

ЦИФРОВАЯ ОБРАБОТКА СИГНАЛОВ

Научно-технический журнал
№ 3/2023

Издается с 1999 года
Выходит четыре раза в год

И.О. ГЛАВНОГО РЕДАКТОРА
д.т.н., член-корр. РАН А.В. ДВОРКОВИЧ

ЗАМЕСТИТЕЛЬ ГЛАВНОГО РЕДАКТОРА:
д.т.н., проф. В.В. ВИТЯЗЕВ

РЕДАКЦИОННАЯ КОЛЛЕГИЯ:

д.т.н., проф. Ар.С. Аджемов, д.т.н., проф. Б.А. Алпатов,
д.т.н., проф. В.Г. Бартенов, д.т.н. Ю.И. Борисов,
д.т.н., проф. Ю.А. Брюханов, д.т.н., проф. 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., акад. РАН Н.А. Кузнецов,
В.Г. Мистюков, д.т.н., проф. С.Л. Мишенков,
д.т.н., проф. Ю.Н. Прохоров,
д.т.н. А.Л. Приоров, д.т.н., проф. В.Г. Санников,
д.т.н., проф. Чиров Д.С.,
к.т.н., проф. В.С. Сперанский.

Адрес редакции:

г. 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.dspsa.ru

Издатель:

Российское научно-техническое общество
радиотехники, электроники и связи им. А.С. Попова
Компьютерная верстка: И.А. Благодарова
Дизайн: М.В. Аверин
Подписано в печать 04.09.23 г.
Формат 60x90/8.

Гарнитура «Arial». Печать офсетная.
Бумага офсетная. Печ.л. 6,5. Тираж 200 экз.
Заказ № 3688. 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.

УЧРЕДИТЕЛИ:

АО «Инструментальные системы»
ФГУП «НИИ радио»

ООО «Российское научно-техническое общество
радиотехники, электроники
и связи им. А.С. Попова»

В НОМЕРЕ:

- Еремеев В.В., Егюшкин Н.А.,
Москвитин А.Э., Соловьев А.В.**
Математические модели и программы
мониторинга качества обработки изображений
от высокоорбитальных космических систем 3
- Москвитин А.Э., Ушенкин В.А., Ларюков С.А.**
Алгоритм и программный комплекс высокоскоростной
нейросетевой сегментации облачности на панхроматических
изображениях от космических аппаратов «Ресурс-П» 8
- Ушенкин В.А.**
Адаптивное нейросетевое сжатие мультиспектральных
космических изображений поверхности Земли 18
- Хрящев В.В., Приоров А.Л., Лебедев А.А., Котов Н.А.**
Обнаружение полипов на колоноскопических изображениях
при помощи алгоритмов
на базе нейросетевой архитектуры YOLO 27
- Кузнецов В.П.**
Комплексная нейронная сеть 32
- Смирнов А.А., Смирнов А.В.**
Двухэтапный алгоритм параметрической оптимизации
весовой пространственно-временной ранговой
фильтрации изображений 37
- Асаф М., Пономарев О.Г.**
Итеративное гибридное прекодирование
с низкой сложностью для массовых MIMO систем
в миллиметровом диапазоне радиоволн 41
- Шарамет А.В., Азаров И.С.**
Особенности работы приемного тракта
в многоканальной радиолокационной станции
с временной автоматической регулировкой усиления 47
- Егоров В.В., Клионский Д.М.**
Некоторые вопросы применения адаптивного метода
декомпозиции нестационарных сигналов на эмпирические моды .. 52
- Жестерев А.Е., Ипатов В.П.**
Оптимальное измерение запаздывания эхо-сигнала
спутникового высотомера по коррелированным отсчетам
наблюдения в условиях морского волнения 59
- Синицын И.А., Галаева К.И.**
Разработка метода и алгоритмов оценки сдвигов ветра
и турбулентности в метеорологическом радиолокационном
комплексе аэродромной зоны 64
- Андреев А.Г., Сенченко А.А.**
Оптимальный КИХ фильтр с дробной задержкой 74
- Дворянков Д.А., Андросов В.В.,
Волченков В.А., Витязев С.**
Восстановление траекторного сигнала
в условиях периодических затенений методом
авторегрессионного анализа 78

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

Журнал «Цифровая обработка сигналов» включен
в базу данных Web of Science –
Russian Science Citation Index (RSCI)

Digital Signal Processing

Science & Technical Magazine
Issue 3, 2023 year

Is published quarterly since 1999

ACTING EDITOR-IN-CHIEF:

Dr. Sci (Tech), Professor
Corresponding Member of RAS A.V. Dvorkovich

DEPUTY EDITORS-IN-CHIEF:

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

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
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.,
Kuznetsov N.A., Dr.Sci.(Tech.), Academician
Mistyukov V.G.,
Mishenkov S.L., Dr.Sci.(Tech.), Professor
Priorov A.L., Dr. Sci (Tech)
Prokhorov Y.N., Dr.Sci.(Tech.), Professor
Sannikov V.G., Dr.Sci.(Tech.), Professor
Chirov D.S., Dr.Sci.(Tech.), Professor
Speranskii V.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.dspsa.ru

Publisher:

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

Computer makeup:

I.A. Blagodarova

FOUNDERS: InSys

Radio Research and Development Institute

CONTENTS:

| | |
|--|----|
| Eremeev V.V., Egoshkin N.A., Moskvitin A.E., Solovev A.V. <i>Peculiarities of the arktika-m spacecraft target information quality monitoring mathematical models and programs.....</i> | 3 |
| Moskrvitin A.E., Ushenkin V.A., Laryukov S.A. <i>Algorithm and software of High-speed neural network cloud segmentation of Resurs-P satellite's panchromatic images.....</i> | 8 |
| Ushenkin V.A. <i>Adaptive neural network compression of multispectral satellite images of the earth's surface.....</i> | 18 |
| Khryashchev V.V., Priorov A.L., Lebedev A.A., Kotov N.A. <i>Detection of polyps on colonoscopic images using algorithms based on the YOLO neural network architecture</i> | 27 |
| Kuznetsov V.P. <i>Complex neural network.....</i> | 32 |
| Smirnov A.A., Smirnov A.V. <i>Two-stage algorithm for parametric optimization of weighted space-time rank filtering of images.....</i> | 37 |
| Assaf M., Ponomarev O.G. <i>Low Complexity Iterative Hybrid Precoding for Millimeter-Wave Massive MIMO Systems</i> | 41 |
| Sheremet A.V., Azarov I.S. <i>Features of operation of the receiving path in a multichannel radar station with temporary automatic gain control</i> | 47 |
| Egorov V.V., Klionskiy D.M. <i>Some aspects of applying the adaptive technique of empirical mode decomposition of non-stationary signals</i> | 52 |
| Zhesterev A.E., Ipatov V.P. <i>optimal estimate of the satellite altimeter echo-signal delay from correlated observation samples under rough sea state</i> | 59 |
| Sinitsyn I.A., Galaeva K.I. <i>Development of a method and algorithms for estimating wind shear and turbulence in the meteorological radar complex of the aerodrome zone</i> | 64 |
| Andreev A.G., Senchenko A.A. <i>Optimal Fractional Delay FIR Filters</i> | 74 |
| Dvoryankov D.A., Androsov V.V., Volchenkov V.A., Vityazev S.V. <i>Trajectory signal restoring in conditions of periodic shading by autoregressive analysis method</i> | 78 |



УДК 004.932

МАТЕМАТИЧЕСКИЕ МОДЕЛИ И ПРОГРАММЫ МОНИТОРИНГА КАЧЕСТВА ОБРАБОТКИ ИЗОБРАЖЕНИЙ ОТ ВЫСОКООРБИТАЛЬНЫХ КОСМИЧЕСКИХ СИСТЕМ

Еремеев В.В., д.т.н., профессор, директор НИИ «ФОТОН» Рязанского государственного радиотехнического университета им. В.Ф. Уткина, e-mail: foton@rsreu.ru

Егошкин Н.А., д.т.н., в.н.с. НИИ «ФОТОН» Рязанского государственного радиотехнического университета им. В.Ф. Уткина, e-mail: foton@rsreu.ru

Москвитин А.Э., д.т.н., доцент, в.н.с. НИИ «ФОТОН» Рязанского государственного радиотехнического университета им. В.Ф. Уткина, e-mail: foton@rsreu.ru

Соловьев А.В., инженер РГРТУ им. В.Ф. Уткина.

PECULIARITIES OF THE ARKTIKA-M SPACECRAFT TARGET INFORMATION QUALITY MONITORING MATHEMATICAL MODELS AND PROGRAMS

Eremeev V.V., Egoshkin N.A., Moskvitin A.E., Solovev A.V.

The paper discusses the features of periodic quality control of information coming from the Arktika-M spacecraft. Monitoring is performed by statistical service information analysis and visual images inspection. Based on the monitoring results, the calibration of the software parameters are refined in order to improve the quality of target information processing.

Key words: quality control of information, monitoring, visual images inspection, calibration of the software parameters.

Ключевые слова мониторинг качества, телеметрические данные, дистанционное зондирование Земли, высокоуровневая обработка спутниковых снимков, контроль функционирования программных средств.

Введение

Для успешного функционирования спутников дистанционного зондирования Земли необходимо регулярно проводить анализ качества информационных продуктов, формируемых космической системой. В состав системы ДЗЗ входят космический аппарат (КА) и наземный комплекс приема и обработки спутниковой информации.

Обработку данных с высокоорбитальных КА на Центрах приема спутниковых данных Научно-исследовательского центра «Планета» осуществляет программное обеспечение разработки Рязанского государственного радиотехнического университета. Специалисты РГРТУ регулярно проводят мониторинг качества работы спутниковой аппаратуры и правильности функционирования наземных средств. По результатам этой работы вводятся поправки в модели, алгоритмы и технологии формирования стандартных информационных продуктов, повышающие их качество, например, точность радиометрических измерений, точность геометрических преобразований, точность геодезической привязки.

Постановка задачи

Программное обеспечение наземной обработки данных от высокоорбитальных осуществляет формирование изображений уровней обработки 1.0 и 1.5. Уровень

Рассматриваются особенности систематического контроля качества информации, поступающей с высокоэллиптических и геостационарных космических аппаратов. Мониторинг проводится путем построения моделей и статистического анализа изображений и служебной информации. По результатам мониторинга производится уточнение параметров моделей сканирующих устройств, заложенных в программном обеспечении обработки спутниковых изображений.

обработки 1.0 подразумевает распаковку, декодирование принимаемого потока данных, его разделение на видеoinформацию и служебную информацию, представление данных в нужных форматах.

Уровень обработки 1.5 подразумевает выполнение геопривязки, а также коррекцию возможных искажений, обусловленных конструктивными особенностями космического аппарата и съемочной аппаратуры [1, 2].

Высокоорбитальные гидрометеорологические комплексы осуществляют съемку Земли в трех диапазонах видимого спектра и семи зонах инфракрасного диапазона спектра электромагнитных волн. Результатом обработки каждого сеанса съемки является формирование набора файлов, содержащих выходные продукты уровней 1.0 и 1.5 отдельно. Эти файлы можно открыть только специализированным программным обеспечением. Ставится задача контроля правильности функционирования программных средств обработки данных высокоорбитальных КА путем визуального анализа изображений, контроля полноты и целостности служебных данных и путем анализа работы бортовой аппаратуры спутника по телеметрической информации.

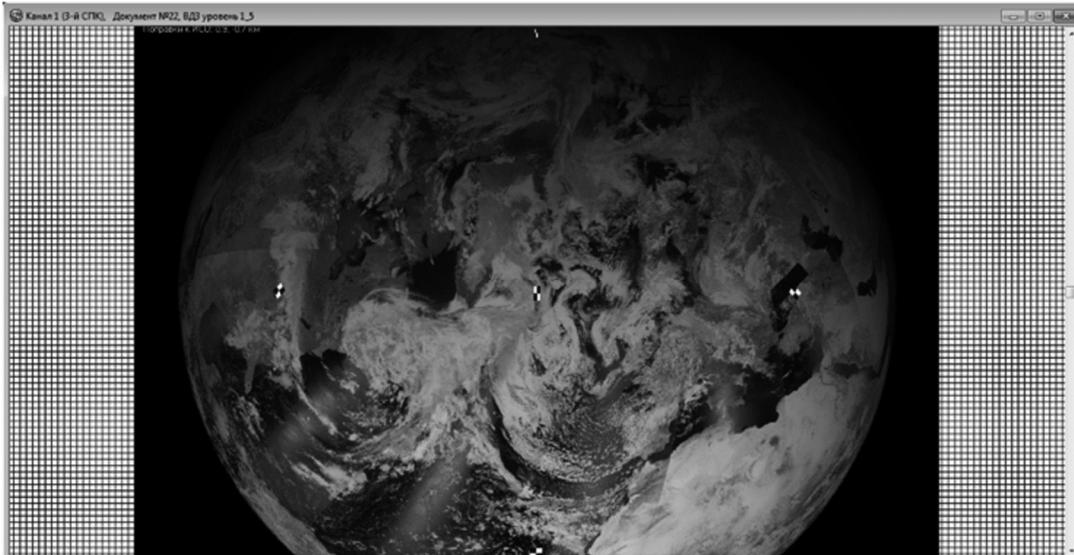


Рис. 1. Пример изображения уровня 1.5

Визуальный контроль спутниковых данных

Осуществляя визуальный контроль сформированных выходных продуктов, можно оперативно сделать вывод о правильности работы программных средств за длительный период.

Для быстрой визуальной оценки содержания выходных продуктов формируются так называемые «квиклуки». «Квиклука» – это копия первоначального продукта с низким разрешением и в распространенном формате, которая сохраняется для быстрого доступа в нужный момент. «Квиклуки» позволяют оператору мгновенно открыть и показать продукты обработки, без необходимости применять специализированные инструменты для просмотра снимков.

Однако часто «квиклуков» недостаточно для адекватной визуальной оценки качества выходных информационных продуктов. Поэтому в инструментарий анализа качества обработки данных ДЗЗ целесообразно включить возможность открытия выходных изображений. Файл, сформированный в результате обработки уровня 1.5, представляет собой черно-белое изображение, каждый пиксель которого занимает 10 бит, что позволяет хранить 1024 градации серого. Пример такого изображения приведен на рис. 1.

Анализ полноты и целостности служебных данных

Полнота информации, полученной в результате приема данных сеанса связи со спутником, анализируется автоматически путем контроля присутствия всех видов информации и ее объема. Для изображений определяются: имя и размер файла, номер канала, число строк в тестовых наборах и в видеоданных, общий размер строки, количество элементов видеоданных в строке, число сканов, число строк телеметрической и служебной информации в скане и другие. Вывод о полноте данных формируется путем сопоставления значений этих параметров с протоколом.

Целостность информации определяется путем оценки вероятности сбоев параметров служебной информа-

ции, значения которых заранее известны, и путем определения среднего значения, СКО и вероятности выхода за допустимый диапазон динамически изменяющихся служебных данных.

Оценка вероятности сбоев выполняется для параметров трех групп:

– первая группа – это параметры, принимающие только одно заранее известное значение (X_{const}). Для таких параметров вероятность сбоев оценивается как

$$v = N_{сбоев} / N,$$

$$\text{где } N_{сбоев} = \sum_{i=1}^N s_i, \quad s_i = \begin{cases} 1, & \text{если } X_i \neq X_{const}; \\ 0, & \text{в противном случае;} \end{cases} \quad X_i,$$

$i = \overline{1, N}$, – значения анализируемого параметра;

– вторая группа – это параметры, принимающие несколько заранее известных значений $W = \{w_k\}$; $k = \overline{1, K}$. Здесь сначала определяется истинное значение параметра как ближайшее к его среднему значению, т.е.

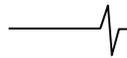
$$X_{const} = \min_k \left| \bar{X} - w_k \right|,$$

где $\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$, $i = \overline{1, N}$, $k = \overline{1, K}$. Затем оценивается

вероятность сбоя также, как и для параметра первой группы;

– третья группа – это параметры, для которых заранее известен только закон изменения, например, «Номер строки». При анализе параметра «Номер строки» полагается, что при сбоях возможен как пропуск строк, так и запись ложных значений. Работа алгоритма начинается с поиска индекса j , с которого начинается корректный рост номера строки, т.е. $X_j + N_p = X_{j+1} + N_p - 1 = \dots = X_{j+N}$, где $N_p + 1$ – размер участка с корректным ростом. Затем, начиная с этого места, выполняется подсчет верных значений параметра:

$$N_{кopp} = \sum_{k=1}^j a_k + \sum_{k=j+1}^N b_k,$$



где $a_k = \begin{cases} 1, & \text{если } X_{j-k+1} - X_{j-k+2} = 1; \\ 0, & \text{в противном случае;} \end{cases}$

$b_k = \begin{cases} 1, & \text{если } X_k - X_{k+1} = 1; \\ 0, & \text{в противном случае.} \end{cases}$ Искомая вероятность

сбоя оценивается как $v = (N - N_{\text{копп}}) / N$.

Анализ динамики сигналов служебной информации осуществляется путем вычисления:

– среднего значения $M = \frac{1}{N} \sum_{i=1}^N X_i$, где $X_i, i = \overline{1, N}$,

значение анализируемого сигнала;

– среднеквадратического отклонения

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - M)^2};$$

– вероятности выхода за допустимый диапазон

$(X_{\min}, X_{\max}) \quad v = N_{\text{вых}} / N, \quad \text{где} \quad N_{\text{вых}} = \sum_{i=1}^N h_i,$

$h_i = \begin{cases} 0, & \text{если } X_i > X_{\max} \text{ и } X_i < X_{\min}; \\ 1, & \text{в противном случае.} \end{cases}$

Контроль работы бортовой спутниковой аппаратуры

Помимо собственно изображений интерес для контроля представляет также информация о параметрах работы спутника и съёмочной аппаратуры. Эта информация по мере наземной обработки извлекается из изображений и другой сопроводительной информации и «бережно» накапливается и архивируется для дальнейшего анализа. Её анализ менее трудоёмок, чем изображений, и зачастую более информативен, поскольку информация содержится в более концентрированном виде.

Обозначим $v(t)$ зависимость вектора поправок, который рассчитывает программа нормализации путем анализа изображений, от времени t (номера сеанса). Компоненты $v(t)$ включают:

– поправки к углам ориентации спутника (к навигационной информации ИСО), которые оцениваются в процессе уточнения геопривязки по опорным ориентирам – контурным точкам диска Земли [3] и электронным картам [4];

– оценки геометрического сдвига одного спектрального канала относительно другого (формируются путём корреляционно-экстремального сопоставления изображений);

– оценки сдвига между собой сканов изображений, в том числе формируемых двумя ПЗС-линейками в модуле видимого диапазона.

Обозначим $s(t)$ – зависимость от t вектора, характеризующего работу бортовой аппаратуры. Компоненты $s(t)$ включают:

– измерения температур в ряде ключевых точек камеры и спутника;

– оценки плавности хода сканирующих зеркал по датчикам углового положения (ибо высокий уровень флуктуаций угла снижает качество изображений и тре-

бует специальной обработки);

– нормы кватернионов ориентации от системы навигации борта (в идеале должно быть 1);

– целостность и полнота исходной информации от станции космической связи как на уровне абстрактных пакетов от бортовой системы сбора и обработки данных, так и по количеству успешно принятых строк изображений и полноте навигационной информации (количество сообщений с борта и их соответствие по времени интервалу съёмки).

Поправки $v(t)$ и измерения $s(t)$ являются исходной информацией для контроля качества функционирования спутника, а также для потенциального улучшения качества геометрической нормализации.

Получение $s(t)$ связано с обработкой измерительной информации с борта КА и не требует высоких вычислительных затрат. Определение $v(t)$ наоборот, требует обработки объёмных изображений и весьма трудоёмко. Чем больше диапазон возможного изменения компонент $v(t)$, тем более высоки затраты (больше зона поиска коррелятора). Поэтому необходимо контролировать значения поправок, не допуская их приближения к границам зоны корреляционного поиска: если $|v_i(t)| > P_i$, где P_i – заданный порог, i – индекс компоненты вектора, то выдаётся предупреждение о необходимости калибровки параметров аппаратуры или о возникшей нештатной ситуации на борту спутника.

Также важная информация может быть извлечена путем совместного анализа $v(t)$ и $s(t)$ для нескольких моментов времени сразу. Для визуального анализа разработан инструментарий для отображения компонент $v(t)$ и $s(t)$ в виде графиков. На графиках одна точка соответствует одному сеансу съёмки (рис. 2).

На графиках можно видеть суточно-периодические изменения, а также «вековые». Можно видеть связь между поправками (т.е. геометрией съёмочной аппаратуры) и температурами, которая особо ярко начинает проявляться после уменьшения периодичности съёмки с 30 на 15 минут, что заметно ухудшает термостабильность спутника. Аномалии на графиках (например, резкие выбросы и нехарактерные изменения) свидетельствуют о необходимости более детально проанализировать изображения.

Для автоматизации выявления аномалий применяется критерий 3σ : по группе измерений $v_i(t)$ за сутки находится оценка σ_i и если $|v_i(t)| > 3\sigma_i$, то фиксируется наличие аномалии.

При обработке изображений корреляционно-экстремальным подходом оценивается геометрическое рассогласование между спектральными каналами. Это несколько повышает время обработки, однако даёт крайне важную информацию для изучения суточно-периодического и «векового» изменения рассогласований. Так было обнаружено постепенное расхождение каналов со временем (рис. 3). В настоящее время выход рассогласований за проектные значения является сигналом для перекалибровки параметров совмещения каналов в комплексе нормализации. Для этого помимо

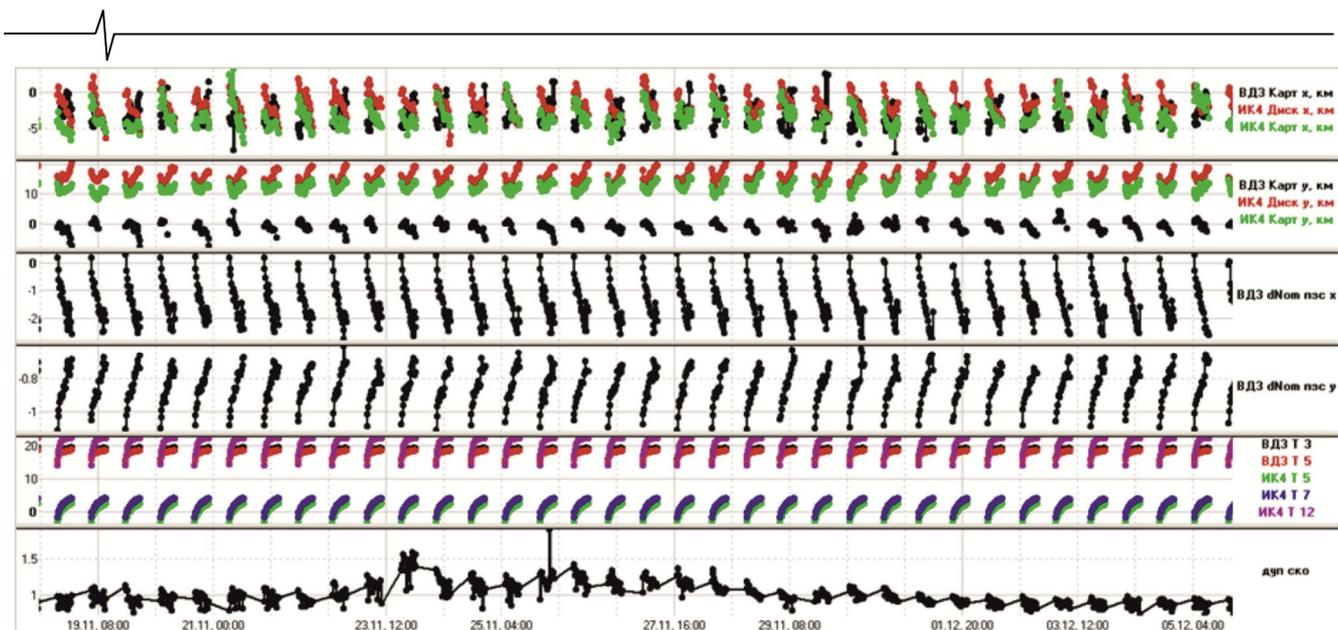


Рис. 2. Эволюция контролируемых параметров во времени

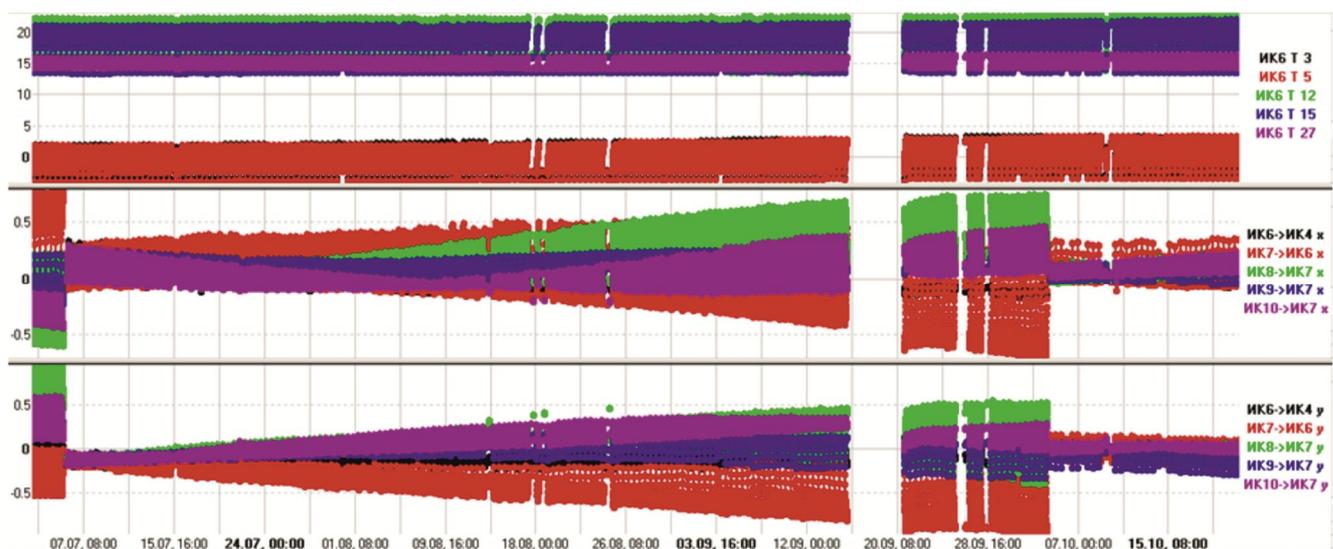


Рис. 3. Эволюция сдвига спектральных каналов во времени

обобщенных оценок точности совмещения накапливаются массивы одноименных точек (появление рассогласования говорит об изменении геометрии съёмки аппаратуры по сравнению с заложенной моделью, но самих интегральных оценок рассогласования может быть недостаточно для уточнения моделей). В результате калибровки погрешность совмещения каналов резко снижается (см. рис. 3).

Накопление поправок $\mathbf{v}(t)$ и измерений $\mathbf{s}(t)$ позволяет не только проводить контроль, но и улучшить обработку. Наиболее фундаментально это может быть сформулировано в терминах предсказателей: поставим задачу предсказать $\mathbf{v}(t_0)$, т.е. найти некоторое $\mathbf{v}^*(t_0)$, зная только поправки $\mathbf{v}(t)$ для $t < t_0$ и $\mathbf{s}(t)$ для $t \leq t_0$. Критерием построения предсказателя может служить условие качества предсказания $\varepsilon = \|\mathbf{v}(t_0) - \mathbf{v}^*(t_0)\|_2^2 = \min$.

Знание прогноза позволяет, во-первых, упростить анализ изображения за счет сужения области корреля-

ционного поиска, центром которого естественно выбрать предсказанное значение. Во-вторых, упрощается выявление аномалий в функционировании системы: высокие значения $|\mathbf{v}(t_0) - \mathbf{v}^*(t_0)|$ говорят о непрогнозируемом изменении в функционировании спутника и съёмочной аппаратуры.

В-третьих, наличие предсказанного значения может позволить улучшить точность геометрической нормализации: для этого в программе нормализации вместо оценки поправки по изображению $\mathbf{v}(t_0)$ используется оптимальная комбинация с предсказанным значением $(1 - \alpha)\mathbf{v}(t_0) + \alpha \mathbf{v}^*(t_0)$, где α – некий весовой коэффициент. Дело в том, что сама поправка $\mathbf{v}(t_0)$ является зашумленной оценкой некоторого истинного значения $\mathbf{v}_{\text{ист}}(t_0)$. Если значения $\mathbf{v}_{\text{ист}}(t)$ коррелированы между сеансами, то для оптимального уменьшения ошибок в геометрии (например, в геодезической привязке) следует использовать ненулевой вес α .

Например, для оценки скоростей ветра по серии

изображений они должны иметь хорошую относительную геопривязку. Обычно это обеспечивается автоматически, поскольку ошибки в навигационной информации изменяются медленно. Однако без использования опорных ориентиров геопривязка оказывается не удовлетворительной. Если же для каждого сеанса вносить поправку к ориентации КА независимо по контурным точкам диска и электронным картам, то оказывается, что не смотря на проектную точность геопривязки каждого снимка, точность относительной привязки сеансов оказывается не достаточной для определения скоростей ветров. Эту проблему позволяет решить выбор $\alpha > 0$ даже при простейшем предсказателе вида $\mathbf{v}^*(t) = \mathbf{v}(t-1)$, где для простоты выбор масштаба t предполагается таким, что $t-1$ соответствует предыдущему сеансу.

В тех случаях, когда поправка $\mathbf{v}_{\text{ист}}(t_0)$ однозначно определяется по температурам (например, для сдвига линеек ПЗС1 и ПЗС2 в видимом диапазоне), то вместо зашумленной оценки поправки выгоднее использовать чистый прогноз, т.е. принять $\alpha = 1$.

На практике достаточное хорошее качество обеспечивает линейный предсказатель вида:

$$\mathbf{v}_i^*(t_0) = \sum_{j=1}^{N_i} A_{ij} \mathbf{v}_i(t_0 - j) + \sum_j B_{ij} s_j(t_0),$$

где N_i – количество используемых предыдущих поправок; A_{ij} , B_{ij} – некоторые коэффициенты; суммирование идёт по тем индексам, для которых переменные определены. Найти коэффициенты A_{ij} , B_{ij} можно по МНК из условия $\varepsilon = \min$. При этом каждое N_i выбирается минимально возможным, при котором наблюдается значимое уменьшение ошибки ε . За счет накопления больших объемов данных (за несколько дней и недель) определение коэффициентов можно выполнить с высокой надежностью. На практике обучаются два предсказателя: один для поиска аномальных ситуаций, а второй – для повышения геометрического качества при штатной обработке [5].

Следует отметить, что знание коэффициентов A_{ij} , B_{ij} позволяет также упростить выбор оптимального значения параметра α [6], хотя на практике хорошие результаты даёт даже эмпирический подбор.

Заключение

Предложенные в работе подходы по мониторингу качества целевой информации прошли апробация на космических аппаратах «Электро-Л» и «Арктика-М». Они позволяют эффективно проводить авторский надзор правильности функционирования наземных средств обработки спутниковых изображений, а так же постоянно удерживать на должном уровне и повышать качество выходных информационных продуктов.

Литература

1. Современные технологии обработки данных дистанционного зондирования Земли / под ред. В.В. Еремеева. М.: ФИЗМАТЛИТ, 2015. 460 с.
2. Егошкин Н.А., Еремеев В.В., Москвитин А.Э., Ушенкин В.А. Обработка информации от современных космических систем радиолокационного наблюдения Земли – Москва: Издательство ФИЗМАТЛИТ, 2019.
3. Егошкин Н.А., Еремеев В.В., Москвитин А.Э. Координатная привязка изображений от геостационарных спутников по контурным точкам диска Земли. Вестник РГРТУ. 2007. № 22. С. 10-16.
4. Воронин А.А., Егошкин Н.А., Еремеев В.В., Москвитин И.В. Геометрическая обработка данных космических систем глобального наблюдения Земли. Вестник РГРТУ. 2009. № 1 (выпуск 27). С. 12-17.
5. Егошкин Н.А., Москвитин А.Э. Повышение точности коррекции изображений на основе фильтрации измерений углового положения сканирующего зеркала. Вестник РГРТУ. 2010. № 3 (выпуск 33). С. 7-11.
6. Егошкин Н.А. Динамические модели геометрической обработки изображений в системах дистанционного зондирования Земли. Цифровая обработка сигналов. 2017. № 1. С. 3-7.

НОВЫЕ КНИГИ

Поборчая Н.Е.

Методы и алгоритмы оценивания параметров канала связи в условиях априорной неопределенности в системах с приемником прямого преобразования: Учебное издание для вузов, М.: Изд-во «Горячая линия-Телеком», 2023 г. 240 с.: ил.

Изложены методы и алгоритмы совместного оценивания параметров сигнала (канала связи) в условиях априорной неопределенности относительно статистических характеристик канала связи и законов распределения шумов. Особое внимание уделено системам с приемником прямого преобразования. Рассмотрены вопросы синтеза и анализа процедур оценивания для систем связи с одной передающей и приемной антенной (SISO), с несколькими передающими и приемными антеннами (MIMO), а также для систем с ортогональным частотным мультиплексированием (OFDM).

Предложенные алгоритмы способствуют повышению помехоустойчивости приема информации или понижению вычислительной сложности процедур обработки сигнала.

Для научных работников, инженеров и аспирантов. Может быть полезна студентам старших курсов и магистрантам, обучающимся по направлению подготовки «Радиотехника» и «Инфокоммуникационные технологии и системы связи».



АЛГОРИТМ И ПРОГРАММНЫЙ КОМПЛЕКС ВЫСОКОСКОРОСТНОЙ НЕЙРОСЕТЕВОЙ СЕГМЕНТАЦИИ ОБЛАЧНОСТИ НА ПАНХРОМАТИЧЕСКИХ ИЗОБРАЖЕНИЯХ ОТ КОСМИЧЕСКИХ АППАРАТОВ «РЕСУРС-П»

Москвитин А.Э., в.н.с. НИИ «Фотон» РГРТУ, д.т.н., foton@rsreu.ru

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

Ларюков С.А., м.н.с. НИИ «Фотон» РГРТУ, foton@rsreu.ru

ALGORITHM AND SOFTWARE OF HIGH-SPEED NEURAL NETWORK CLOUD SEGMENTATION OF RESURS-P SATELLITE'S PANCHROMATIC IMAGES

Moskvitin A.E., Ushenkin V.A., Laryukov S.A.

The paper considers the problem of cloud segmentation on panchromatic images obtained by the Geoton highly detailed sensor of the Resurs-P Earth remote sensing satellite. The architecture of an artificial neural network, which is characterized by lower computational costs and provides the same segmentation quality as more complex known architectures, is proposed. The question of choosing the optimal level of preliminary radiometric processing of segmented images is considered. The results of developed software experimental studies and the achieved values of segmentation quality measures are presented.

Key words: panchromatic image, cloud segmentation, remote sensing, artificial neural network.

Ключевые слова: панхроматическое изображение, сегментация облачности, дистанционное зондирование Земли, искусственная нейронная сеть.

Введение

Наличие облачности является фактором, ограничивающим применение данных дистанционного зондирования Земли (ДЗЗ) видимого и инфракрасного диапазонов. Сплошная облачность загораживает от космического аппарата объекты земной поверхности, не давая их наблюдать, а полупрозрачная дымка искажает яркость и контрастность наблюдаемых объектов, существенно затрудняя их дешифрацию и измерение физических свойств. При стереофотограмметрической обработке [1, 2] пар космических снимков наличие облачности приводит к тому, что измеряется высота не рельефа земной поверхности, а высота облака. В связи с этим при обработке космических снимков видимого и инфракрасного диапазонов [3] необходимы выявление и отбраковка участков изображения, на которых представлены не наземные объекты, а облачность.

Ручное выделение облачности на изображениях является крайне трудоемкой задачей и существенно ограничивает оперативность получения обработанных космических снимков. Поэтому актуально развитие автоматических методов сегментации облачности.

Для обнаружения облачности на изображениях могут использоваться различные признаки: яркостной (облака обычно имеют высокую яркость), колориметрический (облака обычно имеют серый или белый цвет), текстурные (облака обычно имеют гладкую текстуру и плавные границы), стереофотограмметрический (при наличии нескольких снимков, снятых со стереобазой, облака из-за возвышения над поверхностью Земли имеют большие геометрические рассогласования на таких снимках, чем наземные объекты) [4, 5]. При этом наиболее сложной выглядит задача сегментации об-

Рассматривается задача сегментации облачности на панхроматических изображениях, полученных высокодетальной съемочной аппаратурой «Геотон» космических аппаратов дистанционного зондирования Земли «Ресурс-П». Предлагается архитектура искусственной нейронной сети, позволяющая с меньшими вычислительными затратами достичь качества сегментации, не уступающего более сложным известным архитектурам. Рассматривается вопрос выбора оптимального уровня предварительной радиометрической обработки сегментируемых изображений. Приводятся результаты экспериментальных исследований разработанного программного комплекса и достигнутые значения показателей качества сегментации.

лачности на панхроматических изображениях, для которых доступны только два вида признаков: яркостной и текстурные. Задача дополнительно усложняется, если панхроматические изображения сформированы отечественной аппаратурой «Геотон», устанавливаемой на космические аппараты серии «Ресурс-П». Данная аппаратура имеет ограниченный динамический диапазон. Перед съемкой участка земной поверхности предсказываются его энергетические характеристики и выбирается соответствующий коэффициент усиления. При этом все объекты, энергетическая яркость которых превышает некоторый порог насыщения, оказываются засвеченными. За счет того, что облака, как правило, намного ярче наземных объектов, в большинстве случаев на панхроматических снимках аппаратуры «Геотон» они оказываются засвеченными, а для их автоматической сегментации не остается практически никаких признаков. Текстура облачности теряется, а яркость искажается. Лишь по наличию тени, характерной форме, площади и более темным граничным участкам, не подвергшимся искажению, с той или иной долей вероятности человек может распознать на снимке облако и отличить его от других засвеченных ярких объектов.

Для автоматического решения такой сложной задачи детерминированные алгоритмы оказываются мало пригодны. Однако развитие за последние годы методов искусственного интеллекта и в частности искусственных нейронных сетей (ИНС) позволяет получить достаточно качественное решение даже таких сложных задач.

Имеется ряд работ по применению ИНС в задаче сегментации облачности на мультиспектральных изображениях [6-9]. В работе [10] ИНС успешно применены для сегментации облачности на панхроматических изображениях. Однако известные архитектуры ИНС характеризуются большим числом весовых коэффициентов и относительно большим объемом вычислений, что существенно затрудняет их обучение и ограничивает скорость нейросетевой обработки.

В настоящей работе ставится задача разработки архитектуры ИНС, характеризующейся меньшим количеством весовых коэффициентов и меньшим объемом вычислений, при сохранении качества сегментации облачности.

Выбор показателей качества сегментации

В настоящее время для оценки качества бинарной сегментации используется множество различных показателей. Рассмотрим основные из них.

Все показатели предполагают, что имеется результат сегментации в виде бинарного растрового изображения, где один код пикселя соответствует искомому объекту, а другой – окружающему фону. В задаче сегментации облачности искомым объектом являются сплошная облачность и полупрозрачная дымка, а окружающим фоном – наземные объекты. Также должен иметься эталонный результат, полученный, например, в результате ручной сегментации.

Первый показатель – общая точность сегментации a , определяемая как отношение числа пикселей, коды которых совпадают на эталонном и полученном результатах сегментации, к общему числу сегментируемых пикселей.

До единицы показатель a дополняют ошибки первого и второго рода. Под ошибкой первого рода понимается ложное выделение искомого объекта (на результате сегментации представлен объект, а на эталоне – фон). Под ошибкой второго рода понимается пропуск объекта (на результате сегментации представлен фон, а на эталоне – объект).

Абсолютное значение числа ошибок первого и второго рода в пикселях изображения обычно не используется. Оно заменяется относительными значениями, нормированными общей площадью окружающего фона или искомым объектом. Таким образом, вероятность возникновения ошибки первого рода α – это отношение числа пикселей изображения, ложно распознанных как объект, к числу пикселей, соответствующих окружающему фону на эталоне. Вероятность возникновения ошибки второго рода β – это отношение числа пикселей изображения, ошибочно отнесенных к окружающему фону, к числу пикселей, соответствующих искомому объекту на эталоне.

Пусть эталонная доля пикселей изображения, принадлежащих искомому объекту, составляет δ , тогда общая точность сегментации может быть выражена через ошибки первого и второго рода как

$$a = 1 - (1 - \delta)\alpha - \delta\beta. \quad (1)$$

Недостатком показателя a является тот факт, что при нулевых значениях δ он вырождается в $a = 1 - \alpha$, т.е., когда искомым объектов крайне мало, оказывается неважным, пропускаются они или нет; главное, чтобы не было ложных выделений. В обратном случае при $\delta = 1$ показатель вырождается в $a = 1 - \beta$, т.е. учитываются только пропуски искомого объекта, а ложные выделения игнорируются.

В ряде задач сегментации, особенно в области медицины, приведенное поведение показателя a является недопустимым, поэтому для таких случаев были разработаны другие показатели.

Мера Жаккара J – отношение числа пикселей, верно отнесенных к объекту, к сумме эталонного числа пикселей объекта и числа пикселей, ошибочно отнесенных к объекту. Мера Жаккара при $\delta \rightarrow 0$ ведет себя еще менее приемлемым образом. Если выразить меру Жаккара через α , β и δ , получится:

$$J = \frac{\delta(1 - \beta)}{\delta + (1 - \delta)\alpha}. \quad (2)$$

При $\delta \rightarrow 0$ мера Жаккара вырождается в $J = \begin{cases} 0, & \forall \alpha > 0, \\ \text{неопределенность}, & \alpha = 0. \end{cases}$ Таким образом, на значение показателя перестают влиять не только ошибки второго рода, но и положительные значения ошибок первого рода. В отсутствие ошибок первого рода и в отсутствие искомым объектов мера Жаккара вырождается в полную неопределенность, поскольку и числитель, и знаменатель в (2) обращаются в ноль. В обратном случае (т.е. при $\delta \rightarrow 1$) $J = 1 - \beta$, т.е. ведет себя аналогично показателю a .

В качестве альтернативы a , α и β введены показатели точности обнаружения объекта p , полноты обнаружения r и F-меры.

Точность обнаружения объекта схожа с мерой Жаккара и определяется как отношение числа пикселей, верно отнесенных к искомому объекту, к общему числу пикселей, отнесенных на результате сегментации к искомому объекту:

$$p = \frac{\delta(1 - \beta)}{\delta(1 - \beta) + (1 - \delta)\alpha}. \quad (3)$$

Поведение меры p при $\delta \rightarrow 0$ точно такое же, как у меры Жаккара. При $\delta \rightarrow 1$ мера вырождается в единицу независимо от значений α и β .

Полнота обнаружения объекта определяется как отношение числа пикселей, верно отнесенных к искомому объекту, к эталонному числу пикселей объекта, т.е. фактически дополняет β до единицы:

$$r = 1 - \beta. \quad (4)$$

F-мера является средним гармоническим мер p и r :

$$F = \frac{2}{p^{-1} + r^{-1}}. \quad (5)$$

Также известна φ -мера (или коэффициент корреляции Мэттью), определяемая как:

$$\varphi = \frac{\delta(1-\beta)(1-\delta)(1-\alpha) - (1-\delta)\alpha\delta\beta}{\sqrt{\delta(1-\delta)(\delta(1-\beta) + (1-\delta)\alpha)((1-\delta)(1-\alpha) + \delta\beta)}}. \quad (6)$$

При $\delta \rightarrow 0$ и $\delta \rightarrow 1$ φ -мера вырождается в неопределенность (числитель и знаменатель в (6) обнуляются).

Таким образом, из приведенного анализа следует, что, несмотря на свои недостатки, в экстремальных случаях, когда $\delta \rightarrow 0$ или $\delta \rightarrow 1$, наименее неадекватно ведет себя мера общей точности сегментации a .

Исправить ее недостатки на несбалансированных данных, когда $\delta \ll 0,5$ или $\delta \gg 0,5$, можно путем введения поправочных коэффициентов k_α и k_β :

$$\hat{a} = 1 - k_\alpha(1-\delta)\alpha - k_\beta\delta\beta. \quad (7)$$

В экстремальных ситуациях, когда $\delta = 0$ или $\delta = 1$, показатель \hat{a} будет также игнорировать одну из величин α или β , которая в этот момент является неопределенной, но по мере приближения к экстремальным ситуациям за счет коэффициентов k_α и k_β влияние величин α и β на \hat{a} можно сделать более сбалансированным.

Применительно к задаче сегментации облачности введение коэффициентов k_α и k_β видится нецелесообразным. Ситуация пропуска облачности так же плоха, как ситуация ложного ее обнаружения. В первом случае облачность не отбракуется и негативно отразится на последующих видах обработки изображения, а во втором случае будет ошибочно проигнорирована полезная информация. Показатель a оценивает суммарное количество ситуаций обоих видов.

Переход от показателя качества сегментации к дифференцируемой функции потерь при машинном обучении

Наиболее распространенной функцией потерь при машинном обучении в задаче бинарной сегментации является бинарная перекрестная энтропия:

$$H = -\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (e_{ij} \log_2 m_{ij} + (1-e_{ij}) \log_2 (1-m_{ij})), \quad (8)$$

где $(e_{ij})_{M \times N}$ – матричное представление эталонного результата сегментации размером $M \times N$ пикселей, где $e_{ij} = 1$ соответствует наличию искомого объекта, а $e_{ij} = 0$ – отсутствию; $(m_{ij})_{M \times N}$ – матричное представление полученного результата сегментации.

Основным достоинством такой функции потерь является ее дифференцируемость, что позволяет обучать ИНС методами на основе градиентного спуска. Однако при $m_{ij} = 0$ или $m_{ij} = 1$ из-за операции логарифмирования бинарная перекрестная энтропия устремляется в бесконечность. Поэтому растровые маски искомого

объекта m_{ij} , получаемые в результате машинного обучения с данной функцией потерь, содержат лишь промежуточные значения $0 < m_{ij} < 1$, не позволяющие принять однозначное решение, соответствует данный пиксель искомому объекту или нет. Требуется дополнительный поиск некоторого порога бинаризации маски $(m_{ij})_{M \times N}$. Кроме того, поскольку для оценки качества сегментации используются совершенно другие показатели, рассмотренные ранее, целесообразно при машинном обучении применять функцию потерь, максимально близкую к этим показателям.

Напрямую использовать показатель a в качестве функции потерь невозможно, поскольку он не является дифференцируемым. При его расчете подсчитывается количество пикселей, которое является дискретной недифференцируемой величиной. Однако можно рассчитать a как частный случай дифференцируемого выражения. Покажем это.

Пусть результат сегментации $(m_{ij})_{M \times N}$ бинаризован, т.е. m_{ij} может быть только нулем или единицей. Эталонный результат сегментации $(e_{ij})_{M \times N}$ предполагается бинаризованным изначально. В этом случае число верно выявленных пикселей объекта можно записать как $\sum_{i=1}^M \sum_{j=1}^N e_{ij} m_{ij}$, а число верно сегментированных пикселей фона как $\sum_{i=1}^M \sum_{j=1}^N (1-e_{ij})(1-m_{ij})$. Таким образом, если $(m_{ij})_{M \times N}$ бинаризована, общая точность сегментации может быть выражена как

$$\tilde{a} = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (e_{ij} m_{ij} + (1-e_{ij})(1-m_{ij})). \quad (9)$$

Выражение (9) является дифференцируемым и его дополнение до единицы может использоваться в качестве функции потерь $1 - \tilde{a}$. Если $(m_{ij})_{M \times N}$ содержит только нули и единицы, \tilde{a} строго соответствует общей точности сегментации. Если же $(m_{ij})_{M \times N}$ содержит промежуточные значения, \tilde{a} является некоторой приближенной оценкой показателя a .

Проведенные экспериментальные исследования показали, что на практике в ходе машинного обучения ИНС ей оказывается выгодно сформировать маску $(m_{ij})_{M \times N}$, на которой практически не содержится промежуточных значений m_{ij} , поэтому функция \tilde{a} с высокой точностью соответствует целевому показателю качества сегментации a . При этом отпадает необходимость выбора порога для бинаризации $(m_{ij})_{M \times N}$, поскольку маска на выходе ИНС уже практически бинаризована. Для тех редких пикселей, в которых $0 \ll m_{ij} \ll 1$, может быть выбран произвольный порог бинаризации, например, равный 0,5, поскольку на итоговой точности сегментации он почти не отражается.

Такое поведение ИНС может быть достигнуто следующим образом. Функцией активации выходного слоя

ИНС, используемых для бинарной сегментации, как правило, является сигмоида. Для того чтобы сигмоида была близка к нулю, ей на вход нужно подать большое по модулю отрицательное значение. Чтобы сигмоида была близка к единице, ей на вход должно поступить большое положительное значение. Последний слой ИНС обычно выполняет некоторое линейное преобразование яркости пикселей изображения. Чтобы обеспечить такой выход сигмоиды, он должен за счет больших мультипликативных коэффициентов значительно усилить сигнал. Затем к усиленному сигналу прибавляется некоторый аддитивный коэффициент, в результате чего сигнал оказывается положительным или отрицательным. По сути этот аддитивный коэффициент последнего слоя ИНС и является порогом, по которому бинаризуется результат сегментации. Он вместе с другими весовыми коэффициентами ИНС автоматически подбирается в ходе машинного обучения.

Другим достоинством функции потерь $1 - \tilde{a}$ является ее линейность. Машинное обучение ИНС обычно выполняется в пакетном режиме, когда вся обучающая выборка делится на пакеты из небольшого числа изображений. Одна итерация метода градиентного спуска и коррекции весов ИНС путем обратного распространения ошибки выполняется в пределах пакета. Соответственно функция потерь рассчитывается только по небольшому числу изображений, представленных в пакете. Перебор всех пакетов обучающей выборки составляет одну эпоху обучения. По окончании эпохи рассчитываются потери на всей выборке путем усреднения значений, полученных для каждого пакета. Результаты усреднения используются для проверки, улучшила ли свою работу ИНС в ходе эпохи обучения. Если функция потерь является линейной, то усреднение ее значений, рассчитанных для отдельных пакетов, дает строгое значение функции потерь для всей обучающей выборки. Если функция потерь является нелинейной, то для получения истинного ее значения для всей обучающей выборки необходимы более сложные алгоритмы комбинирования значений потерь, рассчитанных по отдельным пакетам. Существующие широко используемые средства машинного обучения такие сложные алгоритмы обычно не поддерживают, а усреднение потерь, рассчитанных по отдельным пакетам, искажает нелинейные целевые функции, оптимизируемые при обучении.

Для сверточных ИНС, которые по сути эквивалентны некоторому нелинейному фильтру конечной апертуры, при расчете потерь нужно учитывать наличие краевых эффектов фильтрации. Для этого либо размеры выхода ИНС должны быть уменьшены по сравнению с входом на апертуру фильтра, либо расчет потерь должен вестись лишь по центральной части выходного изображения. Если вовлекать области краевых эффектов в расчет потерь, ИНС при обучении будет вынуждена уменьшать размер окрестности, которую она анализирует, обнуляя часть своих коэффициентов, чтобы ослабить краевые эффекты. В этом случае архитектура ИНС будет использоваться не полностью, что может негативно отразиться на качестве результата.

Выбор архитектуры ИНС

Наиболее распространенной архитектурой ИНС, применяемой в задачах бинарной сегментации, является «U-Net» [11]. Это сверточная ИНС, основанная на пирамидальной обработке. Каждый следующий уровень пирамиды формируется путем выполнения двух сверток в окне 3×3 с активационной функцией типа «линейный выпрямитель» (англ. rectified linear unit, ReLU) и увеличения шага дискретизации в 2 раза с выбором максимального значения в окне 2×2 . На каждом следующем уровне пирамиды число каналов изображения увеличивается в 2 раза. Пирамида содержит 6 уровней. После достижения верхнего уровня начинается обратная операция. Шаг дискретизации изображения уменьшается в 2 раза с выполнением интерполяции и сокращением в 2 раза числа каналов. К полученному изображению в качестве дополнительных каналов добавляется предыдущий слой пирамиды, затем выполняются две свертки в окне 3×3 с активационной функцией типа «линейный выпрямитель» и снова сокращается в 2 раза число каналов. По достижении нижнего слоя пирамиды выполняется линейное преобразование яркости изображения с сигмоидой в качестве активационной функции и сокращением числа каналов до одного. Графическое представление архитектуры ИНС (рис. 1) напоминает букву «U», что определило название архитектуры. Число каналов, формируемых на нижнем слое пирамиды, в классической архитектуре «U-Net» составляет 64.

Авторы архитектуры «U-Net» применили ее для сегментации медицинских изображений. В их задаче положительного результата удалось достичь на небольшой обучающей выборке.

В задаче сегментации облачности на панхроматических изображениях обучающую выборку приходится делать больших размеров. Это связано с большим разнообразием объектов, которые могут быть представлены на космических снимках. Обучающая выборка должна включать все основные виды объектов, чтобы все из них ИНС научилась отделять от облачности. Кроме того, сама облачность отличается крайне широким разнообразием по форме, структуре, яркости, прозрачности. Одни облака могут находиться в тени других облаков и быть хуже освещенными. Полупрозрачная дымка может располагаться над любыми другими видами объектов.

Машинное обучение сложных многослойных архитектур ИНС на больших обучающих выборках крайне затруднительно, поскольку требует огромного объема вычислений. При большом числе слоев из-за накапливающихся погрешностей алгоритма обратного распространения ошибки сходимость машинного обучения крайне низкая. Требуются десятки и сотни тысяч эпох обучения для минимизации целевой функции, а если обучающая выборка большая, то одна эпоха может длиться несколько часов даже на современных графических ускорителях вычислений.

