Цифровая Обработка Сигналов №3/2018

## ЦИФРОВАЯ ОБРАБОТКА СИГНАЛОВ

Научно-технический журнал № 3/2018

Издается с 1999 года Выходит четыре раза в год

**ГЛАВНЫЙ РЕДАКТОР** д.т.н., чл. кор. РАН Ю.Б. ЗУБАРЕВ

# **ЗАМЕСТИТЕЛИ ГЛАВНОГО РЕДАКТОРА:** д.т.н., проф. В.В. ВИТЯЗЕВ,

д.т.н., проф. В.П. ДВОРКОВИЧ

#### РЕДАКЦИОННАЯ КОЛЛЕГИЯ:

д.т.н., проф. Ар.С. Аджемов, д.т.н., проф. Б.А. Алпатов,
д.т.н., проф. В.Г. Бартенев, д.т.н. Ю.И. Борисов,
д.т.н., проф. Ю.А. Брюханов, д.т.н., член-корр. РАН
А.В. Дворкович, д.т.н., профессор В.И. Джиган,
д.т.н., проф. В.В. Еремеев, д.т.н. Г.В. Зайцев,
Р.В. Зубарев, А.П. Кирпичников,
д.т.н., проф. М.И. Кривошеев, д.т.н., акад. РАН Н.А. Кузнецов,

В.Г. Мистюков, д.т.н., проф. С.Л. Мишенков, д.т.н., проф. А.А. Петровский, д.т.н., проф. Ю.Н. Прохоров, д.т.н. А.Л. Приоров, д.т.н., проф. В.Г. Санников, к.т.н., проф. В.С. Сперанский, д.т.н., проф. Ю.С. Шинаков

#### Адрес редакции:

r. Москва, ул. Авиамоторная, д. 8 Научный центр МТУСИ Тел.: (+7) 903-201-53-33 E-mail: rntores@mail.ru vityazev.v.v@rsreu.ru http://www.dspa.ru

#### Издатель:

Российское научно-техническое общество радиотехники, электроники и связи им. А.С. Попова Компьютерная верстка: И.А. Благодарова

Дизайн: М.В. Аверин

Подписной индекс по каталогу ОАО «Роспечать» – 82185

Подписано в печать 11.10.18 г. Формат 60х90/8.

Гарнитура «Arial». Печать офсетная. Бумага офсетная. Печ.л. 6,5. Тираж 500 экз.

Заказ № 5876. Отпечатано в ООО НПЦ «Информационные технологии» Рязань, ул. Островского, д. 21/1 тел.: (4912) 98-69-84

Издание зарегистрировано в Министерстве Российской Федерации по делам печати, телерадиовещания и средств массовых коммуникаций.

Свидетельство о регистрации ПИ № 77-1488 от 14.01.2000 г.

Журнал «Цифровая обработка сигналов» включен в Перечень ведущих рецензируемых научных изданий, в которых по рекомендации Минобрнауки РФ, должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени доктора и кандидата наук.

#### УЧРЕДИТЕЛИ:

#### АО «Инструментальные системы» ФГУП «НИИ радио»

#### B HOMEPE:

Бобровский А.И.	
Адаптация прикладной телевизионной системы	
к динамике сюжета	3
Москвитин А.Э., Ушенкин В.А.	
Комплексирование радиолокационных изображений	
от восходящего и нисходящего витков орбиты	
на основе нечеткой логики1	0
Егошкин Н.А.,Ушенкин В.А.	
Эффективная организация массовых координатных	
преобразований при геометрической обработке	
радиолокационных изображений1	6
Ушенкин В.А.	
Математическая модель синтеза радиолокационных	
изображений дециметрового разрешения	
из радиоголограмм от космических РСА 2	1
Григорьева О.В.	
Субпиксельная идентификация объектов местности	
по гиперспектральным данным	
на основе последовательного квадратичного	
программирования и метода	
анализа формы спектральных компонент	6
Дам Чонг Нам	
Применения комплексного вейвлет-преобразования	
в задаче сжатия видео	2
Лам Чонг Нам	
Оптимизация метода квантования	
для вейвлет-видеокодека	8
Пихачёв А В	
Новый апгоритм повышения контраста	
повый алеориным повышения контраста мелкомасцитабных деталей изображения	
на основе разложения по вейвлетам	4
показапнель качества цафровых попутоновых изображений 4	q
муравьев В.С., Смирнов С.АСтротов В.В.	
Оонаружение и распознавание возоушных объектов	
	6
Ерохин Д.Ю., Ершов М.Д.	
Современные сверточные неиронные сети	
оля оонаружения и распознавания ооъектов	7
Хрящев В.В., Ганин А.Н., Лебедев А.А., Степанова О.А., Кашин С.В., Куваев Р.О.	
Разработка и анализ алгоритма детектирования патологий	
на эндоскопических изображениях желудка	
на основе сверточной нейронной сети70	)
Беляков П.В., Никифоров М.Б.	
Вариационный метод вычисления оптического потока	
в системе-на-кристалле70	5
Золотарёв В.В.	
О реальностях теории кодирования и её приложений	
(ответ на замечания читателей)8	2

Подписной индекс по каталогу ОАО «Роспечать» – **82185** 

### Digital Signal Processing

Science & Technical Magazine Issue 3, 2018 year

Is published quarterly since 1999

#### THE EDITOR-IN-CHIEF:

Dr.Sci. (Tech.), Professor, Corresponding Member of Russian Academy of Sciences U.N. Zubarev

#### **DEPUTY EDITORS-IN-CHIEF:**

Dr.Sci.(Tech.), Professor V.V. Vityazev, Dr.Sci.(Tech.), Professor V.P. Dvorkovich

#### EDITORIAL BOARD:

Adzhemov A.S., Dr.Sci.(Tech.), Professor Alpatov B.A., Dr.Sci.(Tech.), Professor Bartenev V.G., Dr.Sci.(Tech.), Professor Borisov Y.I., Dr.Sci.(Tech.) Bruchanov Y.A., Dr.Sci.(Tech.), Professor Dvorkovich A.V. Dr. Sci (Tech), Professor Corresponding Member of RAS Djigan V.I., Dr. Sci (Tech), Professor Eremeyev V.V., Dr.Sci.(Tech.), Professor Zaitsev G.V., Dr.Sci.(Tech.) Zubarev R.V., Kirpichnikov A.P., Krivocheev M.I., Dr.Sci.(Tech.), Professor Kuznetsov N.A., Dr.Sci.(Tech.), Academician Mistyukov V.G., Mishenkov S.L., Dr.Sci.(Tech.), Professor Petrovsky A.A., Dr.Sci.(Tech.), Professor Priorov A.L., Dr. Sci (Tech) Prokhorov Y.N., Dr.Sci.(Tech.), Professor Sannikov V.G., Dr.Sci.(Tech.), Professor Speranskii V.S., Dr.Sci.(Tech.), Professor Shinakov Y.S., Dr.Sci.(Tech.), Professor

#### Editorial office address:

Aviamotornaya, 8, Moscow, Russia Research Center of MTUCI Phone: (+7) 903-201-53-33 E-mail: rntores@mail.ru, vityazev.v.v@rsreu.ru Web: http://www.dspa.ru

#### Publisher:

Russian A.S. Popov Society for Radioengineering, Electronics & Communications

#### Computer makeup:

I.A. Blagodarova

InSys	
Radio Research and Development Institute	
CONTENTS:	
Bobrovsky A.I.	
Applied television system adaptation to the plots dynamic	3
Moskvitin A.E., Ushenkin V.A.	
Ascending and descending pass SAR image fusion based on fuzzy logic	10
Egoshkin N.A., Ushenkin V.A.	
Effective organization of mass coordinate transformations in the geometric processing of SAR images	16
Ushenkin V.A.	
Decimeter-resolution spaceborne SAR raw data focusing model	21
Grigorieva O.V.	
Subpixel identification of objects by multi- and hyperspectral data on the sequential quadratic programming and the method of analyses spectral components	26
Dam Trong Nam	
Applications of complex wavelet transform in video compression	32
Dam Trong Nam	
Optimization of quantization method for wavelet-based video codec	38
Likhachov A.V.	
A new algorithm for increasing contrast of small image details based on the decomposition on wavelets	44
Svchev A.S. Kholopov I.S.	
No-reference integrated-multiplicative quality index for digital grayscale images	49
Muraviev M.S., Smirnov S.A., Strotov V.V.	
Aerial object detection and recognition based on multispectral image fusion and processing	56
Erokhin D.Y., Ershov M.D.	
Modern convolutional neural networks for object detection and recognition	64
Khryashchev V.V., Ganin A.N., Lebedev A.A., Stepanova O.A., Kashin S.V., Kuvaev R.O.	
Development and analysis of algorithm of pathology detection	
in endoscopic images of gastric based on convolutional neural network	70
Belyakov P.V., Nikiforov M.B.	
System-on-chip variational optical flow computation	76

FOUNDERS:

Subscription index: 82185 ("Rospechat") ISSN: 1684-2634 9 771684 263005

#### УДК 621.397.3

#### АДАПТАЦИЯ ПРИКЛАДНОЙ ТЕЛЕВИЗИОННОЙ СИСТЕМЫ К ДИНАМИКЕ СЮЖЕТА

Бобровский А.И., к.т.н., старший научный сотрудник, начальник научно-исследовательского центра ФГУП «ГосНИИПП», e-mail: albob@mail.ru.

#### APPLIED TELEVISION SYSTEM ADAPTATION TO THE PLOTS DYNAMIC

#### Bobrovsky A.I.

The main methods of adaptation of the image decomposition parameters in the applied television system to the dynamics of the plot and the criteria of optimization of its control system are considered.

Key words: optimization, adaptation, control, information processing.

Ключевые слова: оптимальность, адаптация, управление, обработка информации.

#### Рассмотрены основные методы адаптации параметров разложения изображения в прикладной телевизионной системе к динамике сюжета и критерии оптимизации её системы управления.

#### Введение

Адаптация прикладной телевизионной системы к динамике сюжета ниже рассматривается на примере контроля сближения космических аппаратов (КА) с использованием оптико-электронных (телевизионных) измерительных систем, предназначенных для измерения пространственно-временных координат кооперируемых КА. С. П. Королёв в самом начале космической эры сформулировал «Предложения по созданию средств для орбитальной сборки» [1], где выдвинул проблему обеспечения сближения и стыковки КА на орбитах спутников Земли, и наметил несколько этапов «создания на базе корабля «Восток» полуавтоматического комплекса сборки» (интерактивного управления). Эта методология С. П. Королёва в основном используется до настоящего времени: в процессе сближения и сейчас используется принцип итераций [2], [3].

Для обеспечения контроля сближения КА с помощью телевизионных систем уже более полувека измеряются координаты кооперируемого КА [4]. При этом итерационный принцип и два диапазона дальностей наблюдения сразу привели к применению двух телекамер - с широким и узким углом поля зрения. При этом в силу простоты телевизионных систем считалось, что «Первый этап (автоматический поиск и «захват» спутников радиотехническими средствами) не имеет существенных особенностей» [5]. Рождение твердотельного телевидения открыло новые возможности телевидения, в том числе перестройки параметров телевизионных систем [6] и распространения итерационного принципа С. П. Королёва на изменение параметров разложения изображения в ходе изменения дистанции между КА [7-10]. При разработке теоретических основ и методов адаптации телевизионных систем к динамике сюжета было установлено, что:

на класс измерительных телевизионных систем распространяется принцип совместного кодирования сигналов в телекамере и цифровом кодере [11];

 – система управления параметрами разложения изображения в фотоприёмной матрице так же, как и другие системы управления КА [2], [3], [5], должна обладать инерционностью и гистерезисом [7], [8], [9];

– синтез телевизионной системы (рис. 1) должен опираться на принципы построения проблемно-ориентированных систем управления и принятия решения, а также на методы оптимизации технических объектов и критерии оценки эффективности управления и обработки информации.

Цель указанного синтеза – преодоление противоречия между стремлением разработчика если не к нулевым ошибкам, то к оптимальности системы при ограниченном отношении сигнал/шум и при широком диапазоне априорной неопределённости сигналов. Сложность системы обработки информации связана также с существенными изменениями видимой яркости, размеров и скорости объекта и ограничениями – фотоприёмник имеет не только конечное число  $N_{\rm макс}$  пикселов, но и конечную скорость считывания информации  $F_{\rm cq}$ , т.е. максимальную кадровую частоту  $F_{\rm K}$  при этом числе пикселов:  $F_{\rm K} \times N_{\rm макс} \leq F_{\rm cq}$ .

Взгляд на кодирование видеоинформации в фотоприёмнике телекамеры как на инструмент повышения качества информации [6] здесь дополняется акцентом на специфику цифровой обработки сигналов в цифровом процессоре проблемно-ориентированной системы управления.

# Этапы адаптации и оптимизации телевизионной системы

Адаптация параметров разложения для разных дистанций имеет свои важные особенности. На очень больших дальностях после обнаружения КА и оценки его пространственных координат осуществляется переход:

 от широкоугольной телекамеры к соосной узкоугольной телекамере [4] (широкоугольная продолжает работу для предотвращения срыва слежения);

– от наблюдения при большом времени накопления (низкой кадровой частоте, необходимой для повышения дальности обнаружения) к кадровой частоте *F*<sub>к0</sub>, предельно возможной при полной чёткости *N*<sub>макс</sub> пикселов применённой фотоприёмной матрицы [6], [8] (рис. 2).



Рис. 1. Структурная схема адаптивной телевизионной системы с вычислителем контура проблемно-ориентированной системы управления; кодер канала и декодер понимаются обобщённо, с учётом усилителей; схема отражает не только управление параметрами разложения в зависимости от принятия решений об обнаружении объекта и его динамике, но и взгляд С. П. Королёва [1] на системы контроля сближения КА: «Одновременно с подачей изображения с наружной телекамеры на бортовое видеоконтрольное устройство информация может передаваться по радиоканалу на Землю»



Широкоугольная камера, обнаружение Стабилизация по звёздам, сигнал КА смазан Максимум чувствительности  $F_{\kappa} \leq F_{\kappa 0}$  $N = N_{\text{макс}}$ 



Широкоугольная камера, сопровождение Стабилизация по КА, сигналы звёзд смазаны Минимум ошибки измерения азимута  $F_{\kappa} \leq F_{\kappa 0}$  $N = N_{\text{макс}}$ 





Узкоугольная камера, большая дистанция Наведение на кооперируемый КА

Узкоугольная камера. малая дистанция

Минимум ошибки измерения	
пространственно-временных координат	1
$F_{\kappa} < F_{\kappa 0}$ $F_{\kappa} = 2F_{\kappa}$	0
$N = N_{\text{макс}}$ $N = N_{\text{макс}}$	/4

Рис. 2. Этапы адаптации телевизионной системы

При измерении пространственно-временных координат КА, наблюдаемого узкоугольной телекамерой, имеется несколько этапов, характеризуемых различной динамикой сюжета. Так как при освещении КА Солнцем ошибки измерения этих координат определяются пространственно-временной дискретизацией изображения в фотоприёмной матрице, то параметры разложения должны оптимизироваться по критерию минимума их суммы (при ограничении на скорость считывания сигнала). Эта задача относится к изопериметрическим, в которой в роли переменных выступают дисперсии парциальных ошибок оценивания  $\varepsilon_i = D_{\Delta i}$ , а в роли «периметра» – доступная скорость F<sub>сч</sub> считывания информации с фотоприёмной матрицы.

Решение уравнения Эйлера с неопределёнными множителями Лагранжа приводит к правилу равенства дисперсий приращений сигнала по всем аргументам, т.е. равенства дисперсий межэлементных (по горизонтали и по вертикали) и межкадровой разностей видеосигнала [7], [9], [10]:

$$D\Delta_x = D\Delta_y = D\Delta_t . \tag{1}$$

Такое выравнивание дисперсий ошибок позволяет минимизировать *промах* при измерении пространственно-временных координат КА. Наличие этого условного оптимума параметров разложения интуитивно ясно, так как на большой дальности видимые размеры и скорость объекта малы, и актуальна предельная чёткость, а на малой дистанции размеры и скорость объекта велики, и актуальна высокая кадровая частота. Вместе с тем абсолютно точное выравнивание этих статистик в твердотельных телекамерах невозможно в силу дискретного изменения числа эффективных пикселов, и система управления должна опираться на решающую статистику (*z*-статистику) отношения указанных дисперсий приращений сигналов по разным аргументам.

Система измерения координат, реализующая правило (1) с помощью *взаимообмена* чёткости (числа элементов разложения N) и кадровой частоты  $F_{\kappa}$ , включает подсистему измерения статистики  $z_k$  (равной текущему отношению оценок дисперсий межэлементной и межкадровой разностей  $z_k = D_{\Delta xk}/D_{\Delta tk}$ ). Если считать, что видеосигнал – случайный процесс с нормальным распределением (включающий некоррелированные сигнальную и шумовую компоненты), то статистика z является модифицированной статистикой Фишера-Снедекора [7], [9], которую можно аппроксимировать гауссовским законом распределения [11].

Оптимизация проблемно-ориентированной системы управления адаптивной измерительной телевизионной системой основана на концепции условного максимума информации об объекте в задаче контроля сближения, (минимума динамического промаха), который достигается в несколько этапов на основе группы критериев (перечисляемых по мере нарастания общности для данной системы) несколькими средствами.

1. Минимум ошибок пространственно-временной дискретизации в фотоприёмнике, обеспечиваемый адаптацией кадровой частоты и чёткости (эффективного числа элементов разложения, определяющих интервалы дискретизации Δx по полю и Δt по времени), в условиях ограничения скорости считывания информации с матрицы – условный минимум, достигаемый взаимообменом в соответствии с решением изопериметрической задачи:

$$\min \sum_{i} \varepsilon_{i} \rightarrow \{F_{\kappa}, N\} = \arg \min_{\Delta_{t}, \Delta_{x}} \sum_{i} \varepsilon_{i}$$
при  $F_{\kappa} \times N \leq F_{c''}$ (2)

В силу дискретности растра матричных фотоприёмников изменение чёткости изображения может осуществляться только дискретно, что в совокупности с флуктуациями статистики  $z_k$  из-за шумов видеосигнала вносит свою специфику в синтез системы управления. В общем случае система может иметь множество состояний с различным числом эффективных пикселов (полученных группированием исходных пикселов по 2×2, 3×3, 4×4 и т. д.) с соответствующим правилу (1) увеличением кадровой частоты (рис. 3).



Динамика изменения параметров разложения, качественно намеченная на рис. 2, иллюстрируется фазовым портретом системы (рис. 3), показывающим изменение *состояния* системы по мере сближения КА: состояние 1 – низкая кадровая частота, высокая чёткость; состояние 2 – максимальная при полной чёткости кадровая частота  $F_{\kappa 0}$ ; состояние 3 – повышенная в 2 раза кадровая частота, состояние 4 – повышенная в 3 раза кадровая частота, состояние 5 – повышенная в 4 раза кадровая частота. Пунктир – переход системы в исходное состояние по завершении стыковки КА. В общем случае таких состояний может быть и больше, но ниже рассмотрим наиболее простой случай системы с двумя состояниями 2 и 3.

Экспериментальная проверка обмена пространственной чёткости и кадровой частоты (рис. 4) [7], [10] показала реализуемость и эффективность разработанной концепции.

2. Минимум ошибок линейной фильтрации решающей статистики, обеспечивающий минимум ошибок принятия решений об изменении состояния системы. Линейная фильтрация, т.е. инерционность системы управления, уменьшает шумовую ошибку, но увеличивает задержку. Минимум ошибок фильтрации обеспечивается оптимизацией (адаптацией) инерционности сглаживающего фильтра, импульсная характеристика которого  $h_k$  позволяет сформировать оптимальную (квазиоптимальную) оценку  $\hat{z}_k$  флуктуирующей в силу наличия шумов в видеосигнале *z*-статистики путём свёртки наблюдения  $z_k$  с импульсной характеристикой  $h_k$  фильтра:  $\hat{z}_k = z_k \otimes h_k$ :

$$\min \sum_{i} \varepsilon_{i} \to h_{k} = \arg \min \sum_{k} (\varphi_{k} - \hat{z}_{k})^{2} =$$
  
= 
$$\arg \min \sum_{k} (\varphi_{k} - z_{k} \otimes h_{k})^{2}.$$
 (3)

Функцию  $\phi(t)$  можно моделировать кусочнополиномиальной зависимостью (в реальных условиях стыковки КА работают корректирующие двигатели, вносящие аддитивные скачки в закон изменения  $\phi(t)$ ). Учитывая в среднем монотонный вид функции  $\phi(t)$ , можно получать её оценку с использованием оптимального фильтра Калмана, вводящего переменную во времени инерционность в процесс оценивания. Упрощённый вариант такой фильтрации может быть реализован с по-





межкадровая разность, бинирование и удвоение кадровой частоте  $D_{\Delta t}$  = 251

Рис. 4 Статистики изображения приближающегося КА

мощью нелинейного накопления, т.е. с помощью принятия решения после превышения статистикой z уменьшенного порога подряд в нескольких кадрах. В этом простом случае существует оптимальное число *т* совместно анализируемых последовательных кадров, зависящее от текущего значения производной математического ожидания  $\phi$  статистики z по времени  $d\phi(t)/dt$  [6]. Промежуточный по точности и сложности метод субоптимальной фильтрации использует рекурсию в соответствии с разностным уравнением, включающим постоянную обратной связи α, определяющую инерционность фильтра и пороговое устройство обнаружения аномальных выбросов фильтруемого процесса, соответствующих импульсной коррекции положения наблюдаемого кооперируемого КА, по сигналам которого память цифрового фильтра очищается [13]:

$$\hat{z}_{k} = \alpha z_{k} + (1 - \alpha) \hat{z}_{k-1}$$
 (4)

Результаты моделирования системы управления адаптивной телекамерой с использованием критерия [12] показали его практическую эффективность.

3. Максимум времени наблюдения при нахождении системы в потенциально минимальной зоне нечувствительности подсистемы управления, необходимой для устойчивости системы, обеспечиваемый оптимизацией порогов  $\gamma_{\rm B}$  и  $\gamma_{\rm H}$  переключения её состояния. Этот принцип является следствием и обобщением необходимости достижения ряда экстремумов при синтезе рассматриваемой системы управления. С учётом частных критериев (2) и (3) можно поставить целью синтеза адаптивной системы определение совокупности её параметров:

$$\left\{F_{\kappa}, N, \gamma_{\mu}, \gamma_{\mu}, h_{k}\right\} = \arg\min_{\Delta_{i}, \Delta_{x}} \min_{\gamma_{\mu}, \gamma_{\mu}} \min_{h_{k/i}} \int_{T} \sum_{i} \varepsilon_{i}.$$
(5)

Анализ функционирования системы управления при оценке её эффективности по критерию (4) показал, что методологически более общим является критерий максимума времени (maxT) наблюдения объекта при нахождении системы в зоне приемлемых значений решающей статистики  $\hat{z}_{\mu}$ . Критерий maxT (имеющий очень широкое применение [13]) формализует метод обработки информации, при котором система управления путём выбора совокупности оптимальных параметров - кадровой частоты F<sub>к</sub>, чёткости N (числа пикселов), порогов гистерезиса  $\gamma_{\scriptscriptstyle \rm B}$  и  $\gamma_{\scriptscriptstyle \rm H}$ , импульсной характеристики  $h_k$  фильтра формирования оценки  $\hat{z}_k$  решающей статистики – обеспечивает максимум времени T, когда оценка  $\hat{z}_k$  решающей статистики zk принадлежит потенциально минимальной зоне нечувствительности системы управления, т.е.  $\hat{z}_{k} \in [1/2, 2]$ . При общем времени наблюдения  $T_{\rm H}$ , кратном времени кадра  $T_{\rm K}$ ,  $T_{\rm H} = KT_{\rm K}$ , K >> 1, пренебрегая переходными процессами при включении системы, критерий max *T* принимает вид:

$$\max T \to \{F_{\kappa}, N, \gamma_{\scriptscriptstyle B}, \gamma_{\scriptscriptstyle H}, h_k\} =$$
  
=  $\arg \max \sum_{k=1}^{K} \operatorname{sign} \{1 - |\log_2 \hat{z}_k|\}.$  (6)

#### Результаты моделирования

Результаты моделирования системы управления адаптивной телекамерой и определения оптимальных порогов гистерезиса системы управления с использованием этого критерия [14] показали его практическую эффективность для простейшего случая, когда адаптивная система наблюдения должна иметь два состояния  $A_{k-1}^1$  с максимальной чёткостью и  $A_{k-1}^2$  с максимальной кадровой частотой. В этом случае гистерезис определяется верхним  $\gamma_{\scriptscriptstyle B}$  и нижним  $\gamma_{\scriptscriptstyle H}$  порогами  $(\gamma_{\rm B} = 1/\gamma_{\rm H})$ . Минимальное отношение порогов  $d = \gamma_{\rm H}/\gamma_{\rm H}$ соответствует изменению отношений интервалов пространственной и временной дискретизации, что в простейшем случае системы с двумя состояниями даёт *d* = 4. Вместе с тем из-за неизбежного наличия в видеосигнале шумов для обеспечения устойчивости адаптивной системы необходимо увеличение «расстояния» d между порогами.

Гистерезис системы проявляется в том, что в каждом k-м кадре она может иметь два текущих *состояния*: – при  $\hat{z}_k > \gamma_e$  переход из состояния  $A_{k-1}^2$  с макси-

мальной кадровой частотой в состояние  $A_k^1$  с максимальной чёткостью;

– при  $\hat{z}_k < \gamma_{\mu}$  переход из состояния  $A_{k-1}^1$  с максимальной чёткостью в состояние  $A_k^2$  с максимальной кадровой частотой;

- при  $\gamma_{\rm H} \le z \le \gamma_{\rm B}$  — сохранение имеющегося состояния  $A_k = A_{k-1}$ .

Такая динамическая система функционирует в соответствии с *уравнением состояния*, реализуемым с помощью *JK* триггера, функционирующего в соответствии с таблицей истинности [12]

K	J	$Q_k$	$Q_{k+1}$	Функция F	
0	0	0	0	Y	
0	0	1	1	хранение информации	
0	1	0	1	Установка 1,	
0	1	1	1	переход в состояние $A^1$	
1	0	0	0	Установка 0,	
1	1	1	0	переход в состояние $A^2$	

 $Q_{k+1} = F(Y_{Jk+1}, Y_{Kk+1})$ 

при соответствующем *уравнении управления* его входами *J* и *K* (рис. 5):

(7)

$$Y_{k+1} = \begin{cases} Y_{Jk+1} = \operatorname{sign}(\hat{z}_k - \gamma_{\scriptscriptstyle B}) \\ Y_{Kk+1} = \operatorname{sign}(\gamma_{\scriptscriptstyle H} - \hat{z}_k) \end{cases}.$$
(8)

Извлечение из наблюдений максимума информации о пространственно-временных координатах объекта в приложении к системе, адаптирующей параметры при синтезе системы в целом и в каждой частной задаче должна быть формализована опорная триада синтеза – априорная информация, критерий качества и ограничения. Например, в частной задаче оптимизации порогов (гистерезиса) в априорную информацию входит результат решения другой частной задачи оптимизации параметров разложения (изопериметрическая задача привела к правилу равенства дисперсии приращений сигнала по всем аргументам) [6], [8-10]. Решение задачи оптимизации гистерезиса возможно, например, при следующей опорной триаде статистического синтеза:





– изменяющееся во времени математическое ожидание  $\varphi(t)$  *z*-статистики имеет распределение в интервале [a, 1/a], a > 2 вида p(z) = 1/2z, обеспечивающее равенство вероятностей значений *z*-статистики, больших и меньших 1;

- при фиксированном значении математического ожидания  $\phi_k = \phi$  *z*-статистики при гауссовском распределении видеосигнала функция плотности вероятности описывается распределением Фишера-Снедекора [6] и при ряде условий может аппроксимироваться гауссовским законом распределения [10]. Эксперимент [12] показал, что в широком диапазоне изменения параметров сюжетов функция плотности вероятности статистики может аппроксимироваться гауссовским законом распределения. Для такой аппроксимации справедлива связь математического ожидания ф и дисперсии D<sub>z</sub>:  $D_z = 2\phi^2/N$ . При сравнительно небольших порогах, обеспечивающих малую вероятность ложной тревоги, и достаточно больших значениях числа пикселов N<sup>2</sup> можно использовать формулу, связывающую значения математического ожидания ф z-статистики и используемого порога у:

$$\gamma - \varphi \approx \frac{\varphi}{p_{\rm nr}} \sqrt{\pi N}; \tag{9}$$

система для обеспечения устойчивости обладает гистерезисом, т.е. имеет не один, а два порога изменения состояния: пороги принятия решения о переходе из состояния  $A_{k-1}^2$  с максимальной кадровой частотой в состояние  $A_k^1$  с максимальной чёткостью и обратно симметричны относительно единичного значения:  $\gamma_{\rm B} = 1/\gamma_{\rm H}$  (далее для упрощения записей упоминаем только один порог  $\gamma_{\rm B}$ , опуская нижний индекс); требование устойчивости из-за изменения при переключении состояния вдвое и чёткости, и кадровой частоты, приводит к неравенству  $\gamma > 2$  [6], [8-10].

Критерий качества (охарактеризованы выше):

 критерий минимума за всё время наблюдения интеграла суммы ошибок по всем аргументам;  – критерий максимума времени *T* нахождения *z*-статистики в интервале [2, ½];

#### Ограничения:

– число N пикселов фотоприёмной матрицы и скорость  $F_{\rm cч}$  считывания информации с неё ограничены;

– доступны два состояния системы в k-м кадре работы:
 – состояние A<sup>1</sup><sub>k</sub> с N пикселами и кадровой частотой

 $F_{\kappa} = F_{cu}/N;$ – состояние  $A_k^2$  с увеличенной вдвое кадровой частотой  $2F_{\kappa}$  при сокращённом до *N*/4 (вдвое по горизонтали и по вертикали) числом пикселов;

#### Решение задачи.

Оптимальное значение порога  $\gamma_0$  для достижения max T при переключении состояния системы определяется двумя противоположными тенденциями (рис. 6):

– при  $\gamma \rightarrow 2$  из-за флуктуаций *z*-статистики возрастает вероятность  $p_{\rm лrr}$  ложного переключения состояния, которая определяет первую компоненту  $T_1$  аномального времени вне интервала [2, ½], в первом приближении вычисляемую как

$$T_1 = 2\int_{1}^{\gamma} d\varphi \int_{\gamma}^{\infty} p(z \mid \varphi = 2) dz = \frac{\ln \gamma}{\ln a} \int_{\gamma}^{\infty} p(z \mid \varphi = 2) dz ; \qquad (10)$$

– при γ >> 2 вероятностью p<sub>1</sub> можно пренебречь, но время нахождения системы в интервале [2, ½] сокращается на величину T<sub>2</sub>, пропорциональную превышению порога γ над его предельным значением 2.

Для принятой априорной плотности вероятности *z*-статистики доля времени нахождения системы в заданном интервале [2, ½] при отсутствии адаптации составляет  $T/T_{\rm H} = \ln 2/\ln T_{\rm H}$ , при включении управления при  $\gamma = 2$  и  $p_1 = 0$  предельное значение max T составит  $T/T_{\rm H} = \ln 2/(\ln T_{\rm H} - \ln 2)$ , а при  $\gamma > 2$ 

$$\frac{T}{T_{\rm H}} = \frac{3\ln 2 - \ln \gamma}{2(\ln a - \ln 2)} \,. \tag{11}$$

С учётом (1) и (2) формула для критерия  $\max T$  принимает вид



Рис. 6. Пример изменения во времени *z*-статистики при функции плотности вероятности вида p(z) = 1/2z, [a, 1/a], a > 2; на первом «периоде» штрих-пунктиром показано изменение *z*-статистики при отсутствии управления, на втором «периоде» пунктиром показаны флуктуации времени переключения состояния системы из-за флуктуаций *z*-статистики («ложных тревог»)

$$\frac{T}{T_{\rm H}} = \frac{3\ln 2 - \ln \gamma \left(1 + \int_{\gamma}^{\infty} p(z \mid \varphi = 2) dz\right)}{2(\ln a - \ln 2)}$$
(12)

Оптимальное значение порога  $\gamma_{\rm o} = \arg \max \frac{T}{T_{\rm H}}$  суще-

ственно зависит от функции плотности вероятности *p*(*z*|φ=2), которая, в свою очередь, зависит от объёма *N* выборки, т.е. числа пикселов, относящихся к сопровождаемому объекту [9], [12], рис. 7.



Рис. 7. К определению оптимального значения порога по критерию тах T

Численное значение порога (параметров гистерезиса системы управления), полученные при использовании критерия  $\max T$  оказались несколько менее (ближе к предельному значению 2), чем «пессимистические» оценки при заданной вероятности ложного переключения  $10^{-3}$ , принятой при предварительном анализе динамики функционирования адаптивной телевизионной системы [8].

Исследования статистических свойств z-статистики показали зависимость её дисперсии (среднеквадратического отклонения) как от размеров выборки (видимых размеров сопровождаемого объекта, т.е. дальности него), так и от среднего значения сигнала объекта (его освещённости), дав хорошее согласие с использованной аппроксимацией функции распределения решающей статистики гауссовским законом с дисперсией, связанной с математическим ожиданием и количеством пикселов наблюдаемого объекта.

Это означает, что в телевизионной системе, адаптивной к динамике сюжета, перестраиваются не только параметры разложения изображения в фотоприёмнике, но параметры обработки информации в цепи управления состоянием системы.

#### Заключение

Адаптация телевизионной системы к динамике сюжета позволяет найти решение задачи минимизации ошибок измерения изменяющихся во времени координат объектов при ограничении скорости считывания информации с фотоприёмных матриц. В решение задачи входят условные оптимальные чёткость и кадровая частота при их дискретно переключаемом взаимообмене, осуществляемом на основе тенденции к равенству дисперсий межэлементных и межкадровых приращений. Обеспечения устойчивости системы автоматического управления параметрами разложения измерительной системы требует введения доверительного интервала при измерении статистик сигнала, превышающего минимально допустимое значение 4, определяемое дискретностью растра.

Изложенная концепция, опирающаяся на правило (1), критерии (2), (3), (5), (6) и уравнение управления (8), является отражением новой парадигмы в теории прикладных телевизионных систем, пришедшей на смену старой парадигме пассивного учёта деградации разрешающей способности при движении наблюдаемого объекта. Оптимизация рассмотренной системы космического телевидения имеет целью извлечение информации максимального качества, без оглядки на ограниченность зрения человека-наблюдателя, с учётом ограничения пропускной способности каналов связи телекамеры и бортового компьютера и/или ЦУП, отражает важный эффект влияния твердотельной технологии на изменение методов системного анализа и синтеза, оптимизации, управления, принятия решений и обработки информации в прикладном телевидении.

#### Литература

1. Творческое наследие академика Сергея Павловича Королёва. Избранные труды и документы / Под ред. М. В. Келдыша. М., Наука, 1980. – 592 с.

2. Легостаев В. П., Раушенбах Б. В. Автоматическая сборка в космосе/ Космические исследования, 1969, вып. 6. С. 803-813.

3. Микрин Е. А. Бортовые комплексы управления космических аппаратов. М., МГТУ им. Н. Э. Баумана, 2014. – 245 с.

4. Брацлавец П. Ф., Росселевич И. А., Хромов Л. И. Космическое телевидение. М.: Связь, 1973. – 248 с. Легостаев В. П., Шмыглевский И. П. Управление сближением космических аппаратов на этапе причаливания. Управление в космосе. Т. 2. М., Наука, 1972. С. 218-228.

5. Твердотельная революция в телевидении: Телевизионные системы на основе приборов с зарядовой связью, систем на кристалле и видеосистем на кристалле / Под ред. А. А. Умбиталиева и А.К. Цыцулина. – М.: Радио и связь, 2006. – 312 с.

6. Теория и практика космического телевидения / Умбиталиев А. А., Пятков В. В., Бобровский А. И. и др. // Под ред. Умбиталиева А. А., Цыцулина А. К. СПб, НИИ телевидения, 2017. – 368 с.

7. Обнаружение объектов на звёздном фоне / Левко Г. В., Бобровский А. И., Морозов А. В., Цыцулин А. К. // Вопросы радиоэлектроники, серия Техника телевидения, 2016, вып. 2. С. 29-38.

8. Статистический синтез управления телевизионной системой, адаптивной к динамике сюжета / Умбиталиев А. А., Пятков В. В., Морозов А. В. и др. // Вопросы радиоэлектроники, серия Техника телевидения, 2016, вып. 1. С. 3-11.

9. Адаптация параметров разложения телевизионного координатора целей / Умбиталиев А. А., Пятков В. В., Бобровский А. И. и др.// Военно-научная конференция «Актуальные научно-технические аспекты разработки, испытаний и эксплуатации средств ракетно-космической обороны», СПб, ВКА им. А. Ф. Можайского, 20 октября 2017 г. С. 183-188.

10. Вероятностные характеристики решающей статистики в телевизионной системе, адаптивной к динамике сюжета/ Рогачёв В. А., Морозов А. В., Бобровский А. И. и др. Вопросы радиоэлектроники, серия Техника телевидения, 2018, вып. 1. – С. 71-78.

11. Хромов Л. И., Цыцулин А. К., Куликов А. Н. Видеоинформатика. М., Радио и связь, 1991. – 192 с.

12. www.digiteh.ru/digital/JK\_trigg.php

13. Рекурсивная фильтрация опорной статистики адаптивной телевизионной системы / Морозов А. В., Чепелев А. Г., Бобровский А. И. и др. // Труды 14-й Междунар. конф. «Телевидение: передача и обработка изображений», СПб, 26-27 июня 2018 г. СПб: Изд-во «Технолит», 2018. С. 199-203.

14. Шамис А. Л. Вектор эволюции. Жизнь, эволюция. Мышление с точки зрения программиста. М.: Книжный дом «ЛИБРОКОМ», 2013. – 200 с.

15. Оптимизация гистерезиса системы управления телекамерой, адаптивной к динамике сюжета/ Бобровский А. И., Рогачёв В. А., Морозов А. В. И др. // Труды 14-й Междунар. конф. «Телевидение: передача и обработка изображений», СПб, 26-27 июня 2018 г. СПб: Издво «Технолит», 2018. С. 40-44.



Цифровые процессоры обработки сигналов



## НОВЫЕ КНИГИ

# Витязев С.В. Цифровые процессоры обработки сигналов / Курс лекций – М.: Изд-во «Горячая линия-Телеком», 2017 г. – 100 с.: ил.

Рассмотрены основы построения архитектур и оптимизации программного обеспечения цифровых сигнальных процессоров. Сформулированы основные задачи цифровой обработки сигналов на сигнальных процессорах. Представлено описание инструментальных и программных средств работы с цифровыми сигнальными процессорами.

Для студентов технических вузов радиотехнических и инфокоммуникационных специальностей, будет полезна преподавателям, читающим соответствующие курсы.

#### УДК 004.932

#### КОМПЛЕКСИРОВАНИЕ РАДИОЛОКАЦИОННЫХ ИЗОБРАЖЕНИЙ ОТ ВОСХОДЯЩЕГО И НИСХОДЯЩЕГО ВИТКОВ ОРБИТЫ НА ОСНОВЕ НЕЧЕТКОЙ ЛОГИКИ

Москвитин А.Э., в.н.с. НИИ «Фотон» РГРТУ, к.т.н., foton@rsreu.ru; Ушенкин В.А., с.н.с. НИИ «Фотон» РГРТУ, к.т.н., foton@rsreu.ru.

#### ASCENDING AND DESCENDING PASS SAR IMAGE FUSION BASED ON FUZZY LOGIC

#### Moskvitin A.E., Ushenkin V.A.

The task of ascending and descending pass SAR image fusion is considered. These images correspond to the observation of the scene from different sides. The goal of this fusion is to reduce amount of geometric distortions and shading. An algorithm of fusion, based on fuzzy logic and strict computation of layover and shadow masks, is proposed.

Key words: SAR image, ascending pass, descending pass, orbit, fuzzy logic, fusion.

Ключевые слова: радиолокационное изображение, восходящий виток, нисходящий виток, орбита, нечеткая логика, комплексирование.

#### Введение

При радиолокационном наблюдении Земли из космоса неизбежен боковой обзор, приводящий к геометрическим искажениям и затенению объектов на изображениях. К геометрическим

искажениям относятся разрывы и переналожения. Под разрывом понимается ситуация, когда узкая линия на исходном радиолокационном изображении (РЛИ) в процессе ортотрансформирования растягивается и становится широкой областью с крайне низким пространственным разрешением по одному из измерений. Под переналожением понимается ситуация, когда отклики на высокие объекты, имеющие большую наземную дальность, накладываются на низкие объекты, имеющие меньшую наземную дальность. Разрывы и переналожения характерны для восходящих склонов рельефа (высота которых увеличивается с дальностью), а затенения – для нисходящих.

При съемке на восходящем и нисходящем витках орбиты космический аппарат, как правило, оказывается с противоположных сторон (слева и справа) от наблюдаемого участка земной поверхности. Следовательно, восходящие склоны рельефа на одном РЛИ становятся нисходящими на другом и наоборот. Из-за этого в горной местности на изображениях наблюдается инверсия радиояркости, а также разное расположение разрывов, переналожений и областей радиотени (рис. 1), что позволяет проблемные фрагменты одного РЛИ заполнить данными другого РЛИ, на котором эти проблемы не наблюдаются. Однако, довольно часто встречаются и ситуации, когда затенение на одном РЛИ соответствует разрыву на другом. В этом случае на обоих РЛИ не содержится сигнала, адекватно описывающего наблюдаемую сцену. Тем не менее, таких участков на результате комплексирования оказывается гораздо меньше, чем на исходных РЛИ. Это позволяет представить на нем

Рассматривается задача объединения радиолокационных изображений, полученных на восходящем и нисходящем витках орбиты космического аппарата и соответствующих наблюдению сцены с разных сторон. Целью объединения является уменьшение количества геометрических искажений и затенений. Предлагается алгоритм комплексирования таких изображений, основанный на нечеткой логике и строгом расчете масок разрывов, переналожений и затенений.

> наблюдаемую сцену достаточно полно. Таким образом, основным назначением комплексирования РЛИ, полученных на восходящем и нисходящем витках, является формирование такого РЛИ, на котором наблюдается минимум теней, разрывов и переналожений, обусловленных трехмерной структурой наблюдаемой сцены.

#### Анализ известных алгоритмов комплексирования

Исходными данными для комплексирования выступают два ортотрансформированных РЛИ одной местности, снятых на восходящем и нисходящем витках орбиты КА, и трехмерная модель наблюдаемой сцены (цифровая модель рельефа или местности, ЦМР или ЦММ), с помощью которой были ортотрансформированы оба изображения. Ортотрансформированные РЛИ должны сопровождаться информацией об исходной геометрии съемки, на основе которой восстанавливается обратный закон координатного соответствия, использовавшийся при ортотрансформировании.

С помощью указанного обратного закона и опорной трехмерной модели для каждого из РЛИ строится маска разрывов (включая переналожения) и затенений. Одно из РЛИ выбирается базовым, а другое – дополнительным. Обратим внимание, что результат комплексирования зависит от этого выбора, поскольку в него помещаются, в первую очередь, данные из основного РЛИ, а данные дополнительного РЛИ используется лишь в участках, где основные данные не обладают требуемым качеством. В [1] рекомендуют в качестве основного выбирать РЛИ, снятое с большим отклонением на надира, поскольку на нем содержится меньше геометрических искажений (разрывов и переналожений). Дополнительное РЛИ и соот-



Рис. 1. Ортотрансформированные РЛИ одной местности, полученные на восходящем и нисходящем витках орбиты

ветствующие ему маски трансформируются в геометрию основного РЛИ, а результат комплексирования формируется исходя из следующих правил [1]:

 если пиксель основного РЛИ не относится к областям разрыва и затенения, то он переносится на результат комплексирования;

 если пиксель основного РЛИ попадает в область разрыва или тени, то анализируется дополнительное РЛИ:

 – если на дополнительном РЛИ наблюдается разрыв, то результат комплексирования формируется из основного РЛИ;

 – если на дополнительном РЛИ наблюдается тень, то результат комплексирования может быть сформирован из любого РЛИ (с предпочтением тени перед разрывом);

 если на дополнительном РЛИ нет ни тени, ни разрыва, то результат комплексирования формируется из дополнительного РЛИ.

В работе [2] предлагается модификация правил комплексирования, делающая его инвариантным к выбору основного и дополнительного РЛИ: когда на обоих РЛИ не наблюдается ни разрывов, ни теней, выполняется их дополнительное объединение. Для этого оценивается степень заметности точечного объекта на каждом из РЛИ. Если на одном РЛИ пиксель оказывается значительно более заметным, чем на другом, то он помещается в результат комплексирования. В противном случае сигналы изображений усредняются. Однако указанный подход справедлив лишь тогда, когда детальность опорной трехмерной модели существенно ниже, чем у комплексируемых РЛИ. Иначе практически все наиболее заметные пиксели изображений, обычно обусловленные уголковыми отражателями на местности, будут соответствовать разрывам, а попытка сохранить наиболее заметные пиксели будет противоречить первоначальной цели комплексирования – устранению разрывов и теней на РЛИ. В настоящей работе мы ограничимся исходной целью комплексирования и не будем пытаться объединять фрагменты РЛИ, не попавшие в области теней и разрывов. Тем не менее, в дальнейшем будет показано, как это можно сделать, если возникнет такая необходимость.

Рассмотрим задачу построения масок разрывов и затенений. В работе [1] предлагается примитивный подход к такому построению. Для каждого пикселя ортотрансформированного РЛИ вычисляется вектор его визирования и наклон рельефа в направлении дальности. Если склон рельефа восходящий (высота возрастает с увеличением дальности), то разрыв (переналожение) образуется, когда сумма углов наклона рельефа и угла визирования оказывается равна (больше) 90°. Если склон нисходящий, то тень образуется, когда угол наклона рельефа больше угла визирования. При этом не рассматривается важный вопрос, как корректно оценить локальный угол наклона земной поверхности в условиях низкой детальности и возможного наличия высокочастотного шума на опорной трехмерной модели. Кроме того, не учитываются два аспекта. Во-первых, разрыв в виде области на ортотрансформированном РЛИ, сильно размытой в направлении дальности, может образоваться, не только когда направление визирования строго перпендикулярно локальному участку земной поверхности, но и когда угол между ними близок к 90°. Во-вторых, область тени может быть на участке с любым наклоном рельефа, если он загорожен высотным объектом. Низкое качество построенных масок в [1] пытаются компенсировать их последующей обработкой с помощью морфологических фильтров. В настоящей работе предлагается более корректно оценивать затенение по методу трассировки лучей, а наличие разрывов – по степени расширения/сжатия пикселей в направлении дальности в процессе ортотрансформирования.

