ЦИФРОВАЯ ОБРАБОТКА СИГНАЛОВ

Научно-технический журнал
№ 1/2025

Издается с 1999 года
Выходит четыре раза в год

ГЛАВНЫЙ РЕДАКТОР
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. Бартенев, д.т.н. Ю.И. Борисов, к.т.н. П.В. Бабаян, д.т.н., проф. Ю.А. Брюханов, к.т.н. С.В. Витязев, д.т.н., проф. К.Ю. Гаврилов, 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. Сергиенко А.В. д.т.н., проф. Чиров Д.С., к.т.н., проф. В.С. Сперанский, д.т.н., проф. В.Ф. Фёдоров.

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.dspa.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 09.12.24 г.
Формат 60x90/8.
Гарнитура «Arial». Печать офсетная.
Бумага офсетная. Печ.л. 6,5. Тираж 200 экз.

Заказ № 1164. 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:

Костров В.В., Храмов К.К.
Характеристики обнаружения двухканальной системы СДЦ с остановленным фазовым центром 3

Хоанг Ван Туан, Дворкович А.В.
Исследование эффективности цифрового предыскажения на основе полиномиальной модели памяти и алгоритма Левенберга-Марквардта для усилителей мощности 10

Купряшкин И.Ф.
Сравнение точности классификации радиолокационных изображений объектов сверточной нейронной сетью и сетью с архитектурой vision transformer 17

Тимошенко А.В., Захаров А.С., Байкалов И.А., Урсегов А.К., Савчук А.М.
Оценка искажений диаграммы направленности радиолокационной станции при нестационарном воздействии тепловых процессов на малошумящий усилитель приемного тракта..... 22

Рязанцев Л.Б., Бабич О.А., Маклаков Д.Ю.
Моделирование дальностно-доплеровских отражений от земной поверхности при радиолокационном зондировании с борта летательного аппарата 27

Рогожников Е.В., Савенко К.В., Дмитриев К.Э., Иванников Н.А., Порваткин Р.А., Гридин И.В., Жапова Д.Ж.
Обнаружение prach-сигнала стандарта GSM на фоне помех 34

Алимагадов К.А., Умняшкин С.В.
Нейросетевая винеровская фильтрация изображений в области дискретного вейвлет-преобразования 39

Егоров В.В., Клионский Д.М.
Декомпозиция на эмпирические моды: тематический обзор (часть 1) 47

Ушенкин В.А.
Метод высокоточной обработки топографических интерферометрических изображений с привлечением низкодетальной стереофотограмметрической информации от оптико-электронных систем ДЗЗ 59

Попов Д.И.
Анализ режекторных фильтров рекурсивного типа..... 67

Полушин П.А., Корнеева Н.Н., Архипов Н.А., Крамаренко Н.И.
Использование асимметричных «созвездий» при фазовой модуляции цифровых сигналов..... 72

Ушенина И.В., Данилов Е.А.
Реализация на плис модуля искусственного нейрона для последовательно-параллельных архитектур нейронных сетей с прямой связью..... 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 – Russion Science Citation Index (RSCI)

Digital Signal Processing

Science & Technical Magazine
Issue 1, 2025 year

Is published quarterly since 1999

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.)
Babayan P.V.
Bruchanov Y.A., Dr.Sci.(Tech.), Professor
Vityazev S.V.
Gavrilov K.U. 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
Sergienko A.V. Dr.Sci.(Tech.), Professor
Chirov D.S., Dr.Sci.(Tech.), Professor
Speranskii V.S., Dr.Sci.(Tech.), Professor
Fedorov V.F. Dr.Sci.(Tech.), Professor

Editorial office address:

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

Publisher:

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

Computer makeup:

I.A. Blagodarova

FOUNDERS:
InSys
Radio Research and Development Institute

CONTENTS:

Kostrov V.V., Khramov K.K.
Detection characteristics of a two-channel displaced phase center antenna MTI system 3

Hoang Van Tuan, Dvorkovich A.V.
Study of the Efficiency of Digital Predistortion Based on the Polynomial Memory Model and the Levenberg-Marquardt Algorithm for Power Amplifiers 10

Kupryashkin I.F.
Convolutional neural network and a network with vision transformer architecture accuracy comparison in radar images classification task 17

Timoshenko A.V., Zakharov A.S., Baikalov I.A., Ursegov A.K., Savchuk A.M.
Estimation of radiation pattern distortions of a radar station under the nonstationary effect of thermal processes on a low-noise amplifier of the receiving path 22

Ryazantsev L.B., Babich O.A., Maklakov D.Ju.
Modeling of range-doppler reflections from the earth's surface during radar sensing from an aircraft 27

Rogozhnikov E.V., Savenko K.V., Dmitriyev E., Ivannikov N.A., Porvatkin R.A., Gridin I.V., Zhapova D.Z.
Detection of gsm standard signal in the background of interference 34

Alimagadov K.A., Umnyashkin S.V.
Neural network Wiener filtering of images in the domain of discrete wavelet transform 39

Egorov V.V., Klionskiy D.M.
Empirical mode decomposition: thematic review (part 1) 47

Ushenkin V.A.
Method for high-precision topographic interferogram processing with use of low-detail stereophotogrammetric information obtained from optical remote sensing systems 59

Popov D.I.
Analysis of recursive rejection filters 67

Polushin P.A., Korneeva N.N., Arhipov N.A., Kramarenko N.I.
Using of asymmetrical «eyes diagram» in digital signals phase modulation 72

Ushenina I.V., Danilov E.A.
FPGA implementation of the artifical neuron module for seriesparallel architectures of feedforward neural networks 78



УДК 621.396.96

ХАРАКТЕРИСТИКИ ОБНАРУЖЕНИЯ ДВУХКАНАЛЬНОЙ СИСТЕМЫ СДЦ С ОСТАНОВЛЕННЫМ ФАЗОВЫМ ЦЕНТРОМ

Костров В.В., *д.т.н., профессор кафедры радиотехники Муромского института (филиала) ВлГУ, e-mail: kostrov.46@mail.ru*
Храмов К.К., *к.т.н., доцент кафедры радиотехники Муромского института (филиала) ВлГУ, e-mail: hramovkk@gmail.com*

DETECTION CHARACTERISTICS OF A TWO-CHANNEL DISPLACED PHASE CENTER ANTENNA MTI SYSTEM

Kostrov V.V., Khramov K.K.

One of the effective methods of moving target indication (MTI) in space-borne synthetic aperture radars (SARs) is the displaced phase center antenna technique. The paper describes the MTI algorithm based on a digital two-channel interferometer. The results of processing the trajectory signals for the X-band space-borne SAR are obtained. Graphs of the change in the amplitude and position of the response of a moving target on a radar image are given, as well as graphs of the phase difference of the input signals of the MTI system. Statistical modeling of a two-channel displaced phase center MTI system is performed. Detection characteristics are obtained that allows us to evaluate the potential capabilities of the MTI system.

Key words: synthetic aperture radar, SAR, moving target indication, MTI, displaced phase center antenna, digital two-channel interferometer, radial velocity, detection characteristics.

Ключевые слова радиолокатор с синтезированием апертуры, РСА, селекция движущихся целей, СДЦ, алгоритм с остановленным фазовым центром, цифровой двухканальный интерферометр, радиальная скорость, характеристики обнаружения.

Одним из эффективных методов селекции движущихся целей (СДЦ) в радиолокаторах с синтезированием апертуры антенны (РСА) космического базирования является метод, использующий алгоритм с остановленным фазовым центром. В работе описан алгоритм СДЦ на базе цифрового двухканального интерферометра. Выполнено моделирование алгоритма, получены результаты обработки траекторных сигналов для РСА космического базирования Х-диапазона, приведены зависимости изменения амплитуды и положения отклика движущейся цели на РЛИ, разности фаз входных сигналов системы СДЦ. Выполнено статистическое моделирование двухканальной системы СДЦ с остановленным фазовым центром, получены характеристики обнаружения, которые позволяют оценить потенциальные возможности такой системы СДЦ.

Введение

В радиолокаторах с синтезированием апертуры антенны (РСА) космического базирования одними из эффективных систем селекции движущихся целей (СДЦ) являются системы, функционирующие с использованием алгоритмов СДЦ с остановленным фазовым центром [1-5]. В таких системах используется интерферометр с антеннами или фазовыми центрами, которые разнесены вдоль линии пути носителя РСА. Рисунок 1 иллюстрирует геометрию обзора в РСА с двумя приемными антеннами (двухканальный интерферометр). Антенна 0 подключена к передатчику и формирует диаграмму направленности на передачу. Антенны 1 и 2 разнесены вдоль линии пути на расстояния $\pm d$ относительно точки 0 и подключены, соответственно, к первому и второму приемными каналам. В этом случае положение фазовых центров может быть аппроксимировано средними точками между передающей и приемными антеннами.

Платформа РСА движется с постоянной скоростью V_0 вдоль оси *x* на высоте H над поверхностью Земли. Визирование участка поверхности осуществляется под углом β. Движущаяся цель *O* является точечным объектом, находится в плоскости визирования и имеет угловое положение $\theta_ц$ относительно вектора V_0. Положим, что полный вектор скорости $V_ц$ движущейся цели содержит только постоянную радиальную составляющую V_r. Точка (x_0, y_0) задает положение движущейся цели при $t = 0$, где t – азимутальное (медленное) время, отнесенное к моменту, когда цель находится в центре луча.

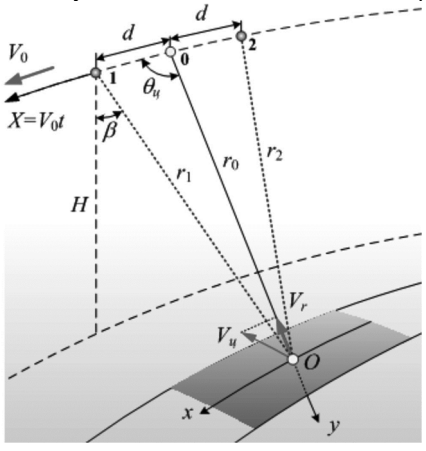


Рисунок 1 – Геометрия обзора в РСА с остановленным фазовым центром

Для подавления сигналов от неподвижных объектов и фона в методе СДЦ с остановленным фазовым цент-

ром необходимы: компенсация радиального движения носителя РСА, задержка сигнала первого по ходу движения приемного канала (рисунок 1) для обеспечения синфазности сигналов, отраженных от неподвижных объектов, и вычитание выходных сигналов приемных каналов [2, 3]. Дальнейшая обработка сводится к когерентному накоплению сигнала цели на фоне шума приемника. При нормальном боковом обзоре такой алгоритм СДЦ может быть записан в виде

$$S \sim \left| \int_0^{T_{ca}} \dot{s}_1(t-\Delta t)\overset{*}{\dot{h}}(t)\mathrm{e}^{-j2\pi f_\phi t}dt - \int_0^{T_{ca}} \dot{s}_2(t)\overset{*}{\dot{h}}(t)\mathrm{e}^{-j2\pi f_\phi t}dt \right| = \left| \int_0^{T_{ca}} \left[\dot{s}_1(t-\Delta t)-\dot{s}_2(t)\right]\overset{*}{\dot{h}}(t)\mathrm{e}^{-j2\pi f_\phi t}dt \right|, \quad (1)$$

где $s_1(t)$ и $s_2(t)$ – траекторные сигналы в приемных каналах; Δt – задержка сигнала первого по ходу движения платформы РСА приемного канала; $\dot{h}(t) = \exp\left(j\frac{4\pi}{\lambda}\frac{V_0^2 t^2}{2R_0} \right)$ – опорная функция, компенсирующая квадратичный фазовый набег траекторных сигналов; f_ϕ – частота пространственного канала, на котором происходит когерентное накопление сигналов; T_{ca} – время синтезирования апертуры.

Разность в квадратных скобках выражения (1) определяет работу компенсатора, на выходе которого сигнал от движущейся цели при обеспечении условия совмещения фазовых центров $V_0\Delta t = d$ примет вид [2]

$$\dot{s}_к(t) = j2\dot{A}_ц G(\theta_ц)\sin\left(\frac{2\pi}{\lambda}V_r\Delta t\right)\times \quad (2)$$

$$\times\exp\left[j\frac{2\pi}{\lambda}\left(2R_0 - 2V_{0r}t - 2V_0\left(t-\frac{\Delta t}{2}\right)\cos\theta_ц + 2V_r\left(t-\frac{\Delta t}{2}\right)\right)\right],$$

где $A_ц$ – случайная амплитуда отраженного от цели сигнала, которую можно считать постоянной за время T_{ca}; V_{0r} и V_0 – радиальная и тангенциальная составляющие скорости движения носителя РСА относительно направления на опорную точку кадра; $G(\theta_ц)$ – передаточная характеристика диаграммы направленности (ДН) антенны в направлении $\theta_ц$; R_0 – дальность элемента разрешения, в котором находится цель.

Из выражения (2) следует, что амплитуда $A_{СДЦ}$ разностного сигнала является функцией V_r

$$\left|A_{СДЦ}(V_r)\right| = 2A_ц\sin\left(\frac{2\pi d}{\lambda}\frac{V_r}{V_0}\right), \quad (3)$$

а рассматриваемый метод СДЦ эффективен для радиальных скоростей

$$V_r < \frac{\lambda}{4\Delta t} = \frac{\lambda V_0}{4d}.$$

Разность фаз выходных сигналов $\dot{s}_1(t-\Delta t)$, $\dot{s}_2(t)$ приемных каналов

$$\Delta\varphi(V_r) = \varphi_1 - \varphi_2 = \frac{4\pi d}{\lambda}\frac{V_r}{V_0} = \frac{4\pi}{\lambda}V_r\Delta t \quad (4)$$

определяется смещением движущейся цели в радиальном направлении за время Δt [2, 6].

Соотношения (1)-(4) составляют основу алгоритма СДЦ с остановленным фазовым центром, достаточно хорошо изучены и представлены в известных источниках, в частности [1-7]. Вместе с тем, недостаточное внимание уделено исследованию характеристик обнаружения и ряда других параметров таких систем при их использовании в радиолокаторах с синтезированием апертуры космического базирования.

Целью работы является получение и анализ характеристик обнаружения системы СДЦ на базе цифрового двухканального интерферометра в космических комплексах с РСА.

Алгоритм СДЦ

Модель формирования радиолокационного изображения (РЛИ), реализующая алгоритм (1), представлена на рисунке 2. Она содержит два канала приема, соответствующих антенной системе с разнесенными фазовыми центрами, в которых выполняется синтез комплексных РЛИ (КРЛИ), формируются векторы $\dot{S}_1$ и $\dot{S}_2$ комплексных отсчетов изображений, и осуществляется задержка сигнала первого приемного канала [2-4]. Сформированные сигналы подаются в суммарный и разностный каналы. Суммарный сигнал используется для синтеза КРЛИ фона и целей, разностный – для селекции движущихся целей (ДЦ) с последующим вычислением их радиальных скоростей и поправок для компенсации смещения отметок на РЛИ по азимутальной координате.

Задержка траекторного сигнала может быть осуществлена в частотной области. В этом случае алгоритм (1) запишется в виде [2]

$$S \sim \left|\dot{S}_2(f_\phi) - \mathrm{e}^{j2\pi f_\phi \Delta t}\dot{S}_1(f_\phi)\right|.$$

Пусть с помощью схемы (рисунок 2) выполняется обработка голограмм, сформированных в приемных каналах траекторными сигналами трех точечных целей – неподвижной и двух движущихся с постоянными радиальными скоростями. Угловое положение этих объектов зададим равным $\theta_{ц1} = \theta_{ц2} = \theta_{ц3} = 90°$. Моделирование выполнено для нормального бокового обзора с параметрами РСА космического базирования, приведенными в таблице 1.

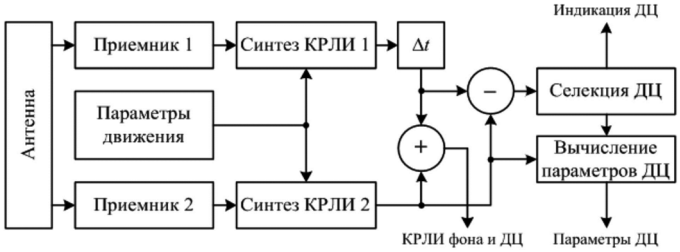


Рисунок 2 – Схема алгоритма, реализующая метод СДЦ с остановленным фазовым центром

Таблица 1 – Параметры РСА для моделирования

Параметр	Значение
Высота орбиты космического аппарата, H, км	510
Угломестное положение максимума ДН антенны, β_0, градусы	45
Ширина луча ДН антенны, $\theta_{x-3дБ} = \theta_{y-3дБ}$, градусы	0,5
Зондирующий сигнал	ЛЧМ
Длина волны зондирующего сигнала, λ, см	3,1
Ширина полосы частот зондирующего сигнала, ΔF, МГц	110
Длительность зондирующих импульсов, τ, мкс	20
Режим работы	Маршрутный

Результаты моделирования при $V_{r1} = -3$ м/с (соответствует отдаляющейся цели), $V_{r2} = 0$ и $V_{r3} = 6$ м/с (приближающаяся цель) приведены на рисунке 3, из которого видно, что пространственное положение неподвижного объекта *2* на РЛИ в каналах приема не изменилось, а отклики от движущихся целей *1* и *3* имеют азимутальное смещение, соответственно, $\Delta x_1 = -165$ и $\Delta x_3 = 330$ отсчетов (рисунок 3,*б*). Такое смещение обусловлено доплеровским сдвигом частоты

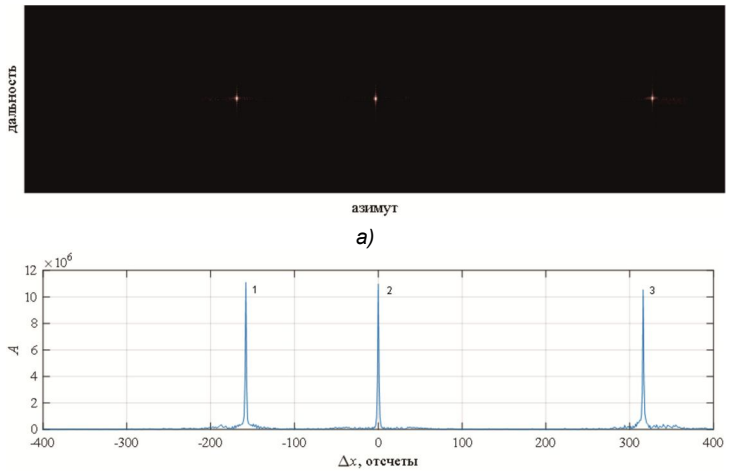
$$f_\partial = \frac{2V_0}{\lambda}\cos\theta_ц + \frac{2V_r}{\lambda},$$

зависящим от величины и знака радиальной составляющей скорости объекта, и при $\theta_ц = 90°$ определяется выражением [1-5]

$$\Delta x(V_r) = 2V_r T_{ca}/\lambda. \quad (5)$$

На РЛИ разностного канала (рисунок 4,*а*) видно, что отклик неподвижного объекта оказался скомпенсирован, а отклики движущихся объектов присутствуют. При этом их амплитуда (рисунок 4,*б*) зависит от радиальной скорости, что соответствует выражению (3). Таким образом, яркостное РЛИ на выходе системы СДЦ позволяет выполнить пороговую селекцию движущейся цели.

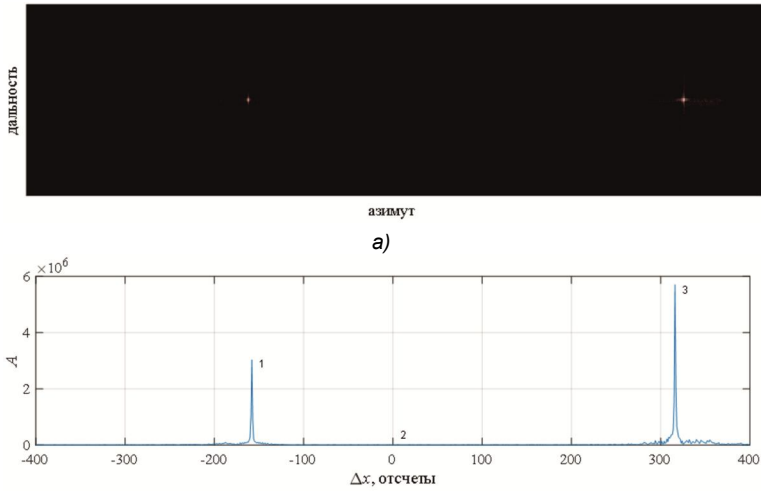
Поскольку доплеровская частота f_∂ сигналов движущихся целей на выходе вычитающего устройства такая же, как на его входе, на выходе системы СДЦ сохранилась неопределённость «угловое положение – радиальная скорость» движущегося объекта. Поэтому отклики целей, имеющих радиальную скорость, на РЛИ сохранили смещение относительно своего исходного азимутального положения на угол $\theta_x = V_r/V_0$, пропорциональный значению V_r.



а)

б)

Рисунок 3 – Фрагмент РЛИ (а) и его сечение по дальности (б) на выходе блока синтеза КРЛИ при моделировании трех точечных целей: неподвижной (2) и движущихся с радиальными скоростями V_{r1} = –3 м/с (1) и V_{r3} = 6 м/с (3)



а)

б)

Рисунок 4 – Фрагмент РЛИ (а) и его сечение по дальности (б) на выходе блока СДЦ при моделировании трех точечных целей: неподвижной (2) и движущихся с радиальными скоростями V_{r1} = –3 м/с (1) и V_{r3} = 6 м/с (3)

Выполним анализ характеристик и параметров двухканальной системы СДЦ с остановленным фазовым центром для принятых параметров РСА.

На рисунке 5,*а* приведена рассчитанная по выражению (5) зависимость азимутального смещения отклика $\Delta x = x_i - x_0$, выраженная в отсчетах относительно исходного положения x_0 цели на РЛИ, от величины радиальной скорости движения V_r (линия 1). Крутизна этой зависимости, имеющей линейный характер, пропорциональна разрешению РСА по азимуту [5-10]. На этом же рисунке приведены результаты моделирования значений $\Delta x(V_r)$ (линия 2). Относительное отклонение полученных значений от расчетных не превышает 1 %.

В РСА высокого разрешения движущаяся цель за время синтезирования перемещается по нескольким элементам разрешения по дальности. На рисунке 5,*б* линия 1 соответствует расчетному значению изменения положения отклика движущейся цели по дальности в зависимости от ее радиальной скорости:

$$\Delta y(V_r) = 2V_r T_{ca} \Delta F / c,$$

где c – скорость света; Δy – смещение в отсчетах.

Однако необходимость коррекции миграции дальности (она реализуется блоками синтеза КРЛИ на рисунке 2) в космических РСА высокого разрешения приводит к дополнительному смещению основной энергии отклика движущейся цели по дальности относительно расчетного значения Δy. Сказанное иллюстрирует кривая 2 (рисунок 5,*б*), полученная при моделировании. Изменение угломестного положения радиально движущейся цели на РЛИ вызывает размытие отклика и соответствующее снижение амплитуды.

Нормированная амплитуда разностного сигнала приемных каналов, полученная с помощью равенства (3), приведена на рисунке 6,*а* (пунктирная линия 1). Предельные однозначно измеряемые значения радиальной скорости $V_{r\max}$ соответствуют величине $\left|A_{СДЦ}/2A_ц\right| = 1$. Для рассматриваемых параметров РСА имеем $V_{r\max} = \pm 33{,}2$ м/с. Как следует из (3) эффективность системы СДЦ возрастает при увеличении разноса d фазовых центров антенны и уменьшении радиальной скорости носителя. Из полученных результатов моделирования видно, что в РСА высокого разрешения амплитуда отклика движущейся цели на выходе СДЦ (кривая 2 на рисунке 6,*а*) меньше расчетной величины, особенно при значениях V_r, близких к максимальным. Причина, как указывалось выше, заключается в расфокусировке отклика из-за его смещения по каналам дальности. Флуктуации амплитуды при изменении V_r обусловлены дискретизацией и «растеканием» отклика по соседним каналам дальности и азимута, если положение эхо-сигнала цели не соответствует в точности моменту дискретизации.

На рисунке 6,*б* представлены два графика разности фаз $\Delta\varphi$ выходных сигналов приемных каналов в зависимости от V_r: рассчитанный по формуле (4) (линия 1) и полученный при моделировании (линия 2). При этом относительное отклонение линии 2 от расчетных значений составляет менее 0,2 %. Как следует из рисунка, зависимость $\Delta\varphi(V_r)$ имеет линейный характер.

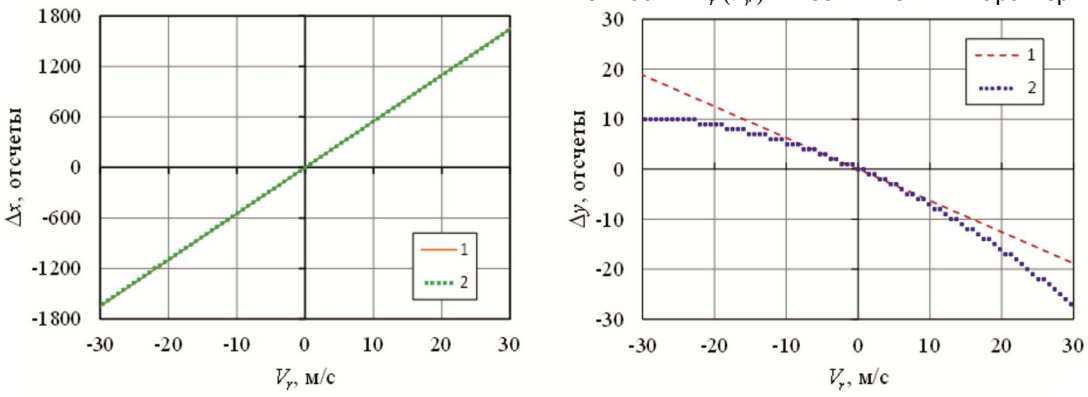


Рисунок 5 – Изменение азимутального положения (а) и дальности (б) отклика движущейся цели в зависимости от ее радиальной скорости



Рисунок 6 – Нормированная амплитуда отклика движущейся цели (а) и разность фаз выходных сигналов приемных каналов системы СДЦ (б) в зависимости от радиальной скорости

Линейность функций $\Delta x(V_r)$ и $\Delta\varphi(V_r)$ позволяет использовать их для оценки радиальной скорости цели [6, 9]. Однако сдвиг функции $\Delta x(V_r)$ при изменении азимута

движущегося объекта может привести к неоднозначности оценки V_r из-за неопределённости «азимут – радиальная скорость». Вместе с тем, разность фаз $\Delta\varphi(V_r)$ выходных сигналов приемных каналов инвариантна к азимутальному положению движущегося объекта. При этом даже если отклик цели на РЛИ окажется размытым из-за ее движения, величина $\Delta\varphi$ сохранит свое значение при выполнении условия $d << r_0$ [8].

Статистическое моделирование

Для анализа статистических свойств и оценки вероятностных характеристик обнаружения рассматриваемой системы СДЦ (рисунок 2) выполним ее моделирование при наличии на входах каналов синтеза КРЛИ сигналов $\dot{s}_1$, $\dot{s}_2$ в виде аддитивной смеси эхо-сигналов и случайной составляющей $\dot{\xi}_1$, $\dot{\xi}_2$ – шума с нормальным распределением и среднеквадратическим отклонением σ_0:

$$\dot{s}_{1,2}(m,n) = A_{ц}(m,n)\exp\left[j\varphi_{1,2}(m,n)\right] + \dot{\xi}_{1,2}(m,n),$$

где $m = 1...M$; $n = 1...N$; M, N – количество отсчетов сигнала, соответственно, по дальности и азимуту; $A_{ц}$ – амплитуда, неизменная на интервале синтезирования; $\varphi_{1,2}$ – фазы входных сигналов. Шумы приемных каналов $\dot{\xi}_1$, $\dot{\xi}_2$ являются независимыми. В приведенном выше выражении не учтен уровень фона от подстилающей поверхности, который добавлялся к сигналу в процессе экспериментов.

Отношение сигнал-шум на входе системы СДЦ определим как

$$q = \frac{A_{ц}^2}{2\sigma_0^2} M\,N. \qquad (6)$$

Оценка радиальной скорости цели может быть выполнена по оценкам разности фаз $\Delta\hat{\varphi}$ входных сигналов системы СДЦ по формуле [6]

$$\hat{V}_r = \frac{\lambda}{4\pi\Delta t}\Delta\hat{\varphi}.$$

Для получения значения $\Delta\hat{\varphi}$ достаточно знать номер доплеровского канала, в котором произошло обнаружение сигнала.

При цифровой обработке радиолокационных изображений, получаемых с помощью РСА, широко используются алгоритмы обнаружения с постоянным уровнем ложных тревог (ПУЛТ) [10]. Стабилизация вероятности ложной тревоги P_f достигается адаптивным формированием порога на основании анализа пикселей, окружающих анализируемый элемент изображения. Как правило, области анализа фона, шума и сигнала представляют собой окна квадратной формы, разделенные защитной зоной (рисунок 7).

Ключевым вопросом обеспечения постоянного уровня ложных тревог является расчет адаптивного порога обнаружения, причем алгоритмы его расчета существенно зависят от функции распределения помех.

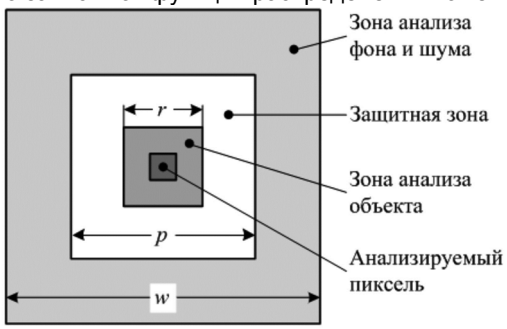


Рисунок 7 – Области объекта и фона для вычисления статистик

Обобщенная структурная схема устройства стабилизации вероятности ложной тревоги, полученная с использованием параметрического подхода, представлена на рисунке 8, где использованы обозначения: K_i – поправочные коэффициенты, зависящие от алгоритма адаптации порога; D – решение об обнаружении пикселя с целью. Такая структура реализует квазиоптимальную оценочно-корреляционно-компенсационную обработку сигналов и изображений [11].



Рисунок 8 – Обобщенная структурная схема ПУЛТ процессора

При использовании модели гауссовского шума и рэлеевских флуктуаций фона порог Неймана-Пирсона T_{NP} рассчитывается для заданной вероятности ложной тревоги P_f по соотношению [12]

$$T_{NP} = \hat{m}_1 + k_{kor} h_{norm} \hat{\sigma},$$

где $\hat{m}_1 = N_f^{-1}\sum_{ij}\rho_{ij}$ и $\hat{\sigma}^2 = N_f^{-1}\sum_{ij}\rho_{ij}^2 - \hat{m}_1^2$ – оценки математического ожидания и дисперсии, получаемые путем усреднения пикселей зоны анализа фона и шума; ρ_{ij} – значение пикселя с номером (i, j); N_f – число пикселей при усреднении; $k_{kor} = 1{,}131$ – корректирующий коэффициент, учитывающий объем выборки в окне анализа фона и шума; $h_{norm} = -1{,}25 + 1{,}061\sqrt{0{,}947 + 5{,}58\lg(P_f^{-1})}$.

Для проведения моделирования использовалось окно с параметрами $r = 3$, $p = 11$ и $w = 21$. Общее количество точек в окне анализа фона и шума составило 320.

В качестве примера обработки РСА изображения рассмотрим случай, когда РЛИ содержит точечную цель, движущуюся с постоянной скоростью на равномерном фоне. Фрагмент выходного изображения системы СДЦ представлен на рисунке 9,*а*. Размер изображения составляет 200×300 пикселей. Экспериментальное исследование ПУЛТ процессора показало, что он обеспечивает приемлемую стабилизацию P_f до значений отношения фон-шум примерно 10…15 дБ.

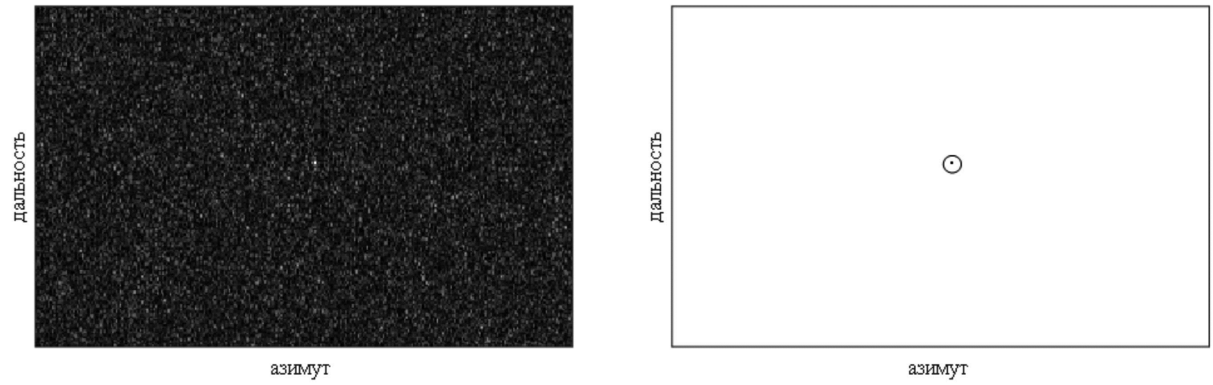
Результат обработки данного изображения в ПУЛТ

процессоре при номинальном значении $P_f = 5 \cdot 10^{-7}$ представлен на рисунке 9,*б*. Видно, что на карте обнаружений имеется автоматически обнаруженная цель (выделена кружком). Ложные объекты на карте обнаружений отсутствуют.

При уменьшении порога обнаружения появляются ложные цели, что хорошо заметно по рисунку 10, где приведены результаты обработки РЛИ при $P_f = 10^{-5}$. На карте обнаружений присутствует отклик движущейся цели и две яркостные отметки (выделены пунктирными кружками), которые представляют собой ложные цели.

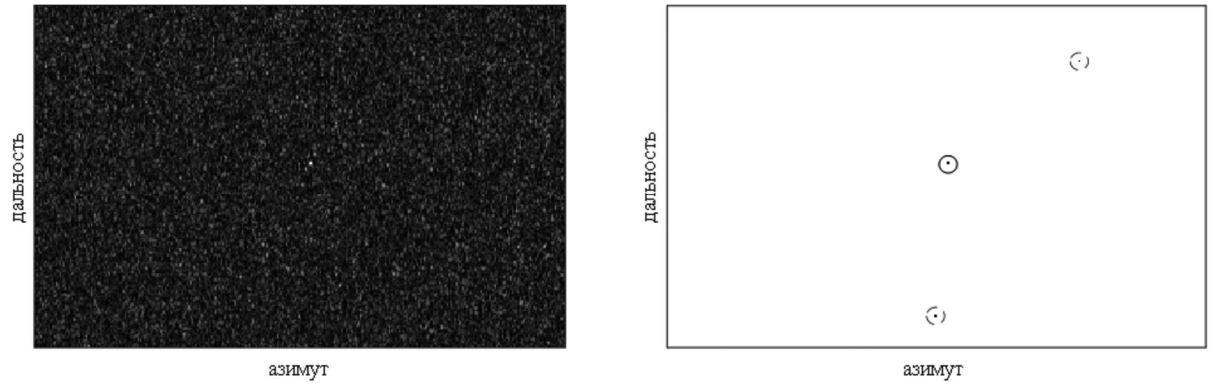
На основе анализа рисунков 9 и 10 можно заключить, что уровень шума на выходе системы СДЦ становится выше, чем уровень шума в каждом из каналов приема. Это связано с тем, что при вычитании выходных сигналов приемных каналов в блоке СДЦ дисперсия шума возрастает в два раза, т.е. $\sigma^2_{СДЦ} = 2\sigma^2_0$.

Для построения характеристик обнаружения было выполнено статистическое моделирование алгоритма при наличии движущейся точечной цели и приведенных выше расфокусировкой отклика движущейся цели при синтезе КРЛИ; 0,3…0,5 дБ составляют потери в ПУЛТ процессоре. Снижению отношения сигнал-шум также способствует удвоение дисперсии шума в блоке СДЦ. Поэтому минимальная радиальная скорость, которую способен селектировать с требуемой вероятностью P_d адаптивный обнаружитель с системой обеспечения ПУЛТ, определя-



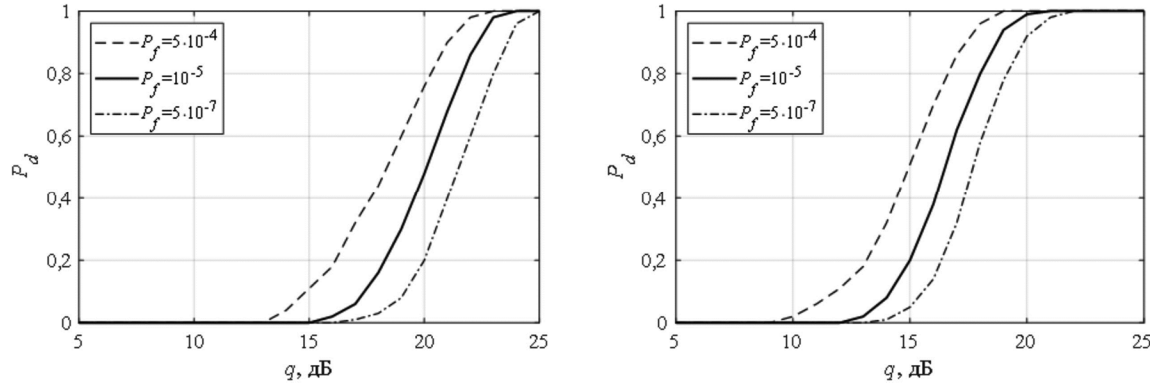
а) *б)*

Рисунок 9 – Фрагмент выходного изображения системы СДЦ (а) и карта обнаружений (б) при наличии точечной цели (V_r = 18,25 м/с, q = 17 дБ, $P_f = 5 \cdot 10^{-7}$)

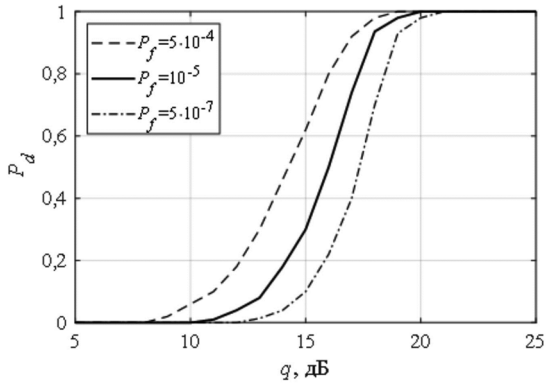


а) *б)*

Рисунок 10 – Фрагмент выходного изображения системы СДЦ (а) и карта обнаружений (б) при наличии точечной цели (V_r = 18,25 м/с, q = 17 дБ, $P_f = 10^{-5}$)



а) *б)*



в)

Рисунок 11 – Характеристики обнаружения двухканальной системы СДЦ с остановленным фазовым центром при наличии движущейся цели с V_r = 9,12 м/с (а), V_r = 18,25 м/с (б), V_r = 27,37 м/с (в)

ется максимально достижимым значением q на входе системы СДЦ. Для значений $V_r < 0,5 \cdot |V_{r\max}|$ выявлена следующая зависимость. При изменении радиальной скорости движущейся цели в k раз вероятность $P_d|_{P_f}$ сохранит свое значение, если пороговое отношение сигнал-шум изменится в 1/k раз. Так, для уверенного обнаружения объектов, имеющих радиальную скорость V_r = 5 м/с, при рассматриваемых параметрах РСА необходимо обеспечить q > 27 дБ $(P_d > 0,95;\ P_f = 10^{-5})$,

Заключение

В работе описан алгоритм, реализующий метод СДЦ с остановленным фазовым центром на базе двухканального интерферометра. Выполнено моделирование алгоритма, получены результаты обработки траекторных сигналов для РСА космического базирования X-диапазона, приведены зависимости изменения амплитуды и положения отклика движущейся цели на РЛИ, разности фаз входных сигналов системы СДЦ. Выполнено статистическое моделирование системы СДЦ с остановленным фазовым центром, построены зависимости вероятности правильного обнаружения от отношения сигнал-шум. Общей закономерностью рассматриваемой системы СДЦ является повышение достоверности оценки радиальной скорости при увеличении последней и возрастании отношения сигнал-шум. Полученные характеристики обнаружения позволяют оценить потенциальные возможности двухканальной системы СДЦ при заданных значениях сигнал-шум и предъявить требования к ее параметрам.

Работа выполнена при финансовой поддержке Российского научного фонда, проект № 23-29-10114.

Литература

1. Груздов В.В., Колковский Ю.В., Криштопов А.В., Кудря А.И. Новые технологии дистанционного зондирования Земли из космоса. М.: Техносфера, 2019. 482 с.

2. Авиационные системы радиовидения / Под ред. Г.С. Кондратенкова. М: Радиотехника, 2015. 648 с.

3. W. Xu, Y. Chen, P. Huang, W. Tan, Y. Qi, «Using the Displaced Phase Center Azimuth Multiple Beams Technique with Spaceborne Synthetic Aperture Radar Systems for Multichannel Reconstruction of Accelerated Moving Targets», Electronics. 2023; 12(24):4954. https://doi.org/10.3390/elect-ronics12244954

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

5. G. Krieger, N. Gebert and A. Moreira, «Unambiguous SAR signal reconstruction from nonuniform displaced phase center sampling», in IEEE Geoscience and Remote Sensing Letters, vol. 1, no. 4, pp. 260-264, Oct. 2004, doi: 10.1109/ LGRS.2004.832700.

6. Попов Д.И. Измерение радиальной скорости цели. Известия вузов России. Радиоэлектроника. 2015, Вып. 1. С. 53-59.

7. K. Suwa, K. Yamamoto, M. Tsuchida, S. Nakamura, T. Wakayama, and T. Hara «Image-Based Target Detection and Radial Velocity Estimation Methods for Multichannel SAR-GMTI», IEEE Transactions on Geoscience and Remote Sensing, vol. 55, no. 3, pp. 1325-1338, March 2017, doi: 10.1109/TGRS.2016.2622712.

8. Витязев В.В., Колодько Г.Н., Витязев С.В. Селекция наземных движущихся целей на основе многоско-ростной адаптивной обработки траекторного сигнала. Цифровая обработка сигналов. 2007, № 1. С. 41-50.

9. V.V. Kostrov and K.K. Khramov, «Statistical Characteristics of the Moving Target Indication in Space-borne Interferometry Synthetic Aperture Radar», 2019 Dynamics of Systems, Mechanisms and Machines (Dynamics), Omsk, Russia, 2019, pp. 1-6, doi: 10.1109/Dynamics47113.2019. 8944594.

10. El-Darymli, Khalid, et al. «Target detection in synthetic aperture radar imagery: A state-of-the-art survey», Journal of Applied Remote Sensing 7.1 (2013): 071598-071598, doi: 10.1117/1.JRS.7.071598.

11. Сосулин Ю.Г., Костров В.В., Паршин Ю.Н. Оценочно-корреляционная обработка сигналов и компенсация помех. М: Радиотехника, 2014. 632 с.

12. V. Kostrov, E. Tolstov and K. Khramov, «Automatic Detection of Thread-like Objects in SAR Using a Modified Radon Transform», 2024 26th International Conference on Digital Signal Processing and its Applications (DSPA), Moscow, Russian Federation, 2024, pp. 1-6, doi: 10.1109/DSPA 60853.2024.10510127.

13. Лихарев В.А. Цифровые методы и устройства в радиолокации. М.: Советское радио, 1973. 456 с.

УДК 621.396, УДК 621.391

ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ ЦИФРОВОГО ПРЕДЫСКАЖЕНИЯ НА ОСНОВЕ ПОЛИНОМИАЛЬНОЙ МОДЕЛИ ПАМЯТИ И АЛГОРИТМА ЛЕВЕНБЕРГА-МАРКВАРДТА ДЛЯ УСИЛИТЕЛЕЙ МОЩНОСТИ

Хоанг Ван Туан, *аспирант кафедры мультимедийных технологий и телекоммуникаций МФТИ, e-mail: khoang.vt@phystech.edu*
Дворкович А.В., *д.т.н., чл.-корр. РАН, зав. кафедрой мультимедийных технологий и телекоммуникаций МФТИ, e-mail: dvork.alex@gmail.com*

STUDY OF THE EFFICIENCY OF DIGITAL PREDISTORTION BASED ON THE POLYNOMIAL MEMORY MODEL AND THE LEVENBERG-MARQUARDT ALGORITHM FOR POWER AMPLIFIERS

Hoang Van Tuan, Dvorkovich A.V.

This study investigates the performance of digital predistortion based on polynomial memory model and Levenberg-Marquardt algorithm for power amplifiers. OFDM technology, known for its high spectral efficiency and robustness to multipath fading, is increasingly adopted in modern radio systems such as DVB-T2 and ATSC 3.0. However, OFDM signals are highly sensitive to nonlinear distortion due to the high peak-to-average power ratio. This study addresses this problem using DPD techniques to compensate for the amplifier nonlinearity and improve its performance, as well as reduce interference in adjacent radio channels. We propose a polynomial memory model optimized using Levenberg-Marquardt and BFGS algorithms to adaptively estimate and minimize the distortion effects. The results show that our approach reduces the normalized mean square error (NMSE) and adjacent channel power ratio (ACPR), which improves signal

Key words: Digital Predistortion, Power Amplifier, Levenberg–Marquardt Algorithm, Memory Polynomial Model, OFDM.

Ключевые слова: цифровое предыскажение, усилитель мощности, алгоритм Левенберга-Марквардта, полиномиальная модель памяти, OFDM.

Рассматривается эффективность цифрового предыскажения на основе полиномиальной модели памяти и алгоритма Левенберга-Марквардта для усилителей мощности. Технология ортогонального частотного мультиплексирования OFDM, известная своей высокой спектральной эффективностью и устойчивостью к замираниям при многолучевом распространении, всё чаще применяется в современных радиосистемах, например, в DVB-T2 и ATSC 3.0. Однако сигналы OFDM крайне чувствительны к нелинейным искажениям из-за высокого отношения пикового значения мощности к средней мощности. В данном исследовании эта проблема решается с помощью методов цифрового предыскажения DPD, компенсирующих нелинейность усилителя и увеличивающих его эффективность, а также снижающих помехи в соседних радиоканалах. Мы предлагаем полиномиальную модель памяти, оптимизированную с использованием алгоритмов Левенберга-Марквардта и BFGS для адаптивной оценки и минимизации эффектов искажений. Результаты показывают, что наш подход снижает нормализованную среднеквадратическую ошибку (NMSE) и уровень мощности в соседнем канале (ACPR), что улучшает качество сигнала и характеристики усилителя.

Введение

Благодаря высокой спектральной эффективности в выделенном частотном диапазоне, способности противостоять эффектам многолучевого распространения и невосприимчивости к частотно-селективным замираниям, технология ортогонального частотного мультиплексирования (OFDM) продолжает привлекать внимание ведущих исследователей по всему миру. Большие успехи в цифровой электронике, особенно в области полупроводниковых интегральных микросхем высокой плотности, вдохновили ученых на разработку и внедрение радиосистем на основе технологии OFDM. В настоящее время OFDM используется в новейших радиосистемах, таких как системы наземного цифрового телевидения второго поколения стандартов DVB-T2 [1] и ATSC 3.0 [2].

Тем не менее, у OFDM также имеются некоторые ограничения при применении в этих системах, среди которых можно выделить несколько особо значимых. Во-первых, требование строгой синхронизации для поддержания ортогональности на приёмной стороне увеличивает сложность системы. Во-вторых, сигнал OFDM имеет высокое отношение пикового значения мощности к средней мощности (Peak-to-Average Power Ratio – PAPR), что делает его крайне чувствительным к нелинейным искажениям, вызванным мощными усилителями на передающей стороне. Это требует наличия широкодинамических усилителей на передатчике, что увеличивает энергопотребление и стоимость оборудования. В-третьих, рост потребности в передаче информации, особенно в услугах реального времени, создаёт вызов для систем обработки сигналов на базе ПЛИС (Field-Programmable Gate Array – FPGA) [3,4]. Это требует, чтобы процессоры обработки сигналов одновременно удовлетворяли два важных, но противоречивых кри-

терия: высокую плотность интеграции (большие ресурсы) и высокую скорость обработки сигналов.

Можно отметить, что из упомянутых ограничений первое и третье могут быть уменьшены благодаря прогрессу в области крупноинтегрированных микросхем. Одновременно это позволит увеличить способность обработки сигналов на высоких скоростях. Усилитель мощности (УМ) является неотъемлемой частью передающего тракта радиосистемы. УМ – это компонент, которому следует уделить особое внимание, так как он не только является одним из наиболее энергозатратных устройств, но и является основным источником нелинейных искажений в передатчике. Выбор усилителя мощности с широкой линейной динамикой, способного обеспечить высокий уровень PAPR OFDM сигнала, является дорогостоящим. При использовании УМ с ограниченной динамикой, сигнал с высоким PAPR заставляет усилитель работать в режиме насыщения (рисунок 1). Нелинейность вызывает появление гармоник высокого порядка за пределами полосы частот, что приводит к помехам в соседних радиоканалах. Она также вызывает искажения внутри полосы, что снижает эффективность передачи данных и выражается в увеличении уровня битовых ошибок (BER).



Рисунок 1 – Характеристика усилителя мощности

Снижение PAPR сигнала OFDM – одно из решений для повышения эффективности работы усилителя. Когда PAPR сигнала уменьшается, это позволяет сдвинуть рабочую точку УМ ближе к области насыщения.

Таким образом, есть несколько подходов к тому, чтобы УМ работал в линейной области: это выбор усилителя с широкой линейной динамикой, что сопряжено с высокими затратами; либо смещение рабочей точки далеко от области насыщения, что приводит к низкой эффективности; либо снижение PAPR сигнала. Тем не менее, на практике не существует полностью линейных усилителей мощности, кроме того, они обладают эффектом памяти. В результате этого характеристика усиления расширяется вокруг своей основной линии (рисунок 2). Чем шире полоса сигнала и выше частота системы, тем сильнее проявляется эффект памяти. Причины эффекта памяти могут включать ёмкость, индуктивность и тепловые постоянные активных элементов, чье поведение зависит от частоты. В итоге выходной сигнал усилителя мощности зависит не только от текущего входного сигнала, но и от предыдущих значений входного сигнала.

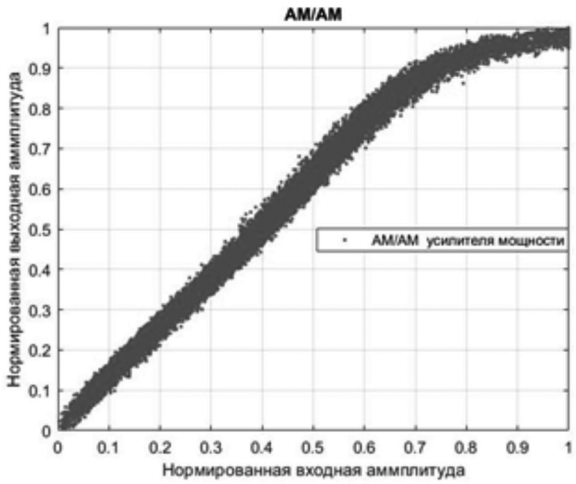


Рисунок 2 – Характеристика усилителя мощности с эффектом памяти

Компенсация нелинейных искажений усилителя мощности с помощью цифрового предыскажения (Digital Predistortion, DPD), как показано на рисунке 3, расширяет линейный диапазон работы усилителя мощности. Правильное сочетание DPD с техниками снижения PAPR позволяет усилителю мощности работать ближе к насыщению, при этом оставаясь в пределах допустимой линейности (рисунок 3). Линейный DPD может устранить или, по крайней мере, уменьшить компромисс между эффективностью и линейностью усилителя мощности. Блок DPD на основе базовой полосы создаёт дополнительную нелинейность, которая компенсирует характеристики компрессии усилителя мощности. С использованием предыскажения УМ может работать в точке, близкой к насыщению, сохраняя хорошую линейность, что значительно увеличивает его эффективность.

На практике характеристики усилителя мощности могут изменяться со временем из-за температурных изменений, старения компонентов и других факторов. Следовательно, система цифрового предыскажения должна иметь способность адаптироваться к этим изменениям.

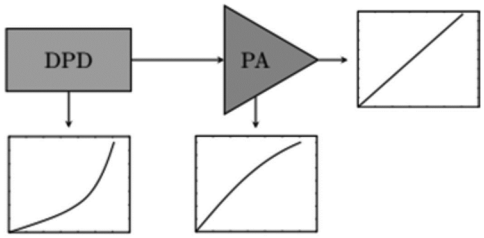


Рисунок 3 – Блок-схема линеаризации характеристики усилителя мощности на основе цифрового предыскажения

Цифровое предыскажение является одним из наиболее эффективных решений для повышения общей эффективности системы OFDM. По сути, DPD – это процесс, направленный на минимизацию искажений сигнала, вызванных усилителем мощности, путем компенсации его нелинейных характеристик, что позволяет улучшить выходную производительность. Введение прототипов DPD, работающих в режиме реального времени, которые могут обеспечить непрерывную передачу данных, имеет важное значение для быстрого внедрения систем радиосвязи нового поколения.

Работа DPD заключается в применении корректирующей функции к входному сигналу, заранее искажая его, чтобы компенсировать искажения системы на выходе. DPD может быть реализовано как в программном, так и в аппаратном обеспечении и должно адаптироваться к изменениям условий эксплуатации системы. Тем не менее, точное моделирование представляет собой сложную задачу на практике.

DPD с использованием таблиц соответствия (LUT) [5, 6, 7] давно признано эффективной техникой для улучшения линейности и эффективности системы за счет предварительной обработки входного сигнала усилителя мощности, компенсирующей нежелательные эффекты. LUT-структуры реализуются путем оценки характеристик УМ с равномерно распределенными значениями амплитуды входного сигнала и последующего хранения этих значений. Однако реализация DPD на основе LUT требует значительных затрат памяти и дорогих цифровых сигнальных процессоров. Кроме того, хранение характеристик в LUT учитывает только амплитуду, в то время как в реальности сигнал обрабатывается как комплексный, что затрудняет компенсацию фазовых искажений. Для компенсации фазовых искажений и быстрой адаптации DPD необходимо использовать несколько LUT, что приводит к значительным требованиям к памяти.

Широко применяется описание на основе рядов Вольтерры, особенно для усилителей мощности с сильным эффектом памяти. Однако из-за большого количества полиномиальных членов трудно использовать его для прямой реализации. Более простой и популярный метод в инженерных приложениях для упрощения рядов Вольтерры – это использование архитектур обобщенной полиномиальной модели памяти (Generalized memory polynomial – GMP) и полиномиальной модели памяти (Memory Polynomial – MP) [8, 9, 10, 11, 12]. Архитектура MP представляет собой альтернативный подход, который упрощает структуру GMP. MP [14, 15] является предпочтительным выбором для поведенческого моделирования УМ благодаря её низкой вычислительной сложности, удовлетворительной точности и простоте аппаратной реализации. Она широко применяется в системах, где усилители демонстрируют нелинейные эффекты памяти [8, 9, 13].

В данном исследовании основное внимание уделяется цифровому предыскажению, основанному на архитектуре косвенного обучения в сочетании с алгоритмами оптимизации параметров Левенберга-Марквардта (LM) и BFGS (алгоритм Бройдена-Флетчера-Гольдфарба-Шанно). Применение этих алгоритмов увеличивает скорость сходимости и оценивается как эффективное решение для нелинейных систем.