В работе [12] был предложен механизм существенно ускорения обучения многослойных ИНС – пакетная нормализация. В рамках этого механизма в ИНС после каждого сверточного слоя (кроме выходного) добавля-

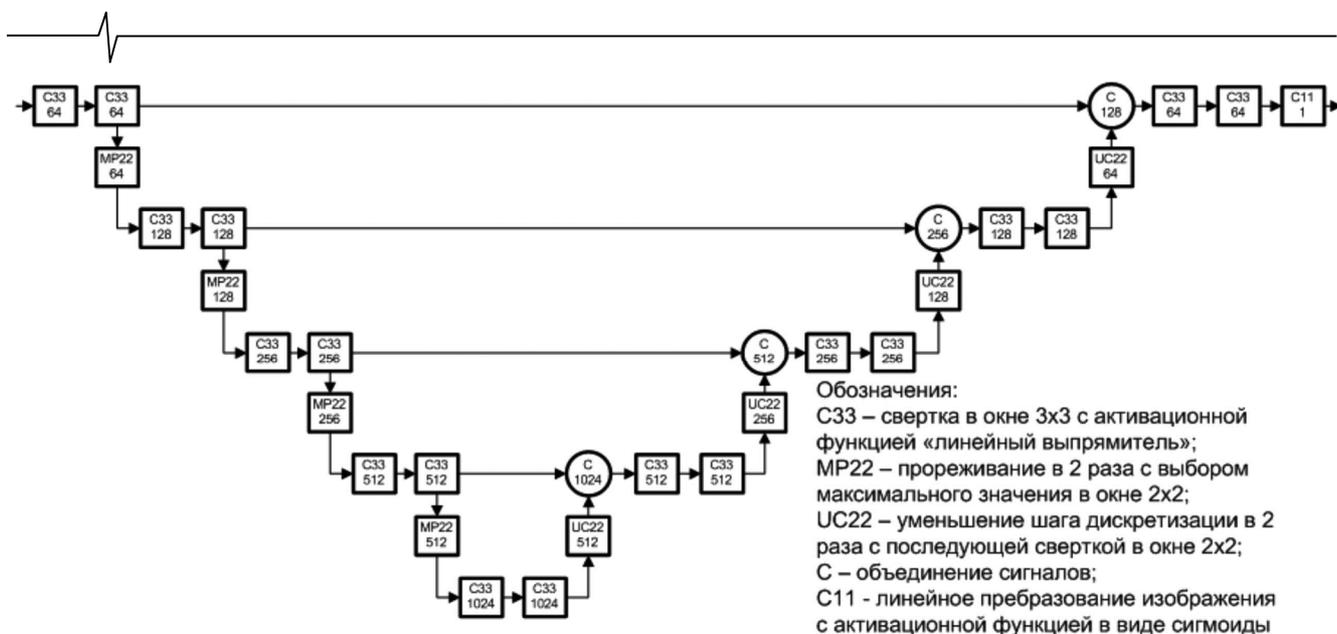


Рис. 1. Классическая архитектура ИНС «U-Net»

ются слои нормализации, которые в ходе обучения в пределах пакета рассчитывают среднее значение и дисперсию сигнала, а затем выполняют линейное преобразование сигнала, обнуляя его среднее и делая дисперсию единичной. Коэффициенты линейного преобразования оказываются уникальными для каждого пакета, однако в ходе обучения запоминаются усредненные значения этих коэффициентов, которые используются при штатной работе обученной ИНС, когда уже нет никаких пакетов, а обрабатывается то или иное отдельное изображение.

Пакетная нормализация позволяет сократить машинное обучение ИНС до нескольких десятков или сотен эпох, однако потери, рассчитываемые в ходе обучения и получаемые в ходе последующего штатного функционирования, разнятся из-за различной работы слоев пакетной нормализации в этих режимах. Поэтому для корректного обучения необходимо в пределах каждой эпохи рассчитывать функцию потерь дважды: один раз в ходе итераций градиентного спуска в режиме обучения и второй раз в конце эпохи в режиме штатной работы для оценки истинного значения потерь.

Другой проблемой машинного обучения ИНС является так называемое переобучение. С точки зрения математики минимизация целевой функции в пределах обучающей выборки не гарантирует, что целевая функция будет принимать малые значения за пределами этой выборки. Поэтому иногда случается, что обученная ИНС хорошо работает в пределах обучающей выборки, но выдает совершенно неприемлемые результаты на других изображениях. Особенно часто это происходит, когда выборка непредставительна (т.е. не охватывает все типовые варианты, которые в дальнейшем могут быть поданы на вход ИНС) или мала по объему, когда ИНС в своих коэффициентах может просто запомнить таблицу преобразования небольшого количества комбинаций на входе в требуемый выход.

Чтобы избежать переобучения, обычно выделяют две выборки: обучающую и валидационную. Обучающая выборка делится на пакеты, по которым выполняются итерации градиентного спуска, а по окончании

эпохи выполняется расчет потерь по валидационной выборке. Т.е. для коррекции весов ИНС используются значения потерь, полученные по обучающей выборке, в т.ч. с применением пакетной нормализации, а минимизируется целевая функция, рассчитанная по валидационной выборке в режиме штатной работы ИНС.

Чтобы пакетная нормализация работала корректно, необходимо выбрать достаточно большой размер пакета и более менее равномерно размещать обучающую выборку, чтобы в каждом пакете на изображениях было представлено как можно больше различных видов объектов. Тогда статистические характеристики сигналов на выходе сверточных слоев в рамках каждого пакета будут достаточно стабильными, и их усредненные значения, которые используются при штатной работе ИНС, будут близкими к ним. В этом случае ИНС обучается быстро и способна достичь малых значений функции потерь как на обучающей, так и на валидационной выборках. Если же размер пакета недостаточно большой, потери на обучающей выборке снижаются достаточно быстро, но в режиме штатной работы на валидационной выборке потери остаются стабильно высокими. Однако размер пакета, который можно выбрать, ограничен объемом памяти графического ускорителя. Нормализация требует одновременного доступа к данным всего пакета изображений для расчета их статистических характеристик. Число слоев нормализации в многослойных сетях измеряется десятками. Если все данные не будут размещены в быстросействующей памяти, время обработки одного пакета катастрофически увеличивается, что лишает смысла пакетную нормализацию как способ ускорения машинного обучения.

Классическая архитектура «U-Net» была разработана без учета применения в сочетании с пакетной нормализацией. В то же время в задаче сегментации облачности без применения пакетной нормализации обучить ИНС за приемлемое время не представляется возможным. Поэтому в архитектуру «U-Net» были добавлены слои нормализации. Однако из-за большого числа слоев и большого числа каналов на выходе промежуточных слоев ИНС «U-Net» в 11 Гбайт памяти графического ускорителя

NVIDIA GeForce RTX 2080 Ti удается разместить пакет только из 3 изображений 512×512 пикселей. Это слишком малый размер пакета, чтобы в нем удалось представить все возможные виды объектов. В результате значение функции потерь $1-\tilde{a}$ на валидационной выборке не опускается ниже 0,0584.

В связи с этим было принято решение в 2 раза сократить число каналов во всех промежуточных слоях ИНС «U-Net», за счет чего становится возможным увеличение размера пакета до 6 изображений. В этом случае удалось достичь значения функции потерь $1-\tilde{a}$ на валидационной выборке, равного 0,0228.

Однако при этом времена обучения ИНС и нейросетевой сегментации остались достаточно высокими. Поэтому были проведены эксперименты по дальнейшему упрощению архитектуры «U-Net». Основным достоинством «U-Net» является пирамидальная организация обработки, позволяющая относительно быстро анализировать широкую окрестность пикселя (радиусом порядка 100 пикселей). Было решено развить идею пирамидальной обработки, увеличив на единицу высоту пирамиды и сократив в 2 раза число сверточных слоев на каждом уровне пирамиды. Размер анализируемой окрестности при этом остается таким же, но существенно сокращается число слоев ИНС. Число каналов на каждом уровне пирамиды дополнительно сокращено в 2 раза. На нижнем уровне пирамиды, таким образом, формируется 16 каналов. При этом за счет большей высоты на верхнем уровне пирамиды число каналов оказывается лишь в 2 раза меньше, чем в классической архитектуре «U-Net». Для обработки одного изображения 512×512 пикселей с помощью модифицированной архитектуры требуется порядка 10 млрд сложений и умножений, в то время как для классической архитектуры «U-Net» – порядка

218 млрд, а для архитектуры «U-Net» с в 2 раза уменьшенным числом каналов в промежуточных слоях – порядка 55 млрд.

Модифицированная архитектура ИНС представлена на рис. 2. После модификации она напоминает более высокую и более узкую букву «U», поэтому назовем модифицированную архитектуру «Lanky U-Net». Размер пакета для «Lanky U-Net» удалось увеличить до 13 изображений. По результатам машинного обучения «Lanky U-Net» было достигнуто значение функции потерь $1-\tilde{a}$ на валидационной выборке, равное 0,0219. Таким образом, «Lanky U-Net» в задаче сегментации облачности на панхроматических изображениях аппаратуры «Геотон» космических аппаратов «Ресурс-П» обеспечивает чуть более высокое качество сегментации, чем архитектура «U-Net» с уменьшенным в 2 раза числом каналов, при в 5,5 раз меньшем объеме вычислений. Классическая архитектура «U-Net» потенциально способна обучиться до сопоставимых результатов, но лишь при развитии графических ускорителей вычислений, когда объем их памяти позволит задать приемлемый размер пакета при обучении «U-Net».

Дополнительным недостатком пакетной нормализации является небольшое увеличение объема вычислений при штатной работе ИНС из-за затрат на линейное преобразование яркостей изображения после каждого сверточного слоя. В то же время, если разместить слой нормализации не после активационной функции сверточного слоя, а между сверткой и активацией, то этих затрат при штатной работе можно избежать путем применения линейного преобразования не к свернутому изображению, а к весовым коэффициентам свертки. Если функция активации существенно нелинейная, такая перестановка положения слоев нормализации не имеет

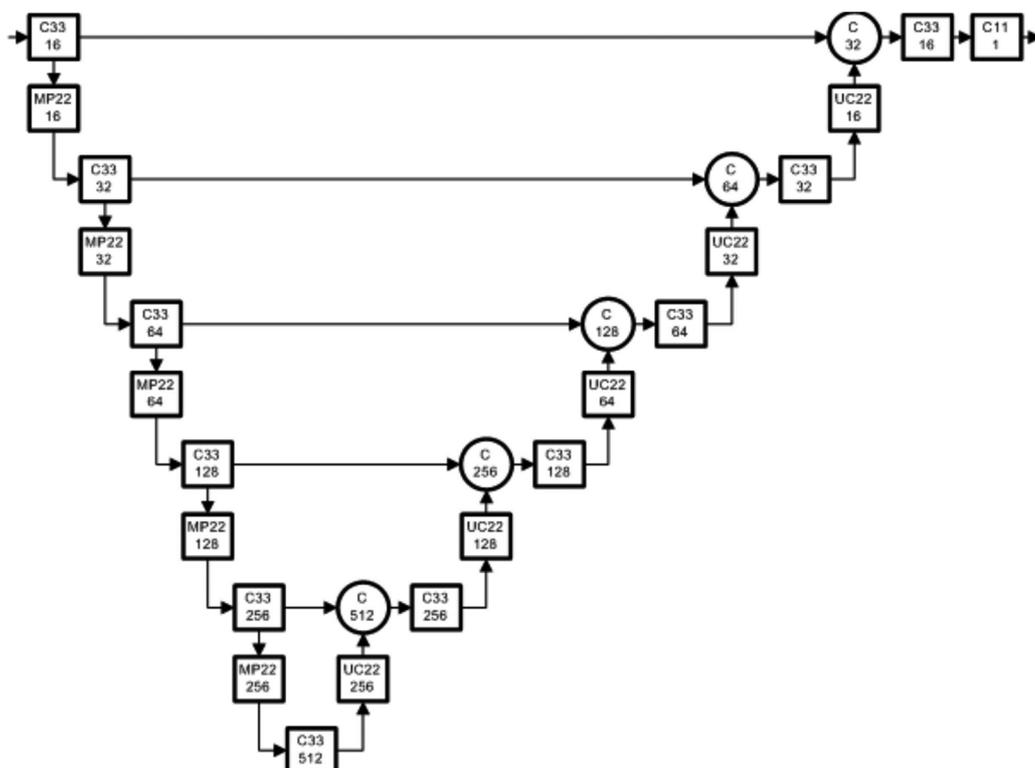


Рис. 2. Предлагаемая архитектура ИНС «Lanky U-Net»



смысла. Однако для активационной функции «линейный выпрямитель», являющейся кусочно-линейной, такая перестановка может быть допустима. Архитектура «Lanky U-Net» была обучена с измененным положением слоев пакетной нормализации. Сходимость процесса обучения от этого незначительно ухудшилась. Достигнутое значение функции потерь $1-\tilde{a}$ на валидационной выборке составило 0,0224. Таким образом, за счет перестановки слоев пакетную нормализацию можно применять при обучении ИНС без снижения скорости штатной работы, достигая при этом практически такого же качества сегментации.

Выбор уровня радиометрической обработки входных данных ИНС

В работе [10] для сегментации облачности предлагалось выполнять радиометрическую нормализацию панхроматических изображений. Космические снимки могут быть сняты с различным коэффициентом усиления и в разных условиях освещенности. Для устранения первого фактора коды яркости изображения необходимо пересчитать в единицы энергетической яркости, используя коэффициенты градуировочной характеристики съемочного устройства. Для устранения второго фактора можно предположить, что поверхность наблюдаемых объектов является горизонтально расположенной ламбертовой поверхностью, и разделить значение полученной энергетической яркости на косинус зенитного угла Солнца в момент съемки, пренебрегая поглощением и рассеянием света в атмосфере. Именно преобразованные таким образом изображения использовались для выбора и обучения архитектуры ИНС в настоящей работе.

Однако выполненная таким образом радиометрическая нормализация является приближенной.

Во-первых, для облачности, в большинстве случаев засвеченной, пересчет кодов яркости в единицы энергетической яркости является некорректным. Значения энергетической яркости получаются существенно заниженными, и облачность потенциально может быть спутана с менее яркими объектами. И если в исходных кодах яркости по абсолютному значению кода можно понять, что это область насыщения, то после преобразования получаема энергетическая яркость может как принадлежать области насыщения при большом коэффициенте усиления, так и не принадлежать ей при малом коэффициенте.

Во-вторых, облачность не является горизонтальной ламбертовой поверхностью. Облака – это взвешенные в воздухе скопления капель воды. Рассеяние излучения этими каплями происходит не на плоской границе облака, а в некоторой его толще. Поэтому деление на косинус зенитного угла Солнца для облачности не применимо.

Таким образом, радиометрическая нормализация позволяет существенно снизить вариации яркости наземных объектов, но при этом существенно искажает яркость искомой облачности. Поэтому заранее предсказать, повышает или понижает качество сегментации облачности такая радиометрическая обработка исходного изображения, не представляется возможным. Сле-

довательно, необходимо провести соответствующие экспериментальные исследования, в которых ИНС обучаются сегментации облачности на полностью нормализованных изображениях, изображениях, пересчитанных в единицы энергетической яркости, и исходных ненормализованных изображениях.

Экспериментальные исследования показали, что ИНС «Lanky U-Net» со слоями нормализации, расположенными между сверткой и активационной функцией, обучается до следующих значений функции потерь на валидационной выборке:

- 0,0224 для полностью нормализованных изображений;
- 0,0226 для изображений, пересчитанных в единицы энергетической яркости;
- 0,0199 для исходных изображений.

Из полученных оценок можно сделать вывод, что любая нормализация в среднем снижает качество сегментации облачности. Визуальный анализ результатов сегментации показал, что на полностью или частично нормализованных изображениях плохо выделяются засвеченные облака, но достаточно неплохо отделяются от облачности другие яркие объекты: пески, снега, крыши зданий. При сегментации по исходным изображениям засвеченные облака почти не пропускаются, зато существенно учащаются случаи ложного распознавания в качестве облачности других ярких объектов. В связи с этим возникла идея подать на вход ИНС многоканальное изображение, в каналах которого содержится поразному нормализованный исходный снимок, и поручить ИНС самой определить, по какому из этих каналов ей лучше выделить облачность.

Были рассмотрены следующие варианты входных данных ИНС:

- исходное и полностью нормализованное изображение (2 канала);
- исходное и частично нормализованное (до энергетических яркостей) изображение (2 канала);
- исходное, частично нормализованное и полностью нормализованное изображения (3 канала).

В результате обучения ИНС были достигнуты следующие значения функции потерь на валидационной выборке: 0,0189 для первого варианта, 0,0181 для второго варианта и 0,0170 для третьего варианта. Видно, что добавление только полностью нормализованных изображений относительно слабо помогает лучше выделить облачность, а более существенный положительный эффект наблюдается от добавления частично нормализованных изображений. Максимальный же эффект дает добавление полностью нормализованных изображений в сочетании с частично нормализованными изображениями.

Таким образом, на вход нейронной сети наиболее целесообразно подавать в трех каналах:

- исходное изображение в кодах яркости, по которому ИНС может обнаружить засвеченные объекты;
- частично нормализованное изображение в единицах энергетической яркости, по которому ИНС наилучшим образом может обнаружить незасвеченную облачность;

– полностью нормализованное изображение, на котором минимальны возможные вариации яркости наземных объектов, что позволяет лучше отделить их от облачности.

Экспериментальные исследования

Обучающая и валидационная выборки составлены из фрагментов размером 512×512 пикселей 32 обзорных панхроматических снимков аппаратуры «Геотон» космических аппаратов «Ресурс-П» № 1-3. Снимки получены с различными коэффициентами усиления при различных зенитных углах Солнца и включают все основные типы наблюдаемых объектов: городскую застройку, водоемы, поля, леса, горы со снежными шапками, пески, засвеченную и незасвеченную сплошную облачность, а также полупрозрачную дымку. Шаг дискретизации обзорного изображения увеличен в 8 раз по сравнению с детальным и составляет около 6,5 м на земной поверхности. Это позволяет охватывать нейронной сетью в 64 раза большую площадь земной поверхности при малом объеме вычислений. За счет большого размера облаков и относительно гладкой их текстуры вся основная информация для отделения об-

лачности от других объектов на обзорном изображении сохраняется. Фрагментам, составляющим обучающую и валидационную выборки, соответствует примерно половина суммарной площади снимков, при этом объем обучающей выборки в 9 раз больше (13837 фрагментов), чем валидационной (1537 фрагментов).

Все результаты обучения, приведенные ранее, получены на описанных выборках. Сами исходные снимки составляют тестовую выборку, для которой вручную получены эталонные маски облачности. Для выполнения исследований разработан программный комплекс, реализующий рассмотренную нейросетевую сегментацию. При тестах снимки разбиваются на фрагменты с перекрытием 256 пикселей. Перекрытие выбрано так, чтобы оно было больше или равно диаметру окрестности пикселя, анализируемой ИНС, а также кратно 32. Кратность 32 обусловлена тем, что слои прореживания с выбором максимального значения инвариантны к сдвигу сигнала только на целое число своих апертур. Таким образом гарантируется отсутствие блочных артефактов. На краях тестовые изображения зеркально дополняются на 128 пикселей.

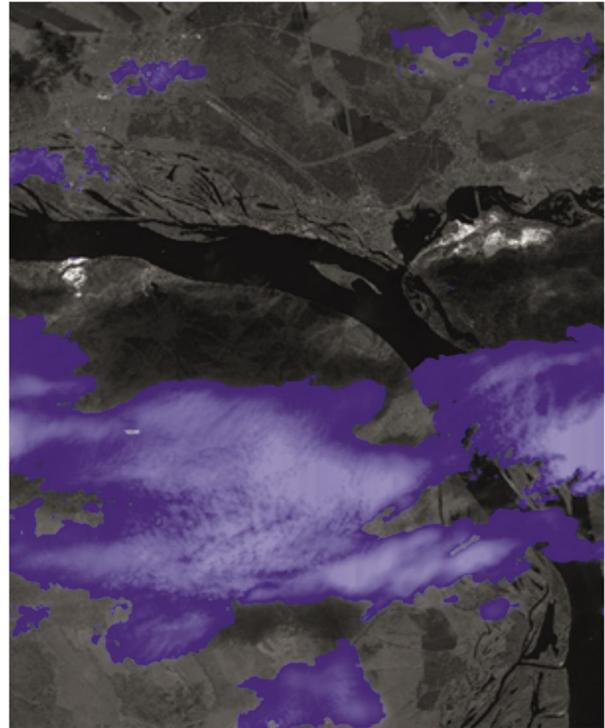
В таблице приведены значения показателей δ , α и

Показатели качества сегментации облачности на тестовых изображениях

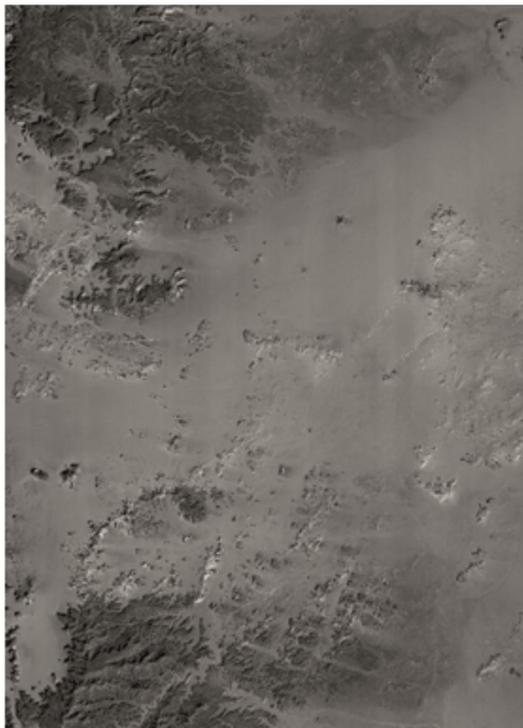
| № п/п | Ширина изображения, пикс. | Высота изображения, пикс. | δ | α | β | a |
|-------|---------------------------|---------------------------|----------|----------|---------|-------|
| 1 | 5625 | 11298 | 0,024 | 0,001 | 0,304 | 0,992 |
| 2 | 4500 | 11406 | 0,125 | 0,014 | 0,017 | 0,985 |
| 3 | 4500 | 11172 | 0,735 | 0,093 | 0,012 | 0,966 |
| 4 | 5625 | 16014 | 0,306 | 0,005 | 0,029 | 0,987 |
| 5 | 5625 | 8094 | 0,058 | 0,009 | 0,206 | 0,980 |
| 6 | 4500 | 8850 | 0,059 | 0,013 | 0,104 | 0,981 |
| 7 | 4500 | 11244 | 0,495 | 0,028 | 0,042 | 0,965 |
| 8 | 5625 | 18678 | 0,922 | 0,258 | 0,011 | 0,970 |
| 9 | 5625 | 13818 | 0,000 | 0,000 | – | 1,000 |
| 10 | 5625 | 13782 | 0,000 | 0,001 | – | 0,999 |
| 11 | 5625 | 10380 | 0,704 | 0,306 | 0,009 | 0,903 |
| 12 | 5625 | 13800 | 0,432 | 0,024 | 0,024 | 0,976 |
| 13 | 5625 | 19956 | 0,131 | 0,021 | 0,088 | 0,970 |
| 14 | 3375 | 8688 | 0,000 | 0,433 | – | 0,567 |
| 15 | 5625 | 20028 | 0,136 | 0,025 | 0,048 | 0,972 |
| 16 | 4500 | 10470 | 0,915 | 0,699 | 0,004 | 0,937 |
| 17 | 4500 | 8688 | 0,584 | 0,254 | 0,023 | 0,881 |
| 18 | 4096 | 8704 | 0,071 | 0,019 | 0,046 | 0,979 |
| 19 | 4096 | 9216 | 0,321 | 0,035 | 0,011 | 0,973 |
| 20 | 5625 | 6870 | 0,960 | 0,498 | 0,001 | 0,979 |
| 21 | 5625 | 7446 | 0,184 | 0,021 | 0,042 | 0,975 |
| 22 | 6204 | 4572 | 0,284 | 0,014 | 0,080 | 0,967 |
| 23 | 6580 | 10873 | 0,155 | 0,014 | 0,045 | 0,981 |
| 24 | 1453 | 4359 | 0,017 | 0,005 | 0,322 | 0,989 |
| 25 | 6258 | 34244 | 0,456 | 0,036 | 0,035 | 0,965 |
| 26 | 5145 | 17923 | 0,000 | 0,000 | – | 1,000 |
| 27 | 5469 | 8464 | 0,000 | 0,001 | – | 0,999 |
| 28 | 6305 | 30814 | 0,000 | 0,001 | – | 0,999 |
| 29 | 6289 | 7842 | 0,000 | 0,009 | – | 0,991 |
| 30 | 6304 | 10753 | 0,000 | 0,007 | – | 0,993 |
| 31 | 6338 | 10287 | 0,000 | 0,002 | – | 0,998 |
| 32 | 5183 | 8319 | 0,000 | 0,000 | – | 1,000 |
| Итого | | | 0,253 | 0,029 | 0,025 | 0,972 |



а)



б)



в)



г)



д)

Рис. 3. Примеры результатов сегментации облачности:

- а) выделение сплошной облачности и ложное обнаружение дымки на мелководном озере;*
- б) корректное отделение облачности от песчаных карьеров; в) песок не выделен как облачность;*
- г) наиболее яркий песок с гладкой текстурой ошибочно выделен как облачность;*
- д) снежные шапки гор практически не выделяются как облачность*

β для всех тестовых снимков, а также итоговые значения показателей, рассчитанные с учетом площади каждого снимка и доли облачности на нем, достигнутые нейронной сетью «Lanku U-Net» со слоями нормализации, расположенными между сверткой и активационной

функцией, на вход которой подаются в трех каналах различные виды нормализации яркостей. Итоговое значение показателя α составляет 0,972. При этом результаты сегментации оказались неплохо сбалансированы по вероятностям ошибок первого и второго рода, со-

ставляющим 0,029 и 0,025 соответственно. Скорость обработки на графическом ускорителе NVIDIA GeForce RTX 2080 Ti составила 7,4 Мпикс./с.

На рис. 3 приведены примеры сегментации облачности. Визуальный анализ показывает, что ИНС в основном удается отличить сплошную облачность от других ярких объектов: песков, крыш зданий, снегов. При этом полупрозрачная дымка выделяется плохо, и иногда за нее могут быть ошибочно приняты мелководные водоемы с просвечивающим дном и пески.

Заключение

В настоящей работе предложена архитектура ИНС «Lanky U-Net», позволяющая при в 5,5 раз меньшем объеме вычислений, чем у архитектуры «U-Net» с уменьшенным в 2 раза числом каналов, достигнуть того же качества сегментации облачности на панхроматических изображениях. Проведены исследования по выбору уровня радиометрической обработки исходных данных ИНС. Предложено подавать на вход ИНС в трех каналах: изображение, полностью нормализованное с учетом зенитного угла Солнца; изображение, частично нормализованное путем пересчета в энергетические яркости на входном зрачке съёмочного устройства; исходное изображение в кодах яркости. Достигнуты общая точность сегментации 0,972, вероятность возникновения ошибок первого рода 0,029, вероятность возникновения ошибок второго рода 0,025 и скорость обработки на графическом ускорителе NVIDIA GeForce RTX 2080 Ti 7,4 Мпикс./с.

Литература

1. Гомозов О.А., Кузнецов А.Е., Побаруев В.И., Пошехонов В.И. Программно-математическое обеспечение системы обработки космических стереоизображений. Вестник Рязанского государственного радиотехнического университета. 2009. № 27. С. 17-22.
2. Кузнецов А.Е., Пошехонов В.И. Структурно-параметрический синтез компонентов малого космического аппарата картографического назначения. Вестник Рязанского государственного радиотехнического университета. 2019. № 69. С. 185-192.

3. Кузнецов А.Е. Системы и технологии обработки аэрокосмической информации. Вестник Рязанского государственного радиотехнического университета. 2012. № 39-2. С. 7-14.

4. Ветров А.А., Кузнецов А.Е. Сегментация облачных объектов на панхроматических изображениях земной поверхности. Цифровая обработка сигналов. 2011. № 3. С. 32-36.

5. Ereemeev V.V., Kochergin A.M., Kuznetsov A.E. Automatic detection of clouds in multispectral images subjected to interchannel parallax. IEEE International Geoscience and Remote Sensing Symposium. 2015, pp. 4928-4930.

6. Li Z., Shen H., Cheng Q., Liu Y., You S., He Z. Deep learning based cloud detection for medium and high resolution remote sensing images of different sensors. ISPRS J. of Photogram. and Rem. Sen. 2019, vol. 150, pp. 197-212.

7. Jiao L., Huo L., Hu C., Tang P. Refined UNet: UNet-Based Refinement Network for Cloud and Shadow Precise Segmentation // Remote Sensing. 2020, vol. 12(12).

8. Mohajerani S., Saeedi P. Cloud-Net: An End-to-End Cloud Detection Algorithm for Landsat 8 Imagery. IEEE International Geoscience and Remote Sensing Symposium. 2019, pp. 1029-1032.

9. Mohajerani S., Saeedi P. Cloud and Cloud Shadow Segmentation for Remote Sensing Imagery via Filtered Jacard Loss Function and Parametric Augmentation. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2021, vol. 14, pp. 4254-4266.

10. Ereemeev V., Kuznetsov A., Kochergin A., Makarenkov A. Clouds segmentation on panchromatic high spatial resolution remote sensing images using convolutional neural networks. Proceedings of the SPIE. 2019, vol. 11155.

11. Ronneberger O., Fischer P. and Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention. 2015, pp. 234-241.

12. Ioffe S., Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. Proceedings of the 32nd International Conference on International Conference on Machine Learning. 2015, vol. 37, pp. 448-456.

НОВЫЕ КНИГИ

Шарамет А.В.

Информационное обеспечение систем защиты летательных аппаратов от управляющих средств поражения: Научное издание, М.: Изд-во «Горячая линия-Телеком», 2023 г. 178 с.: ил.

Рассмотрены вопросы повышения информативности бортового комплекса обороны летательного аппарата для увеличения эффективности его защиты от управляемого ракетного оружия за счет согласованного помехового, маневренного и огневого противодействия. Изложены и систематизированы основные проблемы по защите летательного аппарата от управляемых средств поражения, а также рассмотрены основные пути повышения эффективности его защиты от них. Даны оценочные расчеты отражательных свойств головок самонаведения управляемых ракет класса «воздух-воздух» и «земля-воздух», как объектов радиолокационного наблюдения. Представлено описание и проведен анализ результатов полунатурного экспериментального исследования. На основе векторно-алгебраического подхода к формализации задач определения координат объектов в многопозиционной измерительной системе проведен синтез алгоритма оценки декартовых координат атакующей ракеты по суммарно-дальномерной информации.



Для специалистов, занимающихся вопросами защиты летательных аппаратов, научных работников и инженеров. Может быть полезна аспирантам и студентам вузов.

АДАПТИВНОЕ НЕЙРОСЕТЕВОЕ СЖАТИЕ МУЛЬТИСПЕКТРАЛЬНЫХ КОСМИЧЕСКИХ ИЗОБРАЖЕНИЙ ПОВЕРХНОСТИ ЗЕМЛИ

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

ADAPTIVE NEURAL NETWORK COMPRESSION OF MULTISPECTRAL SATELLITE IMAGES OF THE EARTH'S SURFACE

Ushenkin V.A.

The paper considers the problem of compressing multispectral images on board of remote sensing satellites in order to reduce the amount of information transmitted over the radio line. A simple architecture of artificial neural networks family, that allows adaptive compression of images with controlled distortions, is proposed. The question of training such family of networks is considered. The results of experimental studies are presented in comparison with known deterministic and neural network compression algorithms. In this paper, it was possible to obtain a two-stage neural network implementation of adaptive lossy image compression, which retains the advantages of known artificial neural networks over deterministic algorithms at high compression rates, but has significantly lower requirements for the speed and memory capacity of computing devices, which potentially allows it to be used on board of remote sensing satellites.

Key words: multispectral image, adaptive compression, remote sensing, artificial neural network.

Ключевые слова: мультиспектральное изображение, адаптивное сжатие, дистанционное зондирование Земли, искусственная нейронная сеть.

Введение

В последние годы с увеличением детальности съемки наблюдается стабильный рост объемов целевой информации, формируемой космическими аппаратами (КА) дистанционного зондирования Земли (ДЗЗ). Это создает проблемную ситуацию, связанную с обеспечением оперативности передачи на Землю большого по объему массива изображений при ограниченной пропускной способности радиолиний. Разрешение данной проблемы может быть осуществлено за счет бортового сжатия изображений перед их передачей по радиолинии.

Известен ряд классических алгоритмов, позволяющих уменьшить объем изображений. Алгоритмы сжатия без потерь (Deflate, RLE, LZW, LZMA, Lossless JPEG, Lossless JPEG2000) не вносят искажений в изображения, однако для космических снимков отличаются сравнительно невысокой степенью сжатия. Объем данных после применения указанных алгоритмов уменьшается примерно в 2 – 4 раза. В большей степени уменьшают объем данных алгоритмы сжатия с потерями (JPEG и JPEG2000), однако значительная степень сжатия достигается только при существенных искажениях изображения.

При передаче изображений с борта на Землю целесообразно прийти к компромиссу в виде сжатия с контролируруемыми потерями, при которых основные характеристики изображения деградируют незначительно, зато при этом повышается оперативность передачи изображений, а также ослабляются ограничения на производительность КА. При контролируемых потерях с помощью алгоритмов JPEG и JPEG2000 можно уменьшить объем данных примерно в 5–30 раз.

Рассматривается задача сжатия мультиспектральных изображений на борту космических аппаратов дистанционного зондирования Земли с целью уменьшения объема информации, передаваемой по радиолинии. Предлагается простая архитектура семейства искусственных нейронных сетей, позволяющего адаптивно сжимать изображения с контролируемыми искажениями. Рассматривается вопрос обучения такого семейства сетей. Приводятся результаты экспериментальных исследований в сравнении с известными детерминированными и нейросетевыми алгоритмами сжатия.

Также целесообразно применение адаптивного сжатия. На спутниковых снимках часто присутствуют малоинформативные объекты: облачность, водная поверхность и др. Фрагменты изображений с такими объектами можно сжимать с существенными потерями, дополнительно уменьшая объем передаваемых данных.

Помимо классических детерминированных алгоритмов [1, 2], в последние годы опубликован ряд работ по применению искусственных нейронных сетей (ИНС) в задаче сжатия данных, в т.ч. изображений. Среди всех работ также можно выделить два крупных направления: сжатие изображений с потерями и сжатие данных без потерь. При этом подходы в рамках первого направления не исключают последующее применение подходов из второго направления к результатам сжатия с потерями, за счет чего степень сжатия увеличивается дополнительно. Результаты экспериментальных исследований, приводимые в опубликованных работах, показывают, что применение ИНС позволяет достичь при значительных степенях сжатия меньших искажений, чем в алгоритмах JPEG и JPEG2000. Однако известные нейросетевые подходы отличаются высокой вычислительной сложностью, что не позволяет их применять в условиях ограниченных вычислительных мощностей борта КА. Поэтому актуальна разработка нейросетевых подходов, обладающих существенно меньшей вычислительной сложностью.

стью, но сохраняющих преимущества своих прототипов. Такая разработка является целью настоящей работы.

Анализ известных нейросетевых подходов к сжатию изображений

В основе абсолютного большинства работ по нейросетевому сжатию изображений с потерями лежит использование так называемых автокодировщиков, представляющих собой ИНС, состоящую из двух частей: кодировщика и декодировщика [3–6]. При этом вход и выход ИНС идентичны по объему и структуре данных, а данные, передаваемые от кодировщика к декодировщику, имеют существенно меньший объем, чем данные на входе и выходе ИНС. При обучении автокодировщика добиваются наибольшего сходства данных на входе и выходе ИНС, при этом их отличие составляет потери при сжатии. Степень сжатия определяется структурой сети, а именно отношением объема данных на входе/выходе ИНС к объему данных, проходящих через центр сети. После обучения сеть разбивается на кодировщик и декодировщик, которые затем используются независимо, первый – при передаче данных, второй – при их приеме и распаковке.

Для управления степенью сжатия изображений с потерями используется два основных подхода. В первом подходе [3, 6] обучается несколько автокодировщиков, отличающихся объемом данных, проходящих через центр ИНС, а непосредственно при сжатии в соответствии с некоторым признаком для каждого сжимаемого фрагмента изображения адаптивно выбирается кодировщик с наиболее подходящей степенью сжатия. Во втором подходе [4, 5] используется рекуррентная схема. Обучается автокодировщик, характеризующийся большой степенью сжатия и значительными потерями. При выполнении сжатия изображение проходит через весь автокодировщик, после чего путем вычитания выходного и входного сигнала автокодировщика вычисляются потери. Сигнал, прошедший через центр автокодировщика, записывается в выходной поток. Если потери оказываются выше требуемых, то на вход автокодировщика подаются уже потери, результат сжатия которых записывается в выходной поток. Опять сравниваются сигналы на выходе и входе автокодировщика, и если они снова сильно отличаются друг от друга, рекуррентная схема сжатия продолжается.

Величина потерь при фиксированной степени сжатия существенно зависит от структуры автокодировщика. В опубликованных работах применяются многослойные перцептроны [3], сверточные ИНС [6], рекуррентные ИНС [4, 5], слои нормализации данных [5, 6] и их комбинация. Многослойные перцептроны не учитывают растровую структуру изображения и не позволяют добиться малых потерь. Рекуррентные ИНС хорошо подходят для рекуррентной схемы управления степенью сжатия, поскольку обладают памятью и по-разному преобразуют сигнал в зависимости от своего состояния. Однако они характеризуются наиболее высокой вычислительной сложностью. Поэтому для реализации на борту КА ДЗЗ наиболее подходят сверточные ИНС, учитывающие растровую структуру изображения и обладающие уме-

ренной вычислительной сложностью.

Результат сжатия с потерями может быть подвергнут дополнительному сжатию без потерь, которое может быть выполнено как с помощью традиционных алгоритмов, так и с помощью ИНС. В задаче сжатия без потерь ИНС используются для предсказания следующего значения сигнала по его предыдущим значениям. Результатом предсказания являются условные вероятности появления следующего символа после известных текущих. Переход от безусловных вероятностей появления символа в потоке к условным вероятностям позволяет существенно уменьшить энтропию сигнала, в результате чего последующее традиционное энтропийное кодирование существенно уменьшает объем данных. Результаты экспериментов, выполненных Мэттом Махони [7], показывают, что применение ИНС для предсказания следующего значения сигнала позволяет добиться значительно большего сжатия данных без потерь, чем в традиционных алгоритмах, ценой существенного увеличения времени сжатия и затрат оперативной памяти. Это недопустимо для бортовой реализации, однако возможно уменьшить затраты времени и памяти без существенного ухудшения степени сжатия. Основные затраты обусловлены необходимостью обучать ИНС при каждом сжатии новых данных. Их можно уменьшить, если удастся ИНС, используемую для предсказания, обучить заранее с обеспечением приемлемой точности предсказания для всех сжимаемых данных. Поскольку при бортовой обработке космических снимков сжатие без потерь применяется не к произвольным данным, а к результатам сжатия с потерями, всегда имеющим одну и ту же структуру и сходные зависимости между символами, такая задача может быть успешно решена.

Архитектура семейства автокодировщиков для адаптивного сжатия мультиспектральных изображений с потерями

Известные автокодировщики на базе сверточных ИНС имеют следующие общие черты. В них выполняется несколько последовательных сверток сигнала. Результаты сверток постепенно прореживаются по строкам и столбцам, а число каналов в изображении увеличивается. После прохождения нескольких сверточных слоев получается в несколько раз прореженное изображение с большим числом каналов. При этом в конце концов отношение текущего числа каналов к исходному оказывается меньше, чем квадрат коэффициента прореживания. За счет этого достигается уменьшение объема изображения. Для дополнительного уменьшения объема обычно выполняется квантование значений сигнала в пикселях сжатого изображения. Описанная часть ИНС представляет собой кодировщик. Декодировщик включает те же сверточные слои, расположенные в обратном порядке и чередующиеся с операцией «Depth to space». Указанная операция увеличивает число строк и столбцов изображения и уменьшает число каналов, заполняя новые строки данными отбрасываемых каналов. В результате восстанавливается изображение с исходным числом строк, столбцов и каналов. Выполнение в сверточных нейронных сетях нескольких последовательных



сверток с прореживанием позволяет построить пирамиду изображений, что имеет некоторое сходство с вейвлет-преобразованием, используемым в алгоритме JPEG2000 и учитывает растровую структуру изображения.

В алгоритме JPEG2000 используются одномерные фильтры из 7 и 9 элементов (вейвлеты Козна-Добешифово 9/7), применяемые сначала к строкам, а затем к столбцам изображения. Таким образом, преобразование фактически сводится к свертке в окне 9x9, которая описывается всего $9+7=16$ коэффициентами. Свертка в пределах окна большей площади позволяет снизить влияние погрешностей квантования сигнала на верхнем уровне пирамиды на сигнал в отдельном пикселе на нижнем уровне пирамиды после восстановления сжатого изображения. Известные архитектуры автокодировщиков на каждом уровне пирамиды используют существенно большее число коэффициентов, но выполняют свертку с фильтрами меньшей апертуры (3x3 или 5x5). Для сокращения объема вычислений решено операцию «Conv2D» (свертки) заменить двумя операциями «DepthwiseConv2D» (поканальной свертки) в окнах 9x1 и 1x9 соответственно. Операция «DepthwiseConv2D» для формирования выходного канала изображения использует только один канал входного изображения, что существенно сокращает число весовых коэффициентов.

Кроме того, выполнен отказ от избыточного числа каналов на промежуточных уровнях пирамидального преобразования. На каждом последующем уровне число каналов увеличивается в 4 раза с одновременным прореживанием по строкам и столбцам в 2 раза и сохранением общего объема данных. По окончании пирамидального преобразования на верхнем уровне пирамиды выполняется дополнительная свертка сигнала с уменьшением числа каналов и последующим квантованием с целью уменьшения объема данных.

Поскольку при сжатии с потерями в первую очередь деградирует высокочастотная составляющая сигнала изображения, решено добавить в архитектуру автокодировщика слой коррекции этой составляющей. Он заключается в поканальной свертке сигнала в окне 5x5 с фильтром, сумма коэффициентов которого равна нулю, и прибавлении результата свертки к исходному сигналу.

Перед кодированием выполняется нормализация яркостей изображения, а в конце декодирования – денормализация. Также в начале и в конце автокодировщика выполняется операция «Conv2D» в окне 1x1, по сути выполняющая некоторое преобразование цветового пространства изображения по аналогии с алгоритмами JPEG и JPEG2000. Полученная архитектура автокодировщика приведена на рис. 1.

Недостатком известных адаптивных подходов к сжатию является возникновение блочных артефактов при переключении степени сжатия. Это связано с тем, что для каждой степени сжатия автокодировщики работают с разными весовыми коэффициентами во всех слоях. Чтобы избежать влияния краевых эффектов на свертки, выполняемые при декодировании сжатого изображения, требуется записывать в сжатый поток избыточные данные, что отрицательно сказывается на степени сжатия.

В противном случае на границах блоков возникают краевые эффекты, увеличивающие уровень искажений при сжатии и хорошо заметные глазу.

Для устранения этого эффекта предлагается принципиально новый подход к обучению ИНС для сжатия: обучение семейства взаимосогласованных по весовым коэффициентам кодировщиков для каждой степени сжатия с единым декодировщиком.

Все кодировщики в рамках семейства имеют одни и те же весовые коэффициенты для пирамидального преобразования изображения и отличаются друг от друга лишь на последнем уровне. Последние уровни кодировщиков формируют разное число X выходных каналов. При этом весовые коэффициенты кодировщика с большей степенью сжатия, формирующего меньшее число выходных каналов, целиком включаются в состав выходного слоя кодировщика с меньшей степенью сжатия, формирующего те же выходные каналы, а также ряд дополнительных выходных каналов.

количество квантованных каналов сжатого изображения. Единый декодировщик позволяет из всех сжатых блоков собрать единое сжатое изображение. Если при некоторой степени сжатия в блоке формируется меньшее число каналов, недостающие каналы заполняются нулями. В этом также прослеживается сходство с алгоритмами JPEG/JPEG2000, в которых обнуляется часть коэффициентов. Единое сжатое изображение преобразуется декодировщиком в результат распаковки, на котором отсутствуют блочные артефакты, поскольку резкая граница между нулевым и ненулевым сигналом соседних блоков размывается в ходе свертки, выполняемых декодировщиком.

Основные трудности возникают с обучением такого семейства нейронных сетей. Добиться минимума потерь на всех степенях сжатия крайне затруднительно, поскольку для этого требуется многокритериальная оптимизация по нескольким выходам нейронной сети. В связи с этим предлагается «жадный» алгоритм обучения.

Сначала подбираются коэффициенты пирамидального преобразования изображения, обеспечивающие минимальное влияние шумов квантования сигнала верхнего уровня пирамиды на восстановление изображения на нижнем уровне. Для этого строится базовый автокодировщик, не выполняющий сжатие изображения за счет уменьшения числа каналов. Сжатие в нем выполняется только лишь за счет квантования сигнала верхнего уровня пирамиды с минимальным числом бит. В рамках настоящей работы в качестве такого числа бит выбрано 4.

Для обучения семейства кодировщиков последовательно составляются и обучаются автокодировщики с разным числом выходных каналов в слое перед блочно-адаптивным квантованием. Первым обучается автокодировщик с максимальной степенью сжатия (минимальным числом каналов, равным 12, каждый из которых адаптивно квантуется 4 битами на отрезке от минимального до максимального значения сигнала в пределах блока). Границы адаптивного квантования в пределах блока вместе с собственно квантованными значениями записываются в выходной поток.

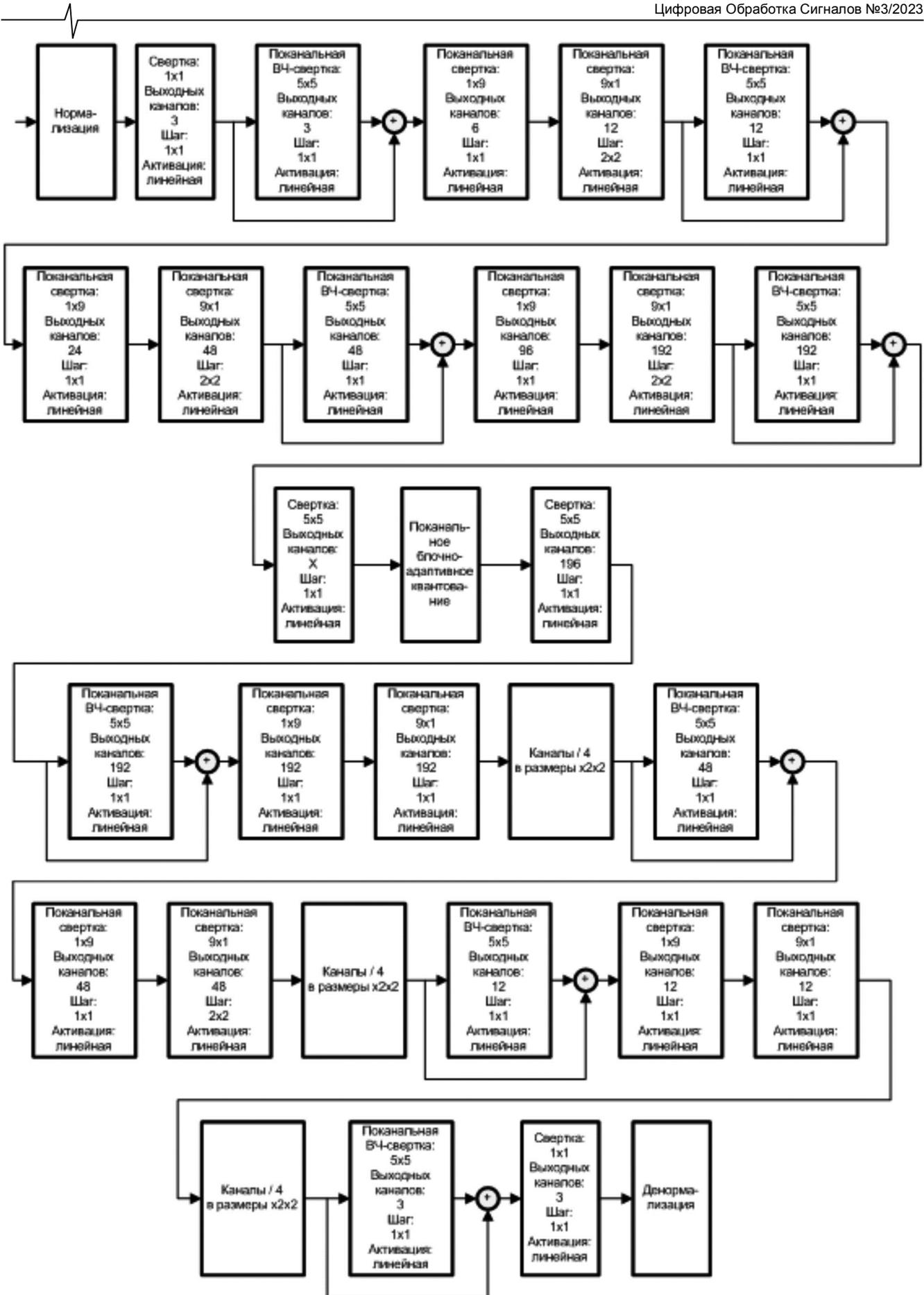
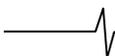


Рис. 1. Предлагаемая архитектура автокодировщика



При обучении такого автокодировщика в начальные и конечные слои сети загружаются и фиксируются весовые коэффициенты, полученные для автокодировщика, обученного ранее, а обучение выполняется только для двух средних слоев, соседних со слоем квантования.

Декодировщик формируется единый для всех кодировщиков и принимает на вход максимальное

Для следующего автокодировщика увеличивается до 24 число каналов на выходе слоя перед квантованием и на входе слоя после квантования. Каждый канал по-прежнему квантуется 4 битами. В слои данного автокодировщика загружаются и фиксируются весовые коэффициенты от предыдущей сети. В средних слоях новой сети число весовых коэффициентов увеличивается. Поэтому в них загружается и фиксируется лишь часть весов, а оставшаяся часть подбирается в ходе обучения. Таким образом, новый кодировщик формирует те же 12 каналов сжатого изображения, что и предыдущий, а также 12 новых каналов, позволяющих более точно восстановить исходное изображение из сжатого.

Для следующего автокодировщика число каналов составляет 36. Первые 24 канала квантуются 5 битами и формируются с теми же весами, что и в предыдущей сети, а дополнительные 12 каналов формируются с весами, подбираемыми в ходе обучения, и квантуются 6 битами.

Для последнего автокодировщика число каналов составляет 60. Первые 36 каналов квантуются 6 битами, а дополнительные 24 канала – 7 битами. Таким образом, сигнал от каждого следующего кодировщика целиком включает в себя сигналы от всех предыдущих кодировщиков, лишь менее грубо их квантуя и дополняя новыми каналами, содержащими поправки для более точного восстановления исходного изображения. Вторая половина слоев последнего автокодировщика образует единый декодировщик адаптивно сжатого изображения.

Объемы данных на выходе кодировщиков составляют 0,75, 1,5, 3 и 6 бит на пиксель исходного мультиспектрального 3-канального изображения.

Для адаптивного выбора степени сжатия с приемлемым уровнем искажений может использоваться следующий подход. В рамках этого подхода сигнал проходит через последний обученный кодировщик, выполняющий минимальное сжатие. Затем выполняется отбрасывание (обнуление) части каналов на выходе кодировщика и квантование сигнала в оставшихся каналах. Сначала оставляются и квантуются 4 битами 12 каналов, соответствующих максимальному сжатию. Полученный сигнал подается на вход декодировщика, сигнал на выходе которого сопоставляется с блоком исходного изображения с вычислением потерь при сжатии. Если потери слишком велики, то от сигнала, сформированного кодировщиком, оставляется большее количество каналов и вновь выполняется декодирование с оценкой потерь. Процесс повторяется либо до достижения при некоторой степени сжатия приемлемых потерь, либо до достижения минимально предусмотренной степени сжатия. Эксперименты показали, что среднеквадратичная ошибка восстановления исходного изображения по сжатому для каждого следующего автокодировщика уменьшается

примерно в 1,4 раза по сравнению с предыдущим.

В дополнение к описанному процессу может выполняться учет дополнительных признаков. Если имеется маска малоинформативных объектов на снимке, а сигнал сжимаемого блока целиком соответствует малоинформативным объектам, для этого блока выбирается максимальная степень сжатия без контроля потерь.

Архитектура семейства нейросетевых предсказателей для дополнительного сжатия изображений без потерь

Результаты нейросетевого сжатия с потерями могут быть подвергнуты дополнительному сжатию без потерь на основе энтропийного кодирования и нейросетевого предсказания значения следующего отсчета сигнала по предыдущим значениям. Такой подход позволяет добиться большего сжатия без потерь результатов нейросетевого сжатия с потерями, чем детерминированные алгоритмы. Экспериментальные исследования показали, что традиционные архиваторы WinRAR и 7 Zip уменьшают объем результатов нейросетевого сжатия с потерями всего на ~5 %, а лучшие из имеющихся детерминированных алгоритмов (MCM и BSC) – на 10-11 %. В то же время с помощью открытого алгоритма нейросетевого сжатия без потерь DeepZIP [8] удалось уменьшить объем на ~20 %.

При этом был поставлен эксперимент, в рамках которого ИНС-предсказатель в DeepZIP была обучена по одной пятой части изображения, а затем применена ко всему изображению. Достигнутая степень сжатия оказалась сопоставима как для пятой части файла, по которой выполнялось обучение, так и для файла целиком. Это позволило сделать вывод, что нейронную сеть-предсказатель достаточно обучить один раз, а затем многократно использовать для сжатия без потерь результатов сжатия изображений с потерями.

В составе нейронных сетей-предсказателей в настоящее время используются следующие слои: рекуррентные слои (LSTM, GRU) и полносвязные слои. По результатам экспериментальной отработки с помощью алгоритма DeepZIP, поддерживающего несколько различных архитектур сети-предсказателя, установлено, что наилучшее предсказание следующего пикселя результатов сжатия с потерями обеспечивает нейронная сеть на базе полносвязных слоев. Достигнутый объем данных получается на 1 % ниже, чем при использовании слоев LSTM и GRU, при этом объем вычислений оказывается в 2 раза меньше.

Следующим шагом по повышению степени сжатия без потерь стало обучение отдельной сети на предсказание следующего пикселя в каждом канале изображения, сжатого нейронной сетью с потерями. Нейронная сеть использует значения каналов предыдущего пикселя изображения и значения предыдущих каналов текущего пикселя, чтобы предсказать значение следующего канала текущего пикселя. Число нейронных сетей-предсказателей равняется числу каналов. В результате с помощью набора нейронных сетей на базе полносвязных слоев, каждая из которых была обучена по отдельному каналу изображения, достигнуто на несколько про-

центров большее уменьшение объема данных, чем у DeepZIP. При этом в каждом отдельном предсказателе удалось уменьшить число нейронов без существенного снижения достигаемой степени сжатия. В итоге, наилучшим образом учтя структуру сжимаемых данных, удалось добиться высокой степени сжатия без потерь при небольших вычислительных затратах и умеренном потреблении оперативной памяти. Достигнутая степень сжатия лишь немногим уступает лучшему на сегодняшний день алгоритму сжатия без потерь NNCP [9] на основе ИНС-трансформера, требующему на несколько порядков больших затрат.

Архитектура семейства поканальных предсказателей на основе полносвязных ИНС представлена на рис. 2, где b – число бит, которыми квантуется сигнал на выходе сжатия с потерями. Размерность сигнала, подаваемого на вход предсказателя, равняется числу каналов при заданной степени сжатия.