Другой проблемой комплексирования является наличие «швов» – резких границ между фрагментами, заполненными данными из разных исходных РЛИ, что иллюстрирует рис. 2. Для решения указанной проблемы в настоящей работе предлагается изменить правила комплексирования, построив их на базе нечеткой логики.



Рис. 2. «Швы» на результате комплексирования РЛИ, полученных на восходящем и нисходящем витках орбиты

Таким образом, цель настоящей работы составляет разработка алгоритма комплексирования радиолокационных изображений от восходящего и нисходящего витков орбиты, основанного на нечеткой логике и строгом расчете масок затенений, разрывов и переналожений и позволяющего повысить качество комплексирования по сравнению с известными алгоритмами.

#### Построение маски разрывов и переналожений

Для получения маски разрывов построим карту растяжения пикселей исходного РЛИ по отношению к ортотрансформированному. Карту будем получать в картографической проекции, а шаг дискретизации выберем в соответствии с пространственным разрешением опорной трехмерной модели.

В каждой точке с картографическими координатами  $(N_{\nu}, E_{\nu})$  степень растяжения определяется следующим образом. Координаты  $(N_{\kappa}, E_{\kappa}, h(N_{\kappa}, E_{\kappa}))$  по обратному закону (с использованием аппроксимирующей сетки) пересчитываются в исходную геометрию радиолокационного изображения «дальность – азимут»  $(R_{\perp}, \eta_{\perp})$ . По прямому закону (тоже с использованием аппроксимирующей сетки) координаты  $(R_{\perp} + \Delta R, \eta_{\perp}, h(N_{\kappa}, E_{\kappa})),$  где  $\Delta R$  – шаг дискретизации дальности, пересчитываются в  $(\overline{N}_{\kappa}, \overline{E}_{\kappa})$ . Расстояние  $d_0$  между  $(\overline{N}_{\kappa}, \overline{E}_{\kappa})$  и  $(N_{\kappa}, E_{\kappa})$ соответствует исходному размеру пикселя РЛИ без растяжения и сжатия, а вектор  $(\overline{N}_{\kappa} - N_{\nu}, \overline{E}_{\kappa} - E_{\nu})$  показывает локальное направление дальности в заданной картографической проекции. Отступим от точки (N<sub>к</sub>, E<sub>к</sub>) на величину, равную разрешению трехмерной модели *d*<sub>DEM</sub>, в двух противоположных направлениях, задаваемых вектором  $(\overline{N}_{\kappa} - N_{\kappa}, \overline{E}_{\kappa} - E_{\kappa})$ . В результате получим две точки  $(N_{\kappa 1}, E_{\kappa 1})$  и  $(N_{\kappa 2}, E_{\kappa 2})$ . Переведем  $(N_{\kappa 1}, E_{\kappa 1})$  $h(N_{\kappa 1}, E_{\kappa 1}))$  и  $(N_{\kappa 2}, E_{\kappa 2}, h(N_{\kappa 2}, E_{\kappa 2}))$  по обратному закону в координаты  $(R_{\perp 1}, \eta_{\perp 1})$  и  $(R_{\perp 2}, \eta_{\perp 2})$ . Степень растяжения пикселя определяется отношением:

$$k_d(N_{\kappa}, E_{\kappa}) = \frac{d_0 |R_{\perp 2} - R_{\perp 1}|}{2d_{DEM}}.$$
 (1)

Формула (1) не учитывает случай, когда ортотрансформирование выполняется с понижением пространственного разрешения относительно исходного РЛИ. В этом случае небольшие сжатия пикселей становятся несущественными, что учитывает следующая формула:

$$k_d(N_{\kappa}, E_{\kappa}) = \frac{\max(d_0, d_{opmo}) |R_{\perp 2} - R_{\perp 1}|}{2d_{DEM}},$$
(2)

где *d<sub>opmo</sub>* – шаг пикселя ортотрансформированного изображения.

Значения  $k_d(N_{\kappa}, E_{\kappa})$ , существенно меньшие единицы (сильное сжатие), соответствуют разрывам, а отрицательные значения  $k_d(N_{\kappa}, E_{\kappa})$  – переналожениям.

Отметим, что в оценке  $k_d(N_{\kappa}, E_{\kappa})$  косвенным образом учитывается наклон рельефа в виде сопоставления двух высот  $h(N_{\kappa 1}, E_{\kappa 1})$  и  $h(N_{\kappa 2}, E_{\kappa 2})$ . Поскольку рас-

стояние между точками  $(N_{\kappa 1}, E_{\kappa 1})$  и  $(N_{\kappa 2}, E_{\kappa 2})$  составляет удвоенное пространственное разрешение опорной трехмерной модели  $d_{DEM}$ , оценка наклона содержит минимум ошибок, вызванных интерполяцией отсчетов опорной высоты. Если на опорной трехмерной модели содержится высокочастотный шум, он может негативно сказаться на качестве построенной маски разрывов. В этом случае требуется предварительное сглаживание опорной трехмерной модели.

Для построения маски разрывов и переналожений в разрешении ортотрансформированного изображения выполняется интерполяция полученной карты растяжения пикселей с помощью фильтра Ланцоша, обеспечивающего гладкость результата интерполяции. Четкая маска разрывов и переналожений может быть получена из интерполированной карты растяжения пикселей путем пороговой обработки. К разрывам (включая переналожения) относятся все пиксели  $(N_{\kappa}, E_{\kappa})$ , для которых  $k_d(N_{\kappa}, E_{\kappa})$  оказывается меньше выбранного порога.

Чтобы получить нечеткую маску, необходимо задать два порога, один из которых соответствует полному отсутствию разрыва и переналожения, другой – ярко выраженному разрыву или переналожению. Между выбранными порогами значение нечеткой маски будет линейно изменяться от нуля до единицы:

$$\mu(L) = \begin{cases} 1, & k_d \le 0.5, \\ 1 - 4(k_d - 0.5), & 0.5 < k_d < 0.75, \\ 0, & k_d \ge 0.75. \end{cases}$$
(3)

#### Построение маски затенений

Для получения маски затенений сформируем карту возвышений объектов наблюдаемой сцены над границей тени, полученной по методу трассировки лучей. Геометрический смысл такого возвышения иллюстрирует рис. 3. Карта возвышений над границей тени формируется в картографической проекции с шагом дискретизации, соответствующим детальности опорной трехмерной модели, как и рассмотренная выше карта растяжения пикселей.



Рис. 3. Геометрический смысл возвышения над границей тени

Трассировку луча удобнее всего выполнять в геометрии исходного РЛИ, поскольку в этом случае луч направлен строго вдоль строки изображения. Однако здесь возникает проблема, связанная с трансформированием опорной трехмерной модели в данную геометрию, которое сопровождается для каждого пикселя сложными итерационными вычислениями и требует много времени. В то же время трансформирование трехмерной модели в картографическую проекцию осу-

ществляется гораздо проще и быстрее. К сожалению, в разных точках картографической проекции трассируемый луч может быть направлен произвольным образом. что в общем случае не позволяет эффективно организовать трассировку множества лучей. Однако для ряда картографических проекций (например, для проекций Гаусса-Крюгера и UTM) обеспечивается условие, что проекция направления трассировки луча на ось N<sub>v</sub> не изменяет знак по всему полю снимка. Знак этой проекции определяется наклонением орбиты КА и направлением бокового обзора (влево или вправо). В рамках одного сеанса съемки не изменяются ни наклонение орбиты, ни направление бокового обзора, именно поэтому знак проекции направления трассировки на ось  $N_{\nu}$ остается постоянным. Таким образом, для наиболее распространенных картографических проекций Гаусса-Крюгера и UTM можно эффективно организовать одновременную трассировку сразу всех лучей, двигаясь по растру сверху вниз или снизу вверх (в зависимости от проекции направления трассировки на ось N<sub>к</sub>).

Для каждого пикселя  $(N_{\kappa}, E_{\kappa})$  аналогично построению карты расширения пикселей определяется направление трассировки луча в виде вектора  $(\overline{N}_{\kappa} - N_{\kappa}, \overline{E}_{\kappa} - E_{\kappa})$ . В противоположном направлении ищется точка  $(\widetilde{N}_{\kappa 0}, \widetilde{E}_{\kappa 0})$  предыдущей обработанной строки, для которой известна высота границы тени. Отрезок между точками  $(\widetilde{N}_{\kappa 0}, \widetilde{E}_{\kappa 0})$  и  $(N_{\kappa}, E_{\kappa})$  разбивается на несколько равных частей, длина которых  $d_m$  не превышает  $d_{DEM}$ , с помощью точек  $(\widetilde{N}_{\kappa i}, \widetilde{E}_{\kappa i})$ ,  $i = \overline{1, I_m - 1}$ . Для каждой i-й точки находится высота границы тени:

$$h_m(\widetilde{N}_{\kappa i}, \widetilde{E}_{\kappa i}) = \max\left(h(\widetilde{N}_{\kappa i}, \widetilde{E}_{\kappa i}), h_m(\widetilde{N}_{\kappa i-1}, \widetilde{E}_{\kappa i-1}) - \frac{d_m}{\operatorname{tg}\alpha_{\mu a \partial}}\right), \quad (4)$$

где  $\alpha_{_{Had}}$  – угол отклонения направления зондирования от надира.

Для точки  $(N_{\kappa}, E_{\kappa})$  возвышение над границей тени определяется как

$$\Delta h_m(N_\kappa, E_\kappa) = h(N_\kappa, E_\kappa) - h_m(\widetilde{N}_{\kappa I_m - 1}, \widetilde{E}_{\kappa I_m - 1}) + \frac{d_m}{\operatorname{tg} \alpha_{\mu a \partial}}.$$
 (5)

Отрицательные значения  $\Delta h_m$  соответствуют областям радиотени. Полученная карта значений  $\Delta h_m(N_\kappa, E_\kappa)$  интерполируется с помощью фильтра Ланцоша, а затем путем пороговой обработки из нее формируется нечеткая маска теней:

$$\mu(S) = \begin{cases} 1, & \frac{\Delta h_m \operatorname{tg} \alpha_{\mu a \partial}}{d_m} \leq 0, \\ 1 - \frac{10\Delta h_m \operatorname{tg} \alpha_{\mu a \partial}}{3d_m}, & 0 < \frac{\Delta h_m \operatorname{tg} \alpha_{\mu a \partial}}{d_m} < 0, 3, \\ 0, & \frac{\Delta h_m \operatorname{tg} \alpha_{\mu a \partial}}{d_m} \geq 0, 3. \end{cases}$$
(6)

#### Нечеткие правила комплексирования

Для построения нечетких правил комплексирования формализуем с помощью математической логики пра-

вила, предложенные в [1]. Пусть  $L_1$  и  $L_2$  – булевы признаки наличия разрыва на исходных РЛИ, а  $S_1$  и  $S_2$  – булевы признаки затенения, тогда:

$$\begin{split} s_{\kappa o M n} &= w(\overline{L_2} \wedge (S_1 \wedge \overline{S_2} \vee L_1))s_1 + w(\overline{L_2} \wedge (S_1 \wedge \overline{S_2} \vee L_1))s_2, \quad \textbf{(7)} \\ \textbf{где } s_{\kappa o M n} &- \textbf{сигнал результата комплексирования, } s_1 - \textbf{сигнал основного РЛИ, } s_2 - \textbf{сигнал дополнительного} \\ \textbf{РЛИ, } w(x) &= \begin{cases} 1, & x - u c m u h a, \\ 0, & x - n o \mathcal{H} c \textbf{.} \end{cases} \end{split}$$

Для перехода к нечеткой логике заменим булевы признаки  $L_1$ ,  $L_2$ ,  $S_1$  и  $S_2$  нечеткими переменными, степени истинности которых задаются в соответствии с (3) и (6). Значения порогов в (3) и (6) можно варьировать с учетом соотношения пространственных разрешений опорной трехмерной модели и РЛИ, а также с учетом желаемой плавности переходов от одного РЛИ к другому на результате их комплексирования.

Конъюнкции и дизъюнкции в (7) заменим нечеткими аналогами на основе следующей *T* -нормы:

$$T(\mu_1, \mu_2) = \mu_1 \mu_2.$$
(8)

*T* -норма определяет степень истинности результата нечеткой конъюнкции. Степень истинности результата дизъюнкции определяется *T* -конормой, которая для *T* -нормы (8) имеет вид:

$$S(\mu_1, \mu_2) = \mu_1 + \mu_2 - \mu_1 \mu_2.$$
(9)

Наконец, строго единичные и нулевые веса в (7) заменим степенями истинности соответствующих нечетких логических выражений:

$$s_{\kappa o M n} = \mu (L_2 \wedge (S_1 \wedge S_2 \vee L_1))s_1 + + \mu (\overline{L_2} \wedge (S_1 \wedge \overline{S_2} \vee L_1))s_2,$$
(10)

За счет того, что степени истинности вычисляются для двух взаимно противоречивых логических выражений, сумма весов, с которыми суммируются  $s_1$  и  $s_2$ , всегда оказывается равной единице.

#### Экспериментальные исследования

Работу предложенного алгоритма проиллюстрируем на паре радиолокационных изображений, представленных на рис. 1. Маска разрывов и переналожений для первого комплексируемого РЛИ представлена на рис. 4, а маска затенений – на рис. 5. Градациями серого на рисунках представлены значения  $\mu(L)$  и  $\mu(S)$  в диапазоне от 0 до 1.

Из рис. 4 видно, что все области, сильно размытые на ортотрансформированном РЛИ, оказались под маской, что говорит о высоком ее качестве. В то же время сильное размытие и полученная маска неидеально соответствуют ярким участкам РЛИ. Это связано с недостаточной детальностью опорной трехмерной модели, использовавшейся при ортотрансформировании и построении маски. Из рис. 5 видно, что выделенные участки тени приходятся на темные области РЛИ, однако не все темные области РЛИ помечены как тень. Это связано с особенностями отражения в этих областях и с недостаточной детальностью опорной трехмерной модели.

В целом из рисунков видно, что построенные маски отличаются достаточно высоким качеством и не требуют дополнительной морфологической обработки. Неточные



Рис. 4. РЛИ и нечеткая маска разрывов и переналожений на нем



Рис. 5. РЛИ и нечеткая маска затенений на нем

границы масок связаны лишь с низкой детальностью опорной трехмерной модели. А в случае, когда детальность опорного рельефа соответствует пространственному разрешению обрабатываемых изображений, маски разрывов и затенений строятся с высокой точностью.

Результат нечеткого комплексирования с использованием полученных масок приведен на рис. 6, из которого видно, что резкие границы (рис. 2) на стыках фрагментов, заполненных данными из разных РЛИ, превратились в плавные переходы. Более того, плавность перехода обусловлена не произвольным искусственным размытием масок разрывов и теней, а плавностью изменений высот рельефа, поскольку значения  $\mu(L)$  и  $\mu(S)$  определяются, в первую очередь, крутизной склонов рельефа. Поэтому резкие границы высотных объектов останутся резкими и на результате комплексирования, а плавные изменения высоты будут сопровождаться плавным переходом радиояркости от одного РЛИ к другому.



Рис. 6. Плавные переходы на результате комплексирования на основе нечеткой логики

#### Модификации предложенного алгоритма

В случае низкой детальности опорной трехмерной модели положение выявленных разрывов и теней может не строго совпадать с сюжетом РЛИ. Из геометрии радиолокационной съемки следует, что разрывы и переналожения должны быть очень яркими, а в области теней сигнал должен быть на уровне фонового шума. Этим можно воспользоваться для уточнения положения разрывов и теней на основе анализа сигнала РЛИ. Однако следует помнить, что в участках, где выделенные разрывы и тени не согласуются с сюжетом, на комплекссируемых РЛИ наблюдаются значительные геометрические искажения, вызванные неточным ортотрансформированием, из-за которых два РЛИ могут быть неточно совмещены друг с другом.

Для уточнения положения разрывов и теней по сигналу РЛИ необходимо оценить средние яркости пикселей, для которых  $\mu(L) = 1$  (средняя яркость разрыва/переналожения) или  $\mu(S) = 1$  (средняя яркость тени). Чтобы полученные оценки яркостей были адекватными, опорная трехмерная модель должна быть достаточно точной. Для очень грубого опорного рельефа уточнить положение разрывов и теней, скорее всего, будет невозможно. После получения средних яркостей выполняется анализ яркости пикселей РЛИ в окрестности выделенных областей разрыва и тени. Пикселям, яркость которых близка к средней яркости разрыва/тени, назначается  $\mu(L)$  и  $\mu(S)$ , близкая к единице. Чтобы уменьшить влияние спекл-шума, можно анализировать не яркость отдельного пикселя, а среднюю яркость в его малой окрестности.

Предложенный выше подход к комплексированию на основе нечеткой логики при необходимости можно рас-

пространить и на случай, когда исходные РЛИ смешиваются в отсутствие разрывов и теней, например, как было предложено в [2]. В указанном случае правила комплексирования несколько изменяются, а следовательно, необходимо модифицировать логические выражения, их описывающие:

$$s_{\kappa o M n} = \mu (L_1 \land (S_2 \land S_1 \lor L_2))s_1 + + \mu (\overline{L_2} \land (S_1 \land \overline{S_2} \lor L_1))s_2 + + \mu (L_1 \land L_2 \lor S_1 \land S_2) \frac{s_1 + s_2}{2} + + \mu (\overline{L_1} \land \overline{L_2} \land \overline{S_1} \land \overline{S_2})F(s_1, s_2),$$
(11)

где  $F(s_1, s_2) = F(s_2, s_1)$  – некоторая функция, выполняющая смешивание сигналов двух РЛИ.

В постановке (11) результат комплексирования не зависит от того, какое из объединяемых РЛИ выбрано основным, а какое – дополнительным. Выбор конкретной функции  $F(s_1, s_2)$  в настоящей работе не рассматривается. Ее можно реализовать, основываясь на работе [2],

а можно – на основе одного из универсальных алгоритмов комплексирования изображений. В любом случае *F*(*s*<sub>1</sub>, *s*<sub>2</sub>) уже не учитывает геометрический аспект, а работает исключительно с яркостью пикселей объединяемых изображений.

#### Заключение

Предложенный алгоритм позволяет качественно построить маски геометрических искажений и затенений объектов на РЛИ, а затем выполнить их комплексирование с обеспечением плавности перехода между участками, заполненными данными из разных РЛИ. При этом плавность перехода определяется характером изменения высот объектов наблюдаемой сцены.

#### Литература

1. Zhang J., Wei J., Huang G., Zhang Y. Fusion of ascending and descending polarimetric SAR data for colour orthophoto generation // International Archives of Photogrammetry and Remote Sensing. 2010. Vol. 38. Part 7A. Pp. 323-328.

2. Chandrakanth R., Saibaba J., Varadan G., Ananth Raj P. Fusion of ascending and descending pass high resolution SAR data // Journal of Geomatics. 2014. Vol. 8(2). Pp. 164-169.



#### ЦИФРОВАЯ ОБРАБОТКА СИГНАЛОВ И ЕЕ ПРИМЕНЕНИЕ

*The 21<sup>th</sup> International Conference* DIGITAL SIGNAL PROCESSING AND ITS APPLICATIONS

Москва, ИПУ РАН, март 2019 года

Уважаемые коллеги!

#### ПРИГЛАШАЕМ ВАС ПРИНЯТЬ УЧАСТИЕ В РАБОТЕ КОНФЕРЕНЦИИ

#### МЕЖДУНАРОДНЫЙ ОРГАНИЗАЦИОННЫЙ КОМИТЕТ КОНФЕРЕНЦИИ:

Председатель Международного комитета: академик РАН Гуляев Юрий Васильевич

Сопредседатели: академик РАН Васильев С.Н., академик РАН Каляев И.А., Кирпичников А.П.,

академик РАН Сигов, академик РАН Федоров И.Б.

#### Программная группа Международного оргкомитета:

#### Председатель Программной группы: член-корр. РАН Зубарев Юрий Борисович

**Члены Программной группы:** проф. Аджемов А.С. (Россия), член-корр. РАН Борисов В.И. (Россия), д.т.н. Борисов Ю.И. (Россия), академик РАН Бугаев А.С. (Россия), проф. Витязев В.В., проф. Лукин В.В. (Украина), проф. Митра М. (США), проф. Петровский А.А. (Белоруссия), проф. Сарамаки Т. (Финляндия), член-корр. РАН Черепенин В.А. (Россия).

Административная группа Международного оргкомитета: проф. Бартенев В.Г., проф. Брюханов Ю.А., проф. Васильев К.К., проф. Дворкович В.П., проф. Дворкович А.В., проф. Джиган В.И., Зубарев Р.В., проф. Пожидаев В.Н., д.т.н. Приоров А. Л., проф. Прохоров Ю.Н., доц. Самсонов Г.А., проф. Санников В.Г., доц. Сергиенко А.Б., проф. Скачко Ю.В., проф. Солонина А.И., проф. Соколов И.А., проф. Сперанский В.С., доц. Стешенко В.Б., проф. Фаворская М.Н., проф. Чобану М.К., проф. Шинаков Ю.С., проф. Ямпурин Н.П.

#### РАБОТА КОНФЕРЕНЦИИ ПЛАНИРУЕТСЯ ПО 9 СЕКЦИЯМ:

- Теория сигналов и систем
- Теория и методы цифровой обработки сигналов
- Обработка сигналов в системах телекоммуникаций
- Обработка сигналов в радиотехнических системах
- Обработка и передача изображений

- Обработка сигналов в биотехнических системах
- Проектирование и техническая реализация систем ЦОС
- Системы «технического зрения», искусственного инте-
- лека и робототехники
- Нейрокомпьютерная обработка сигналов и изображений

Подробную информацию о работе конференции смотрите на сайте РНТО РЭС им. А.С. Попова: www.rntores.ru

#### УДК 004.932

#### ЭФФЕКТИВНАЯ ОРГАНИЗАЦИЯ МАССОВЫХ КООРДИНАТНЫХ ПРЕОБРАЗОВАНИЙ ПРИ ГЕОМЕТРИЧЕСКОЙ ОБРАБОТКЕ РАДИОЛОКАЦИОННЫХ ИЗОБРАЖЕНИЙ

Егошкин Н.А., в.н.с. НИИ «Фотон» РГРТУ, к.т.н., foton@rsreu.ru; Ушенкин В.А., с.н.с. НИИ «Фотон» РГРТУ, к.т.н., foton@rsreu.ru.

# **EFFECTIVE ORGANIZATION OF MASS COORDINATE TRANSFORMATIONS IN THE GEOMETRIC PROCESSING OF SAR IMAGES**

#### Egoshkin N.A., Ushenkin V.A.

The problem of mass coordinate transformations effective organization in the geometric processing of SAR images is considered in order to achieve high processing speed with low RAM costs. The approximate regular grid with piecewise parabolic interpolation between nodes is proposed. Its efficiency in terms of speed and memory is substantiated and confirmed experimentally.

Key words: SAR image, geometric processing, interpolation, geocoding, orthotransformation.

Ключевые слова: радиолокационное изображение, геометрическая обработка, интерполяция, геокодирование, ортотрансформирование.

#### Введение

Геометрическая обработка радиолокационных изображений (РЛИ) является одним из основных этапов наземной обработки данных космических радиолокационных систем и включает три основные операции:

1) геореференцирование – преобразование РЛИ из геометрии «наклонная дальность – ази-

мут» в геометрию «наземная дальность – азимут» с выравниванием шага дискретизации по дальности и по азимуту;

 геокодирование – преобразование РЛИ в картографическую проекцию с учетом средней высоты наблюдаемой сцены;

 ортотрансформирование – преобразование РЛИ в картографическую проекцию с учетом цифровой модели пространственного распределения высот объектов наблюдаемой сцены.

Пересчет планарных координат пикселя на РЛИ в картографические координаты может быть выполнен на основе геодезической привязки изображения. Номер столбца на РЛИ определяет значение наклонной дальности на траверзе  $R_{\perp}$ , а номер строки – азимутальное время  $\eta_{\perp}$  максимального сближения радиолокатора с синтезированной апертурой антенны (РСА) и наблюдаемой цели. Для каждой строки РЛИ известны координаты фазового центра антенны РСА  $\mathbf{r}_{_{p}}(\eta_{\perp}) = (x_{_{p}}, y_{_{p}}, z_{_{p}})$  и вектор линейной скорости его движения  $\mathbf{V}_{n}(\eta_{\perp})$  =  $=(V_{_{DX}},V_{_{DY}},V_{_{DZ}})$ , задаваемые в Гринвичской системе координат (трехмерной прямоугольной геоцентрической системе, ось Z которой направлена по земной оси, а оси Х и У лежат в плоскости экватора и вращаются вместе с Землей). Для каждой пары  $(R_{\perp}, \eta_{\perp})$  может быть составлена система уравнений:

Рассматривается проблема эффективной организации массовых координатных преобразований при геометрической обработке изображений от космических радиолокаторов с синтезированной апертурой с целью достижения высокой скорости обработки при небольших затратах оперативной памяти. Предлагается аппроксимационная регулярная сетка с кусочно-параболической интерполяцией между узлами. Обосновывается и экспериментально подтверждается ее эффективность по быстродействию и затратам оперативной памяти.

$$\begin{cases} V_{px}(\eta_{\perp})(x - x_{p}(\eta_{\perp})) + V_{py}(\eta_{\perp})(y - y_{p}(\eta_{\perp})) + \\ + V_{pz}(\eta_{\perp})(z - z_{p}(\eta_{\perp})) = 0, \\ \frac{x^{2} + y^{2}}{(a_{3} + h)^{2}} + \frac{z^{2}}{(b_{3} + h)^{2}} = 1, \\ (x - x_{p}(\eta_{\perp}))^{2} + (y - y_{p}(\eta_{\perp}))^{2} + (z - z_{p}(\eta_{\perp}))^{2} = R_{\perp}^{2}, \end{cases}$$
(1)

где  $a_3$  и  $b_3$  – полуоси земного эллипсоида, h – априорно известная геодезическая высота наблюдаемой цели, (x, y, z) – координаты наблюдаемой цели в Гринвичской системе координат.

Найденные координаты (x, y, z) могут быть пересчитаны в геодезические широту B и долготу L в соответствии с ГОСТ 32453-2017 [1], а по ним в свою очередь с помощью правил проецирования могут быть найдены картографические координаты  $(N_x, E_x)$ .

Решение системы (1) находится итерационно по методу Ньютона и совместно с преобразованиями  $(x, y, z) \rightarrow (B, L) \rightarrow (N_{\kappa}, E_{\kappa})$  требует большого количества вычислений в каждом пикселе изображения, что увеличивает временные затраты на геометрическую обработку.

Тем не менее, от пикселя к пикселю результаты преобразований изменяются плавно, практически по линейному закону. Это позволяет строго рассчитать их лишь в узлах некоторой редкой сетки, а между узлами оценивать приближенно на основе интерполяции. Цель настоящей работы – выбрать структуру сетки и вид интерполяции между ее узлами, обеспечивающие разумный компромисс между вычислительной эффективностью массовых координатных преобразований и затратами оперативной памяти на хранение сетки.

# Прямой и обратный законы геометрических преобразований

Геометрическое трансформирование можно выполнять по прямому либо по обратному закону. Если обработка ведется по прямому закону, то для каждого пикселя исходного РЛИ находится соответствующий пиксель на результирующем РЛИ, в который помещается код яркости исходного пикселя. Недостатком обработки по прямому закону является высокая вероятность получить пропуск на результирующем РЛИ, когда выполняется ортотрансформирование для местности со сложным рельефом или когда шаг дискретизации на результате меньше, чем на исходном РЛИ. Кроме того, крайне проблематично учесть субпиксельные эффекты и вычислительно сложно трансформировать опорные высоты в геометрию исходного РЛИ. Трансформация высот в этом случае представляет собой итерационный поиск для каждого пикселя РЛИ высоты h, при которой решение системы (1) даст такие (B, L), что h = h(B, L), где h(B,L) – интерполированная цифровая модель рельефа в геодезической системе координат. Иначе говоря, объем вычислений для каждого пикселя увеличивается на порядок по сравнению с вычислением геодезической привязки.

Данных недостатков лишена обработка по обратному закону. В этом случае для каждого пикселя результирующего РЛИ находится соответствующий ему пиксель на исходном РЛИ. Если полученные координаты получаются нецелочисленными, выполняется интерполяция исходного РЛИ. Интерполированное значение радиояркости помещается в пиксель результирующего РЛИ.

Тем не менее, обработка по обратному закону порождает две проблемы: во-первых, она не позволяет определить размеры результирующего РЛИ, а вовторых, геодезическая привязка РЛИ и правила проецирования определяют прямой закон, а не обратный. Первая проблема обычно решается путем выборочного применения прямого закона к пикселям исходного РЛИ (например, к граничным пикселям).

Вторая проблема для радиолокационной съемки решается достаточно просто. Во-первых, для наиболее распространенных картографических проекций (UTM, проекция Гаусса-Крюгера, стереографическая проекция) имеются аналитические выражения для обратного преобразования картографических координат в геодезические, сопоставимые по объему вычислений с правилами проецирования. Во-вторых, обратная геодезическая привязка РЛИ, в отличие от прямой привязки, является однозначной на интервале съемки и имеет вполне ясный геометрический смысл: опускание перпендикуляра из наблюдаемой точки с координатами (x, y, z), полученными из (B, L, h) по ГОСТ 32453-2017, на траекторию движения РСА и вычисление длины этого перпендикуляра  $R_{\perp}$ , а также момента времени  $\eta_{\perp}$ , когда РСА

оказывается в основании перпендикуляра. Фактически обращение геодезической привязки РЛИ сводится к решению уравнения:

$$V_{px}(\eta_{\perp})(x - x_{p}(\eta_{\perp})) + V_{py}(\eta_{\perp})(y - y_{p}(\eta_{\perp})) + V_{pz}(\eta_{\perp})(z - z_{p}(\eta_{\perp})) = 0.$$
(2)

Уравнение (2) может быть решено методом половинного деления или методом хорд после предварительной локализации корня на выбранном временном интервале (например, интервале съемки). После нахождения  $\eta_{\perp}$  наклонная дальность на траверзе выражается следующим образом:

$$R_{\perp} = \sqrt{(x - x_p(\eta_{\perp}))^2 + (y - y_p(\eta_{\perp}))^2 + (z - z_p(\eta_{\perp}))^2}.$$
 (3)

В то же время при обработке по обратному закону существенно упрощается учет опорной трехмерной модели наблюдаемой сцены h(B,L), изначально представленной в геодезической системе координат. Ее необходимо предварительно трансформировать не в геометрию исходного РЛИ, а в картографическую проекцию, как и результирующее РЛИ. Для этого не требуется сложных итерационных вычислений в каждом пикселе, а достаточно лишь применить правила проецирования к широте и долготе, игнорируя высоту.

Тем не менее, даже при геометрической обработке по обратному закону строгие координатные преобразования требуют достаточно много вычислений, что негативно отражается на скорости геометрической обработки. Поэтому вопрос выбора аппроксимирующих сеток для массовых преобразований координат пикселей остается актуальным.

#### Выбор аппроксимирующей сетки

В оптике основную проблему составляет обращение геодезической привязки. Оно требует гораздо большего количества вычислений, чем решение уравнения (2) в радиолокации, поскольку там участвует не только линейное, но и угловое движение космического аппарата, а также дисторсия оптики. Из-за этого обратить привязку даже в узлах сетки может быть вычислительно накладно. Поэтому в [2] предлагается сначала построить сетку для прямого закона, для каждой ячейки прямой сетки определить, какие узлы обратной сетки попадают в нее, и вычислить значения в узлах сетки, аппроксимирующей обратный закон. Чтобы было просто определить, какие узлы обратной сетки попадают в ячейку прямой сетки, интерполяция между узлами ограничивается кусочноаффинной или кусочно-билинейной. Это в свою очередь требует достаточно большого количества узлов в сетке, чтобы обеспечить приемлемую точность аппроксимации. Особенно это касается ортотрансформирования, когда аппроксимирующая сетка становится трехмерной.

Поскольку в радиолокации неизбежен боковой обзор, влияние рельефа оказывается сильным и приводит к большому количеству узлов сетки по третьему измерению. Типичный шаг кусочно-линейной сетки составляет около 50 пикселей по широте и долготе и около 60 м по высоте. В каждом узле сетки хранится пара чисел, которые для больших изображений должны иметь формат с плавающей точкой двойной точности (по 8 байтов на каждое число). Максимально возможный перепад высот рельефа составляет около 10 км (Гималаи). Следовательно, при ортотрансформирования одноканального 16-битного радиолокационного изображения объем памяти для хранения аппроксимирующей сетки в худшем случае составляет примерно половину объема изображения. Такие накладные расходы оперативной памяти неприемлемы. Поэтому целесообразно рассмотреть интерполяцию между узлами сетки более высокого порядка, отличающуюся большей точностью и соответственно требующую существенно меньшего количества узлов в сетке.

В настоящей работе предлагается использовать в сетке трехмерную кусочную аппроксимацию параболами. По одному измерению кусочно-параболическая сетка, построенная на некотором отрезке, включает N+1 основных и N промежуточных узлов. Основные и промежуточные узлы чередуются друг с другом, причем на краях отрезка располагаются основные узлы. Через каждую пару соседних основных узлов и промежуточный узел между ними (так называемая ячейка сетки) проводится парабола (рис. 1). Одномерные сетки по  $E_{\kappa}$ , построенные для различных значений  $N_{\kappa}$ , образуют двумерную сетку. Одномерные сетки по Е, также разделяются на основные и промежуточные. Через соответствующие точки двух соседних основных одномерных сеток и промежуточной сетки между ними также проводятся параболы. Ячейку двумерной сетки при этом образуют 3×3 соседних узлов. Третье измерение в сетке добавляется аналогичным образом. Под шагом сетки по одному из измерений будем понимать расстояние между двумя соседними основными узлами.



Рис. 1. Кусочно-параболическая сетка в одномерном варианте

При геокодировании РЛИ по обратному закону координатные преобразования записываются в виде  $(N_{\kappa}, E_{\kappa}) \rightarrow (R_{\perp}, \eta_{\perp})$ . Для их описания требуется векторная двумерная сетка. При ортотрансформировании РЛИ координатные преобразования уже имеют вид  $(N_{\kappa}, E_{\kappa}, h) \rightarrow (R_{\perp}, \eta_{\perp})$  и описываются векторной трехмерной сеткой.

При геометрических преобразованиях РЛИ расстояние между основными узлами кусочно-параболической сетки по горизонтали будет составлять от нескольких сотен до нескольких десятков тысяч пикселей, а по высоте – от одного до нескольких километров (в зависимости от конкретных условий съемки). В среднем можно считать, что шаг сетки по широте и долготе составляет около 1000 пикселей, а по высоте – около 2 км. В итоге объем памяти для хранения сетки составит сотые доли процента от объема изображения, что вполне приемлемо.

Поскольку в радиолокации геодезическая привязка обращается гораздо легче, чем в оптике, сетки для обратного закона могут строиться напрямую. Это снимает ограничение на интерполяцию между узлами и позволя-

ет использовать кусочно-параболическую. Кроме того, более высокий порядок интерполяции позволяет сократить вычислительные расходы на построение сетки за счет уменьшения количества узлов. Основной проблемой интерполяций высокого порядка является увеличение объема вычислений при массовых координатных преобразованиях с использованием сетки. Тем не менее, покажем, что при использовании предлагаемого вида интерполяции это увеличение незначительно в рамках всей геометрической обработки. При этом рассмотрим два случая: двумерный и трехмерный варианты сетки.

В двумерном варианте сетки массовые координатные преобразования последовательно вызываются для параметров  $(N_{\kappa 0}, E_{\kappa 0}), (N_{\kappa 0}, E_{\kappa 0} + \Delta E_{\kappa}), (N_{\kappa 0}, E_{\kappa 0} + + 2\Delta E_{\kappa}), \dots, (N_{\kappa 0} + \Delta N_{\kappa}, E_{\kappa 0}), (N_{\kappa 0} + \Delta N_{\kappa}, E_{\kappa 0} + \Delta E_{\kappa}), (N_{\kappa 0} + \Delta N_{\kappa}, E_{\kappa 0} + 2\Delta E_{\kappa})$  и т.д. То есть фактически для фиксированного  $N_{\kappa}$  последовательно вызываются преобразования для разных  $E_{\kappa}$  с постоянным шагом, и лишь затем изменяется  $N_{\kappa}$ .

Пусть некоторое количество значений вида  $E_{\kappa 0}$ ,  $E_{\kappa 0} + \Delta E_{\kappa}, \ E_{\kappa 0} + 2\Delta E_{\kappa}$  и т.д. попадает в пределы одной ячейки двумерной кусочно-линейной сетки по горизонтали. Поскольку все преобразования ведутся для фиксированного  $N_{\kappa}$ , билинейную интерполяцию можно заменить на одномерную, предварительно линейно интерполировав значения в двух парах соседних узлов сетки по вертикали. Через полученные интерполированные узловые значения проводится прямая вида

$$(R_{\perp},\eta_{\perp}) = \frac{E_{\kappa}}{\Delta E_{\kappa}}(k_{\text{nunR}},k_{\text{nunq}}) + (b_{\text{nunR}},b_{\text{nunq}}). \quad \mathsf{B} \quad \mathsf{pesynptate}$$

линейную интерполяцию нужно применять напрямую лишь для расчетов в начальной точке  $E_{\kappa 0}$ , а остальные результаты координатных преобразований можно быстро вычислять по предыдущим значениям:

$$R_{\perp i} = R_{\perp i-1} + k_{\text{nuhR}}, \quad \eta_{\perp i} = \eta_{\perp i-1} + k_{\text{nuhn}}.$$
 (4)

Таким образом, вычислительная сложность одного координатного преобразования по двумерной сетке с билинейным интерполятором асимптотически стремится к двум сложениям.

Если линейную интерполяцию заменить на параболическую, то последовательные вычисления для регулярно изменяющихся  $E_{\kappa}$  тоже можно свести к сложениям. На самом деле эту задачу можно свести к сложениям для любого полинома, поскольку у него конечно число ненулевых производных, причем старшая ненулевая производная постоянна. Количество сложений при этом будет в два раза больше степени полинома (за счет того, что коэффициенты полинома – векторы из двух

элементов). Для параболы 
$$(R_{\perp},\eta_{\perp})=0,5{\left(rac{E_{\kappa}}{\Delta E_{\kappa}}
ight)}^2 imes$$

$$\times (a_{\textit{napR}}, a_{\textit{nap\eta}}) + \frac{E_{\kappa}}{\Delta E_{\kappa}} (b_{\textit{napR}}, b_{\textit{nap\eta}}) + (c_{\textit{napR}}, c_{\textit{nap\eta}}) \quad \text{последова-}$$

тельные координатные преобразования записываются в виде:

$$k_{\text{numRi}} = k_{\text{numRi}-1} + a_{\text{napR}}, \quad k_{\text{numni}} = k_{\text{numni}-1} + a_{\text{napn}}, \quad (5)$$

$$R_{\perp i} = R_{\perp i-1} + k_{nunRi-1}, \quad \eta_{\perp i} = \eta_{\perp i-1} + k_{nunni-1}, \tag{6}$$

где 
$$(k_{\text{линRi}}, k_{\text{линqi}}) = 0, 5(a_{\text{napR}}, a_{\text{napq}}) + (b_{\text{napR}}, b_{\text{napq}}) + E_{\kappa 0} \times$$

 $\times (a_{napR}, a_{nap\eta}).$ 

Таким образом, вычислительная сложность одного координатного преобразования по двумерной сетке с бипараболическим интерполятором асимптотически стремится к четырем сложениям. С одной стороны, это в два раза больше, чем при билинейной интерполяции, но с другой стороны, всё равно намного меньше, чем затраты на передискретизацию сигнала при геометрических преобразованиях.

В случае же трехмерной сетки, учитывающей высоту рельефа, объем вычислений существенно возрастает как для трилинейной, так и для трипараболической интерполяции. Это связано с тем, что значения высоты  $h(N_{v}, E_{v})$  изменяются произвольным образом при движении вдоль строк и столбцов двумерной проекции сетки, причем разрешение трехмерной модели выше, чем шаги сетки по  $N_{\kappa}$  и  $E_{\kappa}$ . В результате приходится поступать следующим образом. В пределах нескольких последовательных преобразований. приходяшихся с учетом вариации высоты на одну ячейку трехмерной сетки, выполняется последовательная двумерная интерполяция в двух (для прямой) или трех (для параболы) сечениях сетки для соседних узловых значений высоты, а затем независимо для каждого преобразования выполняется линейная или параболическая интерполяция по высоте. В результате для трилинейного интерполятора каждое преобразование асимптотически сводится к четырем сложениям и одномерной линейной интерполяции, а для трипараболического интерполятора – к двенадцати сложениям и одномерной интерполяции параболой. Таким образом, применение трипараболического интерполятора увеличивает асимптотическую сложность координатных преобразований по трехмерной сетке примерно в три раза. Однако для горной местности со сложным рельефом проигрыш по скорости оказывается гораздо меньше за счет более крупного размера ячеек сетки и более низкой вероятности, что очередное изменение высоты приведет к переходу в другую ячейку. Кроме того, вычислительные затраты на координатные преобразования все равно остаются примерно в полтора раза меньше, чем на передискретизацию сигнала изображения на основе билинейной интерполяции. С учетом же того, что для РЛИ передискретизацию обычно выполняют с помощью фильтров более высокого порядка (бикубического или на основе sinc-функции), увеличение затрат на координатные преобразования практически не отражается на скорости ортотрансформирования.

#### Алгоритм построения аппроксимирующей сетки, обеспечивающей заданную точность преобразования координат

Рассмотрим процесс построения сетки, обеспечивающей приемлемую точность координатных преобразований. При этом ограничимся трехмерной кусочнопараболической сеткой как более общим случаем.

Предельная ошибка координатных преобразований должна составлять не более десятой доли пикселя, чтобы не быть визуально заметной на результате ортотрансформирования. В то же время для автоматического выбора шага сетки нельзя задавать слишком малую предельную ошибку, поскольку расчеты прямой и обратной геодезической привязки выполняются численно, а их результаты содержат вычислительные погрешности. Задаваемая предельная ошибка должна быть больше этих погрешностей.

Выбирать шаг сетки будем независимо для каждого из измерений. Сначала установим величины шагов по измерениям, соответствующие ширине, высоте результирующего РЛИ и диапазону высот. Затем будем уменьшать шаг сетки по отдельным измерениям в 2 раза, до тех пор пока ошибка интерполяции между узлами не снизится до приемлемой.

Наибольшую проблему здесь составляет эффективное оценивание максимальной ошибки интерполяции независимо по каждому из измерений. В общем виде решить эту задачу крайне затруднительно, поэтому будем получать лишь приближенную оценку, основываясь на следующих предположениях:

 ошибка интерполяции по каждому измерению монотонно изменяется при движении вдоль других измерений;

2) гиперповерхности  $R_{\perp}(N_{\kappa}, E_{\kappa}, h)$  и  $\eta_{\perp}(N_{\kappa}, E_{\kappa}, h)$  раскладываются по каждому из измерений в ряд Тейлора, в котором коэффициент перед каждым следующим членом оказывается намного меньше предыдущего коэффициента.

В соответствии с первым предположением будем оценивать ошибку интерполяции по каждому измерению лишь для граничных значений двух других параметров. В соответствии со вторым положением ошибка интерполяции параболой близка к кубической функции с двумя точками экстремума посередине между промежуточным и одним из основных узлов. Именно в этих точках для каждой ячейки одномерной сетки и оценивается ошибка интерполяции.

В реальности сделанные предположения могут выполняться нестрого, из-за чего реальная максимальная ошибка построенной сетки может оказаться в несколько раз больше, чем оцененная. В связи с этим предельную ошибку координатных преобразований для построения сетки следует задавать с некоторым запасом.

#### Экспериментальные исследования

Для экспериментального подтверждения полученных теоретических оценок эффективности предложенной аппроксимационной сетки выполним ортотрансформирование РЛИ (рис. 2) с использованием кусочнолинейной и кусочно-параболической сеток для массовых координатных преобразований. Трансформируемое РЛИ выбрано небольшого размера (примерно 11000 × 6000), поэтому кусочно-линейная сетка, хотя и содержит большое количество узлов, всё же помещается в оперативной памяти.



Рис. 2. Исходное (слева) и ортотрансформированное (справа) РЛИ

На персональном компьютере с процессором Intel Core i7-6800К и 32 ГБ оперативной памяти скорость ортотрансформирования составила:

1) с использованием кусочно-линейной сетки – 72,1 МБ/с;

 с использованием кусочно-параболической сетки – 72,5 МБ/с.

Таким образом, разница в скорости обработки оказалась даже меньше, чем следовало из теоретических оценок. Это можно объяснить тем, что кусочно-линейная сетка занимает существенно больший объем памяти и доступ к ее узлам замедляется из-за более частых кэшпромахов.

Шаги построенных аппроксимационных сеток составили: 1) для кусочно-линейной сетки: по  $E_v$  – 61 пиксель,

по  $N_{\kappa}$  – 53 пикселя, по высоте – 97 м;

2) для кусочно-параболической сетки: по  $E_{\kappa}$  и  $N_{\kappa}$  – 1031 пиксель, по высоте – 3200 м.

Объем памяти для хранения кусочно-линейной сетки составил 14,5 МБ (12,3 % от объема изображения), а

для кусочно-параболической – 19 кБ (0,02 % от объема изображения). Несколько меньший объем сеток обусловлен тем, что перепад высот для исследуемого изображения в 3 раза меньше предельно возможного и составляет около 3 км.

#### Заключение

Таким образом, предложенная аппроксимационная сетка для массовых координатных преобразований приводит к ничтожно малому снижению скорости геометрической обработки, но при этом позволяет существенно сократить затраты оперативной памяти по сравнению с кусочно-линейной сеткой.

#### Литература

1. ГОСТ 32453-2017. Глобальная навигационная спутниковая система. Системы координат. Методы преобразований координат определяемых точек. М.: Стандартинформ, 2017. 23 с.

2. Злобин В.К., Еремеев В.В. Обработка аэрокосмических изображений. М.: ФИЗМАТЛИТ, 2006. 288 с.



# В. В. Витязев МНОГОСКОРОСТНАЯ ОБРАБОТКА СИГНАЛОВ

# Витязев В.В. Многоскоростная обработка сигналов – М.: Изд-во «Горячая линия-Телеком», 2017 г. – 336 с.: ил.

Рассмотрена эволюция теории и технологий многоскоростной обработки сигналов в период с начала 70-х гг. прошлого столетия до наших дней с позиции вклада, который внесли в их развитие работы российских ученых и специалистов в области цифровых информационных технологий реального времени. Описаны методы и алгоритмы многоступенчатой и многокаскадной реализаций цифровых узкополосных фильтров и банков цифровых фильтров на основе эффектов прореживания по времени (децимация во временной области) и по частоте (децимация в частотной области).

Приведена методика оптимизации параметров многоступенчатых и многокаскадных структур цифровых полосовых фильтров. Построение оптимальных структур и расчет параметров фильтров частотной селекции иллюстрируется многочисленными примерами.

