  $k$  с центром в точке  $x + v_k^n(x)$ .

С использованием введенных величин доверие к векторам движения вычисляется по формуле:

$$K_{t-1}^t = \exp\left(-\frac{\mathbf{FBC}_{t-1}^t}{\sigma_1} - \frac{E_{t-1}^t}{\sigma_2}\right), \quad (16)$$

где  $\sigma_1, \sigma_2$  – константы, выбранные с использованием двух видеопоследовательностей с известной эталонной картой прозрачности из набора [3]. Для этого были вычислены вектора движения для этих последовательностей. Затем исключены из рассмотрения вектора, принадлежащие областям с малой дисперсией значений эталонной прозрачности. Для оценки качества вектора  $v_k^n(x)$  используется пиковое отношение сигнала к шуму  $P_k^n(x)$  между блоком 8x8 пикселей с центром в точке  $x$  эталонной карты прозрачности кадра  $n$  и аналогичным блоком кадра  $k$  с центром в точке  $x + v_k^n(x)$ . Путем визуального контроля было установлено, что значения  $P$  менее 12 децибел соответствуют ошибочно найденным векторам.  $\sigma_1, \sigma_2$  были выбраны как:

$$\langle \sigma_1, \sigma_2 \rangle = \arg \min_{\sigma_1, \sigma_2} \sum_{P_k^n(x)} \exp\left(-\frac{\mathbf{FBC}_{t-1}^t}{\sigma_1} - \frac{E_{t-1}^t}{\sigma_2}\right) (12 - P_k^n(x)). \quad (17)$$

Для решения последней задачи был использован метод внутренней точки. Найденные значения:  $\sigma_1 = 3,9$ ;  $\sigma_2 = 29$ .

Рассмотрим шаг матирования. Под операцией компенсации кадра  $k$  к кадру  $n$  будем понимать отображение  $R$ :

$$\Gamma = R(I, v_k^n), \quad (18)$$

такое, что

$$\Gamma(x + v_k^n(x)) = \Gamma(x). \quad (19)$$

Компенсируем предыдущий кадр, следующий кадр и их карты прозрачности (вычисленные на предыдущей итерации) к текущему кадру, будем обозначать их  $I_B$ ,  $I_F$ ,  $\alpha_B$  и  $\alpha_F$  соответственно. Аналогично описанному ранее способу расширим для каждого окна  $\omega$  матрицу  $X_\omega$  пикселями из  $I_B$  и  $I_F$ . Коэффициенты доверия к этим образцам вычисляются по формуле (16), а затем умножаются на 0,2 (экспериментально выбранное значение). В качестве значений прозрачности добавленных пикселей используются значения прозрачности в окне  $\omega$  карт  $\alpha_B$  и  $\alpha_F$ . Добавим в уравнение (11) новые слагаемые:

$$\begin{aligned} & \psi[(\alpha - \alpha_B)^T \text{Diag}(K_{t-1}^t)(\alpha - \alpha_B) + \\ & + (\alpha - \alpha_F)^T \text{Diag}(K_{t+1}^t)(\alpha - \alpha_F)], \end{aligned} \quad (20)$$

где  $\text{Diag}(K_{t-1}^t)$  – диагональная матрица, образованная значениями карты доверия  $K_{t-1}^t$ ;  $\psi = 0,1$ .

Предложенный метод выполняет две итерации сглаживания для вычисления итоговой карты прозрачности.

#### Экспериментальная оценка

Описанный метод был сравнен с другими методами матирования изображений и видео на 10 тестовых видео с известными эталонными картами прозрачности из набора, предложенного в [3] (примеры двух видеопоследовательностей из набора изображены рис. 1). Критериями оценки выступали:  $L_2$  расстояние до эталона и **meSSDdt** (метод сравнения, предложенный в работе

[22], имеющий высокую корреляцию с результатами экспертной оценки). Величина **meSSDdt** выражает среднее различие между ошибкой метода на текущем кадре и скомпенсированной, согласно векторам движения ошибкой на следующем кадре. Иными словами, **meSSDdt** штрафует изменения ошибки между соседними кадрами и тем самым позволяет оценить стабильность результата во времени. Результаты сравнения предложенного метода с 10 ранее существовавшими приведены в табл. 1 и 2. Для каждой тестовой последовательности методы были упорядочены согласно критерию сравнения, а затем было вычислено среднее место метода во всех тестах. Как видно из таблиц, предложенный метод превзошел аналоги по данному показателю. Рассмотрим также сравнение предложенного метода с алгоритмом матирования видео, предложенным в 2015 году [18]. Сравнение было выполнено на устаревшей версии набора тестовых видео (к сожалению, авторы метода не предоставили результатов тестирования своего метода на обновленном наборе). Результаты представлены на рис. 1, согласно которому предложенный метод превзошел аналог.

В табл. 1 представлены результаты сравнения предложенного алгоритма с 10 аналогами на 10 тестовых видеопоследовательностях из набора, описанного в [3]. Пространственные ошибки методов были вычислены как  $L_2$  расстояния между результатами работы алгоритмов и эталоном (приведены в столбцах 3-12). Затем методы были упорядочены по близости результата к эталону (надстрочные числа в столбцах 3-12). Для каждого метода было вычислено среднее место во всех тестах (второй столбец). Меньшие значения соответствуют более высоким местам в рейтинге. Предложенный метод превзошел аналоги по среднему месту во всех тестах.



а) Исходная видеопоследовательность

б) Эталонная карта прозрачности

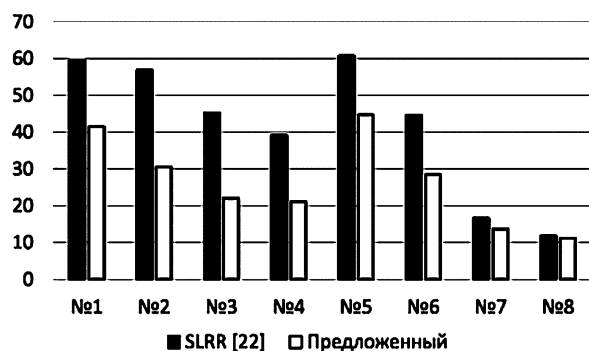
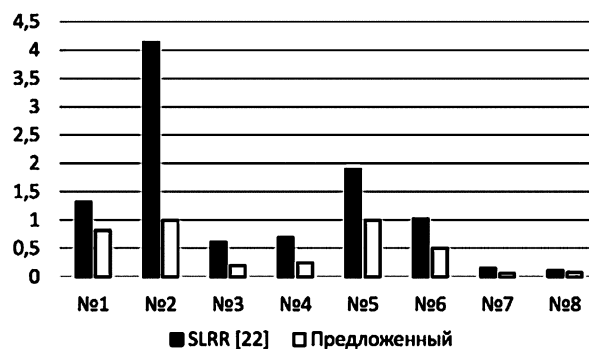
Рис. 1. Пример двух видеопоследовательностей (а) с эталонными картами прозрачности (б) из набора тестовых данных [3]

Таблица 1.

Метод	Ср. место	Тестовая последовательность									
		№1	№2	№3	№4	№5	№6	№7	№8	№9	№10
Предложенный	1,9	29,39 <sup>1</sup>	27,79 <sup>2</sup>	8,83 <sup>2</sup>	122,22 <sup>5</sup>	13,96 <sup>1</sup>	21,01 <sup>1</sup>	31,83 <sup>3</sup>	27,66 <sup>1</sup>	49,26 <sup>1</sup>	59,67 <sup>2</sup>
Learning Based [1]	3,1	29,98 <sup>2</sup>	33,08 <sup>5</sup>	10,84 <sup>6</sup>	134,82 <sup>7</sup>	15,23 <sup>2</sup>	23,35 <sup>2</sup>	31,46 <sup>2</sup>	30,20 <sup>2</sup>	58,38 <sup>2</sup>	56,60 <sup>1</sup>
Closed Form [11]	4,3	31,34 <sup>3</sup>	31,80 <sup>4</sup>	10,96 <sup>7</sup>	149,48 <sup>9</sup>	15,86 <sup>3</sup>	23,67 <sup>3</sup>	30,15 <sup>1</sup>	30,60 <sup>3</sup>	69,94 <sup>4</sup>	82,13 <sup>6</sup>
Comprehensive Sampling [23]	4,3	34,48 <sup>4</sup>	30,18 <sup>3</sup>	12,10 <sup>8</sup>	79,60 <sup>1</sup>	16,17 <sup>4</sup>	25,02 <sup>4</sup>	35,25 <sup>5</sup>	41,05 <sup>5</sup>	77,72 <sup>5</sup>	70,18 <sup>4</sup>
Robust Matting [20]	4,6	37,54 <sup>6</sup>	26,15 <sup>1</sup>	9,61 <sup>3</sup>	88,15 <sup>3</sup>	18,52 <sup>5</sup>	33,31 <sup>6</sup>	34,15 <sup>4</sup>	46,91 <sup>7</sup>	78,56 <sup>6</sup>	80,41 <sup>5</sup>
Shared Matting [24]	4,9	36,79 <sup>5</sup>	46,72 <sup>7</sup>	10,30 <sup>4</sup>	87,64 <sup>2</sup>	24,41 <sup>6</sup>	35,18 <sup>7</sup>	35,26 <sup>6</sup>	45,23 <sup>6</sup>	68,52 <sup>3</sup>	67,17 <sup>3</sup>
Refine Edge [7]	5,8	37,78 <sup>7</sup>	38,48 <sup>6</sup>	8,64 <sup>1</sup>	129,56 <sup>6</sup>	27,47 <sup>7</sup>	31,36 <sup>5</sup>	36,65 <sup>7</sup>	38,75 <sup>4</sup>	102,02 <sup>8</sup>	94,50 <sup>7</sup>
Bayesian Matting [25]	7,9	58,76 <sup>9</sup>	51,66 <sup>8</sup>	10,79 <sup>5</sup>	144,53 <sup>8</sup>	30,36 <sup>8</sup>	53,28 <sup>8</sup>	43,70 <sup>8</sup>	70,50 <sup>9</sup>	96,69 <sup>7</sup>	140,12 <sup>9</sup>
KNN Matting [26]	8,8	67,83 <sup>10</sup>	92,22 <sup>9</sup>	64,73 <sup>11</sup>	112,46 <sup>4</sup>	32,96 <sup>9</sup>	61,50 <sup>9</sup>	50,87 <sup>9</sup>	79,20 <sup>10</sup>	113,41 <sup>9</sup>	123,63 <sup>8</sup>
Nonlocal matting [27]	9,7	54,21 <sup>8</sup>	92,37 <sup>10</sup>	33,86 <sup>10</sup>	243,47 <sup>11</sup>	45,13 <sup>10</sup>	64,05 <sup>10</sup>	61,75 <sup>10</sup>	67,30 <sup>8</sup>	135,08 <sup>10</sup>	171,79 <sup>10</sup>
Spectral Matting [28]	10,7	86,32 <sup>11</sup>	105,24 <sup>11</sup>	13,15 <sup>9</sup>	210,04 <sup>10</sup>	48,76 <sup>11</sup>	65,72 <sup>11</sup>	68,83 <sup>11</sup>	109,44 <sup>11</sup>	157,18 <sup>11</sup>	182,82 <sup>11</sup>

Таблица 2.

Метод	Ср. место	Тестовая последовательность									
		№1	№2	№3	№4	№5	№6	№7	№8	№9	№10
Предложенный	1,3	0,44 <sup>1</sup>	0,941 <sup>2</sup>	0,07 <sup>2</sup>	2,65 <sup>1</sup>	0,15 <sup>1</sup>	0,25 <sup>1</sup>	0,66 <sup>2</sup>	0,51 <sup>1</sup>	1,60 <sup>1</sup>	2,20 <sup>1</sup>
Closed Form [11]	3,1	0,49 <sup>2</sup>	1,35 <sup>4</sup>	0,09 <sup>6</sup>	2,88 <sup>2</sup>	0,19 <sup>3</sup>	0,35 <sup>3</sup>	0,61 <sup>1</sup>	0,61 <sup>2</sup>	2,75 <sup>3</sup>	3,38 <sup>5</sup>
Learning Based [1]	3,2	0,50 <sup>3</sup>	1,42 <sup>5</sup>	0,09 <sup>7</sup>	2,88 <sup>3</sup>	0,18 <sup>2</sup>	0,34 <sup>2</sup>	0,68 <sup>3</sup>	0,64 <sup>3</sup>	2,37 <sup>2</sup>	2,43 <sup>2</sup>
Refine Edge [29]	4,6	0,55 <sup>4</sup>	1,48 <sup>6</sup>	0,06 <sup>1</sup>	3,99 <sup>6</sup>	0,34 <sup>6</sup>	0,54 <sup>5</sup>	0,70 <sup>4</sup>	0,81 <sup>4</sup>	3,80 <sup>6</sup>	3,16 <sup>4</sup>
Comprehensive Sampling [23]	5,1	0,74 <sup>5</sup>	1,06 <sup>3</sup>	0,19 <sup>9</sup>	3,15 <sup>4</sup>	0,22 <sup>4</sup>	0,39 <sup>4</sup>	0,86 <sup>6</sup>	1,24 <sup>5</sup>	3,65 <sup>5</sup>	3,40 <sup>6</sup>
Robust Matting [20]	5,6	0,89 <sup>7</sup>	0,936 <sup>1</sup>	0,07 <sup>3</sup>	4,82 <sup>8</sup>	0,25 <sup>5</sup>	0,88 <sup>6</sup>	0,77 <sup>5</sup>	1,58 <sup>7</sup>	3,89 <sup>7</sup>	3,92 <sup>7</sup>
Shared Matting [24]	5,9	0,79 <sup>6</sup>	2,73 <sup>7</sup>	0,09 <sup>5</sup>	4,50 <sup>7</sup>	0,35 <sup>7</sup>	1,00 <sup>7</sup>	0,91 <sup>7</sup>	1,43 <sup>6</sup>	2,94 <sup>4</sup>	2,94 <sup>3</sup>
KNN Matting [26]	8,4	2,28 <sup>10</sup>	6,40 <sup>9</sup>	1,22 <sup>11</sup>	3,47 <sup>5</sup>	0,61 <sup>8</sup>	1,02 <sup>8</sup>	1,73 <sup>9</sup>	1,94 <sup>8</sup>	6,42 <sup>8</sup>	7,73 <sup>8</sup>
Bayesian Matting [25]	9,1	1,85 <sup>9</sup>	3,58 <sup>8</sup>	0,12 <sup>8</sup>	13,33 <sup>9</sup>	0,74 <sup>9</sup>	2,31 <sup>11</sup>	1,47 <sup>8</sup>	4,51 <sup>10</sup>	8,01 <sup>9</sup>	16,23 <sup>10</sup>
Spectral Matting [28]	9,8	7,03 <sup>11</sup>	13,25 <sup>11</sup>	0,08 <sup>4</sup>	15,36 <sup>10</sup>	2,09 <sup>11</sup>	1,84 <sup>9</sup>	2,45 <sup>10</sup>	6,59 <sup>11</sup>	8,71 <sup>10</sup>	19,97 <sup>11</sup>
Nonlocal matting [27]	9,9	1,56 <sup>8</sup>	8,17 <sup>10</sup>	0,70 <sup>10</sup>	20,97 <sup>11</sup>	1,11 <sup>10</sup>	1,88 <sup>10</sup>	2,51 <sup>11</sup>	2,35 <sup>9</sup>	11,17 <sup>11</sup>	13,37 <sup>9</sup>

а)  $L_2$  расстояние до эталона

б) meSSDdt (предложена в [22])

Рис. 2. Сравнение предложенного алгоритма с алгоритмом, описанным в [18], на 8 тестовых видео последовательностях (каждая пара столбцов соответствует результату для одной тестовой последовательности)

с использованием  $L_2$  расстояния до эталона (а) для оценки пространственной ошибки и meSSDdt (б) для оценки временной стабильности результата

В табл. 2 представлены результаты сравнения предложенного алгоритма с 10 аналогами на 10 тестовых видеопоследовательностях из набора, описанного в [3]. Для каждой видеопоследовательности были вычислены значения квазиметрики meSSDdt (в работе [22] показана высокая корреляция данного способа оценки с результатами экспертной оценки) между результатами работы алгоритмов и эталоном (приведены в столбцах

3-12) для оценки стабильности во времени результатов. Затем методы были упорядочены по близости результата к эталону (надстрочные числа в столбцах 3-12). Для каждого метода было вычислено среднее место во всех тестах (второй столбец). Меньшие значения соответствуют более высоким местам в рейтинге. Предложенный метод превзошел аналоги по среднему месту во всех тестах.



## Заключение

В работе был предложен алгоритм матирования видеопоследовательностей, использующий восстановленный фон в качестве дополнительных данных. К основным особенностям можно отнести:

- модификацию базового алгоритма матирования, позволившую использовать фон в качестве дополнительных данных;

- итеративный подход к пространственно-временному сглаживанию карт прозрачности, позволивший избежать накопления ошибки и вычислять несколько кадров параллельно;

- способ вычисления доверия к векторам движения, выбранный путем анализа двух видео с эталонной картой прозрачности.

Преимущества предложенного метода были продемонстрированы путем объективного сравнения результатов его работы с 11 аналогами.

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта 15-01-08632 а.

Данная статья поступила в редакцию журнала «Цифровая обработка сигналов» 22.03.2016 г.

## Литература

1. Zheng Yuanjie, Kambhamettu C. Learning based digital matting // International Conference on Computer Vision (ICCV). – 2009. – P. 889-896.
2. Зачесов А. А., Ерофеев М. В., Ватолин Д. С. Использование карт глубины при восстановлении фона в видеопоследовательностях // Новые информационные технологии в автоматизированных системах: материалы научно-практического семинара. – М.: МИЭМ НИУ ВШЭ, 2015.
3. Методика объективного сравнения алгоритмов матирования видео/М. В. Ерофеев, Ю. А. Гитман, Д. С. Ватолин, А. А. Федоров // Цифровая обработка сигналов. – 2015. – № 3. – С. 53–59.
4. Video matting of complex scenes / Yung-Yu Chuang, Aseem Agarwala, Brian Curless et al.//ACM Transactions on Graphics (TOG). – 2002. – Vol. 21, no. 3. – P. 243-248.
5. Apostoloff N., Fitzgibbon A. Bayesian video matting using learnt image priors // Computer Vision Pattern Recognition (CVPR). – Vol. 1. – 2004. – P. 1-407-1-414.
6. Lee Sun-Young, Yoon Jong-Chul, Lee In-Kwon. Temporally coherent video matting // Graphical Models. – 2010. – Vol. 72, no. 3. – P. 25–33.
7. Bai Xue, Wang Jue, Simons David. Towards temporally coherent video matting // International Conference on Computer Vision (ICCV). – 2011. – P. 63-74.
8. Video matting via opacity propagation / Z. Tang, Z. Miao, Y. Wan, D. Zhang // The Visual Computer. – 2012. – Vol. 28, no. 1. – P. 47-61.
9. Corrigan D., Robinson S, Kokaram A. Video matting using motion extended grabcut // European Conference on Visual Media Production (CVMP). – 2008. – P. 3–3(1).
10. Rother C., Kolmogorov V., Blake A. Grabcut: Interactive foreground extraction using iterated graph cuts // ACM transactions on graphics (TOG) / ACM. – Vol. 23. – 2004. – P. 309–314.

11. Levin A., Lischinski D., Weiss Y. A closed-form solution to natural image matting // IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI). – 2008. – Vol. 30, no. 2. – P. 228–242.
12. Video snapcut: Robust video object cutout using localized classifiers / X. Bai, J. Wang, D. Simons, G. Sapiro // ACM Transactions on Graphics (TOG). – 2009. – Vol. 28, no. 3. – P. 70:1–70:11.
13. Temporally coherent and spatially accurate video matting / Ehsan Shahrian, Brian Price, Scott Cohen, Deepu Rajan // Computer Graphics Forum. – 2014. – Vol. 33, no. 2. – P. 381–390.
14. Sindeev M., Konushin A., Rother C. Alpha-flow for video matting // Computer Vision-ACCV 2012. – Springer, 2012. – P. 438-452.
15. Choi I., Lee M., Tai Y.-W. Video matting using multi-frame nonlocal matting laplacian // European Conference on Computer Vision (ECCV). – 2012. – P. 540-553.
16. Spatio-temporally coherent interactive video object segmentation via efficient filtering / N. Brosch, A. Hosni, C. Rhemann, M. Gelautz // Pattern Recognition. – Vol. 7476. – 2012. – P. 418-427.
17. He Kaiming, Sun Jian, Tang Xiaoou. Guided image filtering // European Conference on Computer Vision (ECCV). – 2010. – P. 1–14.
18. Video matting via sparse and low-rank representation / D. Zou, X. Chen, G. Cao, X. Wang // Proceedings of the IEEE International Conference on Computer Vision. – 2015. – P. 1564-1572.
19. Linear time euclidean distance transform algorithms / H. Breu, J. Gil, D. Kirkpatrick, M. Werman // Pattern Analysis and Machine Intelligence, IEEE Transactions on. – 1995. – Vol. 17, no. 5. – P. 529–533.
20. Wang J., Cohen M. F. Optimized color sampling for robust matting // Computer Vision Pattern Recognition (CVPR). – 2007. – P. 1–8.
21. Fast video super-resolution via classification / K. Simonyan, S. Grishin, D. Vatolin, D. Popov // International Conference on Image Processing (ICIP). – 2008. – P. 349–352.
22. Perceptually motivated benchmark for video matting / M. Erofeev, Y. Gitman, D. Vatolin et al. // Proceedings of the British Machine Vision Conference (BMVC). – BMVA Press, 2015. – September. – P. 99.1–99.12. – URL: <https://dx.doi.org/10.5244/C.29.99>.
23. Improving image matting using comprehensive sampling sets / E. Shahrian, D. Rajan, B. Price, S. Cohen // Computer Vision Pattern Recognition (CVPR). – 2013. – P. 636–643.
24. Gastal Eduardo S.L., Oliveira Manuel M. Shared sampling for real-time alpha matting // Computer Graphics Forum. – 2010. – Vol. 29, no. 2. – P. 575–584.
25. A bayesian approach to digital matting / Y.-Y. Chuang, B. Curless, D. H. Salesin, R. Szeliski // Computer Vision Pattern Recognition (CVPR). – Vol. 2. – 2001. – P. 11-264-11-271.
26. Chen Qifeng, Li Dingzeyu, Tang Chi-Keung. KNN matting // IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI). – 2013. – Vol. 35, no. 9. – P. 2175-2188.
27. Lee Philip, Wu Ying. Nonlocal matting // Computer Vision Pattern Recognition (CVPR). – 2011. – P. 2193–2200.
28. Levin A., Rav Acha A., Lischinski D. Spectral matting // IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI). – 2008. – Vol. 30, no. 10. – P. 1699-1712.
29. <http://www.adobe.com/en/products/aftereffects.html>.

## МЕТОДИКА ОБЪЕКТИВНОЙ ОЦЕНКИ КАЧЕСТВА ВОССТАНОВЛЕНИЯ ФОНА В ВИДЕО

*Боков А.А., аспирант Московского государственного университета им. М.В. Ломоносова, e-mail: abokov@graphics.cs.msu.ru;*

*Ватолин Д.С., к.ф.-м.н., с.н.с. Московского государственного университета им. М.В. Ломоносова, e-mail: dmitriy@graphics.cs.msu.ru.*

## OBJECTIVE QUALITY ASSESSMENT METHODOLOGY FOR VIDEO BACKGROUND RECONSTRUCTION

*Bokov A.A., Vatolin D.S.*

*In its general form, video background reconstruction is usually defined as a task of plausible video region reconstruction that is marked with an input mask. Object removal is a typical example of background reconstruction. Several new methods were introduced over the past few years; however, no standard benchmark has yet been established. In this work we propose an objective background reconstruction quality benchmark that consists of several metrics that we demonstrate to have higher correlation with perceptual quality compared to prior approaches. Perceptual background reconstruction quality is quantitatively evaluated based on pairwise comparison of background reconstruction methods performed by over 300 human subjects.*

**Key words:** background reconstruction, video processing, objective quality assessment, quality benchmark.

**Ключевые слова:** восстановление фона, обработка видео, объективная оценка качества, методика сравнения.

### Введение

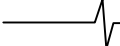
Одной из важных задач в области обработки видео является восстановление фона. Ее можно определить как задачу максимально правдоподобной реконструкции области видео, помеченной входной маской  $\Omega$ , на базе известной части видеопоследовательности, лежащей вне маски:  $V/\Omega$ . Такая постановка находит применение в ряде прикладных областей, включающих восстановление от ошибок видеокодека, возникающих при потере пакетов в процессе передачи видео по сети, восстановление архивных видеозаписей, содержащих различные дефекты пленки, удаление нежелательных объектов из видео, заполнение областей открытия в процессе конвертации видео в стереоскопический формат и многое другое.

Многие алгоритмы накладывают существенные ограничения на входные видеопоследовательности и маски областей восстановления или требуют определенные дополнительные данные для работы, ограничивая таким образом область своей применимости [27, 28] или же ориентируясь на конкретный узкий сценарий работы [4-9]. Однако ряд алгоритмов, предложенных в последние годы, позволяют решать задачу восстановления фона в достаточно широкой ее постановке. Например, алгоритм, предложенный в работе [1], позволяет работать со свободно движущейся камерой и восстанавливать динамические текстуры в видео, такие как поверхность воды, огонь, толпа людей. Алгоритм, предложенный в [2], накладывает только требования статичности восстанавливаемого фона и видимости восстанавливаемых

*В общей постановке задача восстановления фона в видео состоит в максимально правдоподобной реконструкции области видеопоследовательности, отмеченной входной маской, на базе известной части видеопоследовательности, лежащей вне маски. Однако большинство авторов ограничиваются публикацией результатов работы предложенного алгоритма на нескольких видеопоследовательностях, которые зачастую различны для разных подходов. Предлагается методика для объективного сравнения алгоритмов восстановления фона в видео. Вводится ряд показателей качества восстановления фона, превосходящих предшествующие подходы по корреляции с экспертной оценкой, полученной путем попарного сравнения результатов различных алгоритмов с участием более 300 респондентов.*

областей хотя бы в одном кадре входного видео. Но авторы отмечают, что время работы алгоритма на одной последовательности длительностью 100 кадров достигало четырех часов на сервере с 64 логическими процессорами. Аналогичная ситуация характерна для большинства современных алгоритмов восстановления видео: время работы и потребление памяти серьезно ограничивают их применимость к длинным видеопоследовательностям высокого разрешения. В работе [3] авторы предлагают ограничить вычислительную сложность и потребление памяти за счет ограничения пространства поиска до фиксированной временной окрестности текущего кадра.

В целом, прогресс в данной области во многом ограничивается отсутствием общепризнанной методики оценки качества и сравнения различных алгоритмов восстановления фона в видео. Большинство авторов ограничиваются публикацией результатов работы алгоритма на нескольких видео, иногда вместе с результатами работы подходов-конкурентов на тех же видеопоследовательностях. Применение объективных методов



оценки качества затруднено даже при наличии эталонных результатов восстановления, так как алгоритмы восстановления видео, как правило, оцениваются по визуальному качеству восстановления с точки зрения наблюдателя, а не по близости к некоторому эталонному результату. Эта проблема приобретает особую актуальность в случаях, когда область восстановления имеет большой размер как в смысле пространственных размеров, так и временной продолжительности (например, удаление объекта из видео).

В данной работе предлагается ряд методов, позволяющих производить объективную оценку качества восстановления фона в видео, в предположении, что доступен эталонный результат восстановления. Адекватность предложенных методов оценки качества была подтверждена следующим образом:

Построены 7 тестовых видеопоследовательностей, включающих эталонный фон, который является идеальным результатом восстановления (§ 3.1).

Проведена автоматическая оценка качества результатов работы 6 алгоритмов восстановления фона на построенных видеопоследовательностях с помощью предложенных объективных показателей качества (§ 3.2).

Проведена экспертная оценка качества путем попарного сравнения результатов различных алгоритмов восстановления фона с участием более 300 респондентов (§ 3.3).

Вычислена корреляция между результатами экспертной и автоматической оценок (§ 4).

#### **Обзор области исследований**

В данном разделе представлен краткий обзор существующих подходов к оценке качества и получения эталонных результатов восстановления фона в видеопоследовательностях. Рассматриваются методы и наборы данных, которые использовались авторами алгоритмов восстановления фона для проведения объективной оценки качества и сравнения с конкурентами.

Количественная оценка качества восстановления часто используется в случаях, когда восстанавливаемый регион видео имеет относительно малый размер либо в смысле пространственных размеров, либо временной продолжительности. Это верно, например, для задачи скрытия ошибок, вызванных потерей пакетов в процессе передачи видеопоследовательности по сети [4-6]. Как правило, используются традиционные методы оценки качества видео, такие как PSNR и SSIM. Эталонный результат восстановления получается за счет того, что ошибки, эмулирующие эффект потери пакетов разной интенсивности, вносятся в исходный видеопоток искусственным образом. Авторы алгоритма, предложенного в работе [3], проводят его апробацию в ряде различных сценариев применения, однако количественная оценка качества работы проводится только для сценария скрытия ошибок, вызванных потерей пакетов при передаче видеопотока. Аналогичным образом оценка качества производится для задач автоматического устранения логотипов [7, 8] и удаления текста [9] из видеопоследовательности. Авторы работы [10] используют меру

RMSE (Root-Mean-Square Error – корень из среднеквадратичной ошибки) для оценки качества восстановления достаточно крупных областей, но продолжающихся не более 5 кадров. В некоторых тестовых последовательностях целые кадры включались в область восстановления, что, фактически, сводит задачу к межкадровой интерполяции. Предложенный алгоритм восстановления был сравнен с одним альтернативным подходом на 10 последовательностях низкого разрешения (порядка 352×240) с продолжительностью от 35 до 100 кадров.

В работе [11] для построения тестовых последовательностей на каждом кадре исходной видеопоследовательности вырезается случайным образом выбранный блок. Предложенный алгоритм сравнивается с пятью альтернативными алгоритмами. Авторы работ [12] и [13] проводят сравнение на единственной последовательности, где небольшой блок вырезается из одной и той же позиции на каждом кадре. Для оценки качества используются PSNR и среднеквадратичная ошибка. В работе [14] используется простая сумма абсолютных разностей (SAD) между результатом реконструкции и оригинальной видеопоследовательностью для демонстрации высокой точности предложенного подхода. Он сравнивается с двумя альтернативными подходами на одной последовательности и трех отдельных областях восстановления, которые характеризуются различным движением (быстрое движение, медленное движение и его полное отсутствие). Однако стоит отметить, что рассматриваемые авторами области являются относительно крупными как в пространственном, так и во временном направлениях, но никакого обоснования используемой методики оценки качества не приводится. Авторы работы [15] проводят подсчет индекса SSIM на одной из тестовых последовательностей, однако приведенное сравнение с конкурирующими подходами имеет качественный, а не количественный характер. В недавнем обзоре алгоритмов восстановления фона 2014 года [16] отмечается, что авторам неизвестны работы, которые бы рассматривали проблему количественной оценки качества восстановления фона видеопоследовательностей в общей постановке.

По результатам проведенного обзора можно сделать вывод, что проблема количественной оценки качества восстановления больших пространственно-временных областей в видеопоследовательностях явным образом не рассматривалась в предшествующих работах. Классические методы оценки качества, такие как PSNR и SSIM, хорошо подходят для оценки качества восстановления областей, которые либо являются небольшими в смысле пространственных размеров, либо имеют малую продолжительность, не превышающую нескольких кадров. Однако эти методы становятся значительно менее надежными в случаях областей восстановления, которые являются крупными как в смысле пространственных размеров, так и временной продолжительности, так как соответствие результата восстановления некоторому эталону уже не обязательно для достижения высокого визуального качества.

## Методика количественного сравнения алгоритмов восстановления фона

### Построение тестового набора данных

Для количественного сравнения различных подходов, в первую очередь, требуется определение набора данных, на котором оно будет проводиться. При построении тестовых видеопоследовательностей принимался во внимание ряд принципов. Во-первых, каждая тестовая видеопоследовательность должна представлять некоторую сложность для существующих алгоритмов восстановления фона: тривиальные примеры не позволяют отделить высококачественные алгоритмы от низкокачественных. Также предпринимается попытка охватить как можно больше практических случаев восстановления фона, включая последовательности со статичной и свободно перемещающейся камерой, статический и динамический фон, динамические текстуры в видео (например, поверхность воды, огонь). Некоторые из построенных видеопоследовательностей имеют области восстановления, которые включают фрагменты фона, не видимые ни в одном из кадров входного видео, что нарушает предположение, используемое множеством алгоритмов восстановления фона (например, [2, 22]). И, наконец, все предлагаемые тестовые видеопоследовательности имеют разрешение  $1920 \times 1080$  и продолжительность от 150 до 200 кадров. Это важно для выделе-

ния более практичных подходов, так как многие предлагаемые алгоритмы имеют серьезные ограничения в плане обработки продолжительных видеопоследовательностей высокого разрешения из-за низкой скорости работы и/или слишком высокого потребления памяти.

Все построенные тестовые видеопоследовательности относятся к одному способу применения алгоритмов восстановления фона – удаление объектов из видео. Для получения видеопоследовательностей с эталонным восстановленным фоном проводится искусственное наложение различных объектов переднего плана на фоновые видеопоследовательности средствами компьютерной графики. Для максимально правдоподобной интеграции объектов переднего плана в фоновое видео с учетом движения камеры используется программный пакет Blender [17]. Пример построенного видео с маской области восстановления приведен на рис. 1.

Таким образом, каждый алгоритм восстановления фона получает на вход видеопоследовательность с наложенным объектом переднего плана и соответствующей маской восстановления. Затем проводится оценка качества результата восстановления с учетом исходной видеопоследовательности, содержащей эталонные изображения фона за наложенным объектом. Всего было построено 7 тестовых видеопоследовательностей для сравнения качества различных алгоритмов восстановления фона.



а) Кадр с наложенным объектом переднего плана



б) Маска восстановления

Рис. 1. Пример кадра из тестовой видеопоследовательности для оценки качества алгоритмов восстановления фона в видео.

Целью восстановления является максимально правдоподобное удаление объекта из входного видео



а) Результат восстановления фона алгоритмом [1]



б) Результат восстановления фона алгоритмом [24]

Рис. 2. Иллюстрация проблем традиционных методов в контексте оценки качества восстановления фона в видео.

Результат слева очевидно обладает более высоким визуальным качеством, однако метрика MSE указывает на обратное

### Методы оценки качества восстановления фона

В качестве исходной точки для сравнения разумно использовать традиционные для области оценки качества видео методы, такие как:

$$\text{MSE}(V, V_r) = \frac{1}{|\Omega|} \sum_{x \in \Omega} \text{MSE}(P(x), P_r(x)),$$

$$\text{DSSIM}(V, V_r) = \frac{1}{|\Omega|} \sum_{x \in \Omega} 1 - \text{SSIM}(P(x), P_r(x)),$$

где  $V$  и  $V_r$  обозначают видео, содержащее результат восстановления, и видео с эталонным фоном соответственно.  $P(x)$  и  $P_r(x)$  – это блоки результата восстановления и эталона соответственно, с центром в пикселе  $x$  и размером  $n \times n \times 1$  (то есть, двумерные блоки имеющие пространственные размеры  $n \times n$  и продолжительность в 1 кадр, в рамках данной работы  $n$  предполагается равным 9 пикселям). MSE – среднеквадратичная разница между соответствующими блоками, SSIM – индекс структурного сходства между блоками [18]. В рамках данной работы все показатели, основанные на использовании 2D-блоков, вычисляются для яркостной компоненты видеопоследовательности, использование цветовой компонент не приводило к значимому повышению корреляции с экспертной оценкой в проведенных экспериментах.  $\Omega$  обозначает пространственно-временной регион видео, полученный из исходной области восстановления  $\Omega_s$  путем ее расширения на  $n/2$  пикселей в пространственных направлениях, чтобы все блоки  $P(x)$ , содержащие хотя бы один пиксел из  $\Omega_s$ , были целиком включены в  $\Omega$ :  $\Omega = \{x \mid P(x) \cap \Omega_s \neq \emptyset\}$ .

Однако рассмотренные выше традиционные методы обладают рядом недостатков в контексте оценки качества восстановления фона в видео (см. рис. 2). Далее данные недостатки будут рассмотрены и будут приведены возможные способы их устранения. Во-первых, внутри восстановленной области могут присутствовать сдвиги относительно эталона, к которым показатели MSE и DSSIM имеют высокую чувствительность. Наиболее простой способ устранения данного ограничения – это проведение анализа сразу на нескольких масштабах:

$$\text{MSMSE}(V, V_r) = \sum_{i=0}^{M-1} w_i^{\text{MSE}} \text{MSE}(V^i, V_r^i),$$

$$\text{MSDSSIM}(V, V_r) = \sum_{i=0}^{M-1} w_i^{\text{DSSIM}} \text{DSSIM}(V^i, V_r^i).$$

Верхний индекс  $i$  обозначает уровень Гауссовой пирамиды. То есть,  $V_r^0$  – это исходное эталонное видео,  $V_r^1$  – это то же самое видео, но уменьшенное в два раза по обоим пространственным направлениям (продолжительность видеопоследовательности не изменяется).  $M$  – константа, обозначающая общее количество уровней рассматриваемой Гауссовой пирамиды,  $w_i$  – веса соответствующих уровней пирамиды. Присвоение больших весов более высоким уровням пирамиды увеличивает устойчивость метода к мелким сдвигам между результатом восстановления и эталоном. Конкретные значения весов  $w_i$  вычисляются на основе проведенной

экспертной оценки различных алгоритмов восстановления фона.

Человеческое зрение, как правило, более чувствительно к нестабильным во времени искажениям, однако рассмотренные до сих пор методы не пригодны для измерения такого вида искажений. Для устранения этой проблемы предлагаются следующие показатели, явным образом учитывающие стабильность во времени:

$$\text{MSEdt}(V, V_r) = \frac{1}{|\Omega|} \sum_{x \in \Omega} \max \left( \begin{array}{l} \text{MSE}(P(x), P(x + s_x)) - \\ - \text{MSE}(P_r(x), P_r(x + s_x)), 0 \end{array} \right),$$

$$\text{DSSIMdt}(V, V_r) = \frac{1}{|\Omega|} \sum_{x \in \Omega} \max \left( \begin{array}{l} \text{SSIM}(P_r(x), P_r(x + s_x)) - \\ - \text{SSIM}(P(x), P(x + s_x)), 0 \end{array} \right),$$

где  $s_x = (v_x, v_y, -1)$  обозначает вектор движения от текущего кадра к предыдущему на эталонной видеопоследовательности в пикселе  $x$ . Было опробовано несколько алгоритмов оптического потока для вычисления векторов движения на эталонной видеопоследовательности. Однако наилучший результат был получен при использовании алгоритма PatchMatch [19] с радиусом поиска ограниченным до  $1/20$  ширины кадра и с использованием блоков  $9 \times 9$  пикселей. Такой подход обнаруживает нестабильность результата восстановления вдоль векторов движения эталона. Однако это неявно предполагает полную выравненность результата восстановления и эталона. Для смягчения этого требования вводится одновременный учет нескольких масштабов аналогично MSMSE и MSDSSIM:

$$\text{MSMSEdt}(V, V_r) = \sum_{i=0}^{M-1} w_i^{\text{MSEdt}} \text{MSEdt}(V^i, V_r^i),$$

$$\text{MSDSSIMdt}(V, V_r) = \sum_{i=0}^{M-1} w_i^{\text{DSSIMdt}} \text{DSSIMdt}(V^i, V_r^i).$$

В целом, устойчивость рассмотренных показателей к сдвигам между результатом восстановления и эталоном ограничивается лишь пространственными сдвигами небольшой величины. Для повышения устойчивости к сдвигам возможно применение более сложных методов, которые явным образом находят сдвиги, минимизирующие разницу между результатом восстановления и эталоном, причем поиск может не ограничиваться текущим кадром, а включать в себя нахождение наиболее похожего блока во всей эталонной видеопоследовательности. Более конкретно, для каждого блока результата восстановления находится наиболее похожий блок в эталонном видео и расстояния между ними суммируются:

$$C_{\#}^{3D}(V, V_r) = \frac{1}{|\Omega|} \sum_{x \in \Omega} \min_y \#(Q(x), Q_r(y)),$$

$$C_{\#}(V, V_r) = \frac{1}{|\Omega|} \sum_{x \in \Omega} \min_y \#(P(x), P_r(y)).$$

Разница между этими двумя показателями заключается в том, что  $C_{\#}^{3D}$  использует пространственно-временные 3D-блоки, обозначаемые как  $Q(x)$  (используются  $5 \times 5 \times 5$  RGB блоки в соответствии с [1]), а не традиционные 2D-блоки  $P(x)$ .  $\Omega'$ , соответственно, обозначает область восстановления, расширенную на 3 пик-

села во всех направлениях, включая временное (это гарантирует включение всех 3D-блоков, содержащих хотя бы 1 пиксел области восстановления). # – это функция, которая используется для вычисления степени сходства блоков (или, другими словами, расстояния между блоками).