Рис. 2. Предлагаемая архитектура предсказателя

Обучение ИНС

Для обучения автокодировщиков необходимы разрезанные на фрагменты мультиспектральные космические снимки с пространственным разрешением, соответствующим характеристикам заданного КА ДЗЗ. Снимки должны содержать большинство типовых наблюдаемых объектов в пределах всего возможного диапазона яркостей: водную поверхность, облака, поля, леса, городскую застройку. Расчет целевой функции обучения выполняется лишь по центральной части каждого фрагмента для исключения влияния краевых эффектов. В качестве целевой функции могут использоваться среднеквадратичное отклонение яркостей изображений на входе и выходе автокодировщика, одномасштабный и многомасштабный индексы структурного сходства и другие показатели. В настоящей работе при обучении использовалось среднеквадратичное отклонение.

Из-за наличия слоя квантования, отличающегося разрывным градиентом, сходимость процесса минимизации целевой функции крайне нестабильная. При этом применение пакетной нормализации [10] в данном случае сходимость практически не улучшает, поэтому ее использование нецелесообразно. За ходом обучения необходимо постоянно следить и снижать скорость градиентного спуска, если процесс начал расходиться или целевая функция перестала уменьшаться. Особенно сложно выполняется обучение базового автокодировщика, поскольку подбираются коэффициенты для всех слоев нейронной сети (так называемое глубокое обучение).

Для обучения нейронных сетей-предсказателей генерируются отдельные выборки путем неадаптивного сжатия с потерями мультиспектральных космических снимков. Снимки должны быть сжаты с каждой из возможных степеней сжатия, фиксированной в пределах всего снимка. На вход каждого предсказателя значения в k -м канале сжатых данных (k изменяется от 1 до N_k) подаются значения в последних ($N_k - k + 1$) каналах предыдущего пикселя сжатого изображения и в первых ($k - 1$) каналах текущего пикселя. Эталонным выходом нейронной сети является сигнал из $2b$ отсчетов, из которых один, соответствующий b -битному коду в k -м канале текущего пикселя, является единичным, а остальные нулевыми. Целевой функцией, минимизируемой при обучении, является перекрестная энтропия между выходом нейронной сети и эталонным выходом. В отличие от обучения автокодировщиков, включающего от нескольких сотен до нескольких тысяч эпох, обучение каждого предсказателя выполняется быстро: за несколько эпох.

Экспериментальные исследования

Проведены экспериментальные исследования по сравнению потерь при реализованном нейросетевом сжатии по сравнению с детерминированными алгоритмами сжатия изображений с потерями JPEG и JPEG2000, а также с одним из известных нейросетевых алгоритмов [4], для которого в открытом доступе имеется обученная рекуррентная ИНС. Сравнение с другими нейросетевыми алгоритмами сжатия с потерями затруднительно ввиду отсутствия обученных ИНС.

В качестве меры потерь при сжатии использовались следующие показатели:

- среднеквадратичное отклонение яркостей в каналах пикселей исходного и распакованного сжатого изображений;
- индекс структурного сходства и его три составляющие: сходство яркостей, сходство контраста и сходство структуры [11];
- интегральный показатель сходства яркостей, контраста и структуры.

Среднеквадратичное отклонение рассчитывается как

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (b_{ucxi} - b_{cxi})^2}, \text{ где } b_{ucxi} - \text{ код яркости } i\text{-го отсчета}$$

исходного изображения, b_{cxi} – код яркости i -го отсчета распакованного сжатого изображения, N – общее число отсчетов, равное произведению числа пикселей и числа каналов. За счет большого числа отсчетов смещением на единицу числа степеней свободы при оценке выборочной дисперсии пренебрегается. Отсутствию искажений соответствует нулевое значение среднеквадратичного отклонения.

Индекс структурного сходства рассчитывается для каждого канала в скользящем окне 8×8 пикселей, а в качестве итогового значения берется результат усреднения по всем положениям скользящего окна во всех каналах изображений. В пределах каждого окна вычисляются статистические характеристики яркостей в выбранном канале исходного и сжатого изображений: выборочные средние и дисперсии, а также ковариация между яркостями исходного и сжатого изображений. Индекс структурного

сходства в пределах окна рассчитывается как $SSIM = \frac{(2\mu_{исх}\mu_{сж} + (0,01 \cdot 255)^2)(2cov_{исх,сж} + (0,03 \cdot 255)^2)}{(\mu_{исх}^2 + \mu_{сж}^2 + (0,01 \cdot 255)^2)(\sigma_{исх}^2 + \sigma_{сж}^2 + (0,03 \cdot 255)^2)}$

где $\mu_{исх}$ и $\sigma_{исх}^2$ – выборочные среднее и дисперсия яркостей в канале исходного изображения, $\mu_{сж}$ и $\sigma_{сж}^2$ – выборочные среднее и дисперсия яркостей в канале распакованного сжатого изображения, $cov_{исх,сж}$ – ковариация между яркостями исходного и сжатого изображений.

Значение индекса структурного сходства в пределах каждого окна для каждого канала может быть представлено в виде произведения трех составляющих: сходства яркостей $l = \frac{2\mu_{исх}\mu_{сж} + (0,01 \cdot 255)^2}{\mu_{исх}^2 + \mu_{сж}^2 + (0,01 \cdot 255)^2}$, сходства контра-

ста $c = \frac{2\sigma_{исх}\sigma_{сж} + (0,03 \cdot 255)^2}{\sigma_{исх}^2 + \sigma_{сж}^2 + (0,03 \cdot 255)^2}$ и сходства структуры

$s = \frac{cov_{исх,сж} + 0,5 \cdot (0,03 \cdot 255)^2}{\sigma_{исх}\sigma_{сж} + 0,5 \cdot (0,03 \cdot 255)^2}$. Последний показатель

представляет собой коэффициент корреляции Пирсона, в числитель и знаменатель которого добавлены регуляризирующие коэффициенты, позволяющие избежать деления на нуль. Перечисленные составляющие также усредняются по всем положениям скользящего окна и всем каналам с получением трех частных показателей сходства изображений. Индекс структурного сходства и три его составляющие при отсутствии искажений принимают единичное значение, а с увеличением искажений при сжатии снижаются. Поэтому в качестве меры потерь могут использоваться величины $1 - SSIM$, $1 - l$, $1 - c$ и $1 - s$.

Индекс структурного сходства является широко распространенной мерой оценки качества нейросетевого сжатия, однако разложение его на составляющие показывает его недостатки. Показатель сходства яркостей сопоставляет средние яркости, но при наличии искажений оказывается не инвариантен к одинаковому аддитивному сдвигу яркостей сравниваемых изображений: $l(a, b) \neq l(a + c, b + c)$. Если средние яркости оказываются малы, то их небольшие различия начинают сильно снижать значение показателя, в то время как такие искажения практически не заметны человеческому глазу и практически не ухудшают измерительные свойства изображения. Частично этот недостаток компенсирует наличие регуляризирующего коэффициента в числителе и знаменателе показателя, однако негативное влияние всё равно присутствует. Аналогичные рассуждения можно провести и для показателя сходства контрастов. В случае сопоставления изображений однородных по яркости объектов, когда $\sigma_{исх}^2$ и $\sigma_{сж}^2$ малы, значение показателя получается заниженным.

С учетом вышесказанного в рамках настоящей работы предлагается новый интегральный показатель сходства яркостей, контраста и структуры, выступающий альтернативой индексу структурного сходства. В пределах каждого положения скользящего окна в каждом канале он

рассчитывается как $LCS = \frac{cov_{исх,сж} + 0,5 \cdot (0,03 \cdot 255)^2}{\sigma_{исх}\sigma_{сж} + 0,5 \cdot (0,03 \cdot 255)^2}$

$\frac{|\mu_{исх} - \mu_{сж}|}{255} - \frac{6 \cdot |\sigma_{исх} - \sigma_{сж}|}{255}$. Коэффициент 6 предпола-

гает, что яркости наиболее контрастных изображений имеют нормальное распределение на отрезке $[0, 255]$.

При выполнении экспериментальных исследований выполнялось сжатие изображений алгоритмом JPEG с разными значениями параметра качества, алгоритмом JPEG2000 с разными значениями степени сжатия и предложенным адаптивным нейросетевым алгоритмом с разными значениями допустимых потерь в двух конфигурациях: без дополнительного сжатия без потерь и с дополнительным сжатием без потерь на основе нейросетевого предсказания и энтропийного кодирования. Графики полученных значений показателей приведены на рис. 3-8.

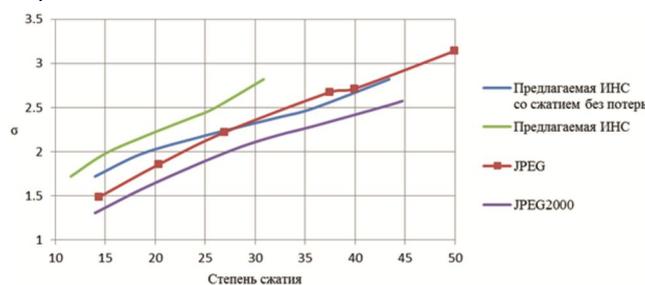


Рис. 3. Зависимость среднеквадратичного отклонения от степени сжатия мультиспектрального изображения

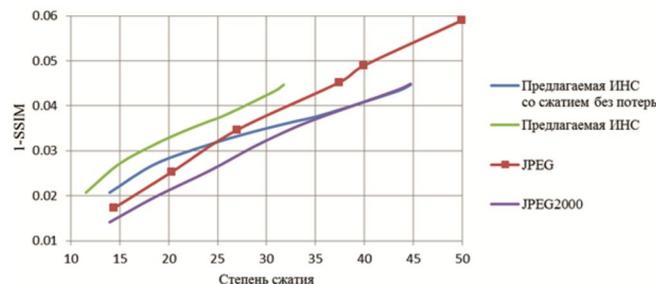


Рис. 4. Зависимость индекса структурного сходства от степени сжатия мультиспектрального изображения

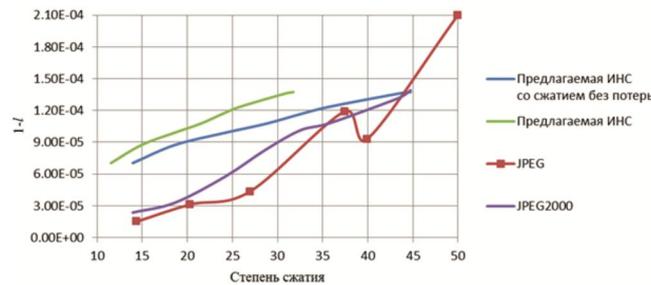


Рис. 5. Зависимость сходства яркостей от степени сжатия мультиспектрального изображения

Из рисунков видно, что нейросетевое сжатие мультиспектральных изображений имеет преимущество над детерминированными алгоритмами по части показателей при высоких степенях сжатия. При этом преимущество достигается только в сочетании с последующим сжатием без потерь на основе нейросетевого предсказания. При степенях сжатия, больших 28, нейросетевое

сжатие превосходит алгоритм JPEG по критерию среднеквадратичного отклонения яркостей, но всё равно уступает по данному критерию алгоритму JPEG2000. Нейросетевое сжатие лучше показывает себя по индексу структурного сходства. Начиная со степени сжатия 25, оно существенно превосходит алгоритм JPEG, а со степени сжатия 38 начинает незначительно превосходить алгоритм JPEG2000. Наибольший выигрыш нейросетевого сжатия наблюдается по введённому интегральному показателю сходства яркостей, контраста и структуры. Оно существенно превосходит алгоритм JPEG, начиная со степени сжатия 29, и примерно на 1 % превосходит алгоритм JPEG2000, начиная со степени сжатия 33.

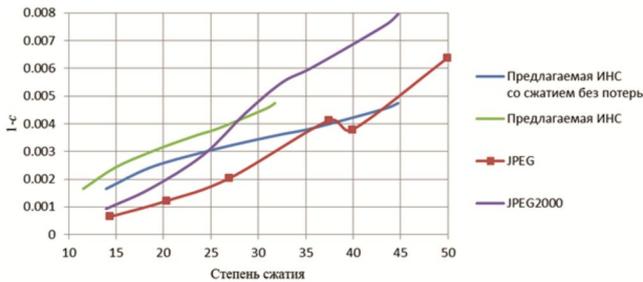


Рис. 6. Зависимость сходства контраста от степени сжатия мультиспектрального изображения

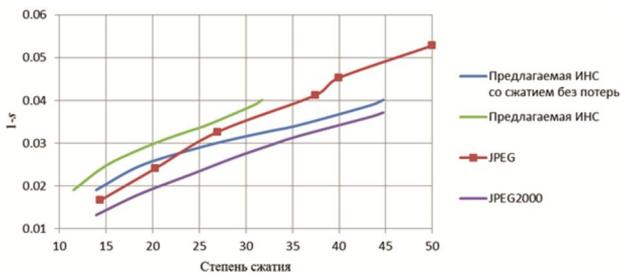


Рис. 7. Зависимость сходства структуры от степени сжатия мультиспектрального изображения

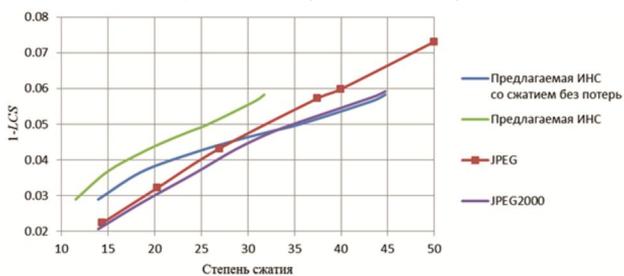


Рис. 8. Зависимость интегрального показателя сходства яркостей, контраста и структуры от степени сжатия мультиспектрального изображения

Частные показатели показывают лишь отдельные аспекты сходства исходного и сжатого изображений и позволяют выявить, какого типа искажения в основном вносит тот или иной алгоритм сжатия. Слабым местом алгоритма JPEG2000 является искажение локального контраста изображения. Слабым местом ИНС – искажение среднего уровня яркости. Слабым местом алгоритма JPEG – искажение структуры изображения.

Скорость предлагаемого нейросетевого сжатия на процессоре Intel Core i7-6800K составила около 0,32 Мпикс/с, что в ~12 раз медленнее алгоритма JPEG2000, но уже потенциально реализуемо на борту КА ДЗЗ.

Сравнение с рекуррентной ИНС [4] выполнено на небольшом фрагменте мультиспектрального спутникового снимка городской территории (размером 4000x2000 пикселей) из-за огромной вычислительной сложности обработки с помощью этой ИНС. Время обработки указанного фрагмента рекуррентной нейронной сетью на процессоре Intel Core i7-6800K составило около 3 часов, а затраты оперативной памяти – около 45 Гбайт. В то же время обработка с помощью предложенной в настоящей работе архитектуры заняла 18 секунд при затратах оперативной памяти в десятки Мбайт. Поскольку для рекуррентной ИНС [4] в открытом доступе имеется полноценная реализация лишь первой части сжатия (сжатие с потерями на базе автокодировщика), а для последующего сжатия без потерь реализация отсутствует, при сравнении с ней предлагаемая архитектура автокодировщиков также использовалась без последующего дополнительного сжатия без потерь. Результаты сравнения приведены на рис. 9-14. Из рисунков видно, что нейронные сети очень близки по значениям основных показателей. Рекуррентная ИНС [4] лучше по среднеквадратичному отклонению и сходству яркостей, а предложенная сеть лучше по индексу структурного сходства и сходству контрастов. По сходству структуры и введённому интегральному показателю сети практически одинаковы.

Заключение

В настоящей работе удалось получить двухэтапную нейросетевую реализацию адаптивного сжатия изображений с потерями, которая сохраняет преимущества известных ИНС перед детерминированными алгоритмами при высоких степенях сжатия, но обладает существенно меньшими требованиями к быстродействию и объему памяти вычислительных устройств, что потенциально позволяет ее использовать на борту КА ДЗЗ.

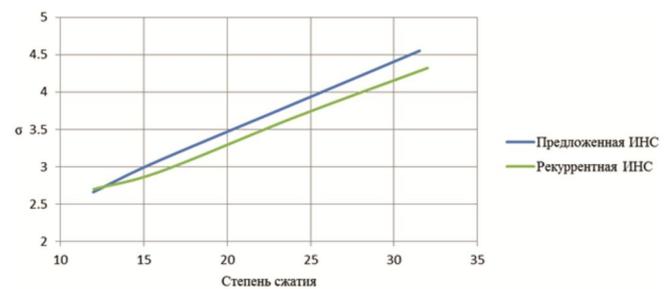


Рис. 9. Зависимость среднеквадратичного отклонения от степени сжатия тестового фрагмента мультиспектрального изображения

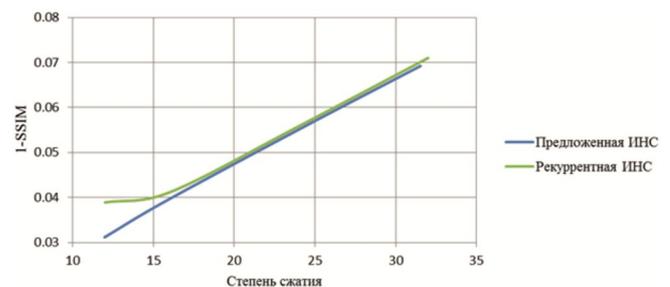


Рис. 10. Зависимость индекса структурного сходства от степени сжатия тестового фрагмента мультиспектрального изображения

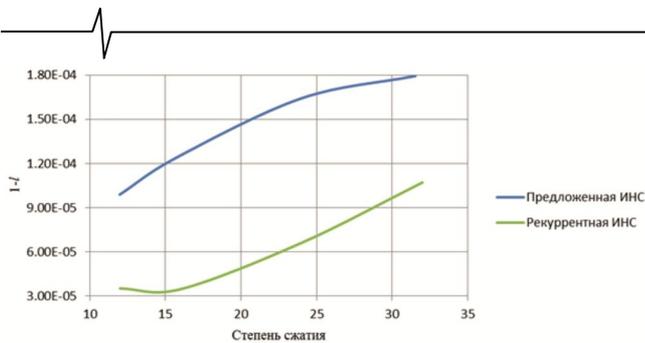


Рис. 11. Зависимость сходства яркостей от степени сжатия тестового фрагмента мультиспектрального изображения

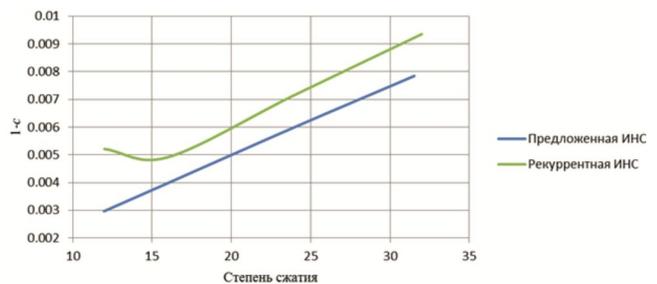


Рис. 12. Зависимость сходства контраста от степени сжатия тестового фрагмента мультиспектрального изображения

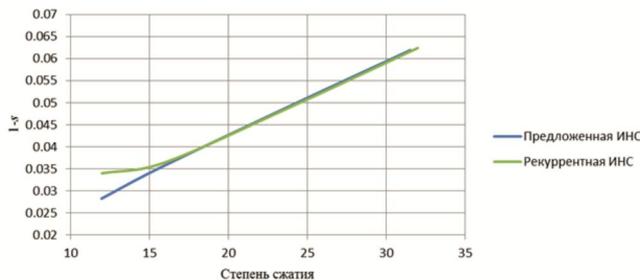


Рис. 13. Зависимость сходства структуры от степени сжатия тестового фрагмента мультиспектрального изображения

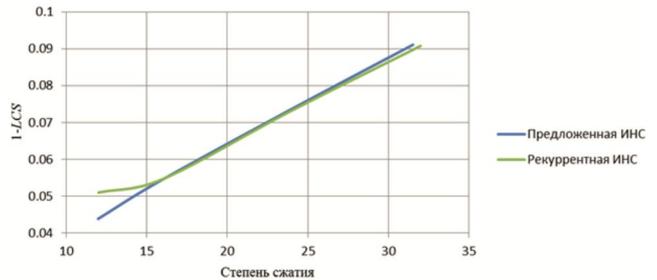


Рис. 14. Зависимость интегрального показателя сходства яркостей, контраста и структуры от степени сжатия тестового фрагмента мультиспектрального изображения

Литература

1. Косткин И.В. Алгоритм вейвлет-сжатия неподвижных цифровых изображений. Вестник Рязанского государственного радиотехнического университета. 2007. № 20. С. 114-117.
2. Васильев А.М. Алгоритм сжатия изображений на основе модифицированного дискретного косинусного преобразования. Вестник Рязанского государственного радиотехнического университета. 2013. № 46-2. С. 16-21.
3. Sahoolizadeh H., Suratgar A.A. Adaptive image compression using neural networks. 5th International Conference: Science of Electronics, Technologies of Information and Telecommunications. 2009.
4. Toderici G., Vincent D., Johnston N., Hwang S.J., Minnen D., Shor J., Covell M. Full resolution image compression with recurrent neural networks. IEEE Conference on Computer Vision and Pattern Recognition. 2017.
5. Islam K., Dang M., Lee S., Moon H. Image compression with recurrent neural networks and generalized divisive normalization. IEEE Conference on Computer Vision and

Pattern Recognition. 2021.

6. Theis L., Shi W., Cunningham A., Huszar F. Lossy image compression with compressive autoencoders. International Conference on Learning Representations. 2017.
7. Mahoney M. Large Text Compression Benchmark. URL: <http://www.mattmahoney.net/dc/text.html> (дата обращения 01.02.2023).
8. Goyal M., Tatwawadi K., Chandak S., Ochoa I. DeepZip: Lossless data compression using recurrent neural networks. Data Compression Conference. 2019.
9. Ballard F. Lossless Data Compression with Neural Networks. URL: <https://ballard.org/nncp/nncp.pdf> (дата обращения 01.02.2023).
10. Ioffe S., Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. Proceedings of the 32nd International Conference on International Conference on Machine Learning. 2015, vol. 37, pp. 448-456.
11. Wang Z., Bovik A.C., Sheikh H.R., Simoncelli E.P. Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing. 2004.

УДК 004.891.2

ОБНАРУЖЕНИЕ ПОЛИПОВ НА КОЛОНОСКОПИЧЕСКИХ ИЗОБРАЖЕНИЯХ ПРИ ПОМОЩИ АЛГОРИТМОВ НА БАЗЕ НЕЙРОСЕТЕВОЙ АРХИТЕКТУРЫ YOLO

Хрящев В.В., к.т.н., доцент кафедры цифровых технологий и машинного обучения Ярославского государственного университета им. П.Г. Демидова, e-mail: v.khryashchev@uniyar.ac.ru

Приоров А.Л., д.т.н., профессор кафедры цифровых технологий и машинного обучения Ярославского государственного университета им. П.Г. Демидова, e-mail: pri@uniyar.ac.ru

Лебедев А.А., к.т.н., ведущий инженер центра искусственного интеллекта и цифровой экономики Ярославского государственного университета им. П.Г. Демидова, e-mail: lebedevdes@gmail.com

Котов Н.А., аспирант кафедры цифровых технологий и машинного обучения Ярославского государственного университета им. П.Г. Демидова, e-mail: n.kotov2@uniyar.ac.ru

DETECTION OF POLYPS ON COLONOSCOPIC IMAGES USING ALGORITHMS BASED ON THE YOLO NEURAL NETWORK ARCHITECTURE

Khryashchev V.V., Priorov A.L., Lebedev A.A., Kotov N.A.

Neural network algorithms for detecting colon polyps in images obtained during colonoscopic examination is investigated. The Kvasir-SEG image database was used to train and test deep machine learning algorithms. YOLOv6, YOLOR, YOLOv7, YOLOv7X, and YOLOv8 networks previously trained on the basis of MS COCO images were used as neural network architectures. Due to the small volume of images in the Kvasir-SEG database, data augmentation was actively used. As a result of the application of detection algorithms to the test set of endoscopic images, the highest values of the metrics $AP@[0.25..0.75]$ – equal to 98.4, and $AP@0.50$ – equal to 98.6, were obtained for a neural network detector based on the YOLOv8 network. The obtained results can be used in the development of a video stream analysis system in an endoscopic system operating in real time, during colonoscopic examinations.

Key words: endoscopy, colonoscopic images, intestinal polyps, neural network detection algorithms, deep learning, YOLO neural network architecture.

Ключевые слова: эндоскопия, колоноскопические изображения, полипы кишечника, алгоритмы нейросетевого детектирования, глубокое обучение, нейросетевая архитектура YOLO.

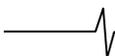
Введение

Алгоритмы и методы анализа видеоизображений в медицине находят применение в такой перспективной области диагностики, как эндоскопические исследования [1-3]. Одной из эндоскопических процедур является колоноскопия, представляющая собой дорогостоящую и достаточно трудоемкую инвазивную процедуру, требующую высокой квалификации врача, проводящего исследование.

Одним из маркеров патологий являются полипы толстой кишки. Фактически это аномальные разрастания ткани, выступающие над слизистой оболочкой. Они часто являются предвестниками достаточно распространенного колоректального рака. Полипы могут возникать в любом месте желудочно-кишечного тракта (ЖКТ), однако чаще всего их обнаруживают в колоректальной области. В толстом кишечнике выделяют несколько видов классификации полипов: по количеству, по патоморфологическим данным, по особенностям прикрепления к стенке кишки [1, 2]. Размер колоректальных полипов может существенно различаться: от совсем небольших (≤ 5 мм) до весьма значительных (≥ 10 мм).

Исследовано использование нейросетевых алгоритмов детектирования полипов толстой кишки на изображениях, полученных при колоноскопическом обследовании. Для обучения и тестирования алгоритмов глубокого машинного обучения использовалась база изображений Kvasir-SEG. В качестве нейросетевых архитектур использовались сети YOLOv6, YOLOR, YOLOv7, YOLOv7X, YOLOv8, предварительно обученные на базе изображений MS COCO. В связи с малым объемом изображений в базе Kvasir-SEG активно использовалась аугментация данных. В результате применения алгоритмов детектирования к тестовому набору эндоскопических изображений получены наивысшие значения метрик $AP@[0.25..0.75]$ – равное 98,4; и $AP@0.50$ – равное 98,6; для нейросетевого детектора на базе сети YOLOv8. Полученные результаты могут использоваться при разработке системы анализа видеопотока в эндоскопической системе, работающей в режиме реального времени, при проведении колоноскопических исследований.

Статистика эндоскопических исследований показывает, что уровень обнаружения полипов врачом высокой квалификации составляет примерно 80 % [3-5]. Новые эндоскопические устройства и диагностические инструменты улучшили точность обнаружения полипов, но проблема их пропуска остается по-прежнему актуальной [6-9]. Таким образом, задача разработки автоматических методов обнаружения полипов на видеопотоке с эндоскопа, работающих в режиме реального времени и дающих «второе мнение» для врачей, проводящих диагностику, является актуальной на данном этапе развития медицинских приборов.



Для решения этой задачи использовались классические методы машинного обучения, основанные на анализе цвета и текстуры полипов [10]. Позднее их заменили методы глубокого машинного обучения, основанные на сверточных нейронных сетях (СНС) [11-14], с автоматическим формированием признаков во время процедуры обучения. Их использование стало возможным при появлении общедоступных наборов размеченных эндоскопических изображений [15-17].

Обобщенная структурная схема разработки нейросетевого алгоритма для решения задачи обнаружения полипов на эндоскопических изображениях кишечника приведена на рис. 1. Кроме самой эндоскопической системы она включает размеченную медицинскими специалистами базу изображений, полученных с эндоскопической системы в виде отдельных кадров соответствующих видеопоследовательностей. Также в схеме используются стандартные или модифицированные алгоритмы аугментации, значительно расширяющие исходную базу изображений. Важной составляющей является также выбор нейросетевой архитектуры и процедура обучения соответствующего нейросетевого алгоритма. После выполнения указанных этапов на выходе алгоритма получаем прямоугольную область, соответствующую области нахождения полипа на исследуемом эндоскопическом изображении.



Рис. 1. Обобщенная структурная схема разработки нейросетевого алгоритма обнаружения полипов кишечника

Целью настоящей работы является исследование эффективности применения алгоритмов обнаружения

полипов при проведении колоноскопических исследований на базе современной нейросетевой архитектуры YOLO [18-20].

Обзор баз колоноскопических видеоизображений

Рассмотрим существующие открытые базы колоноскопических изображений и видеопоследовательностей с разметкой данных:

- Kvasir-SEG [21] – содержит 1000 изображений толстой кишки с экспертной разметкой полипов для задач обнаружения и сегментации;
- Kvasir [22] – содержит 8000 изображений всего ЖКТ с разметкой нескольких классов объектов и патологий;
- Nerthus [23] – содержит 21 видеопоследовательность исследования толстой кишки с разметкой классов: очищенный кишечник/не очищенный кишечник;
- HyperKvasir [24] – содержит 110079 изображений и 373 видеопоследовательности всего ЖКТ с разметкой 40 различных классов;
- CVC-Clinic [25] – содержит 612 изображений толстой кишки с разметкой полипов для задачи сегментации;
- Kvasir-Instrument [26] – содержит 590 изображений толстой кишки с разметкой медицинских инструментов для задач обнаружения и сегментации изображений.

Для проведения настоящего исследования выбрана база Kvasir-SEG, содержащая 1072 полипа и предлагающая одновременно ограничивающие прямоугольные рамки и бинарные маски для сегментации. В рассматриваемом наборе данных разрешение изображений варьируется от 332×487 до 1920×1072 пикселей. Такое число обучающих данных недостаточно для современных нейросетевых методов обнаружения объектов, поэтому здесь активно используется процедура аугментации (размножения, обогащения) данных, которая производится с учетом специфики получения эндоскопических изображений [8-9].

Архитектура сверточной нейронной сети YOLO

В рамках исследования выбраны последние версии детекторов семейства YOLO, а именно: YOLOv6, YOLOR, YOLOv7, YOLOv7X, YOLOv8. Важно отметить, что каждая последующая архитектура содержит различные модификации и самой архитектуры, и методов процедуры обучения. Рассмотрим общую архитектуру такой нейронной сети на примере YOLOv7. Процесс обработки эндоскопического изображения на основе алгоритма, использующего эту архитектуру, представлен на рис. 2.

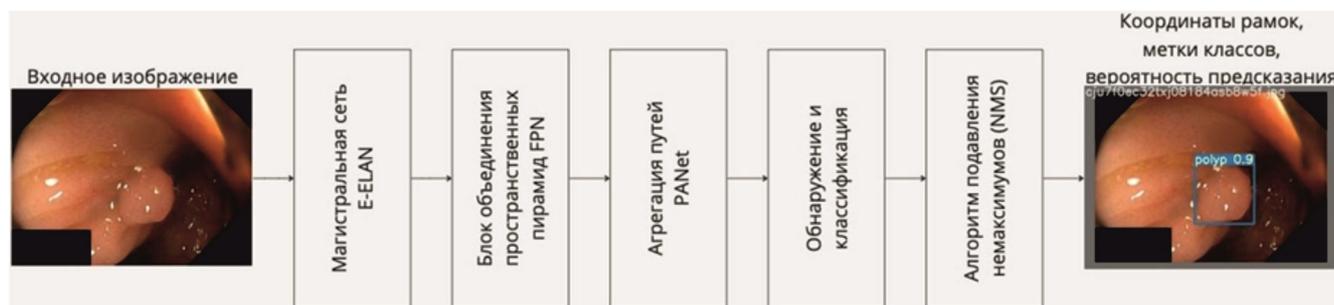


Рис. 2. Архитектура сверточной нейронной сети YOLO

Нейронная сеть семейства YOLO состоит из трех основных частей – магистральной, объединения пространственных пирамид и предсказания [19].

В качестве магистральной части сети применяется нейронная сеть с использованием блоков E-ELAN, служащая для извлечения признаков из входного изображения. Этот блок контролирует кратчайший и самый длинный градиентный пути для лучшей сходимости модели, используя элементы расширения, перетасовки и слияния. В нем входные данные подаются непосредственно через блок свертки 1×1 , в то время как два других подключения к блоку получены путем свертки входных данных двумя блоками свертки 3×3 с тем же канальным множителем. Далее все функции объединяются, и для изучения обширной информации вновь применяется свертка 1×1 .

К магистральной части добавляется часть объединения пространственных пирамид, выделяющая наиболее важные характерные особенности с разных уровней магистралей через сверточные слои CBS и SPPFCSPC. В конечных блоках пирамиды признаков свертки заменены блоками RepConv, необходимыми для обновления параметров модели.

Далее следуют слои обнаружения и классификации, дающие предсказание модели. Также используется механизм назначения меток, использующий главную и вспомогательные ветки предсказания, а также программный определитель меток. Тем самым на этапе тренировки модели используются признаки в промежуточных слоях, улучшающие итоговую предсказательную способность модели.

Сравнительный анализ нейросетевых алгоритмов обнаружения полипов

Для алгоритмов на основе архитектур YOLO использовались следующие параметры обучения: размер батча (последовательности изображений) составлял 16 изображений, в качестве алгоритма оптимизации использовался алгоритм Adam (Adaptive Moment Estimation), сочетающий

в себе идеи накопления движения и более слабого обновления весов для типичных признаков. Начальная скорость обучения -10^{-3} , общее число эпох обучения – 250. Другие параметры данного алгоритма выбирались следующим образом: $\beta_1 = 0,9$; $\beta_2 = 0,999$; $\varepsilon = 10^{-8}$.

Во время обучения использовались функции потерь с бинарной кросс-энтропией для сетей YOLOv6, YOLOv7 и YOLOR, а также функция потерь, основанная на индексе Жаккара (IoU), – для сети YOLOv8.

Для проведения сравнения с алгоритмами, приведенными в [9, 27], база Kvasir-SEG разделялась на тестовый и обучающий наборы (120 – тестовый и 880 – обучающий). Для обучения и тестирования нейронных сетей, координаты ограничивающих рамок в файлах разметки базы пересчитывались из формата Pascal-VOC в формат YOLO.

Для оценивания качества работы алгоритмов в [27] предложено использовать метрики средней точности, рассчитанные при значениях порога IoU 0,25, 0,50 и 0,75. Далее данные метрики обозначены как $AP@0,25$, $AP@0,50$ и $AP@0,75$ соответственно. Кроме того, для комплексной оценки по множеству порогов IoU предложена метрика $AP@[0,25..0,75]$, которая рассчитывается как среднее 11-ти значений AP , рассчитанных при 11 различных порогах IoU в пределах от 0,25 до 0,75 с шагом 0,05.

В табл. 1 приведены результаты тестирования алгоритмов для всех указанных нейросетевых архитектур на базе эндоскопических изображений Kvasir-SEG.

На рис. 3 представлены графики зависимости точности (P) от полноты (R) на тестовых изображениях для исследованных в работе нейросетевых алгоритмов детектирования. Видно, что лучшие результаты по обнаружению полипов показал алгоритм на основе сети YOLOv8. Значение $AP@0,50$ для него достигает 98,6; а усредненная по диапазону порогов метрика $AP@[0,25..0,75]$ – 98,4; что говорит о высокой способности детектирования полипов в широком диапазоне порогов детектирования.

Таблица 1. Тестирование нейросетевых алгоритмов обнаружения полипов

| Метрика | $AP@[0,25..0,75]$ | $AP@0,25$ | $AP@0,5$ | $AP@0,75$ |
|---|-------------------|-------------|-------------|-------------|
| Известные аналоги из [27] | | | | |
| EfficientDet-D0 | 47,6 | 68,5 | 50,5 | 22,8 |
| Faster R-CNN | 78,7 | 89,5 | 84,2 | 56,6 |
| RetinaNet50 | 87,0 | 93,9 | 90,9 | 69,7 |
| RetinaNet101 | 87,4 | 94,8 | 90,9 | 71,3 |
| YOLOv3+spp | 81,0 | 88,6 | 85,3 | 75,9 |
| YOLOv4 | 85,1 | 91,2 | 82,3 | 75,9 |
| ColonSegNet | 80,0 | 90,0 | 81,7 | 67,1 |
| Предложенные алгоритмы на базе YOLOv5 из [9] | | | | |
| HAO _{SV512} | 90,2 | 96,6 | 93,8 | 74,9 |
| HAO _{YS} | 88,6 | 94,0 | 90,8 | 73,8 |
| HAO _{YM} | 91,8 | 95,9 | 93,3 | 82,8 |
| HAO _{YL} | 92,5 | 96,7 | 93,9 | 85,1 |
| Алгоритмы, исследованные в настоящей работе | | | | |
| YOLOv6 | 89,2 | 93 | 90 | 84,1 |
| YOLOR | 87,8 | 93,2 | 88,2 | 79,3 |
| YOLOv7 | 80,6 | 89,5 | 80,6 | 66,4 |
| YOLOv7X | 80,9 | 88,7 | 80,9 | 70 |
| YOLOv8 | 98,4 | 98,7 | 98,6 | 97,6 |

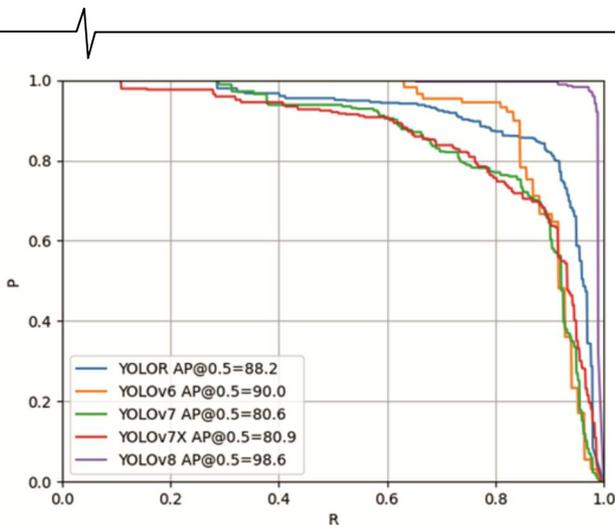


Рис. 3. Кривые на плоскости точность-полнота для исследованных нейросетевых алгоритмов

Можно отметить, что тестирование сетей YOLOv7, YOLOv7X и YOLOR показало значение метрик $AP@0,50$ меньше, чем для YOLOv6, несмотря на обратные показатели по данной метрике на больших эталонных базах. Это связано с большим числом параметров сети относительно количества обучающих примеров в базе Kvasir-SEG, несмотря на применение методов аугментации во время процесса обучения.

Таким образом, установлено, что алгоритм на базе нейронной сети YOLOv8 достигает наилучшего результата по метрикам средней точности среди всех рассматриваемых алгоритмов обнаружения полипов, причем это преимущество особенно существенно по метрике $AP@0,75$.

Заключение

Результаты проведенных исследований нейросетевых алгоритмов детектирования полипов кишечника позволяют сформулировать следующие выводы и закономерности:

- в настоящее время наблюдается существенное отставание размера имеющихся баз эндоскопических изображений от требований современных нейросетевых алгоритмов и методов глубокого машинного обучения;

- в результате тестирования нейросетевых алгоритмов обнаружения полипов на изображениях из базы Kvasir-SEG наилучший результат получен для нейросетевого алгоритма YOLOv8, для которого значение метрик $AP@[0,25..0,75]$ равно 98,4 и $AP@0,50$ – 98,6;

- при сравнении результатов работы нейросетевых алгоритмов YOLOv8 и YOLOv5 оказывается, что для первых они лучше по метрикам $AP@[0,25..0,75]$ – на 5,9 – 9,8 %; $AP@0,25$ – на 2 – 4,7 %; $AP@0,50$ – на 4,7 – 7,8 % и $AP@0,75$ – на 12,5 – 23,8 %;

- исследования, проведенные в работе, будут использованы для построения на основе нейросетевого алгоритма обнаружения полипов модуля анализа видеопотока в реальной эндоскопической системе.

Литература

1. Палевская С.А., Короткевич А.Г. Эндоскопия желудочно-кишечного тракта. М.: ГЭОТАР-Медиа, 2020. 752 с.

2. Никонов Е.Л., Аксенов В.А., Кашин С.В., Нехайкова Н.В. Международный опыт скрининга колоректального рака. Доказательная гастроэнтерология. 2017. Т. 6. № 3. С. 30-35.

3. Lee J. Resection of diminutive and small colorectal polyps: what is the optimal technique? Clinical endoscopy. 2016, vol. 49 (4). 355 p.

4. Lebedev A., Khryashchev V., Stefanidi A., Stepanova O., Kashin S., Kuvaev R. Convolutional neural network for early detection of gastric cancer by endoscopic video analysis. Proc. SPIE 11433. Twelfth International Conference on Machine Vision (ICMV 2019).

5. Батухтин Д.М., Пеганова Е.В., Митракова Н.Н. 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. Коваленко Д.А., Гнатюк В.С. Ассоциация сцен в эндоскопических видео. GraphiCon 2017: Обработка и анализ биомедицинских изображений. Пермь. 2017. 269-274 с.

7. Matyja M., Pasternak A., Szura M., Wysocki M., Pędzwiatr M., Rembiasz K. How to improve the adenoma detection rate in colorectal cancer screening? Clinical factors and technological advancements. Archives of medical science: AMS. 2019, vol. 15 (2). 424 p.

8. Лебедев А.А., Хрящев В.В., Среднякова А.С., Казина Е.М. Разработка алгоритма детектирования полипов на эндоскопических изображениях с использованием сверточных нейронных сетей. Цифровая обработка сигналов. 2021. № 2. 55-60 С.

9. Лебедев А. Исследование нейросетевых алгоритмов обнаружения объектов на видеоизображениях в медицинских системах прикладного телевидения: диссертация на соискание ученой степени кандидата технических наук. Ярославль. 2022.

10. Karkanis S.A., Iakovidis D.K., Maroulis D.E., Karras D.A., Tzivras M. Computer-aided tumor detection in endoscopic video using color wavelet features. IEEE Trans. on Information Technology in Biomedicine. 2003, vol. 7 (3), pp. 141-152.

11. Николенко С.И., Кадурич А.А., Архангельская Е.О. Глубокое обучение. Питер: СПб. 2018. 480 с.

12. Tajbakhsh N., Shin J.Y., Gurudu S., Hurst R., Kendall C. et al. Convolutional neural networks for medical image analysis: Full training or fine tuning? IEEE transactions on medical imaging. 2016, vol. 35(5), pp. 1299-1312.

13. Shin H.-C., Roth H., Gao M., Lu L., Xu Z. et al. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. IEEE Trans. on Medical Imaging. 2016, vol. 35 (5), pp. 1285-1298.

14. Wang Y., Tavanapong W., Wong J., Oh J.H., De Groen P.C. Polyp-alert: Near real-time feedback during colonoscopy. Computer Methods and Programs in Biomedicine. 2015, vol. 120 (3), pp. 164-179.

15. Shin Y., Qadir H.A., Aabakken L., Bergsland J., Balasingham I. Automatic Colon Polyp Detection Using Region Based Deep CNN and Post Learning Approaches. IEEE Access. 2018, vol. 6, pp. 40950-40962.

16. Lee J., Jeong J., Song E., Ha C., Lee H. et al. Real-time detection of colon polyps during colonoscopy using deep learning: systematic validation with four independent datasets. *Scientific Reports*. 2020, vol. 10 (1), pp. 1-9.
17. Yamada M., Saito Y., Imaoka H., Saiko M., Yamada S. et al. Development of a real-time endoscopic image diagnosis support system using deep learning technology in colonoscopy. *Scientific reports*. 2019, vol. 9 (1), pp. 1-9.
18. Li C., Li L., Jiang H., Weng K., Geng Y. et al. YOLOv6: A single-stage object detection framework for industrial applications. *arXiv preprint arXiv:2209.02976*, 2022.
19. Wang C.-Y., Bochkovskiy A., Liao H.-Y. M. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv preprint arXiv:2207.02696*, 2022.
20. Jocher G., Chaurasia A., Qiu J. YOLO by Ultralytics. <https://github.com/ultralytics/ultralytics>, 2023. Accessed: February 30, 2023.
21. Jha D., Smedsrud P., Riegler M., Halvorsen P., Lange T. et al. Kvasir-seg: A Segmented Polyp Dataset. *Proc. of International Conference on Multimedia Modeling*. 2020, pp. 451-462.
22. Pogorelov K., Randel K.R., Griwodz C., Eskeland S.L., Lange T. et al. Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection. *Proc. of the 8th ACM on Multimedia Systems Conference*. 2017, pp. 164-169.
23. Pogorelov K., Randel K.R., Lange T., Eskeland S.L., Griwodz C. et al. Nerthus: A bowel preparation quality video dataset. *Proc. of the ACM on Multimedia Systems Conference*. 2017, pp. 170-174.
24. Borgli H., Thambawita V., Smedsrud P.H., Hicks S., Jha D. et al. Hyperkvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy. *Scientific Data*. 2020, vol. 7 (1), pp. 1-14.
25. Bernal J., Sánchez F.J., Fernández-Esparrach G., Gil D., Rodríguez C., Vilarriño F. Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. *Computerized Medical Imaging and Graphics*. 2015, vol. 43, pp. 99-111.
26. Jha D., Ali S., Emanuelsen K., Hicks S.A., Thambawita V. et al. Kvasir-Instrument: Diagnostic and therapeutic tool segmentation dataset in gastrointestinal endoscopy. *International Conference on Multimedia Modeling*. 2021, pp. 218-229.
27. Jha D., Ali S., Tomar N.K., Johansen H.D., Johansen D. et al. Real-Time Polyp Detection, Localization and Segmentation in Colonoscopy Using Deep Learning. *Computer Vision and Pattern Recognition*. 2021. pp. 40496-40510.

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 200 экз.

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. Журнал «Цифровая обработка сигналов» включен в базу данных Web of Science - Russian Science Citation Index.

Планируемые сроки издания отдельных номеров журнала:

- № 3 сентябрь 2023 г. Тематический выпуск: «Цифровая обработка изображений».
- № 4 декабрь 2023 г. Тематический выпуск: «ЦОС в радиотехнике и системах телекоммуникаций».
- № 1 март 2024 г. Тематический выпуск: «ЦОС в инфокоммуникационных системах».
- № 2 июнь 2024 г. Тематический выпуск по материалам 26-й Международной научно-технической конференции «Цифровая обработка сигналов и ее применение-DSPA».

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

КОМПЛЕКСНАЯ НЕЙРОННАЯ СЕТЬ

Кузнецов В.П., к.т.н., доцент РГРТУ им В.Ф. Уткина, e-mail: kuznetsovaitu@yandex.ru

COMPLEX NEURAL NETWORK

Kuznetsov V.P.

The conception of complex neural networks construction on example of a two-layer feed-forward network is proposed. The network contains the complex weights and the complex activation functions as a complex variable function. Complex neural network is meant for complex signals processing, which use widely for telecommunication systems analysis. Such neural networks may be used for identification, pre-distortion, prediction, control. The results of experimental research of complex neural network for identification of nonlinear power amplifier are presented.

Key words: neural network, weight coefficient, activation function, error signal, learning algorithm, target function, gradient, partial derivative.

Ключевые слова: нейронная сеть, весовой коэффициент, функция активации, сигнал ошибки, алгоритм обучения, целевая функция, градиент, частная производная.

Введение

Предметом данной работы является методика построения комплексных нейронных сетей, содержащих комплексные весовые коэффициенты и функции активации в виде функций комплексной переменной. Областью применения таких сетей являются системы телекоммуникаций, при анализе которых широкое использование находят аналитические сигналы, основанные на преобразовании Гильберта [1]. Аналитический сигнал является комплексным сигналом и должен обрабатываться системой как единое целое без разделения на вещественную и мнимую составляющие. Такая же задача ставится перед нейронной сетью, определяя ее комплексный характер. Одной из задач в области телекоммуникаций, которая может быть решена на основе нейронной сети, является идентификация нелинейной динамической системы. Известны подходы к решению этой задачи для вещественных сигналов на основе сетей прямого распространения и рекуррентных сетей [2]. В данной работе используется сеть прямого распространения (рис. 1). Особенность работы состоит в комплексном характере входной и выходной последовательностей нелинейной динамической системы.



Рис. 1. Идентификация нелинейной системы

Предлагается концепция построения комплексных нейронных сетей на примере двухслойной сети прямого распространения. Нейронная сеть содержит комплексные весовые коэффициенты и функции активации в виде функций комплексной переменной. Комплексная нейронная сеть предназначена для обработки комплексных сигналов, которые находят широкое применение при анализе телекоммуникационных систем с различными видами модуляции. Такие нейронные сети могут использоваться для решения задач идентификации, предсказания, прогнозирования, управления. Приводятся результаты экспериментального исследования комплексной нейронной сети в задаче идентификации нелинейного усилителя мощности.

Структура нейронной сети

Решение задачи предлагается искать на основе двухслойной нейронной сети прямого распространения (рис. 2). В первом слое стоят нелинейные нейроны с комплексными весовыми коэффициентами и функцией активации в виде функции комплексной переменной функции активации из множества функций комплексной переменной $f(q)$. Выходной слой представлен одним линейным нейроном. Основная проблема построения комплексной нейронной сети состоит в выборе функции активации из множества функций комплексной переменной. Широкое применение в вещественных сетях находит биполярная логистическая функция, которая, сохраняя знак входного сигнала, преобразует его в диапазон $(-1, 1)$. В данной работе в качестве функции активации предлагается использовать такую функцию комплексной переменной, которая, сохраняя фазу входного сигнала, преобразует его модуль внутри окружности единичного радиуса с центром в начале координат комплексной плоскости. Комплексную функцию активации с такими свойствами можно сконструировать на основе известных вещественных функций активации. Это, во-первых, алгебраическая сигмоида [3]: $y = x / \sqrt{1 + x^2}$, где x и y – вещественные переменные. Создадим на ее основе функцию комплексной переменной $z = \frac{q}{\sqrt{1 + |q|^2}} =$

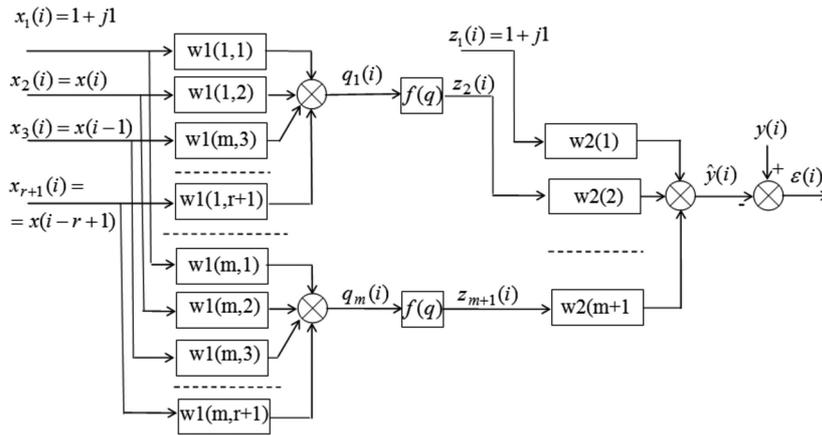


Рис. 2. Структурная схема нейронной сети

$$= \frac{q_u + jq_v}{\sqrt{1+q_u^2+q_v^2}} = \frac{q_u}{\sqrt{1+q_u^2+q_v^2}} + j \frac{q_v}{\sqrt{1+q_u^2+q_v^2}} = z_u + jz_v.$$

Во-вторых, можно взять за основу функцию SoftSigm [4]: $y = x / (1 + |x|)$ и превратить ее в функцию комплексной переменной

$$z = \frac{q}{1+|q|} = \frac{q_u + jq_v}{1 + \sqrt{q_u^2 + q_v^2}} = \frac{q_u}{1 + \sqrt{q_u^2 + q_v^2}} + j \frac{q_v}{1 + \sqrt{q_u^2 + q_v^2}} = z_u + jz_v.$$

Экспериментальные исследования показали, что предложенные функции активации сопровождаются примерно одинаковой эффективностью обучения. Однако функция SoftSigm приводит к меньшему объему вычислений на этапе нахождения ее частных производных.

Выходной сигнал нейронной сети $\hat{y}(i)$ в настоящий момент времени t_i формируется на основе входного отсчета $x(i)$ и $(r-1)$ входных отсчетов (задержек) в предыдущие моменты времени. Таким образом, выходной сигнал сети описывается функцией регрессии $\hat{y}(i) = \varphi(x(i), x(i-1), x(i-2), \dots, x(i-r+1))$, причем функция φ конструируется структурой и параметрами сети.

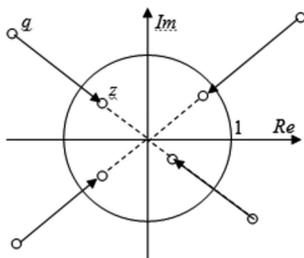


Рис. 3. Преобразование сигнала функцией активации

Алгоритм обучения

При построении нейронной сети экспериментальный материал, которым располагает исследователь, делится на обучающую и тестовую части. Из первой части формируется множество обучающих примеров, которое многократно предъявляется нейронной сети. После обучения нейронная сеть проверяется на тестовом множестве примеров.

В работе используется пакетный режим обучения, основанный на минимизации суммарного квадрата модуля ошибки $\varepsilon(i) = y(i) - \hat{y}(i) = \varepsilon_u(i) + j\varepsilon_v(i)$:

$$E = \frac{1}{2} \sum_{i=1}^N |\varepsilon(i)|^2 = \frac{1}{2} \sum_{i=1}^N (\varepsilon_u^2(i) + \varepsilon_v^2(i)) \rightarrow \min_{W_u, W_v}.$$

Здесь: $\varepsilon_u, \varepsilon_v$ – вещественная и мнимая составляющие сигнала ошибки, N – число примеров обучения в пакете, W_u, W_v – векторы вещественных и мнимых частей весовых коэффициентов. Таким образом, целевая функция рассматривается как вещественная функция многих переменных, в качестве которых выступают вещественные и мнимые части весовых коэффициентов первого и второго слоев.

Минимизацию целевой функции проведем методом градиента:

$$W(k+1) = W(k) - h \cdot \text{grad } E(k), \quad k = 0, 1, 2, \dots$$

Здесь W – вектор всех аргументов целевой функции, k – номер эпохи, h – параметр обучения, grad – вектор-градиент частных производных целевой функции по каждому аргументу.