Для специалистов, научных работников, преподавателей вузов, аспирантов; будет полезна студентам информационных и инфокоммуникационных направлений подготовки.

#### УДК 004.932

#### МАТЕМАТИЧЕСКАЯ МОДЕЛЬ СИНТЕЗА РАДИОЛОКАЦИОННЫХ ИЗОБРАЖЕНИЙ ДЕЦИМЕТРОВОГО РАЗРЕШЕНИЯ ИЗ РАДИОГОЛОГРАММ ОТ КОСМИЧЕСКИХ РСА

Ушенкин В.А., с.н.с. НИИ «Фотон» РГРТУ, к.т.н., foton@rsreu.ru.

#### DECIMETER-RESOLUTION SPACEBORNE SAR RAW DATA FOCUSING MODEL

#### Ushenkin V.A.

The new factors, which have to be taken into account in spaceborne SAR raw data focusing, when SAR image spatial resolution becomes sub-meter and approaches to decimeter, are considered in the paper. The mathematical model of focusing, that provides higher quality in this case, is proposed. It is confirmed by the results of simulation modeling.

Key words: SAR, raw data, focusing, SAR image.

Ключевые слова: РСА, радиоголограмма, фокусировка, радиолокационное изображение.

#### Введение

Радиолокаторы с синтезированной апертурой антенны в ходе съемки формируют не детальное изображение, а специфический сигнал – радиоголограмму, содержащую пространственно

протяженные отклики на точечные объекты. Детальное радиолокационное изображение с компактными откликами на точечные цели получают путем цифровой обработки радиоголограммы в ходе ее фокусировки.

Фокусировка радиоголограммы заключается в согласованной фильтрации ее сигнала и состоит из трех основных этапов:

1) сжатия импульсного отклика PCA по дальности (первичного и вторичного);

 устранения миграции импульсного отклика РСА по отсчетам дальности (выпрямления отклика);

3) сжатия импульсного отклика РСА по азимуту.

Указанная декомпозиция на этапы позволяет эффективно реализовать согласованную двумерную фильтрацию за счет разделения ее по измерениям и выполнения путем умножения спектров сигнала и фильтра.

Если параметры фильтра первичного сжатия по дальности определяются лишь модуляцией излучаемых радиоимпульсов, то параметры остальных фильтров зависят от модели относительного движения фазового центра антенны РСА и наблюдаемой точечной цели.

При формировании изображений с разрешением 1–10 м традиционно применяется модель на основе аппроксимации реального движения космического РСА на интервале синтеза апертуры прямолинейным равномерным движением, при котором зависимость наклонной дальности цели *R* от азимутального времени  $\eta$  имеет вид:

$$R(\eta) = \sqrt{R_{\perp}^{2} + V_{_{\pi_{9}}}^{2} (\eta - \eta_{\perp})^{2}}, \qquad (1)$$

где  $R_{\perp}$  – дальность на траверзе,  $\eta_{\perp}$  – время максимального сближения РСА и цели,  $V_{_{R2}}$  – линейная эквивалентная скорость (параметр модели, вычисляемый по

Анализируются новые факторы, которые необходимо учитывать при фокусировке радиоголограмм от космических радиолокаторов с синтезированной апертурой, когда пространственное разрешение получаемых радиолокационных изображений становится субметровым и приближается к дециметру. Предлагается математическая модель, обеспечивающая в этом случае более высокое качество фокусировки, что подтверждается результатами моделирования.

методу наименьших квадратов (МНК)).

Основным достоинством указанной модели является возможность с помощью метода стационарной фазы [1] аналитически получить простые приближенные выражения для построения фильтров в частотной области, что позволяет относительно быстро выполнить фокусировку. Однако в случае получения изображений субметрового разрешения длина синтезируемой апертуры становится настолько большой, что в ее пределах аппроксимация (1) начинает отклоняться от реального закона изменения наклонной дальности на единицы миллиметров. Таким образом, ошибка приближается по величине к длине волны РСА и начинает оказывать существенное влияние на фазовую составляющую комплексного сигнала радиолокационного изображения.

Цель настоящей работы – разработка новой модели изменения наклонной дальности, на основе которой можно все так же эффективно построить в частотной области фильтры фокусировки и которая обеспечивает субмиллиметровую точность аппроксимации  $R(\eta)$ .

#### Анализ существующих подходов к формированию высокодетальных изображений из радиоголограмм от космических РСА

Вид ошибки аппроксимации (1) на временном интервале синтеза апертуры в 8 с, соответствующем разрешению изображения 0,2 м, при высоте РСА 700 км и наклонной дальности на траверзе 830 км, полученный для реальной орбиты космического аппарата «Sentinel-1A», показан на рис. 1.

В первом приближении ошибка аппроксимации похожа на кубическую параболу с вершиной в точке  $\eta_{\perp}$ , поэтому в работе [2] предложено в (1) ввести дополни-



#### Рис. 1. Изменение погрешности аппроксимации (1) с азимутальным временем

тельный член, пропорциональный  $(\eta - \eta_{\perp})^3$ . Однако к такому закону изменения наклонной дальности крайне затруднительно строго аналитически применить метод стационарной фазы, т.к. он требует решения уравнения шестой степени. Поэтому в [2] найден лишь весьма приближенный результат этого метода, что негативно сказывается на точности фокусировки. Кроме того, из рис. 1 видно, что ошибка аппроксимации не является нечетной функцией, как кубическая парабола. Для указанных значений высоты орбиты и дальности на траверзе в отрицательной области модуль ошибки аппроксимации оказывается в среднем в 2 раза больше, чем в положительной области. Это означает, что введение члена, пропорционального  $(\eta - \eta_{\perp})^3$ , даже в случае строгого применения метода стационарной фазы уменьшит максимальную ошибку аппроксимации не более чем в 4 раза.

Учтя все трудности аналитического описания ошибки аппроксимации (1), в работе [3] предложили численно находить и ошибку аппроксимации, и изменения в частотных фильтрах, которые она вносит. Однако параметры частотных фильтров изменяются по полю радиоголограммы, а нахождение поправки для каждого пикселя требует чрезвычайно много времени. Поэтому поправка к частотным фильтрам находится лишь для центрального пикселя, а затем предполагается, что она слабо изменяется по полю радиоголограммы. Это негативно отражается на точности такого подхода, когда съемка ведется в широкой полосе. На краях полосы съемки шириной 50 км максимальная ошибка в законе изменения наклонной дальности приближается к миллиметру.

#### Кусочная модель изменения наклонной дальности

Чтобы обеспечить приемлемую точность аппроксимации временной зависимости наклонной дальности без введения в (1) каких-либо дополнительных членов и без выполнения численных расчетов спектра, в настоящей работе предлагается рассмотреть ошибку аппроксимации как кусочную функцию, состоящую из состыкованных ветвей двух парабол: с положительным и отрицательным старшим коэффициентом. При этом вершина обеих парабол приходится на точку  $\eta_{\perp}$  максимального сближения РСА и цели. Модули старшего коэффициента парабол выбираются различными, чтобы учесть нестрогую нечетность ошибки аппроксимации (1). Стыковка двух фрагментов кусочной функции в точке  $\eta_{\perp}$  позволяет элементарно найти точку их стыка в спектральной области: она приходится на нулевую доплеровскую частоту. При этом положительные доплеровские частоты соответствуют левой параболе, а отрицательные частоты – правой.

Как известно, аппроксимирующее выражение (1) само является квазипараболой с вершиной в точке максимального сближения. Поэтому уменьшить ошибку аппроксимации можно простой коррекцией параметра  $V_{xy}$ , который теперь будет иметь различные значения до и после прохождения точки максимального сближения РСА и цели:

$$V_{_{J3}}(\eta) = \begin{cases} V_{_{J31}}, & \eta - \eta_{\perp} \le 0, \\ V_{_{J32}}, & \eta - \eta_{\perp} > 0 \end{cases}$$
(2)

или в частотной области по азимуту:

$$V_{_{\Lambda_{9}}}(f_{\eta}) = \begin{cases} V_{_{\Lambda_{92}}}, & f_{\eta} < 0, \\ V_{_{\Lambda_{91}}}, & f_{\eta} \ge 0, \end{cases}$$
(3)

где *f<sub>n</sub>* – доплеровская частота.

Значения  $V_{x_{391}}$  и  $V_{x_{392}}$  находятся по МНК для соответствующих частей интервала синтеза апертуры. В результате ошибка аппроксимации уменьшается практически на порядок, что иллюстрирует рис. 2, и в пределе составляет десятые доли миллиметра. Таким образом, при доступном в настоящее время разрешении 2 дециметра и хуже предложенная кусочная аппроксимация вполне применима, в том числе при широкой полосе съемки, в отличие от подхода [3]. При дальнейшем же улучшении детальности радиолокационной съемки предложенная аппроксимации окажется недостаточно точной, но, если ее скомбинировать с подходом [3], необходимая точность будет обеспечиваться в широкой полосе съемки.

Основным преимуществом предлагаемой кусочной аппроксимации является сохранение неизменными всех частотных фильтров фокусировки и выражений для оценки миграции по отсчетам дальности. Единственное отличие фокусировки будет в том, что в положительной и отрицательной областях азимутального спектра в указанные выражения будут подставляться разные значения линейной эквивалентной скорости.



Рис. 2. Изменение погрешности кусочной аппроксимации с азимутальным временем

#### Учет движения РСА между излучением и приемом зондирующего сигнала

При синтезе радиолокационных изображений разрешением 1 м и хуже дальность наблюдаемой цели традиционно считалась одинаковой при излучении радиоимпульса и приеме эхо-сигнала, а пространственно протяженный импульсный отклик РСА на радиоголограмме представлялся как [4]

$$s_{u}(\tau,\eta) \approx w_{\tau} \left(\tau - \frac{2R(\eta)}{c}\right) w_{\eta}(\eta - \eta_{c}) \times \\ \times \exp\left(j\pi K_{\tau} \left(\tau - \frac{2R(\eta)}{c}\right)^{2}\right) \exp\left(-j \cdot 2\pi f_{0} \frac{2R(\eta)}{c}\right),$$
(4)

где  $w_{\tau}$  – огибающая сигнала по дальности,  $w_{\eta}$  – огибающая сигнала по азимуту,  $\eta_c$  – азимутальное время, когда антенна РСА направлена строго на цель,  $\tau$  – быстрое время,  $K_r$  – скорость изменения частоты излученного радиоимпульса,  $f_0$  – несущая частота, c – скорость света.

Именно из (4) и аппроксимации (1) выводились выражения для частотных фильтров, традиционно используемых при фокусировке. Однако при субметровом разрешении синтезируемых изображений аппроксимация (4) становится недопустимой, поэтому  $R(\eta)$  в ней необходимо заменить на среднее арифметическое дальностей в моменты излучения и приема:

$$\overline{R}(\eta) = \frac{R(\eta) + R\left(\eta - \frac{2R(\eta)}{c}\right)}{2}.$$
(5)

В (5) подразумевается, что шкала азимутального времени *η* соответствует приему сигнала.

В результате замены  $R(\eta)$  на  $\overline{R}(\eta)$  импульсный отклик РСА приобретает вид:

$$(\tau,\eta) \approx w_{\tau} \left(\tau - \frac{2R(\eta)}{c}\right) w_{\eta} (\eta - \eta_{c}) \times \\ \times \exp\left(j\pi K_{\tau} \left(\tau - \frac{2\overline{R}(\eta)}{c}\right)^{2}\right) \times \exp\left(-j \cdot 2\pi f_{0} \frac{2\overline{R}(\eta)}{c}\right),$$
(6)

Разность величин  $R(\eta)$  и  $\overline{R}(\eta)$  квазилинейно зависит от  $\eta$ , однако обращается в нуль не строго в момент максимального сближения РСА и цели, хотя и недалеко от него. Это приводит к азимутальному смещению и некоторому расширению сжатого отклика РСА на изображении, если строить фильтры фокусировки на базе (4), а не (6).

Выведем аналитическое выражение для зависимости  $\overline{R}(\eta)$  в случае, когда исходная зависимость  $R(\eta)$  аппроксимируется предложенной кусочной моделью.

Поскольку выражение (5) определяет  $\overline{R}(\eta)$  рекурсивно, модифицируем его:

$$\overline{R}(\eta) \approx \frac{R(\eta) + R\left(\eta - \frac{2R(\eta)}{c}\right)}{2}.$$
(7)

Погрешность аппроксимации (7) не превышает единиц микронов и на несколько порядков меньше длины волны РСА.

Для начала запишем выражение для  $\overline{R}(\eta)$ , если  $R(\eta)$  изменяется по закону (1):

$$\begin{split} \overline{R}(\eta) &= 0, 5 \left( \sqrt{R_{\perp}^{2} + V_{\scriptscriptstyle A3}^{2}} (\eta - \eta_{\perp})^{2} + \sqrt{R_{\perp}^{2} + V_{\scriptscriptstyle A3}^{2}} (\eta - \eta_{\perp})^{2} + \sqrt{R_{\perp}^{2} + V_{\scriptscriptstyle A3}^{2}} (\eta - \eta_{\perp})^{2} - \eta_{\perp} \right)^{2} \bigg) &= \\ &= \frac{R_{\perp}}{2} \left[ \sqrt{1 + \frac{V_{\scriptscriptstyle A3}^{2}}{R_{\perp}^{2}} (\eta - \eta_{\perp})^{2}} + \left( 1 + \frac{4V_{\scriptscriptstyle A3}^{2}}{c^{2}} - \frac{4V_{\scriptscriptstyle A3}^{2}}{cR_{\perp}} \sqrt{1 + \frac{V_{\scriptscriptstyle A3}^{2}}{R_{\perp}^{2}}} (\eta - \eta_{\perp})^{2} (\eta - \eta_{\perp}) + \left( 1 + \frac{4V_{\scriptscriptstyle A3}^{2}}{c^{2}} - \frac{4V_{\scriptscriptstyle A3}^{2}}{cR_{\perp}} \sqrt{1 + \frac{V_{\scriptscriptstyle A3}^{2}}{R_{\perp}^{2}}} (\eta - \eta_{\perp})^{2} \right)^{2} \right]. \end{split}$$

$$(8)$$

Приближенно просуммируем квадратные корни, разложив их в ряд Маклорена и отбросив все члены ряда, кроме нулевого и первого:

Дополним (9) недостающими членами до ряда Маклорена, перейдем от него обратно к квадратному корню и внесем множитель *R*<sub>1</sub> под знак корня:

$$\overline{R}(\eta) \approx \left( R_{\perp}^{2} - \frac{2V_{n_{2}}^{2}R_{\perp}}{c} \left( \eta - \eta_{\perp} - \frac{R_{\perp}}{c} \right) + V_{n_{2}}^{2} \left( 1 + \frac{2V_{n_{2}}^{2}}{c^{2}} \right) (\eta - \eta_{\perp})^{2} - \frac{V_{n_{2}}^{4}}{cR_{\perp}} (\eta - \eta_{\perp})^{3} \right)^{\frac{1}{2}}.$$
(10)

Погрешность полученной аппроксимации (10) составляет десятые доли микрона. Аналогичный вклад в значение  $\overline{R}(\eta)$  вносит множитель  $(1+2V_{_{RS}}^2/c^2)$ , поэтому его вполне можно опустить.

Следующим по малости является кубический член подкоренного выражения в (10). Его вклад в значение  $\overline{R}(\eta)$  составляет десятые доли миллиметра. Следовательно, пренебрегать им нежелательно. Однако, если его оставить, к выражению (10) не удастся строго применить метод стационарной фазы. Поэтому заменим указанный

кубический член близким по значению квадратичным членом, коэффициент которого найдем по МНК:

$$\overline{R}(\eta) \approx \left( R_{\perp}^{2} - \frac{2V_{\scriptscriptstyle R}^{2}R_{\perp}}{c} \left( \eta - \eta_{\perp} - \frac{R_{\perp}}{c} \right) + V_{\scriptscriptstyle R}^{2} \left( 1 \pm \frac{5V_{\scriptscriptstyle R}^{2}\Delta\eta}{6cR_{\perp}} \right) (\eta - \eta_{\perp})^{2} \right)^{\frac{1}{2}},$$
(11)

где  $\Delta \eta = T_a / 2$  – половина временного интервала синтеза апертуры  $T_a$ . Знак « $\pm$ » раскрывается в плюс при отрицательных значениях ( $\eta - \eta_{\perp}$ ) и в минус – при положительных.

Ошибка аппроксимации (11) не превышает десятой доли миллиметра (при разрешении 2 дециметра и хуже).

Выражение (11) отличается от (1) наличием линейного члена, влияющего лишь на положение сжатого импульсного отклика PCA, а также наличием множителя

 $\left(1\pm \frac{5V_{x_2}^2\Delta\eta}{6cR_{\perp}}
ight)$ , изменяющего частотную модуляцию сиг-

нала по азимуту, а следовательно, и выражения для построения фильтров фокусировки.

По аналогии с предложенной кусочной аппроксимацией для  $R(\eta)$  можно ввести кусочную аппроксимацию для  $\overline{R}(\eta)$ . При этом точка стыка сместится в положение

 $\underset{\eta}{\operatorname{argmin}} \overline{R}(\eta) = \eta_{\perp} + \frac{R}{c}.$  Одновременно учтем множитель

 $\left(1\pm \frac{5V_{_{\!\!M\!S}}^2\Delta\eta}{6cR_{_{\!\!L}}}
ight)$ , сместив для него точку перемены знака с

 $\eta_{\perp}$  на  $\eta_{\perp} + R/c$ . Такое смещение не приводит к значительной погрешности, поскольку при  $\eta$ , близких к  $\eta_{\perp}$ ,

влияние множителя 
$$\left(1\pm\frac{3V_{xy}\Delta\eta}{6cR_{\perp}}\right)$$
 пренебрежимо мало.

Введем заданное кусочно скорректированное значение линейной эквивалентной скорости:

(

$$\tilde{V}_{n,1}(\eta) = \begin{cases} V_{n,1}\sqrt{1 + \frac{5V_{n,2}^2\Delta\eta}{6cR_{\perp}}}, & \eta - \eta_{\perp} - \frac{R}{c} \le 0, \\ V_{n,2}\sqrt{1 - \frac{5V_{n,2}^2\Delta\eta}{6cR_{\perp}}}, & \eta - \eta_{\perp} - \frac{R}{c} > 0. \end{cases}$$
(12)

В частотной же области точка стыка останется на нуле Доплера:

$$\tilde{V}_{_{A3}}(f_{\eta}) = \begin{cases} V_{_{A32}}\sqrt{1 - \frac{5V_{_{A3}}^{2}\Delta\eta}{6cR_{\perp}}}, & f_{\eta} < 0, \\ V_{_{A31}}\sqrt{1 + \frac{5V_{_{A3}}^{2}\Delta\eta}{6cR_{\perp}}}, & f_{\eta} \ge 0. \end{cases}$$
(13)

Отличие  $\tilde{V}_{_{n_2}}(\eta)$  от  $V_{_{n_2}}(\eta)$  составляет тысячные доли метра в секунду и имеет значение только для квадратичного члена.

$$\overline{R}(\eta) \approx \sqrt{\begin{matrix} R_{\perp}^2 - \frac{2\widetilde{V}_{\scriptscriptstyle B^3}^2(\eta)R_{\perp}}{c} \times \\ \times \left(\eta - \eta_{\perp} - \frac{R_{\perp}}{c}\right) + \widetilde{V}_{\scriptscriptstyle B^3}^2(\eta)(\eta - \eta_{\perp})^2. \end{matrix}}$$
(14)

К выражению (14) применим метод стационарной фазы, с помощью которого можно найти выражения для частотных фильтров фокусировки. Рассматриваемые в настоящей работе эффекты практические не отражаются на сжатии по дальности и устранении ее миграции, а существенны лишь при сжатии по азимуту. Частотный фильтр сжатия по азимуту, полученный по методу стационарной фазы с учетом (6) и (14), имеет вид:

$$H_{a}(f_{\eta}) = \frac{1}{T_{a}\sqrt{|K_{a}|}} \exp\left(j\frac{\pi}{4}\right) \Pi\left(\frac{f_{\eta} - f_{\partial u}}{K_{a}T_{a}}\right) \times \exp\left(j \cdot 4\pi \frac{R_{\perp}\sqrt{c^{2} + \tilde{V}_{x3}^{2}(f_{\eta})}f_{0}\sqrt{1 - \left(\frac{cf_{\eta}}{2f_{0}\tilde{V}_{x3}(f_{\eta})}\right)^{2}}}{c^{2}}\right) \times (15) \times \exp\left(j \cdot 2\pi \frac{R_{\perp}}{c}f_{\eta}\right),$$

где  $K_a$  – средняя доплеровская скорость (первая производная доплеровской частоты по времени),  $f_{\partial y}$  – доплеровский центроид (средняя доплеровская частота).

#### Экспериментальные исследования

В настоящее время радиолокационные изображения с разрешением 2 дециметра позволяет получать только космический аппарат «TerraSAR-X». Исходные радиоголограммы для таких изображений не распространяются, поэтому экспериментальные исследования в настоящей работе выполняются на модельных данных. Поскольку эффекты миллиметровой погрешности  $R(\eta)$  отражаются лишь на сжатии по азимуту, исследуется лишь этот этап фокусировки. Таким образом, исходными данными выступают модельные импульсные отклики РСА, прошедшие процедуры сжатия по дальности и устранения ее миграции.

Сжатие по азимуту выполнялось:

а) с использованием классической модели (1);

б) с использованием точных фильтров, полученных численно;

в) с использованием фильтров, численная поправка к которым получена для точки, смещенной на 25 км по дальности;

г) с использованием предложенной модели.

Исходное положение модельного отклика выбиралось таким образом, чтобы после фокусировки он оказался в нулевом отсчете получаемого изображения. В этом случае фазовая составляющая азимутального спектра идеально сжатого отклика должна быть постоянной. Вариации фазы в азимутальном спектре говорят о неточной фокусировке.



Рис. 3. Графики фазы азимутального спектра импульсного отклика РСА, сжатого различными фильтрами

Графики фазы азимутального спектра импульсных откликов РСА, сжатых по азимуту с помощью исследуемых моделей представлены на рис. 1, буквенные обозначения на котором соответствуют списку моделей фокусировки, приведенному выше. Среднеквадратичные отклонения фазы от нуля, соответствующие графикам, следующие:

а) для классической модели (1) – 7,47°;

б) для точных фильтров, полученных численно, – 0°;

в) для фильтров, численная поправка к которым получена для точки, смещенной на 25 км по дальности, – 4,15°;

г) для предложенной модели – 2,22°.

#### Заключение

Таким образом, предложенная модель, оставаясь аппроксимационной, вносит небольшую ошибку в результат фокусировки, но эта ошибка оказывается в несколько раз меньше, чем для классической модели и для подхода [3] на краях широкой полосы съемки. Это позволяет ее применять для более точной фокусировки радиоголограмм с получением радиолокационных изображений дециметрового разрешения в случае съемки в широкой полосе захвата.

#### Литература

1. Кошляков Н.С., Глинер Э.Б., Смирнов М.М. Уравнения в частных производных математической физики. М.: Высшая школа, 1970. 712 с.

2. Kuang H., Chen J., Yang W., Zhu Y., Zhou J., Li Ch. Accurate compensation of stop-go approximation for high resolution spaceborne SAR using modified hyperbolic range equation // IEEE International Geoscience and Remote Sensing Symposium. 2014. Pp. 462–465.

3. Prats-Iraola P., Scheiber R., Rodriguez-Cassola M., Wollstadt S., Mittermayer J., Brautigam B., Schwerdt M., Reig-ber A., Moreira A. High precision SAR focusing of TerraSAR-X experimental staring spotlight data // IEEE International Geoscience and Remote Sensing Symposium. 2012. Pp. 3576-3579.

4. Cumming I.G., Wong F.H. Digital processing of synthetic aperture radar data: algorithms and implementation. Artech House, 2005. 660 p.

#### УДК 004.932

#### СУБПИКСЕЛЬНАЯ ИДЕНТИФИКАЦИЯ ОБЪЕКТОВ МЕСТНОСТИ ПО ГИПЕРСПЕКТРАЛЬНЫМ ДАННЫМ НА ОСНОВЕ ПОСЛЕДОВАТЕЛЬНОГО КВАДРАТИЧНОГО ПРОГРАММИРОВАНИЯ И МЕТОДА АНАЛИЗА ФОРМЫ СПЕКТРАЛЬНЫХ КОМПОНЕНТ

Григорьева О.В., старший научный сотрудник Военно-космической академии имени А.Ф. Можайского, к.т.н., alenka12003@mail.ru.

#### SUBPIXEL IDENTIFICATION OF OBJECTS BY MULTI-AND HYPERSPECTRAL DATA ON THE SEQUENTIAL QUADRATIC PROGRAMMING AND THE METHOD OF ANALYSES SPECTRAL COMPONENTS

#### Grigorieva O.V.

An paper show original ensemble algorithm of thematic treatment of hyperspectral data of remote sensing. The algorithm based on the use of sequential quadratic programming method to define spectral components of hyperspectral data, analysis of which allows to identification of objects of interest on subpixel level. For preparation informative features that needs for identification of objects make convolution of etalon spectra свертка by the methods of spatial-scalable filtering and sequence analysis. Demonstrate the example of verification algorithm for treatment of experiments aircraft hyperspectral data.

**Key word:** hyperspectral data, treatment, sequential quadratic programming, spectral components, spatial-scalable filtering, sequence analysis.

Ключевые слова: гиперспектральные данные, обработка, метод последовательного квадратичного программирования, спектральные компоненты, пространственно-масштабируемая фильтрация, секвентный анализ

#### Введение

Важным преимуществом гиперспектральных (ГС) данных дистанционного зондирования Земли является возможность использования при их тематической обработке не только топологических, но и спектрально-энергетических характеристик объектов интереса. Это обстоятельство

позволяет не только автоматизировать процесс анализа массивов ГС изображений, но и снизить требования к детальности съемки, так как дешифрирование при определенных условиях может производиться даже на «субпиксельном» уровне, когда в пиксель снимка попадает смесь объектов местности, размеры которых меньше, чем разрешение съемочной аппаратуры.

Как правило, в таких случаях предполагается, что спектральная смесь объектов линейная:

$$\mathbf{R} = \mathbf{M} \mathbf{S} + \boldsymbol{\varepsilon},\tag{1}$$

где  $M = \{m_{ij}\}, i = \overline{l, l}, j = \overline{l, p}$  – матрица смешивания, каждый столбец которой содержит спектральный вектор конечных членов (объектов), l – количество спектральных каналов, p – количество объектов;

 $S = \{s_{ij}\}, i = \overline{l,p}$  и  $j = \overline{l,n}$  – матрица, столбцы которой являются относительными распространенностями объектов, имеющих спектральные сигнатуры  $M_j$ , то есть элементы  $s_{ij}$  являются вероятностью отнесения

Рассматривается оригинальный ансамблевый алгоритм тематической обработки материалов гиперспектральной аэрокосмической съемки. Алгоритм предусматривает применение метода последовательного квадратичного программирования для определения спектральных компонент массива гиперспектральных данных, анализ которых позволяет распознавать объекты интереса на субпиксельном уровне. Для подготовки информативных признаков, необходимых для идентификации объектов, производится свертка эталонных спектральных сигнатур методами пространственномасштабируемой фильтрации и секвентного анализа. Приведены примеры верификации алгоритма при обработке экспериментальных гиперспектральных аэроснимков.

j-го пикселя к спектральной сигнатуре  $M_j$ , n – количество пикселей в анализируемом фрагменте ГС изображения;

 $R = \{r_{ij}\}, i = \overline{l,l}$  и  $j = \overline{l,n}$  – матрица, столбцы которой являются спектральными векторами анализируемого фрагмента ГС изображения;

ε – доля аддитивного шума.

В соответствии с формулой (1) субпиксельная идентификация заключается в определении весовых коэффициентов разложения, то есть доли объекта (фракции) в пикселе. Если для искомого объекта этот коэффициент превышает определенный порог, то пиксель идентифицируется как объект. При этом на коэффициенты смеси накладываются ограничения (сумма коэффициентов полагается равной единице, каждый из коэффициентов должен быть неотрицательным):

$$s_{ij} \ge 0 \ \text{in} \sum_{j=1}^{p} S_j = 1.$$
 (2)

Для решения уравнения (1) относительно коэффициентов линейной спектральной смеси используют методы линейного спектрального разложения (англ. linear spectral unmixing): метод минимизации среднеквадратического отклонения, метод ортогональной проекции подпространства и другие, различающиеся по критериям и методам оптимизации [1, 2]. Основные сложности, ограничивающие применение этих методов, связаны с необходимостью знания априорной информации о спектральных сигнатурах большинства объектов, присутствующих на ГС изображении.

Для преодоления этой проблемы автором предложен альтернативный вариант, предусматривающий использование ансамблевого алгоритма обработки с определением априорно неизвестных спектров объектов в результате анализа компонент ГС данных с помощью метода последовательного квадратичного программирования (англ. sequential quadratic programming SQP) [3, 4]. Для распознавания среди спектров той сигнатуры, которая соответствует объекту поиска, был специально разработан универсальный метод формализации эталонных спектральных характеристик (CX) объектов местности, полученных ГС съемочной аппаратурой, на основе пространственного-масштабируемой фильтрации (LOG-фильтрации) и секвентного анализа.

![](_page_26_Figure_3.jpeg)

Рис.1. Структурная схема алгоритма субпиксельной идентификации объектов местности по ГС данным

Структурная схема предложенного алгоритма приведена на рис. 1. На первом этапе обработки с целью сокращения анализируемого массива данных производится сегментация изображения с помощью различных индексов и метода пороговой обработки. Пороговые значения индексов выбираются в зависимости от того, к какому классу объектов окружающей среды относится объект (антропогенный, растительный или почвенный). Далее для получения матрицы спектральных сигнатур компонентов M, составляющих изображение, реализуется алгоритм SQP. Основной идеей метода SQP является последовательное решение задач квадратичного программирования, аппроксимирующих данную задачу оптимизации. SQP метод является одним из самых современных методов в области нелинейного программирования, позволяющим учесть полные ограничения, накладываемые на коэффициенты разложения смеси. Schittkowski K. успешно реализовал и провел тестовые расчеты с использованием SQP, показав его высокую эффективность для решения большинство задач [5].

На последнем этапе непосредственная идентификация спектральной компоненты, соответствующей искомому объекту, осуществляется с помощью разработанного метода формализации СХ. Преимуществом этого метода по сравнению с традиционно используемыми индексами является выявление информативных признаков объектов, основанных на смещении и локальных точках перегиба спектральной кривой целей, нехарактерной для фоновых объектов, и сокращение объемов обрабатываемой информации без уменьшения вероятности обнаружения цели.

#### Получение матрицы спектральных сигнатур компонентов изображения с помощью алгоритма последовательного квадратичного программирования

Для формирования задачи последовательного квадратичного программирования SQP формируется симплекс, содержащий спектральные векторы ГС изображения R. Вершинами симплекса являются спектральные сигнатуры каждого объекта  $M_j$ , которые оцениваются подгонкой симплекса минимального объема к набору спектральных векторов при ограничениях (2). Это формулируется как задача оптимизации:

$$-|\det Q| \to \min,$$
 (3)

при условиях  $QR \ge 0$  и  $(1_p)^T QR = 1_N$ , где  $Q = M^{-1}$ .

Важным преимуществом преобразования с помощью симплекса минимального объема является высокая устойчивость даже при отсутствии чистых спектральных векторов на изображении, что позволяет для выделения конечных членов фракции получить симплекс, очень близкий к истинному [6].

Для решения задачи нелинейной оптимизации с линейными ограничениями (3) с помощью SQP матрицу М необходимо привести к квадратному виду (размеру p×p) посредством сокращения спектральной размерности ГС данных или выполнить псевдообращение матрицы. Для уменьшения размерности может использоваться метод главных компонент, где идентификация информативных спектральных каналов в подпространстве сигнала более низкой размерности выбирается по максимуму вариации данных, что позволяет не только выделить полезный сигнал, но и сократить шум.

При этом решение будет единственным при равенстве ранга матрицы R значению p. C этой целью исходная матрица R также преобразуется в квадратную матрицу R<sub>p</sub>, которая будет состоять из p линейно независимых векторов. Для этого должен быть рассчитан базис матрицы для

ранга р или использована псевдообратная матрица R<sup>-1</sup>. В этом случае ограничение абсолютной суммируемости преобразуется следующим образом:

$$(l_p)^T Q = q$$
, rge  $q = (l_p)^T (R_p^{-1}) = (l_N)^T R^T (RR^T)^{-1}$  (4)

При реализации SQP условия ограничения в виде линейных равенств и неравенств должны быть формализованы в следующем виде:

 $Ax \leq b$  и  $A_{eq}x = b_{eq}$ .

В соответствии с этим, условие равенства преобразуется следующим образом:

$$\mathbf{A}_{eq} \cdot \operatorname{vec}(\mathbf{Q}) = b_{eq},\tag{5}$$

где  $A_{eq} = I_p \otimes (l_p)^T$ ,  $b_{eq} = q^T$ , vec(Q) – вектор, элементы которого являются элементами столбцов матрицы Q друг за другом;  $I_p$  – единичная матрица;  $\otimes$  – оператор тензорного произведения Кронекера для матриц.

Условие неравенства приводится к виду:

$$-A \cdot \operatorname{vec}(\mathbf{Q}) \le b, \tag{6}$$
  
rge  $\mathbf{A} = R^T \otimes I_p$  is  $\mathbf{b} = \mathbf{0}_{p^* p}.$ 

На каждой главной итерации метода SQP осуществляется квадратичная аппроксимация для функций Лагранжа при помощи квазиньютоновского модифицированного метода, для чего рассчитывается градиент  $\nabla f(Q)$  и Гессиан H целевой функции  $f(Q) = |\det Q|$ :

$$\nabla f(Q) = -(\mathbf{Q}^{-1})^{\mathrm{T}};$$

 $\mathbf{H} = (\nabla f(Q))^{\mathrm{T}} \nabla f(Q).$ 

Далее методом проекций решается задача квадратичного программирования на основе полученной квадратичной аппроксимации функции Лагранжа в точке  $(x_k, \lambda_k)$ :

$$\min_{\mathbf{d}} \frac{1}{2} \mathbf{d}^{\mathrm{T}} \mathbf{H}_{\mathbf{k}} \mathbf{d} + \nabla \mathbf{f} (\mathbf{x}_{\mathbf{k}})^{\mathrm{T}} \mathbf{d},$$
при условии
$$\nabla g_{1} (\mathbf{x}_{\mathbf{k}})^{\mathrm{T}} \mathbf{d} + g_{1} (\mathbf{x}_{\mathbf{k}}) = 0;$$
(7)

 $\nabla g_2(x_k)^T d + g_2(x_k) \le 0;$ 

где *H*<sub>*k*</sub> – матрица Гессе на k-ой основной итерации;

 $\nabla f(x_k)$  – градиент целевой функции на k-ой основной итерации;

 $g_1(x_k) = (1_p)^T x_k - (1_p)^T (R_p)^{-1}$  – значение функции ограничений в виде равенства в точке  $x_k$ ;

 $abla g_1(x_k) = \mathbf{A}_{eq} = I_p \otimes (\mathbf{1}_p)^{\mathrm{T}}$  – градиент функции  $g_1(x_k);$ 

 $g_2(x_k) = -x_k R$  – значение функции ограничений в виде неравенства в точке  $x_k$ ;

 $\nabla g_2(x_k) = \mathbf{A} = \mathbf{R}^T \otimes \mathbf{I}_p$  – градиент функции  $g_2(x_k)$ ;

d = x<sub>k</sub> – направление поиска на этапе решения задачи квадратичного программирования.

Поскольку ограничения  $g_1(x_k)$  и  $g_2(x_k)$  уже являются линейными, то для них линейная аппроксимация не нужна и условия оптимизации на данном этапе можно переписать в первоначальном виде (5) и (6).

На последнем этапе по полученному вектору поиска

d осуществляется расчет новой точки для следующей главной итерации с помощью функции выгоды.

В результате получаем матрицу спектральных сигнатур объектов, присутствующих на изображении *M*.

#### Методы анализа формы спектральных сигнатур

После получения матрицы М осуществляется идентификация спектральной компоненты, соответствующей искомому объекту, с помощью специальных методов анализа эталонных спектральных сигнатур, содержащихся в специализированной базе данных [7]. Для распознавания используются признаки, которые несут информацию о локальных изменениях спектральных характеристик (экстремумах, перегибах и др.), являющихся уникальными для рассматриваемого объекта.

Следует отметить, что производные спектральные признаки отличаются от исходных СХ не только меньшей размерностью, но и большей устойчивостью, так как на них не оказывают существенного влияния возможные вариации состояния (степени запыленности, увлажненности, шероховатости и др.) и, соответственно, отражательных свойств поверхности искомых объектов.

Минимизация размерности признакового пространства при обработке ГС данных реализуется двумя способами:

 выбором спектральных каналов в количестве, близком к существенной размерности функций, описывающих спектральные отражательные свойства объекта интереса;

 сверткой (формализацией) исходных СХ объектов в ограниченный набор признаков, характеризующих физическую природу потока излучения, отраженного от этого объекта.

В качестве методов формализации эталонных СХ, позволяющих сформировать новый вектор информативных признаков для идентификации объектов по данным ГС съемки, автором предложено использовать пространственно-масштабируемую фильтрацию и секвентный анализ исходных спектральных сигнатур [8].

#### Пространственно-масштабируемая фильтрация

Основная идея пространственно-масштабируемой фильтрации состоит в выявлении спектральных признаков, характеризующих количество и положение смещений и локальных точек перегиба (максимального наклона) в спектре исследуемых объектов. Расчет этих признаков осуществляется с помощью свертки спектра с фильтрами различной ширины.

После свертки функции исходного спектра R(x) с гауссовым фильтром g(x) заданной ширины (так называемым масштабом фильтрации о) выполняется вычисление второй производной, нулевые значения которой соответствуют наиболее значительным изменениям функции. Одновременно значения, которые относятся к более слабым изменениям, подверженным шумам, исчезают.

Свертка двух функций определяется как:

$$s(x) = R(x)g(x) = R(x)\left(\frac{1}{\sqrt{2\pi}}\right)e^{-x^2/2\sigma^2}$$
 (8)

Используя теорию линейных систем, можно показать, что вторая производная от свертки будет равна: s''(x) = R(x)g''(x), (9)

где g"(x) = 
$$\frac{x^2 - \sigma^2}{\sigma^4}$$
g(x).

٨

Далее производится нанесение переходов через ноль на графике зависимости величин среднеквадратического отклонения фильтра от длины волны.

Новыми спектральными признаками являются длины волн, на которых происходит переход графика второй производной свертки функции исходного спектра с гауссовым фильтром заданной ширины s"(x) через ноль. Как правило, при увеличении масштаба число пересечений графика второй производной свертки с осью x уменьшается. В качестве информативного признака при обнаружении объекта может рассматриваться также значение σ, при котором переход через ноль s"(x) объекта больше не наблюдается, если для фона этот переход прекращается при другом масштабе.

#### Секвентный анализ

В задаче подготовки формализованных спектральных признаков объектов и фонов перспективным является использование методов улучшения качества изображений, применяемых в пространственной и частотной областях. Особый интерес представляет преобразование исходной СХ методом секвентного анализа при помощи кусочно-постоянных функций. При определении информативных ГС признаков объектов секвентный анализ может быть применен к исходной спектральной сигнатуре объекта для получения частотного спектра, параметры которого будут являться новыми признаками для идентификации объекта.

В основе секвентного анализа лежит понятие секвенты, которая определяется как число изменений знака несинусоидальных функций за единицу пространства, времени или длины волны. В секвентном анализе вычислительно более простым является преобразование, построенное на системах кусочно-постоянных функций, в частности преобразование Уолша-Адамара, которое получается с помощью функционально идентичных ядер [9]:

$$wal(w,x) = (-1)^{\sum_{k=1}^{n} (w_{n-k+1} \otimes w_{n-k})x_k},$$
 (10)

где n – показатель степени в выражении  $N = 2^{n}$ ; N – количество отсчетов в секвенте; w – номер секвенты;

 $w_{_{n-k+1}} \otimes w_{_{n-k}}$  – сложение по модулю 2 номеров в двоичной системе;

x<sub>к</sub> – k-й разряд в представлении номера отсчета х в

двоичной системе счисления: 
$$x = \sum_{k=1}^{n} x_k 2^{n-k}$$

Приведем пример сложения по модулю 2 номеров в двоичной системе.

Пусть n = 1 и k = 3 для wal(i, x) = wal(n, x)wal(k, x), где  $i = n \otimes k$ , тогда  $n \otimes k = 01_2 \otimes 11_2 = 10_2 = 2$ .

В результате умножения получим:

				n
1	1	-1	-1	wal(1, x)
1	-1	1	-1	wal(3, x)
1	-1	-1	1	wal(2, x)

Это определение дает систему функций, упорядоченную по возрастанию секвенты (по Уолшу), то есть область нижних секвент спектра будут характеризовать медленно меняющиеся компоненты исходного сигнала, а область высоких секвент – резкие перепады яркости, как в случае гармонического анализа изображений (рис. 2) [10].

![](_page_28_Figure_17.jpeg)

Рис. 2. Функции, упорядоченные по Уолшу [11] Дискретное преобразование Уолша определяется как:

$$R_{w} = \sum_{i=0}^{n-1} R_{i} wal(w, i),$$
 (11)

 R<sub>i</sub> – исходные значения яркости спектра на i-ой длине волны;

R<sub>w</sub> – коэффициенты преобразования, число которых соответствует количеству длин волн.

Совокупность дискретных функций Уолша представляется в виде матрицы, в каждой строке которой указывается, сколько дискретных чисел берется при дискретизации точек на оси x. Например, матрица H(n), полученная при восьмиточечной дискретизации соответствующих непрерывных функций, где n – показатель степени в выражении  $N = 2^n$ , имеет вид:

	1	1	1	1	1	1	1	1	
	1	1	1	1	-1	-1	-1	-1	
	1	1	-1	-1	-1	-1	1	1	
I(n-2)	1	1	-1	-1	1	1	-1	-1	
1(1=5) =	1	-1	-1	1	1	-1	-1	-1	
	1	-1	-1	1	-1	1	1	-1	
	1	-1	1	-1	-1	1	-1	1	
	1	-1	1	-1	1	-1	1	-1	

ŀ

В матричном виде дискретное преобразование Уолша-Адамара описывается выражением:

 $R_{\rm w} = H(n)R$  . (12) В основе метода выбора спектральных признаков с ис-

В основе метода выоора спектральных признаков с использованием преобразования спектральной сигнатуры Уолша-Адамара лежит получение детализирующей информации о локальных особенностях функции спектрального образа в результате анализа коэффициентов R<sub>w</sub>.

Анализируя абсолютные значения коэффициентов дискретного преобразования, вычисленных на различных масштабах N, можно получить информацию о наличии и величине существенного изменения (скачка) функции спектрального образа R. В результате пространственно-масштабируемой фильтрации и секвентного анализа СХ эталона формируются информативные признаки объекта идентификации в виде следующих показателей:

 длины волн, на которых происходит переход графика второй производной свертки функции исходного спектра с гауссовым фильтром заданной ширины через ноль;

 – значение ширины фильтра, при котором переход через ноль графика объекта больше не наблюдается;

 абсолютные значения коэффициентов дискретного преобразования, вычисленные на различных масштабах секвенты.

По рассчитанным признакам осуществляется определение спектральной сигнатуры компонента изображения  $M_j$ , соответствующей объекту. Идентификация объекта осуществляется непосредственно на изображении найденной спектральной компоненты объекта методом пороговой обработки по значению коэффициента разложения смеси. При этом в программной реализации алгоритма предусмотрено, что оператор получает доступ к спектральным компонентам, которые являются относительными распространенностями всех объектов, присутствующих на ГС изображении.

#### Результаты апробации алгоритма и выводы

Для реализации рассмотренного в статье алгоритма субпиксельной идентификации был разработан специализированный модуль программы обработки данных аэрокосмической ГС съемки. На рис. 3 показан результат идентификации объекта в виде свалок бытовых отходов, отдельные из которых по причине небольших размеров визуально не обнаруживаются на панхроматическом изображении (см. рис. 3 г и рис. 3 д). В качестве исходных ГС изображений использовались данные, полученные с помощью видеоспектрометра разработки НПО «Лептон» (спектральное разрешение 0,4-13 нм, спектральный диапазон 402-1031 нм).

Результаты обработки этих данных показали, что применение предложенного ансамблевого алгоритма позволяет с вероятностью 0,88 определить свалки по идентифицируемой спектральной компоненте (см. рис. 2 а и рис. 2 б). При этом уменьшается количество ложных целей и пропусков в 1,5-2 раза по сравнению с традиционными методами линейного спектрального разложения (например, ортогональной проекции подпространства).

Идентификация объекта осуществляется непосредственно на изображении найденной спектральной компоненты объекта (рис. 3 б) методом пороговой обработки по значению коэффициента разложения смеси.

Спектральные признаки, по которым была найдена спектральная компонента, отображающая относительную распространенность искомого объекта, приведены на рис. 1 и в табл. 1. В качестве эталонных спектральных сигнатур для формирования значения признаков использовались СХ свалок различного морфологического состава, полученные в результате наземных измерений с помощью полевого спектрорадиометра FieldSpec®4

![](_page_29_Picture_11.jpeg)

в – график коэффициента спектральной яркости свалки, г – увеличенный фрагмент результата классификации,
 д – панхроматическое изображение увеличенного фрагмента)

![](_page_30_Figure_1.jpeg)

Рис. 4. Результат преобразования спектральной сигнатуры эталона и фона с помощью функции Уолша-Адамара

Таблица 1. Спектральные признаки эталона и фона, полученные

в результате пространственно-масштабируемой фильтрации при различных масштабах

Масштаб фильтрации	Свалка	Фон
$\sigma = 1$	539, 562, 572, 596, 613, 644, 706, 746, 767	557, 644, 674, 707, 735, 760
$\sigma = 1,5$	536, 649, 706, 749, 763	553, 644, 673, 715, 724, 762
$\sigma = 1,8$	671, 695	534

ASD. Пример эталонной сигнатуры приведен на рис. 3 в. Видно, что свалки и фон хорошо отличаются по значениям коэффициентов дискретного преобразования, вычисленных на различных масштабах секвенты (рис. 4), а также различаются по длинам волн, соответствующим ширине фильтра 1,8 при преобразовании исходных спектров методом пространственно-масштабируемой фильтрации (табл. 1).

В разработанной программе оператор в процессе подготовки эталонной информации для автоматического дешифрирования ГС снимков может самостоятельно подбирать наиболее информативный масштаб пространственно-масштабируемой фильтрации или при использовании секвентного анализа задавать пороговое значение для исключения амплитуд, соответствующих малозначимых изменениям исходной сигнатуры.