Модель предыскажателя

Структура модели

Существует два основных типа архитектур, широко применяемых для идентификации параметров предыскажения: архитектура косвенного обучения (Indirect Learning Architecture – ILA) [14] и архитектура прямого обучения (direct learning architecture – DLA) [15]. Несмотря на то, что DLA более устойчива к шуму на выходе УМ и обеспечивает несмещённые оценки параметров, процесс расчета более сложен, так как адаптивные алгоритмы, используемые в DLA, требуют множества итераций для нахождения оптимального набора параметров [16]. В связи с этим адаптивная ILA чаще всего применяется для оценки параметров предыскажения в УМ [4].

Рисунок 4 показывает блок-схему DPD на основе архитектуры ILA, где используется блок пост-предыскажателя для идентификации обратной модели УМ. Сигнал $x(n)$ подается на предыскажатель, который генерирует сигнал $u(n)$, являющийся входом УМ. Выходной сигнал УМ $y(n)$ нормализуется линейным коэффициентом усиления G, что приводит к выходу $z(n)$, т.е. $z(n) = y(n)/G$. Блок пост-предыскажателя принимает на вход сигнал $z(n)$ и выдает сигнал $z_p(n)$. Параметры блока корректируются путем минимизации сигнала ошибки $e(n) = x(n) - z_p(n)$ с использованием адаптивных алгоритмов. Следует отметить, что модели предыскажателя и пост-предыскажателя идентичны. Таким образом, когда коэффициенты c_{km} пост-предыскажателя (см. ниже) определены, они напрямую копируются в предыскажатель.



Рисунок 4 – Архитектура косвенного обучения (ILA)

Модель MP [14, 15] является предпочтительным выбором для поведенческого моделирования УМ и DPD благодаря её низкой вычислительной сложности, удовлетворительной точности и простоте аппаратной реализации. Она широко применяется в системах, где усилители демонстрируют нелинейные эффекты памяти. Модель MP используется как для предыскажения, так и для пост-предыскажения с параметрами K (порядок нелинейности) и M (глубина памяти). Взаимосвязь между входным и выходным сигналом модели предыскажителя математически описывается следующим образом:

$$u(n) = \sum_{k=1}^{K} \sum_{m=0}^{M} c_{km} x(n-m) \,|\, u(n-m) \,|^{k-1}, \quad (1)$$

где $c_{km} = [c_{10}, ..., c_{K0}, ..., c_{1M}, ..., c_{KM}]^T$.

Вход и выход модели пост-предыскажателя УМ можно выразить следующими уравнениями:

$$z_p(n) = \sum_{k=1}^{K} \sum_{m=0}^{M} c_{km} z(n-m) \,|\, u(n-m) \,|^{k-1}, \quad (2)$$

$$y(n) = \sum_{k=0}^{K-1} \sum_{m=0}^{M-1} a_{km} u(n-m) \,|\, u(n-m) \,|^{k}, \quad (3)$$

где a_{km} – коэффициенты MP модели усилителя мощности.

Сигналы $u(n)$ и $z_p(n)$ используются для оценки коэф-

фициентов модели c_{km} с помощью адаптивного алгоритма, который минимизирует ошибку, основанную на разнице между $u(n)$ и $z_p(n)$.

$$e(n) = x(n) - z_p(n). \quad (4)$$

Для ускорения сходимости адаптации был разработан рекурсивный алгоритм наименьших квадратов (RLS) [17] для предыскажения. Алгоритм RLS выводится путем минимизации взвешенной суммы квадратов ошибок, где недавние ошибки получают больший вес по сравнению с более старыми. Это позволяет алгоритму быстро корректировать коэффициенты предыскажения в ответ на изменения в системе, улучшая его способность компенсировать нелинейности усилителя мощности.

Алгоритм RLS часто используется в задачах, таких как DPD, для оптимизации и корректировки параметров системы с целью минимизации нелинейных искажений, вызванных усилителем мощности. Алгоритм RLS, использующий экспоненциальный фактор забывания, минимизирует целевую функцию:

$$\xi(n) = \sum_{i=1}^{n} \lambda^{n-i} e^2(i). \quad (5)$$

где $\lambda \le 1$ – фактор забывания. Чем меньше значение λ, тем быстрее информация о предыдущих данных будет забыта. Таким образом, выбор λ контролирует способность алгоритма отслеживать изменяющиеся во времени параметры. Общий алгоритм RLS описывается следующим образом:

1. Определить мгновенную ошибку $e(n) = x(n) - z_p(n)$.

2. Обновить матрицу весов

$$k(n) = \frac{P(n-1)z(n)}{\lambda + z^T(n)P(n-1)z(n)}. \quad (6)$$

3. Обновить ковариационную матрицу

$$P(n) = \frac{1}{\lambda}\left[P(n-1) - k(n)z^T(n)P(n-1)\right]. \quad (7)$$

4. Обновить коэффициенты предыскажителя

$$c(n) = c(n-1) + k(n)e(n). \quad (8)$$

Алгоритм RLS имеет ограничения, связанные с его чувствительностью к коэффициенту забывания λ. При малых значениях λ возникают значительные колебания параметров и снижение стабильности, что затрудняет точную линеаризацию УМ. Быстрое «забывание» прошлых данных также делает алгоритм уязвимым к шумам и небольшим изменениям. Эти недостатки указывают на необходимость использования более оптимизированных методов, таких как LM и BFGS, для обеспечения лучшей адаптации и стабильности системы.

Предлагаемый адаптивный алгоритм

Метод Левенберга-Марквардта (LM) является мощным методом для решения нелинейных задач наименьших квадратов и широко используется в задачах оптимизации. В приложениях цифрового предыскажения LM помогает оптимизировать коэффициенты нелинейных моделей, таких как модель памяти или модель Вольтерры, для компенсации искажений, вызванных усилителями мощности. LM комбинирует преимущества метода градиентного спуска и метода Гаусса-Ньютона, обеспечивая быструю и стабильную сходимость в нелинейных системах.

Параметр с предыскажителя можно оценить с помощью метода LM путем минимизации целевой функции:

$$\xi(c) = 0.5\sum_{i=1}^{n} e^2(i) = 0.5\sum_{i=1}^{n}[x(i) - z_p(i)]^2 = \\ = 0.5\sum_{i=1}^{n}[x(i) - f(z(i), c)]^2, \quad (9)$$

$$\hat{c} = \arg\min_c \sum_{i=1}^{n}[x(i) - f(z(i), c)]^2. \quad (10)$$

Адаптивный алгоритм предыскажения по LM описывается следующим образом:

1. Определить мгновенную ошибку $e(n) = x(n) - z_p(n)$.

2. Определить матрицу Якоби: матрица Якоби J содержит частные производные ошибки $e(n)$ по параметрам предыскажителя $c(n)$

$$J = \frac{\partial e(n)}{\partial c(n)}. \quad (11)$$

Якобиан позволяет вычислить, как изменение коэффициентов предыскажителя повлияет на ошибку.

3. Обновить коэффициенты предыскажителя: обновление коэффициентов происходит с использованием матрицы Якоби и инверсии гессиана:

$$c(n+1) = c(n) - \left(J^T J + \lambda I\right)^{-1} J^T e(n), \quad (12)$$

где, J – матрица Якоби; λ – параметр регуляции, который контролирует баланс между методом Гаусса-Ньютона и градиентным спуском; I – единичная матрица.

Алгоритм на каждом шаге: рассчитывается ошибка $e(n)$, вычисляется матрица Якоби J, на основе ошибки и матрицы Якоби обновляются коэффициенты $c(n+1)$.

Каждая итерация сопровождается пересчетом λ. Если ошибка уменьшается, значение λ уменьшается, что приближает метод к методу Гаусса-Ньютона и увеличивает шаги. Если ошибка увеличивается, λ увеличивается, и метод приближается к градиентному спуску, чтобы сделать шаги более осторожными.

Метод BFGS (Бройдена-Флетчера-Голдфарба-Шанно) – это один из наиболее популярных методов для решения нелинейных задач оптимизации. Он относится к классу методов квази-Ньютона, которые используют аппроксимацию матрицы гессиана (вторая производная функции) на основе градиента и обновления параметров на каждом шаге. Можно определить параметр с с помощью (4).

Адаптивный алгоритм предыскажения по BFGS описывается следующим образом:

Определить целевую функцию:

$$\xi(c) = 0.5\sum_{i=1}^{n}[x(i) - f(z(i), c)]^2. \quad (13)$$

Вычислить градиент целевой функции:

$$\nabla\xi(c) = \frac{\partial\xi(c)}{\partial c}. \quad (14)$$

Обновить параметры предыскажителя:

$$c(n+1) = c(n) - B(n)^{-1}\nabla\xi(c(n)) \ . \quad (15)$$

Обновить аппроксимацию гессиана:

$$B(n+1) = B(n) + \frac{\Delta g(n)\Delta g(n)^T}{\Delta g(n)^T \Delta c(n)} -$$

$$-\frac{B(n)\Delta c(n)\Delta c(n)^T B(n)}{\Delta c(n)^T B(n)\Delta c(n)},$$

$$\Delta g(n) = \nabla e(n+1) - \nabla e(n),$$

$$\Delta c(n) = c(n+1) - c(n). \quad (16)$$

Результаты моделирования

Характеристика УМ зависит как от нелинейных эффектов, так и от эффектов памяти, которые проявляются в процессе работы устройства. Для моделирования характеристик усилителя мощности в данном исследовании применяется MP с параметрами K = 7 и M = 5. Эта модель позволяет учесть сложные взаимодействия и нелинейности, а также влияние предыдущих состояний сигнала на текущий выход.

На рисунке 5 представлена характеристика УМ и модель MP. Видно, что модель MP хорошо описывает амплитудно-фазовые характеристики усилителя мощности. Она способна точно моделировать как изменения амплитуды, так и фазовые отклонения, обеспечивая высокую точность и эффективность при корректировке искажений. Благодаря этому модель MP может быть использована для разработки и оптимизации DPD систем, направленных на улучшение линейности и стабильности УМ.

Проводя моделирование усилителя мощности на основе модели MP, использующей те же значения параметров K = 7 и M = 5, как и у усилителя, мы разрабатываем DPD, противоположный характеристикам усилителя. Для оценки эффективности модели используются два показателя: NMSE (Normalised Mean Square Error) для оценки точности моделирования усилителя и ACPR (Adjacent Channel Power Ratio) [10, 11, 12] для оценки эффективности после применения DPD. В таблице 1 подробно представлены результаты модели при использовании трех оптимизационных алгоритмов: LM, BFGS и RLS. Видно, что уровень потерь энергии значительно снижается после использования DPD.

Формула для вычисления NMSE:

$$NMSE[dB] = 10\log_{10}\left(\frac{\sum_{i=1}^{N}|y_i - \hat{y}_i|^2}{\sum_{i=1}^{N}|y_i|^2}\right)\text{dB}, \quad (17)$$

где y_i – реальное значение, а $\hat{y}_i$ – прогнозируемое значение.

Формула для вычисления ACPR:

$$ACPR = 10\log_{10}\left(\frac{P_{\text{adj}}}{P_{\text{main}}}\right)\text{dB}, \quad (18)$$

где P_{adj} – мощность в соседнем канале, P_{main} – мощность в основном канале.

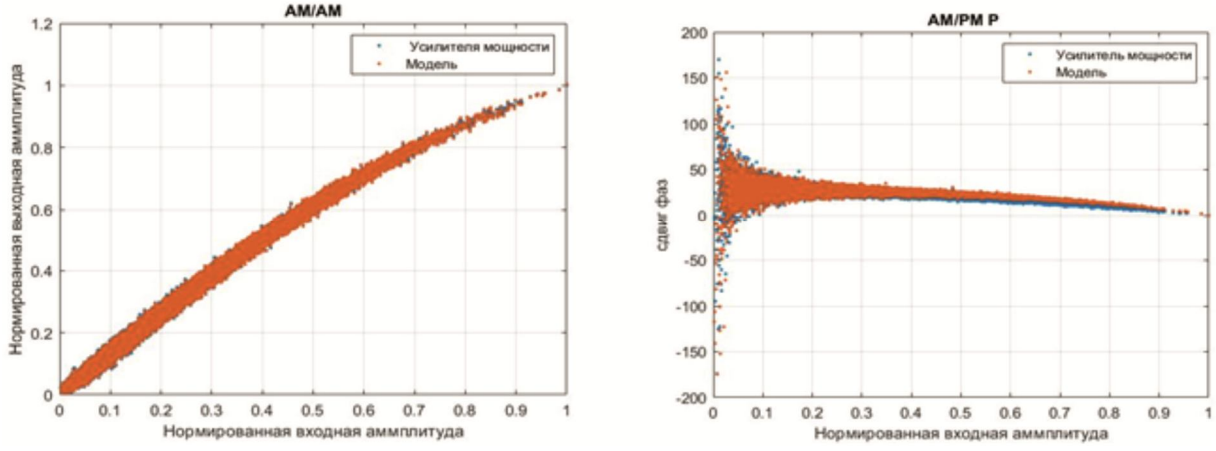
Таблица 1 – Сравнение NMSE [dB], ACPR [dB] при применении методов LM, GFGS, RLS (сигнал OFDM, ширина спектра 100 МГц, QAM 16)

Метод	NMSE [dB]	ACPR [dB]
LM	-44.66	-46.112
BFGS	-44.67	-46.114
RLS	-43.93	-45.93

Рисунки 6 и 7 демонстрируют характеристики усилительного блока при применении трех методов: LM, BFGS и RLS. Видно, что после применения алгоритма DPD удалось улучшать характеристики сигнала после прохождения через УМ.

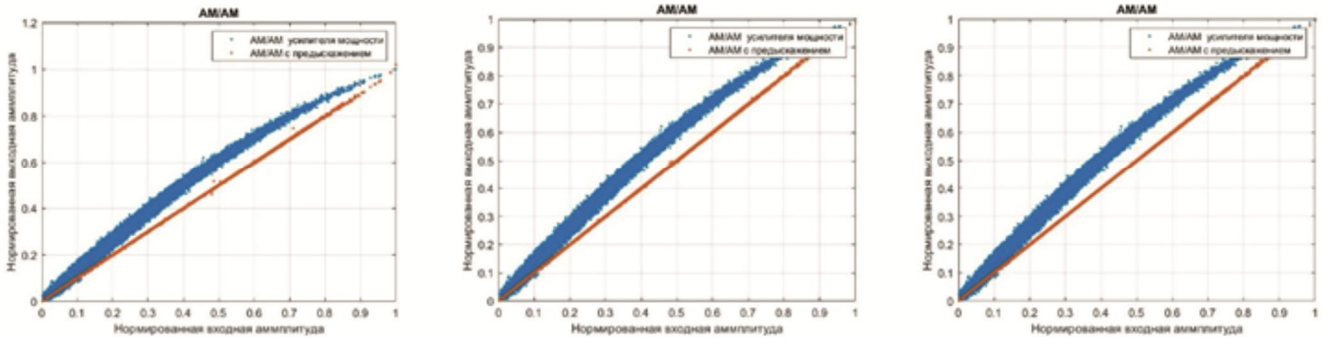
Методы LM и BFGS обладают значительными преимуществами по сравнению с методом RLS при применении к задачам корректировки входных данных в системах DPD. Во-первых, LM и BFGS являются мощными методами нелинейной оптимизации, которые позволяют эффективно решать сложные задачи благодаря сочетанию поиска по градиенту и различных улучшений. LM представляет собой гибрид градиентного спуска и метода Гаусса-Ньютона, что обеспечивает сбалансированное решение для минимизации нелинейных искажений без чрезмерного усложнения системы.

Одним из основных преимуществ методов LM и BFGS по сравнению с RLS является их способность эффективно оптимизировать в условиях сложных нелинейностей, что особенно важно при обработке данных от УМ в системах DPD. Эти методы позволяют моделировать и компенсировать искажения сигнала с высокой



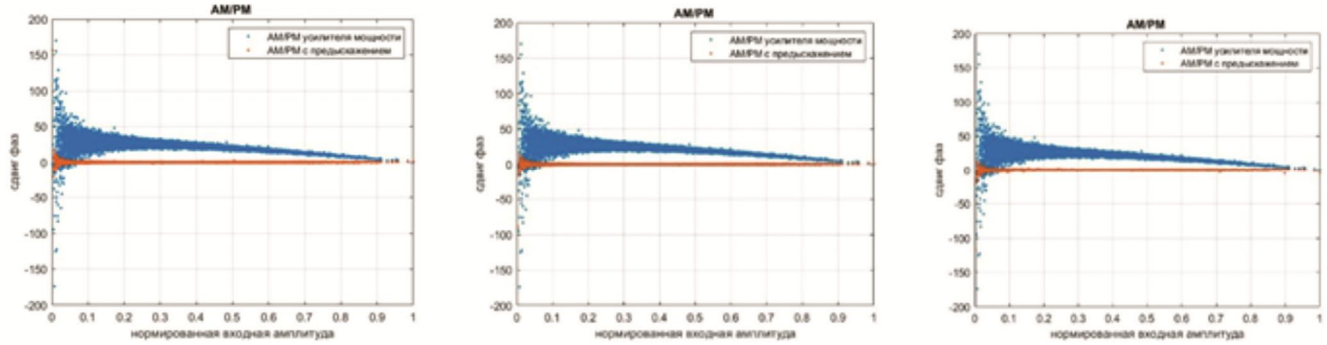
а) *б)*

Рисунок 5 – Характеристика усилителя мощности и модели – а) AM/AM; б) AM/PM



а) LM *б) BFGS* *с) RLS*

Рисунок 6 – Характеристика AM/AM усилителя мощности при применении предыскажения



а) LM *б) BFGS* *с) RLS*

Рисунок 7 – Характеристика AM/PM усилителя мощности при применении предыскажения

точностью, что способствует улучшению качества сигнала и общей производительности системы, в то время как метод RLS может сталкиваться с трудностями при обработке резких изменений или быстрых колебаний входного сигнала.

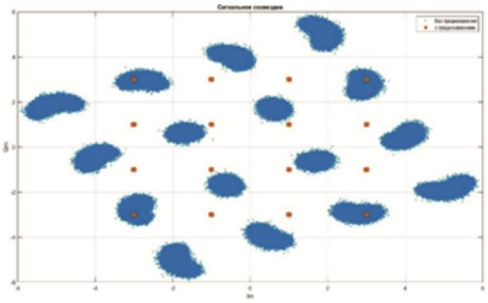


Рисунок 8 – Сигнальное созвездие при работе вблизи области насыщения с использованием и без использования DPD

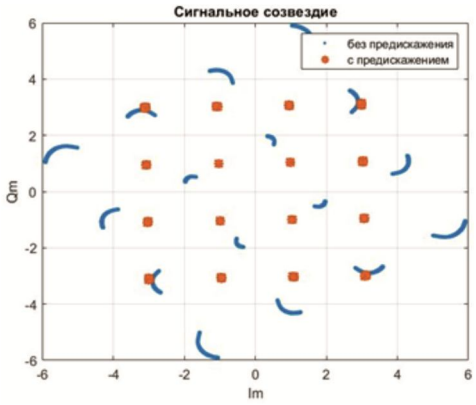


Рисунок 9 – Сигнальное созвездие при работе вдали от области насыщения с использованием и без использования DPD

Из рисунков 8 и 9 видно влияние нелинейности и эффектов памяти УМ. Они не только изменяют амплитуду сигнала, но и искажают его фазу. Применение DPD позволяет полностью устранить искажения сигнала, вызванные усилителем, и способствует более эффективной работе УМ.

Заключение

В данном исследовании рассмотрены моделирование УМ и DPD. В работе используются полиномиальная модель памяти для оценки реальных характеристик усилителя мощности. Модель УМ была интегрирована с алгоритмами DPD. DPD смоделирован как система полиномиальной памяти. Коэффициенты адаптивного DPD оцениваются с использованием алгоритмов RLS, LM и BFGS. Результаты моделирования показывают, что алгоритмы DPD достигают значительного снижения нелинейных искажений и и показывает высокую эффективность при применении алгоритмов LM и BFGS.

Литература

1. Kadhim A.A., Mohammed S.J., Al-Gayem Q. Digital Video Broadcasting T2 Lite Performance Evaluation Based on Rotated Constellation Rates. Journal of Internet Services and Information Security. 2023. Т. 13. №. 3. С. 127-137.

2. Hong Z. H. et al. In-band full-duplex communications in ATSC 3.0 single frequency network. IEEE Transactions on Broadcasting. 2023. Т. 69. №. 2. С. 560-568.

3. Behdad N. et al. PAPR Reduction in OFDM System Using Metaheuristic Algorithm. Full Length Article. 2023. Т. 4. №. 2. С. 28-8-35.

4. Abed G.A. A New Approach to Improve Transmitting and Receiving Timing in Orthogonal Frequency Division

Multiplexing (OFDM) Systems. Iraqi Journal For Computer Science and Mathematics. 2023. Т. 4. №. 2. С. 83-96.

5. H. Jiang et al. «Aggressive Look-Up Table Compression by Principal Component Analysis for Digital Pre-Distortion in Optical Communication» in Journal of Lightwave Technology, vol. 40, no. 24, pp. 7718-7726, 15 Dec. 2022.

6. J. Ren. «Digital Predistorter for Short-Wave Power Amplifier With Improving Index Accuracy of Lookup Table Based on FPGA», in IEEE Access, vol. 7, pp. 182881-182885, 2019, doi: 10.1109/ACCESS.2019.2960092.

7. L. Sun et al. «A Low Complexity LUT-Based Digital Predistortion Block With New Pruning Method», in IEEE Microwave and Wireless Components Letters, vol. 32, no. 9, pp. 1131-1134, Sept. 2022, doi: 10.1109/LMWC.2022.3163890.

8. Lv, G. Xu, C. Zhang, H. Zhang and T. Liu. «Full-Wave Digital Predistortion Linearization Based on Memory Polynomials for HF Power Amplifiers», in IEEE Access, vol. 12, pp. 29362-29372, 2024, doi: 10.1109/ACCESS.2024.3368858.

9. J. Ren. «A New Digital Predistortion Algorithms Scheme of Feedback FIR Cross-Term Memory Polynomial Model for Short-Wave Power Amplifier», in IEEE Access, vol. 8, pp. 38327-38332, 2020, doi: 10.1109/ACCESS.2020.2976152.

10. Ahmed, Serien, Majid Ahmed, Souheil Bensmida, and Oualid Hammi. 2024. «Power Amplifier Predistortion Using Reduced Sampling Rates in the Forward and Feedback Умths» Sensors, vol. 24, no. 11: 3439. https://doi.org/10.3390/s24113439.

11. Somasundaram, Gokul, Jill C. Mayeda, Clint Sweeney, Donald Y. C. Lie, and Jerry Lopez. 2023. «Effective Digital Predistortion (DPD) on a Broadband Millimeter-Wave GaN Power Amplifier Using LTE 64-QAM Waveforms» Electronics, vol. 12, no. 13: 2869. https://doi.org/10.3390/ electronics 12132869.

12. S. Boumaiza and F. Ghannouchi. «Thermal memory effects modeling and compensation in RF power amplifiers and predistortion linearizers», IEEE Trans. Microw. Theory Tech., vol. 51, pp. 2427-2433, Dec. 2003.

13. Abdelrahman AE, Hammi O, Kwan AK, Zerguine A, Ghannouchi FM (2016) A novel weighted memory polynomial for behavioral modeling and digital predistortion of nonlinear wireless transmitters. IEEE Trans Industrial Electron 63(3):1745-1753

14. Le D. H. et al. Linearization of RF power amplifiers in wideband communication systems by adaptive indirect learning using RPEM algorithm. Mobile Networks and Applications. 2020. Т. 25. С. 1988-1997.

15. Hussein M.A., Bohara V.A., Venard O. (2012) On the system level convergence of ILA and DLA for digital predistortion. In: Proc. 2012 international symposium on wire less communication systems (ISWCS), pp. 870-874.

16. Chani-Cahuana J, Landin PN, Fager C, Eriksson T (2016) Iterative learning control for RF power amplifier linearization. IEEE Trans Microw Theory Tech 64(9):2778-2789.

17. Vu V.K., Tran V.N. Adaptive digital predistortion of RF power amplifiers based on memory polynomial model and indirect learning architecture. 2020 International Conference Engineering and Telecommunication (En&T). IEEE, 2020. С. 1-5.

ООО «Лаборатория Сфера»

SPHERA LAB

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, цифрового вещания и оборудования для электромобилей.

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.
- **Обучение пользователей.** Проводим обучающие семинары и тренинги. При необходимости возможно проведение углубленного повышения квалификации.
- **Консультации.** Осуществляем консультативную деятельность в области компьютерных технологий и систем связи.

Наши клиенты: ООО «Т8», АвтоВАЗ, LADA Sport, ООО «ЗВ-Сервис», ООО «Цифровые решения», ООО «НПФ «Сад-Ком», ООО НПП «ТЕПЛОВОДОХРАН», ООО «Силовая электроника», МТУСИ, МФТИ и др.

Контакты:

Адрес: 390000, г. Рязань, ул. Каширина, д.1т, второй этаж, офис 1
Тел.: +7(995)540-75-03
E-mail: info@labsphera.ru сайт: https://labsphera.ru/

УДК 621.396.9

СРАВНЕНИЕ ТОЧНОСТИ КЛАССИФИКАЦИИ РАДИОЛОКАЦИОННЫХ ИЗОБРАЖЕНИЙ ОБЪЕКТОВ СВЕРТОЧНОЙ НЕЙРОННОЙ СЕТЬЮ И СЕТЬЮ С АРХИТЕКТУРОЙ VISION TRANSFORMER

Купряшкин И.Ф., *д.т.н., доцент, начальник кафедры «Военного учебно-научного центра Военно-воздушных сил «Военно-воздушная академия им. проф. Н.Е. Жуковского и Ю.А. Гагарина» (г. Воронеж), e-mail: ifk78@mail.ru*

CONVOLUTIONAL NEURAL NETWORK AND A NETWORK WITH VISION TRANSFORMER ARCHITECTURE ACCURACY COMPARISON IN RADAR IMAGES CLASSIFICATION TASK

Kupryashkin I.F.

Training and test data preparing procedure is described, a structures of a deep convolutional neural network and a network with the Vision Transformer are given. Their accuracy in classifying MSTAR dataset radar images at different noise levels was compared. It is shown that the use of the Vision Transformer architecture in this task does not lead to an increase in classification accuracy even with a significant change in the hyperparameters.

Key words: deep convolutional neural network, transformer, radar image, classification accuracy.

Ключевые слова: глубокая сверточная нейронная сеть, трансформер, радиолокационное изображение, точность классификации.

Описан порядок подготовки обучающих и тестовых данных, приведены структуры глубокой сверточной нейронной сети и сети с архитектурой Vision Transformer. Проведена их сравнительная оценка по точности классификации отметок объектов на радиолокационных изображениях набора MSTAR при различном уровне шума. Показано, что применение архитектуры Vision Transformer в рассматриваемой задаче не приводит к увеличению точности классификации даже при существенном изменении гиперпараметров.

Введение

Вопросы обработки радиолокационных изображений (РЛИ) земной поверхности и объектов с использованием технологий глубокого обучения продолжают вызывать значительный интерес и во многом уже достаточно хорошо проработаны [1-8]. Как правило, в числе наиболее перспективных инструментов для решения задачи классификации РЛИ объектов до недавнего времени рассматривались глубокие сверточные нейронные сети (ГСНС).

Однако начиная с 2017 года, с момента публикации работы [9] активное развитие получили сети, основанные на архитектуре Transformer, использующей механизмы внимания (self-attention). На начальном этапе новая архитектура продемонстрировала наиболее впечатляющие результаты в задачах обработки естественного языка, и достаточно отметить, что все современные большие языковые модели основаны именно на ней. Эти успехи вполне ожидаемо стимулировали применение архитектуры Transformer в других задачах, в том числе задачах компьютерного зрения.

Одно из первых описаний эффективного варианта трансформера для обработки изображений приведено в [10]. Предложенная архитектура получила название Vision Transformer (ViT) и в настоящий момент сеть OmniVec [11], основанная на ViT, демонстрирует наилучшую точность (92,4 % Top-1 Accuracy) на наборе ImageNet. К настоящему времени уже известны различные варианты реализации трансформеров для обработки РЛИ [12-14], однако исследование их возможностей при различном уровне шума на изображении не проводилось.

В связи с этим сравнение возможностей ViT-архитектуры и ГСНС в задачах обработки радиолокационных изображений представляет безусловный интерес, и целью настоящей работы является сравнение точности классификации малоразмерных объектов на детальных радиолокационных изображениях, достигаемой при использовании указанных архитектур, при различном уровне шума.

Характеристика обучающих и тестовых наборов данных

В [15] подробно описан порядок подготовки обучающих и тестовых наборов данных, и в настоящей работе он является полностью аналогичным. Так же в качестве исходных данных использована открытая часть набора MSTAR [16], включающего радиолокационные изображения X-диапазона десяти образцов военной техники с разрешением 0,3 м (рисунок 1).

Весь имеющийся набор изображений разбит на обучающий (1923 изображения, угол визирования при съемке 17 градусов), проверочный (891 изображение, угол визирования 17 градусов) и тестовый (2503 изображения, угол визирования 15 градусов) наборы.

Каждое изображение набора MSTAR хранится в отдельном файле в форме двумерного массива комплексных отсчетов $\dot{\mathbf{I}}$. Предварительная обработка этих изображений включает подавление фона местности вблизи отметок объектов, понижение размерности изображений до 44×44 отсчетов, добавление аддитивного гауссовского шума для имитации воздействия шумовой

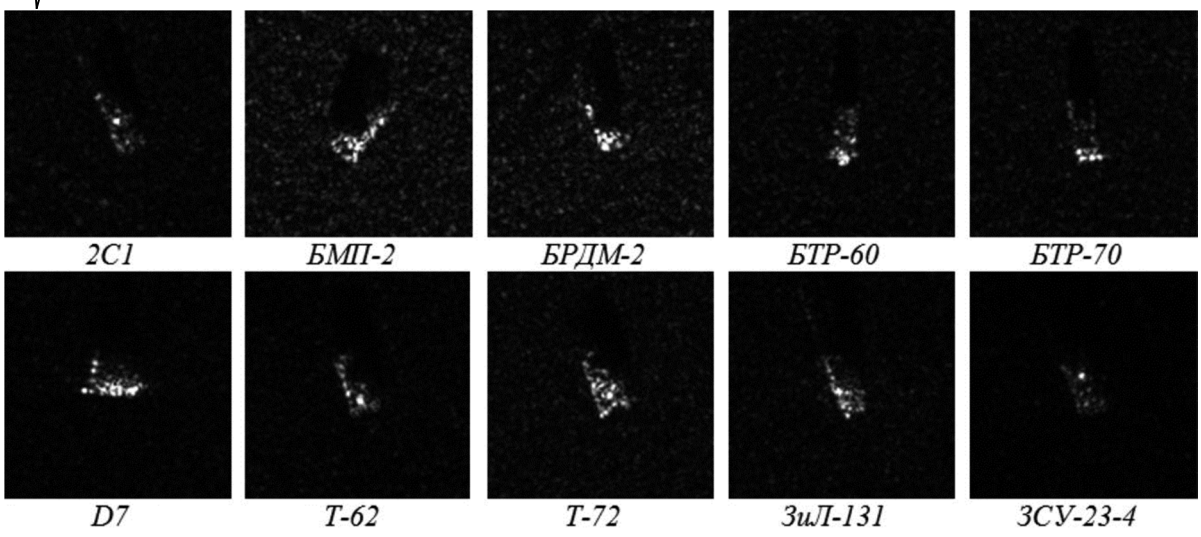


Рисунок 1 – Радиолокационные изображения объектов набора MSTAR

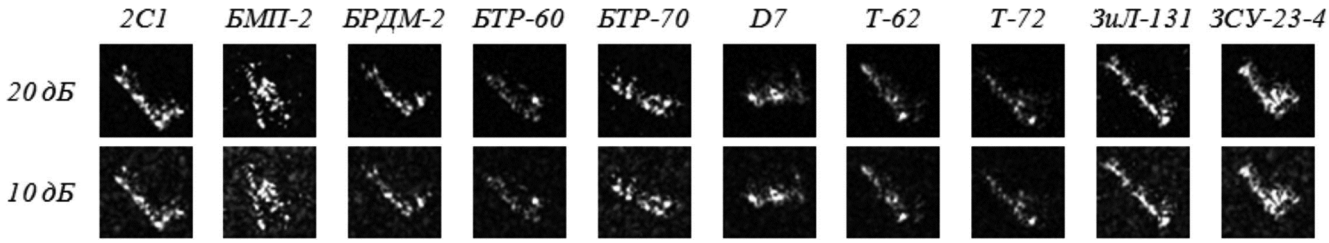


Рисунок 2 – Примеры изображений обучающих наборов

помехи на РЛС, пространственную фильтрацию и преобразование в целочисленный беззнаковый восьмибитный формат [15, 17].

Сформированы два одинаковых набора изображений, отличающихся только средним уровнем шумовой помехи. В первом наборе уровень шума выбирался таким, чтобы отношение среднего уровня отметок объектов к среднему уровню помехи на изображении составило $q = 20$ дБ, во втором наборе – $q = 10$ дБ. Примеры изображений приведены на рисунке 2.

Архитектура нейронных сетей и результаты их обучения

В качестве инструмента для разработки и обучения сетей использовался фреймворк TensorFlow 2.0 и библиотека Keras [18].

Для сверточной сети использовалась архитектура VGG-типа [19], включающая четыре пары сверточных слоев с одинаковыми ядрами 3×3, единичным шагом свертки и количеством фильтров 32, 64, 128 и 256 в каждой паре соответственно. После каждой пары включены слои подвыборок 2×2. Классификатор включает полносвязный входной слой на 4096 входов и 256 выходов, слой 50-процентного прореживания и выходной полносвязный слой с 10 выходами по числу классов объектов. В качестве функции активации сверточных слоев и входного слоя классификатора используется ReLU. Размерность входного слоя составляла 128×128 отсчетов. Количество настраиваемых параметров сети составило 2'223'082. Архитектура сверточной сети представлена на рисунок 3,*а*.

Архитектура трансформера включает входной слой с размерностью 128×128, к которому подключен слой, обеспечивающий преобразование входного изображения в последовательность его фрагментов (patches). Длина последовательности определяется исходя из заданной размерности фрагмента, так, при заданной в рассматриваемом случае размерности отдельного фрагмента 32×32 отсчета их количество составляет 16.

Далее каждый фрагмент преобразуется в одномерный вектор длиной 1024 элемента, и подается на позиционный кодировщик (patch encoder), включающий полносвязный слой и слой кодирования позиции (эмбеддинга). Входом полносвязного слоя является собственно вектор, входом слоя эмбеддинга – номер соответствующего фрагмента в поступающей последовательности. Сумма их выходов является общим выходом позиционного кодировщика. Размерность выхода является гиперпараметром сети, в данном случае ее значение задавалось равным 128.

Закодированная последовательность фрагментов изображения поступает на стек из восьми одинаковых блоков трансформера. Каждый блок трансформера включает слой нормализации, слой многоголового внимания (с размерностью векторов запроса, ключа и значения, равной 96, и четырьмя «головами» внимания), полносвязные слои с 10-и процентным прореживанием. Количество полносвязных слоев составляет 16 и определяется длиной последовательности, в каждом из них осуществляется обработка отдельных кодированных векторов последовательности независимо друг от друга. В качестве функции активации полносвязных слоев применяется geLU [20].

Выходом стека блоков трансформера является последовательность из 16 векторов с длиной 128 элементов каждый, которая перед подачей на вход полно-

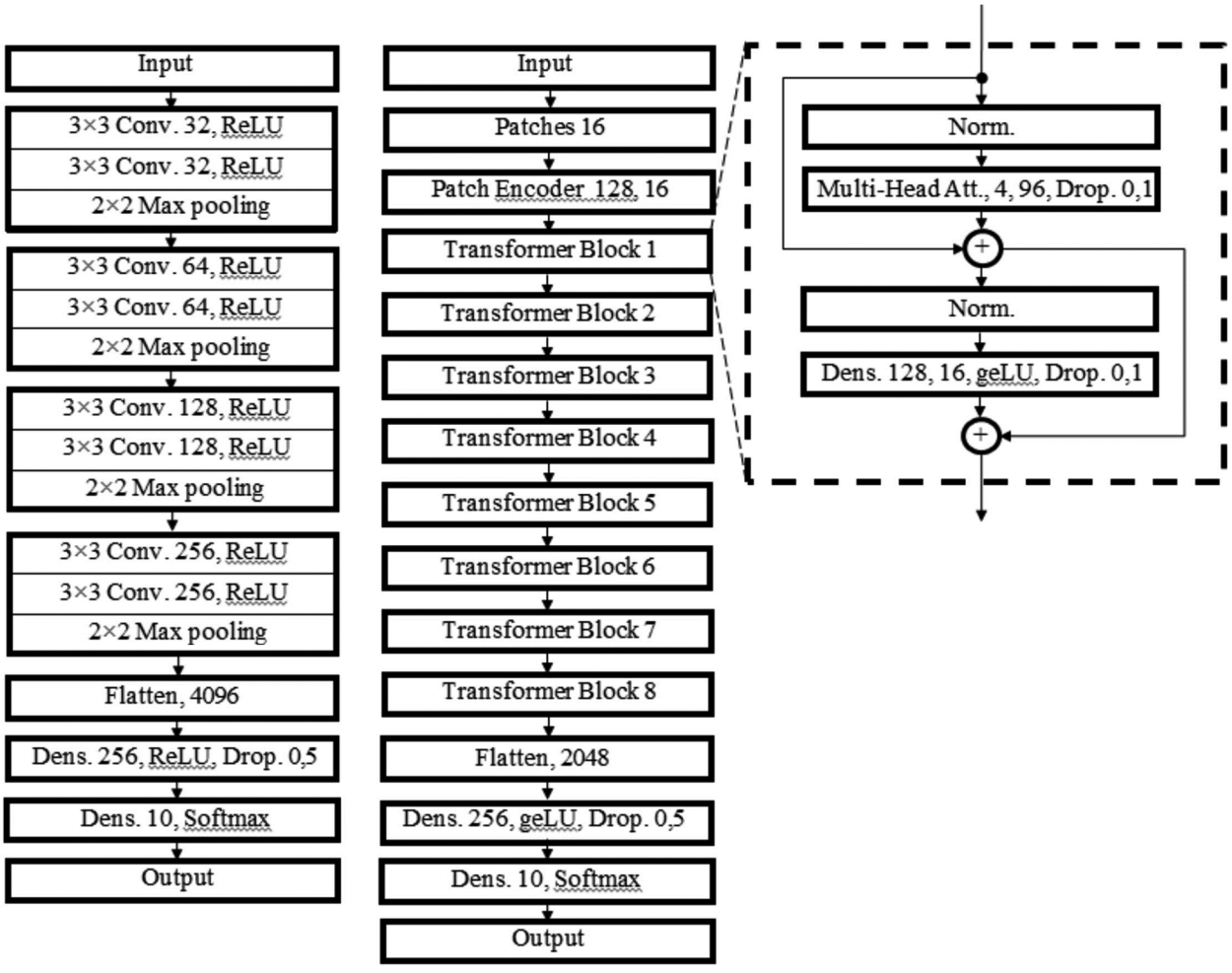
связного классификатора преобразуется в единый вектор длиной 2048. Классификатор включает входной полносвязный слой на 2048 входов и 256 выходов с функцией активации geLU, слой 50-процентного прореживания и выходной полносвязный слой с 10 выходами по числу классов объектов с функцией активации softmax.

Гиперпараметры трансформера выбирались таким образом, чтобы общее количество настраиваемых параметров (обучаемых весов), равное 2'379'914, примерно соответствовало количеству настраиваемых весов сверточной нейросети (2'223'082) для обеспечения возможности объективного сравнения получаемых результатов. Архитектура трансформера представлена на рисунке 3,*б*.

Условия обучения сверточной сети и трансформера являются одинаковыми. В качестве оптимизатора используется адаптивный алгоритм Adam с параметром скорости обучения 10^{-4}, обучение осуществляется в течение четырехсот (для наборов с $q = 20$ дБ) и шестисот (для наборов с $q = 10$ дБ) эпох пакетами по 64 изображения. В качестве функции потерь применена перекрестная энтропия [21]. Общее время обучения с использованием GPU (GeForce GTX1650) составляет примерно 27 минут ($q = 20$ дБ, 400 эпох) и 40 минут ($q = 10$ дБ, 600 эпох) как для сверточной сети, так и для трансформера, т.е. около четырех секунд на эпоху во всех случаях. В качестве дополнительного приема предотвращения переобучения применяется расширение данных (augmentation) путем смещения каждого изображения обучающего набора по вертикали и горизонтали на случайное число пикселей от одного до пяти. Этот же прием при подаче изображений валидационного и тестового наборов не применялся. Перед подачей на вход нейронной сети целочисленные значения диапазона [0; 255] преобразуются в вещественные диапазона [0; 1] путем деления на 255.

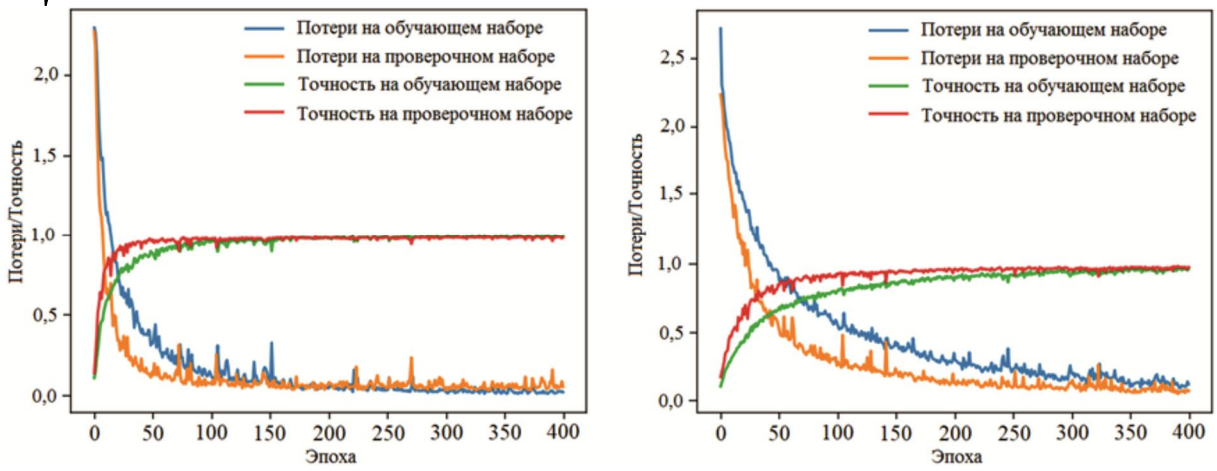
Ход процесса обучения на наборе с $q = 20$ дБ иллюстрируют зависимости функции потерь и точности классификации на рисунке 4, на наборе с $q = 10$ дБ – аналогичные зависимости на рисунке 5.

На тестовом наборе с $q = 20$ дБ сверточная сеть продемонстрировала точность 97,84 %, а трансформер – 95,43 %. Полученная оценка точности трансформера хорошо согласуется с результатами 95,96 % и 95,61 %, полученными в [13] и [14] соответственно при сходных, хотя и не полностью аналогичных, условиях обучения. На наборе с $q = 10$ дБ результат составил 95,59 % и 92,79 % соответственно. То есть трансформер с сопоставимым количеством весов на обоих наборах продемонстрировал несколько худшие результаты. Кроме того, из анализа зависимостей на рисунках 4 и 5 следует, что сверточная сеть характеризуется еще и более



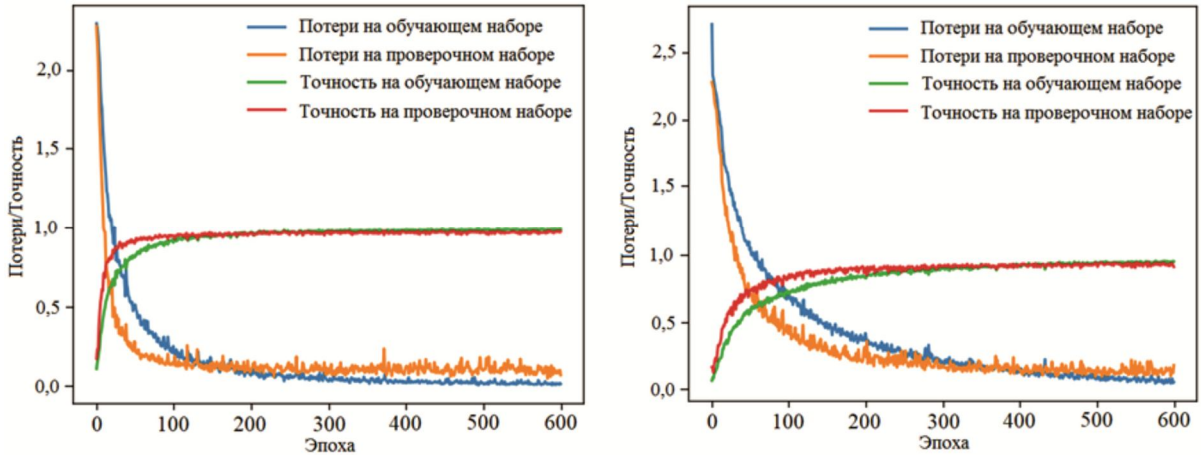
а) *б)*

Рисунок 3 – Архитектура сверточной нейросети (а) и трансформера (б)



а) *б)*

Рисунок 4 – Ход процесса обучения сверточной (а) сети и трансформера (б) при q = 20 дБ



а) *б)*

Рисунок 5 – Ход процесса обучения сверточной (а) сети и трансформера (б) при q = 10 дБ

высокой скоростью обучения (скоростью изменения точности и потерь). Можно также отметить несколько лучшую стабильность процесса обучения обоих сетей на наборе с $q = 10$ дБ, что обусловлено, по всей видимости, регуляризирующим эффектом шумовой помехи. Особенность зависимостей, состоящая в более быстром нарастании точности не на обучающем, а на валидационном (проверочном) наборе, обусловлена отсутствием случайных сдвигов изображений последнего при подаче на вход сети.

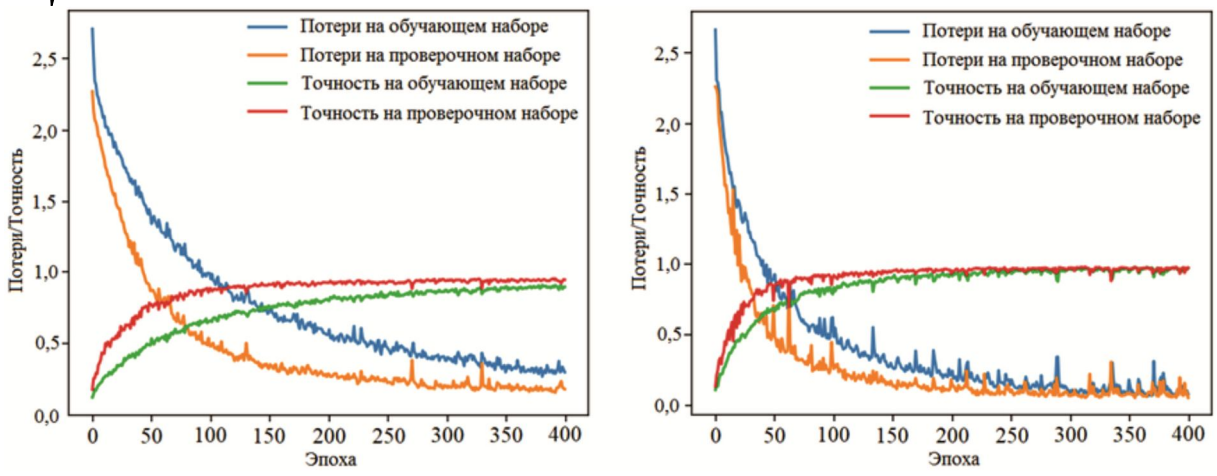
Для оценки влияния гиперпараметров трансформера на получаемые результаты были обучены и протестированы еще два его варианта, кратно отличающиеся количеством настраиваемых параметров от описанного выше.

В первом случае размерность фрагментов была увеличена с 32×32 до 64×64, то есть длина обрабатываемой последовательности уменьшилась с 16-ти до четырех, размерность слоев позиционного кодировщика снижена со 128-ти до 64-х, количество блоков трансформера уменьшено с восьми до четырех. Количество «голов» внимания не изменилось и осталось равным четырем, однако размерность векторов запроса, ключа и значения снижена с 96-ти до 64-х. Общее количество настраиваемых параметров этого варианта трансформера составило 614’090, то есть уменьшилось практически в четыре раза. Ход процесса обучения на наборе с $q = 20$ дБ иллюстрируют зависимости функции потерь и точности классификации на рисунке 6,*а*. Точность на тестовом наборе составила 91,79 %.

Во втором случае размерности фрагментов и слоев позиционного кодировщика не изменялись и составили 32×32 и 128 соответственно. Количество блоков трансформера увеличено с восьми до 16-ти. Количество «голов» внимания увеличено с четырех до восьми, размерность векторов запроса, ключа и значения увеличена с 96-ти до 128-ти. Общее количество настраиваемых параметров этого варианта сети составило 9’372’810, то есть увеличилось практически в четыре раза. Ход процесса обучения на наборе с $q = 20$ дБ иллюстрируют зависимости на рисунке 6,*б*. Точность на тестовом наборе составила 95,03 %.

Заключение

Таким образом, сравнение результатов классификации радиолокационных изображений набора MSTAR классической сверточной нейронной сетью VGG-типа и сетью с ViT-архитектурой продемонстрировало более низкое качество работы последней во всех случаях. При этом как четырехкратное увеличение, так и четырехкратное уменьшение количества настраиваемых параметров (обучаемых весов) трансформера не привело к



а) *б)*

Рисунок 6 – Ход процесса обучения сетей-трансформеров с уменьшенным до 614’090 (а) и увеличенным до 9’372’810 (б) количеством настраиваемых параметро

увеличению точности классификации по сравнению с исходным вариантом, сопоставимым по количеству весов со сверточной сетью. Сверточная сеть также продемонстрировала несколько более высокую скорость обучения.

Полученный результат, разумеется, не свидетельствует о более низком качестве работы трансформеров в задаче классификации радиолокационных изображений вообще. Но, по всей видимости, раскрытие преимуществ ViT-архитектуры по сравнению со сверточной сетью, как и в случае оптических изображений, возможно при использовании датасетов с существенно большим количеством и разнообразием примеров по сравнению с хорошо сбалансированным, однородным и относительно небольшим базовым набором MSTAR.

Литература

1. Zhu X., Montazeri S., Ali M., Hua Yu., Wang Yu., Mou L., Shi Yi., Xu F., Bamler R. Deep Learning Meets SAR. arXiv:2006.10027.

2. Chen S., Wang H., Xu F., Jin Y.-Q. Target Classification Using the Deep Convolutional Networks for SAR Images. IEEE Transaction Geoscience and Remote Sensing, 2016, vol. 54, no. 8, pp. 4806-4817. doi: 10.1109/TGRS.2016.2551720.

3. Anas H., Majdoulayne H., Chaimae A., Nabil S.M. Deep Learning for SAR Image Classification. 2020. doi: 10.1007/978-3-030-29516-5_67.

4. Chen S., Wang H. SAR Target Recognition Based on Deep Learning. 2014 International Conference on Data Science and Advanced Analytics (DSAA), 2014, pp. 541-547. doi: 10.1109/DSAA.2014.7058124.

5. Coman C., Thaens R. A Deep Learning SAR Target Classification Experiment on MSTAR Dataset. 2018 19th International Radar Symposium (IRS), 2018, pp. 1-6. doi: 10.23919/IRS.2018.
8448048.

6. Furukawa H. Deep Learning for End-to-End Automatic Target Recognition from Synthetic Aperture Radar Imagery. arXiv:1801.08558.

7. Profeta A., Rodriguez A., Clouse H.S. Convolutional Neural Networks for Synthetic Aperture Radar Classification. Proc. SPIE 9843, Algorithms for Synthetic Aperture Radar Imagery XXIII, 98430M (14 May 2016). https://doi.org/10.1117/12.2225934.

8. Amrani M., Bey A., Amamra A. New SAR Target Recognition Based on YOLO and Very Deep Multi-Canonical Correlation Analysis. arXiv:2110.15383 https://doi.org/10.48550/arXiv.2110.15383.

9. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser L., Polosukhin I. Attention Is All You Need. arXiv:1706.03762.

10. Dosovitskiy A., Beyer L., Kolesnikov A., Weissenborn D., Zhai X., Unterthiner T., Dehghani M., Minderer M., Heigold G., Gelly S., Uszkoreit J., Houlsby N. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv:2010.11929.

11. Srivastava S., Sharma G. OmniVec: Learning robust representations with cross modal sharing. arXiv:2311.05709.

12. Li K., Zhang M., Xu M., Tang R., Wang L., Wang H. Ship Detection in SAR Images Based on Feature Enhancement Swin Transformer and Adjacent Feature Fusion. Remote Sensing 2022, 14, 3186. https://doi.org/10.3390/rs14133186.

13. Wickramasinghe S., Parikh D., Zhang B., Kannan R., Prasanna V., Busart C. VTR: An Optimized Vision Transformer for SAR ATR Acceleration on FPGA. arXiv:2404.04527.

14. Fein-Ashley J., Ye T., Kannan R., Prasanna V., Busart C. Benchmarking Deep Learning Classifiers for SAR Automatic Target Recognition. arXiv:2312.06940.

15. Купряшкин И.Ф. Сравнительные результаты точности классификации радиолокационных изображений объектов набора MSTAR сверточными нейронными сетями с различными архитектурами. Журнал радиоэлектроники, 2021, №11. https://doi.org/10.30898/1684-1719.2021.11.14.

16. https://www.sdms.afrl.af.mil/index.php?collection=mstar.

17. Шолле Ф. Глубокое обучение на Python. СПб:Питер, 2018. 400 с.

18. Simonyan K., Zisserman A. Very Deep Convolutional Networks For Large-Scale Image Recognition. arXiv:1409.1556. https://arxiv.org/abs/1409.1556.

19. Hendrycks D., Gimpel K. Gaussian Error Linear Units (GELUs). arXiv:1606.08415. https://doi.org/10.48550/arXiv.1606.08415.

20. Николенко С., Кадурин А., Архангельская Е. Глубокое обучение. СПб.: Питер, 2018. 480 с.

УДК 621.396.621

ОЦЕНКА ИСКАЖЕНИЙ ДИАГРАММЫ НАПРАВЛЕННОСТИ РАДИОЛОКАЦИОННОЙ СТАНЦИИ ПРИ НЕСТАЦИОНАРНОМ ВОЗДЕЙСТВИИ ТЕПЛОВЫХ ПРОЦЕССОВ НА МАЛОШУМЯЩИЙ УСИЛИТЕЛЬ ПРИЕМНОГО ТРАКТА

Тимошенко А.В., *д.т.н., профессор, старший преподаватель Военного Учебного Центра МГУ им. М.В. Ломоносова, e-mail: u567ku78@gmail.com*
Захаров А.С., *лаборант факультета космических исследований МГУ им. М.В. Ломоносова, e-mail: zakharov.as17@physics.msu.ru*
Байкалов И.А., *инженер 1 категории лаборатории 341 АО «НПО дальней радиолокации им. А.Л.Минца», e-mail: ibaikalov@npodr.ru*
Урсегов А.К., *инженер лаборатории 422 АО «НПО дальней радиолокации им. А.Л. Минца», e-mail: aursegov@npodr.ru*
Савчук А.М., *д.ф-м.н., профессор факультета космических исследований МГУ им. М.В. Ломоносова, e-mail: savchuk@cosmos.msu.ru*

ESTIMATION OF RADIATION PATTERN DISTORTIONS OF A RADAR STATION UNDER THE NONSTATIONARY EFFECT OF THERMAL PROCESSES ON A LOW-NOISE AMPLIFIER OF THE RECEIVING PATH

Timoshenko A.V., Zakharov A.S., Baikalov I.A., Ursegov A.K., Savchuk A.M.

