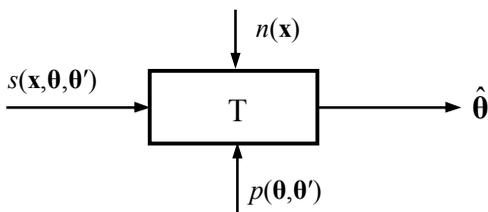


И.П. Гуров
ОСНОВЫ ТЕОРИИ СИСТЕМ И
СИСТЕМНОГО АНАЛИЗА В ОПТИКЕ

Учебно-методическое пособие



Санкт-Петербург
2020

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

УНИВЕРСИТЕТ ИТМО

И.П. Гуров

**ОСНОВЫ ТЕОРИИ СИСТЕМ И
СИСТЕМНОГО АНАЛИЗА В ОПТИКЕ**

Учебно-методическое пособие

РЕКОМЕНДОВАНО К ИСПОЛЬЗОВАНИЮ В УНИВЕРСИТЕТЕ ИТМО
по направлению подготовки 12.03.03 Фотоника и оптоинформатика
в качестве учебно-методического пособия для реализации основных
профессиональных образовательных программ высшего образования
бакалавриата

 **УНИВЕРСИТЕТ ИТМО**

Санкт-Петербург

2020

Гуров И. П. Основы теории систем и системного анализа в оптике. Учебно-методическое пособие. – СПб: Университет ИТМО, 2020. – 110 с.

Рецензент: Тропченко Александр Ювенальевич, доктор технических наук, профессор

Рассмотрены основы системного анализа с привлечением теоретико-информационного подхода, принципы построения адаптивных оптических систем и интерферометрических систем, особенности систем биомедицинской диагностики, принципы и критерии информационной оптимизации оптико-цифровых систем при решении задач бесконтактного контроля объектов.

Пособие адресовано прежде всего студентам бакалавриата, обучающимся по направлению подготовки 12.03.03 "Фотоника и оптоинформатика", и может быть полезно магистрантам, аспирантам (в части материалов разделов 1 и 5) и специалистам в области оптики и фотоники.

Рекомендовано к печати Ученым советом факультета Прикладной оптики, протокол № 12 от 10 марта 2020 г.



Университет ИТМО – ведущий вуз России в области информационных и фотонных технологий, один из немногих российских вузов, получивших в 2009 году статус национального исследовательского университета. С 2013 года Университет ИТМО – участник программы повышения конкурентоспособности российских университетов среди ведущих мировых научно-образовательных центров, известной как проект «5 в 100». Цель Университета ИТМО – становление исследовательского университета мирового уровня, предпринимательского по типу, ориентированного на интернационализацию всех направлений деятельности.

© Университет ИТМО, 2020

© Гуров Игорь Петрович, 2020

Оглавление

Введение	4
Раздел 1. Основы системного анализа в оптике	6
1.1 Обобщенная структура систем в оптике	6
1.2 Основные понятия теории систем и системного анализа	8
1.3 Априорная информация и математическое моделирование объектов и систем	13
1.4 Идентификация систем	16
Список литературы	19
Раздел 2. Адаптивные системы	20
2.1 Принципы построения адаптивных систем	20
2.2 Адаптивные системы в оптике	24
2.3 Адаптивный сумматор и его применение в оптике	38
Список литературы	43
Раздел 3. Интерферометрические системы	45
3.1 Аппаратная функция интерферометрической системы	45
3.2 Структура интерферометрических систем	47
3.3 Характеристики интерферометрических систем и сигналов ..	53
3.4 Динамические интерферометрические системы	56
Список литературы	62
Раздел 4. Оптические системы биомедицинской диагностики	63
4.1 Системы бесконтактного контроля биологических объектов ..	63
4.2 Оптическая диагностика состояния кровеносной системы ...	66
4.3 Системы оптической когерентной микроскопии	78
Список литературы	87
Раздел 5. Информационные характеристики систем в оптике	89
5.1 Представление характеристик оптических полей на основе теоретико-информационного подхода	89
5.2 Информационная оптимизация оптико-цифровых систем ...	91
5.3 Информационный критерий качества систем бесконтактного контроля объектов	100
Список литературы	109

Введение

Изучение принципов формирования, преобразования, регистрации и анализа характеристик световых полей составляет основу подготовки специалистов в области фотоники и проводится при освоении учебных дисциплин «Оптическая физика», «Основы фотоники», «Основы оптоинформатики» и др.

Системный подход к описанию и анализу световых полей различной степени когерентности получил развитие в связи с созданием и широким использованием лазеров. В первой монографии [1] по данной проблематике обобщены результаты ставших классическими работ в области Фурье-оптики [2], системного рассмотрения задач статистической оптики [3], исследования оптических систем обработки информации [4] и др. Опубликованы многочисленные статьи, монографии и учебные пособия, в том числе отражающие аналогии в области оптических и радиосистем (см., например, [5]). При этом основное внимание уделено теории линейных систем в оптике применительно к описанию дифракции световых полей, интегральным преобразованиям, пространственной фильтрации.

В связи с развитием вычислительной техники проведено детальное рассмотрение методов обработки сигналов, зарегистрированных в оптических системах (см., например, [6]). Особенности дискретного представления и обработки сигналов изложены в учебном пособии [7].

Системный подход основывается на целостном рассмотрении проблемы от постановки задачи до получения и анализа результатов. При этом требуется выполнить синтез системы как совокупности взаимосвязанных элементов (подсистем), которые образуют структуру системы, на основе заданного критерия качества.

Настоящее учебное пособие ориентировано на формирование компетенций в области системного анализа в оптике. В первом разделе рассмотрены основные понятия теории систем и системного анализа, роль априорной информации и особенности математического моделирования объектов и систем. Второй раздел посвящен адаптивным системам, рассматриваются принципы построения адаптивных систем в оптике и особенности современных адаптивных оптических систем. В третьем разделе проводится рассмотрение интерферометрических систем и анализ их структуры при решении различных задач бесконтактного контроля объектов. Особое внимание уделяется особенностям динамических интерферометрических систем.

Четвертый раздел содержит сведения об оптических системах биомедицинской диагностики. Обсуждаются свойства исследуемых объектов и особенности построения систем оптической диагностики с некогерентным освещением и систем оптической когерентной томографии. В пятом разделе приведено описание методов определения информационных характеристик систем в оптике с позиций теоретико-информационного подхода с учетом свойств изучаемых объектов, получаемых данных и пропускной способности оптического и электронного каналов передачи информации.

Материалы учебного пособия ориентированы на формирование личностных и общепрофессиональных компетенций в использовании системного подхода и его применения к решению актуальных научно-технических задач в области оптики и фотоники.

Список литературы

1. Папулис А. Теория систем и преобразований в оптике. – М.: Мир, 1971.
2. Гудмен Дж. Введение в Фурье-оптику. – М.: Мир, 1970.
3. О'Нейл Э. Введение в статистическую оптику. – М.: Мир, 1966.
4. Shulman A.R. Optical data processing. – NY: Wiley, 1970.
5. Литвиненко О.Н. Основы радиооптики. – Киев: Техніка, 1974.
6. Васильев В.Н., Гуров И.П. Компьютерная обработка сигналов в приложении к интерферометрическим системам. – СПб: ВХВ-Санкт-Петербург, 1998.
7. Гуров И.П. Формирование и анализ сигналов в системах компьютерной фотоники. – СПб: Университет ИТМО, 2018.

1 Основы системного анализа в оптике

1.1 Обобщенная структура систем в оптике

Оптические системы широко используются при исследованиях и бесконтактном контроле различных объектов с высокой разрешающей способностью. Структура системы определяется её назначением. Назначение системы в оптике состоит в получении сведений об объектах и передаче их получателю (см. рисунок 1.1).



Рисунок 1.1. Блок-схема системы бесконтактного контроля объектов

Оптическая система регистрирует первичную информацию об объекте в форме отраженного (прошедшего) излучения. Характеристики оптической системы определяют разрешающую способность и другие основные параметры системы в целом. Фотоэлектрический преобразователь формирует сигналы в определенном спектральном диапазоне при заданном быстродействии. Блок сопряжения преобразует сигналы для передачи их в вычислительную систему для обработки по заданному алгоритму при помощи программно-алгоритмического обеспечения (ПАО).

В зависимости от вида источника излучения системы можно разделить на пассивные и активные. В пассивных системах используется естественное внешнее освещение, в активных – искусственные источники излучения.

Источники классифицируются как некогерентные и когерентные. Следует отметить, что даже естественный свет имеет небольшую длину когерентности порядка длины волны. Длина когерентности зависит от ширины спектра и средней длины волны.

Активные системы могут быть построены на основе использования источников когерентного и некогерентного излучения, например, в форме структурированного освещения с заданным пространственным распределением интенсивности, проецируемым на объект.

Современные фотоэлектрические преобразователи используют принципы дискретизации и квантования сигналов по уровню. При передаче информации дискретизация осуществляется во времени. Многоэлементные приемники излучения осуществляют дискретизацию в пространственной области. При этом имеет место также временная дискретизация при формировании отдельных видеок кадров.

Ввиду сложности и разнообразия задач бесконтактного контроля объектов целесообразно обратиться к теоретико-информационному подходу [1.1] и рассматривать этапы от формирования световых полей до их визуализации в информационных терминах. При автоматическом анализе и преобразовании световых полей этот подход оказывается эффективным и приобретает все большую популярность. Действительно, в методах автоматического анализа всегда используются некоторые представления, или способы описания объекта или системы, при этом к представлениям могут применяться информационные критерии качества, что позволяет описать систему и ее элементы с помощью единой системы понятий.

Обобщенная структура информационной системы иллюстрируется на рисунке 1.2, а. В информационной системе источник порождает сообщения, которые должны быть переданы через канал передателю. Во многих случаях сообщения не могут быть переданы непосредственно, и требуется кодирование сообщения для передачи по каналу с последующим декодированием. В оптико-информационной системе (рисунок 1.2, б) кодирование осуществляется при взаимодействии оптического излучения с объектом, и для декодирования используются специальные технические средства и алгоритмы, обеспечивающие извлечение информации для получателя.



а)



б)

Рисунок 1.2. Блок-схемы информационной системы (а) и передачи информации в оптико-информационной системе (б)

Пример: Определить среднее количество информации, содержащейся в одном пикселе изображения.

Построим гистограмму значений интенсивности изображения (рисунок 1.3), которая представляет собой оценку плотности вероятности $p(I)$ значений интенсивности I и из которой можно определить вероятность $P(I_i)$, $i = 1, \dots, m$, значений интенсивности в изображении. При этом энтропия, т.е. среднее количество информации, содержащейся в одном пикселе изображения, определяется известной формулой Шеннона

$$H = -\sum_{i=1}^m P(I_i) \log_2 P(I_i) \text{ [бит/пиксель]},$$

где m – количество уровней квантования.

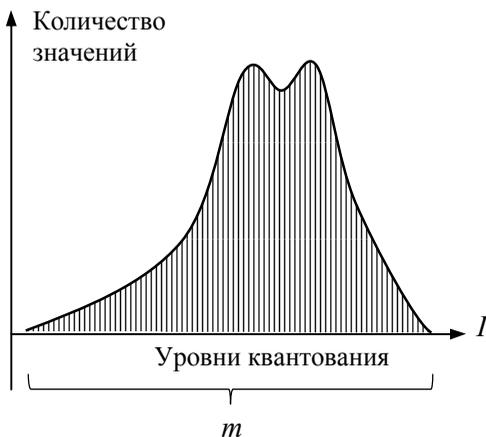


Рисунок 1.3. Гистограмма значений интенсивности

Особенности применения теоретико-информационного подхода для описания объекта как источника информации и оптико-информационной системы для извлечения и передачи информации рассматриваются в разделе 5 настоящего пособия.

1.2 Основные понятия теории систем и системного анализа

Системный подход основывается на рассмотрении системы, определяемой как *целостное* понятие. Система включает связанные между собой элементы или подсистемы. В схеме на рисунке 1.2, б система включает в себя исследуемый объект, который в свою очередь

может рассматриваться как подсистема: объект может иметь индивидуальную структуру, составные части и т. п. Свойства системы в целом отличаются от простой суммы свойств отдельных составных элементов и зависят от взаимосвязей элементов. Свойства создаваемой системы определяются ее назначением, т.е. конкретной решаемой задачей.

Системный подход составляет общую основу системного анализа [1.2], [1.3]. Совокупность взаимосвязанных элементов (подсистем) образует структуру системы. Система как совокупность элементов может иметь иерархическую структуру, когда элементы состоят из составляющих элементов более низкого уровня. Под *иерархией* понимают многоуровневое представление анализируемой системы с учетом соотношения между уровнями и элементами внутри уровня. Совокупность элементов и связей между ними определяют *морфологию* системы.

Описание системы возможно в форме внешнего и внутреннего описания. Внешнее описание системы, или описание «вход-выход» (см. рисунок 1.4), основывается на использовании преобразований, определяемых оператором системы $T\{s\}$, в которых принимаются в расчет входной сигнал и результат преобразования (например, преобразования Фурье). Внутреннее описание представляет собой параметрическую модель, т. е. отражает зависимость характеристик системы от параметров, знание которых не используется в явной форме при внешнем описании системы.

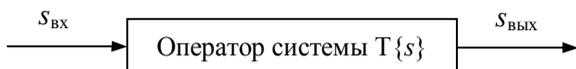


Рисунок 1.4. Система «вход–выход»

По структуре системы принято классифицировать прежде всего по количеству входов и выходов. Система с точечным источником излучения и точечным фотодетектором представляет собой систему *с одним входом и одним выходом*. Система с точечным источником и многоэлементным приемником излучения может рассматриваться как *система с одним входом и многими выходами*. Использование нескольких источников и приемников излучения интерпретируется как *система с многими входами и многими выходами*. Следует отметить, что структурное выделение входов и выходов определяется видом решаемой задачи и целью функционирования системы.

Важной структурной характеристикой системы является наличие (отсутствие) обратной связи между элементами системы. При наличии обратной связи осуществляется *функция управления* системой в целях оптимизации функционирования в соответствии с заданным критерием. Системы без обратной связи называют *разомкнутыми*.

Системы подразделяются на *линейные* и *нелинейные*. Линейные системы подчиняются принципу суперпозиции, выражаемому соотношением

$$T\{\alpha s_1 + \beta s_2\} = \alpha T\{s_1\} + \beta T\{s_2\}, \quad (1.1)$$

где s_1 и s_2 – входные воздействия, α и β – постоянные. Некогерентные системы линейны по *интенсивности* излучения.

Когерентные системы линейны по *напряженности поля*, но не по интенсивности, а именно, при сложении двух электромагнитных колебаний $E = E_1 + E_2$ значения интенсивности

$$I = |E|^2 = |E_1|^2 + |E_2|^2 + 2|E_1||E_2|\cos\varphi,$$

где φ – разность фаз колебаний. Два первых слагаемых представляют собой сумму интенсивностей. Третье слагаемое выражает эффект нелинейности системы по интенсивности. При некогерентном освещении значения фазы φ – случайная величина с равномерным распределением в интервале $(-\pi, \pi]$. При этом наблюдаемое среднее значение третьего слагаемого равно нулю.

По характеру функционирования системы подразделяют на детерминированные, подчиняющиеся функциональным закономерностям, и стохастические, описание которых основано на положениях теории вероятностей и теории случайных процессов.

Системные преобразования сигналов

В зависимости от вида и объема априорной информации о характеристиках полезного сигнала и помех можно выделить три основных подхода к синтезу систем [1.4].

При первом подходе предполагается детерминированный характер значений сигнала $s(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\theta}')$, где \mathbf{x} – вектор независимых переменных, $\boldsymbol{\theta}$ – вектор искомым детерминированных параметров, $\boldsymbol{\theta}'$ – вектор сопутствующих детерминированных параметров. Задача состоит в получении оценки $\hat{\boldsymbol{\theta}}$ вектора параметров. Операторное уравнение системы в этом случае имеет вид

$$\hat{\boldsymbol{\theta}} = T\{s(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\theta}')\}, \quad (1.2)$$

При этом задача сводится к синтезу сравнительно простого оператора системы T , реализующего обратное функциональное преобразование и инвариантного к вектору параметров θ' (см. рис. 1.5).

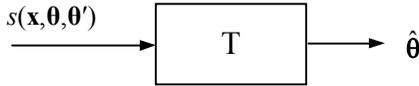


Рисунок 1.5. Системное преобразования с детерминированным оператором системы

Пример. Пусть на вход системы (рисунок 1.5) поступает комплексный сигнал $z(\theta, \theta') = \rho(a + jb)$, где $\theta = (a \ b)^T$, $\theta' = \rho$, и требуется определить фазу φ сигнала. Оператор системы осуществляет детерминированное функциональное преобразование в форме вычисления обратной функции $\varphi = \arctg(\text{Im } z / \text{Re } z)$, инвариантное к сопутствующему параметру $\theta' = \rho$.

Второй подход (см. рисунок 1.6) основывается на предположении о сигнале как реализации случайного процесса $\{s(\mathbf{x}, \theta, \theta')\}$ с известной априорной плотностью вероятности параметров $p(\theta, \theta')$. Необходимо с допустимой погрешностью получить оценку

$$\hat{\theta} = T\{s(\mathbf{x}, \theta, \theta') + n(\mathbf{x})\}, \quad (1.3)$$

при обработке зарегистрированной реализации случайного процесса на фоне шума наблюдений $n(\mathbf{x})$ с известными статистическими характеристиками.

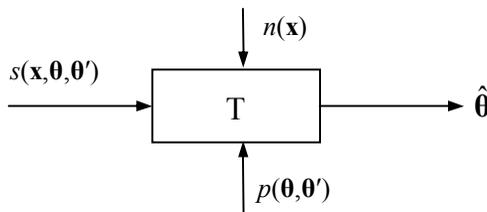


Рисунок 1.6. Системное преобразование со стохастическими параметрами

Синтез требуемого оператора системы в (1.3) может быть выполнен на основе достаточно общих критериев, например, критерия

минимума среднего риска, правила максимума функционала правдоподобия или других критериев, асимптотически приближающихся по характеристикам к критерию максимума апостериорной плотности вероятности.