#### Заключение

Предложен ансамблевый алгоритм субпиксельной идентификации объектов местности по гиперспектральным данным с использованием LOG-фильтрации и секвентного анализа спектральных компонент ГС данных, инициализированных в результате линейного разделения смеси методом SQP. Метод позволяет снять ограничения о присутствии чистых спектральных векторов на изображении, учесть полные ограничения, накладываемые на коэффициенты разложения смеси, и требует знания априорной информации о спектральной сигнатуре только искомого объекта. Оценка точности идентификации объектов также повышается за счет использования информативных признаков формы спектральных компонент объекта.

#### Литература

1. J.C. Harsanyi, C.-I. Chang Hyperspectral Image Classification and Dimensionality Reduction: An Orthogonal Subspace Projection Ap-proach // IEEE Trans. Geosci. Remote Sens. 32 (4), 1994. PP. 779-785. 2. S. Kraut, L.L. Scharf, L.T. McWhorter Adaptive Subspace Detectors // IEEE Trans. Signal Process. 49 (1). 2001. Pp. 1-16.

 Гладких Б.А. Методы оптимизации и исследование операций для бакалавров информатики. Ч. II. Нелинейное и динамическое программирование: учебное пособие. – Томск: Изд-во НТЛ, 2011. – 264 с.

4. Klaus Schittkowski, Ya-xiang Yuan Sequential Quadratic Programming Methods. Wiley Encyclopedia of Operations Research and Management Science. 2011.

5. K.Schittkowski More test examples for nonlinear programming // Lecture notes in Economics and Mathematical Systems, Vol. 282, Springer. 1987.

6. Li J., Bioucas-Dias J.M. Minimum volume simplex ana-lysis: a fast algorithm to unmix hyperspectral data // Proc. of IEEE International Conference on Geoscience and Remote Sensing Simposium (IGARSS). Boston, USA: IEEE. 2008. V. 3. P. 250-253.

7. Свидетельство о государственной регистрации базы данных № 2012621165 «База данных КСЯ в спектральном диапазоне 400...1000 нм для основных типов подстилающей поверхности (почв, растительности, ледовых и снежных покровов, водной поверхности, объектов искусственного происхождения), встречающихся в районах съемки КА «Ресурс-П», авторы: Чапурский Л.И., Марков А.В., Григорьева О.В., Саидов А.Г., Астахова Е.И., Жуков Д.В., заявка № 2012620611, дата поступления 21.06.2012 г., дата гос. регистрации в Реестре баз данных – 13.11.2012 г.

8. Григорьева О.В., Иванец М.О., Марков А.В., Жуков Д.В. Методы подготовки формализованных эталонных признаков для идентификации объектов по данным гиперспектральной космической съемки // Материалы V Всероссийской научно-технической конференции «Актуальные проблемы ракетно-космической техники» (V Козловские чтения) / под общ. ред. А.Н. Кирилина. – Самара: СамНЦ РАН, 2017. Т. 1. С. 281-286.

9. Гонсалес Р., Вудс Р. Цифровая обработка изображений. – М., 2005. – 1072 с.

10. Костров Б.В. Теория и методология применения секвентного анализа для обработки аэрокосмических изображений. Автореферат диссертации на соискание ученой степени доктора технических наук.

11. Залманзон Л.А. Преобразования Фурье, Уолша, Хаара и их применение в управлении, связи и других областях. – М.: Наука. Гл. ред.физ.-мат.лит., 1989. – 496 с.

#### УДК 004.627

#### ПРИМЕНЕНИЯ КОМПЛЕКСНОГО ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЯ В ЗАДАЧЕ СЖАТИЯ ВИДЕО

Дам Чонг Нам, аспирант Московского физико-технического института (государственного университета), chong.dam@phystech.edu.

#### APPLICATIONS OF COMPLEX WAVELET TRANSFORM IN VIDEO COMPRESSION

#### **Dam Trong Nam**

In presented paper, the principles of inter-frame coding are shortly given on the example of wavelet-based video coder. The construction and properties of discrete wavelet transform (CWT) and complex wavelet transform (CWT) which is based on the dual tree complex wavelet transform (DTCWT) are considered. Motion compensation method for real videos using CWT is proposed. Comparisons between proposed method and reference methods which are based on using value of luma component (block matching (BM) and overlapped block matching (OBM)) are presented. The experiments with real high-definition videos show that proposed method gives higher PSNR of prediction error (from 0,2 to 0,5 dB) and smaller compressed output for quantized prediction error (from 6 % to 25 %) over reference methods even in conditions when reference methods work well.

**Key words:** inter-frame coding, DWT, CWT, DTCWT, motion compensation, block matching, overlapped block matching.

Ключевые слова: межкадровое кодирование, ДВП, КВП, ДДКВП, компенсация движения, метод сопоставления блоков, метод сопоставления перекрывающихся блоков.

#### Введение

В большинстве современных видеокодеков используется дискретное косинусное преобразование (ДКП), характерными недостатками которого являются блокинг-эффект и мозаичный эффект. В отличие от ДКП, дискретное вейвлет-преобразование (ДВП) работает не с блоками маленького размера, а с тайлами и даже с кадрами, вследствие чего ДВП лишено таких дефектов [1]. Следовательно, примене-

ние ДВП для компрессии статических и подвижных изображений перспективно. Действительно, на данный момент алгоритмы ДВП представлены в широко распространенных системах компьютерной математики, таких как Mathcad, MATLAB, Mathematica, более того, ДВП применяется в международном стандарте сжатия статических изображений JPEG2000 [2] и стандарте сжатия подвижных изображений MPEG-4 [3]. При этом используется двухканальное вейвлет-разложение. Применение многоканального вейвлет-разложения [4] обеспечит более высокие степени сжатия при заданном качестве восстановленного изображения.

Однако, при сжатии подвижных изображений и двухканальное, и многоканальное ДВП, основанные на действительных вейвлет-коэффициентах, имеют ряд недостатков. Первым недостатком является неинвариантность к сдвигу, состоящая в том, что малейшие смещения изображения в кадре приводят к значительным изменениям значений вейвлет-коэффициентов, что не позволяет использовать вейвлет-коэффициенты для процедуры анализа и компенсации движения, являю-

Приведен принцип межкадрового кодирования на примере вейвлет-видеокодера. Рассмотрены построение и свойства дискретного вейвлет-преобразования (ДВП) и комплексного вейвлетпреобразования (КВП), основанного на двойном дереве комплексного вейвлет-преобразования (ДДКВП). Предложен метод компенсации движения реальных видео с помощью комплексного вейвлетпреобразования (КВП). Приведено сравнение предложенного метода с эталонными методами – методом сопоставления блоков (СБ) и методом сопоставления перекрывающих блоков (СПБ). Экспериментами с реальными видео высокого разрешения показано, что предложенный метод выигрывает у эталонных методов от 0,2 до 0,5 дБ PSNR остатка предсказания и от 6 % до 25 % объема сжатого потока для квантованного остатка предсказания в режимах, когда эталонные методы хорошо работают.

> щей очень важной в видеокодеках для уменьшения временной избыточности, и тем самым увеличения степени сжатия. Вторым недостатком является отсутствие концепции фазовых соотношений, как в случае преобразования Фурье, которая может предоставить более точное описание и возможность компенсации движения. К счастью, широко известное в задаче обработки сигналов преобразование Фурье имеет привлекательные свойства для задачи сжатия изображений, такие как простое свойство линейности фазовых соотношений, кодирующих смещения исходного сигнала, и гладкость амплитуды, инвариантной к сдвигу [5]. Комплексное вейвлетпреобразование (КВП), вдохновленное преобразованием Фурье, является мощным фактором решения упомянутых проблем ДВП в задаче сжатия подвижных изображений.

> Имеются исследования ортогональных или биортогональных КВП [6, 7]. В работе [8] предложен метод анализа движения с использованием КВП, не являющегося обратимым, что не позволяет использовать его для по

строения предсказанного кадра. В [9] указано, что проводились исследования [10, 11] по построению обратимого КВП из полиномов Добеши. Однако, полученные вейвлет-фильтры в [10, 11] не обладают хорошими импульсными характеристиками, обеспечивающими инвариантность к сдвигу. Также в [9] показано, что очень трудно построить обратимое КВП, основанное на комплексных фильтрах, с привлекательными импульсными характеристиками для всех уровней разложения. Более того, легко заметить, что вычисления с комплексными коэффициентами сложнее и дольше, так что за последнее двадцатилетие существует единственное активное направление исследования КВП – КВП на основе двойного дерева комплексного вейвлет-преобразования (ДДКВП) [12, 13], которое в одномерном случае состоит из двух действительных ортогональных или биортогональных вейвлет-базисов, один из которых отвечает за «действительную часть», а второй за «мнимую часть», то есть, с одной стороны используются действительные вейвлет-коэффициенты, а с другой - вейвлет-преобразование носит комплексный характер.

Для обработки изображений КВП, в основном, применяется в задаче шумоподавления [14, 15]. КВП также находит применение в задачах классификации [16] и сегментации [17]. В задаче сжатия видео имеется крайне мало работ по использованию КВП [18], что объясняет новизну исследования КВП для задачи сжатия видео.

В данной работе рассмотрено применение КВП для компенсации движения реальных видео.

Цель работы – предложить метод построения предсказания кадров реальных видео в пространстве КВП (амплитуды и фазы) по найденному другим методом набору векторов движения и оценить эффективность предложенного метода в сравнении с эталонными методами, основанными на использовании яркостной составляющей кадра – методом сопоставления блоков (СБ) (block matching – BM) [1] и методом сопоставления перекрывающих блоков (СПБ) (overlapped block matc-hing – OBM), используемым в видеокодеке Dirac [19].

Для достижения поставленной цели в работе реализованы на языке программирования C++ решения следующих задач – чтения видео, поиска векторов движения с точностью до четверти пиксела методом полного перебора, построения предсказанных кадров разными методами с помощью опорного кадра и одним найденным набором векторов движения, записи предсказанного кадров и остатков предсказания для визуального сравнения методов предсказания, и сравнения эффективности методов разными метриками.

На реальных видео разного типа показано, что по сравнению с эталонными методами предложенный метод обеспечивает точное предсказание и меньший сжатый поток данных при одинаковом качестве восстановленного кадра. Преимущества данного метода заключаются в переходе из пространства яркости в пространство ДДКВП (амплитуды и фазы), легкой реализации с возможностью быстрого вычисления как в случае ДВП.

#### Общий принцип межкадрового кодирования

Для уменьшения временной избыточности в современных видеокодеках используется процедура анализа и компенсации движения. В качестве примера на рис. 1 приведена общая схема вейвлет-видеокодера.

![](_page_32_Figure_9.jpeg)

Рис. 1. Общая схема вейвлет-видеокодера

Текущий кадр (Y cur) поступает на вход блока анализа движения (далее БАД). Получив опорный кадр (Y ref) из буфера опорных кадров и текущий кадр, БАД определяет набор векторов движения (MV), описывающих перемещения блоков текущего кадра относительно опорного кадра и передающихся на блок арифметического кодирования. С помощью набора векторов движения и опорного кадра блок компенсации движения (далее БКД) строит предсказанный кадр, который нужно вычесть из текущего кадра для получения остатка предсказания, необходимого для дальнейшей обработки и передачи. По сравнению с полным текущим кадром, энергия остатка предсказания небольшая, однако, мы можем еще больше сократить объем передаваемых данных путем последовательного применения к остатку предсказания преобразования (дискретного косинусного преобразования, дискретного синусного преобразования или других ортогональных преобразований [1] в общем случае или ДВП в данном случае) и квантования. Затем квантованный остаток предсказания (далее КОП) и векторы движения поступают на блок арифметического кодирования для получения сжатого потока данных, который нужно передавать в канал связи или сохранять на каком-либо носителе.

На стороне декодера полученный сжатый поток последовательно проходит через блок арифметического декодирования, блок обратного квантования и блок обратного преобразования (ДВП в этом случае) для получения набора векторов движения и остатка предсказания, искаженного из-за квантования. В декодере недоступны исходные видеокадры, вследствие чего декодированные заранее кадры используются как опорные. Получив набор векторов движения, на основе буфера декодированных ранее кадров декодер реконструирует предсказанный кадр и прибавляет декодированный искаженный остаток предсказания для получения восстановленного кадра, который записывается в буфер декодированных кадров. Нужно отметить, что восстановленный кадр отличается от текущего кадра, доступного в кодере, вследствие этого в кодере необходимо повторить такой же процесс декодирования для получения таких восстановленных опорных кадров вместо текущих кадров.

Практической целью данной работы является разработка нового метода компенсации движения в БКД с использованием КВП. Для сравнения эффективности предложенного метода с существующими методами также реализованы в БКД наиболее популярные методы компенсации движения – метод СБ и метод СПБ.

#### Метод построения ДВП

На практике ДВП реализуется с помощью банка фильтров, основанных на двух фильтрах: нижних (ФНЧ) и верхних частот (ФВЧ) [1, 2, 5]. Схема построения ДВП для двумерного случая представлена на рис. 2 (Здесь  $h_0(n) - \Phi$ HЧ,  $h_1(n) - \Phi$ BЧ,  $\Phi$  – масштабирующая функция,  $\Psi$  – вейвлет-функция).

![](_page_33_Figure_4.jpeg)

![](_page_33_Figure_5.jpeg)

Стандартное действительное двумерное ДВП получается с помощью тензорного произведения одномерных ДВП по горизонтали и по вертикали, в результате чего получены масштабирующая функция  $\Phi(x, y) = \Phi(x)\Phi(y)$  и три вейвлет-функции  $\Psi_1(x, y) = \Phi(x)\Psi(y)$ ,  $\Psi_2(x, y) = \Psi(x)\Phi(y)$ ,  $\Psi_3(x, y) = \Psi(x)\Psi(y)$  (см. рис. 2).

При сжатии подвижных изображений ДВП, основанное на действительных вейвлет-коэффициентах, имеет недостатки – отсутствие концепции фазовых соотношений и неинвариантность к сдвигу, состоящая в том, что малейшие смещения изображения в кадре приводят к значительным изменениям значений вейвлет-коэффициентов, что не позволяет использовать их для процедуры анализа и компенсации движения, являющей очень важной в видеокодеках для уменьшения временной избыточности, и тем самым увеличения степени сжатия. В отличии от ДВП, КВП может решить эти проблемы и КВП легко реализуется на основе ДДКВП [8, 12].

#### Метод построения КВП, основанного на ДДКВП

Одномерное ДДКВП обеспечивает разложение действительного сигнала с помощью двух комплексных масштабирующих и вейвлет-функций, которые получаются из двух независимых банков фильтров [8, 12] (см. рис. 3). Обозначим  $\Phi_h(t)$  и  $\Psi_h(t)$  масштабирующую и вейвлет-функцию соответственно,  $c_{h_{L,s}}$  и  $d_{h_{l,s}}$  – соответствующие коэффициенты, где h указывает определенный набор вейвлет-фильтров. Вейвлет-функции  $\Psi_h(t)$  и  $\Psi_g(t)$  из двойного дерева играют роль действительной и мнимой частей комплексного аналитического вейвлета  $\Psi^C(t) = \Psi_h(t) + j\Psi_g(t)$ . Мнимый вейвлет  $\Psi_g(t)$  приблизительно является преобразованием Гильберта действительного вейвлета  $\Psi_h(t)$ .

![](_page_33_Figure_12.jpeg)

на основе пары банков фильтров

Схема построения двумерного ДДКВП представлена на рис. 4.

![](_page_33_Figure_15.jpeg)

Рис. 4. Схема построения двумерного ДДКВП первого уровня Здесь h0(n), h1(n) – ФНЧ и ФВЧ для верхней ветви, g0(n), g1(n) – ФНЧ и ФВЧ для нижней ветви. Нужно отметить, что для первого уровня может быть использован любой дискретный банк вейвлет-фильтров, т.е. h0(n), h1(n) и g0(n), g1(n) совпадают с условием, что прорежи-

вание второй ветви смещено на один отчет по сравнению с первой ветвью – выполняется условие задержки одного отчета [12]. В работе рассмотрено ДДКВП первого уровня в связи с избыточностью и сложной реализацией для следующих уровней. Сначала применяются ФНЧ и ФВЧ для верхней и нижней ветви со смещением прореживания к строкам данных, в результате чего получено два массива данных, к столбцам каждого из которых применив еще раз такую процедуру мы получим четыре массива данных (Y1, Y2, Y3, Y4) (см. рис. 4).

По аналогии с ДВП, ДДКВП может быть построено двумя масштабирующими функциями  $\Phi(x)\Phi(y)$ ,  $\Phi(x)\Phi^{*}(y)$ , и шестью следующими вейвлетфункциями:  $\Phi(x)\Psi(y)$ ,  $\Psi(x)\Phi(y)$ ,  $\Psi(x)\Psi(y)$ ,  $\Psi(x)\Psi^{*}(y)$ , где  $\Phi^{*}(y)$  и  $\Phi(x)\Psi^{*}(y)$  $\Psi(x)\Phi^{*}(y),$  $\Psi^*(y)$  комплексно-сопряженные к  $\Phi(y)$  и  $\Psi(y)$ , и  $\Phi(x)$ ,  $\Phi(y)$ ,  $\Psi(x)$ ,  $\Psi(y)$ комплексны, т.е.  $\Phi(x) = \Phi_{\mu}(x) + j\Phi_{\sigma}(x),$  $\Phi(y) = = \Phi_h(y) + j\Phi_g(y),$  $\Psi(x) = \Psi_{\mu}(x) + j\Psi_{\mu}(x), \quad \Psi(y) = -\Psi_{\mu}(y) + j\Psi_{\mu}(y).$ Подставив комплексное представление и выделив действительную и мнимую части каждого из шести вейвлетов, получим из четырех массивов данных (У1, У2, У3, Y4) все комплексные коэффициенты ДДКВП следую-

щим образом:  $Y_re1=(Y1 + Y4)$   $Y_im1=(Y3 - Y2)$   $Y_re2=(Y1 - Y4)$   $Y_im2=(Y3 + Y2)$ После нормировки  $Y_re1=(Y1 + Y4)/2$   $Y_im1=(Y3 - Y2)/2$  $Y_re2=(Y1 - Y4)/2$   $Y_im2=(Y3 + Y2)/2$ 

Из двух комплексных коэффициентов (Y\_re1, Y\_im1) и (Y\_re2, Y\_im2) вычисляются амплитуды и фазы (A1, Ph1) и (A2, Ph2), с которыми выполняется процесс предсказания кадра в предложенном методе.

#### Предложенный метод компенсации движения

Схема предложенного алгоритма построения предсказанного кадра реальных видео показана на рис. 5.

![](_page_34_Figure_6.jpeg)

Рис. 5. Предложенный алгоритм построения предсказанного кадра

На вход поступают кадры, если этот кадр – опорный, то сохраняются амплитуды и фазы всех возможных смещенных кадров с точностью до четверти пиксела, иначе выполняется процесс нахождения векторов движения с точностью до четверти пиксела методом СБ с помощью алгоритма полного перебора. Затем на основе найденных векторов движения и буфера наборов амплитуд и фаз определяются амплитуды и фазы каждого блока предсказанного кадра. После чего вычисляется предсказанный кадр обратным ДДКВП. Блок «оценки качества предсказания» получает предсказанный кадр, вычисляет метрики для сравнения и записывает результаты в файл оценок.

#### Результаты экспериментальных исследований

В этой работе кадр номера 0 выбран в качестве опорного кадра, предсказание кадра выполняется для следующих кадров и качество предсказания оценивается метрикой PSNR остатка предсказания и энтропии КОП.

Метод СБ реализован с алгоритмом полного перебора с маленьким размером блока 4х4 и 8х8, что позволяет определить точные векторы движения с точностью до четверти пиксела. При этом яркость смешенного на долю пиксела определяется с помощью возьмиточечной линейной интерполяции, используемой в стандарте H.265 для интерполяции яркости (H265L) [20, 21].

Метод СПБ [19] реализован с перекрывающими блоками размера 6x6 и 12x12 с соответствующими окнами [3 5 8 8 5 3]/8 и [1 3 5 7 8 8 8 8 7 5 3 1]/8.

Как сказано выше, в предложенном методе можно использовать любой вейвлет-фильтр для построения ДДКВП. В работе исследованы разные вейвлеты для построения ДДКВП, такие как Legal5\_3, DD9\_7, CDF9\_7, DD13\_7 и другие. Результаты исследования показывают, что на качество предсказания очень мало влияет тип выбранного вейвлета и оказывается, что для качества предсказания и быстродействия лучше всего выбрать короткий фильтр Legal5\_3.

Если в качестве вейвлета выбран наилучший для предложенного метода фильтр Legal5\_3, то проводя оценку предсказания для нескольких кадров получим зависимость PSNR остатка предсказания от номера предсказанного кадра. В качестве примера на рис. 6 приведена эта зависимость для режима разбиения блока 4х4 и точности вектора движения четверти пиксела для видео «City» и «Stockholm».

Как показано на рис. 6, предложенный метод лучше двух эталонных методов (от 0,2 до 0,5 дБ по энергии остатка предсказания) даже в режиме высокой точности вектора движения (четверти пиксела) и маленького размера разбиваемых блоков (4х4), когда отлично работают эталонные методы.

На степень сжатия также влияет распределение КОП. Чтобы определить предел сжатия КОП арифметическим кодированием в разных методах, была исследована зависимость энтропии КОП от PSNR остаточного кадра, являющего разностью текущего и восстановленного кадров. В качестве примеров на рис. 7 и 8 показана эта зависимость для видео «City» и «Stockholm» и кадров номер 1 и 2, когда оба эталонных метода хорошо работают.

![](_page_35_Figure_0.jpeg)

Рис. 6. Сравнение PSNR остатка предсказания в режиме разбиения блока 4x4 и точности вектора движения четверти пиксела для видео «City» (левый) «Stockholm» (правый)

![](_page_35_Figure_2.jpeg)

Рис. 7. Зависимость энтропии КОП от PSNR остаточного кадра для 1-ого кадра в режиме разбиения блока 4x4, точности четверти пиксела для видео «City» (левый) и «Stockholm» (правый)

![](_page_35_Figure_4.jpeg)

Рис. 8. Зависимость энтропии КОП от PSNR остаточного кадра для 2-ого кадра в режиме разбиения блока 4x4, точности четверти пиксела для видео «City» (левый) и «Stockholm» (правый)

На рис. 7, 8 видно, что при одинаковом качестве восстановленного кадра, т.е. при одном значении PSNR остаточного кадра, КОП по предложенному методу обладает самой минимальной энтропией.

Для количественного сравнения в работе при одинаковом качестве восстановления кадра также вычислено отношение энтропии квантованного остатка предсказания в предложенном методе к такой энтропии в эталонном методе, который дает минимальную энтропию из двух эталонных методов. Сравнительный анализ показывает, что при маленьком отклонении между текущим (кадр 1 или кадр 2) и опорным кадром (кадр 0), когда отлично работают эталонные методы, предложенный метод может уменьшить объем сжатого потока данных для квантованного остатка предсказания от 20 до 25 % для видео «City» и от 6 до 8 % для видео «Stockholm». Следовательно, предложенный метод лучше подходит для задачи сжатия видео, чем эталонные методы.
### Заключение

В работе предложен метод предсказания кадра с использованием ДДКВП. Для проверки работоспособности и оценки эффективности предложенного метода была разработана экспериментальная программа на C++, выполняющая чтение видео, поиск векторов движения с точностью до четверти пиксела методом полного перебора, построение предсказанных кадров разными методами с помощью опорного кадра и одним найденным набором векторов движения, запись предсказанного кадров и остатков предсказания для визуального сравнения методов предсказания, и сравнение эффективности методов разными метриками.

В ходе работы с видео высокого разрешения «city» и «stockholm» (https://media.xiph.org/video/derf/) были получены следующие основные результаты:

 Качество предсказания предложенного метода мало зависит от типа вейвлета, что позволит построить ДДКВП любым вейвлет-фильтром и легко адаптировать предложенный метод в существующих вейвлетвидеокодеках.

– Предложенный метод обеспечивает меньшую энергию остатка предсказания. Конкретно, если отклонение между текущим и опорным кадром небольшое, то предложенный метод выигрывает у эталонных методов от 0,2 до 0,5 дБ PSNR остатка предсказания, в других случаях предложенный метод выигрывает больше 0,5 дБ PSNR остатка предсказания.

– При одинаковом качестве восстановленного кадра предложенный метод снижает объем сжатого потока для КОП от 20 до 25 % видео «City» и от 6 до 8 % видео «Stockholm». Следовательно, предложенный метод лучше подходит для сжатия чем эталонные методы.

## Литература

1. Дворкович В.П., Дворкович А.В. Цифровые видеоинформационные системы (теория и практика). Москва: Техносфера, 2012. – 1008 с. ISBN 978-5-94836-336-3.

2. ITU-T Recommendation T.800 (2002-08) – Information technology – JPEG 2000 image coding system: Core coding system.

3. ISO/IEC 14496-2 (Second edition 2001-12-01) – Information technology – Coding of audio-visual objects – Part 2: Visual.

4. Дворкович А.В., Дворкович В.П. Оконные функции для гармонического анализа сигналов. Москва: Техносфера, 2016. – 208 с. ISBN 987-5-94836-432-2.

5. Сергиенко А.Б. Цифровая обработка сигналов. – СПб.: Питер, 2002. – 608 с. ISBN 5-318-00666-9.

6. Lina J.M., Mayrand M. Complex Daubechies wavelets // Appl. Comput. Harmon. Anal., vol. 2, no. 3, pp. 219–229, 1995.

7. Spaendonck R.V., Blu T., Baraniuk R. and Vetterli M.

«Orthogonal Hilbert transform filter banks and wavelets».// Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, Apr. 6-10, 2003, vol. 6.

8. Magarey J.and Kingsbury N.G. Motion Estimation Using a Complex-Valued Wavelet Transform. // IEEE Trans. on Signal Processing, 46(4):1069-1084, 1998.

9. Kingsbury N.G. «Image processing with complex wavelets». // Phil. Trans. Royal Society London A, September 1999, Special issue for the discussion meeting on «Wavelets: the key to intermittent information?» (held Feb 24-25, 1999), 357, pp 2543-2560

10. Lawton W. Applications of complex valued wavelet transforms to subband decomposition. IEEE Trans. Signal Proc. 41, 3566-3568.

11. Belzer B., Lina J.M., and Villasenor J. Complex, linear-phase filters for efficient image coding. // IEEE Transactions on Signal Processing, 40(4):2425–2427.

12. Selesnick, R. Baraniuk, and N. Kingsbury «The dualtree complex wavelet transform». // IEEE Signal Process. Mag., vol. 22, no. 6, pp. 123-151, Nov. 2005.

13. Selesnick I. «Matlab Implementation of Wavelet Transforms. Dual-Tree Complex Wavelet Transform» // http://taco.poly.edu/WaveletSoftware.

14. Naga Prudhvi Raj, V., Venkateswarlu. Denoising of medical images using dual tree complex Wavelet transform. // Proc. Technol.4, 238–244, C3IT- 2012.

15. Shi F. and Selesnick I.W. Video denoising using oriented complex wavelet transforms. // Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP), June 2004, vol.2, pp.949-952.

16. Abdul Rehman, Yang Gao, Jiheng Wang, Zhou Wang, Image Classication Based on Complex Wavelet Structural Similarity, IEEE International Conference on Image Processing, Brussels, Belgium, Sept. 2011.

17. An Vo, Soontorn Oraintara, A study of relative phase in complex wavelet domain: Property, statistics and applications in texture image retrieval and segmentation. // Signal Processing: Image Communication 25, 2010, pp.28-46.

18. Unan Y. Oktiawati and Vooi Voon Yap, A Motion Estimation Algorithm Using DTCWT. // ITB J. ICT, vol.6, no.1, 2012, pp.82-101.

19. Jianguo Zhang Ling Shao Lei Zhang Graeme A. Jones. Intelligent video event analysis and understanding. Springer, 2010 edition. - 251 pages. ISBN 978-3-642-17554-1.

Sullivan G.J., Jens-Rainer Ohm, Woo-Jin Han, Thomas Wiegand. Overview of the high efficiency video coding (HEVC) standard. // IEEE Trans. Circ. Syst. Video Technol. 22(12). 1649-1668 (2012).

20. Jill M. Boyce, Yan Ye, Jianle Chen, Adarsh K. Ramasubramonian. Overview of SHVC: Scalable Extensions of the High Efficiency Video Coding Standard. // IEEE Trans. Circuits Syst. Video Techn. 26(1): 20-34 (2016). УДК 004.627

# ОПТИМИЗАЦИЯ МЕТОДА КВАНТОВАНИЯ ДЛЯ ВЕЙВЛЕТ-ВИДЕОКОДЕКА

Дам Чонг Нам, аспирант Московского физико-технического института (государственного университета), chong.dam@phystech.edu.

# **OPTIMIZATION OF QUANTIZATION METHOD FOR WAVELET-BASED VIDEO CODEC**

# **Dam Trong Nam**

This paper is dedicated to the research and development of quantization methods for wavelet-based video codec. The presented paper investigates uniform, nonlinear and entropy-constrained quantization methods and their potential applications for wavelet-based video codec Dirac. Based on analyzing the processing results of various video types using quantization methods mentioned above, the optimized quantization method for wavelet-based video codec was proposed. The output data rate for given distortion levels is reduced noticeably by using the proposed method. The processing results for various types of video show that the proposed method provides up to 4,2 % bitrate saving for low distortion level in the intra-frame coding mode, up to 9,8 % bitrate reduction for low distortion level and up to 11,2 % bitrate saving for medium (acceptable) distortion level in the inter-frame coding mode compared to the uniform quantization method implemented in this video codec.

**Key words:** video codec Dirac, video coding, wavelet decomposition, quantization, uniform quantization, nonlinear quantization, Lloyd-Max algorithm, entropy-constrained quantization.

Ключевые слова видеокодек Dirac, видеокодирование, вейвлет-декомпозиция, квантование, равномерное квантование, нелинейное квантование, алгоритм Ллойда-Макса, алгоритм энтропийного ограниченного квантования.

## Введение

Для уменьшения визуальной и статистической избыточности в видеокодеках используется групповое кодирование кадров, основанное на их разбиении на участки и одновременном кодировании сразу группы пикселов. При групповом кодировании обычно используют преобразование исходных данных для увеличения количества незначимых коэффициентов после преобразования, которые можно отбросить без заметного изменения качества восстановленного кадра, и для декорреляции исходных данных.

После преобразования полученные коэффициенты повергаются квантованию для уменьшения объема данных. Задача оптимального квантования коэффициентов преобразования является важной для сжатия изображений. Исследования показывают, что коэффициенты ортогонального преобразования телевизионных кадров имеют распределение, подчиняющееся либо нормальному, либо Лапласовому законам. Для этих случаев при большом количестве уровней квантования близким к оптимальному является равномерное квантование [1].

Возможен другой поход к выбору алгоритма квантования коэффициентов преобразований, связанный с учетом субъективной чувствительности глаза к различными пространственным частотам, учитывая, что при передаче более высокочастотных коэффициентов допустима большая погрешность [1]. Группой JPEG эмпирически получен ряд таблиц для квантования коэффи-

Данная статья посвящена исследованию и разработке методов квантования для вейвлет-видеокодека. В статье исследованы методы равномерного, нелинейного и энтропийного ограниченного квантования и их потенциальные применения для вейвлет-видеокодека Dirac. На основе анализа результатов обработки видео разных типов с помощью перечислительных методов квантования предложен оптимизированный метод квантования для вейвлетвидеокодека. Его применение в вейвлет-видеокодеке Dirac позволяет заметно снизить скорость выходного потока при заданном уровне искажений. Результаты обработки видео разных типов показывают, что по сравнению с методом равномерного квантования, реализованного в рассматриваемом видеокодеке, предложенный метод для режима внутрикадрового кодирования уменьшает битовую скорость до 4,2 % при низком уровне искажений, а для режима межкадрового кодирования уменьшает битовую скорость до 9,8 % при низком уровне искажений и до 11,2 % при среднем (приемлемом) уровне искажений.

циентов дискретного косинусного преобразования (ДКП), которые приняты в стандарте JPEG [1, 2].

В отличие от ДКП, применяемого на практике для квадратных блоков малого размера, что вызывает блокинг эффект на границе блоков, дискретное вейвлетпреобразование (ДВП) может применяться для всего кадра. В стандарте JPEG2000 [1, 3] для сжатия изображения коэффициенты вейвлет-преобразования разбиваются на частотные области. Для каждой области могут использоваться различные коэффициенты квантования [1].

В рассматриваемом ниже вейвлет-видеокодеке Dirac коэффициенты вейвлет-преобразования разделяются на диапазоны разных пространственных частот и квантуются простым равномерным квантованием [4, 5]. Однако, как ортогональные преобразования, коэффициенты ДВП могут квантоваться с учетом субъективной чувствительности глаза к различными пространственным частотам. Так что можно квантовать коэффициенты разных диапазонов по-разному, как в стандарте JPEG2000. Более того, имеются некоторые известные алгоритмы квантования [5-7], которые, в основном, используются в преобразовании аналогового сигнала в цифровой, откуда следует возможность их исследования для задачи сжатия видео. В связи с простотой метода квантования, реализованного в рассматриваемом видеокодеке, существует потенциальная возможность нахождения лучшего метода, который позволяет уменьшить скорость выходного потока при заданном уровне искажений, что объясняет актуальность и новизну исследуемой темы.

Цель работы – исследование и разработка оптимизированного метода квантования для вейвлет-видеокодека. Для достижения поставленных целей в работе исследованы метод равномерного квантования, реализованный в вейвлет-видеокодеке Dirac, а также методы нелинейного и энтропийного ограниченного квантования, их потенциальные применения для вейвлет-видеокодека. В этой работе предложены и реализованы некоторые методы улучшения метода квантования для видеокодека Dirac. На основе полученных результатов предложен оптимизированный метод квантования для рассматриваемого вейвлет-видеокодека.

Применение предложенного метода в вейвлетвидеокодеке Dirac позволяет заметно снизить скорость потока кодированных данных при сохранении качества восстановленного изображения для видео разных типов.

## Постановка задачи

На рис. 1 показана схема вейвлет-видеокодера. В оригинальной открытой версии видеокодека Dirac, разработанного компанией BBC [4, 8], использовано двухканальное ДВП. В этот видеокодек внедрен ряд улучшений, одним из которых является замена двухканального ДВП на трехканальное ДВП [9]. Банки фильтров трехканального ДВП вычислены и тестированы для видео разных типов, и для этой работы используется оптимизированный для сжатия банк фильтров, обеспечивающий наибольшую степень сжатия при заданном качестве восстановленного видео. Однако, на данный момент рассматриваемый видеокодек использует простой метод равномерного квантования. Практической целью работы является разработка оптимизированного метода квантования в блок «Квантование» для модифицированного видеокодека Dirac (см. рис. 1).

Данная статья структурирована следующим образом. Сначала рассмотрена попытка использования особенности ДВП, заключающейся в том, что точности квантования вейвлет-коэффициентов в разных диапазонах имеют различные вклады в искажение восстановленного кадра. Далее рассмотрена возможность применения алгоритма Ллойда-Макса, обеспечивающего минимальную энергию остатка деквантования при заданном количестве уровней. В статье также рассмотрен модифицированный алгоритм Ллойда-Макса, так называемый энтропийным ограниченным квантованием. Затем, на основе анализа полученных результатов предложен оптимизированный метод квантования для рассматриваемого вейвлетвидеокодека. Результаты обработки видео разных типов с предложенным методом квантования и сделанные выводы приведены в конце статьи.



Рис. 1. Схема вейвлет-видеокодера

## Использование особенности дискретного вейвлет-преобразования

Один уровень трехканального ДВП позволяет разложить одномерный сигнал на три диапазона, а двумерный сигнал – на девять диапазонов [9, 10]. На рис. 2 показано разделение двумерного сигнала на диапазоны для одного уровня трехканального вейвлет-разложения. Из рис. 2 видно, что после фильтрации большая часть энергии сигнала сосредоточится в низкочастотном диапазоне по горизонтали и по вертикали (9-LL), а остальная энергия сигнала распределится между остальными диапазонами неравномерно, из чего следует различное влияние искажения вейвлет-коэффициентов разных диапазонов на восстановленный сигнал.

9-LL	8-LM	7-LH
6-ML	5-MM	4-MH
3-HL	2-HM	1-HH

## Рис. 2. Разделение сигнала на диапазоны с помощью трехканального вейвлет-разложения для первого уровня

Целью этого раздела является нахождение оптимального шага для каждого диапазона на каждом участке RD кривой (rate-distortion curve, кривая зависимости уровня искажений от скорости выходного потока). Для этого проводится один уровень трехканального вейвлет-разложения. Сначала диапазон (9-LL) не квантуется, так как далее этот диапазон может подвергаться трехканальному вейвлет-разложению, а остальные диапазоны квантуются равномерно с индексом QF (20, 16, 12, 8, 4, 0), соответствующими шагам квантования Q (32, 16, 8, 4, 2, 1). Таким образом, получится RD кривая, соответствующая равномерному квантованию, использованному в рассматриваемом видеокодеке, и являющаяся эталонной для оценки предложенного метода (см. REF на рис. 3).

Исследование состоит в том, что для каждого диапазона отдельно (для диапазона 5-ML, например) вместо QF ставится фиксированное значение QF-band-test (20, 16, 12, 8, 4, 0), а значение QF из набора (20, 16, 12, 8, 4, 0) – для всех остальных (кроме LL), тогда получатся новые RD кривые. Если для рассматриваемого диапазона, найдется какая-то точка, находящаяся выше эталонной кривой, то такое значение QF-band-test для этого диапазона будет использовано на соответствующем участке RD кривой. Таким образом, можно получить оптимальный шаг линейного квантования для всех диапазонов на всех участках RD кривой. Проведенный анализ показал, что результаты для всех диапазонов имеют одинаковый характер для разных видео в разных режимах. В качестве примера на рис. 3 приведены RD кривые для диапазона 6-ML для видео «city\_4cif» в режиме внутрикадрового кодирования.



Рис. 3. RD кривые для диапазонов 6-ML для видео «city\_4cif» в режиме внутрикадрового кодирования

Из рис. З видно, что полученные кривые всегда находятся ниже эталонной, что означает превосходство равномерного квантования по отношению к предложенному методу использования различных шагов линейного квантования для разных диапазонов. Это может быть связано с тем, что использованный банк фильтров специально вычислен с учетом минимизации ошибки равномерного квантования в разных диапазонах, т.е., использованный банк фильтров оптимизирован для метода равномерного квантования. Можно сделать вывод, что если приходится использовать линейное квантование для группы каких-либо диапазонов, то оптимальным вариантом для этих диапазонов является использование одинакового шага квантования.

### Применение алгоритма Ллойда-Макса

Алгоритм Ллойда-Макса [5, 6] для заданного количества уровней (L) выполняется следующими шагами:

Шаг 1 – Задать начальный набор деквантованных значений

$$R_a, q = 0, 1, \dots, L-1$$

Шаг 2 – Вычислить уровни квантования

$$T_q = \frac{R_{q-1} + R_q}{2}, q = 1, 2, \dots, L-1$$

Шаг 3 – Вычислить новые деквантованные значения

$$R_q = \frac{\sum_{n=T_q}^{T_{q+1}} nP[n]}{\sum_{n=T_q}^{T_{q+1}} P[n]}, q = 0, 1, \dots, L-1$$

Шаг 4 – Обновить фактор искажения

Шаг 4.1 – Обновить среднеквадратичное отклонение деквантованных и исходных данных MSE

Шаг 4.2 – Обновить фактор искажения 
$$\varepsilon = \frac{MSE_{old} - MSE_{new}}{2}$$
.

Шаг 5 – Повторять шаги 2, 3 и 4 до тех пор, пока не будет выполняться неравенство  $\varepsilon < \varepsilon_{eiven}$ .

Суть алгоритма заключает в том, что из статистики данных и заданного количества уровней квантования вычисляются оптимальные уровни квантования и деквантованные значения (центры тяжести), обеспечивающие минимальное среднеквадратичное отклонение деквантованных данных и исходных данных. Так что применение данного алгоритма гарантирует улучшение восстановленного кадра, однако при этом неизвестно, как меняется количество битов для сжатия. Целью этого раздела является нахождение RD кривой для случая с использованием алгоритма Ллойда-Макса.

Проведенный анализ показал, что для разных видео в разных режимах, результаты имеют общий характер. В качестве примера, приведены сравнения результатов применения алгоритма Ллойда-Макса (см. nonlinear) с результатами эталонного равномерного квантования (см. linear\_ref) для видео «city\_4cif» для режима внутрикадрового кодирования (см. рис. 4) и режима межкадрового кодирования (см. рис. 5). При этом для алгоритма Ллойда-Макса используется несколько итераций (см. 1 iter, 2 iter, 5 iter) и много итераций (см. Мах PSNR), когда почти нет расхождения значений MSE между двумя последующими итерациями.



Рис. 4. Сравнение алгоритма Ллойда-Макса с равномерным квантованием для видео «city\_4cif» в режиме внутрикадрового кодирования

Из рис. 4 видно, что чем больше количество итераций, тем лучше PSNR, но при этом также быстро растет количество битов, и полученные RD кривые находятся чуть ниже эталонной для режима внутрикадрового кодирования. Для режима межкадрового кодирования (см. рис. 5) получается выигрыш при нескольких первых итерациях. Для обоих режимов наблюдается нарушение монотонности RD кривой в точке, соответствующей шагу квантования Q = 8. Это может быть связано с тем, что для этой точки равномерное квантование настолько хорошо, что при использовании алгоритма Ллойда-Макса PSNR мало увеличивается, однако при этом количество битов сильно увеличится, так что в итоге рассматриваемая точка сдвигается вправо много быстрее, чем вверх.





Чтобы выяснить причины проявления этой ситуации, в работе проведено исследование, заключающее в том, что для каждого рассматриваемого диапазона применяется алгоритм Ллойда-Макса, а для остальных диапазонов применяется равномерное квантование. Проведенным анализом показано, что при Q = 8, для диапазонов (1-НН, 2-НМ, 4-МН) энтропия квантованных коэффициентов или количество битов для сжатия быстро увеличивается при использовании алгоритма Ллойда-Макса, при этом значение PSNR почти не меняется в связи с малым влиянием искажения коэффициентов в рассматриваемых диапазонах на качество восстановленного кадра. Таким образом, при Q = 8, наблюдается ситуация, которой можно избежать, если для рассматриваемых диапазонов применять не алгоритм Ллойда-Макса, а равномерное квантование. Аналогичный анализ для режима межкадрового сжатия также приводит к аналогичному выводу. Нужно отметить, что вывод второго раздела не позволяет квантовать коэффициенты диапазонов (1-НН, 2-НМ, 4-МН) по разным шагам линейного квантования. Таким образом, можно сделать общий вывод, что для любого видео и для обоих режимов работы лучше всего использовать равномерное квантование для диапазонов (1-НН, 2-HM, 4-MH) и применять алгоритм нелинейного квантования Ллойда-Макса к другим диапазонам.

# Применение алгоритма энтропийного ограниченного квантования

Недостаток алгоритма Лоида-Макса заключается в том, что этот алгоритм дает оптимальный вариант квантования только со стороны энергии остатка деквантования, и не уделяет внимание объему данных, необходимому для хранения квантованных данных в сжатом виде. Для задачи сжатия данных более подходит модифицированный вариант алгоритма Ллойда-Макса, так называемый энтропийным ограниченным квантованием [7].

Алгоритм энтропийного ограниченного квантования для заданного количества уровней (L) выполняется следующими шагами: Шаг 1 – Задать начальный набор деквантованных значений:

$$R_q, q = 0, 1, \ldots, L - 1$$
.

Шаг 2 – Вычислить уровни квантования:

$$T_q = \frac{R_{q-1} + R_q}{2} - \lambda \frac{\log_2 p_{q-1} - \log_2 p_q}{R_{q-1} - R_q}, q = 1, 2, \dots, L-1.$$

Шаг 3 – Вычислить новые деквантованные значения:

$$R_q = \frac{\sum_{n=T_q}^{T_{q+1}} nP[n]}{\sum_{n=T_r}^{T_{q+1}} P[n]}, q = 0, 1, \dots, L-1.$$

Шаг 4 – Обновить фактор искажения:

Шаг 4.1 – Обновить среднеквадратичное отклонение деквантованных и исходных данных MSE.

Шаг 4.2 – Обновить энтропию деквантованных данных H(R).

Шаг 4.3 — Обновить функцию стоимости  $J = MSE + \lambda H(R)$ .

Шаг 4.4 – Обновить фактор искажения
$$arepsilon = rac{J_{old} - J_{new}}{J_{old}}.$$

Шаг 5 – Повторять шаги 2, 3 и 4 до тех пор, пока не будет выполняться неравенство  $\varepsilon < \varepsilon_{einen}$ .

В отличие от алгоритма Ллойда-Макса, в этом случае для контроля качества восстановленного кадра и количества нужных битов для хранения квантованных данных в сжатом виде добавляется Лагранжиан ( $\lambda$ ) в функции стоимости (см. шаг 4.3). Тогда уровни квантования (см. шаг 2) определяются с учетом параметра  $\lambda$ . Трудностью в этом методе является выбор подходящего значения  $\lambda$ .

Для определения параметра λ при выполнении алгоритма Ллойда-Макса записывается в файл значение среднеквадратичного отклонения деквантованных и исходных данных MSE и значение энтропии деквантованных данных H(R) после каждой итерации. Затем можно взять λ, равное отношению разности MSE к разности H(R) при нескольких начальных итерациях алгоритма Ллойда-Макса.

Проведенный анализ показал, что алгоритм энтропийного ограниченного квантования дает примерно такой же результат, как алгоритм Ллойда-Макса для первой итерации. Кроме того, трудность с определением параметра λ трудно обойти, так что достаточно и целесообразно остановиться на использовании алгоритма Ллойда-Макса. Для применения алгоритма энтропийного ограниченного квантования в задаче сжатия видео нужны дополнительные исследования, ориентированные на выбор подходящего Лагранжиана (λ).

## Предложенный метод оптимизированного квантования

На основе выводов предыдущих разделов для поставленной задачи предложен метод квантования, заключающийся в совместном использовании равномерного квантования для высокочастотных диапазонов (1-HH, 2-HM, 4-MH) и алгоритма Ллойда-Макса для других диапазонов.



Рис. 6. Оценка предложенного метода для видео «ice\_4cif» режим INTRA



Рис. 7. Оценка предложенного метода для видео «rush\_hour\_1080p» режим INTRA



Рис. 8. Оценка предложенного метода для видео «soccer\_4cif» режим INTER



Рис. 9. Оценка предложенного метода для видео «city\_4cif» режим INTER

На рис. 6-7 приведены результаты применения предложенного метода с различными видео разного типа (4cif и 1080p) для режима внутрикадрового кодирования (INTRA).