Modern radar stations strive to increase the resolution to accurately determine the coordinates and characteristics of objects. One of the promising methods to achieve this goal is the use of a super-resolution mode. However, thermal distortions occur during the operation of the radar, which significantly reduce the efficiency of superresolution. This article presents a new model for the analysis and compensation of these distortions, based on the use of the original integro-differential equation.

It is shown that the unsteady behavior of thermal processes leads to distortions of the amplitude-phase distribution, amplitude-frequency response and radiation pattern, which changes the resolution characteristics. It is proposed to compensate for the influence of thermal processes by means of operational calibration of amplitude and phase, taking into account the distortions of the radiation pattern.

A demonstration experiment was conducted to simulate the radiation pattern at different temperatures of low-noise amplifiers of the receiving path and the same heating rate.

Key words: radar station, receiving path, thermal conductivity, radiation pattern.

Ключевые слова: радиолокационная станция, приемный тракт, теплопроводность, диаграмма направленности.

Современные радиолокационные станции стремятся к повышению разрешающей способности для точного определения координат и характеристик объектов. Одним из перспективных методов для достижения этой цели является применение режима сверхразрешения. Однако в процессе работы радиолокационной станции возникают тепловые искажения, которые значительно снижают эффективность сверхразрешения. В данной статье представлена новая модель для анализа и компенсации этих искажений, основанная на использовании оригинального интегро-дифференциального уравнения.

Показано, что нестационарное поведение тепловых процессов приводит к искажениям амплитудно-фазового распределения, амплитудно-частотной характеристики и диаграммы направленности, что изменяет характеристики разрешения. Предложено компенсировать влияние тепловых процессов с помощью оперативной калибровки амплитуды и фазы с учетом искажений диаграммы направленности.

Проведен демонстрационный эксперимент по моделированию диаграммы направленности при разных температурах малошумящих усилителей приемного тракта и одинаковой скорости нагрева.

Введение

Современные требования к радиолокационным станциям мониторинга космического пространства (РЛС МКП) [1-5] помимо традиционных задач обнаружения и сопровождения космических объектов включают режим сверхраз решения для малоразмерных космических объектов (КО), целью которого является создание радиолокационных портретов (РЛП) [6]. Это возможно даже при использовании квазинепрерывных сигналов с длинными импульсами и низкой скважностью для радиолокационного сопровождения [7-11]. Для работы в режиме сверхразрешения требуется обеспечение равномерности амплитудно-частотной характеристики (АЧХ) системы передачи и приема сигнала [12, 13].

Выбор сигнала для создания радиолокационного портрета зависит от возможностей аппаратно-программного комплекса (АПК) РЛС, равномерности АЧХ системы и функции неопределенности (ФН), которая помогает восстанавливать профиль по дальности и скорости. ФН играет ключевую роль при анализе разрешающей способности сигналов в этих координатах.



Рисунок 1 – Обобщенная схема приемо-передающего тракта РЛС [15]

В настоящее время приемный тракт РЛС МКП состоит из малошумящего усилителей и блока цифровой обработки информации [14]. Общую структурную схему РЛС МКП можно представить в виде совокупности функциональных узлов приемной и передающей систем соответственно [15]. Обобщенная схема ключевых функциональных узлов приемо-передающего тракта МКП приведена на рисунке 1.

Указанные на схеме функциональные узлы обеспе--чивают, как формирование, так и прием и обработку высокочастотных сигналов, что обуславливает обеспечения требуемых значений функциональных характеристик. Одной из ключевых особенностей РЛС МКП, обуславливающих необходимость высоких требований в части разрешающей способности к функциональным узлам являются протяженные приемо-передающие тракты. Так же стоит отметить, что число данных трактов для одного образца может достигать нескольких тысяч, что создает предпосылки к возникновению взаимного влияния аппаратуры при непрерывном функционировании РЛС в энергонапряженных режимах. Одним из наиболее значимых с точки зрения прохождения сигналов элементом приемной системы РЛС МКП является малошумящий усилитель (МШУ), характеризующий в значительной степени шумовую температуру приемного тракта [15]. Неправильное изменение амплитуды и фазы сигнала после прохождения МШУ приведет к неточным результатам обработки и снижению разрешающей способности.

Тем самым, обеспечение требуемых характеристик сверхразрешения в широком диапазоне фаз и амплитуд достигается разработкой более точной и детальной модели изменения диаграммы направленности, которая является основой аппарата коррекции амплитудно-фазового распределения путем введения дополнитель--ных поправок к амплитуде и фазе сигнала в части цифровой обработки сигнала.

Модель изменения амплитуды и фазы

Как правило [16], моделирование тепловых процессов проводится с помощью уравнения теплопроводности, в то время как расчет электродинамических характеристик проводится на основе решений уравнений Максвелла в решении объединенной задачи.

Для описания модели изменения ФН введем уравнение теплопроводности для АФАР с электромагнитной мощностью в виде функции теплового источника представим в виде [14]:

$$c\rho\frac{\partial \mathrm{T}}{\partial t}=\lambda\nabla^2 T+P_d,$$

где c – удельная теплоемкость материала, ρ – плотность материала, λ – коэффициент теплопроводности, P_d – электромагнитная мощность, рассеиваемая в единице объема пространства, $\nabla^2=\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}$ – оператор Лапласса (оператор набла в квадрате).

Электромагнитную мощность сигнала на частоте ω, можно представить как:

$$P=\left|\sum_{n=1}^{N}S_n^{'}\right|^2\omega C_R\, tg\delta\, d^2,$$

где tg – тангенс угла потерь в каналах приемного тракта, d – длина линии передачи, CR – емкость линий передачи. Представив объем и емкость как:

$$V=Sd,\ C_R=\frac{\varepsilon\varepsilon_0 S}{d}.$$

Можно получить электромагнитную мощность, рассеиваемую в единицу объема:

$$P_d=\omega\varepsilon_0\varepsilon\left|\sum_{n=1}^{N}S_n^{'}\right|^2\ \mathrm{tg}\delta.$$

где ε_0 – электрическая постоянная, ε – диэлектрическая проницаемость пространства в каналах приемного тракта.

Для проведения частотного анализа умножим обе части уравнения теплопроводности на оператор прямого преобразования Фурье без нормировочных множителей:

$$F_+\left[f(t)\right]=\int_{-\infty}^{\infty}f(t)e^{-j\omega t}dt.$$

Применяя оператор прямого преобразования Фурье к правой и левой части уравнения, получим:

$$F_+\left[c\rho\frac{\partial \mathrm{T}}{\partial t}\right]=F_+\left[\lambda\nabla^2 T\right]+F_+\left[\omega\varepsilon_0\varepsilon\left|S^*\left(\omega,T\right)\right|^2\right].$$

Учитывая, что все константы и операторы выносятся из-под знака интеграла, то:

$$c\rho F_+\left[\frac{\partial \mathrm{T}}{\partial t}\right]=\lambda\nabla^2 F_+\left[T\right]+\omega\varepsilon_0\varepsilon F_+\left[\left|S^*\left(\omega,T\right)\right|^2\right].$$

Воспользуемся следующим свойством прямого преобразования Фурье:

$$F_+\left[\frac{\partial T}{\partial t}\right]=j\omega F_+\left[T\right].$$

При подстановке данного свойства мы переходим от дифференциального уравнения к интегральному термочастотному уравнению для определения искажений сигнала при обработке:

$$j\omega c\rho F_+\left[T\right]=\lambda\nabla^2 F_+\left[T\right]+\omega\varepsilon_0\varepsilon F_+\left[\left|S^*\left(\omega,T\right)\right|^2\right].$$

Задачу можно записать в операторном виде. Процедура сводится к нахождению температурного распределения с помощью уравнения теплопроводности и подстановки в операторное уравнение на поиск искажений сигнала:

$$F_+\left[T\right]*\left\{\frac{j\omega c\rho-\lambda\nabla^2}{\omega\varepsilon_0\varepsilon}\right\}=F_+\left[\left|S^*\left(\omega,T\right)\right|^2\right].$$

Таким образом, можно получить искаженный сигнал после прохождения его через МШУ по формуле:

$$S^*\left(\omega,T\right)=\sqrt{F_-\left[F_+\left[T\right]*\left\{\frac{j\omega c\rho-\lambda\nabla^2}{\omega\varepsilon_0\varepsilon}\right\}\right]}.$$

Для расчета диаграммы направленности гексагональной АФАР перейдем в обобщенные сферические координаты u и v, тогда общая формула:

$$D(A,\phi)=\sum_i Ae^{j\phi}e^{j\frac{2\pi}{l}\left[x_i(u_i-a)+y_i(v_i-b)\right]}=$$

$$=\sum_i Ae^{j\phi}F(\lambda,x_i,y_i,v_i,u_i,b),$$

где F – множитель решетки; x, y – координаты излучателя; a, b – центр луча в обобщенных координатах, l – длина волны.

Тогда измененная диаграммы направленности будет равна:

$$D_{иск}(A,\phi)=\sum_i(A+\Delta A(T))e^{j(\phi+\Delta\phi)}F(\lambda,x_i,y_i,v_i,u_i,b).$$

Задав определенный сигнал или сочетание сигналов для режима сверхразрешения посчитать искажения сигнала и посчитать изменения диаграммы направленности, что ведет к повышению вероятности срыва сопровождения космического аппарата.

Демонстрационный эксперимент

В качестве демонстрационного эксперимента рассмотрим воздействие тепловых процессов на малошумящие усилители, температура по которым распределена равномерно, но при этом происходит линейное увеличение температуры во времени. В качестве сигнала сверхразрешения возьмем дискретно-кодированный по частоте (ДКЧ) сигнал и разбиением на 198 кодов по диаграмме Костаса. Моделирование проходило в среде MatLab.

Изначально были получены функции неопределенности на полотне время-частота, после чего снимались срезы и графики переводились в плоскость расстояние и радиальная скорость для анализа разрешающей способности. На рисунке 2 показаны срезы функции неопределенности по дальности и радиальной скорости при самой низкой температуре 20˚С и нулевой скорости нагрева.

Теперь увеличим скорость нагрева до 2˚С/мин, после чего проведем моделирование ДН на 40˚С, которое показано на рисунке 4.

Как мы видим из рисунка 4, помимо увеличения боковых лепестков и появление шумовых максимумов произошло небольшое расширение главного лепестка, что повышает разрешающую способность по дальности и углу места. Теперь рассмотрим максимальную температуру, при которой не происходит отключений аппаратуры, а именно 60˚С, при той же скорости нагрева. Срез ФН по скорости представлен на рисунке 5, а диаграмма направленности на рисунках 6 и 7.

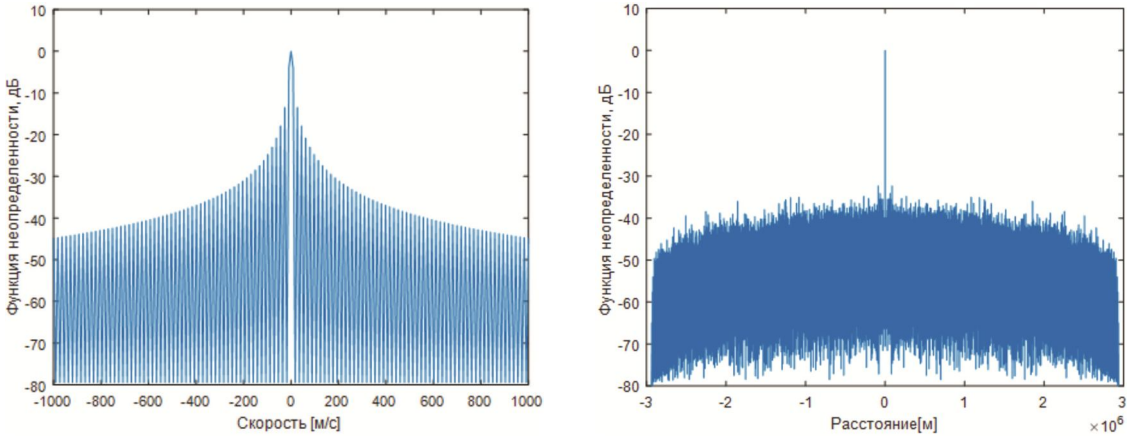


Рисунок 2 – Функции неопределенности в срезе на полотне расстояние/скорость

Диаграмма направленности в таких условиях показана на рисунке 3

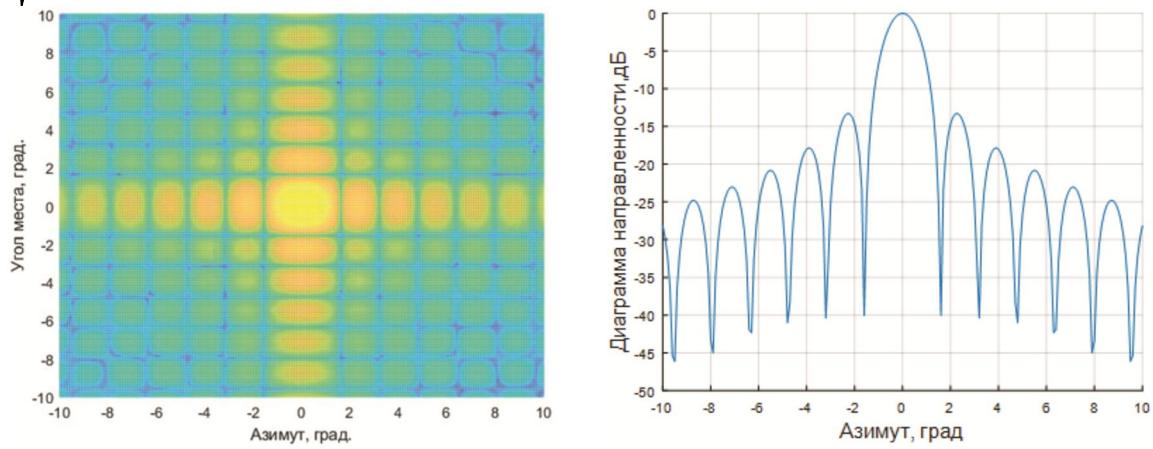


Рисунок 3 – Диаграмма направленности при температуре всех МШУ в тракте равной 20˚С

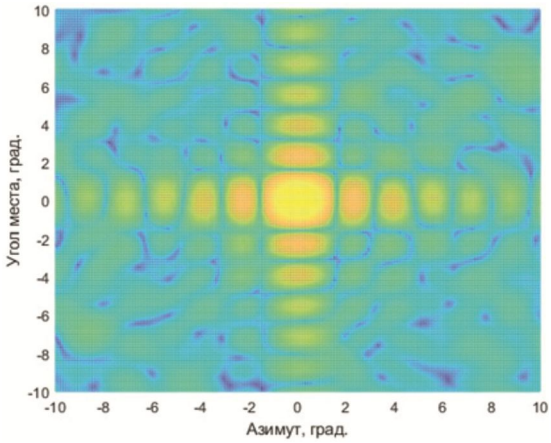


Рисунок 4 – Диаграмма направленности АФАР при нагреве всех МШУ до 40˚С и скорости нагрева 2˚С/мин

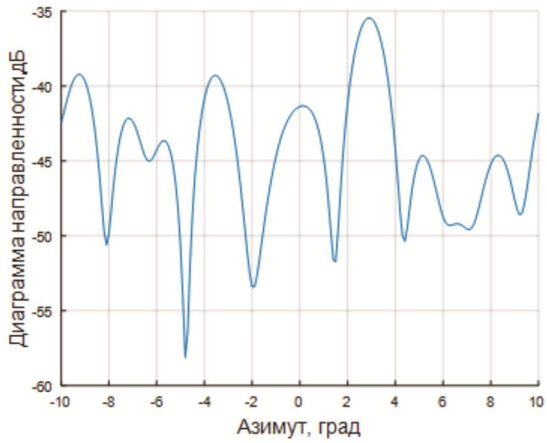


Рисунок 6 – Диаграмма направленности по азимуту при нагреве 2˚С/мин и равномерной температуре 60˚С

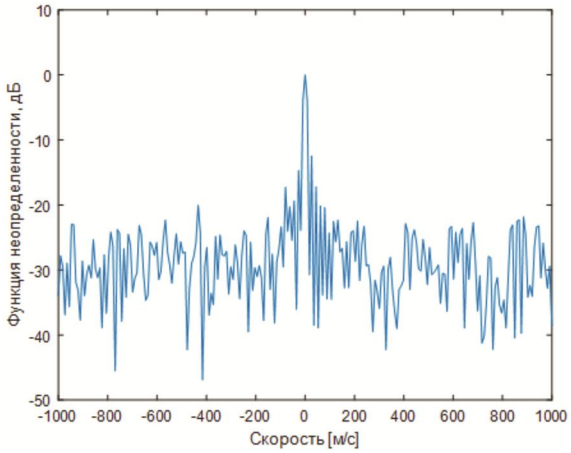


Рисунок 5 – Срез ФН по скорости при нагреве 2˚С/мин и равномерной температуре 60˚С

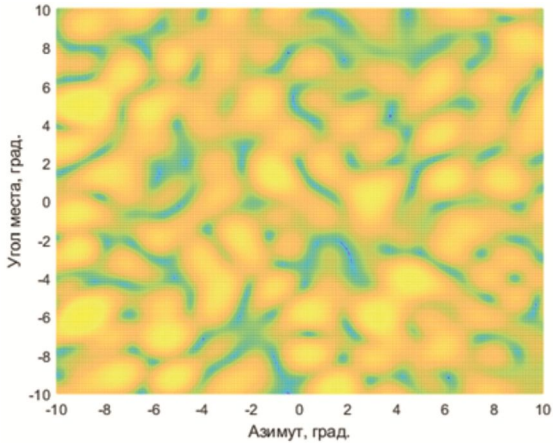


Рисунок 7 – Диаграмма направленности при нагреве 2˚С/мин и равномерной температуре 60˚С

В данном случае при перегреве МШУ происходит полное искривление диаграммы направленности, при которой невозможно определить угловые координаты космического аппарата.

Таким образом, при увеличении температуры и отсутствии калибровочных импульсов функционального контроля необходимо вводить поправки по фазе и амплитуде сигнала, позволяющие компенсировать тепловые искажения. Поправки можно вводить при обработке в части цифрового диагромообразования для уменьшения ошибок при обработке.

Заключение

В статье представлена модель искажений диаграммы направленности из-за теплового воздействия на МШУ приемного тракта РЛС для реализации режима сверхразрешения, отличающаяся от известных тем, что использовано оригинальное интегро-дифференциальное уравнение для расчета искажений ДН. Данное уравнение получено в результате формализации связи Фурье-образа температурного распределения в пространстве и времени с изменением амплитуды сигнала. Показано, что предложенная модель позволяет рассчитать харак-

теристики диаграммы направленности в широком диапазоне температур и скоростей нагрева МШУ. В дальнейших исследованиях предлагается использовать ананализ Фурье для расчета интегрального показателя эффективности режима сверхразрешения.

Проведен демонстрационный эксперимент при линейном и равномерном нагреве МШУ и показано, что после нагрев всех МШУ происходит ухудшение и полное искажение диаграммы направленности. В качестве будущих исследованиях построить алгоритм компенсации тепловых искажений на МШУ с целью увеличения разрешающей способности.

Работа выполнена при поддержке Программы развития МГУ, проект № 24-Ш01-04.

Литература

1. Мацеевич С.В., Владко У.А., Зюзина А.Д., Мочалов М.Н., Захаров А.С. Применение показателя когнитивной нагрузки графического элемента для обоснования требований к системе визуализации РЛС дальнего обнаружения. Научная визуализация, 2024 16.3: 87-96, DOI: 10.26583/sv.16.3.09

2. Шевцов В.А., Тимошенко А.В., Разиньков С.Н. Оценка состояния безопасности полета воздушного судна на основе анализа рисков авиационных инцидентов. Известия высших учебных заведений. Авиационная техника. 2024. № 1. С. 39-44. EDN GELBAJ.

3. Зверев Г.П., Тимошенко А. В., Перлов А.Ю. и др. Применение теории нечетких множеств при решении задач управления временным ресурсом радиолокационной станции мониторинга космического пространства. Научно-технический вестник информационных технологий, механики и оптики. 2024. Т. 24, № 3. С. 513-519. DOI 10.17586/2226-1494-2024-24-3-513-519. EDN STVNST.

4. D.S. Chirov and Y.A. Kochetkov. Assessment of the Accuracy of Determining the Coordinates and Speed of Small-Size UAV of a Multi-Position Radar with Omnidirectional Antenna Elements. 2020 Systems of Signals Generating and Processing in the Field of on Board Communications, Moscow, Russia, 2020, pp. 1-6. DOI: 10.1109/ IEEECONF48371.2020.9078571. EDN STFMIM.

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 поляризационном приеме. Цифровая обработка сигналов. 2023. № 1. С. 54-61. EDN VGWTVC.

6. Порсев В.И., Гелесев А.И., Красько А.Г. Угловое сверхразрешение сигналов с использованием «виртуальных» антенных решеток. Вестник Концерна ВКО «Алмаз – Антей». 2019;(4):24-34. https://doi. org/10.38013/2542-0542-2019-4-24-34

7. Чижов А.А. Сверхразрешение радиолокационных целей при воздействии активных шумовых помех по основному и ближним боковым лепесткам диаграммы направленности антенны РЛС. Информационно-управляющие системы. 2016. №1. С. 94-98. DOI: 10.15217/issn1684-8853.2016.1.88

8. Миронов В.А. Филонович А.В. К вопросу о сверхразрешении стохастических сигналов в радиолокационных системах. «Воздушно-космические силы. Теория и практика». № 16, декабрь 2020. С. 272-281.

9. Haoxuan Y., Qiangyu Z., Jianxin H., Weather Radar Image Superresolution Using a Nonlocal Residual Network. Journal of Mathematics, 2021, 4483907, 11 p., 2021.doi.org/10.1155/2021/4483907

10. Muñoz-Ferreras, J.M., Pérez-Martínez, F. Superresolution versus Motion Compensation-Based Techniques for Radar Imaging Defense Applications. EURASIP J. Adv. Signal Process. 2010, 308379 (2010). https://doi.org/10.1155/2010/308379

11. Ильчук А.Р., Каргашин Ю.Д., Меркулов В.И., Чернов В.С. Угловое сверхразрешение в бортовых радиолокационных системах воздушного базирования. Журнал радиоэлектроники [электронный журнал]. 2021. №11. https://doi.org/10.30898/1684-1719.2021.12.3.

12. Wang, Y., Zhang C., Yongshun H. etc. Space Precession Target Classification Based on Radar High-Resolution Range Profiles. International Journal of Antennas and Propagation. 2019. 1-9. 10.1155/2019/8151620.

13. Чапурский В.В. Избранные задачи теории сверхширокополосных радиолокационных систем. Москва Издательство МГТУ им. Н.Э. Баумана. 2012. 279 с

14. Zakharov A.S., Perlov A.Y., Timoshenko A.V., Sazonov V.V. Model for calculating changes in the radiation pattern and amplitude-phase distribution in the subarray of a large-aperture apaa based on a modified thermal conductivity equation. 2024 Systems of Signals Generating and Processing in the Field of on Board Communications. IEEE: 2024, pp. 1-4, Moscow, Russia. DOI:10.1109/ieeeconf60226.2024.10496770

15. Пилков А.В., Перлов А.Ю., Третьякова Т.Д., Буханец Д.И. Расчетно-аналитическая оценка влияния шумовой температуры приемного тракта на энергетический потенциал РЛС дальнего обнаружения. Электромагнитные волны и электронные системы. 2023. Т. 28, № 2. С. 36-42. DOI 10.18127/j5604128-202302-05

16. Шафир Р.С., Перлов А.Ю., Тимошенко А.В., Корпусов М.О. Моделирование тепловых процессов в высокопотенциальных радиолокационных станциях с плотной компоновкой радиоэлектронной аппаратуры с учетом неравномерного распределения температуры внутри блоков. Вычислительные технологии. 2023. Т. 28. № 5. С. 33-44.

УДК 621.396.96

МОДЕЛИРОВАНИЕ ДАЛЬНОСТНО-ДОПЛЕРОВСКИХ ОТРАЖЕНИЙ ОТ ЗЕМНОЙ ПОВЕРХНОСТИ ПРИ РАДИОЛОКАЦИОННОМ ЗОНДИРОВАНИИ С БОРТА ЛЕТАТЕЛЬНОГО АППАРАТА

Рязанцев Л.Б., *д.т.н. доцент, доцент Военного учебно-научного центра Военно-воздушных сил «Военно-воздушная академия им. проф. Н.Е. Жуковского и Ю.А. Гагарина» (г. Воронеж), e-mail: kernel386@mail.ru*
Бабич О.А., *адъюнкт Военного учебно-научного центра Военно-воздушных сил «Военно-воздушная академия им. проф. Н.Е. Жуковского и Ю.А. Гагарина» (г. Воронеж), e-mail: oleg9mufc5fan@mail.ru*
Маклаков Д.Ю., *ведущий программист АО «НПП «Радар ммс» (г. Санкт-Петербург), e-mail: maklakov_dju@radar-mms.com*

MODELING OF RANGE-DOPPLER REFLECTIONS FROM THE EARTH'S SURFACE DURING RADAR SENSING FROM AN AIRCRAFT

Ryazantsev L.B., Babich O.A., Maklakov D.Ju.

The article is devoted to the study of Range-Doppler portraits of the Earth's surface when probing with coherent signals of a radar sensor placed on board an aircraft. The patterns of portrait formation with different characteristics and orientation of antennas in space, parameters of probing radiation and reflective properties of the surface are revealed. It is proposed to use developed model to evaluate the flight parameters of the radar carrier in the interests of solving the problems of autonomous navigation.

Key words: unmanned aerial vehicle, autonomous navigation, Doppler speed meters, altitude estimation, Earth's surface sensing.

Ключевые слова: корреляционная матрица, радиотехника, разладка, собственные числа, спектр, цифровая обработка сигналов.

Статья посвящена исследованию дальностно-доплеровских портретов земной поверхности при зондировании когерентными сигналами радиолокатора, размещенного на борту летательного аппарата. Выявлены закономерности формирования портретов при различных характеристиках и ориентации антенн в пространстве, параметрах зондирующих излучений и отражательных свойств поверхности. Предложено использовать разработанную модель для оценки параметров полета носителя РЛС в интересах решения задач автономной навигации.

Введение

В последние десятилетия наметилась стойкая тенденция к использованию приемников глобальной навигационной спутниковой системы (ГНСС) для измерения параметров собственного движения беспилотных летательных аппаратов (БпЛА) и их навигации. Большая часть БпЛА не может осуществлять полет без использования таких приемников. Слабая помехозащищенность и подверженность уводящим помехам зачастую не позволяют их использование для решения навигационных задач, а существующие радионавигационные корреляционно-экстремальные системы навигации являются весьма дорогими. Кроме того, их массогабаритные характеристики не позволяют обеспечить установку на БпЛА малого класса.

Для определения высоты полета БпЛА существуют миниатюрные радиовысотомеры [1], позволяющие в отличие от барометрических высотомеров и приемников ГНСС с высокой точностью определять высоту полета носителя относительно земной поверхности. Для определения скорости полета и угла сноса летательных аппаратов применяются радиолокационные доплеровские измерители скорости и угла сноса (ДИСС) [2]. Такие измерители, как правило, строятся по принципу радиолокационных станций (РЛС) с непрерывным частотно-модулированным излучением, что позволяет снизить требования к аппаратуре и уменьшить массогабаритные характеристики. Однако для исключения влияния изменений углового положения носителя в пространстве измерители делаются трех и четырехлучевыми, что увеличивает их габариты и затрудняет установку на малоразмерные носители.

В то же время активное развитие аналоговой и цифровой микроэлектроники, появление миниатюрных СВЧ интегральных схем, реализующих принципы программно-определяемого радио, и цифровых систем на кристалле, обладающими высокими вычислительными возможностями позволило перейти к созданию РЛС с высокой разрешающей способностью, характеризующихся небольшими массой и габаритами [3, 4], и позволяющих решать задачи определения собственных координат и параметров движения носителя в интересах его автономной навигации по характеру доплеровских отражений от подстилающей поверхности.

Таким образом, целью данной работы является разработка модели дальностно-доплеровских отражений от подстилающей поверхности в интересах обоснования способов оценки собственной скорости и высоты полета беспилотного летательного аппарата с использованием установленной на него малогабаритной РЛС.

Модель сигналов, отраженных от подстилающей поверхности

Вид спектра доплеровских частот сигнала, отражённого от подстилающей поверхности зависит от таких

факторов как форма диаграммы направленности антенны (ДНА) и уровень ее боковых лепестков, тип подстилающей поверхности и углы ее облучения, скорость полета носителя и параметры сигналов РЛС [5, 6]. При зондировании подстилающей поверхности РЛС с непрерывным линейно-частотномодулированным (ЛЧМ) сигналом, установленной на воздушном носителе (рисунок 1), сигнал разностной частоты на выходе приемного устройства, представляющий результат перемножения зондирующего сигнала с отраженным от подстилающей поверхности сигналом, будет иметь вид [3]

$$\dot{S}_p(x_э, y_э, t) = A_э(\varepsilon(x_э, y_э), \beta(x_э, y_э)) \times \\ \times \exp\left(j2\pi \frac{2R_{эн}(x_э, y_э, t)}{c}(f_0 + \mu t) \right) + \dot{\eta}(t), \quad (1)$$

где $A_э$ – амплитуда сигнала, отраженного от элемента разрешения на подстилающей поверхности; $R_{эн}(x_э, y_э, t) = \sqrt{h^2 + y_э^2 + (x_э - Vt)^2}$ – наклонная дальность до элемента разрешения; h и V – высота и путевая скорость полета носителя РЛС; f_0 – начальная частота ЛЧМ сигнала; $\mu = \Delta f_c / T_м$ – скорость изменения частоты ЛЧМ сигнала; Δf_c и $T_м$ – ширина спектра и период модуляции зондирующего ЛЧМ сигнала соответственно; $x_э$ и $y_э$ – координаты элемента разрешения на подстилающей поверхности; $\dot{\eta}(t)$ – белый гауссовский шум; $t = t_м + kT_м$; $t_м \in [0, T_м)$; k – номер периода зондирования; $\varepsilon(x, y) = arctg(y/x)$ и $\beta(x, y) = arctg(\sqrt{x^2 + y^2}/h)$ – азимут и угол места соответственно.

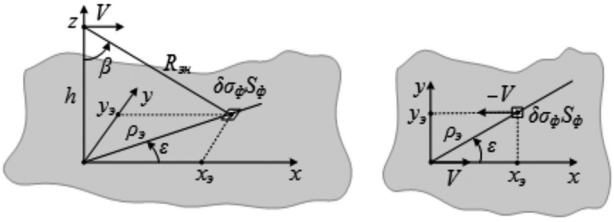


Рисунок 1 – Геометрия зондирования подстилающей поверхности

Амплитуда сигнала может быть аппроксимирована зависимостью

$$A_э(\varepsilon, \beta) = \frac{G(\varepsilon, \beta)\lambda}{R_{иэ}^2} \sqrt{\frac{P_{прд}\delta\sigma_\phi S_\phi(\varepsilon, \beta)}{(4\pi)^3 L}\cos(\beta)}, \text{ где} \quad (1)$$

$$G(\varepsilon, \beta) = G_0 f_{бл} + G_0(1 - f_{бл}) \exp\left(-2,78\left(\frac{(\varepsilon - \varepsilon_a)^2}{\Theta_\varepsilon^2} + \frac{(\beta - \beta_a)^2}{\Theta_\beta^2} \right) \right)$$

– аппроксимированная ДНА; G_0 – коэффициент усиления антенны РЛС по главному лепестку диаграммы направленности; ε_a и β_a – углы установки антенны в азимутальной и угломестной плоскости относительно вертикали; Θ_ε и Θ_β – ширина диаграммы направленности по мощности на уровне 0,5 по азимуту и углу места соответственно; $f_{бл}$ – усредненный уровень боковых лепестков относительно максимума ДНА; $P_{прд}$ – импульсная мощность передатчика; L – суммарные потери на трассе распространения и в приемно-передающем тракте РЛС; $\delta\sigma_\phi$ – удельная ЭПР фона; $S_\phi(\varepsilon, \beta)$ – площадь элемента разрешения на подстилающей поверхности; $\cos(\beta)$ – множитель, характеризующий изменение амплитуды сигнала от угла облучения земной поверхности.

Для построения зависимости амплитуды сигнала (2) в координатах «наклонная дальность-доплеровская частота» необходимо рассчитать площади элементов разрешения S_ϕ, ограниченных линиями равных наклонных дальностей и доплеровских частот (рисунок 2). При расчете площадей элементов разрешения для упрощения вычислений целесообразно учитывать симметрию линий доплеровских частот и наклонных дальностей относительно осей системы координат и ограничиться расчетами только лишь для первого квадранта с последующим использованием этих расчетов для определения площадей элементов разрешения остальных трех квадрантов.

Уравнение m-й линии равных наклонных дальностей на плоской земной поверхности в полярной системе координат описывается выражением

$$\rho_r(m, \varepsilon) = \rho_r(m) = \sqrt{(h + m\delta r)^2 - h^2}, \quad (3)$$

где $\delta r = c/(2\Delta f_c)$ – разрешающая способность РЛС по наклонной дальности; c – скорость света; Δf_c – ширина

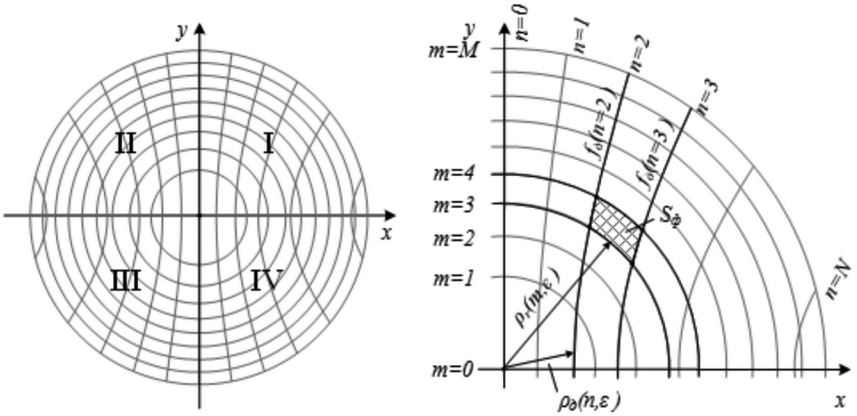


Рисунок 2 – Иллюстрация формы линий равных наклонных дальностей и доплеровских частот на плоской земной поверхности

спектра зондирующего сигнала РЛС; $m = 0,1,\ldots,M$; $M = \text{int}[R_{\max}/\delta r]$; $\text{int}[\cdot]$ – оператор выделения целой части; $R_{\max}$ – максимальная дальность действия РЛС, определяемая ее энергетическими и конструктивными характеристиками.

Исходя из того, что характер изменения доплеровских частот в зоне обзора определяется проекцией максимальной доплеровской частоты на линию визирования

$$f_д(n) = f_{д\max}\sin(\beta(n))\cos(\varepsilon), \quad (4)$$

где $\beta(n) = arc\,\text{t}\,g(\rho_д(n)/h)$, то уравнение для n-й линии равных доплеровских частот на плоской поверхности в полярной системе координат может быть получено из уравнения (4) и представлено в виде

$$\rho_д(n,\varepsilon) = \frac{f_д(n)h}{\sqrt{f^2_{д\max}\cos^2(\varepsilon) - f^2_д(n)}},\ |\varepsilon| < \arccos\left(\frac{f_д(n)}{f_{д\max}}\right), \quad (5)$$

где $f_{д\max} = 2V/\lambda$ – максимальное значение доплеровской частоты; $f_д(n) = n\,\delta f_д$; $n = 0,1,\ldots,N$; $N = \text{int}[f_{д\max}/\delta f_д]$; $\delta f_д = 1/T_c$ – разрешение РЛС по доплеровской частоте; T_c – время когерентного накопления сигналов.

Для нахождения площади элемента разрешения можно воспользоваться интегрированием по области, ограниченной кривыми равных наклонных дальностей и доплеровских частот. Для этого необходимо найти точки пересечения этих кривых. Расчет координат точек пересечения между n-й линией равных доплеровских частот и m-й линией равных наклонных дальностей может быть осуществлен по выражениям

$$\varepsilon(m,n) = \arccos\left(\frac{f_д(n)}{f_{д\max}}\frac{(h+m\delta r}{\sqrt{(h+m\delta r)^2 - h^2}}\right), \quad (6)$$

$$\beta(m,n) = \arccos\left(\frac{h}{h+m\delta r}\right), \quad (7)$$

$$\rho(m,n) = \sqrt{2hm\delta r + m^2\delta r^2}. \quad (8)$$

Выражения (6) и (7) могут быть использованы для пересчета индексов m и n в координаты цилиндрической системы координат.

Анализ формы фигур элементов разрешения (рисунок 2) показывает наличие пяти различных вариантов, которые характеризуются различным количеством (двух, четырех и шести) точек пересечения. Варианты фигур, соответствующих пяти типам элементов разрешения, представлены на рисунке 3. Для корректного расчета амплитуд отражений в зоне обзора необходимо также учитывать и шестой вариант, при котором отсутствуют точки пересечения. Это соответствует нулевой площади поверхности и, как следствие, отсутствию сигналов из этой области пространства.

Для реализации алгоритма автоматического расчета площади элементов разрешения необходимо определить признаки, на основе которых будет осуществляться нахождение координат точек пересечения и выбор дальнейшего порядка расчета площади каждого $S_ф(m,n)$-го элемента разрешения в зоне обзора. В качестве таких признаков можно рассматривать наличие пересечений между кривыми $\rho_r(m)$ и $f_д(n)$, $\rho_r(m)$ и $f_д(n+1)$, $\rho_r(m-1)$ и $f_д(n+1)$, $\rho_r(m-1)$ и $f_д(n)$. В таблице 1 приведены логические условия пересечений кривых, по совместному выполнению которых осуществляется определение типа поверхности элемента разрешения и выбор выражений для вычисления элементов вектора координат точек пересечений $\mathbf{P}_{m,n} = \{p_0(\varepsilon_0,\rho_0), p_1(\varepsilon_1,\rho_1),\ldots\}$ для дальнейшего вычисления площади элементов разрешения.

Сформированный вектор $\mathbf{P}_{m,n}$ позволяет найти площадь элемента разрешения через интегрирование по области $R_{m,n}$ [7], заданной точками, являющимися компонентами этого вектора, и ограниченной кривыми равных доплеровских частот и наклонных дальностей

$$S_ф(m,n) = \iint_{R_{m,n}} \rho\alpha_{m,n}(\varepsilon)d\rho d\varepsilon, \quad (9)$$

где в качестве функции $\alpha_{m,n}(\varepsilon)$ в процессе интегрирования выбирается одна из двух функций $\rho_r(m)$ или $\rho_д(n,\varepsilon)$ в зависимости от пределов интегрирования и типа поверхности элемента разрешения.

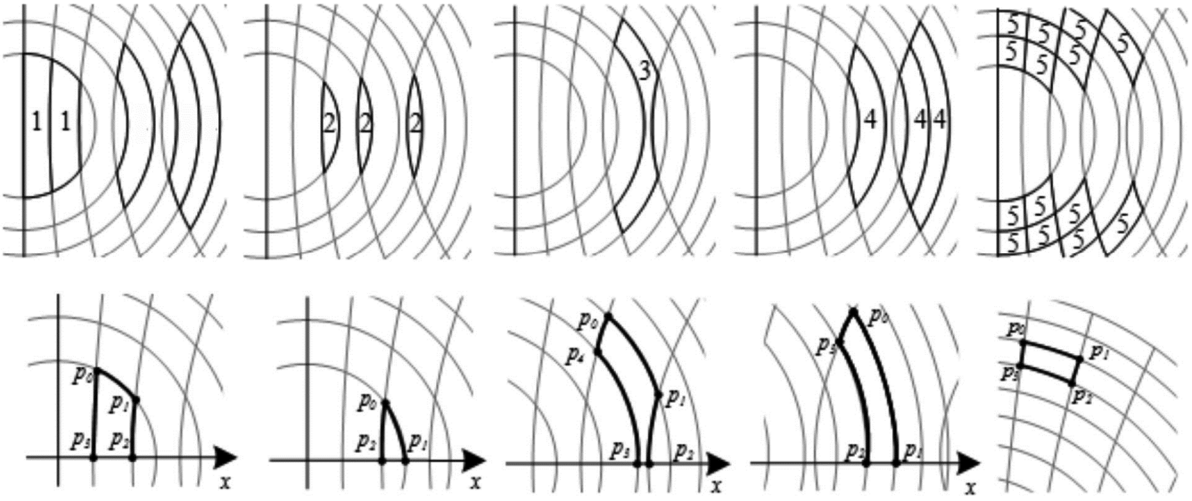


Рисунок 3 – Варианты фигур элементов разрешения и расположение точек пересечения линий для последующего вычисления площади

Таблица 1 – Логические условия для вычисления точек пересечения

Тип поверхности	Условия пересечений	Координаты точек
1	$\rho_r(m) \cap f_\partial(n) = True$, $\rho_r(m) \cap f_\partial(n+1) = True$, $m \geq 0$	$\varepsilon_0 = \varepsilon(m,n)$, $\rho_0 = \rho(m,n)$ $\varepsilon_1 = \varepsilon(m,n+1)$, $\rho_1 = \rho(m,n+1)$ $\varepsilon_2 = 0$, $\rho_2 = \rho_\partial(n+1, \varepsilon = 0)$ $\varepsilon_3 = 0$, $\rho_3 = \rho_\partial(n, \varepsilon = 0)$
2	$\rho_r(m) \cap f_\partial(n) = True$, $\rho_r(m) \cap f_\partial(n+1) = True$, $\rho_r(m-1) \cap f_\partial(n+1) = False$, $\rho_r(m-1) \cap f_\partial(n) = False$	$\varepsilon_0 = \varepsilon(m,n)$, $\rho_0 = \rho(m,n)$ $\varepsilon_1 = 0$, $\rho_1 = \rho_0$ $\varepsilon_2 = 0$, $\rho_3 = \rho_\partial(n, \varepsilon = 0)$
3	$\rho_r(m) \cap f_\partial(n) = True$, $\rho_r(m) \cap f_\partial(n+1) = True$, $\rho_r(m-1) \cap f_\partial(n+1) = False$, $\rho_r(m-1) \cap f_\partial(n) = True$	$\varepsilon_0 = \varepsilon(m,n)$, $\rho_0 = \rho(m,n)$ $\varepsilon_1 = \varepsilon(m,n+1)$, $\rho_1 = \rho(m,n+1)$ $\varepsilon_2 = 0$, $\rho_2 = \rho_\partial(n+1, \varepsilon = 0)$ $\varepsilon_3 = 0$, $\rho_3 = \rho(m-1,n)$ $\varepsilon_4 = \varepsilon(m-1,n)$, $\rho_4 = \rho(m-1,n)$
4	$\rho_r(m) \cap f_\partial(n) = True$, $\rho_r(m) \cap f_\partial(n+1) = False$, $\rho_r(m-1) \cap f_\partial(n+1) = False$, $\rho_r(m-1) \cap f_\partial(n) = True$	$\varepsilon_0 = \varepsilon(n,m)$, $\rho_0 = \rho(n,m)$ $\varepsilon_1 = 0$, $\rho_1 = \rho_0$ $\varepsilon_2 = 0$, $\rho_2 = \rho_3$ $\varepsilon_3 = \varepsilon(n,m-1)$, $\rho_3 = \rho(n,m-1)$
5	$\rho_r(m) \cap f_\partial(n) = True$, $\rho_r(m) \cap f_\partial(n+1) = True$, $\rho_r(m-1) \cap f_\partial(n+1) = True$, $\rho_r(m-1) \cap f_\partial(n) = True$	$\varepsilon_0 = \varepsilon(n,m)$, $\rho_0 = \rho(n,m)$ $\varepsilon_1 = 0$, $\rho_1 = \rho(n+1,m)$ $\varepsilon_2 = \varepsilon(n+1,m-1)$, $\rho_2 = \rho(n+1,m-1)$ $\varepsilon_3 = \varepsilon(n,m-1)$, $\rho_3 = \rho(n,m-1)$
6	$\rho_r(m) \cap f_\partial(n) = False$	$\mathbf{P}_{m,n} = \varnothing$

Порядок расчета амплитуд отраженных сигналов

Учитывая, что количество элементов разрешения при облучении земной поверхности современными радиолокаторами может достигать десятков и сотен тысяч, то целесообразно рассмотреть порядок расчета амплитуды отражений для его реализации на ЭВМ с использованием специализированных программ или программных пакетов математического моделирования. Структурно порядок расчета может быть представлен в виде схемы, изображенной на рисунке 4.

Нахождение интеграла (9) может быть основано на применении формулы площади Гаусса (алгоритма «шнурования») [8], при котором фигуры поверхностей представляются в форме простых многоугольников, вершины которых заданы точками, являющихся компонентами вектора $\mathbf{P}_{m,n}$. В этом случае

$$S_\phi(m,n) = \frac{1}{2}\sum_{i=1}^{P}\left(x_i y_{i+1} - x_{i+1} y_i\right), \qquad (10)$$

где P – количество компонентов вектора $\mathbf{P}_{m,n}$; $x_P = x_0$; $y_P = y_0$; $x_{P+1} = x_1$; $y_{P+1} = y_1$; $x_i = \rho_i \cos(\varepsilon_i)$; $y_i = \rho_i \sin(\varepsilon_i)$.

На рисунке 5 представлен результат вычисления площадей элементов разрешения в соответствии с представленным на рисунке 4 порядком при следующих исходных данных: $h = 500$ м; $R_{\max} = 1$ км; $\delta r = 20$ м; $\delta f_\partial = 20$ Гц; $V = 120$ км/ч; $\lambda = 5{,}5$ см.

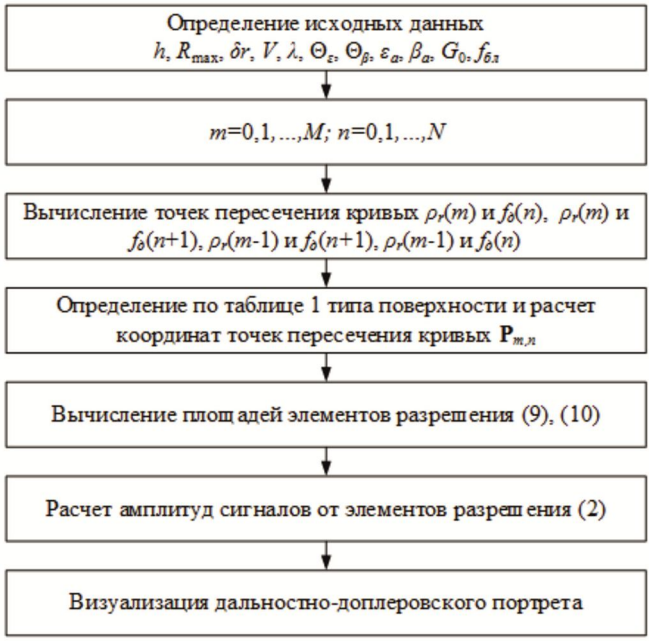


Рисунок 4 – Порядок расчета амплитуд сигналов в координатах «наклонная дальность – доплеровская частота»

Анализ полученных результатов показывает следующее.

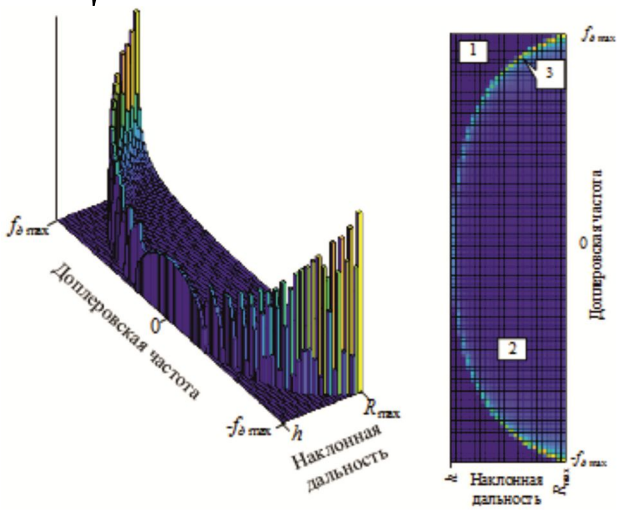


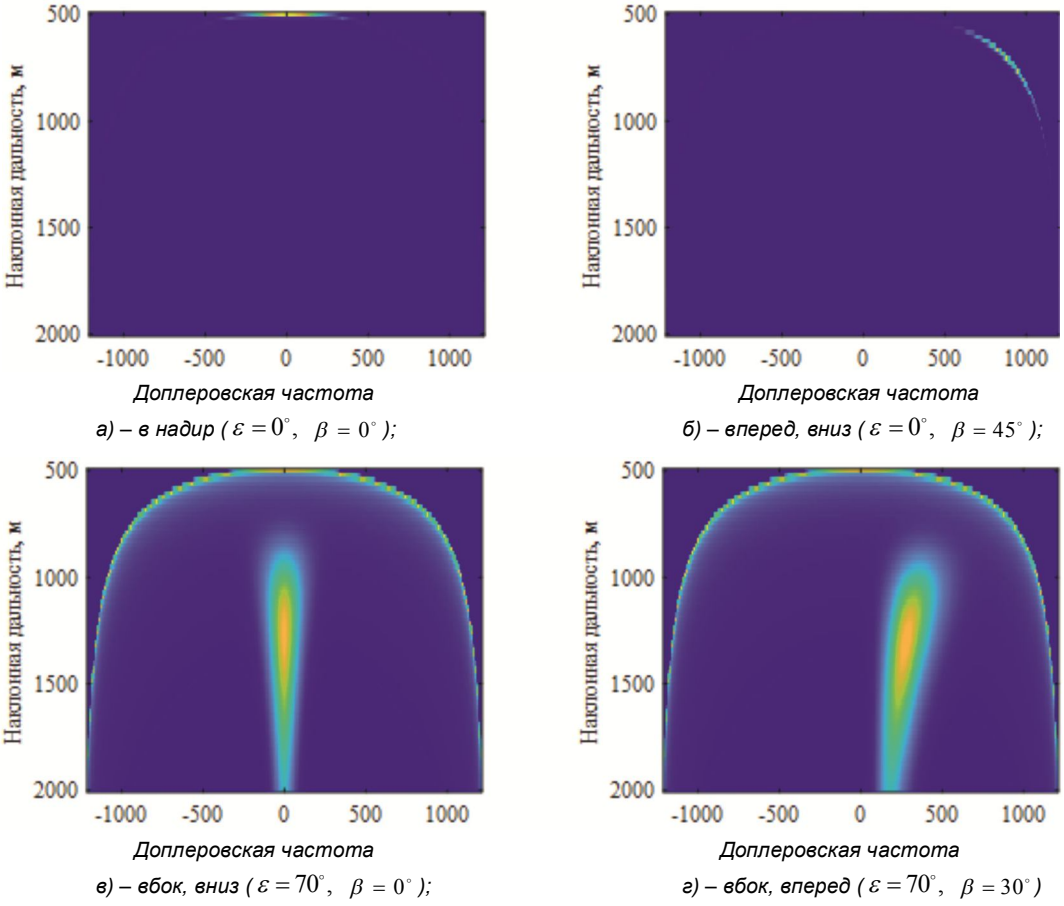
Рисунок 5 – Результат расчета площадей элементов разрешения

Первое. Площадь элементов разрешения с увеличением наклонной дальности возрастает (рисунок 5). Однако учитывая, что амплитуда сигналов убывает пропорциональна квадрату наклонной дальности и значению $\cos(\beta)$, то амплитуда отражений с увеличением наклонной дальности достаточно быстро снижается.

Второе. Дальностно-доплеровский портрет имеет три характерные области (рисунок 5). В области 1 отсутствуют отражения от подстилающей поверхности. Эта область соответствуют наклонным дальностям, величина которых не превышает расстояния до подстилающей поверхности. Для элементов разрешения этой области отсутствуют пересечения линий доплеровских частот и наклонных дальностей. Область 2 представлена пятым типом поверхности (рисунок 3) и имеет наибольшее количество элементов разрешения. Область 3 представляет собой границу «раздела сред» и характеризуется мощными всплесками амплитуд сигналов. Отражения в этой области имеют значительный уровень даже при приеме по боковым лепесткам диаграммы направленности антенны (рисунки 6, *в*, *г*). Данная граница формируется отражениями от элементов разрешения первых четырех типов. Именно поэтому при разработке модели доплеровских отражений описанию таких элементов разрешения было уделено наибольшее внимание. Выражение для формы границы «раздела сред» может быть получено из выражения (4) при условии $\varepsilon = 0,$ т.е.

$$R(f_\partial) = h \Big/ \sqrt{1-\left(\frac{\lambda f_\partial}{2V}\right)^2}, \quad f_\partial \le f_{\partial\max}. \qquad (11)$$

На рисунке 7 представлены полученные на основе (11) зависимости $R(f_\partial)$ при различных V. Анализ зависимостей показывает, что с увеличением скорости носителя форма кривой расширяется по координате доплеровской частоты, при этом ее положение по наклонной дальности не изменяется. Это свойство может служить информативным признаком для определения путевой скорости полета носителя по форме отражений от «границы сред». Крутизна формы кривой увеличивается с увеличением доплеровской частоты. Поэтому для более точной оценки путевой скорости по ее форме целесообразно использовать участки с большей крутизной.



а) – в надир ($\varepsilon = 0^\circ,\ \beta = 0^\circ$);

б) – вперед, вниз ($\varepsilon = 0^\circ,\ \beta = 45^\circ$);

в) – вбок, вниз ($\varepsilon = 70^\circ,\ \beta = 0^\circ$);

г) – вбок, вперед ($\varepsilon = 70^\circ,\ \beta = 30^\circ$)

Рисунок 6 – Иллюстрация зависимости амплитуды сигнала в координатах «наклонная дальность-доплеровская частота»

при различных ориентациях антенн

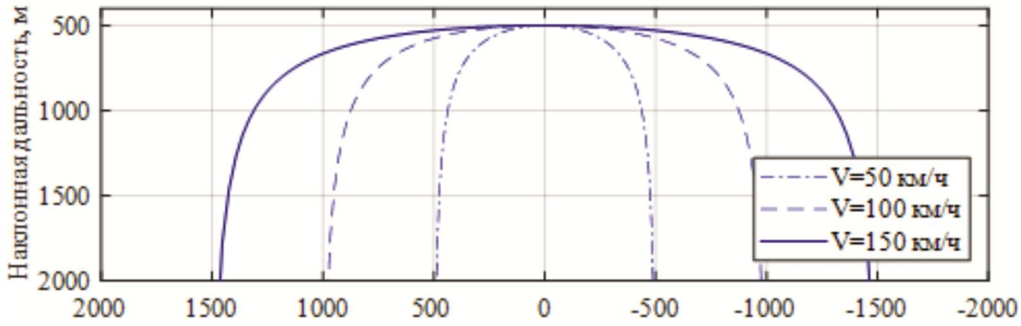


Рисунок 7 – Зависимости $R(f_д)$ при различных V

Третье. Форма и ориентация диаграммы направленности определяет вид доплеровских отражений, который может использоваться для извлечения различных признаков об ориентации и параметрах движения летательного аппарата относительно подстилающей поверхности. Так, на рисунке 6 представлены иллюстрации зависимости амплитуды сигнала в координатах «наклонная дальность-доплеровская частота» при различных ориентациях антенн. Моделирование дальностно-доплеровских портретов осуществлено при следующих условиях: $h = 500$ м; $R_{\max} = 1$ км; $\delta r = 20$ м; $\delta f_д = 5$ Гц; $V = 120$ км/ч; $\lambda = 5{,}5$ см; $G_0 = 1$; $f_{бл} = -13$ дБ; $\Theta_\varepsilon = \Theta_\beta = 20^\circ$.