Третий подход (см. рисунок 1.7) применяется в тех случаях, когда априорно не известны даже статистические характеристики сигнала и помехи. При этом необходимо использовать принципы *адаптации* систем.

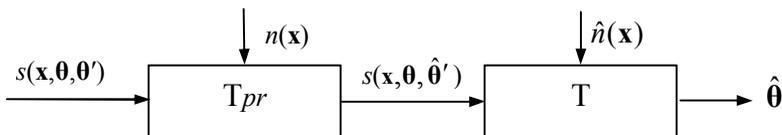


Рисунок 1.7. Преобразования в адаптивной системе

Согласно этому подходу, в процессе обработки по зарегистрированным данным должны быть найдены параметры, характеризующие реализацию $s(\mathbf{x}, \boldsymbol{\theta})$, причем параметры заменяются их оценочными значениями. После получения оценочных значений задача сводится к предыдущей, а именно

$$\hat{\boldsymbol{\theta}} = T\{T_{pr}[s(\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\theta}') + n(\mathbf{x})]\} = T\{s(\mathbf{x}, \boldsymbol{\theta}, \hat{\boldsymbol{\theta}}') + \hat{n}(\mathbf{x})\}, \quad (1.4)$$

где $\hat{\boldsymbol{\theta}}'$ и $\hat{n}(\mathbf{x})$ – оценки вектора сопутствующих параметров сигнала и шума, соответственно, получаемые на стадии предварительной обработки, определяемой оператором $T_{pr}\{\cdot\}$.

Во всех случаях достигаемая точность и сложность обработки в значительной мере определяются выбором адекватных моделей сигналов и помех в (1.2) – (1.4) на основе рассмотрения физических принципов и особенностей построения систем в оптике и фотонике.

Синтез оператора системы T может осуществляться в частотной области или в области независимых переменных.

Различают дискретные и непрерывные сигналы и системы. В современных системах оперируют дискретными сигналами, при этом непрерывные сигналы подвергают дискретизации по независимой переменной и квантованию по уровню.

1.3 Априорная информация и математическое моделирование объектов и систем

Реальный объект характеризуется многообразными свойствами. Изучение объекта возможно на основе использования ограниченной эмпирической (полученной путем наблюдения) информации, которая необходима для достижения поставленной цели. При этом создается упрощенное подобие объекта – модель.

Моделирование заключается в замене *реальной системы* (объекта, процесса) *моделью*, которая находится с ней (с ними) в некотором соответствии и способна воспроизводить интересующие исследователя свойства. В выборе свойств системы, которые должны быть включены в модель, решающую роль играет неформальный анализ решаемой задачи на основе учитываемой исследователем априорной информации. Причиной применения моделей является замещение непосредственного наблюдения функционирования систем их моделированием, что позволяет эффективно решать конкретные задачи исследования.

Модель как упрощенное описание объекта или системы обеспечивает выполнение классификации, т.е. выделение сущности, и отбрасывание свойств, которые не существенны с учетом поставленной цели. Математическая модель позволяет проводить исследования вариантов развития исследуемого процесса и функционирования системы и обеспечивает возможность прогнозирования.

Возможности моделирования определяются тем, что модель включает не только информацию о конкретном объекте или системе, но и информацию теоретическую, т.е. представление, существующее априори относительно решаемой задачи. Теория опирается на предыдущие наблюдения и включает накопленный опыт.

Компьютерное моделирование и вычислительные эксперименты

С помощью моделей, реализованных программными средствами, можно изучать новые явления, решать задачи анализа и проектирования сложных систем, осуществлять выбор наилучших вариантов решений, выполнять анализ и прогнозирование поведения технических систем и решать множество других задач.

Компьютерное моделирование заключается в проведении серии вычислительных экспериментов, целью которых является анализ, интерпретация и сопоставление результатов моделирования с реальным поведением изучаемого объекта или системы.

Правила и этапы моделирования

Можно выделить следующие правила моделирования при исследовании технических систем.

Первое правило моделирования заключается в наличии различных компетенций для разработки и анализа моделей (например, компетенции в оптике и фотонике, математике, программировании).

Второе правило состоит в том, что разработчикам моделей нужно знать как общие законы функционирования технических систем, так и частные соотношения в предметной области (оптике и фотонике), которые обычно представляются математическими соотношениями.

Третье правило заключается в том, что объективная сложность технических систем исключает возможность их всестороннего изучения с помощью только одной какой-либо модели. Это означает, что обобщенная модель обычно включает несколько разнотипных взаимосвязанных подмоделей.

Процесс построения обобщенной модели состоит из следующих основных этапов.

1. Содержательная постановка задачи – исследование системы (объекта, процесса) и формулировка технического задания на разработку модели.
2. Концептуальная постановка задачи – целеполагание моделирования системы (объекта, процесса).
3. Проверка корректности полученной модели и ее предварительный качественный анализ.
4. Математическая постановка задачи и обоснование метода ее решения.
5. Подбор или разработка вычислительного алгоритма решения задачи.
6. Исследование модели и проверка ее адекватности.
7. Результат процесса моделирования технической системы – создание функционирующей адекватной модели, удовлетворяющей заданным критериям (количественной мере качества модели).

Качество математических моделей

При проведении исследований требуется оценивать качество моделей и проводить сравнение моделей между собой.

Идентификация (отождествление) математических моделей проводится с точки зрения их сравнения с данными с целью получения моделей *обоснованного* качества.

Идентификация *в широком смысле* необходима в случаях, когда априорная информация отсутствует или является очень бедной, поэтому приходится решать задачи выбора структуры системы и задания класса моделей, определения степени линейности и стационарности процесса, определения вида влияния входных переменных на выходные, выбора информативных переменных и т.д. Методы идентификации в широком смысле активно разрабатываются в последние годы при использовании возрастающих вычислительных возможностей компьютерной техники.

Идентификация *в узком смысле* состоит в *оценивании параметров* модели по данным наблюдения. Структура системы предполагается известной и класс моделей – априорно заданным.

Качество модели наиболее последовательно и строго оценивается методами теории вероятностей и случайных процессов. Основная методика состоит в следующем.

1. Данные представляются суммой регулярной составляющей процесса и случайной помехи (шума) с некоторой функцией распределения вероятности, а именно

$$\xi = s(\mathbf{x}, \boldsymbol{\theta}) + \mathbf{n}, \quad \mathbf{n} \in \mathbf{F}(\boldsymbol{\gamma}), \quad (1.5)$$

где ξ – вектор измеряемых выходных переменных (значений сигнала), $s(\mathbf{x}, \boldsymbol{\theta})$ – векторная функция, представляющая идеализированную модель полезных данных, \mathbf{n} – реализация случайного вектора шума, $\mathbf{F}(\boldsymbol{\gamma})$ – векторная функция многомерного распределения шума с параметрами $\boldsymbol{\gamma}$, \mathbf{x} – вектор входных (независимых) переменных, $\boldsymbol{\theta}$ – вектор параметров модели. Во многих случаях в качестве независимой переменной принимают текущее время t . При моделировании систем в оптике независимые переменные представляют пространственные координаты.

2. На основе предварительных исследований устанавливается вид функции $\mathbf{F}(\boldsymbol{\gamma})$ и, возможно, вектор параметров шума $\boldsymbol{\gamma}$.

3. Исходя из требований несмещенности, эффективности, состоятельности формируется *критерий идентификации* Q модели, являющейся для *конкретной* выборки опытных данных функцией параметров модели $\boldsymbol{\theta}$.

4. Минимизация (максимизация) критерия Q приводит к получению оценки $\hat{\boldsymbol{\theta}}$ вектора параметров, наиболее приближающей расчетные значения модели к данным.

5. Проверяется статистическая гипотеза о соответствии полученных экспериментальных данных *отклонений* модели от

выбранной функции $F(\gamma)$. Если эта гипотеза не отвергается, то говорят об адекватности модели.

При использовании этой методики нужно учитывать следующие особенности. При выборе функции $F(\gamma)$ в (1.5) неизвестна полезная составляющая, поскольку обычно полезный сигнал и шум отдельно не наблюдаемы. Следует иметь в виду также, что проводится замена вопроса об адекватности модели вопросом об адекватности модели эксперимента (модель может быть неправильной, но это незаметно вследствие ошибок или искажений в реальных условиях эксперимента). Если статистическая гипотеза по п. 5 отвергается, то не ясно, систематическая погрешность больше, чем влияние шума, или апостериорная функция распределения разности «модель-эксперимент» не соответствует принятому допущению для $F(\gamma)$.

Можно сделать следующий вывод: для преодоления упомянутых сложностей целесообразно рассматривать тонкую структуру разности расчетных и фактических значений, так как обычно постулируется, что именно эта разность является мерой сходства оригинала и модели.

1.4 Идентификация систем

Идентификация означает отождествление модели с реальным объектом. Идентификация систем в общем случае состоит в определении их структуры и параметров по наблюдаемым данным – входному и выходному сигналам. Идентификация осуществляется при помощи настраиваемой математической модели (см. рисунок 1.8), параметры которой могут изменяться.



Рисунок 1.8. Идентификация на основе настраиваемой модели

С помощью модели получают оценку выходных данных $\hat{y}(x)$ и сравнивают с выходными данными $y(x)$ реальной системы. Модель подстраивается так, чтобы минимизировать разность между данными модели и данными системы. В результате полученная модель описывает систему.

По наблюдаемым входным воздействиям и выходным величинам подбираются параметры модели, обеспечивающие выполнение

заданного критерия, характеризующего качество идентификации. Изменение параметров модели осуществляется в соответствии с алгоритмом идентификации.

Рассмотрим процедуру идентификации более подробно, следуя методике, подробно рассматриваемой в [1.5]. На рисунке 1.9 показана функциональная схема идентификации исследуемой системы. На каждом дискретном отсчете $k = 1, 2, \dots$ на вход системы и настраиваемой модели поступает сигнал $s(k)$. Учитывается влияние помехи $n(k)$. Выходная величина $y(k)$ определяется значениями $s(k)$, $n(k)$ и априорно не известного вектора параметров θ^* . Выход $\hat{y}(k)$ настраиваемой модели зависит от настраиваемых параметров θ и пересчитывается согласно алгоритму обработки всех доступных наблюдаемых данных $\mathbf{z}(k)$. Набор этих данных зависит от конкретного вида задачи идентификации.

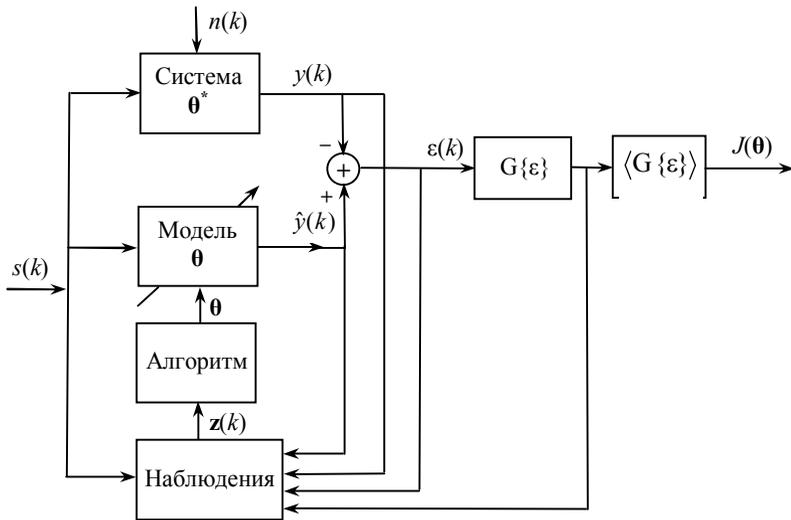


Рисунок 1.9. Функциональная схема выполнения идентификации

Разность значений $\hat{y}(k)$ и $y(k)$ образует невязку

$$\varepsilon[\mathbf{z}(k), \boldsymbol{\theta}] = \hat{y}(k) - y(k), \quad (1.6)$$

которая поступает на вход функционального преобразователя с оператором $G\{\varepsilon\}$. Предполагается, что система функционирует в

стационарном режиме, т.е. вероятностные характеристики $\hat{y}(k)$, $y(k)$ и, следовательно, $\mathbf{z}(k)$ не зависят от отсчета k . Такой режим принято называть *режимом нормальной работы*.

Соответствие настраиваемой модели оценивается критерием качества идентификации

$$J(\boldsymbol{\theta}) = \langle G \{ \varepsilon[\mathbf{z}(k), \boldsymbol{\theta}] \} \rangle,$$

где угловые скобки означают оценку математического ожидания. Оператор G в данном случае интерпретируется в виде функции потерь, определяющей значимость размера невязки и характер учета невязки в данных $\mathbf{z}(k)$. В этом смысле критерий качества $J(\boldsymbol{\theta})$ выражает *средние потери*. Чем меньше средние потери, тем выше качество идентификации. Улучшение качества идентификации достигается надлежащим выбором настраиваемой модели и изменением ее параметров, которое осуществляется *алгоритмом идентификации*. Алгоритм идентификации определяется функцией потерь и структурой настраиваемой модели.

Таким образом, для решения задачи идентификации необходимо:

1. Очертить класс систем, т.е. выполнить идентификацию в широком смысле.
2. Выбрать адекватную настраиваемую модель, параметры которой можно изменять.
3. Выбрать критерий качества идентификации, т.е. допустимые средние потери, которые характеризуют различие между $\hat{y}(k)$ и $y(k)$.
4. Сформировать алгоритм идентификации, который, используя доступные для наблюдения данные $\mathbf{z}(k)$, изменяет параметры настраиваемой модели так, чтобы средние потери с ростом k достигали минимума.

Контрольные вопросы

1. Объясните аналогию информационной системы общего вида и оптико-информационной системы. В чем состоит сходство и отличие этих двух видов систем?
2. Сформулируйте основные понятия теории систем и системного анализа. В чем состоят особенности внешнего и внутреннего описания системы?
3. Приведите примеры классификации систем в оптике и фотонике по количеству входов и выходов.

4. В чем выражается свойство линейности системы? Приведите примеры линейных и нелинейных систем в оптике.
5. В чем состоят особенности представления реальных объектов и систем математической моделью? Перечислите основные правила и этапы моделирования.
6. В чем заключаются процедуры идентификации в широком смысле и в узком смысле?
7. Объясните процедуры оценивания качества модели на основе теории вероятностей и случайных процессов.
8. Раскройте процедуру идентификации исследуемой системы с использованием функциональной схемы идентификации.

Список литературы

- 1.1. Васильев В. Н., Гуров И. П., Потапов А. С. Современная видеоинформатика: проблемы и перспективы // Оптический журнал. 2012. Т. 79. №11. С. 5–15.
- 1.2. Перегудов Ф.П., Тарасенко Ф.П. Введение в системный анализ. – М.: Высшая школа, 1989.
- 1.3. Качала В.В. Основы теории систем и системного анализа. Учебное пособие для вузов. – М.: Горячая линия – Телеком, 2016.
- 1.4. Гуров И.П. Формирование и анализ сигналов в системах компьютерной фотоники. – СПб: Университет ИТМО, 2018.
- 1.5. Ципкин Я.З. Информационная теория идентификации. – М.: Наука. Физматлит, 1995.

2 Адаптивные системы

2.1 Принципы построения адаптивных систем

В адаптивных системах структура и/или параметры системы подстраиваются так, чтобы улучшить функционирование в соответствии с заданным критерием с учетом взаимодействия с внешней средой (изменениями характеристик среды).

Адаптивные системы могут иметь структуру без обратной связи и с обратной связью [2.1], как иллюстрируется схемами на рисунке 2.1.

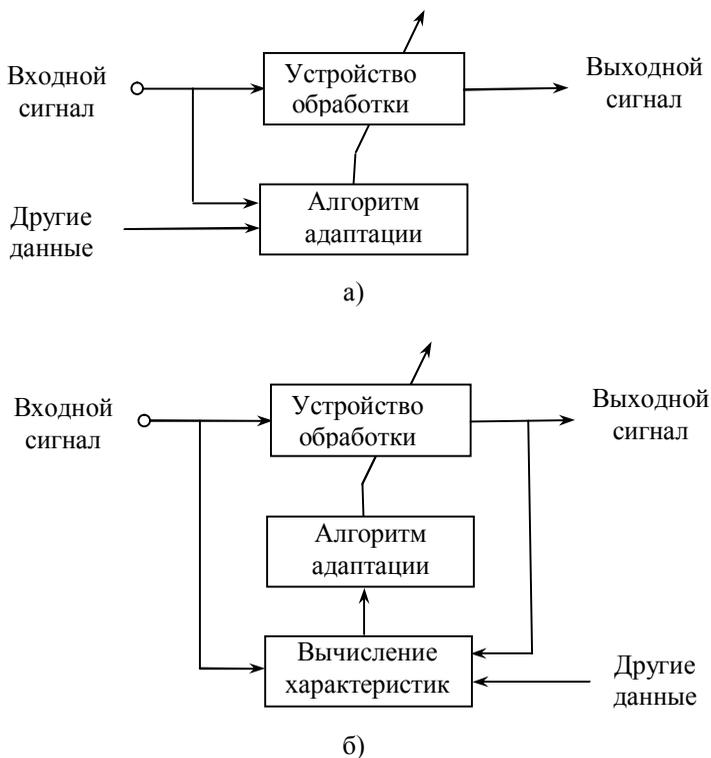


Рисунок 2.1. Структура адаптивных систем без обратной связи (а) и с обратной связью (б)

При отсутствии обратной связи (рисунок 2.1, а) система адаптируется по входному сигналу и другим данным (например, об условиях внешней среды) таким образом, чтобы обеспечить функционирование системы в заданном диапазоне изменения параметров и выходного сигнала на основании ранее изученного функционирования системы при различных входных воздействиях.