Рис. 10. Оценка предложенного метода для видео «pedestrian\_1080p» режим INTER



Рис. 11. Оценка предложенного метода для видео «sunflower\_1080p» режим INTER



Puc. 12. Оценка предложенного метода для видео «station2\_1080p» режим INTER

Из рис. 6-7 можно сделать вывод, что для режима внутрикадрового кодирования предложенный метод дает примерно такой же результат на областях высокого качества RD кривой и немного лучше для областей низкого качества RD кривой по сравнению с эталонным методом – методом равномерного квантования, реализованного в вейвлет-видеокодеке Dirac.

На рис. 8-12 приведены результаты применения предложенного метода с различными видео разного типа (4cif и 1080р) для режима межкадрового кодирования (INTER).

Из рис. 8-12 можно сделать вывод, что по сравнению с эталонным методом равномерного квантования, реа-

лизованного в вейвлет-видеокодеке Dirac, предложенный метод всегда дает заметно лучший результат и на областях низкого качества RD кривой, и для областей высокого качества RD кривой. При этом особенный интерес для приложения реального времени уделяется областям приемлемого качества. Если видео достаточно простое для кодирования и межкадровое предсказание хорошо работает, то остаточный кадр после трехканального преобразования и квантования с большим шагом состоит из много нулей, и число значимых битов составляет совсем маленький процент. Например, для видео «pedestrian 1080p» (см. рис. 10), это составляет 0,03 % для шага квантования Q = 32, 0,46 % для Q = 16, 2,35 % для O = 8. Тогда в этих областях неважно, какой метод квантования использовать, предложенный метод там практически не отличается от эталонного метода (см. рис. 10, 11). Если видео сложное, то остаток предсказания будет больше и даже в областях низкого качества RD кривой, когда шаг квантования большой, предложенный метод в этом случае работает намного лучше эталонного. Из рис. 8, 9, 12 можно сделать вывод, что чем сложнее видео или чем менее точно работает межкадровое предсказание, тем лучше предложенный метод работает по сравнению с методом равномерного квантования.

Численным сравнением предложенного метода с эталонным получены следующие результаты:

– для режима внутрикадрового кодирования, на областях высокого качества RD кривой предложенный метод сэкономил до 4,2 % битовой скорости при качестве восстановленного кадра PSNR = 45 для видео «rush hour 1080p».

– для режима межкадрового кодирования, на областях с приемлемым качеством RD кривой предложенный метод уменьшил 13,5 % битовой скорости при качестве восстановленного кадра PSNR = 41,7 для видео «station\_2\_1080p», 11,2 % битовой скорости при качестве восстановленного кадра PSNR = 37 для видео «city\_4cif», а на областях высокого качества предложенный метод сэкономил до 9,8 % битовой скорости при качестве восстановленного кадра PSNR = 45 для видео «station 2 1080p».

#### Заключение

В работе исследованы разные методы квантования и их потенциальное применение для задачи сжатия видео. На основе результатов исследований предложен оптимизированный метод квантования для вейвлетвидеокодека Dirac. В ходе работы с разными видео разных типов получены следующие выводы:

 если приходится использовать линейное квантование для группы каких-то диапазонов, то наилучшим вариантом для этих диапазонов является использование равномерного квантования, т.е., использование линейного квантования с одинаковым шагом квантования для рассматриваемых диапазонов;

 алгоритм энтропийного ограниченного квантования дает примерно такой же результат, как алгоритм ЛлойдаМакса для первой итерации, так что целесообразно остановиться на использовании алгоритма Ллойда-Макса;

 предложен оптимизированный метод квантования на основе совместного использования известного метода равномерного квантования для высокочастотных диапазонов (1-HH, 2-HM, 4-MH) и алгоритма Ллойда-Макса для других диапазонов;

– для режима внутрикадрового кодирования предложенный метод дает примерно такой же результат на областях низкого качества RD кривой и немного лучший результат для областей высокого качества RD кривой по сравнению с методом равномерного квантования, реализованного в рассматриваемом видеокодеке Dirac, конкретно, предложенный метод уменьшает до 4,2 % битовой скорости;

– для режима межкадрового кодирования предложенный метод всегда дает лучший результат по сравнению с равномерным квантованием, предложенный метод уменьшает до 11,2 % битовой скорости на участке с приемлемым качеством (PSNR = 37) RD кривой, и до 9,8 % на участке высокого качества RD кривой;

для сложных видео, когда межкадровое предсказание не очень хорошо работает, предложенный метод намного лучше эталонного на областях с приемлемым PSNR, что показывает целесообразность использования предложенного метода по сравнению с эталонным методом.

## Литература

1. Дворкович В.П., Дворкович А.В. Цифровые видеоинформационные системы (теория и практика). Москва: Техносфера, 2012. – 1008 с. ISBN 978-5-94836-336-3.

2. William B. Pennebaker, Joan L. Mitchell, JPEG still image data compression standard (3rd ed.), Springer, 1993, 291 p. ISBN 978-0-442-01272-4.

3. ITU-T Recommendation T.800 (2002-08) - Information technology – JPEG 2000 image coding system: Core coding system.

4. Dirac video codec // https://sourceforge.net/projects/ Dirac/

5. Lloyd S. Least squares quantization in PCM, IEEE Transactions on Information Theory, IT-28, 129–137, March 1982.

6. Max J. Quantizing for minimum distortion, IRE Transactions on Information Theory, IT-6, 7–12, 1960.

7. Chou P.A., Lookabaugh T. and Gray R.M. «Entropyconstrained vector quantization». // IEEE Trans. Acoust., Speech, Signal Processing, vol. 37, pp. 31–42, Jan. 1989.

 Jianguo Zhang Ling Shao Lei Zhang Graeme A. Jones.
 Intelligent video event analysis and understanding. Springer, 2010 edition. - 251 pages. ISBN 978-3-642-17554-1.

9. Gryzov G.Y, Dvorkovich A.V. «Three-Channel Wavelet Transform for Video Compression Applications». // 6<sup>th</sup> Mediterranean Conference on Embedded Computing MECO 2017, 11-15 June 2017, pp. 1-4.

10. Дворкович А.В., Дворкович В.П. Оконные функции для гармонического анализа сигналов. Москва: Техносфера, 2016. – 208 с. ISBN 987-5-94836-432-2.

#### УДК 004.932

# НОВЫЙ АЛГОРИТМ ПОВЫШЕНИЯ КОНТРАСТА МЕЛКОМАСШТАБНЫХ ДЕТАЛЕЙ ИЗОБРАЖЕНИЯ НА ОСНОВЕ РАЗЛОЖЕНИЯ ПО ВЕЙВЛЕТАМ

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

# A NEW ALGORITHM FOR INCREASING CONTRAST OF SMALL IMAGE DETAILS BASED ON THE DECOMPOSITION ON WAVELETS

## Likhachov A.V.

Based on the multiresolution signal representation the algorithm for processing of noisy images containing small structures is suggested. Detailing coefficients of the decomposition, with the order corresponding to the size of the investigated elements of the image are multiplied by a constant that is computed taking into account the level of noise and the difference between the amplitudes of the structure and the surrounding background. A computational experiment results confirmed the efficiency of the method, in particular, the contrast of small objects risen in 1,5-2 times.

Key words: contrast enhancement of the image elements, multiresolution analysis of signals.

Ключевые слова: повышение контраста элементов изображения, кратномасштабный анализ сигналов.

## Введение

Работа посвящена методам анализа полутоновых изображений, содержащих элементы малого размера, контраст которых относительно окружающего фона не позволяет их надежного обнаружения в присутствии интенсивного шума. Таковыми, в частности, являются рентгенограммы, имеющие важнейшее значение, в первую оче-

редь, для медицинской диагностики. Именно в этой области очень часто применяются методы локализации и идентификации небольших низкоконтрастных объектов, таких как опухоли, особенно на ранних стадиях развития. В качестве математической модели здесь обычно используются функции двух переменных, значения которых определяют яркость, обозначаемую ниже как *G*(*x*, *y*).

Для многих методов анализа и преобразования сигналов основой является разложение в ряд Фурье [1]. В одномерном случае оно имеет вид

$$g(x) = \sum_{n=-\infty}^{n=\infty} \left\langle g, \psi_n^F \right\rangle \psi_n^F(x).$$
(1)

Здесь  $\psi_n^F(x) = \exp(inx\Delta\omega)$ ,  $\Delta\omega = 2\pi/(x_2 - x_1)$ ; угловыми скобками обозначено скалярное произведение в Гильбертовом пространстве  $H_{[x_1;x_2]}$ , состоящем из квадратично интегрируемых функций, определённых на отрезке  $[x_1;x_2]$ :

$$\left\langle g, \psi_n^F \right\rangle = \frac{1}{x_2 - x_1} \int_{x_1}^{x_2} g(x) (\psi_n^F)^*(x) dx$$
, (2)

где  $(\psi_n^F)^*(x)$  – комплексно сопряжённая функция. Несмотря на достигнутые с его помощью успехи, базис  $\psi_n^F(x)$  имеет ряд существенных недостатков [2]. Один из них, наиболее важный с точки зрения излагаемого

На основе кратномасштабного представления сигналов предложен новый алгоритм обработки зашумлённых изображений, содержащих мелкие структуры. Детализирующие коэффициенты разложения, имеющие порядок, соответствующий размеру исследуемых элементов изображения, умножаются на константу, которая вычисляется с учётом уровня шума и разности амплитуд структуры и окружающего фона. Проведён вычислительный эксперимент, результаты которого подтвердили эффективность метода, в частности, было получено повышение контраста малоразмерных объектов в 1,5 – 2 раза.

> здесь метода, состоит в том, что совокупность коэффициентов разложения (спектр) практически не даёт возможности определить локализацию особенностей, в частности разрывов функции g(x), поскольку информация о них распределяется по всем произведениям (2). До некоторой степени это можно преодолеть, выделяя определённые участки g(x) оконными функциями, см. например, [3]. При этом для каждого положения окна также получается набор коэффициентов (2), анализируя которые можно делать выводы о свойствах g(x) в рассматриваемой области. Однако в случае, когда ставится относительно простая задача обнаружения включения с амплитудой, значительно отличающейся от среднего, сложный спектральный анализ представляется нецелесообразным. Альтернативой является разложение по вейвлетам, развитие теории которых было стимулировано ограничениями, присущими преобразованию Фурье [2, 4, 5]. Они получаются из функции  $\psi(x)$ , обладающей определёнными свойствами [5], путём масштабирования и сдвига:

$$\psi_{m,k}(x) = a^{m/2} \psi(a^m x - k),$$
 (3)

где *а* – вещественное число; *m* и *k* – целые.

Базисы (3), ввиду их разнообразия, предоставляют более широкие исследовательские возможности. В частности, для решения упомянутой выше задачи обнаружения достаточно перенести начало координат и подобрать значение *m* таким образом, чтобы носитель  $\psi_{m,0}(x)$  полностью покрывал предполагаемую область неоднородности, а носители  $\psi_{m,k}(x)$  с  $k \neq 0$  не пересекались с ней. Тогда, если скалярное произведение  $\langle (g - \overline{g}), \psi_{m,0} \rangle$  будет заметно превосходить остальные  $\langle (g - \overline{g}), \psi_{m,k} \rangle$  с  $k \neq 0$ , то это будет свидетельствовать в пользу наличия неоднородности. Здесь  $\overline{g}$  – это среднее значение функции g(x).

Современные способы хранения, преобразования и воспроизведения сигналов, основанные на компьютерных технологиях, подразумевают, что они представляются в виде упорядоченного набора чисел. Применительно к рассматриваемому случаю это означает, что G(x, y) задаётся в узлах сетки с координатами  $(x_p, y_q)$ . Каждый узел соответствует пикселю изображения, а значение  $G(x_p, y_q)$  – его яркости. Кроме того, сама функция G(x, y) может принимать лишь определённый ряд значений (так называемые уровни квантования). Последним обстоятельством будем пренебрегать, считая G(x, y) функцией, непрерывной по амплитуде.



Рис. 1. Функции, порождающие используемые базисы

Обработка оцифрованных данных подробно излагается, например, в работах [3, 6, 7]. Одним из её инструментов является кратномасштабный анализ, обеспечивающий быстрое и эффективное решение ряда задач, связанных с выявлением частотно-координатной структуры [8]. При этом используются два взаимосвязанных ортонормированных автомодельных базиса  $\varphi_{m,k}(x)$  и  $\psi_{m,k}(x)$ , построенных согласно (3) с константой *а*, равной 2. Пусть получено  $N = 2^M$  отсчётов. Имеет место следующее разложение, [8]:

$$g(x) = \sum_{k=0}^{2^{n}-1} c_{n,k} \varphi_{n,k}(x) + \sum_{m=n}^{M-1} \left( \sum_{k=0}^{2^{m}-1} d_{m,k} \psi_{m,k}(x) \right).$$
(4)

Здесь  $c_{n,k}=\left\langle g, \varphi_{n,k} \right
angle$  и  $d_{m,k}=\left\langle g, \psi_{m,k} \right
angle$  – масштаби-

рующие и детализирующие коэффициенты соответственно; n может принимать значения от нуля до M, однако при M = n второй член в (4) отсутствует. Первый из индексов, нумерующих  $c_{m,k}$  и  $d_{m,k}$ , определяет их уровень. Практической ценности выражения (4) во многом способствует тот факт, что, независимо от базисных функций, выполняются соотношения [8]

$$c_{m-1,k} = (1/\sqrt{2})(c_{m,2k} + c_{m,2k+1}),$$
  

$$d_{m-1,k} = (1/\sqrt{2})(c_{m,2k} - c_{m,2k+1}),$$
(5 a)

$$c_{m,2k} = (1/\sqrt{2})(c_{m-1,k} + d_{m-1,k}),$$
  

$$c_{m,2k+1} = (1/\sqrt{2})(c_{m-1,k} - d_{m-1,k}).$$
(5 6)

Таким образом, зная высшие масштабирующие коэффициенты  $c_{M,k}$ , по формулам (5а) можно найти все  $c_{m,k}$  и  $d_{m,k}$  в разложении (4) до любого уровня n, не используя явного вида функций  $\varphi(x)$  и  $\psi(x)$ . При проведении исследований часто порождающей функцией базиса  $\varphi_{m,k}(x)$  выбирается прямоугольный импульс, тогда базис  $\psi_{m,k}(x)$  образуется из вейвлета Хаара, см. рис. 1. Обычно в этом случае в качестве  $c_{M,k}$  используются значения самого сигнала, т.е. полагается  $c_{M,k} = g(x_k)$ .

# Предлагаемый метод

Пример, приведённый во введении, показывает, что, применяя кратномасштабный анализ, неоднородность на квазипостоянном фоне можно охарактеризовать единственным параметром – скалярным произведением g(x)на соответствующую базисную функцию. Это наблюдение явилось идейной основой предлагаемого в работе метода, суть которого состоит в следующем. Каждая строка изображения представляется в виде (4), где л, определяется размером исследуемых включений. Предположим, что одному из них соответствует коэффициент  $d_{n,k'}$ , при этом  $|d_{n,k'} - d_{n,k'-1}| pprox |d_{n,k'} - d_{n,k'+1}| pprox eta_{n,k'}$ . Величина  $\beta_{n.k'}$  зависит, в частности, от разности амплитуд включения и окружающего фона, которая будет обозначаться как б. По формуле (4) проведём сборку строк, используя вместо каждого  $d_{n,k}$  его произведение на некоторое число  $\alpha_n > 1$ . В полученном изображении рассматриваемые разности детализирующих коэффициентов будут порядка  $\alpha_n \beta_{n,k'}$ . Поскольку базис остаётся прежним, это изменение будет обусловлено, в первую очередь, увеличением  $\delta$ . Следовательно, после такого преобразования контрастность включений заданного размера повысится, что будет способствовать улучшению условий их обнаружения и идентификации. Далее в этом разделе поэтапно описывается алгоритм, реализующий метод, изложенный выше на качественном уровне.

Пусть изображение  $N \times N$  пикселей, содержащее достаточно большой фрагмент фона с амплитудой  $G_0$ , на котором располагаются L включений  $\Delta_j$ , j = 1, ..., L, искажено аддитивным стационарным центрированным некоррелированным шумом с дисперсией  $\sigma^2$ . Обозначим через  $l_j$  характерный размер в пикселях области  $\Delta_j$ , а через  $\delta_j$  – величину её превышения над фоном. Таким образом, яркость в пределах  $\Delta_j$  определяется значениями функции G(x, y) равными  $G_0 + \delta_j$ . Упорядочим индексы следующим образом:  $l_1 \leq l_2 \leq ... \leq l_L$ . При этом предполагается, что включения занимают небольшую площадь, так что выполняется условие  $l_L << N$ .

На первом этапе разработанного алгоритма вычисляются коэффициенты разложения (4), при этом полагается

$$\begin{cases} n = M - [\log_2 l_L] - 2; \ \log_2 l_L - [\log_2 l_L] > 0, 5, \\ n = M - [\log_2 l_L] - 1; \ \log_2 l_L - [\log_2 l_L] \le 0, 5, \end{cases}$$
(6)

где квадратные скобки означают целую часть числа. В соответствии со сказанным в конце предыдущего раздела, за высшие масштабирующие коэффициенты  $c_{M,k}$  полагаются значения функции G(x, y), т.е. яркости пикселей исходного изображения. Например, в разложении для q-ой строки  $c_{M,k} = G(x_k, y_q), \quad k = 0, 1, ..., N-1$ . Коэффициенты низших уровней находятся по формулам (5а). В частности,

$$c_{M-1,k} = (1/\sqrt{2})(G(x_{2k}, y_q) + G(x_{2k+1}, y_q)),$$
  

$$d_{M-1,k} = (1/\sqrt{2})(G(x_{2k}, y_q) - G(x_{2k+1}, y_q)).$$
(7)

и т.д. В выражениях (7) индекс k изменяется от 0 до N/2-1, таким образом, при понижении порядка на единицу количество коэффициентов уменьшается в два раза. Заметим, что в (4) из масштабирующих коэффициент входят только  $c_{n,k}$ , однако для того чтобы найти  $d_{m,k}$  с  $n \le m < M$  по второму из уравнений (5а), требуется также вычислить все  $c_{M,k}$  при  $n+1 \le m < M$ .

На втором этапе путём замены в (6)  $l_L$  на  $l_j$  для каждого  $\Delta_j$  определяется уровень детализирующих коэффициентов  $m_j$ , отвечающих его размеру, причём, как это нетрудно видеть из (6), имеет место  $2^{M-m_j-3/2} < l_j \leq 2^{M-m_j-1/2}$ . Далее выбираются один или несколько масштабов неоднородностей (соответствующие им индексы будут обозначаться через  $m_s$ ), контрастность которых нужно повысить. Все коэффициенты  $d_{m_s,k}$  каждого из выбранных уровней преобразуются путём умножения на константу бо́льшую единицы:  $d_{m_s,k} \rightarrow \tilde{d}_{m_s,k} \equiv \alpha_{m_s} d_{m_s,k}$ . При этом числа  $\alpha_{m_s}$  в общем случае различны для различных уровней.

Чтобы получить оценки параметров  $\alpha_{m_s}$  допустим, что рассматриваемая строка пересекает единственную неоднородность  $\Delta_j$  по отрезку, имеющему длину  $l_j = 2^{M-m_s-1}$ , левый край которого приходится на пиксель с номером  $k'l_j - 1$ . Остальную часть занимает фон. При таких условиях, с учётом того, что  $c_{M,k}^0$  равны либо  $G_0$ , либо  $G_0 + \delta_j$ , прямое вычисление по (5а) даёт  $c_{m_s+1,2k'}^0 = 2^{(M-m_s)/2}(G_0 + \delta_j)$ ,  $c_{m_s+1,2k'+1}^0 = 2^{(M-m_s)/2}G_0$ ;  $d_{m_s,k'}^0 = (1/\sqrt{2})(c_{m_s+1,2k'}^0 - c_{m_s+1,2k'+1}^0) = 2^{(M-m_s-1)/2}\delta_j$ . (8)

Здесь верхний индекс «0» указывает, что значения коэффициентов определяются в отсутствие шума. При его наличии  $c_{m_s,k'}$  и  $d_{m_s,k'}$  являются случайными величинами, причём, учитывая предположенные выше свойства шума,  $M\{c_{m_s,k'}\} = c_{m_s,k'}^0$  и  $M\{d_{m_s,k'}\} = d_{m_s,k'}^0$ , где  $M\{\circ\}$  – математическое ожидание. Для вычисления дисперсии, воспользуемся формулами (5а). Поскольку

 $D\{c_{Mk}\} = \sigma^2$ , получаем

$$D\{c_{M-1,k}\} = (1/2)(D\{c_{M,2k}\} + D\{c_{M,2k+1}\}) = (1/2)(\sigma^2 + \sigma^2) = \sigma^2.$$
(9)

Отсюда по индукции имеем  $D\{c_{m,k}\} = \sigma^2$  для любого *m*. Тогда

$$D\{d_{m_{s},k'}\} = D\{(1/\sqrt{2})(c_{m_{s}+1,2k'} + (-1)c_{m_{s}+1,2k'+1})\} = (1/2)(\sigma^{2} + (-1)^{2}\sigma^{2}) = \sigma^{2}.$$
(10)

Исходя из последнего равенства (8), введём случайную величину *ξ*, характеризующую среднее отклонение амплитуды от фона:

$$\xi = 2^{(m_s - M + 1)/2} d_{m_s, k'}.$$
 (11)

Она является несмещённой оценкой параметра  $\delta_j$ , так как из сказанного выше видно, что  $M\{\xi\} = \delta_j$  для всех *m*. Дисперсия величины (11) равна  $D\{\xi\} = 2^{m_s - M + 1}\sigma^2$ . При умножении  $d_{m_s,k'}$  на  $\alpha_{m_s}$ , приходим к случайной величине  $\xi_{\alpha}$  с  $M\{\xi_{\alpha}\} = \alpha_{m_s}\delta_j$  и  $D\{\xi_{\alpha}\} = 2^{m_s - M + 1}\alpha_{m_s}^2\sigma^2$ . Значение  $\alpha_{m_s}$  следует определить из условия

$$M\{\xi\} + \sqrt{D\{\xi\}} = M\{\xi_{\alpha}\} - \sqrt{D\{\xi_{\alpha}\}} , \qquad (12)$$

откуда получаем

$$\alpha_{m_s} = (M\{\xi\} + \sqrt{D}\{\xi\}) / (M\{\xi\} - \sqrt{D}\{\xi\}) =$$
  
=  $(\delta_j + 2^{(m_s - M + 1)/2} \sigma) / (\delta_j - 2^{(m_s - M + 1)/2} \sigma).$  (13)

Согласно (13), с ростом уровня  $m_s$  (что соответствует уменьшению размера неоднородности) и при увеличении шума, множитель  $\alpha_{m_s}$  растёт. По мере приближения  $\delta_j \kappa 2^{(m_s - M + 1)/2} \sigma$  он стремится к бесконечности, а в случае  $2^{(m_s - M + 1)/2} \sigma > \delta_j$  величина  $\alpha_{m_s}$  становится отрицательной. В то же время вычислительный эксперимент показал, что при  $\alpha_{m_s} > 4 \div 6$  (в зависимости от  $m_s$ ) изображение сильно искажается. На нём появляются характерные артефакты в виде чередования светлых и ярких сегментов, занимающих  $2^{M-m_s-1}$  пикселя. В связи с этим, формулу (13) предлагается использовать, пока полученные по ней  $\alpha_{m_s}$  не превосходят некоторого числа  $\alpha^c$ , которое каждый раз определяется по результа-

там расчетов (в частности, в данной работе было выбрано  $\alpha^c = 5$ ). Если же оказывается  $\alpha_{m_s} > \alpha^c$ , то знаменатель нужно заменить на  $\delta_j$ .

Следует заметить, что  $\alpha_{m_s}$  вычисляется по формуле (13) для неоднородности с фиксированной амплитудой  $G_0 = \delta_j$ . Когда на изображении имеются структуры одинакового масштаба, но различной яркости, для каждойиз них, вообще говоря, требуется свой множитель. Если эта разница незначительна, то точные  $\delta_j$  можно заменить некоторым единым усреднённым значением. В противном случае для эффективного применения метода нужна более подробная информация о локализации включений. В

частности, если известны положения и амплитуды всех неоднородностей, то  $\alpha_{m_s}$  можно находить отдельно для каждой строки. Однако такие знания об исследуемом объекте на практике, как правило, недоступны.



а) без шума



б) при наличии аддитивного белого шума, σ = 0,1
 Рис. 2. Модельное изображение

Наконец, на последнем этапе алгоритма проводится сборка строк изображения после преобразования  $d_{m_s,k} \rightarrow \tilde{d}_{m_s,k}$ . Для этого по формулам (5b) последовательно находятся масштабирующие коэффициенты вплоть до высшего порядка M. При этом для выбранных уровней  $m_s$  берутся новые детализирующие коэффициенты  $\tilde{d}_{m_s,k}$ . Полученные таким образом  $\tilde{c}_{M,k}$  принимаются за новые значения функции G(x, y), характеризующей яркость изображения.

#### Вычислительный эксперимент

Численное моделирование проводилось с математическим фантомом, изображённым на рис. 2,а. Разрешение составляет 1024 × 1024. На фоне с  $G_0 = 1$  располагаются пять квадратов со стороной 9 и три круга диаметром 16 (здесь и далее размеры указываются в пикселях). Для того чтобы было удобнее визуально различить эти структуры, на рис. 2а слева от кругов поставлены крестики. Значение  $\delta$  для всех включений одинаково и составляет 0,1. Таким образом, M = 10, для квадратов m = 6, а для кругов m = 5. Формально фантом задаётся уравнением:

$$G(x, y) = G_0 \sum_{j=1}^{2} \chi_j^e(x, y) + \delta \sum_{j=1}^{3} \chi_j^c(x, y) + \delta \sum_{j=1}^{5} \chi_j^s(x, y),$$
(14)

где  $\chi^e(x,y)$ ,  $\chi^c(x,y)$  и  $\chi^s(x,y)$  характеристические функции эллипса, круга и квадрата. В процессе вычислительного эксперимента к фантому добавлялся белый шум с дисперсией  $\sigma^2$ , имеющий в каждой точке гауссо-

во распределение. Пример представлен на рис. 2 b. Для него величина стандартного отклонения *о* равна 0,1.



Рис. 3. Зависимости контраста от стандартного отклонения шума

Определим для каждого вида неоднородностей контраст Вебера

$$K_{c} = (\overline{I}_{c} - \overline{B})/\overline{I}_{c}, \quad K_{c} = (\overline{I}_{s} - \overline{B})/\overline{I}_{s}, \quad (15)$$

где  $\overline{I}_c$ ,  $\overline{I}_s$  и  $\overline{B}$  – средние значения:

$$\overline{I}_{c} = \sum_{(x_{p}, y_{q}) \in S_{c}} \frac{G(x_{p}, y_{q})}{N_{c}}, \quad \overline{I}_{s} = \sum_{(x_{p}, y_{q}) \in S_{s}} \frac{G(x_{p}, y_{q})}{N_{s}},$$
$$\overline{B} = \sum_{(x_{p}, y_{q}) \in S_{b}} \frac{G(x_{p}, y_{q})}{N_{b}}.$$
(16)

В (16) суммирование ведётся в областях, занятых кругами, квадратами и эллипсами:

$$S_c = \bigcup_{j=1}^{3} \chi_j^c, \quad S_s = \bigcup_{j=1}^{3} \chi_j^s, \quad S_b = (\chi_1^e \cup \chi_2^e) \setminus (S_c \cup S_s).$$
(17)

Количества пикселей, содержащихся в  $S_c$ ,  $S_s$  и  $S_b$ , обозначены как  $N_c$ ,  $N_s$  и  $N_b$ . Из описания фантома и формул (15), (16) следует, что в отсутствии искажений

$$K_c = K_s = (\delta + G_0 - G_0) / (\delta + G_0) = \delta / (\delta + G_0) \approx 0,09. (18)$$

На рис. 3 а,b приведены зависимости значений  $K_c$  и  $K_s$  от  $\sigma$ . Кривые 1 показывают контрасты на исходном (зашумленном) изображении, все остальные – на обработанном. По характеру кривых 1 видно, что наличие шума не влияет существенно на усреднённые величины, вычисленные согласно (15). Они испытывают случайные колебания, амплитуда которых больше для рис. 3 b, поскольку число пикселей занимаемых квадратами недостаточно велико – 405 (на круги их приходится 603). Сопоставление кривых 1 с кривыми 2 и 3 показывает, что применение предлагаемого метода поднимает  $K_c$  и  $K_s$  пропорционально дисперсии  $\sigma^2$ .



Как было указано в предыдущем разделе, разработанный алгоритм позволяет одновременно изменять  $d_{mk}$ для нескольких т. Именно этому случаю соответствуют кривые 2. Для их получения преобразовывались коэффициенты двух уровней:  $d_{6,k} \rightarrow \alpha_6 d_{6,k}, \quad d_{5,k} \rightarrow \alpha_5 d_{5,k}.$ При этом для больших о, как это видно из (13), множитель α<sub>6</sub> существенно превосходит α<sub>5</sub>, см. также рис. 4. Следовательно, контраст мелких структур, отвечающих шестому уровню, увеличивается сильнее. Более крупные детали, масштаб которых определяется значением индекса m = 5, напротив, будут вуалироваться из-за усиления флуктуаций, имеющих размер в два раза меньший. Поэтому если шум достаточно интенсивный, то для исследования изображения на уровне m = 5 целесообразно преобразовать только коэффициенты  $d_{5,k}$ , что иллюстрирует кривая 3 на рис. 3 а. Она проходит выше кривой 2. Таким образом, одновременный рост  $d_{6k}$  и  $d_{5k}$  приводит к тому, что контраст кругов, имеющих больший размер по сравнению с квадратами, усиливается слабее, чем при увеличении лишь  $d_{5k}$ .





при использовании предлагаемого метода, шум  $\sigma = 0,1$ Зависимости множителей  $\alpha_5$  и  $\alpha_6$ , полученных по формуле (13), от стандартного отклонения шума представлены на рис. 4, кривые 1 и 2 соответственно. Видно, что пока шум незначительный (до  $\sigma \approx 0,05$ ) величины  $\alpha_5$ и  $\alpha_6$  мало отличаются друг от друга. Однако при дальнейшем увеличении его дисперсии  $\alpha_6$  растёт существенно быстрее, чем  $\alpha_5$ . В частности, при  $\sigma$  = 0,2 параметр  $\alpha_6$  становится почти в три раза больше, чем  $\alpha_5$ . Он оказывается приблизительно равным 6, напомним, что это превышает выбранное критическое значение  $\alpha^c = 5$ , см. текст за формулой (13).

Примеры обработки изображения, представленного на рис. 2 b, даны на рис. 5 a,b. Первый из них получен посредством преобразования  $d_{6,k} \rightarrow \alpha_6 d_{6,k}$ ,  $d_{5,k} \rightarrow \alpha_5 d_{5,k}$ , второй – путём изменения только коэффициентов с m = 5:  $d_{5,k} \rightarrow \alpha_5 d_{5,k}$ . Круги (структуры уровня m = 5) на рис. 5 b проявляются отчётливее, нежели на рис. 5 а. Это согласуется со сделанным выше выводом о том, что для анализа более крупных включений следует увеличивать лишь коэффициенты уровня, соответствующего их размеру.

#### Заключение

Работа посвящена актуальной проблеме обнаружения деталей изображения, занимающих несколько сотых процента от его площади. Предлагается новый метод повышения их амплитуды относительно окружающего фона, в основе которого лежит кратномасштабный анализ одномерных сигналов. В разложении строк коэффициенты при вейвлетах, имеющих носитель порядка размеров рассматриваемых объектов, умножаются на константу. Для её вычисления разработан алгоритм, использующий статистические характеристики, обусловленные наличием шума. В процессе численного моделирования для включений, превышающих фон на 10 % и имеющих протяженность 9 ÷ 16 пикселей, при разрешении 1024 × 1024, было достигнуто увеличение контраста в 1,5 – 2 раза.

Говоря о перспективах метода, следует отметить, что наиболее естественным его развитием является переход от построчного анализа изображения к его разложению по двумерным вейвлетам. Одно из преимуществ такого подхода состоит в том, что он позволяет использование базисных функций с различными масштабами по координатным осям, т.е. имеющих не квадратный, а прямоугольный носитель. Другое направление усовершенствования связано с применением более сложных процедур преобразования коэффициентов, в том числе масштабирующих, например, умножение каждого из них на определённое именно для него число. Это могло бы оказаться полезным при наличии нестационарного, а также коррелированного шума.

#### Литература

1. Толстов Г. П. Ряды Фурье. М.: Наука, 1980. 384 с.

2. Чуи К. Введение в вэйвлеты. М.: Мир, 2001. 412 с.

3. Дворкович В.П., Дворкович А.В. Оконные функции для гармонического анализа сигналов. Издание второе, переработанное и дополненное. М.: Техносфера, 2016. 208 с.

4. Астафьева Н. М. Вейвлет-анализ: основы теории и примеры применения // УФН. 1996. Т. 166, № 11. С. 1145-1170.

5. Добеши И. Десять лекций по вейвлетам. Ижевск: РХД, 2001. 464 с.

6. Рабинер Л., Гоулд Б. Теория и применение цифровой обработки сигналов. М.: Мир, 1978. 848

7. Сергиенко А. Б. Цифровая обработка сигналов. 2-е изд. СПб.: Питер, 2007. 751 с.

8. Малла С. Вэйвлеты в обработке сигналов. М.: Мир, 2005. 672 с.

## УДК 004.93

# БЕЗЭТАЛОННЫЙ ИНТЕГРАЛЬНО-МУЛЬТИПЛИКАТИВНЫЙ ПОКАЗАТЕЛЬ КАЧЕСТВА ЦИФРОВЫХ ПОЛУТОНОВЫХ ИЗОБРАЖЕНИЙ

Сычев А.С., магистрант кафедры РТС РГРТУ; alekseisichev2203@gmail.com; Холопов И.С., к.т.н., доцент кафедры РТС РГРТУ; kholopov.i.s@rsreu.ru.

# NO-REFERENCE INTEGRATED-MULTIPLICATIVE QUALITY INDEX FOR DIGITAL GRAYSCALE IMAGES

## Sychev A.S, Kholopov I.S.

The aim of the work is the development of a no-reference integrated-multiplicative normalized quality index of digital grayscale images for evaluating the efficiency of algorithms for improving vision and selecting channels for fusion of images from different sensors. The introduced index operates with estimates of the average brightness, standard deviation of the high-frequency components brightness, standard deviation of noise, mean values of the local contrasts of analyzed image and local contrasts of its low-frequency component. The results of the semi-real experiment showed that, unlike the known integral quality index, the proposed integral-multiplicative index for images decreases rather than increases with a high power of additive white noise.

Key words: no-reference quality index, integral quality index, brightness, standard deviation, histogram, contrast, brightness levels, entropy.

Ключевые слова: безэталонный критерий качества, интегральный показатель качества, яркость, среднеквадратическое отклонение, гистограмма, контраст, уровни яркости, энтропия.

## Введение

В различных информационных системах применяется представление результатов обработки данных в виде изображения, выводимого на устройство отображения для использования наблюдателем (человеком-оператором) – визуализация. Выводимому изображению при этом желательно обеспечить такие качества, благодаря которым его восприятие человеком было бы по возможности комфортным, усилив информативные

особенности наблюдаемой сцены с целью улучшения ее восприятия [1]. Под сценой понимают все объекты, попадающие в кадр фото- и/или видеокамеры [2].

Субъективность восприятия качества изображений затрудняет разработку формализованного критерия для его количественной оценки. Поэтому при обработке изображений с целью их визуализации получили распространение методы, в которых часто отсутствуют строгие математические критерии оптимальности: их заменяют качественные представления о целесообразности тех или иных преобразований, опирающиеся на субъективные оценки результата [1].

Качество изображения зависит от ряда факторов [2]:

- освещенности и контрастности сцены,

 смаза изображения в результате движения камеры или объекта съемки,

– глубины резкости,

– разрядности квантования зарегистрированного сигнала,

 – способа кодирования и степени сжатия при записи в файл.

Цель работы – разработка безэталонного интегрально-мультипликативного нормированного показателя качества цифровых полутоновых изображений для оценки эффективности работы алгоритмов улучшения видения и выбора каналов при комплексировании информации от разноспектральных сенсоров. Введенный показатель оперирует оценками средней яркости, среднеквадратического отклонения яркости высокочастотных компонент и шума, а также средними значениями локальных контрастов анализируемого кадра и его низкочастотной составляющей. Результаты полунатурного эксперимента показали, что в отличие от известного интегрального показателя качества предложенный интегральномультипликативный показатель для изображений с высокой мощностью аддитивного белого шума снижается, а не повышается.

> Численные оценки показателей качества цифровых полутоновых изображений, помимо собственно ранжирования, могут применяться для выбора алгоритмов и параметров преобразований, ориентированных на улучшение визуализации изображений [3, 4], а также выбора наиболее информативных каналов в разноспектральных системах улучшенного видения [5, 6].

## Меры оценки качества цифровых изображений

Показатели качества изображений можно условно разделить на два класса [2]:

– меры сравнения с эталоном, в которых одно изображение *X* считается эталонным, а второе *Y* – преобразованным;

 – безэталонные меры, оперирующие статистиками только текущего анализируемого изображения.

К мерам первого класса относят:

– пиковое отношение сигнал-шум (ПОСШ),

$$q_{\rm m} = 20\log(L_{\rm max}/\sigma_{\rm m}),$$

где  $L_{\max}$  – максимальное значение яркости,  $\sigma_{\max}$  – сред-

неквадратическое отклонение (СКО) шума,

$$\sigma_{\rm ui} = \left[\frac{1}{WH} \sum_{i=1}^{W} \sum_{j=1}^{H} (Y(i,j) - X(i,j))^2\right]^{0.5}$$

У – изображение с аддитивным белым гауссовским шумом (БГШ) с нулевым математическим ожиданием, *W* и *H* – ширина и высота изображения;

- норму Минковского,
- меру структурного подобия SSIM [7],
- статистику Пьеллы [8].

Меры сравнения с эталоном применимы при анализе алгоритмов сжатия изображений и их искусственного искажения при имитационном моделировании, поскольку в обоих случаях имеется исходное изображениеоригинал (эталон).

Во многих практических приложениях эталонное изображение отсутствует: например, в обзорных оптикоэлектронных системах необходимо выполнить преобразования для улучшения визуализации зафиксированного камерой изображения и сразу вывести его на устройство отображения. В этом случае количественными показателями качества изображения могут выступать [9]:

1) средняя яркость

$$\overline{L} = \sum_{i=0}^{255} ih(i) \tag{1}$$

2) СКО яркости по полю изображения

$$\sigma = \sum_{i=0}^{255} (i - \overline{L})^2 h(i),$$
(2)

3) нормированный контраст L<sub>cn</sub>

где  $L_{\min}$  и  $L_{\max}$  – соответственно минимальное и максимальное значения яркости изображения,  $L_m = 2^8 - 1 = 255$  – максимально возможная яркость полутонового восьмибитного изображения;

4) количество информационных уровней яркости

$$N_{\pi} = \sum_{i=0}^{255} \delta(i), \, \delta(i) = \begin{cases} 1, \, h(i) > 0, \\ 0, \, h(i) = 0, \end{cases}$$
(4)

5) энтропия

$$\varepsilon = -\sum_{i=0}^{255} \log_2 \left[ h(i) \right] h(i) \tag{5}$$

и другие [2, 10]. Показатели (1) – (5) могут быть оценены по гистограмме изображения h, где h(i) – значения гистограммы, соответствующие уровню яркости i.

Под нормированным контрастом изображения могут также пониматься:

1) контраст Вебера (Вебера-Фехнера)

 $K_{nW} = |L_{o} - L_{\phi}| / L_{\phi},$ 

где  $L_{o}$  и  $L_{\phi}$  – яркости объекта и фона, на котором объект наблюдается:

2) контраст Михельсона [11]

$$K_{nM} = (L_{max} - L_{min}) / (L_{max} + L_{min}),$$
(6)

где  $L_{\min}$  и  $L_{\max}$  – минимальное и максимальное значения яркости изображения;

3) контраст Воробеля [12]

$$K_{nV} = \frac{\sum_{i=0}^{255} h(i) \left| 2L_i + L_m - \left| 2L_i - L_m \right| \right|}{2L_m},$$
(7)

где  $L_i = i - \overline{L}$ .

В то же время оценка визуального качества изображения по отдельным мерам не является объективной. Так, в [13] отмечается, что ПОСШ при оценке качества изображений может давать отклонения, выявляемые при субъективной оценке изображения, поскольку не учитывает особенностей визуального восприятия. В [9] также делается заключение, что отношение сигнал/шум, кроме наличия ярко выраженного структурного шума, слабо влияет на визуальное восприятие изображений за счет реализуемой человеческим глазом низкочастотной фильтрации. Авторы [14] отмечают, что СКО яркости может слабо меняться при существенном ухудшении субъективно воспринимаемого качества при сжатии изображения, так как, наряду с ПОСШ, не соответствует системе визуального восприятия человека, и потому не может являться объективным критерием качества. Увеличение энтропии реального изображения не всегда эквивалентно увеличению его информативности [15], так как может быть вызвано, например, увеличением СКО шума и не учитывает пространственных зависимостей яркости элементов изображения и особенностей их зрительного восприятия.

По этим причинам целесообразно при оценке визуального качества оперировать комплексными критериями.

# Комплексные безэталонные критерии качества цифровых полутоновых изображений

Комплексные критерии качества формируют единственный (как правило, нормированный на интервале [0, 1]) показатель качества Q из n частных показателей  $\alpha_i$ , i = 1..n, и могут представлять собой их произведение (мультипликативный показатель качества, МПК)

$$Q = w_n \prod_{i=1}^n \alpha_i,$$

либо весовую сумму (интегральный показатель качества, ИПК)

$$Q=\sum_{i=1}^n w_i\alpha_i,$$

где нормирующий множитель  $w_n$  приводит Q к диапазону [0, 1], а для весовых коэффициентов  $w_i$  справедливо

равенство 
$$\sum_{i=1}^{n} w_i = 1$$
.

Примером МПК является эмпирический критерий качества Воробеля [12]:

$$Q_V = w_n L_Q K_Q R_Q K_{nV},$$

где  $w_n$  – весовой коэффициент, нормирующий  $Q_V$ ,  $Q_V \in [0, 1]$ ,  $L_Q$  – оценка уровня адаптации зрительной системы,

$$L_{Q} = 1 - \frac{L - L_{\rm m} / 2}{L_{\rm m} / 2};$$
(8)

Ко-полнота градаций яркости,

 $V_{Q} = S / L_{\rm m},$ 

где S – количество уровней яркостей, для каждого из которых на изображении присутствует не менее чем  $b_{WH}$  количество пикселей с данной яркостью, b – некоторая константа;  $R_Q$  – резкость изображения RO [16], нормированная к максимальной яркости:

$$R_Q = RO / L_{\max},$$

*К<sub>nV</sub>* – контраст Воробеля (7).

В [12, 16] приводится формула для вычисления резкости непрерывной функции яркости *f*(*x*):

$$RO = \frac{\int_{a}^{b} \left(\frac{df}{dx}\right)^{2} dx}{f(a) - f(b)},$$

где *а* и *b* – точки, которые расположены на противоположных краях перепада яркости.

Для цифрового изображения вычисление второй производной яркости реализуется с помощью лапла-

сиана [17] с маской 
$$\begin{vmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{vmatrix}$$
 или  $\begin{vmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{vmatrix}$ .

Недостатком МПК Воробеля является равный вес всех входящих в него частных показателей и близость к нулю  $Q_V$  даже при величине каждого показателя качества около 0,5 (действительно, неравенство  $Q_V < 0,1$  выполняется уже при  $L_Q = K_Q = R_Q = K_{nV} < 0,5$ ).

В работе [9] проблема выбора весов пяти частных показателей (1) – (5) решена предварительным их разбиением на три группы с убывающим приоритетом (математическое ожидание и СКО яркости, контраст, количество градаций яркости и энтропия) и последующим применением метода экспертных оценок и критерия Фишберна. В результате в [9] получено следующее эмпирическое выражение для ИПК:

$$Q_{IIIIK} = 0.33L_n + 0.27s_n + 0.2K_n + 0.13N_n + 0.07\varepsilon_n,$$
(9)

где  $K_n$  – нормированный контраст (3), а  $L_n$ ,  $\sigma_n$ ,  $N_n$  и  $\varepsilon_n$  – соответственно нормированные средняя яркость, СКО, количество уровней и энтропия:

$$L_n = \begin{cases} \overline{L} / 128, \, \overline{L} \le 107, \\ (255 - \overline{L}) / 128, \, \overline{L} > 147, \\ 1, \, \overline{L} \in (107...147). \end{cases}$$
(10)

$$N_n = N_g / 255, \tag{11}$$

$$\varepsilon_n = \varepsilon / 8, \tag{12}$$

$$[\sigma/50, \sigma \le 50,$$

$$\sigma_{n} = \begin{cases} (100 - \sigma) / 50, 50 < \sigma \le 100, \\ 0, \sigma > 100. \end{cases}$$
(13)

ИПК применяется в работах [3, 4, 9] для оценки качества результатов работы алгоритмов улучшения видения.

# Модифицированный интегральномультипликативный показатель качества полутоновых изображений

Значения ИПК при анализе изображений с низким уровнем шума в целом совпадают с субъективными оценками их качества. Однако в случае увеличения СКО шума на изображении статистика (9) стремится к значениям 0,9...1 независимо от сюжета сцены, так как шум увеличивает все входящие в неё частные показатели. В частности, для аддитивной смеси белого гауссовского шума (БГШ) с нулевым математическим ожиданием и СКО  $\sigma$  = 50 и серого фона с яркостью L = 128, не несущих никакой информации, выполняется  $Q_{MTIK}$  = 1.

Вторым недостатком ИПК является то, что оценка СКО (2) является показателем резкости изображения только в том случае, если форма его гистограммы хорошо аппроксимируется гауссовской кривой.

Также к недостаткам ИПК можно отнести разрыв функции  $L_n(\overline{L})$  (10) в точках  $\overline{L} = 107$  и  $\overline{L} = 147$  и статистическую зависимость частных показателей  $N_n$  и  $\varepsilon_n$  (с ростом количества уровней яркости будет увеличиваться и энтропия).

Авторы предлагают оценивать качество полутоновых изображений по частным показателям, для которых можно минимизировать влияние шума сенсора. Для этого принимаются следующие гипотезы:

 шум сенсора описывается математической моделью аддитивного БГШ с нулевым математическим ожиданием;

 формат изображений – полутоновые с глубиной цвета 8 бит.

При принятии данных гипотез вводится интегральномультипликативный показатель качества (ИМПК), который записывается в виде:

$$Q_{HMIIK} = L_{Q}(w_{1}w(q)\sigma_{B^{\prime}n}^{*} + w_{2}\overline{K_{\Lambda\sigma\kappa}^{*}} + w_{3}\overline{K_{\Lambda\sigma\kappa}^{*}})$$
(14)

Частные показатели в (14) рассчитываются следующим образом.