Использование пространственно-временных блоков позволяет также учитывать сходство характера движения сравниваемых областей видеопоследовательности. Авторы работы [1] предлагают модифицированный метод измерения сходства пространственно-временных блоков, который позволил улучшить визуальное качество восстановленного фона в рамках предложенного ими алгоритма:

$$TMSE(Q(x), Q_r(y)) = \frac{1}{N} (\|Q(x) - Q_r(y)\|_2^2 + \lambda \|T(x) - T_r(y)\|_2^2),$$

где  $T(x)$  и  $T_r(y)$  – это 3D-блоки текстурных признаков, определенных в [1], из результата восстановления и эталона соответственно,  $N$  – количество пикселей в блоке.  $TMSE$ , наряду с простой среднеквадратичной разницей  $MSE$ , используется как мера сходства блоков в  $C_{\#}^{3D}$ .  $MSE$  и  $DSSIM$  используются как меры сходства блоков в  $C_{\#}$ .

Таким образом, можно определить многомасштабные показатели  $C_{MSE}^{MS3D}$ ,  $C_{TMSE}^{MS3D}$ ,  $C_{MSE}^{MS}$ ,  $C_{DSSIM}^{MS}$  аналогично определению  $MSMSE$ . Использование результатов на разных масштабах в рамках подходов, основанных на нахождении наиболее похожего блока во всем эталонном видео, позволяет отдельно учитывать искажения разного масштаба, которые, как правило, имеют разную заметность для зрительной системы человека [20].

В то время как  $C_{\#}^{3D}$  позволяет учитывать сходство движения между результатом восстановления и эталонном за счет использования 3D-блоков,  $C_{\#}$  учитывает лишь пространственные искажения. Для устранения этого ограничения предлагается измерить, насколько расстояние до наиболее похожего блока в эталонном видео изменяется от кадра к кадру. Более конкретно, для каждого блока на текущем кадре результата восстановления фона находится наиболее похожий блок на предыдущем кадре в рамках окна ограниченных размеров, а затем проводится сравнение расстояния до наиболее похожего блока эталона между ними:

$$C_{\#dt}(V, V_r) = \frac{1}{|\Omega|} \sum_{x \in \Omega} \left| \min_y (\#(P(x), P_r(y))) - \min_y (\#(P(x_{prev}), P_r(y))) \right|,$$

$$x_{prev} = \underset{y \in \Omega_{prev}^{w \times w}(x)}{\operatorname{argmin}} \#(P(x), P_r(y)), \# = DSSIM, MSE.$$

$\Omega_{prev}^{w \times w}(x)$  – это окно размера  $w \times w$  пикселей (в данной работе используется  $w$  равное 1/10 ширины кадра), которое пространственно центрировано по пикселу  $x$ , но находится на предыдущем кадре по отношению к нему. Версии, учитывающие результаты на разных масштабах,  $C_{MSEdt}^{MS}$  и  $C_{DSSIMdt}^{MS}$  определяются аналогично  $MSMSEdt$ . В отличие от показателя  $MSMSEdt$ , который позволяет оценивать стабильность результата восста-

новления вдоль векторов движения эталонного видео (это предполагает необходимость полного соответствия движения в результате восстановления и движения в эталоне, что может быть слишком сильным предположением в ряде случаев), данные оценки измеряют то, насколько искажения блоков результата восстановления варьируются от кадра к кадру.

Точное вычисление показателей, основанных на нахождении наиболее похожего блока в эталонном видео, затруднено чрезмерно высокой вычислительной сложностью полного перебора всех блоков эталонной видеопоследовательности. Гораздо более эффективным образом можно вычислить приближительные оценки с помощью алгоритма PatchMatch [19]. В случае 3D-блоков данный алгоритм применяется в полной аналогии с работой [1]. Для нахождения наиболее похожего 2D-блоков алгоритм PatchMatch применяется в покадровом режиме, результат предыдущего кадра используется как начальное приближение в текущем для ускорения сходимости. Также, для сравнения блоков всегда используется метрика SSD (сумма квадратов разностей) в цветовом пространстве RGB. Явная минимизация индекса SSIM в проведенных экспериментах не приводила к существенному повышению корреляции с результатами экспертной оценки, в то время как ее использование многократно увеличивает вычислительную сложность алгоритма оценки качества восстановления.

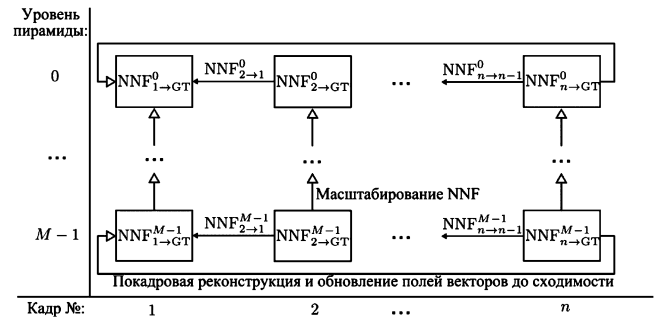


Рис. 3. Общая схема простого жадного алгоритма, используемого для совместной минимизации показателей  $C_{MSE}^{MS}$  и  $C_{MSEdt}^{MS}$  на базе предоставленного начального приближения. В ходе алгоритма итеративно вычисляются межкадровые поля векторов  $NNF_{t \rightarrow (t-1)}^{M-1}$ , соединяющие наиболее похожие блоки кадров с номерами  $t$  и  $t-1$  на нижнем  $(M-1)$ -ом уровне Гауссовой пирамиды, и поля векторов до наиболее похожих блоков эталона  $NNF_{t \rightarrow GT}^{M-1}$  для каждого кадра с помощью алгоритма PatchMatch [19]. На их основе производится покадровая реконструкция. Результат реконструкции используется для вычисления обновленных полей векторов, которые, в свою очередь, используются для проведения следующей итерации реконструкции. После окончания обработки текущего уровня пирамиды происходит переход на следующий уровень, где процесс повторяется

### Проведение экспертной оценки качества

Немногие работы в области восстановления фона в видео предоставляют исходный код предложенных алгоритмов. Поэтому для увеличения количества данных для проведения экспертной оценки качества работы различных алгоритмов и измерения корреляции пред-



ложенных объективных показателей с визуальным качеством в данной работе также оцениваются коммерческие инструменты восстановления фона в видео и алгоритмы восстановления фона в изображениях. Итого в сравнении приняло участие 6 алгоритмов:

- Video Inpainting of Complex Scenes [1].
- Алгоритм BGR, описанный в работе [22].
- Инструмент F\_RigRemoval из программного пакета Nuke [21].
- Инструмент Remove Rig из программного пакета PFClean [23].
- Простой алгоритм восстановления для изображений Telea Inpainting [24].
- Более сложный алгоритм восстановления для изображений Image Completion using Planar Structure Guidance [25].

Для дополнительного расширения используемого набора алгоритмов также добавляется ряд синтетических результатов, получаемых путем прямой совместной минимизации показателей  $C_{MSE}^{MS}$  и  $C_{MSEdt}^{MS}$ . Это является дополнительным способом проверки того, что более низкие значения  $C_{MSE}^{MS}$  и  $C_{MSEdt}^{MS}$  соответствуют более высокому визуальному качеству. Это не очевидно, так как подход, на котором основаны эти показатели, явным образом не использует сходство движения между результатом восстановления и эталоном. Для достижения такой минимизации используется модификация алгоритма реконструкции, предложенного в работе [1]. Общая схема алгоритма представлена на рис. 3. Основная модификация заключается в использовании двух полей векторов в процессе реконструкции – поля  $NNF_{t \rightarrow (t-1)}$ , которое соединяет наиболее похожие блоки между текущим и предыдущим кадрами результата реконструкции, и поля  $NNF_{t \rightarrow GT}$ , которое соединяет блоки текущего кадра результата реконструкции с наиболее похожими блоками во всем эталонном видео. Тогда, одна итерация реконструкции пиксела  $x$  на текущем кадре результата восстановления  $V$  имеет следующий вид:

$$V[x] = \frac{\left( (1-\tau) \sum_{y \in P(x)} w^s(y) V_r[x + NNF_{t \rightarrow GT}[y]] + \tau \sum_{y \in P(x)} w^f(y) V[x + NNF_{t \rightarrow (t-1)}[y]] \right)}{\left( (1-\tau) \sum_{y \in P(x)} w^s(y) + \tau \sum_{y \in P(x)} w^f(y) \right)},$$

где  $\tau$  – это константа, обозначающая вес темпоральной компоненты, которая использует результат реконструкции предыдущего кадра для реконструкции текущего (в данной работе используется  $\tau = 0,4$ ).  $w^s(y)$  и  $w^f(y)$  обозначают весовые функции, которые присваивают большие веса тем векторам полей  $NNF_{t \rightarrow GT}$  и  $NNF_{t \rightarrow (t-1)}$ , соответственно, которые соединяют более похожие блоки (с меньшим значением функции расстояния SSD между блоками). Используются экспоненциальные весовые функции по аналогии с работой [1].

Такой подход позволяет находить локальные оптимумы, которые могут зависеть от начального приближения. Результаты всех 6 алгоритмов использовались в

качестве начальных приближений: во всех случаях предложенный подход позволил существенно уменьшить значения обоих показателей (рис. 4). Добавление таким образом полученных синтетических результатов удваивает количество доступных данных.

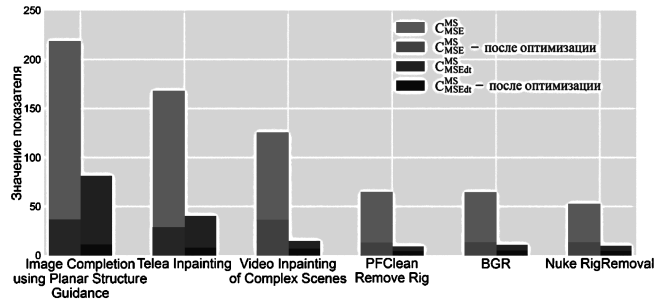


Рис. 4. Значения показателей  $C_{MSE}^{MS}$  и  $C_{MSEdt}^{MS}$

до и после предложенного алгоритма оптимизации, усредненные по всем тестовым видеопоследовательностям

Для количественной оценки визуального качества было проведено исследование, в рамках которого участники попарно сравнивали результаты различных алгоритмов и выбирали результат с наилучшим, по их мнению, качеством. Каждый участник проводил сравнение 28 пар результатов восстановления фона, которые включали 3 контрольных вопроса, где требовалось сравнить эталонный результат с результатом заведомо низкокачественного алгоритма. Для успешного прохождения требовалось ответить на все вопросы, включая правильные ответы на все контрольные вопросы. Общее количество пар, требующих сравнения, составило 2964. Это число составлено из 6 оригинальных алгоритмов, 6 соответствующих синтетических алгоритмов, состоящих в применении вышеописанной процедуры совместной минимизации показателей  $C_{MSE}^{MS}$  и  $C_{MSEdt}^{MS}$  к результатам оригинальных алгоритмов, и 19 видеопоследовательностей, на которых производилось их сравнение (использовались различные фрагменты фиксированного размера из исходных 7 видео). Итого было собрано 8533 попарных сравнений, произведенных 341 участником, которые затем были преобразованы в субъективные ранги с помощью модели Тёрстоуна [26] как для каждой последовательности по отдельности, так и для всех последовательностей сразу (рис. 5).

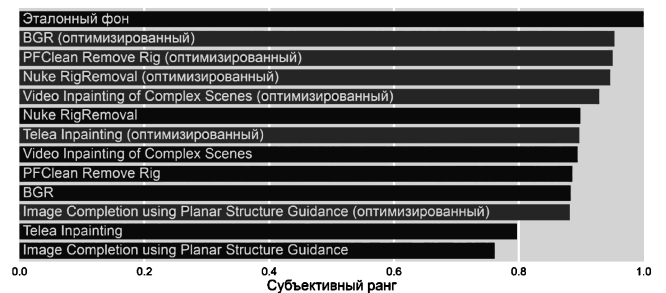


Рис. 5. Общее ранжирование протестированных алгоритмов восстановления фона в видео, вычисленное на базе попарных сравнений, проведенных более чем 300 респондентами

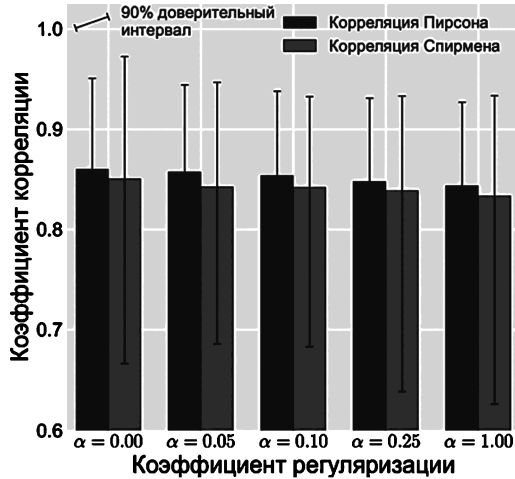
### Результаты исследований

Для получения оптимальных весов в показателях, включающих результаты сразу на нескольких масшта-

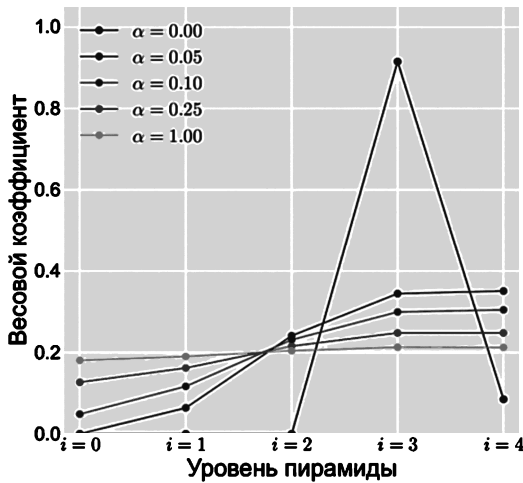
бах, производится максимизация корреляции с субъективными рангами. Сумма весов фиксируется равной единице, и используется регуляризация Тихонова:

$$w = \operatorname{argmax}_{w=[w_0, \dots, w_s]} \sum_{s=0}^S \operatorname{corr}(-\log(wM_s), r_s) - \alpha \|w_2\|,$$

где  $M_s = \{m_{ij}^s\}, m_{ij}^s$  – значение показателя для  $i$ -ого уровня пирамиды,  $j$ -ого алгоритма и  $s$ -ой видеопоследовательности ( $S$  – общее количество последовательностей),  $r_s$  – соответствующий набор субъективных рангов. Эффект регуляризации проиллюстрирован на рис. 6.



а) Корреляция показателя MSDSSIM с субъективными рангами



б) Распределение весов между разными уровнями пирамиды (масштабами) для показателя MSDSSIM

Рис. 6. Иллюстрация эффекта регуляризации при выборе оптимальных весов в ходе максимизации корреляции с субъективными рангами

Она повышает надежность соответствующих показателей за счет покрытия большего числа масштабов, но ценой небольшой потери в значениях корреляции с субъективными рангами. Более корректным способом выбора весов является явное использование синтетических примеров с искажениями на разных масштабах, как это, например, делается авторами метода MSSSIM [20]. Однако такой подход является слишком трудозатратным в рамках рассматриваемой задачи, поэтому итоговые веса получаются простой максимизацией корреляции на всех имеющихся данных с коэффициентом

регуляризации  $\alpha = 0,1$ . Результирующие значения корреляции показаны на рис. 7.

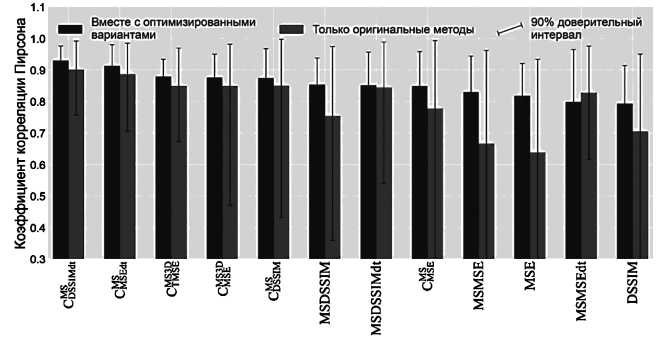


Рис. 7. Итоговые корреляции рассмотренных объективных показателей качества восстановления фона с субъективными рангами

Оценка корреляции производилась на двух различных наборах данных: полном наборе, включающем синтетические результаты, полученные путем предложенной процедуры оптимизации, и наборе, включающем только результаты 6 оригинальных алгоритмов восстановления фона.

В первую очередь, стоит отметить, что практически все объективные методы оценки качества показывают худший результат на наборе данных, включающем только 6 оригинальных алгоритмов, особенно показатели, которые базируются на прямом сравнении с эталоном (MSE, DSSIM, MSMSE, MSDSSIM). Это можно объяснить тем, что предложенная процедура оптимизации уменьшает различия между результатом восстановления и эталоном (что неудивительно, т.к. эталонный результат в процессе оптимизации напрямую используется) и в то же время увеличивает субъективные ранги (рис. 4). Сочетание этих двух факторов естественным образом приводит к повышению корреляции таких простых показателей как MSE и DSSIM на полном наборе данных. Независимо от используемого набора данных подходы, которые базируются на индексе структурного сходства (SSIM), стабильно показывают лучший результат по сравнению с подходами, использующими среднеквадратичную ошибку (MSE). Также, оценки временной стабильности, как правило, показывают более высокую корреляцию с субъективными рангами по сравнению с оценками пространственной ошибки, что и следовало ожидать. С другой стороны, показатели  $C_{MSEdt}^{MS}$  и  $C_{DSSIMdt}^{MS}$  показывают на удивление высокий результат, учитывая, что явным образом сходство движения между результатом восстановления и эталоном в них не используется. Оценки, использующие 3D-блоки, также показывают неплохую корреляцию с субъективными рангами, но они имеют более высокую вычислительную сложность и большое потребление памяти по сравнению с подходами, основанными на использовании 2D-блоков. Таким образом, следующие показатели были выбраны для объективной оценки качества восстановления фона в видео:

$$C_{DSSIMdt}^{MS}, \{w_i^{CDSSIMdt}\} = [0.00; 0.08; 0.25; 0.30; 0.37],$$

$$C_{DSSIM}^{MS}, \{w_i^{CDSSIM}\} = [0.04; 0.11; 0.21; 0.29; 0.35],$$

$$MSDSSIMdt, \{w_i^{DSSIMdt}\} = [0.00; 0.00; 0.30; 0.32; 0.38],$$

MSDSSIM,  $\{w_i^{DSSIM}\} = [0.05; 0.12; 0.23; 0.30; 0.30]$ .

## Заключение

В данной работе представлена методика объективной оценки качества восстановления фона в видеопоследовательностях, состоящая из четырех показателей, которые превзошли традиционные подходы к оценке качества видео. Это продемонстрировано на результатах экспертной оценки 6 алгоритмов восстановления фона на 7 построенных тестовых видеопоследовательностях с эталонным фоном, проведенной с участием более 300 респондентов. Предложенные показатели позволяют проводить объективное сравнение различных алгоритмов восстановления фона даже в случаях крупных и продолжительных по времени областей восстановления (например, удаление объекта из видео), что представляло серьезные проблемы для предшествующих методик количественного сравнения качества восстановления. Предложенные объективные оценки существенно различаются по корреляции с результатами экспертной оценки, однако имеют отдельные интуитивные интерпретации и позволяют более точно оценить слабые и сильные стороны того или иного алгоритма восстановления фона в видео.

Исследование выполнено при поддержке РФФИ в рамках научного проекта 15-01-08632 а.

## Литература

1. Newson A., Almansa A., Fradet M., Gousseau Y. and Pérez P., «Video inpainting of complex scenes», *SIAM Journal on Imaging Sciences*, pp. 1993–2019, 2014.
2. Granados M., Tompkin J., Kim K., Grau O., Kautz J., and C. Theobalt, «How not to be seen-object removal from videos of crowded scenes», *Computer Graphics Forum*, volume 31, – pp. 219–228, 2012.
3. Ebdelli M., Meur O. Le and Guillemot C. «Video inpainting with short-term windows: application to object removal and error concealment», *IEEE Transactions on Image Processing*, – pp. 3034–3047, 2015.
4. Chen Y., Hu Y., Au O.C., Li H. and Chen C.W. «Video error concealment using spatio-temporal boundary matching and partial differential equation», *IEEE Transactions on Multimedia*, pp. 2-15, 2008.
5. Koloda J., Ostergaard J., Jensen S.H., Peinado A.M. and Sanchez V., «Sequential error concealment for video/images by weighted template matching», *IEEE Data Compression Conference*, – pp. 159–168, 2012.
6. Koloda J., Ostergaard J., Jensen S.H., Sanchez V. and Peinado A.M. «Sequential error concealment for video/images by sparse linear prediction», *IEEE Transactions on Multimedia*, pp. 957–969, 2013.
7. Erofeev M and Vatolin D. «Automatic logo removal for semitransparent and animated logos», *Proceedings of Graph-Con 2011*, pp. 26–30, 2011.
8. Yan W.Q., Wang J. and Kankanhalli M.S. «Automatic video logo detection and removal», *Multimedia Systems*, pp. 379-391, – 2005.
9. Mosleh A., Bouguila N. and Hamza A.B. «Automatic

inpainting scheme for video text detection and removal», *IEEE Transactions on Image Processing*, pp. 4460-4472, 2013.

10. Shiratori T., Matsushita Y., Tang X. and Kang S.B. «Video completion by motion field transfer», *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 411–418, 2006.

11. Hu W., Tao D., Zhang W., Xie Y. and Yan Y. «A new low-rank tensor model for video completion», *arXiv preprint arXiv:1509.02027*, 2015.

12. Mosleh A., Bouguila N. and Hamza A.B. «Video completion using bandlet transform», *IEEE Transactions on Multimedia*, pp. 1591–1601, 2012.

13. Mosleh A., Bouguila N. and Hamza A.B. «Bandlet-based sparsity regularization in video inpainting», *Journal of Visual Communication and Image Representation*, pp. 855–863, 2014.

14. You S., Tan R.T., Kawakami R. and Ikeuchi K. «Robust and fast motion estimation for video completion», *IAPR International Conference on Machine Vision Applications (MVA)*, – pp. 181–184, 2013.

15. Benoit J. and Paquette E. «Localized search for high definition video completion», *Journal of WSCG*, pp. 45–54, 2015.

16. Ilan S. and Shamir A. «A survey on data-driven video completion», *Computer Graphics Forum*, pp. 60–85, 2015.

17. Blender. <https://www.blender.org/>.

18. Wang Z., Bovik A.C., Sheikh H.R. and Simoncelli E.P. «Image quality assessment: from error visibility to structural similarity», *IEEE Transactions on Image Processing*, pp. 600–612, 2004.

19. Barnes C., Shechtman E., Finkelstein A. and Goldman D. «Patchmatch: A randomized correspondence algorithm for structural image editing», *ACM Transactions on Graphics (TOG)*, 2009.

20. Wang Z., Simoncelli E.P. and Bovik A.C. «Multiscale structural similarity for image quality assessment», *Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers*, pp. 1398–1402, 2003.

21. The Foundry Nuke. <https://www.thefoundry.co.uk/products/nuke/>.

22. Зачесов А.А., Ерофеев М.В., Ватолин Д.С. «Использование карт глубины при восстановлении фона в видеопоследовательностях», *Новые информационные технологии в автоматизированных системах: материалы научно-практического семинара*, 2015.

23. Pixel Farm PFClean. <http://www.thepixelfarm.co.uk/pfclean/>.

24. Telea A. «An image inpainting technique based on the fast marching method», *Journal of graphics tools*, pp. 23–34, 2004.

25. Huang J.-B., Kang S.B., Ahuja N. and Kopf J. «Image completion using planar structure guidance», *ACM Transactions on Graphics (TOG)*, 2014.

26. Thurstone L.L. «A law of comparative judgment», *Psychological review*, 1927.

27. Herling J. and Broll W. «High-quality real-time video inpainting with PixMix», *IEEE Transactions on Visualization and Computer Graphics*, pp. 866–879, 2014.

28. Patwardhan K. and Sapiro G. «Video inpainting under constrained camera motion», *IEEE Transactions on Image Processing*, pp. 545–553, 2007.

## АЛГОРИТМ ФОРМИРОВАНИЯ ТЕПЛООВОГО ИЗОБРАЖЕНИЯ ОБЪЕКТОВ ПРИ РАДИОМЕТРИЧЕСКОМ НАБЛЮДЕНИИ

*Клочко В.К., д.т.н., зав. кафедрой Рязанского государственного радиотехнического университета (РГРТУ), e-mail: klochkovk@mail.ru;*

*Макарова О.Н., инженер РГРТУ;*

*Гудков С.М., инженер-конструктор 2 кат. Рязанской радиоэлектронной компании (РРК);*

*Коселев А.А., ведущий инженер-программист РРК.*

## ALGORITHM OF THE OBJECTS THERMAL IMAGE FORMATION AT RADIOMETRIC OBSERVATION

*Klochko V.K., Makarova O.N., Gudkov S.M., Koshelev A.A.*

*The algorithm of joint processing of radiometric and optical images allowing to receive the radio thermal image of objects with the spatial resolution corresponding to the size of the required image element of sampling is offered. The results of pilot study showing a possibility of formation of the thermal objects image on the area by means of the offered algorithm are given.*

**Key words:** radiometer, radiometric image, optical image, resolution.

**Ключевые слова:** радиометр, радиометрическое изображение, оптическое изображение, разрешающая способность.

### Введение

Современные системы пассивного наблюдения объектов на местности с целью формирования их теплового изображения активно используют тепловизоры [1]. В тепловизоре оптические линзы проецируют поток электромагнитного излучения инфракрасного диапазона длин волн на матрицу светочувствительных элементов. В результате получается матрица амплитудного (теплового) изображения участка местности, в которой амплитуды, пропорциональные значениям температуры, переводятся в уровни цветности. Данная матрица имеет пространственное разрешение, близкое к разрешению оптического изображения. Однако в инфракрасном диапазоне электромагнитная волна обладает меньшей проникающей способностью, чем волна миллиметрового радиолокационного диапазона, используемая в сканирующих радиометрах [2]. В отличие от тепловизоров, принимающих излучение поверхностного слоя земной поверхности, радиометры позволяют фиксировать радиотепловое излучение объектов, скрытых под поверхностным слоем [3], и в меньшей степени подвержены атмосферным влияниям. Однако пространственная разрешающая способность радиометрических систем значительно уступает разрешающей способности тепловизионных и оптических систем.

Целью работы является разработка алгоритма формирования теплового изображения объектов на местности в радиолокационном миллиметровом диапазоне длин волн с повышенным пространственным разрешением.

Цель достигается за счет совместной обработки радиометрического и оптического изображений. Для разработки алгоритма используются следующие модели изображений.

*Предложен алгоритм совместной обработки радиометрических и оптических изображений, позволяющий получать радиотепловое изображение объектов с пространственным разрешением, соответствующим размеру элемента дискретизации искомого изображения. Приводятся результаты экспериментального исследования, показывающие возможность формирования теплового изображения объектов на местности с помощью предложенного алгоритма.*

### Модель радиометрического изображения

Антенна радиометра построчно сканирует зону обзора. Съём данных в каждой строке производится с шагом дискретизации  $\Delta\varphi$  по азимуту (по  $j$ ) и переход к другой строке осуществляется с шагом  $\Delta\theta$  по углу места (по  $i$ ). Принятый на каждом  $i$ -м,  $j$ -м шаге сканирования сигнал после прохождения тракта первичной обработки регистрируется в цифровой форме величиной  $y(i, j)$ , подчиненной модели радиометрического изображения (РИ):

$$y(i, j) = \mu \sum_{i_1=-m}^m \sum_{j_1=-n}^n \alpha(i_1, j_1) x(i+i_1, j+j_1) + p(i, j), \quad (1)$$

$$i = \overline{1, M}, \quad j = \overline{1, N},$$

где  $Y = \{y(i, j)\}$  –  $M \times N$ -матрица РИ области  $D$  обзора;  $\mu$  – коэффициент радиометрического преобразования;  $\alpha(i, j)$  – аппаратная функция (АФ) радиометра в  $i$ -х,  $j$ -х элементах дискретизации угла места  $\theta_i$  и азимута  $\varphi_j$ , учитывающая действие диаграммы направленности антенны (ДНА) и тракта первичной обработки радиометра на входной сигнал;  $x(i, j)$  – элементы искомого теплового изображения, представляющие интенсивность электромагнитного поля излучения в  $i$ -м,  $j$ -м угловом направлении в температурной шкале;  $p(i, j)$  – шумы аппаратуры;  $2m+1$  и  $2n+1$  – соответственно ширина ДНА по углу места и азимуту в количестве элементов дискретизации; числа  $M$  и  $N$  определяют размеры области  $D$  в

количестве строк и столбцов искомой матрицы  $X = \{x(i, j)\}$ ,  $i = \overline{1, M}$ ,  $j = \overline{1, N}$ .

Задача заключается в оценивании матрицы  $X$  по наблюдениям (1).

В радиометрической системе с целью повышения разрешающей способности изображения по угловым координатам матрица  $Y$  может подвергаться операциям восстановления изображения с помощью оператора восстановления  $R_1$  в пространственной или частотной областях, то есть оценивания ненаблюдаемых величин  $x(i, j)$  на основе наблюдений  $y(i, j)$  и известной АФ  $A = \{\alpha(i, j)\}$ :  $X_1 = R_1[Y, A]$ . Результатом восстановления изображения области  $D$  будет  $M \times N$ -матрица  $Y_1 = \{y_1(i, j)\}$  оценок  $y_1(i, j)$  искомых величин  $x(i, j)$ :

$$y_1(i, j) = \mu_1 \sum_{i_1=-m_1}^{m_1} \sum_{j_1=-n_1}^{n_1} \beta(i_1, j_1) x(i+i_1, j+j_1) + \varepsilon(i, j), \quad (2)$$

$$i = \overline{1, M}, \quad j = \overline{1, N}, \quad m_1 < m, \quad n_1 < n,$$

где  $\mu_1$  – коэффициент остаточного радиометрического преобразования;  $\beta(i, j)$  – функция рассеяния точки (ФРТ), описывающая остаточное искажение  $x(i, j)$  в  $(2m_1 + 1) \times (2n_1 + 1)$ -окрестности  $i$ -го,  $j$ -го элемента матрицы  $Y_1$  в силу ограниченной точности восстановления при низком контрасте РИ  $Y$ ;  $\varepsilon(i, j)$  – остаточные шумы.

Задача заключается в оценивании матрицы  $X$  по наблюдениям (2).

### Модель оптического изображения

Оптическое изображение (ОИ) области  $D$ , полученное с помощью фото или видео камеры с учетом геометрических совмещений, дает  $M_1 \times N_1$ -матрицу  $X_1 = \{x_1(i, j)\}$ , элементы  $x_1(i, j)$  которой удовлетворяют модели:

$$x_1(i, j) = \mu_0 x_0(i, j) + \xi(i, j), \quad i = \overline{1, M_1}, \quad j = \overline{1, N_1}, \quad (3)$$

где  $\mu_0$  – коэффициент оптического преобразования;  $x_0(i, j)$  – амплитуда (интенсивность) оптического излучения в  $i$ -м,  $j$ -м угловом направлении;  $\xi(i, j)$  – шум аппаратуры;  $M_1$  и  $N_1$  – размеры матрицы ОИ, в  $k$  раз превышающие размеры  $M$  и  $N$ -матрицы РИ за счет меньшего в  $k$  раз элемента дискретизации:  $M_1 = k \cdot M$ ,  $N_1 = k \cdot N$ ,  $k$  – масштабный множитель.

### Алгоритм совместной обработки РИ и ОИ

Алгоритм сводится к следующим операциям:

1. В результате сканирования антенной радиометра зоны обзора формируется матрица РИ  $Y$  модели (1).

2. Матрица  $Y$  может подвергаться операциям восстановления с помощью оператора восстановления  $R_1$  [4, 5]. В этом случае формируется матрица восстановленного РИ  $Y_1 = R_1[Y, A]$  модели (2).

3. С помощью оптической системы получается матрица ОИ  $X_1$  модели (3).

4. Матрица  $X_1$  приводится в соответствие масштабу матрицы  $Y$  или  $Y_1$  пересчетом значений ее элементов по формуле:

$$x_2(i, j) = (1/k^2) \sum_{i_1=1}^k \sum_{j_1=1}^k x_1(k(i-1)+i_1, k(j-1)+j_1),$$

$$i = \overline{1, M}, \quad j = \overline{1, N}.$$

Получается матрица ОИ  $X_2$  с разрешением искомой матрицы  $X$ .

5. Матрица  $X_2$  разбивается на  $K$  непересекающихся однородных по амплитуде подобластей  $D_1, D_2, \dots, D_K$  с помощью оператора  $R_2$  сегментации [6]:  $S = R_2[X_2]$ . Например, оператором пороговой обработки и сшивания граничных областей с одним уровнем амплитуды [7, 8]. В результате получается матрица меток  $S = \{S(i, j)\}$ ,  $i = \overline{1, M}$ ,  $j = \overline{1, N}$ , где  $S$  – номер сегмента, которому принадлежит  $i$ -й,  $j$ -й элемент матрицы  $X_2$ .

6. Для каждого  $s$ -го сегмента матрицы  $X_2$  вычисляется средняя нормированная радиометрическая амплитуда  $\tilde{y}_2(s)$  усреднением амплитуд соответствующих  $i$ -х,  $j$ -х элементов  $y(i, j)$  матрицы  $Y$  или элементов  $y_1(i, j)$  матрицы  $Y_1$  с теми же метками  $S(i, j)$ :

$$\tilde{y}_2(s) = (1/n_s) \sum_i \sum_j y(i, j), \quad \forall i, j: S(i, j) = s, \quad (4)$$

где  $n_s$  – количество элементов с меткой  $s$ .

7. Амплитуда  $\tilde{y}_2(s)$  присваивается элементам матрицы  $X_2$  с меткой  $S(i, j) = s$ :  $x_2(i, j) = \tilde{y}_2(s)$ ,  $\forall i, j: S(i, j) = s$ , то есть каждой подобласти матрицы  $X_2$  присваивается средняя радиометрическая амплитуда. Получается матрица  $X_2 = \{x_2(i, j)\}$ ,  $i = \overline{1, M}$ ,  $j = \overline{1, N}$ , с пространственным разрешением, равным размерам искомой матрицы теплового изображения, амплитуды элементов которой соответствуют радиометрической температуре  $i$ -х,  $j$ -х элементов местности.

8. Амплитуды элементов  $x_2(i, j)$  полученной матрицы  $X_2$  переводятся в уровни цветности. В результате получается цветное изображение контролируемого участка местности с информацией о температуре излучающих элементов поверхности.

### Замечание

Возможна другая реализация п. 4 алгоритма, а именно:

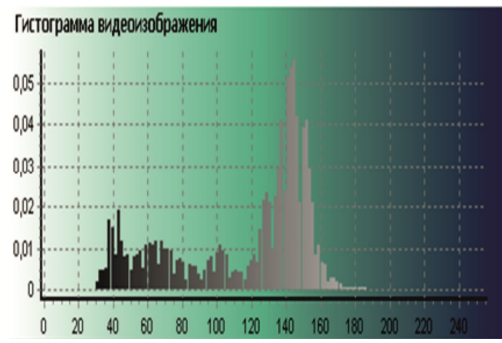
4. Матрица  $Y$  или  $Y_1$  приводится соответствие масштабу матрицы  $X_1$  включением в их состав дополнительных строк и столбцов. Получается матрица  $Y = \{y(i, j)\}$  или  $Y_1 = \{y_1(i, j)\}$ ,  $i = \overline{1, M_1}$ ,  $j = \overline{1, N_1}$ . Элементы дополнительных строк и столбцов вычисляются с помощью линейной интерполяции соответствующих элементов соседних строк и столбцов. При дальнейшей обработке формируется тепловое изображение  $X_2$  с разрешающей способностью оптического изображения  $X_1$ .

### Результаты экспериментального исследования

Натурные испытания проводились с помощью радиометра в 3 мм диапазоне длин волн при наблюдении объектов на местности на расстоянии 100 м. На рис. 1,а показано видео изображение наблюдаемого участка местности, приведенное к масштабу искомого теплового изображения – матрица  $X_2$ . На рис. 1.б – гистограмма амплитудного видео изображения, используемая для определения двух порогов при трехуровневой сегментации. На рис. 1,в – отсегментированное видео изображе-



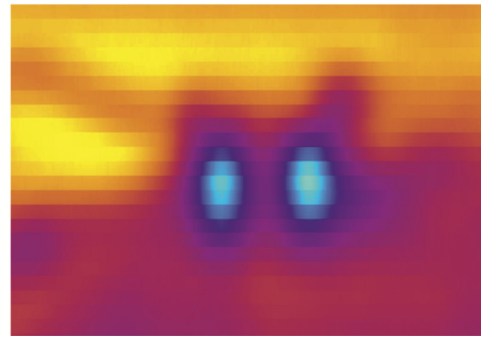
а)



б)



в)



г)



д)

Рис. 1. Результаты экспериментального исследования

ние. На рис. 1,г показано в цвете радиометрическое изображение того же участка местности – матрица  $Y$ . На рис. 1,д – радиометрическое изображение, приведенное к отсегментированному видео изображению в соответствии с предложенным алгоритмом. Видно более четкое тепловое изображение на рис. 1,д по сравнению с тепловым изображением на рис. 1,г.

### Заключение

Предложенный алгоритм, в отличие от существующих алгоритмов формирования тепловых изображений, позволяет наблюдать тепловое изображение объектов на отсегментированном видео изображении, отвечающее радиолокационному миллиметровому диапазону длин волн. Это дает дополнительную информацию о тепловых свойствах объектов в указанном диапазоне.

Проводятся дополнительные исследования, направленные на повышение разрешающей способности тепловых изображений, а также формирование тепловых изображений объектов, скрытых на местности.

### Литература

1. Обработка изображений в геоинформационных сис-

темах: учеб. пособие / В.К. Злобин, В.В. Еремеев, А.Е. Кузнецов. Рязан. гос. радиотехн. ун-т. Рязань, 2006. 264 с.

2. Шарков Е.А. Радиотепловое дистанционное зондирование Земли: физические основы: в 2 т. / Т. 1. М.: ИКИ РАН, 2014. 544 с.

3. Кутуза Б.Г., Данилычев М.В., Яковлев О.И. Спутниковый мониторинг Земли: Микроволновая радиометрия атмосферы и поверхности / М.: Ленард, 2016. 338 с.

4. Василенко Г.И., Тараторин А.М. Восстановление изображений. М.: Радио и связь, 1986. 304 с.

5. Ключко В.К., Кузнецов В.П. Восстановление изображений объектов по прореженной матрице наблюдений // Вестник Рязанского государственного радиотехнического университета. 2016. № 55. С. 111-117.

6. Гонсалес Р., Вудс Р., Эддинс С. Цифровая обработка изображений в среде MATLAB. М.: Техносфера, 2006. 616 с.

7. Математические методы восстановления и обработки изображений в радиотеплооптоэлектронных системах / В.К. Ключко; Ряз. гос. радиотехн. ун-т. Рязань, 2009. – 228 с.

8. Обработка изображений и управление в системах автоматического сопровождения объектов: учеб. пособие / Б.А. Алпатов, П.В. Бабаян, О.Е. Балашов, А.И. Степашкин; Ряз. гос. радиотехн. ун-т. Рязань, 2011. 236 с.

УДК 004.932

## КОРРЕКЦИЯ СМАЗА И РАСФОКУСИРОВКИ СПУТНИКОВЫХ ИЗОБРАЖЕНИЙ С УЧЕТОМ ГЕОМЕТРИЧЕСКИХ ИСКАЖЕНИЙ

*Егошкин Н.А., к.т.н., в.н.с. НИИ «Фотон» РГРТУ, e-mail: foton@rsreu.ru.*

### SATELLITE IMAGES BLURRING AND DEFOCUS CORRECTION IN THE CASE OF GEOMETRIC DISTORTION

*Egoshkin N.A.*

*The problem of speed blurring correction and defocus of satellite images from modern scanning sensors is considered. It is shown that it's need to consider the geometric distortion of images. Speed blurring correction is based on the analytical description of the signal acquisition process. Defocus correction is based on point spread function evaluation using point objects and multi-channel images of the same area. Methods of wavelet transform are used to improve the correction quality and simplify configuration of the algorithms.*

**Key words:** image improvement, defocus correction, blurring correction, remote sensing, point spread function, deconvolution.