Вид дальностно-доплеровского портрета, представленного на рисунке 6, *а* соответствует ориентации антенн в надир. Такая ориентация может быть использована для измерения высоты полета носителя и оценки его путевой скорости по форме кривой доплеровских частот. Рисунок 6, *б* соответствует ориентации антенн с установкой под углом 45 градусов относительно вертикали вперед по направлению движения носителя. Такая ориентация антенн может использоваться для измерения скорости полета носителя по значению средней доплеровской частоты спектра отраженных от поверхности сигналов. Рисунок 6, *в* соответствует установке антенн, характерного для режима синтезирования апертуры антенны при боковом обзоре, а рисунок 6, *г* – для режима синтезирования апертуры при переднебоковом обзоре или при наличии угла сноса.

Результаты экспериментальной проверки

Для подтверждения полученных результатов проведена экспериментальная проверка на образце малогабаритной РЛС С-диапазона «Квазар», разработанной АО «НПП «Радар ммс» при научно-техническом сопровождении ВУНЦ ВВС «Военно-воздушная академия имени проф. Н.Е. Жуковского и Ю.А. Гагарина» (г. Воронеж).

При проведении эксперимента характеристики РЛС и параметры ее зондирующего сигнала были следующие: ширина диаграммы направленности антенн – 20х20 град., ширина спектра сигнала – 300 МГц, период повторения – 1,5 мс, время когерентного накопления сигналов – 0,2 с. РЛС устанавливалась на беспилотном вертолете БВС ВТ-45 (рисунок 8), при этом антенны ориентировались в надир. Полет вертолета осуществлялся на высоте около 100…130 м над земной поверхностью со скоростью 54 км/ч.

Дальностно-доплеровский портрет, сформированный в процессе полета, представлен на рисунки 9, *а* и *б*. Изображение на рисунке 9, *а* соответствует участку местности с ровной песчаной поверхностью. Рисунок 9, *б* отображает участок местности с лесным массивом и неровностями рельефа. Анализ кривизны дальностно-доплеровского портрета, представленного на рисунок 9, *а* показывает его соответствие с кривой, рассчитанной по выражению (11) для условий $V = 54$ км/ч и $h = 112$ м, что подтверждает соответствие практических результатов теоретическим зависимостям. Маршрут полета и вид земной поверхности показан на рисунке 9, *в*.

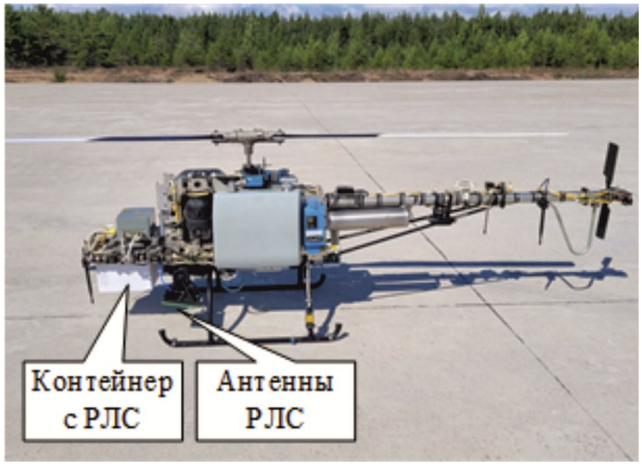


Рис. 8. Вид носителя БВС ВТ-45 (АО «НПП «Радар ммс») с установленной РЛС

Заключение

Таким образом, разработанная модель позволяет рассчитать вид доплеровских отражений от подстилающей поверхности для различных параметров полета носителя РЛС (скорость и высота), условий облучения поверхности (ориентация диаграммы направленности антенны в пространстве и ее форма), характеристиках фона (удельная ЭПР), разрешающей способности РЛС и времени когерентного накопления сигналов. Использование разработанной модели в дальнейшем позволит провести обоснование способов обработки сигналов в когерентных малогабаритных РЛС для получения оценок высоты, путевой скорости и углов сноса в интересах решения задач автономной навигации малоразмерных беспилотных летательных аппаратов.

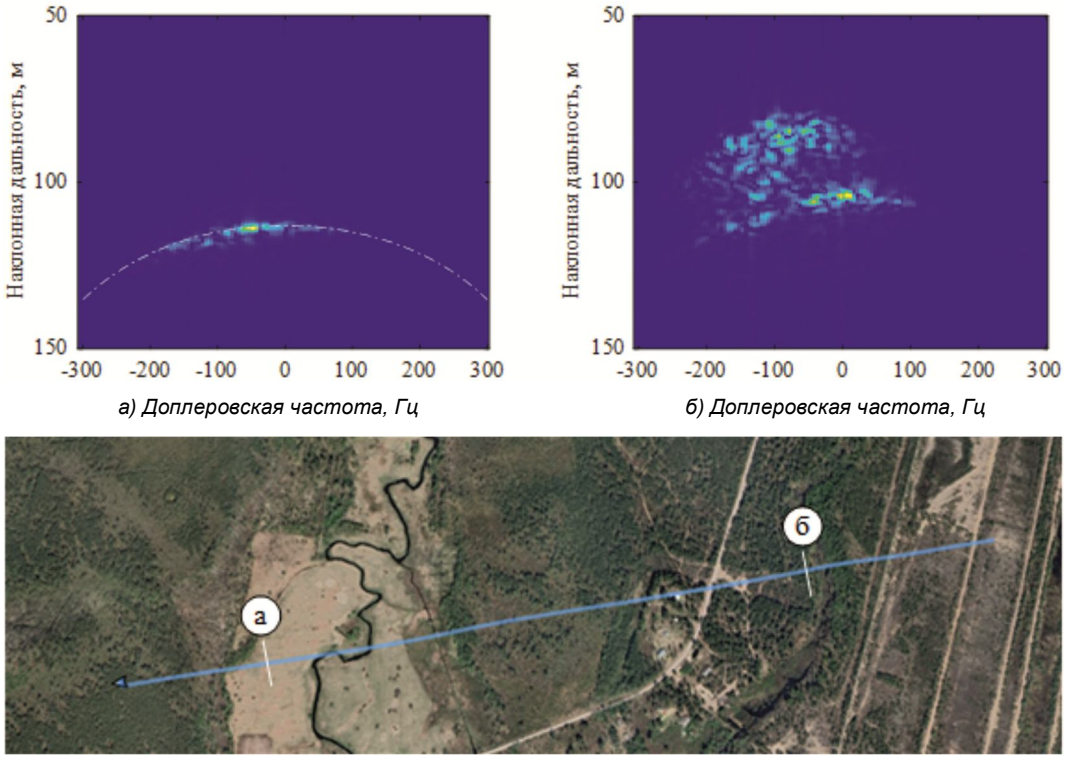


Рисунок 9 – Вид дальностно-доплеровских портретов (а и б) и соответствующие им участки местности (в)

Литература

1. 24GHz Alimeter Radar NRA24 [Электронный ресурс]. http://en.nanoradar.cn/Article/detail/id/372.html (дата обращения: 04.02.2024).

2. Колчинский В.Е., Мандуровский И.А., Константиновский М.И. Автономные допплеровские устройства и системы навигации летательных аппаратов. М.: «Советское радио», 1975. 430 с.

3. Купряшкин И.Ф., Лихачев В.П., Рязанцев Л.Б. Малогабаритные многофункциональные РЛС с непрерывным частотно-модулированным излучением. М.: Радиотехника, 2020. 288 с.

4. Иванников К.С., Купряшкин И.Ф., Рязанцев Л.Б. Алгоритм формирования профиля рельефа местности с использованием малогабаритной РЛС с синтезированной апертурой антенны высокого разрешения. Цифровая обработка сигналов, 2019. № 1. С.31-36.

5. Дудник П.И., Ильчук А.Р., Татарский Б.Г. Многофункциональные радиолокационные системы. М.: Дрофа, 2007. 283 с.

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

7. George B. Thomas, Jr. Thomas' Calculus Early Transcendentals. Pearson Education, 2014. 1206 pp.

8. Rhoad R., Milauskas G., Whipple R. Geometry for Enjoyment and Challenge. McDougal Littell, 1991, pp. 717-718.

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.

УДК 621.391.82

ОБНАРУЖЕНИЕ PRACH-СИГНАЛА СТАНДАРТА GSM НА ФОНЕ ПОМЕХ

Рогожников Е.В., *к.т.н., заведующий кафедрой ТУСУР, e-mail: evgenii.v.rogozhnikov@tusur.ru*
Савенко К.В., *ТУСУР, младший научный сотрудник, e-mail: kirill.savenko@tusur.ru*
Дмитриев Э.М., *к.т.н., ТУСУР, старший преподаватель, e-mail: edgar.dmitriev@tusur.ru*
Иванников Н.А., *НТИ ЦКБСИВ, м.н.с., e-mail: ivannikoy0201@mail.ru*
Порваткин Р.А., *НТИ ЦКБСИВ, м.н.с, e-mail: porvatkin2001@gmail.com*
Гридин И.В., *НТИ ЦКБСИВ, м.н.с., e-mail: vanyagridin2001@bk.ru*
Жапова Д.Ж., *НТИ ЦКБСИВ, м.н.с., e-mail: dragondarima2001@mail.ru*

DETECTION OF GSM STANDARD SIGNAL IN THE BACKGROUND OF INTERFERENCE

Rogozhnikov E.V., Savenko K.V., Dmitriyev E., Ivannikov N.A., Porvatkin R.A., Gridin I.V., Zhapova D.Z.

The task of detecting signals against a background of noise remains one of the key problems in the field of digital communications. Various methods of correlation processing are used to solve this problem. The paper considers three types of correlation receivers: classical, iconic, and a receiver that calculates a cross-correlation function in the frequency domain. In order to study correlation receivers, a simulation was performed in the MATLAB programming environment and the probability of correct detection was compared with different signal-to-noise ratios for each approach.

Key words: correlation, correlation receiver, GSM, PRACH, adaptive threshold, DSP, FPGA.

Ключевые слова: корреляция, корреляционный приемник, GSM, PRACH, адаптивный порог, DSP, ПЛИС.

Задача обнаружения сигналов на фоне шумов остается одной из ключевых проблем в области цифровых коммуникаций. Для решения данной задачи используются различные методы корреляционной обработки. В работе рассматриваются три типа корреляционных приемников: классический, знаковый, а также приемник, вычисляющий взаимную корреляционную функцию в частотной области. С целью исследования корреляционных приёмников было проведено моделирование в среде программирования MATLAB и проведено сравнение вероятности правильного обнаружения при различных отношениях сигнал-шум для каждого из подходов.

Введение

В рамках научно-исследовательской работы «Разработка программно-аппаратного комплекса для поиска в лесном массиве потерявшегося человека с устройством мобильной связи», проводимой молодежной проектной командой Передовой Инженерной Школы «Электронное приборостроение и системы связи» имени А.В. Кобзева исследуется способ использования беспилотных летательных аппаратов (БПЛА) с базовой станцией стандарта GSM 900 на борту для помощи поисковым службам. Суть данного способа заключается в следующем: БПЛА пролетает над лесным массивом, позволяя мобильным устройствам, не зарегистрированным в сети, подключиться к этой временной базовой станции [1]. Чтобы определить наличие абонента и зарегистрировать его, вычисляется корреляция между полученным сигналом и известным сигналом PRACH (Physical Random Access Channel, физический канал случайного доступа). Это позволяет эффективно обнаружить потерявшихся людей в лесном массиве через их мобильные устройства связи.

Дополнительное обнаружение PRACH сигнала посредством ПЛИС (программируемая логическая интегральная схема) используется для предотвращения ситуаций, когда мобильное устройство отправляет запрос на подключение, но не получает ответа. Если БПЛА пролетел, а абонент не успел подключиться к сети при детектировании PRACH сигнала на физическом уровне, БПЛА замедляется или возвращается обратно и повторно проверяет связь с абонентом. Это значительно повышает шансы на успешное установление контакта с потерявшимся человеком.

Как известно, системы мобильной связи функционируют в условиях воздействия разнообразных мешающих факторов, таких как шумы, многолучевое распространение и другие виды помех. Эти факторы снижают качество принимаемого сигнала, ухудшая общую эффективность обнаружения сигнала PRACH. Устранение помех является одной из главных задач, в которой применяются методы, основанные на различии структуры опорного сигнала и шумовых компонентов.

В работе рассматривалось решение задачи посредством корреляционного приёмника.

Корреляционный приемник сравнивает принятый сигнал с заранее известным опорным сигналом. Сравнение осуществляется путем вычисления коэффициента взаимной корреляции $R(t)$ принятого сигнала $s(t)$ и опорного сигнала $s_p(t)$ за время передачи одного символа T:

$$R(t) = \int_0^T [s(t) \cdot s_p(t) dt]. \quad (1)$$

Хотя математическое выражение выглядит простым,

реализация алгоритма в аппаратном обеспечении сопряжена с определенными трудностями. Согласно выражению (1) необходимо выполнять операцию умножения, которая требует значительного количества специализированных блоков умножения или большого объема логических элементов в случае использования ПЛИС, что увеличивает затраты на оборудование и снижает его энергоэффективность. Для решения задачи в ПЛИС применяются альтернативные алгоритмы корреляционного анализа, требующие меньшего количества аппаратных ресурсов. В данном исследовании рассматриваются их ключевые достоинства и недостатки.

Описание канала PRACH

В современном мире мобильной связи стандарт GSM 900 играет важную роль в обеспечении надежной и эффективной передачи данных. Одним из аспектов стандарта является использование общего транспортного физического канала в восходящей линии связи, известного как PRACH, который используется для инициализации соединения между мобильным телефоном и базовой станцией. Канал предназначен для передачи коротких контрольных сообщений, таких как запросы на соединение или индикация готовности к передаче данных. PRACH канал используется для установления начального соединения между мобильным устройством и сетью, что позволяет мобильным устройствам получить доступ к базовой станции для передачи и приема данных. Следует отметить, что в данной статье под выражением «PRACH сигнал» подразумевается совокупность сигналов, которые передаются в PRACH канале.

Процесс обработки сигнала в этом канале представлен на рисунке 1.

Блок 1 в данной схеме содержит в себе 8 информационных бит и 6 контрольных бит, блок 2 отвечает за коррекцию ошибок (перемежение), блок 3, используя сверточный код, кодирует всю последовательность с кодовой скоростью 1/2; в блоке 4 производится формирования пакета; блок 5 отвечает за модуляцию [2]-[4].

Сообщения PRACH могут быть переданы в нулевом слоте каждого кадра в пределах мультикадра, согласно структуре, представленной на рисунке 2. Так как длительность одного кадра составляет 4,615 мс, то длительность мультикадра равна: $4,615 \cdot 51 = 235,365$ мс.

Таким образом сообщение канала PRACH передается мобильной станцией один раз в 235,365 мс.



Рисунок 1 – Процесс обработки сигнала в PRACH канале

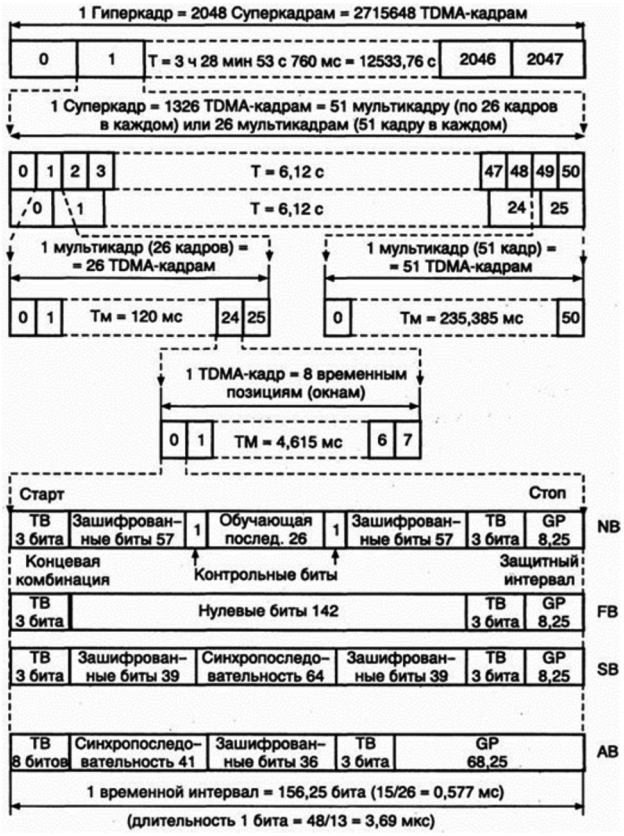


Рисунок 2 – Структура кадра стандарта GSM 900

Описание методов корреляции

Наиболее часто встречающимся видом корреляционных приемников является корреляционный приемник, который основан на умножении входного сигнала на отсчеты опорного сигнала. Этот приемник принято называть классическим. Структурная схема подобного корреляционного приемника при реализации на ПЛИС представлена на рисунке 3 [6].

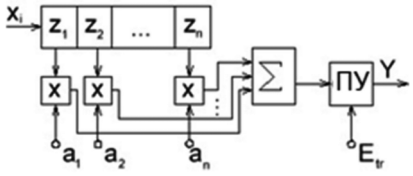


Рисунок 3 – Классический корреляционный приемник

Отчеты сигнала x_i подаются на вход сдвигового регистра z. Затем каждый отчет регистра z умножается на соответствующие отсчеты PRACH сигнала a_n. Результаты этих произведений направляются на вход сумматора. После этого вычисленное значение передается на пороговое устройство (ПУ), которое принимает решение о наличии или отсутствии сигнала в каждый такт работы устройства.

Схема реализует расчет корреляции без упрощений (согласно формуле 1), благодаря чему такой тип приемника отличается высокой точностью. Однако, как видно на рисунке 3, он требует большого количества блоков умножения, зависящего от размера преамбулы, что существенно увеличивает требования к аппаратной части. Для данного способа преамбула (PRACH) имеет длину 54 значения, разрядностью по 4 бита каждое.

Для реализации данного метода корреляционного приемника на ПЛИС требуется 208 четырехразрядных регистров и 54 байта памяти для хранения преамбулы, также для дальнейших расчетов понадобиться 209 DSP блоков.

Также возможна реализация корреляционного приемника без применения блоков умножения. Схема такого приемника приведена на рисунке 4.

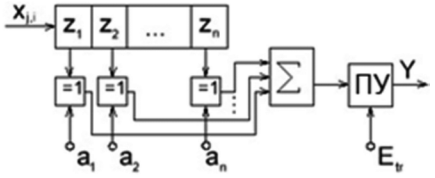


Рисунок 4 – Знаковый корреляционный приемник

Как показано на рисунке 4, на вход знакового корреляционного приемника поступают знаковые разряды отчетов сигнала $x_{j,i}$. Затем выполняется сравнение знаковых отчетов PRACH сигнала a_n с соответствующими значениями регистра z, результаты этого сравнения поступают на вход сумматора. Подобно классическому корреляционному приемнику, итог суммы передается на ПУ, где, исходя из установленного порога E_{tr}, принимается решение о наличие или отсутствии сигнала PRACH.

Такое упрощение приводит к существенному снижению точности обнаружения корреляционного пика, что может вызывать ошибки обнаружения. Тем не менее, этот подход позволяет реализовать корреляционный приемник на более доступной аппаратной плат форме, что особенно полезно в условиях ограниченных ресурсов и невысоких требований к точности.

Для реализации этого приемника на ПЛИС требуется 208 регистров для записи только знаковой части преамбулы и 14 байт памяти, кроме того, для вычислений не требуются DSP блоки.

Так же имеется корреляционный приемник, построенный по структурной схеме, которая представлена на рисунке 5 [7]. Отличительной особенностью данного корреляционного приемника является вычисление взаимной корреляционной функции не во временной, а в частотной области.

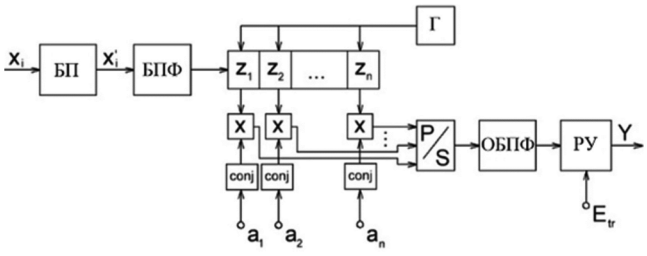


Рисунок 5 – Структурная схема корреляционного приемника, работающего в частотной области

На вход корреляционного приемника (рисунок 5) поступают отсчеты входного сигнала x_i, которые накапливаются в буферной памяти (БП), далее производится операция быстрого преобразования Фурье (БПФ). Полученные значения поступают в ячейки z_i сдвигового регистра z. Значение каждой ячейки перемножается с комплексно-сопряженным значением a_i^* отсчетов опорного сигнала a_i. В отличие от корреляционного приемника, построенного по схеме, которая представлена на рисунке 3, отсчеты a_i являются представлением опорного сигнала в частотной области. Далее происходит параллельно-последовательное преобразование результатов перемножения, и данные отсчеты поступают на вход блока обратного БПФ. Значения с выхода данного блока поступают на решающее устройство, в котором принимается решение о наличии или отсутствии сигнала на входе корреляционного приемника.

$$Y = \begin{cases} 1, \sum_{k=0}^{N-1} (a_i^* \cdot z_i) \cdot e^{\frac{j2\pi nk}{N}} \geq E_{tr} \\ 0, \sum_{k=0}^{N-1} (a_i^* \cdot z_i) \cdot e^{\frac{j2\pi nk}{N}} < E_{tr} \end{cases}. \quad (2)$$

Недостатком данного корреляционного приемника является большая задержка при вычислении взаимной корреляционной функции и, в связи с этим, необходимость применения буферной памяти. Её размер определяется количеством отсчетов в БПФ. Также наличие блоков прямого и обратного преобразования Фурье увеличивает количество используемых DSP блоков до 256, 208 четырехразрядных регистров, а для хранения преамбулы и буфера входных значений используется и 208 байт памяти.

Моделирование корреляционных приёмников

Для ПУ был реализован алгоритм, который определяет порог, позволяющий уменьшить количество неверных срабатываний коррелятора, а также позволяет динамически подстраиваться к уровню входного сигнала. Суть алгоритма определения порога заключается в следующем: фильтр рассчитывает значение порога как среднее значение корреляции $y(n)$, полученное путем суммирования всех значений корреляций в каждый момент времени x_{n-i} и деления результата на размер фильтра M, а после полученное значение умножается на коэффициент W, который был получен экспериментальным путем:

$$y(n) = W \cdot \left(\frac{1}{M} \sum_{i=0}^{M-1} x(n-i) \right). \quad (3)$$

Для определения порогов исследуемых корреляторов была создана модель, структурная схема которой представленная на рисунке 6.

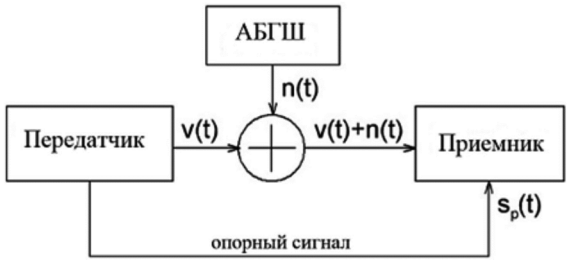


Рисунок 6 – Структурная схема модели исследования

Параметры сигналов, используемых при моделировании, имеют следующие значения: диапазон отношений сигнал-шум от -5 дБ до 5 дБ, количество PRACH

сигналов, переданных при каждом отношении сигнал шум – 10000, размер опорного сигнала – 54 отсчета.

Результат расчета корреляции с адаптивным порогом представлен на рисунке 7.

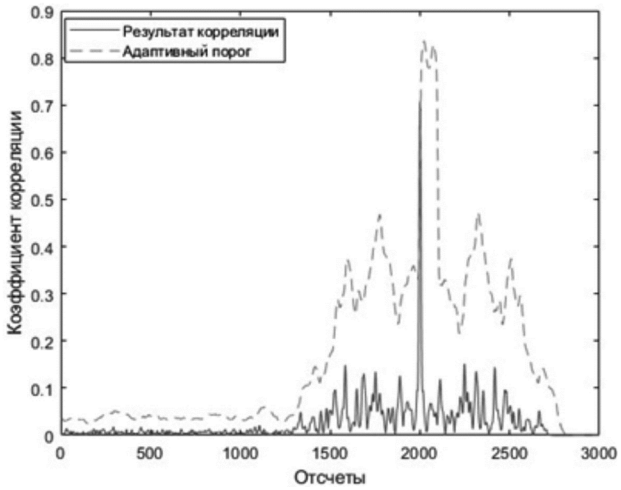


Рисунок 7 – Результат пика корреляции относительно адаптивного порога

В передатчике была сформирована последовательность из N PRACH сигналов, а для передачи этой последовательности использовался канал с АБГШ (аддитивный белый гауссовский шум). При этом последовательность была сформирована таким образом, что события наличия на входе приемника суммы PRACH сигнала с шумом либо только шума были равновероятны. В приемнике выносились решения о наличии или отсутствии сигнала при разных значениях порога, затем данные решения сравнивались с истинными. В результате для каждого отношения сигнал шум были определены вероятности правильного обнаружения. Результаты проведенных исследований представлены на рисунках 8 и 9 и в таблицах 1 и 2.

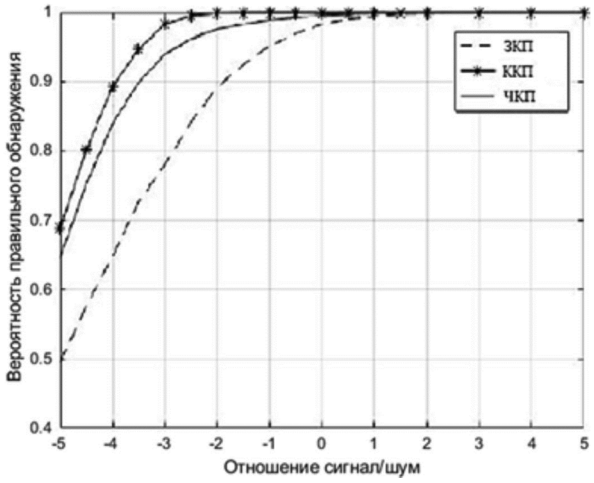


Рисунок 8 – Зависимости вероятности правильного обнаружения от отношения сигнал-шум для классического корреляционного приемника (ККП), знакового корреляционного приемника (ЗКП), частотного корреляционного приемника (ЧКП)

Как показали результаты моделирования (рисунок 8) и (таблица 1), вероятность правильного обнаружения сигнала превышает 95 % при отношениях сигнал-шум более 0 дБ для ЗКП, -1.5 дБ для ЧКП, -2 дБ для ККП.

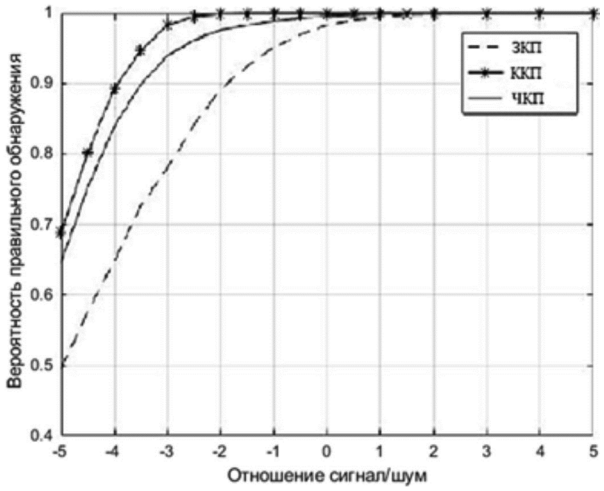


Таблица 1 – Зависимости правильного обнаружения от отношения сигнал-шум

SNR, дБ	-5	0	5
ЗКП	4993	369	0
ККП	3022	0	0
ЧКП	3472	94	0

Вероятности ложного срабатывания представлены на рисунке 9 и в таблице 2.

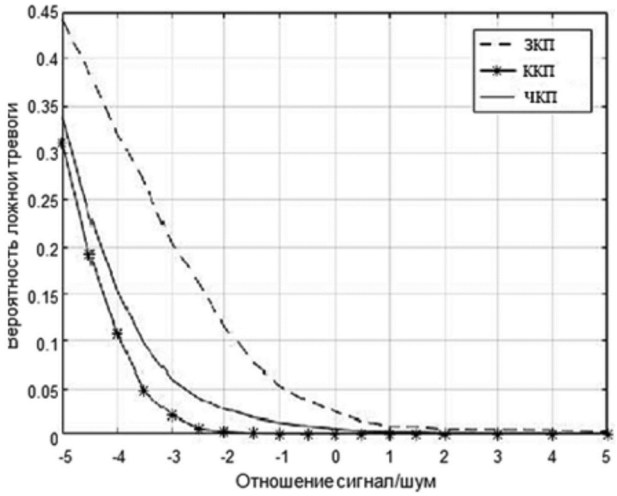


Рисунок 9 – Зависимости вероятности ложного срабатывания от отношения сигнал-шум для классического корреляционного приемника (ККП), знакового корреляционного приемника (ЗКП), частотного корреляционного приемника (ЧКП)

Таблица 2 – Зависимости ложного срабатывания от отношения сигнал-шум

SNR, дБ	-5	0	5	SNR, дБ
ЗКП	4406	237	3	ЗКП
ККП	3201	0	0	ККП
ЧКП	3398	18	0	ЧКП

Сравнение характеристик корреляционных приемников

Когда на вход корреляционного приемника попадает детерминированный сигнал, значение на выходе коррелятора превышает установленный порог. Если же шум ошибочно распознаётся как сигнал, происходит ложное

срабатывание. Основные показатели эффективности корреляционного приёмника включают вероятность правильного обнаружения сигнала (P_{cd}) и вероятность ложного срабатывания (P_{at}).

Чтобы определить характеристики изучаемых корреляторов, в системе MATLAB была разработана модель, параметры которой представлены в таблице 3.

Таблица 3 – Параметры сигналов, используемых при моделировании

Параметр	Значение
Диапазон SNR	от -5 дБ до 5 дБ
Количество PRACH сигналов (N)	10000
Количество отсчетов в опорном сигнале	54

В передатчике было сформировано N PRACH-сигналов. Затем передавали сигналы через канал с АБГШ. В приемнике принимались решения о присутствии или отсутствии сигнала, после чего эти решения сопоставлялись с данными, записанными с реальной базовой станции стандарта GSM 900. В итоге для каждого соотношения сигнал-шум были определены вероятности правильного обнаружения (P_{cd}) и ложного срабатывания (P_{at}).

Математическое моделирование показало, что классический корреляционный приемник обладает наилучшими характеристиками, тогда как значения для корреляторов со знаковым методом уступают примерно на 2 дБ, что заметно снижает их эффективность. Корреляционный приемник, работающий в частотной области, показал результаты, близкие к классическому варианту, отличаясь всего на 1 дБ, что делает его пригодным для применения в аналогичных условиях.

Заключение

В заключение следует отметить, что в ходе исследования были проанализированы три типа корреляционных приемников для обнаружения PRACH сигнала стандарта GSM на фоне помех. Описаны их структура, принципы работы. В контексте проекта разработки программно-аппаратного комплекса для поиска потерявшихся людей в лесном массиве, использующего БПЛА, особое значение приобретает расчет корреляции между принятым сигналом и PRACH сигналом. Это позволяет определять наличие мобильных устройств в сложных условиях, когда устройство могло не успеть подключиться к временной базовой станции на борту БПЛА.

На основе полученных результатов, наиболее подходящим корреляционным приемником для программно-аппаратного комплекса является классический корреляционный приемник. Выбрано он был по следующим причинам: знаковый приемник требует меньше ресурсов, но не обеспечивает необходимую вероятность правильного обнаружения, частотный приемник, хотя и точен, но сложнее в реализации и вносит задержку. Классический корреляционный приемник, учитывая небольшую длину преамбулы (PRACH сигнала) в 54 отсчета, затрачивает умеренное количество ресурсов, что делает его оптимальным выбором.

Литература

1. Разработка программно-аппаратного комплекса для поиска в лесном массиве потерявшегося человека с устройством мобильной связи [Online]. Available: https://engineers.tusur.ru/projects/2cb392b9-ab42-445f-a6fd-ac77acc0c721
2. Digital cellular telecommunications system (Phase 2+); Multiplexing and multiple access on the radio path (GSM 05.02) [Online]. Available: https://www.etsi.org/deliver/etsi_gts/05/0502/05.00.00_60/gsmts_0502v050000p.pdf
3. Digital cellular telecommunications system (Phase 2+); Physical layer on the radio path; General description (GSM 05.01 version 5.3.0) [Online]. Available: https://www.etsi.org/deliver/etsi_gts/05/0501/05.03.00_60/gsmts_0501v050300p.pdf.
4. Digital cellular telecommunications system (Phase 2+); Modulation (GSM 05.04 version 7.0.0) [Online]: Available: https://www.etsi.org/deliver/etsi_en/300900_300999/300959/07.00.00_40/en_300959v070000o.pdf
5. Как работает радиоинтерфейс в GSM-сетях [Online]. Available: https://habr.com/ru/articles/268127/ (дата обращения 11.12.2023).
6. Зюко, А.Г., et al. Теория передачи сигналов. М.: Радио и связь 355 (1986).
7. Коростелев, А.А., Клюев Н.Ф. and Мельник Ю.А.. Теоретические основы радиолокации / Под ред. В.Е. Дулевича. М.: Сов. радио (1978).

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».

УДК 004.932.4

НЕЙРОСЕТЕВАЯ ВИНЕРОВСКАЯ ФИЛЬТРАЦИЯ ИЗОБРАЖЕНИЙ В ОБЛАСТИ ДИСКРЕТНОГО ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЯ

Алимагадов К.А., *аспирант Национального исследовательского университета «МИЭТ», e-mail: alimagadovk@yandex.ru*
Умняшкин С.В., *д.ф-м.н., профессор Национального исследовательского университета «МИЭТ», e-mail: vrinf@miee.ru*

NEURAL NETWORK WIENER FILTERING OF IMAGES IN THE DOMAIN OF DISCRETE WAVELET TRANSFORM

Alimagadov K.A., Umnyashkin S.V.

The research is devoted to quality improvement of additive white Gaussian noise suppression in images. The architecture of neural network for image denoising based on Wiener filtering in the domain of discrete wavelet transform (DWT) has been developed. The key point of the processing scheme is the adaptation of image local filtering technique based on contour analysis in the domain of DWT. The number of trainable parameters of the proposed network is an order of magnitude smaller as compared with the modern machine learning models used to solve the same task.
The neural network was tested on «Berkeley segmentation data set 500». Estimation of filtering quality was based on analysis of peak signal to noise ratio and structure similarity index values. It was found that the developed approach provides higher quality of noise suppression than other considered algorithms based on Wiener filtering.

Key words: noise suppression, image processing, discrete wavelet transform, neural networks, Wiener filter.

Ключевые слова: подавление шумов, обработка изображений, дискретное вейвлет-преобразование, нейронные сети, фильтр Винера.

Исследование посвящено улучшению качества подавления аддитивного белого гауссового шума в изображениях. Разработана архитектура нейронной сети, выполняющей подавление шумов на основе винеровской фильтрации в области дискретного вейвлет-преобразования (ДВП). Ключевым моментом схемы обработки является адаптация способа локальной фильтрации изображения на основе контурного анализа в области ДВП. Предлагаемая архитектура сети имеет на порядок меньшее количество обучаемых параметров, чем у современных моделей машинного обучения, решающих аналогичную задачу.

Для сравнения качества обработки нейронная сеть была протестирована на наборе данных «Berkeley segmentation data set 500». Оценка качества фильтрации выполнялась на основе анализа значений мер пикового отношения сигнал/шум и индекса структурного сходства. Установлено, что разработанный подход обеспечивает более качественное подавление шумов, по сравнению с другими рассмотренными алгоритмами на основе винеровской фильтрации.

Предварительные сведения

Искажения и шумы, возникающие в процессах съёмки и передачи цифровых изображений, затрудняют восприятие фото- и видео данных человеком и могут приводить к некорректной работе систем компьютерного зрения. Подавление шумов способствует улучшению визуального качества изображений, а также снижению доли ошибок распознавания при последующем применении алгоритмов интеллектуального анализа данных [1].

На сегодняшний день наиболее качественные результаты подавления шумов демонстрируют методы обработки на основе нейронных сетей [2-4]. Однако следует отметить, что применение глубоких нейросетевых моделей, как правило, требует бóльших вычислительных ресурсов, по сравнению с классическими алгоритмами фильтрации изображений. Современные нейронные сети содержат сотни тысяч, а иногда и миллионы параметров [2-4]. Это приводит к необходимости сбора значительного объёма обучающих данных, искусственного расширения обучающей выборки за счёт аугментаций и использования регуляризации для того, чтобы избежать переобучения и сохранить способность модели к обобщению [5]. Помимо этого, глубокие нейронные сети являются плохо интерпретируемыми (работают по принципу «чёрного ящика») и подвержены влиянию спектрального сдвига в область низких частот, который приводит к деградации полезной высокочастотной информации при обработке [6, 7], выражающейся в потере резкости контуров на изображении.

Одним из направлений исследований в области нейросетевой фильтрации изображений является разработка алгоритмов, сочетающих преимущества машинного обучения и классических методов цифровой обработки, что позволяет снизить количество параметров моделей и повысить их интерпретируемость [3, 8].

Целью работы является улучшение качества подавления шумов в изображениях и развитие идей, описанных в статье [8].

Предлагаемый подход основан на нейросетевой винеровской фильтрации в базисе дискретного вейвлет-преобразования (ДВП) с учётом контурной информации

и позволяет обеспечить больший прирост метрик качества после обработки, по сравнению с большинством рассмотренных в работе аналогов.

Модель шума

В рамках исследования рассматривается модель белого аддитивного не коррелированного с изображением шума, распределённого по нормальному закону с нулевым математическим ожиданием μ = 0 и среднеквадратическим отклонением (СКО) σ.

Дискретные полутоновые изображения будем представлять в виде матриц, элементы которых определяют яркость пикселей. Искажение изображения в формализованном виде описывается выражением:

$$\mathbf{X} = \mathbf{Y} + \mathbf{N}, \quad (1)$$

где $\mathbf{X} = (x_{m,n})$, $\mathbf{Y} = (y_{m,n})$ и $\mathbf{N} = (\eta_{m,n})$ – матрицы зашумлённого изображения, исходного изображения и шума соответственно.

Дискретное вейвлет-преобразование

Фильтрацию изображений будем выполнять в базисе двумерного дискретного вейвлет-преобразования (ДВП). Вычисление ДВП-спектра входного изображения реализовано с применением биортогональных фильтров 9/7 [9], импульсные характеристики (ИХ) которых задают одномерные свёртки, применяющиеся последовательно вдоль строк, затем вдоль столбцов матрицы изображения.

В процессе обработки выполняется 3-х уровневое ДВП без прореживания откликов ВЧ и НЧ-фильтров, задающих преобразование [9]. При вычислении каждого i-го ($i \in \{1, 2, 3\}$) уровня ДВП получаются четыре набора вейвлет-коэффициентов (саббэндов): один саббэнд, соответствующий низкочастотной составляющей ДВП-спектра, $\mathbf{V}_i^A$, и три вейвлет-саббэнда $\mathbf{V}_i^H$, $\mathbf{V}_i^V$, $\mathbf{V}_i^D$, соответствующие горизонтальной, вертикальной и диагональной высокочастотным составляющим. Фильтрация применяется только к саббэндам $\mathbf{V}_i^H$, $\mathbf{V}_i^V$, $\mathbf{V}_i^D$, а саббэнд $\mathbf{V}_i^A$ используется для вычисления (i + 1)-го уровня ДВП.

Фильтр Винера

Для подавления шума предлагается использовать фильтр Винера, который минимизирует математическое ожидание квадрата ошибки для оценки полезного сигнала [10]. Двумерные ДВП-спектры (т.е. наборы коэффициентов ДВП) изображений $\mathbf{X}$, $\mathbf{Y}$ и шума $\mathbf{N}$ обозначим $\mathbf{V} = (v_{k,l})$, $\mathbf{U} = (u_{k,l})$ и $\mathbf{G} = (g_{k,l})$ соответственно. Из предположения о некоррелированности шума с полезным сигналом следует, что спектр мощности зашумлённого изображения представляет собой сумму спектров мощности исходного изображения и шумовой составляющей. Тогда коэффициенты фильтра Винера вычисляются по формуле [8]

$$w_{k,l} = \frac{\hat{s}_{k,l}}{\hat{s}_{k,l} + \hat{p}_{k,l}}, \quad (2)$$

где $\hat{s}_{k,l}$ и $\hat{p}_{k,l}$ – оценки коэффициентов ДВП-спектров мощности полезного сигнала $s_{k,l} = \mathrm{M}\left[u_{k,l}^2\right]$ и шума $p_{k,l} = \mathrm{M}\left[g_{k,l}^2\right]$ соответственно (обозначение M используется для математического ожидания).

Фильтрация осуществляется путём поэлементного перемножения коэффициентов фильтра Винера $w_{k,l}$ и коэффициентов ДВП $v_{k,l}$ зашумлённого изображения: $\hat{\mathbf{U}} = (\hat{u}_{k,l} = w_{k,l} v_{k,l})$. После этого для получения восстановленного изображения выполняется обратное ДВП (ОДВП): $\hat{\mathbf{Y}} = \text{ОДВП}(\hat{\mathbf{U}})$.

Общая схема метода подавления шума

Предлагаемый алгоритм подавления шумов перед фильтрацией выполняет нечёткую классификацию коэффициентов ДВП-спектра $\mathbf{V}$ на два класса в зависимости от локализации соответствующих базисных функций (двумерных вейвлетов) на изображении: 1) «контурные», соответствующие контурам на изображении, 2) «фоновые», соответствующие фоновым областям. Такая классификация позволяет более гибко подстраивать параметры $\mathbf{W} = (w_{k,l})$ фильтра Винера: необходимые в (2) оценки спектров мощности $\hat{u}_{k,l}^2$ формируются по-разному для каждого из двух классов вейвлет--коэффициентов.

Общая схема разработанного алгоритма подавления шума состоит из следующих шагов. Входными данными являются изображение $\mathbf{X}$ и заданная величина СКО σ шума $\mathbf{N}$.

Шаг 1. Вычисление ДВП-спектра входного изображения: $\mathbf{V} = \text{ДВП}(\mathbf{X})$.

Шаг 2. Оценка дисперсии шума в саббэндах $(\mathbf{V}_i^H, \mathbf{V}_i^V, \mathbf{V}_i^D)_{i=1}^3$ по величине σ.

Шаг 3. Определение степени принадлежности вейвлет-коэффициентов саббэндов $(\mathbf{V}_i^H, \mathbf{V}_i^V, \mathbf{V}_i^D)_{i=1}^3$ к классам «контурные» и «фоновые».

Шаг 4. Нахождение оценок спектров мощности полезного сигнала $\hat{\mathbf{S}}_{\mathbf{Y}} = (\hat{s}_{k,l})$ и шума $\hat{\mathbf{S}}_{\mathbf{N}} = (\hat{p}_{k,l})$ для каждого класса («контурных»/«фоновых») вейвлет-коэффициентов.

Шаг 5. Вычисление коэффициентов $\mathbf{W} = (w_{k,l})$ фильтра Винера для каждого класса вейвлет-коэффициентов по формуле (2).

Шаг 6. Фильтрация ДВП-спектра, состоящая в покомпонентном перемножении коэффициентов фильтра Винера и соответствующих им коэффициентов ДВП-спектра: $\hat{\mathbf{U}} = (\hat{u}_{k,l} = w_{k,l} \cdot v_{k,l})$.

Шаг 7. Восстановление изображения путём выполнения обратного ДВП: $\hat{\mathbf{Y}} = \text{ОДВП}(\hat{\mathbf{U}})$.

Шаг 8. Компенсация артефактов, возникших в результате винеровской фильтрации.

Перечисленные выше шаги были реализованы как операции, выполняемые слоями нейронной сети с архитектурой (см. рисунок 1), описание которой приведено в последующих разделах.

Дисперсия шума в саббэндах ДВП-спектра

Помимо входного изображения, нейросетевая мо-

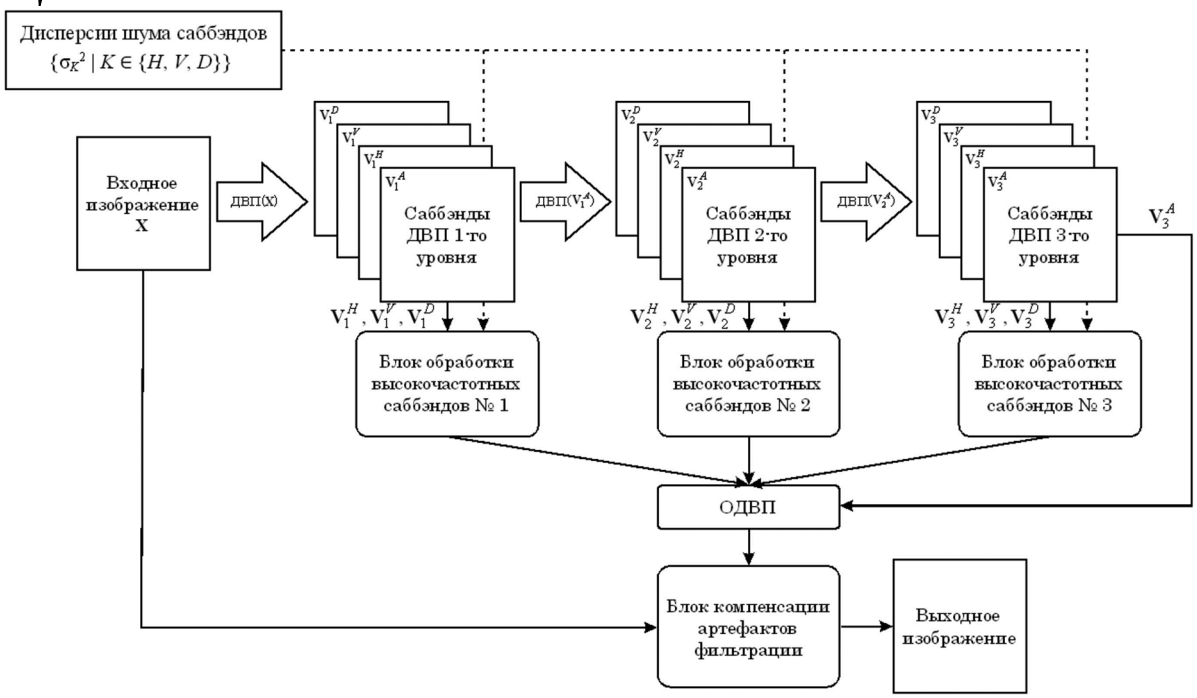


Рисунок 1 – Блок-схема архитектуры нейронной сети

дель также принимает значение СКО шума σ. Поскольку выбранный базис ДВП является биортогональным, в результате ДВП не выполняется сохранение евклидовой нормы сигнала, и дисперсия шумовой составляющей $\mathbf{G}$ саббэндов ДВП-спектра имеет значение, несколько отличающееся от дисперсии σ^2 исходного шума $\mathbf{N}$.

По заданному значению σ^2 дисперсии шумовой составляющей σ_K^2, $K \in \{A, H, V, D\}$, для саббэндов первого уровня ДВП $\mathbf{V}_1^K$ рассчитываются как [10]

$$\sigma_K^2 = \sigma^2 C_h C_f, \tag{3}$$

где $C_h = \sum_{i=1}^{L} h_i^2$, $C_f = \sum_{j=1}^{P} f_j^2$ – константы, $\{h_i\}_{i=1}^{L}$ и $\{f_j\}_{j=1}^{P}$ – КИХ фильтров, задающих ДВП и применяемых соответственно вдоль строк и столбцов изображения для вычисления коэффициентов саббэнда.

Для саббэндов второго и третьего уровней ДВП (i = 2,3) формула (3) применяется рекуррентно, где вместо σ^2 используется дисперсия σ_A^2, найденная для низкочастотного саббэнда предыдущего уровня преобразования $\mathbf{V}_{i-1}^A$.

Значения дисперсий σ_K^2 и соответствующие им саббэнды $\mathbf{V}^K$ передаются в блоки обработки, выполняющие обнаружение контуров и фильтрацию ДВП-спектра (см. рисунок 1).

Алгоритм обработки саббэндов ДВП-спектра

Блоки обработки высокочастотной информации, отвечающие за шаги 3-6 разработанного алгоритма, реализуют схему обработки, представленную на рисунке 2 и состоящую из нечёткой классификации вейвлет-коэффициентов, нахождения оценок ДВП-спектров мощности полезного сигнала и шума, формирования коэффициентов фильтра Винера $\mathbf{W}$ и собственно фильтрации ДВП-спектра $\mathbf{V}$. Далее поясним содержание схемы рисунка 2 более детально.

Нечёткая классификация коэффициентов ДВП

Учёт информации о положении контуров на изображении в процессе подавления шумов позволяет добиться более качественных результатов обработки [3, 8]. Поэтому перед фильтрацией ДВП-спектра изображения выполняется классификация его вейвлет-коэффициентов на «контурные» и «фоновые». Классификация применяется ко всем высокочастотным саббэндам всех уровней ДВП и включает в себя следующие этапы.

1. На основе значений вейвлет-коэффициентов саббэндов $\mathbf{V}^K$, $K \in \{H, V, D\}$, формируются массивы (карты признаков) $\mathbf{R}_1^K = \varphi_K(\mathbf{V}_K)$ с помощью функций $\varphi_K(x)$, имеющих вид

$$\varphi_K(x) = \left| \frac{x^3}{x^2 + (\sigma_K \lambda_K)^2} \right|, \tag{4}$$

где коэффициенты λ_K являются обучаемыми параметрами модели, σ_K находятся из (3). По аналогии с идеей, предложенной в работе [3], функция (4) используется для подавления шумовых вейвлет-коэффициентов с малыми абсолютными значениями, что можно интерпретировать как некоторую пороговую фильтрацию.

2. С целью морфологического сглаживания к полученным картам признаков $\mathbf{R}_1^K$, $K \in \{H, V, D\}$, последовательно применяются операции полутонового размыкания и замыкания [11] с примитивами:

$$\mathbf{Q}^H = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{Q}^V = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix},$$

$$\mathbf{Q}^{D_1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{Q}^{D_2} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

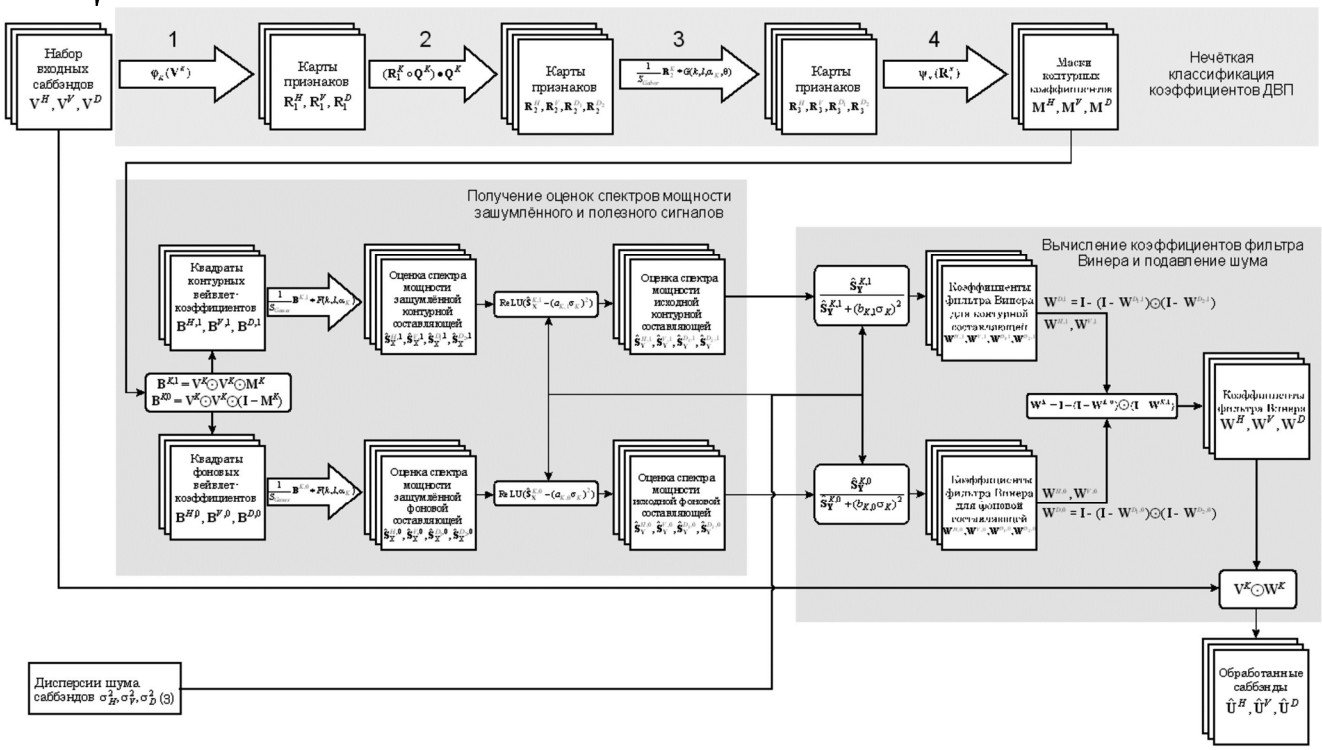


Рисунок 2 – Блок обработки высокочастотных саббэндов $\mathbf{V}^H$, $\mathbf{V}^V$, $\mathbf{V}^D$

При этом карты признаков $\mathbf{R}_1^H$ и $\mathbf{R}_1^V$ подвергаются обработке с примитивами $\mathbf{Q}^H$ и $\mathbf{Q}^V$ соответственно, а для обработки карты $\mathbf{R}_1^D$ используются два примитива $\mathbf{Q}^{D_1}$ и $\mathbf{Q}^{D_2}$ (по одному для каждого направления диагональных контуров). В результате выполнения данного этапа получаются четыре карты признаков: $\mathbf{R}_2^H$, $\mathbf{R}_2^V$, $\mathbf{R}_2^{D_1}$ и $\mathbf{R}_2^{D_2}$; в общем виде процедуру можно записать как

$$\mathbf{R}_2^K = (\mathbf{R}_1^K \circ \mathbf{Q}^K) \bullet \mathbf{Q}^K, \qquad (5)$$

где символы $\circ$ и $\bullet$ используются для обозначения морфологических операций полутонового размыкания и замыкания соответственно; $K \in \{H, V, D_1, D_2\}$.