СКО высокочастотной составляющей  $\sigma_{\rm BY}$  оценивается по разности исходного изображения и оценки фоновой составляющей, полученной путем свертки с весовой функцией окна большой апертуры, после чего корректируется с учётом оценки СКО собственного шума  $\sigma_{\rm m}$ :

$$\sigma_{\mathrm{BY}}^{*} = \begin{cases} (\sigma_{\mathrm{BY}}^{2} - \sigma_{\mathrm{u}}^{2})^{0,5}, \sigma_{\mathrm{BY}} > \sigma_{\mathrm{u}}, \\ 0, \sigma_{\mathrm{BY}} < \sigma_{\mathrm{u}}. \end{cases}$$

Для оценки яркости фоновой составляющей в работе применяется процедура быстрого сглаживания исходного изображения ВОХ-фильтром [18] с апертурой 128×128.

Для вычисления нормированного показателя  $\sigma^*_{B^{q_n}}$  по аналогии с [9] применяется формула (13), где в качестве о подставляют  $\sigma^*_{B^{q_n}}$ .

СКО шума  $\sigma_{\rm m}$  оценивается по минимуму локального СКО в секторах *N*×*N* пикселей и умножается на поправочный коэффициент  $w_{\rm m}$ :

$$\sigma_{\rm m} = w_{\rm m} \min_{i} \{\sigma_{\rm loc\,}_{ij}\} \tag{15}$$

где *i* = 1, 2, ..., *W*/*N*, *j* = 1, 2, ..., *H*/*N*. При вычислении (15) из двумерного массива олок іј для предотвращения формирования заниженного значения СКО о<sub>ш</sub> исклю-

чаются локальные оценки s<sub>лок ij</sub>, не удовлетворяющие одному из условий:

$$\begin{split} & 2,5 \mathbf{S}_{_{\textit{\textit{AOK}}\,ij}} < m_{_{\textit{\textit{AOK}}\,ij}} < L_{\mathrm{m}}\text{-}2,5 \mathbf{S}_{_{\textit{AOK}\,ij}} \\ & m_{_{\textit{AOK}\,ij}} \neq 0, \quad m_{_{\textit{AOK}\,ij}} \neq L_{\mathrm{m}}. \end{split}$$

где  $m_{\rm локіј}$  – локальное математическое ожидание яркости,  $L_{\rm nop1}$ ,  $L_{\rm nop2}$  – соответственно нижний и верхний пороги яркости. Необходимость введения дополнительных условий продиктована ограниченным диапазоном значений яркости сенсоров с глубиной цвета 8 бит: от 0 до  $L_{\rm m}$ .

В работе оценка  $\sigma_{\rm m}$  выполнялась в секторах 16×16 пикселей с  $w_{\rm m}$  = 1,2, а пороги яркости  $L_{\rm nop1}$ , и  $L_{\rm nop2}$  были заданы как 0,1 и 0,9 от  $L_{\rm max}$ .

Весовой коэффициент w(q), зависящий от глобального отношения сигнал-шум (ОСШ)  $q = \sigma_{Bq}^* / \sigma_{u}$ , учитывает степень искажения шумом малоразмерных деталей изображения. В работе он задан монотонно возрастающей сигмоидальной функцией:

 $w(q)=1-\exp(-0,2q^2)$ .

Среднее значение локального контраста  $\overline{K_{\text{лок}}}$  определяется по аналогии с [19]:

$$\overline{K_{\rm JOK}} = \overline{LC_{ij}} , \qquad (16)$$

где  $\overline{LC_{ij}}$  – значение локального контраста, вычисляемое по формуле (3), в секторе с номером *ij* размером  $N \times N$  пикселей, *i* = 1, 2, ..., *W*/*N*, *j* = 1, 2, ..., *H*/*N*.

При расчете  $\overline{LC_{ij}}$  локальные контрасты  $LC_{ij} < LC_{nop}$ не учитываются (в работе [19] принят порог  $LC_{nop} = 0,1$ ). При расчете  $\overline{LC_{ij}}$  на изображениях от камер с широкоугольными объективами, для которых характерно виньетирование, также не учитываются секторы, расположенные вблизи границ кадра.

Для изображений с низким контрастом фона, но содержащих в то же время малоразмерные высококонтрастные объекты, (16) можно модифицировать:

$$\overline{K_{_{\pi \text{or}}}} = \begin{cases} \max_{i,j} \{LC_{ij}\}, \max_{i,j} \{LC_{ij}\} < L_{_{\text{nop}}}, \\ [\overline{LC_{ij}} \max_{i,j} \{LC_{ij}\}]^{0.5}, \max_{i,j} \{LC_{ij}\} \ge L_{_{\text{nop}}}. \end{cases}$$
(17)

Поскольку БГШ с нулевым математическим ожиданием увеличивает контраст (16) на  $6\sigma_u / L_m$ , необходимо дополнительно выполнять коррекцию оценок локальных контрастов:

$$\overline{K_{_{\rm JOK}}^*} = \overline{LC_{ij}^*},$$

где  $LC_{ij}^* = LC_{ij} - 6\sigma_{ui} / L_m$ .

Известно [20], что пространственно-частотные характеристики глаза объясняются частично оптическими и частично нервными механизмами. Как оптический инструмент глаз имеет ограниченную разрешающую способность из-за конечных размеров апертуры линзы, оптических аберраций и конечных размеров палочек и колбочек. Эти эффекты в модели одноцветного зрения могут быть представлены фильтром нижних пространственных частот. поскольку наибольший вклад в частотную характеристику глаза вносит механизм латерального торможения [21]. Этим объясняется то, что даже при низких отношениях сигнал-шум наблюдатель хорошо различает протяженные объекты на изображении [9, 22]. По этой причине в ИМПК вводится средний контраст низкочастотной (НЧ) составляющей изображения К<sub>локНЧ</sub>, которую в свою очередь получают кратномасштабным разложением по базисным функциям Хаара [17]. Поскольку НЧ составляющая кратности М имеет в 2<sup>м</sup> раз меньшие по сравнению с исходным изображением ширину и высоту, то локальный контраст LC<sub>ii</sub> для подстановки в (16) или (17) оценивается только для соседних пикселей (по сектору 2×2 пикселя).

При  $\sigma_{\rm BY}^* = 0$  собственный шум будет увеличивать локальный контраст НЧ составляющей кратномасштабного разложения. Преобразование Хаара кратности M уменьшит СКО шума в  $2^M$  раз, а вычисление разности яркостей соседних пикселей для оценки контраста по (3) – увеличит в  $\sqrt{2}$  раз. Предполагая также, что в секторе 2×2 пикселя максимальное значение разности числителя (3) не превысит (4...5)  $\sigma_{\rm m}$ , получим следующую приближенную формулу для  $\overline{K_{\rm лок}}$  с учетом коррекции воздействия шума:

$$\overline{K^{*}_{_{\rm локHY}}} = \begin{cases} \overline{K_{_{\rm локHY}}}, \sigma^{*}_{\rm BY} > 0, \\ \\ \overline{K_{_{\rm локHY}}} - \frac{(4...5)\sqrt{2}\sigma_{_{\rm III}}}{255 \cdot 2^{M}}, \sigma^{*}_{\rm BY} = 0. \end{cases}$$

Для анализируемого в работе уровня кратномасштабного разложения *M* = 3 справедливо приближённое равенство:

$$\overline{K^{*}_{\text{локНЧ}}} \approx \begin{cases} \overline{K_{\text{локНЧ}}}, \sigma^{*}_{\text{BY}} > 0, \\ \overline{K_{\text{локНЧ}}} - 0,003\sigma_{\text{III}}, \sigma^{*}_{\text{BY}} = 0. \end{cases}$$

Коэффициенты  $w_1$ ,  $w_2$  и  $w_3$  определяют вес частных показателей в ИМПК (14). При проведении экспериментов авторами приняты значения  $w_1 = 0.5$ ,  $w_2 = w_3 = 0.25$ .

Принятие гипотезы об аддитивном характере БГШ ограничивает применение ИМПК к изображениям, полученным при малом ОСШ от сенсоров инфракрасного (ИК) диапазона длин волн. Для них характерен регулярный геометрический шум [23, 24], обусловленный неоднородным усилением сигнала в столбцах (строках) матрицы сенсора и образующий характерный рисунок, состоящий из темных или светлых вертикальных (горизонтальных) полос. ИМПК дает для таких изображений завышенную оценку качества из-за увеличения СКО  $\sigma^*_{BY}$ . Завышенное значение ИМПК также характерно при анализе изображений, полученных в результате нелинейных преобразований яркости (например, Multiscale Retinex [25, 26]): для них оценка СКО шума по формуле (15) является заниженной.

## Результаты полунатурных экспериментов

Одной из областей применения безэталонных показателей качества является динамический выбор того



53

или иного метода улучшения визуализации изображений и/или объединения кадров различных спектральных диапазонов [3-6]. Поэтому при проведении экспериментальных исследований выполнялось сравнение ИМПК (14) как с ИПК (9), так и с нормированным показателем информативности (НПИ) разноспектральных изображений  $Q_{HПИ}$  из работ [5, 6], основанным на анализе протяжённых перепадов яркости.

Сравнение ИМПК с ИПК выполнялось по тестовым изображениям «Лена», серому фону с яркостью 128 и изображению взлетно-посадочной полосы (рис. 1) с наложенным на них аддитивным БГШ с СКО  $\sigma_{u.}$ . Численные значения показателей качества, а также меры SSIM приведены под изображениями.

Из полученных результатов видно, что с ростом СКО шума ИМПК уменьшается. Исключение составляет серый фон, на котором шум ошибочно принимается за полезную ВЧ составляющую, поэтому значение ИМПК медленно растёт с ростом СКО шума, оставаясь близким к нулю.

Сравнение ИМПК как с ИПК, так и с НПИ, выполнялось по наборам изображений от телевизионной камеры (TB) и двух камер ИК диапазонов (коротковолнового – SWIR и длинноволнового – LWIR) с совмещенными полями зрения (рис. 2), полученных в ходе лётных испытаний мультиспектральной системы улучшенного видения, выполненных ФГУП «ГосНИИАС» [5].

Из полученных результатов видно, что значения и НПИ, и ИМПК приблизительно совпадают с субъективным восприятием качества. Также следует отметить, что отношение значений ИМПК для кадров различных спектральных диапазонов одной и той же сцены лучше коррелирует с субъективным восприятием изменения качества: например, для ТВ и SWIR изображений первой строки рис. 2  $Q_{HПИ_SWIR}/Q_{HПИ_TB} = 2,21$ , в то время как  $Q_{IMПIK_SWIR}/Q_{IMПIK_TB} = 5,75$ . В то же время для ИК изображений без предварительной обработки (наличие «битых» пикселей, выраженный геометрический шум) значение ИМПК относительно субъективного восприятия качества является завышенным (рис. 2, вторая строка, SWIR диапазон).

#### Заключение

При гауссовском распределении шума результаты оценки качества тестовых изображений с использованием разработанного безэталонного интегральномультипликативного показателя приблизительно совпадают с субъективной оценкой информативности изображения и позволяет ранжировать цифровые изображения по данному критерию. Показано, что с увеличением СКО шума значение ИМПК, в отличие от ИПК, снижается, что также соответствует субъективному восприятию.

В то же время для кадров с низким отношением сигнал-шум, полученных от сенсоров ИК диапазона длин волн, ИМПК дает завышенные по сравнению с НПИ значения. Это связано с отнесением пикселей характерного для ИК сенсоров геометрического шума к высокочастотной составляющей полезного сигнала.

Направлением дальнейшего развития ИМПК являет-

ся применение его для обработки данных с использованием машинного интеллекта: распознавания классов имеющихся на изображении объектов, что потребует введения целевых функций в зависимости от решаемой задачи. Это позволит дополнительно корректировать ИМПК, увеличивая его для изображений, содержащих представляющие целевой интерес объекты: например, людей и автомобили – для поисково-спасательных задач; взлетно-посадочные полосы, трубы, высотные здания и линии электропередач – для систем повышения безопасности полетов и т.п.

#### Благодарности

Авторы выражают благодарность научному сотруднику ООО «РАСТР ТЕХНОЛОДЖИ» Бондаренко М. А. за предоставленные наборы изображений видимого и инфракрасного диапазонов для проведения сопоставительного анализа ИМПК с нормированным показателем информативности [5, 6].

## Литература

1. Грузман И. С., Киричук В. С., Косых В. П., Перетягин Г. И., Спектор А. А. Цифровая обработка изображений в информационных системах: учеб. пособие. Новосибирск: Изд-во НГТУ, 2002. 352 с.

2. Старовойтов В. В., Старовойтов Ф. В. Сравнительный анализ безэталонных мер оценки качества цифровых изображений // Системный анализ и прикладная информатика. 2017. № 1. С. 24-32.

3. Колчаев Д. А., Муратов Е. Р., Никифоров М. Б. Математическое обеспечение системы динамического выбора метода улучшения изображений в реальном времени // Известия ТулГУ. Технические науки. 2017. Вып. 2. С. 83-89.

4. Колчаев Д. А., Муратов Е. Р., Никифоров М. Б. Автоматическая настройка конвейера обработки изображений // Информационные технологии и нанотехнологии: сборник трудов III международной конференции и молодежной школы. Самара: Новая техника, 2017. С. 624-628.

5. Бондаренко А., Бондаренко М. Аппаратнопрограммная реализация мультиспектральной системы улучшенного видения // Современная электроника. 2017. № 1. С. 32-37.

6. Бондаренко М. А., Дрынкин В. Н., Набоков С. А., Павлов Ю. В. Адаптивный алгоритм выбора информативных каналов в бортовых мультиспектральных видеосистемах // Программные системы и вычислительные методы. 2017. № 1. С. 46-52.

7. Wang Z., Bovik A. C., Sheikh H. R., Simoncelli E. P. Image quality assessment: from error visibility to structural similarity // IEEE Trans. on Image Processing. 2004. Vol. 13, Is. 4. P. 600-612.

8. Piella G., Heijmans H. A new quality metric for image fusion // Proc. IEEE International Conference in Image Processing ICIP–2003. V. 3. P. 173-176.

9. Богданов А. П., Романов Ю. Н. Оценка качества цифровых изображений // Техническое зрение в системах управления – 2012: тезисы докладов. М.: ИКИ РАН, 2012. С. 218-226.

10. Pertuz S., Puig D., Garcia M. A. Analysis of focus measure operators for shape-from-focus // Pattern Recognition.

2013. Vol. 46, No. 5. P. 1415-1432.

11. Michelson A. A. Studies in optics. Chicago: The University of Chicago Press, 1927. 164 p.

12. Воробель Р. А., Журавель И. М., Опыр Н. В., Попов Б. О., Дереча В. Я., Равлик Я. М. Метод количественной оценки качества рентгенографических изображений // Неразрушающий контроль и техническая диагностика: труды 3-й украинской научно-технической конференции. Днепропетровск, 2000. С. 233-236.

13. Илюшкина Н., Чобану М. Применение новых критериев оценки качества изображений после их сжатия с потерями // Современная электроника. 2007. № 3. С. 66-69.

14. Конюхов А. Л., Костевич А. Г., Курячий М. И. Критерии оценки отношения сигнал/шум в активно-импульсных телевизионно-вычислительных системах // Доклады ТУСУРа. 2012. № 2. Ч. 1. С. 111-115.

15. Космическое землеведение: информационноматематические основы / под ред. Садовничего В. А. М.: МГУ, 1998. 576 с.

16. Розенфельд А. Распознавание и обработка изображений: пер. с англ. под. ред. Л. С. Лебедева. М.: Мир, 1972. 232 с.

17. Гонсалес Р., Вудс Р. Цифровая обработка изображений. М.: Техносфера, 2006. 1072 с.

18. Jähne B. Digital image processing / 6th ed, revised

and extended. New York: Springer, 2005. 585 p.

19. Tai Yu-W., Brown M. S. Single image defocus map estimation using local contrast prior // 16<sup>th</sup> IEEE Int. Conf. on image processing (ICIP-2009). Cairo, 2009. P. 1797-1800.

20. Прэтт У. Цифровая обработка изображений: пер. с англ. Кн.1. М.: Мир, 1982. 312 с.

21. Ratliff F., Hartline H. K., Miller W. H. Spatial and temporal aspects of retinal inhibitory interaction // J. Opt. Soc. Am. 1963. Vol. 53, No. 1. P. 110-120.

22. Травникова Н. П. Эффективность визуального поиска. М.: Машиностроение, 1985. 128 с.

23. Грузевич Ю. К. Оптико-электронные приборы ночного видения. М.: ФИЗМАТЛИТ, 2014. 276 с.

24. Бехтин Ю. С. Комплексирование зашумленных мультиспектральных изображений с использованием пространственно-ориентированных деревьев вейвлет-преобразования // Цифровая обработка сигналов. 2012. № 1. С. 27-31.

25. Jobson D. J., Rahman Z., Woodell G. A. A multiscale retinex for bridging the gap between color images and the human observation of scenes // IEEE Trans. on image processing. 1997. Vol. 6, No. 7. P. 965-976.

26. Jobson D. J., Rahman Z., Woodell G. A. Properties and performance of a center/surround retinex // IEEE Trans. on image processing. 1997. Vol. 6, Is. 3. P. 451-462.

# Уважаемые коллеги!

Приглашаем Вас принять участие в формировании тематических выпусков журнала «Цифровая обработка сигналов» и размещению рекламы продукции (услуг) Вашей организации на его страницах. В случае положительного решения просим представить в редакцию журнала Ваши предложения по плановому размещению информационных материалов и макет рекламы продукции (услуг) с указанием желаемого её месторасположения: обложка (2-я, 3-я или 4-я стр.), цветная внутренняя полоса (объем полосы).

Журнал «Цифровая обработка сигналов» издается с 1999 года. Выходит ежеквартально, тиражом – 700 экз. Распространяется по подписке через агентство «Роспечать» в России (индекс 82185), СНГ и странах Балтии (индекс 20630), а также на Конференции: «Цифровая обработка сигналов и ее применение – DSPA'».

Научно-технический журнал «Цифровая обработка сигналов» включен в Перечень изданий, рекомендуемый ВАК РФ для публикации результатов научных исследований соискателями ученой степени доктора и кандидата технических наук в области радиотехники, связи, вычислительной техники, электроники, приборостроения, информационных технологий, информационно-измерительных и управляющих систем. По предварительным итогам за 2016 год по рейтингу Science Index базы РИНЦ (3,394) журнал «Цифровая обработка сигналов» занимает 344-ю позицию из почти 3000 представленных изданий. Импакт-фактор журнала за 5-летний период цитируемости составил 0,535!

# Планируемые сроки издания отдельных номеров журнала:

- № 4 декабрь 2018 г. Тематический выпуск: «ЦОС в радиотехнике и системах телекоммуникаций».

– № 1 март 2019 г. Тематический выпуск: «ЦОС в инфокоммуникационных системах».

– № 2 июнь 2019 г. Тематический выпуск по материалам 20-й Международной научно-технической конференции «Цифровая обработка сигналов и ее применение-DSPA»

- № 3 сентябрь 2019 г. Тематический выпуск: «Цифровая обработка изображений».

# Ориентировочная стоимость рекламных услуг:

- 4-я (внешняя) страница цветной обложки - 25 тысяч рублей.

- 2-я и 3-я (внутренние) страницы цветной обложки - 15 тысяч рублей.

- 1\2 цветной внутренней полосы - 8 тысяч рублей.

Ждем Ваших предложений.

С наилучшими пожеланиями, зам. главного редактора

д.т.н., профессор Витязев Владимир Викторович, телефон 8-903-834-81-81. Предложения прошу направлять по адресу: *E-mail*: vityazev.v.v@rsreu.ru или info@dspa.ru

## УДК 004.932

# ОБНАРУЖЕНИЕ И РАСПОЗНАВАНИЕ ВОЗДУШНЫХ ОБЪЕКТОВ НА ОСНОВЕ СОВМЕСТНОЙ ОБРАБОТКИ ИЗОБРАЖЕНИЙ, ФОРМИРУЕМЫХ В РАЗЛИЧНЫХ СПЕКТРАЛЬНЫХ КАНАЛАХ

Муравьев В.С., к.т.н., доцент 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.

# AERIAL OBJECT DETECTION AND RECOGNITION BASED ON MULTISPECTRAL IMAGE FUSION AND PROCESSING

## Muraviev M.S., Smirnov S.A., Strotov V.V.

In this work the approach for aerial object detection, position estimation and recognition based on multispectral image fusion and processing is proposed. Object detection and position estimation is based on image spatial filtering. This approach is expanded on multispectral imaging. The object recognition algorithm is based on outer contour descriptor matching. The source data for the recognition algorithm is a set of binary images that produced by object detection algorithm. The results on the experimental examinations are given. The experimental examinations are performed using a set of natural multispectral video sequences.

Key words: aerial object detection, object position estimation, object recognition, outer contour descriptor, multispectral image processing.

Ключевые слова: обнаружение воздушных объектов, измерение координат объекта, распознавание, дескриптор внешнего контура, мультиспектральное наблюдение.

#### Введение

Основными задачами, решаемыми современными системами технического зрения (СТЗ), содержащими один или несколько датчиков изображений и устройство обработки информации, являются задачи обнаружения, измерения координат и распознавания объектов. Ранее для решения этих задач использовался комплекс методов и алгоритмов, ориентированных

на применение датчика, работающего в одном спектральном диапазоне. Возможности подходов, связанные с анализом информации от одного видеодатчика, во многом приблизились к своему пределу, поэтому естественным является стремление использовать несколько датчиков изображений, работающих в разных спектральных диапазонах. В качестве объектов интереса в работе выступают летательные аппараты, наблюдаемые на фоне ясного или облачного неба.

Для решения задач обнаружения, измерения координат и распознавания объектов широкое распространение получили системы, получающие информацию от монохроматических датчиков видимого и инфракрасного диапазонов. Применение двух датчиков, наблюдающих в разных спектральных диапазонах, позволяет снизить зависимость СТЗ от недостатков датчиков, работающих в одном спектральном диапазоне. При этом использование мультиспектральных датчиков неизбежно приводит к увеличению объема исходных данных и влечет за собой необходимость объединения информа-

Рассмотрен подход к обнаружению с последующим измерением координат и распознаванием воздушных объектов на основе совместной обработки изображений, формируемых в различных спектральных каналах. Обнаружение и измерение координат объектов базируется на пространственной фильтрации изображений и является обобщением известного алгоритма на случай мультиспектрального наблюдения. Предлагаемый подход также ориентирован на распознавание воздушных объектов на основе описания изображения объекта с помощью дескрипторов внешнего контура. Исходными данными для алгоритма распознавания являются бинарные изображения, формируемые алгоритмом обнаружения. Представлены результаты экспериментальных исследований, выполненные с использованием натурных видеосюжетов.

> ции для извлечения более полезных сведений об объекте интереса.

Объединение информации может осуществляться на трех уровнях обработки изображений: комплексирование изображений, объединение признаков и объединение решений [1, 2]. При комплексировании объединённое изображение используется для дальнейшей обработки с использованием традиционных подходов к анализу изображений. Для комплексирования могут использоваться подходы на основе преобразования главных компонент [3, 4], вейвлет-преобразования [5], марковских моделей [6] и т.д. На втором уровне обработки осуществляется объединение признаков, полученных в каждом спектральном диапазоне, после чего решаются задачи обнаружения и распознавания объектов. Объединение информации может осуществляться на основе подходов, базирующихся на байесовской теории обнаружения, теории Демпстера – Шафера, теории нечетких множеств, теории искусственных нейронных сетей [7]. На третьем уровне обработки анализ

изображений выполняется независимо в каждом спектральном канале с последующим объединением результатов. Основные методики, применяемые в данном случае, являются развитием и обобщением методик, используемых при обнаружении и распознавании объектов при моноспектральном наблюдении [8,9].

Анализ отечественной и иностранной литературы показывает, что наиболее перспективными направлениями являются комплексирование информации на втором и третьем уровнях обработки изображений. В данной работе рассматривается подход к объединению информации от нескольких датчиков на этапе обнаружения объектов. Для обнаружения и измерения координат воздушного объекта при мультиспектральном наблюдении предлагается использовать алгоритм, полученный при учете модели наблюдения и ряда ограничений, вытекающих из предположений о характере фоноцелевой обстановки. Описываемый подход является дальнейшим развитием идей, изложенных в [10, 11]. После обнаружения объекта, формирования бинарной маски, задающей положение точек объекта, и измерения параметров полученного бинарного сегмента, решается задача распознавания. Для распознавания предлагается использовать подход на основе описания изображения объекта с помощью дескрипторов внешнего контура [12].

Алгоритм обнаружения и измерения координат воздушного объекта. Пусть наблюдение за объектом производится в видимом и инфракрасном спектральных диапазонах. Датчики съюстированы по углам зрения и имеют одинаковое разрешение  $N_x$  по горизонтали и  $N_y$  по вертикали, а информация поступает с одинаковой частотой и синхронизирована по времени. На каждом кадре наблюдаемое изображение в первом канале будет описываться функцией яркости  $l_1(i, j)$ , а во втором канале – функцией яркости  $l_2(i, j)$ . При решении задачи обнаружения требуется сделать выбор между гипотезой  $X_1$  – объект присутствует в кадре, и альтернативой  $X_0$  – объект отсутствует.

При использовании одного датчика изображений в [10, 14] была предложена пространственная модель наблюдения воздушных объектов, которая может быть расширена на случай мультиспектрального наблюдения. Будем считать, что формирование изображений происходит независимо в разных каналах наблюдения. Запишем модель формирования мультиспектрального изображения:

$$l_{q}(i, j) = h_{q}(i, j)r(i, j) + g_{q}(i, j)(1 - r(i, j)) + \xi_{q}(i, j), q = 1, 2,$$

$$(i, j) \in R, R = \{(i, j) : i = \overline{0, N_{x} - 1}, j = \overline{0, N_{y} - 1}\},$$
(1)

где функции  $h_q(i, j)$ ,  $g_q(i, j)$ ,  $\xi_q(i, j)$  – задают соответственно эталонное изображение объекта, фона и шумовой процесс в *q*-м спектральном канале; r(i, j) – индикаторная функция, значения которой равны 1 в точках, принадлежащий объекту, и 0 в противном слу-

чае. Значения шумового процесса  $\xi_q(i,j)$  некоррелированы по пространственным координатам и подчиняются нормальному распределению с нулевым средним и дисперсией  $\sigma_q^2$ .

При наблюдении удаленного объекта на фоне неба яркости объекта и фона можно приближенно считать случайными величинами, не зависящими от пространственных координат [10]. Априорные плотности распределения яркостей объекта  $h_q$  и фона  $g_q$  в канале q зададим в виде  $p(h_q) = p(g_q) = 1/(l_{q \max} - l_{q \min})$ , при  $h_q = [l_{q \min}, l_{q \max}]$ ,  $g_q = [l_{q \min}, l_{q \max}]$ , где  $l_{q \min}$  и  $l_{q \max}$  – нижняя и верхняя границы диапазона яркостей наблюдаемого изображения в q-м канале. Наблюдение производится в условиях достаточной освещенности, поэтому яркость точек фона и объекта много больше значений  $\sigma_q^2$ .

Местоположение объекта на изображении неизвестно, но допустим, что известна его конфигурация, определяемая множеством  $H_0$ . Координаты точек объекта на наблюдаемом изображении зададим множеством  $H_{(\alpha,\beta)} = \{(i,j) \mid (i-\alpha,j-\beta) \in H_0\},$  где координаты  $(\alpha,\beta)$  определяют возможное местоположение объекта.

С целью упрощения изложения, наблюдаемые изображения удобно представить в виде векторов  $\mathbf{l}_i(n), i = \overline{\mathbf{l}, N}$ , содержащих элементы, записанные в лексикографическом порядке. В соответствии с байесовской теорией обнаружения и оценивания [13] правило принятия решения о наличии объекта состоит в сравнении отношения правдоподобия  $\Lambda = p_{cp}(\mathbf{l}_1, \mathbf{l}_2 \mid X_1) \div p_{cp}(\mathbf{l}_1, \mathbf{l}_2 \mid X_0)$  с заданным порогом *C*, где  $p_{cp}(\mathbf{l}_1, \mathbf{l}_2 \mid X_1), \quad p_{cp}(\mathbf{l}_1, \mathbf{l}_2 \mid X_0)$  – усредненные по неизвестным параметрам плотности распределения процессов  $\mathbf{l}_1 = [l_1(i, j)], \quad \mathbf{l}_2 = [l_2(i, j)], i = \overline{\mathbf{0}, N_x - \mathbf{l}}, j = \overline{\mathbf{0}, N_y - \mathbf{l}}$  при наличии и отсутствии объекта.

Принимая во внимание модель наблюдения, характер функций яркости объекта и фона и условие независимости, можно явно задать функции правдоподобия  $p(\mathbf{l}_q | X_1), \quad p(\mathbf{l}_q | X_0), \quad q = 1,2$  при условии наличия и отсутствия объекта в аналитическом виде:

$$p(\mathbf{l}_{q} \mid X_{1}) = k_{q} \exp\left(-\frac{1}{2\sigma_{q}^{2}} \times \left(\sum_{(i,j)\in G_{(\alpha,\beta)}} (l_{q}(i,j) - g_{q})^{2} + \sum_{(i,j)\in H_{(\alpha,\beta)}} (l_{q}(i,j) - h_{q})^{2}\right)\right); \quad (2)$$

$$p(\mathbf{l}_{q} \mid X_{0}) = k_{q} \exp\left(-\frac{1}{2\sigma_{q}^{2}} \sum_{(i,j)\in R} (l_{q}(i,j) - g_{q})^{2}\right), q = 1, 2,$$

где  $k_q = (2\pi\sigma_q^2)^{-N_xN_y/2}$ ,  $G_{(\alpha,\beta)} = R \setminus H_{(\alpha,\beta)}$  – множество точек фона.

Максимальная вероятность правильного обнаружения объекта при фиксированной вероятности ложной тревоги, не превышающей величины  $P_{un}$ , достигается при использовании критерия Неймана-Пирсона [13]. Опуская детали математического вывода, в соответствии с данным критерием правило принятия решения об обнаружении объекта при объединении мультиспектральных изображений запишется в виде:

$$\frac{1}{N_{o\delta}} \sum_{\alpha} \sum_{\beta} \exp\left(\frac{S_g S_h}{2N_x N_y} (SNR_1^2(\alpha, \beta) + +SNR_2^2(\alpha, \beta))\right) > C, \qquad (3)$$

где  $SNR_q(\alpha, \beta) = |\hat{h}_q(\alpha, \beta) - \hat{g}_q(\alpha, \beta)| / \sigma_q$  — оценка отношения сигнал/шум в *q*-м канале при наличии объекта в координатах  $(\alpha, \beta)$ ,  $C = C(P_{xm})$  — некоторая константа, зависящая от допустимой величины ложной тревоги,

$$\hat{h}_q(\alpha, \beta) = \sum_{(i,j) \in H_{(\alpha,\beta)}} l_q(i,j) / S_h,$$
  
 $\hat{g}_q(\alpha, \beta) = \sum_{(i,j) \in G_{(\alpha,\beta)}} l_q(i,j) / S_g$  — средние значения ярко-

стей объекта и фона для координат ( $\alpha$ ,  $\beta$ ),  $S_h$ ,  $S_g$  – число точек объекта и фона соответственно,  $N_{oo}$  – количество возможных положений объекта на изображении.

При решении задачи обнаружения объекта проверяется гипотеза о наличии или отсутствии объекта на изображении. В практической ситуации необходимо также знать конкретное местоположение объекта с целью измерения его параметров и распознавания.

Для нахождения оценок координат объекта  $(\hat{\alpha}_{o\delta}, \hat{\beta}_{o\delta})$  может использоваться алгоритм, построенный на основе критерия максимального правдоподобия. Сложная гипотеза  $X_1$  о наличии объекта в кадре при условии независимости наблюдений принимается в том случае, если выполняется неравенство:

$$\frac{\max_{a,\beta,g_1,h_1,g_2,h_2} p(\mathbf{l}_1 \mid X_1) p(\mathbf{l}_2 \mid X_1)}{\max_{g_1,g_2} p(\mathbf{l}_1, \mid X_0) p(\mathbf{l}_2 \mid X_0)} > C.$$
(4)

При подстановке в (4) выражений (2) и оценок максимального правдоподобия неравенство можно преобразовать к виду:

$$\max_{\alpha,\beta}(SNR_1^2(\alpha,\beta) + SNR_2^2(\alpha,\beta)) > C_1, \qquad (5)$$

где коэффициент  $C_1 = 2 \ln(C) N_x N_v / S_h S_g$ .

Таким образом, алгоритм обнаружения объекта состоит в максимизации некоторой функции, стоящей в левой части (5), и сравнении значений с порогом при всех возможных положениях объекта. В случае наличия объекта его местоположение будет определяться в координатах, доставляющих максимум этой функции.

Заметим, что алгоритм обнаружения объектов получен на основании предположения о постоянстве фоновой составляющей. Однако данный алгоритм допускает модификацию, при которой возможна его эффективная работа, если  $g_q(i,j)$  является изменяющимся процессом. При этом условно можно считать, что фон постоянен в пределах ограниченной области. Размеры этой области будут зависеть от яркостной изменчивости фоновой составляющей и площади изображения объекта.

В общем случае размеры объекта и область постоянства фона могут отличаться в разных спектральных каналах наблюдения. Форма удаленных объектов может быть принята близкой к прямоугольной. В большинстве случаев значения  $\sigma_q^2$  меняются с течением времени и необходимо вычислять их оценки на каждом кадре.

Рассмотрим последовательность действий, выполняемую на этапе обнаружения воздушных объектов при мультиспектральном наблюдении, с учетом выводов теоретических исследований и приведенных выше замечаний более подробно.

Изображение  $l_q(i,j)$ , сформированное в спектральном канале q, обрабатывается усредняющим фильтром, имеющим маску  $w_{o \delta}^{(q)}$  размерностью  $a_{o \delta}^{(q)} \times a_{o \delta}^{(q)}$ . Параллельно с этим  $l_q(i,j)$  сглаживается фильтром с маской  $w_{\phi}^{(q)}$  размерностью  $a_{\phi}^{(q)} \times a_{\phi}^{(q)}$ ,  $a_{\phi}^{(q)} > a_{o \delta}^{(q)}$  вида:

$$w_{\phi}^{(q)}(i_{m}, j_{m}) = \begin{cases} 0, \text{ при } i_{m}, j_{m} = \\ = -(a_{o\delta}^{(q)} - 1)/2, (a_{o\delta}^{(q)} - 1)/2, \\ 1/(a_{\phi}^{(q)} - a_{o\delta}^{(q)})^{2}, \text{ иначе.} \end{cases}$$
(6)

Результат фильтрации с маской  $w_{\phi}^{(q)}$  представляет собой оценку фоновой составляющей наблюдаемого изображения в соответствующем канале. Сглаживание с маской  $w_{oo}^{(q)}$  служит целью усреднения яркости изображения объекта и уменьшения степени зашумления. Затем в спектральном канале q формируется изображение  $d_q(i, j)$ , формируемое путем вычитания результата фильтрации фона из результата операции сглаживания яркости объекта. Во многих случаях математическое ожидание  $d_q(i, j)$  после компенсации фона будет близко к нулю. Итоговое решающее правило задается в виде:

$$\hat{r}(i,j) = \begin{cases} 1, \ \frac{d_1^2(i,j)}{\hat{\sigma}_1^2} + \frac{d_2^2(i,j)}{\hat{\sigma}_2^2} > k^2, \\ 0, \ uhave, \end{cases}$$
(7)

где k – пороговый коэффициент,  $\hat{r}(i,j)$  – бинарное

изображение, 
$$\hat{\sigma}_q^2 = rac{1}{N_x N_y} \sum_{i=0}^{N_x-1} \sum_{j=0}^{N_y-1} d_q^2(i,j)$$
 – дисперсия

изображения, вычисленная во всем кадре. Таким образом, разности  $d_q(i,j)$  в каждом канале возводятся в квадрат и нормируются на соответствующую дисперсию  $\sigma_q^2$ , после чего выполняется операция сложения. В случае если полученное значение в точке (i, j) превосходит порог, то принимается решение о принадлежности этой точки объекту, иначе – фону.

Зачастую на изображениях, формируемых датчиками ИК-диапазона, интересующие потребителя воздушные объекты имеют большую яркость по сравнению с окружающим фоном. Эту особенность можно учесть путем наложения дополнительного ограничения на решающее правило (7). При моноспектральном наблюдении описанный подход сводится к известному пространственному алгоритму обнаружения воздушных объектов, описанному в [10]. Для измерения координат объектов на бинарном изображении  $\hat{r}(i, j)$ , являющейся оценкой r(i, j), можно использовать известные алгоритмы разметки и параметризации [14].

Описанный алгоритм обнаружения и измерения координат объектов получен на основе предположения о том, что форма удаленных объектов близка к прямоугольной. Однако проведенные исследования показали, что данный подход работоспособен и в том случае, когда маска объекта заранее неизвестна, а его размеры неточечные. При этом алгоритм показывает хорошие результаты для малоразмерных и среднеразмерных летательных аппаратов различных конструкций. В случае, когда размеры объекта значительно превышают точечные, актуальной является задача распознавания типа объекта.

# Критерий работоспособности алгоритма обнаружения воздушных объектов

В статье [15] предложен критерий работоспособности для алгоритма обнаружения при моноспектральном наблюдении, позволяющий оценить его эффективность при текущих условиях наблюдения. Критерий основан на оценке амплитуды яркостного импульса, представляющего объект на изображении:

$$KF_{ss} = \frac{\max_{(i,j)\in H} |d(i,j)|}{\hat{\sigma}},$$
(8)

где d(i, j) – яркость разностного изображения в точке (i, j), принадлежащей области, ограничивающей объект,  $\hat{\sigma}$  – оценка СКО фона, H – множество точек объекта.

По аналогии с моноспектральным случаем критерий работоспособности для алгоритма обнаружения воздушных объектов на основе совместной обработки изображений, формируемых в двух спектральных каналах, можно записать в виде:

$$KF_{ms} = \frac{\max_{(i,j)\in H} |d_1(i,j)|}{\hat{\sigma}_1} + \frac{\max_{(i,j)\in H} d_2(i,j)}{\hat{\sigma}_2}.$$
 (9)

Алгоритм распознавания воздушных объектов. Пусть известно бинарное изображение воздушного объекта. Необходимо распознать наблюдаемый объект, т.е. определить тип воздушного объекта из известного набора объектов: самолёты, вертолёты, беспилотные аппараты. В качестве бинарного изображения может выступать изображение  $\hat{r}(i, j)$ , сформированное алгоритмом, описанном в предыдущем разделе.

Рассматриваемый алгоритм распознавания воздушных объектов состоит из двух этапов. Первый этап – подготовка базы данных эталонных объектов. На этом этапе происходит расчет дескрипторов эталонных изображений. Эталонные изображения формируются на основе 3D-модели исследуемого объекта. Предполагается, что для всех объектов, из рассматриваемого множества, известна 3D-модель, которая используется в дальнейших вычислениях. На втором этапе для определения типа объекта находится дескриптор эталонного изображения, который в наибольшей степени соответствует дескриптору наблюдаемого изображения.

Изображения объекта могут меняться в зависимости от ориентации объекта, таким образом, для решения задачи распознавания объектов по их изображениям необходимо рассматривать объект на дискретном множестве ориентаций, максимально полно охватывающим все возможные ракурсы наблюдения. Размещая 3Dмодель объекта в центр дискретизированной сферы, а наблюдателя на ее поверхность, то для каждой точки дискретизированной сферы можно сформировать соответствующую пространственную ориентацию объекта интереса. При проведении исследований в [12] было показано, что для получения равномерного распределения точек на сфере необходимо использовать метод дискретизации по принципу геосферы. Таким образом, для каждой 3D-модели объекта формируется заданное количество изображений.

В качестве дескриптора изображения в разработанном алгоритме используется дескриптор внешнего контура объекта, так как он обладает минимальными вычислительными требованиями, а также инвариантен к повороту изображения. Данный дескриптор рассчитывается на основании бинарного изображения объекта или его контура. В первом случае необходимо извлечь внешний контур из бинарного изображения. Точки контура преобразуются в полярную систему координат с началом в центре бинарного изображения объекта. Далее к координатам преобразованного контура применяются процедуры линейной интерполяции и медианной фильтрации [12, 16]. В результате чего формируется дискретный дескриптор внешнего контура. На рис. 1 приведены: эталонное изображение летательного аппарата, внешний контур объекта и соответствующий ему дескриптор.

Дескриптор контура можно вычислить, используя следующее уравнение:

$$D(i) = F_{med} \left\{ \max\left( d(P^{center}, P\left(\frac{2\pi i}{N_D}, \frac{\pi}{N_D}\right) \right) \right\},$$
(10)

где  $i = \overline{1, N_D}$  – номер текущего дескриптора;  $N_D$  – размер дескриптора;  $d(P_1, P_2)$  – евклидово расстояние между  $P_1$  и  $P_2$ ;  $P^{center}$  – координаты центра тяжести объекта;  $P(\alpha, \Delta \alpha)$  – точка контура объекта, расположенная в секторе круга, которая ограничена углами  $\alpha \pm \Delta \alpha$  (окружность центрирована относительно  $P^{center}$ );  $F_{med} \{ \cdots \}$  – символьная запись операции медианной фильтрации.

Для всех 3D-моделей объектов формируется эталонные дескрипторы, составляющие базу данных. В данной базе каждому дескриптору внешнего контура поставлена в соответствие метка класса воздушного объекта (самолет, вертолет или беспилотный летательный аппарат).

Первый этап алгоритма распознавания является вычислительно затратным в виду того, что включает мно-



Рис. 1. а) Эталонное изображение, б) внешний контур, в) дескриптор внешнего контура тестового летательного аппарата







a)



Рис. 2. Характерные кадры тестовых пар видеосюжетов

жество сложных операций, связанных с обработкой 3Dмоделей (например, рендеринг изображений). Однако это обстоятельство не накладывает ограничений на аппаратную часть, так как данный этап осуществляется заранее и может быть выполнен на стационарном компьютере.

На втором этапе алгоритма непосредственно осуществляется процесс распознавания воздушных объектов. Данный этап также базируется на описании выделенных контуров изображения с помощью дескрипторов. Этот этап выполняется в режиме реального времени. Определение сходства дескриптора наблюдаемого изображения с эталонными дескрипторами из базы данных выполняется на основе вычисления критериальной функции:

$$f_{crit}(j) = \min_{s \in S} \left[ \sum_{i=1}^{N_D} (D_0(i-s) - D_j(i))^2 \right],$$

$$S = \{s : s = \overline{1, N_D} \}$$
(11)

где  $D_0$  – дескриптор контура наблюдаемого объекта,

*D<sub>i</sub>* – текущий эталонный дескриптор из базы данных, и

*s* – величина циклического сдвига дескриптора. Использование данного критерия позволяет добиться инвариантности к повороту изображения.

В результате вычисления функции (11) для каждого эталонного дескриптора из базы данных формируется вектор значений:



Рис. 3. Графики зависимостей  $P_{no} = f(P_{ne})$  и  $K_{n} = f(P_{ne})$ для мультиспектральных пар: а) сюжет № 1, б) сюжет № 2

$$M = (f_{crit}(j) | j = \overline{1, N_g})$$
 (12)  
Величиной  $R_{\iota}$ , характеризирующей степень соот-

ветствия объекта интереса k -му объекту из базы данных, будем считать минимальную величину расстояния (11) между дескриптором наблюдаемого объекта и эталонным дескриптором k -го эталонного объекта, т.е. значение минимального элемента вектора M:

$$R_{k} = \min(f_{crit}(j)), \quad j = 1, N_{g}.$$
 (13)

Распознавание типа объекта осуществляется путем поиска номера k, которому соответствует минимальное значение величины  $R_k$ :

$$match = \arg\min(R_k), k = 1, K, \qquad (14)$$

где *match* – индекс эталона, наиболее схожего с объектом интереса; *К* – общее число рассматриваемых эталонных объектов. Таким образом, наблюдаемый объект следует отнести к классу объектов, имеющих в базе эталонов индекс *match*.

# Экспериментальные исследования алгоритмов обнаружения и распознавания

Первая цель экспериментальных исследований состояла в сравнительном анализе эффективности работы мультиспектрального алгоритма обнаружения объектов и алгоритма, описанного в [10], ориентированного на моноспектральный случай. Для проведения сравнительных исследований была сформирована база данных, содержащая 4 мультиспектральных видеопоследовательности. В базу данных вошли мультиспектральные пары сюжетов общей продолжительностью 2х40350 кадров. Сюжеты были съюстированы и синхронизированы во времени. Размер объектов варьировался от примерно 10х10 пикселей до 70х25 пикселей. В сюжетах наблюдались воздушные объекты на фоне ясного неба или облачного неба с перепадами яркости. На рис. 2 приведены кадры из двух пар видеосюжетов.

В целях оценки точности обнаружения объектов использовались кривые зависимости частоты правильного обнаружения точек изображения  $P_{no}$  от частоты ложно-

го выделения точек изображения  $P_{\rm ref}$ . Указанная зависимость строились при варьировании порогового коэффициента k алгоритмов обнаружения. Чем больше значение  $P_{\rm ref}$  при фиксированной величине  $P_{\rm ref}$ , тем выше эффективность сравниваемого алгоритма.

На рис. 3, а и 3, б приведены графики изменения показателей эффективности, полученные при обработке пары сюжетов № 1 и № 2 с помощью мультиспектрального алгоритма обнаружения объектов (MS) и алгоритма обнаружения, описанного в [10], в видимом (TV) и ИК (IR) диапазонах.



Рис. 4. Примеры кадров из расширенной базы видеосюжетов

Анализ полученных результатов показал, что на рассматриваемом наборе сюжетов частота правильного обнаружения  $P_{no}$  для мультиспектрального алгоритма обнаружения объектов была выше по сравнению с алгоритмом обнаружения, ориентированного на моноспектральный случай наблюдения. Частота правильного обнаружения объектов в целом не хуже 0,9 при частоте ложных тревог 0,05. На паре сюжетов № 1 эффективность работы мультиспектрального алгоритма обнаружения была заметно выше. Малые значения частоты ложного выделения





Рис. 5. а) Изображение самолета №5, б) бинарное изображение самолета, сформированное алгоритмом обнаружения,
 в) дескриптор наблюдаемого изображения, г) дескриптор эталонного изображения из базы данных,
 д) исходное эталонное изображение, е) эталонное изображение, повернутое на 70° против часовой стрелки

объясняются небольшим числом белых пикселей, отнесенных к площади фоновой области. На остальных сюжетах характеристики обнаружения были очень близкими с приростом частоты правильного обнаружения на несколько процентов при *k* >4,5. К числу достоинств алгоритма можно причислить выраженное улучшение качества связности бинарной маски объекта.