В системах с обратной связью (рисунок 2.1, б) адаптация осуществляется с использованием выходного сигнала и других данных при формировании характеристик алгоритма адаптации на основании этих данных.

Преимущества адаптивных систем состоят в возможности работы при неточно известных закономерностях, когда отсутствует точное аналитическое описание функционирования при различных внешних условиях. При этом системы устойчивы к влиянию ошибок (помех).

К недостаткам адаптивных систем следует отнести отсутствие единственного оптимума при подстройке параметров и неустойчивость (расходимость) при реальных условиях, значительно отличающихся от принятых в расчет при синтезе системы.

Общие свойства адаптивных систем:

1. Способность к самонастройке при изменениях внешних условий (в том числе нестационарных) и требований к системе.
2. Способность обучения на данных (их изменениях) заданного вида в целях принятия наилучшего режима функционирования.
3. Возможность автоматического синтеза оптимальной структуры и наилучших значений параметров на основе обучения.
4. Не требуется тщательная проработка метода синтеза (как в неадаптивных системах) ввиду возможности самоорганизации.
5. Возможность экстраполяции модели для функционирования в новых условиях после обучения на конечном и часто небольшом числе обучающих ситуаций.
6. Возможность в некоторой степени восстанавливаться, адаптироваться к определенным внутренним дефектам.
7. Возможность расширения области функционирования системы, когда параметры входных сигналов неизвестны или изменяются во времени.
8. Реализация нелинейных алгоритмов с нелинейно изменяющимися во времени параметрами.

Адаптивные системы нелинейны ввиду зависимости их параметров от входных сигналов, имеют изменяющиеся параметры и удовлетворяют одному или нескольким из вышеперечисленных свойств.

Примеры применения адаптивных систем с обратной связью

Принцип работы адаптивной системы с обратной связью иллюстрируется схемой на рисунке 2.2. Входной сигнал s преобразуется в выходной сигнал y при помощи устройства обработки, параметры которого изменяются с использованием алгоритма адаптации. На вход алгоритма поступают данные о сигнале ошибки ε , формируемом как разность выходного сигнала устройства обработки и требуемого (опорного) сигнала d .

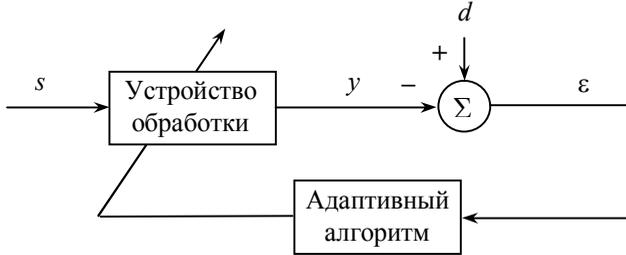


Рисунок 2.2. Общая структурная схема адаптивной системы с обратной связью

На основе минимизации сигнала ошибки ε по заданному критерию алгоритм позволяет изменять характеристики системы.

На рисунке 2.3 приведена в качестве примера схема адаптивного устройства предсказания значений сигнала. При линейном предсказании осуществляется процедура, позволяющая по некоторому набору предыдущих значений предсказать последующий отсчет сигнала в анализируемой последовательности отсчетов $s(k)$.

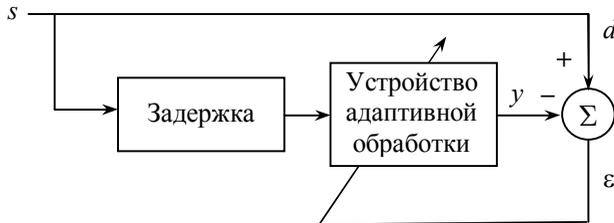


Рисунок 2.3. Схема адаптивного устройства предсказания

Полезным является входной сигнал s , поступающий после задержки в устройство адаптивной обработки. При этом выходной

сигнал y компенсирует опорный сигнал d с минимизацией сигнала ошибки.

Оценка при предсказании формируется в виде

$$\hat{s}(k) = \sum_{m=1}^M w_m s(k-m). \quad (2.1)$$

При этом задача состоит в том, чтобы найти набор весовых коэффициентов $\{w_m\}$, при котором минимизируется дисперсия ошибки $\varepsilon^2 = \langle [\hat{s}(k) - s(k)]^2 \rangle \rightarrow \min$.

Отметим, что если на вход поступает смесь полезного сигнала и шума, то задержка позволяет отделить коррелированную (предсказуемую) составляющую полезного сигнала от некоррелированной (непредсказуемой) составляющей шума.

В ряде случаев требуется выполнить компенсацию влияния внешних факторов с формированием характеристик системы, обратных влиянию этих факторов. Структура системы адаптивной компенсации иллюстрируется на рисунке 2.4.

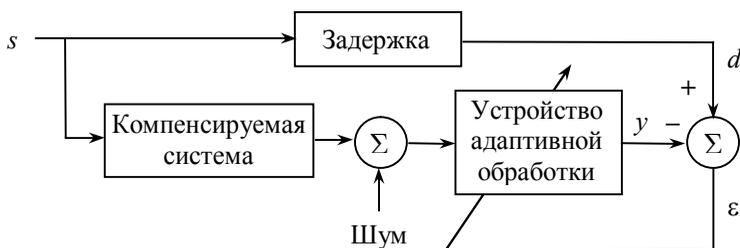


Рисунок 2.4. Схема адаптивной системы компенсации

В адаптивной системе формируется характеристика, обратная влиянию компенсируемой системы (например, в случае вносимых ей искажений) и дополнительно осуществляется компенсация влияния шума. Блок задержки формирует опорный сигнал d в виде смещенного во времени входного сигнала с учетом влияния компенсируемой системы.

Известны другие виды и применения адаптивных систем, характеристики которых детально рассмотрены в [2.1]. Пример реализации адаптивной системы в интерферометрии приведен в разделе 3.4.

2.2 Адаптивные системы в оптике

Методы адаптивной оптики используются для компенсации внешних возмущений оптического поля и коррекции неидеальности характеристик оптических компонентов.

Примеры:

- распространение света через турбулентную атмосферу;
 - биомедицинская оптика (распространение излучения в неоднородной среде с изменяющимися параметрами);
 - компенсация деформаций оптико-механических компонентов при изменяющихся механических нагрузках и внешних воздействиях;
 - компенсация тепловых эффектов при воздействии мощного лазерного излучения
- и т.д.

Оптические свойства атмосферы, принимаемые в расчет при создании адаптивных оптических систем, подробно рассмотрены в [2.2] и в учебном пособии [2.3]. Особенности влияния деформаций оптических элементов и тепловых эффектов обсуждаются в [2.4]. Применения адаптивных оптических систем в современной микроскопии и биомедицинской оптике рассмотрены в [2.5].

Обратимся к основным особенностям применения адаптивных оптических систем.

Напряженность поля электромагнитной волны определяется комплексной амплитудой

$$E(\mathbf{p}) = A(\mathbf{p}) \exp[j\Phi(\mathbf{p})], \quad (2.2)$$

где \mathbf{p} – вектор, обозначающий пространственные координаты (x, y) точек волнового фронта с амплитудой волны $A(\mathbf{p})$ и фазой $\Phi(\mathbf{p})$.

Волновой фронт – геометрическое место точек, в которых фаза электромагнитной волны постоянна. Волновой вектор перпендикулярен волновому фронту в каждой точке.

На рисунке 2.5 иллюстрируются волновой фронт точечного источника (рисунк 2.5, а) и волновой фронт плоской волны (рисунк 2.5, б). Наличие неоднородностей показателя преломления искажает волновой фронт плоской волны (рисунк 2.5, в). В отличие от плоского волнового фронта искажения волнового фронта не позволяют сфокусировать излучение в одной точке (рисунк 2.5, д).

При освещении объекта излучением с комплексной амплитудой

$$E_0(\mathbf{p}_1) = A_0(\mathbf{p}_1) \exp[j\Phi_0(\mathbf{p}_1)] \quad (2.3)$$

в точках ρ_1 плоскости объекта $z = 0$ электрическое поле в точках ρ_2 в плоскости наблюдения z определяется на основе принципа Гюйгенса-Френеля суперпозицией вторичных сферических волн, исходящих от каждой точки объекта, и выражается в форме

$$E(z, \rho_2) = \frac{k}{j2\pi z} \int_{-L}^L E_0(\rho_1) \exp\left[\frac{jk}{2z}(\rho_2 - \rho_1)^2\right] d\rho_1, \quad (2.4)$$

где $k = 2\pi/\lambda$ – волновое число, $2L$ – протяженность объекта.

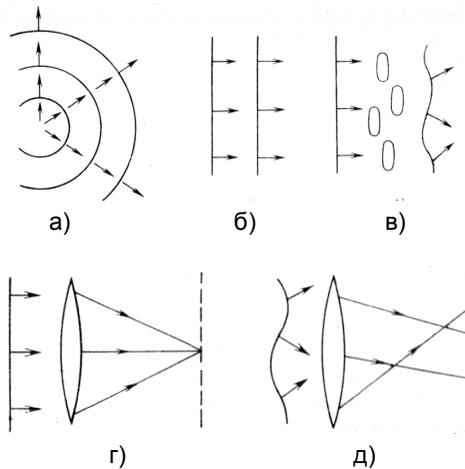


Рисунок 2.5. Искажения волнового фронта и их проявления [2.4]

Выражение (2.4) можно записать в более компактной форме

$$E(z, \rho_2) = \int_{-L}^L E_0(\rho_1) H_0(\rho_2 - \rho_1) d\rho_1, \quad (2.5)$$

где $H_0(\rho_2 - \rho_1)$ – ядро интегрального преобразования (функция Грина), описывающее распространение сферической волны в свободном пространстве (вакууме),

$$H_0(\rho_2 - \rho_1) = \frac{k}{j2\pi z} \exp\left[\frac{jk}{2z}(\rho_2 - \rho_1)^2\right]. \quad (2.6)$$

При распространении излучения через случайно неоднородную среду функция (2.6) испытывает возмущения фазы $\Psi(\rho_2, \rho_1)$ вследствие чего выражение (2.5) принимает вид

$$E(z, \rho_2) = \int_{-L}^L E_0(\rho_1) H(\rho_2, \rho_1) d\rho_1, \quad (2.7)$$

где

$$H(\rho_2, \rho_1) = \frac{k}{j2\pi z} \exp\left[\frac{jk}{2z}(\rho_2 - \rho_1)^2 + j\Psi(\rho_2, \rho_1)\right]. \quad (2.8)$$

Выражение (2.8), очевидно, можно записать в форме произведения

$$H(\rho_2, \rho_1) = H_0(\rho_2 - \rho_1) h(\rho_2, \rho_1), \quad (2.9)$$

где $h(\rho_2, \rho_1) = \exp[j\Psi(\rho_2, \rho_1)]$.

Из (2.5), (2.7) с учетом (2.9) следует, что возмущения $\Psi(\rho_2, \rho_1)$ можно скомпенсировать умножением (2.8) на комплексно-сопряженный множитель $h^*(\rho_2, \rho_1) = \exp[-j\Psi(\rho_2, \rho_1)]$.

Другой возможный подход состоит в том, чтобы ввести предсказания $E'_0(\rho_1)$ поля (2.3), освещающего объект, таким

$$\begin{aligned} \text{образом, чтобы в (2.7) выполнялось условие } & \int_{-L}^L E'_0(\rho_1) H(\rho_2, \rho_1) d\rho_1 \approx \\ & \approx \int_{-L}^L E_0(\rho_1) H_0(\rho_2 - \rho_1) d\rho_1. \end{aligned}$$

Заметим, что в обоих случаях для компенсации нужно иметь информацию о возмущениях волнового фронта. Кроме этого, следует учитывать, что проведенное выше аналитическое рассмотрение усложняется для случаев использования немонохроматического излучения. Для активных оптических систем существенное значение имеет степень пространственно-временной когерентности излучения.

Отметим также, что искажения волнового фронта могут быть связаны как с турбулентностью атмосферы или неоднородностью объекта на пути распространения излучения, так и с влиянием aberrаций, обусловленных неидеальностью формы элементов оптической системы и погрешностями при юстировке, которые могут быть компенсированы в адаптивной оптической системе.

Искажения волнового фронта могут изменяться во времени вследствие влияния различных факторов, отмеченных на рисунке 2.6 (см. также [2.2] – [2.4]). На рисунке приведены типичные диапазоны временных частот (ν , Гц) и пространственных частот (f_ρ , м^{-1}) для различных внешних факторов.

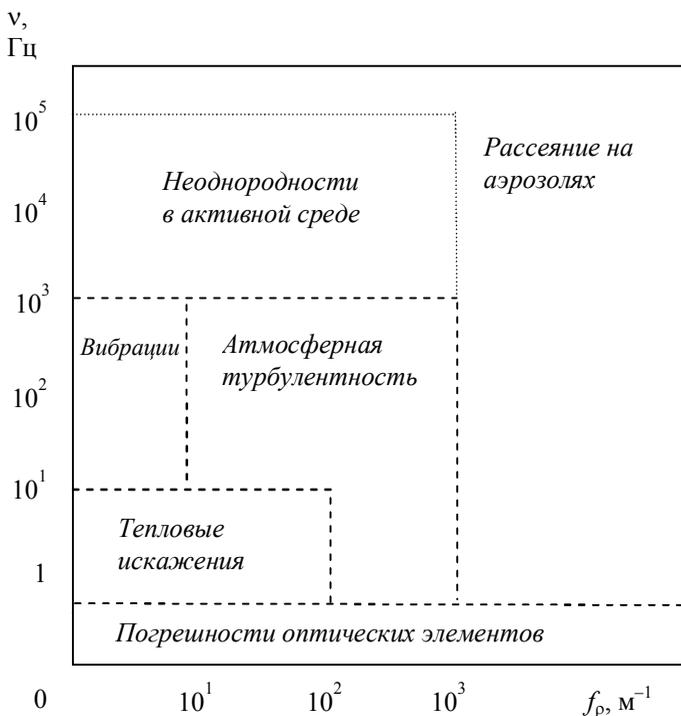


Рисунок 2.6. Распределение типичных видов искажений по пространственно-временным частотам [2.4]

Используются два варианта адаптивных оптических систем: с выходящей волной, когда корректируется волновой фронт и/или распределение интенсивности источника излучения, и системы с принимаемой волной, в которых корректируется световое поле, распространяющееся от объекта.

Рассмотрим схемы адаптивных оптических систем на основе материалов, приведенных в [2.4].

На рисунке 2.7 показана схема с корректировкой волнового фронта выходящего излучения.

Излучение от источника 1 поступает на устройство управления волновым фронтом 4. Управляющие сигналы формируются при помощи датчика волнового фронта 2 с дополнительным опорным зеркалом и устройства обработки 3 данных датчика.

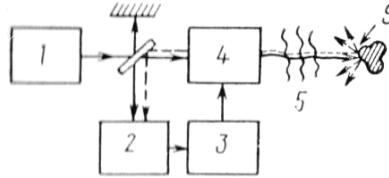


Рисунок 2.7. Схема адаптивной оптической системы с выходящей волной

Схема адаптивной системы с корректировкой принимаемой световой волны показана на рисунке 2.8.

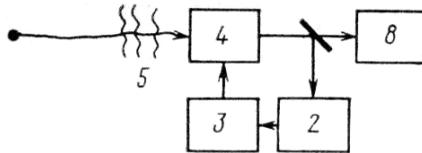


Рисунок 2.8. Схема адаптивной оптической системы с принимаемой волной

Излучение от объекта с возмущениями 5 волнового фронта поступает на устройство управления волновым фронтом 4. Управляющие сигналы формируются при помощи датчика волнового фронта 2 и устройства обработки данных 3 датчика. Скорректированный волновой фронт регистрируется фотодетектором 8.

Как отмечено выше, если изменения фазы волны обусловлены влиянием внешних возмущений $\Psi(\rho_2, \rho_1)$, то умножение (2.9) на комплексно сопряженный множитель $\exp[-j\Psi(\rho_2, \rho_1)]$ позволяет скомпенсировать эти искажения. Такие системы носят название адаптивных систем с *фазовым сопряжением*.

Результаты коррекции плоского и сферического волнового фронта методом фазового сопряжения иллюстрируются на рисунке 2.9, а, б соответственно.

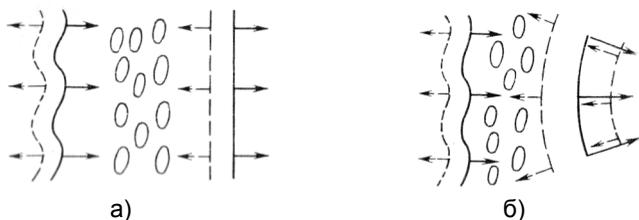


Рисунок 2.9. Коррекция волнового фронта методом фазового сопряжения: плоский волновой фронт (а) и сферический волновой фронт (б)

На рисунке 2.10 представлен пример изображений объекта до и после коррекции волнового фронта. На рисунке 2.10, б ясно заметны детали структуры объекта, которые не разрешаются без коррекции (рисунок 2.10, а).

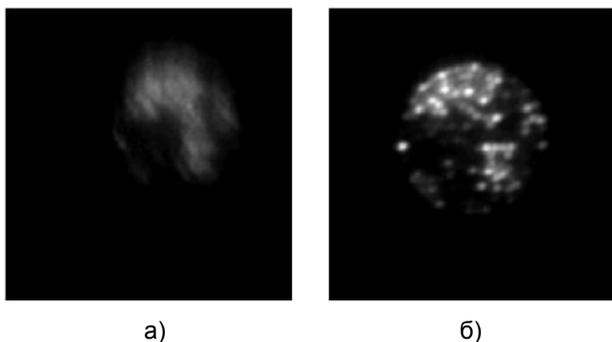


Рисунок 2.10. Изображение до и после коррекции

В адаптивных системах с фазовым сопряжением используется датчик волнового фронта, который в некоторых случаях представляет собой технически сложное устройство (например, интерферометр) с высокими требованиями к точности функционирования для обеспечения коррекции фазы с погрешностью в доли фазового цикла (т.е. в доли длины волны излучения).