**Ключевые слова:** улучшение изображений, коррекция расфокусировки, коррекция смаза, дистанционное зондирование земли, функция рассеяния точки, обратная свёртка.

#### Введение

Современные космические систем дистанционного зондирования Земли (ДЗЗ) используют достаточно сложные механизмы формирования изображений [1]. Излучение от подстилающей поверхности проходит через оптическую систему и регистрируется линейками приборов с зарядовой связью (ПЗС) или матрицами, работающими в режиме временной задержки и накопления сигнала (ВЗН). Развертка изображений осуществляется за счет полета спутника (для высокодетального ДЗЗ), либо с использованием сканирующего зеркала (для систем глобального ДЗЗ). Изображения формируются в виде отдельных частей, так называемых сканов.

Для организации накопления сигнала скорость опроса ВЗН-матрицы должна быть строго согласована со скоростью и направлением бега изображения в фокальной плоскости, в противном случае возникает скоростной смаз, снижающий пространственное разрешение съемки. Для минимизации скоростного смаза при съемке осуществляется маневрирование углами ориентации спутника, частоты опроса для ПЗС-матриц задаются в зависимости от их расположения в фокальной плоскости. Однако полностью избежать скоростного смаза очень трудно, а зачастую принципиально не возможно в пределах всего поля зрения датчика.

В процессе вывода спутника на орбиту неизбежные деформации фокальной плоскости и оптической системы. Иногда предусматривается возможность настройки объектива на орбите, однако возможности такой настройки не беспредельны. В результате изображения оказываются в различной степени расфокусированными, причем степень расфокусировки может зависеть от

*Рассматривается проблема коррекции скоростного смаза и расфокусировки спутниковых изображений от современных сканирующих датчиков. Показано, что необходимо учитывать геометрические искажения снимков. Устранение скоростного смаза основано на аналитическом описании процесса накопления сигнала. Коррекция расфокусировки основана на оценивании функция рассеяния точки по точечным объектам и по многоканальным изображениям одной и той же местности. Для повышения качества коррекции и упрощения настройки алгоритмов используется методы вейвлет-преобразования.*

спектрального канала или оптического модуля. Таким образом, весьма актуальна задача коррекции скоростного смаза и расфокусировки изображений ДЗЗ.

Явления смаза и расфокусировки традиционно описываются сверткой [2] изображений с некоторой функцией рассеяния точки (ФРТ). Соответственно, задача коррекции является известной задачей обратной свертки. Однако для систем ДЗЗ применение известных подходов часто оказывается затруднительным. Во-первых, характер смаза и расфокусировки существенно зависят от режима и геометрии съёмки: при ДЗЗ форма ФРТ может изменяться от изображения к изображению и в поля снимка. Подавляющее большинство методов оценивания ФРТ и обратной свертки требуют постоянства ФРТ. Во-вторых, из-за больших и сверхбольших объемов информации ДЗЗ на практике возможно применение только высокоэффективных методов обратной свертки, «быстрых» преобразований. В третьих, скановая структура изображений осложняет обработку из-за многочисленных краевых эффектов, возникающих на границах сканов.

Что касается коррекции смаза и расфокусировки применительно к системам ДЗЗ, то обычно в литературе рассматриваются случаи незначительной расфокусировки [3-5], этом задача коррекции решается достаточно просто. В практике отечественного ДЗЗ, особенно для новых экспериментальных видеодатчиков, приходится

иметь дело с гораздо более сильными искажениями, при этом задача обратной свертки становится не корректно поставленной. В статье рассматривается подход к коррекции смаза и сильной расфокусировки в системах ДЗЗ с учетом скановой структуры современной съёмочной аппаратуры инвариантно к принципу развертки изображений, как для высокодетальных, так и для глобальных систем наблюдения.

### Модель формирования изображений

Рассмотрим формирование изображения одного скана. Изображение подстилающей поверхности  $S$  подается на вход оптической системы при помощи некоторого механизма развертки (например, сканирующего зеркала). Оптическая система формирует изображение и проецирует его на ВЗН матрицу. В плоскости ВЗН – матрицы введем систему координат  $xu$  так, что ось  $x$  расположена вдоль строки, а ось  $y$  перпендикулярна  $x$  и направлена вдоль переноса заряда в ВЗН, рис. 1.

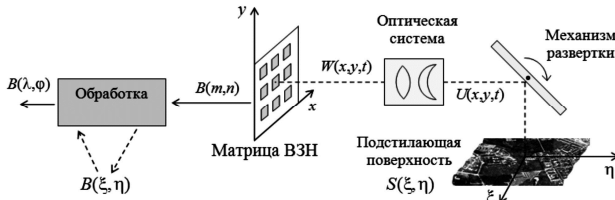


Рис. 1. Обобщенная схема формирования и обработки изображений ДЗЗ

Обозначим  $U(x, y, t)$  – изображение на входе оптической системы в виде функции яркости  $U$  координат  $x, y$  и времени  $t$ . Для низкоорбитальных систем развертка осуществляется за счет полета спутника, при этом изображение на входе оптической системы с некоторой скоростью движется поступательно. Для систем глобального наблюдения сканирование осуществляется вращением зеркала и закон сканирования более сложен [6]. Однако если зеркало плоское, то отражение от него – изометрия, поэтому развертка зеркалом равносильна изменению углов ориентации спутника. Поэтому можно выбрать некоторую криволинейную системы координат  $(\xi, \eta)$  на подстилающей поверхности и записать для обоих типов съёмочных систем:

$$U(x, y, t) = S(\xi, \eta), \quad \xi = x + V_x t, \quad \eta = y + V_y t, \quad (1)$$

где  $V_x, V_y$  – компоненты вектора скорости бега изображения в системе координат  $xu$ .

Оптическая система проецирует на матрицу изображение  $W(x, y, t) = U(x, y, t) \otimes F(x, y)$ , где  $F(x, y)$  – не зависящая от времени ФРТ точки оптической системы.

В процессе опроса матрицы ВЗН формируется изображение  $B(m, n)$ , где  $m, n$  – номера строк и столбцов. Процесс ВЗН можно математически описать в виде

$$B(m, n) = \sum_{v=0}^{N-1} W(\Delta_x n - \Delta_y v, (m - v) / f), \quad \text{где } \Delta_x, \Delta_y -$$

расстояния между фотоприёмными элементами матрицы в соответствующих направлениях;  $f$  – частота опроса строк матрицы,  $N$  – число строк накопления.

Также обычно в ходе последующей обработки материалов ДЗЗ осуществляют геометрическую трансформацию исходного изображения  $B(m, n)$  в некоторую картографическую систему координат  $\lambda, \varphi$ .

Для фиксированного момента времени  $t$  операции развертки и расфокусировки перестановочны. Поэтому с учетом (1) можно записать:

$$B'(\xi, \eta) = G(\xi, \eta) \otimes F(\xi, \eta) \otimes S(\xi, \eta), \quad (2)$$

$$G(\xi, \eta) = \sum_{v=0}^{N-1} \delta(\xi - vV_x / f) \delta(\eta - v(\Delta_y + V_y / f)), \quad (3)$$

$$B(m, n) = B'(\xi, \eta), \quad \xi = \Delta_x n + V_x m / f, \quad \eta = V_y m / f, \quad (4)$$

где  $\delta$  – дельта-функция.

Таким образом, выбор системы координат  $\xi\eta$  таким образом, чтобы выполнялось (1), позволяет описать процессы расфокусировки и скоростного смаза в виде свертки  $S(\xi, \eta)$  с функциями  $F$  и  $G$ . Формулы (4) при этом связываются результат такой свертки и изображение на выходе матрицы. Показательно, что вообще говоря, система координат  $\xi\eta$  не совпадает ни с картографической системой координат  $\lambda\varphi$ , ни с системой координат регистрируемого изображения  $mn$ . Поэтому не оптимально решать задачу обратной свертки непосредственно по изображениям  $B(m, n)$  или  $B(\lambda, \varphi)$ . Например, для систем глобального наблюдения и исходные и обработанные изображения существенно искажены по геометрии [6].

Таким образом, для решения задачи коррекции смаза и расфокусировки необходимо, во-первых, оптимальным образом ввести систему координат  $\xi\eta$ . Во-вторых, определить ФРТ  $F(\xi, \eta)$  и  $G(\xi, \eta)$ . В-третьих, сформировать новое исправное изображение в координатах  $mn$  или  $\lambda\varphi$ .

### Технология коррекции расфокусировки и смаза

Первым делом для конкретной съёмочной аппаратуры на основе (1) вводится оптимальная система координат  $\xi\eta$ . Далее определяются параметры  $V_x, V_y, f$  связи  $\xi\eta$  и  $mn$  в формуле (4). Параметры  $V_x, V_y, f$  вообще говоря, зависят от времени, но обычно меняются либо достаточно медленно, либо кусочно-постоянно в пределах скана. На практике  $V_x, V_y, f$  определяют, исходя из геометрии съёмки с использованием служебной информации и параметров внутреннего и внешнего ориентирования съёмочной аппаратуры [1]. Также для современных систем ДЗЗ имеется возможность уточнения этих параметров по многокановому изображению на основе сопоставления изображений отдельных сканов в областях перекрытия [7].

Далее возможны два подхода. Первый заключается в геометрической трансформации изображения  $B(m, n)$  в оптимальную систему координат:  $B(m, n) \rightarrow B(\xi, \eta)$ .

Для  $B(\xi, \eta)$  расфокусировка не зависит от параметров съёмки и задача коррекции решается более эффективно. В процессе последующей обработки формирование



$B(\lambda, \varphi)$  осуществляется уже по скорректированному  $B(\xi, \eta)$ . На этапе формирования  $B(\xi, \eta)$  также удобно решить задачу совмещения сканов, при этом автоматически решается проблема краевых эффектов, возникающих на границах сканов. Такой подход особенно целесообразен для инфракрасных датчиков, в которых имеется большое количество узких сканов и имеется значительный радиометрический шум. В процессе формирования  $B(\xi, \eta)$  шумы могут быть уменьшены [8], что очень важно из-за чувствительности последующих алгоритмов к шумам. Наконец, формирование изображения  $B(\xi, \eta)$  необходимо, чтобы набрать достаточно статистики для оценивания ФРТ какими либо известными «слепыми» методами.

Второй подход не требует формирования раstra  $B(\xi, \eta)$ . Вместо этого  $B(m, n)$  обрабатывается достаточно маленькими фрагментами, причем для каждого фрагмента ФРТ  $F(\xi, \eta), S(\xi, \eta)$  переводятся в координаты  $m, n$  с использованием (4) и задача обратной свертки решается в плоскости  $mn$ . Такой подход может увеличить скорость обработки и уменьшить затраты памяти, однако  $F(\xi, \eta), S(\xi, \eta)$  должны быть уже известными, их достоверное определение по отдельным фрагментам затруднительно.

Рассмотрим теперь коррекцию смаза и расфокусировки изображения  $B(\xi, \eta)$ . В принципе, для решения этой задачи можно использовать различные методы обратной свертки. Большие объемы информации ДЗЗ требуют применения вычислительно эффективных методов, как правило основанных на переходе в спектральную область с использованием алгоритмов быстрого преобразования Фурье. В спектральной области (2) принимает вид  $B(\omega, \theta) = G(\omega, \theta) F(\omega, \theta) S(\omega, \theta)$ , где  $\omega, \theta$  – частоты Фурье по осям  $\xi, \eta$ , а  $B, G, F, S$  обозначены спектры Фурье соответствующих функций.

При наличии шумов оптимальное восстановление изображения  $S(\omega, \theta)$ , рассматриваемого как стационарный случайный процесс, обеспечивает фильтр Винера [9]:

$$S^*(\omega, \theta) = \frac{B(\omega, \theta) \overline{G(\omega, \theta)} \overline{F(\omega, \theta)}}{(|G(\omega, \theta) F(\omega, \theta)|^2 + \varepsilon(\omega, \theta))}, \quad (5)$$

где  $\varepsilon(\omega, \theta)$  – отношение энергий шума к энергии изображения на частотах  $\omega, \theta$ . Известны также различные модификации фильтра Винера [5, 10, 11], лучше согласованные с особенностями человеческого восприятия скорректированных изображений. На практике обычно  $\varepsilon(\omega, \theta)$  не известно и принимают  $\varepsilon(\omega, \theta) = \varepsilon$ , где  $\varepsilon$  подбирается исходя из наилучшего качества восстановленного изображения.

Основной проблемой при фильтрации (5) является рост шумов на изображении. Реальные изображения не являются стационарными случайными процессами, они обычно имеют участки относительно однородной яркости (например, водная поверхность). На таких участках возросший шум становится хорошо заметен. Это затрудняет подбор параметра  $\varepsilon$ , так как его оптимальное

значение зависит от условий съемки (кол-во строк накопления  $N$ ) и особенностей сюжета.

В ряде работ предлагается для решения задачи обратной свертки использовать методы вейвлет – преобразования [12-14]. Они основаны на разложении изображения по базису вейвлетов, например,  $B(\xi, \eta) =$

$$= \sum_k \sum_j a_{k,j} \varphi_{k,j}(\xi, \eta), \text{ где } \varphi_{k,j}(\xi, \eta) \text{ – базисные функ-}$$

ции,  $k \in Z$  – масштаб,  $j \in Z$  – сдвиг,  $a_{k,j}$  – коэффициенты вейвлет-разложения. Для реальных изображений, содержащих объекты с резкими границами и относительно однородные области, базис вейвлетов квази-оптимален в том смысле, что основная энергия изображения  $B(\xi, \eta)$  сосредоточена в малой части коэффициентов  $a_{k,j}$ . Поэтому оптимальная фильтрация белого шума осуществляется обнулением или уменьшением коэффициентов  $a_{k,j}$ , которые  $|a_{k,j}| < \sigma$ , где  $\sigma$  – порог, пропорциональный уровню шума на изображении.

Вейвлеты достаточно хорошо локализованы в частотной области. Поэтому коррекция некоторых классов ФРТ может быть решена на основе модификации коэффициентов  $a_{k,j}$  [12,13]. К сожалению, в общем случае

такой подход не применим. В [14] предложено комбинировать Винеровскую фильтрацию (5) и последующую вейвлет-обработку. Например, будем ограничивать вейвлет-коэффициенты из условия  $|a_{k,j}| < \sigma \lambda_k$ , где  $\lambda_k$  – ко-

эффициенты, учитывающие усиление шума фильтром (4) для масштаба  $k$ . В результате в (5) параметр  $\varepsilon$  можно увеличить по сравнению с только Винеровской фильтрацией, а параметр  $\sigma$  можно выбирать не зависимо от сюжета в соответствии с оценкой шума датчика. Это позволяет настроить однократно оптимальные параметры фильтрации под конкретный датчик, а не подстраивать  $\varepsilon$  под конкретный сюжет.

Рассмотрим теперь задачу оценивания  $G(\omega, \theta)$  и  $F(\omega, \theta)$ . Что касается  $G(\omega, \theta)$ , то она согласно (3) зависит от параметров  $V_x, V_y, f, N$  даже в координатах  $\xi, \eta$ .

Можно показать, что в частотной области для дискретного изображения (3) можно достаточно качественно аппроксимировать как

$$G(\omega, \theta) = \frac{\sin 0,5(\omega b + \xi c)}{0,5(\omega b + \xi c)}, \quad (6)$$

$$b = N V_x / f, \quad c = N (\Delta_y + V_y / f).$$

На рис. 2 показан пример изображения российского КА «Ресурс-ДК», полученное в конце срока эксплуатации при наличии проблем с системой управления ВЗН (длина вектора смаза ~13 пикселей, угол к направлению полета ~62°), а справа – результат коррекции соответствующего скоростного смаза.

Определение ФРТ расфокусировки  $F$  – наиболее трудоемкая операция. В принципе, после формирования изображения  $B(\xi, \eta)$  для него можно применять различные известные методы оценивания ФРТ [3, 11], в том числе «слепые». Однако в практике ДЗЗ наиболее до-

стоверная оценка ФРТ может быть получена в специально калибровочном режиме, например, путем съёмки полигона с отдельным точечным объектом или яркой звезд. В этом случае образ точечного объекта в координатах  $\xi, \eta$  представляет собой искомую ФРТ. Для отделения точечного объекта от других объектов используется параметрическая коррекция:  $\hat{F}(\xi, \eta) = (0,5(B(\xi, \eta) - a + |B(\xi, \eta) - a|))^b$ , где  $a$  – параметр отсечения фона,  $b \geq 1$  – «степенной» параметр. Значения  $a, b$  подбираются с учетом реального качества коррекции изображений с использованием указанной ФРТ.

При отсутствии полигона можно использовать следующий простой подход: на  $B(\xi, \eta)$  визуально ищутся отдельные точечные объекты в виде характерных пятен (форма пятна совпадает с формой ФРТ, а, их расфокусированные образы представляют собой ФРТ. Если на изображении наблюдается группа характерных пятен, то самое яркое можно использовать в качестве оценки

ФРТ. Такой подход особенно эффективен, если ФРТ имеет характерную геометрическую форму. На рис. 3 слева приведено модельное изображение, полученное путем расфокусировки высокодетального изображения Земли с использованием ФРТ характерной формы (такая ФРТ наблюдалась у реальной системы ДЗЗ). Справа – результат геометрической коррекции с использованием оценки ФРТ по точечному объекту.

Также для современных систем ДЗЗ перспективен следующий подход к оцениванию ФРТ. Обычно на борту спутника имеется многочисленная аппаратура, при этом изображение одной и той же подстилающей поверхности может быть получено различными телескопами или в различных спектральных каналах. Если одно из таких изображений, например,  $C(m, n)$  достаточно четкое, то можно оценить  $F$  для  $B$  по паре изображений  $B$  и  $C$ . Для этого  $C(m, n)$  геометрически трансформируется в систему координат  $\xi, \eta$ :  $C(m, n) \rightarrow C(\xi, \eta)$ . В идеале в спектральной области  $F(\omega, \theta) = B(\omega, \theta) / C(\omega, \theta)$ . Более

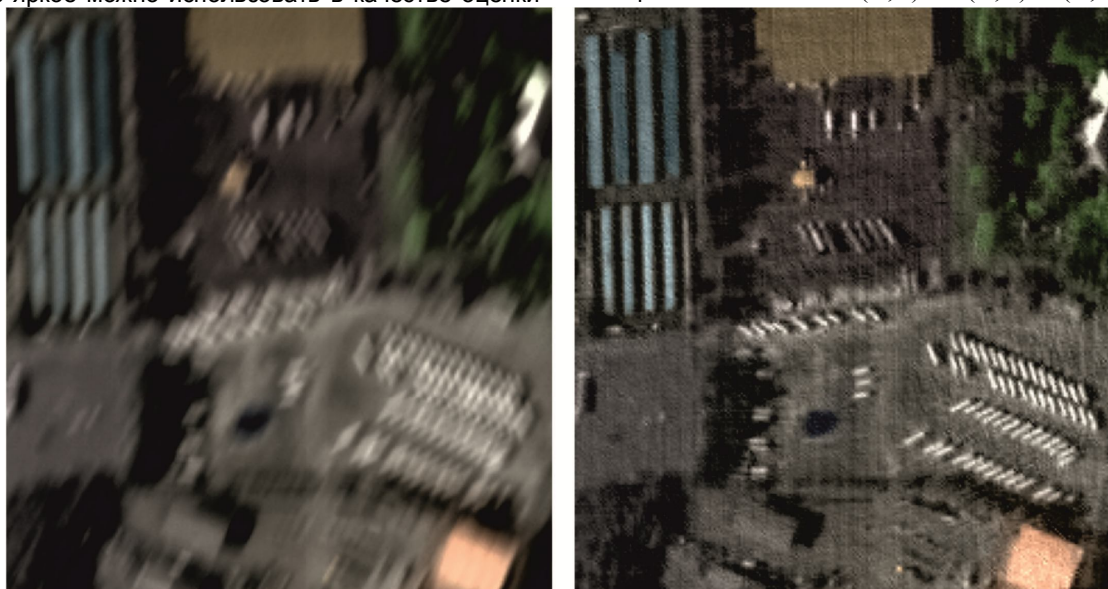


Рис. 2. Скоростной смаз изображения КА «Ресурс-ДК» (слева) и результат его коррекции (справа)

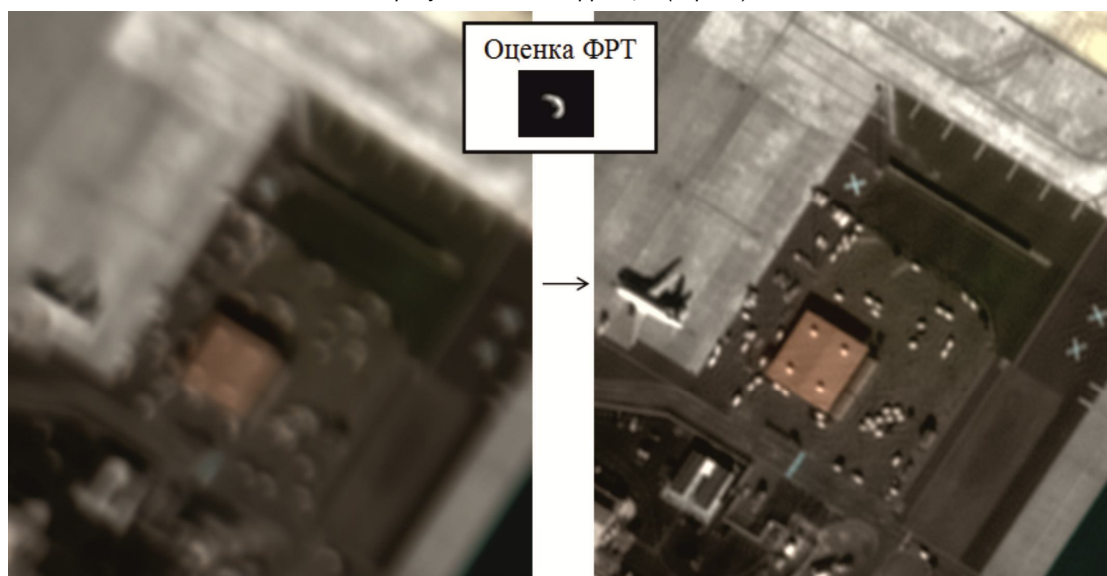


Рис. 3. Фрагмент расфокусированного изображения (слева), оценка ФРТ по точечному объекту и результат коррекции расфокусировки (справа)

строгое действие шумов и свойства подобной оценки изучались [15]. Однако на практике более важно учесть наличие сюжетных различий изображений  $B, C$ . Для их учета выполним оценку по методу наименьших квадратов (МНК) с использованием множества фрагментов изображений  $B$  и  $C$ , при этом

$$\hat{F}(\omega, \theta) = \sum_k w_k B_k(\omega, \theta) \bar{C}_k(\omega, \theta) / \sum_k |C_k(\omega, \theta)|^2, \quad (7)$$

где  $k$  – индекс используемого фрагмента,  $w_k$  – веса. На первой итерации все  $w_k = 1$ . Затем проводится вторая итерация, причем веса  $w_k$  обратно пропорциональны невязкам в МНК. Это позволяет снизить влияние сюжетных различий между изображениями, исключив сильно отличающиеся фрагменты, которые вообще говоря, получены в разных спектральных каналах, в разное время и с разных точек наблюдения.

### Заключение

По результатам выполненных исследований можно сделать следующие выводы.

1. Процесс возникновения скоростного смаза и расфокусировки изображений в современных системах дистанционного зондирования земли целесообразно описывать инвариантно к процессу сканирования, единообразно для высокдетальных и глобальных систем наблюдения земли. За счет специального выбора системы координат изображения удаётся устранить влияние режима съёмки на форму функции рассеяния точки.

2. Показана целесообразность коррекция смаза одновременно с расфокусировкой, при этом смаз аналитически определяется из геометрии съёмки с использованием служебной информации от съёмочной аппаратуры и при необходимости уточняется на основе сопоставления изображений отдельных сканов в областях перекрытия.

3. Для коррекция расфокусировки предложено сперва геометрически трансформировать изображение для исключения влияния режима съёмки на ФРТ. Это позволяет применять известные методы оценивания ФРТ. Наиболее надёжно ФРТ оценивается по полигонам, содержащих изображения практически точечных объектов. Также предложена технология оценивания ФРТ по визуально наблюдаемым точечным объектам и по многоканальным изображениям одной и той же местности.

4. Целесообразно для повышения качества коррекции и уменьшения шумов используются методы вейвлет-преобразования. Показано, что за счет их использования можно упростить подстройку параметров алгоритмов к особенностям определенных сюжетов. Примеры коррекции показывают, что визуальное качество изображений ДЗЗ удаётся значительно повысить, при

сильном смазе и расфокусировке качество съёмки увеличивается в разы.

### Литература

1. Современные технологии обработки данных дистанционного зондирования Земли / Под. ред. В.В. Еремеева. М.: Физматлит, 2015. 460 с
2. Schowengerdt R.A. Remote Sensing: Models and Methods for Image Processing. Academic Press, 2007. 558 p.
3. Storey J.C. Landsat 7 on-orbit modulation transfer function estimation // Proceedings of the SPIE, 2001. Vol. 4540. pp. 50-61.
4. Xin M., Shuyan X., Guangxin L. Remote sensing image restoration with modulation transfer function compensation technology in-orbit // Proceedings of SPIE, 2013. Vol. 8768.
5. Eunsong O., Ki-Beom A., Seongick C., Joo-Hyung R. A modulation transfer function compensation for the geostationary ocean color imager based on the wiener filter // Journal of Astronomy and Space Sciences, 2013. Vol. 30, issue 4. pp. 321-326.
6. Воронин А.А., Егошкин Н.А., Еремеев В.В., Москатиньев И.В. Геометрическая обработка данных космических систем глобального наблюдения земли // Вестник РГПУ, 2009. №1. С. 12-17.
7. Кузнецов П.К., Семавин В.И., Солодуха А.А. Алгоритм компенсации скорости смаза изображения подстилающей поверхности, получаемого при наблюдении Земли из космоса // Вестник Самарского ГТУ, 2005. №37. С. 150-157.
8. Егошкин Н.А., Еремеев В.В., Москвитин А.Э. Комплексирование изображений от линеек фотоприёмников в условиях геометрических искажений // Цифровая обработка сигналов, 2012. №3. – С. 40 - 44.
9. Pratt W.K. Digital Image Processing. Hoboken, New Jersey, 2007. 807pp.
10. Егошкин Н.А., Еремеев В.В. Коррекция смаза изображений в системах космического наблюдения земли // Цифровая обработка сигналов, 2010. №4. С. 28 - 32.
11. Jeon B, Kim H., Chang Y. A MTF compensation for satellite image using L-curve-based modified Wiener filter // Korean Journal of Remote Sensing, 2012. Vol 28, pp. 561-571.
12. Mallat S. A Wavelet Tour of Signal Processing. Academic Press, 1998. 550p.
13. Donoho D.L. Nonlinear solution of linear inverse problems by Wavelet-Vaguelette Decomposition // Appl. Comp. Harm. Anal., 1995. Vol. 2, pp. 101-126.
14. Neelamani R., Choi H., Baraniuk R. Wavelet-based deconvolution for ill-conditioned systems // IEEE Transactions on Image Processing, 1999.
15. Ljung L. System Identification. Theory for the User. PTR Prentice Hall, Upper Saddle River, 1999. 609 p.

## УТОЧНЕНИЕ ПРОСТРАНСТВЕННОЙ БАЗЫ ПРИ КОСМИЧЕСКОЙ БИСТАТИЧЕСКОЙ РАДИОЛОКАЦИОННОЙ СЪЕМКЕ ЗЕМЛИ ПО СИГНАЛУ ИНТЕРФЕРОГРАММЫ

*Егошкин Н.А., в.н.с. НИИ «Фотон» РГРТУ, к.т.н., e-mail: foton@rsreu.ru;*

*Ушенкин В.А., м.н.с. НИИ «Фотон» РГРТУ, e-mail: foton@rsreu.ru.*

### BISTATIC SAR BASELINE ESTIMATION BY INTERFEROGRAM ANALYSIS

*Egoshkin N.A., Ushenkin V.A.*

*The problem of spaceborne InSAR baseline estimation by interferogram analysis is considered. Two main parameters, which depend on baseline and need to be estimated, are selected. They are multiplicative factor of phase dependence on relief height and multiplicative factor of flat relief phase. The algorithms of these parameters high-precision estimation, which don't require interferogram phase unwrapping, are proposed.*

**Key words:** interferogram, SAR imaging, digital elevation model, phase unwrapping, interferometric baseline.

**Ключевые слова:** интерферограмма, радиолокационная съемка, цифровая модель рельефа, развертывание фазы, пространственная интерферометрическая база.

#### Введение

Бистатическая радиолокационная съемка Земли из Космоса подразумевает использование двух антенн, разнесенных в пространстве. Вектор, соединяющий положения фазовых центров антенн в моменты времени, когда они снимают одну и ту же точку на поверхности Земли, принято называть пространственной базой.

В зависимости от ориентации вектора базы относительно траектории движения радиолокатора выделяют продольную и поперечную бистатические съемки. Продольная бистатическая съемка предназначена для селекции движущихся целей и в данной работе не рассматривается. Поперечная бистатическая радиолокационная съемка предназначена для построения высокодетальных цифровых моделей рельефа (ЦМР) земной поверхности путем интерферометрической обработки изображений, полученных от двух антенн [1].

Существует три варианта проведения поперечной бистатической съемки (далее – интерферометрической съемки).

Первый вариант предусматривает использование одного радиолокатора с двумя антеннами на жесткой базе. Такая съемка выполнялась в 2000 г. шаттлом «Индевор» для построения глобальной ЦМР SRTM. Модуль и направление вектора базы в этом случае известны с высокой точностью, и их уточнение не требуется.

Второй вариант предусматривает использование двух радиолокаторов, расположенных на двух космических аппаратах, движущихся рядом друг с другом по параллельным траекториям. Указанный вариант съемки называется «тандемным» и используется в связке космических аппаратов TerraSAR-X и TanDEM-X. Модуль и направление вектора базы при «тандемной» съемке

*Рассматривается задача уточнения пространственной базы при интерферометрической радиолокационной съемке Земли из Космоса путем анализа сигнала интерферограммы. Выделяется два основных параметра, зависящих от базы и подлежащих уточнению: мультипликативный коэффициент зависимости фазы на интерферограмме от высоты рельефа земной поверхности и коэффициент квазилинейного набега фазы плоского рельефа. Предлагаются алгоритмы высокоточной оценки значений указанных параметров, не требующие развертывания фазы на интерферограмме.*

также известны с высокой точностью, поскольку для их определения используется дифференциальный режим навигационных систем.

Третий вариант подразумевает съемку одним радиолокатором при двух его пролетах над одной и той же местностью на разных витках орбиты (межвитковая интерферометрическая съемка). При этом траектории двух пролетов немного смещены друг от друга в Гринвичской системе координат, образуя пространственную базу. Межвитковая интерферометрическая съемка наиболее распространена, благодаря относительной простоте и дешевизне ее организации, однако при ней модуль и направление вектора пространственной базы известны недостаточно точно, поскольку из-за появившегося интервала времени между съемками двух изображений невозможно использовать дифференциальный режим навигационных систем.

Погрешности определения пространственной базы негативно сказываются на точности конечных и промежуточных результатов интерферометрической обработки.

Первым этапом интерферометрической обработки является пространственное совмещение двух изображений, полученных при съемке. Погрешность величины проекции вектора базы на вектор визирования, называемой параллельной базой, приводит к ошибке совмещения изображений, которая, тем не менее, может быть устранена путем корреляционно-экстремального уточнения параметров совмещения [2].

На втором этапе формируется интерферограмма –

матрица разностей фаз сигнала в одноименных пикселях совмещенных изображений. Фазоразностный сигнал интерферограммы принято для простоты называть фазой. Фаза на интерферограмме включает пять составляющих: фазовый шум, атмосферные искажения, фазу смещений земной поверхности (только при межвитковой интерферометрии), фазу плоского рельефа и топографическую фазу. Для построения ЦМР из фазы на интерферограмме вычлняют топографическую составляющую. При этом наличием атмосферных искажений и смещений земной поверхности пренебрегают, а шум устраняют путем фильтрации интерферограммы [1].

Фаза плоского рельефа  $\varphi_{\text{плоск}}$  квазилинейно зависит от дальности  $R$ , пропорциональной номеру пикселя в строке интерферограммы:

$$\varphi_{\text{плоск}}(R) \approx K_{\text{плоск}}(B_{\perp}, f_1, f_1 - f_2)R, \quad (1)$$

где  $K_{\text{плоск}}$  – коэффициент квазилинейной зависимости, зависящий от рабочих частот  $f_1$  и  $f_2$  радиолокатора, используемых при съемке двух изображений, и перпендикулярной базы  $B_{\perp}$  – проекции вектора пространственной базы на нормаль к вектору визирования.

Фаза плоского рельефа вычитается из сигнала интерферограммы для вычленения топографической составляющей фазы. При этом, если  $K_{\text{плоск}}$  определен неточно, на интерферограмме остается паразитный квазилинейный набег фазы в направлении дальности. Паразитный набег фазы может также наблюдаться в направлении азимута (вдоль столбца интерферограммы). Он вызван малыми (сантиметровыми) изменениями величины абсолютной погрешности определения параллельной пространственной базы.

Топографическая фаза  $\varphi_{\text{моно}}$  квазипропорциональна высоте  $h$  рельефа земной поверхности. При этом коэффициент пропорциональности  $K_{\text{моно}}$  определяется величиной перпендикулярной базы  $B_{\perp}$  и рабочей частотой  $f_2$  радиолокатора при съемке второго изображения:

$$\varphi_{\text{моно}}(h) \approx K_{\text{моно}}(B_{\perp}, f_2)h. \quad (2)$$

Таким образом, в случае неточного знания величины пространственной базы ЦМР в результате интерферометрической обработки содержит мультипликативную ошибку и паразитный набег высоты вдоль направлений дальности и азимута.

Основным подходом к уточнению пространственной базы является сравнение полученной ЦМР с опорной информацией о рельефе. Паразитный набег высоты оценивается путем соотнесения с опорой высот в пикселях ЦМР, имеющих различное положение по дальности и азимуту. Мультипликативная ошибка выявляется путем сравнения ширины диапазона высот на полученной ЦМР с шириной диапазона высот опоры [3].

Указанный подход имеет недостаток, связанный с тем, что отличия полученной ЦМР от опоры могут быть вызваны не только погрешностью базы, но и ошибками развертывания фазы интерферограммы, которое заключается в восстановлении истинного диапазона значений топографической фазы из отсчетов интерферограммы, принадлежащих диапазону от  $-\pi$  до  $\pi$  радиан,

путем прибавления к каждому из них некоторого целого количества фазовых периодов ( $2\pi$  радиан) [4]. Кроме того, в работе [5] показано, что в наиболее сложных случаях для правильного развертывания фазы может понадобиться привлечение опорного рельефа и пересчет его в опорную топографическую фазу. Следовательно, при развертывании фазы уже необходимо с высокой точностью знать пространственную базу и  $K_{\text{моно}}$ .

**Цель настоящей работы** – построение алгоритмов высокоточной оценки  $K_{\text{плоск}}$  и  $K_{\text{моно}}$  по интерферограмме, не прошедшей процедуру развертывания фазы. При этом уточнение  $K_{\text{плоск}}$  и  $K_{\text{моно}}$  должно выполняться независимо, поскольку на точности  $K_{\text{моно}}$  отражаются только погрешности  $B_{\perp}$ , а на точности  $K_{\text{плоск}}$  – еще и нестабильность рабочей частоты радиолокатора, из-за которой высока погрешность определения  $f_1 - f_2$ .

### Предварительное уточнение фазы плоского рельефа

Перед уточнением  $B_{\perp}$  и  $K_{\text{моно}}$  нужно добиться, чтобы хотя бы на небольшом фрагменте интерферограммы величина паразитного набег фазы по дальности и азимуту была пренебрежимо мала. Поэтому необходимо выполнить предварительное уточнение фазы плоского рельефа.

Для уточнения следует вычлнить из сигнала интерферограммы остаточную фазу опорного рельефа, определяющую паразитный набег. Для этого авторы предлагают размыть и проредить интерферограмму до разрешения опорной низкодетальной ЦМР, а затем выполнить вычитание фазы опорного рельефа.

Для размытия может применяться любой сглаживающий фильтр, например, фильтр Гаусса. Ключевой особенностью является его применение не к фазе  $\varphi$ , а к комплексному сигналу  $e^{j\varphi}$ :

$$\varphi_{\text{разм } mn} = \arg \left( \sum_k \sum_l G_{kl} e^{j\varphi_{m+k, n+l}} \right), \quad (3)$$

где  $G$  – фильтр Гаусса.

Прореженная интерферограмма с вычтенной фазой опорного рельефа формируется следующим образом:

$$\hat{\varphi}_{mn} = \arg e^{j(\varphi_{\text{разм } [mK_y], [nK_x]} - \varphi_{\text{моно}}(h_{on mn}))}, \quad (4)$$

где  $\hat{\varphi}$  – фаза на прореженной интерферограмме,  $[x]$  – операция округления  $x$  до ближайшего целого,  $\varphi_{\text{моно}}(h_{on mn})$  – фаза опорного рельефа  $h_{on mn}$ .

Размытие интерферограммы позволяет удалить топографическую фазу, вызванную мелкими деталями рельефа, а вычитание  $\varphi_{\text{моно}}(h_{on mn})$  – удалить топографическую фазу, вызванную крупными элементами рельефа. В результате сигнал на прореженной интерферограмме представляет собой сумму атмосферных искажений, фазы смещений, остаточной фазы плоского рельефа и остаточной низкодетальной топографической фазы.

Остаточная низкодетальная топографическая фаза вызвана погрешностью  $\Delta K_{\text{моно}}$  коэффициента пропор-

циональности топографической фазы высоте и представляет собой значительно уменьшенную по амплитуде фазу опорного рельефа  $\frac{\Delta K_{mono}}{K_{mono}} \varphi_{mono}(h_{on mn})$ . Ее наличие на интерферограмме не позволяет выполнить высокоточную оценку  $K_{плоск}$ , но в то же время она достаточно мала для того, чтобы выполнить приближенную оценку  $K_{плоск}$ .

Поскольку развертывание фазы на прореженной интерферограмме не выполнялось, то непрерывная фазовая поверхность не получена, а следовательно, нет возможности выполнить ее аппроксимацию плоскостью. В то же время, если предположить, что разность фаз в соседних пикселях прореженной интерферограммы не превышает половину периода, то можно оценить градиент фазы как:

$$\hat{G}_{mnx} = \arg e^{j(\hat{\varphi}_{m,n+1} - \hat{\varphi}_{mn})}, \quad \hat{G}_{mny} = \arg e^{j(\hat{\varphi}_{m+1,n} - \hat{\varphi}_{mn})}. \quad (5)$$

Поскольку атмосферные искажения, остаточная фаза плоского рельефа и большая часть остаточной топографической фазы носит плавный характер, то оценки  $\hat{G}_{mnx}$  и  $\hat{G}_{mny}$  будут совпадать с истинным градиентом фазы практически во всех пикселях прореженной интерферограммы, и лишь в малом количестве пикселей оценки будут отличаться от истинного градиента на целое количество периодов. Следовательно, величину паразитного набега фазы можно приближенно оценить по средневзвешенной оценке градиента фазы:

$$G_{cp x} = \frac{\sum_m \sum_n w_{mn} \hat{G}_{mnx}}{\sum_m \sum_n w_{mn}}, \quad G_{cp y} = \frac{\sum_m \sum_n w_{mn} \hat{G}_{mny}}{\sum_m \sum_n w_{mn}}. \quad (6)$$

Веса  $w_{mn}$  предлагается задавать в соответствии со средней когерентностью на фрагменте исходной интерферограммы, приходящемся на  $mn$ -й пиксель прореженной интерферограммы. Это позволяет отбраковать участки низкой когерентности, в которых вместо сигнала содержится только коррелированный шум, оставшийся после фильтрации. Для большей точности оценки паразитного набега фазы можно считать нулевой когерентность сигнала в участках на исходной интерферограмме, в которых наиболее вероятен фазовый разрыв (т.е. изменение фазы за один пиксель более чем на половину периода). Критерий для выявления таких участков предложен в работе [4].

Предварительная поправка к  $K_{плоск}$  определяется следующим образом:

$$\Delta K_{плоск} = \frac{G_{cp x}}{\Delta R}, \quad (7)$$

где  $\Delta R$  – шаг по дальности на прореженной интерферограмме.