3. Выполняется свёртка карт признаков, которые были получены на предыдущем этапе, с ИХ модифицированных фильтров Габора [12]:

$$G(k,l,\alpha,\theta) = \frac{1}{2}\exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}\right)\left(1+\cos\left(2\pi\theta x\right)\right), \qquad (6)$$

где $x = k\cos(\alpha) + l\sin(\alpha)$, $y = -k\sin(\alpha) + l\cos(\alpha)$, α – угол, задающий направление фильтра; σ_x, σ_y – параметры, отвечающие за затухание ИХ фильтра вдоль осей координат Ox и Oy соответственно; θ – пространственная частота гармонических колебаний. Параметры σ_x, σ_y, θ настраиваются в процессе обучения сети, а значение α выбирается из набора $\alpha_K \in \{0, \pm\pi/4, \pi/2\}$ на основе информации о типе саббэнда $\mathbf{V}^K$. $\alpha_H = 0$ – для горизонтальной составляющей $\mathbf{V}^H$, $\alpha_V = \pi/2$ – для вертикальной составляющей $\mathbf{V}^V$. Для диагональной составляющей $\mathbf{V}^D$ используются два значения угла фильтрации $\alpha_{D_1} = \pi/4$, $\alpha_{D_2} = -\pi/4$ и вычисляются две карты признаков (для каждого направления диагональных контуров). Размеры масок фильтров (6) выбраны равными 9×9, в соответствии с размерами ИХ используемых двумерных ДВП-фильтров. Результатом выполнения данного этапа являются карты признаков:

$$\mathbf{R}_3^K = \frac{1}{S_{Gabor}}\mathbf{R}_2^K * G(k,l,\alpha_K,\theta), \qquad (7)$$

где символ $*$ используется для обозначения операции двумерной дискретной свёртки, $K \in \{H, V, D_1, D_2\}$; нормировочный параметр $S_{Gabor} = \sum_{i=-4}^{4}\sum_{j=-4}^{4} G(i,j,\alpha_K,\theta)$.

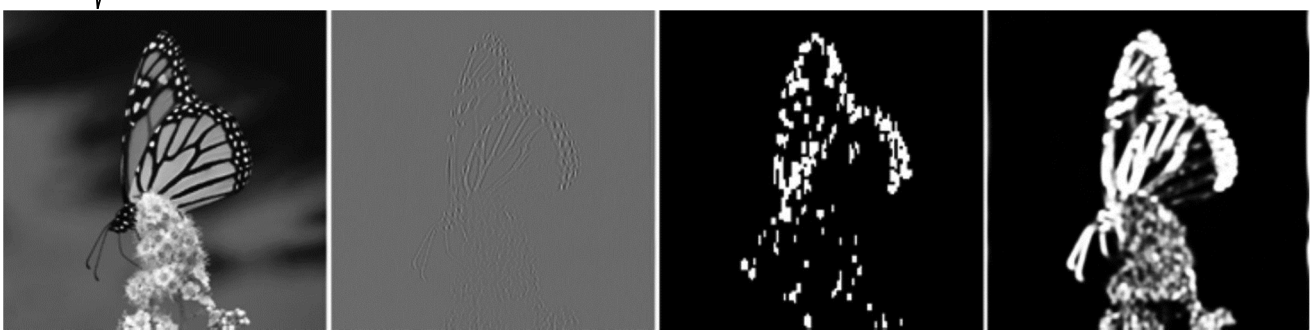
4. Найденные на третьем этапе карты признаков (7) проходят через нелинейную функцию активации $\psi_K(x)$, которая приводит входные значения к диапазону [0; 1) и имеет вид:

$$\psi_K(x) = (1 - \exp(-\mu_K^2 x^2))^2, \qquad (8)$$

где μ_K – обучаемый параметр, определяющий ширину полосы, при попадании в которую значение функции будет близко к нулю; $K \in \{H, V, D\}$. Для диагональных саббэндов полагаем $\psi_D(x) = \psi_{D_1}(x) = \psi_{D_2}(x)$. На основе откликов функции (8) по матрицам $\mathbf{R}_3^K = (r_{k,l}^K)$ формируются маски $\mathbf{M}^K = (m_{k,l} = \psi_K(r_{k,l}^K))$, задающие для каждого вейвлет-коэффициента $v_{k,l} \in \mathbf{V}^K$ степень его принадлежности к контурной составляющей ДВП-спектра изображения как непрерывную величину $m_{k,l} \in [0; 1)$:

$$\mathbf{M}^K = \psi_K(\mathbf{R}_3^K), \qquad (9)$$

где $K \in \{H, V, D_1, D_2\}$. Маски $\mathbf{M}^{D_1}$ и $\mathbf{M}^{D_2}$, соответствующие диагональным контурам различных направлений, объединяются в одну общую маску $\mathbf{M}^D = \mathbf{I} - -(\mathbf{I} - \mathbf{M}^{D_1}) \odot (\mathbf{I} - \mathbf{M}^{D_2})$, где $\mathbf{I}$ – матрица, состоящая из единиц, а символ $\odot$ обозначает операцию покомпонентного перемножения матриц.



а) *б)* *в)* *г)*

Рисунок 3 – Изображение Butterfly (а); его саббэнд $\mathbf{V}_2^V$ (б); помеченные «контурные» вейвлет-коэффициенты (алгоритм из статьи [8]) (в); помеченные «контурные» вейвлет-коэффициенты (разработанный алгоритм) (г)

Для каждого саббэнда $\mathbf{V}^K$ соответствующая ему маска $\mathbf{M}^K$ позволяет определить совокупность «контурных» вейвлет-коэффициентов как нечёткое множество [11], являющееся подмножеством $\mathbf{V}^K$: чем ближе значение $m_{k,l} = \psi_K(r_{k,l}^K)$ к единице, тем выше степень принадлежности $v_{k,l}$ к множеству контурных коэффициентов. Разработанный алгоритм «более мягко» выделяет структуру и положение контуров по высокочастотным саббэндам ДВП-спектра, в отличие от аналогичного детектора из работы [8], использующего бинарные признаки принадлежности для контурных точек {0; 1} (см. рисунок 3).

Получение оценок спектров мощности зашумлённого и полезного сигналов

Оценка ДВП-спектра мощности зашумлённого сигнала находится как свёртка квадратов значений коэффициентов ДВП-спектра с гауссовой ИХ вида:

$$F(k,l,\alpha) = \exp\left(-\frac{x^2}{\sigma_x^2} - \frac{y^2}{\sigma_y^2}\right), \quad (10)$$

где $x = k\cos(\alpha) + l\sin(\alpha)$, $y = -k\sin(\alpha) + l\cos(\alpha)$, α – угол, задающий поворот ИХ; параметры σ_x, σ_y отвечают за затухание ИХ фильтра вдоль осей координат Ox и Oy соответственно и настраиваются в процессе обучения сети. Как и ранее для фильтров (6), размеры масок фильтров (10) были выбраны равными 9×9.

Выражения для оценки спектров мощности контурной и фоновой составляющих зашумлённого изображения имеют вид

$$\hat{\mathbf{S}}_{\mathbf{X}}^{K,C} = \frac{1}{S_{Gauss}}\mathbf{B}^{K,C} * F(k,l,\alpha_K), \quad (11)$$

где $K \in \{H, V, D_1, D_2\}$; $C \in \{0,1\}$ – признак контурной (C = 1) либо фоновой (C = 0) составляющей; $\mathbf{B}^{K,1} = \mathbf{V}^K \odot \mathbf{V}^K \odot \mathbf{M}^K$, $\mathbf{B}^{K,0} = \mathbf{V}^K \odot \mathbf{V}^K \odot (\mathbf{I} - \mathbf{M}^K)$ – квадраты значений «контурных» и «фоновых» вейвлет-коэффициентов соответственно; $\alpha_K \in \{0, \pm\pi/4, \pi/2\}$ – параметр, значения которого выбирались, как и в (7), исходя из информации о типе саббэнда $\mathbf{V}^K$; $S_{Gauss} = \sum_{i=-4}^{4}\sum_{j=-4}^{4} F(i,j,\alpha)$. Для саббэнда $\mathbf{V}^D$ вычисляются две различных оценки $\hat{\mathbf{S}}_{\mathbf{X}}^{D_1,C}$ и $\hat{\mathbf{S}}_{\mathbf{X}}^{D_2,C}$ (при этом в (11) полагается, что $\mathbf{B}^{D_1,C} = \mathbf{B}^{D_2,C} = \mathbf{B}^{D,C}$).

Поскольку в рассматриваемой модели шум является белым и не коррелированным с полезным сигналом, его спектр мощности является константой. С учётом этого оценку спектра мощности полезного сигнала $\hat{\mathbf{S}}_{\mathbf{Y}}^{K,C} = (\hat{s}_{k,l}^{K,C})$ для контурной и фоновой составляющих предлагается формировать как

$$\hat{\mathbf{S}}_{\mathbf{Y}}^{K,C} = \mathrm{ReLU}\,(\hat{\mathbf{S}}_{\mathbf{X}}^{K,C} - (a_{K,C}\sigma_K)^2), \quad (12)$$

где $\mathrm{ReLU}(x) = \max(0, x)$ – функция активации, $K \in \{H, V, D_1, D_2\}$, $C \in \{0,1\}$; $a_{K,C}$ – обучаемый параметр, причём $a_{D_1,C} = a_{D_2,C}$.

Вычисление коэффициентов фильтра Винера и подавление шума

Значения коэффициентов фильтра Винера $\mathbf{W}^{K,C}$ для контурной и фоновой составляющих ДВП-спектра рассчитываются в соответствии с формулой (2) следующим образом:

$$w_{k,l}^{K,C} = \frac{\hat{s}_{k,l}^{K,C}}{\hat{s}_{k,l}^{K,C} + (b_{K,C}\sigma_K)^2}, \quad (13)$$

где $K \in \{H, V, D_1, D_2\}$, $C \in \{0,1\}$, $\hat{s}_{k,l}^{K,C}$ – оценки коэффициентов ДВП-спектра мощности $\hat{\mathbf{S}}_{\mathbf{Y}}^{K,C}$, $b_{K,C}$ – обучаемый параметр. Набор коэффициентов фильтра $\mathbf{W}^{D,C} = (w_{k,l}^{D,C})$ для обработки саббэнда $\mathbf{V}^D$ вычисляется как $w_{k,l}^{D,C} = 1 - (1 - w_{k,l}^{D_2,C})(1 - w_{k,l}^{D_1,C})$.

На основе полученных значений $w_{k,l}^{K,C}$ (13) для каждого саббэнда $\mathbf{V}^K$, $K \in \{H, V, D\}$, вычисляются коэффициенты фильтра $\mathbf{W}^K = (w_{k,l}^K)$, учитывающие степень принадлежности ДВП-коэффициентов к контурам и фону:

$$w_{k,l}^K = 1 - (1 - w_{k,l}^{K,0})(1 - w_{k,l}^{K,1}). \quad (14)$$

Фильтрация ДВП-спектра производится путём покомпонентного перемножения матриц коэффициентов фильтра (14) и вейвлет-коэффициентов саббэндов: $\hat{\mathbf{U}}^K = \mathbf{W}^K \odot \mathbf{V}^K$. Так преобразованные саббэнды ДВП-спектра используются для получения восстановленного изображения с помощью ОДВП.

Компенсация артефактов фильтрации

В процессе винеровской фильтрации вместе с шумовыми коэффициентами ДВП-спектра также могут быть подавлены вейвлет-коэффициенты, соответству-

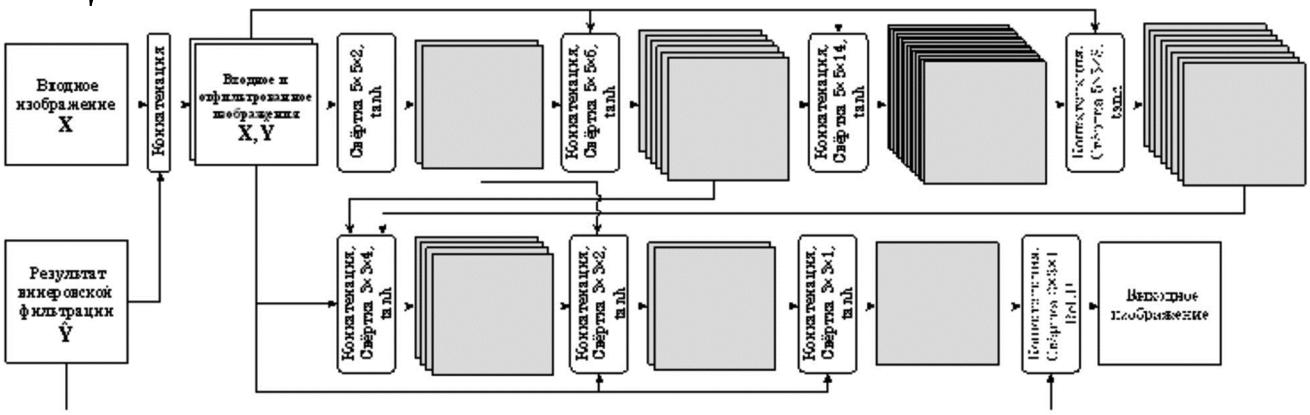


Рис. 4. Блок компенсации артефактов фильтрации

ющие полезной высокочастотной информации. Для того чтобы снизить влияние возникающих вследствие этого искажений и артефактов, в архитектуру модели дополнительно был добавлен блок (см. рисунок 4), разработанный на основе идей, изложенных авторами нейронной сети U-Net в работе [13]. Серым цветом на рисунке выделены карты признаков нейронной сети, а выполняемые операции представлены блоками, где под конкатенацией подразумевается объединение матриц карт признаков и изображений в 3-х мерный тензор (матрицы конкатенируются как каналы тензора); рядом с каждой свёрткой подписан размер её ядра в формате (высота×ширина×количество выходных каналов); обозначения tanh (гиперболический тангенс) и ReLU соответствуют использованию нелинейных функций активации после вычисления свёрток.

Блок компенсации артефактов фильтрации состоит из 8 свёрточных слоёв и имеет «короткие соединения» на основе конкатенации карт признаков. Количество слоёв и размеры свёрток в блоке, а также карты признаков, между которыми выполняется конкатенация, были определены в ходе экспериментов таким образом, чтобы результат обработки обеспечивал наибольший прирост пикового отношения сигнал/шум (PSNR) [11]. Двумерные свёрточные слои были реализованы как разделимые свёртки [2]. Перед вычислением свёрток выполняется операция дополнения отсчётами, симметрично отражёнными относительно границ изображений и карт признаков, чтобы размеры (высота и ширина) откликов всех слоёв оставались равными размерам изображения, поступающего на вход.

Как показали проведённые нами эксперименты, на тестовых данных при среднем ($\sigma = 15$) и высоком ($\sigma = 30$) уровнях шума добавление блока компенсации даёт прирост PSNR в среднем на 0,46 дБ.

Результаты экспериментов

Обучение нейронной сети

Для обучения описанной нейросетевой модели шумоподавления использовался набор данных ILSVRC [14], состоящий из 1 281 167 изображений с разрядностью 8 бит/пиксель и диапазоном значений яркости пикселей {0, …, 255}. Исходные цветные изображения были преобразованы к полутоновому виду. Для каждой эпохи из тренировочного набора случайным образом отбирались 500 изображений, затем из случайно выбранной области каждого изображения вырезался фрагмент размера 128×128 пикселей. Данные фрагменты использовались в качестве эталонных меток в последующем обучении. При формировании входного изображения фрагмент-метка подвергался искажению белым аддитивным нормальным шумом с нулевым математическим ожиданием и СКО σ, значение которого выбиралось как реализация случайной величины, имеющей равномерное распределение U(0, 30). Зашумлённый таким способом фрагмент изображения использовался затем в качестве образа обучающей выборки.

В качестве функции потерь использовался квадратичный штраф. Обучение модели выполнялось с помощью библиотеки PyTorch версии 2.4.1 [15]. Для минимизации выборочного штрафа использовался алгоритм Adam [16] со значениями параметров: $\alpha = 10^{-2}$; $\beta_1 = 0{,}9$; $\beta_2 = 0{,}999$; $\varepsilon = 10^{-8}$, количество эпох было выбрано равным 600 (при дальнейшем обучении рост метрик PSNR и SSIM не наблюдался), размер батча выбран равным 8. Для всех параметров алгоритма Adam, за исключением скорости обучения α, использовались значения по умолчанию. Параметр α, размеры фрагментов-образов обучающей выборки, а также количество обучающих примеров в батче были подобраны эмпирически как показавшие наибольшие значения метрик качества фильтрации.

Тестирование методов шумоподавления

Тестирование сети выполнялось на наборе данных BSDS500 [17], состоящем из 500 полутоновых изображений. Оценка качества обработки выполнялась на основе сравнения значений мер PSNR и индекса структурного сходства (SSIM) [18] до и после обработки.

Проведено сравнение характеристик разработанной нейронной сети с существующими алгоритмами подавления шумов: винеровской фильтрацией в области ДВП [19], винеровской фильтрацией с применением ансамбля нейросетевых моделей [8], нейросетевой моделью DCT2net [3]. Рассмотрены три фиксированных уровня шума, соответствующие различным значениям СКО: низкий уровень шума ($\sigma = 7$), средний ($\sigma = 15$) и высокий ($\sigma = 30$). Значения метрик, рассчитанные по результатам экспериментов и усреднённые по тестовому набору, представлены в таблице 1.

Таблица 1 – Средние значения PSNR и SSIM до и после фильтрации при разных значениях СКО шума, вычисленные по набору BSDS500

Метрика	PSNR (средн.) [дБ]	SSIM (средн.)	PSNR (средн.) [дБ]	SSIM (средн.)	PSNR (средн.) [дБ]	SSIM (средн.)
СКО шума	σ = 7		σ = 15		σ = 30	
Значение до обработки	31,26	0,80	24,75	0,56	18,94	0,32
Фильтр Винера [19]	34,59	*0,93*	29,84	0,83	25,93	0,67
Фильтр Винера + нейросетевой ансамбль [8]	34,75	*0,93*	30,10	0,85	26,43	0,72
Разработанный алгоритм	*35,20*	**0,94**	*31,17*	*0,87*	*27,49*	*0,76*
DCT2net [3]	**35,38**	**0,94**	**31,22**	**0,88**	**27,72**	**0,77**

Примечание. Наибольшие значения в каждом столбце выделены жирным шрифтом, курсивом – следующие по убыванию после наибольших.

Таблица 2 – Значения PSNR (дБ) до и после фильтрации при разных значениях СКО шума

Изображение	Barbara	Boat	Lena	Goldhill	Mandrill
СКО шума	σ = 7				
PSNR до обработки	31,22	31,22	31,26	31,22	31,22
Фильтр Винера [19]	34,87	34,49	36,19	34,95	*35,64*
Фильтр Винера + нейросетевой ансамбль [8]	35,00	34,64	36,49	35,11	**35,77**
Разработанный алгоритм	*35,32*	*34,93*	*36,60*	*35,38*	34,62
DCT2net [3]	**35,78**	**35,22**	**36,97**	**35,79**	34,71
СКО шума	σ = 15				
PSNR до обработки	24,61	24,64	24,77	24,62	24,60
Фильтр Винера [19]	30,21	30,35	31,77	30,61	30,20
Фильтр Винера + нейросетевой ансамбль [8]	30,47	30,76	32,26	31,02	*30,41*
Разработанный алгоритм	*31,62*	*31,75*	**33,39**	*31,81*	**30,52**
DCT2net [3]	**31,95**	**31,92**	*33,23*	**32,07**	30,40
СКО шума	σ = 30				
PSNR до обработки	18,76	18,74	19,00	18,76	18,64
Фильтр Винера [19]	26,29	26,66	27,67	27,14	25,76
Фильтр Винера + нейросетевой ансамбль [8]	26,64	27,43	28,34	28,02	*26,16*
Разработанный алгоритм	*27,70*	*28,53*	**29,72**	*28,76*	**26,70**
DCT2net [3]	**28,30**	**28,89**	*29,66*	**29,13**	**26,70**



а) *б)* *в)* *г)*

Рисунок 5 – Фрагменты изображений, искажённых шумом при различных значения СКО, до (сверху) и после фильтрации (снизу): σ = 0 (а), σ = 7 (б), σ = 15 (в), σ = 30 (г)

Полученные значения метрик позволяют сделать вывод, что предложенная нейронная сеть обеспечивает более качественное шумоподавление по сравнению с другими алгоритмами винеровской фильтрации, уступая лишь сети DCT2net. При сравнении с последней разница в приросте PSNR в среднем не превышает 0,18 дБ для низкого, 0,05 дБ для среднего и 0,23 дБ для высокого уровня шума, а разница в приросте SSIM в среднем не превышает 0,01.

Рассмотренные алгоритмы также были протестированы на полутоновых изображениях Barbara, Boat, Lena, Goldhill и Mandrill, широко применяющихся для проверки алгоритмов цифровой обработки (см. таблицу 2).

Найденные значения PSNR согласуются с результатами из таблицы 1, при этом можно видеть, что в отдельных случаях при среднем и высоком уровнях шума

предложенный алгоритм демонстрирует лучшее качество обработки по сравнению с нейросетевой моделью DCT2net. Следует отметить, что DCT2net содержит 28 561 обучаемый параметр [3], нейросетевой ансамбль из работы [8] содержит 2772 параметра, в то время как у предложенной в настоящем исследовании модели их всего 1491. Примеры зашумлённых и отфильтрованных изображений для различных значений СКО шума представлены на рисунке 5.

Заключение

Значения метрик качества, полученные в экспериментах, свидетельствуют о том, что предложенный алгоритм нейросетевой фильтрации обеспечивает лучшее подавление шумов в сравнении с рассмотренными альтернативными реализациями фильтра Винера. Разработанная для подавления белого шума в изображениях нейросеть в среднем несколько уступает DCT2net по метрике PSNR, однако в некоторых случаях демонстрирует сопоставимые, а иногда и превосходящие результаты обработки.

Количество параметров предлагаемой нами сети шумоподавления примерно в 19 раз меньше чем DCT2net и в 1,8 раз меньше, чем у ансамбля нейронных сетей из нашей предыдущей работы [8]. Таким образом, описанный в настоящей работе алгоритм нейросетевой винеровской фильтрации в области ДВП, представляющий собой развитие подходов из [8], является более предпочтительным решением с точки зрения баланса сложности и качества обработки.

Литература

1. Алимагадов К.А., Умняшкин С.В. Аугментация данных на основе вейвлет-фильтрации при обучении нейронных сетей. Труды Международной конференции по компьютерной графике и зрению «Графикон». М.: Институт прикладной математики им. М.В. Келдыша РАН, 2023. С. 437-442. DOI 10.20948/graphicon-2023-437-442.

2. Huang J.-J., Dragotti P.L. WINNet: Wavelet-inspired invertible network for image denoising. IEEE Transactions on Image Processing. IEEE, 2022, pp. 4377-4392.

3. Herbreteau S., Kervrann C. DCT2net: An interpretable shallow CNN for image denoising. IEEE Trans. Image Process. IEEE, 2022, pp. 4292-4305.

4. Tu Z., Talebi H., Zhang H., Yang F., Milanfar P., Bovik A., Li Y. MAXIM: Multi-Axis MLP for image processing. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022, pp. 5759-5770.

5. Haykin S. Neural networks and learning machines. 3rd ed. New York: Pearson, 2009. 936 p.

6. Rahaman N., Baratin A., Arpit D., Draxler F., Lin M., Hamprecht F., Bengio Y., Courville A. On the spectral bias of neural networks. In International Conference on Machine Learning. PMLR. 2019, pp. 5301-5310.

7. Gal R., Cohen D., Bermano A., Cohen-Or D. SWAGAN: A style based wavelet-driven generative model. CoRR. 2021, pp. 1-11.

8. Алимагадов К.А., Умняшкин С.В. Подавление белого шума на изображениях на основе винеровской фильтрации в области дискретного вейвлет-преобразования с применением нейросетевых технологий. // Известия высших учебных заведений. Электроника. 2022. Т. 27, № 6. С. 807-818. DOI 10.24151/1561-5405-2022-27-6-807-818.

9. Guo T. et al. A review of wavelet analysis and its applications: Challenges and opportunities. IEEE Access. 2022, no. 10, pp. 58869-58903. DOI 10.1109/ACCESS.2022.3179517.

10. Умняшкин С.В. Основы теории цифровой обработки сигналов. М.: Техносфера, 2025. 550 с.

11. Gonzalez R.C., Woods R.E. Digital image processing. 4th ed. New York: Pearson Education, 2018. 1168 p.

12. Feichtinger H.G., Strohmer T. Gabor analysis and algorithms: Theory and applications. Springer Science & Business Media, 2012.

13. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention. Munich: Springer International Publishing, 2015, pp. 234-241.

14. Russakovsky O. et al. Imagenet large scale visual recognition challenge // International journal of computer vision. - 2015. - №115. - P. 211-252.

15. PyTorch. GitHub URL: https://github.com/pytorch/pytorch (дата обращения: 21.09.2024).

16. Kingma D.P., Ba J.L. Adam: a method for stochastic optimization. arXiv URL: https://arxiv.org/abs/1412.6980v1 (дата обращения: 21.09.2024).

17. Berkeley segmentation data set and benchmarks 500 (BSDS500). Berkeley Computer Vision Group URL: https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/resources.html (дата обращения: 15.09.2024).

18. 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, no. 13, pp. 600-612.

19. Alimagadov K.A., Umnyashkin S.V. Application of Wiener filter to suppress white noise in images: wavelet vs Fourier basis. 2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus). St. Petersburg, Moscow: IEEE, 2021, pp. 2059-2063. DOI 10.1109/ ElConRus51938.2021.9396470.

УДК 519.254, 621.391

ДЕКОМПОЗИЦИЯ НА ЭМПИРИЧЕСКИЕ МОДЫ: ТЕМАТИЧЕСКИЙ ОБЗОР (ЧАСТЬ 1)

Егоров В.В., *д.т.н., старший научный сотрудник, заведующий кафедрой радиостроения и средств связи Санкт-Петербургского государственного университета аэрокосмического приборостроения, e-mail: egorovrimr@mail.ru*
Клионский Д.М., *к.т.н., доцент кафедры «Информационные системы» Санкт-Петербургского государственного электротехнического университета «ЛЭТИ», e-mail: klio2003@list.ru*

EMPIRICAL MODE DECOMPOSITION: THEMATIC REVIEW (PART 1)

Egorov V.V., Klionskiy D.M.

Empirical mode decomposition (EMD) and the more general Hilbert-Huang technology first appeared in 1998. In the present thematic review we provide and analyze the main known modifications of the EMD method: classical EMD, EMD with parabolic envelope interpolation, EMD based on wavelet packet decomposition, ensemble EMD and the multidimensional modification of EMD. The main ideas and application features are described in the paper. We also provide connection between EMD and ordinary differential equations. Some features of the practical application of EMD modifications are specified along with the applications of the Hilbert-Huang technology.

Key words: empirical mode decomposition, envelope, sifting process, Hilbert-Huang technology, EMD modification, multicomponent signal.

Ключевые слова: декомпозиция на эмпирические моды, огибающая, процесс отсеивания, технология Гильберта-Хуанга, модификации ДЭМ, многокомпонентный сигнал.

Метод декомпозиции на эмпирические моды (ДЭМ) и обобщающая его технология Гильберта-Хуанга впервые вышли в свет в 1998 году. В данном тематическом обзоре приведены и проанализированы основные известные модификации метода ДЭМ: классический метод ДЭМ, ДЭМ с параболической интерполяцией огибающих, ДЭМ на основе вейвлет-пакетного разложения сигналов, ансамблевая (множественная) ДЭМ и многомерная модификация метода. Описаны основные идеи и особенности их применения. Приведена связь ДЭМ с обыкновенными дифференциальными уравнениями. Указаны некоторые области практического применения модификаций ДЭМ и всей технологии Гильберта-Хуанга.

Введение

Метод ДЭМ был впервые опубликован в 1998 году группой ученых во главе с Н. Хуангом и впервые применен при анализе поверхностных волн тайфунов [1]. С тех пор идеи метода были многократно использованы в отечественных и зарубежных публикациях и различных научно-технических приложениях.

Исторически, задолго до появления ДЭМ, Фурье-анализ был одним из основных методов обработки данных и сигналов и использовался для представления глобальных частотно-временных распределений, учитывающих всю выборку данных. Во многом из-за относительной простоты вычислений, наличия быстрых вычислительных алгоритмов, а также физической и математической интерпретации разложения [2], Фурье-анализ стал доминировать над всеми методами обработки сигналов. Для корректной применимости преобразования Фурье к сигналу $s(k)$ должны быть выполнены следующие условия (условия стационарности):

$$E\{|\, s^2(k)\,|\} < \infty,\ \ E\{|\, s(k)\,|\} = m = const\,,$$

$$C\{s(k_1), s(k_2)\} = C\{s(k_1+\tau), s(k_2+\tau)\},$$

где $E\{...\}$ – оператор вычисления математического ожидания, $C\{...\}$ – оператор вычисления автоковариационной функции.

Однако на практике анализ Фурье частично или полностью теряет свою эффективность при исследовании нестационарных и сигналов с частотной эволюцией (изменяющейся по некоторому сложному закону мгновенной частотой). Классическая Фурье-спектрограмма [2] не всегда обеспечивает требуемое разрешение в частотно-временной области и зависит от выбора окна и его параметров. Кроме того, преобразование Фурье определено на всей временной оси, поэтому из-за отсутствия локализационных свойств во временной области возникают сложности определения основных диапазонов концентрации энергии (на основе классической Фурье-периодограммы и ее модификаций), которая рассеивается по достаточно широкому диапазону частот, создавая порой ложную картину распределения энергии по частотам. Также Фурье-анализ может оказаться неэффективным при обработке сигналов с резкими изменениями ввиду гладкости тригонометрических функций.

Статья посвящена рассмотрению и обзору адаптивного мультимасштабного метода обработки и анализа сигналов – декомпозиции на эмпирические моды (ДЭМ, англ. EMD – Empirical Mode Decomposition) [1, 3-13]. ДЭМ, обладая подобными вейвлет-анализу свойствами, не требует априорного выбора базиса для разложения (декомпозиции) сигнала. Разложение по системе функций, называемых эмпирическими модами (ЭМ, англ. IMF – Intrinsic Mode Function) [1, 3, 11, 12, 14-17], с целью последующего анализа структуры сигнала проводится с учетом локальных особенностей (экстремумов и нулей) и структуры (наличия в сигнале трех основных типов компонент – шумовых, трендовых и так называе-

мых компонент-шаблонов) конкретного анализируемого сигнала. Мультимасштабность данного подхода состоит в том, что исходный сигнал $s(k)$ представляется конечным набором компонент $\{f_i(k)\}$ с разным временным масштабом (временным разрешением):

$$s(k) = f_1(k) + f_2(k) + ... + f_i(k) + ... + f_M(k),$$

где k – номер дискретного отсчета (дискретное нормированное время), i – номер компоненты в разложении, M – общее число компонент в разложении.

Сами функции извлекаются непосредственно из исходного сигнала, следовательно, такой базис всегда уникален (не повторяется в точности для других сигналов), апостериорен (становится полностью известным лишь после разложения данных, а не до него, как, например, в случае с тригонометрическим базисом, где функции известны заранее) и адаптивен (то есть приспосабливается к особенностям и свойствам данных) [1, 3, 6, 11, 12].

Применительно к ДЭМ термин «адаптивность» имеет иной смысл, чем в его классической трактовке в цифровой обработке сигналов, где встречается понятие «адаптивный фильтр» [18-22]. Как известно, адаптивный фильтр – система с обратной связью, параметры которой адаптируются (подстраиваются) к сигналу с заранее неопределенной статистической моделью в процессе его обработки. При этом под «адаптацией» понимается настройка параметров (коэффициентов) фильтра в соответствии с заданным критерием. Адаптивные фильтры используются во многих практических задачах [18-22].

Цель обзора – провести анализ возможностей ДЭМ и ее известных модификаций в составе технологии Гильберта-Хуанга и указать известные области применения метода ДЭМ и его модификаций.

Основные идеи метода ДЭМ

Рассмотрим соотношение, определяющее декомпозицию сигнала в случае априорно известного базиса (например, вейвлет-базиса или Фурье-базиса) [23, 24]:

$$s(k) = \sum_i g_i f_{i\,\mathrm{det}}(k), \qquad (1)$$

где $s(k)$ – анализируемый (обрабатываемый) сигнал; g_i – коэффициенты разложения сигнала по базису; $f_{i\,\mathrm{det}}(k)$ – детерминированные (априорно известные) функции. При разложении сигнала по базису $\{f_{i\,\mathrm{det}}(k)\}$ некоторые базисные функции могут исключаться из рассмотрения.

Для ДЭМ формула (1) приобретает следующий вид:

$$s(k) = \sum_i g_i f_i(k, s(k)), \qquad (2)$$

где g_i – коэффициенты адаптивного разложения; $f_i(k, s(k)$ – априорно неизвестные (апостериорно известные) функции, извлекаемые из исходного сигнала $s(k)$ в соответствии с определенной процедурой, что отражено в зависимости вида базиса от исходного сигнала. В данном случае базисные функции формируются апостериорно, что является преимуществом с точки зрения учета индивидуальных свойств сигнала.

Вышеприведенные формулы наглядно иллюстрируют тот факт, что ДЭМ наиболее эффективно применять для обработки многокомпонентных сигналов, описываемых формулами (1) или (2). Многокомпонентный сигнал – сигнал, состоящий из конечного аддитивного набора разномасштабных компонент различной физической природы, занимающих различные частотные полосы. Под разномасштабностью компонент подразумевается их различное разрешение по времени. По причине разномасштабности целесообразно применение именно мультимасштабного подхода к обработке и дальнейшему анализу таких данных.

На каждом этапе разложения ДЭМ представляет собой суперпозицию медленно меняющейся низкочастотной компоненты и быстро меняющейся высокочастотной компоненты [1, 3, 5, 6, 11, 12].

Представление функции в виде быстро меняющейся и медленно меняющейся компонент проиллюстрировано на рисунке 1.

При этом ДЭМ не делает допущения о гармонической природе осцилляций. Следствием этого является более компактное представление сигнала – с меньшим количеством ЭМ, чем при Фурье или вейвлет-декомпозиции. Аналитически представление сигнала на основе низкочастотной и высокочастотной компонент можно представить в виде:

$$s(k) = d_1(k) + m_1(k),\ m_1(k) = d_2(k) + m_2(k), ...$$

$$s(k) = \sum_{m=1}^{M-1} d_m(k) + m_M(k),$$

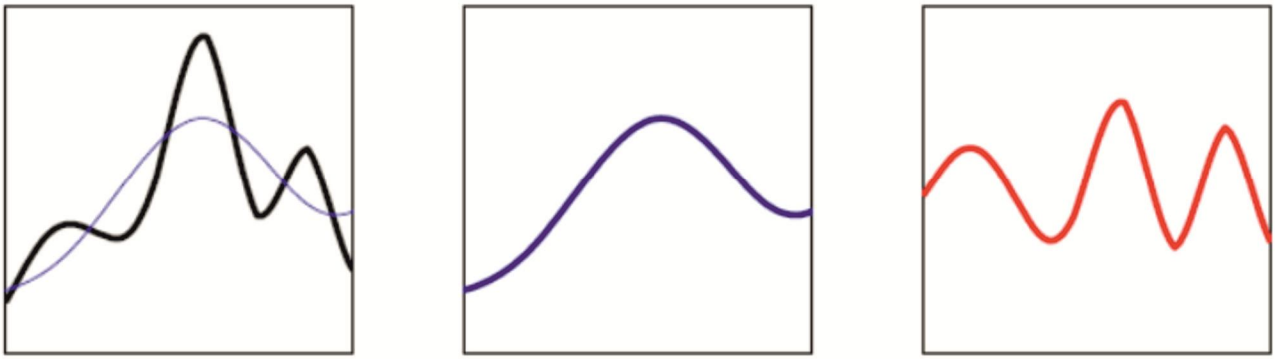


Рисунок 1 – Иллюстрация представления сигнала (слева) в виде суммы низкочастотной, медленно меняющейся компоненты (в центре) и высокочастотной, быстро меняющейся (справа)

где $m_1(k), m_2(k), \ldots$ – медленно меняющиеся низкочастотные компоненты, $d_1(k), d_2(k), \ldots$ – быстро меняющиеся высокочастотные компоненты. Число экстремумов для каждой компоненты $m_1(k), m_2(k), \ldots$ убывает при переходе от одного остатка к другому (остаток – сигнал, из которого извлекается очередная компонента), что гарантирует сходимость алгоритма.

Вычислительная сложность ДЭМ определяется соотношением $O(\log_2 N)$, где N – длина сигнала.

Технология Гильберта-Хуанга

Технология обработки сигналов на основе ДЭМ, называемая технологией Гильберта-Хуанга [1, 3, 11, 12], в настоящее время получила широкое распространение при решении следующих основных задач:

– предварительная обработка сигналов;

– анализ структуры многокомпонентных сигналов;

– частотно-временной анализ сигналов.

Предварительная обработка – комплекс математических операций, направленных на первичную обработку сигнала, его очистку от различного рода шумов и помех, вычисление ряда характеристик (статистические характеристики, фрактальные характеристики, энтропия и пр.). Предварительная обработка включает следующие основные задачи [1, 3, 6, 8-12, 25]:

– очистка сигналов от шумов различной природы;

– выделение тренда в сигналах.

Анализ структуры сигналов включает в себя следующие основные задачи [1, 3, 7, 11, 12]:

– анализ степени хаотичности сигналов на основе оценивания показателя Херста и др. фрактальных характеристик (показателя Гельдера, фрактальной размерности и др.) [1, 3, 11, 12];

– проведение мультиразрешающего анализа (исследование тонкой структуры сигналов и их глобального поведения в целом, анализ свойств отдельных компонент) и мультиполосного анализа (исследование компонент сигнала в различных частотных полосах);

– выделение шаблонов в сигналах и их классификация на основе физической и математической интерпретаций (шумовые, трендовые компоненты, характерные шаблоны и др.).

Процесс формирования эмпирических базисных функций на основе ДЭМ является первым этапом технологии. Далее по апостериорно найденным ЭМ формируется частотно-временное распределение сигналов – так называемый спектр (спектральное представление) Гильберта-Хуанга.

Основные задачи, решаемые путем исследования построенного спектра Гильберта-Хуанга [1, 3, 11, 12]:

– выявление скрытых в шумах модуляций сигнала (амплитудных, частотных);

– идентификация частотных и временных диапазонов концентрации энергии;

– классификация множества сигналов по спектру Гильберта-Хуанга;

– расчет на основе спектра Гильберта-Хуанга так называемого маргинального спектра путем численного интегрирования по времени спектра Гильберта-Хуанга, являющегося аналогом частотного спектра Фурье для нестационарных сигналов;

– расчет на основе спектра Гильберта-Хуанга так называемой мгновенной плотности энергии путем численного интегрирования по частоте квадрата спектра Гильберта-Хуанга, позволяющей наглядно выявлять изменения энергетических соотношений в сигналах.

Понятие эмпирической моды

Эмпирическая мода (ЭМ) – функция, заданная непрерывно на интервале существования сигнала или дискретно в виде вектора отсчетов, имеющая в общем случае произвольную форму и произвольную аналитическую запись (ЭМ во многих случаях не подчиняется аналитической зависимости), но при этом удовлетворяющая двум необходимым условиям [1, 3-15]:

1) общее число максимумов и минимумов такой функции (общее число экстремумов) должно быть строго равно числу нулей функции либо отличаться от числа нулей исходной функции по модулю не более, чем на единицу:

$$N_{\max} + N_{\min} = N_{zero} \text{ или } N_{\max} + N_{\min} = N_{zero} \pm 1, \quad (3)$$

где $N_{\max}, N_{\min}, N_{zero}$ – число максимумов, минимумов и нулей функции соответственно. Величины $N_{\max}, N_{\min}$ не учитывают краевые отсчеты сигнала, которые в некоторых случаях могут оказаться единственными экстремумами сигнала. Нули N_{zero} в дискретном случае могут быть найдены с использованием различных алгоритмов интерполяции функций. В общем случае, величины $N_{\max}, N_{\min}, N_{zero}$ должны определяться численно, поскольку аналитическая запись ЭМ может не существовать. Данное условие также подразумевает, что между двумя экстремумами различного типа (максимумами и минимумами) расположен один нуль кратности 1. Другими словами, имеем следующую структуру:

… минимум – нуль – максимум …;

… максимум – нуль – минимум …;

2) локальное (мгновенное) среднее значение ЭМ, определенное в виде полусуммы верхней и нижней огибающих, должно удовлетворять условию:

$$\frac{U(k)+L(k)}{2} \le \eta,\ k = \overline{0, N-1}, \quad (4)$$

где k – дискретное нормированное время (номер отсчета сигнала); $U(k)$ и $L(k)$ – значения верхней и нижней огибающих сигнала соответственно в k -й дискретный момент времени (огибающие получены путем интерполяции локальных максимумов и минимумов соответственно); N – общее количество отсчетов (длина сигнала); η – некоторый порог, устанавливаемый обработчиком-экспертом и обычно принимающий близкие к нулю значения.

На рисунке 2 проиллюстрированы четыре ЭМ, а также верхняя и нижняя огибающие и их полусумма. Цифрой 1 обозначена верхняя огибающая, 2 – нижняя огибающая, 3 – локальное среднее (полусумма верхней и нижней огибающих), 4 – сама ЭМ. Этот метод построения огибающей был предложен Хуангом [1, 3] и отличается от классического понятия «огибающей». Огибаю-

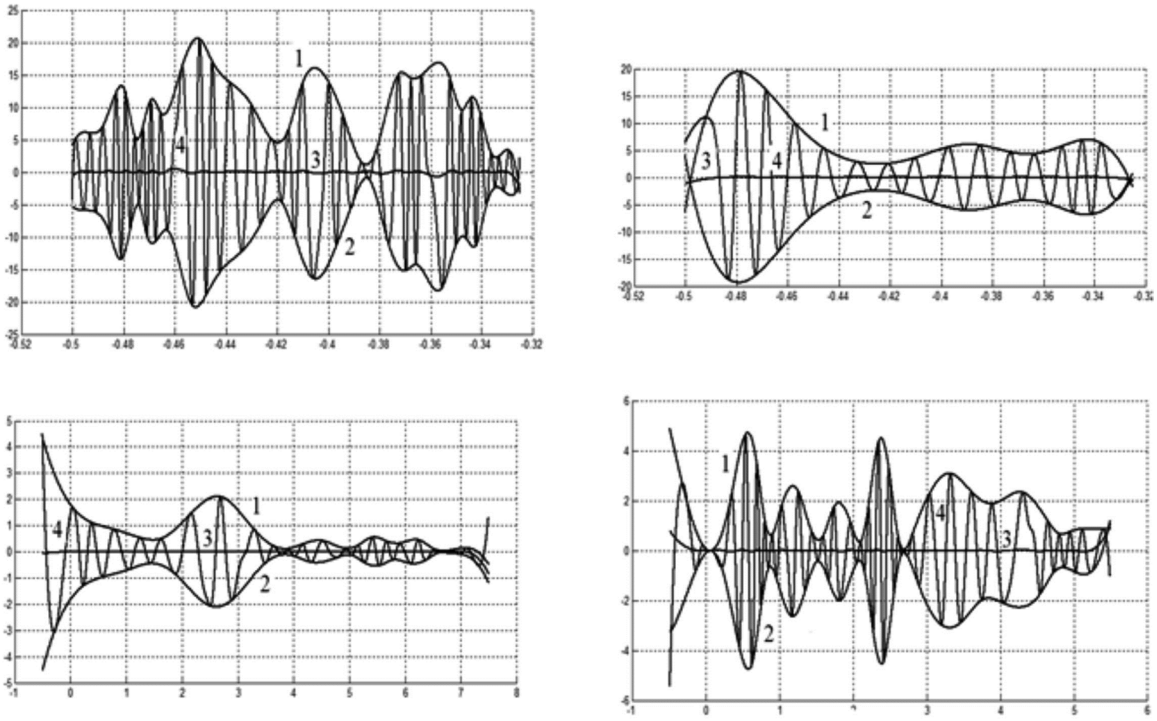


Рисунок 2 – Примеры функций (эмпирических мод) и их огибающих

щие рассчитываются посредством интерполяции экстремумов кубическими сплайнами, однако использованы могут быть и сплайны более высоких порядков, чем 3-й. Необходимо помнить, что высокие порядки сплайнов приводят к значительной вычислительной сложности и поэтому применяются редко.

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

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

где $c(k)$ – ЭМ. Условие равенства нулю локального среднего накладывает ограничения на симметрию огибающих относительно оси абсцисс:

$$U(k) = -L(k);\ \ k = \overline{0, N-1},$$

что для реальных сигналов может быть заменено на

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

где γ - некоторая малая константа.

Функция, удовлетворяющая 1-у условию, но не удовлетворяющая 2-у, является «слабой» ЭМ [11, 12, 16]. Слабая ЭМ должна иметь число максимумов $N_{\max}$ и число минимумов $N_{\min}$, совпадающее с числом нулей с точностью до 1, но при этом может обладать несимметричными огибающими:

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

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

ЭМ, которая может быть задана аналитически, называется также характеристической модой.

Описание эмпирических мод на основе дифференциальных уравнений

Как было отмечено ранее, среди класса ЭМ можно выделить «слабые» ЭМ [16]. При этом «слабая» ЭМ удовлетворяет дифференциальному уравнению [16]:

$$\frac{d}{dt}\left(P(t)\frac{df(t)}{dt} \right) + Q(t)f(t) = 0, \quad (5)$$

где время $t \in (a;b)$, a, b – границы интервала, $P(t) > 0$, $Q(t)$ – непрерывные функции. Более общим является дифференциальное уравнение Штурма-Лиувилля, которому также должна удовлетворять «слабая» ЭМ [16]:

$$\frac{d}{dt}\left(p(t)\frac{df(t)}{dt} \right) + \left(\lambda\rho(t) - q(t) \right) f(t) = 0, \quad (6)$$

Если $Q(t) > 0$, то решение уравнения (5) имеет ровно один максимум или один минимум между двумя последовательно расположенными нулями, а, следовательно, решения дифференциальных уравнений (5) и (6) удовлетворяют определению «слабой» ЭМ.

В том случае, если ЭМ $f(t)$ принадлежит множеству дважды дифференцируемых функций на $[a;b]$ и функция $f(t)$ и ее первая производная $f'(t)$ имеют только простые нули, следующие утверждения являются эквивалентными:

– число нулей и экстремумов функции $f(t)$ на интервале $[a;b]$ отличаются по модулю не более, чем на 1.

– существуют положительные дважды дифференцируемые функции $P(t)$, $Q(t)$ такие, что $f(t)$ является решением дифференциального уравнения.

Классический метод ДЭМ и его отдельные применения

Алгоритм разложения сигнала и извлечения из него ЭМ можно представить в виде следующей последовательности шагов [1, 3, 11, 12, 26, 27].

Шаг 1. Рассматривается текущий остаток $r_p(k)$, $p = 1, 2, \ldots$.

Определяются экстремумы и нули $r_p(k)$ и формируются два конечных множества:

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

где $\{M_i\}, \{m_i\}$ – множества максимумов и минимумов функции $r_p(k)$ соответственно.

Если экстремумы отсутствуют в сигнале, то их можно искусственным образом сформировать путем применения операции численного дифференцирования и последующего возврата к исходному множеству значений:

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

$U_j(k) = f_U(M_i, k), k = 0,1,...,N-1;$

$L_j(k) = f_L(m_i, k), k = 0,1,...,N-1;$

где $U_j(k)$ и $L_j(k)$ – верхняя и нижняя огибающие, построенные соответственно по найденным локальным максимумам и минимумам с помощью сплайн-интерполяции; j – номер итерации отсеивания, который будет рассмотрен ниже.

Построенные огибающие должны заключать в себе текущий остаток:

$L_j(k) \le r_p(k) \le U_j(k), \ \forall k.$

В некоторых случаях наблюдаются так называемые «всплески» на k-м отсчете, описываемые формулами:

$r_p(k) \le L_j(k)$ или $r_p(k) \ge U_j(k), \ \forall k.$

После этого определяется полусумма двух огибающих (локальное среднее значение):

$e_j(k) = 0.5\left[U_j(k) + L_j(k)\right].$

Шаг 2. Найденное мгновенное среднее значение $e_j(k)$ вычитается из текущего остатка:

$h_1(k) = r_p(k), \ldots, h_j(k) = h_{j-1}(k) - e_j(k), \ j > 1.$

Далее проверяется выполнение условий (3) и (4). Если оба условия выполняются, то осуществляется переход к шагу 3. Если какое-либо из них нарушено, то осуществляется возврат к шагу 1 и начинается так называемый процесс отсеивания (англ. – Sifting Process):

$h_{j+1}(k) = h_j(k) - e_j(k), j \ge 1, \ldots., c_p(k) = h_{iter_p}(k),$

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):

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

где $c_p(k)$ – полученная ЭМ; $r_p(k)$ – текущий остаток; $r_{p+1}(k)$ – новый остаток.

Шаг 4. Далее осуществляется переход к шагу 1, где в качестве функции, из которой будут извлекаться ЭМ с более высокими номерами, то есть в качестве текущего остатка, выступит тот остаток, который был получен на третьем шаге, то есть $r_{p+1}(k)$.

Представленный алгоритм является классическим алгоритмом ДЭМ.

Ниже приведен псевдокод алгоритма ДЭМ, на базе которого осуществлена его программная реализация.

$r(k) = s(k)$ % Инициализация текущего остатка (на первом этапе – исходный сигнал)

$I_1(k) = r(k)$ % Инициализация процесса отсеивания

$i = 1; p = 1;$ % Инициализация переменных-счетчиков для извлечения ЭМ и формирования текущего остатка в ходе процесса отсеивания

while $r(k) \ne 0$ или $r(k)$ не является монотонной функцией

  while $I_1(k)$ имеет среднее значение, отличное от нуля

    $U(k) =$ кубический сплайн через локальные максимумы $I_i(k)$

    $L(k) =$ кубический сплайн через локальные минимумы $I_i(k)$

    $Av(k) = \dfrac{L(k) + U(k)}{2}$ % Локальное (мгновенное) среднее верхней и нижней огибающих

    $I_i(k) = I_i(k) - Av(k)$ % Итерация процесса отсеивания: вычитание локального (мгновенного) среднего значения

    $i = i + 1$ % модификация переменной-счетчика выполнения процесса отсеивания

  end

$IMF_p(k) = I_i(k)$ % Извлечение очередной *p*-й ЭМ

$r(k) = r(k) - IMF_p(k)$ % Определение текущего остатка

$p = p + 1$ % Модификация переменной-счетчика для извлечения ЭМ

End

Используя результат разложения сигнала $s(k)$ на основе ДЭМ, можно записать связь между всеми извлеченными компонентами и исходным сигналом:

$$s(k) = c_1(k) + c_2(k) + ... + c_M(k),$$

где M – общее количество компонент.

Разложение может быть представлено в виде суммы низкочастотных и высокочастотных компонент, которые следуют в соответствии с представлением ДЭМ в виде диадического банка фильтров [1, 3, 11, 12]:

$$s(k) = c_{1\text{ВЧ}}(k) + ... + c_{m\text{ВЧ}}(k) + c_{(m+1)\text{НЧ}}(k) + ... + c_{Q\text{НЧ}}(k),$$

где $c_{1\text{ВЧ}}(k) + ... + c_{m\text{ВЧ}}(k)$ и $c_{(m+1)\text{НЧ}}(k) + ... + c_{Q\text{НЧ}}(k)$ – высокочастотные и низкочастотные компоненты соответственно.

ДЭМ раскладывает сигнал на набор компонент с амплитудной и частотной модуляциями [1, 3, 11, 12]:

$$s(k) = \text{Re}\left\{\sum_{i=1}^{M-1} a_i(k) e^{j\theta_i(k)}\right\} + r(k),$$

где $a_i(k)$ – амплитудная огибающая, определяющая закон амплитудной модуляции, $\theta_i(k)$ – фазовая функция, определяющая закон фазовой модуляции, $r(k)$ – результирующий остаток. Для отслеживания характера данных модуляций применяется частотно-временное распределение Гильберта-Хуанга [1, 2].

Процесс отсеивания играет важную роль при извлечении конкретной ЭМ из исходного сигнала. Он является итерационным, и для его остановки применяются специальные критерии. Один из наиболее часто используемых определяется выражением [1, 3, 11, 12]:

$$SD = \sum_{k=0}^{N-1} \frac{\left[h_{1(p-1)}(k) - h_{1p}(k)\right]^2}{\left[h_{1(p-1)}(k)\right]^2},$$

где $h_{1(p-1)}(k), h_{1p}(k)$ – две последовательные итерации при извлечении ЭМ на итерациях с номерами $(p-1)$ и p. Значение критерия SD, как правило, должно находиться в пределах от 0.2 до 0.3 для завершения процесса отсеивания. Основным недостатком данного критерия является то, что он не связан напрямую с определением ЭМ, т.е. не учитывает ни число максимумов и минимумов, ни равенство нулю полусуммы верхней и нижней огибающих.

Другой, альтернативный, критерий [11, 12] остановки процесса отсеивания можно записать в следующем виде:

$$\sigma(k) = \left|\frac{m(k)}{a(k)}\right|, \; a(k) = \frac{U(k) - L(k)}{2}; \; m(k) = \frac{U(k) + L(k)}{2};$$

$$\sigma(k) < \theta_1 \text{ для доли } (1-\alpha); \; \sigma(k) < \theta_2 \text{ для остальной доли.}$$

Доля определяется от общей длительности сигнала. Как правило, тройка значений этих параметров следующая: $(\alpha, \theta_1, \theta_2) = (0.05, 0.05, 0.5)$.

Еще один критерий остановки процесса отсеивания, который был апробирован при обработке сейсмических данных [28], имеет вид:

$$SD_1 = \frac{1}{N}\left(\left(\sqrt{\sum_{k=1}^{N}\left(s_{n-1}(k)\right)^2}\right) - \left(\sqrt{\sum_{k=1}^{N}\left(s_n(k)\right)^2}\right)\right) < tol,$$

$$s_{n+1}(k) = s_n(k) - m_n(k),$$

где tol – соответствующий порог.

Пример выполнения ДЭМ мультигармонического сигнала с шумом показан на рисунке 3. Исходный сигнал показан в верхней части рисунка, шумовые ЭМ имеют номера с 1-й по 6-ю, информативные ЭМ-шаблоны (дискретные гармоники) 7-я и 8-я, трендовые компоненты – с 9-й по 11-ю.