Рассмотрим обучение выходного нейрона. Найдем частные производные целевой функции по каждому весовому коэффициенту, представленному вещественной и мнимой частями. Частную производную по вещественной части коэффициента $W2(l), l = \overline{1, m+1}$, найдем следующим образом:

$$\frac{\partial E}{\partial W2_u(l)} = \sum_{i=1}^N \left[\varepsilon_u(i) \frac{\partial \varepsilon_u(i)}{\partial W2_u(l)} + \varepsilon_v(i) \frac{\partial \varepsilon_v(i)}{\partial W2_u(l)} \right], \quad l = \overline{1, m+1}.$$

Вещественная и мнимая части сигнала ошибки имеют вид:

$$\varepsilon_u(i) = y_u(i) - \hat{y}_u(i), \quad \varepsilon_v(i) = y_v(i) - \hat{y}_v(i).$$

Поэтому частная производная

$$\frac{\partial E}{\partial W2_u(l)} = - \sum_{i=1}^N \left[\varepsilon_u(i) \frac{\partial \hat{y}_u(i)}{\partial W2_u(l)} + \varepsilon_v(i) \frac{\partial \hat{y}_v(i)}{\partial W2_u(l)} \right], \quad l = \overline{1, m+1} \tag{1}$$

Выходной сигнал нейронной сети

$$\hat{y}(i) = \sum_{n=1}^{m+1} z_n(i) \cdot W2(n) =$$

$$= \sum_{n=1}^{m+1} [z_{n,u}(i) + jz_{n,v}(i)] \cdot [W2_u(n) + jW2_v(n)]. \quad (2)$$

После очевидных преобразований получим:

$$\hat{y}(i) = \sum_{n=1}^{m+1} \{ [z_{n,u}(i)W2_u(n) - z_{n,v}(i)W2_v(n)] + j[z_{n,u}(i)W2_v(n) + z_{n,v}(i)W2_u(n)] \}. \quad (3)$$

От весового коэффициента $W2(l)$ зависит только одно слагаемое в этой сумме при $n=l$. Поэтому частные производные

$$\frac{\partial \hat{y}_u(i)}{\partial W2_u(l)} = z_{l,u}(i), \quad \frac{\partial \hat{y}_v(i)}{\partial W2_u(l)} = z_{l,v}(i).$$

В итоге получим:

$$\frac{\partial E}{\partial W2_u(l)} = -\sum_{i=1}^N [\varepsilon_u(i)z_{l,u}(i) + \varepsilon_v(i)z_{l,v}(i)], \quad l = \overline{1, m+1}.$$

Частная производная по мнимой части коэффициента $W2(l)$, $l = \overline{1, m+1}$, вычисляется аналогично:

$$\frac{\partial E}{\partial W2_v(l)} = \sum_{i=1}^N \left[\varepsilon_u(i) \frac{\partial \varepsilon_u(i)}{\partial W2_v(l)} + \varepsilon_v(i) \frac{\partial \varepsilon_v(i)}{\partial W2_v(l)} \right], \quad l = \overline{1, m+1}.$$

Повторяя предыдущие преобразования, установим:

$$\frac{\partial E}{\partial W2_v(l)} = -\sum_{i=1}^N [-\varepsilon_u(i)z_{l,v}(i) + \varepsilon_v(i)z_{l,u}(i)], \quad l = \overline{1, m+1}.$$

Перейдем к обучению нейронов первого слоя. Для большей наглядности вычислим частные производные целевой функции по конкретному коэффициенту, например, $W1(1,1) = W1_u(1,1) + jW1_v(1,1)$, а затем обобщим результаты на остальные коэффициенты первого слоя. Найдем частную производную по вещественной части этого коэффициента, отталкиваясь от соотношения (1):

$$\frac{\partial E}{\partial W1_u(1,1)} = -\sum_{i=1}^N \left[\varepsilon_u(i) \frac{\partial \hat{y}_u(i)}{\partial W1_u(1,1)} + \varepsilon_v(i) \frac{\partial \hat{y}_v(i)}{\partial W1_u(1,1)} \right]. \quad (4)$$

Дальнейшие рассуждения проведем в русле соотношений (2) и (3), а затем обратим внимание на то, что от коэффициента $W1(1,1)$ зависит только сигнал $z_2(i)$.

Поэтому частные производные

$$\frac{\partial \hat{y}_u(i)}{\partial W1_u(1,1)} = W2_u(2) \frac{\partial z_{2,u}(i)}{\partial W1_u(1,1)} - W2_v(2) \frac{\partial z_{2,v}(i)}{\partial W1_u(1,1)},$$

$$\frac{\partial \hat{y}_v(i)}{\partial W1_u(1,1)} = W2_v(2) \frac{\partial z_{2,u}(i)}{\partial W1_u(1,1)} + W2_u(2) \frac{\partial z_{2,v}(i)}{\partial W1_u(1,1)}.$$

Подставим этот результат в (4) и сгруппируем слагаемые следующим образом:

$$\frac{\partial E}{\partial W1_u(1,1)} = -\sum_{i=1}^N [\varepsilon_u(i)W2_u(2) + \varepsilon_v(i)W2_v(2)] \frac{\partial z_{2,u}(i)}{\partial W1_u(1,1)} - \sum_{i=1}^N [-\varepsilon_u(i)W2_v(2) + \varepsilon_v(i)W2_u(2)] \frac{\partial z_{2,v}(i)}{\partial W1_u(1,1)}. \quad (5)$$

Перейдем к вычислению производных от сигнала $z_2(i)$, который также представим в виде вещественной

и мнимой частей (аргумент i опустим):

$$z_2 = f(q_1) = f(q_{1,u} + jq_{1,v}) = f_u(q_{1,u}, q_{1,v}) + j f_v(q_{1,u}, q_{1,v}) = z_{2,u} + jz_{2,v}.$$

Рассмотрим производные

$$\frac{\partial z_{2,u}}{\partial W1_u(1,1)} = \frac{\partial f_u}{\partial q_{1,u}} \cdot \frac{\partial q_{1,u}}{\partial W1_u(1,1)} + \frac{\partial f_u}{\partial q_{1,v}} \cdot \frac{\partial q_{1,v}}{\partial W1_u(1,1)},$$

$$\frac{\partial z_{2,v}}{\partial W1_u(1,1)} = \frac{\partial f_v}{\partial q_{1,u}} \cdot \frac{\partial q_{1,u}}{\partial W1_u(1,1)} + \frac{\partial f_v}{\partial q_{1,v}} \cdot \frac{\partial q_{1,v}}{\partial W1_u(1,1)}. \quad (6)$$

Здесь первые множители зависят от конкретной функции активации, а вторые обусловлены линейной структурой сигнала q_1 :

$$q_1 = \sum_{n=1}^{r+1} x_n W1(1, n) = \sum_{n=1}^{r+1} (x_{n,u} + jx_{n,v}) [W1_u(1, n) + jW1_v(1, n)].$$

Представим сигнал q_1 в виде вещественной и мнимой частей:

$$q_1 = \sum_{n=1}^{r+1} [x_{n,u} W1_u(1, n) - x_{n,v} W1_v(1, n)] + j \sum_{n=1}^{r+1} [x_{n,u} W1_v(1, n) + x_{n,v} W1_u(1, n)].$$

От коэффициента $W1(1,1)$ зависят только первые слагаемые в каждой сумме, поэтому частные производные

$$\frac{\partial q_{1,u}}{\partial W1_u(1,1)} = x_{1,u}, \quad \frac{\partial q_{1,v}}{\partial W1_u(1,1)} = x_{1,v}. \quad (7)$$

С учетом этого результата выражения (6) примут более компактный вид:

$$\frac{\partial z_{2,u}}{\partial W1_u(1,1)} = \frac{\partial f_u}{\partial q_{1,u}} x_{1,u} + \frac{\partial f_u}{\partial q_{1,v}} x_{1,v},$$

$$\frac{\partial z_{2,v}}{\partial W1_u(1,1)} = \frac{\partial f_v}{\partial q_{1,u}} x_{1,u} + \frac{\partial f_v}{\partial q_{1,v}} x_{1,v}.$$

Теперь частная производная (5) может быть представлена в окончательном виде:

$$\frac{\partial E}{\partial W1_u(1,1)} = -\sum_{i=1}^N [\varepsilon_u(i)W2_u(2) + \varepsilon_v(i)W2_v(2)] \times \left[\frac{\partial f_u}{\partial q_{1,u}} x_{1,u}(i) + \frac{\partial f_u}{\partial q_{1,v}} x_{1,v}(i) \right] - \sum_{i=1}^N [-\varepsilon_u(i)W2_v(2) + \varepsilon_v(i)W2_u(2)] \times \left[\frac{\partial f_v}{\partial q_{1,u}} x_{1,u}(i) + \frac{\partial f_v}{\partial q_{1,v}} x_{1,v}(i) \right].$$

Обобщим полученный результат на частные производные по вещественной части каждого весового коэффициента первого слоя:

$$\frac{\partial E}{\partial W1_u(l, n)} = -\sum_{i=1}^N [\varepsilon_u(i)W2_u(l+1) + \varepsilon_v(i)W2_v(l+1)] \times \left[\frac{\partial f_u}{\partial q_{l,u}} x_{n,u}(i) + \frac{\partial f_u}{\partial q_{l,v}} x_{n,v}(i) \right] - \sum_{i=1}^N [-\varepsilon_u(i)W2_v(l+1) + \varepsilon_v(i)W2_u(l+1)] \times \left[\frac{\partial f_v}{\partial q_{l,u}} x_{n,u}(i) + \frac{\partial f_v}{\partial q_{l,v}} x_{n,v}(i) \right]. \quad (8)$$

$$-\sum_{i=1}^N [-\varepsilon_u(i)W2_v(l+1) + \varepsilon_v(i)W2_u(l+1)] \times \left[\frac{\partial f_v}{\partial q_{l,u}} x_{n,u}(i) + \frac{\partial f_v}{\partial q_{l,v}} x_{n,v}(i) \right], \quad l = \overline{1, m}, \quad n = \overline{1, r+1}.$$

Частные производные по мнимой части каждого весового коэффициента первого слоя вычисляются аналогичным образом. В соотношениях (4), (5), (6) придется поменять только переменную дифференцирования $W1_u(1,1)$ на $W1_v(1,1)$, а соотношения (7) примут другой вид:

$$\frac{\partial q_{1,u}}{\partial W1_v(1,1)} = -x_{1,v}, \quad \frac{\partial q_{1,v}}{\partial W1_v(1,1)} = x_{1,u}.$$

В итоге, частные производные по мнимой части каждого весового коэффициента первого слоя вычисляются таким образом:

$$\begin{aligned} \frac{\partial E}{\partial W1_v(l,n)} &= -\sum_{i=1}^N [\varepsilon_u(i)W2_u(l+1) + \varepsilon_v(i)W2_v(l+1)] \times \\ &\times \left[\frac{\partial f_u}{\partial q_{l,u}} (-x_{n,v}(i)) + \frac{\partial f_u}{\partial q_{l,v}} x_{n,u}(i) \right] - \\ &- \sum_{i=1}^N [-\varepsilon_u(i)W2_v(l+1) + \varepsilon_v(i)W2_u(l+1)] \times \\ &\times \left[\frac{\partial f_v}{\partial q_{l,u}} (-x_{n,v}(i)) + \frac{\partial f_v}{\partial q_{l,v}} x_{n,u}(i) \right], \quad l = \overline{1, m}, \quad n = \overline{1, r+1}. \end{aligned} \quad (9)$$

Соотношения (8), (9) содержат частные производные функций активации. Рассмотрим вычисление этих производных для функции активации SoftSign, описываемой выражением:

$$z = \frac{q}{1+|q|} = \frac{q_u + jq_v}{1 + \sqrt{q_u^2 + q_v^2}} = \frac{q_u}{1 + \sqrt{q_u^2 + q_v^2}} + j \frac{q_v}{1 + \sqrt{q_u^2 + q_v^2}} = f_u(q_u, q_v) + j f_v(q_u, q_v).$$

Частные производные имеют вид:

$$\begin{aligned} \frac{\partial f_u}{\partial q_u} &= \frac{\sqrt{q_u^2 + q_v^2} + q_u^2}{(1 + \sqrt{q_u^2 + q_v^2})^2 \sqrt{q_u^2 + q_v^2}} = \frac{|q| + q_u^2}{(1 + |q|)^2 \cdot |q|}, \\ \frac{\partial f_v}{\partial q_v} &= \frac{\sqrt{q_u^2 + q_v^2} + q_v^2}{(1 + \sqrt{q_u^2 + q_v^2})^2 \sqrt{q_u^2 + q_v^2}} = \frac{|q| + q_v^2}{(1 + |q|)^2 \cdot |q|}, \\ \frac{\partial f_u}{\partial q_v} &= \frac{\partial f_v}{\partial q_u} = -\frac{q_u \cdot q_v}{(1 + \sqrt{q_u^2 + q_v^2})^2 \sqrt{q_u^2 + q_v^2}} = \frac{q_u \cdot q_v}{(1 + |q|)^2 \cdot |q|}. \end{aligned}$$

Экспериментальная часть

Предложенная концепция построения комплексной нейронной сети использована в задаче идентификации нелинейного усилителя мощности, входной и выходной сигналы которого имеют комплексный характер. На рис. 4 показаны фрагменты входной и выходной последовательностей исследуемого объекта длиной 500 комплексных отсчетов, приведенных в интервал (-1; 1), а также сигнал рассогласования. Сигналы разделены на вещественные и мнимые составляющие. На визуальном уровне отличие между входным и выходным сигналами

практически неразличимо, однако сигнал рассогласования вскрывает это различие; он располагается в диапазоне (-0,1; 0,1). В системах телекоммуникаций, использующих комплексные сигналы, отличие между ними принято оценивать показателем NMSE (Normalized Mean Square Error):

$$NMSE = 10 \lg \left(\frac{|e|^2}{|x|^2} \right).$$

В данном случае $NMSE = -21,5$ дБ.

Разработанное математическое обеспечение реализовано в виде программы, написанной на языке системы MATLAB. Для нейронной сети, представленной на рис. 2, есть два основных параметра, определяющих точность аппроксимации желаемых откликов. Это число отсчетов входного сигнала r и число нейронов первого слоя m . На этапе обучения к этим основным параметрам добавляются объем обучающей выборки, параметр обучения h и число эпох, необходимых для отыскания минимума целевой функции. Точность обучения будем характеризовать показателем $NMSE$ в виде:

$$NMSE = 10 \lg \left(\frac{|e|^2}{|y|^2} \right).$$

Сигнал рассогласования обусловлен нелинейными эффектами в каскадах усилителя (нечувствительность, насыщение) и внутренним шумом. При отсутствии шума задача идентификации сводится к задаче аппроксимации желаемых откликов. В задаче аппроксимации усложнение модели позволяет получить необходимую точность приближения для любого обучающего множества, представленного набором входов и желаемых откликов. Применительно к двухслойной нейронной сети, это утверждение базируется на теореме об универсальной аппроксимации [3]. При наличии шума задача идентификации сводится к задаче регрессионного анализа. В этой задаче усложнение модели за счет большего числа отсчетов входного сигнала r и большего числа нейро-

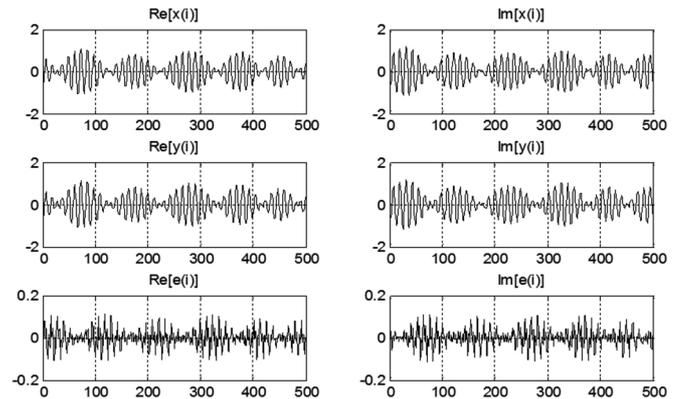


Рис. 4. Фрагменты вещественных и мнимых составляющих входного, выходного сигналов и сигнала рассогласования

нов первого слоя m приведет к некоторому точностному порогу, который не удастся преодолеть. Это объясняется тем, что при одном и том же входном векторе нейронной сети отклик объекта будет принимать разные значения, обусловленные шумом. Как только сигнал ошибки обучения начинает отождествляться с дискретным белым шумом, обучение останавливается.

Таблица 1. Результаты экспериментального исследования

| m | NMSE, дБ | | | | | | | |
|-----|----------|--------|----------|--------|----------|--------|----------|--------|
| | $r=4$ | | $r=6$ | | $r=8$ | | $r=10$ | |
| | Обучение | Тест | Обучение | Тест | Обучение | Тест | Обучение | Тест |
| 2 | -25,90 | -22,93 | -26,16 | -22,97 | -26,24 | -23,01 | -26,45 | -23,21 |
| 4 | -27,27 | -24,57 | -27,89 | -24,67 | -27,98 | -25,73 | -28,95 | -26,63 |
| 6 | -30,32 | -28,26 | -31,14 | -29,27 | -31,81 | -30,29 | -32,33 | -30,63 |
| 8 | -32,71 | -30,62 | -33,17 | -31,30 | -33,50 | -31,55 | -33,67 | -31,72 |
| 10 | -33,28 | -31,14 | -33,71 | -31,40 | -33,74 | -31,47 | -33,88 | -31,68 |
| 12 | -33,61 | -31,51 | -34,16 | -31,77 | -34,28 | -31,85 | -34,52 | -31,95 |
| 14 | -33,92 | -31,73 | -34,17 | -31,85 | -34,57 | -31,77 | -34,86 | -32,07 |
| 16 | -34,22 | -32,07 | -34,42 | -32,17 | -34,67 | -32,22 | -35,15 | -32,27 |
| 20 | -34,26 | -32,18 | -34,78 | -32,41 | -35,16 | -32,65 | -35,38 | -32,45 |
| 30 | -34,45 | -32,31 | -34,82 | -32,72 | -35,47 | -32,86 | -36,04 | -33,16 |

Результаты экспериментального исследования (лучшие реализации обучения) представлены в табл. 1. Для обучения нейронной сети использовалась 1000 комплексных отсчетов входного и выходного сигналов усилителя. Тестирование обученной сети проводилось на следующей 1000 отсчетов. Проведем анализ полученных результатов для процедуры обучения нейронной сети. Для большей наглядности эти результаты оформлены в виде графиков (рис. 5).

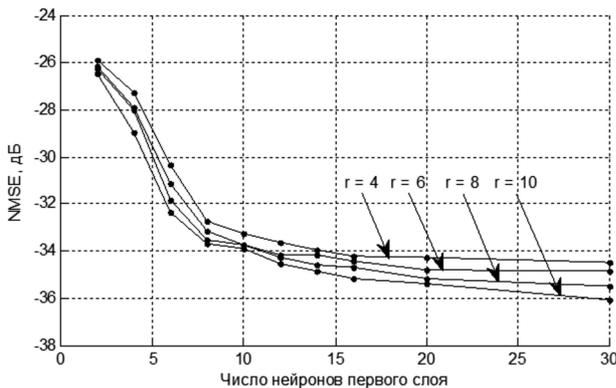


Рис. 5. Точность обучения в зависимости от числа нейронов первого слоя m при разном числе отсчетов входного сигнала r

Как и следовало ожидать, точность обучения нейронной сети повышается с увеличением числа нейронов первого слоя. Причем на начальном этапе увеличение числа нейронов с двух до восьми приводит к заметному повышению точности обучения. При дальнейшем увеличении числа нейронов эффективность обучения снижается.

Точность обучения в зависимости от числа отсчетов входного сигнала r увеличивается более равномерно (рис. 6).

На этапе тестирования нейронной сети предъявлялись новые отсчеты входного сигнала, которые не участвовали в обучении. Поэтому точность приближения выходного сигнала нейронной сети \hat{y} к выходному

сигналу усилителя y снизилась примерно на 10 %.

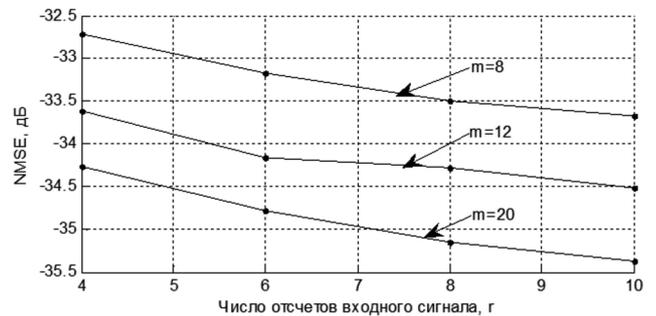


Рис. 6. Точность обучения в зависимости от числа отсчетов входного сигнала r при разном числе нейронов первого слоя m

Заключение

В данной работе предложена концепция построения комплексных нейронных сетей, которые могут найти применение при анализе и проектировании телекоммуникационных систем с комплексными сигналами. Эта концепция реализована на примере сравнительно простой двухслойной нейронной сети и показала свою работоспособность. Дальнейшие исследования в этой области могут быть направлены на разработку многослойных сетей прямого распространения и рекуррентных сетей. В плане практического применения интерес представляет задача предсказания, когда нейронная сеть ставится перед нелинейной системой с целью коррекции ее нелинейных эффектов.

Литература

1. Прокис Дж. Цифровая связь. М.: Радио и связь, 2000.
2. Осовский С. Нейронные сети для обработки информации. М.: Финансы и статистика, 2004.
3. Хайкин С. Нейронные сети: полный курс, 2-е издание. М.: Изд. дом «Вильямс», 2008.
4. Николенко С., Кадури А., Архангельская Е. Глубокое обучение. СПб.: Питер, 2018.

УДК 621.391.26

ДВУХЭТАПНЫЙ АЛГОРИТМ ПАРАМЕТРИЧЕСКОЙ ОПТИМИЗАЦИИ ВЕСОВОЙ ПРОСТРАНСТВЕННО-ВРЕМЕННОЙ РАНГОВОЙ ФИЛЬТРАЦИИ ИЗОБРАЖЕНИЙ

Смирнов А.А., к.т.н., доцент кафедры Радиоуправления и связи ФГБОУ ВО «РГРТУ», e-mail: a.l.an.smirnov@yandex.ru

Смирнов А.В., к.т.н., доцент, e-mail: ant.vas.smirnov@yandex.ru

TWO-STAGE ALGORITHM FOR PARAMETRIC OPTIMIZATION OF WEIGHTED SPACE-TIME RANK FILTERING OF IMAGES

Smirnov A.A., Smirnov A.V.

The problem of weighted space-time image filtering using rank statistics is considered. A two-stage algorithm for parametric optimization of weighted space-time image filtering is proposed, at the first stage of which dynamic programming methods were used, and at the second stage – effective local optimization algorithms of the quasi-Newtonian type were used. It is shown that the application of the proposed two-stage algorithm of weighted space-time image filtering provides high efficiency of pulse noise suppression.

Key words: parametric optimization, dynamic programming method, quasi-Newtonian methods, space-time image filtering.

Ключевые слова: параметрическая оптимизация, метод динамического программирования, квазиньютоновские методы, пространственно-временная фильтрация изображений.

Введение

Современные системы ТВ и ИК диапазона мониторинга Земли, воздушного и космического пространств должны обеспечивать обнаружение объектов, измерение их параметров, сопровождение, распознавание, управление в составе наземных и бортовых комплексов при наличии помех и шумов [1, 2].

Решение этих задач диктует необходимость совершенствования математических и программных средств пространственно-временной фильтрации изображений.

Пространственно-временная фильтрация изображений – совокупность операций по преобразованию изображения, позволяющая осуществить обнаружение искомого объекта среди других мешающих факторов при наличии помех и шумов, в том числе и преднамеренных [1, 2, 5].

С более общей точки зрения, пространственно-временная фильтрация изображений представляет собой важнейшую подготовительную операцию максимизации отношения сигнал/помеха при распознавании объекта, облегчающую принятие решения о наличии и положении искомого объекта в поле изображения [2, 3].

Основу современных методов пространственно-временной фильтрации изображений составляют эффективные нелинейные алгоритмы фильтрации при наличии помех и шумов. Алгоритмы нелинейной фильтрации изображений делятся на две большие группы: ранговые и морфологические [1, 4].

Целью работы является исследование эффективности применения численных методов оптимизации весовой ранговой пространственно-временной фильтрации изображений при наличии импульсных помех.

Рассмотрена задача весовой пространственно-временной фильтрации изображений с использованием ранговых статистик. Предложен двухэтапный алгоритм параметрической оптимизации весовой пространственно-временной фильтрации изображений, на первом этапе которого использовались методы динамического программирования, а на втором – эффективные алгоритмы локальной оптимизации квазиньютоновского типа. Показано, что применение предложенного двухэтапного алгоритма весовой пространственно-временной фильтрации изображений обеспечивает высокую эффективность подавления импульсного шума.

Постановка задачи

Пространственная модель изображения размера $N_1 \times N_2$ представляет собой двумерную дискретную функцию $f(x, y)$, координаты которой (x, y) принимают дискретные значения: $x = 0, \dots, N_1 - 1$, $y = 0, \dots, N_2 - 1$. При этом за начало координат принимается левый верхний угол изображения. Каждое дискретное значение $f(x, y)$ называется элементом изображения или пикселем.

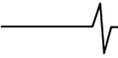
Изображения в процессе формирования и передачи по каналам связи обычно подвергаются воздействию различных помех и шумов. Для описания случайных воздействий используют модели аддитивного белого и импульсного шумов. Тогда линейная модель наблюдения изображения в условиях помех принимает вид [1, 2]:

$$g(x, y) = f(x, y) + n(x, y),$$

где $n(x, y)$ – аддитивный шум.

Целью восстановления исходного изображения $f(x, y)$ является получение из искаженного изображения $g(x, y)$, при помощи некоторой обработки, изображения $\hat{f}(x, y)$, которое было бы близко к $f(x, y)$ по заданному критерию.

Поскольку исходное, искаженное и восстановленное



изображения – суть матрицы размерностью $N_1 \times N_2$, в работе для оценки эффективности качества восстановления исходного изображения предлагается использовать минимум нормы матрицы невязки $\Delta(x, y) = f(x, y) - \hat{f}(x, y)$ исходного и восстановленного изображений:

$$\varepsilon = \|\Delta(x, y)\|^P = \left\| f(x, y) - \hat{f}(x, y) \right\|^P \rightarrow \min, \quad (1)$$

где P – показатель степени, а $\|\cdot\|$ – некоторая матричная норма. При этом конкретный вид матричной нормы, фигурирующей в (1), будет определять конкретный вид критерия качества восстановления исходного изображения. Так, если в выражении (1) использовать квадрат нормы Фробениуса с соответствующей нормировкой, то получится хорошо известный критерий минимума квадрата средней квадратичной ошибки (СКО) [2]:

$$\sigma^2 = \frac{1}{N_1 N_2} \sum_{x=0}^{N_1-1} \sum_{y=0}^{N_2-1} \left(f(x, y) - \hat{f}(x, y) \right)^2 \rightarrow \min.$$

Данный критерий является наиболее универсальным и распространенным на практике критерием качества восстановления при проектировании алгоритмов фильтрации изображений.

Пространственные методы фильтрации в общем случае описываются выражением вида [1, 4]:

$$\hat{f}(x, y) = A\{g(x, y)\},$$

где $g(x, y)$ – искаженное изображение на входе фильтра, $\hat{f}(x, y)$ – восстановленное изображение на выходе фильтра, а $A\{\}$ – оператор фильтрации, определенный в некоторой окрестности S_{xy} точки (x, y) .

Определение окрестности S_{xy} вокруг точки (x, y) заключается в использовании квадратной или прямоугольной области размерностью $n = n_1 \times n_2$ пикселей – подмножества изображения, центрированного в точке (x, y) . Процесс фильтрации заключается в том, что центр окрестности S_{xy} передвигается от пикселя к пикселю, начиная с верхнего левого угла. Оператор $A\{\}$ выполняется в каждой точке (x, y) , давая в результате выходное значение $\hat{f}(x, y)$ для данной точки – отклик фильтра.

Решение задачи оптимизации весовой пространственно-временной фильтрации изображений

С целью повышения эффективности фильтрации, т.е. для уменьшения квадрата СКО, в работе предлагается использовать весовую пространственную фильтрацию, которая задается выражением следующего вида:

$$\hat{f}(\vec{w}, x, y) = A\{\vec{w}, g(x, y)\}. \quad (2)$$

где $\vec{w} = [w_1 \dots w_n]^T$ – n -мерный вектор весовых коэффициентов, однозначно определенный в окрестности S_{xy} .

Согласно выражению (2) каждому элементу апертуры фильтра соответствует определенное число, назы-

ваемое весовым коэффициентом $w_i, i = \overline{1, n}$. При этом апертуру фильтра вместе с заданным на ней вектором \vec{w} назовем маской.

Весовая пространственная фильтрация осуществляется перемещением маски по изображению. В каждом положении маски вектор весовых коэффициентов \vec{w} поэлементно умножается на значения соответствующих пикселей апертуры фильтруемого изображения $g(x, y)$. Полученные произведения подвергаются затем действию оператора фильтрации $A\{\}$, который и определяет отклик фильтра $\hat{f}(\vec{w}, x, y)$ для данной точки. При этом вектор \vec{w} в процессе перемещения маски остается неизменным.

При временной фильтрации изображений рассматривается последовательность кадров изображения $g(x, y, 1), g(x, y, 2), \dots, g(x, y, n)$, полученных в дискретные моменты времени. Апертура S_{xy} , в этом случае, принимает временной характер, т.е. включает в себя соответственные пиксели всех n кадров.

Процесс временной фильтрации состоит в том, что апертура S_{xy} передвигается от пикселя к пикселю, начиная с верхнего левого угла, при этом в каждой точке (x, y) выполняется оператор фильтрации $A\{\}$, давая отклик фильтра

$$\hat{f}(x, y) = A\{g(x, y, k), k = \overline{1, n}\}.$$

С целью повышения эффективности временной фильтрации в работе предлагается использовать весовую временную фильтрацию, которая задается выражением вида:

$$\hat{f}(\vec{w}, x, y) = A\{\vec{w}, g(x, y, k), k = \overline{1, n}\}.$$

Весовая временная фильтрация осуществляется таким же способом, что и пространственная.

В общем случае при весовой пространственно-временной фильтрации выражение квадрата СКО ошибки приближения восстановленного $\hat{f}(\vec{w}, x, y)$ и исходного $f(x, y)$ изображений принимает вид нелинейной функции от \vec{w} .

$$\sigma^2(\vec{w}) = F(\vec{w}) = \frac{1}{N_1 N_2} \sum_{x=0}^{N_1-1} \sum_{y=0}^{N_2-1} \left(f(x, y) - \hat{f}(\vec{w}, x, y) \right)^2. \quad (3)$$

Таким образом, при весовой пространственно-временной фильтрации изображений возникает задача нахождения такого оптимального вектора весовых коэффициентов \vec{w}^* , который обеспечивал бы минимум квадрата СКО вида (3), т.е. минимум нелинейной целевой функции $F(\vec{w})$:

$$\vec{w}^* = \arg \min_{\vec{w} \in W} F(\vec{w}). \quad (4)$$

на ограниченном замкнутом множестве допустимых весовых коэффициентов $W = \{\vec{w} \in R_n : a_i \leq w_i \leq b_i, i = \overline{1, \dots, n}\}$ где a_i, b_i – значения i -й компоненты вектора весовых коэффициентов, характеризующие область ее возможных значений. В общем случае это нелинейная задача условной оптимизации [6].

В связи с особенностями целевой функции вида (3) (многоэкстремальность, не выпуклость, неаналитичность, плохая обусловленность) в работе для решения задачи (4) предлагается использовать алгоритм оптимизации, состоящий из двух этапов – глобального и локального [7-11]. На первом этапе с использованием одного из методов динамического программирования находится грубое приближение оптимального вектора весовых коэффициентов \bar{w}^* (область глобального экстремума). А на втором этапе компоненты вектора \bar{w}^* определяются с заданной точностью одним из эффективных локальных методов квазиньютоновского типа.

Экспериментальные исследования

В работе, с использованием разработанной на языке Си программы [12] весовой пространственно-временной фильтрации изображений, исследована эффективность применения предложенного двухэтапного метода оптимизации при весовой пространственно-временной фильтрации изображений, основанной на использовании порядковых статистик среднего ранга (медианная фильтрация). Медианный фильтр сохраняет перепады яркости без искажений и особенно эффективен при фильтрации импульсных шумов, которые выглядят как наложение на изображение случайных черных и белых точек с соответствующими вероятностями P_a и P_b [2].

Исследования проводились на наборе из 25 тестовых черно-белых изображений с различным характером плотности распределения яркости элементов. Размеры тестовых изображений составляли 360x240 пикселей. Фильтрация изображений осуществлялась при воздействии «белого» импульсного шума с вероятностями:

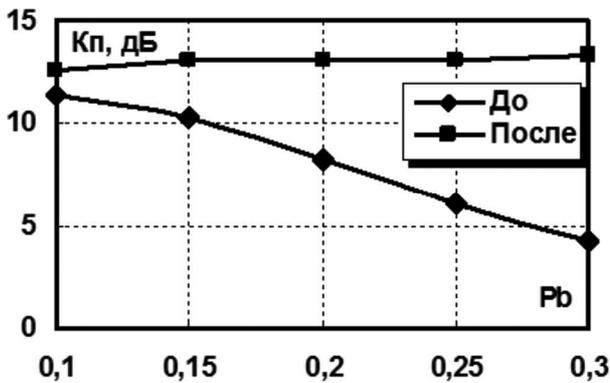
$$P_b = 0.1, 0.15, 0.2, 0.25, 0.3.$$

В качестве критерия эффективности использовался коэффициент подавления фильтра, определяемый по формуле: $K_{\Pi} = 20 \lg(\sigma_0 / \sigma_1)$, где σ_0 – СКО ошибки на входе фильтра (СКО ошибки между исходным $f(x, y)$ и искаженным $g(x, y)$ изображениями), σ_1 – СКО ошибки на выходе фильтра (СКО ошибки между исходным $f(x, y)$ и восстановленным $\hat{f}(\bar{w}, x, y)$ изображениями).

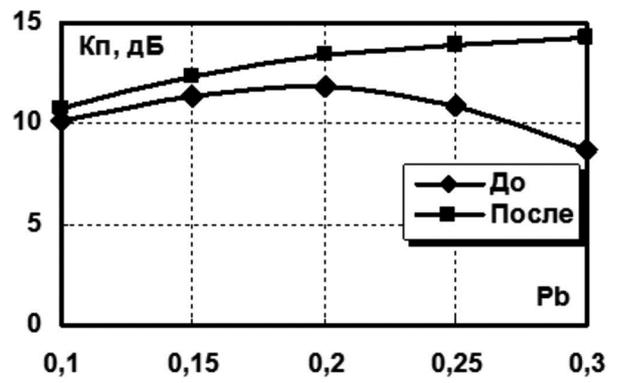
Результаты применения двухэтапного метода оптимизации (модифицированного метода динамического программирования и квазиньютоновского метода с преобразованной BFGS-формулой) при медианной фильтрации приведены на рис. 1 и рис. 2. На рис. 1 приведены зависимости среднего коэффициента подавления (усредненного по набору тестовых изображений) от вероятности P_b до и после оптимальной весовой пространственной фильтрации. На 2 – до и после оптимальной весовой временной фильтрации.

На рис. 3 в качестве примеров оптимальной весовой фильтрации с использованием двухэтапного метода оптимизации приведены результаты медианной пространственной ($n=3 \times 3$) и временной ($n=5$) фильтрации тестового изображения. На рисунке представлены: исходное изображение, изображение пораженное импульсным шумом ($P_b = 0.2$) и восстановленные изображения.

Следует отметить, что аналогичные результаты получаются и при воздействии «черного» импульсного шума.

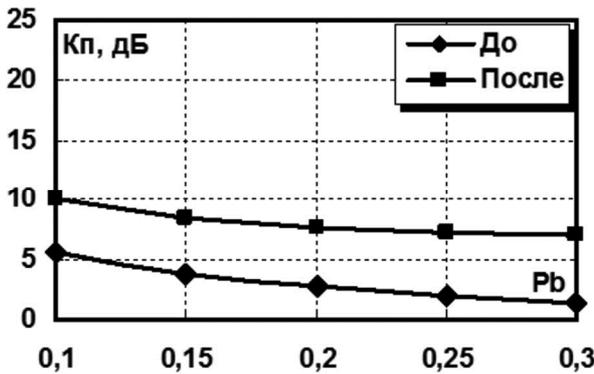


а) $n = 3 \times 3$

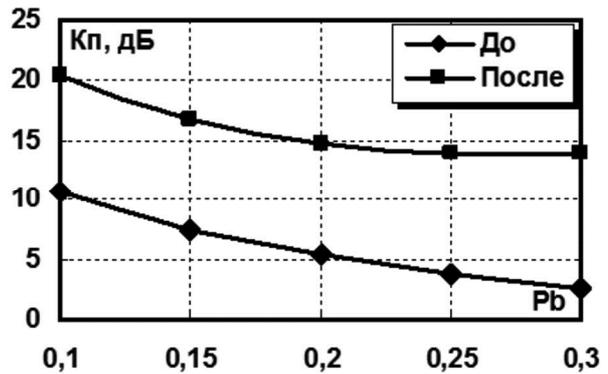


б) $n = 5 \times 5$

Рис. 1. Пространственная фильтрация



а) $n = 3$



б) $n = 5$

Рис. 2. Временная фильтрация



а) исходное изображение



б) зашумленное изображение



в) пространственная фильтрация



г) временная фильтрация

Рис. 3. Примеры весовой пространственно-временной фильтрации

Заключение

Полученные результаты позволяют сделать следующие выводы:

1. Отсутствие весовой обработки приводит к ухудшению эффективности подавления импульсного шума с ростом вероятностей P_a и P_b (средний K_{Π} уменьшается).

2. Применение двухэтапного метода оптимизации при весовой пространственно-временной фильтрации изображений, основанной на использовании медианной фильтрации, обеспечивает высокую эффективность подавления во всем диапазоне вероятностей импульсного шума (средний K_{Π} практически не изменяется), что показывает робастность алгоритма к статистическим характеристикам изображений и импульсного шума.

3. С точки зрения вычислительных затрат и эффективности подавления импульсных шумов наиболее эффективным является использование оптимальной весовой временной медианной фильтрации.

Литература

1. Гонсалес Р., Вудс Р. Цифровая обработка изображений. М: Техносфера, 2012. 1104 с.
2. Pratt W.K. Digital image processing, 4th edition. Wiley, 2007. 807 p.
3. Хуанг Т.С. (ред.) Быстрые алгоритмы в цифровой обработке изображений. Преобразования и медианные фильтры. М: Радио и связь, 1984. 221 с.
4. Яшин В.В., Калинин Г.А. Обработка изображений на языке Си для IBM PC: Алгоритмы и программы. М.: Мир, 1994. 241 с.

5. Graboskwi S., Bienieck W. A two-pass median-like filter for impulse noise removal in multi-channel images. KOSYR, 2003, pp. 195-200.

6. Рейклейтис Г., Рейвиндран А., Рэгсдел К. Оптимизация в технике. В 2-х книгах. М.: Мир. 1986. 670 с.

7. Маторин А.В., Смирнов А.А. Алгоритм параметрического синтеза многоэлементных тонкопроволочных антенн и устройств СВЧ. Вестник РГРТА. 1997. Вып. 2. С. 85-92.

8. Маторин А.В., Смирнов А.А. Оценка эффективности двухэтапного метода нелокальной оптимизации на основе решения тестовых задач. Вестник РГРТА. 1998. Вып. 5. С. 42-45.

9. Маторин А.В., Смирнов А.А. Результаты разработки методики и учебно-исследовательского программного комплекса параметрического синтеза и статистического анализа устройств СВЧ. Вестник РГРТА. 1998. Вып. 4. С. 71-82.

10. Смирнов А.А. Исследование и разработка алгоритмов параметрического синтеза устройств СВЧ в радиотехнических системах: Автореферат дис. канд. техн. наук. Рязань: РГРТА, 2000. 24 с.

11. Маторин А.В., Смирнов А.А. Двухэтапный численный метод решения задач синтеза многоэлементных тонко-проволочных антенн и устройств сверхвысоких частот. Радиотехника и электроника. 2001. Т. 46. № 6.

12. Маторин А.В., Смирнов А.А. Оценка влияния современных Си-компиляторов на производительность вычислительных программ. Вестник РГРТА. 2000. Вып. 6.

УДК 621.376

ИТЕРАТИВНОЕ ГИБРИДНОЕ ПРЕКОДИРОВАНИЕ С НИЗКОЙ СЛОЖНОСТЬЮ ДЛЯ МАССОВЫХ MIMO СИСТЕМ В МИЛЛИМЕТРОВОМ ДИАПАЗОНЕ РАДИОВОЛН

*Асаф М., аспирант каф. радиофизики Томского государственного университета, e-mail: md.moh1600@gmail.com
Пономарев О.Г., к.ф.-м.н., доцент каф. радиофизики Томского государственного университета, e-mail: ponomarev@phys.tsu.ru*

LOW COMPLEXITY ITERATIVE HYBRID PRECODING FOR MILLIMETER-WAVE MASSIVE MIMO SYSTEMS

Assaf M., Ponomarev O.G.

In this paper, we provide a hybrid precoding method that may significantly improve the performance of MIMO transceiver systems operating at millimeter-wave frequencies. We show that the angular information of the analog phase shifters can be directly derived from the optimal unconstrained precoder without the need for any additional information or decomposition techniques, which reduces complexity and the amount of feedback information. The suggested method employs an iterative process to reduce residual error between the product precoding matrices and the optimum matrix. When the number of streams is less than the number of RF chains, simulation results show that the proposed scheme can almost match the performance of the upper bound obtained by the optimal full-baseband design.

Key words: analog/ digital precoding, Millimeter wave communication, massive MIMO, hybrid beamforming.

Ключевые слова: аналоговое/цифровое прекодирование, связь на миллиметровых радиоволнах, массовое MIMO, гибридное формирование луча.

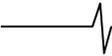
Введение

Одной из важных технологий для достижения высокой производительности в мобильных сетях пятого поколения 5G (от англ. Fifth Generation) является использование диапазона миллиметровых (мм) волн (от 30 до 300 ГГц). Он может поддерживать большое количество пользователей со скоростью передачи данных несколько гигабит в секунду [1]. Небольшая длина волны в миллиметровом диапазоне позволяет использовать технологию mMIMO (от англ. massive multiple-input multiple-output), при которой большое количество антенн размещается на относительно небольшом пространстве, что позволяет создавать антенные решетки с высоким коэффициентом усиления, которые могут помочь компенсировать увеличение потерь при распространении радиоволн [2]. В обычных системах MIMO, также известных как полно-цифровые системы MIMO, передаваемые символы прекодируются в базовой полосе, т.е. до гетеродинирования на несущую радиочастоту. Как следствие, для цифрового формирования луча каждый антенный элемент в системе должен иметь собственную радиочастотную цепь, состоящую из АЦП/ЦАП, смесителя и усилителя мощности и все эти цепи должны быть синхронизованы. Все это приводит к тому, что использование цифровых систем формирования луча для реализации массового MIMO на данном технологическом этапе практически невозможно. Уменьшения количества радиочастотных цепей в mMIMO системах можно добиться, используя концепцию гибридной обработки [3], в которой цифровое прекодирование в базовой полосе сочетается с аналоговым прекодированием, выполняе-

Предложен гибридный метод прекодирования, при котором возможно значительное улучшение спектральной эффективности приемопередающих систем MIMO, работающих в миллиметровом диапазоне. Показано, что угловая информация для аналоговых фазовращателей может быть непосредственно получена из оптимального неограниченного прекодера без необходимости использования какой-либо дополнительной информации или методов разложения, что снижает сложность и количество информации обратной связи. Предложенный метод использует итерационный процесс для уменьшения оставшейся ошибки между произведением матриц прекодирования и оптимальной матрицей. Когда количество потоков данных меньше, чем количество радиочастотных цепей, результаты моделирования показывают, что предложенная схема обеспечивает спектральную эффективность, близкую к максимально возможной, т.е. той, которая получается в полностью цифровой системе.

мым с помощью экономически эффективных управляемых фазовращателей. Проблема, возникающая при этом при построении гибридной системы, заключается в том, как настроить весовые коэффициенты прекодирования и комбинирования в различных ситуациях, чтобы максимизировать спектральную эффективность при минимизации помех.

За последние несколько лет было проведено множество исследований по оптимизации весов прекодирования и комбинирования для улучшения производительности гибридного формирования луча. В [4] предложен метод, называемый ортогональным поиском соответствия (OMP, от англ. Orthogonal Matching Pursuit), в котором характеристики рассеяния миллиметровых волн используются для проектирования радио (RF) и базовых прекодеров. В [5], [6] была предложена итеративная гибридная техника прекодирования на основе подмассива с низкой сложностью и близкой к оптимальной производительностью. Чтобы уменьшить взвешен-



ную сумму квадратов остатков между оптимальной и гибридной конструкциями, предложенная в [6] техника итеративно обновляет фазы фазовращателей в RF-прекодере и комбинаторе. Схема, предложенная в [7], использует корреляционные значения между собственными векторами матрицы отклика решетки и оптимальным прекодером для выбора наиболее релевантных базисных векторов. Для снижения вычислительной сложности авторы [8] предложили проектировать матрицу прекодирования RF непосредственно на основе правых сингулярных векторов матрицы канала. Согласно [9], матрица прекодирования RF разбивается на две части, первая из которых содержит число векторов, равное числу потоков, и основана на собственных векторах левой матрицы разложения сингулярных значений (SVD, от англ. Singular Value Decomposition). Второй компонент получается путем использования информации о зависимости, которая существует между левой в сингулярном разложении матрицей и оптимальной цифровой матрицей прекодирования. Концепция «эквивалентного канала» была использована в [10] для разработки гибридного прекодера и комбинатора совместно для улучшения спектральной эффективности системы.

Общей проблемой существующих методов прекодирования является то, что они либо сложны и требуют множества вычислений [5], [6], либо требуют дополнительной информации о радиоканале [4], [8], которая нежелательна, либо используют матричное разложение [7], [10], [11], которое также не является простой операцией и не обеспечивает высокой производительности в случаях, когда число радиочастотных цепей больше числа потоков данных.

Предлагается итеративный гибридный подход к прекодированию. Его целью является минимизация разрыва в спектральной эффективности между оптимальной полностью цифровой схемой и существующими схемами, когда количество потоков меньше количество радиочастотных цепей, и сохранение тех же показателей спектральной эффективности при меньшей сложности, когда количество потоков равно количеству радиочастотных цепей. Наш гибридный прекодер/комбайнер разработан таким образом, чтобы использовать преимущества дополнительных радиочастотных цепей для минимизации расстояния между матрицей прекодирования и произведениями гибридных радиочастотных и базовых матриц прекодирования/комбинирования, что приводит к приближению к оптимальной спектральной эффективности. Предлагаемый нами метод не требует полного знания характеристик радиоканала и сложных методов декомпозиции, что снижает сложность и количество информации обратной связи.

Обозначения: в данной работе матрицы и векторы выделены жирным шрифтом, для матриц используются прописные буквы, а для векторов – строчные. Надстрочные знаки H и $^{-1}$ обозначают сопряженное транспонирование и инверсию, соответственно $\|\cdot\|_F$ обозначает норму Фробениуса матрицы. \mathbf{I}_N – матрица идентичности $N \times N$; $\mathbf{0}_{M \times N}$ – матрица всех нулей $M \times N$. $\text{tr}(A)$ означает след матрицы A .

Системная модель гибридного прекодирования

На рис. 1 показана гибридная система MIMO для одного пользователя в миллиметровом диапазоне. Предполагается, что передатчик имеет N_s пространственных потоков, передаваемые через N_t антенны с помощью N_t^{RF} радиочастотных цепей, а приемник использует N_r антенны и N_r^{RF} радиочастотных цепей для приема N_s потока данных. Для достижения основной цели гибридной структуры – низкой сложности – количество радиочастотных цепей должно быть меньше количества антенн, и, следовательно $N_s \leq N_t^{RF} (N_r^{RF} < N_t(N_r))$.

Символьный вектор $\sqrt{\rho}\mathbf{s}$ размерностью $N_s \times 1$ сначала прекодируется базовой матрицей, а затем радиочастотным прекодером \mathbf{F}_{RF} . ρ обозначает среднюю мощность принятого сигнала. Здесь элементы вектора \mathbf{s} предполагаются независимыми и с нормированной энергией, равной $1/N_s$. Размерности двух матриц \mathbf{F}_{BB} и \mathbf{F}_{RF} равны $N_t^{RF} \times N_s$ и $N_t \times N_t^{RF}$ соответственно. Затем вектор прекодированного сигнала $\sqrt{\rho}\mathbf{F}_{RF}\mathbf{F}_{BB}\mathbf{s}$ передается по узкополосному каналу распространения радиоволн H , чтобы получить принятый вектор \mathbf{y} , заданный как:

$$\mathbf{y} = \sqrt{\rho}\mathbf{H}\mathbf{F}_{RF}\mathbf{F}_{BB}\mathbf{s} + \mathbf{n}, \quad (1)$$

где вектор шума $\mathbf{n} \in \mathbb{C}^{N_r \times 1}$ предполагается с нулевым средним значением и ковариационной матрицей $\sigma^2\mathbf{I}_{N_r}$.

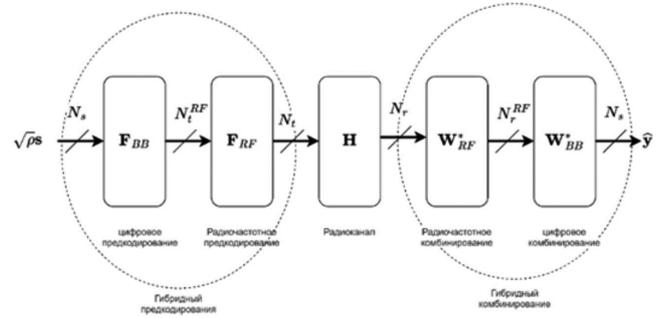
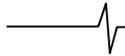


Рис. 1. Гибридное прекодирование/комбинирование MIMO
Блок-схема

Мы используем расширенную геометрическую модель Салеха-Валенсуэлы, как в [4], для моделирования однопользовательского узкополосного канала. Эта модель широко используется для построения mMIMO систем в миллиметровом диапазоне. Она основана на идее, что N_{cl} кластеров рассеяния, каждый из которых содержит N_{ray} путей распространения, объединяются вместе для создания матрицы канала \mathbf{H} . В результате, матрица узкополосного канала \mathbf{H} может быть выражена как:

$$\mathbf{H} = \sqrt{\frac{N_t N_r}{N_{cl} N_{ray}}} \sum_{i=1}^{N_{cl}} \sum_{l=1}^{N_{ray}} \alpha_{i,l} \mathbf{a}_r(\phi_{i,l}^r) \mathbf{a}_t(\phi_{i,l}^t)^*, \quad (2)$$

где $\alpha_{i,l}$ – комплексный коэффициент усиления l -го луча в i -м кластере. $\mathbf{a}_r(\phi_{i,l}^r)$ и $\mathbf{a}_t(\phi_{i,l}^t)$ – векторы откли-



ка массива антенн на приемнике и передатчике для l -го луча в i -м кластере с азимутальными углами $\phi_{i,l}^t$ и $\phi_{i,l}^r$. Матрица канала $\mathbf{H} \in \mathbb{C}^{N_r \times N_s}$ и нормирована так, что $E\{\|\mathbf{H}\|_F^2\} = N_r N_s$.

Если предположить, что передатчик получает информацию о матрице канала с помощью явного метода [12], в котором приемник оценивает канал и затем передает эту информацию обратно передатчику, то для любого метода прекодирования важно минимизировать количество передаваемой информации.

На стороне приемника вектор данных \mathbf{y} обрабатывается матрицей комбинатора $\mathbf{W}^* = \mathbf{W}_{\text{BB}}^* \mathbf{W}_{\text{RF}}^*$, где $\mathbf{W}_{\text{BB}} \in \mathbb{C}^{N_r^{\text{RF}} \times N_s}$ – базовой комбинатор, а $\mathbf{W}_{\text{RF}} \in \mathbb{C}^{N_r \times N_r^{\text{RF}}}$ – RF комбинатор. В результате на приемнике получается сигнал:

$$\hat{\mathbf{y}} = \sqrt{\rho} \mathbf{W}_{\text{BB}}^* \mathbf{W}_{\text{RF}}^* \mathbf{H} \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}} \mathbf{s} + \mathbf{W}_{\text{BB}}^* \mathbf{W}_{\text{RF}}^* \mathbf{n}. \quad (3)$$

Как мы видим, гибридная система выполняет формирование луча как в цифровой, так и в аналоговой областях. Поскольку аналоговые фазовращатели регулируют только фазу сигнала антенных элементов, следует предположить, что элементы двух RF матриц имеют постоянный модуль.

Одним из наиболее важных критериев производительности, который должен быть максимизирован при разработке матриц прекодирования и комбинирования, является спектральная эффективность [1]. Спектральная эффективность может быть выражена в следующей форме [4]:

$$R = \log_2 \left(\left| \mathbf{I}_{N_s} + \frac{\rho}{N_s} \mathbf{R}_n^{-1} \mathbf{W}_{\text{BB}}^* \mathbf{W}_{\text{RF}}^* \mathbf{H} \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}} \right. \right. \\ \left. \left. \times \mathbf{F}_{\text{BB}}^* \mathbf{F}_{\text{RF}}^* \mathbf{H}^* \mathbf{W}_{\text{RF}} \mathbf{W}_{\text{BB}} \right| \right), \quad (4)$$

где $\mathbf{R}_n = \sigma_n^2 \mathbf{W}_{\text{BB}}^* \mathbf{W}_{\text{RF}}^* \mathbf{W}_{\text{RF}} \mathbf{W}_{\text{BB}}$ – ковариационная матрица шума после комбинирования.

Ключевой вопрос здесь заключается в том, как найти глобальный максимум величины R при известной канальной матрице \mathbf{H} , варьируя матрицы прекодирования $\mathbf{F}_{\text{RF}}, \mathbf{F}_{\text{BB}}$ и комбинирования $\mathbf{W}_{\text{RF}}, \mathbf{W}_{\text{BB}}$.

Предлагаемый метод

Чтобы максимизировать спектральную эффективность R в (4), необходимо совместно разработать RF и базовые матрицы прекодирования и комбинирования. К сожалению, оптимизировать все четыре матрицы одновременно очень трудно [4]. Поэтому предлагаются различные стратегии для достижения субоптимального решения при разумных вычислительных затратах. Двумя наиболее известными методами для решения этой задачи оптимизации являются следующие:

- разделение проблемы оптимизации R между передатчиком и приемником на два этапа: первый этап фокусируется на проектировании матриц передатчика $\mathbf{F}_{\text{RF}}, \mathbf{F}_{\text{BB}}$ для максимизации спектральной эффективности, а второй использует этот результат для оптимизации матриц приема;

- разделение проблемы оптимизации R между радиочастотной и цифровой областями: первый этап заключается в максимизации спектральной эффективности с использованием $\mathbf{W}_{\text{RF}}, \mathbf{F}_{\text{RF}}$, а второй – в оптимизации цифровых матриц $\mathbf{F}_{\text{BB}}, \mathbf{W}_{\text{BB}}$.

В нашей работе мы примем первую стратегию, которая разделяет оптимизацию между прекодированием передатчика и комбинированием приемника.