По результатам предварительной оценки паразитного набега фазы прореженная интерферограмма корректируется следующим образом:

$$\hat{\varphi}_{mn}^* = \arg e^{j(\hat{\varphi}_{mn} - G_{cp x} n - G_{cp y} m + \varphi_{mono}(h_{on mn}))}, \quad (8)$$

где  $\hat{\varphi}^*$  – фаза на скорректированной интерферограмме, основной составляющей которой является низкодета-

льная топографическая фаза. Также  $\hat{\varphi}^*$  включает атмосферные искажения, фазу смещений и малый паразитный набег, оставшийся после предварительной оценки.

### Уточнение перпендикулярной пространственной базы

Уточнение  $K_{mono}$  предлагается выполнять на основе сравнения статистических характеристик низкодетаальной топографической фазы  $\hat{\varphi}^*$  и фазы опорного рельефа.

В результате выполнения предварительного уточнения фазы плоского рельефа паразитный набег фазы  $\hat{\varphi}^*$  вдоль строки (столбца) интерферограммы, как правило, не превышает 1-2 периодов. Такая величина набега не позволяет достоверно собрать статистику по всей прореженной интерферограмме, но в малой области радиусом в несколько пикселей набег влияет только на среднее значение фазы  $\hat{\varphi}^*$  и практически не влияет на ее среднеквадратическое отклонение (СКО). Аналогичные рассуждения можно провести относительно влияния атмосферных искажений. Поэтому в настоящей работе предлагается сравнивать оценки СКО низкодетаальной топографической фазы и фазы опорного рельефа по малым фрагментам на прореженной интерферограмме.

При вычислении СКО предполагается, что в рамках малого фрагмента разность развернутой низкодетаальной топографической фазы и фазы опорного рельефа колеблется в пределах одного периода. Исходя из этого предположения в рамках каждого малого фрагмента оценивается с точностью до постоянной развернутая топографическая фаза:

$$\hat{\varphi}_{раз kl}^* = \varphi_{mono}(h_{on kl}) + \arg e^{j(\hat{\varphi}_{kl}^* - \varphi_{mono}(h_{on kl}) - \varphi_{cp mn})}, \quad (9)$$

где  $\varphi_{cp mn} = \arg \left( \sum_{k=m-r}^{m+r} \sum_{l=n-r}^{n+r} e^{j(\hat{\varphi}_{kl}^* - \varphi_{mono}(h_{on kl}))} \right)$ ,  $r$  – радиус

малого фрагмента.

Для сравнения оценок СКО низкодетаальной топографической фазы  $\hat{\varphi}^*$  и фазы опорного рельефа вычисляется их отношение:

$$\delta K_{mono mn} = \frac{\sigma_{\hat{\varphi}^* mn}}{\sigma_{on mn}} = \quad (10)$$

$$= \sqrt{\frac{\sum_{k=m-r}^{m+r} \sum_{l=n-r}^{n+r} \hat{\varphi}_{раз kl}^{*2}}{(2r+1)^2} - \left( \frac{\sum_{k=m-r}^{m+r} \sum_{l=n-r}^{n+r} \hat{\varphi}_{раз kl}^*}{(2r+1)^2} \right)^2} \div \sqrt{\frac{\sum_{k=m-r}^{m+r} \sum_{l=n-r}^{n+r} \varphi_{mono}^2(h_{on kl})}{(2r+1)^2} - \left( \frac{\sum_{k=m-r}^{m+r} \sum_{l=n-r}^{n+r} \varphi_{mono}(h_{on kl})}{(2r+1)^2} \right)^2},$$

где  $\sigma_{\hat{\varphi}^* mn}$  и  $\sigma_{on mn}$  – оценки СКО низкодетаальной топографической фазы  $\hat{\varphi}^*$  и фазы опорного рельефа, вычисленные по фрагменту с центром в  $mn$ -м пикселе прореженной интерферограммы.

Поскольку  $\sigma_{\hat{\varphi}^* mn}$  и  $\sigma_{on mn}$  оцениваются по малой выборке, погрешность  $\delta K_{mono mn}$  достаточно высока. Получить оценку отношения СКО с приемлемой точностью предлагается путем усреднения  $\delta K_{mono mn}$  по большому количеству фрагментов. Усреднение допустимо, поскольку  $B_{\perp}$  и ее относительная погрешность практически не изменяются за время съемки. При этом следует отбраковать те фрагменты, в которых мало  $\sigma_{on mn}$  (меньше порога  $\sigma_{min}$ ), поскольку при этом становится велико влияние атмосферных искажений и фазы смещений на значение  $\delta K_{mono mn}$ . Для повышения точности оценки при усреднении можно учитывать веса  $w_{mn}$ , использованные ранее в (6), и значение  $\sigma_{on mn}$ :

$$\delta K_{mono} = \frac{\sum_{mn, \sigma_{on mn} \geq \sigma_{min}} w_{mn} \sigma_{on mn} \delta K_{mono mn}}{\sum_{mn, \sigma_{on mn} \geq \sigma_{min}} w_{mn} \sigma_{on mn}}. \quad (11)$$

Уточненные значения перпендикулярной пространственной базы  $B_{\perp}^*$  и зависящего от нее коэффициента  $K_{mono}^*$  пропорциональности топографической фазы высоте рельефа вычисляются следующим образом:

$$B_{\perp}^* = B_{\perp} \delta K_{mono}, \quad K_{mono}^* = K_{mono} \delta K_{mono}. \quad (12)$$

После уточнения  $B_{\perp}$  и  $K_{mono}$  следует скорректировать фазу опорного рельефа:

$$\varphi_{mono}^*(h_{on kl}) = \varphi_{mono}(h_{on kl}) \delta K_{mono}. \quad (13)$$

При больших относительных погрешностях  $B_{\perp}$  и больших перепадах высоты рельефа ширина диапазона значений остаточной топографической фазы в ряде фрагментов может превышать один период. В этом случае неправильно оценивается развернутая низкодетальная топографическая фаза  $\hat{\varphi}_{разв}^*$ , что негативно отражается на точности  $\delta K_{mono}$ . Поэтому описанную процедуру уточнения  $B_{\perp}$  и  $K_{mono}$  целесообразно выполнять итерационно, до тех пор, пока значение  $\delta K_{mono}$  не станет близко к единице.

### Окончательное уточнение фазы плоского рельефа

Для окончательного уточнения необходимо вычле-

нить из прореженной интерферограммы остаточную фазу плоского рельефа:

$$\hat{\varphi}_{mn}^{**} = \arg e^{j(\hat{\varphi}_{mn}^* - \varphi_{mono}^*(h_{on mn}))}. \quad (14)$$

Помимо остаточной фазы плоского рельефа,  $\hat{\varphi}^{**}$  содержит только атмосферные искажения и фазу смещений, поскольку остаточная топографическая фаза устранена на предыдущем этапе. Поэтому  $\hat{\varphi}^{**}$  может использоваться для высокоточной оценки величины паразитного набега фазы.

Оценка величины набега на основе усреднения градиента фазы недостаточно точна, поскольку аппроксимация сигнала линейной функцией по среднему градиенту слишком чувствительна к шумам в крайних отсчетах, что для одномерного случая иллюстрирует рис. 1. Гораздо большую устойчивость к шумам в крайних отсчетах проявляет аппроксимация по методу наименьших квадратов (МНК), однако ее реализация для неразвернутой фазы связана с рядом проблем.

В случае развернутой фазы  $\hat{\varphi}_{разв}^{**}$  оценка величины паразитного набега по МНК сводится к нахождению псевдорешения системы линейных алгебраических уравнений (СЛАУ):

$$\begin{cases} \varphi_0 + G_x n + G_y m = \hat{\varphi}_{разв mn}^{**} \end{cases} \quad (15)$$

Нахождение псевдорешения в этом случае тривиально и выполняется путем однократного набора статистики по интерферограмме и решения СЛАУ из трех уравнений с тремя неизвестными  $\varphi_0$ ,  $G_x$  и  $G_y$ .

В случае же неразвернутой фазы  $\hat{\varphi}^{**}$  оценка величины паразитного набега по МНК сводится к нахождению псевдорешения системы трансцендентных уравнений:

$$\begin{cases} \arg e^{j(\varphi_0 + G_x n + G_y m - \hat{\varphi}_{mn}^{**})} = 0. \end{cases} \quad (16)$$

Нахождение псевдорешения системы (16) возможно только с помощью численных итерационных методов. При этом на каждой итерации для расчета невязок необходим сбор статистики по прореженной интерферограмме, требующий достаточно большого объема вычислений.

Еще одной проблемой является то, что невязки в системе (16) начинают монотонно уменьшаться только тогда, когда паразитный набег фазы вдоль строки (столбца) интерферограммы становится меньше периода. При большей величине набега невязки остаются практически постоянными из-за периодичности функции  $e^{j\varphi}$ .

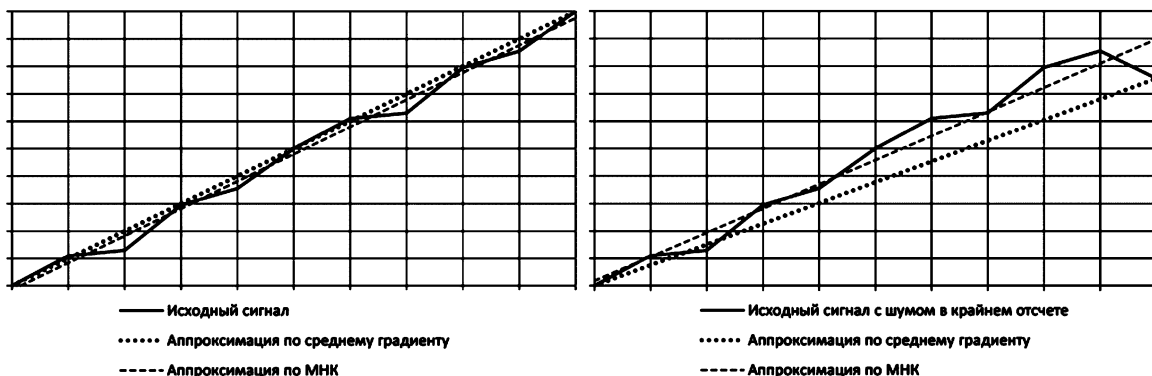
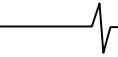


Рис. 1. Влияние шумов в крайних точках на точность аппроксимации по среднему градиенту и по МНК





Таким образом, если паразитный набег фазы превышает период, его оценка с погрешностью, меньшей периода, возможна только путем полного перебора всех возможных значений набег в целое количество периодов.

Если предположить, что в результате предварительной оценки паразитный набег фазы вдоль строки (столбца) интерферограммы не превышает по модулю  $K_{\max}$  периодов (как правило, это 1–2 периода), то количество вариантов для перебора составляет  $(2K_{\max} + 1)^2$ .

Значения  $G_x$  и  $G_y$  для каждого варианта определяются как:

$$G_x(K_x) = \frac{2\pi K_x}{N-1}, \quad G_y(K_y) = \frac{2\pi K_y}{M-1}, \quad (17)$$

где  $K_x, K_y = -K_{\max}, K_{\max}$  – величина паразитного набег вдоль строки (столбца) в периодах,  $M$  и  $N$  – размеры раstra прореженной интерферограммы.

Значения  $\varphi_0$  для каждой комбинации  $G_x(K_x)$  и  $G_y(K_y)$  определяются следующим образом:

$$\varphi_0(K_x, K_y) = \arg \left( \sum_m \sum_n e^{j(\varphi_{mn}^{**} - G_x(K_x)n - G_y(K_y)m)} \right). \quad (18)$$

Среди всех полученных троек  $\varphi_0(K_x, K_y)$ ,  $G_x(K_x)$  и  $G_y(K_y)$  выбирается такая тройка  $\varphi_{0\min 0}$ ,  $G_{x\min 0}$  и  $G_{y\min 0}$ , для которой минимальна сумма квадратов невязок системы (16). При этом целесообразно проконтролировать, что для выбранной тройки средний квадрат невязок существенно меньше  $\pi^2/3$ . Значение среднего квадрата невязок, сопоставимое с  $\pi^2/3$ , говорит о том, что ни один из вариантов не соответствует истинному паразитному наклону (с точностью, меньшей периода) и следует повторить перебор для большего значения  $K_{\max}$ .

После успешного выполнения полного перебора становится возможным итерационное приближение к истинному значению паразитного набег.

Итерационное приближение предлагается осуществлять следующим образом. На каждой  $i$ -й итерации формируется 9 возможных троек:

$$G_{xi}(K_{xi}) = G_{x\min i-1} + \frac{2\pi K_{xi}}{2^i(N-1)}, \quad (19)$$

$$G_{yi}(K_{yi}) = G_{y\min i-1} + \frac{2\pi K_{yi}}{2^i(M-1)},$$

$$\varphi_{0i}(K_{xi}, K_{yi}) = \arg \left( \sum_m \sum_n e^{j(\varphi_{mn}^{**} - G_{xi}(K_{xi})n - G_{yi}(K_{yi})m)} \right),$$

где  $K_{xi}, K_{yi} = -1, 1$ .

Из указанных троек выбирается тройка  $\varphi_{0\min i}$ ,  $G_{x\min i}$  и  $G_{y\min i}$ , которая обеспечивает наименьшую сумму квадратов невязок системы (16).

Максимальная величина уточнения паразитного

набег фазы вдоль строки и столбца на каждой последующей итерации сокращается в 2 раза. Итерации прекращаются на  $i_{\max}$ -й итерации, когда  $2\pi / (2^{i_{\max}} |K_{\text{моно}}^*|)$  становится меньше заданного порога, например, равно  $0,2\pi$ .

Окончательная поправка к  $K_{\text{плоск}}$  определяется как:

$$\Delta K_{\text{плоск}}^* = \Delta K_{\text{плоск}} + \frac{G_{x\min i_{\max}}}{\Delta R}, \quad (20)$$

а уточненное значение коэффициента квазилинейной зависимости фазы плоской земли от дальности определяется как:

$$K_{\text{плоск}}^* = K_{\text{плоск}} + \Delta K_{\text{плоск}}^*. \quad (21)$$

При необходимости можно оценить величину относительной нестабильности рабочей частоты радиолокатора, вычислив  $(f_1 - f_2) / f_1$  путем обращения зависимости

$$K_{\text{плоск}}^*(B_{\perp}^*, f_1, f_1 - f_2).$$

Для большей точности оценки  $\Delta K_{\text{плоск}}^*$  классический МНК может быть заменен на весовой МНК. Алгоритм окончательного уточнения фазы плоского рельефа при этом не изменится, просто вместо обычной суммы квадратов невязок системы (16) будет вычисляться взвешенная сумма. В качестве весов целесообразно взять  $w_{mn}$ , использовавшиеся в (6) при предварительной оценке.

### Экспериментальные исследования

Выполнены экспериментальные исследования предложенных выше алгоритмов на натурной информации от космических аппаратов (КА) RadarSat-2 и TerraSAR-X. Примеры результатов работы предложенных алгоритмов представлены на рис. 2-4, где:

а – сигнал  $\varphi$  исходной интерферограммы;

б – сигнал  $\hat{\varphi}$  прореженной интерферограммы;

в – сигнал  $\arg e^{j(\hat{\varphi}^* - \varphi_{\text{моно}}(h_{\text{он}}))}$  прореженной интерферограммы после предварительного уточнения фазы плоского рельефа;

г – сигнал  $\hat{\varphi}^{**}$  прореженной интерферограммы после уточнения перпендикулярной пространственной базы;

д – сигнал прореженной интерферограммы после окончательного уточнения и устранения фазы плоского рельефа (основные составляющие сигнала: фаза смещений и атмосферные искажения).

Цветовая шкала, в которой представлены сигналы, приведена на рис. 5. Черным цветом помечены участки, получившие нулевые веса  $w_{mn}$ .

Выполнена оценка точности полученных результатов. Для этого выполнялось высокоточное развертывание фазы и формирование ЦМР с учетом полученных оценок  $K_{\text{моно}}^*$  и  $K_{\text{плоск}}^*$ , а затем традиционным способом оценивались и пересчитывались в фазу паразитный набег высоты на полученной ЦМР и отличие СКО полученной высоты от СКО опорного рельефа. Результаты оценки приведены в табл. 1, из которой видно, что абсолютная погрешность оценки паразитного набег фазы предложенным алгоритмом не превышает  $0,12\pi$  радиан, или  $0,06$  фазового периода, а погрешность оценки  $B_{\perp}$



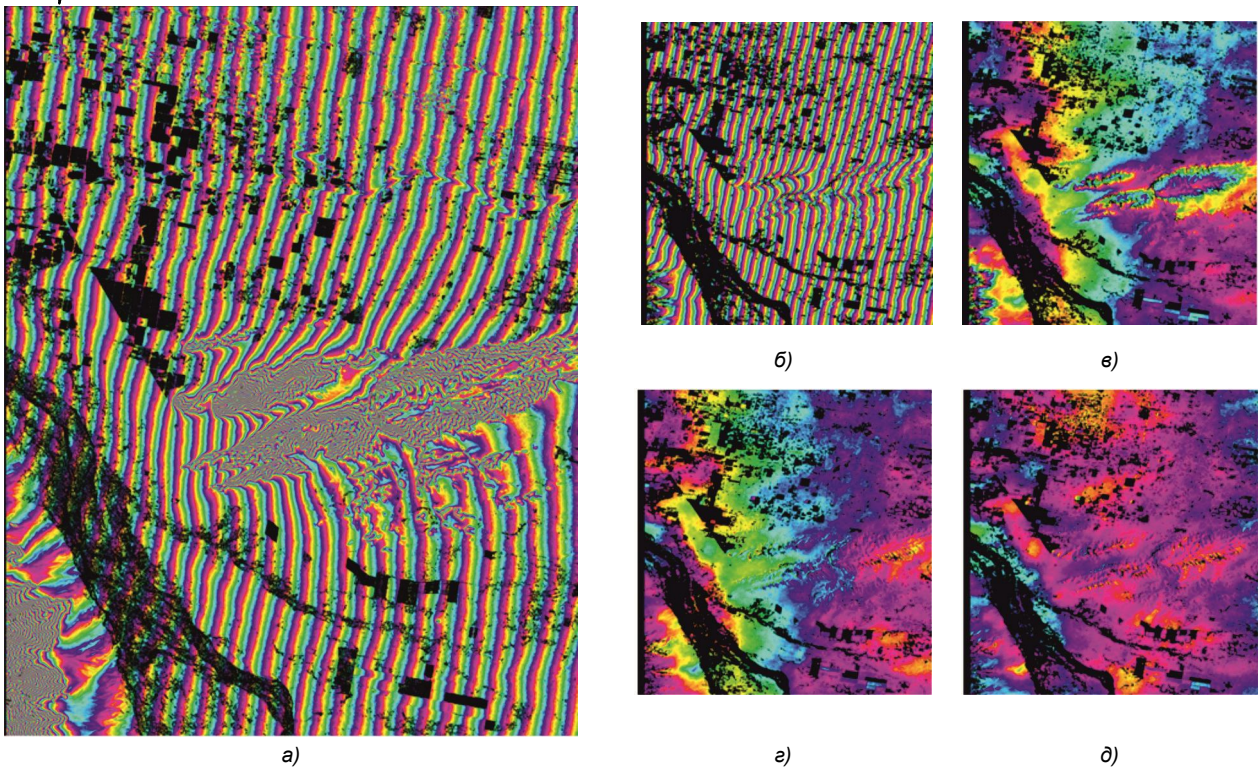


Рис. 2. Результаты уточнения фазы плоского рельефа и перпендикулярной пространственной базы для интерферограммы города Финикс, США, КА RadarSat-2, 4–28 мая 2008 г.

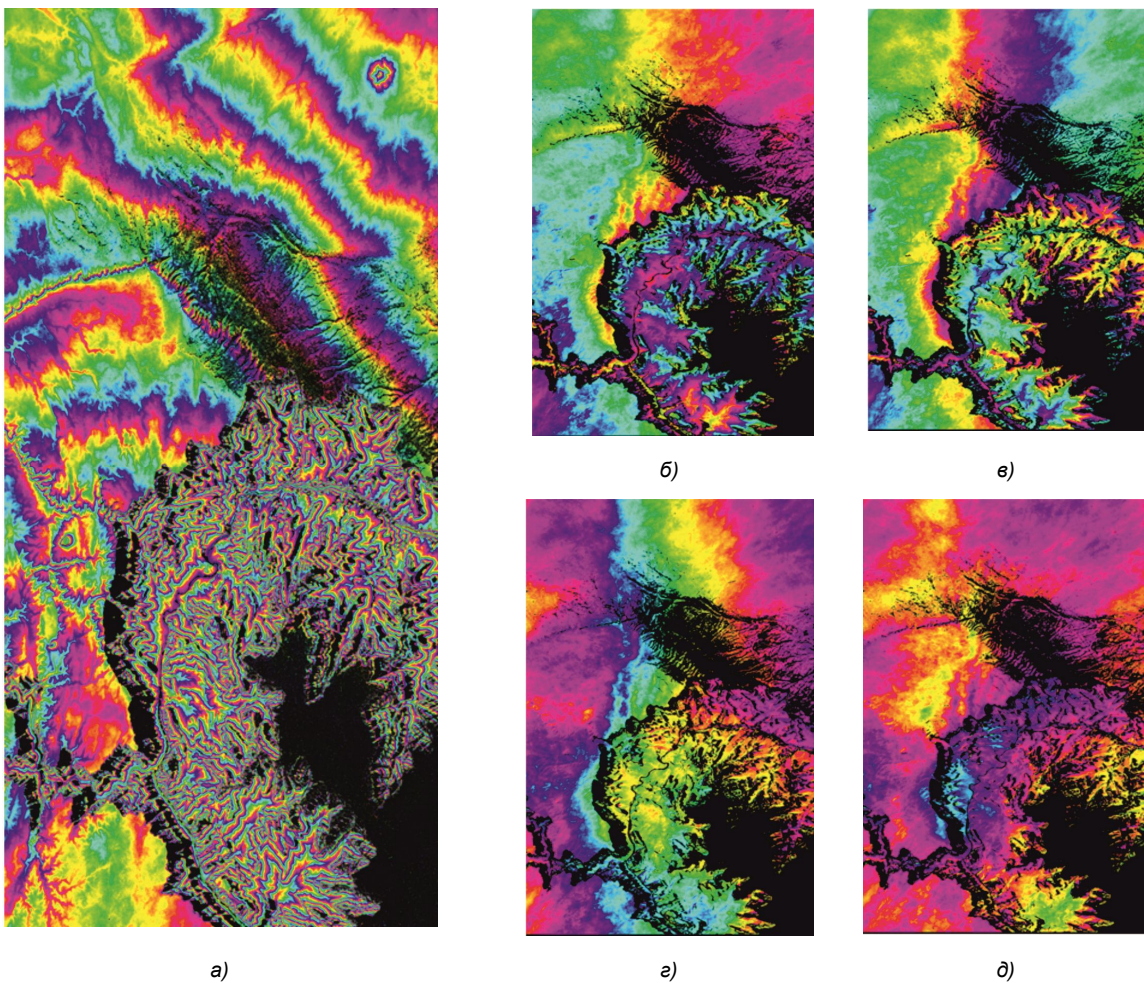
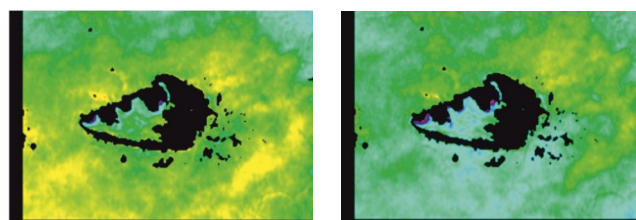
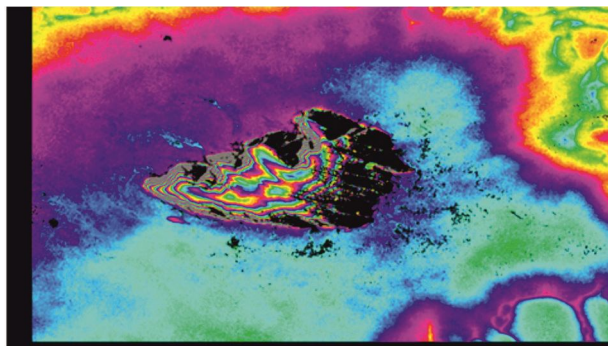


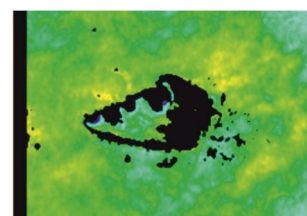
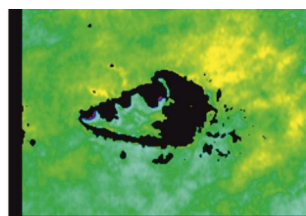
Рис. 3. Результаты уточнения фазы плоского рельефа и перпендикулярной пространственной базы для интерферограммы Большого Каньона, США, КА TerraSAR-X, 10-21 марта 2008 г.





б)

е)



г)

д)

а)

Рис. 4. Результаты уточнения фазы плоского рельефа и перпендикулярной пространственной базы для интерферограммы заповедника Улуру, Австралия, КА TerraSAR-X, 12–23 февраля 2009 г.



Рис. 5. Цветовая шкала представления сигнала интерферограммы

Таблица 1. Остаточный паразитный набег высоты и отличия СКО высоты от СКО опорного рельефа из-за погрешности оценки перпендикулярной интерферометрической базы

Параметр	Интерферограмма		
	Финикс	Большой Каньон	Улуру
Остаточный паразитный набег высоты по дальности (вдоль строки), м	1,8	-0,2	1,5
Остаточный паразитный набег высоты по азимуту (вдоль столбца), м	-2,2	-3,1	0,6
Остаточный паразитный набег фазы по дальности, радиан	-0,097π	0,003π	-0,081π
Остаточный паразитный набег фазы по азимуту, радиан	0,119π	0,043π	-0,032π
СКО высоты опорного рельефа, м	81,5	393,8	65,9
СКО высоты на полученной ЦМР, м	81,9	397,8	66,5
СКО фазы опорного рельефа, радиан	4,41π	5,43π	3,56π
СКО развернутой топографической фазы, радиан	4,43π	5,48π	3,59π

приводит к отличию СКО топографической фазы от СКО фазы опорного рельефа не более чем на 0,05π радиан, или 0,025 фазового периода.

### Заключение

Предложенные алгоритмы позволяют до этапа развертывания фазы интерферограммы оценить величину фазы плоского рельефа и устранить ее паразитный набег с точностью до 0,06 фазового периода, а также оценить перпендикулярную интерферометрическую базу и уравнивать СКО топографической фазы и фазы опорного рельефа с точностью до 0,025 фазового периода. За счет этого становится возможным привлечение опорного рельефа для более точного развертывания фазы.

### Литература

1. Верба В.С., Неронский Л.Б., Осипов И.Г., Турук В.Э. Радиолокационные системы землеобзора космического базирования / под. ред. В.С. Вербы. М.: Радиотехника, 2010. 680 с.

2. Егошкин Н.А., Ушенкин В.А. Совмещение высокодетальных изображений с использованием опорной цифровой модели рельефа при интерферометрической обработке радиолокационной информации // Вестник Рязанского государственного радиотехнического университета. 2015. № 51. С. 72-79.

3. Ren K., Prinnet V., Shi X., Wang F. Comparison of satellite baseline estimation methods for interferometry applications // Proc. of IGARSS'03. 2003. Vol. 6. P. 3821-3823.

4. Егошкин Н.А., Ушенкин В.А. Интерферометрическая обработка радиолокационной информации на основе комбинации методов развертывания фазы // Вестник Рязанского государственного радиотехнического университета. 2015. № 54-2. С. 21-31.

5. Ушенкин В.А. Интерферометрическая обработка высокодетальной информации от космических систем радиолокационного наблюдения Земли // Тезисы докладов Четвертой международной научно-технической конференции «Актуальные проблемы создания космических систем дистанционного зондирования Земли». М.: АО «Корпорация «ВНИИЭМ», 2016. С. 206-207.

УДК 004.93'11

## МОДЕЛЬ АТАК И МЕТОД ПОДТВЕРЖДЕНИЯ ПОДЛИННОСТИ ОБЪЕКТОВ В СИСТЕМАХ КОНТРОЛЯ ДОСТУПА ПО ИЗОБРАЖЕНИЮ ЛИЦА<sup>1</sup>

*Ефимов И.Н., аспирант, ФГБОУ ВО «СамГТУ», e-mail: Mr.Efimov.IN@gmail.com;*

*Косолапов А.М., д.т.н., профессор, в.н.с., ФГБОУ ВО «СамГТУ»;*

*Ефимов Н.А., к.т.н., доцент, ФГБОУ ВО «СамГУПС».*

### METHOD AND SYSTEM FOR SPOOFING DETECTION AND APPARATUS OF VIDEO IMAGE

*Efimov I.N., Kosolapov A.M., Efimov N.A.*

*A description, advantages and disadvantages of the original method of authentication recognizable object, based on the evaluation of repeated dispersal of brightness conjugate pixels of the object image. The model of attacks on a person's biometric identification system. Possible action by malicious deception systems by spoofing.*

**Key words:** method of authentication recognizable, dispersal of brightness, model of attacks, person's biometric identification system.

**Ключевые слова:** компьютерное зрение, методы подтверждения подлинности распознаваемого объекта, распознавание образов, системы контроля доступа, спуфинг атаки.

#### Введение

Большинство систем биометрического распознавания не устойчиво к атакам спуфинга. Системы контроля в большей степени нацелены на распознавание лица человека и не защищают от подмены распознаваемого объекта. В [11] авторы продемонстрировали, успешный взлом коммерческих систем распознавания лиц, используя фотографию или видеозапись зарегистрированного пользователя. Системы биометрического распознавания нуждаются во внедрении эффективных методов подтверждения подлинности распознаваемого объекта.

В следующих публикациях [4, 5, 9, 10, 12-14] описан ряд подходов, используемых исследователями для подтверждения подлинности распознаваемого объекта. Однако их существенными недостатками являются высокая вычислительная сложность, неудобство для конечного пользователя, использование специализированного оборудования или отсутствие защиты перед использованием злоумышленниками фотографий, видеозаписей или фотомасок лица зарегистрированного пользователя.

В работе описаны математическая модель, метод и алгоритм подтверждения подлинности распознаваемого объекта. Метод подтверждения подлинности объекта относится к классам, использующим воздействие на пользователя (классификация представлена в [2]).

*Приводятся описание, достоинства и недостатки оригинального метода подтверждения подлинности распознаваемого объекта, на основе многократной оценки рассеивания яркостей сопряжённых пикселей изображений объекта. Представлена модель атак на биометрическую систему распознавания человека по изображению лица. Рассмотрены возможные действия злоумышленников по обману систем распознавания с помощью подмены распознаваемого объекта (атаки спуфинга). Описаны требования к системе, способной предотвратить подмену распознаваемого объекта.*

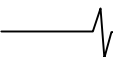
Предложенное решение не использует специализированное оборудование, что является преимуществом по сравнению с аналогами из данного класса, удобнее для конечного пользователя и может быть легко интегрирован в существующие системы распознавания лиц. Метод, математическая модель и алгоритм найдут применение в системах распознавания изображений лиц человека в качестве самостоятельного решения или в виде дополнительного средства предотвращения спуфинг атак. Результаты исследования защищены патентом РФ [8].

#### Модель атак на биометрическую систему

При разработке методов защиты биометрических систем необходимо определить все возможные виды угроз и описать модель атак. На данный момент уже составлены модели атак на различные биометрические системы [7, 16]. В данном пункте описана модель атак на системы распознавания по изображению лица. Определены девять наиболее уязвимых мест для атак злоумышленников в общей схеме биометрической системы, представленной на рис.1.

Для осуществления успешных атак на системы биометрического распознавания, злоумышленник должен обладать навыками в различных специализированных областях, а также иметь информацию о изъянах в оборудовании и аппаратной реализации, структуре и способе организации базы данных, методах расчёта и сопо-

<sup>1</sup> Работа поддержана грантом фонда содействия развитию малых форм предприятий («У.М.Н.И.К.» I полугодие, Самара, 2014).



ставления информативных признаков, о подсистеме взаимодействия с базой данных и о других моделях и методах, заложенных в реализованную систему биометрического распознавания [7]. Возможными целями злоумышленника могут являться:

- нелегальный доступ одного или нескольких незарегистрированных пользователей к охраняемой информации или на объекты предприятия;
- запрет в доступе зарегистрированному пользователю;
- полный отказ в работе всей системы биометрического распознавания.

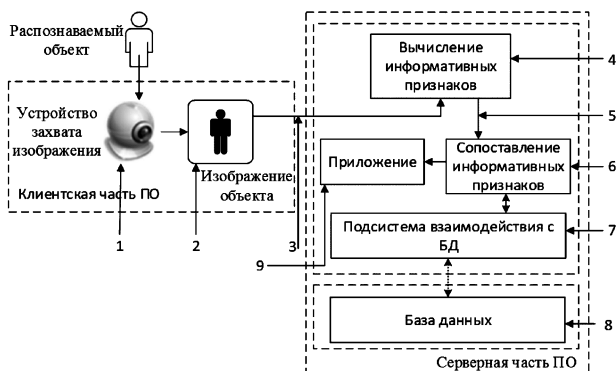


Рис. 1. Общая схема биометрической системы распознавания пользователей с обозначенными атаками

На рис. 1 приведены места для возможных атак на систему распознавания по изображению лица, которые также могут быть применены и для других существующих биометрических систем. Цифры на рис. 1 соответствуют пунктам в следующем списке:

1. **Устройство захвата изображения.** Представление устройству захвата изображения поддельных биометрических данных. Для различных систем биометрического распознавания используют такие объекты, как искусственный палец, копия подписи, маска лица и т.д.

2. **Канал связи между устройством захвата изображения и клиентской частью биометрической системы.** Замена сигнала, полученного с устройства захвата изображения, и передача для дальнейшей обработки в клиентскую часть биометрической системы.

3. **Канал связи между клиентской и серверной частями системы биометрического распознавания.** Для совершения атаки поддельная версия сигнала, подготовленная злоумышленником, передаётся в систему биометрического распознавания, минуя устройство захвата изображения и клиентскую часть программного обеспечения.

4. **Модуль вычисления информативных признаков.** Атака производится на модуль вычисления информативных признаков, таким образом, чтобы модуль генерировал информативные признаки, предварительно выбранные злоумышленником.

5. **Канал передачи информативных признаков.** Результат работы модуля вычисления информативных признаков подменяются злоумышленником (предполагается, что метод вычисления признаков известен злоумышленнику). Данный вид атаки актуален в случае если информативный признак передаётся удалённо.

6. **Модуль сопоставления информативных признаков.** Модуль сопоставления информативных признаков

повреждается таким образом, чтобы формировать заранее выбранные злоумышленником оценки соответствия.

7. **Канал связи и подсистема взаимодействия с БД.** Целью данной атаки является перехват и модификация информативных признаков, полученных от БД через канал связи и подсистему взаимодействия.

8. **Элемент базы данных, содержащий информативные признаки.** Злоумышленником модифицируется один или несколько элементов в базе данных. Подобная модификация может привести либо к ошибочной выдаче доступа к системе злоумышленнику, либо к отказу в доступе зарегистрированному пользователю.

9. **Приложение.** Производится подмена окончательного решения о выдаче доступа пользователю к системе. Это означает, что система распознавания была полностью отключена, даже если база информативных признаков и система распознавания образов находятся в работоспособном состоянии.

Все выше перечисленные атаки, кроме атаки на устройство захвата изображения, являются общими для всех биометрических систем. Для защиты систем биометрического распознавания от подобных атак необходимо использовать методы шифрования и цифрового кодирования. Наибольшую угрозу для биометрических систем, использующих изображение лица в роли объекта распознавания, представляют атаки на устройство захвата изображений (спуфинг атаки). Злоумышленник имеет непосредственный доступ к видеокамере, а использовать методы цифрового кодирования и шифрования не представляется возможным. По этой причине необходимо более подробно остановиться на спуфинг атаках. Далее представлены возможные виды атак на устройство захвата изображения.

### Постановка задачи подтверждения подлинности распознаваемого объекта

В отличие от систем, где требуется дополнительное оборудование (распознавание по пальцу, сетчатке глаза, кровеносным сосудам и др.), при распознавании человека по изображению лица очень легко создать копию распознаваемого объекта. Все, что необходимо, это фотография человека, которую легко можно обнаружить в интернете или сфотографировать объект на расстоянии. Одной из задач исследования является детектирование подмены представленного видеокамере объекта. Данная задача является не всегда лёгкой даже для человека. Решение должно обладать низкой вычислительной сложностью и высокой вероятностью верного обнаружения подмены распознаваемого объекта, без использования дополнительного специализированного оборудования. В зависимости от направленности биометрической системы, спуфинг атаки могут иметь различные уровни сложности. В таблице 1 рассмотрены возможные виды угроз и меры защиты от подмены распознаваемого объекта. В данном исследовании не рассматриваются такие виды атак, как грим или силиконовая маска. В соответствии с приведёнными в табл. 1 угрозами и представленными мерами защиты, решение задачи подтверждения подлинности распознаваемого объекта должно содержать следующие действия:

Таблица 1. Виды угроз и меры защиты от подмены объекта

Угроза	Описания	Мера защиты
Фото	Злоумышленник предъявляет видеокамере распечатанные фотографии высокого разрешения, либо цифровые фотографии с экрана какого-либо устройства.	Человеческое лицо это 3D объект, поэтому различным регионам поверхности лица присущи различные коэффициенты отражения, рассеяния, преломления и поглощения. Бумага и экран устройства указанным свойством не обладают. Необходимо анализировать изображения объекта с различной интенсивностью освещения.
Видеозапись	Системе распознавания предъявляется видеозапись объекта, подготовленная в реальной жизни, либо серия искусственно сформированных изображений.	Необходимо сформировать серию изображений объекта с различной интенсивностью освещения в случайные моменты времени, чтобы избежать возможного совпадения.
Устройство с экраном + ПО	Изображение объекта, изменённое специализированной программой, предъявляется системе распознавания. Программа способна изменять заранее подготовленное изображение объекта, имитируя моргание глаз, падение света как на 3D объект (подсвечиваются соответствующие регионы) и т.д.	Следует производить анализ изображений после формирования серии изображений, чтобы сократить время формирования решения о подмене. Тем самым предотвратить возможность предъявления сенсору подложных изображений.
Фото маска	Видеокамере предъявляется распечатанное изображение лица человека с прорезями для глаз и губ.	Для предотвращения атаки необходимо анализировать изображения с различной интенсивностью освещения объекта.
3D муляж	Сенсору предъявляют 3D муляж лица человека.	Следует анализировать биометрическую активность объекта.
Грим/ Силиконовая маска	Изменение внешности злоумышленника с целью обмана системы защиты.	Необходимо использовать дополнительное дорогостоящее оборудование.

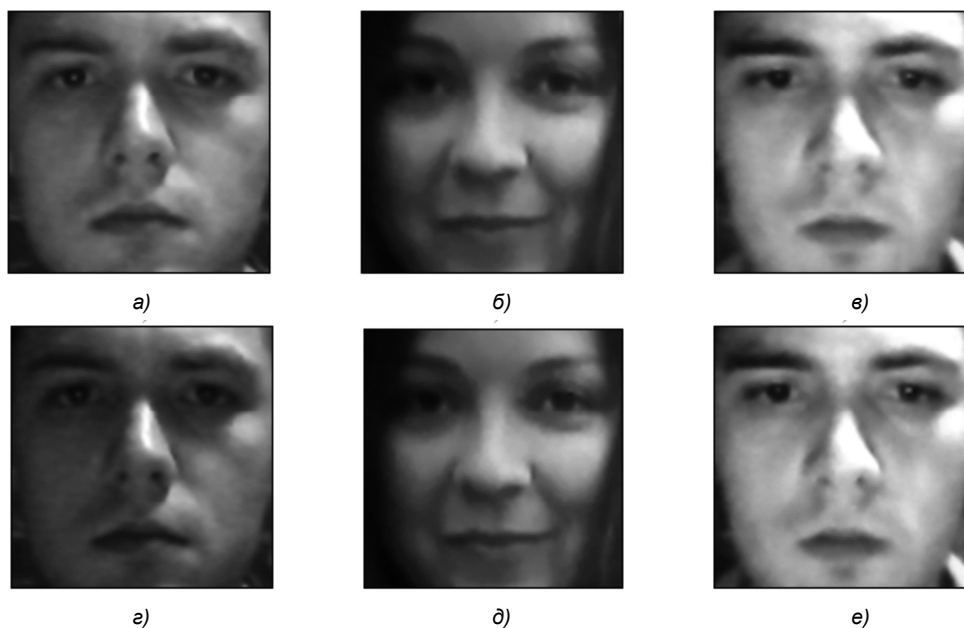


Рис. 2. Примеры изображений с различной интенсивностью освещения:  
а), г) реальный человек; б), д) фотография; в), е) экран телефона

– формирование серии изображений объекта с различной интенсивностью освещения объекта в случайные моменты времени;

– анализ изображений распознаваемого объекта с различной интенсивностью освещения. Анализ следует производить после проведения всех подготовительных операций по подготовке серии изображений, чтобы сократить время формирования решения о подмене. Тем самым предотвратить возможность предъявления сенсору подложных объектов;

– оценка биометрической активности объекта, для обнаружения случаев предъявления 3D муляжа лица зарегистрированного человека.