Более простое техническое решение состоит в использовании принципа *апертурного зондирования*, который заключается во внесении в волновой фронт задаваемых пробных возмущений, которые преобразуются в отклонения регистрируемого сигнала. В результате пробных возмущений в системах с выходящей волной

можно определить условия освещения объекта (форму зондирующего волнового фронта), при которых достигается наилучшая фокусировка излучения на объекте.

На рисунке 2.11 представлена схема устройства апертурного зондирования.

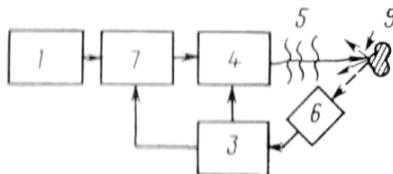


Рисунок 2.11. Схема системы апертурного зондирования с выходящей волной

Излучение от источника 1 поступает в систему апертурного зондирования 7, на выходе которой установлено устройство управления 4 волновым фронтом. Блоки 4 и 7 управляются от устройства 3, анализирующего сигналы приемника излучения 6, пропорциональные интенсивности принятого излучения. Можно показать [2.3], что интенсивность максимальна при точной фокусировке зондирующего излучения на объекте.

Метод апертурного зондирования можно реализовать также в режиме с принимаемой волной, когда корректируется световое поле, распространяющееся от объекта (см. рисунок 2.12).

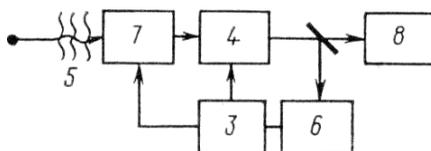


Рисунок 2.12. Схема системы апертурного зондирования с принимаемой волной

Излучение от объекта с возмущениями 5 волнового фронта поступает в систему апертурного зондирования 7, содержащую устройство управления волновым фронтом 4. Как и в случае рассмотренной выше системы с выходящей волной, блоки 4 и 7 управляются от устройства 3, анализирующего сигналы фотодетектора 6 по критерию максимума интенсивности

регистрируемого излучения. Скорректированный волновой фронт регистрируется приемником излучения 8.

Критерии качества в системах адаптивной оптики

В адаптивных системах фазового сопряжения используются следующие основные критерии.

Дисперсия отклонений волнового фронта определяется выражением

$$\sigma^2 = \frac{1}{\Omega} \iint_{\Omega} [\hat{\Phi}(\rho) - \Phi(\rho)]^2 d^2\rho \rightarrow \min, \quad (2.10)$$

где $\hat{\Phi}(\rho)$ – сформированный волновой фронт, $\Phi(\rho)$ – расчетный волновой фронт, Ω – площадь апертуры (в последующих формулах нормировку на площадь апертуры опускаем для упрощения записи).

Амплитудный критерий

$$J_a = \iint A(\rho) \hat{A}(\rho) d^2\rho \rightarrow \max \quad (2.11)$$

характеризует соответствие излученной $A(\rho)$ и отраженной $\hat{A}(\rho)$ волн в смысле их совпадения (перекрытия) в пределах апертуры.

В системах апертурного зондирования применяются следующие критерии.

Функционал фокусировки

$$J_f = \iint M(\rho) |A(\rho)|^2 d^2\rho \rightarrow \max \quad (2.12)$$

характеризует интенсивность излучения, попадающего в апертуру $M(\rho)$. При $M(\rho)$ в форме дельта-функции, т.е. $M(\rho) = \delta(\rho)$, этот критерий выражает нормированную осевую интенсивность (которая носит названия числа Штреля).

Эффективная ширина распределения интенсивности

$$J_{\omega} = \iint |\rho|^2 |A(\rho)|^2 d^2\rho \rightarrow \min \quad (2.13)$$

учитывает ширину пучка, в которой содержится определенная доля интенсивности излучения (возрастание ρ к краям влияет в пределах апертуры).

Линейная комбинация критериев

$$J_L = J_f - cJ_{\omega} \rightarrow \max \quad (2.14)$$

подбором коэффициента c позволяет получить фокусировку заданной доли излучения в минимальной площади.

Функционал пространственного распределения интенсивности

$$J_p = \iint [\hat{I}(\rho) - I(\rho)]^p d^2\rho \rightarrow \min, \quad (2.15)$$

где $p = 2, 4, 6, \dots$, выражает отличие полученного изображения с распределением интенсивности $\hat{I}(\rho)$ от известного (идеального) изображения $I(\rho)$.

Функционал резкости изображения

$$J_d = \iint I^n(\rho) d^2\rho \rightarrow \max, \quad (2.16)$$

где $n \geq 2$, обеспечивает улучшение условий наблюдения при формировании изображений.

Устройства и элементы адаптивных оптических систем

Адаптивная оптическая система обычно имеет в своем составе оптическое приемное и/или передающее устройство, а также воздействующий на волновой фронт активный элемент, датчик волнового фронта и ветвь обратной связи. Приемная и передающая апертуры могут быть разнесены, либо частично или полностью совмещены.

В качестве активных элементов чаще всего используются управляемые отражатели (см. рисунок 2.13): секционированные зеркала и зеркала с непрерывной деформируемой поверхностью.

При деформации зеркальной поверхности изменяется фаза (оптический путь) волн, падающих на данный участок поверхности. При этом возможна компенсация различных видов возмущений, в частности, локальных наклонов волнового фронта, расфокусировки и т.д.

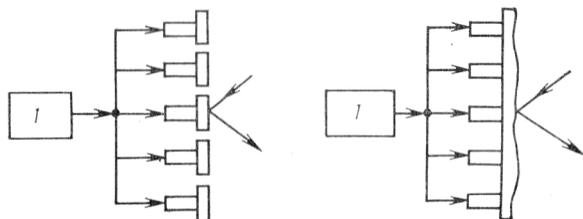


Рисунок 2.13. Управляемые отражатели: секционированное зеркало (а) и зеркало с деформируемой поверхностью (б)

Один из важнейших элементов адаптивной оптической системы – датчик волнового фронта. Устройство предназначено для определения характеристик световых волн, поступающих на вход датчика, после

чего выполняется оценивание и передача параметров волнового фронта в управляющее устройство. Погрешности датчика волнового фронта определяют погрешность всей системы так как дальнейшие процедуры адаптивного управления выполняются в соответствие с данными, полученными с помощью датчика.

На рисунке 2.14 приведен пример структуры системы с датчиком волнового фронта. Поступающий волновой фронт с искажениями разделяется светоделителем и направляется на анализатор волнового фронта, который управляет устройством корректирования фазы волнового фронта. Анализатор волнового фронта формирует сигналы, используемые высокоскоростным процессором, который вычисляет оценки, такие как дрожания фазы или наклон волнового фронта, а также выделяет информацию об искажениях формы волнового фронта более высокого порядка. Эти сигналы затем используются в управляющей ветви обратной связи для получения близкого к плоскому фронту волны в фокальной плоскости видеокамеры и скорректированного изображения объекта с высоким разрешением.

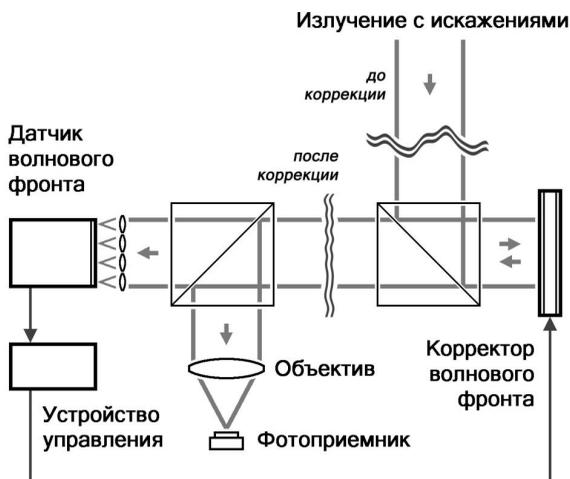


Рисунок 2.14. Общая схема адаптивной оптической системы

Сигналы наклона приводят в движение управляемые зеркала, чтобы удерживать данный объект в центре поля изображения; сигналы управления формой волнового фронта поступают к корректору волнового фронта (деформируемому зеркалу).

В зависимости от решаемой задачи находят применение датчики, предназначенные для оценивания следующих основных оптических параметров:

- Волновые аберрации.
- Распределение интенсивности в изображении точки.
- Интенсивность отраженного света.

Эти параметры характеризуют качество изображения в соответствии с рассмотренными выше критериями.

Как видно из вышеизложенного, принцип действия интерференционных датчиков основан на когерентном сложении двух световых волн и формировании интерференционной картины с зависящим от измеряемого волнового фронта распределением интенсивности света в картине интерференционных полос. При воздействии возмущений волнового фронта наблюдаются отклонения интерференционных полос от их идеального положения при плоском волновом фронте. Обработка картины интерференционных полос позволяет получить оценки отклонений волнового фронта. Методы обработки рассмотрены, например, в [2.6], [2.7].

Специфика систем адаптивной оптики требует, чтобы регистрация и обработка результатов происходили в режиме реального времени. Для повышения быстродействия целесообразно использовать только часть информации, необходимую для функционирования адаптивной системы. При адаптации важна не абсолютная форма волнового фронта, а ее отклонения от номинальной формы. Для определения отклонений широко используются интерферометры сдвига, в которых интерференционная картина формируется при сложении двух волн, образованных из исходной волны, с взаимным поперечным сдвигом. Как показано в разделе 5.2, такое решение повышает информационную эффективность за счет отображения только отклонений, но не полного волнового фронта.

Анализируемая световая волна делится на две волны с помощью светоделителя, при этом в одном плече интерферометра осуществляется сдвиг волнового фронта. Интерференционная картина формируется в результате сложения волн со сдвинутыми волновыми фронтами и проецируется на светочувствительную матрицу. Выделив электронными средствами значения фазы, можно оценить наклон и другие характеристики анализируемого волнового фронта (отдельные виды аберраций). Примеры интерференционных картин при различных искажениях приведены на рисунке 2.15.

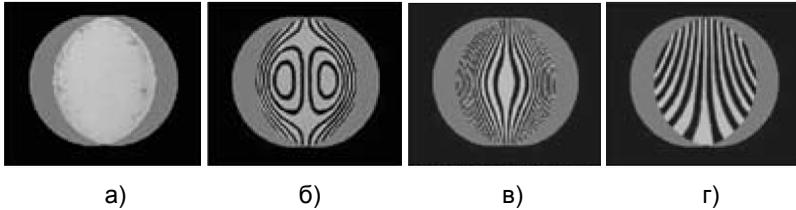


Рисунок 2.15. Примеры картин полос в интерферометре бокового сдвига в случае плоского волнового фронта (а), при сферической aberrации (б), при сферической aberrации и дефокусировке (в), при aberrации комы и дефокусировке (г) [2.7]

Рассмотрим характеристики интерферометра сдвига в области пространственных частот. На рисунке 2.16 показана схема формирования отклонений волнового фронта.

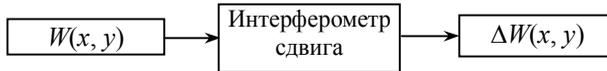


Рисунок 2.16. Структура «вход – выход» интерферометра сдвига

Пусть на вход интерферометра поступает волна с единичной амплитудой и волновым фронтом $W(x, y)$. В интерферометре сдвига формируется разностный волновой фронт

$$\Delta W(x, y) = W(x, y) - W(x - \chi, y), \quad (2.17)$$

где χ – боковой сдвиг волнового фронта по координате X . Отметим, что интерферометр – линейная система по напряженности электрического поля.

Применим к соотношению (2.17) преобразование Фурье $F\{\cdot\}$, при этом в соответствии с известной теоремой сдвига получим

$$F\{\Delta W(x, y)\} = F\{W(x, y)\} - F\{W(x, y)\} \exp(-j2\pi f_x \chi), \quad (2.18)$$

где f_x – пространственная частота для направления X . Выражение (2.18) можно записать в форме

$$F\{\Delta W(x, y)\} = j2 \sin(\pi f_x \chi) F\{W(x, y)\} \exp(-j\pi f_x \chi), \quad (2.19)$$

откуда получим частотную характеристику интерферометра сдвига

$$\begin{aligned} H(f_x) &= \frac{F\{\Delta W(x, y)\}}{F\{W(x, y)\}} = j2 \sin(\pi f_x \chi) \exp(-j\pi f_x \chi) = \\ &= 2 \sin(\pi f_x \chi) \exp[-j\pi(f_x \chi - 1/2)]. \end{aligned} \quad (2.20)$$

Из выражения (2.20) следует, что в интерферометре сдвига осуществляется подавление составляющей на нулевой частоте,

$H(0) = 0$, т.е. устраняется информация о *среднем значении* фазы волнового фронта (знание которого не требуется в адаптивной системе при коррекции *формы* волнового фронта) и происходит модификация спектра периодического характера с частотой $f = 1/\chi$. Как известно (и дополнительно обсуждается в разделе 5.3) оба указанные эффекта повышают информационную эффективность системы. Максимальная эффективность достигается в случае равномерного спектра в широком диапазоне частот. Однако согласно (2.20) частотная характеристика имеет нули на частотах $f = n/\chi$, где n – целое число. Поэтому интерферометр сдвига обладает неравномерной чувствительностью на различных пространственных частотах отклонений волнового фронта. Для преодоления этого недостатка требуется формировать интерференционные картины на наборе нескольких сдвигов, однако при этом усложняется практическая реализация и возрастает вычислительную сложность.

Особенности построения оптических систем интерферометров сдвига подробно рассмотрены в [2.3] и [2.7].

Датчик Шека-Гартмана

Датчик волнового фронта основывается на использовании некогерентного освещения и экрана с рядом отверстий, равномерно расположенных по полю (такой экран называется диафрагмой Гартмана). Принцип контроля волнового фронта состоит в том, что диафрагму располагают перед объективом и делают два снимка: один при положении фотопластины перед фокальной плоскостью объектива (предфокальный снимок), а второй – в положении за нею (зафокальный снимок). По смещению световых пятен судят об изменениях волнового фронта, вносимых объективом.

В дальнейшем диафрагма Гартмана использовалась при проецировании набора пятен на контролируемую поверхность. При отражении направление обратного хода лучей, очевидно, определяется локальным наклоном отражающей поверхности, и зарегистрированный видеокamerой набор пятен позволяет по их смещениям от номинальных положений определить отклонения формы поверхности.

Однако предложенная И. Гартманом методика требовала усовершенствования, прежде всего для повышения эффективности использования энергии источника излучения. Для этого Р. Шек использовал набор линз (линзовый растр).

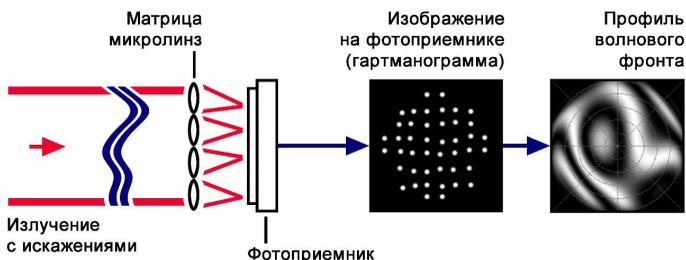


Рисунок 2.17. Датчик Шека-Гартмана

Принцип работы датчика иллюстрируется на рисунке 2.17. Световая волна падает на массив (растр) линз. Линзы разбивают основную апертуру на субапертуры, формирующие фокальные пятна. Направление локальной нормали волнового фронта определяет положение фокального пятна каждой субапертуры. Регистрация набора пятен с помощью матричного фотоприемника позволяет определить положение каждого пятна, и на основе полученных координат пятен вычисляются наклон и другие характеристики отклонений волнового фронта. Восстановление формы волнового фронта осуществляется методом интегрирования отклонений (по нескольким независимым траекториям).

Преимущество датчика Шека-Гартмана состоит в возможности его использования в сравнительно широком диапазоне наклонов волнового фронта, в том числе при контроле объектов больших размеров, когда другие методы (например, интерференционный) не применимы. Кроме того, датчик отличается сравнительно малой чувствительностью к механическим вибрациям, что также является существенным достоинством при практическом применении.

Недостаток датчика заключается в получении информации только на ограниченном наборе точек, что требует использования процедур интерполяции, и в неоднозначности результатов, когда волновой фронт сильно искажен, и при отклонениях требуется идентифицировать положение каждого пятна, которое может выйти за пределы предусмотренного подмассива для субапертуры (см. рис. 2.17) и попасть на соседний подмассив матрицы (относящийся к соседнему пятну). Однозначность достигается при использовании специальных алгоритмов, использующих априорную информацию (ограничения) для отбора корректного решения.

Методики вычисления параметров волнового фронта приведены в [2.3] и [2.7].

2.3 Адаптивный сумматор и его применение в оптике

Адаптивный сумматор в общем случае представляет собой систему с несколькими входами и одним выходом (см. рисунок 2.18). Сигналы на входе суммируются с весовыми коэффициентами, значения которых изменяются (адаптируются) таким образом, чтобы результирующий сигнал на выходе соответствовал заданному опорному сигналу согласно определенному критерию.

Структура адаптивной системы этого вида может использоваться для решения многих задач, в частности, для подавления помех, выделения заданных сигналов, обработки сигналов решетки сенсоров [2.1], в том числе в оптике.

На рисунке 2.18 показана схема адаптивного сумматора.