Оптимальные характеристики любой гибридной системы могут быть определены при отсутствии ограничений на RF матрицу, другими словами, когда сигнал, подаваемый на антенну, может управляться по величине и фазе. Оптимальный прекодер \mathbf{F}_{opt} на передатчике определяется N_s столбцами левых сингулярных векторов канальной матрицы \mathbf{H} , которые соответствуют наибольшему сингулярным значениям. Пусть сингулярное разложение SVD (от англ. Singular Value Decomposition) матрицы \mathbf{H} имеет вид:

$$\mathbf{H} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^*, \quad (5)$$

где \mathbf{U} и \mathbf{V} – $N_r \times N_r$ и $N_t \times N_t$ унитарные матрицы, соответственно, а $\mathbf{\Sigma}$ – $N_r \times N_t$ диагональная матрица с диагональными элементами, расположенными в порядке убывания. Тогда $\mathbf{F}_{\text{opt}} = \mathbf{V}(:, 1:N_s)$. Видно, что оптимальная матрица \mathbf{F}_{opt} является полуунитарной матрицей, где $\mathbf{F}_{\text{opt}}^* \mathbf{F}_{\text{opt}} = \mathbf{I}_{N_s}$.

На стороне передатчика задача оптимизации заключается в поиске двух матриц \mathbf{F}_{BB} и \mathbf{F}_{RF} при их совместном приближении к оптимальной матрице \mathbf{F}_{opt} так, что:

$$(\mathbf{F}_{\text{RF}}^{\text{opt}}, \mathbf{F}_{\text{BB}}^{\text{opt}}) = \arg \max_{\mathbf{F}_{\text{RF}}, \mathbf{F}_{\text{BB}}} \text{tr}(\mathbf{F}_{\text{opt}}^* \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}), \\ \text{при условиях } |\mathbf{F}_{\text{RF}}(i, j)| = 1, \quad \forall i, j, \quad (6) \\ \|\mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}\|_F^2 = N_s$$

Согласно [4], задача оптимизации в (6) эквивалентна минимизации $\|\mathbf{F}_{\text{opt}} - \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}\|_F^2$. В результате, цель гибридного прекодера – сделать произведение $\mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}$ как можно ближе к оптимальной матрице \mathbf{F}_{opt} . Основная проблема при попытке вычислять \mathbf{F}_{RF} и \mathbf{F}_{BB} заключается в ограничении $|\mathbf{F}_{\text{RF}}(i, j)| = 1 \forall i, j$, поэтому можно сказать, что в производительности гибридного прекодирования доминирует аналоговые фазовращатели.

Методы на основе OMP в [4], [6], [7], используют векторы отклика антенной решетки для создания RF матрицы \mathbf{F}_{RF} , методы в [9], [10], [11], используют информацию о фазе главных левых сингулярных векторов оптимальной матрицы \mathbf{F}_{opt} . В данной работе, информация о фазе элементов оптимальной матрицы \mathbf{F}_{opt} будет непосредственно использоваться для получения элементов RF матрицы \mathbf{F}_{RF} . После получения матрицы \mathbf{F}_{RF} , матрица прекодирования базовой полосы может быть получена с помощью метода наименьших квадратов:

$$\mathbf{F}_{\text{BB}} = (\mathbf{F}_{\text{RF}}^* \mathbf{F}_{\text{RF}})^{-1} \mathbf{F}_{\text{RF}}^* \mathbf{F}_{\text{opt}}. \quad (7)$$

В системах массового MIMO количество радиочастотных цепей может быть больше, чем N_s , поэтому проблема выбора $\mathbf{F}_{\text{RF}} = \mathbf{F}_{\text{opt}} ./ \text{abs}(\mathbf{F}_{\text{opt}})$ заключается в том, что размерность оптимальной матрицы \mathbf{F}_{opt} составляет $N_t \times N_s$, а размерность RF матрицы \mathbf{F}_{RF} равно $N_t \times N_t^{RF}$, где $N_s \leq N_t^{RF}$. Для достижения максимальной спектральной эффективности необходимо выбрать $N_t^{RF} - N_s$ столбцов. По этой причине многие исследователи используют векторы отклика антенной решетки или левые сингулярные векторы. В предложенном алгоритме остальные столбцы $N_t^{RF} - N_s$ будут использоваться итерационно для улучшения спектральной эффективности через минимизацию остаточной ошибки между оптимальной матрицей \mathbf{F}_{opt} и произведением $\mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}$. Алгоритм 1 показывает псевдокод для предлагаемого прекодера. Базовая матрица \mathbf{F}_{BB} и RF матрица \mathbf{F}_{RF} разделены на L частей, каждая часть состоит из N_s столбцов для RF матрицы и N_s строк для матрицы базовой полосы. Вычисления производятся циклически для каждой из L частей. В конце цикла произведение матриц прекодирования будет равно сумме произведений этих частей $\mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}} = \sum_{i=1}^L \mathbf{F}_{\text{RF}}^i \mathbf{F}_{\text{BB}}^i$, где \mathbf{F}_{RF}^i и \mathbf{F}_{BB}^i – это i -я часть RF и базовой матрицы, соответственно.

Алгоритм 1. Предлагаемый гибридный алгоритм прекодирования

Требуется \mathbf{F}_{opt}

1. $\mathbf{F}_{\text{RF}}, \mathbf{F}_{\text{BB}}$ пустые матрицы, $L = \left\lceil \frac{N_t^{RF}}{N_s} \right\rceil$
2. **for** $i = 1 : L$
3. $\mathbf{F}_{\text{res}} = \mathbf{F}_{\text{opt}} - \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}$
4. $\hat{\mathbf{F}}_{\text{RF}} = \mathbf{F}_{\text{res}} ./ \text{abs}(\mathbf{F}_{\text{res}})$
5. $\hat{\mathbf{F}}_{\text{BB}} = (\hat{\mathbf{F}}_{\text{RF}}^* \hat{\mathbf{F}}_{\text{RF}})^{-1} \hat{\mathbf{F}}_{\text{RF}}^* \mathbf{F}_{\text{res}}$
6. $\mathbf{F}_{\text{RF}} = [\mathbf{F}_{\text{RF}} | \hat{\mathbf{F}}_{\text{RF}}], \mathbf{F}_{\text{BB}} = [\mathbf{F}_{\text{BB}} | \hat{\mathbf{F}}_{\text{BB}}]$
7. **end for**
8. $\mathbf{F}_{\text{BB}} = \sqrt{N_s} \frac{\mathbf{F}_{\text{BB}}}{\|\mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}\|_F}$
9. **return** $\mathbf{F}_{\text{RF}}, \mathbf{F}_{\text{BB}}$

На первой итерации $\mathbf{F}_{\text{res}} = \mathbf{F}_{\text{opt}}$. Первый столбец радиочастотного прекодера вычисляется на шаге 4 в соответствии с выражением $\hat{\mathbf{F}}_{\text{RF}} = \mathbf{F}_{\text{opt}} ./ \text{abs}(\mathbf{F}_{\text{opt}})$. Затем первые N_s строк \mathbf{F}_{BB} вычисляются на шаге 5. На следующей итерации остаточная ошибка между матрицей \mathbf{F}_{opt} и произведением $\mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}$ вычисляется на шаге 3

как матрица \mathbf{F}_{res} , затем вычисляются следующие N_s столбцов \mathbf{F}_{RF} и N_s строк \mathbf{F}_{BB} для минимизации остаточной ошибки.

В предлагаемом подходе не требуется, чтобы N_t^{RF} было кратным N_s . В случае, когда N_t^{RF} не кратен N_s , необходимые столбцы и строки $\hat{\mathbf{F}}_{\text{BB}}$ и $\hat{\mathbf{F}}_{\text{RF}}$ вычитаются из двух матриц на последней итерации, чтобы получить требуемые размеры для обеих матриц \mathbf{F}_{RF} и \mathbf{F}_{BB} , а оставшиеся векторы можно игнорировать.

Стоит отметить, что гибридные RF и базовый комбинатор на стороне приемника могут быть получены тем же способом, что и гибридный прекодер в алгоритме 1.

Общая сложность предложенного нами метода пропорциональна $o(N_s^2(N_t + N_r) + N_s^3L + 3N_s^2N_tL)$, где первый терм $o(N_s^2(N_t + N_r))$ представляет собой количество операций, необходимое для вычисления матрицы нацеливания \mathbf{F}_{opt} при использовании алгоритма усеченного SVD (от англ. Truncated SVD, TSVD), а остальные слагаемые $o(N_s^3L + 3N_s^2N_tL)$ – это количество операций, необходимых для вычисления L частей базовой матрицы на шаге 5. Обычно количество цепей устанавливается близким к количеству потоков, чтобы снизить аппаратные затраты и энергопотребление. Поэтому L не будет большим числом, а сложность будет низкой.

Оценка производительности

Чтобы показать улучшение спектральной эффективности предложенной гибридной схемы прекодирования, результаты предложенного алгоритма сравниваются с результатами алгоритма OMP в [4], алгоритма SVD с низкой сложностью в [8], алгоритма гибридного прекодирования по эквивалентному каналу НР-ЕС (от англ. Hybrid precoding equivalent matrix) в [10].

Радиоканал генерируется на основе следующих настроек: количество кластеров $N_{cl} = 8$, количество лучей $N_{ray} = 10$. Моделируются две структуры антенной решетки: первая – 64×16 и вторая – 256×64 , чтобы продемонстрировать влияние увеличения размера антенной решетки на производительность системы. Каждая точка моделирования усредняется по 5000 различным реализациям канала.

На рис. 2 и 3 показана зависимость спектральной эффективности от отношения сигнал/шум (SNR) при $(N_t = 64, N_r = 16, N_t^{RF} = N_r^{RF} = 4)$ и $(N_t = 256, N_r = 64, N_t^{RF} = N_r^{RF} = 8)$, соответственно, для различных значений N_s . Из результатов симуляции на рис. 2 следует, что при $N_s = 1$, предложенная схема работает так же, как и оптимальная схема прекодирования, при $N_s = 2$, разрыв между предложенной схемой и оптимальной схемой увеличивается, но все еще меньше, чем разрыв с текущими схемами. При $N_s = 4$, разрыв увеличивается и достигнутая производительность аналогична текущим схемам. Как показано на рис. 3, увеличение размера антенной решетки дает те же результаты.

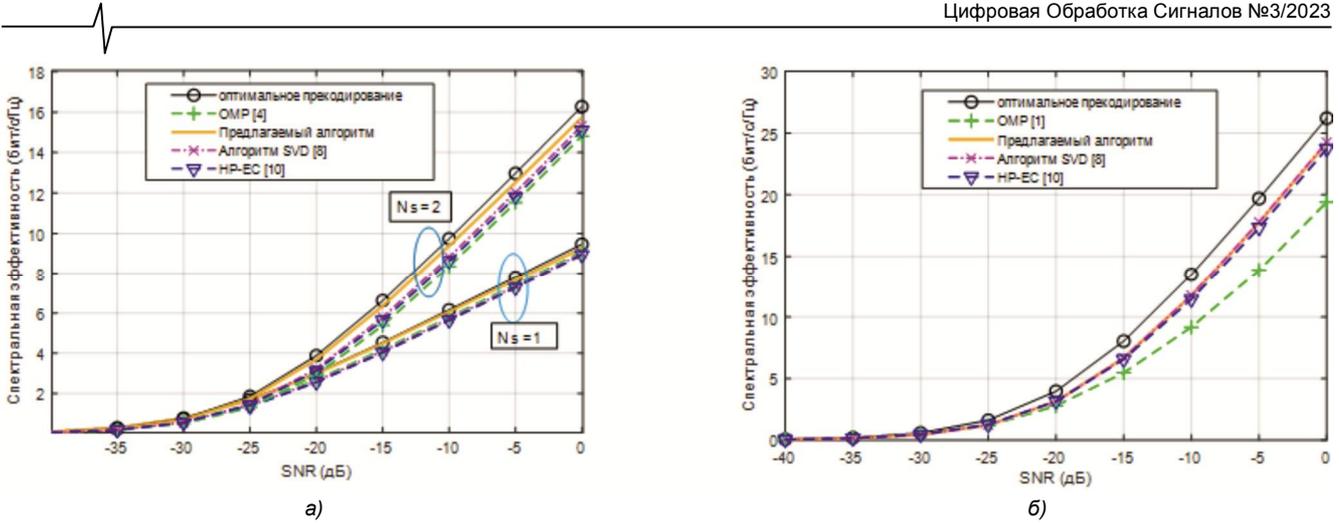


Рис. 2. Спектральная эффективность при различном SNR, когда $N_t = 64$ и $N_r = 16$.
 а) Поток данных $N_s = 1$ и 2. б) Поток данных $N_s = 4$

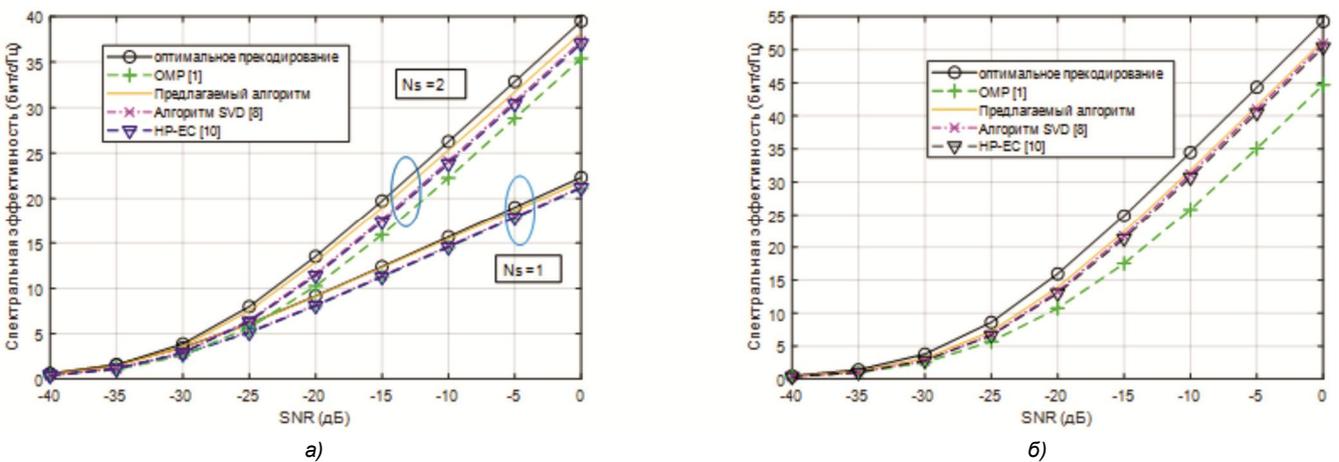


Рис. 3. Спектральная эффективность при различном SNR, когда $N_t = 256$ и $N_r = 64$.
 а) Поток данных $N_s = 1$ и 4. б) Поток данных $N_s = 6$

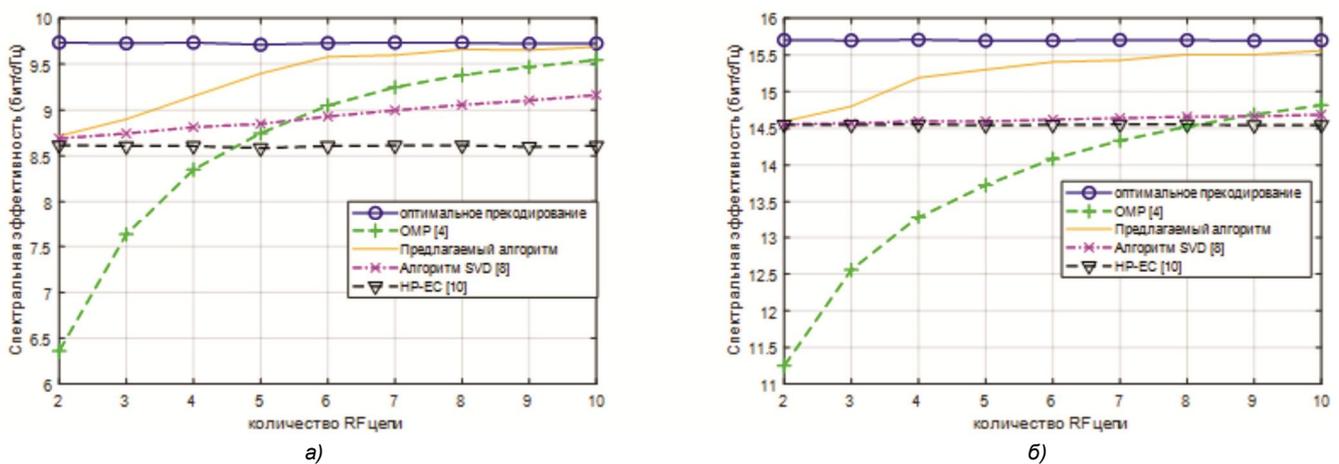


Рис. 4. Спектральная эффективность при различном N_t^{RF} и $SNR = -10$ дБ.
 а) ($N_t = 64$ and $N_r = 16$), б) ($N_t = 256$ and $N_r = 64$)

Из графиков видно, что предложенная схема гибридного прекодирования превосходит существующие методы прекодирования, если $N_s \leq N_t^{RF}$. Когда N_s увеличивается, разрыв в спектральной эффективности между предложенной схемой и оптимальной схемой увеличивается.

На рис. 4. показано, как спектральная эффективность изменяется в зависимости от количества радио-

частотных цепей при $SNR = -10$ дБ, $N_s = 2$ для двух антенных решеток разных размеров. Очевидно, что количество радиочастотных цепей оказывает существенное влияние на производительность системы. Все существующие методы, за исключением OMP [4], не дают существенного повышения спектральной эффективности при увеличении числа радиочастотных цепей. Для предложенного метода – это не так. Здесь имеется яв-

ный рост спектральной эффективности с ростом количества радиочастотных цепей. Как следствие, предложенный алгоритм превосходит существующие для всех значений $N_t^{RF} > N_s$.

Важно отметить, что при $N_s = N_t^{RF}$, количество итераций $L = 1$ и предлагаемый метод менее сложен, чем все исследованные алгоритмы, поскольку угловая информация аналоговых фазовращателей может быть непосредственно получена из матрицы нацеливания без необходимости использования какой-либо дополнительной информации или методов разложения. В результате, когда $N_s = N_t^{RF}$, производительность остается той же, но сложность уменьшается.

Заключение

В работе рассматриваются алгоритмы прекодирования/комбинирования для однопользовательских гибридных систем массового MIMO на миллиметровых волнах, которая обеспечивает разумный компромисс между производительностью и сложностью. Спектральная эффективность была улучшена за счет итеративного построения векторов RF и базовой матриц прекодирования с использованием угловой информации непосредственно из матрицы наведения. Это позволило приблизить произведение двух матриц к оптимальной матрице. Полученные результаты показали, что при $N_s < N_t^{RF}$ разрыв спектральной эффективности между оптимальной полностью цифровой конструкцией и существующими схемами уменьшается, а при $N_s = N_t^{RF}$ достигается та же спектральная эффективность при меньшей вычислительной сложности.

Литература

1. Z. Xiao et al. «A survey on millimeter-wave beamforming enabled UAV communications and networking», IEEE Commun. Surveys Tuts., vol. 24, no. 1, pp. 557-610, 5th Quart., 2022.
2. S. A. Busari, K.M.S. Huq, S. Mumtaz, L. Dai, and J. Rodriguez. «Millimeter-wave massive MIMO communica-

tion for future wireless systems: A survey», IEEE Commun. Surveys Tuts., vol. 20, no. 2, pp. 836-869, 2nd Quart., 2017.

3. A.F. Molisch et al. «Hybrid beamforming for massive MIMO: A survey», IEEE Commun. Mag., vol. 55, no. 9, pp. 134-141, Sep. 2017.

4. O.E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi and R.W. Heath. «Spatially Sparse Precoding in Millimeter Wave MIMO Systems», in IEEE Transactions on Wireless Communications, vol. 13, no. 3, pp. 1499-1513, March 2014.

5. L. Dai, X. Gao, J. Quan, S. Han and C.-L.I. «Near-optimal hybrid analog and digital precoding for downlink mmWave massive MIMO systems», 2015 IEEE International Conference on Communications (ICC), London, UK, 2015, pp. 1334-1339.

6. C.-E. Chen. «An Iterative Hybrid Transceiver Design Algorithm for Millimeter Wave MIMO Systems», in IEEE Wireless Communications Letters, vol. 4, no. 3, pp. 285-288, June 2015.

7. X. Liu et al. «Hybrid Precoding for Massive mmWave MIMO Systems», in IEEE Access, vol. 7, pp. 33577-33586, 2019.

8. D. Zhang, P. Pan, R. You and H. Wang. «SVD-Based Low-Complexity Hybrid Precoding for Millimeter-Wave MIMO Systems», in IEEE Communications Letters, vol. 22, no. 10, pp. 2176-2179, Oct. 2018.

9. S. Wang, L. Li, R. Ruby, and P. Li. «A general hybrid precoding scheme for millimeter wave massive MIMO systems», Wireless Netw., vol. 26, pp. 1331-1345, Mar. 2020.

10. S. Wang, M. He, J. Wang, R. Ran, H. Ji and V.C.M. Leung. «A Family of Hybrid Precoding Schemes for Millimeter-Wave Massive MIMO Systems», in IEEE Systems Journal, vol. 16, no. 3, pp. 4881-4891, Sept. 2022.

11. S. Wang et al. «A Joint Hybrid Precoding/Combining Scheme Based on Equivalent Channel for Massive MIMO Systems», in IEEE Journal on Selected Areas in Communications, vol. 40, no. 10, pp. 2882-2893, Oct. 2022.

12. H. Lou, M. Ghosh, P. Xia, and R. Olesen. «A comparison of implicit and explicit channel feedback methods for MU-MIMO WLAN systems», in Proc. IEEE Pers. Indoor Mobile Radio Commun. (PIMRC), Sep. 2013, pp. 419-424.

НОВЫЕ КНИГИ

Косичкина Т.П., Сперанский В.С.

Цифровые сигнальные процессоры и их применение в системах телекоммуникаций и электроники: Учебное пособие для вузов. М.: Изд-во «Горячая линия-Телеком», 2022 г. 316 с.: ил.

Рассмотрены вопросы теории и практики использования цифровых сигнальных процессоров. Описаны основные операции цифровой обработки сигналов, структура, архитектура, классификация и характеристики современных процессоров. Представлен обзор процессоров ведущих фирм-производителей: Analog Devices, Ceva, Feecsale, Texas Instruments. Отдельная глава посвящена российским цифровым процессорам. Существенное внимание уделено многоядерным процессорам и их характеристикам. В разделе, связанном с программируемыми логическими микросхемами и системами на кристалле показаны возможности их использования в качестве сигнальных процессоров. Даны примеры реализации цифровых устройств с помощью САПР на языке программирования VHDL, рассмотрены его элементы. В приложениях даны примеры реализации формирователей двоичных последовательностей, генераторов сигналов и медианных фильтров.

Для студентов радиотехнических и инфокоммуникационных специальностей, будет полезна аспирантам и специалистам.



УДК.004.383:621.391

ОСОБЕННОСТИ РАБОТЫ ПРИЕМНОГО ТРАКТА В МНОГОКАНАЛЬНОЙ РАДИОЛОКАЦИОННОЙ СТАНЦИИ С ВРЕМЕННОЙ АВТОМАТИЧЕСКОЙ РЕГУЛИРОВКОЙ УСИЛЕНИЯ

Шарамет А.В., к.т.н., доцент, начальник тематического отдела, ОАО «КБ Радар» – управляющая компания холдинга «Системы радиолокации» (г. Минск, Республика Беларусь), e-mail: shandrei@yandex.ru
Азаров И.С., д.т.н., доцент, заведующий кафедрой ЭВС, Белорусский государственный университет информатики и радиоэлектроники (г. Минск, Республика Беларусь), e-mail: azarov@bsuir.by

FEATURES OF OPERATION OF THE RECEIVING PATH IN A MULTICHANNEL RADAR STATION WITH TEMPORARY AUTOMATIC GAIN CONTROL

Sheremet A.V., Azarov I.S.

The power of the radar signal reflected from "distributed" local objects has a significant impact on the detection characteristics of an aerial target. In the conditions of the use of technologies to reduce their radar visibility or the use of small-sized unmanned aerial vehicles, the total differential effective scattering area becomes commensurate, and in some cases predominant over the level of the signal reflected from the target. This is due to the fact that reflections significantly overload the receiving path of the radar station. A feature of modern radar stations is their multi-channel nature. One of the main ways to reduce the power level of reflected signals is the use of individual circuits with gain control of the receiving path in each channel according to a known law. At the same time, there are no recommendations on the procedure for preliminary evaluation of the parameters of the gain control law of the receiving path, taking into account the geometry of the location of the beams, the tactical and technical characteristics of the radar station and the specifics of its application. In the course of the work, such an approach is proposed and its correctness is shown.

Key words: multichannel radar station, automatic gain control, feature of modern radar, detection characteristics of an aerial target.

Ключевые слова: многоканальность, автоматическая регулировка усиления, пространственно-временная обработка, дифференциальная эффективная площадь рассеяния, радиолокационная станция.

Введение

Одним из направлений развития современных радиолокационных станций (РЛС) является обнаружение и сопровождение воздушных объектов на малых и предельно малых высотах. Особенностью их обнаружения [1] является сложная помеховая обстановка:

- в анализируемом воздушном пространстве одновременно находятся различные типы объектов (самолеты, вертолеты, дельтапланы, парпланы, беспилотные летательные аппараты (БПЛА) и т.д.);
- в одном элементе разрешения может оказаться отраженный сигнал от нескольких воздушных объектов, как частный случай рой БПЛА;
- постоянно присутствуют различного рода помехи (индустриальные излучения, отражения от метеорологических образований, от подстилающей поверхности и местных предметов и т.п.), которые образуют естественный фон;
- наличие организованных радиопомех (активных и пассивных).

Для решения задач обнаружения, оценки параметров и распознавания воздушных объектов с учетом этих

Мощность отраженного от «распределенных» местных предметов радиолокационного сигнала оказывает существенное влияние на характеристики обнаружения воздушной цели. В условиях применения технологий снижения их радиолокационной заметности или использования малогабаритных беспилотных летательных аппаратов полная дифференциальная эффективная площадь рассеяния становится соизмеримой, а в ряде случаев и преобладающей над уровнем отраженного от цели сигнала. Это вызвано тем, что отражения существенно перегружают приемный тракт радиолокационной станции. Особенностью современных радиолокационных станций является их многоканальность. Одним из основных способов снижения уровня мощности отраженных сигналов является применение индивидуальных схем с регулировкой усиления приемного тракта в каждом канале по известному закону. При этом рекомендации по порядку предварительной оценки параметров закона управления коэффициентом усиления приемного тракта с учетом геометрии расположения лучей, тактико-технических характеристик радиолокационной станции и особенностей ее применения отсутствуют. В ходе работы предложен такой подход и показана его корректность.

особенностей в радиолокационных станциях традиционно используются системы селекции движущихся целей (СДЦ). В рамках развития данного направления следует особо отметить работы П.А. Бакулева [2, 3, 4] и М. Скольника [5]. Помеховая обстановка оказывает наиболее существенное влияние на процесс обнаружения всех типов летательных аппаратов на ближней дистанции и на всей дистанции БПЛА. Необходимость

устойчивого обнаружения последних на всей дистанции существенно усложняется в виду их массовости, разнообразия и относительно небольшой стоимости, в виду чего последнее время они получили очень широкое распространение в различных сферах деятельности общества. Их основными особенностями являются:

- малое значение эффективной площади рассеивания, которая лежит в диапазоне от $0,1 \text{ м}^2$ до $0,01 \text{ м}^2$, а в отдельных случаях достигает $0,001 \text{ м}^2$;
- высокая маневренность в сочетании с малыми скоростями полета (до 30 м/с), вплоть до зависания в воздухе;
- движение БПЛА идентично полету птиц;
- способность выполнять полеты на малых (пределно малых) высотах (от 0 до 200 м).

Для работы РЛС в сложной помеховой обстановке при обнаружении БПЛА необходимо сформировать автоматически изменяющееся (ручная регулировка усиления (в иностранной литературе – MGC (от англ. Main Gain Control)) далее не рассматривается) управляющее напряжение для дополнительной регулировки коэффициента усиления приемного тракта с учетом дальности обнаружения РЛС. Это обеспечивает поддержание отношения сигнал/шум на постоянном уровне, а также предотвращает ограничение сигнала при его обработке в приемном тракте. Проведенный анализ [2, 4, 7-9] показывает, что в основе различных подходов лежат схемы с временной (ВАРУ, в иностранной литературе – SCT (от англ. Sensitivity Time Control)) или шумовой (ШАРУ, в иностранной литературе – AGC (от англ. Automatic Gain Control)) автоматической регулировкой усиления [11].

По материалам открытых публикаций [2, 4, 10, 12] в основе серийно выпускаемых РЛС в основном используется ВАРУ с известным законом управления [10]. Широкое внедрение многолучевых РЛС привело к тому, что результирующая зона обзора формируется по результатам обработки нескольких лучей, каждый из которых имеет отдельный приемный канал и свои геометрические особенности. При преобладании в канале отражений от «распределенных» местных предметов таких, как дождь, морская поверхность и т.п. применение ВАРУ в каждом канале обеспечивает их дополнительное подавление, что улучшает условия обработки сигналов [6, с.157]. При этом порядок предварительной оценки основных параметров импульсов ВАРУ, которые должны учитывать технические особенности каждого канала РЛС отсутствует.

Основная часть

В процессе радиолокационного обнаружения воздушных целей излучаемые антенной радиолокационной станции сигналы отражаются от воздушных объектов, а также от поверхности земли. Отраженные сигналы от земной поверхности оказывают большое влияние на их обнаружение. Рассмотрим решение данной задачи с учетом следующих ограничений:

высота полета воздушной цели (h_c) и высота фазового центра антенны (h_r) суще-

ственно меньше радиуса земли ($R_e = 6371 \text{ км}$);

- радиогоризонт РЛС рассчитывается с учетом нормальной рефракции в атмосфере. Для повышения точности определения высоты данный параметр уточняется эмпирически для каждой РЛС с учетом ее конструктивных особенностей [10, стр.245];
- законы распределения отраженных от земной поверхности сигналов и внутренних шумов имеют нормальный закон распределения;
- в качестве зоны сбора мешающих отражений поверхности рассматривается ее проекция на направление облучателя [12]. Таким образом, рассматривается дифференциальная площадь проекции поверхности;
- полная дифференциальная эффективная площадь рассеивания (ЭПР) [5] по боковым лепесткам одной ДН многоканальной радиотехнической системы определяется по первому. Это целесообразно при решении большинства практических задач, так как он в большинстве случаев является определяющим по отношению к другим боковым лепесткам и его параметры (такие как направление и уровень первого бокового лепестка) всегда точно известны.

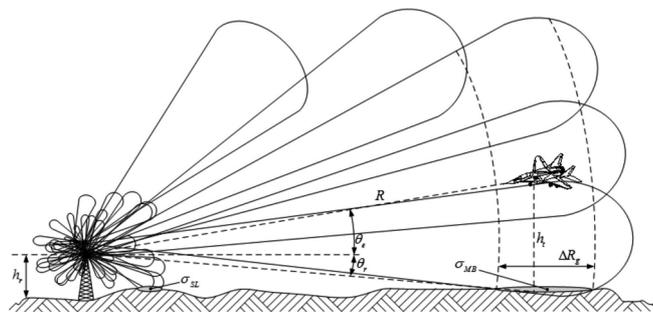


Рис. 1. Геометрия отражений от поверхности земли

Применение в качестве оценки величины дифференциальной ЭПР подчеркивает, что отраженный от земной поверхности сигнал формируется большим количеством элементарных рассеивающих элементов, хаотически расположенных на ограниченной площади. При этом фазы отдельных отраженных сигналов независимы. Это позволяет произвести их суммирование. Тогда полная дифференциальная ЭПР составит:

$$\sigma_c = \sigma_{MB} + \sigma_{SL}, \quad (1)$$

где σ_{MB} – дифференциальная ЭПР по главному лучу ДН (здесь и далее индекс MB означает main beam, от англ. – основной луч); σ_{SL} – дифференциальная ЭПР по боковым лепесткам ДН (здесь и далее индекс SL означает side lobe, от англ. – боковой лепесток).

В качестве примера рассмотрим расчет величины σ_c , от сухого участка земной поверхности (коэффициентом обратного отражения, составляет минус 17.72 дБ) и РЛС имеющей следующие технические характеристики:

- S-диапазон;
- уровень первого бокового лепестка – минус 25 дБ ;
- высота электрического центра антенны – 4 м ;
- высота центра луча на краю – 150 м ;
- ширина ДН по азимуту на прием/передачу – 1.5 град. ;

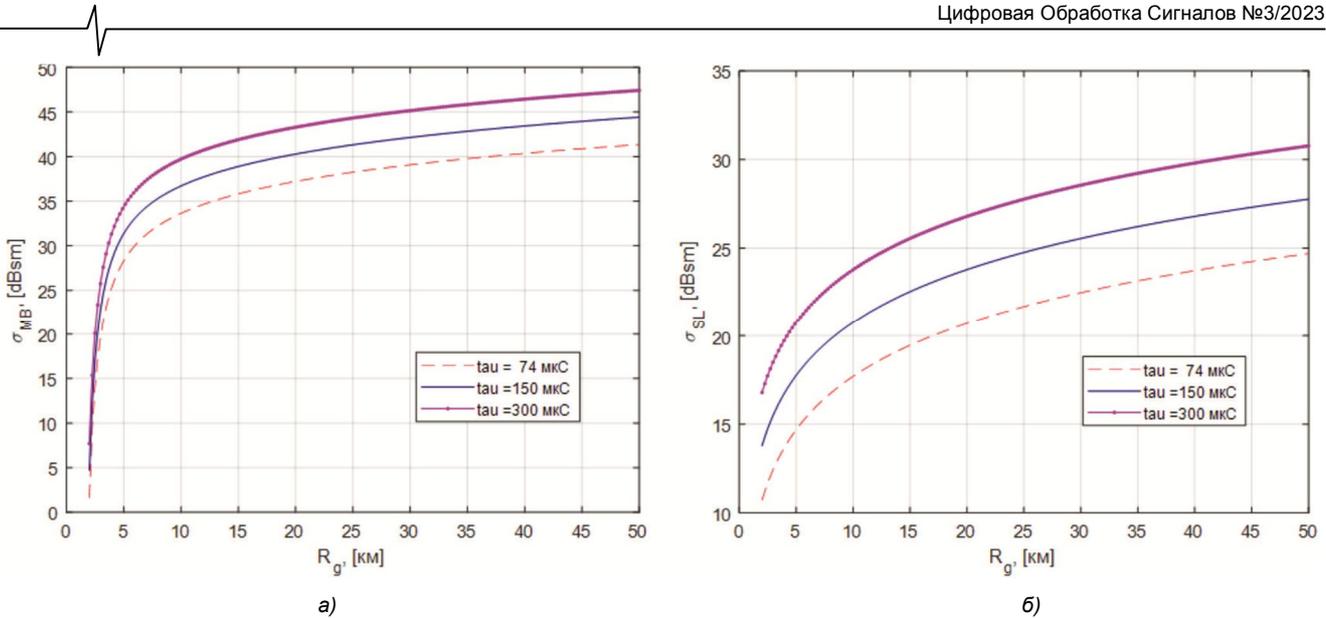


Рис. 2. Величина отражений от поверхности Земли, вносимая диаграммой направленности: а – по основному лепестку; б – по первому боковому лепестку

- ширина ДН по углу места на прием – 1.7 град.;
 - ширина ДН по углу места на передачу – 8.0 град.
- Из рис. 1 видно, что проекция наклонной дальности (R_g) на поверхность земли составляет:

$$R_g = R \cos \theta_r, \tag{2}$$

где R – наклонная дальность.

Основные параметры, определяющие взаимное положение РЛС и воздушной цели могут быть получены с учетом выражений:

$$\theta_r = \arcsin(h_r / R) \text{ и } \theta_e = \arcsin((h_t - h_r) / R),$$

$$\Delta R_g = \Delta R \cos \theta_r, \tag{3}$$

где ΔR – разрешающая способность РЛС по дальности.

Из выражений (2) и (3) следует, что площади сбора мешающих отражений от элементов земной поверхности по главному лучу (A_{MB}) и первому боковому лепестку (A_{SL}) ДН могут быть найдены в соответствии с выражениями:

$$A_{MB} = \Delta R_g \theta_a R_g, \quad A_{SL} = \Delta R_g \pi R_g, \tag{4}$$

где θ_a - ширина диаграммы направленности по азимуту.

Так как, дифференциальная ЭПР зависит от угла наблюдения и расположения отражающих участков местности, то с учетом выражений (4) по основному лепестку диаграммы направленности составит:

$$\sigma_{MB} = \sigma_0 A_{MB} G^2 (\theta_r + \theta_e) = \sigma_0 \Delta R_g R_g \theta_a G^2 (\theta_r + \theta_e), \tag{5}$$

а для первого бокового лепестка ДН с учетом величины его уровня (SL):

$$\sigma_{SL} = \sigma_0 A_{SL} (SL)^2 = \sigma_0 \Delta R_g \pi R_g (SL)^2, \tag{6}$$

где G – коэффициент усиления антенны; σ_0 – коэффициент обратного отражения от земной поверхности.

Результаты моделирования в соответствии с выражениями (5) и (6) представлены на рис. 2. Анализ рис. 2 показывает, что основной лепесток вносит существенно больший вклад в величину дифференциальной ЭПР по

сравнению с первым боковым лепестком. Данная величина зависит от дальности и изменяется в диапазоне от 8 до 17 дБ/см.

Для того чтобы учесть изменение полной дифференциальной ЭПР в зависимости от дальности, выражение (1) в соответствии с [12] может быть преобразовано к виду:

$$\sigma_c(R) = \frac{\sigma_{MB} + \sigma_{SL}}{(1 + (R / R_h)^4)}, \tag{7}$$

где R_h – дальность радиогоризонта.

Дальность радиогоризонта в соответствии с [5], а также принятыми ограничениями определяется высотой полета воздушной цели (h_t) и фазового центра антенны (h_r):

$$R_h = \sqrt{2\kappa h_r h_t}, \tag{8}$$

где $\kappa = 4/3$ – постоянная величина.

Таким образом, зависимость полной дифференциальной ЭПР в соответствии с выражением (1) в графическом виде имеет вид, представленный на рис. 3.

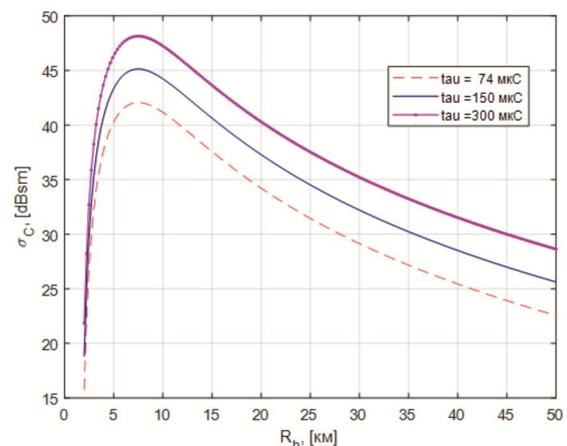


Рис. 3. Полная дифференциальная ЭПР

Анализ рис. 3 показывает, что в S-диапазоне максимум отражений от земной поверхности приходится на диапазон 5-10 км и составляет порядка 48 дБ. С увеличением дальности их величина уменьшается. Стоит

отметить, что данная величина слабо зависит от дли-

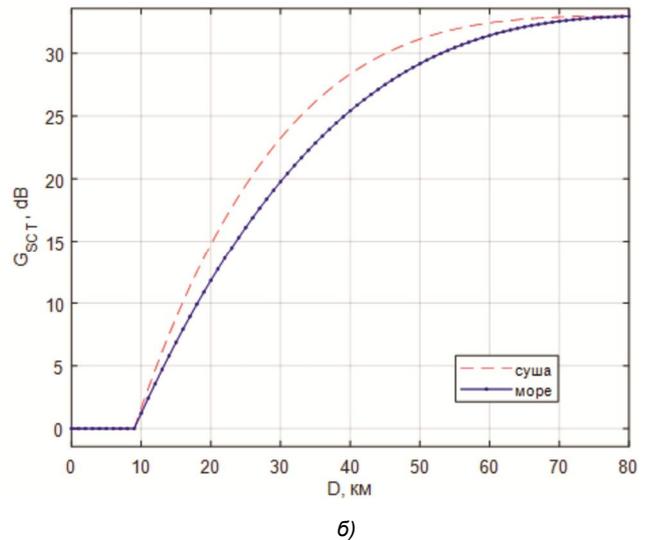
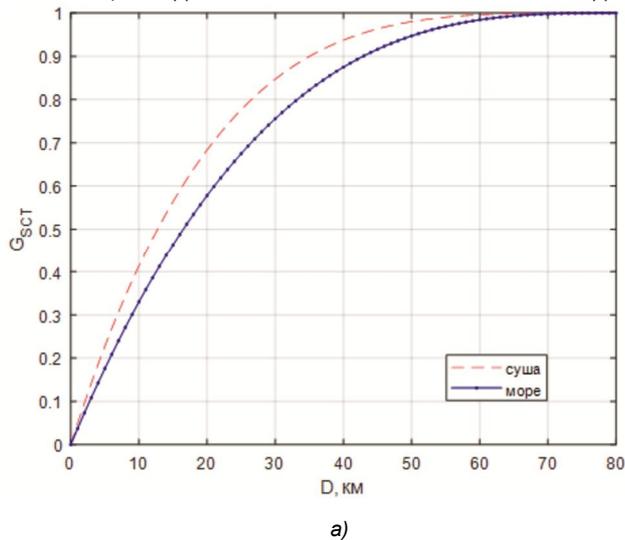


Рис. 4. Формирование управляющего напряжения:

а – закон изменения коэффициента усиления приемного тракта; б – импульс ВАРУ

тельности импульса. Так при ее увеличении в два раза величина полной дифференциальной ЭПР увеличивается не значительно, на 3-4 дБ. Кроме того максимум полной дифференциальной ЭПР в 48 дБ достигает на дальности 9 км

При обнаружении воздушных целей на фоне поверхности земли или в непосредственной близости к морской поверхности законы изменения коэффициента усиления приемного тракта имеют некоторые особенности, при этом в общем виде они определяются в соответствии с выражением:

$$G_{SCT}(D) = 1 - [c_g(D_{max} - D)]^n, \quad (9)$$

где $c_g = (1 - g_{att}) / (D_{max})^4$ – затухание на удалении D_{max} ; g_{att} – затухание в начале дистанции; n – определяется типом поверхности (для суши $n = 4$, для моря $n = 3$).

Закон управления коэффициентом усиления приемного тракта РЛС по результатам анализа рис. 3 и учетом выражения (9) представлены на рис. 4. Величина остатков мешающих отражений на выходе приемного устройства РЛС представлена на рис. 5.

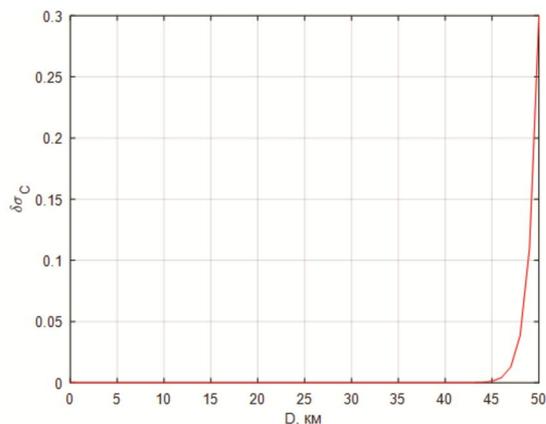


Рис. 5. Остатки мешающих отражений на выходе усилителя

Анализ рис. 5 показывает, что в результате оценки уровня воздействия мешающих отражений на приемный

тракт РЛС (рис. 3) при наличии в приемном тракте системы ВАРУ (рис. 4) на его выходе все же имеется незначительный, не полностью скомпенсированный остаток отражений от поверхности земли, величина которого зависит от дальности.

Заключение

Таким образом, в ходе работы предложен подход к определению параметров закона управления коэффициентом усиления канала приемного тракта многоканальной РЛС на основе оценки мощности отражений, которые поступают на вход приемного тракта РЛС с учетом ее реальных технических характеристик. В его основе лежит оценка полной дифференциальной эффективной площади рассеяния. По результатам проведенного анализа в S-диапазоне можно сделать ряд выводов:

- дифференциальная ЭПР по главному лучу и первому боковому лепестку ДН отличаются на величину от 8 до 17 дБ/см, которая слабо зависит от дальности и в большей степени определяется размерами зоны сбора мешающих отражений. Данная величина является существенной при обнаружении любых воздушных целей, включая БПЛА;

- в ходе проектирования РЛС стоит особо уделить внимание снижению уровня боковых лепестков ДН, т.к. параметры основного луча ДН не всегда могут быть изменены. Это связано с тем, что они в большей степени определяют зону обзора РЛС и ее тактико-технические характеристики;

- уровень отражений в ближней зоне по результатам расчетов составляет порядка 48 дБ, что несомненно существенно ограничивает динамический диапазон радиоприемных устройств любой РЛС;

- значение полной дифференциальной ЭПР может рассматриваться в качестве основы при формировании параметров управляющего напряжения схем с ВАРУ;

- по результатам моделирования получено, что величина не полностью скомпенсированных остатков мешающих отражений в конце дистанции не превышает минус 14 дБ, что говорит о корректности предложенного подхода.

Литература

1. Вовшин Б.М. Принципы построения обзорного радиолокатора обнаружения беспилотного летательного аппарата Вестник ВКО, 2019, № 3. С. 38-46.

2. Бакулев П.А., Степин В.М. Методы и устройства селекции движущихся целей. М.: Радио и связь, 1986. 288 с.

3. Бакулев П.А. Радиолокационные системы. Изд. 2-е, перераб. и доп. Сер. Учебник для ВУЗов. М.: Радиотехника, 2007. 375 с.

4. Разработка первых систем защиты РЛС от пассивных помех (исторический обзор) / П.А. Бакулев [и др.]. Радиотехника. 2016. № 5. С. 142-151.

5. Сколник М. Справочник по радиолокации. Книга 1 / Под ред. М. Сколника. 3-е издание. Перевод с англ. под общей редакцией В.С. Вербы. В 2-х книгах. М.: Сов. радио, 2014. 528 с.

6. Сколник М. Справочник по радиолокации. Том 3 / Под ред. М. Сколника. Перевод с англ. под общей редакцией К.Н. Трофимова. В 4-х томах. М.: Техносфера, 1979. 672 с.

7. Фитасов Е.С. Пространственно-временная обработка сигналов в малогабаритных мобильных радиолокационных системах обнаружения низколетящих воз-

душных объектов: дис. доктора тех. наук. Нижегородский государственный университет им. Н.И. Лобачевского, 2018. 378 с.

8. Степанов М.А. Влияние флуктуаций скорости ветра в турбулентной атмосфере на характеристики обнаружения РЛС с СДЦ: дис. канд. физ.-мат.наук. Красноярск, 2009. 187 с.

9. Межобзорная компенсация дискретных мешающих отражений с формированием карты помех и накоплением решений / А.С. Солонар, С.А. Горшков, П.А. Хмарский, В.А. Вашкевич. Доклады БГУИР. 2015. № 4 (90). С. 74-79.

10. Ляшенко В.А. Радиолокационная станция 19Ж6 (СТ-68У). В.А. Ляшенко, Л.Ю. Боскутис, Г.И. Тылец. Учебное пособие МО СССР. М.: Военное издательство. 1992.

11. Sumanta, Pal. A Novel Scheme of Digital Instantaneous Automatic Gain Control (DIAGC) for Pulse Radars. Pal. Sumanta, Shanmugam. Nirmala, Kumar. Mohit, P. Radhakrishna. Presented at International Symposium of India 2011.

12. Mahafza B.R., Radar Signal Analysis and Processing Using MATLAB, Taylor & Francis Group, Boca Raton, 2009.

Уважаемые авторы!

Редакция научно-технического журнала «Цифровая обработка сигналов» просит Вас соблюдать следующие требования к материалам, направляемым на публикацию:

1) Требования к текстовым материалам и сопроводительным документам:

- Текст – текстовый редактор Microsoft Word, формулы – в редакции MathType.
- Таблицы и рисунки должны быть пронумерованы. На все рисунки, таблицы и библиографические данные указываются ссылки в тексте статьи.
- Объем статьи до 12 стр. (шрифт 12). Для заказных обзорных работ объем может быть увеличен до 20 стр. Название статьи на русском и английском языках.
- Рукопись статьи сопровождается: краткой аннотацией на русском и английском языках; номером УДК; сведениями об авторах (Ф.И.О., организация, должность, ученая степень, телефоны, электронная почта); ключевыми словами на русском и английском языках; актом экспертизы (при наличии в вашей организации экспертной комиссии).

2) Требования к иллюстрациям:

- Векторные (схемы, графики) – в графическом редакторе Corel DRAW.
- Растровые (фотографии, рисунки) – М 1:1, разрешение не менее 300dpi, формат tiff.

НОВЫЕ КНИГИ



Воскресенский Д.И., Добычина Е.М.

Цифровые антенные решетки: Монография. М.: Изд-во Радиотехника, 2020 г. 240 с.: ил.

Рассмотрен новый класс антенных систем – цифровых антенных решеток бортовых радиолокационных комплексов, позволяющих повысить энергетический потенциал за счет использования новых методов оптимизации режима работы высокоэффективных активных устройств и обеспечения высокой точности формирования амплитудно-фазового распределения с помощью предложенной системы автоматической калибровки. Представлены результаты экспериментальных исследований возможностей цифрового диаграммообразования и точностных характеристик калибровки макета цифровой решетки.

Для научных работников и инженеров, занимающихся исследованиями в области разработки, создания и применения цифровых решеток в современных радиоэлектронных системах. Может быть рекомендована в качестве учебного пособия студентам радиотехнических специальностей, а также аспирантам и магистрантам по направлениям «Радиотехника», «Радиофизика и электроника».

НЕКОТОРЫЕ ВОПРОСЫ ПРИМЕНЕНИЯ АДАПТИВНОГО МЕТОДА ДЕКОМПОЗИЦИИ НЕСТАЦИОНАРНЫХ СИГНАЛОВ НА ЭМПИРИЧЕСКИЕ МОДЫ

Егоров В.В., д.т.н., старший научный сотрудник, заведующий кафедрой Санкт-Петербургского государственного университета аэрокосмического приборостроения

Клионский Д.М., к.т.н., доцент, доцент кафедры математического обеспечения и применения ЭВМ СПбГЭТУ «ЛЭТИ», e-mail: klio2003@list.ru

SOME ASPECTS OF APPLYING THE ADAPTIVE TECHNIQUE OF EMPIRICAL MODE DECOMPOSITION OF NON-STATIONARY SIGNALS

Egorov V.V., Klionskiy D.M.

The present paper discusses the issues of adaptive empirical mode decomposition. We represent the modified empirical mode decomposition method that allows us to eliminate decomposition redundancy. The properties of empirical modes are studied. We represent the results of comparing the empirical mode decomposition, Fourier analysis, and wavelet analysis. Some experimental data are shown that deal with studying harmonic signals. The classification of empirical modes is introduced and we suggest the technique of selecting informative components based on information criteria.

Key words: empirical mode decomposition, modified empirical mode decomposition, informative components, information criteria, classification of components.

Ключевые слова: разложение на эмпирические моды, декомпозиция, модифицированный метод декомпозиции на эмпирические моды, информативные компоненты, информационные критерии.

Введение

Обработка и анализ нестационарных сигналов являются неотъемлемой частью таких, например, направлений как речевые технологии, гидроакустика, обработка вибрационных сигналов, геофизические и биомедицинские исследования и пр. Для повышения точности и достоверности анализа этих сигналов желательно использовать специальные подходы и методы, обладающие свойством *адаптивности* к рассматриваемому сигналу.

Из известных на сегодняшний день методов указанному свойству адаптивности наиболее полно соответствует *дискретное вейвлет-преобразование* на основе схемы диадического (двоичного) банка фильтров [1-5]. Для данного преобразования существуют отработанный математический аппарат, в том числе быстрые вычислительные алгоритмы [2-4].

В силу особенностей конструирования вейвлет-базиса на основе масштабных растяжений и сдвигов материнской вейвлет-функции вдоль временной оси появляется возможность *адаптивно*, с учетом свойств гладкости текущего сегмента данных, обрабатывать сигналы путем достаточно точного учета временных особенностей. На рис. 1 представлен пример нестационарного сигнала с локальными временными особенностями.

Следует отметить, что проблемой, порой затрудняющей эффективное практическое применение вейвлет-преобразования, является большое многообразие вейвлетов и неочевидность выбора конкретного вей-

Рассмотрены вопросы адаптивной декомпозиции сигналов на основе разложения на эмпирические моды. Представлен модифицированный метод декомпозиции на эмпирические моды, позволяющий устранить избыточность в разложении. Исследуются свойства эмпирических мод. Представлены сравнительные характеристики методов декомпозиции на эмпирические моды, анализа Фурье и вейвлет-анализа. Приведены некоторые экспериментальные данные по исследованию гармонических сигналов. Также вводится классификация эмпирических мод и предлагается способ отбора информативных компонент в разложении на основе информационных критериев.

лета (его типа, длины вейвлет-фильтра) для решения поставленной задачи, в том случае, когда нет формализованных или эмпирических критериев выбора вейвлет-базиса.

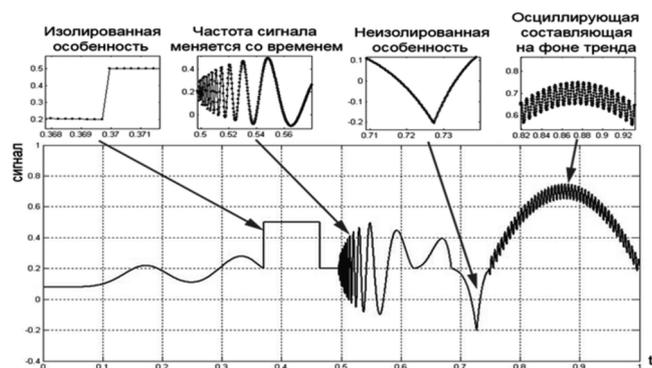


Рис. 1. Пример нестационарного сигнала с локальными особенностями

В статье рассматривается адаптивный мультимасштабный метод обработки и анализа сигналов посредством использования системы функций, называемый *декомпозицией на эмпирические моды (ДЭМ)* [6-11]. Обладая свойствами, подобными вейвлет-разложению, ДЭМ не требует априорного выбора базиса. Таким образом, разложение с использованием такого рода системы

функций проводится с учетом *локальных особенностей* (экстремумы и нули сигнала) и *структуры* конкретного сигнала. При этом структура исследуемого сигнала характеризуется наличием в нем трех основных типов компонент – шумовых, трендовых и компонент-шаблонов. Мультимасштабность подхода состоит в представлении исходного сигнала набором компонент с разным временным масштабом (временным разрешением).