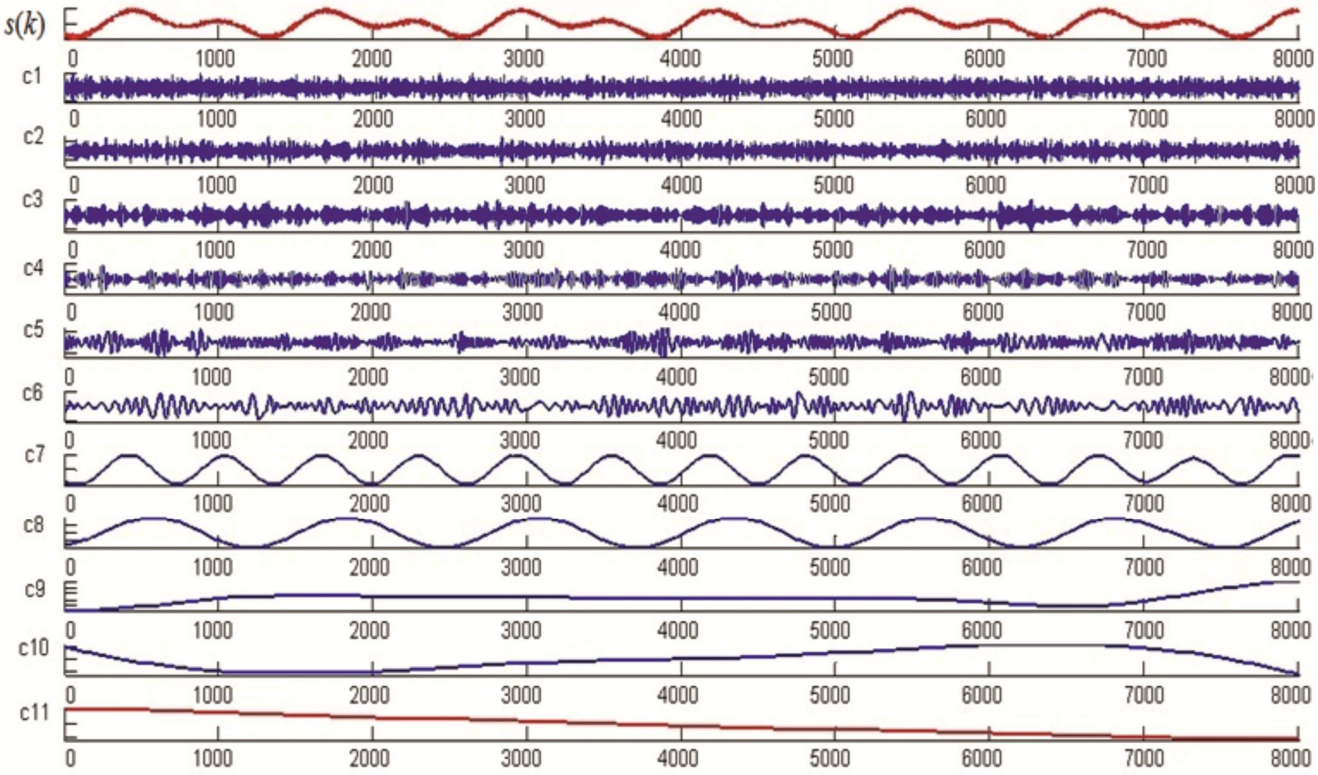


Рисунок 3 – Пример выполнения ДЭМ мультигармонического сигнала с шумом

В [29] предложен похожий на ДЭМ адаптивный метод декомпозиции сигналов на основе экстремальной фильтрации. Экстремальная фильтрация – альтернатива ДЭМ. При этом данный алгоритм осуществляет разложение на знакопеременные составляющие на основе адаптивной фильтрации сигнала по его экстремальным значениям.

Экстремальная фильтрация представляет собой параллельное выполнение низкочастотной и высокочастотной фильтрации [29]. Сглаживание обеспечивается следующим оператором:

$$y_{ci} = 0.25y_{эi-1} + 0.5y_{эi} + 0.25y_{эi+1},$$

где $y_{эi}$ – экстремальные значения в моменты времени $t_{эi}$, $i = 1, K$. Выделение знакопеременной составляющей осуществляется оператором:

$$y_{pi} = -0.25y_{эi-1} + 0.5y_{эi} - 0.25y_{эi+1}.$$

Развитие своего методы авторы привели в [30], где представлен метод выделения компонент на основе дифференцирования и интегрирования, позволяющий извлекать компоненты как в порядке возрастания, так и в порядке убывания частот.

ДЭМ представляет собой математически двойной (вложенный) итерационный процесс. Число внешних итераций можно заранее определить с достаточно высокой точностью, а число внутренних итераций определяется на основе специальных критериев [1, 3, 11, 12]. ДЭМ применяется как к стационарным, так и нестационарным сигналам и временным рядам, что опять-таки обусловило широкое применение метода в различных областях.

Применение метода ДЭМ к данным сопряжено с понятием «огибающей» сигнала. Она строится на основе локальных максимумов и локальных минимумов (таким образом, формируются верхняя и нижняя огибающие сигнала соответственно). Построение огибающих продемонстрировано на рисунке 2.

Метод ДЭМ широко используется в различных областях, включая обработку геофизических, биомедицинских, вибрационных, акустических, музыкальных и других типов сигналов [1, 3, 11, 12, 15, 28, 31].

Одно из применений ДЭМ в области геофизики приведено в [28, 32]. В биомедицинских приложениях ДЭМ применяется в кардиологии при обнаружении QRS-комплекса в данных электрокардиограммы (ЭКГ), а также P и T-волн [33-39]. Выделение QRS-комплекса – наиболее важный шаг при анализе сигналов ЭКГ. ЭМ с начальными номерами в разложении захватывают быстро осциллирующие процессы, а ЭМ с конечными номерами захватывают медленно осциллирующие процессы. Высокочастотные ЭМ используются для выявления QRS-комплекса и его отличия от P и T волн, а также шума и дрифта. Сглаженный ЭКГ-сигнал может быть получен из ЭМ высоких порядков, где могут быть выделены P и T волны. В работе [35] также представлен алгоритм выявления максимума R. Границы QRS-комплекса, являющиеся информативными параметрами, находятся на основе анализа первых 3-х ЭМ. В работе [40] предложен метод оценки качества ЭКГ-записей на основе ДЭМ и машинного обучения, и авторами достигнута высокая точность при использовании предложенной метрики качества и классификатора на основе решающих деревьев.

Актуальные вопросы анализа и обработки речевых сигналов, в том числе для анализа психоэмоционального состояния человека [41], на основе ДЭМ рассмотрены в [42, 43]. Исследована [42] проблема повышения эффективности анализа речевых сигналов в алгоритме оценивания частоты основного тона, что относится к группе задач предварительной обработки. Наилучшие результаты получены на основе метода комплементарной множественной ДЭМ, что находит практическое применение в алгоритмах оценивания частоты основного тона. Кроме того, исследовано распределение мощности речевого сигнала по ЭМ, извлекаемым из сигнала в процессе декомпозиции [41, 43].

Приведем результаты сравнения ДЭМ с вейвлет-технологией обработки сигналов в частотно-масштабном пространстве, а также обработкой сигналов на основе спектрального анализа Фурье.

Таблица 1 – Сравнение ДЭМ с основными методами-аналогами

Метод обработки сигналов / Критерий сравнения	ДЭМ	Вейвлет-технология	Спектральный анализ Фурье
Тип базиса	апостериорный	априорный	априорный
Способ представления сигналов	частотно-временное представление Гильберта-Хуанга / уровни декомпозиции	время-масштабная скалограмма / уровни вейвлет-декомпозиции	спектральная плотность / спектральная плотность мощности / Фурье-спектрограмма
Тип обрабатываемых сигналов	стационарные / нестационарные	стационарные / нестационарные	стационарные
Извлечение локальных особенностей во временной области	присутствует	присутствует	отсутствует
Теоретическая база	частично разработана	полностью разработана	полностью разработана
Наличие компактного носителя	есть	есть	есть (в частотной области)
Сходимость	строго не доказана	доказана аналитически	доказана аналитически
Наличие быстрых вычислительных алгоритмов	есть	есть	есть

Укажем основные преимущества метода ДЭМ:

– возможность формирования базиса разложения адаптивно непосредственно из исходного сигнала с учетом его структуры и локальных особенностей, что особенно важно для многочисленного класса нестационарных сигналов;

– наличие близкого соответствия между компонентами, извлеченными на основе ДЭМ, и реальными физическими процессами и явлениями в различных прикладных задачах;

– возможность проведения адаптивной предварительной обработки сигналов;

– ДЭМ – один из наиболее эффективных методов при проведении мультиразрешающего (кратномасштабного) анализа и исследовании сигналов со сложной многокомпонентной структурой, включающей разно масштабные осциллирующие составляющие на фоне шума;

– ДЭМ позволяет выполнять синтез банка фильтров, используемого для проведения мультиполосного анализа сигналов (анализа в различных частотных полосах);

– возможность расчета частотно-временного распределения Гильберта-Хуанга и решения ряда задач на его основе (классификация сигналов, расчет маргинального спектра и мгновенной плотности энергии для нестационарных сигналов, выявление скрытых в шумах модуляций).

К некоторым недостаткам метода ДЭМ можно отнести следующие:

– довольно значительные вычислительные затраты при обработке сигналов большой длины и при высокой частоте дискретизации ввиду необходимости определения коэффициентов сплайнов при выполнении кубической сплайн-интерполяции. При использовании альтернативных методов интерполяции может проявляться избыточность в разложении (появление дополнительных неинформативных компонент);

– краевые эффекты сплайн-интерполяции могут приводить к осцилляциям на границах сигнала. Их можно уменьшать путем зеркального отображения экстремумов, с помощью использования нейронных сетей, прогнозирования на основе авторегрессионных моделей, вейвлетов, метода, основанного на опорных векторах [44];

– определенная зависимость результата декомпозиции от выбора критерия остановки процесса отсеивания;

– в ряде случаев наблюдается низкое временное разрешение при обработке сигналов с малыми значениями (невозможность выделить осциллирующие компоненты из анализируемого сигнала) и зависимость разрешения по частоте от структуры соответствующего диадического банка фильтров.

ДЭМ с параболической интерполяцией огибающих

Наряду с классическим алгоритмом ДЭМ возможно также использование параболической интерполяции огибающих. Данный метод позволяет определить экстремумы сигнала на каждом этапе извлечения ЭМ с более высокой точностью и избежать тем самым избыточности в разложении [8, 11, 12]. Алгоритм можно представить в виде следующей последовательности шагов:

1) выбираются 3 первых отсчета исходного сигнала $s(k)$;

2) по 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);

3) если коэффициент $a \neq 0$, то среди трех точек есть экстремум, который располагается посередине (вторая точка в последовательности из трех точек). Координата экстремума уточненным способом определяется как координата вершины параболы $x_0 = -b/2a$;

4) Если $a = 0$, то все 3 точки лежат на одной прямой, и среди них нет экстремума. Далее необходимо сдвинуть скользящее окно на 1 точку вправо и повторить ту же последовательность действий.

5) по найденным экстремумам строятся огибающие и далее выполняется процедура в соответствии с классическим алгоритмом ДЭМ.

На рисунке 4 показаны примеры параболической ДЭМ для моногармонического сигнала (рисунок 4, *а*) и мультигармонического сигнала (рисунок 4, *б*), состоящего из 3-х дискретных гармоник.

В результате применения процедуры декомпозиции извлечены, соответственно, 1 гармоника (риснок 4, а) и 3 гармоники (рисунок 4, *б*) и отсутствуют избыточные компоненты.

Следует отметить, что на практике условие алгоритма записывается в виде $|a| \geq \varepsilon$ - наличие экстремума, $|a| < \varepsilon$ - отсутствие экстремума, где ε – порог для определения экстремумов.

В данной модификации точность определения экстремумов повышается за счет определения их местоположения как координаты вершины параболы.

К основным преимуществам параболической ДЭМ по сравнению с классической ДЭМ можно отнести следующие:

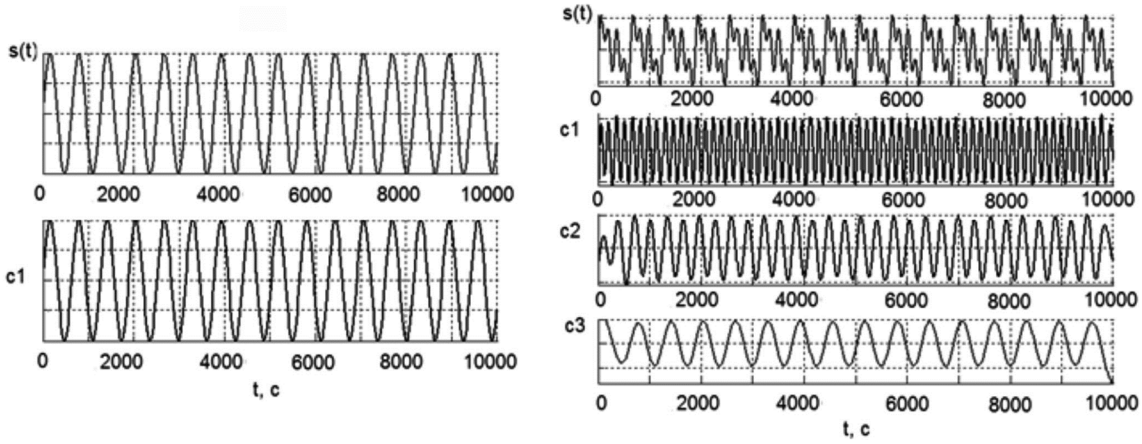
– более точное определение местоположения экстремумов (меньшее влияние ошибок при определении экстремумов) при построении верхней и нижней огибающих и за счет этого уменьшение избыточности в разложении (появления при ДЭМ дополнительных неинформативных составляющих);

– меньшее проявление «всплесков» (резкое превышение значением функции среднего уровня в промежутке между узлами интерполяции) на границах исследуемого сигнала и при извлечении отдельных компонент;

– меньшее влияние краевых эффектов (осцилляции в окрестности точек начала и окончания сигнала) при разложении;

– возможность в ряде случаев более точного решения задач предварительной обработки сигналов, в частности, задачи очистки сигналов от шума, а также задач структурного анализа, в частности, выделения информативных компонент-шаблонов.

К недостаткам параболической ДЭМ можно отнести



а) *б)*

Рисунок 4 – Пример параболической ДЭМ гармонического сигнала: а – с извлечением одной ЭМ; б – с извлечением 3-х ЭМ

более сложный алгоритм поиска экстремумов, по сравнению с классическим методом ДЭМ.

Множественная (ансамблевая) декомпозиция на эмпирические моды

Метод ансамблевой ДЭМ (англ. EEMD – Ensemble Empirical Mode Decomposition) позволяет в ряде случаев повысить точность выполнения разложения и избежать проблемы «смешивания» ЭМ (англ. – Mode Mixing) [45, 46], что может привести к потере физического смысла отдельных извлеченных из сигнала компонент. Эффекты разложения с использованием множественной ДЭМ состоят в том, что реализации белого гауссовского шума нейтрализуют друг друга, а средние ЭМ остаются в пределах интерпретации двоичного (диадического) банка фильтров.

Как правило, методы анализа данных и обработки сигналов направлены на избавление от шума, однако шум специально добавляется для облегчения анализа данных, обнаружения слабых сигналов и выявления скрытых процессов. Исходя из сказанного, алгоритм множественной ДЭМ может быть представлен в виде следующей последовательности шагов:

1) добавление белого гауссовского шума $w(k)$ к исходному сигналу $s(k)$:

$$s_p(k) = s(k) + w_p(k),\ p = \overline{1,P}, \quad (7)$$

где $s(k)$ – p-я реализация смеси полезного сигнала и белого гауссовского шума, P – общее число реализаций шума.

Для ансамбля реализаций белого шума соотношение (7) примет вид:

$$\{s_p(k)\}_{p=1}^{P} = s(k) + \{w_p(k)\}_{p=1}^{P},\ p = \overline{1,P},$$

2) вычисление ДЭМ для каждой сформированной реализации $s_p(k)$ отдельно и получение M_n аддитивных компонент, которые можно обозначить в виде множества $\{c_m^p(k)\}_{m=1}^{M_n}$.

3) расчет средних значений соответствующих ЭМ:

$$\overline{c_m(k)} = \frac{1}{N}\sum_{p=1}^{P} c_m^p(k).$$

К основным преимуществам множественной ДЭМ можно отнести следующие:

– гораздо меньшее «смешивание» различных ЭМ и более четкое разделение извлекаемых компонент по частотным полосам, что обеспечивает физический смысл извлекаемых компонент;

– возможность извлечения компонент из сигналов с малыми значениями, что не всегда достигается при использовании классической и параболической ДЭМ.

К недостаткам множественной ДЭМ можно отнести следующие:

– более высокие вычислительные затраты на построение огибающих за счет увеличения общего числа экстремумов вследствие формирования реализаций с белым шумом;

– в ряде случаев возможно появление дополнительных высокочастотных составляющих вследствие влияния высокочастотного белого шума.

Улучшенный метод ДЭМ на основе вейвлет-пакетного разложения

При использовании классического алгоритма ДЭМ в разложении могут появляться ряд недостатков:

1) нежелательные низкочастотные составляющие (ЭМ), появляющиеся вследствие влияния краевых эффектов, выбора конкретного критерия остановки процесса отсеивания и т.д. Это затрудняет интерпретацию разложения в терминах данной предметной области.

2) у 1-й ЭМ, являющейся самой высокочастотной, может быть слишком широкий частотный диапазон и тем самым не будут выполняться свойства монокомпонентного сигнала.

В связи с этим возможно использование вейвлет-пакетного преобразования [15, 23, 24] для декомпозиции сигнала на набор узкополосных сигналов перед применением ДЭМ. При этом каждая извлеченная ЭМ становится, в свою очередь, монокомпонентной. В этом случае можно определить все мгновенные частоты у нестационарного сигнала.

Улучшенный алгоритм ДЭМ состоит из следующих шагов:

– вейвлет-пакетное разложение исходного сигнала $s(k)$ и получение набора базисных функций $s_i, i = 1,2,...,n;$

– выполнение алгоритма ДЭМ и получение набора ЭМ $\{c_i(k)\}$;

– выбор информативных компонент в разложении в соответствии с определенным критерием;

– вычисление спектра Гильберта-Хуанга и построение единого частотно-временного распределения.

В работе [15] показано, что при использовании классической ДЭМ для мультигармонического сигнала, состоящего из двух дискретных гармоник (двух тонов), получаются 2 ЭМ и 3 псевдокомпоненты в нихкочастотной области, что может привести к ошибочной интерпретации разложения. В свою очередь, алгоритм на основе вейвлет-пакетного разложения показал себя эффективным [15] при исследовании вибраций в системе «роторстатор» двигателя и прогнозировании возможных разрушений. В ульучшенном алгоритме ДЭМ на основе вейвлет-пакетного разложения проявляются скрытые низкочастотные компоненты, которые не видны в исходном алгоритме частотно-временного распределения Гильберта-Хуанга. Они позволяют выявлять скрытые предвестники разрушений, которые не видны при использовании классического ДЭМ.

Таким образом, к основным преимуществам ДЭМ на основе вейвлет-пакетного разложения можно отнести следующие:

– возможность более точного, чем при классической ДЭМ, восстановления (извлечения) из исходного сигнала низкоэнергетических компонент с малыми амплитудами, что является информативным в ряде прикладных задач;

– возможность более «тонкого» деления банка фильтров, соответствующего декомпозиции, на частотные полосы вследствие использования вейвлет-пакетного разложения. Как следствие, возможно более детальное проведение мультиполосного анализа и выделение компонент, соответствующих различным частотным полосам;

Некоторым недостатком данной модификации, по сравнению с классической ДЭМ, является более сложный с вычислительной точки зрения алгоритм за счет дополнительного использования вейвлет-пакетного разложения. Кроме того, данная модификация требует дополнительных исследований в прикладных задачах.

Многомерный метод ДЭМ

Данная модификация алгоритма ДЭМ может быть представлена в следующем виде [47, 48-50]:

1) выбор соответствующего набора точек для дискретизации на $(n-1)$ -мерной сфере.

2) вычисление проекции $\{p^{\theta_k}(t)\}_{t=1}^{T}$ входного сигнала $\{v(t)\}_{t=1}^{T}$ вдоль направляющего вектора X^{θ_k} для всех значений k, при этом рассматривается полный набор направляющих векторов. В данном случае $\{p^{\theta_k}(t)\}_{t=1}^{T}$ – множество проекций.

3) нахождение временных отсчетов $\{t_i^{\theta_k}\}$, соответствующих максимумам множества проектируемых сигналов $\{p^{\theta_k}(t)\}_{k=1}^{K}$.

4) интерполяция $[t_i^{\theta_k}, v(t_i^{\theta_k})]$ для получения многомерных кривых огибающей $\{e^{\theta_k}(t)\}_{k=1}^{K}$.

5) расчет для множества K направляющих векторов среднего значения $m(t)=\frac{1}{K}\sum_{k=1}^{K} e^{\theta_k}(t)$.

6) извлечение детализирующей составляющей $d(t)=x(t)-m(t)$. Если рассчитанные детали удовлетворяют критерию остановки процесса отсеивания для многомерной ЭМ, применить процедуру с шагами 1-5 для сигнала $d(t)=x(t)-m(t)$, иначе применить их к $d(t)$.

Многомерный метод ДЭМ предназначен для обработки многомерных данных, в том числе изображения. Он позволяет строить многомерные огибающие при реализации разложения и решать различные задачи предварительной обработки, структурного и частотно--временного анализа. Основным преимуществом многомерной модификации являются ее адаптивные свойства за счет извлечения многомерных ЭМ непосредственно из исходного многомерного сигнала.

Заключение

Рассмотрен классический метод ДЭМ и его модификации в составе технологии Гильберта-Хуанга, которые могут применяться в различных задачах обработки и анализа одномерных и многомерных данных, включая обработку биомедицинских, геофизических, вибрационных и других видов сигналов. Метод ДЭМ и его модификации обладают высокой адаптивностью, т.к. базисные функции извлекаются непосредственно из анализируемого одномерного или многомерного сигнала и не требуется априорная информация о выборе базиса.

Метод ДЭМ и его модификации дают во многих случаях преимущества в точности анализа и обработки сигналов за счет адаптивных свойств, поскольку базис разложения формируется непосредственно из исходного сигнала. ДЭМ позволяет проводить предварительную обработку стационарных и нестационарных многокомпонентных сигналов и частотно-временной анализ с высоким разрешением. За счет извлечения базисных функций непосредственно из исходного сигнала можно интерпретировать результаты декомпозиции в терминах конкретной предметной области, включая биомедицину, геофизику, обработку вибраций, акустику и пр.

Основные известные модификации метода ДЭМ, а именно классический метод ДЭМ, ДЭМ с параболической интерполяцией огибающих, ДЭМ на основе вейвлет-пакетного разложения сигналов, ансамблевая (множественная) ДЭМ и многомерная модификация метода позволяют осуществлять выбор нужного метода в конкретной задаче с учетом описанных в статье их преимуществ и недостатков.

Литература

1. N. E. Huang, et al. The empirical mode decomposition and the Hilbert spectrum for non-linear and non-stationary time series analysis. Proc. Royal Soc. London, vol. 454, 1998, pp. 903-995.

2. С.Л. Марпл-мл. Цифровой спектральный анализ и его приложения, пер. с англ. Изд. Мир, 1990, 584 с.

3. N.E. Huang, S. S. P. Shen Hilbert-Huang transform and its applications. World Scientific, 2005. 350 p.

4. P. Flandrin, G. Rilling, P. Gonsalves Empirical mode decomposition as a filter bank. IEEE Signal Process. Lett, vol. 11, 2004, pp. 112-114.

5. Д.М. Клионский, Н.И. Орешко, В.В. Геппенер Декомпозиция на эмпирические моды и ее использование при анализе дробного броуновского движения. Цифровая обработка сигналов, № 3, 2008, с. 37-45.

6. D.M. Klionsky, N.I. Oreshko, V.V. Geppener Applications of empirical mode decomposition for processing nonstationary signals. Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications, vol. 18, no. 3, 2008, pp. 390-399.

7. Д.М. Клионский, И.В. Неунывакин, Н.И. Орешко, В.В. Геппенер Декомпозиция на эмпирические моды и ее применение для идентификации информативных компонент и прогнозирования значений сигналов с использованием нейронных сетей. Нейрокомпьютеры: разработка, применение, №6, 2010, с. 43-50.

8. Д. М. Клионский, Н. И. Орешко, В. В. Геппенер Декомпозиция на эмпирические моды с параболической интерполяцией огибающих в задачах очистки сигналов от шума. Цифровая обработка сигналов, № 2, 2011, с. 51-60.

9. В.В. Егоров, Д.М. Клионский Некоторые вопросы применения адаптивного метода декомпозиции нестационарных сигналов на эмпирические моды. Цифровая обработка сигналов, № 3, 2023, с. 52-58.

10. D. Klionskiy, M. Kupriyanov, D. Kaplun Signal denoising based on empirical mode decomposition. Journal of Vibroengineering, vol. 19, no. 7, 2017, p. 5560-5570.

11. Н.И. Орешко, Д.М. Клионский, В.В. Геппенер, А.В. Экало. Декомпозиция на эмпирические моды в цифровой обработке сигналов. Изд. СПбГЭТУ «ЛЭТИ», 2013, 164 с.

12. Д.М. Клионский, В.В. Геппенер Технология Гильберта-Хуанга и ее применение в цифровой обработке сигналов. Изд. СПбГЭТУ «ЛЭТИ», 2019, 150 с.

13. N.E. Huang, Z. Shen, S. R. Long A new view of nonlinear water waves: the Hilbert spectrum. Annual Review of Fluid Mechanics, 1999, рз. 417-457.

14. Y. Chen, M.Q. Feng. A technique to improve the empirical mode decomposition in the Hilbert-Huang transform. Earthquake Engineering and Engineering Vibration, vol. 2, no. 1, 2003, pp. 75-85.

15. Z. K. Peng, P. W. Tse, F. L. Chu An improved Hilbert-Huang transform and its application in vibration signal analysis. Journal of Sound and Vibration, vol. 286, no. 1-2, 2005, pp. 187-205.

16. R.C. Sharpley, V. Vatchev. Analysis of the intrinsic mode functions. Industrial Mathematics Institute, Department of Mathematics, Institute of South California, vol. 12, 2004. 31 p.

17. P. Goncalves, P. Abry, G. Rilling, P. Flandrin Fractal dimension estimation: empirical mode decomposition versus wavelets. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '07), 2007, pp. 1153-1156.

18. А.И. Солонина, Д.М. Клионский, Т.В. Меркучева, С.Н. Перов. Цифровая обработка сигналов и MATLAB. Изд. БХВ-Петербург, 2013, 512 с.

19. А.И. Солонина. Цифровая обработка сигналов в зеркале MATLAB. Изд. БХВ-Петербург, 2018, 560 с.

20. В. Уидроу, С. Стирнз Адаптивная обработка сигналов. Изд. Радио и связь, 1989, 440 с.

21. К.Ф. Коуэн, П.М. Грант. Адаптивные фильтры. Изд. Мир, 1988, 392 с.

22. В.И. Джиган. Адаптивная фильтрация сигналов: теория и алгоритмы. Изд. Техносфера, 2013, 528 с.

23. Н.К. Смоленцев Вейвлет-анализ в Matlab. 3-е изд. Изд. ДМК Пресс, 2010, 448 с.

24. D.B. Percival, A.T. Walden. Wavelet methods for time series analysis. Cambridge University Press, 2000, 594 p.

25. G. Tsolis, T.D. Xenos. Signal denoising using empirical mode decomposition and higher order statistics. International Journal of Signal Processing, Image Processing and Pattern Recognition, vol. 4, no. 2, 2011, pp. 91-106.

26. S. Meignen, V. Perrier. A new formulation for empirical mode decomposition based on constrained optimization. IEEE Signal Processing Letters, vol. 14, no. 12, 2007, pp. 932-935.

27. P. Flandrin, P. Goncalves, G. Rilling Detrending and denoising with empirical mode decomposition, 12th European Signal Processing Conference, 2004, pp. 1581-1584.

28. B.M. Battista, C. Knapp, T. McGee, V. Goebel. Application of the empirical mode decomposition and Hilbert-Huang transform to seismic reflection data. Geophysics, vol. 72, no. 2, 2007, pp. H29-H37.

29. Н.В. Мясникова, М.П. Берестень. Разложение на эмпирические моды на основе экстремальной фильтрации. Цифровая обработка сигналов, № 4, 2014, с. 13-17.

30. Н.В. Мясникова, Л.А. Долгих, М.Г. Мясникова. Применение разложения по эмпирическим модам в задачах цифровой обработки сигналов. Датчики и системы, № 5, 2011, с. 8-10.

31. P. Heydarian, J.D. Reiss. Extraction of long-term rhythmic structures in musical signals using the empirical mode decomposition. Proc. of the 8th Int. Conference on Digital Audio Effects (DAFX-05), 2005, pp. 258-261.

32. А.С. Долгаль, Л.А. Христенко. Применение эмпирической модовой декомпозиции при обработке геофизических данных. Известия Томского политехнического университета. Инжиниринг георесурсов, Т. 328, № 1, 2017, с. 100-108.

33. E.P. Souza Neto, M.A. Custaud, J.C. Cejka, P. Abry, J. Frutoso, C. Gharib, P. Flandrin Assessment of cardiovascular autonomic control by the empirical mode decomposetion. Methods of Information in Medicine, vol. 43, no. 1, pp. 60-65.

34. Z.E.H. Slimane, A. Naït-Ali QRS complex detection using empirical mode decomposition. Digital Signal Processing, vol. 20, no. 4, 2010, pp. 1221-1228.

35. K.M. Chang. Arrhythmia ECG noise reduction by ensemble empirical mode decomposition. Sensors, no. 10, 2010, pp. 6063-6080.

36. В.В. Бороноев, В.Д. Омпоков, В.А. Козин. Эмпирическая модовая декомпозиция пульсовых сигналов.

Вестник ВСГУТУ. № 1, 2015, с. 40-43.

37. Л.Ю. Кривоногов, А.Ю. Тычков. Подавление помех в электрокардиосигналах на основе разложения по эмпирическим модам. Известия ЮФУ. Технические науки. № 8, 2010, с. 127-132.

38. Л.Ю. Кривоногов. Анализ и обработка эмпирических мод с целью подавления помех в электрокардиосигналах. Известия ЮФУ. Технические науки. № 9, 2012, с. 119-125.

39. Л.Ю. Кривоногов. Метод и алгоритмы помехоустойчивой обработки электрокардиосигналов на основе эмпирической модовой декомпозиции. Известия ЮФУ. Технические науки. № 10, 2014, с. 104-114.

40. А.В. Козюра, К.В. Подмастерьев. Контроль качества электрокардиографических записей с использованием эмпирической модовой декомпозиции / Метрология в биомедицинской инженерии. № 5, 2013, с. 54-62.

41. А.К. Алимурадов, А.Ю. Тычков, П.П. Чураков. Оценка психоэмоционального состояния человека на основе декомпозиции на эмпирические моды и кепстрального анализа речевых сигналов. Вестник Пензенского гос. университета. № 2, 2018, с. 89-95.

42. А.К. Алимурадов. Исследование частотно-избирательных свойств методов декомпозиции на эмпирические моды для оценки частоты основного тона речевых сигналов. Труды МФТИ. Т. 7, № 3, 2015, с. 56-68.

43. А.Л. Приоров, П.О. Павловичев. Исследование распределения мощности речевого сигнала при декомпозиции на эмпирические моды. Системы синхронизации, формирования и обработки сигналов. № 4, 2017. с. 70-75.

44. J. Cheng, D. Yu, Yu Yang Application of support vector regression machines to the processing of end effects of Hilbert–Huang transform. Mechanical Systems and Signal Processing, vol. 21, 2007, pp. 1197-1211.

45. Z. Wu, N.E. Huang. Ensemble empirical mode decomposition: a noise-assisted data analysis method. Advances in Adaptive Data Analysis, vol. 1, no. 1, 2009, pp. 1-41.

46. N.U. Rehman, C. Park, N.E. Huang, D.P. Mandic. EMD via MEMD Multivariate noise-aided computation of standard EMD Advances in Adaptive Data Analysis, vol. 5, no. 2, 2013, pp. 1-25.

47. M.R. Thirumalaisamy, P.J. Ansell. Fast and adaptive empirical mode decomposition for multidimensional, multivariate signals. IEEE Signal Processing Letters, vol. 25, no. 10, 2018, pp. 1550-1554.

48. G. Rilling, P. Flandrin, P. Goncalves, J. M. Lilly Bivariate empirical mode decomposition. IEEE Signal Processing Letters, vol. 14, no. 12, 2007, pp. 936-939.

49. N. Rehman, D.P. Mandic. Multivariate empirical mode decomposition. Proc. R. Soc. A, vol. 466, 2010, pp. 1291-1302.

50. Z. Wu, N. Huang, X. Chen The multidimensional ensemble empirical mode decomposition method. Advances in Adaptive Data Analysis, vol. 1, no. 3, 2009, pp. 339-372.

НОВЫЕ КНИГИ



Витязев В.В., Волченков В.А., Овинников А.А., Лихобабин Е.А.

Цифровая обработка сигналов: Учебное издание для вузов, М.: Изд-во «Горячая линия-Телеком», 2023 г. 188 с.: ил.

Рекомендовано Научно-методическим советом ФГБОУ ВО «Рязанский государственный Радиотехнический университет им. В.Ф. Уткина» в качестве учебного пособия для студентов обучающихся по программе высшего образования по направлениям 11.03.01 – «Радиотехника» и 11.03.02 – «Инфокоммуникационные технологии и системы связи», магистров по 09.04.04 – «Программная инженерия» и специалистов по специальности 12.05.01 – «Электронные и оптико-электронные приборы и системы специального назначения».

Изложены общие принципы цифровой обработки сигналов. Основное внимание уделено проектированию цифровых фильтров в классе КИХ- и БИХ-цепей. Приводятся теоретические сведения, касающиеся характеристик фильтров, приводятся подробные примеры их проектирования и анализа их характеристик. Рассматриваются инструменты моделирования прохождения сигналов через цифровые фильтры в среде SimInTech. Использование среды SimInTech для проектирования цифровых фильтров и моделирования прохождения сигналов через них отличает данное учебное издание от имеющейся учебной литературы в данной области.

Для студентов, обучающихся по направлениям 11.03.01 – «Радиотехника» и 11.03.02 – «Инфокоммуникационные технологии и системы связи» и специальности 12.05.01 – «Электронные и оптико-электронные приборы и системы специального назначения», и магистрантов, обучающихся по направлению 09.04.04 – «Программная инженерия», может представлять интерес для специалистов.

УДК 004.932

МЕТОД ВЫСОКОТОЧНОЙ ОБРАБОТКИ ТОПОГРАФИЧЕСКИХ ИНТЕРФЕРОМЕТРИЧЕСКИХ ИЗОБРАЖЕНИЙ С ПРИВЛЕЧЕНИЕМ НИЗКОДЕТАЛЬНОЙ СТЕРЕОФОТОГРАММЕТРИЧЕСКОЙ ИНФОРМАЦИИ ОТ ОПТИКО-ЭЛЕКТРОННЫХ СИСТЕМ ДЗЗ

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

METHOD FOR HIGH-PRECISION TOPOGRAPHIC INTERFEROGRAM PROCESSING WITH USE OF LOW-DETAIL STEREOPHOTOGRAMMETRIC INFORMATION OBTAINED FROM OPTICAL REMOTE SENSING SYSTEMS

Ushenkin V.A.

The paper proposes a new method for high-precision processing of radar topographic interferograms with use of low-detail stereophotogrammetric information obtained from optical remote sensing systems. In comparison with traditional interferometry technologies, additional processing stages are introduced, which are fusion of reference low-detail stereophotogrammetric digital elevation models in order to reduce the number of artifacts and calculation of the phase corresponding to the fusion result, additional median filtering of the interferogram, rejection of areas with radar signal overlapping, identified using the reference stereophotogrammetric data with refinement based on interferogram phase analysis, rejection of shadow areas, identified using the reference stereophotogrammetric data with refinement based on amplitude analysis of the interferometric radar image pair, refinement of the interferometric base and correction of atmospheric phase distortions using the reference stereophotogrammetric data. The results of the proposed method obtained during processing of data from foreign radar satellites are presented.

Key words: synthetic aperture radar, radar image, interferogram, interferometric processing, digital terrain model, digital surface model, digital elevation model.

Ключевые слова: радиолокатор с синтезированной апертурой антенны, радиолокационное изображение, интерферограмма, интерферометрическая обработка, цифровая модель рельефа, цифровая модель местности, цифровая модель высот.

Предлагается новый метод высокоточной обработки радиолокационных топографических интерферограмм с привлечением низкодетальной стереофотограмметрической информации от оптико-электронных систем наблюдения Земли. По сравнению с традиционными технологиями интерферометрии вводятся дополнительные стадии обработки: комплексирование опорных низкодетальных стереофотограмметрических цифровых моделей высот с целью уменьшения количества артефактов и построение фазовой картины результата комплексирования; дополнительная медианная фильтрация интерферограммы; отбраковка областей переналожения радиолокационных сигналов, выявленных по опорным стереофотограмметрическим данным с уточнением на основе фазового анализа интерферограммы; отбраковка областей радиотени, выявленных по опорным стереофотограмметрическим данным с уточнением на основе амплитудного анализа интерферометрической пары изображений; уточнение интерферометрической базы и коррекция атмосферных искажений фазы по опорным стереофотограмметрическим данным. Приводятся результаты предложенного метода, полученные при обработке данных от зарубежных радиолокационных космических аппаратов.

Введение

Обработка топографических интерферограмм, полученных из пары изображений от космических систем радиолокационного наблюдения Земли, позволяет сформировать цифровую модель высот наблюдаемого участка земной поверхности: цифровую модель рельефа при отсутствии искусственных высотных объектов и лесного покрова или цифровую модель местности (поверхности) в противном случае [1].

Традиционно интерферометрическая обработка включает следующие этапы [1]:

1) определение параметров геометрического соответствия двух исходных комплексных радиолокационных изображений;

2) трансформацию дополнительного радиолокационного изображения в координаты основного;

3) формирование фазоразностного изображения – интерферограммы;

4) расчет интерферометрической базы;

5) вычитание из интерферограммы фазы плоского рельефа с учетом интерферометрической базы;

6) оценку когерентности – степени зашумленности фазы на интерферограмме;

7) фильтрацию шумов на интерферограмме;

8) развертывание фазы – восстановление истинных ее значений из свернутых по модулю 2π;

9) пересчет развернутой фазы в высоту с учетом интерферометрической базы;

10) трансформацию полученной матрицы высот в заданную картографическую проекцию.

Погрешности промежуточных результатов на выходе каждого этапа приводят к снижению точности конечной цифровой модели высот или даже к невозможности выполнения дальнейших этапов обработки. Поэтому акту-

альны исследования по модификации традиционной технологии интерферометрической обработки в части введения дополнительных этапов по устранению тех или иных погрешностей. Значительная часть искажающих факторов сильно коррелирует с высотой наблюдаемых объектов, поэтому для устранения этих факторов целесообразно привлечение опорных цифровых моделей высот, более грубых, чем формируемая в результате интерферометрической обработки. В настоящее время имеются два основных метода формирования цифровых моделей высот по спутниковым снимкам: радиолокационная интерферометрия и оптико-электронная стереофотограмметрия [2]. Для использования в качестве опорной информации при интерферометрической обработке предпочтительны стереофотограмметрические цифровые модели высот, поскольку они предоставляют дополнительные сведения об участках земной поверхности, сложных для радиолокационного наблюдения: высотной городской застройке и горной местности.

Цель работы – модификация традиционной технологии интерферометрической обработки с целью повышения точности формируемых цифровых моделей высот.

Комплексирование опорных низкодетальных стереофотограмметрических цифровых моделей высот с целью уменьшения количества артефактов

В настоящее время доступен ряд глобальных цифровых моделей высот низкого разрешения (порядка 30 м), сформированных в том числе путем стереофотограмметрической обработки, например, ASTER GDEM [3] и ALOS World 3D 30 [4]. Данные модели содержат меньше артефактов в горных районах с крутым наклоном рельефа и в участках высотной городской застройки, чем интерферометрические модели. Однако главной проблемой стереофотограмметрических моделей являются искажения высот в участках, закрытых при оптической съемке облачностью или располагающихся в области тени от облака. Такие участки, как правило, маскируются, однако алгоритмы идентификации облачности отрабатывают неидеально, поэтому на распространяемых стереофотограмметрических моделях высот содержится небольшое количество неотбракованных искажений.

Другой особенностью стереофотографических моделей являются высоты малоконтрастных объектов. Например, при формировании ALOS World 3D 30 такие объекты признавались водной поверхностью, и им назначалась некоторая фиксированная высота, даже если в реальности объект содержит перепады высот.

Благодаря тому, что различные модели высот получены из разных источников (съемки разными космическими аппаратами в разное время), расположение артефактов на них в целом не совпадает. Следовательно, появляется возможность повышения точности опорных данных о высоте объектов участка земной поверхности путем комплексирования нескольких цифровых моделей высот.

Алгоритм комплексирования цифровых моделей высот предложен в [5]. В нем выполняется выявление ошибочного выделения водной поверхности по критерию, что на большей части береговой линии высота воды не должна превышать высоту берега. Идентификация неотбракованной облачности осуществляется на основе сопоставления комплексируемых моделей высот и выявления областей их наиболее сильных различий. При этом ошибочными признаются данные той модели высот, для которой градиент высоты на границе выявленной области оказывается больше. После выявления артефактов выполняется объединение неотбракованных участков комплексируемых моделей высот с обеспечением бесшовной стыковки на границах участков. Бесшовная стыковка основана на вычислении разности высот на комплексируемых моделях и ее интерполяции по методу обратных взвешенных расстояний. В результате формируется опорная стереофотограмметрическая модель с существенно уменьшенным числом артефактов. Остаются лишь артефакты, одинаково расположенные на всех комплексируемых моделях.

Результат комплексирования используется в качестве опорной модели высот для совмещения интерферометрической пары радиолокационных изображений [6], а также пересчитывается в фазовую картину опорного рельефа для последующего использования при обработке интерферограммы.

Уточнение перпендикулярной интерферометрической базы

Для вычитания фазы плоского рельефа и пересчета фазы в высоту требуется с высокой точностью знать величину перпендикулярной интерферометрической базы – проекции расстояния между двумя точками радиолокационного наблюдения на ось, перпендикулярную к вектору скорости орбитального движения и оси визирования. В противном случае на интерферограмме остается квазилинейный набег фазы вдоль строки, а на результирующей модели высот, помимо квазилинейного набега, появляется мультипликативная ошибка. Современные отечественные навигационные приборы не позволяют измерить величину перпендикулярной интерферометрической базы с сантиметровой точностью при двухпроходной интерферометрической съемке [7], поэтому требуется уточнение базы в ходе наземной обработки.

В [8] предложен алгоритм уточнения пространственной интерферометрической базы на основе сопоставления сигнала топографической интерферограммы и фазовой картины опорной низкодетальной цифровой модели высот. Уточнение происходит в три этапа: сначала приближенно устраняется основная составляющая квазилинейного набега остаточной фазы плоского рельефа, затем устраняется мультипликативная ошибка пересчета фазы в высоту и, наконец, окончательно устраняется квазилинейный набег фазы. Поскольку алгоритм работает не с окончательной моделью высот, а с интерферограммой, его можно интегрировать в технологическую схему интерферометрической обработки перед этапом развертывания фазы, что позволяет улучшить результаты этого этапа за счет устранения искажений обрабатываемой информации.

Отбраковка областей переналожения радиолокационных сигналов, выявленных по опорным стереофотограмметрическим данным

Развертывание фазы является наиболее сложным этапом интерферометрической обработки и сводится к некорректно поставленной задаче восстановления истинных значений фазы по относительным значениям, свернутым по модулю 2π и отличающимся от истинных на некоторое целое количество периодов волны излучения. Данная задача имеет бесконечное множество решений, только одно из которых соответствует истинным высотам наблюдаемых объектов. Для выбора этого решения применяются различные эвристические предположения, которые редко нарушаются реальным рельефом, но могут часто нарушаться из-за остаточных искажений сигнала интерферограммы, поданного на вход процедуры развертывания фазы. Поэтому необходима наиболее тщательная подготовка интерферограммы перед развертыванием фазы на ней. В работе предлагаются дополнительные этапы отбраковки искажений сигнала интерферограммы, направленные на улучшение точности развертывания фазы и результирующей цифровой модели высот.

Одним из источников искажений является эффект переналожения сигналов из-за перепадов высот наблюдаемых объектов и бокового радиолокационного обзора. Для развертки изображения вдоль строки используется увеличение наклонной дальности, однако для высотных объектов наклонная дальность уменьшается, из-за чего сигнал от более высоких объектов накладывается на сигнал от более низких. Амплитуда сигнала в результате переналожения усиливается, а фаза претерпевает некоторые искажения, характер которых зависит от соотношения радиояркостей высотных и низких объектов.

Для качественного выявления областей переналожения необходимо знание высот наблюдаемых объектов, которые и являются искомыми величинами при интерферометрической обработке. В то же время можно приближенно найти расположение областей переналожения за счет привлечения низкодетальных опорных стереофотограмметрических данных.

Приближенная карта переналожений может быть составлена при трансформации опорной модели высот в систему координат основного радиолокационного изображения интерферометрической пары.

В ходе трансформации для каждого пикселя основного радиолокационного изображения находится высота точки пересечения окружности с радиусом, равным наклонной дальности, и центром в позиции космического аппарата, лежащей в плоскости, перпендикулярной к скорости движения, и поверхности, описываемой опорной моделью высот. Для этого:

– для некоторой высоты h выполняется геодезическая привязка пикселя с получением широты и долготы $(B(h), L(h))$;

– из опорной цифровой модели высот извлекается значение высоты $h_{\text{опор}}(B, L)$;

– выполняется численное решение уравнения:

$$h - h_{\text{опор}}(B(h), L(h)) = 0. \quad (1)$$

В области переналожения уравнение (1) имеет несколько корней. Это можно выявить путем его численного решения с разными начальными условиями.

Таким образом, предлагается следующий алгоритм составления карты переналожений:

1) для каждой строки основного радиолокационного изображения в направлении увеличения наклонной дальности последовательно находятся трансформированные высоты; для пикселя строки с минимальной дальностью в качестве начального значения для решения уравнения (1) выбирается минимально возможное значение высоты, а для последующих пикселей – значение высоты, найденное для предыдущего пикселя;

2) шаг 1 повторяется, но уже в направлении уменьшения наклонной дальности, при этом для пикселя с максимальной дальностью в качестве начального значения для решения уравнения (1) выбирается максимальное возможное значение высоты, а для последующих пикселей – значение высоты, найденное для предыдущего пикселя;

3) пиксели, для которых высоты, найденные на шагах 1 и 2, отличаются сильнее, чем на величину погрешности, признаются принадлежащими области переналожения.

Сигнал в пикселях интерферограммы, принадлежащих выявленным областям переналожения, отбраковывается путем назначения ему нулевых весовых коэффициентов на этапе развертывания фазы. Результирующая цифровая модель высот в данных областях будет содержать пропуски данных. Для получения полных данных о высотах наблюдаемого участка земной поверхности требуется организация радиолокационных интерферометрических съемок с разных ракурсов, например, с восходящего и нисходящего витков орбиты или с разными углами визирования. Результаты интерферометрической обработки данных таких съемок могут быть объединены в единую мозаичную модель высот с минимумом пропусков данных.

Отбраковка областей радиотени

Участкам, попавшим в радиотень, обычно соответствуют слабый уровень сигнала на радиолокационном изображении и низкая когерентность на интерферограмме. Такие участки отбраковываются в традиционной технологии интерферометрической обработки по критерию низкой когерентности.

Однако возможны ситуации, когда сигнал в области радиотени будет иметь достаточно высокую когерентность, при этом не описывая реальные высоты наблюдаемых объектов. Первая ситуация – это коррелированный характер шума. Вторая ситуация – в области радиотени из-за слабого уровня полезного сигнала может быть различим ослабленный сигнал от боковых лепестков антенны радиолокатора, из-за азимутальной неоднозначности соответствующий не текущим наблюдаемым объектам, а объектам, смещенным на несколько километров вдоль трассы космического аппарата.

В рассмотренных ситуациях в части областей радиотени остается неотбракованным сигнал, из-за искажений

плохо согласующийся с окружающим его полезным сигналом, что негативно отражается на точности развертывания фазы, причем из-за сложности этой процедуры ошибки, до развертывания локализованные в пределах области радиотени, могут распространиться и за ее пределы.

Поэтому целесообразно в дополнение к отбраковке по критерию низкой когерентности ввести в технологическую схему интерферометрической обработки этап выявления и отбраковки областей радиотени.

Так же, как и в случае с переналожениями, для выявления затенений можно привлечь низкодетальную опорную стереофотограмметрическую модель высот. Выявление выполняется следующим образом.

После трансформации опорной модели высот в систему координат основного радиолокационного изображения для каждого i -го пикселя становится известна тройка геодезических координат $(B_i, L_i, h_{\text{опор}}(B_i, L_i))$, из которой, зная орбиту космического аппарата, можно рассчитать угол α_i между направлением от объекта на радиолокатор и локальной вертикалью.

Поскольку космические радиолокаторы ведут обзор в боковом направлении с минимальным отклонением от перпендикуляра к вектору скорости, расчет затенений можно выполнять с приемлемой точностью вдоль строки радиолокационного изображения без учета значения доплеровского центроида и положения максимума диаграммы направленности антенны радиолокатора по азимуту.

Поэтому для пикселей каждой строки в направлении увеличения наклонной дальности рассчитывается высота затенения более близкими объектами. Если для i -го пикселя окажется, что его высота $h_{\text{опор}}(B_i, L_i)$ меньше высоты затенения $h_{\text{т}i}$, пиксель признается принадлежащим области радиотени. Высота затенения следующего пикселя вычисляется как

$$h_{\text{т}i+1} = \max(h_{\text{опор}}(B_i, L_i), h_{\text{т}i}) - \Delta R_{\text{н}} \cos\alpha_i, \qquad (2)$$

где $\Delta R_{\text{н}}$ – шаг пикселя по наклонной дальности.

Найденная таким образом маска затенений является приближенной из-за низкой детальности опорной стереофотограмметрической модели высот. Для ее уточнения предлагается амплитудный анализ результата усреднения двух изображений интерферометрической пары.

В пределах пикселей, первоначально отмеченных как теневые, оценивается средняя амплитуда сигнала. Затем выполняется размытие бинарной маски затенений скользящим средним, апертура которого выбирается в соответствии с отношением пространственных разрешений низкодетальной стереофотограмметрической модели высот и радиолокационных изображений.

Для пикселей, в которых размытая маска затенений принимает значение, большее нуля, выполняется фильтрация спекл-шума в малом окне. Если после фильтрации спекл-шума амплитуда сигнала в пикселе не превышает ранее рассчитанную среднюю амплитуду теневого сигнала, умноженную на коэффициент запаса, немного больший единицы, пиксель окончательно признается теневым. Сигнал в таких пикселях интерферограммы отбраковывается путем назначения нулевых весовых коэффициентов на этапе развертывания фазы.

Отбраковка областей переналожения радиолокационных сигналов, выявленных путем фазового анализа сигнала интерферограммы

Маска переналожений, построенная по опорным стереофотограмметрическим данным, является недостаточно точной из-за их низкой детальности, поэтому она может не охватывать некоторую часть реальных переналожений, искажающих сигнал интерферограммы. Часть этих остаточных переналожений, характеризующуюся наиболее сильными искажениями, можно дополнительно выявить путем анализа фазы на интерферограмме.

Не все изменения фазы вдоль строки интерферограммы допустимы с точки зрения геометрической модели интерферометрической съемки. Изменение фазы, соответствующее резкому уменьшению высоты в направлении увеличения наклонной дальности, является недопустимым, поскольку в этом случае более дальний пиксель должен был бы попасть в область радиотени и не содержать полезный сигнал. Если же оба пикселя не отнесены к теневым, то это означает, что фаза на интерферограмме искажена.

Алгоритм дополнительной отбраковки областей переналожения радиолокационных сигналов следующий.

Из (2) следует, что максимально допустимое уменьшение высоты за один пиксель интерферограммы равняется $\Delta R_{\text{н}} \cos\alpha_i$. С помощью геометрической модели интерферометрической съемки, описывающей зависимость фазы от высоты $\varphi(h)$ после уточнения перпендикулярной пространственной базы, находится предельно допустимое приращение фазы за один пиксель:

$$\Delta\varphi_{\text{мин}i} = \varphi(h_{\text{опор}}(B_i, L_i) - \Delta R_{\text{н}} \cos\alpha_i) - \varphi(h_{\text{опор}}(B_i, L_i)).$$

Одновременно по сигналу интерферограммы вычисляется свернутая по модулю 2π конечная разность свернутой фазы $\varphi_{\text{св}}$ в соседних пикселях:

$$G_{xi} = \arg\exp(j \cdot (\varphi_{\text{св}i+1} - \varphi_{\text{св}i})),$$

где j – мнимая единица.

Если $G_{xi} < \Delta\varphi_{\text{мин}i}$ и $\varphi'(h) > 0$ или $G_{xi} > \Delta\varphi_{\text{мин}i}$ и $\varphi'(h) < 0$, пиксель отбраковывается.

Дополнительная медианная фильтрация интерферограммы

Перед выполнением развертывания фазы должен быть максимально отфильтрован фазовый шум на интерферограмме. Наиболее качественное удаление шума среди известных алгоритмов обеспечивает алгоритм Голдштейна [9, 10], основанный на адаптивной полосовой фильтрации. Однако после применения фильтра Голдштейна остается небольшое количество импульсных шумов. Они приводят к увеличению количества так называемых сингулярных точек на интерферограмме, в которых нарушается свойство потенциальности поля фазового градиента, свернутого по модулю 2π. Для восстановления потенциального поля фазового гради-

ента на этапе развертывания фазы выполняется попарное соединение разноименно заряженных сингулярных точек линиями разрыва фазы, вдоль которых истинный фазовый градиент превышает половину периода волны излучения. Чем больше будет сингулярных точек, тем дольше будет выполняться развертывание фазы и тем выше вероятность возникновения ошибки, когда для проведения линии разрыва будет выбрана не та пара точек.

В то же время в [11] предложена адаптация медианного фильтра, позволяющая его применять к периодическому фазовому сигналу интерферограммы. Поэтому в технологическую схему интерферометрической обработки целесообразно добавить этап дополнительной медианной фильтрации интерферограммы, позволяющей устранить остаточный импульсный шум, снизить количество сингулярных точек и тем самым упростить последующее развертывание фазы.

Адаптация медианного фильтра основана на преобразовании периодических значений фазы на интерферограмме к упорядочиваемым значениям. В основу преобразования положено предположение, что в пределах малой окрестности полезный сигнал интерферограммы изменяется практически линейно, т.е. его отклонения от аппроксимирующей линейной зависимости не превышают половину фазового периода. Поэтому медианная фильтрация интерферограммы для каждого малого фрагмента заключается в построении линейной аппроксимации сигнала, вычитании из сигнала полученной аппроксимации, применении медианного фильтра к результату вычитания и обратного прибавления линейной аппроксимации.

Устранение атмосферных искажений и развертывание фазы

В [11, 12] предложен подход к развертыванию фазы на основе анализа сюжета интерферограммы и привлечения опорной низкодетальной модели высот, гарантирующий, что ошибки развертывания фазы будут распространяться не более, чем на несколько пикселей низкодетальных опорных данных. В [13] точность подхода была дополнительно увеличена за счет добавления процедуры уточнения положения фазовых разрывов с целью уменьшения локальных ошибок. Для корректной работы развертывания фазы в соответствии с таким подходом на интерферограмме должны отсутствовать атмосферные искажения фазы.

Если интерферограмма получена по данным синхронной интерферометрической съемки (интерферометрия с жесткой базой или в режиме «тандем»), отсутствие атмосферных искажений фазы гарантируется условиями съемки. Если же интерферограмма получена по данным двухпроходной интерферометрической съемки, требуется дополнительная фильтрация атмосферных искажений [13, 14].

За счет того, что состояние атмосферы изменяется в пространстве плавно, основные атмосферные искажения сосредоточены в низкочастотной составляющей сигнала интерферограммы, которая априорно известна по низкодетальным стереофотограмметрическим моделям высот. Поэтому устранение атмосферных искажений заключается в подавлении низких частот на результате вычитания из интерферограммы фазовой картины опорной стереофотограмметрической модели высот.

Предлагаемый метод интерферометрической обработки

Предлагаемый метод интерферометрической обработки базируется на введении дополнительных этапов, рассмотренных выше, в технологическую схему, представленную на рисунке 1.