#### Метод подтверждения подлинности распознаваемого объекта

Ниже приведены оригинальные метод, математическая модель и алгоритм подтверждения подлинности распознаваемого объекта, не требующие специализированного оборудования, и, как следствие, внедрение не повлечёт за собой существенного удорожания системы контроля доступа. Одним из способов всестороннего изучения и оптимизации параметров объекта, явления или процесса называют математическое моделирование. Математическая модель, представленная в данном подразделе, является совокупностью частных моделей, описывающих процесс подтверждения подлинности объекта.

Визуальный осмотр изображений реального человека и подложного объекта демонстрирует, что изображения могут быть очень похожи, даже человеческому глазу тяжело определить подмену. Ниже представлены изображения реального человека (рис. 2а, г), фотографии на бумажном носителе (рис. 2б, д) и экрана телефона (рис. 2в, е) с различной интенсивностью освещения.

Тем не менее, некоторые различия между изображениями лица человека и подложным объектом могут стать очевидной, если образы перевести в надлежащее пространство признаков. Человеческое лицо это 3D объект, поэтому различным регионам поверхности лица присущи различные коэффициенты отражения, рассеяния, преломления и поглощения. Бумага и экран устройства указанным свойством не обладают. Есть основание предполагать, что текстурные свойства изображений лица реального человека и подложного объекта будут отличаться при различной интенсивности освещения (рис. 2). При изменении интенсивности освещения появятся текстурные артефакты и неравномерно изменится освещённость регионов распознаваемого объекта. Текстурные артефакты будут хорошо заметны в местах однородных областей, например, на лбу или щеках, а неравномерное изменение освещённости в областях глаз и рта. Описанные изменения освещённости будут отчётливо видны при анализе отношений яркости пикселей изображений объекта до подсветки и после (рис. 2):

$$q_i(x, y) = \frac{q_{0,i}(x, y)}{q_{1,i}(x, y) + 1}, \quad (1)$$

где  $i$  – номер текущей проверки подлинности предъявляемого объекта;  $q_{b,i}(x, y)$  – значение яркости пикселя массива  $I_{b,i}$ ; соответствующего изображению лица пользователя;  $b$  – идентификатор массива пикселей, обозначающий режим подсветки, при котором производилась подготовка массива,  $b = \overline{0,1}$ ;  $x$  и  $y$  координаты рассматриваемого пикселя,  $x = \overline{0, W-1}$ ,  $y = \overline{0, H-1}$ ;  $W$  и  $H$  – количество пикселей, соответствующих ширине и высоте массива  $I_{b,i}$ .

Следует оценить меру рассеивания всех значений  $q_i(x, y)$ . Для оценки степени рассеивания относительно среднего значения (математического ожидания), вычисляют дисперсию:

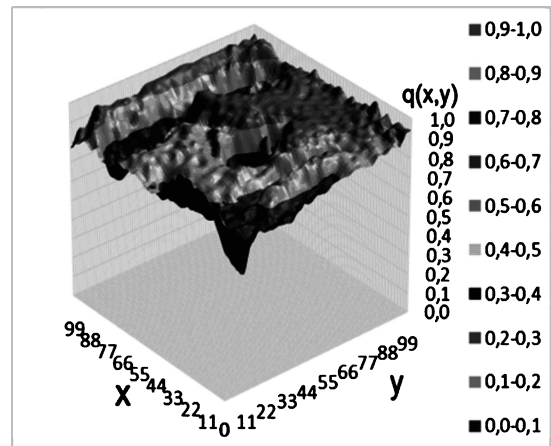
$$D(I_{0,i}, I_{1,i}) = \frac{1}{WH} \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} (q_i(x, y) - \overline{q_i})^2, \quad (2)$$

где  $\overline{q_i}$  – среднее значение или математическое ожидание дискретной случайной величины  $q_i(x, y)$  вычисляется по формуле:

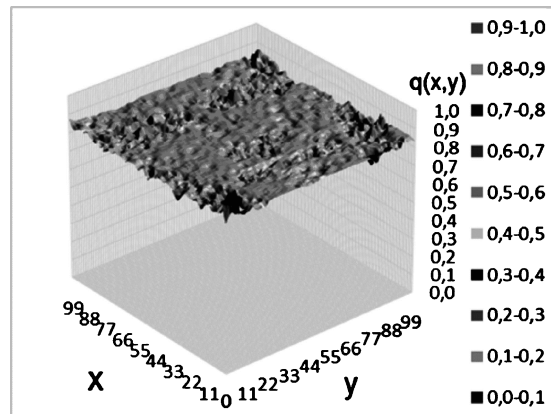
$$\overline{q_i} = \frac{1}{WH} \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} q_i(x, y). \quad (3)$$

Чем больше дисперсия, тем больше рассеяны значения и соответственно, тем более рельефную поверхность имеет распознаваемый объект. На рис. 3 приведены трёхмерные графики, построенные на основе значений  $q_i(x, y)$ , для оценки поверхностей различных объектов (реального лица пользователя, фотографии,

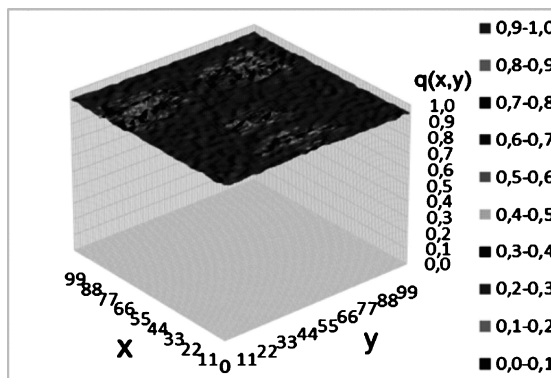
фото маски и экрана устройства).



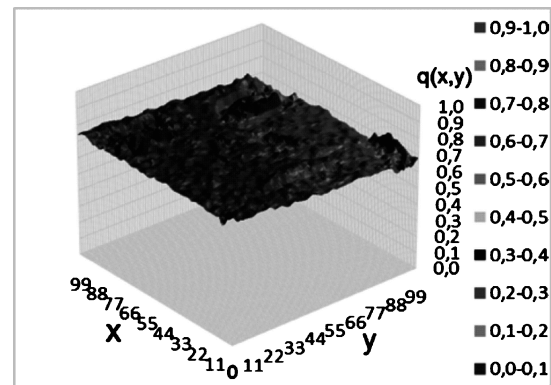
а) поверхность реального объекта  $D = 0,003564$



б) поверхность фотографии на бумажном носителе  $D = 0,000174$



в) поверхность экрана устройства  $D = 0,000039$



г) поверхность фотомаски  $D = 0,000732$

Рис. 3. Графики поверхностей

Оценка меры рассеивания величины  $q_i(x, y)$  способствует уменьшению вероятности взлома системы защиты с помощью плоского объекта (фотографии, фото маски, экрана устройства). Но остаётся опасность взлома подбором, т.е. случайного совпадения вспышки и изменения изображения на экране устройства с помощью соответствующего программного обеспечения. Чтобы исключить подобную возможность, необходимо производить подтверждение подлинности распознаваемого объекта многократно и в различные моменты времени:

$$F = \sum_{i=1}^Q |X(I_{0,i}, I_{1,i}) - Y_i|, \quad (4)$$

где  $i$  – номер текущей проверки подлинности предъявляемого объекта;  $Y_i$  – значение последовательности случайных чисел  $Y$ ;  $Q$  – количество проверок подлинности предъявляемого объекта;  $\sum_{i=1}^Q |X(I_{0,i}, I_{1,i})|$  – значение оценки рельефности распознаваемого объекта имеет вид:

$$X(I_{0,i}, I_{1,i}) = \begin{cases} 1, & D(I_{0,i}, I_{1,i}) \leq \tau_{ROC}; \\ 0, & D(I_{0,i}, I_{1,i}) > \tau_{ROC}, \end{cases} \quad (5)$$

где  $\tau_{ROC}$  – пороговое значение, при котором разница значений ( $TPR - FPR$ ) максимальна;  $TPR$  – процент правильного распознавания случаев представления реального лица видеокамере (True Positives Rate);  $FPR$  – процент пропуска подмены распознаваемого объекта (False Positives Rate). Учитывая вероятность ошибки в расчёте  $X(I_{0,i}, I_{1,i})$ , необходимо ввести пороговое значение  $\tau_{spf}$ . Значение  $\tau_{spf}$  варьируется в зависимости от требуемой степени безопасности системы контроля доступа. Выполнение неравенства, представленного ниже, означает, что на вход системы подано объёмное лицо человека, иначе алгоритм детектирует наличие подмены:

$$F \leq \tau_{spf}, \quad (6)$$

$X(I_{0,i}, I_{1,i})$  и  $Y_i$  могут принимать значения только из множества  $\{0, 1\}$ , где  $Y_i = 0$  соответствует изображению с выключенной подсветкой, а  $Y_i = 1$  – с включенной. Для модели (4) воспользуемся операцией сложения по модулю 2 (XOR). Ниже представлено преобразованное выражение (4):

$$\sum_{i=1}^Q X(I_{0,i}, I_{1,i}) \oplus Y_i \leq \tau_{spf}, \quad (7)$$

В соответствии с методом подтверждения подлинности, разработан алгоритм, состоящий из следующей последовательности действий:

**Шаг 1.** Создаётся последовательность случайных чисел  $Y$ .

**Шаг 2.** Формируется массив пикселей  $I_{0,i}$  без подсветки.

**Шаг 3.** Формируется массив пикселей  $I_{1,i}$ , при этом, если следующее значение в последовательности  $Y_i = 1$ , то массив формируется при включённой подсветке, а

если  $Y_i = 0$ , то при выключенной.

**Шаг 4.** Принимается решение с учётом шага 3, если  $i$ -ая цифра в последовательности  $Y$  не последняя, то следует перейти к шагу 2, иначе к шагу 5.

**Шаг 5.** Формируется совокупность значений оценки рельефности распознаваемого объекта  $X(I_{0,i}, I_{1,i})$ .

**Шаг 6.** Проводится сравнительный анализ последовательности  $Y$  и совокупности значений  $X(I_{0,i}, I_{1,i})$ .

**Шаг 7.** Принимается решение о подлинности объекта.

Представленный алгоритм можно модифицировать: добавлять новые блоки или заменять старые блоки, с условием сохранения основной идеи исследования. Вместо блока подсветки возможно использование лампы накаливания или компьютерного монитора. Для данного алгоритма приемлемы небольшие изменения интенсивности освещения и как следствие снижается воздействие на пользователя.

### Характеристики технических устройств, используемых для подтверждения подлинности объекта

В соответствии с методом подтверждения подлинности имеем: объект, ориентированный нормально к видеокамере, с коэффициентом отражения  $\rho$ , освещаемый световым потоком  $\Phi_b$  и источник света, сопряжённым с видеокамерой. Известны параметры ПЗС-матрицы, параметры оптики видеокамеры и параметры источника света.

Для расчёта яркости источника освещения  $B_b$ , взаимного расположения источника освещения и распознаваемого объекта воспользуемся законом аддитивности освещённости [1]. При воздействии на объект несколькими источниками света, его освещённость равна сумме освещённостей от каждого источника света. Если распознаваемый объект до включения подсветки имеет освещённость  $E_n$ , то закон аддитивности освещённости для случая с пониженным уровнем освещения примет вид:

$$E_{c0} = E_0 + E_n, \quad (8)$$

для случая с повышенным уровнем освещения:

$$E_{c1} = E_1 + E_n, \quad (9)$$

где  $E_{c0}$  и  $E_{c1}$  – общая энергетическая освещённость объекта,  $E_0$  и  $E_1$  – энергетическая освещённость от источника света системы. Вычтем (8) из (9):

$$E_{c1} - E_{c0} = E_1 - E_0. \quad (10)$$

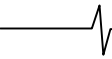
Соотношения для силы света  $J_b$  и энергетической освещённости объекта от точечного источника будет иметь вид [6]:

$$E_b = \frac{\Phi_b}{S_0} = \frac{\Omega \cdot J_b}{S_0} = \frac{J_b \cdot \cos \alpha}{L^2}, \quad (11)$$

$$\Omega = \frac{S_0 \cdot \cos \alpha}{L^2}, \quad (12)$$

где  $\Phi_b$  – световой поток, исходящий в пределах телесного угла  $\Omega$ ;  $S_0$  – площадь поверхности распознаваемого объекта;  $L$  – расстояние от источника света до объекта;  $\alpha$  – угол между направлением источника и перпендикуляром к освещаемой поверхности.





В случае протяжённого источника света, следует говорить о силе света каждого элемента источника и площади его поверхности. Тогда под  $\Phi_b$  следует понимать световой поток, излучаемый каждым элементом поверхности источника. В частности, можно разделить всю поверхность источника света на отдельные  $Z$  участков существенно меньших расстояния до облучаемого объекта и принять их за точечные источники света. Каждый участок источника освещения имеет свою площадь  $S_z$ , яркость  $B_{bz}$ , угол между направлением источника и перпендикуляром к освещаемой поверхности  $\alpha_z$  и расстояние до объекта  $L_z$ . Полная освещённость объекта от всех отдельных участков протяжённого источника света может быть вычислена следующим образом [3]:

$$E_b = \sum_{z=1}^Z \frac{B_{bz} \cdot S_z \cdot \cos \alpha_z}{L_z^2}, \quad (13)$$

где  $z$  – рассматриваемый номер участка.

Угол обзора источника освещения  $\theta_z$  – угол, при котором сохраняется удовлетворительный уровень яркости. Яркость  $B_{bz}$  от каждого региона источника освещения будет одинакова, если  $\alpha_z$  не превышает  $\theta_z$ , а линейные размеры источника освещения не превышают значения  $2 \operatorname{tg}(\theta_z) \cdot L$ . При условии, что яркость  $B_{bz}$  всей поверхности источника освещения одинакова, целесообразно использовать следующую формулу для полной освещённости объекта [3]:

$$E_b = \frac{B_b \cdot S \cdot \cos \alpha}{L^2}, \quad (14)$$

где  $B_b$  – яркость поверхности источника освещения;  $S$  – площадь светящейся поверхности протяжённого источника освещения. Откуда правая часть соотношения (10) равна:

$$E_1 - E_0 = \frac{(B_1 - B_0) \cdot S \cdot \cos \alpha}{L^2} = E_{c1} \quad (15)$$

Выражение для расчёта яркости источника освещения  $B_b$ , взаимного расположения источника освещения и распознаваемого объекта следующее:

$$L = \sqrt{\frac{(B_1 - B_0) \cdot S \cdot \cos \alpha}{E_{c1} - E_{c0}}}. \quad (16)$$

Световое излучения от распознаваемого объекта попадает через объектив на светочувствительный слой ПЗС-матрицы видеокамеры и создаёт на нем оптическое изображение распознаваемого объекта с эквивалентной освещённостью  $E_{sb}$ . Эквивалентная освещённость  $E_{sb}$  показывает степень уменьшения реальной освещённости объекта  $E_{cb}$  при прохождении света через объектив камеры (коэффициент пропускания объектива видеокамеры  $K_a$ ), с учетом спектральной чувствительности видеокамеры  $K_\lambda$ , а также с учетом светосилы объектива и коэффициента диффузного отражения распознаваемого объекта  $\rho$ :

$$E_{sb} = E_{cb} K_\lambda K_a \rho \frac{1}{4F^2}, \quad (16)$$

где  $F$  – отношение фокусного расстояния объектива к диафрагме видеокамеры ( $F$ -число). Выражение для реальной освещённости объекта  $E_{sb}$  с учётом эквивалентной  $E_{cb}$  имеет вид:

$$E_{cb} = \frac{E_{sb} 4F^2}{K_\lambda K_a \rho}. \quad (18)$$

Эквивалентную освещённость оптического изображения объекта  $E_{sb}$  возможно рассчитать, используя массив яркостей оптического изображения объекта  $I_{b,i}$  согласно рекомендациям, описанным в стандарте федеральной комиссии связи (FCC) [15]:

$$E_{sb} = K_c \overline{g_b}, \quad (19)$$

$$\overline{g_b} = \frac{1}{WH} \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} g_b(x, y), \quad (20)$$

где  $K_c$  – коэффициент расчёта энергетической освещённости объекта;  $\overline{g_b}$  – средняя освещённость распознаваемого объекта в плоскости оптического изображения;  $x, y$  – координаты текущего пикселя;  $X \cdot Y$  – размеры изображения;  $g_b(x, y)$  – яркость пикселя, значение рассчитывается с помощью (21). Выражение для яркости пикселя, в соответствии с рекомендациями стандарта Федеральной комиссии связи (FCC) [15], имеет вид:

$$g(x, y) = 0,3 \cdot R_{x,y} + 0,59 \cdot G_{x,y} + 0,11 \cdot B_{x,y}, \quad (21)$$

где  $R_{x,y}$ ,  $G_{x,y}$  и  $B_{x,y}$  – цветовые составляющие пикселя.

Коэффициент  $K_c$  используется для пересчёта освещённости оптического изображения объекта  $\overline{g_b}$  в эквивалентную освещённость объекта  $E_{sb}$ . Для расчёта  $\overline{g_b}$  выбрана область в центре лба, неприкрытая волосами и головным убором. Данная область наиболее плоская и равномерно отражающая свет. Выбор данного региона обусловлен минимизацией погрешности, связанной с неравномерностью распределения освещённости при подсветке трёхмерных объектов. Подставим в выражение (16) соотношения (18) и (19). Итоговое выражение для расчёта яркости источника освещения  $B_b$  и взаимного расположения источника освещения и распознаваемого объекта примет вид:

$$L = \frac{(B_1 - B_0) \cdot S \cdot \cos \alpha \cdot K_\lambda \cdot K_a \cdot \rho}{(g_1 - g_0 + 0,001) \cdot 4F^2 \cdot K_c}. \quad (22)$$

Соотношение (22) необходимо использовать при калибровке системы подтверждения подлинности распознаваемого объекта. Для вычисления коэффициента  $K_c$ , следует на заданном расстоянии  $L$  сформировать два изображения распознаваемого объекта с заданными яркостями источника освещения  $B_0$  и  $B_1$ . Выражение для расчёта  $K_c$  представлено ниже:

$$K_c = \frac{(B_1 - B_0) \cdot S \cdot \cos \alpha \cdot K_\lambda \cdot K_a \cdot \rho}{(g_1 - g_0 + 0,001) \cdot (2FL)^2}. \quad (23)$$

При следующих значениях: яркости источника  $(B_1 - B_0) = 17 \text{ кд/м}^2$ , площади светящейся поверхности протяжённого источника освещения  $S = 0,0825 \text{ м}^2$ , спек-

тральной чувствительности видеокамеры  $K_\lambda = 0,6$ , коэффициента пропускания объектива видеокамеры  $K_a = 0,8$ , коэффициента диффузного отражения кожи лица  $\rho = 0,3$ , средней освещённости распознаваемого объекта в плоскости изображения  $(\overline{g_1} - \overline{g_0}) = 4,5$  лм/м<sup>2</sup>, отношения фокусного расстояния объектива к диафрагме видеокамеры ( $F$ -число)  $F = 0,25$ , расстояние от источника света до объекта  $L = 0,5$  м, значение коэффициента расчёта энергетической освещённости равно  $K_c = 0,72$ .

Для уменьшения воздействия на пользователя, при сохранении эффективности предложенного метода, производится расчёт максимально допустимого уровня вариации освещённости от  $E_0$  до  $E_1$ , в соответствии с коэффициентом пульсации. Коэффициент пульсации освещённости  $K_p$  – критерий оценки относительной глубины колебаний освещённости объекта источником света. Выражение для расчёта  $K_p$  имеет вид:

$$K_p = \frac{E_1 - E_0}{2E_m} 100\% \quad (24)$$

где  $E_1$  и  $E_0$  – максимальное и минимальное значения освещённости соответственно,  $E_m$  – среднее значение освещённости за период колебания.

В соответствии с условиями, налагаемыми на устройство освещения в «Гигиенические требования к естественному, искусственному и совмещённому освещению жилых и общественных зданий»,  $K_p = 20\%$ . Примем  $E_m = E_0$ . Откуда значение освещённости  $E_1$  не должно превышать  $1,4 \cdot E_0$ .

Для обеспечения устойчивости выделяемых признаков к изменениям внешней обстановки и параметров регистрирующих устройств необходимо проводить калибровку системы по объекту с известной спектральной характеристикой коэффициентом отражения и на заданном расстоянии между устройством освещения и объектом. С учётом коэффициента  $K_c$ , расстояния между устройством освещения и объектом  $L$ , характеристиками видеокамеры, а также ограничения  $E_1 < 1,4 \cdot E_0$ , необходимо в ручном или автоматическом режиме подобрать значения освещённостей  $E_0$  и  $E_1$  для источника света.

Общее и текущее время работы системы определяются соотношениями:

$$T = t \cdot N. \quad (25)$$

$$t_i = t \cdot i, \quad (26)$$

где  $t$  – время, затрачиваемое на обработку одного разряда.

$$t = t_{\text{под}} + 2 \cdot t_{\text{кам}} + t_{\text{пад}}, \quad (27)$$

$$t_{\text{омк}} = t_{\text{под}} + t_{\text{пад}}, \quad (28)$$

где  $t_{\text{под}}$  – время «подъема», за которое устройство освещения переходит в режим повышенной яркости,  $t_{\text{кам}}$  – время восприятия изображения видеокамерой,  $t_{\text{пад}}$  – время «падения», за которое устройство освеще-

ния переходит в режим пониженной яркости,  $t_{\text{омк}}$  – время отклика (black-white-black), за которое устройство освещения изменит режим яркости от минимального значения до максимального значения и обратно. Перечисленные временные рамки зависят от технических характеристик устройства освещения и видеокамеры. Тесты системы подтверждения подлинности производились на компьютере с процессором Intel Core i3 тактовая частота – 2,40ГГц и ОЗУ 4,00ГБ, видеокамерой SCB-0350M – 0,3 Мп при скорости работы 11 кадров/с, откуда  $t_{\text{кам}} = 91$  мс. Устройством освещения выбран ЖК-дисплей со следующими характеристиками: диагональ 17,3", значением яркости при максимальном режиме 224 кд/м<sup>2</sup>, а при минимальном режиме – 1,34 кд/м<sup>2</sup> и  $t_{\text{омк}} = 17$  мс. Время для обработки одной проверки  $t = 17 + 182 = 199$  мс. Примем  $t = 0,2$  с. С увеличением количества проверок –  $Q$  увеличивается время работы  $T$  (формула 25) и уменьшается вероятность подбора последовательности  $Y$ . Минимальным значением  $Q$  выбрано 20, откуда общее время  $T = 4$  с.

### Экспериментальные исследования

Для проведения опытов была подготовлена экспериментальная БД (табл. 2). В обучающей выборке находятся 160 случаев предоставления видеокамере лиц реальных людей, а также 400 случаев предоставления видеокамере фотографий, видеозаписей, экранов устройств и фотомасок. В тестовой выборке содержатся 320 случаев предоставления видеокамере лиц реальных людей, а также 700 случаев предоставления видеокамере фотографий, видеозаписей и фотомасок. Каждый случай предоставления объекта видеокамере дублировался при повышенной интенсивности освещения для реализации алгоритмов подтверждения подлинности.

Таблица 2. Экспериментальная БД изображений лиц

	Тестовая выборка	Обучающая выборка
Реальные лица	320 изображений, 40 человек	160 изображений, 40 человек
Попытка обмана	700 изображений различных лиц	400 изображений, 40 человек

Примеры изображений представлены на рис. 2. Существуют различные БД для экспериментальной проверки систем подтверждения подлинности [9, 17], но они не подходят в виду специфики используемого метода подтверждения (необходимо, чтобы каждый случай предоставления объекта видеокамере дублировался при повышенной интенсивности освещения). В БД содержатся изображения лиц людей с различным возрастом, мимикой и ракурсом. Изображения подготавливались в четырёх различных условиях освещения. Распознаваемый объект предъявлялся системе подтверждения подлинности в двух режимах:

Фотографию или экран устройства фиксируют на неподвижной опоре, чтобы избежать движения во время попытки обмана.

Пользователь, используя свои собственные руки, предъявляет сенсору фотографию или экран устройства.

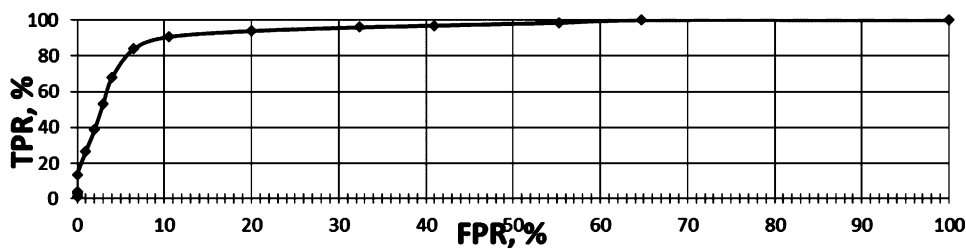


Рис. 4. График экспериментальной зависимости частоты правильного обнаружения подмены объекта от частоты ложных тревог

В данном случае, небольшое движение объекта относительно камеры может внести шум на изображения.

На рис. 4 представлены результаты экспериментальных исследований разработанной системы подтверждения подлинности на основе ROC-анализа (Receiver Operator Characteristic). На графике 4 по оси ординат указан процент правильного распознавания случаев представления реального лица видеокамере (True Positives Rate), а по оси абсцисс указан процент пропуска подмены распознаваемого объекта (False Positives Rate). Для рис. 4 значение площади под криво AUC (Area Under Curve) – 0,944. Значение порога в точке отсечения  $\tau_{ROC}$  (при котором разница значений  $(TPR - FPR)$  максимальна) равно 0,001557. При указанном значении процент правильного распознавания случаев представления реального лица видеокамере не менее 91%, а процент пропуска подмены распознаваемого объекта не более 10%.

### Заключение

Представленное решение найдёт применение в больших информационных средах для обнаружения подмены распознаваемого объекта, например, в дистанционном обучении, при сдаче экзаменов. Для внедрения описанного решения в существующие системы контроля доступа не требуется дорогостоящее специализированное оборудование. Необходимо и достаточно устройства освещения, в роли которого возможно использование источника дневного или инфракрасного света.

### Литература

1. Борн М. Основы оптики / М. Борн, Э. Вольф // М: Наука. – 1973.– 720с
2. Ефимов И.Н. Классификация способов подтверждения подлинности распознаваемого объекта / И.Н. Ефимов, А.М. Косолапов // Сборник материалов международной научно-технической конференции «Перспективные информационные технологии». – Самара: СГАУ. – 2016. – С.109-112.
3. Иванов В.П. Трёхмерная компьютерная графика / В.П. Иванов, А.С. Батраков, Г.М. Полищук // М.: Радио и связь. – 1995.– 224с.
4. Костылев Н.М. Модуль обнаружения витальности лица по спектральным характеристикам отражения кожи человека / Н. М. Костылев, А. В. Горевой // Инженерный журнал наука и инновации. – 2013. – № 9 – С.13.
5. Костылев Н.М. Обнаружение витальности человека

по спектральным характеристикам кожи лица / Н.М. Костылев, Ф. А. Трушкин, В. Я. Колочкин // Вестник МГТУ им. Н.Э. Баумана. Серия «Приборостроение». – 2012. – № 2 – С. 75-85.

6. Пустынский И.Н. К расчету освещенности изображения и числа сигнальных электронов в телевизионном датчике на ПЗС-матрице / И.Н. Пустынский, Е.В. Зайцева // Доклады ТУСУРа. – 2009. – № 2 – С. 5-10.

7. Ручай А.Н. Модель атак и защиты биометрических систем распознавания диктора / А.Н. Ручай // Доклады ТУСУРа. – 2011. – № 1 – С. 96-100.

8. Способ и устройство распознавания рельефности изображения лица: пат. 2518939 Рос. Федерация: МПК А 61 В 3/10 / И.Н. Ефимов, А.М. Косолапов; заявитель и патентообладатель ФГБОУ ВПО «Самарский государственный университет путей сообщения». – № 2013109943 ; заявл. 05.03.2013 ; опубл. 11.04.2014.

9. Chingovska I. On the Effectiveness of Local Binary Patterns in Face Anti-spoofing / I. Chingovska, A. Anjos, S. Marcel // Biometrics Special Interest Group, 2012 BIOSIG - Proceedings of the International Conference of the.– 2012.

10. Chingovska I. Anti-spoofing in action: Joint operation with a verification system / I. Chingovska, M. Switzerland, A. Anjos, S.Marcel // Computer Vision and Pattern Recognition Workshops (CVPRW) . – 2013. – С. 98-104.

11. Duc N.M. Your face is not your password / N.M. Duc // Black Hat Conference. – 2009. – С. 1-16.

12. Freitas P.T. LBP- TOP Based Countermeasure against Face Spoofing Attacks / F. Pereira, T. Anjos, A. Martino, J. Mario, M. Sébastien // International Workshop on Computer Vision With Local Binary Pattern Variants – ACCV. – 2013. – С. 121-132.

13. Housam K.B. Face spoofing detection based on improved local graph structure/ K.B. Housam // IEEE Computer Society. – 2014.

14. Maatta J. Face spoofing detection from single images using micro-texture analysis / J. Maatta, A. Hadid, M. Pietikainen // Biometrics (IJCB), International Joint Conference on Biometrics Compendium, IEEE. – 2011.

15. Plataniotis, K.N. Color image processing and applications/K.N. Plataniotis, A.N. Venetsanopoulos – Engineering Monograph: Springer Science & Business Media. – 2000. – 65 с.

16. Ratha N.K. Enhancing security and privacy in biometrics-based authentication systems / N.K. Ratha, J.H. Connell, R.M. Bolle // IBM Syst. J. – 2001. – Т. 40 – № 3 – С. 614-634.

17. Zhang, Z. A face antispoofing database with diverse attacks / J. Yan, S. Liu, Z. Lei // Biometrics Compendium, IEEE. – 2012. – С. 26-31.

УДК 519.642

## РАЗРАБОТКА СГЛАЖИВАЮЩИХ ФИЛЬТРОВ ПРОЕКЦИОННЫХ ДАННЫХ В ЗАДАЧАХ ДВУМЕРНОЙ ТОМОГРАФИИ

*Лихачёв А.В., д.т.н., старший научный сотрудник Института автоматизации и электрометрии Сибирского отделения РАН, e-mail: ipm1@iae.nsk.su.*

### DESIGN OF PROJECTION DATA SMOOTHING FILTERS FOR TWO-DIMENSIONAL TOMOGRAPHY

*Likhachov A.V.*

*A new method for calculation of the projection data smoothing filter coefficients is proposed. It is based on a previous study of the effect of the distortion type on the reconstruction accuracy. The coefficients are determined by means of minimization on the specified set of the deviation between the correlation function of the smoothed noise and the given Gauss curve. A computational experiment shows, in particular, that the developed algorithm provides a reconstruction error of one and a half times less than the moving average filter.*

**Key words:** Two-dimensional tomography, correlation function of noise, projection data smoothing.

**Ключевые слова:** двумерная томография, корреляционная функция шума, сглаживание проекционных данных.

#### Введение

В работе рассматривается задача двумерной лучевой томографии [1, 2]. С математической точки зрения она сводится к реконструкции функции двух переменных  $g(x, y)$  по набору интегралов от неё вдоль прямых линий. Ниже он будет обозначаться через  $f(p, \varphi)$ , где  $p$  – расстояние от прямой до начала координат,  $\varphi$  – угол наклона её нормали к оси  $X$ . Для определённости возьмём параллельную схему сканирования [1]. В этом случае искомая функция вычисляется по формуле обращения преобразования Радона, [2]:

$$g(x, y) = \frac{1}{4\pi^2} \int_0^\pi \int_{-\infty}^{\infty} |\omega| \tilde{f}(\omega, \varphi) \exp(i\omega(x \cos \varphi + y \sin \varphi)) d\omega d\varphi, \quad (1)$$

Знак «тильда» сверху означает преобразование Фурье по первой переменной. При алгоритмической реализации выражения (1) исходные данные целесообразно представлять в виде множества одномерных проекций  $f_\varphi(p)$ , каждая из которых получается из  $f(p, \varphi)$  если зафиксировать угол  $\varphi$ .

Известно, что описанная задача является некорректно поставленной [2, 3]. Как видно из (1), спектр каждой проекции умножается на модуль частоты (в англоязычной литературе эта операция называется гамп-фильтрацией). Тем самым усиливается высокочастотная составляющая шума, который неизбежно присутствует в зарегистрированных сигналах. Согласно теореме Винера-Хинчина [4], спектральная плотность мощности  $W_\xi(\omega)$  стационарного случайного процесса представляет собой Фурье-образ его корреляционной функции  $R_\xi(p)$ , в связи с чем последняя оказывает суще-

*Предложен новый метод определения коэффициентов сглаживающего фильтра проекционных данных в задачах томографии. В его основе лежит выполненное ранее исследование влияния различных типов помех на точность восстановления. Расчёт ведётся, исходя из минимизации на определённом множестве отклонения корреляционной функции сглаженного шума от гауссовой кривой с заданными параметрами. Проведен вычислительный эксперимент, в котором, в частности, получено, что разработанный алгоритм обеспечивает ошибку реконструкции в полтора раза меньшую, чем фильтр скользящего среднего.*

ственное влияние на результат реконструкции. Этот вопрос рассматривался в [5]. В частности, там было выведено выражение для дисперсии централизованного шума после гамп-фильтрации:

$$D_\xi^{\text{ramp}} = R_\xi^{\text{ramp}}(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{\Omega}^2(\omega) \tilde{R}_\xi(\omega) d\omega. \quad (2)$$

Через  $\tilde{\Omega}(\omega)$  в (2) обозначена частотная характеристика некоторой аппроксимации гамп-фильтра. Методы их проектирования и соответствующие исследования изложены в [1, 2, 6 – 10].

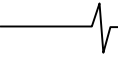
В работе [5] получен следующий важный результат. Для шума с корреляционной функцией в форме кривой Гаусса (гауссов шум) значение правой части (2) оказалось более чем в четыре раза ниже по сравнению с другими рассмотренными случайными сигналами (при одинаковой дисперсии помехи в исходных данных). Это легло в основу предлагаемого метода сглаживания.

#### Теоретическая часть

Будем считать, что статистические свойства помех не зависят от ракурса наблюдения. Примем аддитивную модель, согласно которой каждая зарегистрированная проекция представляет собой следующую сумму:

$$f_\varphi(p) = f_{0,\varphi}(p) + \xi(p). \quad (3)$$

Здесь  $f_{0,\varphi}(p)$  – точные данные, которые в рассмат-



риваемом приближении являются преобразованием Радона от  $g(x, y)$ ; через  $\xi(p)$  обозначена случайная функция с равным нулю математическим ожиданием.

Зададим нерекурсивный симметричный цифровой фильтр посредством выражения

$$f_{\varphi}^F(ih) = w_0 f_{\varphi}(ih) + \sum_{j=1}^n w_j f_{\varphi}((i-j)h) + \sum_{j=1}^n w_j f_{\varphi}((i+j)h), \quad (4)$$

где  $w_j$  – вещественные числа (коэффициенты фильтра),  $h$  – шаг дискретизации одномерных проекций. Ставится задача найти значения  $w_j$ , такие чтобы  $R_{\xi}^F(p)$  – корреляционная функция шума после фильтрации (4) – была близка к заданной функции  $G(p)$ . Учитывая результаты работы [5], положим

$$G(ih) = D_{\xi}^F \exp(-\lambda^2 (ih)^2). \quad (5)$$

Параметры  $D_{\xi}^F$  и  $\lambda$  в (5) не фиксированы, путём их варьирования производится регуляция степени сглаживания.

Пусть проекционные данные (3) содержат центрированный белый шум, для которого, как известно,

$$R_{\xi}((i-j)h) = D_{\xi} \delta_{ij}, \quad (6)$$

где  $D_{\xi}$  – дисперсия,  $\delta_{ij}$  – символ Кронекера. Проиллюстрируем разработанный метод вычисления коэффициентов  $w_j$  для нескольких значений  $n$ . Прежде всего, отметим, что поскольку фильтрации подвергается также и полезный сигнал, для того, чтобы его энергия сохранялась, должно выполняться условие

$$w_0 + 2 \sum_{j=1}^n w_j = 1. \quad (7)$$

При  $n = 1$  уравнение (4), записанное отдельно для шумовой составляющей, принимает вид

$$\xi^F(ih) = w_0 \xi(ih) + w_1 \xi((i-1)h) + w_1 \xi((i+1)h) \quad (8)$$

Возведём (8) в квадрат и возьмём математическое ожидание:

$$M \{ \xi^F(ih) \cdot \xi^F(ih) \} = M \{ (w_0 \xi(ih) + w_1 \xi((i-1)h) + w_1 \xi((i+1)h))^2 \}. \quad (9)$$

По определению левая часть (9) есть  $R_{\xi}^F(p)$  при нулевом значении аргумента. Приравняем её к  $G(0)$ . Преобразуя правую часть, с учётом (6), в конечном итоге получаем

$$G(0) = D_{\xi}^F = D_{\xi} (w_0^2 + 2w_1^2) \quad (10)$$

Решая систему, состоящую из уравнений (7) и (10), имеем

$$\begin{cases} w_0 = 1 - \frac{2}{3} \left( 1 \pm \sqrt{1 - \frac{3}{2}(1-\beta)} \right) \\ w_1 = \frac{1}{3} \left( 1 \pm \sqrt{1 - \frac{3}{2}(1-\beta)} \right) \end{cases}. \quad (11)$$

Через  $\beta$  обозначено отношение  $D_{\xi}^F / D_{\xi}$ .

Будем рассматривать параметр  $D_{\xi}^F$  как дисперсию шума после фильтрации (4). Чтобы числа  $w_0$  и  $w_1$  были

вещественными должно выполняться неравенство  $\beta \geq 1/3$ . Если  $\beta = 1/3$ , то  $w_0 = w_1 = 1/3$ , что соответствует усреднению в скользящем окне. Таким образом, именно оно обеспечивает наименьшую величину дисперсии. Однако значение  $n = 1$  является слишком малым, для того чтобы можно было учитывать форму корреляционной функции.

Перейдём к случаю  $n = 2$ . Аналогично тому, как было получено (10), находим

$$D_{\xi}^F = D_{\xi} (w_0^2 + 2w_1^2 + 2w_2^2). \quad (12)$$

Для определения трёх коэффициентов  $w_0, w_1, w_2$  нужно ещё одно уравнение. Запишем выражение для  $\xi^F((i+1)h)$ :

$$\begin{aligned} \xi^F((i+1)h) &= w_0 \xi((i+1)h) + \\ &+ w_1 \xi(ih) + w_2 \xi((i-1)h) + \\ &+ w_1 \xi((i+2)h) + w_2 \xi((i+3)h). \end{aligned} \quad (13)$$

Умножим (4) при  $n = 2$  на (13) и возьмём математическое ожидание. Заменяя, как и ранее,  $R_{\xi}^F(h)$  на  $G(h)$ , приходим к равенству

$$G(h) = D_{\xi}^F \exp(-\lambda^2 h^2) = 2D_{\xi} (w_0 w_1 + w_1 w_2). \quad (14)$$

В итоге имеем нелинейную систему, состоящую из (7), (12), (14), откуда выводится уравнение четвёртой степени для  $w_1$ :

$$\begin{aligned} 14w_1^4 - 12w_1^3 + \left( 3 + \frac{2G(h)}{D_{\xi}} - \beta \right) w_1^2 - \\ - \frac{6G(h)}{D_{\xi}} w_1 + \frac{3(G(h))^2}{2D_{\xi}^2} = 0. \end{aligned} \quad (15)$$

Если брать  $n > 2$ , то получение алгебраических уравнений, подобных (15), но более высокой степени, также возможно. Однако поиск их аналитических решений сопряжён со значительными трудностями. Поэтому был разработан приближённый метод, который мы поясним на примере  $n = 3$ . Умножая (4) на  $\xi^F((i+j)h)$ ,  $j = 0, 1, 2, 3$ , по аналогии с (10) и (14) получаем

$$G(0) = D_{\xi} (w_0^2 + 2(w_1^2 + w_2^2 + w_3^2)), \quad (16)$$

$$G(h) = 2D_{\xi} (w_0 w_1 + w_1 w_2 + w_2 w_3), \quad (17)$$

$$G(2h) = D_{\xi} (w_1^2 + 2(w_0 w_2 + w_1 w_3)), \quad (18)$$

$$G(3h) = 2D_{\xi} (w_0 w_3 + w_1 w_2). \quad (19)$$

Из последнего равенства следует

$$w_3 = \frac{G(3h)}{2D_{\xi} w_0} - \frac{w_1 w_2}{w_0}. \quad (20)$$

Подставляем (20) в (18) и выражаем отсюда  $w_2$ :

$$w_2 = \frac{G(2h) - D_{\xi} w_1^2 w_0 - G(3h) w_1}{2D_{\xi} (w_0^2 - w_1)}. \quad (21)$$

Далее производится перебор значений  $w_0$  и  $w_1$  с шагом  $dw$ : коэффициент  $w_0$  меняется в интервале  $[1/7 + d_w; 1 - d_w]$ , а  $w_1$  при заданном  $w_0$  – в интервале  $[d_w; w_0 - d_w]$ . Для каждой пары  $w_0, w_1$  коэффициенты  $w_2, w_3$  вычисляются по уравнениям (21), (20) соответ-

ственно. Если имеет место  $0 < w_2$ ,  $0 < w_3$ ,  $w_3 \leq w_2 \leq w_1$ , то рассчитывается невязка между правой и левой частями системы (16) – (19), и выбирается то решение, для которого она минимальна. Определённые путём описанной процедуры  $w_0$ ,  $w_1$ ,  $w_2$ ,  $w_3$  умножаются на нормировочную константу с целью обеспечить равенство (7). Отметим, что при этом будет выполняться условие  $w_3 \leq w_2 \leq w_1 \leq w_0$ . Такое требование объясняется тем, что помимо шума сглаживаются также и проекционные данные. Поэтому представляется разумным, чтобы в результате фильтрации функций  $f_{0,\varphi}(p)$  в некоторой точке с наибольшими весами входили значения из неё самой и её ближайшего окружения.