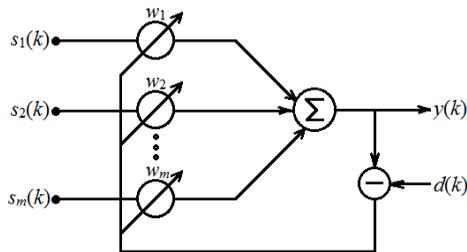


Рисунок 2.18. Общая схема адаптивного линейного сумматора

В этой схеме набор входных сигналов рассматривается как векторный сигнал $\mathbf{s}(k) = [s_1(k), s_2(k), \dots, s_m(k)]^T$, компоненты которого умножаются на регулируемые весовые коэффициенты, рассматриваемые как компоненты вектора $\mathbf{w} = [w_1, w_2, \dots, w_m]^T$, и затем суммируются с формированием выходного сигнала $y(k)$. Выходной сигнал сравнивается с заданным опорным сигналом $d(k)$, и оценка разности сигналов используется для подстройки весовых коэффициентов.

Процесс регулирования или адаптации весовых коэффициентов называют *весовой коррекцией*. Сумматор называют *линейным*, если для некоторого заданного набора весовых коэффициентов выходной сигнал представляет собой линейную комбинацию входных сигналов.

Система, изображённая на рисунке 2.18, – это система с обратной связью (см. рисунок 2.2), в которой вектор весовых коэффициентов зависит от входного сигнала. Поэтому адаптивная система в целом является *нелинейной*.

Адаптивная система может быть реализована в форме адаптивного линейного сумматора с одним входом и элементами задержки (см. рисунок 2.19). Такую структуру принято называть адаптивным *трансверсальным фильтром*.

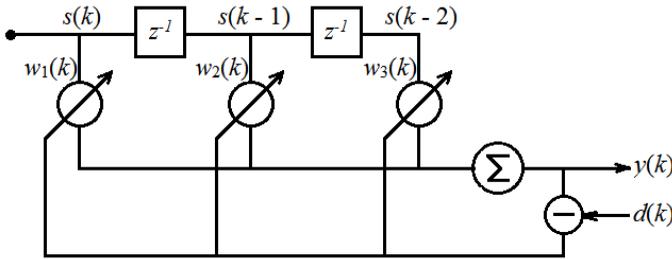


Рисунок 2.19. Схема адаптивного сумматора с одним входом и тремя весовыми коэффициентами. Блоки z^{-1} обозначают задержку на один шаг

Рассмотрим в качестве примера одно из применений адаптивного сумматора (рисунок 2.19) для решения задачи динамического оценивания параметров волнового фронта.

Набор интерференционных картин с фазовым сдвигом γ определяется моделью

$$I(x, y, k) = I_a(x, y) \cos[\gamma k + \varphi(x, y)] + I_b(x, y), \quad (2.21)$$

где $I_a(x, y)$ – амплитудные значения интенсивности интерференционных полос, $I_b(x, y)$ – неинформативная фоновая составляющая интенсивности, $\varphi(x, y)$ – искомые значения фазы интерференционных полос, k – текущий номер интерференционной картины, $k = 0, \dots, K-1$, γ – задаваемый фазовый сдвиг. Значения интенсивности (2.21) при помощи матрицы светочувствительных элементов преобразуются в фотоэлектрические сигналы $s = \mu I$, где μ – коэффициент преобразования, при этом модель (1) принимает вид

$$s(i, j, k) = a(i, j) \cos[\gamma k + \varphi(i, j)] + b(i, j), \quad (2.22)$$

где (i, j) обозначают пиксел с дискретными координатами $x_i = i\Delta x$, $y_j = j\Delta y$, Δx и Δy – шаги пикселов.

Рассмотрим скалярный интерферометрический сигнал в пикселе (i, j) по последовательности зарегистрированных интерференционных картин, который в соответствии с (2.22) определяется выражением

$$s(k) = a \cos(\gamma k + \varphi) + b, \quad (2.23)$$

где $\gamma = 2\pi/P$, P – количество шагов фазового сдвига на одном периоде интерференционных полос. В результате обработки сигнала (2.23) требуется получить точное значение фазы φ при априорно не известных значениях параметров a и b и отличии зарегистрированных значений сигнала от идеальной модели (2.23) вследствие влияния шума и других внешних факторов.

Отсчеты сигнала $s(k)$ взвешиваются с коэффициентами w_{m+1} , $m = 0, \dots, M-1$ и суммируются. При этом сигнал $y(k)$ на выходе фильтра выражается в форме

$$y(k) = \sum_{m=0}^{M-1} w_{m+1} s(k-m) . \quad (2.24)$$

Остаточную ошибку по отношению к заданному опорному сигналу $d(k)$ обозначим $\varepsilon(k) = y(k) - d(k)$. Функционирование адаптивного сумматора основывается на подстройке весовых коэффициентов w_{m+1} согласно критерию минимизации дисперсии отклонения от заданного сигнала

$$\langle \varepsilon^2(k) \rangle \rightarrow \min , \quad (2.25)$$

где угловые скобки обозначают усреднение по ансамблю реализаций сигнала.

Дисперсия ошибки (2.25) при векторном представлении отсчетов сигнала выражается в форме (см., например, [2.1])

$$\varepsilon^2(k) = d^2(k) - \mathbf{w}^T \mathbf{s}(k) \mathbf{s}^T(k) \mathbf{w} - 2d(k) \mathbf{s}^T(k) \mathbf{w} , \quad (2.26)$$

где $\mathbf{w} = [w_1 \ w_2 \ \dots \ w_M]^T$ – вектор весовых коэффициентов фильтра, $\mathbf{s}(k) = [s(k) \ s(k-1) \ \dots \ s(k-M+1)]^T$ – вектор значений сигнала. Вектор $\hat{\mathbf{w}}$, минимизирующий дисперсию ошибки (2.26), определяется условием равенства нулю градиента квадрата ошибки $\partial \varepsilon^2 / \partial \mathbf{w}$, которое выражается известным уравнением Винера-Хопфа в дискретной форме

$$\mathbf{R} \hat{\mathbf{w}} = \mathbf{g} , \quad (2.27)$$

где $\mathbf{R} = \langle \mathbf{s}(k) \mathbf{s}^T(k) \rangle$ – корреляционная (не сингулярная) матрица компонентов сигнала $\mathbf{s}(k)$, имеющая обратную матрицу \mathbf{R}^{-1} , $\mathbf{g} = \langle d(k) \mathbf{s}(k) \rangle$ – вектор взаимной корреляции компонентов сигнала $\mathbf{s}(k)$ и опорного сигнала $d(k)$.

Решение уравнения (2.27) выражается очевидным образом как

$$\hat{\mathbf{w}} = \mathbf{R}^{-1} \mathbf{g} . \quad (2.28)$$

Важно отметить, что в отличие от синтеза фильтра с детерминированной частотной характеристикой адаптивный сумматор представляет собой оптимальный фильтр согласно критерию (2.25), и определение весовых коэффициентов основывается на корреляционных свойствах сигнала (ансамбля сигналов), принадлежащего случайному процессу. При этом учитываются случайные отклонения сигнала, влияние шума и искажений, которые определяют дисперсию ошибки (2.25) в процессе измерений при проведении конкретного эксперимента.

Покажем, что адаптивный сумматор позволяет не только минимизировать дисперсию ошибки (2.26), но и определить значения фазы входного сигнала, которые выражаются через значения коэффициентов фильтра.

Последовательность отсчетов сигнала (2.23) в схеме на рисунке 2.19 определяет выходной сигнал в форме

$$y(k) = w_1[a \cos(\gamma k - \varphi) + b] + w_2[a \cos(\gamma(k-1) - \varphi) + b] + w_3[a \cos(\gamma(k-2) - \varphi) + b]. \quad (2.29)$$

Выражение (2.29) преобразуется к виду

$$y(k) = p \cos(\gamma k + \varphi) + q \sin(\gamma k + \varphi) + c, \quad (2.30)$$

где коэффициенты p , q и c определяются выражениями

$$p = w_1 + w_2 \cos \gamma + w_3 \cos 2\gamma, \quad (2.31)$$

$$q = w_2 \sin \gamma + w_3 \sin 2\gamma, \quad (2.32)$$

$$c = (w_1 + w_2 + w_3)b. \quad (2.33)$$

Выражение (2.30) можно представить в форме

$$y(k) = r \cos(\gamma k + \varphi - \theta) + c, \quad (2.34)$$

где

$$r = \sqrt{p^2 + q^2}, \quad \cos \theta = p/r, \quad \sin \theta = q/r. \quad (2.35)$$

Комбинируя выражения (2.31), (2.32) и (2.35), получаем оценку для сдвига фазы сигнала (2.34) в виде

$$\operatorname{tg} \theta = \frac{q}{p} = \frac{w_2 \sin \gamma + w_3 \sin 2\gamma}{w_1 + w_2 \cos \gamma + w_3 \cos 2\gamma}. \quad (2.36)$$

Априорно заданный опорный сигнал $d(k)$ (см. рисунок 2.19) с известными амплитудой A и фоновой составляющей B определяется выражением

$$d(k) = A \cos(2\pi k / P) + B = A \cos(\gamma k) + B. \quad (2.37)$$

В соответствии с (2.34) минимизация отклонений от опорного сигнала (2.37) достигается при равенстве этих сигналов, а именно при условии

$$r \cos(\gamma k + \varphi - \theta) + c = A \cos(\gamma k) + B, \quad (2.38)$$

из которого следует, что $r = A$, $\varphi = \theta$, $c = B$. При этом в соответствии с (2.36) получаем оценку фазы

$$\varphi = \text{arctg} \frac{w_2 \sin \gamma + w_3 \sin 2\gamma}{w_1 + w_2 \cos \gamma + w_3 \cos 2\gamma} \quad (2.39)$$

Таким образом, при помощи адаптивного сумматора возможно получение оценок фазы интерферометрического сигнала (2.23) по формуле (2.39) с использованием коэффициентов $\hat{\mathbf{w}} = [\hat{w}_1 \hat{w}_2 \hat{w}_3]^T$ адаптивного сумматора.

Оценки фазы (2.39) вычисляются для сигнала (2.22) от каждого пиксела видекамеры при формировании векторного сигнала (2.22).

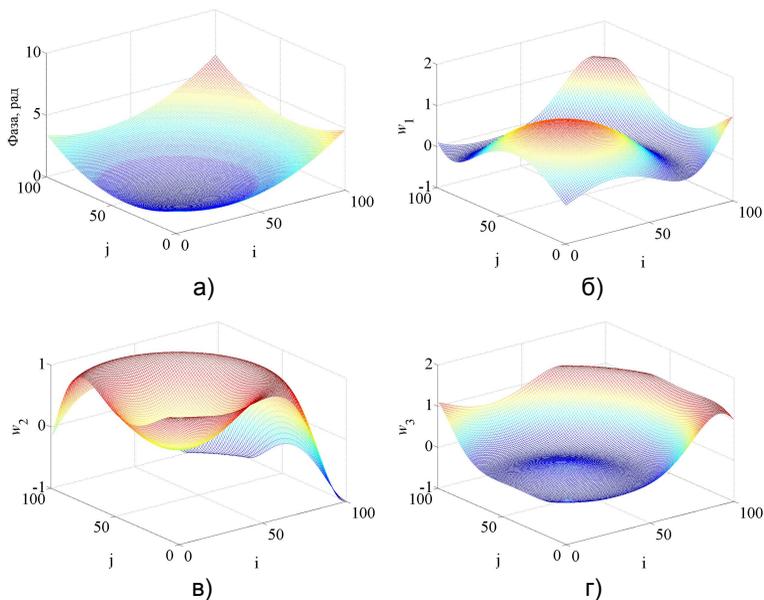


Рисунок 2.20. Распределение фазы волнового фронта форматом 100×100 пикселей (а) и соответствующие значения весовых коэффициентов w_1 , w_2 , w_3 адаптивного сумматора (б–г).

На рисунке 2.20, а иллюстрируется пример пространственного распределения фазы волнового фронта, которому соответствуют

значения весовых коэффициентов $\hat{w}(i, j)$, полученные при пяти шагах фазового сдвига $\gamma = 2\pi/4$ [2.8].

Использование адаптивного сумматора позволяет определять параметры интерференционных полос с высокой точностью в динамическом режиме при малом количестве регистрируемых интерференционных картин. Оценки СКО фазы интерференционных полос находятся на уровне $2\pi/100$ (и менее) [2.8] и соответствуют весьма малой погрешности $\lambda/200$ (λ – длина волны источника излучения) при определении формы волнового фронта.

Контрольные вопросы

1. Объясните особенности построения адаптивных систем без обратной связи и с обратной связью.
2. Перечислите основные свойства адаптивных систем. В чем состоят недостатки адаптивных систем?
3. Приведите примеры адаптивных систем с обратной связью.
4. Перечислите основные виды и источники внешних возмущений оптического поля, компенсируемые в адаптивных оптических системах.
5. В чем состоят особенности адаптивных оптических систем с выходящей волной и с принимаемой волной?
6. Объясните принципы построения и функционирования адаптивных оптических систем с фазовым сопряжением.
7. Объясните принципы построения и функционирования адаптивных оптических систем апертурного зондирования.
8. Объясните формулы, определяющие основные критерии качества в системах адаптивной оптики.
9. Какие известны виды датчиков волнового фронта? В чем состоят их достоинства и недостатки?
10. Приведите основные структурные схемы линейного сумматора. Раскройте принцип оценивания фазы волнового фронта при помощи адаптивного сумматора.

Список литературы

- 2.1. Уидроу Б., Стирнз С. Адаптивная обработка сигналов. – М.: Радио и связь, 1989.
- 2.2. Лукьянов Д.П., Корниенко А.А., Рудницкий Б.Е. Оптические адаптивные системы. – М. Радио и связь, 1989.
- 2.3. Ермолаева Е.В., Зверев В.А., Филатов А.А. Адаптивная оптика. – СПб: НИУ ИТМО, 2012.

- 2.4. Тараненко В.Г., Шанин О.И. Адаптивная оптика. – М.: Радио и связь, 1990.
- 2.5. Adaptive optics for biological imaging / J.A. Kubby, ed. – NY: CRC Press, Taylor & Francis Group, LLC, 2013.
- 2.6. Васильев В.Н., Гуров И.П. Компьютерная обработка сигналов в приложении к интерферометрическим системам. – СПб: БХВ-Санкт-Петербург, 1998.
- 2.7. Malacara D., Servín M., Malacara Z. Interferogram analysis for optical testing. NY: Taylor & Francis, LLC, 2005.
- 2.8. Гуров И. П., Артемьева И. А.; Капанова В. О. Динамическое оценивание фазы в интерферометрии управляемого фазового сдвига методом адаптивной фильтрации Винера // Оптический журнал. 2020. Т. 87. №5. С. 42–53.

3 Интерферометрические системы

3.1 Аппаратная функция интерферометрической системы

Интерферометрические системы основываются на использовании принципа сложения нескольких когерентных электромагнитных колебаний с последующим детектированием суммарного колебания.

Интерферометрическая система включает источник излучения и сформированные с помощью интерферометра тракты распространения электромагнитных волн. По результату взаимодействия электромагнитных волн можно определять свойства источника излучения, либо исследуемого объекта, размещаемого в интерферометре. Наиболее распространены двухлучевые интерферометры, в которых предметная (измерительная) волна складывается с опорной волной, что позволяет определять свойства исследуемого объекта, влияющего на характеристики предметной волны.

В научных исследованиях наиболее важны чувствительность и разрешающая способность интерферометрической системы, в технических приложениях существенны функциональные возможности, помехоустойчивость и быстродействие.

Интерферометрические системы относятся к активным системам, использующим свойство когерентности излучения. Излучение источника разделяется на измерительную волну, взаимодействующую с объектом, и опорную волну (см. рисунок 3.1) при последующей регистрации результирующей интенсивности I излучения.

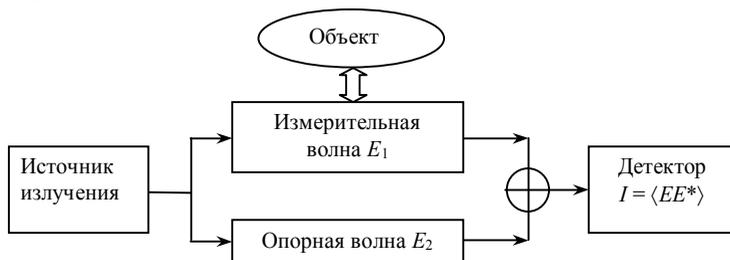


Рисунок 3.1. Принцип построения интерферометрической системы

При сложении двух когерентных монохроматических оптических колебаний с частотой ν и напряженностью электрического поля $E_1 = a_1 \exp[j(2\pi\nu t + \Phi_1)]$ и $E_2 = a_2 \exp[j(2\pi\nu t + \Phi_2)]$ результирующая

интенсивность, как известно, наблюдается при квадратичном детектировании суммарного колебания $E = E_1 + E_2$ и усреднении, обусловленном ограниченным быстродействием фотодетектора по сравнению с периодом колебаний $1/\nu$, а именно

$$I = \langle |E|^2 \rangle = \langle EE^* \rangle = a_1^2 + a_2^2 + 2a_1a_2 \cos \Phi, \quad (3.1)$$

где угловые скобки обозначают операцию усреднения, $\Phi = \Phi_2 - \Phi_1 = 2\pi\Delta/\lambda$ – разность фаз, Δ – оптическая разность хода,

$\lambda = c/\nu$ – длина волны, c – скорость света в вакууме.

Выражение (3.1) можно записать в форме

$$I = T(V, \Phi) I_0, \quad (3.2)$$

где

$$T(V, \Phi) = 1 + V \cos \Phi \quad (3.3)$$

– аппаратная функция интерферометра (оператор системы),

$$I_0 = a_1^2 + a_2^2, \quad V = 2a_1a_2/(a_1^2 + a_2^2). \quad (3.4)$$

Таким образом, интерферометрическая система (см. рисунок 3.2), включающая исследуемый объект, освещается источником когерентного излучения с интенсивностью I_0 , и на выходе наблюдается результат нелинейного преобразования интенсивности (3.2), по которому судят об объекте.