Используются следующие сокращения:

ИП – интерферометрическая пара комплексных радиолокационных изображений;

ПГС – параметры геометрического соответствия изображений;

УПГС – уточненные ПГС;

СИП – совмещенная интерферометрическая пара изображений;

И – интерферограмма;

ПВИ – предварительно выравненная интерферограмма;

ПОТФ – предварительно рассчитанная фазовая картина опорного рельефа (опорная топографическая фаза);

В – матрица весовых коэффициентов;

ВИ – окончательно выравненная интерферограмма;

ОТФ – окончательно рассчитанная опорная топографическая фаза;

ФВИ – фильтрованная выравненная интерферограмма;

ДФВИ – дополнительно фильтрованная выравненная интерферограмма;

ОВ – матрица окончательных весовых коэффициентов;

РИ – разностная интерферограмма;

КРИ – скорректированная разностная интерферграмма (для синхронной съемки совпадает с РИ);

РРФ – разностная развернутая фаза;

РФ – развернутая фаза;

МВ – матрица высот в системе координат основного радиолокационного изображения;

ЦМВ – результирующая цифровая модель высот;

ООВ – ортотрансформированная матрица окончательных весовых коэффициентов, описывающая полноту и качество сформированной ЦМВ.

Классическая схема интерферометрической обработки не включает этапы 1, 12, 13, 15 и отбраковку переналожений и затенений на этапе 8. Этап 10 в классической схеме или не выполняется, или выполняется менее точно без использования опорной информации, или выполняется после интерферометрической обработки по уже сформированной модели высот. В работе [15] предложен метод интерферометрической обработки на основе развертывания фазы в соответствии с [11, 12], добавления в классическую схему этапа 1 и привлечения опорной информации на этапе 10. Предлагаемый в настоящей работе метод отличается этапами 12, 13, 15, отбраковкой переналожений и затенений на этапе 8, и уточнением положения фазовых разрывов с целью уменьшения локальных ошибок на этапе 16.

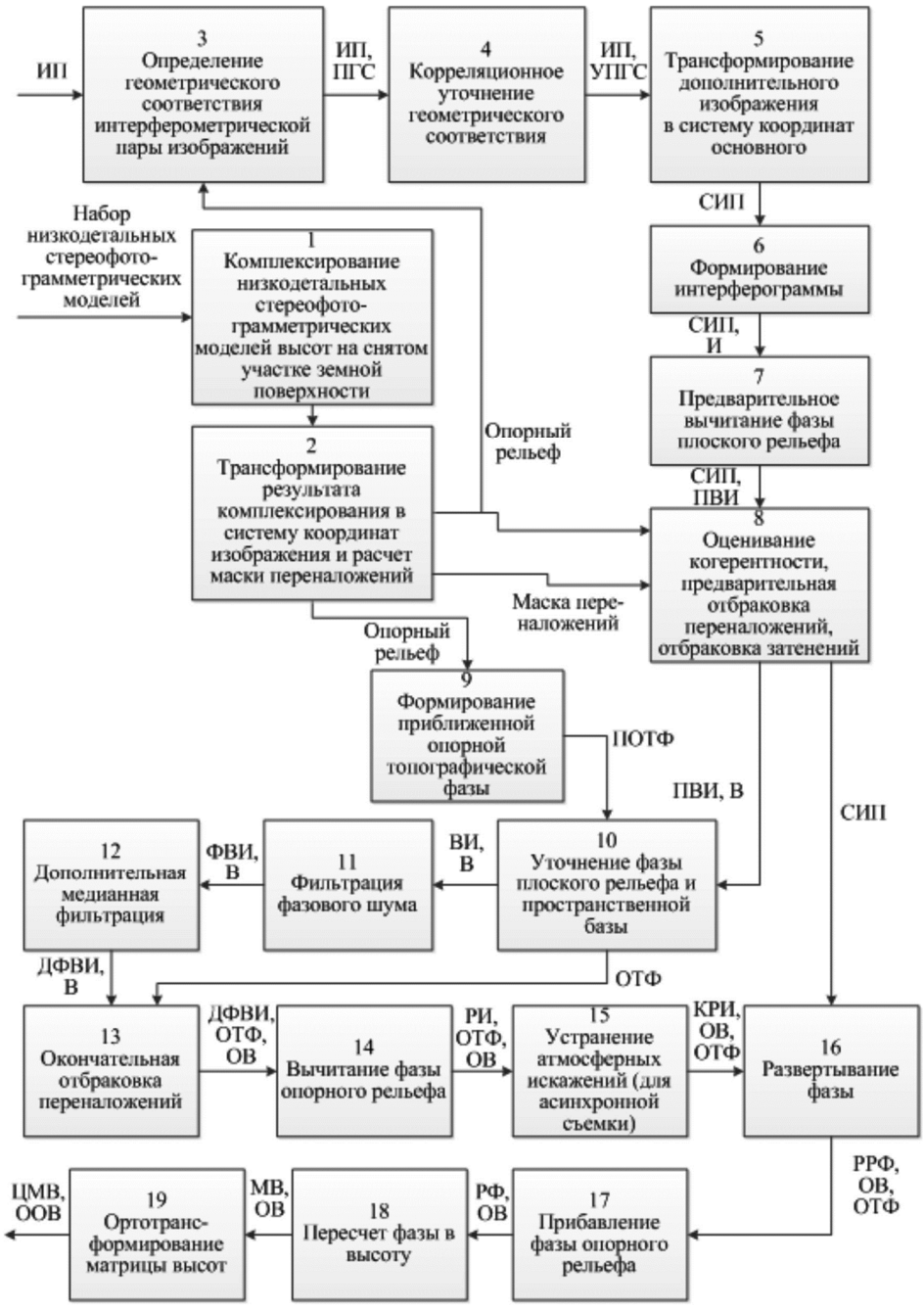


Рисунок 1 – Модифицированная технологическая схема интерферометрической обработки

Экспериментальные исследования

Проведены экспериментальные исследования предложенного метода в сравнении с методом [15], который в среднем в 5 раз превосходит по точности традиционные технологии, с использованием имеющихся в открытом доступе интерферометрических пар от зарубежных космических аппаратов:

1) окрестности скалы Улуру, Австралия, космический аппарат «TerraSAR-X»;

2) г. Финикс, Аризона, США, космический аппарат «Radarsat-2»;

3) Большой Каньон, США, космический аппарат «TerraSAR-X».

В исследовании участвовали 5 вариантов технологической схемы интерферометрической обработки:

1) технологическая схема в соответствии с методом [15] (далее – технология [15]);

2) технология [15] с добавлением этапа 15;

3) технология [15] с добавлением этапа 15 и дополнительной фильтрации (этап 12);

4) технология [15] с добавлением этапа 15 и дополнительных операций по отбраковке переналожений и затенений (этапы 8 и 13);

5) технологическая схема согласно предлагаемому методу, представленная на рисунке 1.

В ходе исследований оценивались:

1) количество сингулярных точек $N_{\text{ст}}$ в сигнале интерферограммы, поданной на вход процедуры развертывания фазы;

2) количество итераций $N_{\text{ит}}$ на самом трудоемком этапе развертывания фазы в рамках подхода [11–13] – этапе окончательной идентификации областей с длин-

Таблица 1 – Результаты экспериментальных исследований

Технологическая схема обработки	Показатель	Номер интерферометрической пары		
		1	2	3
Технология [15]	$N_{ст}$	14516	147675	300769
	σ_h, м	3,342	4,802	18,69
	$N_{ит}$	77	66	417
Технология [15] с коррекцией атмосферных искажений	$N_{ст}$	14479	147526	303416
	σ_h, м	2,369	2,092	6,400
	$N_{ит}$	53	39	82
Технология [15] с коррекцией атмосферных искажений и дополнительной фильтрацией	$N_{ст}$	3024	30575	139371
	σ_h, м	2,396	1,909	6,069
	$N_{ит}$	53	38	87
Технология [15] с коррекцией атмосферных искажений и отбраковкой переналожений и радиотеней	$N_{ст}$	3732	45716	21123
	σ_h, м	1,855	1,880	5,223
	$N_{ит}$	28	23	28
Предлагаемая технологическая схема	$N_{ст}$	83	1941	575
	σ_h, м	1,818	1,724	4,956
	$N_{ит}$	34	27	32

ными линиями разрыва фазы;

3) точность сформированной ЦМВ в виде среднеквадратичного отклонения σ_h высот на ней от опорных высот.

Для обеспечения сопоставимости оценок во всех вариантах технологической схемы для развертывания фазы использовался подход [11, 12], модифицированный в [13].

Величина σ_h обусловлена как ошибками интерферометрической обработки, так и большей детальностью сформированной ЦМВ по сравнению с опорной стереофотограмметрической моделью высот и может рассматриваться как верхняя граница ошибки интерферометрической обработки. Для обеспечения сопоставимости оценок σ_h для разных технологий их расчет выполнялся только по пикселям, не отбракованным в рамках всех сопоставляемых технологий.

В таблице 1 приведены полученные оценки. Из таблицы видно, что предлагаемый метод приводит к повышению точности интерферометрической обработки в 1,8 – 3,8 раза, а также к сокращению вычислений на этапе развертывания фазы в 2,3 – 13 раз по сравнению с методом [15]. Поскольку метод [15] точнее традиционных технологий интерферометрии в среднем в 5 раз, можно считать, что метод, предлагаемый в настоящей работе, превосходит их по точности примерно в 9 – 19 раз. При этом добавление в технологическую схему и коррекции атмосферных искажений, и дополнительной фильтрации, и отбраковки переналожений с радиотенями приводит к большему положительному эффекту, чем добавление только части из этих операций.

Заключение

Предложенный метод интерферометрической обработки за счет добавления дополнительных операций медианной фильтрации шумов, коррекции атмосферных искажений и отбраковки переналожений и затенений с учетом низкодетальной стереофотограмметрической информации позволяет в 1,8 – 3,8 раза повысить точность формирования цифровой матрицы высот и в 2,3 – 13 раз ускорить наиболее трудоемкий этап интерферометрии – развертывание фазы – по сравнению с лучшим из известных методов, а следовательно, и примерно в 9 – 19 раз превзойти по точности традиционные технологии интерферометрии.

Литература

1. Гусев С.И., Егошкин Н.А., Еремеев В.В., Кузнецов А.Е., Москвитин А.Э., Ушенкин В.А. Цифровая обработка данных радиолокационного зондирования Земли из космоса: учебное пособие / под ред. В.В. Еремеева, В.А. Ушенкина. М.: Курс, 2021. 248 с.

2. Кузнецов А.Е., Пошехонов В.И. Информационная технология стереофотограмметрической обработки видеоданных от многоматричных сканирующих устройств. Цифровая обработка сигналов. 2010. № 3. С. 44-49.

3. Cuartero A., Felicisimo A.M., Ariza F.J. Accuracy of DEM generation from Terra-ASTER stereo data. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2004, vol. 35. Part B2, pp. 559-563.

4. Tadono T., Nagai H., Ishida H., Oda F., Naito S., Minakawa K., Iwamoto H. Generation of the 30 m-mesh global digital surface model by Alos PRISM. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2016, vol. 41. Part B4. pp. 157-162.

5. Егошкин Н.А., Ушенкин В.А. Комплексирование цифровых моделей рельефа с целью повышения точности опорной информации о высоте объектов земной поверхности. Цифровая обработка сигналов. 2017. № 1. С. 13-17.

6. Егошкин Н.А., Ушенкин В.А. Совмещение высокодетальных изображений с использованием опорной цифровой модели рельефа при интерферометрической обработке радиолокационной информации. Вестник РГРТУ. 2015. № 51. С. 72–79.

7. Егошкин Н.А. Обработка бортовой навигационной информации для высокоточной оперативной коррекции космических изображений. Цифровая обработка сигналов. 2017. № 4. С. 23-29.

8. Егошкин Н.А., Ушенкин В.А. Уточнение пространственной базы при космической бистатической радиолокационной съемке Земли по сигналу интерферограммы. Цифровая обработка сигналов. 2016. № 3. С. 42-48.

9. Goldstein R.M., Werner C.L. Radar interferogram filtering for geophysical applications. Geophysical Research Letters. 1998, vol. 25 (21). pp. 4035-4038.

10. Baran I., Stewart M.P., Kampes B.M., Perski Z., Lilly P. A modification to the Goldstein radar interferogram filter // IEEE Trans. Geoscience and Remote Sensing. 2003, vol. 41(9), pp. 2114-2118.

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

12. Егошкин Н.А., Еремеев В.В., Москвитин А.Э., Ушенкин В.А. Формирование цифровых моделей рельефа на основе высокоточного развертывания фазы интерферограмм от систем радиолокационной съемки Земли. Радиотехника. 2016. № 11. С. 120-127.

13. Ушенкин В.А. Алгоритмы высокоточной обработки интерферометрической информации от систем дистанционного зондирования Земли на основе 3D-анализа наблюдаемой сцены: дис. … канд. техн. наук: 05.13.01. Рязань, 2017. 165 с.

14. Егошкин Н.А., Еремеев В.В., Москвитин А.Э., Ушенкин В.А. Обработка информации от современных космических систем радиолокационного наблюдения Земли. М.: ФИЗМАТЛИТ, 2019. 320 с.

15. Егошкин Н.А. Методы высокоточной геометрической обработки информации от современных систем космического зондирования Земли: дис. … докт. техн. наук: 05.13.01. Рязань, 2019. 323 с.

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».

У в а ж а е м ы е а в т о р ы !

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) – желательно использование графических редакторов Adobe Illustrator или Corel DRAW.*
- *Растровые (фотографии, рисунки) – М 1:1, разрешение не менее 300dpi, формат tiff.*

УДК 621.391:621.396.96

АНАЛИЗ РЕЖЕКТОРНЫХ ФИЛЬТРОВ РЕКУРСИВНОГО ТИПА

Попов Д.И., *д.т.н., профессор кафедры радиотехнических систем Рязанского государственного радиотехнического университета, e-mail: adop@mail.ru*

ANALYSIS OF RECURSIVE REJECTION FILTERS

Popov D.I.

An improved recursive rejection filter (RF) is proposed, which makes it possible to accelerate the transient process at the filter output by reconstructing its structure. At the same time, the established values of the decorrelated remnants of the rejection of the non-recursive part of the filter are sent to the RF output and to the feedbacks, which practically eliminates the "ringing" in the feedbacks caused by the average value and fluctuations of the clutter samples. The structural scheme of the improved RF is given, in which the restructuring of the RF structure is implemented by switching recursive connections after the establishment of processes in the non-recursive part of the RF. The introduction of clutter rejection residues into recursive connections and to the RF output, corresponding to the established mode of the non-recursive part of the RF, significantly reduces the transition time when compensating for clutter counts. A comparative analysis of the dynamic amplitude-frequency characteristics of the RF of a fixed and tunable structure and the effectiveness of the rejection of passive interference by these filters in the transient mode was carried out. It is shown that the restructuring of the RF structure by switching recursive connections significantly accelerates the processes of establishing a filter at the output, leading to significant (up to tens of decibels) gains in the efficiency of clutter rejection in the transient mode of operation compared to the RF of a fixed structure. Solving the problem of accelerating the transition process of recursive RF allows us to really use the broad possibilities of forming the required characteristics of RF data and their flexible management, which, in the conditions of parametric a priori uncertainty, implies the corresponding adaptation of the RF.

Key words: analysis, amplitude-frequency response, clutter suppression coefficient, restructuring the structure, transient mode, rejection filter, recursion.

Ключевые слова: анализ, амплитудно-частотная характеристика, коэффициент подавления помехи, перестройка структуры, переходный режим, режекторный фильтр, рекурсия.

Предложен усовершенствованный рекурсивный режекторный фильтр (РФ), позволяющий осуществить ускорение переходного процесса на выходе фильтра путем перестройки его структуры. При этом на выход РФ и в обратные связи поступают установившиеся значения декоррелированых остатков режектирования нерекурсивной части фильтра, что практически устраняет «звон» в обратных связях, вызываемый средним значением и флюктуациями отсчетов помехи. Приведена структурная схема усовершенствованного РФ, в котором перестройка структуры РФ реализуется путем коммутации рекурсивных связей после установления процессов в нерекурсивной части РФ. Поступление в рекурсивные связи и на выход РФ остатков режектирования помехи, соответствующих установившемуся режиму нерекурсивной части РФ, существенно сокращает время переходного процесса при компенсации отсчетов пассивной помехи. Проведен сравнительный анализ динамических амплитудно-частотных характеристик РФ фиксированной и перестраиваемой структуры и эффективности режектирования пассивных помех этими фильтрами в переходном режиме. Показано, что перестройка структуры РФ путем коммутации рекурсивных связей существенно ускоряет процессы установления на выходе фильтра, приводя к существенным (до десятков децибел) выигрышам в эффективности режектирования пассивной помехи в переходном режиме работы по сравнению с РФ фиксированной структуры. Решение проблемы ускорения переходного процесса рекурсивных РФ позволяет реально использовать широкие возможности формирования требуемых характеристик данных РФ и гибкого их управления, что в условиях параметрической априорной неопределенности предполагает соответствующую адаптацию РФ.

Введение

При выделении сигналов движущихся целей на фоне пассивных помех основной операцией в когерентно-импульсных радиолокационных системах высокой скважности является режектирование спектральных составляющих помехи [1-4]. Априорная неопределенность спектрально-корреляционных характеристик помехи, а также их неоднородность и нестационарность в зоне обзора дополнительно затрудняют реализацию эффективной защиты от пассивных помех. Преодоление априорной неопределенности параметров помехи основывается на методах адаптации к неизвестным корреляционным параметрам помехи, что приводит, в частности, к алгоритмам адаптивного режектирования помехи с комплексными весовыми коэффициентами и соответствующим адаптивным режекторным фильтрам (АРФ) [5]. Реализация данных АРФ в цифровом виде требует высокого быстродействия выполнения арифметических операций. Избежать указанных трудностей можно путем предварительной компенсации доплеровского сдвига фазы помехи. В работах [6, 7] синтезированы алгоритмы оценивания и предложены принципы построения и структурные схемы автокомпенсаторов доплеровской фазы пассивных помех с прямой и обратной связью. Особенности адаптации к корреляционным свойствам помехи на выходе автокомпенсатора и последующего ее режектирования рассмотрены в рабо-

те [8]. Определенное упрощение процедуры адаптации достигается в АРФ каскадного типа [9]. Другим вариантом упрощения процедуры адаптации является переход от комплексных весовых коэффициентов к действительным, что ограничивает область целесообразного применения соответствующих АРФ при ограниченной и сравнительно малой в зависимости от порядка фильтра и ожидаемых параметров помехи величине ее доплеровской скорости [10]. Компромиссное решение достигается в фильтрах с частичной адаптацией к доплеровской фазе помехи и оптимизацией характеристик режекторных фильтров в априорном диапазоне изменения спектрально-корреляционных параметров помехи [11]. Повышение эффективности нерекурсивных режекторных фильтров высоких порядков достигается при оптимизации их параметров по вероятностному критерию [12].

Известными преимуществами формирования характеристик перед нерекурсивными режекторными фильтрами (РФ) обладают рекурсивные РФ того же порядка, что может быть использовано для повышения эффективности выделения сигналов движущихся целей на фоне коррелированных (пассивных) помех. Развитие аппаратно-программных средств цифровой обработки сигналов стимулирует применение рекурсивных РФ в системах селекции движущихся целей для подавления пассивных помех. Традиционно при синтезе и анализе рекурсивных фильтров использовался частотный подход [13], соответствующий установившемуся режиму работы фильтра. Однако длительность обрабатываемых выборок радиолокационных сигналов, как правило, соизмерима с временем установления процессов в рекурсивных фильтрах [14]. При этом переходный режим является основным режимом работы рекурсивных РФ, что приводит к задачам ускорения переходного процесса и анализа характеристик рекурсивных РФ в переходном режиме.

Характеристики рекурсивных РФ

Рассмотрим рекурсивный РФ в виде представляющего наибольший интерес для практики каскадного соединения нерекурсивного звена 1-го порядка и рекурсивного звена 2-го порядка. Системная функция такого РФ в z-плоскости имеет вид

$$H(z)=\frac{(z-1)(z^2+az+1)}{z(z^2-b_1z-b_2)}=$$

$$=\frac{z^3-(1-a)z^2+(1-a)z-1}{z(z^2-b_1z-b_2)}=\sum_{k=0}^{\infty}g_k z^{-k},$$

где $z=\mathrm{e}^{\mathrm{i}\omega T}$; T – интервал временной дискретизации (для импульсных сигналов соответствует их периоду повторения); а, b_1, b_2 – весовые коэффициенты прямых и обратных связей рекурсивного звена; g_k – коэффициенты импульсной характеристики, определяемые из соотношений:

$$g_0=d_0,\quad g_1=d_1+b_1g_0,\quad g_k=d_k+\sum_{j=1}^{2}b_jg_{k-j}, \qquad (1)$$

а при $k>3$ $\quad g_k=\sum_{j=1}^{2}b_jg_{k-j}, \qquad (2)$

причем, $d_0=-d_3=1,\quad d_1=-d_2=-(1-a)$.

Переходя к циклической частоте $\theta=\omega T$, для квадрата стационарной амплитудно-частотной характеристики (АЧХ) найдем:

$$H^2(\theta)=\left|H(\mathrm{e}^{\mathrm{i}\theta})\right|^2=H(z)H(z^{-1})\Big|_{z=\mathrm{e}^{\mathrm{i}\theta}}=$$

$$=\frac{2(1-\cos\theta)(2+a^2+4a\cos\theta+2\cos2\theta)}{1+b_1^2+b_2^2-2b_1(1-b_2)\cos\theta-2b_2\cos2\theta}. \qquad (3)$$

В переходном режиме характеристики РФ определяются конечным числом n коэффициентов импульсной характеристики g_k, $k=\overline{0,\,n-1}$. Тогда квадрат динамической АЧХ (соответствующей переходному режиму РФ) определяется выражением

$$H^2(\theta,n)=\left|H(\mathrm{e}^{\mathrm{i}\theta},n)\right|^2=\left|\sum_{k=0}^{n-1}g_k\,z^{-k}\right|^2_{z=\mathrm{e}^{\mathrm{i}\theta}}=$$

$$=\sum_{j,k=0}^{n-1}g_jg_k\cos\left[(j-k)\theta\right]. \qquad (4)$$

Соотношения (3), (4) соответствуют фиксированной структуре РФ. Для ускорения переходного процесса необходимо осуществлять перестройку структуры РФ. На рисунке 1 приведена структурная схема рассматриваемого каскадного РФ, перестройка структуры которого реализуется путем коммутации рекурсивных связей $(b_1,\,b_2)$. К приходу 4-го отсчета (импульса), соответствующего установлению процессов в нерекурсивной части РФ, по команде блока управления (БУ) замыкается коммутатор (К). В рекурсивные связи и на выход РФ теперь поступают остатки режектирования помехи, соответствующие установившемуся режиму нерекурсивной части РФ, что существенно сокращает время переходного процесса.

Заметим, что наиболее просто перестройка структуры РФ осуществляется при известном начале обрабатываемой последовательности, имеющем место при дискретном сканировании антенного луча.

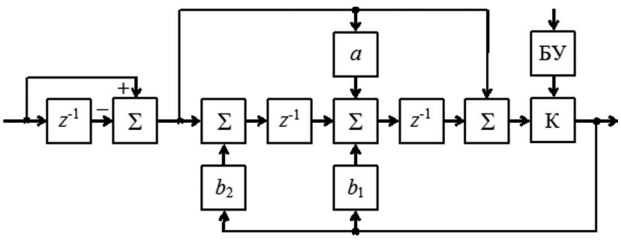


Рисунок 1 – Структурная схема каскадного рекурсивного РФ

При линейном (непрерывном) режиме обзора усовершенствование структуры РФ с целью ускорения его переходного процесса имеет свои особенности, обусловленные модуляцией импульсов помехи на ее кромках диаграммой направленности антенны. Вначале осуществляется обнаружение передней кромки помехи по всем модулированным импульсам кромки к моменту прихода импульса, соответствующего плоскому участку огибающей помехи. Далее задержанные отсчеты кромки помехи взвешиваются с целью восстановления прямоугольности их огибающей, что позволяет при аналогичной рисунку 1 обработке производить эффективное выделение доплеровских сигналов, не дожидаясь поступ-

ления импульсов плоского участка огибающей. При появлении и обнаружении задней кромки помехи также путем взвешивания отсчетов происходит восстановление прямоугольности ее огибающей.

Динамические АЧХ рекурсивных РФ

Нерекурсивной части РФ перестраиваемой структуры (рисунок 1) соответствует системная функция

$$H_{\text{н}}(z) = 1-(1-a)z^{-1}+(1-a)z^{-2}-z^{-3} = \sum_{k=0}^{3} d_k z^{-k},$$

и коэффициенты импульсной характеристики

$$d_0 = -d_3 = 1, \quad d_1 = -d_2 = -(1-a), \tag{5}$$

а рекурсивной части – системная функция

$$H_{\text{р}}(z) = (1-b_1 z^{-1}-b_2 z^{-2})^{-1} = \sum_{k=0}^{\infty} h_k z^{-k},$$

и коэффициенты импульсной характеристики

$$h_0 = 1, \quad h_1 = b_1 h_0, \text{ а при } k \geq 2 \quad h_k = b_1 h_{k-1} + b_2 h_{k-2}. \tag{6}$$

Частотная характеристика РФ в переходном режиме после коммутации рекурсивных связей имеет вид

$$H(\mathrm{e}^{\mathrm{i}\theta},\ n) = \left(\sum_{p=0}^{3} d_p z^{-p}\right)\left(\sum_{k=0}^{n-4} h_k z^{-k}\right)\Bigg|_{z=\mathrm{e}^{\mathrm{i}\theta}} =$$

$$= \sum_{k=0}^{n-4}\sum_{p=0}^{3} h_k d_p z^{-(k+p)}\Bigg|_{z=\mathrm{e}^{\mathrm{i}\theta}}.$$

По аналогии с выражением (4) для квадрата динамической АЧХ РФ перестраиваемой структуры найдем:

$$H^2(\theta, n) = \left|H(\mathrm{e}^{\mathrm{i}\theta}, n)\right|^2 =$$

$$= \sum_{j,k=0}^{n-4}\sum_{l,p=0}^{3} h_j h_k d_l d_p \cos\left[(j+l-k-p)\theta\right]. \tag{7}$$

Рассмотрим конкретные результаты для каскадного РФ эллиптического типа, синтезированного по квадрату стационарной АЧХ со следующими параметрами: циклическая частота смещенного нуля в полосе режекции $\theta_{\text{н}} = 0{,}1\pi$, частота среза полосы пропускания $\theta_{\text{с}} = 0{,}5\pi$ (ширина полосы пропускания $\Delta\theta = \pi$) и неравномерность АЧХ в полосе пропускания $\varepsilon \leq 3$ дБ. При этом весовые коэффициенты рекурсивного звена $a = -1{,}902$, $b_1 = -0{,}186$, $b_2 = -0{,}563$.

На рисунке 2 штриховой кривой изображена стационарная АЧХ РФ, соответствующая выражению (3), а сплошными кривыми – динамические АЧХ РФ фиксированной структуры, соответствующие выражению (4). Как видим, в полосе пропускания динамические АЧХ практически не зависят от количества обрабатываемых отсчетов n и отвечают требованиям к АЧХ в этой области (частоте среза $\theta_{\text{с}}$ и уровню пульсаций ε). Однако в зоне режекции АЧХ имеется существенная зависимость от числа отсчетов n, а приведенные на кривых значения n свидетельствуют о достаточно медленном процессе формирования зоны режекции, приближающейся к установившемуся состоянию только при $n > 16$, что крайне нежелательно для режекторных фильтров, так как приводит к маскировке сигналов движущихся целей нескомпенсированными остатками помехи.

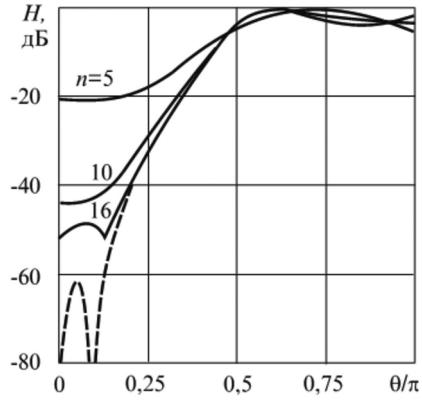


Рисунок 2 – Стационарная и динамические АЧХ рекурсивного РФ фиксированной структуры

На рисунке 3 приведены динамические АЧХ РФ перестраиваемой структуры (рисунок 1), соответствующие выражению (7). В этом случае зона режекции АЧХ остается практически стабильной при объеме обрабатываемой выборки $n \geq 4$. Действительно, к моменту замыкания коммутатора $(n = 4)$ АЧХ нерекурсивного РФ 3-го порядка близка к стационарной (рисунок 2) и имеет нули при $\theta = 0$ и $\theta = \theta_{\text{н}}$, что обеспечивает ее стабильность в окрестности этих точек при последующей обработке поступающих отсчетов. В полосе пропускания АЧХ практически приближается к стационарной при $n \geq 7$. Такая динамика установления для режекторного фильтра является вполне удовлетворительной.

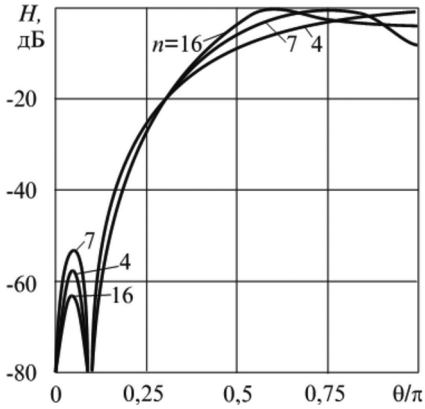


Рисунок 3 – Динамические АЧХ рекурсивного РФ перестраиваемой структуры

Эффективность режектирования рекурсивными РФ

Рассмотрим теперь эффективность режектирования пассивной помехи в переходном режиме рекурсивными РФ фиксированной и перестраиваемой структуры. Полагаем, что отсчеты комплексной огибающей гауссовской узкополосной помехи $U_j = x_j + \mathrm{i}y_j$ при скомпенсированной доплеровской скорости описываются корреляционными моментами

$$R_{jk} = \overline{U_j U_k^*}\big/2 = \sigma_{\text{п}}^2 \rho_{jk} + \sigma_{\text{ш}}^2 \delta_{jk},$$

где $\sigma_{\text{п}}^2$, $\sigma_{\text{ш}}^2$ – дисперсии помехи и собственного шума приемника соответственно, ρ_{jk} – коэффициенты межпериодной корреляции помехи, δ_{jk} – символ Кронекера.

На выходе реализуемого в двух квадратурных кана-

лах РФ фиксированной структуры в переходном режиме имеем

$$V = \sum_{j=0}^{n-1} g_j U_{n-j}.$$

Дисперсия помехи на выходе РФ определяется следующим образом:

$$\sigma^2_{V\text{п}} = \frac{1}{2}\overline{VV^*} = \sigma^2_{\text{п}} \sum_{j,k=0}^{n-1} g_j g_k \rho_{jk} + \sigma^2_{\text{ш}} \sum_{j=0}^{n-1} g_j^2. \quad (8)$$

Эффективность РФ будем оценивать по нормированному относительно прохождения некоррелированного собственного шума коэффициенту подавления помехи [15]

$$\mu = (\sigma^2_{V\text{ш}} / \sigma^2_{\text{ш}}) / (\sigma^2_{V\text{п}} / \sigma^2_{\text{п}}). \quad (9)$$

Для РФ фиксированной структуры с учетом выражений (8) и (9) получим

$$\mu(n) = \sum_{j=0}^{n-1} g_j^2 \Bigg/ \left(\sum_{j,k=0}^{n-1} g_j g_k \rho_{jk} + \lambda \sum_{j=0}^{n-1} g_j^2 \right), \quad (10)$$

где $\lambda = \sigma^2_{\text{ш}} / \sigma^2_{\text{п}}$ – отношение шум/помеха.

Представляя выходную величину РФ перестраиваемой структуры с учетом коммутации рекурсивных связей в виде

$$V = \sum_{j=0}^{n-4} h_j \sum_{l=0}^{3} d_l U_{4+j-l},$$

в результате вычислений, аналогичных предыдущим, получим

$$\mu(n) = \left(\sum_{l,p=0}^{3} \sum_{j=|l-p|}^{n-4} h_j h_{j-|l-p|} d_l d_p \right) \times$$

$$\times \left(\sum_{j,k=0}^{n-4} \sum_{l,p=0}^{3} h_j h_k d_l d_p \rho_{j-l,k-p} + \right. \quad (11)$$

$$\left. + \lambda \sum_{l,p=0}^{3} \sum_{j=|l-p|}^{n-4} h_j h_{j-|l-p|} d_l d_p \right)^{-1}.$$

Результаты расчетов эффективности режектирования в переходном режиме при гауссовской аппроксимации коэффициентов корреляции $\rho_{jk} = \exp\{-[\pi(j-k) \times \beta_{\text{п}}]^2 / 2,8\}$, нормированной ширине спектра помехи $\beta_{\text{п}} = \Delta f_{\text{п}} T = 0,05$ и $\lambda = 10^{-6}$ представлены на рисунке 4. Кривая 1 соответствует РФ фиксированной структуры, кривая 2 – РФ перестраиваемой структуры. В обоих случаях параметры исходной стационарной АЧХ соответствуют $\theta_{\text{н}} = 0,08\pi$, $\theta_{\text{с}} = 0,3\pi$. При этом кривая 1 для

РФ фиксированной структуры рассчитана в соответствии с выражением (10), в котором коэффициенты импульсной характеристики g_k определяются соотношениями (1) и (2). Кривая 2 для РФ перестраиваемой структуры рассчитана в соответствии с выражением (11), в котором коэффициенты импульсной характеристики d_k нерекурсивной части РФ определяются соотношениями (5), а коэффициенты импульсной характеристики h_k рекурсивной части РФ – соотношениями (6).

Как видим, установление процессов в РФ перестраиваемой структуры происходит практически к 5-у отсчету выборки, приводя к существенным (до десятков децибел) выигрышам в эффективности режектирования пассивной помехи в переходном режиме по сравнению с РФ фиксированной структуры.

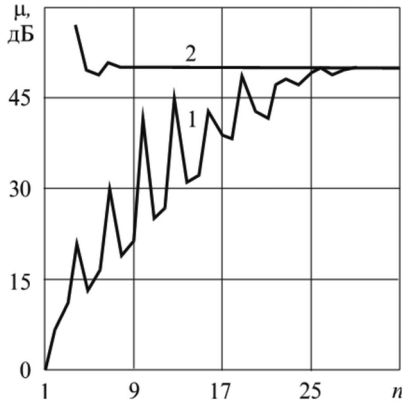


Рисунок 4 – Эффективность режектирования помехи в переходном режиме

В рассмотренном примере нерекурсивный РФ, имеющий место при $n = 4$, обладает весьма высокой эффективностью режектирования помехи и сравнительно узкой полосой пропускания, соответствующей частоте среза $\theta_{\text{с}} = 0,71\pi$. Рекурсивный РФ (при $n > 4$) позволяет существенно расширить полосу пропускания, что со ответствует частоте среза $\theta_{\text{с}} = 0,3\pi$, при сохранении достаточно высокой эффективности режектирования помехи. Наличие собственного (некоррелированного) шума ограничивает предельную эффективность РФ динамическим диапазоном помехи λ^{-1} по отношению к уровню шума, что приводит к выбору минимально необходимой величины $\theta_{\text{с}}$, при которой реализуется предельная или заданная эффективность режектирования в сочетании с максимальной для данных условий полосой пропускания РФ. В конечном счете оптимальная величина $\theta_{\text{с}}$ зависит от параметров помехи.

С учетом решения проблемы ускорения переходного процесса рекурсивные РФ позволяют реально использовать широкие возможности формирования требуемых характеристик и гибкого их управления, что в условиях априорной неопределенности спектрально-корреляционных характеристик пассивных помех создает благоприятные условия для соответствующей адаптации РФ [16].

Заключение

Усовершенствованный рекурсивный РФ позволяет осуществить ускорение переходного процесса путем перестройки его структуры. При этом на выход РФ и в обратные связи поступают установившиеся значения декоррелированых остатков режектирования нерекурсивной части фильтра, что практически устраняет «звон» в обратных связях, вызываемый средним значением и флюктуациями отсчетов помехи.

Предложена структурная схема усовершенствованного РФ, в котором перестройка структуры РФ реализуется путем коммутации рекурсивных связей после установления процессов в нерекурсивной части РФ. Поступление в рекурсивные связи и на выход РФ остатков ре-

жектирования помехи, соответствующих установившемуся режиму нерекурсивной части РФ, существенно сокращает время переходного процесса при компенсации отсчетов пассивной помехи.

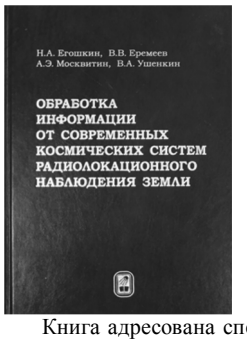
Проведенный сравнительный анализ динамических АЧХ РФ фиксированной и перестраиваемой структуры и эффективности режектирования пассивных помех этими фильтрами показывает, что перестройка структуры РФ путем коммутации рекурсивных связей существенно сокращает длительность процессов установления в фильтре, приводя к существенным (до десятков децибел) выигрышам в эффективности режектирования пассивной помехи в переходном режиме работы по сравнению с РФ фиксированной структуры.

Решение проблемы ускорения переходного процесса рекурсивных РФ позволяет реально использовать широкие возможности формирования требуемых характеристик данных РФ и гибкого их управления, что в условиях параметрической априорной неопределенности предполагает соответствующую адаптацию РФ.

Литература

1. Skolnik M.I. Introduction to Radar System, 3rd ed., New York: McGraw-Hill, 2001. 862 p.
2. Richards M.A., Scheer J.A., Holm W.A. (Eds.). Principles of Modern Radar: Basic Principles. New York: SciTech Publishing, IET, Edison. 2010. 924 p.
3. Melvin W. L., Scheer J.A. (Eds.). Principles of Modern Radar: Advanced Techniques. New York: SciTech Publishing, IET, Edison, 2013. 846 p.
4. Справочник по радиолокации: в 2 кн. Кн. 1 / под ред. М.И. Сколника; пер. с англ. под ред. В.С. Вербы. М.: Техносфера, 2014. 672 с.
5. Попов Д.И. Адаптация нерекурсивных режекторных фильтров. Известия вузов. Радиоэлектроника. 2009. Т. 52. № 4. С. 46-55.
6. Попов Д.И. Автокомпенсация доплеровской фазы пассивных помех. Цифровая обработка сигналов. 2009. № 2. С. 30-33.
7. Попов Д.И. Автокомпенсация доплеровской фазы многочастотных пассивных помех. Вестник Рязанского государственного радиотехнического университета. 2018. № 65. С. 32-37.
8. Попов Д.И. Адаптивное подавление пассивных помех. Цифровая обработка сигналов. 2014. № 4. С. 32-37.
9. Попов Д.И. Адаптивные режекторные фильтры каскадного типа. Цифровая обработка сигналов. 2016. № 2. С. 53-56.
10. Попов Д.И. Адаптивные режекторные фильтры с действительными весовыми коэффициентами. Цифровая обработка сигналов. 2017. № 1. С. 22-26.
11. Попов Д.И. Оптимизация нерекурсивных режекторных фильтров с частичной адаптацией. Цифровая обработка сигналов. 2018. № 1. С. 28-32.
12. Попов Д.И. Оптимизация режекторных фильтров по вероятностному критерию. Цифровая обработка сигналов. 2021. № 1. С. 55-58.
13. Рабинер Л., Гоулд Б. Теория и применение цифровой обработки сигналов. М.: Мир, 1978. 848 с.
14. Попов Д.И. Анализ эффективности режекторных фильтров. Вестник Рязанской государственной радиотехнической академии. 2003. Вып. 11. С. 105-107.
15. Попов Д.И. Анализ адаптивной обработки многочастотных сигналов. Вестник Рязанского государственного радиотехнического университета. 2018. № 63. С. 9-13.
16. Попов Д.И. Адаптивная межпериодная обработка многочастотных сигналов. Вестник Рязанского государственного радиотехнического университета. 2018. № 64. С. 17-22.

НОВЫЕ КНИГИ

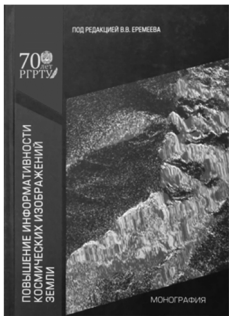


Егошкин Н.А., Еремеев В.В., Москвитин А.Э., Ушенкин В.А.

Обработка информации от современных космических систем радиолокационного наблюдения Земли. М.: ФИЗМАТЛИТ, 2019. 320 с. + 32 с. цв. вкл.

В монографии представлены результаты научных исследований, полученные авторами – сотрудниками НИИ обработки аэрокосмических изображений Рязанского государственного радиотехнического университета – в ходе выполнения работ по созданию средств наземной обработки информации от космических систем радиолокационного наблюдения Земли. В качестве этих результатов выступают новые модели, алгоритмы и технологии первичной, геометрической и интерферометрической обработок космической радиолокационной информации. Впервые в отечественной литературе приводится широкий обзор и сравнительный анализ существующих зарубежных технологий первичной, поляриметрической и интерферометрической обработок радиолокационных данных.

Книга адресована специалистам в области обработки аэрокосмических изображений поверхности Земли, аспирантам и студентам старших курсов вузов.



Егошкин Н.А., Еремеев В.В., Козлов Е.П., Кузнецов А.Е., Макаренков А.А., Москвитин А.Э., Пресняков О.А., Ушенкин В.А.

Повышение информативности космических изображений Земли: монография / под ред. В.В. Еремеева. М.: КУРС, 2022. 368 с.

В монографии представлены результаты научных исследований в области координатной обработки и комплексирования данных космического наблюдения Земли в видимом, тепловом и радиолокационном диапазонах спектра. Дан анализ проблемы повышения информативности изображений от современных систем космического наблюдения Земли, определены пути решения этой проблемы.

Основное внимание в книге уделено вопросам: улучшения геометрического качества изображений, высокоточной привязки материалов космической съемки, повышения информативности изображений путем объединения снимков, полученных в видимом, тепловом и радиолокационном диапазонах спектра, экспериментального подтверждения предложенных подходов.

УДК 621.396.67

ИСПОЛЬЗОВАНИЕ АСИММЕТРИЧНЫХ «СОЗВЕЗДИЙ» ПРИ ФАЗОВОЙ МОДУЛЯЦИИ ЦИФРОВЫХ СИГНАЛОВ

Полушин П.А., *д.т.н., профессор, профессор кафедры радиотехники и радиосистем Владимирского государственного университета, e-mail: polushin.p@mail*
Корнеева Н.Н., *к.т.н., заведующий кафедрой радиотехники и радиосистем Владимирского государственного университета, e-mail: korneevanata@mail.ru*
Архипов Н.А., *аспирант кафедры радиотехники и радиосистем Владимирского государственного университета, e-mail: polushin.p@mail*
Крамаренко Н.И., *магистрант кафедры радиотехники и радиосистем Владимирского государственного университета, e-mail: polushin.p@mail*

USING OF ASYMMETRICAL «EYES DIAGRAM» IN DIGITAL SIGNALS PHASE MODULATION

Polushin P.A., Korneeva N.N., Arhipov N.A., Kramarenko N.I.

Currently, differential types of modulation are widely used in information transmission systems to eliminate the uncertainty of phase adjustment during demodulation in the receiver. However, there is a 1-3 dB loss in noise immunity. It is proposed to use asymmetric «constellations» in phase modulation, which make it possible to abandon differential modulation.

Key words: phase modulation, types of differetial modulation, «eyes diagram» in phase modulation.

Ключевые слова: фазовая модуляция, дифференциальные виды модуляции, «созвездие» при фазовой модуляции.

В настоящее время в системах передачи информации широко используются дифференциальные виды модуляции, позволяющие устранить неопределенность подстройки фазы при демодуляции в приемнике. Однако при этом имеет место проигрыш по помехоустойчивости на 1-3 дБ. Предлагается использовать асимметричные «созвездия» при фазовой модуляции, которые дают возможность отказаться от дифференциальной модуляции.

Введение

При детектировании цифровых фазомодулированных (фазоманипулированных) сигналов обычно используется корреляционная обработка принятого сигнала с помощью опорного генератора (ОГ), расположенного в приемнике [1]. При этом принятый сигнал (при использовании модуляции BPSK – binary phase shift keying, двоичной фазовой манипуляции) перемножается с сигналом ОГ, далее результат перемножения интегрируется на интервале времени, равным длительности символа. На основе анализа знака этого напряжения принимается решение о том, какой информационный символ передавался.

При использовании других видов фазовой модуляции (QPSK, QAM и др.) используются два ОГ, вырабатывающих ортогональные сигналы для корреляционной обработки. Решение о том, какая комбинация информационных символов передавалась, принимается на основе анализа соотношения уровней двух выделенных компонент принятого сигнала.

Для синхронизации фазы сигнала ОГ с целью удаления информационной компоненты могут использоваться различные методы, такие, как умножение частоты принятого сигнала, применение схемы Костаса и др. [1-5]. Однако при этом возникает возможность того, что фаза ОГ настроится не на правильную фазу, а на другое ее значение. Оно может отличаться от правильного значения на 180º, на 90º и др. При этом правильная демодуляция становится невозможной.

Одним из основных методов борьбы с неоднозначностью настройки является переход к дифференциальным видам модуляции, когда передаваемая информация заключается не в абсолютном значении фазы принятого символа, а в ее изменении относительно фазы предыдущего принятого символа.

Однако при этом подобные виды модуляции (DBPSK, DQPSK и др.) обладают существенными недостатками. Один из них заключается в том, что в качестве опорного применяется сигнал предыдущего символа, который также поражен шумами, как и сигнал текущего символа. Это заметно ухудшает отношение «сигнал/шум» и увеличивает вероятность ошибки.

Другой недостаток состоит в возможном размножении ошибок, что также ухудшает среднюю помехоустойчивость передачи. В результате наблюдается проигрыш по помехоустойчивости до 3 дБ.

Однако возможно использовать метод, который позволяет правильно настроить сигнал ОГ без применения дифференциальной модуляции и без существенного снижения помехоустойчивости в результате этого. Метод заключается в определенном изменении вида «созвездия», применяемого в передатчике для осуществления фазовой модуляции.

Теоретическое обоснование метода

Рассмотрим методы модуляции (манипуляции, цифровой модуляции) BPSK и QPSK. «Созвездия» этих ви-

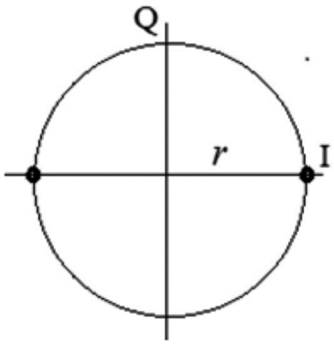


Рисунок 1

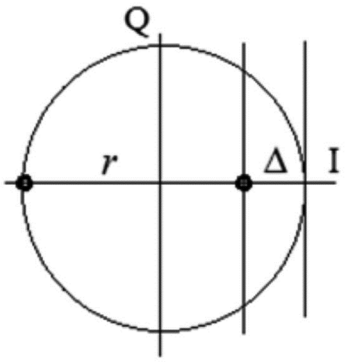


Рисунок 2

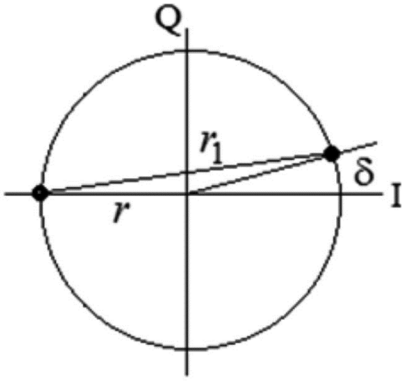


Рисунок 3

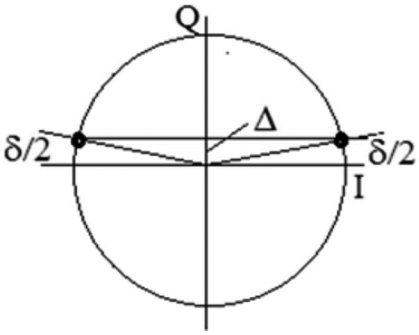


Рисунок 4

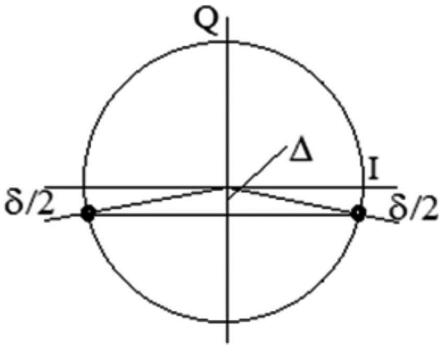


Рисунок 5

дов модуляции образованы точками, расположенными симметрично. В описываемом методе предлагается использовать асимметричное расположение точек в «созвездии».

Для метода BPSK на рисунках 1, 2 и 3 приведены исходный вид «созвездия» (рисунок 1, и два варианта асимметричного созвездия (рисунок 2 с радиальным смещением точки «созвездия» и рисунок 3 с ее угловым смещением). На приемной стороне при классическом виде «созвездия» без использования дифференциальной модуляции начальная фаза ОГ может случайно принимать и значение 0º, и значение 180º. Общий вид «созвездия» из-за его симметричности не изменится и правильная демодуляция затруднительна.

В предлагаемом методе для варианта с радиальным смещением после демодуляции одна из точек созвездия будет расположена ближе к началу координат, чем другая, что укажет на правильное значение фазы ОГ. Для варианта с угловым смещением одной из точек «созвездие» может принимать один из двух видов, приведенных на рисунке 4 и рисунке 5, где первый соответствует «правильной» фазе напряжения ОГ, а второй – его «неправильной» фазе.

Однако применение предлагаемого метода приводит к определенному снижению помехоустойчивости, т.к. расстояние между точками в асимметричном «созвездии» несколько уменьшается. Чем больше выражена асимметрия, тем заметнее ухудшение помехоустойчивости, но тем проще получать правильное значение начальной фазы ОГ.

Определим соответствующие показатели для обоих вариантов асимметричного «созвездия».

Вариант 1 (радиальное смещение точки «созвездия»). Основной причиной возникновения ошибок является воздействие теплового шума, имеющего гауссово распределение и дисперсию, равную σ_0^2. Поскольку при корреляционной демодуляции ошибки возникают из-за одной из двух ортогональных компонент шума, то при классическом виде «созвездия» вероятность возникновения битовой ошибки будет равна

$$P_0(x) = 1 - F[r/\sigma_1] = 0,5 - 0,5erf[r/\sigma_1\sqrt{2}],$$

где $F(\cdot)$ – интегральная гауссова функция; r – амплитуда полезного сигнала; $\sigma_1^2 = 0,5\sigma_0^2$.

При использовании предлагаемого метода она станет равной

$$P_1(x) = 1 - 0,5F[(r-\Delta)/\sigma_1] - 0,5F(r/\sigma_1).$$

Вариант 2 (угловое смещение точки «созвездия»). В данном случае расстояние между противоположными точками «созвездия» будет равно $r_1 = \sqrt{2}r\sqrt{1+\cos\delta}$. Соответственно, вероятность возникновения ошибки будет равна

$$P_2(x) = 1 - F(r_1/2\sigma_1) = 0,5 - 0,5erf[r\sqrt{1+\cos\delta}/2\sigma_1].$$

Теперь определим вероятность неправильной подстройки фазы ОГ для обоих вариантов.

Вариант 1.

Пусть, подстройка фазы ОГ производится по n символам и требуемое значение фазы ОГ выбрано правильно. Устройство должно суммировать значения всех символов после демодуляции. Каждое из них будет равно либо $-r$, либо $r-\Delta$. Поскольку появление обоих вариантов можно считать равновероятным, то среднее арифметическое значение r_M такой суммы будет близким к $-n\Delta/2$. То есть индикатором правильной настройки фазы является отрицательное значение данной суммы.

В результате действия шума каждое из значений имеет гауссово распределение с дисперсией, равной σ_1^2, поэтому и среднее арифметическое значение так-

же будет иметь гауссово распределение с дисперсией, равной $n\sigma_1^2$. Индикатором правильной настройки фазы ОГ служит знак этой суммы. Если он отрицательный, то фаза генератора настроена правильно. Если он положительный, то фаза ОГ выбрана неправильно, и вместо передаваемой информационной последовательности при демодуляции формируется ее инвертированная копия. Это тоже является решением, и проведением последующего дополнительного инвертирования правильная информационная последовательность восстанавливается.

Ошибка при демодуляции будет иметь место, если в первом случае значение величины r_M получится положительным, а во втором случае – отрицательным.

В обеих ситуациях ее величина будет равна

$$P_{D1} = 1 - F(n\Delta/\sigma_1^2 2\sqrt{n}) = 0,5 - 0,5erf(\Delta\sqrt{n}/2\sigma_1^2).$$

Вариант 2. При таком варианте асимметричного «созвездия» фаза ОГ настроится таким образом, что «созвездие» будет расположено либо как показано на рисунке 4 («правильная» настройка фазы ОГ, либо как показано на рисунке 5 («неправильная» настойка фазы генератора).

Для определения правильности настройки совместно с основным сигналом ОГ используется еще один вспомогательный сигнал, ортогональный ему. Его фаза отличается от фазы основного генератора на 90º. С помощью его сигнала также производится демодуляция входного сигнала. Результаты демодуляции усредняются за n символов, и определяется знак результата усреднения.

Результат усреднения представляет собой проекцию величиной $\Delta = r\sin\delta$ точки «созвездия» на вертикальную ось. Если она положительная, то фаза ОГ настроена правильно. Если проекция отрицательная, то фаза настроена неправильно.

Таким образом, аналогично предыдущему варианту, величина ошибки настройки при длине n анализируемой выборки будет равна

$$P_{D2} = 1 - F(nr\sin\delta/\sigma_1^2 2\sqrt{n}) =$$
$$= 0,5 - 0,5erf(\sqrt{n}r\sin\delta/2\sigma_1^2).$$

Теперь рассмотрим модуляцию QPSK. Здесь также возможны радиальная и угловая деформация «созвездия». На рис. 6 приведены оба этих варианта, обозначенные цифрами 1 и 2 соответственно. Кроме того, для удобства все созвездие повернуто на 45º.

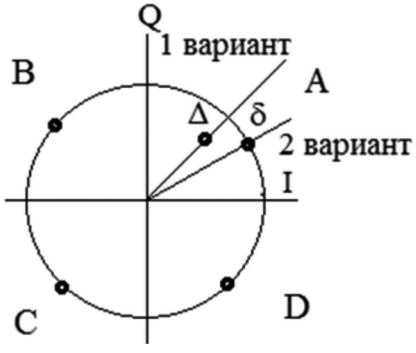


Рисунок 6

Попадание измеренной фазы каждого принятого символа в один из квадрантов (A, B, C, D) будет означать, что принята очередная пара информационных логических сигналов в соответствии с вариантом значений этой пары, приписанным к такому символу.

Из-за воздействия шумов фаза очередного символа может попасть не в нужный квадрант, а в соседний или в противоположный. При использовании кода Грея это приведет, соответственно, либо к ошибке в одном из пары информационных символов, либо в обоих информационных символах.