Можно выделить несколько основных недостатков алгоритма обнаружения объектов при мультиспектральном наблюдении. Во-первых, существует проблема, связанная с объединением информации на этапе пороговой обработки. Она заключается в сложностях подбора размеров окон фильтров в разных спектральных каналах по итоговому бинарному изображению. Вовторых, мультиспектральный алгоритм предъявляет довольно высокие требования к точности юстировки. При этом нельзя полностью исключать возможность постепенного ухода параметров юстировки датчиков в СТЗ вследствие механических воздействий, возникающих в процессе эксплуатации.

Вторая цель экспериментальных исследований заключалась в анализе качества алгоритма распознавания объектов. В качестве критерия, позволяющего оценить качество распознавания, выступала частота правильного распознавания. Используемая для исследования алгоритма обнаружения база содержала сюжеты, в которых наблюдались самолеты, практически не меняющие ракурс во время полета. Поэтому для увеличения количества типов воздушных объектов при проведении исследований был расширен набор видеосюжетов за счет добавления ряда видеопоследовательностей, сформированных одним датчиком изображений (рис. 4). При отсутствии информации в одном из спектральных каналов, алгоритм обнаружения сводится к случаю, описанному в [10]. Размер наблюдаемых воздушных объектов составлял 30х30 пикселей и выше.

Для формирования эталонных изображений и дескрипторов в рамках эксперимента использовались база 3D-моделей 17 воздушных объектов: самолеты, вертолеты и БПЛА. Для каждой 3D-модели генерировались наборы эталонных изображений, расположенных в ориентациях, распределенных по геосфере фактора 3 (92 точки). Генерация изображений происходила в условиях фронтального освещения.

На рис. 5 приведен пример работы алгоритма распознавания. На рис. 5, а представлена часть кадра №100 сюжета, содержащего видеозапись полета самолета. На рис. 5, б приведено бинарное изображение объекта. На рис. 5, в и 5, г приведены дескриптор наблюдаемого изображения и соответствующий эталонный дескриптор для модели самолета № 5. На рис. 5, д представлено эталонное изображение объекта, соответствующее дескриптору, приведенному на рис. 5, г. Сходство данного изображения с изображением на рис. 5, б будет более очевидно, если его повернуть на 70° против часовой стрелки (рис. 5, е).

Результаты экспериментальных исследований для 23 рассматриваемых сюжетов приведены в табл. 1.

Номер	Тип	Наблюдаемый	Частота	
сюжета	сюжета	объект	правильного	
			распознавания, %	
1	TB	Самолет №1	97,2	
2	TB	Самолет №2	93,1	
3	TB	БПЛА №1	87,1	
4	TB	БПЛА №2	80,6	
5	TB	Вертолет №1	85,6	
6	ИК	Самолет №3	100	
7	TB	Самолет №3	95	
8	ТВ+ИК	Самолет №3	100	
9	ИК	Самолет №4	92,3	
10	TB	Самолет №4	83,2	
11	ТВ+ИК	Самолет №4	95,1	
12	ИК	Самолет №3	91,4	
13	TB	Самолет №3	86,7	
14	ТВ+ИК	Самолет №3	93,1	
15	ИК	Самолет №4	89,2	
16	TB	Самолет №4	80,1	
17	ТВ+ИК	Самолет №4	92,1	
18	TB	Самолет №5	100	
19	TB	Самолет №6	95,2	
20	TB	БПЛА №1	88,7	
21	TB	Вертолет №2	100	
22	TB	Вертолет №3	98,1	
23	TB	БПЛА №2	91,3	

Таблица 1 — Результаты исследований алгоритма распознавания воздушных объектов

## Заключение

Предлагаемые в данной работе алгоритмы подходят для обнаружения и распознавания воздушных объектов, наблюдаемых на фоне ясного неба или облачного неба с перепадами яркости. Эксперименты показывают повышение качества решения задачи обнаружения объекта при совместной обработке изображений, формируемых в различных спектральных каналах. Алгоритм распознавания воздушных объектов позволяет определить тип объекта: «самолет», «вертолет» или «БПЛА». Исходными данными для алгоритма распознавания является бинарное изображение, сформированное алгоритмом обнаружения. В целом, точность обнаружения и распознавания объектов в среднем превышает 90 %, но зависит от типа объекта и его ориентации в пространстве.

Исследования выполнены при использовании Стипендии Президента РФ СП-2598.2016.5.

## Литература

1. J. Dong, D. Zhuang, Y. Huang and J. Fu. Advances in Multi-Sensor Data Fusion: Algorithms and Applications // Sensors. – 2009. – №9(10). p. 7771-7784.

2. Lanir J. Maltz M., Rotman S.R. Comparing multispectral image fusion methods for a target detection task // Optical Engineering. – 2007. – Vol. 46(6). – P. 066402-1–066402-8

3. Hailiang Shi., Baohui Tian, Yuanzheng Wang Fusion of multispectral and panchromatic satellite images using Principal Component Analysis and Nonsubsampled Contourlet Transform / Seventh International Conference on Fuzzy Systems and Knowledge Discovery (FSKD). – 2010. – PP. 2312 – 2315.

4. Mitianoudis N., Stathaki T. Adaptive image fusion using ICA bases / Proceedings of the International Conference on Acoustics, Speech and Signal Processing. – Toulouse, 2006. – PP. II-829–II-832.

5. Kaarna A. Integer PCA and wavelet transforms for multispectral image compression / IEEE 2001 International Geoscience and Remote Sensing Symposium (IGARSS). – 2001. – Vol.4. – PP. 1853 – 1855.

6. A. Sarkar et al. A MRF model-based segmentation approach to classification for multispectral imagery // IEEE Transactions on Geoscience and Remote Sensing. – 2002. – Vol. 40, Issue 5. – pp. 1102-1113.

7. F. Samadzadegan. Data integration related to sensors, data and models // ISPRS Congress, Vol. XXXV, Proceedings of Commission IV, Istanbul, Turkey, 2004.– p. 569-574.

8. Vidya Manian, Luis O. Jimenez, Land cover and benthic habitat classification using texture features from hyperspectral and multispectral images // Journal of Electronic Imaging 16(2), 023011 (Apr–Jun 2007), pp. 1-12.

9. 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.

10. Алпатов Б.А. Оптимальное оценивание параметров движущегося объекта в последовательности изображений // Автометрия. – 1994. – № 2. – С. 32-37.

11. Муравьев В.С., Муравьев С.И. Алгоритм выделения и измерения координат объектов, наблюдаемых на облачных фонах // Вестник РГРТУ. – 2007. – № 21– С. 20-24.

12. Алпатов Б.А., Бабаян П.В., Смирнов С.А., Масленников Е.А., Алгоритм предварительного оценивания пространственной ориентации объекта с помощью дескриптора внешнего контура // Цифровая обработка сигналов. – 2014. – № 3. – С. 43-46.

 Репин В.Г., Тартаковский Г.П. Статистический синтез при априорной неопределенности и адаптация информационных систем. – М.: Советское радио, 1977. – 432 с.

14. Методы автоматического обнаружения и сопровождения объектов. Обработка изображений и управление / Алпатов Б.А., Бабаян П.В., Балашов О.Е., Степашкин А.И. – М.: Радиотехника, 2008. – 176 с.

15. Алпатов Б.А., Бабаян П.В., Смирнов С.А. Комбинированный алгоритм слежения за воздушными объектами // Вестник Рязанского государственного радиотехнического университета. – 2011. – № 3 (37). – С.7-12.

16. Muraviev V.S., Smirnov S.A., Strotov V.V. Aerial vehicles detection and recognition for UAV vision system // Computer Optics. – 2017. – Vol. 41. – №. 4. – pp. 545-551.

## УДК 004.932.2

# СОВРЕМЕННЫЕ СВЕРТОЧНЫЕ НЕЙРОННЫЕ СЕТИ ДЛЯ ОБНАРУЖЕНИЯ И РАСПОЗНАВАНИЯ ОБЪЕКТОВ

Ерохин Д.Ю., аспирант Рязанского государственного радиотехнического университета (РГРТУ), e-mail: erokhin.d.y@gmail.com;

Ершов М.Д., аспирант, ассистент кафедры автоматики и информационных технологий в управлении РГРТУ, e-mail: aitu@rsreu.ru.

# MODERN CONVOLUTIONAL NEURAL NETWORKS FOR OBJECT DETECTION AND RECOGNITION

# Erokhin D.Y., Ershov M.D.

This paper contains a comparison of different neural network architectures that are used to solve the problem of object detection and recognition. Modern artificial neural networks are able to detect and localize objects of known classes. This allows them to be used in various technical vision systems. In this work we compare three architectures (YOLO, Faster R-CNN, SSD) by the following criteria: processing speed, mAP, precision and recall.

Key words: intelligent systems, image processing, object detection, pattern recognition, neural networks, machine learning.

Ключевые слова: интеллектуальные системы, обработка изображений, обнаружение объектов, распознавание образов, нейронные сети, машинное обучение.

# Введение

Интеллектуальные системы обработки видеоинформации широко используются в настоящее время в различных областях жизни чело-

века [1, 2]. Развитие подобных систем, с одной стороны, связано с успехами в области вычислительной техники, а с другой – с развитием теории и методов обработки и анализа видеоизображений. Ключевыми задачами, которые лежат в основе большинства приложений систем обработки видеоинформации, являются обнаружение, распознавание и прослеживание объектов [3]. Решение этих задач лежит в основе таких приложений, как системы автоматического обнаружения и сопровождения объектов, робототехника, охранные системы, системы видеоаналитики [4].

В данной работе рассматривается проблема применения искусственных сверточных нейронных сетей в задаче обнаружения и локализации объектов заданных классов.

До появления специализированных нейросетей для обнаружения объектов заданных классов обычно использовался подход, согласно которому изображение проходилось скользящим окном, для каждого положения окна вычислялась карта признаков, например, с помощью гистограммы направленных градиентов или предварительно обученной нейронной сети, которая в последующем поступала на какой-либо классификатор [5]. В качестве классификатора, например, мог использоваться классификатор на основе машин опорных векторов (SVM – support vector machine) [6].

В настоящее время задачу обнаружения и классификации объектов принято решать с помощью искус-

Проводится сравнение различных нейросетевых архитектур для решения задачи обнаружения и распознавания объектов. Современные искусственные нейронные сети способны обнаруживать и локализовывать объекты заранее известных классов, что позволяет их применять в различных системах технического зрения. Проведено сравнение трех архитектур (YOLO, Faster R-CNN, SSD) по следующим критериям: скорость обработки изображения, тАР, точность и полнота.

> ственных сверточных нейронных сетей. Это обусловлено рядом причин:

 – существенный прогресс в области создания графических процессоров;

– большой объем данных для обучения;

 – лучшие результаты по сравнению с классическими подходами;

 – большое количество специализированных программных пакетов для подготовки данных, обучения и использования нейронных сетей.

Нейросетевые архитектуры для обнаружения и распознавания объектов можно разделить на две большие группы:

1. Архитектуры, обрабатывающие регионы на изображении (R-CNN).

2. Архитектуры, обрабатывающие поступившее изображение целиком (YOLO, SSD).

## Архитектура YOLO (You Only Look Once)

В работах [7, 8, 9] представлена архитектура нейросетевого детектора объектов под названием YOLO и ее модификации. Архитектура YOLO изначально разрабатывалась для задач реального времени. В алгоритме YOLO изображение разделяется на ячейки с использованием сетки. Для каждой ячейки сетки оценивается вероятность присутствия объекта вообще, затем строятся несколько наиболее вероятных положений объекта в виде прямоугольников с центром в данной ячейке, после чего для каждого полученного прямоугольника выполняется оценка вероятностей наличия в нем объектов каждого рассматриваемого класса.

В методе YOLO результаты обнаружения представляются в виде тензора размером 7×7×1024. Оценка вероятности нахождения объекта конкретного класса в текущем обрамляющем прямоугольнике – это произведение оценки вероятности нахождения объекта в ячейке и оценки вероятности для конкретного класса.

В случае YOLOv3 [9] для выделения признаков используется сверточная нейронная сеть, которая состоит из 53 слоев, в качестве фильтров используются свертки размером 3×3 и 1×1 и Residual блоки, которые добавляют к выходу текущего слоя значения с выхода предыдущего слоя.

Также стоит отметить, что в YOLOv3 обнаружение объектов выполняется на трех масштабах, что позволило увеличить качество обнаружения небольших объектов. Сеть масштабирует входное изображение пока не достигнет первого уровня обнаружения, на этом этапе шаг фильтров равен 32-м. На последующих сверточных слоях шаг фильтров равен 2. На каждом масштабе обнаружения ячейка предсказывает три обрамляющих прямоугольника, то есть с учетом масштаба каждой ячейке соответствует 9 обрамляющих прямоугольников.

На следующем шаге выполняется фильтрация прямоугольников по вероятности нахождения в них объектов. Потом так же, как и в архитектуре SSD прямоугольники фильтруются с помощью алгоритма подавления ложных максимумов.

Известно, что большинство алгоритмов распознавания предполагают, что выходные метки являются взаимоисключающими. В архитектурах YOLOv1 [7] и YOLOv2 [8] применяют функцию softmax [10] для преобразования оценок в вероятности классов, суммирование которых по всем классам дает единицу. YOLOv3 использует классификацию с несколькими метками. Например, выходные метки могут быть «Пешеход» и «Ребенок», которые не являются взаимоисключающими и сумма выходов может быть больше 1. В YOLOv3 функция активации softmax заменяется на независимые логистические классификаторы для вычисления вероятности выхода, принадлежащей определенной метке. Вместо использования среднеквадратической ошибки при вычислении потери классификации YOLOv3 использует бинарную кросс-энтропий-ную функцию потерь, вычисляемую для каждого класса. Использование этой техники также позволяет сократить объем требуемых вычислений.

# Архитектура Faster R-CNN (Faster Region-based Convolution Neural Network)

Для решения задачи обнаружения объектов Faster R-CNN [11] в настоящее время является одной из часто используемых архитектур на основе глубокого обучения. Предшественниками данной архитектуры являются R-CNN [12] и Fast R-CNN [13].

Работа R-CNN состоит из трех основных этапов:

1. Исходное изображение разбивается на регионы, в которых могут находиться объекты. С этой целью при-

меняется алгоритм Selective Search [14], генерирующий 2000 различных областей, которые с наибольшей вероятностью содержат объекты.

2. Каждый регион подается на вход соответствующей обученной сверточной нейронной сети, которая извлекает вектор признаков для своего региона.

3. Вектора признаков подаются на вход набора SVM, выполняющих функцию классификации. Каждая SVM обучена для определения одного класса объектов. Кроме того, для уточнения параметров охватывающего объект прямоугольника применяется линейная регрессия.

Дополнительным шагом можно считать подавление немаксимумов (алгоритм non-maximum suppression) для исключения избыточного числа прямоугольников, охватывающих один и тот же объект.

Архитектура R-CNN показала высокие показатели точности обнаружения объектов, но были отмечены такие недостатки, как высокие затраты памяти и времени на обучение и обработку изображений. Поэтому были предложены модификации архитектуры, приведшие к созданию Fast R-CNN:

1. Исходное изображение целиком подается на вход одной сверточной нейронной сети, выполняющей извлечение признаков, и на основании полноразмерной карты признаков осуществляется выбор регионовкандидатов.

2. Набор SVM, выполняющих функцию классификации, был заменен слоем softmax.

Таким образом, сверточная нейронная сеть используется один раз для всего изображения вместо обработки 2000 пересекающихся областей, также достаточно обучить одну сеть со слоем softmax без дополнительного обучения множества SVM.

С точки зрения скорости метод Fast R-CNN имеет значительное преимущество перед R-CNN, но еще одним недостатком являлся алгоритм выбора регионовкандидатов (Selective Search). Модификация данного этапа привела к созданию Faster R-CNN.

Алгоритм Selective Search был заменен на сеть выбора регионов-кандидатов (RPN – region proposal network). На вход данной сети подается область размера п×п, взятая из полноразмерной карты признаков, результат передается на два полносвязных слоя: boxregression и box-classification. Регионы-кандидаты, полученные с помощью RPN, представлены координатами описывающего прямоугольника и вероятностью нахождения объекта в данном регионе, вычисленной с применением функции softmax.

Архитектура Faster R-CNN в настоящее время позволяет добиться высокой точности обнаружения объектов и считается относительно быстрой. При этом сохранена главная идея исходной архитектуры R-CNN: выделение на изображении регионов, в которых возможно находятся объекты, и классификация содержимого этих регионов.

## Архитектура SSD (Single Shot MultiBox Detector)

Архитектура SSD [15] обеспечивает значительный прирост скорости обработки по сравнению с Faster R-CNN. Если последняя выполняет выбор регионов-



Рис. 1. Зависимость точности от полноты

кандидатов и классификацию регионов в два отдельных этапа, то SSD выполняет эти действия одновременно при обработке всего изображения. Работу SSD можно описать следующим образом:

1. Исходное изображение проходит через ряд сверточных слоев, что в результате дает набор карт признаков для разных масштабов (например, 19×19, 10×10, 5×5 и т.д.).

2. В каждой точке каждой карты признаков применяется сверточный фильтр размера 3×3 для получения множества описывающих прямоугольников.

3. Для каждого прямоугольника одновременно оцениваются пространственное смещение и вероятность нахождения объекта.

4. В процессе обучения истинные описывающие объект прямоугольники сопоставляются с предсказанными для исключения ложных обнаружений.

В отличие от R-CNN, где в регионах-кандидатах имеется хотя бы минимальная вероятность нахождения

объекта, в SSD шаг фильтрации регионов отсутствует. В результате генерируется гораздо большее количество описывающих прямоугольников на разных масштабах по сравнению с R-CNN, и большая часть не содержит объект. С целью решения данной проблемы в SSD, вопервых, используется подавление немаксимумов для объединения похожих друг на друга прямоугольников в один. Во-вторых, используется техника hard negative mining [16], согласно которой на каждой итерации обучения используется только часть отрицательных примеров, в SSD отношение числа отрицательных примеров к положительным равно 3 к 1.

Выбор регионов-кандидатов и классификация выполняются одновременно: при заданном числе классов С каждый описывающий прямоугольник связан с (4+С)-мерным вектором, который содержит 4 координаты и вероятности для всех классов. На последнем этапе применяется функция softmax для классификации объектов.

#### Экспериментальные исследования

С целью проведения сравнения было обучено пять нейросетевых детекторов:

1. YOLOv3.

2. Faster R-CNN на базе сети InceptionResnet-2, используемой для извлечения признаков.

3. Faster R-CNN на базе сети Resnet-101, используемой для извлечения признаков.

4. SSD на базе сети MobileNet-1, используемой для извлечения признаков.

5. SSD на базе сети MobileNet-2, используемой для извлечения признаков.

При обучении использовалось около 6700 изображений с размеченными объектами классов «пешеход» и «автомобиль». Оценить качество детектора объектов можно путем построения графика зависимости точности (precision) от полноты (recall), а также графиков зависимостей точности, полноты и *F*-меры от заданного порога. Точность, полнота и *F*-мера рассчитываются по формулам:

$$precision = \frac{n_{TP}}{n_{TP} + n_{FP}},$$

$$recall = \frac{n_{TP}}{n_{TP} + n_{FN}},$$

$$F = 2 \frac{precision \cdot recall}{precision + recall},$$

где  $n_{TP}$  (True Positive) – число верно обнаруженных объектов заданного класса;  $n_{FP}$  (False Positive) – число ложных срабатываний;  $n_{FN}$  (False Negative) – число необнаруженных объектов (пропусков).



Рис. 2. Зависимости точности, полноты и F-меры от порога



Рис. 3. Зависимости АР и тАР от номера итерации обучения

При проведении эксперимента обрабатывалось 750 изображений, содержащих объекты классов «пешеход» и «автомобиль». Для построения графиков изменяется пороговый коэффициент в алгоритме обнаружения объектов в диапазоне от 0 до 1 с шагом в 0,01. Под пороговым коэффициентом понимается минимальное значение оценки вероятности, при котором будет принято решение об обнаружении объекта. На рис. 1 представлены графики зависимости точности от полноты для разных детекторов, на рис. 2 – графики зависимостей точности, полноты и *F*-меры от заданного порога.

Для оценки качества работы детекторов в зависимости от номера итерации обучения для каждого класса объектов производился расчет метрики AP (Average Precision) – среднего значения максимальной точности при разных значениях полноты. Графики для разных детекторов представлены на рис. 3. Также на графиках отображена метрика mAP (mean Average Precision), представляющая собой среднее значение АР по всем классам объектов.

В качестве интегральных оценок качества работы детекторов использовались площадь под графиком зависимости точности от полноты (AUC – area under curve) и mAP. Также была проведена оценка вычислительной эффективности каждого детектора, при этом использовалась персональная ЭВМ с графическим процессором NVIDIA GeForce GTX 1070, замерялось среднее время обработки кадра с разрешением 720×468. Названные характеристики представлены в табл. 1.

Следует отметить, что время обработки сильно зависит от конкретной конфигурации оборудования, поэтому результаты несут информативный характер. Также время обработки с использованием графических процессоров персональных ЭВМ не всегда отражает время работы на мобильном устройстве. Например, MobileNet-2 на мобильных устройствах работает быстрее, чем MobileNet-1, но на персональной ЭВМ небольшое преимущество наоборот получила первая версия данной нейронной сети.

Потоктор	Характеристики			
детектор	AUC	mAP, %	Время, мс	
SSD c MobileNet-1	0,534	57,204	56	
SSD c MobileNet-2	0,573	61,249	58	
YOLOv3	0,882	75,740	76	
Faster R-CNN c Resnet-101	0,695	86,411	89	
Faster R-CNN c InceptionResnet-2	0,722	88,376	119	

Таблица 1 – Характеристики обученных детекторов

На рис. 4 приведены примеры обрабатываемых изображений с обнаруженными объектами интереса (автомобили и пешеходы), использовались наборы данных КІТТІ [17] и Cityscapes [18].



Рис. 4. Примеры изображений с обнаруженными объектами

#### Заключение

В данной работе рассмотрены три нейросетевые архитектуры: R-CNN (обработка регионов на изображении), YOLO и SSD (обработка поступившего изображения целиком). Экспериментальные исследования качества и вычислительной эффективности проводились с использованием таких нейросетевых детекторов, как YOLOv3, Faster R-CNN c InceptionResnet-2, Faster R-CNN c Resnet-101, SSD c MobileNet-1, SSD c MobileNet-2. При обучении и обработке на изображениях учитывались два класса объектов: «пешеход» и «автомобиль».

Сети Faster R-CNN показали преимущество в точности. Так, по результатам эксперимента самой высокой точностью обладает Faster R-CNN на базе сети InceptionResnet-2, но и среднее время обработки изображения для данного детектора самое большое. Архитектура SSD является наиболее подходящей для обработки изображений в режиме реального времени (особенно при использовании сетей MobileNet), но необходимо учитывать, что высокие требования к точности не всегда могут быть соблюдены. Нейросетевой детектор YOLOv3 обладает средними показателями точности и вычислительной эффективности по сравнению с другими исследованными детекторами.

Работа выполнена при финансовой поддержке стипендии Президента Российской Федерации молодым ученым и аспирантам (СП-2578.2018.5).

## Литература

1. Лукьяница А.А., Шишкин А.Г. Цифровая обработка видеоизображений. – М.: Ай-Эс-Эс Пресс, 2009. – 518 с.

2. Алпатов Б.А., Бабаян П.В. Технологии обработки и распознавания изображений в бортовых системах тех-

нического зрения // Вестник Рязанского государственного радиотехнического университета. – Рязань. – 2017. – №2. – С. 34-44.

3. 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.

4. Alpatov B.A., Babayan P.V., Ershov M.D. Vehicle Detection and Counting System for Real-Time Traffic Surveillance // Proceedings of 7th Mediterranean Conference on Embedded Computing (MECO). – IEEE, 2018. – pp. 120-123.

5. Gouk H.G.R., Blake A.M. Fast sliding window classification with convolutional neural networks // Proceedings of the 29th International Conference on Image and Vision Computing, New Zealand. – ACM, 2014. – pp. 114-118.

6. Boser B.E., Guyon I.M., Vapnik V.N. A training algorithm for optimal margin classifiers // Proceedings of the fifth annual workshop on Computational learning theory. – ACM, 1992. – pp. 144-152.

7. Redmon J., Divvala S., Girshick R., Farhadi A. You only look once: Unified, real-time object detection // Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. – pp. 779-788.

8. Redmon J., Farhadi A. YOLO9000: better, faster, stronger // arXiv preprint, arXiv:1612.08242. – 2016. – 9 p.

9. Redmon J., Farhadi A. YOLOv3: An incremental improvement // Tech report, arXiv:1804.02767. – 2018. – 6 p.

10. Bishop C.M. Pattern Recognition and Machine Learning. – Springer-Verlag, New York, 2006. – 738 p.

11. Ren S., He K., Girshick R., Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks // Extended tech report, arXiv:1506.01497. – 2016. – 14 p.

12. Girshick R.B., Donahue J., Darrell T., Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2014. – 21 p.

13. Girshick R. Fast R-CNN // IEEE International Conference on Computer Vision (ICCV). – 2015. – 9 p.

14. Uijlings J.R.R., van de Sande K.E.A., Gevers T., Smeulders A.W.M. Selective Search for Object Recognition // International Journal of Computer Vision. – 2013. – Vol. 104. – pp. 154-171.

15. Liu W., Anguelov D., Erhan D., Szegedy C., Reed S., Fu Ch.-Y., Berg A.C. SSD: Single Shot MultiBox Detector // European Conference on Computer Vision (ECCV), Springer, Cham. – 2016. – Vol. 9905. – pp. 21-37.

16. Wan S., Chen Z., Zhang T., Zhang B., Wong K. Bootstrapping Face Detection with Hard Negative Examples // arXiv:1608.02236. – 2016. – 7 p.

17. Geiger A., Lenz P., Urtasun R. Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite // Conference on Computer Vision and Pattern Recognition (CVPR). -2012. -8 p.

18. Cordts M., Omran M., Ramos S., Rehfeld T., Enzweiler M., Benenson R., Franke U., Roth S., Schiele B. The Cityscapes Dataset for Semantic Urban Scene Understanding // Conference on Computer Vision and Pattern Recognition (CVPR). – 2016. – 11 p.

## УДК 621.391

# РАЗРАБОТКА И АНАЛИЗ АЛГОРИТМА ДЕТЕКТИРОВАНИЯ ПАТОЛОГИЙ НА ЭНДОСКОПИЧЕСКИХ ИЗОБРАЖЕНИЯХ ЖЕЛУДКА НА ОСНОВЕ СВЕРТОЧНОЙ НЕЙРОННОЙ СЕТИ

**Хрящев В.В.,** к.т.н., доцент Ярославского государственного университета им. П.Г. Демидова, e-mail: vhr@yandex.ru;

Ганин А.Н., к.т.н., генеральный директор ООО «Точка зрения», г. Ярославль, e-mail: angnn@mail.ru; Лебедев А.А., аспирант Ярославского государственного университета им. П.Г. Демидова, e-mail: lebedevdes@gmail.com;

Степанова О.А., аспирант Ярославского государственного университет им. П.Г. Демидова, e-mail: olga1stepanova@yandex.ru;

Кашин С.В., к.м.н., ГБУЗ Ярославской области «Клиническая онкологическая больница», e-mail: s\_kashin@mail.ru;

*Куваев Р.О., к.м.н., ГБУЗ Ярославской области «Клиническая онкологическая больница», e-mail: kuvaev\_roman@mail.ru.* 

# DEVELOPMENT AND ANALYSIS OF ALGORITHM OF PATHOLOGY DETECTION IN ENDOSCOPIC IMAGES OF GASTRIC BASED ON CONVOLUTIONAL NEURAL NETWORK

## Khryashchev V.V., Ganin A.N., Lebedev A.A., Stepanova O.A., Kashin S.V., Kuvaev R.O.

This paper presents the algorithm of pathology detection in endoscopic images of gastric based on convolutional neural network. Training and testing of the algorithm was carried out on the NVIDIA DGX-1 supercomputer using endoscopic images from the test base, assembled together with the YaroslavI Clinical Oncological Hospital. As a result of experiments, the mAP metric was calculated and the value was 0.875, which is a high result for the task of object detection in images.

Key words: machine learning, convolution neural network, endoscopic image analyses, gastric cancer.

Ключевые слова: машинное обучение, сверточная нейронная сеть, анализ эндоскопических изображений, рак желудка.

#### Введение

С развитием техники и технологий все более реальным становится использование достижений цифровой обработки изображений в области медицины для создания систем автоматической диагностики, а также систем под-

держки принятия решений [1]. Внедрение подобных систем в клиническую медицину нацелено на повышение эффективности диагностики и терапии, сокращение времени и расходов на исследования, проведение контроля качества, а также обучение и совершенствования врачебных навыков специалистов. Одним из актуальных направлений исследований является анализ эндоскопических изображений [2-4].

На сегодняшний день существует ряд российских и зарубежных исследований, посвященных системам автоматического анализа эндоскопических изображений. В работе Батухтина Д.М. и др. [5] представлена система контроля качества эндоскопических исследований и автоматизированного анализа эндоскопических изображений пищевода, основанная на использовании оптимального по критерию максимального правдоподобия классификатора. Тот же коллектив авторов в работе [6] описал систему сегментации изображений слизистой оболочки пищевода, обеспечившую повышение диагно-

Разработан алгоритм детектирования патологий на эндоскопических изображениях желудка на основе глубокой сверточной нейронной сети. Обучение и тестирование алгоритма осуществлялось на суперкомпьютере NVIDIA DGX-1 с использованием эндоскопических изображений из тестовой базы, собранной в Ярославской клинической онкологической больнице. В результате численных экспериментов получено значение метрики тАР, составившее 0,875, что является высоким результатом для задачи детектирования объектов на изображениях.

стической эффективности узкоспектральной эндоскопии по результатам 238 эндоскопических исследований.

Группа исследователей в работе [7] описала прототип системы автоматической сегментации и аннотирования эндоскопических изображений, основанный на классификации векторов топологических признаков исходного изображения.

Коваленко Д.А. и Гнатюк В.С. в работе [8] предложили метод ассоциации сцен для эндоскопических видео, позволяющий выделить ключевые сцены с помощью построения системы соседства похожих кадров и глобального анализа похожих кадров путем применения нейросетевого статистического критерия. Средняя точность предложенного алгоритма составила 57 %.

В работах зарубежных авторов [9, 10] предложены алгоритмы на основе сверточных нейронных сетей для исследования колоректальных эндоскопических изображений, обеспечивающие точность классификации 85-90 %. Авторы [11] рассматривают свёрточные нейронные сети в сочетании с набором низкоуровневых функций для обнаружения патологии в рентгенограмме грудной клетки. В результате получены площади под ROC-кривой (AUC) 0,93 и 0,79 для разных подзадач.

Приведенный анализ литературы показывает, что использование систем компьютерного зрения для анализа эндоскопических изображений на сегодняшний день показывает высокие результаты согласно многочисленным исследованиям; следовательно, разработка подобных систем представляет собой актуальную научно-техническую задачу.

В настоящее время одним из самых распространенных онкологических заболеваний является рак желудка, занимающий второе место по смертности от злокачественных новообразований. Вторичная профилактика рака посредством ранней диагностики и адекватной терапии является наиболее эффективной стратегией повышения выживаемости пациентов [12–14].

Целью данной работы является разработка системы автоматического детектирования патологических изменений слизистой оболочки желудка на эндоскопических изображениях. Использование такой системы в реальной клинической практике позволит существенно повысить качество диагностики различных патологий желудка.

# Создание базы эндоскопических изображений желудка

Задача анализа эндоскопических изображений оказывается весьма сложной в основном из-за разнообразия текстур исследуемой поверхности, ее существенной неоднородности и широкого диапазона масштабов обрабатываемых изображений. Кроме того, эндоскопические изображения характеризуются следующими отличительными особенностями:

 яркость и контрастность изображений сильно зависит от геометрии рассматриваемого участка ткани;

 на изображениях может присутствовать размытие, обусловленное перемещением датчика внутри полости органа;

 изображения тканей могут значительно меняться во время сокращения мышечных волокон органа (перистальтики);

 практически соосное расположение объектива и источника освещения приводит к появлению бликов и артефактов на изображениях [8].

Для повышения уровня диагностики патологических участков слизистой оболочки используются такие высокотехнологичные методы эндоскопической диагностики, как увеличительная эндоскопия и эндоскопия в узком спектре света [13].

Увеличительная эндоскопия позволяет изучить поверхность слизистой оболочки с оптическим увеличением примерно в 115 раз, приближая эту методику к стереоскопической микроскопии. С ее помощью возможно выявление минимальных изменений типичной архитектоники слизистой оболочки в различных отделах желудка [13].

Принцип действия узкоспектральной эндоскопии основан на использовании оптических фильтров, сужающих спектр световой волны: из диапазона длин волн от 400 нм до 800 нм, использующегося при эндоскопии в белом свете, выбираются всего две световых волны длиной 415 нм и 445 нм. Данные волны хорошо поглощаются гемоглобином, что позволяет их использовать в диагностике сосудистых структур слизистого и подслизистого слоев пищеварительного тракта [12, 14].

Таким образом, эффективное использование этих методик требует от специалиста высокого уровня теоретических знаний в области эндоскопии и смежных специальностей, а также практических навыков и опыта клинического применения. По этой причине увеличительная и узкоспектральная эндоскопии остаются методиками экспертного уровня, а обучение специалистов традиционными способами представляет собой длительный процесс, что значительно затрудняет их активное внедрение в широкую клиническую практику [3, 4].

Важным шагом в разработке системы компьютерного анализа медицинских изображений является сбор тестовой базы изображений. В данной работе для анализа использовались эндоскопические изображения с увеличением в узкоспектральном режиме, полученные со следующих эндоскопических систем: эндоскопическая система OLYMPUS EXERA II, видеогастроскоп GIF 160Z; эндоскопическая система OLYMPUS LUCERA SPECTRUM, видеогастроскоп GIF 260Z; эндоскопическая система OLYMPUS EXERA III, видеогастроскоп GIF HQ290; эндоскопическая система PENTAX Medical EPKi7010, видеогастроскоп EG-2990Zi.

База эндоскопических изображений создавалась в Ярославской клинической онкологической больнице. Размер базы составил 437 изображений, из которых 119 изображений содержат раковые патологии и 312 их не содержат. Изображения масштабировались до разрешения 300×300 пикселей таким образом, чтобы фрагмент не содержал технических надписей и регионов, не относящихся к слизистой желудка, после чего база размножалась за счет применения к каждому изображению геометрических преобразований. Далее база данных размножалась путем поворота на 90 и на 180 градусов. Итоговый размер базы данных составил 1293 изображений, из которых 357 относятся к раковым патологиям.

Патологии слизистой оболочки желудка характеризуются изменениями микрососудистого рисунка и рисунка микроструктуры поверхности эпителия. Данные изменения наиболее заметны на границе патологии, которая называется демаркационной линией. В созданной базе данных все участки, содержащие патологии, выделены вручную. Как видно из рис. 1, полученные полигоны представляют собой сложные многоугольники. Для упрощения дальнейшей оценки качества работы алгоритма полигоны, ограниченные демаркационной линией, были описаны прямоугольниками.

# Разработка алгоритма анализа эндоскопических изображений желудка

Стратегия эндоскопической диагностики с применением узкоспектральной и увеличительной эндоскопии зависит от обследуемого органа желудочно-кишечного тракта, определяя так называемый орган-специфичный



l 3/01 EAS D RASS

в) иная патология

г) иная патология



подход в диагностике. Наиболее сложную микроархитектонику имеет слизистая оболочка желудка, поскольку при эндоскопическом осмотре требуется в равной степени оценка как микрососудистого рисунка, так и рисунка микроструктуры поверхности эпителия. Более того, слизистая оболочка желудка имеет разные типы с различным строением, а большой спектр ее патологических изменений (хроническое воспаление, атрофия, кишечная метаплазия, различные гистологические типы рака желудка) определяет разнообразные изменения в их структуре.

В современной классификации патологий, выявляемых при эндоскопической диагностике, выделяют до 8 различных типов, которые можно объединить в 3 группы: раковая патология, патологии, не относящиеся к раковым, а также слизистая без патологических изменений.

В данной работе рассматривается система анализа эндоскопических изображений, производящая детектирование объектов следующих классов:

I класс: слизистая без патологических изменений. Объекты данного класса не выделяются демаркационной линией.

**II класс:** не рак. Данный класс включает в себя участки слизистой с патологиями, не относящимися к раковым, а также фрагменты изображения, которые нельзя отнести к I или III классам (например, просвет полости органа на изображении). Объекты данного класса на изображении выделяются демаркационными прямоугольниками.

**III класс:** рак. Объекты данного класса выделяются демаркационными прямоугольниками.

Для решения задачи анализа эндоскопических изображений желудка выбран аппарат сверточных нейронных сетей, ранее показавших свое превосходство в решении задач распознавания лиц [15], эмоций по изображению лица [16], объектов на спутниковых изображениях [17] и др. Применение данного подхода обусловлено преимуществами сверточных нейронных сетей перед традиционными алгоритмами машинного обучения: формирование признаков в таких сетях происходит автоматически в процессе обучения модели с помощью операции двумерной свертки, что позволяет получить лучшие результаты в задачах классификации, детектирования и распознавания объектов на изображениях. Основой разрабатываемого алгоритма детектирования в рамках данного исследования выбрана сверточная нейронная сеть SSD (Single Shot Multibox Detector) [18], архитектура которой представлена на рис. 2.

Нейронная сеть SSD представляет собой сверточную сеть прямого распространения. На вход сети подается цветное изображение размером 300х300х3 (RGBизображения), на выходе нейронная сеть выдает нормированные координаты детектируемых объектов раз-
личных классов (их число является одним из входных параметров алгоритма).



Ряд начальных слоев сети представляют собой укороченную версию архитектуры VGG-16 [19]. После нее расположены сверточные слои, формирующие своеобразную пирамиду из карт признаков различного масштаба. Эти карты признаков подаются на слои классификации и детектирования, где осуществляется прогнозирование, причем каждая из карт признаков позволяет обнаруживать объекты определенного размера. Карты признаков разбивается на «боксы» прямоугольной формы различного масштаба, далее, на этапе прогнозирования, оценивается вероятность нахождения объекта того или иного класса в данном боксе, после чего происходит корректировка положения бокса. Кроме того, сеть объединяет прогнозы из нескольких карт признаков с различными разрешениями для детектирования и классификации объектов различных размеров. Затем данные о детектированных объектах различных классов обрабатываются быстрым алгоритмом NMS (Non-Maximum Supression, алгоритм подавления немаксимумов) [20] для коррекции границ и фильтрации.

Реализация архитектуры сети осуществлялась с помощью фреймворка Caffe. Для инициализации части весов сверточной нейронной сети SSD использовалась модель укороченной нейронной сети VGG-16 для классификации объектов [19], обученная на базе изображений Imagenet [21]. Веса остальных слоев задавались с помощью инициализации Ксавьера. Смещения инициализировались константами – нулями. Общее количество итераций составило порядка 120 000. В качестве алгоритма численной оптимизации использовался стохастический градиентный спуск с использованием добавочного моментного коэффициента, равного 0,9. Для регуляризации модели в процессе обучения применялось правило обновления весов, в результате чего добавлялся коэффициент распада веса, равный 0,0005. Начальная скорость обучения равнялась 0,001 и затем уменьшалась в 10 раз после 80 000 итераций.

Обучение нейронной сети осуществлялось параллельно на 4 видеокартах суперкомпьютера NVIDIA DGX-1 Volta, предоставленного центром искусственного интеллекта и цифровой экономики ЯрГУ, размер батча составлял 32. Входной размер сети соответствовал разрешению изображений из базы данных: 300х300х3. Из каждого пикселя входного изображения поканально вычитались значения: mean\_value: 104,0; mean\_value: 117,0; mean\_value: 123,0. В процессе обучения изображения зеркалировались.

#### Тестирование алгоритма

Для проведения численного эксперимента из созданной базы эндоскопических изображений случайным образом отбирались 100 изображений, из которых 24 относились к классу «рак».

Для оценки качества работы алгоритма рассчитывались значения метрики AP (Average Precision), а также mAP (mean Average Precision). Значение метрики AP вычисляется как усреднение значений точности при разных значениях порога:

$$AP = \frac{\sum_{r=1}^{N} P_r}{N},\tag{1}$$

где  $P_r$  – значение точности (Presicion) при фиксированном значении порога r, а N – множество значений порога. Метрика mAP используется как основная мера качества работы детекторов объектов и представляет собой усреднение значений AP по всем классам:

$$mAP = \frac{\sum_{i=1}^{K} AP_i}{K},$$
(2)

где К – число классов.

Графики зависимости точности от полноты при варьировании порога изображены на рис. 3.



Рис. 3. Графики зависимости точности P от полноты R при варьировании порога



в) верное обнаружение иной патологии

г) верное обнаружение иной патологии

Рис. 4. Примеры работы алгоритма.

Эталонная демаркационная линия изображена пунктиром, результат работы алгоритма – сплошной линией

Таким образом, АР для класса рак составила 0,827, для класса не рак – 0,923; *mAP* для разработанного алгоритма составляет 0,875.

Примеры детектирования патологий предложенным алгоритмом приведены на рис. 4.

#### Заключение

Предложен и протестирован алгоритм анализа эндоскопических изображений для классификации патологий слизистой оболочки желудка. В его основе лежит реализация сверточной нейронной сети. Обучение и тестирование алгоритма проводилось параллельно на четырех видеокартах суперкомпьютера NVIDIA DGX-1.

В результате численных экспериментов получено значение метрики mAP, составившее 0,875, что является достаточно высоким результатом для задачи детектирования объектов на эндоскопических изображениях желудка. Основной задачей дальнейших исследований является сбор более репрезентативной базы эндоскопических изображений, необходимой для повышения обобщающей способности алгоритма на основе сверточных нейронных сетей. Собирать такую базу планируется в коллаборации с современными медицинскими центрами из Великобритании, Японии, Австрии и Тайваня.

#### Литература

1. Гудфеллоу Я., Бенджио И., Курвилль А. Глубокое обучение / 2017. ДМК-Пресс, 652 с.

2. Bisschops R, Areia M, Coron E, Dobru D., Kaskas B., Kuvaev R., Pech O., Ragunath K., Weusten B., Familiari P., Domagk D., Valori R., Kaminski M.F., Spada C., Bretthauer M., Bennett C., SenoreC., Dinis-Ribeiro M., Rutter M.D. R. Performance measures for upper gastrointestinal endoscopy: a European Society of Gastrointestinal Endoscopy (ESGE) Quality Improvement Initiative // Endoscopy. 48(9). 2016. pp. 843-864.

3. Куваев Р.О., Никонов Е.Л., Кашин С.В., Капранов В.А., Гвоздев А.А. Контроль качества эндоскопических исследований, перспективы автоматизированного анализа эндоскопических изображений // Кремлевская медицина. Клинический вестник. 2013. 2. С. 51–56.

4. Лебедев А.А., Степанова О.А., Юрченко Е.А., Хрящев В.В. Разработка алгоритмов анализа изображений для классификации патологий слизистой оболочки желудка // Цифровая обработка сигналов и ее применение (DSPA-2018): докл. 20-й междунар. конф. М., 2018. Т. 2. С. 644–649.

5. Батухтин Д.М., Пеганова Е.В., Митракова Н.Н., Роженцов А.А., Фурман Я.А. Анализ узкоспектральных эндоскопических изображений на внутренней поверхности пищевода // Вестник Поволжского государственного технологического университета. Серия: радиотехнические и инфокоммуникационные системы. 2014. № 4 (23). С. 45–57.

6. Пеганова Е.В., Батухтин Д.М., Митракова Н.Н. Автоматизированная система сегментации узкоспектральных изображений для оптимизации эндоскопической диагностики при патологии пищевода // ЭиКГ. 2014. № 3 (103).

7. Дунаева О.А., Малкова Д.Б., Мячин М.Л., Эдельсбруннер Х. Сегментация клинических эндоскопических изображений, основанная на классификации векторных топологических признаков // Модел. и анализ информ. систем. 2013. 20:6. С. 162–173.

8. Коваленко Д.А., Гнатюк В.С. Ассоциация сцен в эндоскопических видео // GraphiCon 2017: Обработка и анализ биомедицинских изображений. Пермь. 2017. С. 269–274.

9. Tamaki T., Sonoyama S., Hirakawa T., Raytchev B., Kaneda K., Koide T. Computer-Aided Colorectal Tumor Classification in NBI Endoscopy Using CNN Features // in The Korea-Japan joint workshop on Frontiers of Computer Vision (FCV2016). 2016. pp. 61-65.

10. Ribeiro E., Uhl A., Wimmer G., Häfner M. Exploring Deep Learning and Transfer Learning for Colonic Polyp Classification // Computational and Mathematical Methods in Medicine. vol. 2016.

11. Bar Y., Wolf L., Diamant I., Greenspan H. Deep Learning with Non-Medical Training Used for Chest Pathology Identification // In: SPIE Medical Imaging. 2015.

12. Куваев Р.О., Кашин С.В., Никонов Е.Л., Itoh Т., Gotoda T., Gono K. Ранний рак желудка: методики скрининга, эндоскопической диагностики и малоинвазивного лечения // Доказательная гастроэнтерология. 2014. 3 (3). С. 44–51.

13. Куваев Р.О., Кашин С.В. Современное эндоскопическое исследование желудка с использованием методик узкоспектральной и увеличительной эндоскопии: техника проведения и алгоритмы диагностики // Доказательная гастроэнтерология. 2016. 2 (5). С. 3–13.

14. Куваев Р.О., Никонов Е.Л., Кашин С.В. Helicobacter pylori-ассоциированный хронический гастрит: новые технологии эндоскопической диагностики // Доказательная гастроэнтерология. 2015. 4 (1). С. 19–24.

15. Хрящев В.В., Приоров А.Л., Стефаниди А.Ф., Топников А.И. Разработка и исследование алгоритмов обработки и распознавания речевых сигналов и изображений для систем мультимодальной биометрии // Цифровая обработка сигналов. 2017. № 3. С. 45–49.

16. Khryashchev V., Ivanovsky L., Priorov A. Deep Learning for Real-Time Robust Facial Expression Analysis // Proceedings of the International Conference on Machine Vision and Applications (ICMVA 2018). Singapore, 23-25 April. 2018. pp. 66–70.

17. Khryashchev V., Pavlov V., Priorov A., Kazina E. Convolutional Neural Network for Satellite Imagery // Proceedings of the 22th Conference of Open Innovations Association FRUCT'22. Jyvaskyla, Finland, 15-18 May 2018. pp. 344–347.

18. Liu W., Anguelov D., Erhan D., Szegedy C., and Reed S.E. SSD: Single Shot Multibox Detector. CoRR, abs/1512.02325, 2015.

19. Fully convolutional reduced VGGNet [Электронный pecypc]. Режим доступа: https://gist.github.com/weiliu89/ 2ed6e13bfd5b57cf81d6.