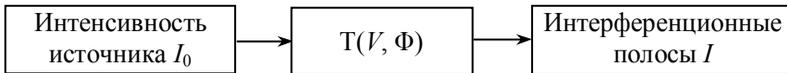


Рисунок 3.2. Структура «вход – выход» интерферометрической системы

Параметр V , называемый *видностью* интерференционных полос, и разность фаз Φ в (3.3) определяются свойствами исследуемого объекта и часто являются искомыми параметрами. С позиций системного подхода получение информации об исследуемом объекте следует рассматривать с точки зрения решения задачи параметрической идентификации аппаратной функции (3.3) интерферометрической системы, включающей исследуемый объект.

Поскольку в уравнения (3.2) и (3.3) входят три величины I_0 , V , Φ для определения параметров аппаратной функции недостаточно единственного измерения интенсивности в (3.2), поэтому требуется использовать определяемые конкретной решаемой задачей методы формирования и обработки регистрируемых интерферометрических сигналов для извлечения полезной информации.

Можно попытаться получить искомую оценку этой функции методом прямого обращения

$$\hat{T}(V, \Phi) = I/I_0, \quad (3.5)$$

после чего из (3) определить разность фаз как

$$\hat{\Phi} = \arccos \{ [\hat{T}(V, \Phi) - 1] / V \}. \quad (3.6)$$

Проблема состоит в том, что значения I_0 , V в (3.5), (3.6) априорно не известны. Кроме этого, формирование интерферометрических данных на практике сопровождается воздействием комплекса шумов и помех. В результате решение (3.6) не удовлетворяет условиям единственности и устойчивости.

Преодолеть указанные трудности можно при помощи методов, которые основываются на использовании априорной информации (например, о непрерывности изменений фазы, плавных изменениях видности по сравнению с шагом интерференционных полос и др.). При этом используют методы обработки интерферометрических сигналов, подробно рассмотренные в [3.1], [3.2].

Практически при получении соотношений вида (3.1) нужно учитывать:

- спектральные (временные и пространственные) характеристики источника излучения и связанные с ними параметры когерентности;

- отклонения формы волнового фронта, направления векторов поляризации, значения показателей преломления среды в рабочей зоне интерферометра;

- геометрические размеры в сечении световых пучков и чувствительной площадки фотодетектора.

Существенны также дополнительные параметры: коэффициенты отражения отдельных рабочих поверхностей и параметры шумов источника излучения, пространственных шумов, вызванных остаточной шероховатостью поверхностей оптических элементов, шумов приемника излучения [3.3].

С другой стороны, при варьировании отдельных параметров можно исследовать широкое многообразие физических эффектов и различных объектов.

3.2 Структура интерферометрических систем

Рассмотрим структуру интерферометрических систем с позиций системного подхода.

Интерферометрические системы можно подразделить на несколько основных видов в зависимости от области применения и решаемых задач.

В современных системах осуществляется фотоэлектрическая регистрация и последующая обработка интерферометрических сигналов. При высокоточном контроле объектов методами интерферометрии обеспечивается стабильность условий проведения эксперимента, и задача состоит в регистрации и обработке *детерминированных* сигналов, близких к идеализированной модели интерференционных полос, что соответствует обобщенной схеме, приведенной на рисунке 1.5. В случае если исследуемый процесс сопровождается случайными изменениями амплитуды (фазы) измерительной волны, следует рассматривать *стохастические* интерферометрические сигналы (см. схему на рисунке 1.6).

Формирование, регистрация и обработка интерферометрических сигналов осуществляются в системах с различной структурой, при этом виды систем отличаются широким разнообразием. Проведем краткое рассмотрение и систематизацию интерферометрических систем с точки зрения их структуры.

Как известно, интерференционная система содержит источник излучения со стороны входного зрачка, приемник излучения, расположенный у выходного люка в поле интерференции, и ветви интерферометра, имеющие общий входной зрачок и общее поле, что выражается структурной схемой, содержащей источник с комплексной амплитудой электрического поля E и в общем случае несколько (n) *трактов* распространения W_1, W_2, \dots, W_n когерентных световых волн с результирующей комплексной амплитудой E_Σ , как это иллюстрируется на схеме, показанной на рисунке 3.3.

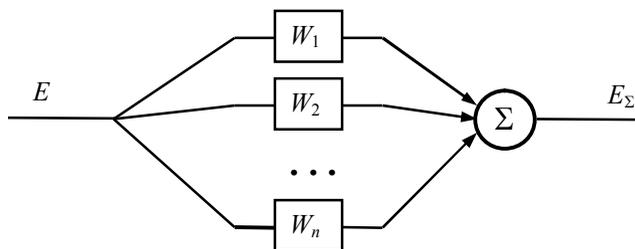


Рисунок 3.3. Тракты распространения интерферирующих волн

Тракты распространения оптических волн в ветвях интерферометра $W_l, l = 1, \dots, n$, характеризуются комплексной

величиной $W = |W| \exp(j \arg W)$. При проведении исследований в зависимости от решаемой задачи могут представлять интерес модуль и/или фаза результирующего поля, характеристики отдельного тракта $|W_i|$, $\arg W_i$, либо разности оптических длин трактов $\Delta \arg W_i = \delta_{ij} / 2\pi$, где δ_{ij} – разность фаз интерферирующих волн. Как уже отмечалось, наиболее распространены двухлучевые интерферометры ($n = 2$), см. рисунок 3.1.

Интерферометрическая система может содержать m независимых взаимно некогерентных измерительных каналов (рисунок 3.4), в каждом из которых распространение световых волн определяется вектором параметров $\mathbf{W}_j = (W_{j1}, W_{j2}, \dots, W_{jn})^T, j = 1, \dots, m$.

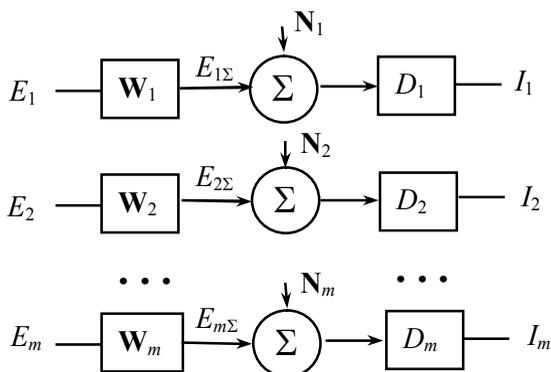


Рисунок 3.4. Структура многоканальной интерферометрической системы

Доступными для измерения являются значения интенсивности света I_j на выходе каждого канала при выполнении квадратичного детектирования D_j результирующего значения $E_{j\Sigma}$ с учетом воздействия помех N_j (см. рисунок 3.4). При этом исследуемый объект или процесс характеризуется некоторой функцией $f(I_1, I_2, \dots, I_m)$ от зарегистрированных значений I_j .

Структура системы, показанная на рисунке 3.4, относится, в том числе, к случаю использования источников излучения с набором отдельных длин волн (спектральное разделение каналов) или с частичной пространственной когерентностью, когда наблюдаемое интерференционное поле можно представить как совокупность

взаимно некогерентных областей (пространственное разделение каналов).

При формировании и регистрации спекл-интерференционных полей интерферометрическую систему можно рассматривать как систему с многими входами и выходами (см. рисунок 3.5). При этом от каждой точки объекта на входе формируются сферические волны E_1, E_2, \dots, E_n , регистрируемые набором m светочувствительных ячеек. При изменении во времени геометрического положения или свойств исследуемого объекта происходит формирование стохастических спекл-интерференционных сигналов, по характеристикам которых судят об исследуемом объекте или процессе.

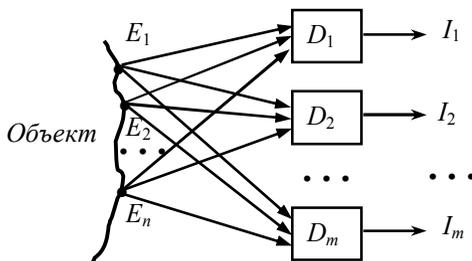


Рисунок 3.5. Система с многими входами и выходами

Интерферометрические каналы можно подразделить на четыре основных вида: спектральные, поляризационные, с пространственным разделением и с временным разделением.

При спектральном разделении интерферометрические сигналы формируются на нескольких длинах волн оптического излучения (см. рисунок 3.6). Спектральное разделение обычно выполняют с помощью спектрального прибора СП, размещенного на выходе интерферометра.

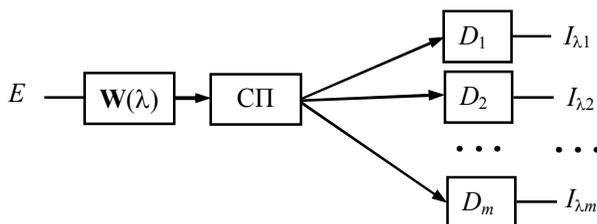


Рисунок 3.6. Спектральная интерферометрическая система

Поляризационное разделение основывается на использовании световых волн с различной (обычно ортогональной) поляризацией. По структуре такие системы подобны показанной на рисунке 3.6, где спектральный прибор заменяется на поляризационные элементы.

При пространственном разделении осуществляется формирование пространственно протяженного интерференционного поля с регистрацией интенсивности независимо на различных участках (см. рисунок 3.4), например, при использовании видеокамеры.

Интенсивность света I на выходе каждого интерферометрического канала определяется выражением

$$I(\mathbf{W}) = E_{\Sigma} E_{\Sigma}^*, \quad (3.7)$$

которое для совокупности интерферометрических каналов принимает векторную форму

$$\mathbf{I}(k, \mathbf{W}) = \begin{pmatrix} I_1(k, \mathbf{W}_1) \\ I_2(k, \mathbf{W}_2) \\ \dots \\ I_m(k, \mathbf{W}_m) \end{pmatrix}, \quad (3.8)$$

где k – временной индекс m -канального процесса, как иллюстрируется схемой на рисунке 3.7 с учетом коэффициентов преобразования K_j интенсивности в электрические сигналы s_j в условиях воздействия шума наблюдения N_{aj} . Таким образом, доступным для обработки и анализа является векторный сигнал с компонентами $s_j(k)$ в дискретном времени.

Структурную схему, показанную на рисунке 3.7, можно интерпретировать как систему с несколькими выходами – компонентами векторного сигнала. Обработка векторного сигнала в отдельном канале ведется в подсистеме с несколькими входами (компонентами векторного сигнала с параметрами W_{jl}) и одним выходом с формированием оценки параметров сигнала s_j . При этом оценка вектора \mathbf{W} может рассматриваться как решение обратной задачи *идентификации трактов* W_{jl} распространения волн.

Временное разделение применяется при формировании набора интерференционных картин в методах управляемого фазового сдвига. При этом значения $s(i, j, k)$ в полученном наборе видеок кадров интерпретируются как векторный сигнал $\mathbf{s}(k)$, компонентами которого являются значения сигнала отдельных пикселей (i, j) регистрирующей матрицы светочувствительных ячеек по последовательности видеок кадров. Выходной сигнал интерферометрической системы в этом случае определяется моделью

$$s(i, j, k, \boldsymbol{\theta}) = B(i, j) + A(i, j, k) \cos[\Phi(i, j, k)], \quad (3.9)$$

где $\boldsymbol{\theta} = (B, A, \Phi)^T$ – вектор параметров.

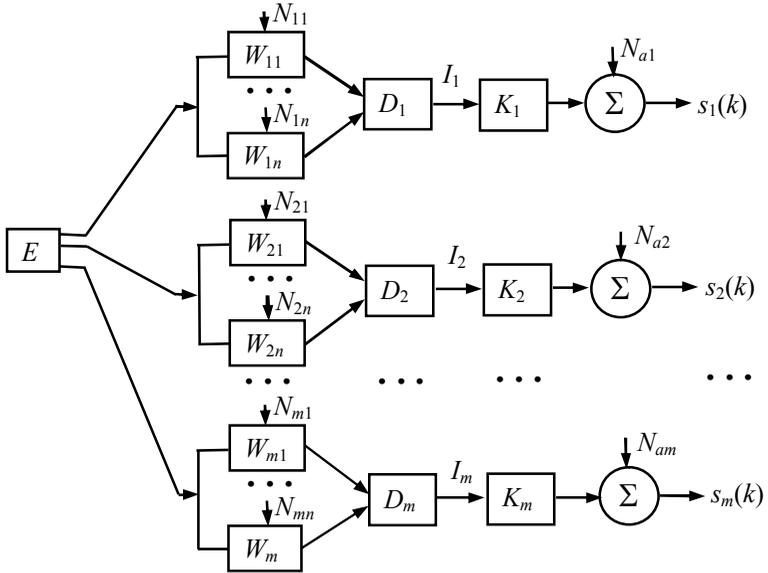


Рисунок 3.7. Обобщенная структурная схема интерферометрической системы

Интерферометрические системы часто подразделяются на две группы: системы с *неподвижной* интерференционной картиной и с *движущейся* интерференционной картиной. В системах первой группы временной индекс k в (3.9) фиксируется, и ведется обработка многомерного сигнала

$$s(i, j, \boldsymbol{\theta}) = B(i, j) + A(i, j) \cos[\Phi(i, j)], \quad (3.10)$$

определяемого в отдельных пикселах (i, j) .

В интерферометрических системах второй группы регистрируют движущуюся интерференционную картину при изменении разности фаз $\Delta\Phi$ интерферирующих волн, при этом $\Phi = \Phi(i, j, k)$ и

$$s(i, j, k, \boldsymbol{\theta}) = B(i, j) + A(i, j) \cos[\Phi(i, j) + k\Delta\Phi(i, j)]. \quad (3.11)$$

В ряде случаев используются комбинированные методы. В частности, спектрально-временное разделение каналов возможно

выполнить с использованием источника излучения с перестраиваемой (переключаемой) во времени длиной волны на входе интерферометра.

Проведенное рассмотрение структуры интерферометрических систем позволяет проводить исследования и разработки с единых позиций описания систем и сигналов на основе многомерного и векторного представления с использованием современных методов и алгоритмов компьютерной обработки интерферометрических сигналов.

3.3 Характеристики интерферометрических систем и сигналов

Интенсивность света (3.2) в целях последующей компьютерной обработки данных подвергается дискретизации по независимой переменной, фотоэлектрическому преобразованию и квантованию по уровню. В реальных условиях данные искажаются воздействием помех, которое характеризуется оператором Ψ , при этом выходная последовательность экспериментальных данных имеет вид

$$\xi(k, \theta) = \Psi \{ \mu [I(k), \Delta s_q] \}, \quad (3.12)$$

где $k = 1, \dots, L$, L – объем выборки, $\theta = (V, \Phi)^T$ – вектор параметров модели, $I(k)$ – значения интенсивности, полученные в результате дискретизации на k -ом шаге, μ – постоянный коэффициент фотоэлектрического преобразования, Δs_q – вектор, компонентами которого являются элементарные (кодовые) квантованные сигналы.

Схема формирования и обработки интерферометрических данных иллюстрируется схемой рисунка 3.8.

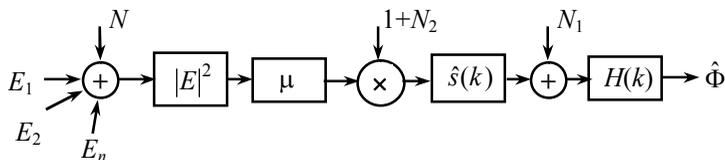


Рисунок 3.8. Одноканальная интерферометрическая система

Учитывая воздействие влияющих факторов, можно выделить три основные варианта модели (3.12), а именно

$$\xi(k, \theta) = s(k, \theta) + N_1(k), \quad (3.13)$$

$$\xi(k, \theta) = s(k, \theta)[1 + N_2(k)], \quad (3.14)$$

$$\xi(k, \theta) = s(k, \theta + \delta\theta_k), \delta\theta = N(k), \quad (3.15)$$

где $s(k, \theta)$ – идеализированная модель данных, N_1, N_2 – случайные шумовые последовательности, \mathbf{N} – вектор, компонентами которого являются шумовые последовательности, соответствующие компонентам вектора θ .