Если имеются решения  $\vec{w} \in R^4$ , обладающие приведёнными выше свойствами, то данным методом будет получено приближение (с точностью порядка  $d_w$ ) к одному из них. В противном случае он приведёт к псевдо решению, минимизирующему на заданном множестве невязку. Сходным образом в работе были вычислены коэффициенты  $w_j$  до  $n = 7$  включительно. Перебирались значения двух – четырёх неизвестных, остальные же выражались через них посредством относительно простых соотношений.

Рассматриваемая задача также может быть решена, исходя из минимизации среднеквадратичного отклонения корреляционной функции шума после фильтрации от функции (5). Однако в работе он не применялся по следующим причинам. Во-первых, условия, накладываемые на  $w_j$ , затрудняют применение этого подхода. Во-вторых, учитывая возможные приложения метода, можно заключить, что число  $n$  не может быть большим. Действительно, уже при  $n = 7$  усреднение происходит по  $2n + 1 = 15$  точкам. При такой ширине окна фильтра мелкомасштабная структура изображения оказывается существенно размытой. В то же время, для  $n \leq 7$  разработанный алгоритм оказался достаточно эффективным.

### Экспериментальная часть

При численном моделировании использовался математический фантом, показанный на рис. 1. Число ракурсов наблюдения, равномерно распределённых по углу в интервале от 0 до  $\pi$ , составляло 180. На каждом из них было 1025 отсчётов. На квадратной области реконструкции задавалась сетка 1025x1025 узлов. Благодаря выбранной форме фантома, максимальное значение на всех проекциях было примерно одинаковым и приблизительно равнялось 1,67. Некоррелированная шумовая составляющая  $\xi(p)$  в каждой точке представляла собой нормально распределённую случайную величину с нулевым средним и дисперсией  $D_\xi = \sigma^2 (f_{\max})^2$ , где  $f_{\max} = 1,67$ . Параметр  $\sigma$  варьировался в вычислительном эксперименте.

Алгоритмическая, реализация формулы обращения (1), проводилась при помощи одной из часто применяемых аппроксимаций *ramp*-фильтра, предложенной Шеппом и Логаном, [7]. Точность реконструкции оценивалась посредством нормированной среднеквадратичной ошиб-

ки  $\Delta$ , которая рассчитывалась как

$$\Delta = \sqrt{\frac{\sum_j (g_{0,j} - g_j)^2}{\sum_j (g_{0,j})^2}}, \quad (22)$$

где  $g_{0,j}$  и  $g_j$  – значения математического фантома и его томограммы в  $j$ -м узле. Суммирование ведётся по всем узлам сетки.

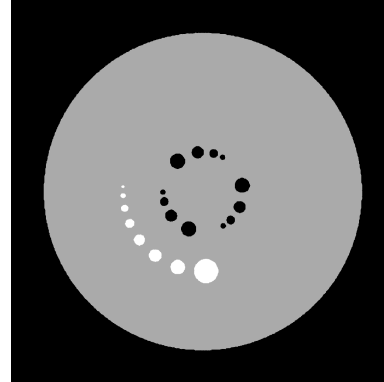


Рис. 1. Математический фантом

Далее приводятся результаты, полученные при  $n = 4$ . В этом случае для сглаживания используются пять коэффициентов  $w_0$ ,  $w_1$ ,  $w_2$ ,  $w_3$ ,  $w_4$  и значения проекции в девяти точках. На рис. 2,а,б представлены зависимости ошибки  $\Delta$  от отношения  $\beta = D_\xi^F / D_\xi$ . Для рис. 2,а величина параметра  $\sigma$  составляет 0,05, а для рис. 2,б – 0,15. Кривым 1, 2 и 3 отвечают  $\lambda = 7,894$ ,  $\lambda = 8,881$  и  $\lambda = 10,149$ . Такой выбор объясняется следующим образом. Амплитуда функции (5) падает вдвое, когда аргумент меняется от нуля до  $\ln 2$ . Число  $\gamma = \sqrt{\ln 2} / (\lambda h)$  принимается за её полуширину (выраженную в шагах сетки). Указанным  $\lambda$  соответствуют  $\gamma$ , равные 3,5, 4 и 4,5. В свою очередь,  $\gamma = 4$  согласуется с рассматриваемым количеством коэффициентов в том смысле, что они вычисляются по пяти значениям  $G(ih)$  в точках, равномерно расположенных на отрезке длиной  $4h$ .

Для построения графиков, изображённых на рис. 2, отношение  $\beta$  бралось из интервала  $]1/9; 1[$ . Если  $\beta = 1/9$ , то имеется решение  $w_j = 1/9$ ,  $\forall j$ , для которого преобразование (4) сводится к фильтру скользящего среднего. При  $\beta = 1$  дисперсия шума после сглаживания не изменяется. В частности, такое получается когда  $w_0 = 1$ , а все остальные коэффициенты равны нулю. В данном контексте оба эти случая интереса не представляют. На рис. 2 минимальное значение  $\beta$  равно 0,125, а максимальное – 0,99.

Как видно из рис. 2, все зависимости  $\Delta(\beta)$  имеют неглубокие минимумы. Для  $\sigma = 0,05$ , они достигаются при  $\beta$  порядка 0,9 ÷ 0,95, а при увеличении  $\sigma$  до 0,15 смещаются к  $\beta \approx 0,3$ . Это означает, что для шума низкой интенсивности нецелесообразно сильно (в несколько раз) понижать его дисперсию, поскольку при этом полезный сигнал (точные проекции) претерпит большие искажения. Желаемый эффект – повышение качества томограммы – может быть достигнут посредством увеличения ширины корреляционной функции помехи в процессе сглажива-

ния. Если же шум значительный, то добиться уменьшения ошибки реконструкции исключительно этим способом нельзя, нужно также подавлять его амплитуду.

Зависимости  $\Delta(\lambda)$  изображены рис. 3,а,б. Кривые 1, получены при  $\beta = 0,3$ , кривые 2 – при  $\beta = 0,6$ , кривые 3 – при  $\beta = 0,9$ ; рис. 3,а соответствует  $\sigma = 0,05$ , а рис. 3,б –  $\sigma = 0,15$ . Рис. 3 показывает, что для любых  $\beta$  и  $\gamma$  минимум ошибки реконструкции достигается при  $\lambda \approx 9$ . Это согласуется с данными, приведёнными на рис. 2.

Рис. 3 также подтверждает наблюдение, сделанное в связи с рис. 2: если уровень шума невысокий, то, применяя рассматриваемый подход, не следует стремиться существенно уменьшать его дисперсию. В частности, на рис. 3,а кривая с  $\beta = 0,9$  проходит ниже остальных, а на рис. 3,б, для которого  $\sigma$  в три раза больше, она располагается выше, чем другие кривые.

В целом по вычислительному эксперименту был сделан вывод о том, что лучший результат получается, когда полуширина функции  $G(ih)$  близка к  $\gamma_{opt} = nh$ . Этому можно дать следующее объяснение. Поскольку исходный шум некоррелированный, и сглаживание проводится по  $2n+1$  точкам (при этом фильтрация является не рекурсивной), эффективный радиус корреляции должен быть порядка

$(2n+1)h/2 = nh$ , т.е. именно  $\gamma_{opt}$ . Далее, если веса  $w_j$  вычисляются в предположении, что  $\gamma > \gamma_{opt}$  задача оказывается не согласованной: фильтр не может обеспечить заявленной величины  $\gamma$ . Если же, наоборот, берётся  $\gamma < \gamma_{opt}$ , то корреляционная функция помех после сглаживания получается более узкой, что, как было показано в работе [5], ведёт к росту ошибки реконструкции.

Посредством численного моделирования предлагаемый метод сравнивался с другими известными процедурами подавления шума. На рис. 4 приведены зависимости  $\Delta$  от величины параметра  $\sigma$ . Для кривой 1 проекции обрабатывались согласно (4), для кривых 2 и 3 – соответственно средним и медианой в скользящем окне, содержащем девять точек, для кривой 4 – регуляризующими сплайнами [11]. Как следует из рис. 4, разработанный алгоритм обеспечивает более высокую точность, чем усреднение и медиана (в частности, при  $\sigma = 0,05$  ошибка реконструкции уменьшается приблизительно в 1,5 и 2,6 раза), но несколько уступает сплайнам. Однако последние значительно сильнее остальных фильтров сглаживают также и точные данные, что приводит к размытию мелких структур на изображении, как это хорошо заметно на рис. 7,д, см. ниже.

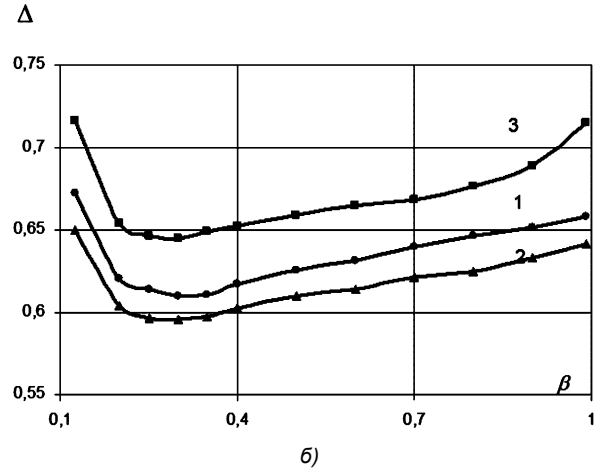
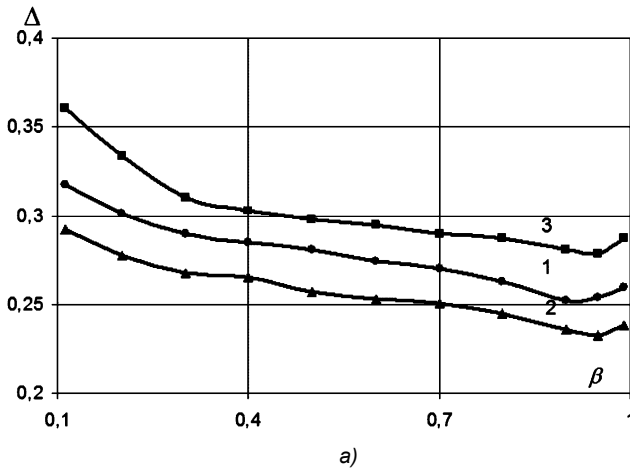


Рис. 2. Зависимости ошибки реконструкции от отношения дисперсии шума после сглаживания к дисперсии исходного шума:

а)  $\sigma = 0,05$ ; б)  $\sigma = 0,15$ . Кривые 1 –  $\gamma = 3,5$ ; 2 –  $\gamma = 4$ ; 3 –  $\gamma = 4,5$  шагов сетки

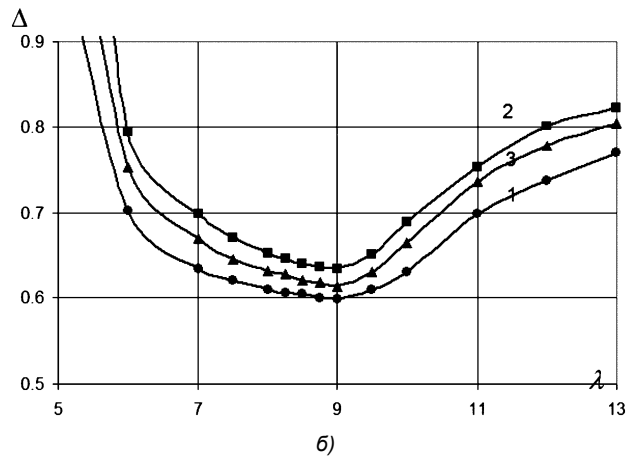
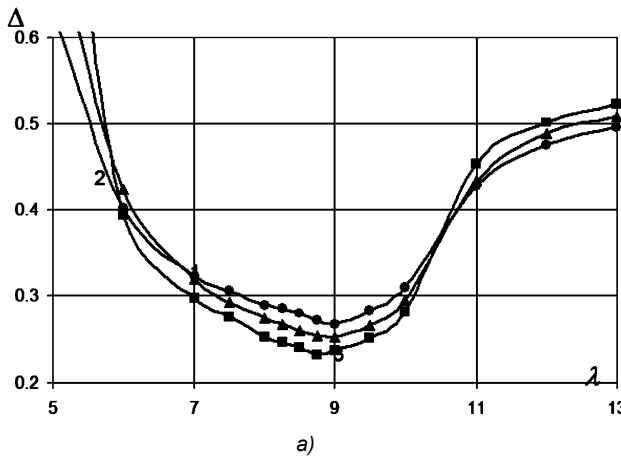


Рис. 3. Зависимости ошибки реконструкции от параметра  $\lambda$ , определяющего ширину корреляционной функции после сглаживания:

а)  $\sigma = 0,05$ ; б)  $\sigma = 0,15$ . Кривые 1 –  $\beta = 0,3$ ; 2 –  $\beta = 0,6$ ; 3 –  $\beta = 0,9$



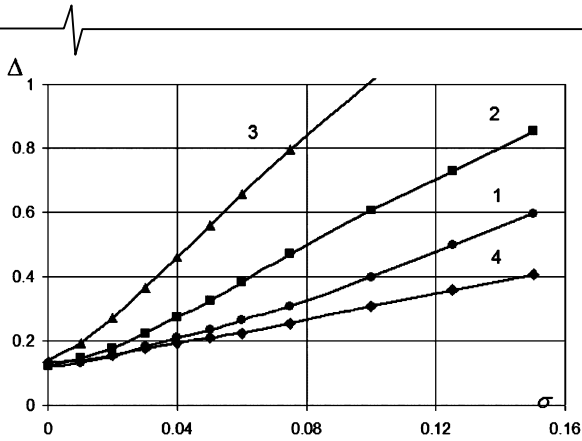


Рис. 4. Зависимости ошибки реконструкции от уровня шума. Кривая 1 – предлагаемый метод, 2 – усреднение, 3 – медиана, 4 – сплайны

Некоторые из полученных томограмм представлены на рис. 5 – рис. 7. Для рис. 5,а и рис. 6 параметр  $\sigma$  равен 0,01, а для рис. 5,б и рис. 7 он составляет 0,05. Известно, что без применения стабилизирующих процедур, реконструкция алгоритмом Шеппа-Логана является неустойчивой [2, 3]. Это иллюстрирует рис. 5, по которому видно, что в отсутствие предварительной обработки данных уже при относительно невысоком уровне шума качество восстановления становится неудовлетворительным.

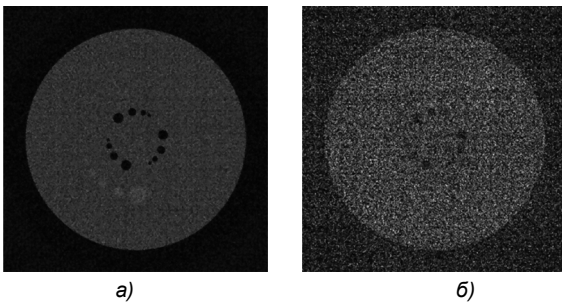


Рис. 5. Томограммы, полученные по зашумлённым данным без сглаживания: а)  $\sigma = 0,01$ ,  $\Delta = 0,553$ ; б)  $\sigma = 0,05$ ,  $\Delta = 2,218$

Для рис. 6, рис. 7 варианты а), б), с) и д) соответствуют сглаживанию проекций фильтром (4), усреднению

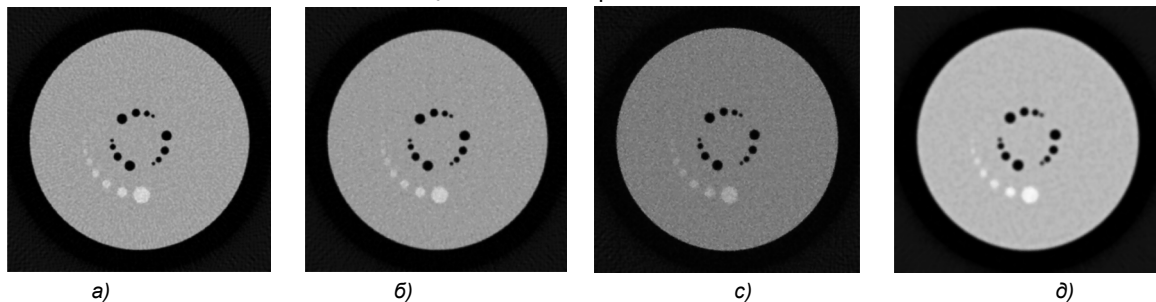


Рис. 6. Томограммы, шум в данных соответствует  $\sigma = 0,01$ : а) предлагаемый метод,  $\Delta = 0,132$ ; б) усреднение,  $\Delta = 0,145$ ; в) медианный фильтр  $\Delta = 0,193$ ; г) сплайны  $\Delta = 0,138$

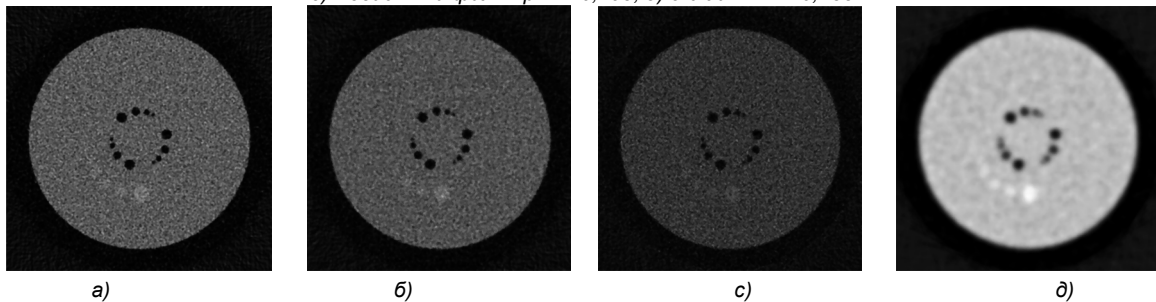


Рис. 7. Томограммы, шум в данных соответствует  $\sigma = 0,05$ : а) предлагаемый метод,  $\Delta = 0,232$ ; б) усреднение,  $\Delta = 0,326$ ; в) медианный фильтр  $\Delta = 0,558$ ; г) сплайны  $\Delta = 0,209$

ем, медианой и сплайнами. Коэффициенты  $w_j$ , определённые описанным в разделе 2 способом, имеют следующие значения  $w_0 = 0,168$ ,  $w_1 = 0,152$ ,  $w_2 = 0,119$ ,  $w_3 = 0,073$ ,  $w_4 = 0,072$  для  $\sigma = 0,01$  и  $w_0 = 0,172$ ,  $w_1 = 0,156$ ,  $w_2 = 0,108$ ,  $w_3 = 0,076$ ,  $w_4 = 0,074$  для  $\sigma = 0,05$ . Небольшое различие объясняется тем, что для  $\sigma = 0,01$  отношение  $\beta$  бралось равным 0,95, а для  $\sigma = 0,05$  оно равнялось 0,9.

**Заключение**

Работа посвящена актуальной проблеме повышения устойчивости решений в задачах томографии. Для этой цели применяется предварительная обработка одномерных проекций посредством симметричного цифрового фильтра. Вычисление его коэффициентов  $w_j$  основывается на предыдущих исследованиях автора [5], где было получено, что среди рассмотренных искажений данных, имеющих одинаковую дисперсию, наиболее высокая точность восстановления достигается для гауссового шума. Исходя из этого, был разработан алгоритм, включающий процедуру перебора, для нахождения набора  $w_j$ , обеспечивающего минимум невязки между заданной кривой Гаусса и корреляционной функцией сглаженной помехи.

В процессе численного моделирования показано преимущество предлагаемого метода перед скользящими средним и медианой. В частности, когда стандартное отклонение шума находится в интервале от 5 до 15% от максимальной величины проекции, он приводит к ошибке в 1,4 ÷ 1,5 раза меньше, чем та, которую даёт усреднение (при одинаковом размере окна обоих фильтров, составляющем девять точек). Кроме того из серии экспериментов, в которых длина фильтра  $2n+1$  варьировалась от пяти до пятнадцати точек, установлено, что лучший результат реконструкции получается, когда полуширина кривой Гаусса, используемой при расчете  $w_j$ , равна  $n$  шагов сетки.



## Литература

1. Хермен Г.Т. Восстановление изображений по проекциям. Основы реконструктивной томографии. Пер. с англ. М.: Мир, 1983. – 342 с.
2. Наттерер Ф. Математические аспекты компьютерной томографии: Пер. с англ. М.: Мир, 1990. – 279 с.
3. Тихонов А. Н., Арсенин В. Я., Тимонов А. А. Математические задачи компьютерной томографии. М: Наука, 1987. – 158 с.
4. Рытов С. М. Введение в статистическую радиофизику. Ч. 1. Случайные процессы. М.: Наука, 1976.– 484 с.
5. Лихачев А. В. Шibaева Ю. А. Зависимость точности томографической реконструкции от корреляционной функции шума в проекционных данных // Цифровая обработка сигналов. 2015. № 2. – С. 28-34.
6. Ramachandran G.N., Lakshminarayanan A.V. Three-dimensional reconstruction from radiographs and electron micrographs: application of convolutions instead of Fourier transforms. // Proc. Nat. Acad. Sci. U.S. 1971. V. 68. – P. 2236-2240.
7. Shepp L. A., Logan B. F. The Fourier reconstruction of a head section. // IEEE Trans. Nucl. Sci. 1974. V. 21, No. 3. – P. 21-43.
8. Ерохин В.А., Шнейдеров В.С. Трехмерная реконструкция (машинная томография). Моделирование на ЭВМ. // Препринт No.23, ЛНИВЦ; Ленинград, 1981.
9. Лихачёв А. В. Исследование  $1/z^2$  фильтрации в алгоритмах томографии. // Автометрия. 2007. Т. 43, № 3. – С. 57-64.
10. Лихачёв А.В. Алгоритм двойной фильтрации для двумерной томографии. // Математическое моделирование. 2009. Т. 21, № 8. – С. 21-29.
11. Василенко В. А. Сплайн-функции: теория, алгоритмы, программы. Новосибирск: Наука, 1983. – 214 с.

## АЛГОРИТМ АДАПТИВНОГО КОНТРАСТИРОВАНИЯ ИЗОБРАЖЕНИЯ

*Завалишин С.С., аспирант кафедры АИТУ Рязанского государственного радиотехнического университета, e-mail: ss.zavalishin@gmail.com.*

## ADAPTIVE IMAGE CONTRASTING ALGORITHM

*Zavalishin S.S.*

*In this paper we propose an adaptive contrasting algorithm that exploits image structure. In contrast to existing methods, proposed one applies preliminary segmentation in order to determine contrast curves for each image area independently, which make it possible to take into account underexposed and overexposed areas. Smooth contrast transitions between nearby regions are provided using a special graph, which store algorithm parameters and adjust them.*

**Key words:** image processing, document processing, contrast enhancement.

**Ключевые слова:** обработка изображений, контрастирование, обработка документов.

**Введение**

В данной работе предлагается алгоритм адаптивного контрастирования изображения с использованием знания о структуре последнего. В отличие от существующих алгоритмов, предлагаемый подход позволяет подобрать оптимальные параметры кривой контраста для каждой области изображения индивидуально. Информация о структуре изображения получается путём проведения предварительной сегментации и последующего анализа сегментов для подбора оптимальных параметров кривой. Полученный граф параметров сглаживается с использованием оценки схожести соседних сегментов таким образом, чтобы переходы между похожими сегментами были незаметны. Подобный подход соответствует современному принципу фотокоррекции, когда схожие по содержанию области фотографии обрабатываются различными способами [1], что позволяет достичь максимальной контрастности изображения на всей площади кадра, не ухудшая визуального восприятия отдельных его частей из-за наличия чрезмерно контрастных областей.

**Методы улучшения контраста**

Рассмотрим алгоритмы улучшения изображений подробнее. Большая их часть корректирует контраст глобально по всему изображению без учёта локальных особенностей. Выделяется три группы различных алгоритмов: первая группа основана на применении физической модели формирования полутонового изображения следующего вида:

$$I = R \times L \quad (1)$$

где  $L$  – освещённость объекта, изменяющаяся медленно по всему изображению,  $R$  – альbedo (отражающая способность) деталей объекта, изменяющаяся с высокой частотой [2]. Заданная модель не принимает во внимание наличие возможных источников света, зеркальных отражений (используются только диффузные

*Предлагается алгоритм адаптивного контрастирования изображения с использованием знания о структуре последнего. В отличие от существующих алгоритмов, предлагаемый подход опирается на предварительную сегментацию изображения на зоны с последующим подбором параметров кривой контраста для каждой зоны в отдельности, что позволяет учесть наличие недоэкспонированных и переэкспонированных участков. Для обеспечения равномерности переходов между зонами предлагается согласовывать параметры соседних зон с помощью построения графа параметров алгоритма специального вида.*

Ламбертовы отражения) и формирования теней от объектов на изображении. Несмотря на ряд недостатков, модель успешно применяется для обработки широкого класса изображений. Отражающая способность измеряется различными способами, но независимо от выбранного метода, она всегда выражается через отношения  $I$  к  $I$ . В некоторых работах, например [3, 4], описывается метод коррекции тёмных тонов на основе различных модификаций алгоритма MultiScale Retinex ( $MSR$ ; разномасштабный ретинекс). Он определяется следующим выражением:

$$I_e = \sum_{n=1}^N \omega_n \times (\log(I) - \log(LPF_n(I))) \quad (2)$$

Здесь  $LFP$  – фильтр Гаусса с переменным значением  $\sigma$ , а  $N$  – число различных масштабов.  $MSR$  позволяет значительно выделить плохо видимые детали изображения, но порождает ореолы и неестественные цвета. Ещё одна проблема связана с тем, что методы на основе классического  $MSR$ , описанного выше, обладают высокой вычислительной сложностью.

Работа [5] описывает метод улучшения тёмных областей, основанный на физической модели следующего вида:

$$I_e = S \times (R + a \times L) \quad (3)$$

где  $a$  – коэффициент ослабления светового потока,  $S$  – линейный масштабирующий оператор, который приводит результат вычислений к масштабу, соответствующему масштабу сигнала,  $L$  – результат применения фильтрации, сохраняющей границы, а  $R$  – функция от

отношения яркости исходного изображения к  $L$ . Фильтрация, сохраняющая границы, применяется к каждой линии изображения, причём, конкретный фильтр, применяемый к изображению, зависит от уровня шума. Шум подавляется в яркостном канале, а цветовые каналы остаются без изменений. Основным недостатком данного метода – снижение насыщенности цветов в средних и тёмных тонах. К достоинствам относится то, что метод не создаёт эффекта гало.

Работа [6] описывает метод коррекции фотографий, содержащих засветку от источника света, направленного в камеру. В данном методе отражающая способность рассчитывается следующим образом:

$$D_s(r, c) = \frac{1.3 \times I(r, c)}{I_f(r, c) + 3 \times \sigma_R} \quad (4)$$

где  $r$  и  $c$  – координаты пикселя,  $I$  – яркость исходного изображения, а  $I_f$  – оценка освещённости, произведённая путём применения билатерального фильтра с разбросом расстояний  $\sigma_R$ . Локальный контраст и яркость изображения повышаются одновременно за счёт смещения изображений  $I$  и  $DS$  по маске:

$$I_e(r, c) = I(r, c) + k_s \times (1 - I_f(r, c)) \times (D_s(r, c) - I(r, c)) \quad (5)$$

В данной формуле  $k_s$  – коэффициент усиления яркости тёмных областей. После применения указанных формул, значения цветовых каналов RGB выравниваются относительно друг друга таким образом, чтобы сохранить оттенок и насыщенность исходного изображения. Описанный алгоритм порождает значительно меньшее число артефактов, чем Retinex, но не может использоваться для улучшения небольших деталей и исправления контраста в светлых областях.

Вторая группа алгоритмов – это алгоритмы локальной коррекции, основанные на применении нелинейных кривых преобразования гистограммы яркостей, где форма и тип применяемых кривых зависят от локального распределения полутонов. Достаточно простой метод коррекции теней и светов описывается в работе [7]: маска областей применения алгоритма находится путём сильного размытия канала яркости, а затем каналы RGB модифицируются с помощью набора кривых воспроизведения тонов (Tone Reproduction Curves), форма которых определяется найденной маской. К примеру, для изображения, яркость которого нормализована в диапазоне [0; 1], кривые принимают следующие значения:

$$R = (R)^{2^{(1-2Mask)}}; \quad G = (G)^{2^{(1-2Mask)}}; \quad B = (B)^{2^{(1-2Mask)}} \quad (6)$$

Патентная заявка [8] описывает метод улучшения изображений, основанный на ортогональной трансформации раstra изображения (orthogonal retino-morphic image transform; ORMIT):

$$I_e = \sum_{i=0}^N a_i (I) \times LPF(P_i(F(I))) \times Q_i(F(I)) + b(I) \quad (7)$$

Здесь  $P_i(x)$  – ортогональный базис функций  $x$ , определённый на интервале (0; 1),  $Q_i$  – отрицательные производные  $P_i$  или их аппроксимация,  $LPF$  – фильтр высоких частот в пространственной области,  $F$  – весовая функция, аналогичная функции гамма-коррекции,  $N$  –

число полос, таких что  $N \ll n$ , где  $n$  – глубина цвета в канале, а  $a$  и  $b$  – константы, определяющиеся для каждой полосы отдельно. Выбор  $P_i$  определяет степень увеличения локального контраста.

Работа [9] описывает алгоритм улучшения тёмных изображений, учитывающих из содержимое. На первом шаге алгоритм конвертирует изображение в цветовое пространство HSV. Коррекция производится в три стадии: разделение канала  $V$  на три плоскости, содержащие тёмные, средние и светлый полутона, коррекция каждой плоскости с помощью кривой воспроизведения тонов и объединение скорректированных плоскостей в один канал, после чего изображение конвертируется обратно в цветовое пространство RGB. Данный алгоритм позволяет получать достаточно качественные изображения, однако малое число плоскостей ведёт к появлению артефактов.

Наконец, третья группа алгоритмов базируется на методе CLAHE (Contrast Limited Adaptive Histogram Equalization; адаптивная эквализация гистограмм с ограничением контраста) [10]. Основное назначение методов адаптивной эквализации гистограмм (Adaptive Histogram Equalization; AHE) – коррекция полутоновых медицинских изображений. Появление большого количества цветных артефактов не позволяет применять данную группу алгоритмов для фото- и видеоизображений.

### Предлагаемый алгоритм

Рассмотрим алгоритм восстановления локального контраста, разработанный специально для работы с документами. Известно, что кроме текста документ может содержать несколько изображений, имеющих разный контраст. В связи с этим восстанавливать контраст необходимо для каждой локальной области отдельно. В простейшем варианте данная операция производится путём выбора индивидуальной функции преобразования для каждого пикселя следующим образом [11]:

- определяем формы  $S$ -образных кривых, которые наилучшим образом восстанавливают контраст в каждой области;

- сглаживаем параметры кривых между соседними пикселями и областями для обеспечения плавных переходов;

- восстанавливаем контраст пикселей с помощью полученных кривых.

В качестве  $S$ -кривой используется сплайн Эрмита, следующего вида:

$$f(t) = (1 - 3t^2 + 2t^3)P_{ox} + (3t^2 - 2t^3)P_{1x} + (t - 2t^2 + t^3)Q_{ox} + (t^3 + t^2)Q_{1x}, \quad (8)$$

В данной формуле параметр  $t \in [0; 1]$ , коэффициенты  $P_{ox}$  и  $P_{1x}$  определяют начальную и конечную точки кривой, а  $Q_{ox}$  и  $Q_{1x}$  представляют собой векторы, определяющие углы наклона кривой в точках  $P_{ox}$  и  $P_{1x}$ , соответственно.

Расстояние между точками  $P_{ox}$  и  $P_{1x}$  влияет на степень растяжения гистограммы. Иными словами, чем ближе друг к другу находятся  $P_{ox}$  и  $P_{1x}$ , тем выше будет линейный контраст, определяемый максимальным и

минимальным её значением, которые мы оставляем при контрастировании. Данные значения называются точкой белого и чёрного, соответственно.  $Q_{ox}$  и  $Q_{lx}$  влияют на угол наклона кривой в той зоне, где необходимо получить максимальный контраст (рис. 1).

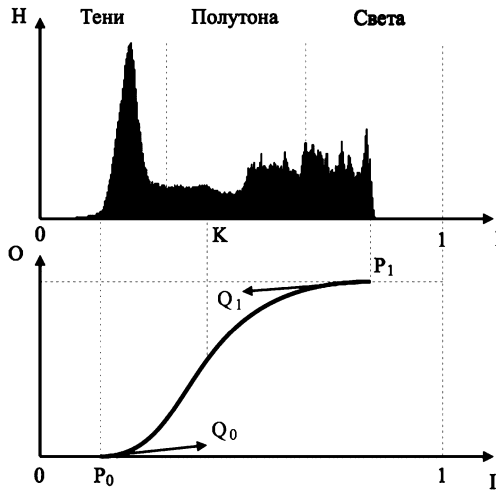


Рис. 1. Пример S-образной кривой, модифицирующей гистограмму яркости изображения

Задача контрастирования сводится к двум подзадачам: максимально расширить гистограмму, сохранив детали в тенях и светах и получить наибольший наклон S-образной кривой в зоне, содержащей объект интереса. Нахождение точек  $P_{ox}$  и  $P_{lx}$  производится следующим образом:

$$P_{ox} = \min \left( \begin{array}{l} D, \min \{i \mid H[i] \geq H_0\}, \\ \min \{i \mid \sum_{k=i}^{255} H[k] \geq C_0\} \end{array} \right),$$

$$P_{lx} = \max \left( \begin{array}{l} \max \{i \mid H[i] \geq H_1\}, \\ \max \{i \mid \sum_{k=i}^{255} H[k] \geq C_1\} \end{array} \right).$$

где  $H_0$  и  $H_1$ ,  $C_0$  и  $C_1$  и  $D$  – пороги, устанавливающие максимальный уровень гистограммы, площадь гистограммы и интенсивность гистограммы, соответственно.  $D$  применяется для недопущения излишнего затемнения изображения.

Значения компонент векторов  $Q_{ox}$  и  $Q_{lx}$  зависят от того, какая часть гистограммы соответствует объекту интереса. В данной работе они определяются путём расчёта порога разбиения гистограммы изображения по яркости методом Отцу [12]:

Если  $K_i > 0,5$ ,

$$\text{то } \begin{bmatrix} Q_{ox} = 1 + A \cdot (K_i - 0,5); Q_{oy} = 0; \\ Q_{lx} = 0; Q_{ly} = 0; \end{bmatrix}, \quad (9)$$

$$\text{иначе } \begin{bmatrix} Q_{ox} = 1; Q_{oy} = 0; \\ Q_{lx} = 1 + A \cdot (0,5 - K_i); Q_{ly} = 0. \end{bmatrix}.$$

Здесь  $A$  – параметр алгоритма в диапазоне от 1 до 6, устанавливающий степень контрастирования, а  $K$  – порог, найденный методом Отцу в интервале  $[0; 1]$ .

В тех случаях, когда ширина гистограммы слишком мала, либо имеет единственный пик, вместо локального значения порога выбирается порог, соответствующий порогам окружающих областей, что будет показано да-

лее. Вычисление ширины гистограммы производится путём нахождения точек чёрного и белого:

$$P_b = \min \{i \mid \sum_{k=0}^i H[k] \geq C_0\}; \quad (10)$$

$$P_w = \max \{i \mid \sum_{k=i}^{255} H[k] \geq C_1\}.$$

Выбор локальных областей, для которых оцениваются параметры, производится двумя способами в зависимости от модификации алгоритма. Первый способ предполагает использование перекрывающихся блоков размером  $100 \times 100$  пикселей с перекрытием 50%. Для каждого блока оценивается ширина гистограммы, а затем, если соблюдается условие (11), где  $T_r$  – минимальная ширина гистограммы, вычисляется локальное значение  $K$ .

$$(P_w - P_b) > T_r, \quad (11)$$

В случае, если условие (11) не соблюдается, используется глобальный  $K$ , рассчитанный по гистограмме всего изображения.

Результатом предыдущих шагов является прямоугольная сетка коэффициентов  $K$ . Для предотвращения резких переходов между частями изображения сетка сглаживается с помощью фильтра низких частот размером  $5 \times 5$ . Затем сетка масштабируется путём применения билинейной интерполяции до размеров изображения, что позволяет сопоставить каждому пикселю изображения собственное значение  $K$  и таким образом изменять в каждой точке индивидуальную кривую преобразования.

Построение S-кривых с помощью сплайна Эрмита для каждой точки изображения является трудоёмкой задачей, поэтому для уменьшения объёма вычислений применяется таблица, содержащая набор заранее посчитанных S-кривых для 256 коэффициентов  $K$  в диапазоне  $[0; 1]$ . Выходные значения яркости вычисляются следующим образом:

$$O = LUT \left( \left[ \begin{array}{c} I - P_{ox} \\ P_{lx} - P_{ox} \end{array} \right], K \right), \quad (12)$$

где  $I$  – входное значение яркости,  $[\cdot]$  – оператор округления, а  $LUT(\cdot, K)$  – функция, возвращающая выходное значение яркости из таблицы.

Применение перекрывающихся блоков позволяет производить локальное контрастирование, которое лишь приблизительно описывает структуру исходного изображения. Вместе с тем, основным достоинством подобного подхода является высокая производительность.

Более точная адаптация к изображению возможна за счёт предварительной сегментации. Алгоритм выглядит следующим образом:

Изображение сегментируется на неперекрывающиеся области.

Для каждой области находятся значения коэффициентов  $P_{ox}$ ,  $P_{lx}$  и  $K$ .

Строится граф  $G = (V := \{P_{ox}, P_{lx}, K\}, E := S_A)$ , вершинами которого являются значения найденных коэффициентов, а рёбрами – коэффициенты схожести соседних областей в диапазоне  $[0; 1]$ .

Значения коэффициентов сглаживаются между похожими областями для предотвращения резких переходов.

Строится сетка, сопоставляющая каждому пикселю изображений свой набор коэффициентов. Сетка сглаживается с помощью фильтра Гаусса размером  $3\sigma$ , где  $\sigma = 3, 5$ , что позволяет сделать переходы на резких границах более естественными.

Для каждого пикселя находится выходное значение яркости с помощью формулы (12).

Опишем перечисленные шаги подробнее. Первоначальная сегментация производится с помощью алгоритма Фельзенсвальба [13]. Параметры алгоритма подбираются таким образом, чтобы число найденных областей не было слишком большим, но каждая область представляла собой значимый участок изображения. В ходе экспериментов оптимальными были признаны следующие параметры:  $\sigma = 1, 0$ ,  $k = 75$  и минимальный размер области 100 пикселей.