Ниже представлены следующие основные положения статьи:

Рассмотрены особенности класса функций, называемых эмпирическими модами, а также некоторые теоретические и практические аспекты метода ДЭМ;

Представлены этапы разложения сигнала по методу ДЭМ и выполнено сравнение анализа Фурье, ДЭМ и вейвлет-анализа;

Предложена модификация алгоритма ДЭМ, которая позволяет избежать избыточности в разложении и выделить информативные компоненты;

Предложена обобщенная классификация эмпирических мод с учетом их физического смысла и появления в разложении сигнала;

Разработан метод отбора информативных компонент в разложении с использованием информационных критериев и предложенной автором классификации эмпирических мод.

Понятие эмпирической моды

Эмпирическая мода (ЭМ), называемая также *внутренней модовой функцией*, является особым типом математических функций [6, 7, 10]. ЭМ – функция, заданная дискретно или непрерывно на интервале существования сигнала. Имеет, в общем случае, произвольную аналитическую запись, но при этом удовлетворяет *двум необходимым условиям*:

1) общее число максимумов и минимумов такой функции должно быть либо равно числу нулей функции, либо отличаться от числа нулей по модулю не более, чем на единицу. Таким образом,

$$N_{\max} + N_{\min} = N_{\text{zero}} \pm 1 \text{ или } N_{\max} + N_{\min} = N_{\text{zero}}, \quad (1)$$

где N_{\max} , N_{\min} , N_{zero} – число максимумов, минимумов и нулей функции соответственно.

2) *локальное (мгновенное) среднее значение ЭМ*, определяемое в виде полусуммы двух *огibaющих* – верхней, полученной путем интерполяции найденных локальных максимумов, и нижней, полученной интерполяцией найденных локальных минимумов, – должно быть меньше или равно заранее определенного порогового значения η . Указанное пороговое значение зависит от машинной точности ε и погрешностей, связанных с получением, преобразованием и передачей сигнальной информации. Аналитически вышеуказанное условие записывается в виде:

$$m(k) = \frac{U(k) + L(k)}{2} \leq \eta, \quad k = \overline{0, N-1}, \quad (2)$$

где k – дискретное нормированное время (номер отсчета сигнала); $m(k)$ – локальное среднее значение,

$U(k)$ и $L(k)$ – значения верхней и нижней *огibaющих* сигнала в k -й дискретный момент времени; N – общее количество отсчетов на длительности сигнала (длина сигнала); η – некоторый порог, принятый обработчиком-экспертом и обычно имеющий близкие к нулю значения.

Огibaющие $U(k)$ и $L(k)$ подчиняются следующему соотношению:

$$U(k) \geq c(k); \quad L(k) \leq c(k); \quad k = \overline{0, N-1}, \quad (3)$$

где $c(k)$ – ЭМ.

Условие равенства нулю локального среднего накладывает ограничения на симметрию *огibaющих* относительно оси абсцисс. Для реальных сигналов соотношение между *огibaющими* имеет вид:

$$U(k) = -L(k) \pm \gamma, \quad k = \overline{0, N-1}, \quad (4)$$

где $\gamma \geq 0$ – некоторая малая константа, связанная с вычислительной точностью.

Из выражений (2) и (4) следует, что константы γ и η аналитически связаны.

Условие равенства нулю локального среднего значения можно также записать в виде:

$$m(k) = \frac{1}{\varepsilon} \int_{k-\varepsilon/2}^{k+\varepsilon/2} c(r) dr = 0, \quad (5)$$

где r – переменная, представляющая собой аналоговое время; величина ε может варьироваться и определяет диапазон, в пределах которого среднее значение ЭМ равно нулю.

Необходимо отметить, что существуют альтернативные методы определения локального среднего значения в окрестности некоторой точки. В одном из них [7] для вычислений используется интегральная функция вида:

$$m_{\delta}[f(x)] = \frac{1}{\delta} \int_{x_{\delta}} f(y) dy, \quad (6)$$

где $m_{\delta}[f(x)]$ – локальное среднее, $f(x)$ – функция, для которой вычисляется локальное среднее, $r_3(k)$ представляет собой δ – окрестность точки с абсциссой x ,

$$\text{задаваемую как } \left[x - \frac{\delta}{2}; x + \frac{\delta}{2} \right].$$

Функция, удовлетворяющая первому, но не удовлетворяющая второму приведенному выше условию, является так называемой «слабой» ЭМ (*англ.* – weak intrinsic mode function). «Слабая» ЭМ имеет суммарное число максимумов N_{\max} и минимумов N_{\min} равное числу нулей N_{zero} или отличающееся от числа нулей на ± 1 , но при этом может обладать несимметричными верхней и нижней *огibaющими*.

На рис. 2 приведены примеры функций, относящихся к классу ЭМ. Здесь цифрой 1 обозначена верхняя *огibaющая* функции, цифрой 2 – нижняя *огibaющая*, цифрой 3 – локальное (мгновенное) среднее, определенное как полусумма обеих *огibaющих*, 4 – ЭМ, для которой построены *огibaющие* 1 и 2.

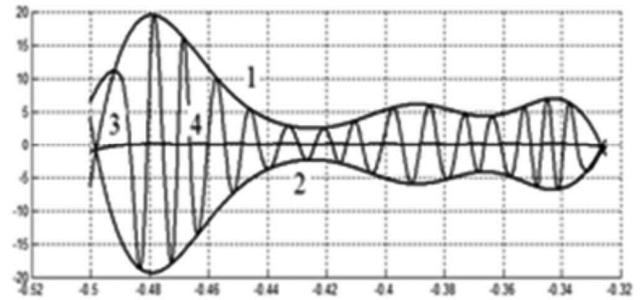
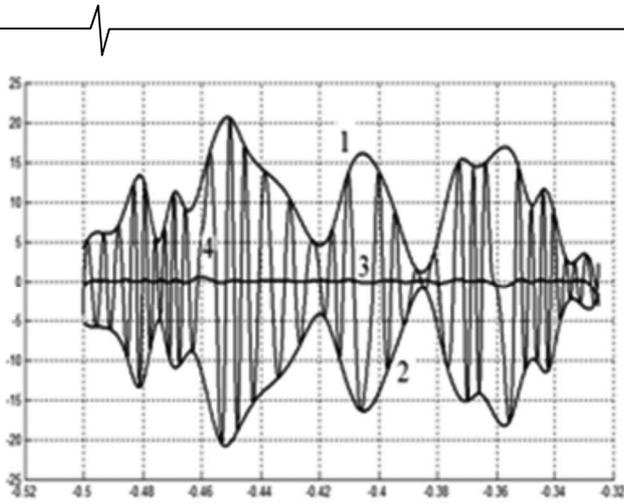


Рис. 2. Примеры ЭМ с иллюстрацией огибающих и их полусуммы

Ниже, в качестве примера, приведены два аналитических выражения для ЭМ, представляющих собой амплитудно-модулированные функции с экспоненциальными огибающими:

- 1) $s(t) = e^{\varepsilon \cos(t)} \sin(\varepsilon \sin(t))$.
- 2) $s(t) = e^{-\varepsilon t} \cos(\omega_0 t)$.

Метод декомпозиции на эмпирические моды

При реализации метода ДЭМ базисные функции-компоненты *извлекаются* непосредственно из исходного сигнала, следовательно, такой базис всегда *уникален, апостериорен* и *адаптивен* [7-10]. В силу отсутствия необходимости выбора априорного базиса для проведения разложения сигнала, данный метод можно отнести к *непараметрическим* методам обработки и анализа данных. Кроме того, метод ДЭМ ориентирован на так называемую *нелинейную обработку данных*, направленную на поиск нелинейных статистических связей. При этом метод ДЭМ более эффективен, чем методы классического корреляционного и спектрального анализа сигналов, ориентированные на поиск линейных закономерностей в сигналах.

Применительно к ДЭМ термин «адаптивность» имеет иной смысл, чем в его классической трактовке в цифровой обработке сигналов, где имеет место понятие «адаптивный фильтр» [12-14]. *Адаптивный* фильтр – нелинейная система с обратной связью, параметры которой *подстраиваются* к сигналу с заранее неопределенной статистической моделью в процессе его обработки. Адаптивность применительно к методу ДЭМ заключается в его приспособлении к локальным особенностям и свойствам данных.

ДЭМ представляет собой математически вложенный итерационный процесс, причем число внешних итераций можно заранее аналитически определить [6] с достаточно высокой точностью, а число внутренних итераций определяется на основе специальных критериев [6, 7].

При реализации алгоритма ДЭМ используются численные методы и алгоритмы теории многочленов, линейной алгебры, математического анализа и т.д.

Сигналы содержат в себе локальные особенности, отражающие их природу и структуру, а метод ДЭМ адаптивно учитывает эти особенности в разложении сигнала. Структура локальных особенностей может

быть представлена на разных масштабах (уровнях разложения) и с различной вычислительной точностью.

В случае разложения произвольного дискретного сигнала по базису Фурье и вейвлет-базису используется следующее выражение:

$$c(k) = \sum_i c_i f_{i\text{det}}(k), \quad (7)$$

где $c(k)$ – исходный сигнал; k – номер отсчета (дискретное нормированное время); c_i – i -й коэффициент разложения по базису; $f_{i\text{det}}(k)$ – i -я детерминированная априорно известная базисная функция.

В общем случае, отдельные компоненты $f_{i\text{det}}(k)$ сигнала $s(k)$ могут исключаться из рассмотрения при его восстановлении в соответствии с выражением (7). Причиной является, например, удаление (после идентификации) шумовых компонент в сигнале либо выделение тренда, который анализируется отдельно.

Для ДЭМ выражение декомпозиции сигнала может быть представлено в виде:

$$s(k) = \sum_i b_i f_i(k), \quad (8)$$

где b_i – i -й коэффициент разложения на основе ДЭМ; $f_i(k)$ – i -я априорно неизвестная функция, извлекаемая из исходного сигнала $s(k)$.

Как отмечено выше, в данном случае базисные функции формируются *апостериорно*, что является преимуществом с точки зрения учета индивидуальных свойств сигналов.

Из выражения (8) следует, что ДЭМ наиболее эффективно применять для обработки *многокомпонентных* сигналов, состоящих из конечного аддитивного набора разномасштабных компонент различной физической природы, занимающих различные частотные полосы. Под разномасштабностью компонент подразумевается их различное разрешение по времени, и именно она определяет целесообразность мультимасштабного подхода к их обработке и анализу.

Широкое использование гармонических функций, базисных вейвлетов и ЭМ делает целесообразным сравнение их свойств. В табл. 1 представлены некоторые сравнительные характеристики методов анализа Фурье, ДЭМ и вейвлет-анализа.

Таблица 1. Сравнение анализа Фурье, ДЭМ и вейвлет-анализа

| Характеристика | Анализ Фурье | ДЭМ | Вейвлет-анализ |
|--|-------------------|--|-------------------|
| Обработка нестационарных сигналов | - | + | + |
| Визуализация частотно-временного об- раза сигнала | + | + | + |
| Способность локализации особенностей сигнала | + | + | + |
| Наличие компактного носителя | + | + | + |
| Наличие аналитического представления | + | + | + |
| Происхождение базисных функций | априорно известны | апостериорно известны | априорно известны |
| Сходимость | + | строго не доказана (подтвер- ждена опытным путем [6-8]) | + |
| Наличие быстрых вычислительных алго- ритмов | + | + | + |

Алгоритм декомпозиции сигнала

Шаг 1. Рассматривается так называемый *текущий остаток* $r_p(k)$. Здесь $r_1(k)$ – первый остаток, который совпадает с исходным сигналом $s(k)$. Определяются его локальные экстремумы и формируются два конечных набора:

$$\{M_i\}, i = 1, 2, 3, \dots; \{m_i\}, i = 1, 2, 3, \dots;$$

где $\{M_i\}$ и $\{m_i\}$ – множество максимумов и минимумов соответственно.

Далее по найденным экстремумам строятся две огибающие посредством интерполяции кубическими сплайнами локальных максимумов и локальных минимумов:

$$U_i(k) = f_U(M_i, k), k = 0, \dots, N - 1; \tag{10}$$

$$L_j(k) = f_L(m_j, k), k = 0, \dots, N - 1; \tag{11}$$

где $U_i(k)$ и $L_j(k)$ – верхняя и нижняя огибающие, построенные соответственно по найденным в (9) локальным максимумам и минимумам с помощью сплайн-интерполяции; j – номер итерации отсеивания.

Затем определяется полусумма двух огибающих (локальное среднее значение, зависящее от времени) по выражению (12):

$$e_j(k) = 0.5(U_j(k) + L_j(k)), \tag{12}$$

и выполняется переход к шагу 2.

Шаг 2. Найденное среднее значение вычитается из текущего остатка согласно выражению (13), и полученный результат $h_j(k)$ оказывается кандидатом на то, чтобы стать очередной ЭМ:

$$\begin{cases} h_1(k) = r_p(k); \\ h_j(k) = h_{j-1}(k) - e_j(k), j > 1. \end{cases} \tag{13}$$

Для перехода к шагу 3 необходимо проверить два аналитических условия отнесения функции к классу ЭМ, приведенные выше. Если оба условия выполняются, то осуществляется переход к шагу 3. Если хотя бы одно из них нарушено, то осуществляется возврат к шагу 1, при этом в качестве текущего остатка выступает полученный на втором шаге результат. Так реализуется *процесс отсеивания*, который может быть записан в следующем виде:

$$\begin{cases} h_{j+1}(k) = h_j(k) - e_j(k), j \geq 1; \\ \dots \\ c_p(k) = h_{iter_p}(k), \end{cases} \tag{14}$$

где $e_j(k)$ – среднее значение функции на j -й итерации процесса отсеивания; $h_j(k)$ – текущий результат на j -й итерации отсеивания; $iter_p$ – общее число итераций отсеивания для p -й ЭМ.

На итерации с номером $iter_p$ процесс извлечения p -й ЭМ прекращается и выполняется переход к шагу 3.

Шаг 3. После извлечения ЭМ в ее окончательном виде осуществляется ее вычитание из текущего остатка для формирования нового (то есть для обновления остатка):

$$r_{p+1}(k) = r_p(k) - c_p(k), p > 1, \tag{15}$$

где $c_p(k)$ – полученная ЭМ; $r_p(k)$ – текущий остаток;

$r_{p+1}(k)$ – новый остаток.

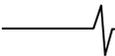
Шаг 4. Осуществляется переход к шагу 1, где в качестве функции, из которой будут извлекаться ЭМ с более высокими номерами, то есть в качестве текущего выступит остаток, полученный на третьем шаге алгоритма, то есть $r_{p+1}(k)$.

Представленный выше алгоритм назовем алгоритмом *классической ДЭМ*.

Модификация алгоритма ДЭМ

В классическом алгоритме ДЭМ применяются кубические сплайны в качестве средства интерполяции обеих огибающих. Кубические сплайны гарантируют сравнительно невысокую вычислительную сложность, по сравнению со сплайнами более высоких порядков. При этом кубические сплайны являются гладкими функциями.

Несмотря на это, при реализации классического алгоритма ДЭМ возможно возникновение ошибок в определении местоположения экстремумов в соотношениях (9)-(11). Неточное определения их местоположения может привести к тому, что в разложении получится *избыточность*. Она может проявиться в том, что будут извлечены дополнительные компоненты, не имеющие физического смысла и связанные с неточным нахождением



экстремальных точек. В связи с этим предлагается модификация алгоритма ДЭМ, который более точно определит экстремумы и позволит избавиться от избыточности.

Очевидно, максимальная ошибка e_{\max} определения экстремума составляет половину интервала дискретизации:

$$e_{\max} = T/2, \quad (16)$$

где T – интервал (период) дискретизации.

С учетом вышесказанного, предлагается новый метод определения экстремумов с использованием *квадратичной интерполяции*, описываемый следующим алгоритмом:

Шаг 1. Выбираются три первых отсчета исходного сигнала $s(k)$;

Шаг 2. По данным трем отсчетам строится квадратичный полином вида $ak^2 + bk + c$, где a, b, c – коэффициенты, подлежащие определению. Для их нахождения необходимо решить систему из трех уравнений с тремя неизвестными;

Шаг 3. Если коэффициент при старшей степени полинома отличен от нуля ($a \neq 0$), то среди трех точек есть экстремум, который располагается посередине. Найденная точка соответствует последовательности из

трех точек. Координата экстремума уточненным способом определяется как координата вершины квадратичной параболы;

Шаг 4. Если коэффициент при старшей степени равен нулю, то все три отсчета лежат на одной прямой с положительным или отрицательным наклоном, и среди них нет экстремума. Далее необходимо на одну точку сдвинуть скользящее окно, включающее в себя в исходном варианте первые три точки, среди которых ищут экстремальную.

Затем шаги 1-4 повторяются.

На рис. 3 приведен пример модификации ДЭМ гармонического сигнала с извлечением одной гармоники (рис. 3, а) и трех гармоник (рис. 3, б).

Из графиков на рис. 3 следует, что в разложении гармонического сигнала, состоящего из одной гармоники, появляется одна ЭМ. Соответственно при декомпозиции мультигармонического сигнала из трех гармоник формируются три компоненты, каждая из которых соответствует одной гармонике, то есть одной ЭМ. Следует отметить, что при разложении сигнала (рис. 3, б) гармоники появляются в соответствии с их частотами и представлением ДЭМ как банка фильтров. В этом случае при использовании классического алгоритма ДЭМ формируется разложение вышеуказанных сигналов, представленное на рис. 4.

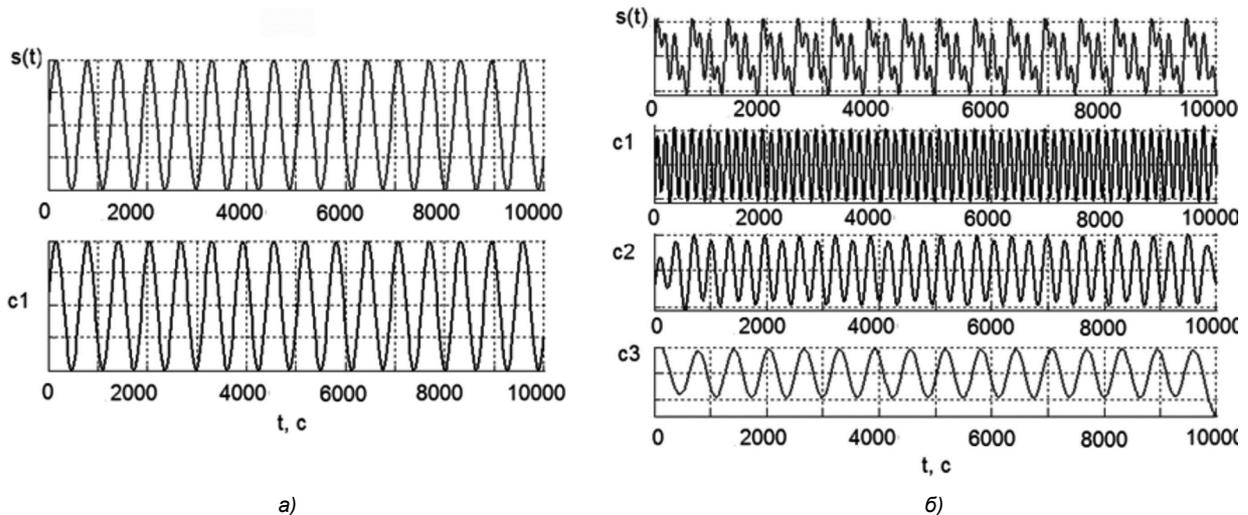


Рис. 3. Пример модификации ДЭМ гармонического сигнала: а – с извлечением одной ЭМ; б – с извлечением 3-х ЭМ

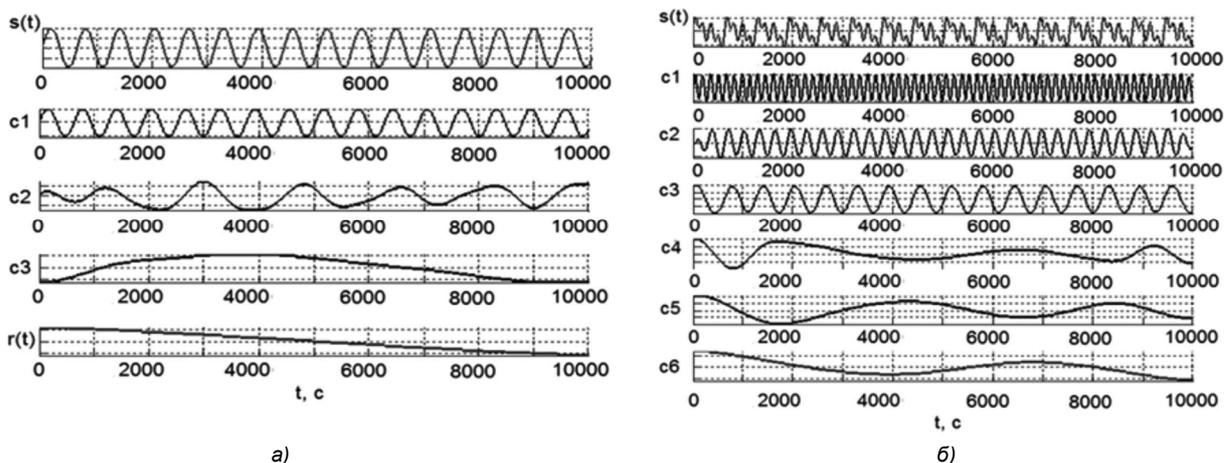


Рис. 4. Пример классической ДЭМ гармонического сигнала: а – с извлечением одной ЭМ; б – с извлечением 3-х ЭМ

В рассматриваемом эксперименте при разложении монокомпонентного сигнала (рис. 4, а) получается четыре компонента. При этом три последних компоненты не имеют физического смысла, и их появление объясняется ошибками в точности определения экстремумов на каждом уровне разложения. Аналогично на рис. 4, б компоненты 4-я, 5-я и 6-я также не имеют физического смысла и появляются за счет избыточности разложения.

Таким образом, приведенные примеры иллюстрируют преимущество предложенной модификации по сравнению с классическим алгоритмом ДЭМ, заключающееся в отсутствии избыточности в разложении.

Классификация эмпирических мод и отбор информативных составляющих в разложении

При решении задач предварительной обработки сигналов (очистка от шума, выделение тренда и пр.) и анализа структуры сигналов с помощью метода ДЭМ необходима предварительная классификация всех извлеченных из сигнала ЭМ. Применительно к методу ДЭМ данная задача ранее не рассматривалась.

В рамках поставленной задачи, для произвольного сигнала все ЭМ могут быть разбиты на две категории:

- 1) *основные ЭМ*;
- 2) *трендовые ЭМ*, включая *компенсирующие* и *истинные трендовые ЭМ*.

Основные ЭМ в разложении отражают структуру и особенности конкретного сигнала. К их числу относятся шумовые ЭМ и ЭМ-шаблоны. Появление шумовых ЭМ в разложении собственно и объясняется наличием в исходном сигнале шума, а ЭМ-шаблоны связаны непосредственно с полезным сигналом и входящими в него компонентами, несущими информацию об исходном сигнале. Введенное здесь понятие «ЭМ-шаблон» описывает некоторую функцию, характерную для данного типа сигналов, например, любую из элементарных гармоник мультигармонического сигнала. В качестве другого примера можно рассмотреть аддитивную смесь ЛЧМ-сигнала, элементарной гармоник и шума, где ЛЧМ-сигнал и гармоника – суть эмпирические моды.

Трендовые ЭМ являются медленно меняющимися функциями, характеризующими изменение во времени локального среднего значения сигнала. Среди них, согласно введенной выше классификации, можно выделить *истинные трендовые ЭМ*, описывающие динамику среднего значения, и так называемые *компенсирующие ЭМ*, которые могут возникать в разложении в том случае, когда тренд в сигнале отсутствует. Истинные трендовые ЭМ появляются, например, при разложении суммы гармонического сигнала и полиномиального тренда. Результат несовершенства самого алгоритма ДЭМ, критериев остановки процесса отсеивания, неточностей вычислений в рамках предлагаемых здесь обозначений является причиной появления компенсирующих ЭМ в разложении сигнала.

Компенсирующие ЭМ создают избыточность в разложении, т.к. проведенные автором экспериментальные вычисления показывают, что в сумме они дают функ-

цию, очень близкую к нулю, то есть компенсируют друг друга. Одной из наиболее типичных причин появления компенсирующих ЭМ являются побочные эффекты интерполяции огибающих, такие как появление «всплесков», сильные осцилляции вблизи концов сигнала (краевые эффекты). Для частичного преодоления вышеперечисленных проблем существуют специальные методы, такие, например, как подавление краевых эффектов с помощью зеркального отображения экстремумов, наиболее близко расположенных к границам сигнала; тестирование различных критериев остановки процесса отсеивания; уточнение местоположения экстремумов для построения огибающих.

На практике часто возникают ситуации, когда необходимо идентифицировать *информативные* компоненты сигнала. Разложение сигнала может интерпретироваться как построение его математической модели в виде конечного набора ЭМ и результирующего неразлагаемого остатка. При этом номер ЭМ представляет собой *порядок* искомой математической модели p . Вторым фактором, влияющим на оптимальный выбор порядка, выступает дисперсия D сигнала ошибки, определенного как разность исходного сигнала и тех ЭМ, что входят в модель.

По аналогии с известными параметрическими методами спектрального оценивания для нахождения порядка модели предлагается применить *информационные критерии* (ИК). В общем случае целый ряд ИК можно записать в следующем обобщенном виде:

$$IC(p, N, D) = N \ln D + g(p, N), \quad (17)$$

где $IC(\dots)$ – информационный критерий; N – число отсчетов в сигнале; p – порядок модели; D – дисперсия ошибки аппроксимации сигнала математической моделью; $g(p, N)$ – некоторая функция двух переменных p и N . На основании выражения (17) был получен и широко используется набор ИК [15]. В качестве примера может быть приведен критерий Акаике:

$$\alpha = 0.1 \quad (18)$$

При использовании критерия Акаике оптимальным считается то значение порядка, которое доставляет *минимум* выражению (18):

$$p_{opt} = \arg \min_p AIC(p, N, D), \quad p = p_{min}, \dots, p_{max}, \quad (19)$$

где p_{min} , p_{max} – соответственно минимально и максимально возможные значения порядка модели. Обычно в отсутствие других априорных сведений полагается $p_{min} = 1$, величина p_{max} определяется исходя из конкретной задачи более сложными способами [15].

В качестве экспериментального примера рассмотрим мультигармонический сигнал, состоящий из трех элементарных гармоник, в смеси с гауссовским шумом (рис. 5).

Среднеквадратическое отклонение шума установлено равным 0.3. Частоты гармоник образуют геометрическую прогрессию со знаменателем 2. При разложении сигнала было получено 11 компонент. В соответствии с [16], медианная оценка СКО шума по первой ЭМ составляет 0.2876.

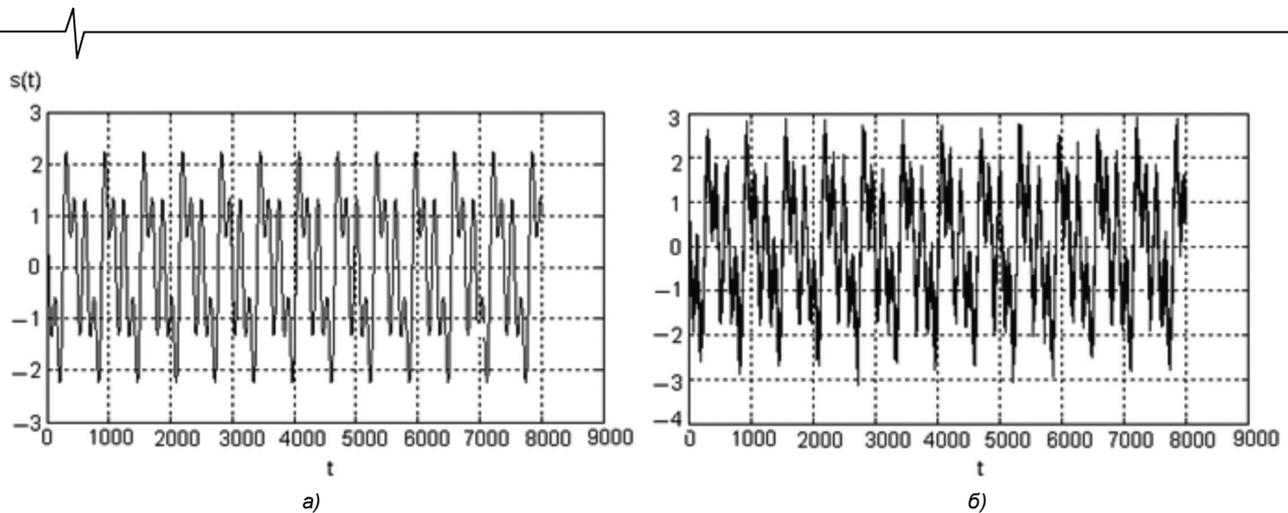


Рис. 5. Мультигармонический сигнал (слева) и его копия с аддитивным гауссовским шумом (справа)

При заданных условиях зависимость критерия Акаике от порядка модели приобретает вид, показанный на рис. 6.

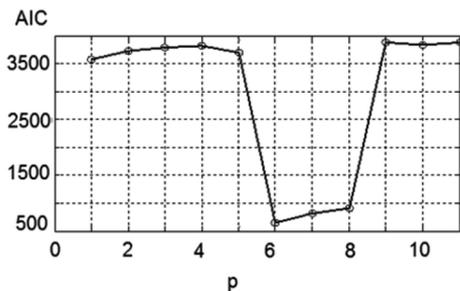


Рис. 6. Зависимость значений критерия АИС от порядка модели p

Нумерация ЭМ в данном примере ведется в порядке возрастания от высокочастотных к низкочастотным компонентам. Компонента 1 на рис. 6 соответствует самой первой компоненте в разложении, компонента 2 – второй и т.д. Представленный метод дает возможность выбора порога для установления категории ЭМ (шумовые, трендовые, ЭМ-шаблоны).

На графике можно видеть три характерные точки, соответствующие 6-й, 7-й, 8-й элементарным гармоникам, значения критерия АИС для которых существенно меньше, чем для всех остальных. Это объясняется малой дисперсией ошибки аппроксимации исходного сигнала любой из этих трех гармоник. Остальные компоненты соответствуют либо *шумовым* (с 1-й по 5-ю), либо *компенсирующим* (с 9-й по 11-ю) ЭМ. Для них соответствующая дисперсия ошибки велика и велико значение критерия Акаике. Для случая, показанного на рис. 6, можно ввести некоторое пороговое значение. В этом случае те ЭМ, для которых значение критерия меньше порога, будут отнесены к *ЭМ-шаблонам*, а остальные – либо к *шумовым*, либо к *компенсирующим* (в более общем случае – к *трендовым*).

Заключение

В статье рассмотрены свойства класса функций, называемых эмпирическими модами. Представлен метод декомпозиции на эмпирические моды. Показано, что метод ДЭМ обладает свойством адаптивности, что, в свою очередь, повышает точность анализа и обработки сигналов. Показано, что данные функции имеют важное значение при реализации метода ДЭМ.

Разработана новая модификация метода ДЭМ, позволяющая избежать избыточности в разложении сигнала. Введена классификация ЭМ и предложен способ отбора информативных компонент в разложении на основе информационных критериев.

Литература

1. К. Чуи Введение в вейвлеты; пер. с англ. М.: Мир. 2001. 412 с.
2. С. Малла Вейвлеты в обработке сигналов; пер. с англ. М.: Мир, 2005. 671 с.
3. Смоленцев Н.К. Вейвлет-анализ в Matlab / 3-е изд. М.: ДМК Пресс, 2010. 448 с.
4. Орешко Н.И., Геппенер В.В., Клионский Д.М. Применение гармонических вейвлетов в задачах обработки осциллирующих сигналов. Цифровая обработка сигналов. 2012. № 2. С. 6-15.
5. Клионский Д.М., Орешко Н.И., Геппенер В.В. Оценка спектральной плотности телеметрических данных на основе теории вейвлетов. Цифровая обработка сигналов и ее применения – DSPA'2011: тез. докл. 13-й Междунар. конф., г. Москва, 30 марта – 1 апр. 2011 г. 2011. С. 123-127.
6. Huang N.E., et al. The empirical mode decomposition and the Hilbert spectrum for non-linear and non-stationary time series analysis. Proc. Royal Soc. London. 1998, vol. 454, pp. 903-995.
7. Huang N.E., Shen S.S.P. Hilbert-Huang Transform and Its Applications. World Scientific, 2005. 350 p.
8. Клионский Д.М., Орешко Н.И., Геппенер В.В. Декомпозиция на эмпирические моды и ее использование при анализе дробного броуновского движения. Цифровая Обработка Сигналов. 2008. № 3. С. 37-45.
9. P. Flandrin, G. Rilling, P. Gonsalves Empirical mode decomposition as a filter bank. IEEE Signal Processing Letters. 2004, vol. 11, no. 2, pp. 112-114.
10. N.E. Huang, Z. Shen, S.R. Long A new view of nonlinear water waves: the Hilbert spectrum / Annu. Rev. Fluid Mech. 1999, vol. 31, pp. 417-457.
11. Мясникова Н.В., Берестень М.П. Разложение на эмпирические моды на основе дифференцирования и интегрирования. Цифровая Обработка Сигналов. 2016. № 1. С. 20-23.
12. Сергиенко А.Б. Цифровая обработка сигналов : учеб. пособие / 3-е изд. СПб.: БХВ-Петербург, 2011. 768 с.
13. Джиган В.И. Адаптивная фильтрация сигналов: теория и алгоритмы. М.: Техносфера, 2013, 529 с.
14. Солонина А.И., Клионский Д.М., Меркучева Т.В., Перов С.Н. Цифровая обработка сигналов и Matlab. СПб.: изд. БХВ-Петербург, 2013, 512 с.
15. Stoica, P. Spectral Analysis of Signals / P. Stoica, R. Moses. Upper Saddle River, New Jersey: Prentice Hall, 2005. 427 p.
16. Хьюбер, Дж. Робастность в статистике. пер. с англ. Дж. Хьюбер. М.: Мир. 1984. 304 с.

УДК 621.396.96

ОПТИМАЛЬНОЕ ИЗМЕРЕНИЕ ЗАПАЗДЫВАНИЯ ЭХО-СИГНАЛА СПУТНИКОВОГО ВЫСОТОМЕРА ПО КОРРЕЛИРОВАННЫМ ОТСЧЕТАМ НАБЛЮДЕНИЯ В УСЛОВИЯХ МОРСКОГО ВОЛНЕНИЯ

Жестерев А.Е., к.т.н., начальник отдела АО «Российский институт радионавигации и времени» (АО «РИРВ»), e-mail: zhesterev_ae@irt.ru.

Ипатов В.П., д.т.н., профессор, главный специалист АО «Российский институт радионавигации и времени» (АО «РИРВ»), e-mail: ival1941@yandex.ru

OPTIMAL ESTIMATE OF THE SATELLITE ALTIMETER ECHO-SIGNAL DELAY FROM CORRELATED OBSERVATION SAMPLES UNDER ROUGH SEA STATE

Zhesterev A.E., Ipatov V.P.

An optimal algorithm for estimating the delay of the satellite altimeter echo signal based on dependent discrete observations in a rough sea condition is synthesized and expressions are obtained for the potential accuracy of the estimate. It is shown that in the presence of sea surface waves, the MSK signal yields in the estimate accuracy to LFM-signal noticeably less than it gains under the calm sea state.

Key words: satellite altimeter, Cramer-Rao bound, LFM-signal, probing signal, MSK.

Ключевые слова: спутниковый высотомер, граница Крамера-Рао, зондирующий сигнал, ЛЧМ-сигнал, минимальная частотная манипуляция.

Введение

Спутниковый радиолокационный высотомер – один из основных измерительных инструментов в составе миссий дистанционного исследования Земли из космоса. Высокая точность оценки текущей высоты с помощью подобных приборов в условиях ограничения на пиковую мощность излучения достигается за счет применения широкополосных сигналов с большим частотно-временным произведением. Традиционно в этой роли выступают сигналы с линейной частотной модуляцией (ЛЧМ) [1-6], основным достоинством которых является относительная простота структуры приемопередатчика. В то же время, как установлено в [7], в отсутствие волнения морской поверхности точностные показатели высотомера можно улучшить, обратившись к форматам фазовой манипуляции с непрерывной фазой, примером которых является минимальная частотная манипуляция (МЧМ).

При рассмотрении названных альтернатив наряду с требованием высокой точности измерения запаздывания в рамках лимитированного энергоресурса приходится учитывать два серьезных дополнительных ограничения: концентрацию спектра в отведенной регламентом ИТУ [8] полосе W_r и умеренное (достаточно близкое к единице) значение пик-фактора излучения. В предлагаемой статье в продолжение [7] синтезирован оптимальный алгоритм и определена потенциальная точность оценки запаздывания эхо-сигнала при использовании зондирующего МЧМ-сигнала в условиях волнения морской поверхности.

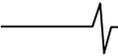
Синтезирован оптимальный алгоритм оценки запаздывания эхо-сигнала спутникового высотомера по зависимым дискретным наблюдениям при взволнованной морской поверхности и получены выражения для потенциальной точности оценки. Показано, что при существенном волнении проигрыш сигнала с минимальной частотной модуляцией в точности измерения запаздывания по отношению к сигналу с линейной частотной модуляцией не превышает долей децибела, т.е. ощутимо меньше выигрыша в условиях спокойного моря.

Корреляционная матрица вектора наблюдений на выходе согласованного фильтра для случая взволнованной морской поверхности

Пусть, как и в [7], $\dot{S}_e(t)$ – комплексная огибающая полезного эхо-сигнала на выходе приемного согласованного фильтра высотомера. Поскольку модель эхо-сигнала традиционно трактуется как суперпозиция элементарных отражений от множества точек освещаемого пятна [9], результирующую комплексную огибающую при волнении морской поверхности в предположении идеальной вертикальной юстировки антенны можно (см. [7]) выразить как:

$$\dot{S}_e(t) = a \sum_i \frac{\dot{S} \left(t + \frac{2(h - r_i - z_i)}{c} \right)}{r_i^2} G(\theta_i) e^{j\psi_i}, \quad (1)$$

где $\dot{S}(t)$ – комплексная огибающая элементарного отражения, обработанного согласованным фильтром, r_i , θ_i и ψ_i – соответственно расстояние от фазового центра антенны, угол видимости с КА и случайная начальная фаза принятого сигнала для i -го элементарного отражателя, h – высота орбиты КА, c – скорость света, $G(\cdot)$ – диаграмма направленности антенны (ДНА), z_i – возвышение i -го индивидуального отражателя над



средним уровнем отражающей поверхности, вызванное волнением, а коэффициент a не зависит от слагаемых суммы. При составлении (1), разумеется, принята во внимание малая ширина луча ДНА. Полагая вслед за [7] колебание $\dot{S}(t)$ действительной функцией и распределение начальных фаз элементарных отражений ψ_i равномерным на отрезке $[0, 2\pi]$, запишем автокорреляционную функцию (АКФ) действительной части комплексной огибающей $U(t)$ эхо-сигнала от всего освещаемого пятна в виде:

$$R_U(t_1, t_2 | \mathbf{z}) = \quad (2)$$

$$= \frac{a^2}{2} \sum_i \frac{S\left(t_1 + \frac{2(h-r_i)}{c} - \frac{2z_i}{c}\right) S\left(t_2 + \frac{2(h-r_i)}{c} - \frac{2z_i}{c}\right)}{r_i^4} G^2(\theta_i),$$

где аргумент за чертой в левой части подчеркивает, что АКФ вычислена при фиксированных возвышениях z_1, z_2, \dots , объединенных в вектор \mathbf{z} , имеющий плотность вероятности $W(\mathbf{z})$. Усредняя (2) по \mathbf{z} , получим

$$R_U(t_1, t_2) = \frac{a^2}{2} \sum_i \int_{-\infty}^{\infty} \frac{S\left(t_1 + \frac{2(h-r_i)}{c} - \frac{2z_i}{c}\right) S\left(t_2 + \frac{2(h-r_i)}{c} - \frac{2z_i}{c}\right)}{r_i^4} \times G^2(\theta_i) W(z_i) dz_i,$$

или при допущении идентичности плотностей вероятности всех возвышений, т.е. $W(z_i) = W(z)$, $z_i = z$, $i = 1, 2, \dots$,

$$R_U(t_1, t_2) = \frac{a^2}{2} \sum_i \int_{-\infty}^{\infty} \frac{S\left(t_1 + \frac{2(h-r_i)}{c} - \frac{2z}{c}\right) S\left(t_2 + \frac{2(h-r_i)}{c} - \frac{2z}{c}\right)}{r_i^4} \times G^2(\theta_i) W(z) dz.$$

После предельного перехода к континуальному множеству отражателей [7] получается

$$R_U(t_1, t_2) = \quad (3)$$

$$= \frac{a^2 \sigma_0}{2} \int_0^{2\pi} \int_0^{\infty} \int_{-\infty}^{\infty} \frac{S\left(t_1 + \frac{2(h-r)}{c} - \frac{2z}{c}\right) S\left(t_2 + \frac{2(h-r)}{c} - \frac{2z}{c}\right)}{r^4} \times W(z) dz G^2(\theta) \rho d\rho d\varphi,$$

где σ_0 – удельная эффективная площадь рассеяния (ЭПР), полагаемая неизменной в пределах засвечиваемого пятна. При аппроксимациях $h-r \approx -\rho^2 / 2h$ и

$$G(\theta) = G \exp\left(-\frac{2}{\gamma} \sin^2 \theta\right), \quad \gamma = \frac{2 \sin^2(\theta_{0.5} / 2)}{\ln 2},$$

где G – коэффициент усиления антенны, а $\theta_{0.5}$ – ширина луча по уровню -3 дБ, (3) приводится к виду

$$R_U(t_1, t_2) = \frac{a^2 G^2}{2h^4} \pi h^2 \sigma_0 \int_0^{\infty} \int_{-\infty}^{\infty} S\left(t_1 - \frac{hx}{c} - \frac{2z}{c}\right) S\left(t_2 - \frac{hx}{c} - \frac{2z}{c}\right) \times$$

$$\frac{\times W(z) dz \exp\left[-\frac{4x}{\gamma(1+x)}\right]}{(1+x)^2} dx,$$

Учтя идентичность АКФ действительной и мнимой компонент комплексной огибающей эхо-сигнала и некоррелированность этих компонент, а также используя для них общее обозначение $R_s(t_1, t_2)$ и переходя к нормированной АКФ зондирующего сигнала $\rho_0(t)$, получим [7]:

$$R_s(t_1, t_2) = \frac{\sigma_n^2 q_0}{2} A_r(t_1, t_2), \quad (4)$$

где σ_n^2 – дисперсия шума на выходе фильтра, q_0 – отношение сигнал-шум по мощности на выходе согласованного фильтра, создаваемое зондирующим сигналом от цели с единичной ЭПР, а функция

$$A_r(t_1, t_2) = \pi h^2 \sigma_0 \int_0^{\infty} \int_{-\infty}^{\infty} \rho_0\left(t_1 - \frac{hx}{c} - \frac{2z}{c}\right) \rho_0\left(t_2 - \frac{hx}{c} - \frac{2z}{c}\right) \times \quad (5)$$

$$\frac{\exp\left[-\frac{4x}{\gamma(1+x)}\right]}{(1+x)^2} dx$$

отличается от аналога в [7] присутствием усреднения по возвышению z .

Следуя основополагающим работам по спутниковой альтиметрии [9-11], в качестве первого приближения плотности вероятности возвышения водной поверхности можно принять гауссовское

$$W(z) = \frac{1}{\sqrt{2\pi}\sigma_z} \exp\left(-\frac{z^2}{2\sigma_z^2}\right),$$

где σ_z – среднеквадратическое отклонение возвышения z . Тогда из (4)-(5)

$$R_s(t_1, t_2) = \frac{\sigma_n^2 q_0}{2} \pi h^2 \sigma_0 \times \int_0^{\infty} \left[\frac{1}{\sqrt{2\pi}\sigma_z} \int_{-\infty}^{\infty} \rho_0\left(t_1 - \frac{hx}{c} - \frac{2z}{c}\right) \rho_0\left(t_2 - \frac{hx}{c} - \frac{2z}{c}\right) \times \quad (6)$$

$$\times \exp\left(-\frac{z^2}{2\sigma_z^2}\right) dz \right] \frac{\exp\left[-\frac{4x}{\gamma(1+x)}\right]}{(1+x)^2} dx.$$

Зависимость этой величины от запаздывания эхо-сигнала τ выявляется после подстановки в (6) значений $t_1 - \tau$ и $t_2 - \tau$ вместо t_1 и t_2 . При взятии n отсчетов действительной/мнимой компоненты на выходе согласованного фильтра с шагом дискретизации δ $t_1 = i\delta - \tau$, $t_2 = j\delta - \tau$, $1 \leq i, j \leq n$, элемент i -й строки и j -го столбца корреляционной матрицы $\mathbf{R}_s(\tau)$ вектора отсчетов действительной/мнимой компоненты комплексной огибающей эхо-сигнала есть $R_s(i\delta - \tau, j\delta - \tau)$. Далее, вследствие независимости эхо-сигнала и шума корреляционная матрица $\mathbf{R}(\tau)$ вектора отсчетов действительной/мнимой компонент комплексной огибающей наблюдаемой смеси эхо-сигнала и шума на выходе согласованного фильтра

$$\mathbf{R}(\tau) = \mathbf{R}_s(\tau) + \mathbf{R}_n,$$

где $\mathbf{R}_n = [R_{n,ij}]$ – симметричная корреляционная матрица шумовых отсчетов и все матрицы имеют размер $n \times n$.

Однопараметрическая оценка запаздывания

Обратимся к гипотетическому сценарию, в рамках которого среднеквадратическое отклонение возвышения σ_z полагается известным и оценке подлежит только время прихода сигнала τ . Согласно [7], граница Крамера-Рао для среднеквадратического отклонения σ_τ оценки $\hat{\tau}$ запаздывания эхо-сигнала, имеющего истинное запаздывание τ_0 ,

$$\sigma_\tau = \sqrt{\text{var}\{\hat{\tau}|\tau_0\}} \geq \frac{1}{\sqrt{Ng}}, \quad (7)$$

где N – число зондирований, используемых при выработке оценки, а точностной фактор

$$g = \text{tr}\left\{\left[\mathbf{R}'(\tau_0) \cdot \mathbf{R}^{-1}(\tau_0)\right]^2\right\} \quad (8)$$

целиком определяется корреляционной матрицей вектора наблюдения $\mathbf{R}(\tau)$ и ее производной в точке τ_0 .

Для сценария невзволнованной морской поверхности при типовом наборе исходных данных (высота орбиты $h = 1000$ км, угловой растров антенны $\theta_0 = 1,2^\circ$, регламентная ширина полосы $W_r = 320$ МГц, число зондирований на одну оценку $N = 100$, шаг дискретизации по времени $1/2W_r$) усредненные по τ_0 кривые зависимости (7) от отношения сигнал-шум q_0 приведены на рис. 1.

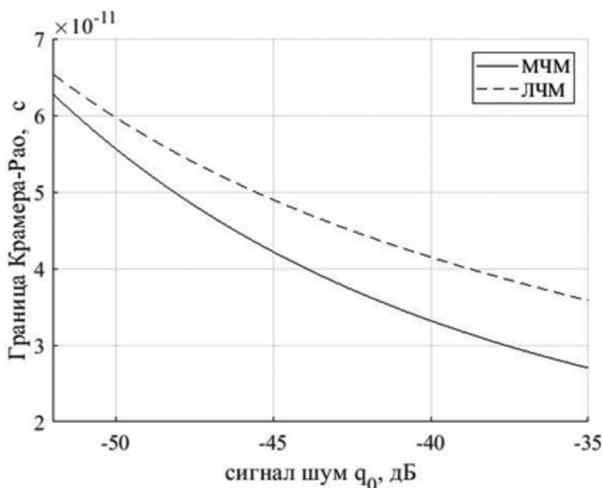


Рис. 1. Усредненные границы Крамера-Рао при отсутствии волнения

Как можно видеть, при спокойной воде, фиксированной регламентной полосе и любых значениях q_0 из рабочего диапазона МЧМ-сигнал реализует более высо-

кую потенциальную точность измерения высоты относительно ЛЧМ-сигнала [7]. В энергетическом выражении этот выигрыш при больших значениях q_0 может достигать 6,5 дБ.

При оговоренном выше однопараметрическом допущении граница Крамера-Рао в случае морского волнения рассчитывается по-прежнему с помощью (7)-(8) с тем отличием, что корреляционная матрица эхо-сигнала определяется из равенства (6). Результаты соответствующих вычислений для ЛЧМ- и МЧМ-сигналов при прежних исходных данных представлены графиками рис. 2, свидетельствующими, что в условиях возмущенной морской поверхности отмеченное выше преимущество МЧМ-сигнала сокращается с увеличением σ_z . Тем не менее, это преимущество сохраняется в значительном диапазоне q_0 и при наличии волнения, уступая место проигрышу только при $\sigma_z \geq 2$ м, не превышающему, однако, 1,5 децибел при $\sigma_z = 4$ м.

За интерпретацией результатов обратимся к АКФ сигналов на рис. 3. В рамках общих регламентных ограничений протяженность АКФ МЧМ-чипа по оси τ составляет примерно $4,73/W_r$ [12], тогда как основной лепесток АКФ ЛЧМ-сигнала, имеющий длительность $2/W_r$, в два с лишним раза короче.

Для спокойной воды влияние длительности сжатого сигнала на точность измерения времени прихода сигнала сводится к противоборству двух эффектов. С одной стороны, с увеличением длительности сжатого сигнала нарастающий фронт профиля становится более пологим, что потенциально снижает точность оценки запаздывания. С другой – в импульсно-ограниченном высотомере площадь отражающего кольца пропорциональна длительности сжатого сигнала [1, 5, 13], что равносильно возрастанию мощности принятого эхо-сигнала, т.е. повышению точности оценки по мере роста длительности. Тот факт, что в актуальной области отношений сигнал-шум точность оценки запаздывания для МЧМ-чипа выше, чем для ЛЧМ-сигнала, говорит о преобладании в названной области второго из названных эффектов над первым.

Для взволнованной водной поверхности картина меняется. Как показано в [13], интегральный эффект волнения можно учесть сверткой сжатого импульса с плотностью вероятности возвышения волны, трактуя ситуацию так, будто зондирование осуществляется при спокойной воде, но не реальным, а неким эквивалентным, более протяженным импульсом. При этом, однако, расширение сигнала за счет волнения сопровождается не ростом, а снижением принятой мощности, так как свертка с вышеназванной плотностью вероятности одновременно с растяжением сигнала снижает его уровень. Смягчения же этих негативных эффектов возрастанием площади отражающего кольца не происходит.

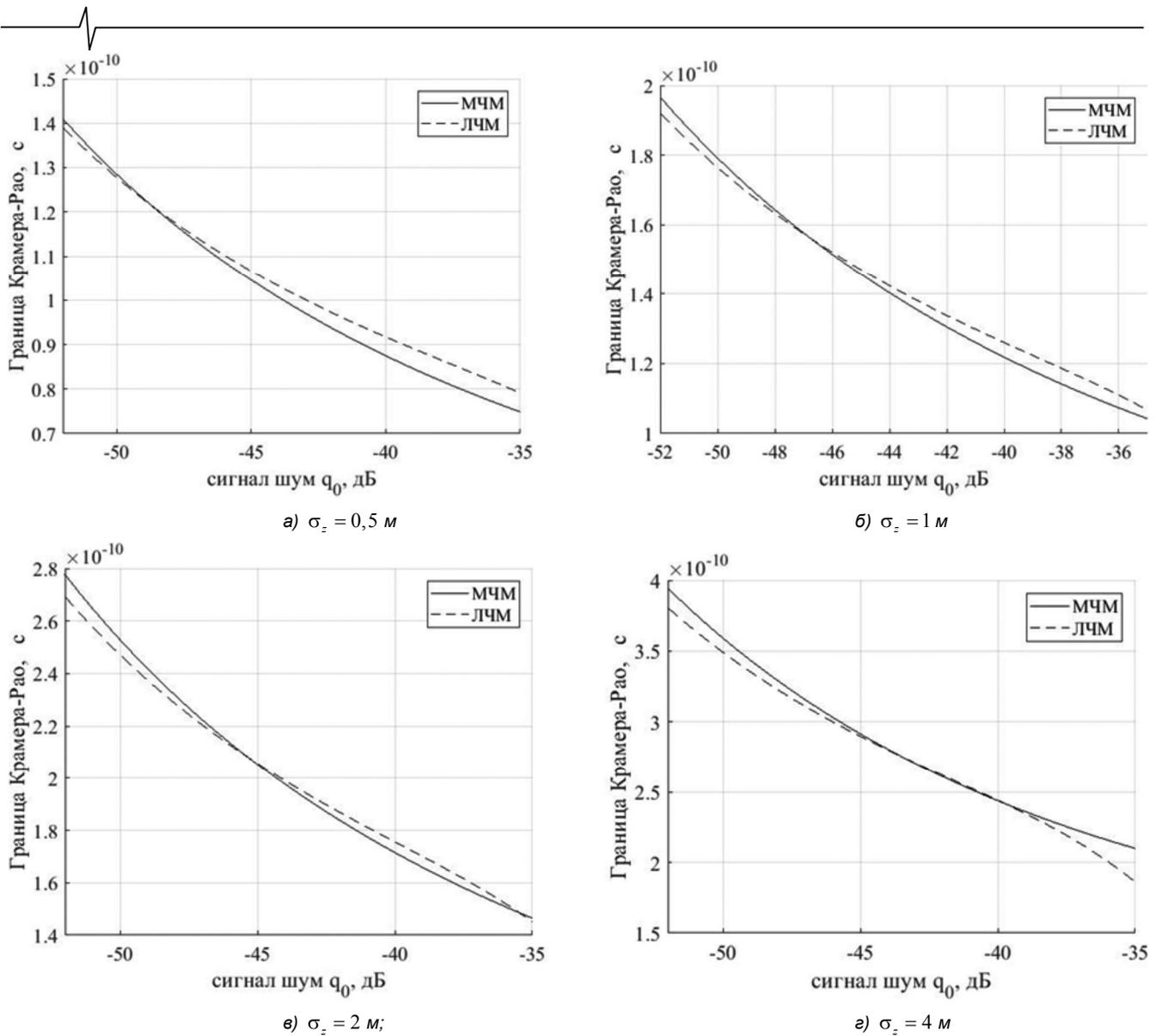


Рис. 2. Усредненные границы Крамера-Рао

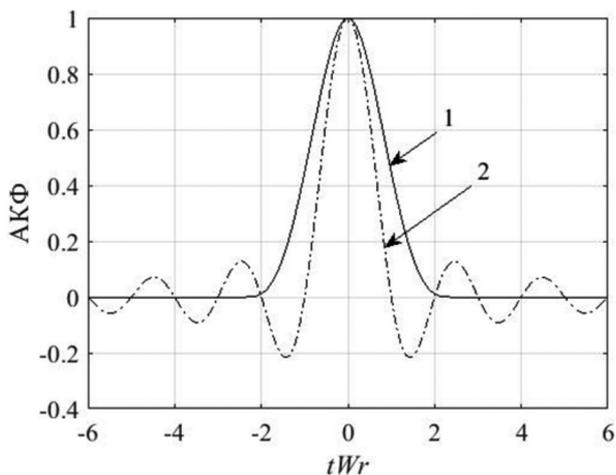


Рис. 3. АКФ сигналов: 1 – МЧМ-сигнал, 2 – ЛЧМ-сигнал

Заключение

В работе получены выражения для потенциальной точности оценки запаздывания эхо-сигнала спутникового высотомера по дискретным зависимым наблюдениям в условиях взволнованной зондируемой поверхности. Расчет границ Крамера-Рао среднеквадратических от-

клонений оценок запаздывания эхо-сигналов для двух альтернативных форматов модуляции (ЛЧМ и МЧМ) выявил границы предпочтительного применения каждого из форматов. Преимущество МЧМ в меньшей погрешности оценки высоты может быть использовано при работе высотомера в условиях слабо возмущенной водной или ледовой поверхности. Если же адаптация формата модуляции к текущему состоянию морской поверхности не предусмотрена, МЧМ-сигнал все равно можно признать приоритетной опцией, так как в условиях сильного волнения его проигрыш в точности ЛЧМ-сигналу меньше его выигрыша при спокойной воде.