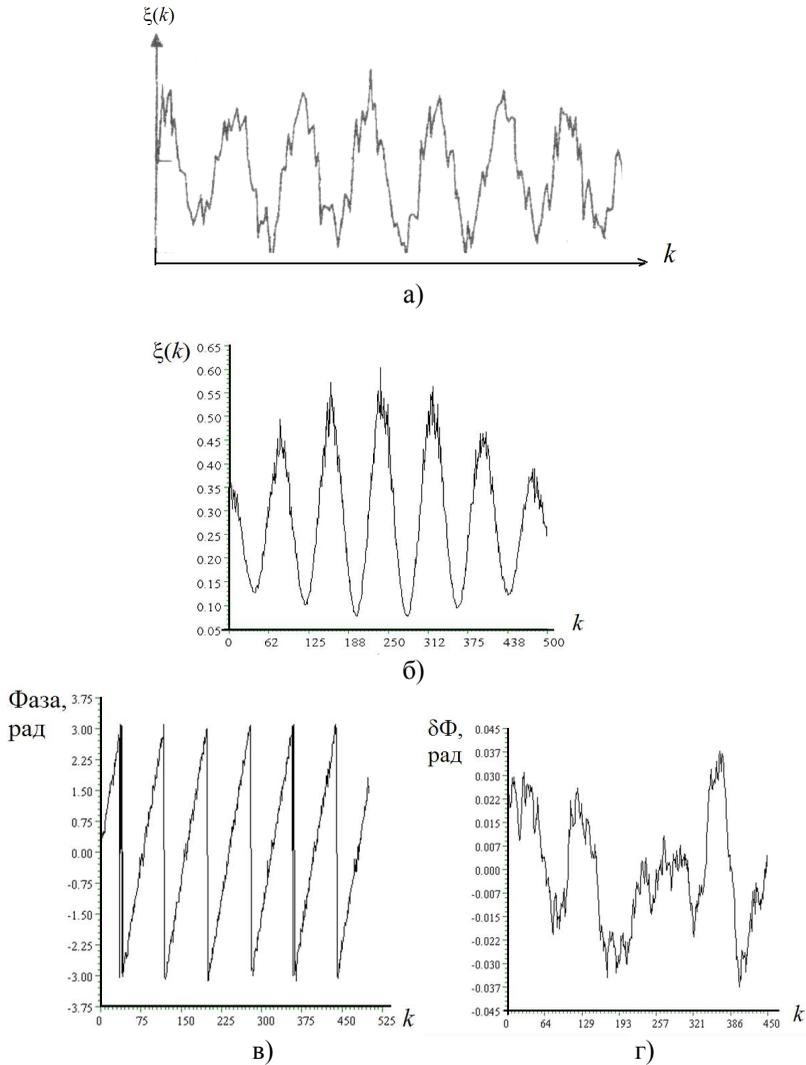


Рисунок 3.9. Примеры влияния шума в интерферометрической системе

Оператор Ψ в (3.12) характеризует в соответствии с (3.13)–(3.15) воздействие аддитивного шума N_1 , не зависящего от полезных данных $s(k, \theta)$ согласно (3.13), как это иллюстрируется на рисунке 3.9, а. Пример влияния мультипликативного шума N_2 , коррелированного с значениями интерферометрического сигнала (выражение (3.14)) приведен на рисунке 3.8, б. Данный вид шума типичен при фотодетектировании сигналов, искаженных дробовым шумом. Случайные отклонения параметров в (3.15), в частности, в виде флуктуаций фазы, показаны на рисунке 3.8, в–г.

Интерферометрические данные исключительно с шумами вида (3.13) практически не встречаются: к этому случаю следует отнести в основном внутренние шумы фотодетектора и шум квантования сигнала по уровню. Дробовой шум источника и приемника излучения, флуктуации направления векторов поляризации, изменения коэффициентов отражения рабочих поверхностей порождают модель вида (3.14), см. рисунок 3.8, б (на рисунке показано воздействие преимущественно дробового шума, дисперсия которого пропорциональна среднему значению полезного сигнала). Модель (3.15) в общем случае учитывает влияние шумов источника излучения, остаточной шероховатости поверхности объекта, флуктуаций показателя преломления и ряда других факторов.

Вышеизложенное позволяет выделить три группы задач:

- методы идентификации адекватного варианта модели (3.12) по эмпирическим данным;
- исследование методов и алгоритмов оценивания параметров или, в общем случае, когда $\theta = \theta(k)$, оптимальной фильтрации полученных данных;
- декомпозиция сложной совокупности влияющих факторов, включая информативные, и определение их статистических характеристик.

Важно отметить, что в (3.13)–(3.15) параметры идеализированной модели интерферометрических данных $s(k, \theta)$, как правило, априорно неизвестны с требуемой точностью. Известен лишь вид модели

$$s(k, \theta) = A(k) \cos \Phi(k) + B(k). \quad (3.16)$$

Следовательно, необходимо дополнительно оценивать или исключать влияние в (3.16) сопутствующих параметров – фоновой составляющей $B(k)$ и огибающей $A(k)$.

Методы и алгоритмы обработки интерферометрических сигналов детально рассмотрены в [3.1], [3.2].

3.4 Динамические интерферометрические системы

Аппаратная функция $T(V, \Phi)$ интерферометрической системы может быть определена не только в терминах «вход-выход», как показано на рисунке 3.2, но и в параметрической форме, выражающей внутреннее описание системы.

В соответствии с (3.2) и (3.3) после преобразования оптического излучения в фотоэлектрический сигнал значения сигнала определяются выражением

$$s = 1 + V \cos \Phi. \quad (3.16)$$

При изменении во времени оптического пути предметной или опорной волны разность фаз Φ изменяется. Следовательно, аппаратная функция при этом не постоянна во времени и характеризует динамическую интерферометрическую систему. Следует отметить, что в данном случае имеется в виду система с изменяющимися параметрами, которые составляют вектор параметров $\theta = (V \Phi)^T$. Описание системы можно выразить через эволюцию параметров, которую обычно рассматривают в дискретном времени $t_k = k\Delta t$, $k = 0, \dots, K$, Δt – шаг дискретизации во времени. При этом

$$s(k) = 1 + V(k) \cos \Phi(k). \quad (3.17)$$

Функции $V(k)$ и $\Phi(k)$ в (3.17) могут быть детерминированными, но на практике вследствие высокой чувствительности интерферометрических систем и влияния внешних факторов представляют собой реализации (выборочные функции) случайных процессов $\{V(k)\}$ и $\{\Phi(k)\}$. При этом сигнал $s(k)$ также представляет значения случайной функции.

Задача состоит в динамическом определении вектора параметров $\theta(k) = [V(k) \Phi(k)]^T$ по наблюдаемым значениям $s(k)$, связанным с вектором параметров. Динамическую систему можно описать двумя уравнениями

$$\theta(k) = \mathbf{g}[\theta(k-1), \mathbf{w}(k)], \quad (3.18)$$

$$s(k) = h[\theta(k), n(k)]. \quad (3.19)$$

Выражение (3.18) представляет собой уравнение *динамической системы*, выражающее *эволюцию* вектора параметров $\theta(k)$ на текущем шаге по отношению к предыдущему шагу $k-1$ с учетом влияния шума системы, выражаемого случайным вектором $\mathbf{w}(k)$. Вид функции $\mathbf{g}[\cdot]$ определяется свойствами конкретной системы.

В свою очередь уравнение (3.19) носит название *уравнения наблюдения*, которое выражает значение сигнала в зависимости от вектора параметров через функциональное преобразование $h[\cdot]$ с учетом влияния шума наблюдения $n(k)$. В интерферометрических системах функция $h[\cdot]$ определяется выражением (3.17).

Один из наиболее продуктивных подходов к описанию эволюции динамической системы состоит в использовании формализма дифференциальных уравнений. Действительно, если фоновая составляющая B и частота интерференционных полос f предполагаются постоянными и фаза интерференционных полос $\Phi = 2\pi f x + \varepsilon$, где ε – начальная фаза, x – независимая переменная, изменяется линейно в данной интерферометрической системе, справедливы соотношения

$$\frac{dB}{dx} = 0, \frac{df}{dx} = 0, \frac{d\Phi}{dx} = 2\pi f. \quad (3.20)$$

Следующий шаг состоит в учете случайных факторов при изменении параметров динамической системы и включении их в модель (3.20) при модификации дифференциальных уравнений. Представление уравнений (3.20) в форме

$$\frac{dB}{dx} = w_B(x), \frac{df}{dx} = w_f(x), \frac{d\Phi}{dx} = 2\pi f + w_\Phi(x), \quad (3.21)$$

где $\mathbf{w} = (w_B \ w_f \ w_\Phi)^T$ – случайный вектор, позволяет учесть стохастические изменения параметров интерференционных полос. Уравнения (3.21) представляют собой стохастические дифференциальные уравнения первого порядка в форме Ланжевена [3.4], которые могут быть записаны в векторном виде

$$\frac{d\boldsymbol{\theta}}{dx} = \mathbf{L}(x, \boldsymbol{\theta}) + \mathbf{w}(x), \quad (3.22)$$

где первое слагаемое относится к детерминированной эволюции параметров, второе слагаемое представляет их случайные отклонения. Априорная информация об эволюции вектора параметров $\boldsymbol{\theta}$ учитывается подходящим выбором соответствующей векторной функции \mathbf{L} и предполагаемыми статистическими свойствами «формирующего» шума $\mathbf{w}(x)$. Важно подчеркнуть, что уравнение (3.22) определяет в общем случае эволюцию нестационарных и нелинейных систем.

В дискретном случае уравнение (3.22) записывается в форме стохастического разностного уравнения, определяющего дискретную последовательность при значениях независимой переменной в точках

$x_k = k\Delta x$, $k = 1, \dots, K$, где Δx – шаг дискретизации. При этом возможно описание эволюции вектора параметров системы в рекуррентной форме на k -ом шаге в виде

$$\boldsymbol{\theta}(k) = \boldsymbol{\theta}(k/k-1) + \mathbf{w}(k), \quad (3.23)$$

где $\boldsymbol{\theta}(k/k-1)$ – предсказанное значение с $(k-1)$ -го шага на k -й шаг с учетом конкретной заданной зависимости $\mathbf{L}(x, \boldsymbol{\theta})$ в уравнении (3.22).

Предсказание в (3.23) содержит ошибку, т.е. отличие априорного знания на шаге $k-1$ от фактического значения на шаге k . Это различие доступно для наблюдения только как отклонения значений интерферометрического сигнала (3.19). Для получения апостериорной оценки параметров $\hat{\boldsymbol{\theta}}(k)$ на k -ом шаге сигнал ошибки нужно преобразовать в поправку к вектору параметров, а именно

$$\hat{\boldsymbol{\theta}}(k) = \boldsymbol{\theta}(k/k-1) + \mathbf{P}(k) \{ \hat{s}(k) - s[k, \boldsymbol{\theta}(k/k-1)] \}, \quad (3.24)$$

где $\mathbf{P}(k)$ – векторная функция, преобразующая скалярную разность (в фигурных скобках) между наблюдаемым отсчетом сигнала $\hat{s}(k)$ и модельным (предсказанным с предыдущего шага) значением $s[k, \boldsymbol{\theta}(k/k-1)]$ в векторную поправку параметров сигнала.

Выражение (3.24) иллюстрируется схемой, показанной на рисунке 3.10.

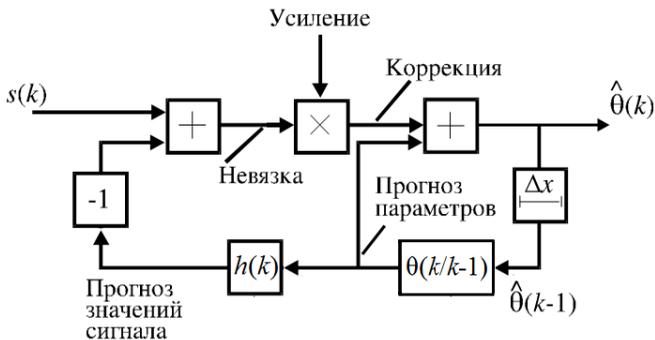


Рисунок 3.10. Идентификация параметров динамической системы

Входные значения $s(k)$ сравниваются с прогнозируемыми согласно выражению (3.19) с учетом (3.17). Полученная невязка усиливается множителем $\mathbf{P}(k)$ в (3.24). В результате на каждом шаге формируется поправка вектора параметров, спрогнозированного с

предыдущего шага, исходя из полученной оценки $\hat{\theta}(k-1)$ и эволюции параметров, определяемой выражением (3.18). При этом осуществляется оперативная идентификация параметров системы $\hat{\theta}(k)$ на каждом шаге дискретизации.

Интерферометрические системы управляемого фазового сдвига

Методы интерферометрии управляемого фазового сдвига (ИФС) основаны на формировании набора интерференционных картин с заданным взаимным фазовым сдвигом, вносимым в опорное плечо интерферометра, и совместной обработке набора картин для оценивания фазы волнового фронта [3.1], [3.2].

В ИФС используется модель интерферометрического сигнала отдельного пиксела (i, j) в форме

$$s(k) = A \cos(2\pi f k \Delta x + \varepsilon) + B, \quad (3.25)$$

где амплитуда сигнала A и фоновая составляющая B – постоянные, значения которых априорно не известны, частота интерференционных полос $f = 1/\lambda$ определяется длиной волны λ , шаг Δx определяет задаваемое изменение оптической разности хода в интерферометре. Количество фазовых сдвигов K на периоде изменения сигнала обычно находится в диапазоне значений $K = 3 \dots 11$ и определяется алгоритмом вычисления фазы (см., например, [3.2]). Задача состоит в точном определении начальной фазы ε в (3.25) при обработке последовательности отсчетов $s(k)$.

Алгоритмы, позволяющие определять значения фазы интерференционных полос в динамическом режиме, представляют особый интерес ввиду возможности оценивания объектов, подверженных малым отклонениям их формы, например, вследствие дрейфа температуры окружающей среды и влияния других внешних факторов в процессе измерений.

Вектор параметров модели (3.25) задается в форме $\theta = (B, A, \Phi, f)^T$, учитывая, что начальная фаза ε согласно (3.25) вычисляется как $\varepsilon = \Phi - 2\pi f k \Delta x$. Если задаваемый шаг по фазе $2\pi f \Delta x$ непостоянен, например, вследствие влияния внешних возмущений на оптическую разность хода в интерферометре, эти изменения выражаются как локальные отклонения частоты интерференционных полос $f = f(k)$. При известных значениях $\Phi(k)$ и $f(k)$ несложно вычислить начальную фазу ε .

На рисунке 3.11 иллюстрируются рекуррентные оценки полной фазы Φ и начальной фазы ε как функции от шага обработки (номера отсчета сигнала k), полученные при использовании алгоритма, схема которого приведена на рисунке 3.10. При моделировании задавались четыре различных значения «истинной» начальной фазы $\varepsilon = \pm 1.5$ рад и ± 2.5 рад и начальное условие (на «нулевом» шаге) $\varepsilon = \pi/2$. Количество шагов на одном периоде интерференционных полос устанавливалось равным $K=100$. На рисунке 3.11 видно, что правильные оценки начальной фазы (смещенные на заданную величину $\varepsilon = \pi/2$) получены на примерно 15-ом шаге, т.е. на 0,15 доле от полного периода.

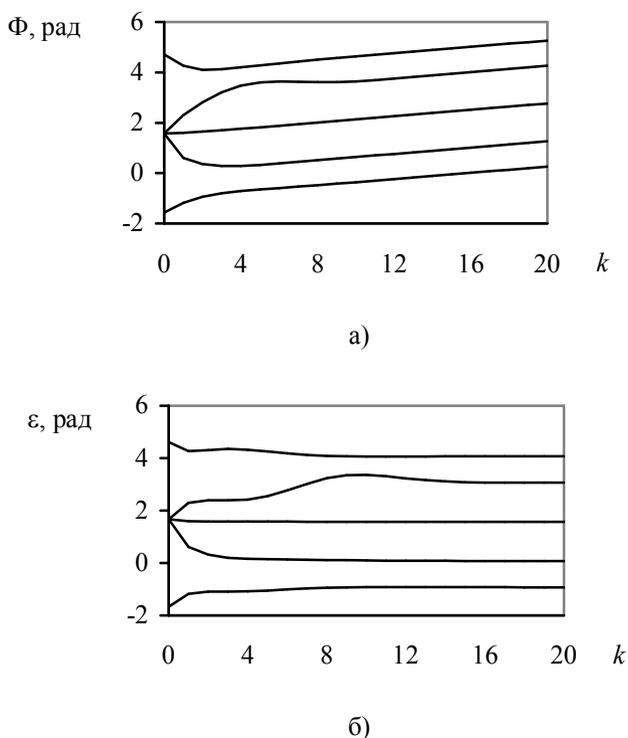


Рисунок 3.11. Рекуррентные оценки полной фазы (а) и начальной фазы (б)

Отметим, что алгоритм автоматически интерпретирует начальную фазу $\varepsilon = \pi/2$ (начальное условие) как $-\pi/2$ или $3\pi/2$ с обеспечением

кратчайшего пути сходимости к истинному значению начальной фазы $\varepsilon = -1,5$ рад и $\pm 2,5$ рад с учетом начального условия.

Погрешности $\delta\varepsilon$ определения начальной фазы показаны на рисунке 3.12. Погрешность фазы становится меньше, чем $2\pi/100$ после примерно 20-ти шагов и становится пренебрежимо малой после обработки отсчетов на половине периода сигнала.

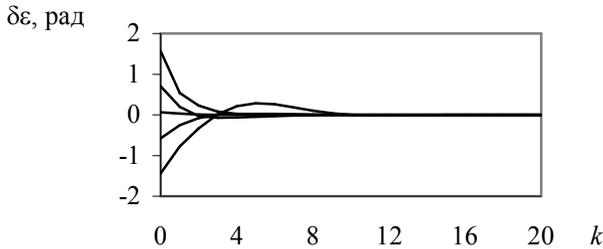


Рисунок 3.12. Зависимость погрешности оценивания начальной фазы от количества шагов обработки

Рекуррентные алгоритмы обработки интерференционных полос в динамических интерферометрических системах основываются на использовании формализма стохастических разностных уравнений, при этом априорное знание о свойствах системы учитывается в модели (3.22). Рекуррентные параметрические методы носят специализированный характер и учетом особенностей конкретной решаемой задачи и поэтому имеют определенные преимущества в точности, помехоустойчивости и быстродействии. Достижимые погрешности оценивания фазы интерференционных полос весьма малы, что соответствует требованиям при решении многих задач высокоточного бесконтактного контроля различных объектов.

Контрольные вопросы

1. Приведите формулу и объясните понятие аппаратной функции интерферометра.
2. В чем состоит принцип построения многоканальной интерферометрической системы?
3. К какому виду систем можно отнести интерферометрическую систему формирования и регистрации спекл-интерференционных полей?

4. Приведите схему и объясните принцип функционирования интерферометрической системы со спектральным разделением каналов.
5. Приведите примеры интерферометрических систем с неподвижной и с движущейся интерференционной картиной.
6. Какие основные процедуры предусматриваются при формировании и обработке интерферометрических данных?
7. Перечислите основные факторы, которые требуется учитывать при построении модели интерферометрического сигнала.
8. Объясните сущность понятия динамической интерферометрической системы.
9. Приведите схему и объясните принцип идентификации параметров динамической системы.
10. В чем состоит принцип построения и функционирования интерферометра управляемого фазового сдвига как динамической системы?