Метод оценки схожести областей основан на идее алгоритма выборочного поиска [14]. Схожесть между двумя соседними областями  $r_i$  и  $r_j$  определяется метрикой:

$$S(r_i, r_j) = \frac{1}{4} [S_c(r_i, r_j) + \alpha_s S_s(r_i, r_j) + \alpha_f S_f(r_i, r_j)] \quad (13)$$

где  $S_c, S_t, S_s$  и  $S_f$  – меры схожести цвета, текстуры, размера и вложенности одной области в другую, соответственно.  $S_c$  вычисляется как пересечение гистограмм яркостей двух областей:

$$S_c(r_i, r_j) = \sum_{k=1}^n \min(c_i^k, c_j^k), \quad (14)$$

где  $C_k$  – значение  $k$ -й ячейки гистограммы.  $S_t$  показывает, насколько текстура одной области похожа на текстуру другой. Для её вычисления изображение размывается с помощью фильтра Гаусса и вычисляются его производные в четырёх направлениях:  $G_x, G_{-x}, G_y, G_{-y}$ . Для каждого направления строится гистограмма значений производных, после чего полученные гистограммы объединяются в единую текстурную гистограмму:  $G = \{G_x; G_{-x}; G_y; G_{-y}\}$ .  $S_t$  находится как пересечение текстурных гистограмм двух областей аналогично (14).

Мера схожести размера  $S_s$  вычисляется как отношение суммы площадей областей  $A(r_i), A(r_j)$  к площади всего изображения  $A(I)$ :

$$S_s(r_i, r_j) = 1 - \frac{A(r_i) + A(r_j)}{A(I)}, \quad (15)$$

Наконец, мера вложенности областей рассчитывается с помощью общего ограничивающего прямоугольника двух областей  $BB(r_i \cup r_j)$  и их площадей:

$$S_f(r_i, r_j) = 1 - \frac{BB(r_i \cup r_j) - A(r_i) - A(r_j)}{A(I)}, \quad (16)$$

Найденные значения  $S(r_i, r_j)$  для каждой пары областей используются для сглаживания переходов между похожими соседними областями:

$$P_{ox}(r_i) = \frac{1}{N} \sum_{j=1}^N S(r_i, r_j) \cdot P_{ox}(r_j),$$

$$P_{ix}(r_i) = \frac{1}{N} \sum_{j=1}^N S(r_i, r_j) \cdot P_{ix}(r_j), \quad (17)$$

$$K(r_i) = \frac{1}{N} \sum_{j=1}^N S(r_i, r_j) \cdot K(r_j),$$

где  $N$  – количество соседних с  $r_i$  областей. Отметим, что между областями, не являющимися соседними, метрика схожести  $S(r_i, r_j)$  равна 0, а метрика схожести области с самой собой  $S(r_i, r_j)$  равна 1, если соблюдается условие (11) и 0 в противном случае. Последнее предотвращает появление артефактов в однотонных областях, не обладающих ярко выраженной структурой.

Сравним сетки коэффициентов, построенные с помощью перекрывающихся блоков и с помощью сегментированных областей (рис. 2). Применение перекрывающихся блоков приводит к построению сетки, обеспечивающей максимально гладкие переходы, в то время, как сетка, построенная на основе сегментированных областей, напротив, состоит исключительно из чётких границ. Вместе с тем, последняя обеспечивает значительно более высокую степень адаптации к структуре изображения.

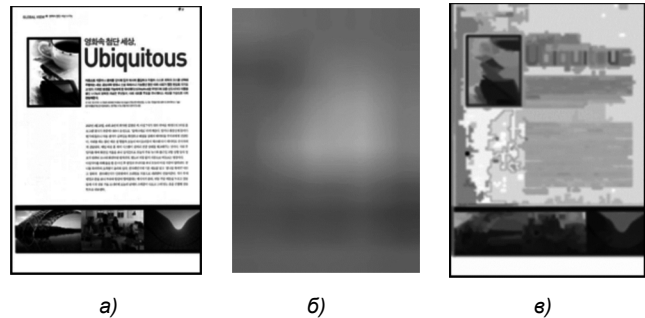


Рис. 2. Пример сетки коэффициентов:  
а) исходное изображение, б) сетка, построенная с помощью перекрывающихся блоков, в) сетка, построенная с помощью сегментированных областей

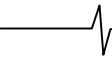
### Оценка эффективности контрастирования

Оценим эффективность повышения контраста с помощью предложенных алгоритмов. Существует ряд метрик для оценки контрастности изображения. В данной диссертационной работе применяется контраст Михельсона [15]:

$$MC = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}, \quad (18)$$

где  $I_{\max}$  и  $I_{\min}$  – наибольшее и наименьшее значение яркости в области, соответственно. Для вычисления МС изображение делится на блоки размером 150x150 пикселей, после чего для каждого блока вычисляется значение МС. Полученные значения усредняются и среди них находится минимум. Последний показывает наилучший контраст, достигнутый при применении алгоритма.

Большинство исследователей в ходе разработки алгоритмов улучшения контраста опираются исключительно на показатели контрастности в качестве метрики качества. Вместе с тем, многие алгоритмы вносят заметные искажения в исходное изображение, что может быть недопустимо в ряде задач. В данной диссертационной



работе предлагается оценивать алгоритмы не только с точки зрения степени улучшения контраста, но и с точки зрения вносимых искажений.

Существует ряд метрик, использующихся для оценки степени искажений, однако большая часть из них приспособлена исключительно для работы с полутоновыми изображениями. Оценку цветных изображений предполагается производить путём поканального применения метрики, что не является верным подходом, так как каналы RGB в большинстве случаев несут одинаковую информацию (особенно, если обрабатывались одним и тем же способом), а однородная обработка цветовых пространств, имеющих отдельные каналы для яркости и цветности (например  $L^*a^*b$ ) не учитывает особенностей человеческого зрения, которое более чувствительно к изменению яркости в сравнении с изменением цвета.

Для устранения этого недостатка была разработана метрика CSSM (Color Structure Similarity Metric; Метрика подобия цветовой структуры), основой которой послужила метрика SSIM (Structure Similarity Index Metric; Метрика индекса структурного подобия) [16]. Главным достоинством SSIM является то, что она учитывает три компонента, отвечающие за схожесть изображений: освещённость  $I(x, y)$ , контрастность  $c(x, y)$  и структуру  $s(x, y)$ :

$$SSIM(x, y) = [I(x, y)]^\alpha \cdot [c(x, y)]^\beta \cdot [s(x, y)]^\gamma, \quad (19)$$

где

$$I(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \quad (20)$$

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \quad (21)$$

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x + \sigma_y + C_3}. \quad (22)$$

Здесь  $\mu_x, \mu_y, \sigma_x, \sigma_y$  и  $\sigma_{xy}$  – локальные средние значения, среднеквадратичные отклонения и кросс-ковариация для изображений  $x$  и  $y$ , а  $C_1, C_2, C_3$  и  $\alpha, \beta, \gamma$  – константы, предотвращающие деление на ноль и степени влияния компонент на результат.

При изучении оценок схожести изображений, получаемых с помощью SSIM, можно заметить, что компоненты освещённости и контрастности влияют на результат в значительно меньшей степени, чем структура, что связано с поведением функции вида  $z(x, y) = 2xy / (x^2 + y^2)$  при значениях  $x$  и  $y$  в интервале  $[0; 1]$  (рис. 3а). Более 50 % всех значений  $z$  лежат выше уровня 0,8, соответственно, влияние  $I(x, y)$  и  $c(x, y)$  проявляется только при очень большой разнице между изображениями по этим двум компонентам. Нормализация значений в диапазоне  $[-1; 1]$  не решает данной проблемы, так как в этом случае при значениях близких к нулю, соответствующих полутонам, даже небольшая разница между изображениями приводит к сильному падению схожести.

Разработанная метрика CSSM учитывает эту особенность. Она состоит из двух компонент: цветоконтрастной  $\Delta I$  и структурной  $\Delta S$ . Отметим, что в отличие от SSIM, комбинирование компонент происходит таким образом, что обе компоненты учитываются в равной пропорции (рис. 3б).

$$CSSM = \frac{2 \cdot \Delta I \cdot \Delta S}{\Delta I + \Delta S}, \quad (23)$$

где  $\Delta S$  рассчитывается как (24), а  $\Delta I$  как (25):

$$\Delta S = \frac{\sigma_{L(x)L(y)} + C}{\sigma_{L(x)} + \sigma_{L(y)} + C}, \quad (24)$$

$$\Delta I = 1 - [ \|LST^*(\bar{x}) - LST^*(\bar{y})\| ]^\alpha, \quad (25)$$

Здесь  $LST^*(\cdot)$  – преобразование изображения из цветового пространства RGB в цветовое пространство  $LST^*$ ,  $L(\cdot)$  – вычисление канала  $L$  цветового пространства  $LST$ , а  $\bar{x}, \bar{y}$  – изображения, обработанные фильтром нижних частот.

Преобразование в пространство  $LST$  применяется в связи с вычислительной простотой в сравнении с преобразованием в  $L^*a^*b$  при сохранении аналогичных свойств. Канал  $L$  хранит значения яркости,  $S$  – насыщенности, а  $T$  – цветового оттенка:

$$L = \frac{1}{3}(R + G + B), \quad S = \frac{R}{R + G}, \quad T = \frac{R}{R + B}. \quad (26)$$

Пространство  $LST^*$  полностью аналогично  $LST$  с той разницей, что в первом каналы насыщенности и цветового оттенка имеют меньшее влияние:  $S^* = S / 2, T^* = T / 2$ , что соответствует модели цветового восприятия человека.

Для тестирования было выбрано 30 случайных изображений документов, содержащих текст и иллюстрации. Каждое из изображений обрабатывалось четырьмя алгоритмами: алгоритмами блочного и структурного адаптивного увеличения контраста, предложенными в данной диссертации, алгоритмом CLAHE [10] и MSR [17].

Табл. 1 содержит результаты сравнения вышеуказанных алгоритмов. Видно, что предложенные алгоритмы (БК – блочное контрастирование; СК – структурное контрастирование) превосходят существующие как с точки зрения среднего контраста, так и с точки зрения вносимых искажений. Отметим, что как алгоритм CLAHE, так и алгоритм MSR, вносят заметные искажения в исходное изображение, выражающиеся в усилении артефактов, имеющихся в исходном изображении, в первом случае и появлении ореолов во втором.

Таблица 1. Сравнение алгоритмов

Метод	Средний МС	Мин. МС	SSIM	CSSM	Артефакты
CLAHE	0,78	0,14	<b>0,92</b>	0,87	Пятна
MSR	0,82	<b>0,17</b>	0,89	0,86	Ореол
<b>БК</b>	<b>0,83</b>	<b>0,17</b>	0,87	<b>0,88</b>	<b>Нет</b>
<b>СК</b>	<b>0,83</b>	0,16	0,87	<b>0,88</b>	<b>Нет</b>
Исх. Изобр.	0,82	<b>0,17</b>	-	-	-



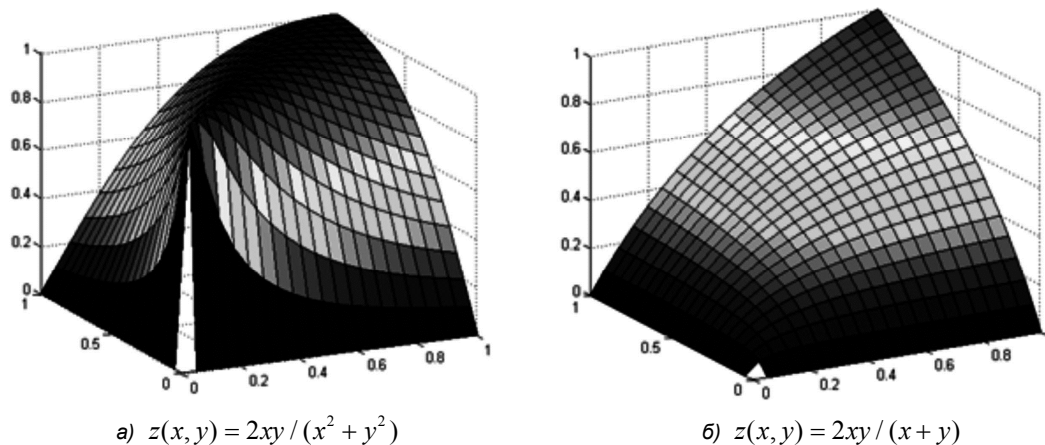


Рис. 3. Графики функций

### Заключение

В статье предложен алгоритм повышения контраста изображения на основе его структуры. Предлагаемый подход позволяет адаптировать набор контрастирующих кривых к конкретному изображению, что даёт возможность более корректно обрабатывать изображения, на которых присутствует фон и один или несколько объектов. Особо стоит отметить случай, когда необходимо обрабатывать изображения документов, содержащих некоторый фон и набор фотографий напечатанных поверх этого фона. В отличие от существующих подходов, предложенный алгоритм позволяет корректно обработать каждую фотографию в отдельности, что невозможно сделать, применяя известные методы.

### Литература

1. Маргулис Д. Photoshop для профессионалов: Классическое руководство по цветокоррекции. 2007.
2. Horn. B. Robot vision // MIT Press. 1986.
3. Sobol. R. Improving the Retinex algorithm for rendering wide dynamic range photographs // Journal of Electronic Imaging, Vol. 13, № 1.
4. Tao L., Asari. V. Modified Luminance based MSR for Fast and Efficient Image enhancement // Proc. of IEEE Applied Imagery Pattern Recognition Workshop. 2003.
5. Brajovic V. Brightness Perception, Dynamic Range and Noise: a Unified Model for Adaptive Image Sensors // Proc. of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2004.
6. Safonov I.V. Automatic correction of exposure problems in photo printer // Proc. of IEEE ISCE'06. 2006.
7. Moroney N. Local color correction using non-linear masking // в Proc. of 8th Color Imaging Conference. 2000.
8. Chesnokov V. Image enhancement method and apparatus therefor // Patent WO 02/089060. 2002.
9. Ramirez A., Rivera B. O. Content-Aware Dark Image Enhancement Through Channel Division // IEEE Transactions on Image Processing. № 21(9), С. 3967-3980. 2012.
10. Zuiderveld K. Contrast Limited Adaptive Histogram Equalization // Graphic Gems IV, С. 474-485. 1994.
11. Kurilin I. et al. Fast algorithm for visibility enhancement of the images with low local contrast // IS&T/SPIE Electronic Imaging. – International Society for Optics and Photonics, 2015. – С. 93950B-93950B-9.
12. Otsu N. A threshold selection method from gray-level histograms // Automatica, № 11(285-296). 1975. С. 23-27.
13. Felzenszwalb P. F., Huttenlocher D. P. Efficient graph-based image segmentation // International Journal of Computer Vision. – 2004. – Т. 59. – №. 2. – С. 167-181.
14. Uijlings J.R.R. et al. Selective search for object recognition // International journal of computer vision. – 2013. – Т. 104. – №. 2. – С. 154-171.
15. Michelson A.A. Studies in Optics // University of Chicago. 1927.
16. Wang Z. et al. Image quality assessment: from error visibility to structural similarity // Image Processing, IEEE Transactions on. – 2004. – Т. 13. – №. 4. – С. 600-612.
17. Funt B., Ciurea F., McCann J. Retinex in MATLAB™ // Journal of electronic imaging. 2004. – Т. 13. – №. 1. – С. 48-57.

## ПОВЫШЕНИЕ БЫСТРОДЕЙСТВИЯ АЛГОРИТМА ФРАКТАЛЬНОГО КОДИРОВАНИЯ ПОЛУТОНОВЫХ ИЗОБРАЖЕНИЙ

*Зыков А.Н., АО «Центр Судоремонта «Звёздочка», г. Северодвинск, e-mail: alexzikov@gmail.com.*

### INCREASING OF THE SPEED OF THE ALGORITHM OF FRACTAL CODING THE HALFTONE IMAGES

*Zykov A.N.*

*The offered to your attention work describes using Local Binary Pattern (LBP) in the task of fractal coding grayscale images. Based on local binary models, local binary patterns offer an effective way to analysis of the image and texture are its effective characteristic. This work is aimed at providing new ideas for reducing the time of fractal coding, in particular, using the classification of areas of an image on the basis of using operator the local binary patterns.*

**Key words:** image compression, fractal coding, image analysis, local binary patterns.

**Ключевые слова:** сжатие изображений, фрактальное кодирование, анализ изображений, локальные бинарные шаблоны.

#### Введение

В нашей жизни изображение представляет нам наиболее полную информацию об окружающем мире и поэтому является важным её носителем. В настоящее время, в связи с резким ростом количества цифровой техники, появилось огромное количество цифровой информации. Значительную долю объема цифровой информации занимают графические цифровые изображения. Поскольку цифровое изображение содержит большое количество данных, то оно должно быть эффективно сжато с тем, чтобы занимать меньше места при хранении, и его можно было передавать с более высокой скоростью по сети.

Основной целью кодирования изображений является представление цифровых изображений как можно минимальным количеством битов, сохраняя при этом уровень качества, необходимого для их использования. Методы сжатия изображений имеют важное значение в реальной жизни в сфере информационных технологий. Эффективные технологии кодирования помогающие убрать избыточность и сохранить при этом качество изображения находятся в центре внимания современных исследований. Фрактальное кодирование изображения [1] является одной из таких новых эффективных технологий кодирования.

Сегодня фрактальное кодирование изображений является одним из наиболее эффективных технологий сжатия графической информации с потерями и, как следствие, достаточно перспективным для использования при разработке компьютерных мультимедиа приложений и обработке цифрового видео. Технология фрактального кодирования представляет собой эффективный метод сжатия изображения, основанная на представлении изображения посредством сжимающих аффинных преобразований в пространстве изображений, для которых

*Описывается использование оператора локальных бинарных шаблонов (LBP – Local Binary Pattern) в задаче фрактального кодирования полутоновых изображений. Основываясь на локальной двоичной модели, локальные бинарные шаблоны предлагают эффективный способ анализа текстуры изображения и являются её эффективной характеристикой. Эта работа нацелена на предоставление новых идей для уменьшения времени фрактального кодирования, в частности, в использовании классификации областей изображения на основе использования оператора локальных бинарных шаблонов.*

неподвижная точка близка к исходному изображению. Этот метод привлек к себе большое внимание в последние годы из-за высокой степени сжатия, которая может быть достигнута при его применении и также скоростью восстановления исходного изображения.

Большим недостатком фрактального кодирования является время его кодирования. В 1989 А. Jасquin был предложен метод [2] автоматического кодирования на основе блоков изображения. Данный метод лёг в основу подавляющего большинства схем фрактального кодирования и имел важные последствия для фрактального кодирования изображений в будущем. Процесс кодирования состоит в использовании системы итерированных кусочно-определенных функций (PIFS – Partitioned Iterated Function Systems) в построении оператора, который будет представлять образ закодированного изображения. Если изображение представлено оператором, то его можно эффективно хранить и передавать. Большая часть времени, в построении такого оператора, уходит на поиск лучшего совпадения между частями подлежащего кодированию изображения. Для сокращения вычислительных затрат используют несколько подходов увеличения скорости кодирования. Одно из направлений состоит в классификации областей изображения, которая может значительно ускорить кодирование за счет уменьшения количества используемых областей, среди которых ведется поиск. Классификация происходит путем выделения определенных характеристических особенностей, описывающих области изображения. Выде-

ление небольшого количества особенностей, характеризующих области изображения, позволяет ускорить процесс фрактального кодирования. Тогда сравнение между областями изображения будет производиться на основании этих характеристик, а не по отдельным пикселям, что существенно сокращает объем работы. Эти характеристики могут быть получены на основании спектрального анализа Фурье, вейвлет-анализа, различных характеристик текстуры изображения.

Текстурный анализ является одним из основных в компьютерном зрении и распознавании образов. Текстура со структурной точки зрения состоит из элементов, которые организованы в соответствии с определенными правилами размещения. В данной работе описывается использование локальных бинарных шаблонов для классификации областей изображения в задаче фрактального кодирования изображений. Изображение кодируется методом поиска наилучшего соответствия между областями только с одними и теми же значениями, полученными с помощью оператора локальных бинарных шаблонов. Экспериментальные результаты показывают, что данная схема является компетентной как в скорости кодирования, так и в качестве восстановления.

**Схема фрактального кодирования изображений**

Рассмотрим классическую схему фрактального блочного кодирования изображений методом А. Jacquin. Возьмем полутоновое изображение  $I$  и представим его в виде двумерного прямоугольного массива пикселей:

$$I = i(x, y), \quad x \in [0, w), \quad y \in [0, h),$$

где  $w$  и  $h$  размеры изображения.

Фрактальное кодирование изображения подразумевает нахождение сжимающего аффинного преобразования  $w_i$  области изображения. Для его поиска необходимо выполнить следующие шаги:

Разбиваем изображение  $I$  на прямоугольные ранговые блоки  $R_i$ . Данные ранговые блоки не перекрываются и покрывают всю область изображения, они могут быть одного размера, но могут также иметь адаптивное разбиение с переменным размером блоков.

$$I = \cup R_i \quad i \in [1, n), \quad R_i \cap R_j = 0, \quad i \neq j$$

где  $n$  – количество ранговых блоков.

Покрываем изображение  $I$  множеством прямоугольных доменных блоков  $D_j$ . Эти доменные блоки могут перекрываться, иметь разный размер и их количество может достигать сотни или тысячи единиц.

$$D_j \subseteq I, \quad j \in [1, m), \quad D_i \cap D_j \neq 0, \quad i \neq j$$

где  $m$  – количество доменных блоков.

Для каждого рангового блока  $R_i$  мы находим сжимающее аффинное преобразование  $w_i$ , которое наилучшим образом аппроксимирует данный ранговый блок в доменный блок  $D_j$  (рис. 1).

$$w_i(D_j) \rightarrow R_i$$

При этом каждый ранговый блок  $R_i$  сравнивается со всеми доменными блоками  $D_j$  для нахождения наи-

лучшего совпадения между ними. Данные этих преобразований сохраняются для дальнейшего восстановления исходного изображения.

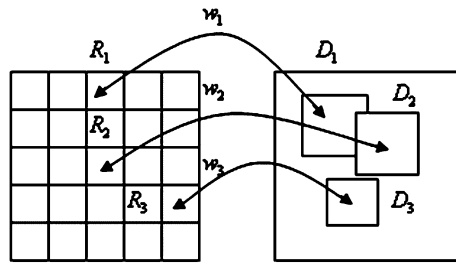


Рис. 1. Отображение ранговых блоков в доменные блоки

Сжимающее аффинное преобразование  $w_i$  – это сочетание геометрических преобразований и преобразований яркости и контраста области изображения. Матрица преобразования  $w_i$  между ранговым и доменным блоком выглядит следующим образом

$$w_i \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} a_i & b_i & 0 \\ c_i & d_i & 0 \\ 0 & 0 & s_i \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} e_i \\ f_i \\ o_i \end{pmatrix},$$

где  $i \in [1, n)$ , коэффициенты  $a_i, b_i, c_i$  и  $d_i$  – геометрическое масштабирование и изометрическое преобразование между ранговым  $R_i$  и доменным  $D_j$  блоками,  $e_i$  и  $f_i$  – координаты смещения между ранговым  $R_i$  и доменным  $D_j$  блоками относительно  $x$  и  $y$  соответственно,  $s_i$  – разность контраста,  $o_i$  – разность яркости между пикселями рангового  $R_i$  и доменного  $D_j$ .

Для того чтобы аффинное отображение доменной области на ранговую область было сжимающим, размеры доменной области должны превышать размеры ранговой области. В процессе данного отображения каждый доменный блок уменьшается в размерах до рангового блока, он также может быть повернут или отражен для лучшего совпадения с ранговым блоком. Также вычисляется разность контраста и яркости между ними. Эти изменения представляют собой сжимающиеся аффинные преобразования доменных блоков (рис. 2).

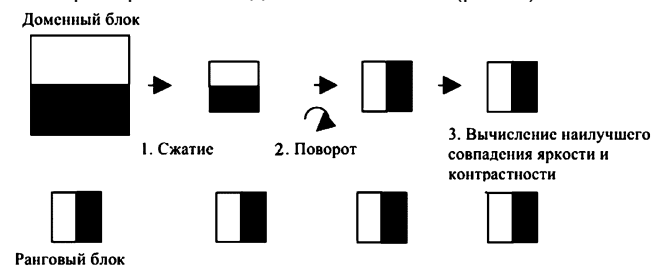


Рис. 2. Схема фрактального кодирования изображения

Для оценки ошибки приближения между ранговым  $R_i$  и доменным  $D_j$  блоками используют среднеквадратичную функцию [3]:

$$E = \sum_{i=1}^N \sum_{j=1}^M (s \cdot d_{i,j} + o - r_{i,j})^2, \quad (1)$$

где  $r_{i,j}$  и  $d_{i,j}$  – соответствующие значения пикселей

ранговой и доменной области,  $N$  и  $M$  – количество строк и столбцов в прямоугольной области.

Контрастность  $s$  и яркость  $o$  определяются следующим образом:

$$s = \frac{a}{b}, \quad o = \bar{r} - \left(\frac{a}{b}\right)\bar{d},$$

где

$$a = \sum_{i=1}^N \sum_{j=1}^M (d_{i,j} - \bar{d}) \cdot (r_{i,j} - \bar{r}),$$

$$b = \sum_{i=1}^N \sum_{j=1}^M (d_{i,j} - \bar{d})^2;$$

$$\bar{d} = \frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M d_{i,j};$$

$$\bar{r} = \frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M r_{i,j}.$$

Хорошее сжатие и качество восстановленного изображения зависят от возможности найти сжимающие аффинное преобразование между доменными и ранговыми блоками с минимальной ошибкой приближения. Совокупность сжимающих аффинных преобразований  $\tau$  представляет собой аффинный коллаж:

$$\tau = \bigcup_{i=1}^N w_i.$$

В ходе фрактального кодирования изображения, для каждой ранговой области, ищется доменная область, которая наилучшим образом аппроксимирует эту ранговую область:

$$E = \min(R_i, \tau(D)).$$

Параметры полученных сжимающих аффинных преобразований представляют собой кодовую книгу, необходимую для восстановления исходного изображения.

Для восстановления исходного изображения из кодовой книги необходимо произвести следующие шаги:

Возьмем произвольное исходное изображение – однотонное или шум. Размеры изображения не имеют значения.

Формируем второе изображение, которое является результатом отображения аффинного коллажа  $\tau$  на исходное изображение.

Меняем местами первое изображение на второе.

Процесс продолжается до тех пор, пока ошибка между этими двумя изображениями не будет меньше определенной величины. На практике для этого вполне достаточно 6-8 циклов.

### Локальные бинарные шаблоны

Оператор локальных бинарных шаблонов позволяет произвести классификацию областей изображения. Впервые данный оператор был предложен Т. Ojala, М. Pietikainen, Т. Maehraa [4] и в дальнейшем развит в работах [5, 6]. Данный оператор преобразует изображение в массив двоичных кодов, описывающих окрестности элементов изображения. Использование данного оператора теоретически простой и одновременно эф-

фективный метод анализа текстуры. Он показывает хорошие результаты во многих эмпирических исследованиях. Может быть представлен как подход, унифицирующий статистические и структурные методы анализа текстуры. Текстура, с его точки зрения, описывается микро-примитивами (textons) и статистикой их размещения.

Сущность оператора локальных бинарных шаблонов заключается в следующем. Рассмотрим окрестность 3x3 вокруг пикселя изображения. Значение центрального пикселя служит пороговой величиной для окружающих его соседних пикселей. Если величина соседнего пикселя имеет значение выше или равное значению центрального пикселя, то ему присваивается значение 1, а если его величина меньше значения центрального пикселя присваивается значение 0. В данной позиции пикселя  $(x_c, y_c)$  оператор локальных бинарных шаблонов определен как упорядоченный набор двоичных сравнений интенсивности пикселя между центральным пикселем и восемью смежными пикселями. Затем, например, по часовой стрелке с левого верхнего угла полученные значения соседних пикселей умножаются на биномиальные веса, приписываемые соответствующим пикселям, и суммируются для получения десятичного кода центрального пикселя (рис. 3).

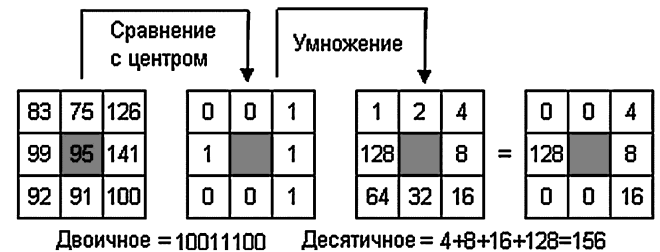


Рис. 3. Вычисление оператора локальных бинарных шаблонов

Окрестность вокруг центрального пикселя состоит из 8 соседних пикселей. Это дает возможность в общей сложности получить  $2^8 = 256$  различных двоичных кодов. Эта возможность зависит от относительных значений тона в центральном пикселе и соседних с ним смежных пикселях. Двоичные числа или их десятичный эквивалент могут быть ассоциированы с центральным пикселем и использоваться как характеристика локальной структуры текстуры изображения вокруг данного пикселя. Следует обратить внимание, что вес каждого бита в двоичном коде, полученном с помощью оператора локальных бинарных шаблонов, имеет одинаковую значимость, а последовательность из двух и более бит может иметь совершенно разный смысл.

Десятичная форма получившегося 8-битного двоичного кода (LBP кода) может быть выражена следующим образом:

$$LBP(x_c, y_c) = \sum_{n=0}^7 s(i_n - i_c) 2^n,$$

где  $i_c$  – значение центрального пикселя  $(x_c, y_c)$ ,  $i_n$  – значения восьми соседних пикселей.

Функция  $s(x)$  определена как:

$$s(x) = \begin{cases} 1 & \text{if } x \geq 0, \\ 0 & \text{if } x < 0. \end{cases}$$

Данный код можно истолковать как индекс ядра структуры текстуры изображения размерностью  $3 \times 3$ , которое суммирует локальную пространственную структуру изображения. По определению оператор локальных бинарных шаблонов не зависит от любого монотонного преобразования тона изображения, которое сохраняет порядок интенсивности пикселя в локальной области и, следовательно, является менее чувствительным к изменениям освещенности. Оператор локальных бинарных шаблонов также инвариантен к масштабированию, вращению и переносу. На рис. 4 изображен циклический поворот двоичного кода центрального пикселя.

0	0	1	10011100 = 156	11001001 = 201
1		1	01001110 = 78	11100100 = 228
0	0	1	00100111 = 39	01110010 = 114
			10010011 = 147	00111001 = 57

Рис. 4. Циклический поворот двоичного кода

Благодаря этим своим отличительным свойствам и своими очень низкими вычислительным затратам, оператор локальных бинарных шаблонов становится довольно эффективным методом для анализа структуры текстуры изображения.

### Эксперимент

При фрактальном кодировании происходит поиск наилучшего соответствия между ранговым и доменным блоками. Этот процесс требует наибольших вычислительных затрат. Есть несколько подходов для сокращения вычислительных затрат и увеличения скорости кодирования. В настоящее время данные работы ведутся по нескольким направлениям исследований. Процесс классификации доменных областей одно из таких направлений исследований. Классификация позволяет значительно ускорить процесс кодирования за счет значительного уменьшения количества доменных областей используемых для поиска подобия.

Рассмотрим предложенную схему фрактального кодирования, использующую классификацию областей изображения с применением оператора локальных бинарных шаблонов.

#### Шаг 1. Предварительные вычисления.

Создание библиотеки ранговых областей.

Изображение разбивается на не перекрывающиеся области размером  $3 \times 3$  пикселя.

Для классификации каждой области, вычисляем двоичный (LBP) код её центрального пикселя, используя оператор локальных бинарных шаблонов.

Создадим библиотеку LBP кодов для каждой ранговой области.

Создание библиотеки доменных областей.

Уменьшим исходное изображение в два раза.

Вычисляем для каждого пикселя изображения, с помощью оператора локальных бинарных шаблонов, его LBP код. Полученные LBP коды будут классифицировать доменные области. Создадим библиотеку LBP кодов для каждой доменной области.

Для крайних и угловых пикселей недостающие соседние пиксели можно принять за 0 или дополнить циклически на противоположный край изображения.

#### Шаг 2. Кодирование ранговых областей.

Из библиотеки ранговых областей берем LBP код выбранной ранговой области.

Из библиотеки доменных областей выбираем только те области, у которых их LBP код равен LBP коду выбранной ранговой области. По формуле (1) производим вычисление ошибки приближения между ранговой областью и всеми доменными областями с тем же LBP кодом. Для доменной области с наименьшей ошибкой приближения сохраняем полученные коэффициенты преобразования для дальнейшего восстановления исходного изображения.

При отсутствии в библиотеке доменных областей области со значением LBP кода ранговой области производим циклический поворот LBP кода ранговой области (рис. 4) и повторяем поиск доменной области в библиотеке с новым LBP кодом. Если в результате полного циклического поворота, отсутствует полное соответствие между LBP кодами ранговой и доменной области, то можно взять доменные области с LBP кодом, близкие по величине LBP коду ранговой области. Находим параметры аффинного преобразования для всех ранговых блоков и сохраняем их в виде аффинного коллажа для последующего восстановления исходного изображения.

Теперь сравним сложность работы данного подхода с классическим методом фрактального кодирования. Предположим, что  $N$  – количество ранговых областей, а  $M$  – количество доменных областей. Исходя из классического алгоритма фрактального кодирования, чтобы найти доменную область, имеющую наименьшую ошибку приближения с ранговой областью, необходимо провести полный перебор всех доменных областей библиотеки. При этом сложность классического метода фрактального кодирования равняется  $O(N * M)$ . В подходе, описанном в данной работе, выбор доменной области ограничивается областями того же класса, что и класс рангового блока. При этом сложность метода фрактального кодирования с классификацией областей изображения с применением локальных бинарных шаблонов равняется  $O(N * \log_2 M)$ . Видно существенное увеличение скорости работы предложенного алгоритма фрактального кодирования на основе оператора локальных бинарных шаблонов.

Для оценки качества восстановленного изображения используем два критерия, наиболее часто используемые для измерения уровня искажений при сжатии изображений:

$PSNR$  – пиковое отношение сигнала к шуму [8]. Данный критерий является инженерным термином, означающим соотношение между максимумом возможного значения сигнала и мощностью шума, искажающего значения сигнала. Проще всего его определить через среднеквадратичную ошибку MSE.

$$MSE = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n (I_{i,j} - K_{i,j})^2,$$

где два монохромных изображения  $I$  и  $K$  размера  $m \times n$ , одно из которых считается зашумленным приближением другого:

$$PSNR = 20 \log_{10} \left( \frac{2^B - 1}{\sqrt{MSE}} \right),$$

где  $B$  разрядность бит изображения.

$SSIM$  – индекс структурного сходства [9]. Является одним из методов измерения схожести между двумя изображениями.  $SSIM$ -индекс это метод полного сопоставления, он проводит измерение качества на основе исходного изображения и учитывает «восприятие ошибки» благодаря учёту структурного изменения информации. Идея данного метода заключается в том, что пиксели имеют сильную взаимосвязь, особенно когда они близки пространственно:

$$SSIM = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)},$$

где  $\mu_x$  – среднее  $x$ ,  $\mu_y$  – среднее  $y$ ,  $\sigma_x^2$  – дисперсия  $x$ ,  $\sigma_y^2$  – дисперсия  $y$ ,  $\sigma_{xy}$  – ковариация  $x$  и  $y$ ,  $c_1 = (k_1L)^2$ ,  $c_2 = (k_2L)^2$  – две переменные,  $L = (2^{(bits \text{ per pixel})} - 1)$  – динамический диапазон пикселей,  $k_1 = 0,01$ ,  $k_2 = 0,03$  – константы.

Полученный  $SSIM$ -индекс лежит в пределах от  $-1$  до  $+1$ . Значение  $+1$  достигается только при полной аутентичности образцов.

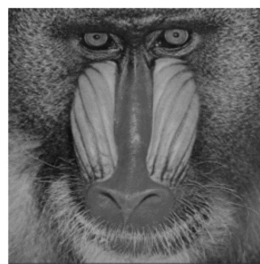
При проведении экспериментов использовался пакет прикладных программ MatLab на базе компьютера с процессором Intel Pentium Dual-Core с тактовой частотой 1,8 ГГц и системной памятью 1 Гб.

Таблица 1. Сравнение методов фрактального кодирования

Изображение	Классический			LBP		
	$T_c$	PSNR	SSIM	$T_c$	PSNR	SSIM
Airplane	61,3	35,03	0,9879	1,8	32,19	0,9859
Baboon	61,2	29,66	0,9594	1,7	27,76	0,9497
Barbara	59,3	32,55	0,9833	1,8	31,04	0,9796
Cameraman	52,6	30,64	0,9780	1,6	28,86	0,9483
House	48,3	37,78	0,9875	1,4	34,57	0,9849
Lena	59,0	35,63	0,9900	1,4	32,35	0,9842
Peppers	60,2	33,34	0,9901	1,3	31,22	0,9834
Sailboat	56,4	33,54	0,9876	1,9	31,46	0,9845



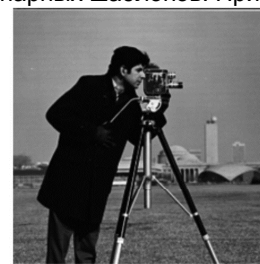
Airplane



Baboon



Barbara



Cameraman

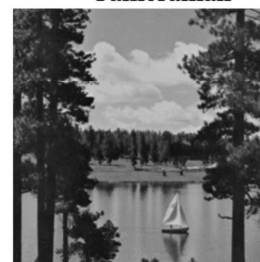
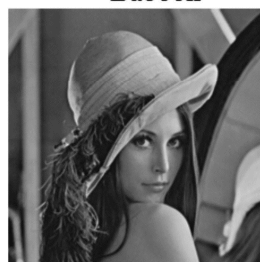
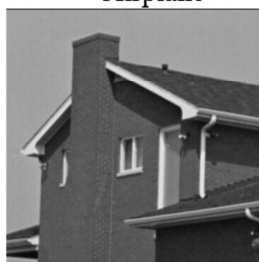


Рис. 5. Набор полутоновых изображений

Результаты экспериментов на полутоновых изображениях [7] (рис. 4), представленных ниже, приведены в табл. 1, где  $T_c$  – обозначает время работы алгоритма.

В результате проведенных экспериментальных исследований, можно сделать следующие выводы:

Использование локальных бинарных шаблонов в качестве классификатора областей изображения, позволяет значительно сократить количество сравнений между ранговыми и доменными областями.

Происходит значительное уменьшение времени работы алгоритма.

Погрешность кодирования сопоставима с погрешностью кодирования при классическом методе.

Незначительная потеря качества восстановленного изображения при использовании предложенного метода фрактального кодирования может быть компенсирована за счет увеличения количества просматриваемых доменных областей.

При расширении количества доменных областей, дополнительных вычислительных затрат почти не требуется, т.к. вычисление LBP кода выполняется достаточно легко.

## Заключение

В работе рассмотрено фрактальное кодирование изображения на основе классификации областей изображения оператором локальных бинарных шаблонов. Оператор локальных бинарных шаблонов представляет простой, эффективный метод классификации текстуры полутоновых изображений, инвариантный к масштабированию, вращению и переносу. Другое важное преимущество данного подхода – вычислительная простота и универсальный подход для анализа текстуры любых полутоновых изображений.

Применение данного подхода позволяет значительно уменьшить временные и сократить вычислительные расходы при фрактальном кодировании изображений и получить хорошие экспериментальные результаты.

Предложен новый алгоритм «неполного» перебора на основе двоичных значений, полученных с помощью оператора локальных бинарных шаблонов. При исполь-

звании этого алгоритма «с потерей качества» ошибка восстановления изображения мало уступает классическому алгоритму, а по скорости превосходит его.

Получено значительное увеличение скорости работы предложенного алгоритма фрактального кодирования при сопоставимой степени сжатия.

Из всего этого следует, что требуется дальнейшее изучение оператора локальных бинарных шаблонов при фрактальном кодировании изображений для его применения при обработке больших объемов графической информации.

#### Литература

1. Barnsley M., and Hurd L., *Fractal Image Compression* // Wellesley, MA: A.K.Peters, Ltd., 1993.
2. Jacquin A.E. *Fractal image coding: A review* // *Proceedings of the IEEE*, 1993. Vol. 81(10). P. 1451-1465.
3. Welstead S., *Fractal and Wavelet Image Compression Techniques* // *Tutorial Texts in Optical Engineering*, SPIE Publications, 1999. Vol 40.
4. Ojala T., Pietikainen M., Maenpaa T. *Multi resolution gray-scale and rotation invariant texture classification with local binary patterns* // *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, Vol.24, №.7. P. 971–987.
5. Ahonen T., Hadid A., and Pietikainen M. *Face recognition with local binary patterns* // *Proc. European Conf. Computer Vision*, 2004, P. 469–481.
6. Takala V., Ahonen T., Pietikainen M. *Block-based methods for image retrieval using local binary patterns* // *Scandinavian Conference on Image Analysis*, Joensuu, Finland, 2005, P. 882–891.
7. <https://sipi.usc.edu/database/database.php?volume=misc>.
8. <https://ru.wikipedia.org/wiki/PSNR>.
9. <https://ru.wikipedia.org/wiki/SSIM>.