При той же длительности символов средняя вероятность возникновения ошибки из-за воздействия шумов будет равна для недеформированного «созвездия»

$$P_0 = 2\cdot 0,5[1 - F(r/\sigma_1\sqrt{2})] + [1 - F(r/\sigma_1\sqrt{2})]^2 =$$
$$= [1 - F(r/\sigma_1\sqrt{2})] + [1 - F(r/\sigma_1\sqrt{2})]^2.$$

Для радиальной деформации

$$P_1 = 0,75\{[1 - F(r/\sigma_1\sqrt{2})] + [1 - F(r/\sigma_1\sqrt{2})]^2\} +$$
$$+0,25\{1 - F[(r-\Delta)/\sigma\sqrt{2}]\} + 0,25\{1 - F[(r-\Delta)/\sigma_1\sqrt{2}]\}^2.$$

Для угловой деформации

$$P_1 = 0,75\{1 - F(\frac{r}{\sigma_1\sqrt{2}}) + [1 - F(\frac{r}{\sigma_1\sqrt{2}})]^2\} +$$
$$+0,25\cdot 0,5\{1 - F[\frac{r\sin(45° - \delta)}{\sigma_1}]\} + 0,25\times$$
$$\times 0,5\{1 - F[\frac{r\cos(45° - \delta)}{\sigma_1}]\} +$$
$$+0,25\{1 - F[\frac{r\sin(45° - \delta)}{\sigma_1}]\}\{1 - F[\frac{r\cos(45° - \delta)}{\sigma_1}]\}.$$

Определим вероятность неправильной подстройки фазы ОГ для обоих вариантов. Рассмотрим «созвездие» без поворота на 45º (рисунок 7).

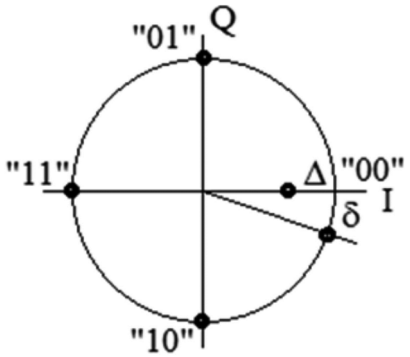


Рисунок 7

При удалении влияния модуляции учетверением частоты с последующей подстройкой генератора и делением его частоты на четыре, или с использованием схемы Костаса, фаза колебания генератора примет одно из четырех значений, близких к 0º, к 90º, к 180º или к 270º. Далее с использованием асимметричности «созвездия» определяется, какой из этих вариантов в текущий момент имеет место.

Пусть при «правильном» значении фазы генератора соответствие значений фазы и пары информационных символов таково, что фазе 0º соответствует пара логических символов «00», фазе 90º соответствует «01», фазе 180º соответствует «11» и фазе 270º соответствует «10».

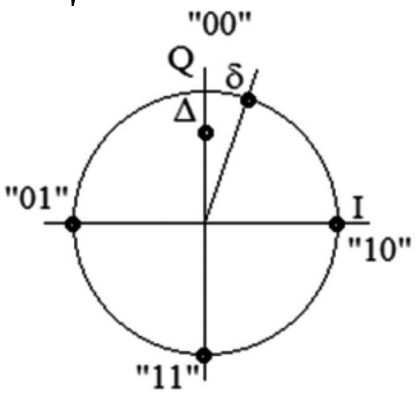


Рисунок 8

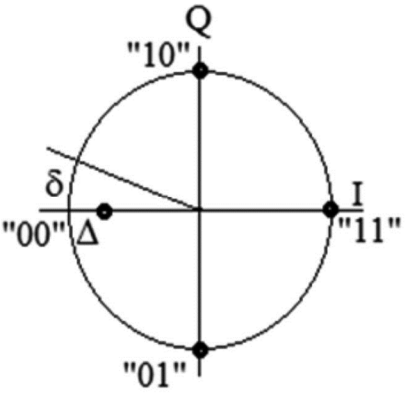


Рисунок 9

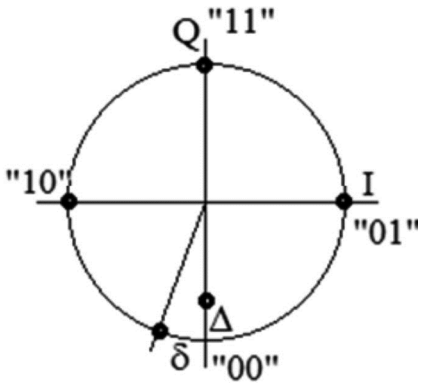


Рисунок. 10

Если же к фазе сигнала ОГ при подстройке дополнительно добавилось одно из других трех возможных, то, если станет известно, какое из них добавилось, в этой ситуации нет необходимости перестраивать фазу ОГ. Тогда достаточно определить, какая из них добавилась к «правильной» фазе, и уже на основе получившейся фазы генератора демодулировать принимаемый сигнал на основе одного из приведенных на рисунках 7-10 вариантов соответствия фазы и значения пары логических сигналов. Например, при правильной фазе генератора (рисунок 7) паре символов «00» соответствуют значения ортогональных составляющих $I > 0$, $Q \approx 0$, паре символов «01» – $I \approx 0$, $Q > 0$, паре символов «11» – $I < 0$, $Q \approx 0$, а паре символов «01» – $I \approx 0$, $Q < 0$.

При добавлении к правильной фазе генератора угла, равного 90º (рисунке 8), паре символов «10» соответствуют значения ортогональных составляющих $I > 0$, $Q \approx 0$, паре символов «00» – $I \approx 0$, $Q > 0$, паре символов «01» – $I < 0$, $Q \approx 0$, а паре символов «11» – $I \approx 0$, $Q < 0$. и т.д.

Для демодуляции QPSK используются два ортогональных напряжения. Пусть одно из них будет опорным, а другое, фаза которого сдвинута относительно опорного на 90º, – дополнительным. Определение, какая фаза сформирована, производится следующим образом. С помощью обоих генераторов производится вспомогательная демодуляция, то есть определяются значения ортогональных составляющих I и Q принимаемых символов. Далее для всех принятых символов складываются все полученные значения Q и I по отдельности каждой из обеих ортогональных компонент и определяются знаки и величины этих сумм, соответственно, r_{IM} и r_{QM}. Поскольку появление логических нулей и единиц в передаваемой последовательности можно считать равновероятным, то сочетание из знаков укажет на то дополнительное значение фазы, которое добавилось к ее правильному значению.

В частности, если будет зафиксировано сочетание: $r_{IM} > 0$, $r_{QM} \approx 0$, то сформирована «правильная» фаза ОГ (рисунок 7), если сочетание $r_{IM} \approx 0$, $r_{QM} > 0$, то при формировании фазы ОГ к нему добавилось значение 90º (рисунок 8). Если будет зафиксировано сочетание $r_{IM} < 0$, $r_{QM} \approx 0$, то при формировании фазы ОГ к нему добавилось значение 180º (рисунок 9). Если будет зафиксировано сочетание $r_{IM} \approx 0$, $r_{QM} < 0$, то при формировании фазы ОГ к нему добавилось значение 270º (рисунок 10). В соответствии с этим будет использован один из вариантов демодуляции.

Вариант 1 (радиальная деформация).

Как и в случае использования модуляции BPSK, подстройка фазы ОГ производится по n символам. Пусть требуемое значение фазы ОГ выбрано правильно. Устройство суммирует значения всех символов по отдельности по каждой ортогональной составляющей. По составляющей I каждое из них будет равно либо $-r$, либо r-Δ. Поскольку появление обоих вариантов можно считать равновероятным, то среднее арифметическое значение r_{IM} такой суммы будет близким к $-n\Delta/2$. По составляющей Q значения будут равны либо r, либо $-r$. Среднее арифметическое значение r_{QM} такой суммы будет близким к нулю.

В результате действия шума каждое из значений имеет гауссово распределение с дисперсией, равной σ_1^2, поэтому и среднее арифметическое значение обеих сумм также будет иметь гауссово распределение с дисперсией, равной $n\sigma_1^2$. Индикатором правильной настройки будет выполнение условий: $r_{IM} < 0$, $|r_{QM}| < |r_{IM}|$.

Вероятность неправильной настройки будет равна:

$$P_{D1} = [1 - F(\Delta\sqrt{n}/2\sigma_1)][F(\Delta\sqrt{n}/2\sigma_1) - F(-\Delta\sqrt{n}/2\sigma_1)].$$

Поскольку все четыре варианта значения сформированной дополнительной фазы равноправны, то и ошибка в трех остальных случаях будет иметь такое же выражение.

Вариант 2 (угловая деформация).

Здесь тоже сравниваются значения двух сумм результатов демодуляции принимаемых символов с ортогональными напряжениями, но значения пар логических символов по-другому соответствуют значениям сумм.

При правильном формировании фазы ОГ дополнительная фаза отсутствует (рисунок 7), соответствующие значения сумм будут $r_{IM} \approx 0$, $r_{QM} < 0$. При формировании дополнительной фазы 90º (рисунок 8) соответствующие значения сумм будут $r_{IM} > 0$, $r_{QM} \approx 0$. При формировании дополнительной фазы 180º (рисунок 9) соответствующие значения сумм будут $r_{IM} \approx 0$, $r_{QM} > 0$. При формировании дополнительной фазы 270º (рисунок 10) соответствующие значения сумм будут $r_{IM} < 0$, $r_{QM} \approx 0$.

Аналогично предыдущему варианту, вероятность неправильной настройки при любом значении дополнительной фазы будет равна:

$$P_{D1} = [1 - F(r\sin\delta\sqrt{n}/2\sigma_1)][F(r\sin\delta\sqrt{n}/2\sigma_1) -$$
$$-F(-r\sin\delta\sqrt{n}/2\sigma_1)].$$

Поскольку здесь также все четыре варианта значения сформированной дополнительной фазы равноправны, то и ошибка в трех остальных случаях будет иметь такое же выражение.

На рисунках 11-13 приведены в качестве примера графики зависимостей, полученных на основе приведенных выражений. Графики на рисунке 11 показывают зависимость вероятности битовой ошибки для модуляции BPSK (графики под номером 2) и QPSK (графики под номером 1) от величины отношения «сигнал/шум». Графики для ситуации отсутствия асимметрии обозначены точками, для радиальной симметрии – сплошными линиями, а для ситуации угловой асимметрии – прерывистыми линиями. Параметры асимметрии: $\Delta = 0{,}3r$, $\delta = 0{,}3$ радиан.

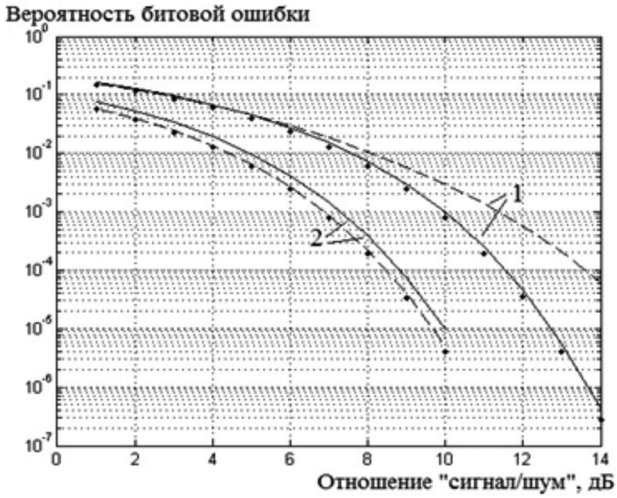


Рисунок 11

Графики на рисунке. 12 относятся к модуляции BPSK, графики на рисунке 13 – к модуляции QPSK. Сплошные линии соответствуют радиальной асимметрии, прерывистые – угловой. Номера графиков на обоих рисунках относятся к количеству усредняемых символов: 1) $n = 10$, 2) $n = 20$, 3) $n = 30$.

Реализация описанных методов введения асимметрии в «созвездие»

На рисунках 14 и 15 приведены структурные схемы для реализации метода введения асимметрии при BPSK (рисунок 14 для радиальной деформации и рисунок 15 для угловой) и QPSK (рисунок 16 для радиальной и для угловой деформации).

На рисунке 14 на основе принятого сигнала блоком автоподстройки фазы (БАПФ) фаза ОГ подстраивается под фазу несущей. В корреляторе (Корр) с помощью напряжения ОГ производится корреляционная обработка принятых сигналов, т.е. их перемножение на напряжение ОГ и усреднение результатов перемножения на интервале длительности символа. В усреднителе (Уср) усредняется определенное количество значений демодулированных символов, а в детекторе знака (ДЗ) определяется знак получившегося результата усреднения.

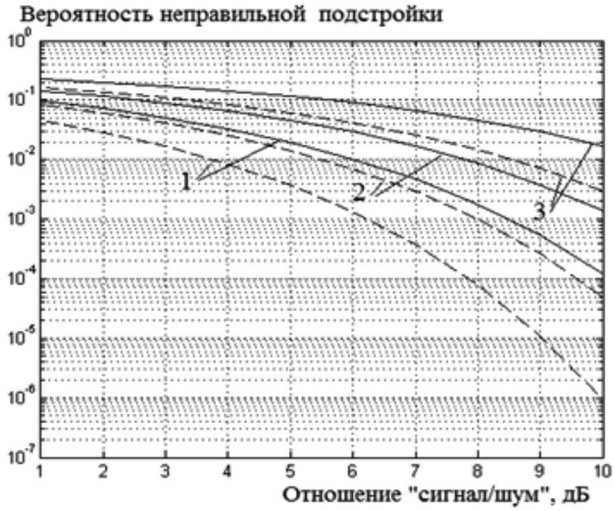


Рисунок 12

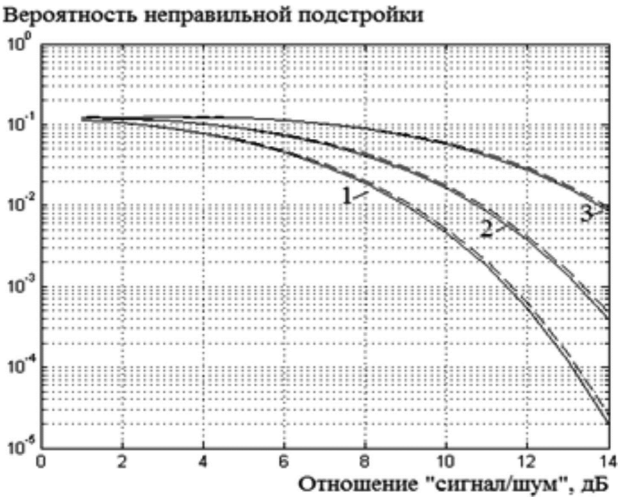


Рисунок 13

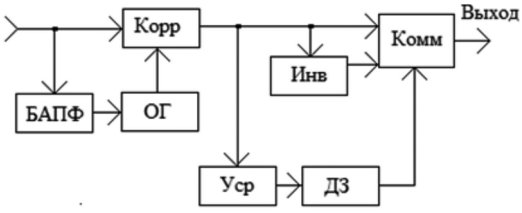


Рисунок 14

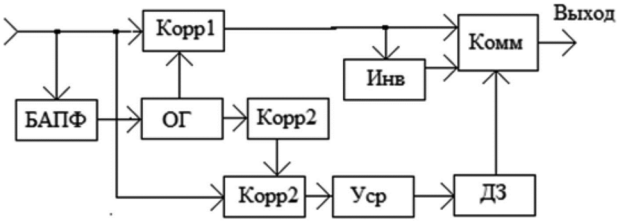


Рисунок 15

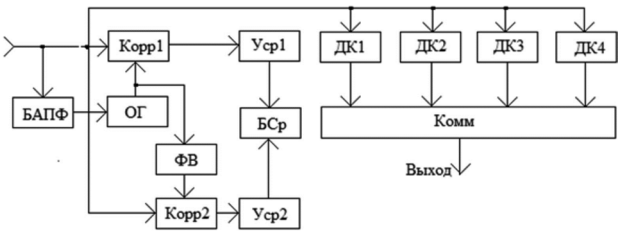


Рисунок 16

Этот знак укажет на наличие или отсутствие добавленной фазы 180º к «правильной» фазе ОГ, то есть последовательность принятых символов демодулирована правильно или она получилась инвертированной. Этот знак управляет работой коммутатора (Комм). В одной ситуации, когда фаза ОГ сформирована правильно, коммутатор подключает на выход схемы для дальнейшей обработки сигнал непосредственно с выхода корреляторa. В другой ситуации, если к «правильной» фазе ОГ добавилась дополнительная фаза 180º, сигнал корреляторa инвертируется в инверторе (Инв). Правильный результат демодуляции восстанавливается и подается коммутатором на выход схемы.

В схеме для угловой деформации (рисунок 15) обозначения блоков такие же, но дополнительно добавлены еще один коррелятор (Корр 2) и блок фазовращателя на 90º (Фв). На второй коррелятор подается напряжение, ортогональное напряжению, подаваемому на первый коррелятор. Далее его выходное напряжение также усредняется за определенное количество символов, и детектор знака определяет знак результата усреднения. На основе его управляется коммутатор, подключая на выход схемы либо сигнал непосредственно с первого корреляторa, либо после инвертора.

На рисунке 16 приведена схема, реализующая метод для радиальной и угловой деформации при модуляции QPSK. Сформированный сигнал ОГ может иметь либо «правильную» фазу, либо к нему может добавиться фаза, равная 90º, 180º, или 270º. Каждый из декодеров ДК1, ДК2, ДК3 и ДК4 настроен для декодирования одного из этих четырех вариантов. Для определения, какой же из них имеет место в текущий момент, входной сигнал демодулируется с помощью двух корреляторов, на которые подаются ортогональные напряжения. В блоке сравнения (БСр) анализируется сочетание значений результатов усреднения выходных напряжений корреляторов и, в соответствии с описанным правилом, управляется коммутатор, подключающий на выход схемы тот из четырех декодеров, который правильно декодирует входной сигнал.

Заключение

Использование деформированного вида «созвездий» при фазовой модуляции позволяет отказаться от применения дифференциальных видов модуляции, что дает возможность повысить помехоустойчивость передачи информации на 1-3 дБ. Вводимая асимметрия может быть различного вида, однако предпочтительнее угловая асимметрия, полнее использующая энергетические возможности передатчика.

Литература

1. Скляр Б. Цифровая связь. Теоретические основы и практическое применение / пер. с англ. М.: Изд. дом «Вильямс», 2003. 1104 с.

2. Полушин П.А., Киселев А.Ю., Шалина В.В. Исследование влияния шумов приемника на систему фазовой автоподстройки с умножением частоты при использовании фазоманипулированных сигналов. Проектирование и технология электронных средств. 2022. № 2. С. 26-29.

3. Полушин П.А., Никитин О.Р. Модификация схемы Костаса для автоподстройки фазы в системах с подавленной несущей. Радиотехнические и телекоммуникационные системы. 2022. № 2. С. 58-66.

4. Патент РФ № 208271 на полезную модель «Устройство фазовой автоподстройки системы с фазовой модуляцией», авторы Полушин П.А., Никитин О.Р., Прасолов А.В., Бобрус С.Ю., Зарегистрирован 01.12.21, Бюл. № 34.

5. Lee-Fang Wei. Rotationally Invariant Convolutional Channel Coding with Expanded Signal Space. IEEE Journal on Selected Areas in Communications. vol. SAC-2, no. 5, September, 1984.

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:

– № 1 март 2025 г. Тематический выпуск: «ЦОС в инфокоммуникационных системах».

– № 2 июнь 2025 г. Тематический выпуск по материалам 27-й Международной научно-технической конференции «Цифровая обработка сигналов и ее применение-DSPA».

– № 3 сентябрь 2025 г. Тематический выпуск: «Цифровая обработка изображений».

– № 4 декабрь 2025 г. Тематический выпуск: «ЦОС в радиотехнике и системах телекоммуникаций».

Ориентировочная стоимость рекламных услуг:

4-я (внешняя) страница цветной обложки – 25 тысяч рублей.

2-я и 3-я (внутренние) страницы цветной обложки – 15 тысяч рублей.

1\2 цветной внутренней полосы – 8 тысяч рублей.

Ждем Ваших предложений.

С наилучшими пожеланиями, зам. главного редактора

д.т.н., профессор Витязев Владимир Викторович, телефон 8-903-834-81-81.

Предложения прошу направлять по адресу: E-mail: vityazev.v.v@rsreu.ru или info@dspa.ru

УДК 621.382

РЕАЛИЗАЦИЯ НА ПЛИС МОДУЛЯ ИСКУССТВЕННОГО НЕЙРОНА ДЛЯ ПОСЛЕДОВАТЕЛЬНО-ПАРАЛЛЕЛЬНЫХ АРХИТЕКТУР НЕЙРОННЫХ СЕТЕЙ С ПРЯМОЙ СВЯЗЬЮ*

Ушенина И.В., *к.т.н., доцент кафедры «Программирование» Пензенского государственного технологического университета, e-mail: ivl23@yandex.ru*
Данилов Е.А., *к.т.н., доцент кафедры «Программирование» Пензенского государственного технологического университета, e-mail: evgenydplus@mail.ru*

FPGA IMPLEMENTATION OF THE ARTIFICAL NEURON MODULE FOR SERIES-PARALLEL ARCHITECTURES OF FEEDFORWARD NEURAL NETWORKS

Ushenina I.V., Danilov E.A.

In this paper, the authors consider the levels of parallelism used in feedforward neural networks and, taking this into account, evaluate the possibilities of implementing artificial neurons using FPGA resources. As a result, the artificial neuron module that supports neuron-level parallelism is implemented on FPGA and analyzed. The module has been designed so that when implementing a neural network layer by connecting several instances of the module into the chain, no external elements are necessary in the layer's circuit. This feature provides predictable timing characteristics for the neural network layer being implemented. In the module, the linear part of calculations and calculation of activation function are made by a DSP block and programmable logic blocks, respectively.

Key words: neuron, neuron-level parallelism, bit-level mapping method, FPGA, DSP block, timing analysis.

Ключевые слова: нейрон, параллелизм на уровне нейронов, метод поразрядного отображения, ПЛИС, блок ЦОС, временной анализ.

Исходя из уровней параллелизма вычислений в нейронных сетях с прямой связью, авторы оценивают возможности построения на ресурсах ПЛИС схем искусственных нейронов. В итоге реализован на ПЛИС и проанализирован модуль искусственного нейрона, поддерживающий параллелизм на уровне нейронов. Модуль содержит все необходимое, чтобы при формировании слоя нейронной сети из нужного количества его экземпляров требовалось только их объединение в цепочку без добавления к ней каких-либо внешних элементов. Это обеспечит предсказуемость временных характеристик реализуемой нейронной сети. Линейная часть вычислений в модуле выполняется блоком цифровой обработки сигналов ПЛИС, вычисление функции активации – блоками программируемой логики.

Введение

Аппаратная реализация нейронных сетей выполняется для ускорения их работы или обучения, а также, если требуется работа нейронной сети в реальном времени. Одним из требований к аппаратной платформе, на которой реализуется нейронная сеть, является возможность выполнения массивно-параллельных вычислений. Этому требованию удовлетворяют несколько видов платформ, в том числе графические процессоры (GPU), специализированные микросхемы и системы на кристалле (ASIC, SoC), ПЛИС (FPGA). Последние могут проигрывать другим платформам в производительности и энергоэффективности, но являются наиболее доступными [1-3]. Согласно обзору [3], около двух третей случаев аппаратной реализации алгоритмов искусственного интеллекта, в том числе нейронных сетей, приходятся на ПЛИС.

Поскольку нейронная сеть – это, по сути, массив соединенных между собой нужным образом одинаковых вычислительных блоков (искусственных нейронов), при реализации нейронной сети на ПЛИС логично оформить описание нейрона в виде модуля и использовать нужное количество его экземпляров. Однако в зависимости от вида параллелизма вычислений в нейронной сети схемы нейронов могут выглядеть по-разному и формировать нейронные сети, работающие с разной производительностью.

С этой точки зрения в работе проанализированы варианты реализации на ПЛИС модулей нейронов для нейронных сетей с прямой связью типа «многослойный персептрон». В итоге выбрана и реализована схема модуля с параллелизмом вычислений на уровне нейронов. При реализации слоя нейронной сети нужное количество экземпляров этого модуля объединяется без каких-либо дополнений получившейся схемы. Модуль реализован на ПЛИС GW5AST от Gowin Semiconductor [4]; оценены его производительность и ресурсоемкость. Для описания, реализации и временного анализа схемы модуля использована среда проектирования Gowin EDA.

Вычисления в нейронах и ресурсы ПЛИС

На рисунке 1 представлено устройство искусственного нейрона. Нейрон под номером k, расположенный в i-м слое нейронной сети, содержащем K_i нейронов, принимает от K_{i-1} нейронов предыдущего слоя сигналы x_{k1}^i … $x_{kK_{i-1}}^i$ и умножает их на весовые коэффициенты w_{k1}^i … $w_{kK_{i-1}}^i$, а полученные произведения складывает и добав-

ляет к ним пороговое значение b_k^i:

$$a_k^i = b_k^i + \sum_{j=1}^{K_{i-1}} x_{kj}^i w_{kj}^j . \quad (1)$$

Полученная сумма (обозначим ее a_k^i) поступает на вычислитель функции активации (ФА), в качестве которой выступает сигмоида:

$$s(a_k^i) = \frac{1}{1+e^{-a_k^i}} . \quad (2)$$

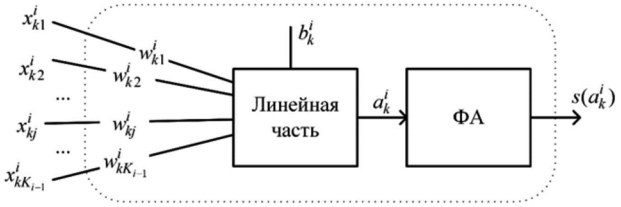


Рисунок 1 – Устройство искусственного нейрона

Умножение с накоплением, выполняемое в (1) – линейная операция, требующая сумматоров и умножителей. Так как она выполняется многократно, скорость ее выполнения оказывает большое влияние на производительность нейронной сети в целом. Очевидным решением для реализации вычислений по ней являются блоки цифровой обработки сигналов (ЦОС), которые присутствуют почти во всех современных ПЛИС хотя бы в количестве нескольких десятков. Типичный блок ЦОС содержит хотя бы один умножитель и сумматор, на базе которого можно реализовать аккумулятор. Блоки ЦОС можно объединять в цепочки [5].

В качестве ФА в нейронных сетях с прямой связью чаще всего используются сигмоида или гиперболический тангенс. Эти функции не только имеют подходящую форму, но и непрерывно дифференцируются, что необходимо для обучения нейронной сети по алгоритму обратного распространения ошибки. Обе функции – нелинейные, и для их непосредственного вычисления ресурс у ПЛИС отсутствует. В данной работе для вычисления сигмоиды (2) используется метод поразрядного отображения [6, 7]. Суть метода в том, что разряды двоичного представления $s(a_k^i)$ вычисляются отдельно друг от друга как булевы функции разрядов a_k^i. Для таких вычислений можно использовать блоки программируемой логики ПЛИС, сохранив менее многочисленные блоки ОЗУ и ЦОС для линейной части вычислений (1). В [7] авторы представили схемы вычисления отдельных разрядов $s(a_k^i)$. В данной работе в вычислителях ФА (рисунок 2) используется по два набора таких схем, принимающих значащие разряды аргумента функции сигмоиды, a_k^i. Первый набор вычисляет ФА, исходя из предположения, что $a_k^i < 0$. Полученный им результат обозначен на рисунке 2 как $sn(a_k^i)$. Второй набор схем предполагает, что $a_k^i \geq 0$, а полученный им результат обозначен как $sp(a_k^i)$. Далее нужное значение ФА выбирается мультиплексором в зависимости от знакового разряда a_k^i.

Таким образом, линейная часть вычислений в разрабатываемом модуле искусственного нейрона выполняется блоками ЦОС, а нелинейная – блоками программируемой логики. Коэффициенты для взвешивания входных сигналов нейронов хранятся в блоках ОЗУ ПЛИС.

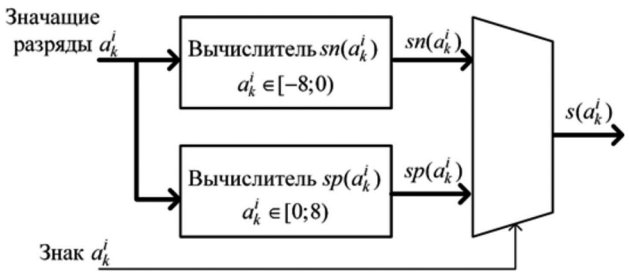


Рисунок 2 – Вычислитель функции активации

В [8] показано, что при аппаратной реализации обученной нейронной сети достаточно представления всех чисел, которыми она оперирует, в формате с фиксированной запятой с точностью 7-8 бит. Реализуемый здесь модуль искусственного нейрона работает с числами, значащая часть которых представлена 8 битами, а знаковый разряд, когда он нужен – дополнительным 9-м битом. Подробности представления чисел в модуле искусственного нейрона приведены в таблице 1.

Уровни параллелизма вычислений в нейронных сетях

При ускорении нейронных сетей или организации их работы в реальном времени важнейшей характеристикой реализуемых схем является производительность, которая во многом зависит от степени параллелизма выполнения вычислений. Выделяют [9-11] несколько

Таблица 1 – Форматы представления чисел в модуле искусственного нейрона

Вид переменной	Обозначение	Диапазон значений	Знаковый разряд, бит	Целая часть, бит	Дробная часть, бит
j-й входной сигнал *k*-го нейрона в i-м слое	x_{kj}^i	[0;1)	0	0	8
Выходной сигнал *k*-го нейрона в *i*-м слое	y_k^i	[0;1)	0	0	8
Весовой коэффициент *j*-го входного сигнала, поступающего на *k*-й нейрон *i*-го слоя	w_{kj}^i	[-8;8)	1	3	5
Аргумент ФА *k*-го нейрона в *i*-м слое	a_k^i	[-8;8)	1	3	5
Значение ФА *k*-го нейрона в *i*-м слое	$s(a_k^i)$	[0;1)	0	0	8

уровней параллелизма вычислений в нейронных сетях, среди которых: 1) уровень слоев, когда каждый слой нейронной сети реализован на отдельных ресурсах, что позволяет организовать конвейерные вычисления; 2) уровень нейронов, когда каждый нейрон занимает отдельный ресурс, и вычисления во всех нейронах как минимум в пределах одного слоя нейронной сети выполняются параллельно; 3) уровень синапсов, то есть поступающих на нейрон входных сигналов, при котором все они умножаются на весовые коэффициенты одновременно, а последующее сложение полученных произведений по возможности распараллеливается.

Нейронные сети могут сочетать несколько уровней параллелизма. Крайними случаями являются полное отсутствие параллелизма, когда вычисления, требуемые от каждого нейрона в каждом слое, выполняются последовательно единственным вычислительным блоком [12], и параллелизм на всех трех уровнях [11]. Тем не менее, при полном отсутствии параллелизма возможности ПЛИС не используются в полной мере, а параллелизм на всех уровнях реализуем на большинстве ПЛИС только при малых размерах сети.

Если ПЛИС содержит небольшое (относительно размеров сети) количество блоков ЦОС, вероятно, придется отказаться от параллелизма на уровне слоев и выбрать один из уровней параллелизма – нейронов, как, например, в [13-15], или синапсов, как в [16]. Сравним эти два варианта.

При параллелизме на уровне нейронов каждый нейрон в слое нейронной сети автономен, так как реализуется на отдельных ресурсах ПЛИС (рисунок 3). На все нейроны слоя в каждый момент времени поступает сигнал только от одного из нейронов предыдущего слоя. То есть, на каждый k-й нейрон i-го слоя поступают с каждым тактом сигналы x_{k1}^i, x_{k2}^i, …, x_{kj}^i, …, x_{kK-1}^i. Весовые коэффициенты w_{k1}^i, w_{k2}^i, …, w_{kj}^i, …, w_{kK-1}^i с каждым тактом считываются из блоков памяти (ОЗУ), подключенных к умножителям. Согласно (1), x_{kj}^i и w_{kj}^i перемножаются, и их произведение поступает в аккумулятор (его образуют сумматор и следующий за ним регистр). Перед началом цикла K_{i-1} операций умножения с накоплением в регистр аккумулятора помещается пороговое значение b_k^i. Результат вычислений (1), a_k^i, поступает на блок вычисления ФА.

При отсутствии параллелизма на уровне слоев выходные сигналы нейронов, y_k^i, должны быть снова по очереди поданы на их входы для выполнения вычислений следующего, i + 1-го слоя сети. Сигналы y_k^i принимает мультиплексор, выход которого синхронизируется регистром 1. Вместо мультиплексора для поочередной подачи сигналов y_k^i на входы нейронов может использоваться регистр сдвига [13, 14].

Структура, представленная на рисунке 3, может использоваться для реализации слоев нейронных сетей, начиная со второго. Нейроны первого слоя не обрабатывают, а только перераспределяют поступающие сигналы. Для выбора между сигналами от первого слоя, x_{kj}^2, и последующих слоев, x_{kj}^i, перед умножителями установлен еще один мультиплексор.

Обозначив T_{clk} период тактового сигнала, поступающего на регистры и ОЗУ схемы на рисунке 3, получим, что на все вычисления в i-м слое нейронной сети требуется время $t_н$:

$$t_н = T_{clk} \cdot (K_{i-1} + 2). \quad (3)$$

Определения параллелизма на уровне синапсов, данные в [9-11], предписывают нейронам выполнять все операции умножения x_{kj}^i на w_{kj}^i одновременно и по возможности распараллеливать последующие операции суммирования полученных произведений. Таким образом, нейрон должен содержать K_{i-1} блоков ЦОС для выполнения умножений.

Для распараллеливания суммирования произведений в [9] предлагается использовать в нейроне «дерево» сумматоров. Тогда выходной сигнал нейрона, y_k^i, будет вычислен за время, пропорциональное $\log_2 K_{i-1}$. Но древовидная организация сумматоров на ПЛИС невыгодна: она требует или повышенного расхода блоков ЦОС, или реализации части сумматоров на блоках программируемой логики с потерей производительности.

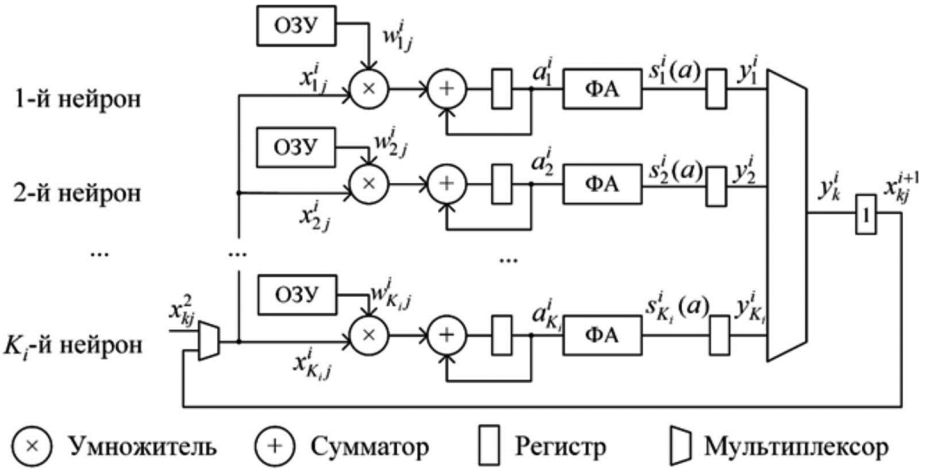


Рисунок 3 – Слой нейронной сети с параллелизмом на уровне нейронов (параллелизм на уровнях синапсов и слоев отсутствует; вместо мультиплексора сигналов y_k^i может использоваться регистр сдвига)

Для ПЛИС больше подходит лестничная организация сумматоров в нейроне, при которой сумматоры соединены последовательно, образуя цепочку. То есть, в каждом из K_{i-1} блоков ЦОС нейрона выполняется взвешивание одного из его входных сигналов и передача произведения на один из входов сумматора. На другой вход сумматора поступает сумма произведений x_{kj}^i и w_{kj}^i, вычисленная предыдущими сумматорами цепочки. На первый сумматор в цепочке поступает, кроме произведения $x_{k1}^i w_{k1}^i$, пороговое значение b_k^i. Значение y_k^i при лестничной организации сумматоров будет вычисляться за время, пропорциональное K_{i-1}.

При отсутствии параллелизма на уровне нейронов и строгом следовании определению параллелизма на уровне синапсов схема с лестничной организацией сумматоров выполняет вычисления для нейронов *i*-го слоя по очереди и проигрывает в производительности схеме с параллелизмом на уровне нейронов (3) примерно в K_i раз.

Если отклониться от определения параллелизма на уровне синапсов и допустить конвейеризацию вычислений в умножителях и сумматорах в пределах слоя нейронной сети, можем получить схему, представленную на рис. 4, где лестничная организация сумматоров совмещена с конвейеризацией вычислений. То есть, часть умножителей и сумматоров, уже выполнивших расчеты для *k*-го нейрона, начинают обслуживать (*k*+1)-й нейрон, (*k*+2)-й и т.д. Входные сигналы $x_{k1}^i \dots x_{kK-1}^i$. поступают на эту схему одновременно, и для любого k это будет один и тот же набор сигналов. При каждом *j*-м ($j \in [1; K_{i-1}]$) блоке ЦОС находится блок памяти (ОЗУ), хранящий набор коэффициентов $w_{1j}^i \dots w_{K_i j}^i$, одним из которых будет взвешиваться *j*-й входной сигнал реализуемого в данный момент времени *k*-го нейрона. По окончании расчетов, соответствующих *k*-у нейрону, формируется выходной сигнал y_k^i, который поступает на цепочку регистров. По окончании расчетов, соответствующих K_i-у нейрону, на выходах регистров этой цепочки содержится набор сигналов $y_1^i \dots y_{K_i}^i$, которые поступают на мультиплексоры. На мультиплексоры также поступают сигналы от первого слоя сети – $x_{k1}^2 \dots x_{kK_1}^2$. При переходе к обслуживанию следующего, *i* + 1-го, слоя нейронной сети выходы мультиплексоров синхронизируются и поступают на умножители.

При использовании схемы на рисунке 4 все вычисления в *i*-м слое нейронной сети займут время t_c

$t_c = T_{clk} \cdot (K_{i-1} + K_i + 1)$.

Для схем на рисунке 3 и рисунке 4 минимальное значение T_{clk} будет определяться временем одной итерации вычислений по (1), выполняемых блоком ЦОС, или временем вычисления ФА. Таким образом, конвейеризация вычислений в пределах одного слоя в схеме с параллелизмом на уровне синапсов и лестничной организацией сумматоров сокращает ее отставание в производительности от схемы с параллелизмом на уровне нейронов.

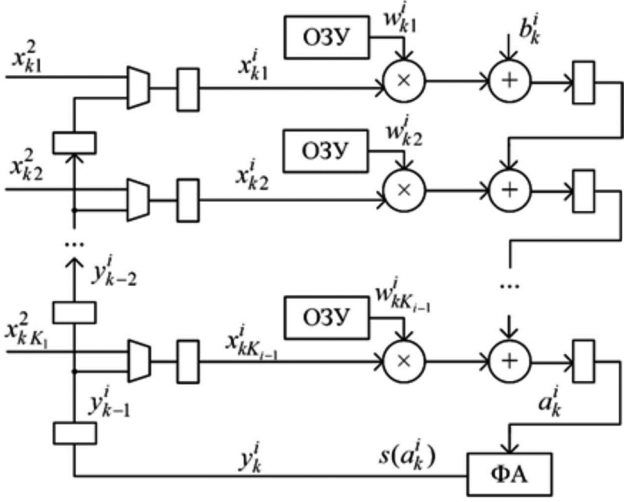


Рисунок 4 – Слой нейронной сети с параллелизмом на уровне синапсов и конвейеризацией вычислений в пределах слоя (параллелизм на уровнях нейронов и слоев отсутствует)

Это отставание можно сократить еще больше при дальнейшей конвейеризации схемы с параллелизмом на уровне синапсов и выполнении в каждый момент времени вычислений, относящихся не только к разным нейронам, но и к разным слоям. Однако по сравнению со схемой с параллелизмом на уровне нейронов (рисунок 3) у полученной схемы будет ряд недостатков.

Во-первых, построение нейронной сети на основе схемы с параллелизмом на уровне синапсов – это не соединение нужного количества экземпляров модуля искусственного нейрона, а настройка размеров одного модуля в зависимости от количества нейронов в предыдущем слое. Это может сделать комплекс мер для сохранения высокой производительности нейронной сети и их эффективность зависимыми от размеров сети.

Во-вторых, при отсутствии параллелизма на уровне слоев размеры схемы (количество в ней блоков ЦОС) определяются размерами слоя с максимальным количеством нейронов. При обслуживании слоев с меньшим количеством нейронов промежуточные суммы произведений проходят более длинные, чем необходимо, цепочки сумматоров, что замедляет работу нейронной сети.

В-третьих, в отличие от параллелизма на уровне нейронов, где все нейроны слоя могут управляться одним и тем же набором сигналов, схема с параллелизмом на уровне синапсов нуждается в индивидуальном управлении каждым блоком ЦОС. Это усложняет управляющий автомат и может увеличить нагрузку на трассировочные ресурсы ПЛИС.

Таким образом, для модуля искусственного нейрона, реализуемого на ПЛИС, предпочтительнее оказывается поддержка параллелизма на уровне нейронов.

Схема модуля искусственного нейрона

Представим схему модуля искусственного нейрона, поддерживающую параллелизм на уровне нейронов и предназначенную для реализации на ПЛИС GW5AST.

При разработке модуля предполагалось его исполь-

зование в нейронных сетях как с параллелизмом на уровне слоев, так и без него. От модуля также требовалось, чтобы его схема не зависела от количества нейронов в слоях сети, а также, чтобы при формировании из его экземпляров нейронной сети дополнительные внешние элементы были не нужны, и сеть представляла бы собой массив одинаковых вычислительных блоков. Все это обеспечивает предсказуемость временных характеристик как отдельного модуля, так и составленной из его экземпляров нейронной сети, а также облегчает описание нейронных сетей различных размеров.

С этой точки зрения рассмотрены два варианта обработки массива выходных сигналов параллельно работающих нейронов, y_k^i, которые при переходе к вычислениям следующего слоя должны по очереди снова поступать на входы нейронов. Первый вариант, показанный на рисунке 3, – использование мультиплексора с K_i входами. Так как ПЛИС не располагают мультиплексорами с большим количеством входов, на практике это будет «дерево» из многих мультиплексоров с намного меньшим, чем K_i, количеством входов у каждого из них. Внизу «дерева» будет находиться несколько мультиплексоров с суммарным количеством входов, равным K_i. На вершине «дерева» будет находиться один мультиплексор, формирующий выходной сигнал. Сколько мультиплексоров пройдут выходные сигналы нейронов и за какое время, зависит от величины K_i и от архитектуры ПЛИС. Мультиплексоры «дерева» не получится равномерно распределить по модулям отдельных нейронов, то есть «дерево» придется оформлять отдельным модулем и настраивать в зависимости от размеров слоя сети, что противоречит выдвинутым выше требованиям.

Второй вариант – использование регистра сдвига с параллельной загрузкой, количество звеньев которого равно K_i. Модуль искусственного нейрона сможет содержать одно звено этого регистра, соединяемое со звеньями других модулей при формировании слоя сети. Чтобы все разряды y_k^i сдвигались за один такт, каждый разряд y_k^i передается по отдельной сдвигающей цепочке триггеров. Таким образом, приходим к схеме регистра сдвига, представленной на рисунке 5. Каждое звено этого регистра представлено набором триггеров с мультиплексорами перед входами данных. На один из входов мультиплексоры принимают разряды выходного сигнала y_k^i «своего» нейрона, на другой – разряды выходного сигнала звена регистра сдвига предыдущего нейрона.

На рисунке 6 представлена схема модуля искусственного нейрона, реализованная на ПЛИС GW5AST. Блоки ЦОС этой ПЛИС содержат по два умножителя, а значит, вычисление выходного сигнала идет почти в два раза быстрее, чем представлено в (3). Весовые коэффициенты для обоих входных сигналов, x_{kj}^i и x_{kj+1}^i, хранятся в общем блоке памяти (ОЗУ). Ячейки ОЗУ разрядностью 18 бит хранят по два весовых коэффициента, считываемых одновременно.

Умножители модуля могут принимать как сигналы от первого слоя нейронной сети, то есть x_{kj}^2 и x_{kj+1}^2, так и сигналы x_{kj}^i и x_{kj+1}^i, рассчитанные ранее этим и другими модулями и поступающие от регистра сдвига. Для выбора нужных входных сигналов перед умножителями установлены мультиплексоры. В отличие от схемы на рисунке 3, входные сигналы умножителей синхронизированы регистрами 5-8.

В связи с тем, что в блоках ЦОС содержится по два умножителя, все модули, составляющие слой нейронной сети, разбиваются на две равные группы с отдельными друг от друга регистрами сдвига выходных сигналов y_k^i.

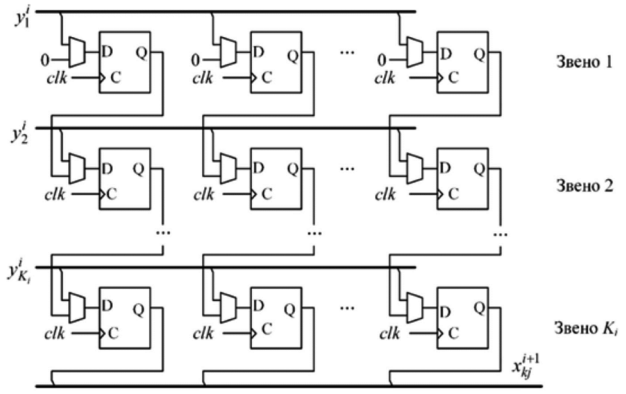


Рисунок 5 – Регистр сдвига со звеньями, распределяемыми по модулям искусственного нейрона

Таким образом, на все вычисления в слое нейронной сети, состоящем из представленных на рисунке 6 модулей, потребуется время $t_н$:

$$t_н = T_{clk} \cdot (\frac{K_{i-1}}{2} + 4) .$$

Единственное дополнение, которое может потребоваться схеме слоя нейронной сети, состоящей из K_i экземпляров представленного модуля, – это регистры, расставленные на путях следования сигналов от регистров сдвига к умножителям. Эти регистры нужны для сохранения производительности (тактовой частоты) на уровне, достигнутом внутри модулей.

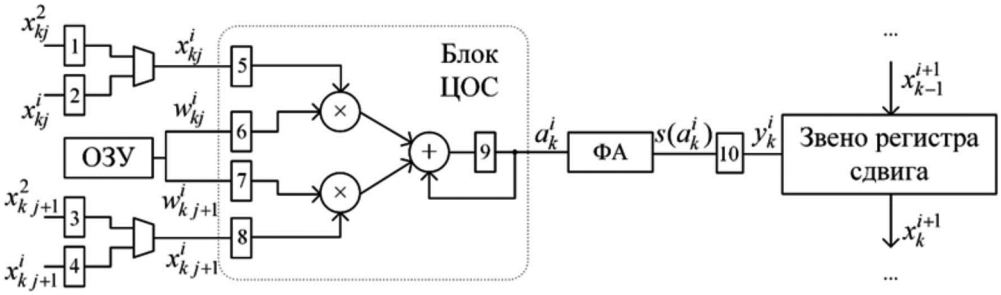


Рисунок 6 – Схема модуля искусственного нейрона, реализованная на ПЛИС GW5AST

Результаты реализации модуля искусственного нейрона

Результаты реализации модуля искусственного нейрона на ПЛИС GW5AST представлены в таблицах 2 и 3. В таблице 2 приведены потребности модуля в различных видах ресурсов ПЛИС. На реализацию модуля требуется один блок ЦОС, один блок ОЗУ и 27 блоков программируемой логики.

ПЛИС GW5AST содержит 298 блоков ЦОС, поэтому при отсутствии параллелизма на уровне слоев на ней можно реализовывать нейронные сети, содержащие в слоях до 298 нейронов.

Общее количество нейронов сети, реализуемой на ПЛИС, ограничивается также ее запасами блочной памяти; в данном случае ПЛИС содержит 340 блоков ОЗУ по 18 кбит. На хранение весовых коэффициентов для сигналов, поступающих от одного слоя из 298 нейронов, требуется 2,62 кбит памяти. Таким образом, максимальный размер нейронной сети, который может быть реализован на выбранной ПЛИС с помощью представленного модуля – 6 слоев по 298 нейронов в каждом слое.

На реализацию модуля также тратится 27 блоков программируемой логики, что составляет менее 0,16 % от их общего запаса в ПЛИС. При реализации нейронной сети с использованием всех имеющихся блоков ЦОС будет занято около 47 % от всего запаса блоков программируемой логики. То есть, блоки программируемой логики в данном случае не являются критическим ресурсом для реализации нейронной сети.

Если синхронизирующие регистры в модуле расположены, как показано на рисунке 6, получаются пять ступеней конвейера, которым соответствуют пять видов операций в слое нейронной сети. В таблице 3 приведены составляющие этих операций и время их выполнения. Видно, что больше всего времени (6,761 нс) занимает вычисление ФА. Это время и будет ограничивать максимальную тактовую частоту модуля $F_{\max}$. С учетом этого времени, качества и условий распространения тактового сигнала и времени установления синхронных элементов временной анализ модуля показал, что его $F_{\max}$ составляет 128 МГц.

При переходе от одного модуля к состоящей из его экземпляров нейронной сети без параллелизма на уровне слоев $F_{\max}$ снижается из-за большой задержки распространения сигналов от регистров сдвига до регистров 2 и 4 каждого нейрона (рисунок 6). При реализации слоя нейронной сети на ПЛИС GW5AST с использованием предложенного модуля эта задержка может составить до 11,67 нс. Чтобы вернуть $F_{\max}$ к уровню, достигнутому для отдельно взятого модуля, можно разместить по два регистра на путях следования x_{kj}^i и x_{kj+1}^i.

Таблица 2 – Ресурсоемкость модуля искусственного нейрона

Ресурсы ПЛИС и реализуемые на них элементы модуля	Требуемый объем ресурсов
Блоки ЦОС (умножители, сумматор, регистры 5-9)	1 блок из 298
Блоки памяти (ОЗУ)	1 блок из 340
Блоки программируемой логики (ФА)	21 блок из 17280
Блоки программируемой логики (звено регистра сдвига, регистры 1-4, регистр 10, мультиплексоры)	6 из 17280

Таблица 3 – Временные характеристики модуля искусственного нейрона

Вид операции	Составляющие операций	Время выполнения, нс
1. Доставка x_{kj}^i и w_{kj}^i к регистрам 5-8	Чтение ОЗУ	1,608
	Распространение сигналов от ОЗУ до регистров 6 и 7	до 1,371
	Запись регистров 1-4	0,306
	Распространение сигналов от регистров 1-4 до регистров 5 и 8 Из них время работы мультиплексоров	до 1,917 0,463
2. Одна итерация в блоке ЦОС	Умножение с накоплением согласно формуле (1)	3,338
3. Вычисление ФА	Запись регистра 9	0,306
	Распространение сигналов от регистра 9 до вычислителя ФА	до 2,922
	Вычисление ФА	3,531
	Распространение сигналов от вычислителя ФА до регистра 10	до 0,002
4. Доставка y_k^i к триггерам регистра сдвига	Запись регистра 10	0,306
	Распространение сигналов от регистра 10 до триггеров звена регистра сдвига (см. рисунок 5) Из них время работы мультиплексоров звена регистра сдвига	до 1,902 0,463
5. Прохождение звена регистра сдвига	Запись триггеров звена регистра сдвига	0,306
	Распространение сигналов от триггеров звена регистра сдвига в составе модуля до триггеров звена регистра сдвига в следующем модуле ((k+1)-й нейрон) Из них время работы мультиплексоров звена регистра сдвига в следующем модуле	до 0,632 0,463

Заключение

Реализован на ПЛИС и проанализирован модуль искусственного нейрона, предназначенный для нейронных сетей с прямой связью и поддерживающий параллелизм вычислений на уровне нейронов. За счет этого время вычислений в слое нейронной сети не зависит от количества нейронов в нем, а за счет наличия в модуле двух умножителей время вычислений в слое пропорционально половине количества нейронов в предыдущем слое. Независимо от размеров нейронной сети, состоящей из экземпляров модуля, ее производительность может быть сохранена на уровне производительности отдельных нейронов.

** Исследование выполнено за счет гранта Российского научного фонда и Пензенской области № 24-21-20100, https://rscf.ru/project/24-21-20100/*

Литература

1. Nurvitadhi E. et al. Can FPGAs beat GPUs in accelerating next-generation deep neural networks? Proceedings of the 2017 ACM/SIGDA international symposium on field-programmable gate arrays (FPGA'17). 2017, pp. 5-14.

2. Шипицин С.П., Ямаев М.И. Развитие аппаратно-ориентированных нейронных сетей на FPGA и ASIC. Вестник Пермского национального исследовательского политехнического университета. Электротехника, информационные технологии, системы управления. 2019. № 31. С. 177-192.

3. Talib M.A. et al. A systematic literature review on hardware implementation of artificial intelligence algorithms. The Journal of Supercomputing. 2021. vol. 77. no. 2. pp. 1897-1938.

4. GW5AST series of FPGA Products. URL: https://cdn.gowinsemi.com.cn/DS1104E.pdf (Дата обращения: 25.10.24).

5. Arora V Digital Signal Processing (DSP) User Guide. URL: https://www.gowinsemi.com/upload/database_doc/2927/document/6671c51a10e8b.pdf (Дата обращения: 25.10.24).

6. Tommiska M.T. Efficient digital implementation of the sigmoid function for reprogrammable logic. IEEE Proceedings-Computers and Digital Techniques. 2003, vol. 150, no. 6, pp. 403-411.

7. Ushenina I.V., Danilov E.A. Implementation of the Sigmoid Function on Current FPGAs Using the Bit-Level Mapping Method. Proceedings of the 2024 International Russian Automation Conference (RusAutoCon). 2024. pp. 295-300.

8. Holt J.L., Hwang J.N. Finite precision error analysis of neural network hardware implementations. IEEE Transactions on Computers. 1993, vol. 42, no. 3, pp. 281-290.

9. Omondi A.R., Rajapakse J.C. FPGA implementations of neural networks. New York: Springer, 2006. 360 p.

10. Nordström T., Svensson B. Using and designing massively parallel computers for artificial neural networks. Journal of parallel and distributed computing. 1992, vol. 14, no. 3, pp. 260-285.

11. Savich, A., Moussa, M., Areibi S. A scalable pipelined architecture for real-time computation of MLP-BP neural networks. Microprocessors and Microsystems. 2012, vol. 36, no. 2, pp. 138-150.

12. Gorbounov Y., Chen H. Optimized FPGA Implementation of an Artificial Neural Network Using a Single Neuron. Proceedings of International Conference on Computer Science and Education in Computer Science. 2023, pp. 233-246.

13. Medus L.D., Iakymchuk T., Frances-Villora J.V., Bataller-Mompeán M., Rosado-Munoz A. A novel systolic parallel hardware architecture for the FPGA acceleration of feedforward neural networks. IEEE Access. 2019, vol. 7, pp. 76084-76103.

14. Thasnimol V.S., George M.A. Hardware Accelerator Implementation of Multilayer Perceptron. Proceedings of Congress on Intelligent Systems (CIS 2020). 2021, vol. 1, pp. 107-119.

15. Vu H.M., Thang H.V. A Customized Hardware Architecture for Multi-layer Artificial Neural Networks on FPGA. Proceedings of Fourth International Conference on Information Systems Design and Intelligent Applications. 2018. pp. 637-644.

16. Gomperts A., Ukil A., Zurfluh F. Development and implementation of parameterized FPGA-based general purpose neural networks for online applications. IEEE Transactions on Industrial Informatics. 2010. vol. 7, no. 1, pp. 78-89.

НОВЫЕ КНИГИ



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-кодоманипулированных 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.