Литература

1. Комплексный спутниковый мониторинг морей России / О.Ю. Лаврова, А.Г. Костяной, С.А. Лебедев, М.И. Митягина, А.И. Гинзбург, Н.А. Шеремет. М.: ИКИ РАН, 2011. 480 с.
2. Баскаков А.И., Жутяева Т.С., Лукашенко Ю.И. Локационные методы исследования объектов и сред. М.: Издательский центр «Академия», 2011, 384 с.
3. Coastal Altimetry; ed. by S. Vignudelli, A.G. Kostianoy, P. Cipollini, J. Benveniste. Berlin: Springer, 2011. 565 p.

4. Martin S. An Introduction to Ocean Remote Sensing. 2 nd ed. Cambridge: Cambridge University Press, 2014. 496 p.

5. Satellite Altimetry and Earth Sciences. A Handbook of Techniques and Applications; ed. by L.-L. Fu and A. Cazenave. San Diego: Academic Press, 2001. 463 p.

6. Vincent P., Steunou N., Caubet E., Phalippou L., Rey L., Thouvenot E., Verron J. AltiKa: a Ka-band altimetry payload and system for operational altimetry during the GMES period. Sensors. 2006, vol. 6, pp. 208-234.

7. Жестерев А.Е., Ипатов В.П. Оптимальное измерение запаздывания эхо-сигнала спутникового высотомера при дискретных зависимых наблюдениях. Новости навигации. 2022, № 2. с. 30-35.

8. Регламент радиосвязи Международного союза электросвязи. 2020.

9. Brown G.S. The average impulse response of a rough

surface and its applications. IEEE Trans. on Antennas and Propagation, 1977, vol 25, no.1, pp. 67-74.

10. Hayne G.S. Radar altimeter mean return waveforms from near-normal-incidence ocean surface scattering. IEEE Trans. on Antennas and Propagation, 1980, vol 28, no. 5, pp. 687-692.

11. Barrick, D. E., and Lipa, B. J. Analysis and interpretation of altimeter sea echo, Adv. Geophys., 1985, vol. 27, pp. 60-99.

12. Варгаузин В.А., Цикин И.А. Методы повышения энергетической и спектральной эффективности цифровой радиосвязи. Санкт-Петербург: «БХВ-Петербург», 2013.

13. Вопросы построения радиоинтерфейса спутникового высотомера / Д.С. Боровицкий, А.Е. Жестерев, В.П. Ипатов, Р.М. Мамчур; под ред. В.П. Ипатова. Санкт-Петербург, «БХВ-Петербург», 2017.

НОВЫЕ КНИГИ

Основы статистической теории радиотехнических систем:

Учебное пособие под ред. А.В. Коренного. М.: Изд-во Радиотехника, 2021 г. 240 с.: ил.

Приведены необходимые сведения из теории вероятностей, случайных процессов и математической статистики, и на их основе рассмотрены статистические методы анализа линейных и нелинейных систем. На базе теории фильтрации изложены современные методы синтеза радиоэлектронных систем различного назначения, основы теории информации и методы статистического моделирования. Методика применения теоретических результатов к решению практических задач проиллюстрирована содержательными примерами.

Предназначено для слушателей и курсантов военных вузов, а также студентов вузов, обучающихся по специальности «Радиоэлектронные системы и комплексы». Может быть аспирантам и преподавателям вузов, занимающихся вопросами синтеза и анализа радиотехнических устройств и систем.



Гаврилов К.Ю., Каменский И.В., Кирдяшкин В.В., Линников О.Н.

Моделирование и обработка радиолокационных сигналов в MATLAB: Учебное пособие. М.: Изд-во Радиотехника, 2020 г. 264 с.: ил.

Рассмотрены методы моделирования радиолокационных сигналов при отражении от сложных целей, принципы моделирования аналоговых и цифровых устройств обработки сигналов, включающие формирование двумерной матрицы цифровых отсчетов, методы согласованной фильтрации, обнаружения и обработки сигналов в импульсно-доплеровских радиолокационных системах.

Показаны примеры обработки наиболее распространенных видов радиолокационных сигналов – импульсных, с линейной частотной модуляцией и фазо-кодированных сигналов. Приведены программы моделирования и обработки сигналов в среде MATLAB.

Для студентов, аспирантов и инженеров, изучающих и использующих теорию радиолокации и методы моделирования и обработки радиолокационных сигналов. Будет полезна научным работникам и разработчикам радиолокационных систем.



РАЗРАБОТКА МЕТОДА И АЛГОРИТМОВ ОЦЕНКИ СДВИГОВ ВЕТРА И ТУРБУЛЕНТНОСТИ В МЕТЕОРОЛОГИЧЕСКОМ РАДИОЛОКАЦИОННОМ КОМПЛЕКСЕ АЭРОДРОМНОЙ ЗОНЫ

Синицын И.А., к.т.н., старший научный сотрудник, доцент АО «Научно-производственное предприятие «Исток» им. А.И. Шокина», e-mail: isinitsyn@yandex.ru.

Галаева К.И., к.т.н., преподаватель кафедры технической эксплуатации радиоэлектронного оборудования воздушного транспорта ФГБОУ ВО МГТУ ГА, e-mail: ks.galaeva@mail.ru

DEVELOPMENT OF A METHOD AND ALGORITHMS FOR ESTIMATING WIND SHEAR AND TURBULENCE IN THE METEOROLOGICAL RADAR COMPLEX OF THE AERODROME ZONE

Sinitsyn I.A., Galaeva K.I.

This article presents the developed method and algorithms for estimating wind shear and turbulence used in the ground-based meteorological radar complex of the aerodrome zone MRLC AZ. Wind shear refers to horizontal wind shear by 600 m and vertical wind shear by 30 m, turbulence refers to the parameter of the eddy dissipation rate of turbulent energy EDR, these terms comply with modern requirements of ICAO, WMO, RF Roshydromet. The method for estimating wind shear and turbulence consists of 3 stages, during which the measured values of the radial velocity, the width of the radial velocity spectrum become the initial data for calculating wind shear and turbulence, followed by classification by hazard degree. Based on the method for estimating wind characteristics, algorithms have been developed for estimating horizontal and vertical wind shear, parameters of the specific eddy dissipation rate of turbulent energy, which is a more correct representation of atmospheric turbulence. The article presents an assessment of the efficiency of the obtained algorithms by the method of statistical tests, it is shown that the obtained values of the standard deviation of estimates of EDR, wind shears meet modern existing requirements.

Key words: meteorological radar, airfield zone, method and algorithms, specific velocity of turbulent energy dissipation, horizontal wind shear, vertical wind shear, hazard gradation, standard deviation of estimates of wind shear and turbulence.

Ключевые слова: метеорологический радиолокатор, аэродромная зона, метод и алгоритмы, удельная скорость диссипации турбулентной энергии, горизонтальный сдвиг ветра, вертикальный сдвиг ветра, градация по степени опасности, среднеквадратичное отклонение оценок сдвигов ветра и турбулентности.

Введение

Актуальной проблемой авиации является влияние метеоусловий на безопасность полётов. Более 20 % случаев авиационных событий были связаны с неблагоприятными метеорологическими условиями [1]. При этом 34,4 % авиационных событий происходили без метеообеспечения полётов гражданской авиации при сложных метеоусловиях [2-6]. Следовательно, причинами авиационных событий является недостаточный уровень метеообеспечения, особенно при обнаружении опасных сдвигов ветра и турбулентности. При этом наземный метеорологический радиолокатор является уникальным источником метео данных с высокой дискретностью в пространстве и времени для анализа метеорологических явлений и характеристик в аэродромной зоне (далее – АЗ).

В настоящей статье приведены разработанные метод и алгоритмы оценки сдвигов ветра и турбулентности, используемые в наземном метеорологическом радиолокаторе аэродромной зоны. Под сдвигами ветра понимаются горизонтальный сдвиг ветра на 600 м и вертикальный сдвиг ветра на 30 м, под турбулентностью понимается параметр удельной скорости диссипации турбулентной энергии EDR, указанные термины соответствуют современным требованиям ИКАО, ВМО, Росгидромета РФ. Метод оценки сдвигов ветра и турбулентности состоит из 3 этапов, в ходе которых измеренные значения радиальной скорости, ширины спектра радиальных скоростей становятся начальными данными для расчёта сдвигов ветра и турбулентности с последующей классификацией по степени опасности. На основе метода оценки ветровых характеристик разработаны алгоритмы оценки горизонтального и вертикального сдвигов ветра, параметров удельной скорости диссипации турбулентной энергии, являющейся более корректным представлением турбулентности атмосферы. В статье приведена оценка эффективности полученных алгоритмов методом статистических испытаний, показано, что полученные значения среднеквадратичного отклонения оценок EDR, сдвигов ветра удовлетворяют современным существующим требованиям.

Метод оценки ветровых характеристик для секторного режима в МРЛК АЗ

Предназначением метеорологического радиолокационного комплекса аэродромной зоны (МРЛК АЗ) является оценка опасных метеоявлений, в том числе опасных ветровых характеристик в аэродромной зоне в секторах взлета и посадки воздушных судов. В МРЛК АЗ помимо основного кругового режима обзора пространства внедрён секторный режим работы, обладающий азимутальной шириной сектора в диапазоне 0...70° и меньшим шагом выдачи информации в пространстве и времени, что актуально для таких сильно флуктуирующий ветровых характеристик как турбулентность и сдвиг ветра. Для оценки атмосферной облачной турбулентности используется параметр *EDR* – единица кубического корня из скорости затухания вихря или скорость диссипации турбулентной энергии [7-8], вертикальный сдвиг ветра (далее – *BCB*) оценивается на 30 или на 100 м, горизонтальный сдвиг ветра (далее – *GCB*) на 600 м согласно требованиям ИКАО, Росгидромета РФ, ВМО [7-11].

Начальными данными для оценки сдвигов ветра и турбулентности являются радиальная скорость отражённого сигнала *V* и ширина спектра радиальных скоростей *W*. Далее по входным сигналам *V* и *W* для конических сечений в каждом канале дальности формируются оценки *GCB*, *BCB*, *EDR* с заданными диапазоном и дискретностью в ячейках выбранного размера в декартовых координатах. После осуществляется градация *GCB*, *BCB* и *EDR* по степени опасности согласно классификации ИКАО [8-11]. Выходная информация отображается на индикаторе кругового обзора МРЛК АЗ в виде опасных контуров *GCB*, *BCB* и *EDR*. На рис. 1 продемонстрирована схема метода оценки опасных ветровых метеорологических явлений: *GCB*, *BCB*, *EDR*, разделённая на три этапа.

На основе разработанного метода оценки турбулентности и сдвигов ветра были разработаны алгоритмы оценки удельной скорости диссипации турбулентной энергии, оценки вертикального и горизонтального сдвигов ветра, описанные ниже.

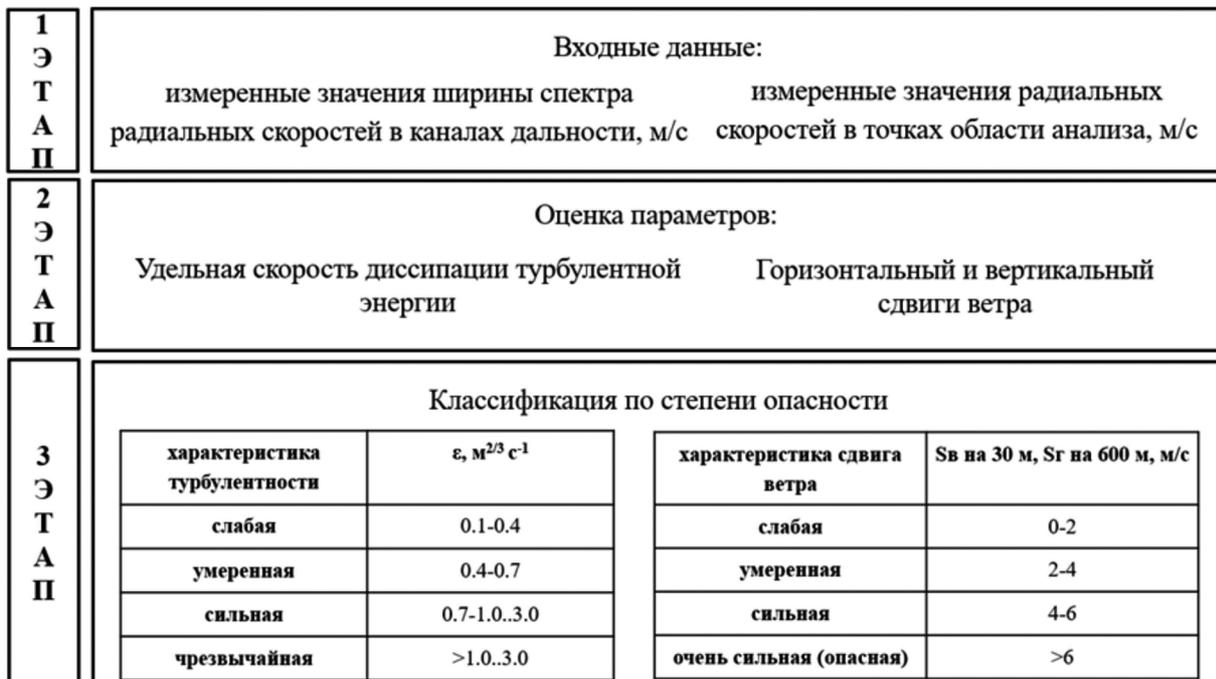


Рис. 1. Схема метода оценки опасных ветровых метеорологических явлений (*GCB*, *BCB*, *EDR*) в МРЛК АЗ

Алгоритм оценки турбулентности для секторного режима в МРЛК АЗ

Начальными данными для алгоритма *EDR* являются: оценки радиальной скорости *V* в каналах дальности $V_m; m = 0; M - 1$, оценки ширины спектра скоростей *W* в каналах дальности $(\sigma_v)_m; m = 0; M - 1$.

Выходными данными алгоритма являются оценки *EDR* в каналах дальности $\epsilon_m; m = 0; M - 1$.

Условиями функционирования алгоритма являются полученные оценки *V* и *W* в каналах дальности, заданные общие параметры секторного режима и параметры рассматриваемого алгоритма, представленные в табл. 1. Блок-схема алгоритма представлена на рис. 2.

Определение ошибки оценки турбулентности *EDR*.

Определение ошибки оценки турбулентности *EDR* методом статистических испытаний «Монте-Карло» реализовано в виде следующих этапов, представленных на рис. 3.

В рамках третьего этапа (цикл по независимым испытаниям) выполняются следующие действия, представленные на рис. 4.

Относительная ошибка оценки турбулентности *EDR* $\delta_{EDR} \%$ рассчитывается по выражению:

$$\delta_{EDR} \% = \sigma_{EDR} \cdot 100\% / EDR, \tag{1}$$

Исходные данные для проведения расчетных испытаний удельной скорости диссипации турбулентной энергии *EDR* методом «Монте Карло» представлены в табл. 2.

Результаты испытаний удельной скорости диссипации турбулентной энергии представлены в табл. 3.

Таблица 1. Параметры алгоритма «Измерение параметров турбулентности»

| Обозначение | Название, размерность | Значение |
|----------------------------------|--|--|
| K_σ | параметр, определяющий среднее квадратичное отклонение σ оценки радиальной скорости | 12 |
| $L^{(0)}$ | ожидаемое среднее значение интегрального масштаба турбулентности, м | 760 |
| | интервал дискретизации коэффициента продольной корреляции скорости | 10^{-2} |
| $f_k, k = \overline{0; K-1}$ | значения коэффициента продольной корреляции скорости | хранятся в заранее подготовленном массиве |
| δM | минимальная длина серии ненулевых оценок радиальной скорости | 10 |
| δD | дискрет дальности, м | |
| ΔD | разрешающая способность по дальности, м | $\Delta D = \frac{c\Delta\tau_0}{2}$ |
| Δ | нормированное разрешение по дальности | $\frac{\Delta D}{\delta D}$ |
| I | коэффициент перекрытия стробов дальности | $truc(\Delta)$ |
| $\rho_m; m = \overline{0; 2M-1}$ | коэффициенты корреляции ошибок определения радиальной скорости | $\left(1 - \frac{m}{\Delta}\right)^2$ при $m < I$; 0 при $m \geq I$ |

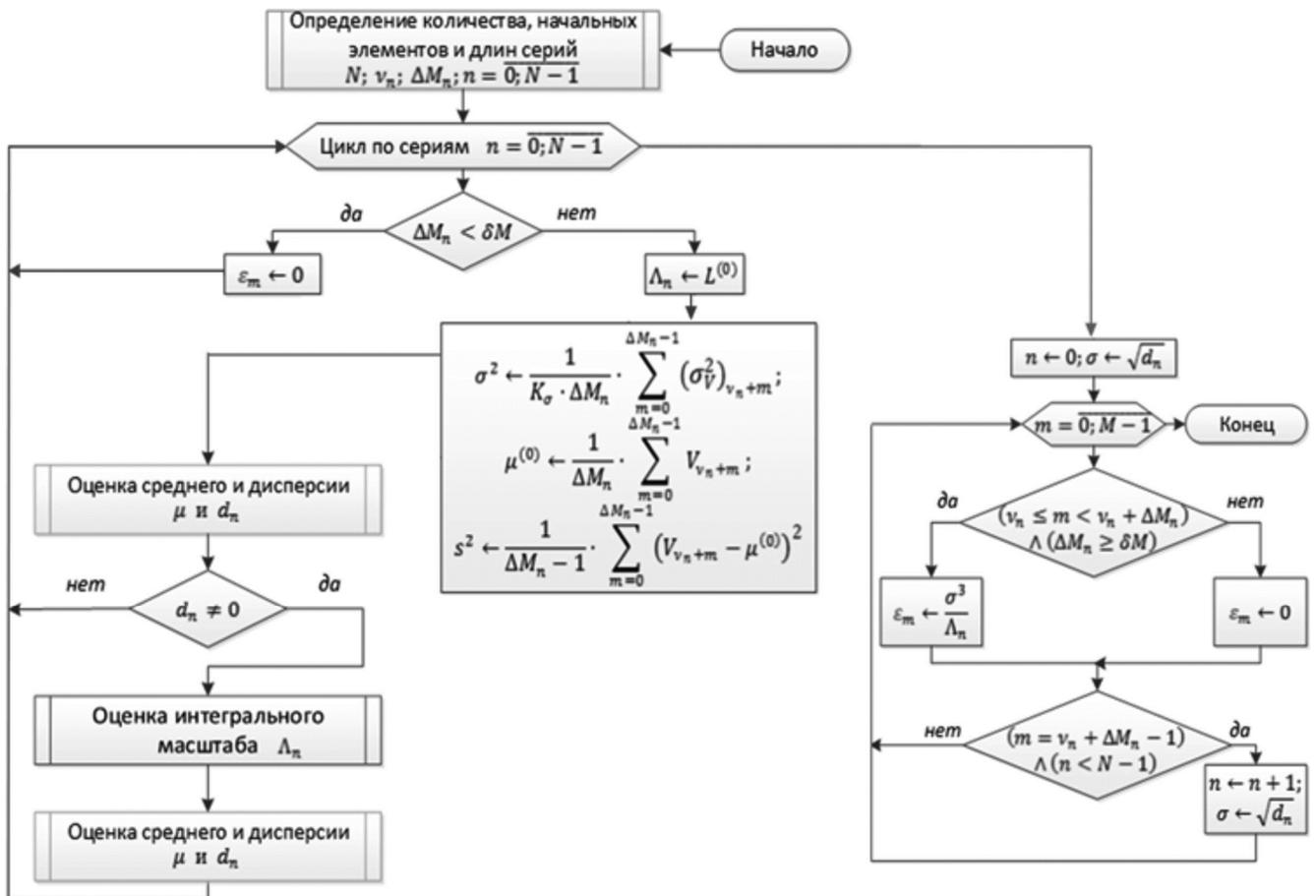


Рис. 2. Блок-схема алгоритма оценки турбулентности в МРЛК АЗ



Рис. 3. Этапы реализации определения характеристик точности оценки параметров турбулентности EDR методом статистических испытаний «Монте-Карло»



Рис. 4. Этапы цикла по независимым испытаниям

Таблица 2. Исходные данные для проведения расчетных испытаний EDR методом «Монте-Карло»

| параметр | Значение |
|---|-----------|
| σ ошибок измерений радиальных скоростей σ_V для горизонтальной дальности 100 км | 0,432 м/с |
| Значение дискретности данных по дальности (до 50 км) | 101,25 м |
| Значение дискретности данных по дальности (до 100 км) | 202,5 |

Таблица 3. Результаты расчетов σ оценки EDR

| Требуемое значение EDR, $m^{2/3}/c$ | δ_{EDR} для дальности до 50 км и дискретности 101.25 м, $m^{2/3}/c$ | Относительная ошибка EDR $\delta_{EDR} \%$ для дальности до 50 км и дискретности 101.25 м, % | σ_{EDR} для дальности до 100 км и дискретности 202.5 м, $m^{2/3}/c$ | Относительная ошибка EDR $\delta_{EDR} \%$ для дальности до 100 км и дискретности 202.5 м, % |
|-------------------------------------|--|--|--|--|
| 0,1 | 0,0304 | 30,4 | 0,0242 | 24,2 |
| 0,2 | 0,0589 | 29,5 | 0,0499 | 25,0 |
| 0,3 | 0,0588 | 19,6 | 0,06 | 20,0 |
| 0,4 | 0,0714 | 17,9 | 0,0775 | 19,4 |
| 0,5 | 0,0873 | 17,5 | 0,0939 | 18,8 |
| 0,6 | 0,1038 | 17,3 | 0,1119 | 18,7 |
| 0,7 | 0,1202 | 17,2 | 0,1313 | 18,8 |
| 0,8 | 0,1360 | 17,0 | 0,1486 | 18,6 |
| 0,9 | 0,1552 | 17,2 | 0,1672 | 18,6 |
| 1 | 0,1707 | 17,1 | 0,1861 | 18,6 |
| 1,1 | 0,1886 | 17,1 | 0,2083 | 18,9 |

Таким образом, дискретность оценки турбулентности по удельной скорости диссипации турбулентной энергии (EDR) соответствует требованиям, указанным в источниках [7-8,12-20] при следующих условиях: до 50 км горизонтальной дальности дискретность оценки EDR не более $0,05 \text{ м}^{2/3}/\text{с}$ при значении требуемой EDR менее $0,2 \text{ м}^{2/3}/\text{с}$; до 100 км горизонтальной дальности дискретность оценки EDR не более $0,1 \text{ м}^{2/3}/\text{с}$ при значении требуемой EDR менее $0,2 \text{ м}^{2/3}/\text{с}$.

Алгоритмы оценки горизонтального и вертикального сдвигов ветра для секторного режима в МРЛК АЗ

Далее приведены алгоритмы оценки сдвигов ветра: в табл. 4 указаны входные данные для функционирования алгоритмов сдвигов ветра, на рис. 5 представлена блок-схема алгоритмов сдвигов ветра.

Таблица 4. Входные данные для функционирования алгоритмов горизонтального и вертикального сдвигов ветра при секторном сканировании (режим «Сектор»)

| Обозначение | Название, размерность |
|---|--|
| M | количество каналов дальности |
| $D[m]; m = \overline{0; M-1}$ | расположение каналов дальности, м |
| N | количество угловых положений главного луча (ГЛ) диаграммы направленности антенны (ДНА) при сканировании по азимуту |
| $\alpha[n]; n = \overline{0; N-1}$ | углы по азимуту ГЛ ДНА при сканировании, рад |
| K | количество положений углов ГЛ ДНА при сканировании по углу места |
| $\beta[k]; k = \overline{0; K-1}$ | углы по месту положений ГЛ ДНА при сканировании, рад |
| $\begin{cases} x[m, n, k] = D[m] \cos \beta[k] \cdot \cos \alpha[n]; \\ y[m, n, k] = D[m] \sin \beta[k]; \\ z[m, n, k] = D[m] \cos \beta[k] \cdot \sin \alpha[n] \end{cases}$ | декартовы координаты точек области анализа, м |
| $V[m, n, k]$ | измеренные значения V в точках области анализа, м/с |
| $\sigma_v[m, n, k]$ | измеренные значения ширины спектра скоростей в точках области анализа, м/с |

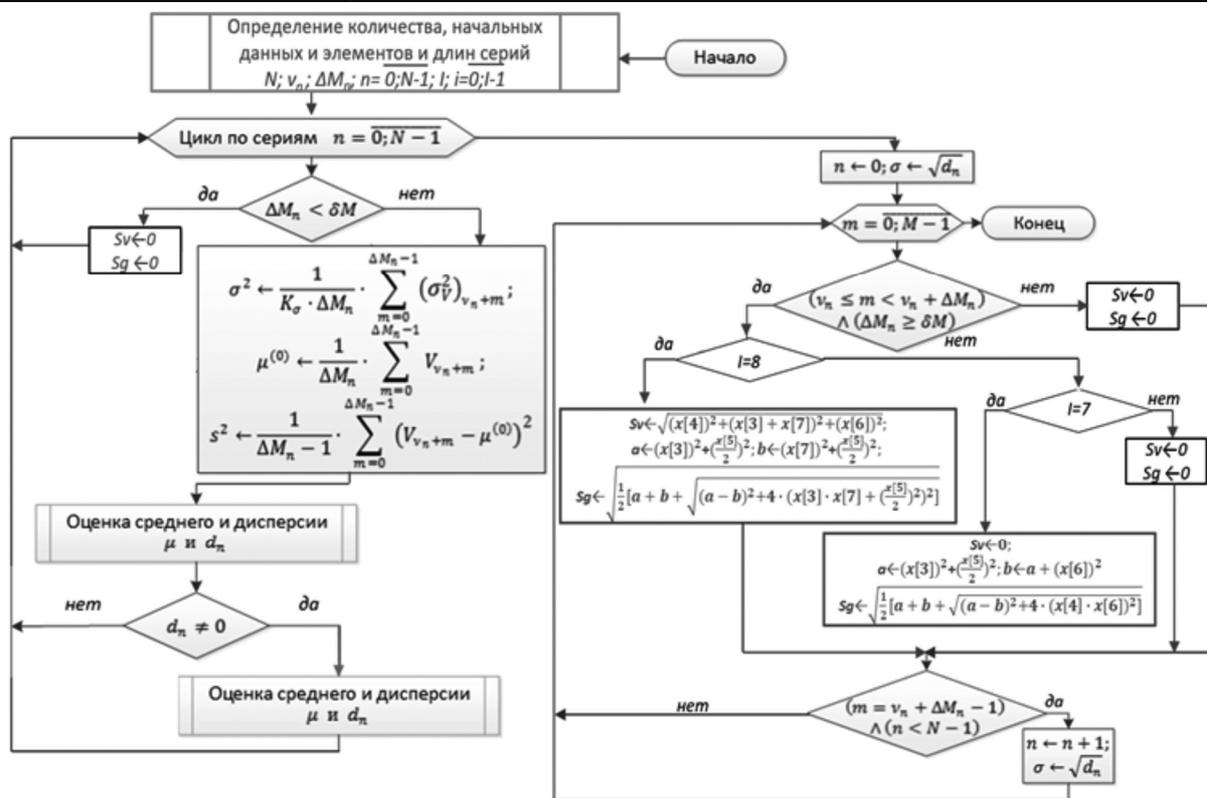


Рис. 5. Блок-схема алгоритма оценки сдвигов ветра в МРЛК АЗ

Определение ошибки оценки вертикального ВСВ и горизонтального ГСВ сдвигов ветра

Для σ формирования сдвигов ветра были получены следующие выражения: для ВСВ S_g :

$$\sigma_g = 30\sqrt{\sigma_4^2 + \sigma_6^2 + \sigma_7^2}, \quad (2)$$

для ГСВ S_z :

$$\sigma_r = 600\sqrt{\sigma_3^2 + \sigma_5^2 + \sigma_8^2}. \quad (3)$$

Результаты испытаний алгоритмов оценки горизон-

тального сдвига ветра представлены в табл. 5.

Таким образом, В МРЛК АЗ дискретность цифровых данных для оценки горизонтального сдвига ветра со значением 0,2 м/с на 600 метров обеспечивает определение уровня градации опасности ГСВ согласно требованиям [7-8,12-20] и рис. 1.

Далее представлен расчёт ошибки оценки вертикального сдвига ветра (таблица 6). Ошибка оценки вертикального сдвига ветра определяется СКО по выражению (2) для объектов анализа (далее – ОА).

Таблица 5. Результаты расчетов ошибки оценки ГСВ по выражению (3)

| Дальность центра ОА D_0 , м | Размер ОА по ΔD , м | Размер ОА по $\Delta \alpha$, градус | Размер ОА по УМ $\Delta \beta$, градус | σ ГСВ σ_z м/с на 600 м |
|--|-----------------------------|---------------------------------------|---|--------------------------------------|
| Значение ГСВ согласно требованиям – 0,5 м/с на 600 м | | | | |
| 1500,0 | 1500,0 | 23,00 | 17,00 | 0,1444 |
| 3000,0 | 1500,0 | 19,00 | 15,00 | 0,1533 |
| 10000,0 | 1500,0 | 17,00 | 14,50 | 0,1010 |
| 25000,0 | 1500,0 | 15,00 | 13,00 | 0,1109 |
| 50000,0 | 1500,0 | 13,00 | 13,00 | 0,1260 |
| Значение ГСВ согласно требованиям – 2,0 м/с на 600 м | | | | |
| 1500,0 | 1500,0 | 23,00 | 17,00 | 0,1444 |
| 3000,0 | 1500,0 | 19,00 | 15,00 | 0,1532 |
| 10000,0 | 1500,0 | 17,00 | 14,50 | 0,1009 |
| 25000,0 | 1500,0 | 15,00 | 13,00 | 0,1110 |
| 50000,0 | 1500,0 | 13,00 | 13,00 | 0,1265 |
| Значение ГСВ согласно требованиям – 4,0 м/с на 600 м | | | | |
| 1500,0 | 1500,0 | 23,00 | 17,00 | 0,1444 |
| 3000,0 | 1500,0 | 19,00 | 15,00 | 0,1532 |
| 10000,0 | 1500,0 | 17,00 | 14,50 | 0,1009 |
| 25000,0 | 1500,0 | 15,00 | 13,00 | 0,1114 |
| 50000,0 | 1500,0 | 13,00 | 13,00 | 0,1288 |
| Значение ГСВ согласно требованиям – 6,0 м/с на 600 м | | | | |
| 1500,0 | 1500,0 | 23,00 | 17,00 | 0,1444 |
| 3000,0 | 1500,0 | 19,00 | 15,00 | 0,1532 |
| 10000,0 | 1500,0 | 17,00 | 14,50 | 0,1011 |
| 25000,0 | 1500,0 | 15,00 | 13,00 | 0,1122 |
| 50000,0 | 1500,0 | 13,00 | 13,00 | 0,1329 |
| Значение ГСВ согласно требованиям – 8,0 м/с на 600 м | | | | |
| 1500,0 | 1500,0 | 23,00 | 17,00 | 0,1445 |
| 3000,0 | 1500,0 | 19,00 | 15,00 | 0,1533 |
| 10000,0 | 1500,0 | 17,00 | 14,50 | 0,1012 |
| 25000,0 | 1500,0 | 15,00 | 13,00 | 0,1135 |
| 50000,0 | 1500,0 | 13,00 | 13,00 | 0,1386 |
| Максимум меньше 0,2 м/с на 600 м | | | | |

Таблица 6. Результаты расчетов ошибки оценки ВСВ

| Дальность центра ОА D_0 , м | Размер ОА по ΔD , м | Размер ОА по $\Delta \alpha$, градус | Размер ОА по УМ $\Delta \beta$, градус | σ ГСВ σ_z м/с на 600 м |
|--|-----------------------------|---------------------------------------|---|--------------------------------------|
| Значение ВСВ согласно требованиям – 0,5 м/с на 600 м | | | | |
| 1500,0 | 1500,0 | 23,00 | 17,00 | 0,0150 |
| 3000,0 | 1500,0 | 19,00 | 15,00 | 0,0145 |
| 10000,0 | 1500,0 | 17,00 | 14,50 | 0,0142 |
| 25000,0 | 1500,0 | 15,00 | 13,00 | 0,0338 |
| 30000,0 | 1500,0 | 13,00 | 13,00 | 0,0422 |
| 35000,0 | 1500,0 | 13,00 | 13,00 | 0,0482 |
| 40000,0 | 1500,0 | 13,00 | 13,00 | 0,0541 |
| 45000,0 | 1500,0 | 13,00 | 13,00 | 0,0600 |
| 50000,0 | 1500,0 | 13,00 | 13,00 | 0,0657 |

| Дальность центра ОА D_0 , м | Размер ОА по $D \Delta D$, м | Размер ОА по $Az \Delta \alpha$, градус | Размер ОА по УМ $\Delta \beta$, градус | σ ГСВ σ_z м/с на 600 м |
|--|----------------------------------|---|--|---|
| Значение ВСВ согласно требованиям – 2,0 м/с на 600 м | | | | |
| 1500,0 | 1500,0 | 23,00 | 17,00 | 0,0249 |
| 3000,0 | 1500,0 | 19,00 | 15,00 | 0,0338 |
| 10000,0 | 1500,0 | 17,00 | 14,50 | 0,0510 |
| 25000,0 | 1500,0 | 15,00 | 13,00 | 0,1322 |
| 30000,0 | 1500,0 | 13,00 | 13,00 | 0,1661 |
| 35000,0 | 1500,0 | 13,00 | 13,00 | 0,1903 |
| 40000,0 | 1500,0 | 13,00 | 13,00 | 0,2142 |
| 45000,0 | 1500,0 | 13,00 | 13,00 | 0,2378 |
| 50000,0 | 1500,0 | 13,00 | 13,00 | 0,2610 |
| Значение ВСВ согласно требованиям – 4,0 м/с на 600 м | | | | |
| 500,0 | 1500,0 | 23,00 | 17,00 | 0,0431 |
| 3000,0 | 1500,0 | 19,00 | 15,00 | 0,0639 |
| 10000,0 | 1500,0 | 17,00 | 14,50 | 0,1013 |
| 25000,0 | 1500,0 | 15,00 | 13,00 | 0,2639 |
| 30000,0 | 1500,0 | 13,00 | 13,00 | 0,3318 |
| 35000,0 | 1500,0 | 13,00 | 13,00 | 0,3802 |
| 40000,0 | 1500,0 | 13,00 | 13,00 | 0,4281 |
| 45000,0 | 1500,0 | 13,00 | 13,00 | 0,4753 |
| 50000,0 | 1500,0 | 13,00 | 13,00 | 0,4918 |
| Значение ВСВ согласно требованиям – 6,0 м/с на 600 м | | | | |
| 1500,0 | 1500,0 | 23,00 | 17,00 | 0,0625 |
| 3000,0 | 1500,0 | 19,00 | 15,00 | 0,0947 |
| 10000,0 | 1500,0 | 17,00 | 14,50 | 0,1517 |
| 25000,0 | 1500,0 | 15,00 | 13,00 | 0,1956 |
| 30000,0 | 1500,0 | 13,00 | 13,00 | 0,2076 |
| 35000,0 | 1500,0 | 13,00 | 13,00 | 0,2701 |
| 40000,0 | 1500,0 | 13,00 | 13,00 | 0,3419 |
| 45000,0 | 1500,0 | 13,00 | 13,00 | 0,4128 |
| 50000,0 | 1500,0 | 13,00 | 13,00 | 0,4824 |
| Значение ВСВ согласно требованиям – 8,0 м/с на 600 м | | | | |
| 1500,0 | 1500,0 | 23,00 | 17,00 | 0,0823 |
| 3000,0 | 1500,0 | 19,00 | 15,00 | 0,1257 |
| 10000,0 | 1500,0 | 17,00 | 14,50 | 0,2022 |
| 25000,0 | 1500,0 | 15,00 | 13,00 | 0,2274 |
| 30000,0 | 1500,0 | 13,00 | 13,00 | 0,2633 |
| 35000,0 | 1500,0 | 13,00 | 13,00 | 0,3100 |
| 40000,0 | 1500,0 | 13,00 | 13,00 | 0,3558 |
| 45000,0 | 1500,0 | 13,00 | 13,00 | 0,3703 |
| 50000,0 | 1500,0 | 13,00 | 13,00 | 0,4431 |

Таким образом, в МРЛК АЗ дискретность оценки вертикального сдвига ветра соответствует рассмотренным требованиям не более 0,5 м/с в диапазоне значений интенсивности ВСВ на 30 м от не более 0,5 м/с до не менее 6,0 м/с. Следовательно, в МРЛК АЗ обеспечивается определение уровней градации опасности вертикального сдвига ветра по его интенсивности в соответствии с рис. 1 и требованиям [7-8, 12-20].

Получение экспериментальных фактических данных об удельной скорости диссипации турбулентной энергии и сдвигах ветра в МРЛК АЗ

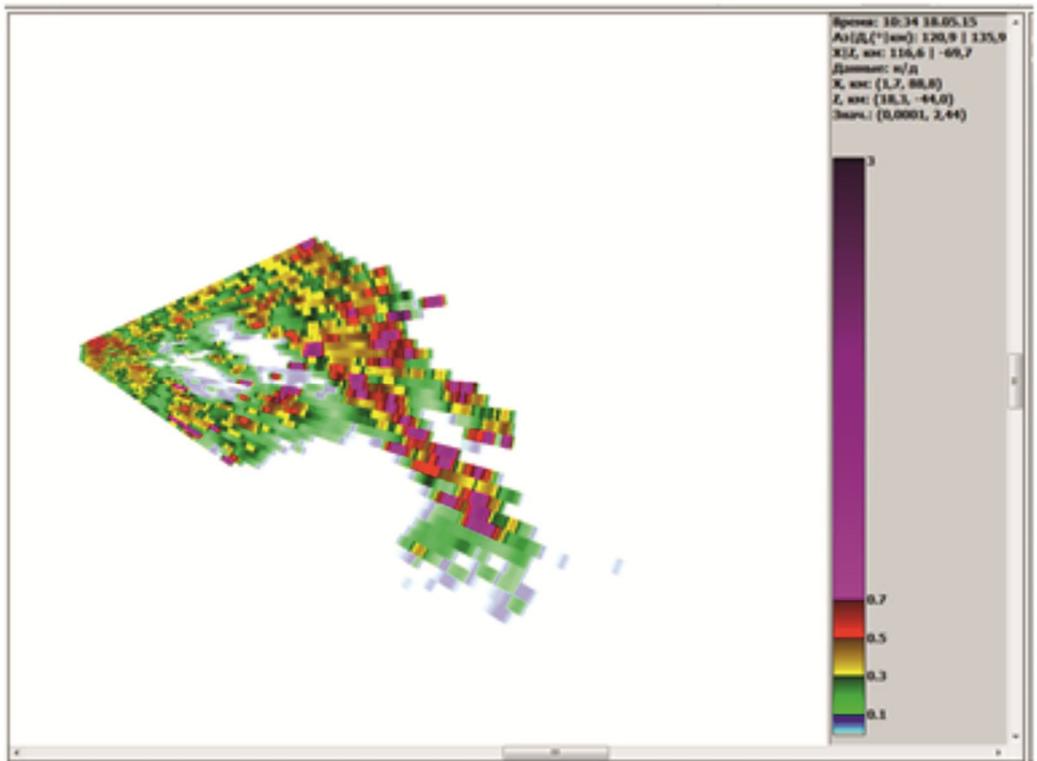
Для подтверждения работоспособности разработанных метода и алгоритмов для всей зоны секторного обзора МРЛК АЗ были получены матрицы значений для удельной скорости диссипации турбулентной энергии EDR (рис. 6), а также горизонтальный ГСВ и вертикаль-

ный ВСВ сдвиги ветра (рис. 7-8 соответственно).

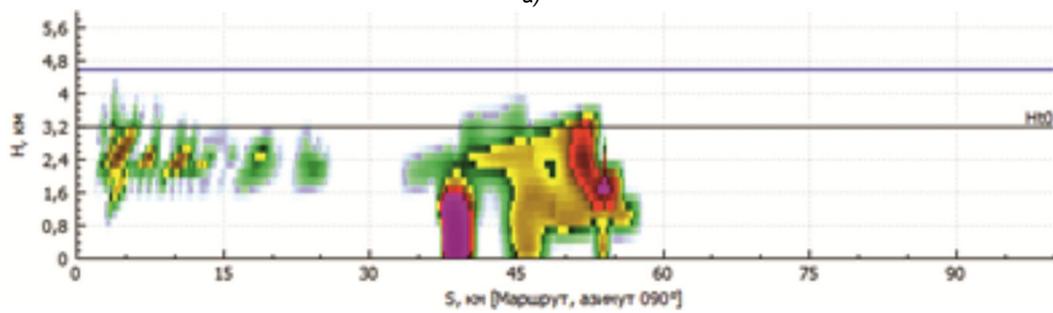
Рис. 6-8 являются результатами серий экспериментальных испытаний по фактической погоде для аэродрома Орловка (Тверская область).

Заключение

В настоящее время разработан метод оценки турбулентности и сдвигов ветра в МРЛК АЗ, в котором поэтапно определяются радиальная скорость и её среднеквадратичное отклонение – ширина спектра радиальных скоростей. Далее рассчитываются турбулентность в единицах удельной скорости диссипации турбулентной энергии EDR, горизонтальный сдвиг ветра на 600 метров, вертикальный сдвиг ветра на 30 метров согласно требованиям ИКАО, с градациями степеней опасности. На основе разработанного метода были синтезированы алгоритмы удельной скорости диссипации турбулентной энергии и вертикального, горизонтального сдвигов вет-



а)



б)

Рис. 6. Горизонтальное (а) и вертикальное (б) сечения EDR в секторном режиме в МРЛК АЗ

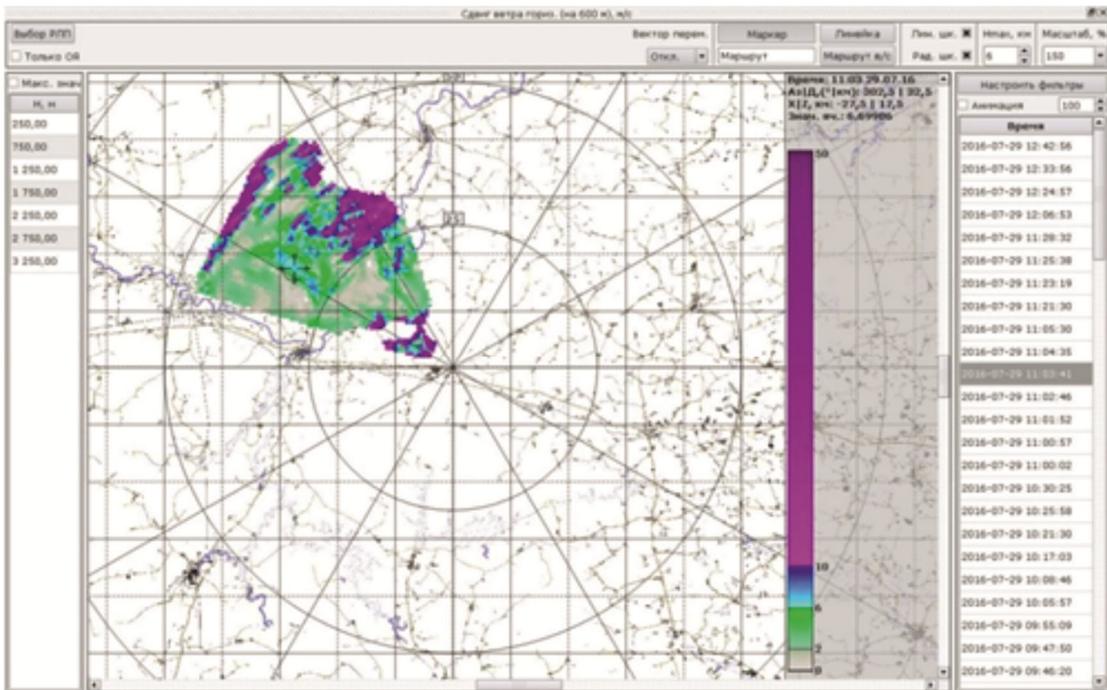


Рис. 7. Карта ГСВ в режиме «Сектор» в МРЛК АЗ

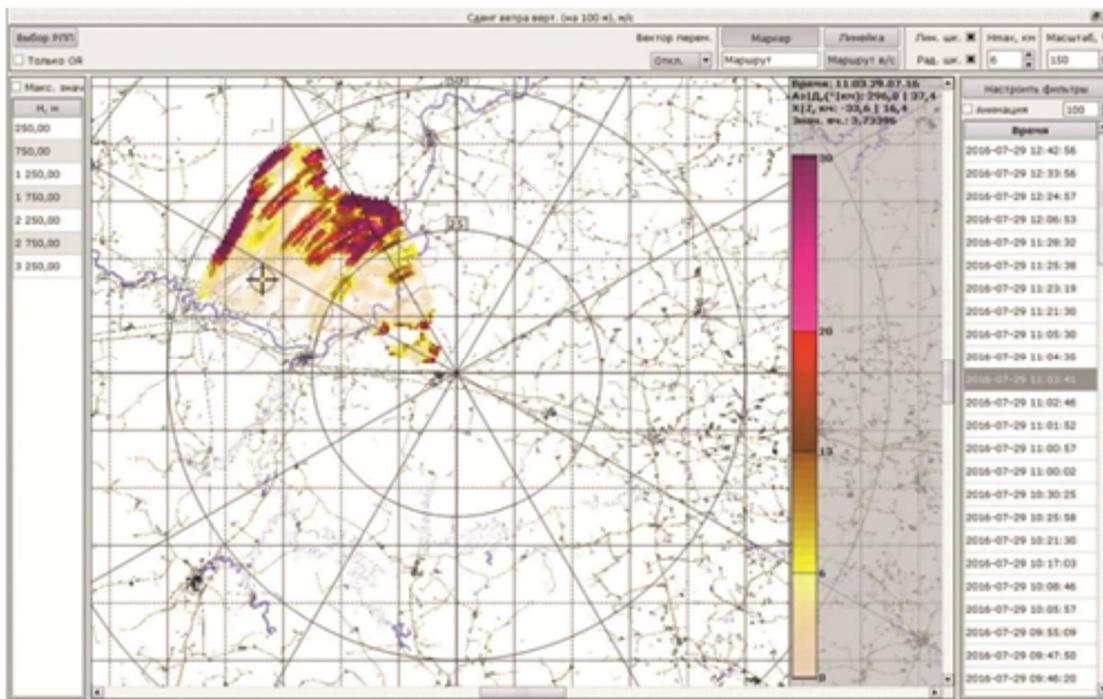


Рис. 8. Карта ВСВ в режиме «Сектор» в МРЛК АЗ

ра. Начальными данными для алгоритмов оценки *EDR*, горизонтального и вертикального сдвигов ветра являются: количество каналов дальности, количество угловых и азимутальных углов, декартовы координаты точек области анализа, значения в каналах дальности V и W . Выходные массивы алгоритмов представляют собой оценки *EDR* и оценки сдвигов ветра в каналах дальности.

Получены оценки для среднеквадратичного отклонения формирования сдвигов ветра, рассчитанные значения шага дискретизации для вертикального сдвига ветра составляет 0,5 м/с и для горизонтального сдвига ветра составляет 0,2 м/с, что соответствует выдвинутым в работе требованиям.

В МРЛК АЗ диапазоны оценки турбулентности *EDR* соответствуют следующим значениям: от не более $0,1 \text{ м}^{2/3}/\text{с}$ до не менее $3 \text{ м}^{2/3}/\text{с}$, что полностью удовлетворяет рассмотренным требованиям.

Статья подготовлена в рамках поддержанного грантом Российского научного фонда проекта № 23-29-00450 «Исследование отражательных и турбулентных характеристик атмосферы с использованием нового отечественного метеорологического радиолокационного комплекса ближней зоны «Монокль» в различных климатических зонах в интересах повышения достоверности автоматической классификации опасных метеоявлений».

Литература

1. Методическое пособие. Расследование авиационных происшествий и инцидентов, связанных с метеорологическими факторами. Изд. 3. Москва, 2009. 57 с.
2. Vasiliev O., Bolelov E., Galaeva K., Gevak N., Zybkin S., Kolesnikov E., Peshko A., Sinitsyn I. The Design and Operation Features of the Near-airfield Zone Weather Radar Complex «Monocle». 2021 XVIII Technical Scientific Conference on Aviation Dedicated to the Memory of

N.E. Zhukovsky (TSCZh). DOI:10.1109/TSCZh53346.2021.9628352.

3. Vasiliev O.V., Boyarenko E.S., Galaeva K.I., Zybkin S.A. Concerning the Issue of Classification of Hazardous Weather Events. 2022 XIX Technical Scientific Conference on Aviation Dedicated to the Memory of NE Zhukovsky (TSCZh). IEEE, 2022. С. 76-78. DOI:10.1109/TSCZh55469.2022.9802491

4. Nanding N., Rico-Ramirez M.A. Precipitation Measurement with Weather Radars. ICT for Smart Water Systems: Measurements and Data Science/Springer Nature. November 2019. pp.1-24. DOI: 10.1007/698_2019_404.

5. Васильев О.В., Зябкин С.А., Никоненко А.В., Чиров Д.С. Функционально-ориентированная модель формирования метеорологического продукта в X-диапазоне от гидрометеоров переохлажденной жидкости при полном поляризованном приеме. Цифровая Обработка Сигналов, №1, 2023. С.57-64.

6. D.S. Chirov, E.A. Bolelov, S.A. Zybkin and O.V. Vasiliev. «Fuzzy-logical Classifier of the Phase State of Hydrometeors in X-band Weather Radars», 2023 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF), St. Petersburg, Russian Federation, 2023, pp. 1-4, doi: 10.1109/WECONF57201.2023.10148003.

7. Руководство по метеорологическим приборам и методам наблюдений. ВМО-№ 8, Женева, 2010.

8. Руководство по сдвигу ветра на малых высотах. Doc 9817 AN/449. ИКАО, Монреаль, 2005.

9. Временные методические указания по использованию информации доплеровского метеорологического радиолокатора ДМРЛ-С в синоптической практике М. 2014. 110 с.

10. Временные методические указания по использованию информации доплеровского метеорологического радиолокатора ДМРЛ-С в синоптической практике М. 2017. 121 с.

11. Методические указания по использованию информации доплеровского метеорологического радиолокатора ДМРЛ-С в синоптической практике М. 2019. 129 с.

12. Авиационные правила. Часть 170. Сертификация оборудования аэродромов и воздушных трасс (АП-170). Том II. Сертификационные требования к оборудованию аэродромов и воздушных трасс. Дата актуализации: 01.02.2020.

13. Метеорологическое обеспечение международной аэронавигации. Прил.3 к Конвенции о международной гражданской авиации. ИКАО, Монреаль, 2013.

14. Основные технические требования к системе обнаружения опасных атмосферных явлений и штормового оповещения на базе метеорологических радиолокаторов. Приказ Росгидромета № 95 от 21.06.2004 г.

15. Руководство по авиационной метеорологии.

Изд.10. ИКАО, Монреаль, 2015.

16. Руководство по Глобальной системе наблюдений. Изд. 3-е. ВМО № 488, Женева, 2010.

17. Руководство по производству наблюдений и применению информации с неавтоматизированных радиолокаторов МРЛ-1, МРЛ-2, МРЛ-5. РД 52.04.320-91. СПб. 1993.

18. Руководство по требованиям к системе организации воздушного движения. Дос 9882. ИКАО, Монреаль, 2008.

19. Федеральные авиационные правила «Использование воздушного пространства Российской Федерации», приказ Минтранса РФ № 138 от 11.03.2010 г.

20. Федеральные авиационные правила «Предоставления метеорологической информации для обеспечения полетов воздушных судов» 3.03.2014 г. № 60.

ООО «Лаборатория Сфера»

Наша компания успешно занимается научными исследованиями и разработками в области связи, цифровой обработки сигналов и помехоустойчивого кодирования. Мы реализуем проекты в области радиосвязи, мобильной связи, радиодоступа и цифрового вещания.

Основные направления деятельности нашей компании:

- **Разработка ПО.** Разрабатываем специализированное компьютерное программное обеспечение.
- **Разработка для ПЛИС.** Разрабатываем сложные функциональные блоки цифровой обработки сигналов.
- **Научные исследования.** Проводим исследования в области интересующих Вас задач.
- **Разработка решений в области связи.** Решаем поставленные Вами задачи в области связи, радиодоступа и телерадиовещания.
- **Действия с интеллектуальной собственностью.** Осуществляем проведение патентных исследований, составление заявок на патенты на изобретения и полезные модели, а также на свидетельства на программы для ЭВМ и баз данных.
- **Консультации.** Осуществляем консультативную деятельность в области компьютерных технологий и систем связи.

Наши клиенты: ООО «Т8», ООО «3В-Сервис», ООО «Цифровые решения», ООО «НПФ «Сад-Ком», ООО НПП «ТЕПЛОВОДОХРАН», ООО «Силовая электроника» и др.

Контакты:

Адрес: 390035, г. Рязань, ул. Гоголя, д. 16, оф. 308

Тел.: +7(995)540-75-03

E-mail: info@labsphera.ru

Сайт: <https://labsphera.ru/>



ОПТИМАЛЬНЫЙ КИХ ФИЛЬТР С ДРОБНОЙ ЗАДЕРЖКОЙ

Андреев А.Г., к.т.н., доцент кафедры радиоэлектронных систем ИИФиРЭ ФГАОУ ВО «Сибирский федеральный университет», e-mail: aandrejev@sfu-kras.ru

Сенченко А.А., старший преподаватель каф. Радиотехники ИИФиРЭ ФГАОУ ВО «Сибирский федеральный университет», e-mail: asenchenko@sfu-kras.ru

OPTIMAL FRACTIONAL DELAY FIR FILTERS

Andreev A.G., Senchenko A.A.