20. Canny J. A Computational Approach to Edge Detection // IEEE Transactions On Pattern Analysis And Machine Intelligence. 1986. vol. Pami-8, no. 6. pp. 679–698.

21. ImageNet Image Database [Электронный ресурс]. Режим доступа: http://www.imagenet.org/.

### Солонина А.И.



www.bhv.ru

Отдел оптовых поставок E-mail: opt@bhv.spb.su



Цифровая обработка сигналов в зеркале MATLAB

**РЕЦЕНЗЕНТЫ:** Е.Б. Соловьева, д-р техн. наук, завкафедрой теоретических основ электротехники Санкт-Петербургского государственного электротехнического университета «ЛЭТИ»; А.А. Монаков, д-р техн. наук, проф. кафедры радиотехнических систем Санкт-Петербургского государственного университета аэрокосмического приборостроения.

Учебное пособие представляет собой базовый курс по цифровой обработке сигналов с традиционными темами: линейные дискретные системы; эффекты квантования в цифровых системах с фиксированной точкой; дискретные сигналы с описанием в частотной области, включая алгоритмы дискретного и быстрого преобразований Фурье; случайные процессы и их статистические характеристики; методы непараметрического и параметрического спектрального анализа; цифровые фильтры общего и специального назначения; многоскоростные системы и их полифазные структуры; банки фильтров; основы адаптивной обработки сигналов.

Теория ЦОС преподносится с позиции моделирования в MATLAB: теоретические аспекты базовых методов и алгоритмов ЦОС увязываются с их моделированием в MATLAB и иллюстрируются соответствующими примерами.

Книга предназначена для студентов и преподавателей вузов, а также специалистов в области цифровой обработки сигналов.

#### УДК 004.932

## ВАРИАЦИОННЫЙ МЕТОД ВЫЧИСЛЕНИЯ ОПТИЧЕСКОГО ПОТОКА В СИСТЕМЕ-НА-КРИСТАЛЛЕ

Беляков П.В., аспирант кафедры ЭВМ Рязанского государственного радиотехнического университета, belyakov.p.v@evm.rsreu.ru;

Никифоров М.Б., к.т.н., доцент кафедры ЭВМ Рязанского государственного радиотехнического университета, nikiforov.m.b@evm.rsreu.ru.

## SYSTEM-ON-CHIP VARIATIONAL OPTICAL FLOW COMPUTATION

#### Belyakov P.V., Nikiforov M.B.

Variational methods of the optical flow computation are widely known and common used methods for motion detection, object tracking in various areas of image processing, 3D reconstruction and autonomous robot navigation. The variational nonlinear method of the optical flow computation is the most accurate, but at the same time the most computationally intensive. Its implementation on a system-on-chip (SoC) is a trade-off between the design difficulty and high performance hardware implementation. The article is devoted to the specific SoC-based solution of the variational optical flow computation method which has been implemented using Verilog hardware description language. The proposed solution is applicable for a dense variational nonlinear optical flow real time computation and can act as a hardware SoC accelerator for the optical flow computation in various image processing tasks.

Key words: optical flow, variational methods, FPGA, system-on-chip.

Ключевые слова: оптический поток, вариационный метод, ПЛИС, система-на-кристалле.

#### Введение

Оптический поток – это изображение видимого движения, представляющее собой сдвиг каждой точки между двумя изображениями [1]. Это векторное поле, где каждый вектор имеет две компоненты, чтобы показать смещение между точками при их перемещении от первого изображения ко второму [2]. По сути, оптический поток представляет собой поле скоростей (т.к. сдвиг эквивалентен мгновенной скорости) и для каждой точки первого изображения  $I_1(x,y)$ находится такой сдвиг (dx, dy), чтобы исходной точке соответствовала точка на втором изображении  $I_2(x + dx, y + dy)$  [3].

Метод, который вычисляет оптический поток для всех пикселей в изображении, называется плотным, тогда как методы, обрабатывающие некоторые отдельные пиксели, называются разреженные [4].

Предлагаемая работа основана на вариационном методе [5], который относится к плотным методам и основывается на нескольких основных предположениях для вычисления оптического потока: предположении о постоянстве значения яркости изображения, предположении о постоянстве значения градиента яркости и предположении гладкости искомого вектора оптического потока. Минимизация функционала, содержащего указанные предположения, основана на решении уравнений в частных производных численными методами.

Использование минимизации функционала с регуляризацией является одним из эффективных методов вычисления оптического потока. Этот подход основан на минимизации заданного функционала путем решения системы линейных уравнений. Для решения такой систе-

Вариационные методы вычисления оптического потока - широко известные и общепринятые в различных областях обработки изображений для обнаружения движения, сопровождения объектов, 3D-реконструкции и автономной навигации роботов. Нелинейная модель вычисления оптического потока вариационным методом является наиболее точной, но вместе с тем наиболее вычислительно сложной. Ее реализация на системе-на-кристалле (SoC -System on Chip) является компромиссом между трудностью проектирования и высокой производительностью аппаратной реализации. Статья посвящена реализации на SoC вариационного метода вычисления оптического потока с использованием языка аппаратного дизайна Verilog. Предлагаемое решение способно производить вычисление плотного нелинейного оптического потока в реальном масштабе времени и может выступать в качестве аппаратного ускорителя на SoC вычисления оптического потока в различных задачах обработки изображений.

> мы уравнений можно использовать один из итерационных численных методов решения системы линейных алгебраических уравнений, например, метод ослабления Гаусса-Зейделя (метод релаксации SOR – Successive-Over Relaxation), обладающий повышенной сходимостью.

> Дополнительным инструментом эффективного вычисления оптического потока является применение подхода масштабирования изображений, когда поиск решения начинается на самом грубом уровне пирамиды сглаженных масштабированных изображений и проходит через всю последовательность пирамиды изображений разного масштаба для поиска глобального минимума функционала.

> В течение долгого времени главным инструментом для вычисления оптического потока в реальном масштабе времени из-за его сложности оставались графические процессоры. Относительно недавно и SoC стали подходить для выполнения аналогичных задач. Хорошо известно, что SoC поддерживают разный уровень па

раллелизма, что, в свою очередь, приводит к повышению производительности с точки зрения вычислительной эффективности [6]. Целью этой работы является разработка соответствующей аппаратной архитектуры для вычисления оптического потока. Для наиболее эффективной аппаратной реализации использовался язык аппаратного дизайна Verilog. Таким образом, в этой статье представлена высокопроизводительная оптимизированная аппаратная архитектура SoC для вычисления оптического потока в реальном времени.

Работа показывает возможность реализации алгоритма вариационного плотного оптического потока на ПЛИС на языке описания аппаратуры Verilog для создания параллельной высокопроизводительной конвейерной архитектуры, обеспечивающей максимальную производительность [7]. Аппаратное ядро оптического потока реализовано с помощью Xilinx Vivado IDE. Система разработки и отладки вариационного оптического потока, состоящая из программного приложения и аппаратной части, реализована на платформе Xilinx Zynq-7000 SoC отладочной платы ZC706.

# Вариационный метод вычисления оптического потока

Одной из наиболее точной, но вместе с тем наиболее вычислительно сложно, является постановка задачи вычисления оптического потока в вариационной формулировке и ее решение через минимизацию функционала, описываемого выражением

$$E(u,v) = \int (\psi((I(\mathbf{x}+\mathbf{w}) - I(\mathbf{x}))^2 + \gamma |\nabla I(\mathbf{x}+\mathbf{w}) - \nabla I(\mathbf{x})|^2 + \alpha \psi(|\nabla u|^2 + |\nabla u|^2)) d\mathbf{x}.$$
(1)

Здесь I – изображение,  $\mathbf{x} = (x, y, t)^T$  – координаты пикселя на изображении,  $\mathbf{w} = (u, v, 1)^T$  – искомый вектор смещения между пикселями двух изображений в момент времени t и момент времени t+1.  $\nabla := (\partial_x, \partial_y)^T$  пространственный градиент,  $\gamma$  – коэффициент между яркостью и градиентом яркости,  $\alpha > 0$  параметр регуляризации, описывает то, насколько важно требование гладкости полученного вектора смещений. Функция регуляризации  $\psi(s^2) = \sqrt{s^2 + \varepsilon^2}$  определяет устойчивость минимизации функционала к изменениям яркости изображения и гладкости поля с параметром  $\varepsilon = 0,001$ .

#### Уравнения Эйлера-Лагранжа

Минимизация функционала (1) сводится к решению уравнений Эйлера-Лагранжа:

$$\psi'(I_{z}^{2} + \gamma(I_{xz}^{2} + I_{yz}^{2})) \cdot (I_{x}I_{z} + \gamma(I_{xx}I_{xz} + I_{xy}I_{yz})) - -\alpha \, div(\psi'(|\nabla u|^{2} + |\nabla v|^{2})\nabla u) = 0, \psi'(I_{z}^{2} + \gamma(I_{xz}^{2} + I_{yz}^{2})) \times (I_{y}I_{z} + \gamma(I_{yy}I_{yz} + I_{xy}I_{xz})) - -\alpha \, div(\psi'(|\nabla u|^{2} + |\nabla v|^{2})\nabla v) = 0.$$
(2)

Здесь использование индекса *z* вместо *t* означает аппроксимацию операции дифференцирования по времени разностью.

$$I_{x} := \partial_{x} I(\mathbf{x} + \mathbf{w}),$$
$$I_{y} := \partial_{y} I(\mathbf{x} + \mathbf{w}),$$

$$I_{z} := I(\mathbf{x} + \mathbf{w}) - I(\mathbf{x}),$$

$$I_{xx} := \partial_{xx}I(\mathbf{x} + \mathbf{w}),$$

$$I_{xy} := \partial_{yy}I(\mathbf{x} + \mathbf{w}),$$

$$I_{yy} := \partial_{yy}I(\mathbf{x} + \mathbf{w}),$$

$$I_{xz} := \partial_{x}I(\mathbf{x} + \mathbf{w}) - \partial_{x}I(\mathbf{x}),$$

$$I_{yz} := \partial_{y}I(\mathbf{x} + \mathbf{w}) - \partial_{y}I(\mathbf{x}),$$

$$|\nabla u|^{2} = u_{x}^{2} + u_{y}^{2}, \quad |\nabla v|^{2} = v_{x}^{2} + v_{y}^{2},$$

$$u_{x} = \frac{\partial u}{\partial x}, \quad u_{y} = \frac{\partial u}{\partial y}, \quad v_{x} = \frac{\partial v}{\partial x}, \quad v_{y} = \frac{\partial v}{\partial y}.$$

#### Минимизация функционала

Для устранения нелинейности в выражении (2) для w = (u, v, 1) первоначально применяется метод последовательных приближений [8]. Пусть k означает итерацию последовательно приближения, тогда  $I_*^k$  обозначает первые и вторые производные в выражении (2) применительно к неизвестному вектору  $w^k$  на k -ой итерации вместо w. Тогда  $w^{k+1}$  будет очередным приближения решения уравнения

$$\psi'((I_z^{k+1})^2 + \gamma((I_{xz}^{k+1})^2 + (I_{yz}^{k+1})^2)) \times \times (I_x^k I_z^{k+1} + \gamma(I_{xx}^k I_{xz}^{k+1} + I_{xy}^k I_{yz}^{k+1})) - (3)$$
  
$$-\alpha \operatorname{div}(\psi'(|\nabla u^{k+1}|^2 + |\nabla v^{k+1}|^2) \nabla u^{k+1}) = 0,$$

Второе выражение в уравнении Эйлера-Лагранжа может быть записано аналогичным образом.

Далее, для устранения нелинейности в выражениях  $I_*^{k+1}$ , применяется разложение в ряд Тейлора:

$$I_{z}^{k+1} \approx I_{z}^{k} + I_{x}^{k} du^{k} + I_{y}^{k} dv^{k},$$

$$I_{xz}^{k+1} \approx I_{xz}^{k} + I_{xx}^{k} du^{k} + I_{xy}^{k} dv^{k},$$

$$I_{yz}^{k+1} \approx I_{xz}^{k} + I_{xy}^{k} du^{k} + I_{yy}^{k} dv^{k},$$

где  $u^{k+1} = u^k + du^k$  и  $v^{k+1} = v^k + dv^k$ . Искомые  $u^{(k+1)}, v^{(k+1)}$  разделяются на решение с предыдущей итерации  $u^k, v^k$  и неизвестное приращение вектора  $du^k, dv^k$ .

Обозначим

$$\begin{split} (\psi_{D}^{'})^{k} &= \psi^{'} ((I_{z}^{k} + I_{x}^{k} du^{k} + I_{y}^{k} dv^{k})^{2} + \\ &+ \gamma ((I_{xz}^{k} + I_{xx}^{k} du^{k} + I_{xy}^{k} dv^{k})^{2} + \\ &+ (I_{yz}^{k} + I_{xy}^{k} du^{k} + I_{yy}^{k} dv^{k})^{2})), \\ (\psi_{s}^{'})^{k} &= \psi^{'} (|\nabla (u^{k} + du^{k})|^{2} + |\nabla (v^{k} + dv^{k})|^{2}) \end{split}$$

где  $\psi_D$  – устойчивость к изменению яркости, and  $\psi_s$  – требование гладкости вектора, тогда выражение (3) может быть записано как

$$\begin{aligned} (\psi_{D}^{'})^{k} \cdot (I_{x}^{k}(I_{z}^{k}+I_{x}^{k}du^{k}+I_{y}^{k}dv^{k})) + \\ +\gamma(\psi_{D}^{'})^{k}(I_{xx}^{k}(I_{xz}^{k}+I_{xx}^{k}du^{k}+I_{xy}^{k}dv^{k}) + \\ +I_{xy}^{k}(I_{yz}^{k}+I_{xy}^{k}du^{k}+I_{yy}^{k}dv^{k})) - \\ -\alpha \, div((\psi_{x}^{'})^{k,l}\nabla(u^{k}+du^{k})) = 0. \end{aligned}$$

$$(4)$$

Аналогично и для второго выражения уравнения Эйлера-Лагранжа. Для устранения остающейся нелинейности в приращениях  $du^k$ ,  $dv^k$  и в выражениях для  $\psi$ , применяется второй внутренний цикл последовательных приближений с индексом *l*. Окончательно выражение для вычисления  $du^{k,l+1}$  записывается

$$(\psi_{D}^{'})^{k,l} \cdot (I_{y}^{k}(I_{z}^{k} + I_{x}^{k}du^{k,l+1} + I_{y}^{k}dv^{k,l+1}) + +\gamma I_{xy}^{k}(I_{xz}^{k} + I_{xx}^{k}du^{k,l+1} + I_{xy}^{k}dv^{k,l+1}) + +\gamma I_{xy}^{k}(I_{yz}^{k} + I_{xy}^{k}du^{k,l+1} + I_{yy}^{k}dv^{k,l+1})) - -\alpha div((\psi_{s}^{'})^{k,l}\nabla(u^{k} + du^{k,l+1})) = 0.$$
(5)

Аналогично для  $du^{k,l+1}$ .

Интерполяция изображения  $I(x + w^k)$  может быть осуществлена билинейной интерполяцией.

#### Численная аппроксимация

Численные методы оказывают определяющие влияние как на точность вычислений, так и на производительность [9].

Аппроксимация дифференциальных уравнений в частных производных (5) операцией конечной разности для искомых приращений вектора смещения du<sup>k</sup> на -ой итерации последовательного приближения записывает-ся в виде

$$\begin{split} &(\psi_{D}^{'})^{k,l} \cdot (I_{z}^{k} + I_{x}^{k} du^{k,l+1} + I_{y}^{k} dv^{k,l+1}) + \\ &+ \gamma I_{xx}^{k} (I_{zz}^{k} + I_{xx}^{k} du^{k,l+1} + I_{xy}^{k} dv^{k,l+1}) + \\ &+ \gamma I_{xy}^{k} (I_{yz}^{k} + I_{xy}^{k} du^{k,l+1} + I_{yy}^{k} dv^{k,l+1})) - \\ &- \alpha \sum_{n=x,y} \sum_{j \in N_{n}(i)} \frac{(\psi_{si}^{'})^{k,l} + (\psi_{si}^{'})^{k,l}}{2} \cdot \frac{u_{j}^{k} + du_{j}^{k,l+1} - u_{i}^{k} - du_{i}^{k,l+1}}{(h_{n}^{k})^{2}} = 0, \end{split}$$

для сетки  $h_x^k \times h_y^k$ . Аналогичным образом и для приращения  $dv^k$ .

Таким образом линеаризация нелинейных уравнений приводит к системе линейных алгебраических уравнений (СЛАУ), которые могут быть численно решены итерационным методом ослабления Гаусса-Зейделя (методом релаксации SOR – *Successive-Over Relaxation*), обладающим повышенной сходимостью.

$$Ax = b,$$
  

$$A = L + D + U,$$
  

$$Dx^{m+1} = (1 - \omega)Dx^m - \omega(Lx^{m+1} + Ux^m) + \omega b.$$

L – нижняя треугольная матрица, U – верхняя треугольная матрица, D – диагональная матрица и  $\omega$  – коэффициент релаксации.

В окончательном виде выражение для итерационного решения системы алгебраических уравнений для приращения *du<sup>k</sup>* запишется:

$$du_{i}^{k,l,m+1} = (1-\omega)du_{i}^{k,l,m} + \omega \frac{F^{-} + F^{+} - F}{R} - \frac{(\psi_{D}^{'})_{i}^{k,l}}{\alpha}(L+\gamma G)}{R},$$
(6)  
rge  $F^{-} = \sum_{j \in N^{-}(i)} (\psi_{s}^{'})_{i-j}^{k,l}(u_{j}^{k} + du_{j}^{k,l,m+1}),$   
 $F^{+} = \sum_{j \in N^{+}(i)} (\psi_{s}^{'})_{i-j}^{k,l}(u_{j}^{k} + du_{j}^{k,l,m}),$ 

$$\begin{split} F &= \sum_{j \in N^{-}(i) \bigcup N^{+}(i)} (\psi_{s}^{'})_{i \sim j}^{k,l} u_{i}^{k}, \\ R &= \sum_{j \in N^{-}(u) \bigcup N^{+}(u)} (\psi_{s}^{'})_{i \sim j}^{k,l} + \\ f &= \frac{(\psi_{D}^{'})_{i}^{k,l}}{\alpha} (((I_{x})_{i}^{k})^{2} + ((I_{xy})_{i}^{k})^{2} + ((I_{xx})_{i}^{k})^{2}), \\ &+ \frac{(\psi_{D}^{'})_{i}^{k,l}}{\alpha} (((I_{x})_{i}^{k})^{2} + ((I_{xy})_{i}^{k})^{2} + ((I_{xx})_{i}^{k})^{2}), \\ L &= (I_{x})_{i}^{k} ((I_{y})_{i}^{k} dv_{i}^{k,l,m} + (I_{z})_{i}^{k}), \\ G &= (I_{xx})_{i}^{k} ((I_{xy})_{i}^{k} dv_{i}^{k,l,m} + (I_{xz})_{i}^{k}) \\ &+ (I_{xy})_{i}^{k} ((I_{yy})_{i}^{k} dv_{i}^{k,l,m} + (I_{yz})_{i}^{k}). \end{split}$$

Аналогично для dv<sup>k</sup>.

Здесь *m* – индекс итерации SOR,  $N^- := \{j \in N_n(i) \mid j < i\}$  означает пикселы, которые только будут обработаны в текущей итерации вычисления SOR,  $N^+ := \{j \in N_n(i) \mid j > i\}$  означает пикселы, которые были вычислены на предыдущей итерации.  $(\psi_s')_{i \sim j}^{k,l}$  коэффициент гладкости вектора между пикселями *i* и *j*.

#### Аппаратная реализация

Декомпозиция задачи вычисления оптического потока на программную (SW-software) и аппаратную (HWhardware) части является естественным для реализации в системе-на-кристалле, содержащую процессорную систему (ARM-процессор) и перепрограммируемую логику (ПЛИС), что вместе образует гибкую вычислительную платформу. На рис. 1 показано распределение задач между программными (SW) и аппаратными (HW) компонентами системы вычисления оптического потока.



Рис. 1. Схема разбиения SoC на программную и аппаратную подсистемы

Основная часть вычисления вариационного оптического потока реализована на ПЛИС как аппаратное ядро для вычисления вектора [10, 11] на текущем уровне пирамиды изображений. Построение пирамид изображений, интерполяция изображения и вектора возлагается на дополнительные программные модули (на ARMпроцессор) из-за дефицита перепрограммируемых логических ресурсов ПЛИС.

Псевдокода обобщенного алгоритма вычисления вариационного оптического потока:

for s = nScales - 1, s > 0, s = s - 1 do // масштабирование изображений Pyramid\_Images = I<sup>s</sup><sub>1,2</sub> \* scaleFactor; while k ≤ outer\_num\_Iterations //итерации по интерполяции изображения //вычисление производных для текущего уровня пирамиды изображения Ix, Ixx, Iyy, Ixy, Iz, Izx, Izy // для I<sup>s</sup><sub>1</sub> и I<sup>s</sup><sub>2</sub> while 1 ≤ inner num Iterations foreach col, row in Images [0, nRows -1]  $\times$  [0, nCols - 1] do // вычисление градиента вектора оптического потока u<sub>x</sub>, u<sub>y</sub>, v<sub>x</sub>, v<sub>y</sub> // функции штрафов для яркости изображений и гладкости потока Vdata, Vsmoth // формирование тензоров движения  $J(col, row) = \nabla_3 I_{1,2} \nabla_3 I_{1,2}^T(col, row);$ // формирование тензоров потока dx2, dxy, dtdx, dy2, dtdy // Successive Over-Relaxation While m ≤ sor num Iterations AND error decrease ≥ MIN ERROR do relax system for du, dv // интерполирование I2 в соответствии с полученными смещениями потока (du, dv).  $I_2$ warp = bilinear interpolation warp( $I_2$ , du, dv); // обновление значения вектора потока u+du,v+dv // интерполирование вектор потока на следующий уровень пирамиды изображений u<sup>s-1</sup>, v<sup>s-1</sup> = bilinear interpolation(u<sup>s</sup>, v<sup>s</sup>,

u<sup>3</sup>, v<sup>3</sup> = bilinear\_interpolation(u scaleFactor)

Разработанная аппаратная реализация алгоритма вычисления вариационного оптического потока орга-низована как конвейерная схема с различными этапами обработки [12], которые изображены на рис. 2.

• На этапе построения разномасштабной пирамиды изображений выполняется построение пирамид исходных изображений с одновременным сглаживанием масштабируемых изображений для подавления шума.

• На этапе вычисление первых и вторых производных исходных изображений происходит вычисление пространственных производных изображений ( $I_x$ ,  $I_y$ ,  $I_{xx}$ ,  $I_{xy}$ ,  $I_x$ ,  $I_y$ ,  $I_{xz}$ ,  $I_{yz}$ ) как компонент функционала оптического потока.

• На этапе вычисление производных вектора оптического потока происходит вычисление скорости изменения (градиента) вектора оптического потока по направлениям ( $u_x$ ,  $u_y$ ,  $v_x$ ,  $v_y$ ).

• Вычисление производных функций отклонения от постоянных значений  $\psi'$  для яркостной и векторной составляющей функционала оптического потока.

• На этапах построения тензоров изменения яркости и градиента (J11, J12, J13, J22, J23) и построения тензоров движения оптического потока (dx2, dxy, dy2, dtdx, dtdy) происходит формирование матриц линейного уравнения (5) вида Ax = b, для вычисления неизвестного смещения оптического потока.

• Этап итерационного решения СЛАУ (5) методом последовательного ослабления Гаусса- Зейделя (метод релаксации SOR).

Таким образом нелинейный вариационный метод реализуется на основе внутреннего и внешнего циклов последовательного приближения, в свою очередь содержащих итерации метода SOR, переопределение  $\psi'$ и обновление тензоров движения и потока соответственно.



Рис. 2. Схема алгоритма вычисления оптического потока На рис. 3 изображена разработанная функциональная схема параллельного вычисления производных, определение функций штрафа и построения тензоров оптического потока. Процессы не зависимы по данным и могут быть вычислены независимо и параллельно. Схема отражает возможность конвейерной организации последовательных этапов вычисления производных и построения тензоров, вплоть до этапа итерационного решения СЛАУ SOR.



Рис. 3. Функциональная схема аппаратной реализации алгоритма вычисления оптического потока на текущем уровне пирамиды изображений

Исходное выражение для вычисления уравнения (6) непосредственно не подходит для аппаратной реализации из-за зависимости данных от каждого *du* and *dv*. Последовательная схема вычисления (SOR, рис. 4 слева) приводит к зависимости по данным, когда для вычисления значения в нечетной позиции, например, 7 (красный элемент в кружке) требуется обновленные в текущей итерации значения двух соседних элементов 2, 6 (красные элементы в кружке) и сохраненные с предыдущей итерации значения двух соседних элементов 8, 12 (черные элементы).

Для возможности параллельной аппаратной реализации модуля решения СЛАУ, применяется представленная на рисунке 4 модификация метода релаксации SOR – метод релаксации «красное-черное» SOR-Red Black (SOR-RB), при котором устраняется зависимость по данным. В схеме параллельного вычислительного процесса (SOR-RB, рис. 4 справа) в текущей итерации для обновления значения в четной позиции, например, 4 (красный элемент в кружке) требуются значения четырех соседних элементов 11, 13, 14, 16 (черные элементы) вычисленных на предыдущей итерации. Т.о. реализуется независимость по данным за счет двухпроходной конвейерной реализации, когда сначала обновляются, например, все красные элементы, а затем все черные [13].

В контексте системы *Ax* = *b* задача переупорядочения сводится к вопросу о перестановках строк и столбцов матрицы А.



Рис. 4. Избавление от зависимости по данным в методе SOR

Эффективная реализация алгоритмов обработки изображений на ПЛИС требует правильного стиля кодирования на языке аппаратного дизайна HDL. При программной реализации алгоритмов обработки изображений изображения представляют собой 2D-массивы, которые находятся в памяти, при аппаратной же реализации на ПЛИС, изображения представляют собой поток пикселей, который не может быть полностью сохранен во внутренней памяти ПЛИС. Как правило, только несколько линий изображения сохраняются в линейных буферах, которые задерживают входящий поток пропорционально ширине линии. В ПЛИС эти линейные буферы отображаются как FIFO, реализованные как двухпортовые блочные ОЗУ (один блок ОЗУ может быть сконфигурирован для хранения 1024 пикселей по 18 бит на пиксель). Для организации двумерного сверточного фильтра с размером ядра 3x3 требуется два линейных буфера (третья строка – это входящий поток) и скользящее окно размером 3х3. Скользящее окно соотносится с окрестностью входных пикселей, которая взвешивается коэффициентами ядра фильтра и суммируется для генерации каждого выходного пикселя. Кроме того, блочные ОЗУ полностью удовлетворяют требованиям линейного буфера к одновременному доступу для чтения и записи. И наоборот, скользящее окно реализуется как параллельные регистры. Такой подход реализуется как в вычислительном модуле SOR-RB, так и в модуле вычисления признаков изображений, а также для синхронизации потоков данных. На рис. 5 показана функциональная схема модуля SOR-RB.

Аппаратная реализация, основанная на аппаратных примитивах и библиотеках элементов современных ПЛИС, позволяет относительно быстро реализовать математические операции в формате с плавающей точкой. Например выражение  $\sqrt{s^2 + \varepsilon^2}$  относительно легко реализуется с помощью функционального генератора операций с плавающей точкой Float Point Arithmetic IP-Соге Generator, создающего аппаратное ядро операций с плавающей точкой на базе ядра DSP IP [14] На рис. 6 показана функциональная схема модуля выделения признаков.







Рис. 6. Функциональная схема аппаратной реализации модуля вычисления признаков

Таким образом, как показано на рис. 7, аппаратные модули для вычисления оптического потока обеспечивают конвейерную и параллельного организацию вычислений на ПЛИС. Основным преимуществом такой конвейерной многоступенчатой организации вычислений является уменьшение количества перемещения данных во внешнюю память и их объема, т.к. не требуется сохранения промежуточных результатов в памяти в связи с тем, что такие данные каждый раз вычисляются «на лету» без сохранения в память. Таким образом, на каждой итерации вычисление тензоров, функций штрафов и производных будет производиться «на лету» и это время вычислений будет намного меньше, чем время доступа к памяти, требующей большую ширину линии данных для чтения и записи промежуточных результатов.



Рис. 7. Функциональная схема аппаратной организации последовательности вычисления оптического потока

Помимо аппаратной оптимизации вычислительного ядра оптического потока требуется организация работы с внешней памятью. Для итерационного вычисления оптического потока на используемой в данной работе архитектуре SoC Zynq [15] предлагается двойная буферизация памяти, которая представлена на рис. 8.



Рис. 8. Организация двойной буферизации памяти

НР0, НР2 – специализированные высокопроизводительные порты доступа к DDR3 памяти системы на кристалле SoC Zynq фирмы Xilinx (HP-high performance port), позволяющие производить обмен данными (входные изображения  $I_1$ ,  $I_2$  и вектор оптического потока u, v с приращениями du, dv посредством прямого доступа к памяти VDMA (Video Direct Memory Access).

Вся внутренняя и внешняя передача данных основана на шине AXI4 (Advanced eXtensible Interface), что позволяет организовать универсальный обмен между различными модулями.

На рис. 9 показана отладочная платформа Xilinx SoC и структурная схема реализации вычисления оптического потока.



Рис. 9. Прототипирование алгоритма на система-на-кристалле SoC Zynq

Отладочная платформа Xilinx ZYNQ7045 SoC на плате ZC706, сочетает двухъядерный процессор ARM Cortex-A9 с программируемой логикой и делает разработку и отладку (в том числе с помощью тестового приложения Chipscope) более гибкой [16].

#### Результаты

Потребление ресурсов ПЛИС для аппаратной части вычисления вариационного оптического потока, состоящей из модуля выделения признаков и 8 объединенных в конвейер модулей SOR-RB, показано на рис. 10. Время выполнения было измерено для одной итерации вычисления приращения вектора *du* и *dv*. Оценка времени вычислений проводилась без учета времени исполнения программной части, т.е. без учета времени интерполяции изображения, интерполяции вектора и построения пирамид изображений.



Рис. 10. Потребление ресурсов ПЛИС

Время выполнения было измерено для отладочной платы с пропускной способностью памяти DDR3 533 МГц x 32 бит x 2 ≈ 4,2 ГБ/с и 150 МГц внутренним тактовым сигналом ПЛИС.

Для наибольшего уровня пирамиды изображения – 685х494 пикселей (изображения  $I_1$ ,  $I_{2warp}$ ) объем считываемых из DDR3-памяти данных  $I_1$ ,  $I_{2warp}$ , u, v,  $du^m$ ,  $dv^m$  – 6x685x494x32 бит ≈ 8 МБ;

Объем записываемых данных в DDR3-память  $du^{m+1}$ ,  $dv^{m+1}$  – 2х685х494х32 bit ≈ 2 MB;

Т.о. время вычисления вектора  $du^{m+1}$ ,  $dv^{m+1}$  составляло ≈ 6 мс.

Для наименьшего уровня пирамиды изображения – 33х24 пикселя (изображения  $I_1$ ,  $I_{2warp}$ ) объем с считываемых из DDR3-памяти данных  $I_1$ ,  $I_{2warp}$ , u, v,  $du^m$ ,  $dv^m$  -6x33x24x32 бит  $\approx$  19 КБ;

Объем записываемых данных в DDR3-память – 2x33x24x32 бит ≈ 6 KB;

Т.о. время вычисления вектора  $du^{m+1}$ ,  $dv^{m+1}$  составляло ≈ 20 us.

В целом время вычисления оптического потока зависит от количества итераций SOR и уровней пирамиды изображения и может варьироваться в зависимости от задачи, но приведенные оценки свидетельствуют о возможности использования разработанной аппаратной реализации для вычисления оптического потока в реальном времени.

#### Заключение

Предложенная аппаратная архитектура нелинейного вариационного алгоритма оптического потока была реализована на языке аппаратного дизайна Verilog для обеспечения высокоэффективного параллелизма вычислений и конвейерной обработки. Использование хорошо развитой специализированной библиотеки компонентов (Xilinx) для организации вычислений в формате с плавающей точкой, позволило реализовать все преимущества нелинейного подхода для вычисления вариационного оптического потока в реальном времени.

Экспериментальные результаты показывают, что высокопроизводительная аппаратно-оптимизированная архитектура для высокоэффективного вычисления вариационного оптического потока может быть реализована как ускоритель на SoC. Для выполнения интерполяции изображения и вектора в реальном времени эти программные модули должны быть также реализованы на аппаратном уровне.

Тем не менее предлагаемое решение из-за высокой производительности и точности в сочетании с энергоэффективностью применимо для вычисления смещения изображений и определения глубины сцены для 3Dреконструкции в задачах автономной навигации роботов (алгоритмы SLAM) [17] когда не требуется высоких скоростей перемещения.

#### Литература

1. B. Lucas and T. Kanade. An iterative image registration technique with an application in stereo vision. In Proc. IEEE Int. Joint Conf., Artificial Intelligence, 1981, pp. 674–679.

2. Елесина С.И., Никифоров М.Б., Логинов А.А., Костяшкин Л.Н. Монография под ред. Л.Н. Костяшкина, М.Б. Никифорова. Совмещение изображений в корреляционно-экстремальных навигационных системах. М.: Радиотехника, 2015. 208 с.

3. B. K. P. Horn and B. G. Schunck. Determining optical flow, Artificial Intelligence, 17:185–203, 1981.

4. Абдухаликов А.А., Беляков П.В., Никифоров М.Б. Реализация на ПЛИС алгоритма поиска ключевых точек на изображении. Международная научно-техническая и научно-методическая конференция «Современные технологии в науке и образовании» СТНО-2016, 2016, с. 103-108.

5. T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In Proc. European Conf., Computer Vision, volume 4, 2004, pp. 25–36.

6. Обработка изображений в авиационных системах технического зрения / Под ред. Л.Н. Костяшкина, М.Б. Никифорова. М.: ФИЗМАТЛИТ, 2016, с. 28-32

7. D. Ustukov, Y. Muratov, M. Nikiforov, V. Gurov. Implementing one of stereovision algorithms on FPGA. Mediterranean Conference on Embedded Computing, Jun 2016.

8. A. Bruhn and J. Weicker. Towards ultimate motion estimation: combing highest accuracy with real-time performance. In Proc. 10th IEEE Int.Conf., Computer Vision, 2005, pp. 749–755.

9. A. Bruhn, J. Weickert, and C. Schnorr. Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. Int. J. Computer Vision, 2005, 61:211–231.

10. M. Kunz, A. Ostrowski, P. Zipf. An FPGA-optimized architecture of Horn and Schunck optical flow algorithm for real-time applications. Field Programmable Logic and Applications (FPL), 2014 24th International Conference.

11. J.L. Martin, A. Zuloaga, C. Cuadrado, J. Lazaro, and U. Bidarte. Hardware implementation of optical flow constraint equation using fpgas. Computer Vision and Image Understanding, 2005, pp 462–490.

12. Z. Chai, H. Zhou, †Z. Wang and ‡D. Wu Using C to implement high-efficient computation of dense optical flow on FPGA-accelerated heterogeneous platforms. IEEE 14 International Conference on Field-Programmable Technology (FPT), 2014.

13. Ortega, James M. Introduction to Parallel and Vector Solution of Linear Systems, 1988.

14. Xilinx. Zynq-7000 SoC. https://www.xilinx.com/support/documentation/user\_guides/ug479\_7Series\_DSP48E 1.pdf.

15. Xilinx. Zynq-7000 SoC. http://www.xilinx.com/products/silicon-devices/soc/zynq-7000/index.htm.

16. Xilinx. Vivado Design Suite. http://www.xilinx.com/ products/design-tools/vivado.

17. Ларкин Е.В. Моделирование процесса дистанционного управления роботом. Извести ТулГУ. Технические науки, 2016, Вып. 12. Ч. 4, с. 202-214.

## Уважаемые коллеги!

Для тех, кто не успел оформить подписку на первое полугодие 2019 года через АО «Роспечать», сохраняется возможность приобретения журналов непосредственно в редакции по адресу: г. Москва, ул. Авиамоторная, дом 8, Научный Центр МТУСИ, ком. 612. Российское научно-техническое общество радиотехники, электроники и связи им. А.С. Попова, метро «Авиамоторная», или оформить Заказ в соответствии с требованиями, выставленными на сайте журнала: www.dspa.ru.

Справки по телефону: (+7 903)201-53-33 (Самсонов Геннадий Андреевич).

E-mail: rntores@mail.ru

## О РЕАЛЬНОСТЯХ ТЕОРИИ КОДИРОВАНИЯ И ЕЁ ПРИЛОЖЕНИЙ

(Ответ на замечания читателей) В.В. Золотарёв

Как автор книги благодарю профессоров М.А. Быховского и В.П. Дворковича за внимание к моей новой монографии в их отзыве, напечатанном в предыдущем номере журнала Цифровая обработка сигналов (ЦОС). В их отзыве имеется ряд критических замечаний, на которые следует ответить.

Во-первых, введённое в книге сокращение «ОТ» воспроизведено везде в отзыве как Основная Теорема, тогда как это значит «Оптимизационная Теория». В книге, как и полагается, этот термин и его сокращение приведены вместе при их самом первом появлении на стр. 6. Эта ошибка оппонентов очень сильно меняет смысл многих выражений отзыва на книгу.

Во-вторых, необходимо отметить, что приведённые рецензентами многие цитаты из моей книги соответствуют оценкам очень большого числа других специалистов, которые они дают нашей Оптимизационной Теории (ОТ). Но тут есть и моя оценка ОТ как руководителя передовой научной школы мира по этой теме. С этим мнением вполне согласны все наши многочисленные рецензенты, сторонники, соавторы монографий, а также многие научные редакторы – члены-корреспонденты и академики РАН. Эти мнения уже вполне устоялись за последние 15..20 лет. Нас поддерживает и зарубежье. Ваш покорный слуга был недавно награждён Золотой медалью Евросоюза (ЕС) «За исключительные достижения», которой могут быть удостоены только учёные, имеющие действительно выдающиеся научные результаты. А после строгой экспертизы Международным союзом электросвязи (МСЭ/ITU) в Женеве по рекомендации его Генерального секретаря издана наша монография по Оптимизационной теории кодирования (Zolotarev V.V., Zubarev Y.B., Ovechkin G.V. Optimization Coding Theory and Multithreshold Algorithms, ITU, 2015).

В ответ на замечание про определение теории кодирования отвечу, что, безусловно, теория кодирования не сводится только к задаче поиска глобального экстремума, решаемой при декодировании принятых сообщений. Она, конечно, гораздо шире (в том числе много шире определения, данного авторами отзыва). Но в монографии рассматриваются основные проблемы, с точки зрения нашей научной школы, теории помехоустойчивого кодирования: разработка декодеров, обеспечивающих близкое к оптимальному декодирование длинных помехоустойчивого кодирования: разработка декодеров, обеспечивающих близкое к оптимальному декодирование длинных помехоустойчивого кодирования: разработка декодеров, обеспечивающих близкое к оптимальному декодирование длинных помехоустойчивых кодов с наименьшей вычислительной сложностью, и построение кодов, наилучшим образом подходящих для этих декодеров. При этом используются предложенные автором монографии оптимизационные процедуры, которые позволяют получить впечатляющие результаты, недостижимые для других методов коррекции ошибок с сопоставимой вычислительной сложностью. Все это по мнению многих специалистов, действительно позволяет считать разработанную оптимизационную теорию «новой квантовой механикой» теории информации. При этом, отмечу, что термин «квантовые каналы», поставленный мне в вину в отзыве, в монографии совсем не используется. Скорее всего авторы отзыва так восприняли стандартные модели гауссовских каналов с квантованием выходного сигнала демодулятора на несколько уровней.

Далее отвечу на пронумерованные замечания.

1. Монография посвящена разработке не систем связи, а методов кодирования/декодирования помехоустойчивых кодов. Поэтому в ней используются не предложенные уважаемыми рецензентами показатели эффективности систем связи, а те, которые традиционно используются при сравнении эффективности различных методов помехоустойчивого кодирования в типичных моделях каналов связи. По этой же причине в монографии нет сравнения с «рядом эффективных методов передачи и приема сообщений», перечисленных рецензентами. Это совсем не нужно. В то же время полученные в монографии характеристики сравниваются с характеристиками, обеспечиваемыми классическим декодером Витерби сверточных кодов, декодерами турбо и низкоплотностных кодов, которые считаются одними из наиболее перспективных для применения в высокоскоростных системах передачи данных, и характеристики декодеровнекоторых других кодов.

2. Под «уровнем шума» в монографии понимается некоторая характеристика, влияющая на вероятность ошибки при передаче цифровых данных по каналу с шумом и зависящая от модели используемого канала. Так для двоичного симметричного канала это вероятность битовой ошибки, для *q*-ичного симметричного канала – вероятность символьной ошибки, для канала со стираниями – вероятность стирания, для гауссовского канала – отношение сигнал-шум на входе демодулятора. Использование единого термина «уровень шума» с нашей точки зрения позволяет несколько упростить изложение материала. Так же здесь отмечу, что конкретные отличия пропускной способности каналов и границ рабочих областей наших декодеров всегда указываются в тексте монографии на многих графиках для разных кодов и каналов.

3. В первой части п. 3 отзыва его авторы предлагают анализировать не конкретные коды, а ансамбли кодов, выявляя общие закономерности изменения их помехоустойчивости с изменением длины. Но дело в том, что проблемы декодирования – это всегда алгоритмические задачи создания только конкретных кодов. Нельзя сделать декодер для ансамбля кодов! При этом конкретный декодер можно применять в миллионах разных систем. Именно поэтому в наших работах внимание уделяется конкретным кодам с указанием характеристик их декодирования. При этом для оценки эффективности конкретного декодера для конкретного кода при большом уровне шума в нашем случае необходимо использовать компьютерное моделирование, при малом же шуме эффективность кода и декодера всегда можно оценить аналитически с помощью представленных в монографии выражений.

4. Утверждение, приведенное в монографии, про то, что рассматриваемые в ней коды позволяют исправлять более, чем *d*/2 ошибок, несмотря на замечание рецензентов, является верным. Это известно уже более 50 лет со времени выхода в свет книги Дж. Месси «Пороговое декодирование». Причем этот факт, подробно обсуждаемый в разделе 2.1 монографии, является отправным моментом всей обсуждаемой оптимизационной теории кодирования. Дополнительно отмечу, что и многие другие помехоустойчивые коды (турбо, низкоплотностные и даже некоторые алгебраические, такие как коды Рида-Соломона) позволяют исправлять больше ошибок, чем гарантируется минимальным кодовым расстоянием. Безусловно, сделать они это могут не всегда и обязательно существуют конфигурации из более, чем *d*/2 ошибок, которые приведут к ошибке декодирования. Но такие конфигурации встречаются гораздо реже исправляемых конфигураций ошибок.

5. Поскольку при большом уровне шума (малом отношении сигнал-шум на выходе демодулятора) на эффективность декодера влияние оказывает не столько минимальное кодовое расстояние, сколько спектр кода и многие другие параметры кодера и декодера, пропадает необходимость сравнения построенных нами кодов с другими с помощью предложенных рецензентами границ. Хотя граница Хэмминга (т.е. сферическая упаковка) обсуждается в монографии в разделе 1.6.

6. Утверждение рецензентов, что «характеристики рассмотренных кодов наверняка уступают характеристикам кодов с максимально достижимым расстоянием (МДР)» выглядит весьма странно и бездоказательно. При этом, для символьных кодов в разделе 3.5.3 достаточно подробно на примере характеристик конкретных кодов показано, что коды Рида-Соломона, относящиеся к МДР-кодам, существенно уступают предлагаемым в монографии символьным кодам при одинаковом размере символа. Безусловно, в монографии отмечено, что это оказалось возможным в том числе за счет того, что наши коды длиннее кодов Рида-Соломона. Но более длинных кодов Рида-Соломона при таком же размере символа, к сожалению, нет и ничего с этим поделать нельзя. Даже использование всевозможных вариантов организации каскадирования на базе кодов Рида-Соломона не позволяет получить лучшие, чем в монографии результаты. Кроме того, сложность алгебраического декодера кода Рида-Соломона пропорциональна квадрату длины кода, в то время как сложность нашего декодера всего лишь линейная. Хотя, справедливости ради, можно отметить, что для своей длины коды Рида-Соломона действительно обеспечивают очень хорошие характеристики.

Что касается замечания про нашу недооценку фундаментальных результатов, полученными многими отечественными и зарубежными специалистами, то ее нет. Наша школа с величайшим почтением изучает и использует труды многих учёных: К. Шеннона, Дж. Месси, А. Витерби, Д. Форни, Дж. Кларка, Дж.П. Робинсона, Д. Салливена, В.М. Готлиба, Э.М. Габидулина, А.Л. Ларина и многих других уважаемых специалистов. И это не запоздалые теперь уже реверансы в адрес моих коллег из-за обвинений рецензентов. Ссылки на разные работы в области теории кодирования сделаны мной в книге на страницах: 21, 24, 96, 155..159 и далее, не считая общего списка литературы из 72-х позиций, где есть не только моя фамилия. Но в процитированных рецензентами со стр. 197 недостатках алгебраической теории ошибок нет. Действительно, все перечисленные проблемы ею не решены, и именно они ещё много десятилетий назад стали причиной её заката во всех прикладных аспектах.

Таким образом, в комментариях рецензентов не выявлено ни одной даже мелкой неточности автора в изложении важнейшей для цифровой связи новой Оптимизационной Теории. Наша научная школа, действительно, решила на высоком научном и технологическом уровнях проблему простого высокодостоверного декодирования вблизи пропускной способности канала, поставленную 70 лет назад К. Шенноном.

В заключении я выражаю редакции журнала свою признательность за терпеливое ожидание моего ответа на критику рецензентов и за резервирование для этого тех ограниченных ресурсов журнала, которые, конечно, в другой ситуации можно было бы выделить для других статей. Поэтому более точный и детальный ответ на замечания оппонентов, восстанавливающий научную репутацию монографии, её автора и уважаемого журнала по современным цифровым технологиям, представлен на портале нашей научной школы по адресу http://www.mtdbest.ru/books.html, где уже есть материалы и про нашу новую книгу, в частности, рецензия, опубликованная в журнале «Электросвязь». Рассмотренный здесь мной текст оппонентов тоже там представлен. Там же мы поместим мнения о нашей монографии других наших читателей.

д.т.н., профессор В.В. Золотарёв

– автор монографии «Теория кодирования как задача поиска глобального экстремума (Оптимизационная Теория помехоустойчивого кодирования – новая «квантовая механика» теории информации)». М., Горячая линия – Телеком, 2018, 221 с.;

– Лауреат премии Правительства РФ по науке и технике и Золотой медали Евросоюза (ЕС) «За исключительные достижения».

zolotasd@yandex.ru , www.mtdbest.iki.rssi.ru , www.mtdbest.ru моб.: 8-916-518-86-28