УДК 004.932

## ОПТИМИЗАЦИЯ АЛГОРИТМОВ СЕГМЕНТАЦИИ С ПРИМЕНЕНИЕМ МЕТОДОВ ПАРАЛЛЕЛЬНЫХ ВЫЧИСЛЕНИЙ ДЛЯ СИСТЕМ АНАЛИЗА ИЗОБРАЖЕНИЙ

*Степанов В.Н., к.т.н., с.н.с. Института проблем управления им. В.А. Трапезникова РАН,  
e-mail: vnstepanov@yandex.ru.*

### SEGMENTATION ALGORITHMS OPTIMIZATION WITH PARALLEL COMPUTING TECHNIQUES FOR IMAGE ANALYSIS SYSTEMS

*Stepanov V.N.*

*The paper presents an analysis of the General Processing on Graphical Processing Unit (GPGPU) in relation to image processing (segmentation) for analyzers of medical and biological micro-objects. Adaptation of processing algorithms for parallelization and increase in execution speed are shown.*

**Key words:** parallel computing, image processing, segmentation, GPGPU.

**Ключевые слова:** параллельные вычисления, обработка изображений, сегментация, общие вычисления на графических процессорах.

#### Введение

За последнее десятилетие широкое распространение получила технология общих вычислений на графических процессорах (GPGPU – General Processing on Graphical Processing Unit). Развитие видеоадаптеров шло благодаря компьютерным играм и в определенный момент быстродействие достигло такого уровня, что позволило рассматривать видеоадаптер как самостоятельный вычислительный модуль, применяемый в любых задачах.

На данный момент второй по вычислительной мощности суперкомпьютер Cray Titan использует специализированные вычислительные блоки Tesla K20X, которые являются, по сути, бытовыми видеоадаптерами NVidia GeForce GTX Titan.

Работа с технологией GPGPU осуществляется с помощью аппаратно-зависимых и аппаратно-независимых реализаций. К первым относятся NVidia CUDA и AMD Firestream, рассчитанные на работу с видеоадаптерами соответственно NVidia и AMD. Эти реализации, как правило, позволяют достичь большего быстродействия, т.к. созданы самими разработчиками видеоадаптеров и используют все их нюансы. Поэтому аппаратно-зависимые реализации применяют обычно для очень ресурсоемких задач на уникальных аппаратных платформах, таких как суперкомпьютеры.

Для менее ресурсоемких, но более массовых задач, таких как обработка изображений, целесообразно использовать универсальные аппаратно-независимые реализации. В данной работе используется аппаратно-независимая реализация OpenCL [1, 2]. Эта реализация основана на языке C и разрабатывается некоммерческим консорциумом KhronosGroup, в который входят большинство основных производителей видеоадаптеров и центральных процессоров.

В данной работе исследования проводились на ап-

*Представлен анализ технологии общих (неграфических) вычислений на графических процессорах (GPGPU) применительно к обработке изображений (сегментации) для систем анализа и обработки изображений медико-биологических микрообъектов. Приводится адаптация алгоритмов обработки для распараллеливания и повышения производительности.*

паратной платформе следующей конфигурации: Процессор Intel Core i7-860, оперативная память DDR3 12Гб, видеоадаптер nVidia GeForce GTX 970. Программная среда EmbarcaderoRADStudio 10 Seattle.

#### Анализ эффективности технологии GPGPU для сегментации изображений на примере системы «Морфолог-Сеть»

Обработка изображений является удобной задачей для применения параллельных вычислений. Изображение состоит из пикселей и, как правило, при обработке каждый пиксель изменяется независимо от результата изменения предыдущих. То есть все пиксели можно изменять одновременно. В компьютерном анализаторе «Морфолог сеть» [3] проводится предварительная фильтрация изображения, сегментация и постобработка полученного семантического описания.

Наиболее ресурсоемкая операция – это сегментация. Уже существуют работы по применению параллельных вычислений к различным методам сегментации изображений. Например, в [4] автор адаптировал алгоритм Mean Shift к выполнению на GPU и получил прирост скорости обработки в 97 раз (на GPU NVidia GeForce GTX 470 относительно процессора Intel Core i5 750). Этот алгоритм применяется для сегментации изображений с неизвестным количеством классов. При анализе изображений медико-биологических препаратов обычно заранее известно количество классов и их признаки, поэтому алгоритм Mean Shift может привести к появлению ложных классов. В [5] показана теоретическая возможность распараллеливания генетического алгоритма сегментации трехмерных изображений. При анализе изображений медико-биологических препаратов обычно име-

ется большой объем исходных данных, помимо самого изображения, которые необходимо использовать в том числе и при сегментации. Это позволит повысить точность анализа. Поэтому целесообразно применение специализированных алгоритмов. В данной работе показана адаптация под параллельные вычисления интерактивного метода пороговой сегментации, разработанного для системы анализа изображений «Морфолог-сеть». Алгоритм интерактивной сегментации подразумевает сравнение каждого пикселя изображения с массивом опорных точек, с вычислением расстояний в трехмерных цветовых пространствах [6]. Интерактивность метода заключается не только в выборе опорных точек, принадлежащих конкретной области сегментации, но и в настройке коэффициентов расстояний, при одновременном наблюдении результатов сегментации. Для этого требуется достаточно большая частота обновления результата. Однопоточный алгоритм на центральном процессоре не может обеспечить достаточно быстрого действия и приходилось отображать результат в маленьком окне предварительного просмотра, что не очень удобно. Применение алгоритмов параллельной обработки на графическом процессоре позволит отображать результат с приемлемой частотой для всего изображения (или увеличить на порядок размер окна предварительного просмотра для очень больших изображений).

Для анализа быстродействия параллельного алгоритма сегментации был сгенерирован ряд синтетических изображений размерами 2000x2000, 4000x2000, 4000x4000 и 8000x4000 пикселей. Пример такого изображения приведен на рис. 1. Изображения создавались на основе реальных изображений препаратов предстательной железы [7].

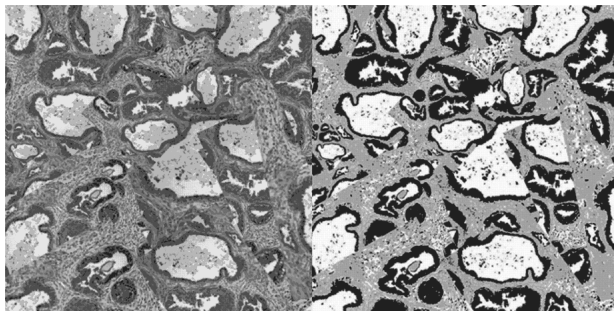


Рис. 1. Пример синтетического изображения на основе препаратов предстательной железы размером 2000x2000 пикселей для анализа производительности: а) исходное, б) сегментированное

В системе «Морфолог-сеть» все расчеты при сегментации могут вестись как в пространстве  $RGB$ , так и  $HSV$  (цвет, насыщенность, яркость). Пространство  $HSV$  лучше подходит для выделения областей на изображении, как показано в [6], но требует дополнительных вычислений.

На рис. 2 показан график увеличения времени сегментации на центральном процессоре в пространстве  $RGB$  и  $HSV$  в зависимости от размера изображения. При размере изображения 32 МРiх падение производительности достигает 24 %.

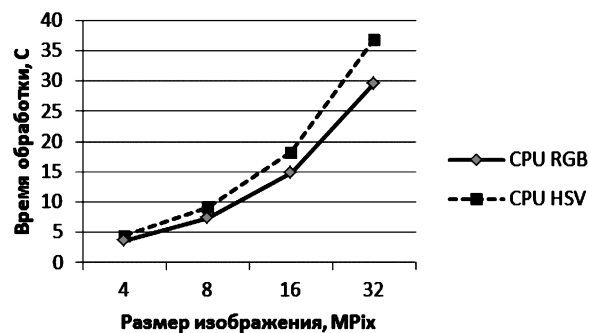


Рис. 2. График зависимости скорости расчета (с) на центральном процессоре от размера изображения (МРiх) для пространства  $RGB$  и  $HSV$

При сегментации с помощью технологии GPGPU преобразование пространства цветов  $RGB$  к  $HSV$  производится по формулам:

$$R < B < G \left\{ \begin{array}{l} H = \frac{1}{3} - \frac{G-B}{6 \times (G-R)}; \\ S = \frac{G-R}{G}; \\ V = G; \end{array} \right.$$

$$G < R < B \left\{ \begin{array}{l} H = \frac{2}{3} - \frac{B-R}{6 \times (B-G)}; \\ S = \frac{B-G}{B}; \\ V = B; \end{array} \right.$$

$$G < B < R \left\{ \begin{array}{l} H = \frac{R-B}{6 \times (R-G)}; \\ S = \frac{R-G}{R}; \\ V = R; \end{array} \right.$$

$$B < R < G \left\{ \begin{array}{l} H = \frac{1}{3} + \frac{G-R}{6 \times (G-B)}; \\ S = \frac{G-B}{G}; \\ V = G; \end{array} \right.$$

$$B < G < R \left\{ \begin{array}{l} H = -\frac{R-G}{6 \times (R-B)}; \\ S = \frac{R-B}{R}; \\ V = R. \end{array} \right.$$

где  $R, G, B$  – соответственно красный, синий и зеленый компоненты цвета, а  $H, S, V$  – цвет, насыщенность и яркость. Все параметры могут принимать значения от 0 до 1. Это не оптимальные формулы с точки зрения математики, но они более удобны для вычисления специализированным графическим процессором по технологии GPGPU. Кроме того, при кодировании формул были

максимально широко использованы векторные операции [2], на которые графический процессор рассчитан. Это позволило сократить разрыв в быстродействии между алгоритмами с RGB и HSV до 5 %. На рис. 3 приведен аналогичный график (рис. 2), но с использованием графического процессора.

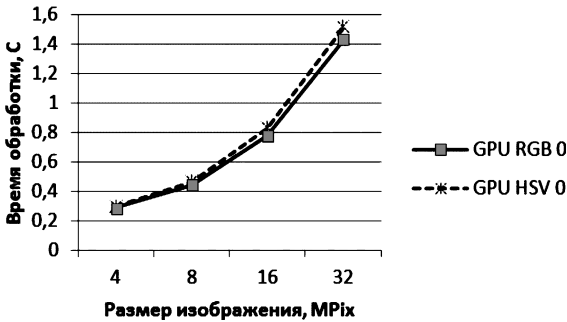


Рис. 3. График зависимости скорости расчета (с) на графическом процессоре от размера изображения (MPix) для пространства RGB и HSV

Расстояние между текущей точкой изображения и опорными точками для пространства RGB вычисляется по формуле:

$$D = \sqrt{(Rn - Ri)^2 + (Gn - Gi)^2 + (Bn - Bi)^2},$$

где  $Rn, Gn, Bn$  – красный, зеленый и синий компонент цвета текущей опорной точки, а  $Ri, Gi, Bi$  – компоненты цвета текущей точки изображения.

Для пространства HSV формула имеет вид:

$$D = \sqrt{(Xn - Xi)^2 + (Yn - Yi)^2 + (Vn - Vi)^2},$$

где  $Vn, Vi$  – значение яркости опорной точки и точки изображения, а  $Xn, Xi, Yn, Yi$  – координаты, преобразованные из полярных координат HS по формулам:

$$X = \cos(H) \times S;$$

$$Y = -\sin(H) \times S.$$

Для вычисления расстояния применена быстрая функция half\_sqrt [2]. Эта функция обладает пониженной точностью, а именно 8192 ULP. ULP – это единица наименьшей точности (unit of least precision), то есть минимальное расстояние между соседними числами с плавающей точкой. Для 32 битных вычислений это 0,00000011920929. Таким образом, функция half\_sqrt выдает результат с шагом  $0,00000011920929 \times 8192 = 0,00097656250368$ . Сравнение результатов сегментации с полной и пониженной точностью показало отсутствие различий.

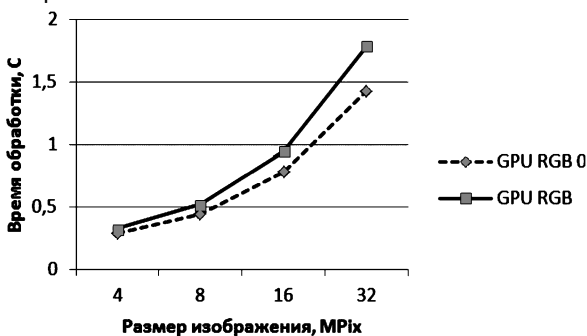


Рис. 4. График зависимости скорости расчета (с) от размера изображения (MPix) для быстрой функции (GPURGB0) и функции с полной точностью (GPURGB)

Применение быстрой функции позволяет увеличить быстродействие в среднем на 15,9 %, причем разница растет вместе с увеличением размера обрабатываемого изображения. Это объясняется тем, что при увеличении размера изображения уменьшается процент накладных расходов, связанных с подготовкой изображения, и увеличивается процент собственно вычислений.

В табл. 1 представлены все результаты изменений времени работы алгоритмов сегментации, для однопоточного алгоритма на центральном процессоре в пространстве RGB (CPURGB), параллельного алгоритма на графическом процессоре (GPURGB), быстрого параллельного алгоритма с пониженной точностью (GPURGB0), однопоточного алгоритма в пространстве HSV (CPUHSV) и быстрого параллельного алгоритма в пространстве HSV (GPUHSV 0).

Таблица 1.

	CPU RGB	GPU RGB 0	GPU RGB	CPU HSV	GPU HSV 0
4	3,658	0,292	0,329	4,567	0,301
8	7,387	0,444	0,522	9,18	0,466
16	14,772	0,778	0,947	18,402	0,83
32	29,646	1,429	1,782	36,92	1,513

Видно, что увеличения быстродействия от применения параллельных вычислений достигает 25 раз. На рис. 5 представлена информация из табл. 1 в графическом виде, три графика для GPU практически слились в один.

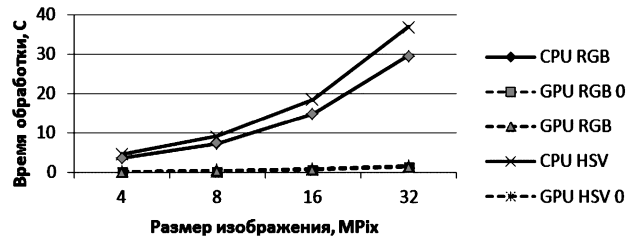


Рис. 5. График зависимости скорости расчета (с) от размера изображения (MPix)

### Заключение

В результате проведенных исследований можно отметить, что обработка изображений является хорошо распараллеливаемой задачей. Применение параллельных алгоритмов с использованием технологии общих вычислений на видеоадаптере позволило ускорить сегментацию изображений до 25 раз.

Использование быстрых функций с пониженной точностью не сказывается на результате сегментации – попиксельное сравнение сегментированных изображений показало, что результат не отличается от такового при расчетах с полной точностью. Ускорение при этом составляет в среднем 15,9 %. Применение оптимизированного алгоритма преобразования пространства RGB в HSV позволило увеличить быстродействие еще на 19 %.

Таким образом, параллельная обработка на используемом в работе оборудовании (видеоадаптер NVidia GeForce 970) позволяет проводить сегментацию реальных изображений (обычно не более 1 MPix) со скоростью, приближенной к реальному времени. Задержка составляет 170 мс. Такая задержка позволяет отказаться от окна предварительного просмотра при настройке порогов сегментации, как было ранее, и выводить результат прямо на изображение.



## Литература

1. AaftabMunshi; Benedict R. Gaster; Timothy G. Mattson; James Fung; Dan Ginsburg. OpenCLProgramming Guide. – Addison-WesleyProfessional, 2011.
2. AaftabMunshi. The OpenCL Specification 1.0.48. KynosGroup. 2009
3. Попова Г.М., Степанов В.Н., Дружинин Ю.О., Дятчина И.Ф. Многофункциональный информационно-вычислительный комплекс анализа и диагностики изображений // Информационные технологии и вычислительные системы. 2010. № 4. С. 25-37.
4. Димашова М.П. Реализация алгоритма сегментации изображений MeanShift на GPU // Сборник трудов X Всероссийской конференции «Высокопроизводительные параллельные вычисления на кластерных системах» 2010. Т.1 – с. 214
5. Дрынчик Р.В., Привалов М.В. Распараллеливание эволюционных методов сегментации трехмерных медицинских изображений.// Информационные управляющие системы и компьютерный мониторинг. – Донецк :ДонНТУ, 2012. - С. 480 - 484.
6. Попова Г.М., Степанов В.Н., Дружинин Ю.О. Интерактивный метод обработки цветных изображений объектов биологической природы для получения их семантического описания // Системный анализ и управление в биомедицинских системах. 2009. Т. 8, № 3. С. 741-746.
7. Попова Г.М., Дружинин Ю.О., Степанов В.Н., Дятчина И.Ф., Чазова Н.Л., Берцанская А.М., Мельникова Н.В. Количественная диагностика онкологических заболеваний предстательной железы с использованием компьютерного анализатора «Морфолог-Сеть» // Системный анализ и управление в биомедицинских системах. 2006. Т. 5, № 4. С. 943-954.

УДК 621.397

## ФИЛЬТРАЦИЯ ЦИФРОВЫХ ИЗОБРАЖЕНИЙ НА ОСНОВЕ АВТОЭНКODERA

*Ипатов А.А., аспирант кафедры динамики электронных систем Ярославского государственного университета им. П. Г. Демидова, e-mail: artoymirator@gmail.com;*

*Волохов В.А., к.т.н., доцент кафедры динамики электронных систем Ярославского государственного университета им. П. Г. Демидова, e-mail: volokhov@piclab.ru;*

*Приоров А.Л., д.т.н., доцент кафедры динамики электронных систем Ярославского государственного университета им. П. Г. Демидова, e-mail: andcat@yandex.ru;*

*Апальков И.В., к.т.н., доцент кафедры динамики электронных систем Ярославского государственного университета им. П. Г. Демидова, e-mail: ilya@apalkoff.ru.*

## FILTRATION OF DIGITAL IMAGES BASED ON AUTOENCODER

*Ipatov A.A., Volokhov V.A., Priorov A.L., Apalkov I.V.*

*This paper presents the implementation and study of noise reduction algorithm, based on autoencoder. Autoencoder is a kind of feedforward neural network, which is unsupervised learning algorithm. Standard dataset were used to test the proposed filtering algorithm. Additive white Gaussian noise is considered as a noise model. This paper presents the numerical and visual results, showing the main features of considered algorithm.*

**Key words:** image filtering, machine learning, feedforward neural network, autoencoder.

**Ключевые слова:** фильтрация изображений, машинное обучение, нейронная сеть прямого распространения, автоэнкодер.

### Введение

На сегодняшний день алгоритмы цифровой обработки изображений, а именно шумоподавления [1], широко применяются в области современной науки и техники и имеют множество практических приложений. Например, можно использовать алгоритмы шумоподавления в качестве этапа предобработки цифровых изображений в системах технического зрения или для обработки медицинских снимков, сделанных в разных диапазонах длин волн. Также для человека наиболее важной частью информации является ее визуальная составляющая, именно поэтому обработка цифровых изображений столь необходима на данный момент времени.

Подавление шума в основном используется с целью улучшения визуального восприятия, то есть предназначено для приближения «искаженного» изображения к «идеальному». В частности, изображения, полученные в неблагоприятных условиях съемки, подвергаются воздействию различных шумов, которые создают электроникой устройств (полупроводниковые сенсоры) фотоаппарата. Вследствие чего, для получения высококачественных изображений необходимо разрабатывать подходы, позволяющие подавлять эти шумы.

Стоит отметить, что алгоритмы фильтрации обычно разрабатываются для подавления какого-то конкретного шума, так как на данный момент не существует универсальных фильтров, способных детектировать и подавлять все возможные виды шумов. Однако многие шумы можно довольно хорошо аппроксимировать моделью аддитивного белого гауссовского шума (АБГШ),

*Представлены реализация и исследование алгоритма шумоподавления, построенного на основе автоэнкодера. Автоэнкодер является разновидностью нейронной сети прямого распространения, обучаемой без учителя. Для тестирования предложенного алгоритма фильтрации использовались стандартные базы тестовых изображений. В качестве модели шума рассматривался аддитивный белый гауссовский шум. Приведены численные и визуальные результаты, демонстрирующие основные особенности предложенного алгоритма.*

что следует из центральной предельной теоремы, в которой говорится о том, что сумма из большого числа случайных величин, имеющих произвольное распределение, наилучшим образом описывается гауссовским распределением. Именно поэтому в настоящей работе при реализации и тестировании алгоритмов фильтрации использовалась модель АБГШ.

Достаточно интересным подходом к задаче фильтрации изображений является использование методов машинного обучения [2]. Машинное обучение – это обширная область науки и техники, подраздел искусственного интеллекта, использующий разделы математической статистики, теории вероятностей, численные методы оптимизации для извлечения полезной информации (знания) из данных. Методы машинного обучения используются для классификации, кластеризации, анализа больших баз данных и для других актуальных на сегодняшний день научно-технических задач. В работе предлагается реализация алгоритма фильтрации цифровых изображений, построенная на основе одного из методов машинного обучения – автоэнкодера.

### Описание алгоритма фильтрации

Автоэнкодер (автоассоциатор) представляет собой специальную нейронную сеть прямого распространения

[2, 3], которая обучается без учителя, то есть, не используя размеченную базу данных. Это нейронная сеть подбирает значения выходов таким образом, чтобы они были как можно ближе к значениям на входе сети. На рис. 1 представлена модель трехслойного автоэнкодера (один входной слой, один скрытый и один выходной). Входной слой содержит шесть входных единиц (нейронов) и одну единицу смещения  $x_0 = 1$ ; скрытый слой содержит три скрытые единицы и одну единицу смещения  $a_0^{(2)} = 1$ ; выходной слой содержит такое же число единиц, что и входной.

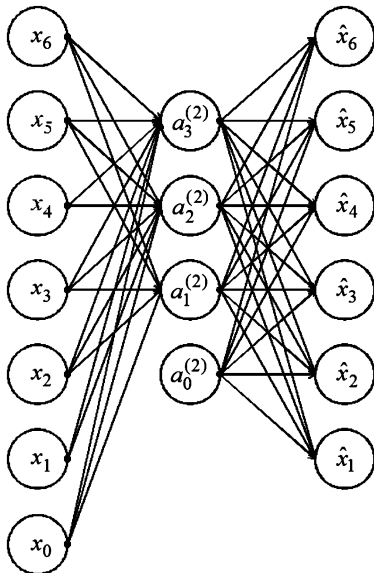


Рис. 1. Схема автоэнкодера, где  $x_0 = 1$  и  $a_0^{(2)} = 1$  – единицы смещения во входном и скрытом слоях.

Кружками обозначены нейроны в слоях, а стрелками показаны соединения между нейронами разных слоев

Автоэнкодер пытается получить функцию (гипотезу)  $\mathbf{h}_{w,b}(\mathbf{x}) \approx \mathbf{x}$ , которая аппроксимирует вход  $\mathbf{x}$  оценкой  $\hat{\mathbf{x}}$  на выходе, как показано на рис. 1. Для того чтобы преобразование не было тривиальным, на скрытый слой автоэнкодера накладывают ограничение. Число единиц в скрытом слое должно быть меньше, чем во входном. Идентичное отображение  $\hat{\mathbf{x}} = \mathbf{x}$  не позволяет получить «интересного» решения.

Ограничение числа нейронов в скрытом слое позволяет получить сжатое представление входных данных, в смысле понижения их размерности. Это означает, что автоэнкодер обучается восстанавливать входные данные из сжатого представления, что может быть полезно для реализации алгоритма шумоподавления.

Обучение такой нейронной сети выполняется с использованием метода обратного распространения ошибки путем минимизации стоимостной функции, например, с помощью градиентного спуска, позволяющего настроить веса (параметры модели) автоэнкодера и получить гипотезу  $\mathbf{h}_{w,b}(\mathbf{x}) \approx \mathbf{x}$ .

Рассмотрим алгоритм шумоподавления на основе модели трехслойного автоэнкодера, пример которого приведен выше. Схема предлагаемого алгоритма фильтрации изображения на основе автоэнкодера представлена на рис. 2.

На вход алгоритма поступает зашумленное АБГШ  $\mathbf{n}$

с нулевым математическим ожиданием и известной дисперсией  $\sigma^2$  изображение  $\mathbf{y} = \mathbf{x} + \mathbf{n}$ . Обозначим через  $m(i)$  число единиц в конкретном слое автоэнкодера, где индекс  $i = 1, 2, 3$  обозначает номер слоя сети.

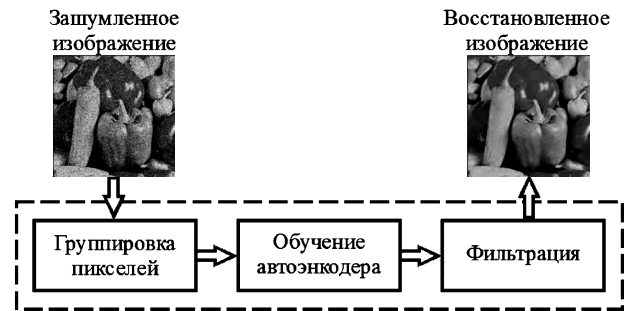


Рис. 2. Схема фильтрации изображения на основе автоэнкодера

Основные этапы работы предлагаемого алгоритма представлены ниже.

1. Выделяем на зашумленном изображении  $y$  совокупность перекрывающихся блоков, каждый из которых состоит из области обучения, области фильтрации и области наложения блоков [4, 5].

2. Для обучения автоэнкодера необходимы данные, взятые из зашумленного изображения, поэтому требуется сформировать матрицу данных  $S_y$ , столбцами которой являются всевозможные вектора обучения, набранные внутри области обучения. При формировании векторов обучения внутри области обучения выбираются всевозможные блоки размером  $l \times l$ , которые затем представляются в виде векторов столбцов размером  $l^2$ . Матрица данных  $S_y$  имеет размер  $l^2 \times n$ , где  $l^2$  – размер векторов обучения, то есть размерность данных,  $n$  – количество набранных векторов обучения.

3. Подготавливаем обрабатываемые данные, состоящие из яркостей пикселей изображения  $y$ , для обучения автоэнкодера, а именно, производим масштабирование данных в  $S_y$  путем деления их на максимальное значение яркости пикселя в зашумленном изображении. Полученную матрицу обозначим через  $\bar{S}_y$ .

4. Используем модель автоэнкодера, у которого число входных единиц  $m^{(1)}$  в первом слое будет равно размеру вектора обучения  $l^2$ , то есть число нейронов во входном слое зависит от размерности данных в матрице  $\bar{S}_y$ . Количество нейронов в скрытом слое должно быть меньше количества нейронов во входном слое ( $m^{(1)} < m^{(2)}$ ). Это условие требуется для того, чтобы автоэнкодер сжал входные данные, отбросив связанные с шумом компоненты. Этим автоэнкодер очень похож на один из часто используемых методов сокращения размерности данных – анализ главных компонент [2].

5. Выполняем алгоритм прямого распространения сигнала, который в данном случае описывается следующими выражениями, представленными в матричной форме:

$$\mathbf{a}_j^{(2)} = f(\mathbf{z}_j^{(2)}) = f(\mathbf{W}^{(1)}(\bar{S}_y)_j + \mathbf{b}^{(1)}), \quad (1)$$

$$\mathbf{a}_j^{(3)} = \mathbf{h}_{w,b}((\bar{S}_y)_j) = \mathbf{z}_j^{(3)} = \mathbf{W}^{(2)}\mathbf{a}_j^{(2)} + \mathbf{b}^{(2)}. \quad (2)$$

Активации единиц скрытого слоя автоэнкодера получают при использовании нелинейного преобразования

$f$ , основанного на сигмоидной функции [2], а активации нейронов на выходе сети – при помощи линейного преобразования, определяющего линейный декодер. В выражениях (1) и (2)  $(\bar{S}_y)_j$  –  $j$ -й вектор-столбец матрицы  $\bar{S}_y$ ,  $\mathbf{a}_j^{(2)}$  – вектор-столбец размерностью  $m^{(2)}$  рассчитанных активаций нейронов скрытого слоя автоэнкодера,  $\mathbf{a}_j^{(3)}$  – вектор-столбец размерностью  $m^{(1)}$  рассчитанных активаций нейронов выходного слоя,  $\mathbf{W}^{(1)}$  – матрица размерностью  $m^{(2)} \times m^{(1)}$ , содержащая параметры модели и обеспечивающая связь между нейронами входного и скрытого слоев автоэнкодера,  $\mathbf{W}^{(2)}$  – матрица размерностью  $m^{(1)} \times m^{(2)}$ , содержащая параметры модели и обеспечивающая связь между нейронами скрытого и выходного слоев автоэнкодера,  $\mathbf{b}^{(1)}$  и  $\mathbf{b}^{(2)}$  – вектора-столбцы размером  $m^{(2)}$  и  $m^{(1)}$ , соответственно, содержащие элементы смещения для нейронов скрытого и выходного слоев.

Необходимо отметить, что исходно веса нейронной сети инициализируются случайным образом, для того чтобы избежать проблемы симметричности весов.

6. Получив начальную гипотезу  $\mathbf{h}_{w,b}((\bar{S}_y)_j)$ , определяем следующую стоимостную функцию, посчитанную по всей выборке векторов обучения в матрице  $\bar{S}_y$ :

$$J(\mathbf{W}, \mathbf{b}) = \left[ \frac{1}{2n} \sum_{j=1}^n \|\mathbf{h}_{w,b}((\bar{S}_y)_j) - (\bar{S}_y)_j\|_2^2 \right] + \frac{\lambda}{2} \left[ \sum_{j=1}^{m^{(2)}} \sum_{k=1}^{m^{(1)}} (W_{jk}^{(1)})^2 + \sum_{j=1}^{m^{(1)}} \sum_{k=1}^{m^{(2)}} (W_{jk}^{(2)})^2 + \sum_{j=1}^{m^{(2)}} (b_j^{(1)})^2 + \sum_{j=1}^{m^{(1)}} (b_j^{(2)})^2 \right], \quad (3)$$

где  $\|\cdot\|_2$  –  $L^2$ -норма, а  $\lambda$  – параметр регуляризации, контролирующей сложность модели.

7. Производим поиск частных производных функции (3) по параметрам  $\mathbf{W}^{(1)}$ ,  $\mathbf{W}^{(2)}$ ,  $\mathbf{b}^{(1)}$ ,  $\mathbf{b}^{(2)}$  с использованием алгоритма обратного распространения ошибки [2].

8. Получив необходимые частные производные

функции (3), осуществляем ее минимизацию по параметрам сети  $\mathbf{W}^{(1)}$ ,  $\mathbf{W}^{(2)}$ ,  $\mathbf{b}^{(1)}$ ,  $\mathbf{b}^{(2)}$  при помощи алгоритма численной оптимизации Бройдена–Флетчера–Гольдфарба–Шанно с ограниченным использованием памяти.

9. После минимизации стоимостной функции получаем параметры обученной сети  $\mathbf{W}^{(1)}$ ,  $\mathbf{W}^{(2)}$ ,  $\mathbf{b}^{(1)}$ ,  $\mathbf{b}^{(2)}$ . Пропускаем через автоэнкодер зашумленные данные из матрицы  $\bar{S}_y$ , выполняя алгоритм прямого распространения (1) и (2) с использованием обновленных параметров сети. В итоге получаем оценку  $\hat{S}_x$  матрицы незашумленных данных  $S_x$ . Далее преобразовываем вектора обучения из столбцов матрицы  $\hat{S}_x$  обратно в блоки размерами  $l \times l$ , вкладывая их в соответствующие пространственные позиции области обучения. Область наложения блоков усредняется арифметически. Получаем восстановленную область обучения, внутри которой выделяем область фильтрации. Обучая автоэнкодер и обрабатывая им все оставшиеся области фильтрации на изображении  $y$ , а затем, вкладывая их в соответствующие пространственные позиции, получим оценку  $\hat{x}$  неискаженного изображения  $x$ . Области фильтрации вкладываются с наложением, поэтому область наложения необходимо арифметически усреднить, чтобы устранить артефакты блочности.

## Результаты моделирования

Рассматриваемый в работе алгоритм реализован в пакете прикладных программ Matlab. Для проведения исследований использовался «классический» набор тестовых полутоновых изображений размером  $256 \times 256$  и  $512 \times 512$  пикселей, предложенных для анализа на интернет-сайте [6]. В качестве численных метрик оценки качества в работе применялись пиковое отношение сигнала к шуму (ПОСШ, дБ) [7] и коэффициент структурного подобия (КСП) [8].

Численные и визуальные результаты восстановления изображений, зашумленных АБГШ, с использованием предложенного алгоритма шумоподавления на основе автоэнкодера приведены на рис. 3 и рис. 4.



а) исходное изображение  
( $\infty$  дБ; 1)



б) зашумленное изображение  
(20,54 дБ; 0,350)



в) восстановленное изображение (27,80 дБ; 0,815)

Рис. 3. Результат восстановления тестового изображения «Фотограф» ( $256 \times 256$ ), искаженного АБГШ с  $\sigma = 25$ , с указанием ПОСШ, дБ и КСП

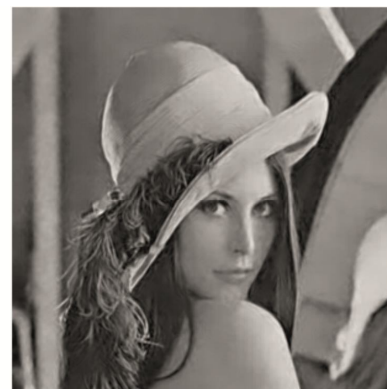




а) исходное изображение (+ дБ; 1)



б) зашумленное изображение  
(20,24 дБ; 0,273)



в) восстановленное изображение  
(31,37 дБ; 0,847)

Рис. 4. Результат восстановления тестового изображения «Лена» (512×512), искаженного АБГШ с  $\sigma = 25$ , с указанием ПОСШ, дБ и КСП



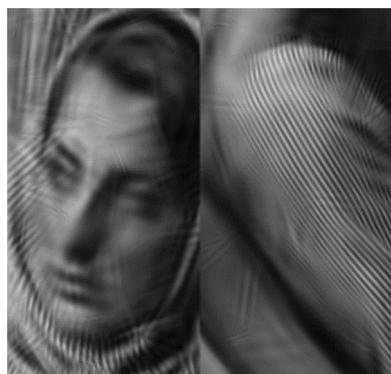
а) исходное изображение ( $\infty$  дБ; 1)



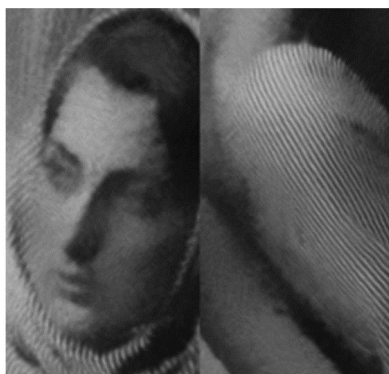
б) зашумленное изображение  
(17,54 дБ; 0,300)



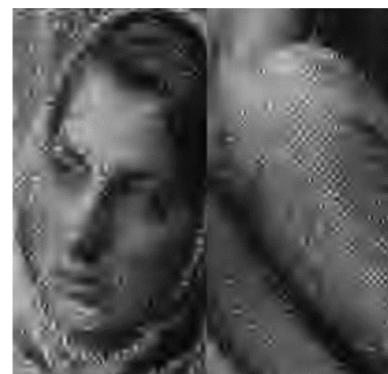
в) предложенный алгоритм  
фильтрации (27,58 дБ; 0,808)



г) курвлет-преобразование [9]  
(27,23 дБ; 0,789)



д) нелокальные средние [10]  
(26,22 дБ; 0,734)



е) вейвлет-преобразование [11, 12]  
(22,91 дБ; 0,608)

Рис. 5. Результаты восстановления фрагментов тестового изображения «Барбара II» (512×512), искаженного АБГШ ( $\sigma = 35$ ), с указанием ПОСШ, дБ и КСП

Таблица 1. ПОСШ восстановленных изображений для  $\sigma = 25$

Тестовое изображение	Предложенный алгоритм	Курвлет-преобразование	Нелокальные средние	Вейвлет-преобразование
«Фотограф»	27,80	27,36	28,05	24,41
«Лена»	31,37	30,72	30,21	27,84
«Лодки»	28,93	28,57	28,06	25,82
«Барбара II»	29,63	28,93	28,40	24,19

Таблица 2. КСП восстановленных изображений для  $\sigma = 25$

Тестовое изображение	Предложенный алгоритм	Курвлет-преобразование	Нелокальные средние	Вейвлет-преобразование
«Фотограф»	0,815	0,767	0,795	0,696
«Лена»	0,847	0,823	0,797	0,753
«Лодки»	0,774	0,757	0,733	0,661
«Барбара II»	0,860	0,837	0,811	0,680

На рис. 5 представлены численные и визуальные результаты восстановления изображений, зашумленных АБГШ, с использованием предложенного алгоритма шумоподавления на основе автоэнкодера и ряда алгоритмов фильтрации, известных из научно-технической литературы [9-12].

Дополнительно в табл. 1 и табл. 2 представлены численные результаты сравнения предложенного алгоритма фильтрации с алгоритмами [9-12] по метрикам ПОСШ и КСП, соответственно. Среднеквадратическое отклонение  $\sigma = 25$ . В качестве тестовых изображений выбраны: «Фотограф» (256×256), «Лена» (512×512), «Лодки» (512×512) и «Барбара II» (512×512)

Анализ визуальных и численных результатов показывает, что для указанных тестовых изображений и степеней зашумления алгоритм фильтрации на основе автоэнкодера показывает наилучшие результаты по сравнению с рассматриваемыми подходами. Отметим, что предложенный алгоритм достаточно качественно восстанавливает высокотекстурированные области (рис. 5в). Однако для изображений, содержащих объекты с высококонтрастными изменениями яркости, при прохождении через их границу, предложенный алгоритм фильтрации приводит к появлению эффекта «ложного оконтуривания» (эффекта «звона») границ изображения (рис. 3в).

### Заключение

В работе рассмотрено применение автоэнкодера в задаче фильтрации цифровых полутоновых изображений. Основным недостатком представленного алгоритма является высокая вычислительная сложность, обусловленная тем, что необходимо проводить обучение автоэнкодера для обработки каждой области фильтрации на зашумленном изображении. Ускорить работу алгоритма можно при помощи распараллеливания кода и его выполнения на графическом процессоре видеокарты. Дополнительно необходимо отметить то, что восстановленные алгоритмом изображения могут содержать артефакты в виде «звона границ». Основным достоинством алгоритма является возможность качест-

венного восстановления высокотекстурированных областей изображения из зашумленных данных.

Работа выполнена при поддержке РФФИ (проект № 15 08-99639-а).

### Литература

1. Katkovnik V., Foi A., Egiazarian K., Astola J. From local kernel to nonlocal multiple-model image denoising // *Int. J. Computer Vision*. 2010. V. 86, № 8. P. 1–32.
2. Marsland S. *Machine learning: an algorithmic perspective*. Chapman and Hall, 2009.
3. Vincent P. Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion // *J. Machine Learning Research*. 2010. V. 11. – P. 3371–3408.
4. Muresan D.D., Parks T.W. Adaptive principal components and image denoising // *Proc. IEEE Int. Conf. Image Processing*. 2003. V. 1. – P. 101–104.
5. Priorov A., Tumanov K., Volokhov V. Efficient denoising algorithms for intelligent recognition systems // In: Favorkaya M., Jain L.C. (eds.) *Computer Vision in Control Systems-2, Intelligent Systems Reference Library*, Springer International Publishing. 2015. V. 75. – P. 251–276.
6. The UGR-DECSAI-CVG image database, <http://decsai.ugr.es/cvg/dbimagenes>.
7. Сэломон Д. *Сжатие данных, изображений и звука*. – М.: Техносфера, 2004.
8. Wang Z., Bovik A.C., Sheikh H.R., Simoncelli E.P. Image quality assessment: from error visibility to structural similarity // *IEEE Trans. Image Processing*. 2004. V. 13, № 4. – P. 600–612.
9. Starck J.-L., Candès E.J., Donoho D.L. The curvelet transform for image denoising // *IEEE Trans. Image Processing*. 2002. V. 11, № 6. P. 670–684.
10. Buades A., Coll B., Morel J.M. Nonlocal image and movie denoising // *Int. J. Computer Vision*. 2008. V. 76, № 2. – P. 123–139.
11. Malla S. *A wavelet tour of signal processing*. Academic Press, 1999.
12. Donoho D.L., Johnstone I.M., Keryacharian G., Picard D. Wavelet Shrinkage: Asymptopia? // *J.R. Statist. Soc. B*. 1995. V. 57, № 2. – P. 301–369.