To model a receiver with antenna array impinging signals must be delayed for a value which is not multiple of signal sampling rate. This problem can be solved using finite impulse response filters. The article discusses a method for modeling and implementation of a wideband delay element. Solution accuracy is studied depending on filter order and coefficients calculation algorithms.

Key words: fractional delay filters, FIR, optimal coefficients, antenna arrays, signal modelling, signal delay.

Ключевые слова: КИХ фильтр с дробной задержкой, оптимальные коэффициенты, антенные решетки, моделирование сигнала, задержка сигнала.

Введение

Существуют задачи, в которых требуется смоделировать сигнал на выходах антенной решетки, пришедший с произвольного направления. Одной из областей применения таких задач является отладка и моделирование адаптивных алгоритмов пространственной фильтрации помех [1, 2]. Фундаментальное свойство любой антенной решетки, которое требуется воспроизвести при создании её компьютерной модели, это задержка сигнала, возникающая из-за геометрической разности хода лучей и индивидуальная для каждого антенного элемента. Данная задержка зависит от направления прихода сигнала и может меняться, если источник перемещается.

Помимо озвученной задачи существуют и другие области, в которых необходимо осуществлять задержку сигнала на произвольную величину, например, в системах связи, в аудиосистемах, при сжатии сигналов, содержащих речь и т.д. [3].

Компьютерная модель антенной решетки работает с дискретизированными по времени сигналами. На вход модели поступают выборки сигнала, а с выходов модели должны идти копии входного сигнала, задержанные на величину разности хода лучей и не кратную периоду частоты дискретизации.

Существует несколько способов задержки цифрового сигнала. Один из наиболее простых заключается в кратном повышении частоты дискретизации, задержке полученного сигнала на необходимое число тактов и последующим понижением частоты дискретизации до прежнего уровня. Данный способ имеет два существенных недостатка: во-первых, задержка может принимать только дискретные значения, во-вторых, большие вычислительные затраты, так как требуется фильтрация сигнала на повышенной частоте дискретизации.

Второй способ основан на применении полиноми-

Для создания моделей обработки сигнала в антенных решетках необходимо иметь возможность задерживать сигнал на произвольную величину, в том числе не кратную периоду дискретизации сигнала. Данная задача может быть решена с помощью КИХ фильтров с дробной задержкой. Приведен способ расчета оптимальных коэффициентов КИХ фильтра с дробной задержкой.

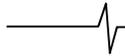
альной интерполяции [4] (фильтр Фарроу) и сводится к определению коэффициентов полинома, проходящего через выборки сигнала, и вычислению интерполированной выборки с нужной задержкой. К недостаткам данного метода относится наличие нелинейных искажений, большая погрешность интерполяции для более высокочастотных спектральных составляющих [5] и необходимость перерасчета коэффициентов полинома для каждой новой выборки. К достоинствам способа стоит отнести независимость коэффициентов интерполяции от требуемой величины задержки, что может быть важным для моделирования прохождения сигнала через антенную решетку.

Третий известный способ задержки сигнала на величину, не кратную частоте дискретизации, заключается в синтезе КИХ фильтра с равномерной АЧХ и линейной ФЧХ (то есть постоянным ГВЗ) в широкой полосе частот. Авторами данной работы был выбран этот способ, так как в нем отсутствуют нелинейные искажения сигнала, уровень частотных искажений может быть уменьшен до требуемого путем увеличения порядка КИХ фильтра, не требуется обработка сигнала на повышенной частоте дискретизации и, как следствие, меньшие вычислительные затраты, в сравнении с первым способом.

Постановка задачи

В соответствии с изложенным выше, требуется найти способ вычисления коэффициентов КИХ фильтра, с постоянной АЧХ и линейной ФЧХ в полосе пропускания, осуществляющего задержку широкополосного сигнала на величину, не кратную периоду дискретизации сигнала, то есть фильтра с дробной задержкой.

Задача задержки сигнала делится на два этапа: задержка сигнала на целое число тактов и задержка сиг-



нала на оставшуюся дробную часть. Таким образом, величина задержки может быть представлена следующим образом:

$$\tau = n + \tau_r, \tag{1}$$

где n – целая часть задержки; τ_r – дробная часть задержки, обе величины выражены в тактах частоты дискретизации.

Так как речь в статье пойдет о реализации дробной задержки сигнала на основе КИХ фильтра, следует указать важную особенность данного способа. Частотные характеристики КИХ фильтра с дробной задержкой будут наиболее близкими к идеальным тогда, когда задержка в этом фильтре удовлетворяет условию:

$$\frac{(N-1)}{2} - 0,5 \leq \tau_\phi \leq \frac{(N-1)}{2} + 0,5, \tag{2}$$

где N – порядок фильтра, нечетное целое число; τ_ϕ – задержка сигнала в фильтре, выраженная в тактах. Следовательно, полная задержка в КИХ фильтре с дробной задержкой определяется выражением:

$$\tau_\phi = \frac{(N-1)}{2} + \tau_r, \tag{3}$$

где дробная часть задержки удовлетворяет условию $|\tau_r| \leq 0,5$. Действительно, дробная часть задержки может быть отрицательной, так как полная задержка τ_ϕ больше нуля. Расчет коэффициентов КИХ фильтра с дробной задержкой везде далее будет проводиться для $\tau_r = 0,5$, так как в этом случае отклонение АЧХ и ФЧХ фильтра от требуемых значений в полосе пропускания будет наибольшим в сравнении с другими значениями τ_r .

Синтез оптимального КИХ фильтра с дробной задержкой

Обозначим вектор-столбцом h набор коэффициентов КИХ-фильтра с дробной задержкой:

$$h = [h_0, h_1, \dots, h_{N-1}]^T, \tag{4}$$

где N – порядок фильтра, $()^T$ – оператор транспонирования.

Тогда частотная характеристика фильтра будет определяться выражением:

$$H(\omega, h) = e^{j\omega} H, \tag{5}$$

где $e(\omega) = [1, e^{j\omega}, e^{j2\omega}, \dots, e^{j(N-1)\omega}]^T$ – вектор-столбец комплексных экспонент, $()^H$ – операция транспонирования и комплексного сопряжения. Определим оптимальные коэффициенты КИХ фильтра с дробной задержкой по критерию минимума среднего квадрата ошибки:

$$e(h) = \int_{\Omega} W(\omega) |E_c(\omega, h)|^2 d\omega, \tag{6}$$

где $E_c(\omega, h)$ – комплексная функция ошибки:

$$E_c(\omega, h) = H(\omega, h) - D(\omega),$$

где $D(\omega)$ – требуемая комплексная частотная характеристика.

В работе [6] показано, что ошибка, вычисленная по формуле (6), будет минимальна для коэффициентов:

$$h_{opt} = R^{-1} s, \tag{7}$$

$$R = \int_{\Omega} W(\omega) e(\omega) e^H(\omega) d\omega, \tag{8}$$

$$s = \int_{\Omega} W(\omega) D(\omega) e(\omega) d\omega. \tag{9}$$

В результате вычисления интеграла в формуле (8) получается квадратная матрица, а в формуле (9) – вектор-столбец. Чтобы вычислить их зададим весовую функцию в следующем виде:

$$W(\omega) = \begin{cases} 0, & \omega < 0 \\ 1, & 0 \leq \omega \leq \lambda, \\ 0, & \lambda < \omega \end{cases} \tag{10}$$

где $\lambda = 2\pi f / f_s$ – нормированная граничная частота в диапазоне $[0; \pi]$, f – верхняя частота полосы пропускания фильтра, f_s – частота дискретизации. Меняя нормированную частоту λ можно задавать частотный диапазон, в пределах которого необходимо минимизировать величину ошибки (6).

Для КИХ фильтра с дробной задержкой требуемая ЧХ $D(\omega)$ определяется выражением [3]:

$$D(\omega) = e^{-j\omega\tau}, \tag{11}$$

где τ – задержка сигнала, нормированная к периоду частоты дискретизации.

Подставив (10) и (11) в выражения (8) и (9), получим формулы для вычисления элементов матриц R и s :

$$R_{k,l} = 2 \frac{\sin[\lambda(k-l)]}{k-l} = 2\lambda \operatorname{sinc}[\lambda(k-l)], \tag{12}$$

$$s_k = 2 \frac{\sin[\lambda(k-\tau)]}{k-\tau} = 2\lambda \operatorname{sinc}[\lambda(k-\tau)]. \tag{13}$$

Как можно заметить, матрица, определяемая (12), является Тёплицевой матрицей, не зависящей от требуемой задержки τ , а значит обратная матрица R может быть вычислена заблаговременно и позже использована для расчетов коэффициентов КИХ фильтра с дробной задержкой.

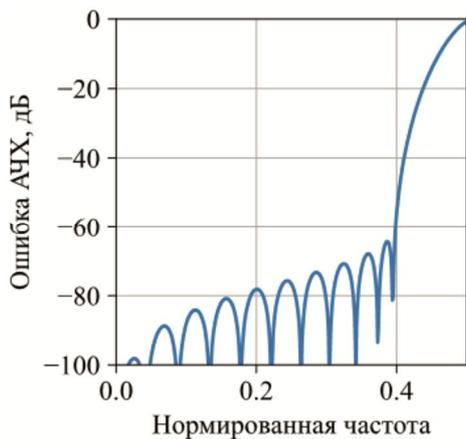
Результаты синтеза оптимальных КИХ фильтров

В качестве примера выполним синтез оптимального КИХ фильтра 21 порядка с полной задержкой $\tau_\phi = 10,5$ тактов и полосой пропускания $\lambda = 2\pi \cdot 0,4$ (т.е. средний квадрат ошибки будет минимизирован в полосе от 0 до 0,4 частоты дискретизации). В результате синтеза получим следующие коэффициенты:

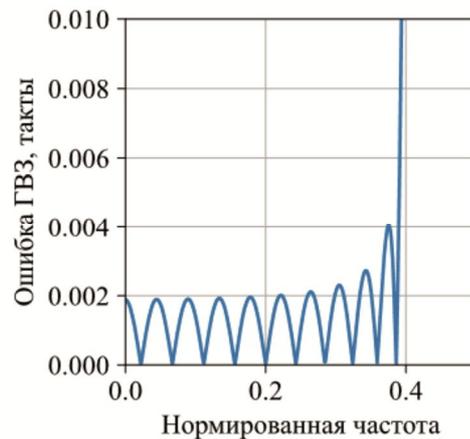
| | | | |
|-----------|-----------|-----------|-----------|
| 0.000561 | -0.001853 | 0.004446 | -0.009012 |
| 0.016403 | -0.027751 | 0.044799 | -0.070917 |
| 0.114698 | -0.208062 | 0.641471 | 0.622899 |
| -0.190370 | 0.098659 | -0.057108 | 0.033549 |
| -0.019126 | 0.010230 | -0.004940 | 0.002022 |
| -0.000602 | | | |

На рис. 1. приведены частотные зависимости ошибок АЧХ и ГВЗ (отклонение частотных характеристик синтезированного фильтра от требуемой $D(\omega)$).

На рис. 2 представлены зависимости максимальной ошибки АЧХ и ГВЗ в заданном при синтезе частотном диапазоне от порядка фильтра N . Представленные зависимости построены для двух нормированных граничных частот: сплошные линии соответствуют $\lambda = 2\pi \cdot 0,40$,

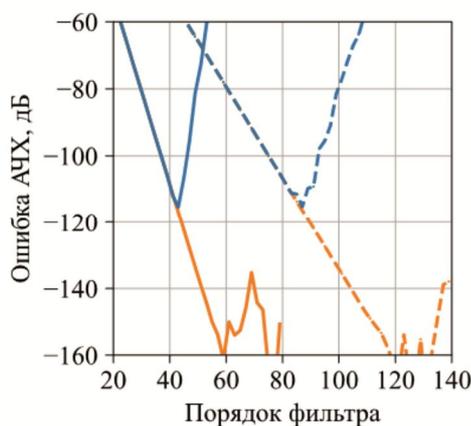


а)

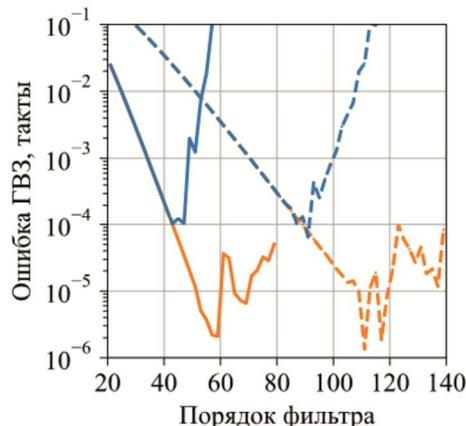


б)

Рис. 1. Ошибка АЧХ и ГВЗ для фильтра 21 порядка с параметрами: $\tau_{\phi} = 10,5$ тактов и полосой пропускания $\lambda = 2\pi \cdot 0,4$



а)



б)

Рис. 2. Зависимости максимальной ошибки АЧХ (а) и ГВЗ (б) от порядка фильтра для разных методов вычисления оптимальных коэффициентов

пунктирные – $\lambda = 2\pi \cdot 0,45$. Графики получены с применением метода `MatrixXd::fullPivLu()` библиотеки Eigen v3.4.90 (библиотека линейной алгебры, написанная на языке программирования C++). Метод позволяет вычислить как обратную матрицу, так и решить систему линейных алгебраических уравнений (СЛАУ) без непосредственного вычисления обратной матрицы. В основе метода используется LU разложение с полным поворотом (LU decomposition with complete pivoting). В документации к библиотеке Eigen данный метод позиционируется как один из самых точных. Зависимость ошибок АЧХ и ГВЗ от порядка фильтра N при расчетах по формуле (7) (с вычислением обратной матрицы) показаны на рисунке синим цветом. Зависимость ошибок АЧХ и ГВЗ от порядка фильтра N при вычислении коэффициентов путем решения уравнения $Rh_{\text{opt}} = s$ (без непосредственного вычисления обратной матрицы) показаны на рисунке оранжевым цветом. Как видно из графиков, предпочтительнее использовать методы решения СЛАУ без непосредственного вычисления обратной матрицы. При вычислении обратной матрицы не удастся получить ошибки АЧХ и ГВЗ меньше -115 дБ и $\sim 10^{-4}$ такта соответственно. При использовании методов без обращения матрицы порядок фильтра можно увеличивать, пока ошибки АЧХ и

ГВЗ не достигнут уровня -160 дБ и $\sim 2 \cdot 10^{-6}$ такта. Все расчеты проводились с использованием чисел с плавающей точкой двойной точности (размером 8 байт).

Заключение

В статье получены и приведены формулы для вычисления оптимальных коэффициентов КИХ фильтра с дробной задержкой. В качестве критерия оптимальности использовался минимум среднего квадрата ошибки (6). Для получения малых ошибок АЧХ и ГВЗ необходимо синтезировать КИХ фильтры с высоким порядком, однако ограничиваясь вычислениями с числами с плавающей точкой двойной точности, не удастся получить ошибки ниже -160 дБ и $\sim 2 \cdot 10^{-6}$ такта соответственно. Для дальнейшего увеличения порядка фильтра и уменьшения величины ошибок АЧХ и ГВЗ синтезированных фильтров необходимо использовать специальные методы решения СЛАУ либо использовать в расчетах числа с большей точностью. Например, можно использовать библиотеку GMP [7], однако в этом случае время синтеза КИХ фильтра сильно увеличивается.

Исследование выполнено в рамках государственного задания ФГАОУ ВО Сибирский федеральный университет (номер FSRZ-2023-0008).

Литература

1. Комплексная модель цифровой антенной решетки спутниковых радионавигационных систем на корпусе летательного аппарата / А.В. Ряполов, Д.А. Гредяев, О.В. Юрченко, Н.В. Фамбулов. Сбор. докл. V Междун. н.-техн. конф. Радиотехника, электроника и связь: Омск., 2019.

2. Сенченко А.А. LCMV алгоритм при воздействии имитационных помех. Перспективы развития фундаментальных наук: сборник трудов XVIII Международной конференции студентов, аспирантов и молодых ученых. В 7 томах. Том 7. / под ред. И.А. Курзиной, Г.А. Вороновой, Томск: Изд-во Томского политехнического университета, 2021. 110-113 с.

3. V. Valimaki, T.I. Laakso. Fractional Delay Filters-Design

and Applications.2001. 10.1007/978-1-4615-1229-5_20.

4. Цифровая передискретизация сигналов на основе полиномиальной интерполяции. Фильтр Фарроу [Электронный ресурс]: https://ru.dsplib.org/content/resampling_lagrange/resampling_lagrange.html, дата обращения 28.08.2023.

5. Оценка искажений при передискретизации цифрового сигнала с использованием фильтра Фарроу / М.И. Спажакин, В.Д. Репников, А.Б. Токарев. Вестник Воронежского государственного технического университета, 2013.

6. Algorithms for the Constrained Design of Digital Filters with Arbitrary Magnitude and Phase Responses, Mathias Lang, 1999.

7. The GNU Multiple Precision Arithmetic Library: <https://gmplib.org/>.



Всероссийская конференция «СОВРЕМЕННЫЕ ТЕХНОЛОГИИ ОБРАБОТКИ СИГНАЛОВ» (СТОС-2023)

12 – 13 декабря 2023 г. Россия, Москва

ALL-RUSSIAN CONFERENCE

"MODERN TECHNOLOGIES OF SIGNAL PROCESSING" (MTSP-2023)

Уважаемые коллеги!



ПРИГЛАШАЕМ ВАС ПРИНЯТЬ УЧАСТИЕ В РАБОТЕ КОНФЕРЕНЦИИ

ОРГАНИЗАТОРЫ:

- Российское научно-техническое общество радиотехники, электроники и связи им. А.С. Попова
- Институт радиотехники и электроники им. В.А. Котельникова РАН
- Московский технический университет связи и информатики
- Министерство образования и науки Российской Федерации

ПРИ УЧАСТИИ:

- ОАО «Концерн радиостроения «Вега»
- ФГУП «НИИР»
- Балтийский федеральный университет им. И. Канта
- Владимирский государственный университет
- ОАО «Концерн «Созвездие»
- Московский энергетический институт
- Московский институт электронной техники
- Московский государственный технический университет им. Н.Э. Баумана
- Военная академия РВСН им. Петра Великого
- Тульский государственный университет
- Пензенский государственный университет
- Нижегородский технический госуниверситет им. Р.Е. Алексеева
- Рязанский государственный радиотехнический университет им. В.Ф. Уткина
- Санкт-Петербургский государственный университет телекоммуникаций им. проф. А.М. Бонч-Бруевича
- Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина)
- Ульяновский государственный технический университет
- Ярославский государственный университет

ПРОГРАММНЫЙ КОМИТЕТ

Председатель: академик РАН Сергей Аполлонович Никитов.

Заместитель Председателя: академик РАН Кузнецов Н.А.

Члены комитета: д.т.н, проф. Витязев В.В., д.т.н., проф. Вишневский В.М., д.т.н., проф. Джиган В.И., член-корр. РАН Дворкович А.В., д.ф.-м.н., проф. Миллер Б.М., д.т.н., проф. Рубинович Е.Я., д.ф.-м.н., проф. Семенович К.В., д.т.н., проф. Сергеев В.В., д.т.н. Чочиа П.А.

ОРГАНИЗАЦИОННЫЙ КОМИТЕТ:

Председатель: д.т.н., профессор Артём Сергеевич Аджемов

Заместитель Председателя: д.т.н., профессор Сергеев В.А.

Члены Оргкомитета: доц. Алёшин В.С., проф. Брюханов Ю.А., проф. Васильев К.К., доц. Иванов А.П., проф. Калошин В.А., проф. Мамон Ю.И., проф. Пахотин В.А., проф. Постников И.И., доц. Самсонов Г.А., проф. Степанов С.Н., доц. Тычков А.Ю., проф. Цимбал В.А., проф. Чиров Д.С., проф. Ямпури Н.П.

Организация работы в форме пленарных и проблемно-тематических заседаний. Пленарные доклады будут представлены по основным направлениям работы Конференции. Доклады, включенные в Программу конференции, будут опубликованы в Сборнике «Доклады Конференции» и размещены на сайте <http://www.rntores.ru/> в формате <pdf>. По опыту прошедших конференций, мероприятие оказалось очень эффективным местом общения для всех участников мероприятия: представителей науки и производства, заказчиков и разработчиков, потребителей и поставщиков, преподавателей и студентов, работодателей и соискателей, научно-технических издательств и читателей.

ВОССТАНОВЛЕНИЕ ТРАЕКТОРНОГО СИГНАЛА В УСЛОВИЯХ ПЕРИОДИЧЕСКИХ ЗАТЕНЕНИЙ МЕТОДОМ АВТОРЕГРЕССИОННОГО АНАЛИЗА

Дворянков Д.А., аспирант кафедры телекоммуникаций и основ радиотехники ФГБОУ ВО «Рязанский государственный радиотехнический университет им. В.Ф.Уткина», e-mail: dvoryankov.d.a@mail.ru

Андросов В.В., главный конструктор по направлению АО «Государственный Рязанский приборный завод»

Волченков В.А., к.т.н., доцент кафедры телекоммуникаций и основ радиотехники ФГБОУ ВО «Рязанский государственный радиотехнический университет им. В.Ф.Уткина», e-mail: volchenkov.v.a@tor.rsreu.ru

Витязев С.В., к.т.н., доцент кафедры телекоммуникаций и основ радиотехники ФГБОУ ВО «Рязанский государственный радиотехнический университет им. В.Ф.Уткина», e-mail: vityazev.s.v@tor.rsreu.ru

TRAJECTORY SIGNAL RESTORING IN CONDITIONS OF PERIODIC SHADING BY AUTOREGRESSIVE ANALYSIS METHOD

Dvoryankov D.A., Androsov V.V., Volchenkov V.A., Vityazev S.V.

The autoregressive analysis algorithm is described in relation to the signal shape restoration, which was distorted due to the helicopter rotor blades influence. Alternative restoring methods are also provided in this paper. The AR analysis method is compared with other recovery methods by modeling the reflected radio signals processing. The ability of AR analysis to recover both simple and more complex distorted signals is shown. The dependences of the recovery quality on the speed and bearing angle of the radar are given, and the computational costs are estimated. The computational costs and algorithm ability to work in real time are also evaluated.

Key words: radar systems, interpolation, AR model, radar image, filling gaps in radar signal.

Ключевые слова: радиовидение, интерполяция, авторегрессия, радиолокационное изображение, восстановление пропусков в сигналах.

Введение

В настоящее время получение максимально доступной информации играет важную роль во многих сферах науки, техники и повседневной реальности. Возрастающие требования к решению всё более и более сложных задач во всех отраслях науки приводят к необходимости собирать и обрабатывать большой объём информации практически непрерывно, в любое время суток и при любых погодных условиях. Возник такой термин как «глобальный мониторинг», который означает получение информации со всех точек земного шара и околоземного пространства в режиме реального времени. Одними из систем, способных решать задачу глобального мониторинга, являются системы радиовидения. Используя подобные системы, можно получить радиолокационные изображения, сравнимые по детальности с фотографиями, полученными при помощи оптических систем.

Во многих малогабаритных самолётах и, преимущественно, в вертолётах задачи получения информации наряду с оптико-электронными и инфракрасными датчиками решают и системы РЛС. Такие системы в меньшей степени чем оптические подвергаются влиянию погодных условий и способны работать с одинаковой эффективностью круглосуточно. Благодаря данным преимуществам бортовая РЛС установлена на большинстве активно используемых летательных средств.

Рассматривается использование авторегрессионного анализа (АР) для восстановления формы траекторного сигнала бортового радиолокационного комплекса (БРЛК) вертолёта, искажённой периодическими «затенениями» лопастями вращающегося винта. Метод АР-анализа сравнивается с другими способами восстановления путем моделирования обработки набора отражённых радиосигналов. Демонстрируется, что метод АР-анализа позволяет восстанавливать как простые, так и более сложные искажённые траекторные сигналы. Приводятся зависимости качества восстановления сигнала от скорости носителя и величины угла обзора. Производится оценка вычислительных затрат. Оценивается способность работы алгоритма в режиме реального времени.

Достаточно важным является выбор варианта расположения РЛС на борту вертолёта. В последнее время большую популярность получил вариант расположения БРЛС над вращающимся винтом. Такое расположение имеет множество преимуществ, однако влечёт за собой и возникновение новых проблем. Одной из таких проблем является искажение радиолокационного сигнала из-за наличия вращающихся лопастей винта.

Постановка задачи

Пусть есть сигнал $S(t)$, отражённый от земной поверхности и принятый антенной РЛС. В силу негативного влияния отражения сигнала от лопастей винта вертолёта значения в некоторых отсчётах времени t отсутствуют (затенения или пропадания сигнала). Необходимо найти оценку $\hat{S}(t)$ значений $S(t)$ в местах пропусков.

Таким образом, ставится задача компенсации эффекта негативного влияния вращающихся лопастей на сигналы, поступающие на приёмную антенну БРЛК вертолёта.

Эффект влияния лопастей вызывает пропадания фрагментов сигнала, вследствие чего возникает проблема восстановления формы траекторного сигнала [1].

Рассматриваются разные способы восстановления сигнала, в том числе с применением интерполяционных методов и авторегрессионного анализа. Данные методы применяются для восстановления смоделированного сигнала отражения от подстилающей поверхности, содержащей точечные отражатели. В качестве критериев оценки рассматриваются точность восстановления и время обработки. Также были оценены вычислительные затраты на реализацию алгоритма, позволяющего решать поставленную задачу.

Применение методов интерполяции

Интерполяция или интерполирование – это приближённое или точное нахождение величины по её отдельным известным значениям или по значениям других величин, связанных с ней [2].

В процессе интерполирования необходимо определить каким образом ведёт себя функция между заданными точками. Рассматриваемые точки могут быть интерполированы при помощи разных функций, поэтому надо иметь тот или иной критерий выбора. Обычно такие критерии основываются на степени гладкости интерполирующей функции или на объёме вычислений [3].

Большую часть интерполирующих функций можно получить путём комбинации двух или более элементарных функций. Линейные комбинации одночленов образуют алгебраические полиномы, комбинации тригонометрических функций – тригонометрические полиномы и так далее.

Самым распространённым типом интерполирующих функций можно назвать кольцо алгебраических многочленов (полиномов). Преимущество полиномов заключается в том, что их значения легко вычислить. Также значения полиномов просто складывать, умножать, дифференцировать и интегрировать. Существует ещё одна особенность полиномов: если c – постоянная величина, а $P_n(x)$ – полином n -го порядка, то полиномами тех же порядков будут являться также $c \cdot P_n(x)$ и $P_n(x+c)$. Существует множество классов функций со схожими свойствами, но в то же время не все они имеют положительные интерполяционные качества [3].

Ниже приведём выражение для канонического полинома степени n :

$$P_n(x) = a_0 + a_1x^1 + a_2x^2 + \dots + a_{n-1}x^{n-1} + a_nx^n. \quad (1)$$

Выбор полинома степени n основан на том, что через $n + 1$ точку проходит единственная кривая степени n . Коэффициенты интерполяционного многочлена $a_0, a_1, a_2, \dots, a_n$ могут быть найдены при решении системы алгебраических уравнений:

$$\begin{cases} a_0 + a_1x_0 + a_2x_0^2 + \dots + a_{n-1}x_0^{n-1} + a_nx_0^n; \\ a_0 + a_1x_1 + a_2x_1^2 + \dots + a_{n-1}x_1^{n-1} + a_nx_1^n; \\ a_0 + a_1x_2 + a_2x_2^2 + \dots + a_{n-1}x_2^{n-1} + a_nx_2^n; \\ a_0 + a_1x_n + a_2x_n^2 + \dots + a_{n-1}x_n^{n-1} + a_nx_n^n. \end{cases} \quad (2)$$

Интерполяционный многочлен можно описать разными выражениями, подходящими для решения той или иной задачи. Однако стоит заметить, что при достаточно большом объёме данных, интерполяционный полином будет иметь слишком высокую степень. В таком случае процесс вычисления интерполяционной формулы будет очень долгим, а погрешность интерполяции на концах интервала станет недопустимо велика в силу накопления ошибок округления, иначе говоря формула будет неустойчивой. Возникает потребность в разбиении интервала интерполяции с некоторым шагом h , с последующим построением своего интерполяционного многочлена на каждом из получившихся отрезков. Такая кусочно-заданная интерполяционная формула, обеспечивающая наиболее высокую степень гладкости для данного класса функций, называется сплайном.

Если рассматривать кубический сплайн, то главный полином на каждом интервале интерполяции можно записать в следующем виде:

$$f_{3i}(x) = a_i x^3 + b_i x^2 + c_i x + d_i. \quad (3)$$

В [1] было показано, что на конкретной модели РЛИ оптимальным с точки зрения точности восстановления является метод интерполяции сплайнами. Однако при тестировании данного метода применительно к другому типу сигнала оказалось, что данный метод не всегда показывает удовлетворительные показатели восстановления. Возникла задача поиска другого, более универсального, метода восстановления фрагментов сигнала. Подходящим оказался метод восстановления сигнала на основе модели авторегрессии.

Авторегрессионный анализ в задаче восстановления

Авторегрессионная модель (АР-модель) – модель временных рядов, в которой значения временного ряда в данный момент времени линейно зависят от предыдущих значений этого же ряда. Авторегрессионный процесс порядка p определяется следующим образом:

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \dots + \phi_p \tilde{z}_{t-p} + a_t, \quad (4)$$

где ϕ_1, \dots, ϕ_p – параметры модели (коэффициенты авторегрессии), a_t – белый шум [4].

Первостепенной задачей при проведении авторегрессионного анализа является нахождение оптимальных значений коэффициентов авторегрессии. Большинство методов предполагает, что ряд \tilde{z}_t является линейным и стационарным. Обычно также предполагается, что ряд \tilde{z}_t имеет нулевое среднее значение, в противном случае в приведенном выше уравнении (4) появляется ненулевой член a_0 .

Существует множество методов оценки коэффициентов АР-модели, среди которых можно выделить два основных: метод наименьших квадратов (МНК) [5] и метод Берга [6]. Каждый из этих методов имеет в свою очередь несколько вариаций.

Самый распространённый метод нахождения коэффициентов включает умножение приведенного выше выражения (4) на \tilde{z}_{t-d} и нормализацию, что даёт набор

линейных уравнений, называемых уравнениями Юла – Уолкера, которые могут быть записаны в матричной форме. В нашем случае мы будем использовать алгоритм нахождения коэффициентов на основе метода Берга, в силу того, что в данном методе при проведении анализа коротких сигналов прослеживается высокая разрешающая способность [6]. Другое название этого метода – алгоритм максимальной энтропии, при помощи него можно произвести точную оценку значений коэффициентов авторегрессии.

Ниже приведём блок-схему рассматриваемого алгоритма (рис. 1), обозначив входной сигнал как $S(n)$. Блок поиска участков затенения используется для нахождения искажённых или потерянных фрагментов сигнала и будет более подробно рассмотрен в разделе, посвящённом моделированию. После определения участков затенения выполняется алгоритм восстановления для каждого фрагмента.

Алгоритм восстановления для каждого участка затенения сигнала состоит из следующих шагов: сначала исходный временной ряд центрируется путем вычитания среднего. Далее оцениваются прямая и обратная авторегрессионные модели. На последнем шаге выполняется оптимальное усреднение результатов прямой и обратной моделей с предварительным прибавлением среднего [7].



Рис. 1. Блок-схема алгоритма восстановления сигнала на основе АР-модели

Рассмотрим более подробно каждый из этапов алгоритма восстановления. Фрагмент сигнала с пропусками проходит через прямую и обратную модели предсказания. В обоих случаях используется функция *arfill* (рис. 2). Приведём математическое описание основных этапов выполнения данной функции, какими являются: расчёт коэффициентов авторегрессии, расчёт начальных условий и непосредственно оценка пропущенных значений.

Оценка коэффициентов авторегрессии производится при помощи метода Берга. Данный метод оперирует с входным сигналом и позволяет рассчитать $m + 1$ коэффициентов отражения, из которых формируется m коэффициентов авторегрессии a , где m – порядок АР-модели. Математическое описание алгоритма поиска

коэффициентов предсказания в виде блок-схемы представлено на рис. 3. Здесь в последнем блоке $conj(\cdot)$ обозначает комплексное сопряжение, т.е. $conj(x) = real(x) - j*imag(x)$.



Рис. 2. Блок-схема функции arfill

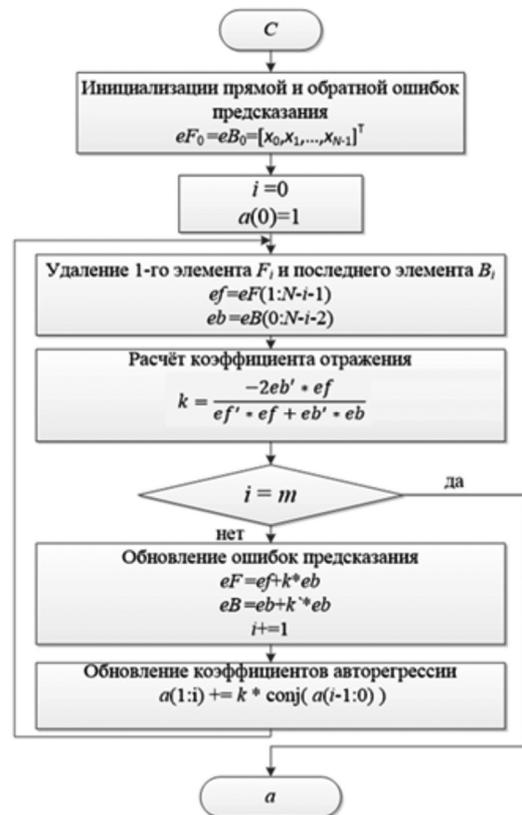


Рис. 3. Блок-схема алгоритма нахождения коэффициентов предсказания

Расчёт начальных условий происходит на основе предыдущих значений сигнала $s(n)$, а также коэффициентов авторегрессии a и реализуется при помощи стандартного КИХ-фильтра:

$$z(n) = \sum_{l=0}^m b(l) \cdot x(n-l). \quad (5)$$

Здесь $b = -a(m+1:2)/a(1)$, x – предыдущие m значений сигнала $s(n)$.

Оценка искажённых значений осуществляется с помощью фильтра с обратной связью (БИХ) и использует в качестве начальных условий ранее рассчитанные значения $z(n)$. Математическое описание данной операции представлено ниже:

$$y(n) = -\sum_{k=0}^m a(k) \cdot y(n-k). \tag{6}$$

Здесь a – коэффициенты авторегрессии, а $y(-(m-1):0) = z$ – рассчитанные выше начальные условия. Также на этом этапе происходит добавление среднего при расчёте окончательной оценки значений y .

Расчёт весов w для последующего оптимального усреднения происходит следующим образом: $w(0:L_{gap}-1) = [L_{gap}, L_{gap}-1, \dots, 0]$, где L_{gap} – размер участка пропадания сигнала, подлежащего восстановлению.

Вышеописанные операции производятся одинаково как для прямой, так и для обратной моделей предсказания. В результате имеем оценки y_f, y_b , а также массивы весов w_f и w_b , при этом y_b и w_b «переворачиваются», то есть выполняются операции: $y_b(L-1:0) = y_b(0:L-1)$, $w_b(L-1:0) = w_b(0:L-1)$, где L – длина обрабатываемого сигнала (в нашем случае сигнал обрабатывается в каждом канале дальности для каждого парциального кадра и $L = 200$ отсчётов).

Последней операцией согласно рис. 1 является оптимальное усреднение, которое выполняется по следующей формуле:

$$y(n) = \frac{s_f(n) \cdot w_f(n) + s_b(n) \cdot w_b(n)}{w_f(n) + w_b(n)}, n = 1..L. \tag{7}$$

Сравнение методов восстановления на примере фрагмента сигнала

Один из способов оценить насколько точно регрессионная модель соответствует набору данных – вычислить среднеквадратичную ошибку, которая сообщает нам среднее расстояние между прогнозируемыми моделью значениями и фактическими значениями в наборе данных [8]. Такую метрику качества восстановленного изображения часто используют, при этом обозначая её аббревиатурой *RMSE*. Формула для нахождения среднеквадратичной ошибки выглядит следующим образом:

$$RMSE = \sqrt{\sum_{i=1}^N \sum_{j=1}^M (S_{i,j} - \bar{S}_{i,j})^2 / N \cdot M}. \tag{8}$$

Таким образом, можно теперь в полной мере оценить качество восстановления сигнала при помощи авторегрессионной модели и сравнить данный подход с

методами на основе интерполяции (табл. 1).

Сравнение методов восстановления было выполнено на примере фрагментов двух различных групп сигналов, они обозначены в таблице 1 как сигнал А и сигнал Б. В фрагменте сигнала А (200 отсчётов) в результате вращения вращающихся лопастей появляются 8 участков пропадания сигнала по 3 отсчёта каждый. В фрагменте сигнала Б (также 200 отсчётов) – 4 участка пропадания сигнала по 5 отсчётов каждый.

Моделирование обработки траекторного сигнала

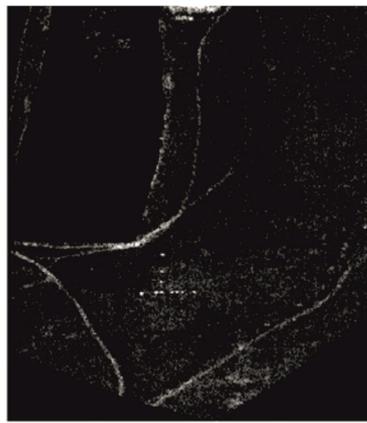
Перейдём от фрагмента сигнала к общему изображению, представляющему собой модель результата зондирования РЛС земной поверхности. Цель моделирования – сымитировать сигнал, отражённый от земной поверхности с учётом влияния лопастей и выполнить его обработку. Известны частота вращения лопастей и время затенения сигнала. В качестве модели земной поверхности используется радиолокационное изображение, полученное в режиме реального луча. На это изображение накладываются дополнительные точечные цели, удобные для проведения исследований. Полученное оптическое изображение преобразуется обратно в сигнал во временной области с использованием известных математических выражений [1]. В модель сигнала закладываются требуемые параметры движения носителя и угла обзора. В частности, скорость носителя выбрана равной 80 м/с, угол обзора – 90 градусов. Цели моделируются путём умножения сигнала на последовательность импульсов, частота которых соответствует частоте вращения лопастей, а дальность – ширине лопасти. Значения импульсов равны 0 и -20 дБ соответственно в моменты между импульсами и на длительности импульсов, имитирующих влияние лопастей.

Описанный выше алгоритм предлагается применять для практических задач, но при этом появляется ещё одна проблема – поиск моментов пропадания сигнала. Алгоритм поиска участков пропадания сигнала предлагается реализовать следующим образом: в цикле по каналам дальности производится суммирование квадратов реальной и мнимой частей сигнала. Полученный массив S сравнивается с адаптивным пороговым значением. Если $S(i)$ меньше порогового значения, индекс i записывается в массив индексов затенений. Следующим этапом является формирование границ фрагментов пропадания сигнала из общего массива индексов затенений.

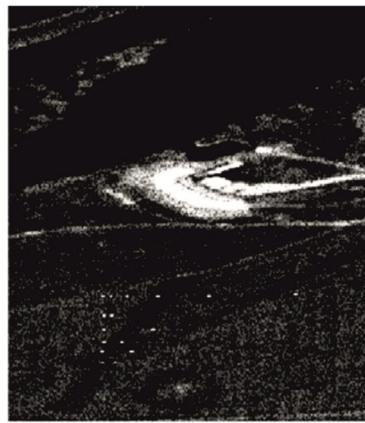
Рассмотрим расчёт порогового значения. Для сигнала в каждом кадре порог рассчитывается по формуле $\hat{S} * h$, где \hat{S} – среднее значение массива S , h – порог, рассчитываемый в данном случае следующим образом:

Таблица 1. Сравнение методов восстановления сигнала

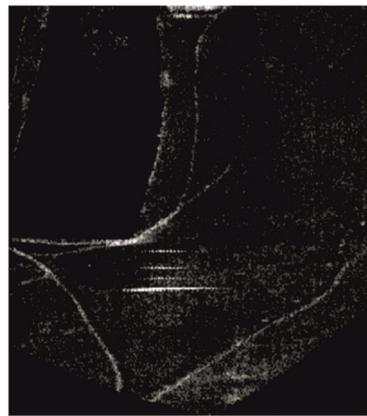
| | Сигнал А | | Сигнал Б | |
|------------------------------------|----------|----------|----------|----------|
| | RMSE | Время, с | RMSE | Время, с |
| Эталон | 0 | 9,260 | 0 | 5,430 |
| Линейная интерполяция | 0,201 | 12,530 | 1,597 | 6,520 |
| Полином Эрмита | 0,148 | 12,621 | 1,573 | 6,550 |
| Метод Акимы | 0,121 | 12,590 | 1,573 | 6,550 |
| Интерполяция кубическими сплайнами | 0,060 | 12,780 | 1,715 | 6,670 |
| Авторегрессионная модель | 0,084 | 37,020 | 0,582 | 17,920 |



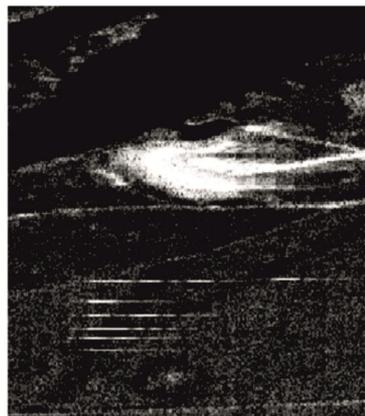
a)



д)



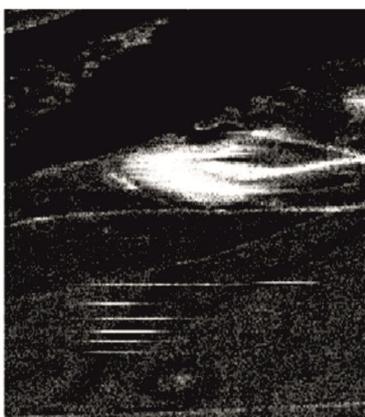
б)



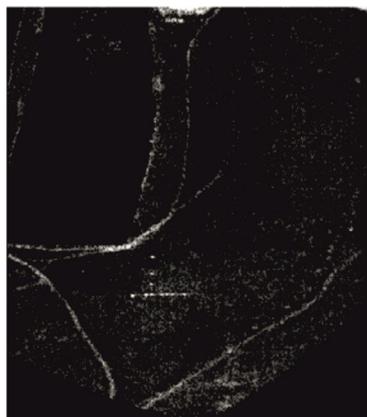
е)



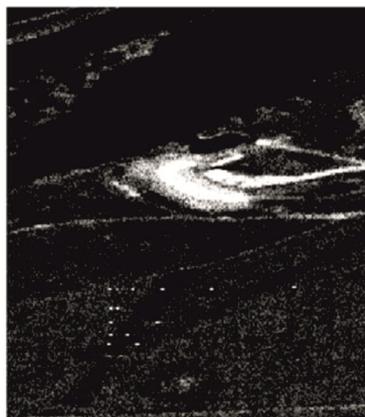
в)



ж)



г)



з)

Рис. 4. Пример восстановления сигналов при помощи сплайн-интерполяции и AP-анализа

$$h = \frac{\hat{S}}{\max(S) * 2}. \quad (9)$$

Таким образом, результатом работы алгоритма поиска участков пропадания являются два массива с начальными и конечными индексами участков пропадания сигнала соответственно. Далее перейдём непосредственно к моделированию восстановления сигнала, искажённого за счёт негативного влияния лопастей винта.

Исходный сигнал А без помех от лопастей представлен на рис. 4, а. Сигнал А, искажённый из-за негативного влияния вращающихся лопастей показан на рис. 4, б, тот же сигнал, но уже восстановленный при помощи интерполяции кубическими сплайнами и методом на основе АР-модели, представлен на рис. 4, в и 4, г соответственно. Иллюстрация восстановления сигнала Б показана аналогично на рис. 4, д – 4, з.

На основе анализа полученных изображений (рис. 4) можно сделать вывод, тождественный выводу из результатов, представленных в табл. 1. Метод АР-анализа позволяет восстанавливать как простые, так и более сложные искажённые траекторные сигналы, в отличие от применяемого ранее метода интерполяции сплайнами.

Заметим, что для суждения об эффективности метода необходимо убедиться в том, что он будет работать при разных параметрах исследуемого сигнала. Поэтому для полноценного сравнения и выбора оптимального метода восстановления формы траекторного сигнала, искажённого вращающимися лопастями винта, были рассчитаны зависимости качества восстановления ($RMSE$) от следующих параметров РЛС: скорости носителя и угла обзора. За основу был взят сигнал Б, так как его параметры наиболее приближены к «худшему» случаю для реального сигнала.

При моделировании были заданы различные значения скорости носителя в диапазоне от 40 до 100 м/с. При этом формировались различные РЛИ и оценивалось качество их восстановления при помощи выбранного параметра $RMSE$. На рис. 5 представлены зависимости $RMSE$ от скорости носителя для восстановления при помощи методов интерполяции и авторегрессии. Также на рис. 5 приведена кривая зависимости значения $RMSE$ от скорости носителя без применения алгоритма восстановления (обозначено как «лопасти»). Можно заметить, что метод на основе авторегрессионного анализа отличается стабильно высоким качеством восстановления, в то время как интерполяционный метод при низких значениях скорости носителя и вообще повышает значение среднеквадратичной ошибки, что говорит о неэффективности применения данного метода при подобных параметрах.

Приведём также зависимость качества восстановления от угла обзора. На рис. 6 представлены зависимости $RMSE$ от угла обзора для восстановления при помощи методов интерполяции и авторегрессии. Также на рис. 6 приведена кривая зависимости значения $RMSE$ от угла обзора без применения алгоритмов восстановления.

Можно заметить, что при значениях угла, близких к 80° качество восстановления снижается. Этот эффект

связан с тем, что при таком угле обзора моделируемые точечные отражатели находятся на самом краю видимого РЛИ. При других углах обзора, когда отражатели полностью находятся в видимой области, наблюдается высокая степень качества восстановления. В то же время метод на основе интерполяции при углах отличных от 90° показывает отрицательный результат и не может быть применён для восстановления сигнала.

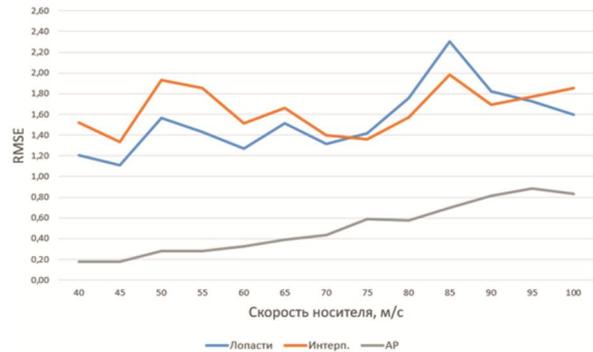


Рис. 5. Зависимость качества восстановления от скорости носителя

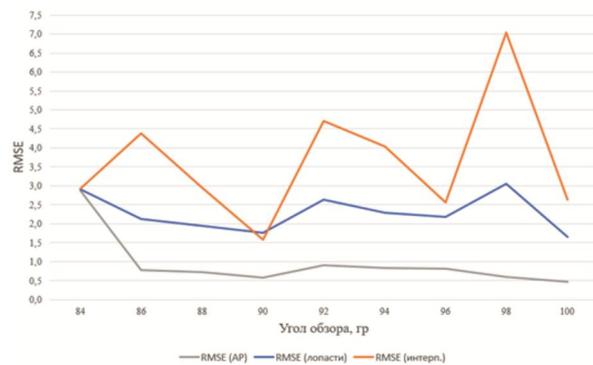
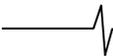


Рис. 6. Зависимость качества восстановления от угла обзора

Оценка затрат

При расчёте вычислительных затрат приблизительно оценим требуемое количество операций комплексного умножения. При этом предварительно оценим затраты на алгоритм поиска участков пропадания сигнала. В рамках данного алгоритма комплексное умножение присутствует в двух операциях: расчёт порога и суммирование квадратов действительной и мнимой частей в цикле по каналам дальности. Следовательно, затраты на весь алгоритм можно приблизительно оценить как $C_{fg} = (N_{rg} * T_i) + 1$, где в скобках приведено умножение количества каналов дальности на длину интервала накопления сигнала в каждом канале дальности в отсчётах.

Основные операции, требующие наибольшего количества вычислений комплексного умножения – это поиск коэффициентов методом Берга и две функции фильтрации в рамках вычисления $arfill$. Сложность стандартного алгоритма Берга можно приблизительно оценить, как $3Nm - m^2$, где N – длина обрабатываемого сигнала (глубина предсказания), m – порядок АР модели. Это будет показано ниже. Как известно КИХ-фильтр с числом звеньев X требует X циклов (операций) умножения с накоплением. Число звеньев фильтра в нашем случае соответствует длине сигнала N .



Рассчитаем количество операций комплексного умножения, затрачиваемое на расчёт коэффициентов авторегрессии. Здесь примем, что операции расчёта коэффициента отражения, обновление ошибок предсказания и коэффициентов авторегрессии повторяются на каждой из m итераций. При этом в расчёте коэффициента отражения с учётом распараллеливания команд на каждой итерации есть 2 операции комплексного умножения массивов размера $N-m$, где N – глубина предсказания.

Обновление ошибок предсказания – довольно затратная операция, но с учётом одновременного обновления прямой и обратной ошибок получается $N-m$ комплексных умножений на каждой итерации.

В рамках обновления коэффициентов авторегрессии на каждой итерации производится перемножение коэффициента отражения на массив АР-коэффициентов из m элементов, то есть с учётом комплексного сопряжения – $2m$ операций комплексного умножения.

Теперь можно легко оценить общее количество операций комплексного умножения при расчёте коэффициентов стандартным методом Берга: $2Nm - 2m^2 + Nm - m^2 + 2m^2 = 3Nm - m^2$, что соответствует оценке затрат данного метода, приведённой в [9].

Таким образом, опустив расчёт весов и вычитание среднего, а также другие не содержащие комплексных умножений операции, получим для каждой функции *arfill* сложность $C = 2N + 3Nm - m^2$. Учитывая, что функция вызывается два раза для прямого и обратного прогнозирования, а также добавив затраты на усреднение $C = 2$, получим общее значение затрат для каждого участка восстановления:

$$C_{gap} = 2(2N + 3Nm - m^2) + 2 = 4N + 2m(3N - m) + 2. \quad (10)$$

Для того чтобы оценить затраты на обработку одного интервала накопления сигнала, воспользуемся выражением:

$$C_{T_i} = \frac{T_i}{P_{gap}} * C_{gap} = \frac{T_i}{\frac{P_n}{F_p}} * C_{gap} = \frac{1000}{50} * C_{gap} = \quad (11)$$

$$= 4 \cdot (4N + 2m(3N - m) + 2).$$

Здесь N – глубина предсказания, m – порядок авторегрессии. При учёте того, что интервал накопления составляет 200 отсчётов, при заданных N и m ($N = 25$, $m = 5$) затраты на обработку сигнала в каждом канале дальности составят: $C_{T_i} = 4 \cdot (4 \cdot 25 + 2 \cdot 5 \cdot (3 \cdot 25 - 5) + 2) = 4 \cdot (100 + 702) = +702) = 3208$ операций комплексного умножения.

Оценим возможность работы всего алгоритма в масштабе реального времени. При заданных параметрах и количестве каналов дальности $N_{rg} = 300$ получим $3208 \cdot 300 = 962400$ операций комплексного умножения на обработку одного парциального кадра. Добавим к этому значению затраты на поиск участков пропадания сигнала $C_{fg} = 962400 + (N_{rg} \cdot T_i) + 1 = 962400 + 300 \cdot 0,2 + 1 = 962461$ операций.

В каждом парциальном кадре мы имеем обработку 200 отсчётов сигнала на тактовой частоте 1000 Гц,

то есть процессор должен успеть выполнить все операции за $200/1000$ Гц = 0,2 с. Процессор 1967ВН028 производства АО «ПКК «МИЛАНДР» работает на частоте 450 МГц и способен выполнить за один такт 2 операции комплексного умножения [10]. Следовательно, такой сигнальный процессор позволяет выполнить все операции разработанного алгоритма за $(962461/2)/450$ МГц = 0,0011 секунд, что примерно в 200 раз меньше чем 0,2 секунды, а значит соблюдены условия работы в реальном масштабе времени.

Заключение

Таким образом, были рассмотрены разные способы решения задачи компенсации негативного влияния лопастей винта вертолёта. Была также проведена оценка точности восстановления сигнала при помощи методов интерполяции и авторегрессии. На основе данных критериев была построена сравнительная таблица методов восстановления и графики зависимостей ошибки восстановления от параметров РЛС. Проведена оценка вычислительных затрат на выполнение предложенного алгоритма. Результаты расчётов показали, что разработанный алгоритм может быть выполнен в реальном масштабе времени. В будущих работах планируется уделить внимание таким вопросам как оптимальный выбор параметров авторегрессии и их зависимость от типа исследуемого сигнала.

Литература

1. Дворянков Д.А., Андросов В.В., Витязев С.В., Витязев В.В. Формирование изображений системой радиовидения в условиях периодических затенений. Труды Международной конференции по компьютерной графике и зрению «Графикон». 2022. С. 774-783.
2. Беклемишев Д.В. Дополнительные главы линейной алгебры. М.: Наука, 1983. 336 с.
3. Ильин М.Е. Аппроксимация и интерполяция. Методы и приложения: Учеб. пособие. Рязань: РГРТА, 2003. 56 с.
4. Box G., Jenkins G. Time series analysis: Forecasting and control. San Francisco: Holden-Day, 1970.
5. S. Orfanidis, Optimum Signal Processing, 2nd Ed. Macmillan, 1988, Chapter 5.
6. J.P. Burg. Maximum entropy spectral analysis. Proc. 37th Meet.Soc. Explorational Geophys., Oklahoma City, OK, 1975.
7. Козионов А.П., Пяйт А.Л., Мохов И.И., Иванов Ю.П. Исследование алгоритмов восстановления пропусков в измеряемых сигналах для системы мониторинга состояния дамб. Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление, 2015, № 2-3(217-222), 93-104 с.
8. V.V. Voronin, V.I. Marchuk, S.P. Petrosov, I. Svirin, S. Agaian and K. Egiazarian. Image restoration using 2D autoregressive texture model and structure curve construction. Proc. SPIE 9497 Mobile Multimedia/Image Processing Security and Applications, 2015.
9. Koen Vos. A Fast Implementation of Burg's Method, 2013.
10. Спецификация на микросхему 1967ВН028. АО «ПКК Миландр», 2021. <https://ic.milandr.ru/upload/iblock/556/g5zyamzsnf6h8c1jmare6avuuds0ov2h/1967%D0%92%D0%9D028.pdf>.