Список литературы

- 3.1. Васильев В.Н., Гуров И.П. Компьютерная обработка сигналов в приложении к интерферометрическим системам. – СПб: БХВ-Санкт-Петербург, 1998.
- 3.2. Malacara D., Servín M., Malacara Z. Interferogram analysis for optical testing. NY: Taylor & Fransis, LLC, 2005.
- 3.3 Гуров И.П. Формирование и анализ сигналов в системах компьютерной фотоники. – СПб: Университет ИТМО, 2018.
- 3.4. Ван Кампен Н.Г. Стохастические процессы в физике и химии. – М.: Высшая школа, 1990.

4 Оптические системы биомедицинской диагностики

4.1 Системы бесконтактного контроля биологических объектов

Методы биомедицинской диагностики основываются на различных физических принципах, при этом оптические методы [4.1] обеспечивают возможность бесконтактного излучения большого объема информации при наблюдениях исследуемого участка объекта с различной степенью увеличения.

Принцип построения системы оптической диагностики иллюстрируется схемой, показанной на рисунке 4.1.

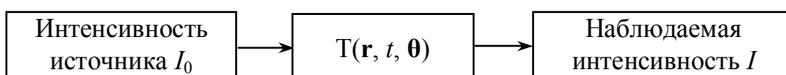


Рисунок 4.1. Структура «вход – выход» системы биомедицинской диагностики

Излучение источника с интенсивностью I_0 преобразуется оператором системы $T(\mathbf{r}, t, \boldsymbol{\theta})$, где \mathbf{r} – вектор, определяющий пространственное положение точки объекта в пространстве, t – время, $\boldsymbol{\theta}$ – вектор параметров, характеризующих свойства объекта, с формированием наблюдаемой интенсивности излучения I на выходе системы. При этом задача состоит в идентификации оператора системы, характеризующего объект.

На начальном этапе исследований проводится идентификация оператора системы в широком смысле, когда осуществляются адекватный выбор вида структуры (оператора) системы и задается класс моделей, описывающих объект, с определением степени линейности и стационарности наблюдаемого процесса, определения вида влияния входных переменных на выходные, выбора информативных переменных и т.д.

Рассмотрим в качестве примера задачу диагностики состояния кровеносной системы методом видеокапилляроскопии [4.2] (см. рисунок 4.2).

Участок объекта (поверхности кожи) освещается источником излучения с равномерным пространственным распределением интенсивности $I(x, y)$, где (x, y) – координаты в плоскости объекта, и энергетическим спектром $G(\lambda)$, λ – длина волны излучения. Отраженное излучение при помощи объектива направляется на матричный фотодетектор с координатами пикселей (i, j) .

Зарегистрированный фотоэлектрический сигнал $s(i, j)$ содержит информацию об исследуемом объекте.

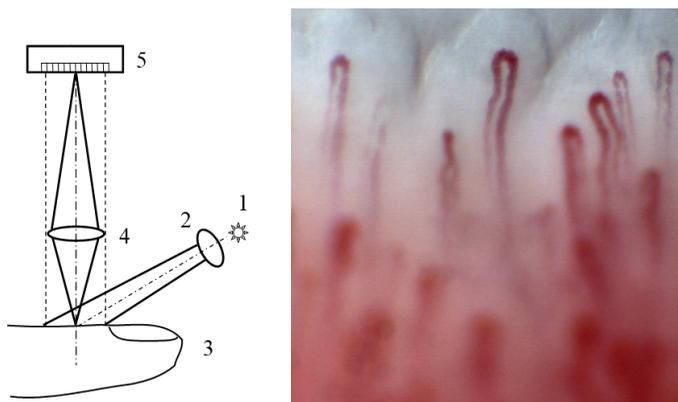


Рисунок 4.2. Оптическая система (а) и пример наблюдаемого изображения капилляров ногтевого ложа (размер 250 мкм × 250 мкм). 1 – источник излучения, 2 – линза, 3 – объект, 4 – проекционный объектив, 5 – светочувствительная матрица [4.3]

Задача идентификации в широком смысле состоит в определении энергетического спектра $G(\lambda)$ излучения, в пределах которого свойства изучаемого объекта наиболее выражены, и описании общих свойств объекта, параметры которого требуется идентифицировать, в форме математической модели.

На рисунке 4.3 показаны кривые, характеризующие степень поглощения оптического излучения в капиллярах в зависимости от длины волн.

Из рисунка видно, что коэффициент поглощения существенно снижается при длинах волн более 600 нм. Отмечены участки спектра, которые используются в серийно выпускаемых оптических устройствах для определения пульса в кровеносной системе.

Применительно к задачам видеокапилляроскопии установлено, что наблюдение капилляров целесообразно проводить при освещении зеленым цветом (длина волны около 530 нм), когда контраст капилляров на фоне окружающей биоткани максимален ввиду поглощения излучения эритроцитами. В диапазоне меньших длин волн (синий цвет) излучение проникает только на весьма малую глубину ввиду сильного поглощения эритроцитами (в несколько раз большего, чем при зеленом цвете, см. рис. 4.3 и [4.4]) и сильного

рассеяния. Поэтому модуляция излучения зеленого цвета более выражена, что определяет его предпочтительное использование в видеокапилляроскопии.

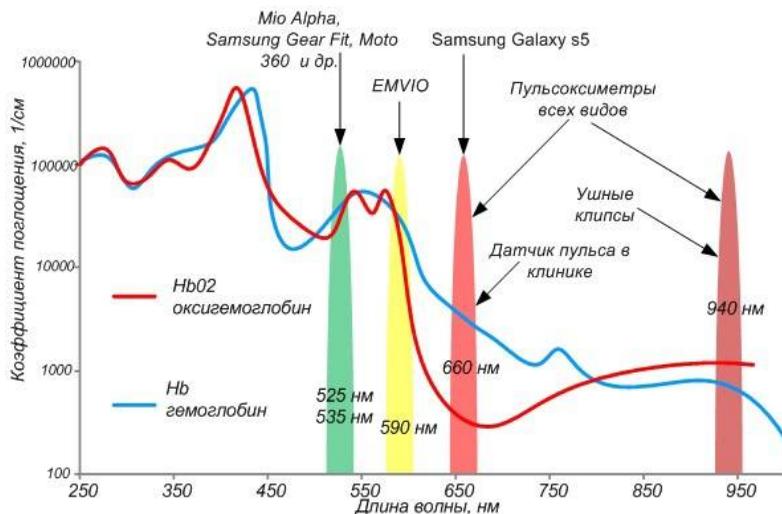


Рисунок 4.3. Зависимость коэффициента поглощения крови от длины волны излучения. Mio ALPHA – устройство определения пульса кровеносных сосудов. EMVIO – браслет для оценок и контроля стресса

Для диагностики ряда заболеваний требуется выделять линии капилляров с определением формы капилляров и плотности капилляров в капиллярной сети [4.5] (см. рисунок 4.4).

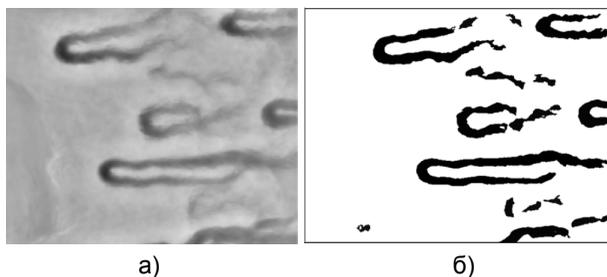


Рисунок 4.4. Примеры изображений структуры капиллярной сети: исходное изображение (а) и выделенные области капилляров (б)

В этом случае осуществляется идентификация оператора $T(x, y, \theta)$, не зависящего от времени, с определением параметров капилляров θ .

Рассмотрение системы «вход-выход» в реальных условиях наблюдений должно учитывать неконтролируемые произвольные смещения наблюдаемого живого объекта, как это представлено схемой на рисунке 4.5.

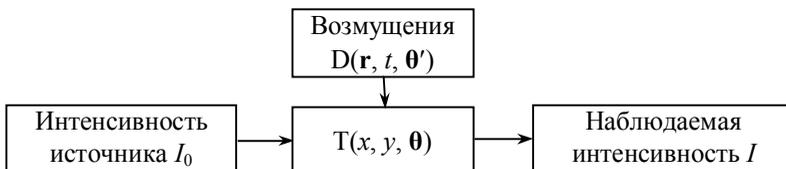


Рисунок 4.5. Структура системы «вход – выход» в условиях возмущений

Оператор возмущений $D(\mathbf{r}, t, \theta')$ учитывает случайные изменения объекта во времени с параметрами θ' . Вектор параметров θ' включает направления и скорость смещений объекта в течение интервала наблюдения. Оператор системы принимает вид

$$T_D(x, y, \theta) = T(x, y, \theta) \otimes d(x, y), \quad (4.1)$$

где $\delta[x - x_d(t), y - y_d(t)]$ – дельта-функция, $x_d(t)$ и $y_d(t)$ – случайные величины с функцией совместной плотности вероятности $p(x_d, y_d)$, характеризующие смещения объекта, \otimes – символ свертки. Ввиду того, что свертка представляет собой интегральное преобразование, приводящее в течение промежутка времени экспозиции к размытию представления исследуемого объекта, для идентификации истинного оператора системы $T(x, y, \theta)$ по наблюдаемому $T_D(x, y, \theta)$ в (4.1) требуется выполнить процедуру обратной свертки. Методы обратной свертки подробно рассмотрены в учебном пособии [4.6].

4.2 Оптическая диагностика состояния кровеносной системы

Важная задача биомедицинской диагностики состоит в определении скорости движения эритроцитов в капиллярах [4.3], [4.7]. При этом система диагностики относится к классу динамических систем.

Зарегистрированная видеопоследовательность $s(i, j, k)$, где (i, j) – координаты пикселя, k – номер видеокadra, $k = 0, \dots, K - 1$, может быть представлена в виде

$$s(i, j, k) = \mu I(i, j, k) \otimes d(i, j, k) + n(i, j), \quad (4.2)$$

где $I(i, j, k)$ – распределение интенсивности в изображении, μ – коэффициент преобразования интенсивности в фотоэлектрический сигнал, $n(i, j)$ – составляющая шума наблюдения с нулевым средним значением, \otimes – символ свертки,

$$d(i, j, k) = \delta[i - i_d(k), j - j_d(k)], \quad (4.3)$$

$\delta[i - i_d(k), j - j_d(k)]$ – дельта-функция, $i_d(k)$ и $j_d(k)$ – случайные величины с функцией совместной плотности вероятности $p(i_d, j_d)$, характеризующие «дрожание» изображений в последовательности видеок кадров.

Интенсивность элемента изображения капиллярной сети содержит две составляющие,

$$I(i, j, k) = \tilde{I}(i, j, k) + I_b(i, j, k), \quad (4.4)$$

где $\tilde{I}(i, j, k)$ – полезная составляющая, соответствующая движущимся эритроцитам, и $I_b(i, j, k)$ – фоновая составляющая. Следовательно, видеопоследовательность (4.2) может быть представлена выражением

$$s(i, j, k) = [\tilde{s}(i, j, k) + b(i, j, k)] \otimes d(i, j, k) + n(i, j), \quad (4.5)$$

где $\tilde{s}(i, j, k) = \mu \tilde{I}(i, j, k)$, $b(i, j, k) = \mu I_b(i, j, k)$. Для выделения полезного сигнала $\tilde{s}(i, j, k)$ фоновая составляющая должна быть удалена, как и составляющая шума наблюдения $n(i, j)$.

Оценка и удаление шума наблюдения основываются на предположении, что пространственно-частотный спектр шума относится главным образом к области высоких частот по отношению к полезной составляющей изображения. При этом шум наблюдения может быть оценен путем вычитания усредненной и обработанной фильтром нижних частот составляющей интенсивности в видеопоследовательности. Полученная оценка шума затем вычитается из каждого кадра записанной видеопоследовательности.

Оценка фоновой составляющей является более сложной задачей, поскольку изменчивость фона при наблюдении живых биологических объектов проявляется не только внутри видеок кадра, но и от кадра к кадру. Однако в целом можно принять, что пространственные изменения фона более медленные по сравнению с полезной составляющей сигнала.

Исследования показали целесообразность использования следующей процедуры итерационной обработки для получения оценки фоновой составляющей. Вначале записанное изображение $s(i, j)$ сглаживается путем свертки с заданной функцией $g(i, j)$ гауссовой формы. Результат принимается как начальное условие для последующих итераций

$$b(i, j, q) = \bar{b}(i, j, q-1) \otimes g(i, j), \quad (4.6)$$

где q – номер итерации. Значение $\bar{b}(i, j, q-1) = \max\{b(i, j, q), b(i, j, q-1)\}$ определяется на каждом шаге итерации. В результате точки с большим значением сигнала остаются неизменными, в то время как значения сигнала в других точках увеличиваются при размытии контрастных деталей изображения, как это иллюстрируется на рисунке 4.6, б–г.

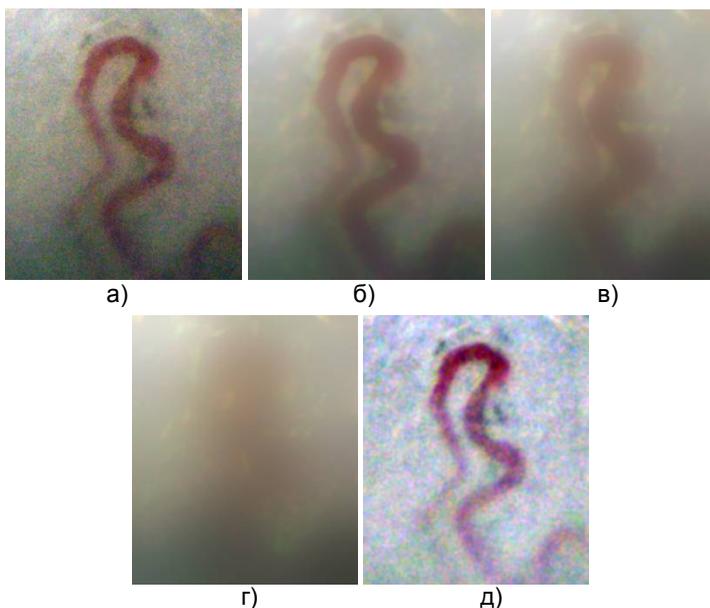


Рисунок 4.6. Исходное изображение капилляра с удалённым шумом наблюдения (а), оценки фона на первой, восьмой и заключительной итерациях (б–г), соответственно, и улучшенное изображение с удалённой фоновой составляющей (д).

Итерационная обработка прекращается при выполнении условия $\sum_{i,j} |b(i, j, q) - b(i, j, q-1)| < \varepsilon$, где ε – малое пороговое значение. После

вычитания фоновой составляющей изображение приобретает вид, показанный на рисунке 4.4, д.

Рассмотренные выше процедуры предварительной обработки каждого кадра в видеопоследовательности необходимы для последующего точного и надежного совмещения изображений.

При определении скорости движения эритроцитов в капиллярах необходимо учитывать следующую априорную информацию.

- Морфология капиллярной сети, форма и размеры объекта (отдельного капилляра), как и его положение, заранее не известны.
- Вследствие высокого рассеяния и поглощения оптического излучения наблюдаемое распределение интенсивности неоднородно и зависит от исследуемого участка.
- Поток эритроцитов внутри капилляра может содержать (или не содержать) промежутки плазмы на различных интервалах наблюдения.
- Эритроциты в капиллярах обычно объединяются в изменчивые агрегаты, при этом отдельный эритроцит перемещается со скоростью, немного отличающейся от скорости соседних эритроцитов.
- Капилляры наблюдаются на фоне окружающей биоткани с различной структурой в зависимости от исследуемого участка.
- Наблюдаемый участок испытывает случайные смещения по отношению к полю зрения оптической системы с произвольными направлением и величиной вектора скорости.
- Пульсации крови приводят к модуляции отраженного излучения от капилляров и от окружающей биоткани на частоте биений сердца.

Ввиду того, что исследуемый объект изменчив по своей природе, в системе диагностики требуется выполнить следующие условия: точное совмещение изображений в зарегистрированной видеопоследовательности; выделение и оценивание геометрических параметров каждого индивидуального капилляра с высоким разрешением и помехоустойчивостью; надежное выделение границ капилляра и определение центральной линии капилляра; высокое быстродействие при записи, достаточное для временного разрешения отдельного движущегося эритроцита.

С позиций системного анализа идентификации подлежит априорно неопределенный объект, наблюдаемый в изменчивых условиях влияния помех. При этом система диагностики осуществляет процедуры *обнаружения* объекта (капилляра) на сложном нестационарном фоне, *компенсацию* изменяющихся внешних возмущений, *оценивание параметров* геометрической структуры капиллярной сети и отдельного капилляра, *измерение* скорости эритроцитов.

Структурная схема системы идентификации движущихся эритроцитов представлена на рисунке 4.7.



Рисунок 4.7. Процедуры идентификации объекта

Особенности взаимодействия оптического излучения с биологическими средами, в том числе в условиях различных влияющих факторов, рассмотрены в [4.1]. Дискретные представления оптических сигналов описаны в [4.6]. Обратимся к рассмотрению процедур обнаружения объектов (капилляров), оценивания их параметров и определения скорости эритроцитов в капиллярах, следуя методике, подробно представленной в [4.7].

Как отмечалось выше, при регистрации видеопоследовательности видеокadres оказываются смещенными друг относительно друга вследствие естественных движений живого биологического объекта. Для наблюдения движущихся эритроцитов в капиллярах осуществляется совмещение изображений в видеопоследовательности методом компенсации взаимных смещений видеокadres. Наиболее эффективные методы совмещения основаны на принципе оценивания локальной взаимной корреляции изображений [4.3].

Обнаружение капилляров

Определение наличия области капилляра на фоне окружающей биоткани возможно при использовании априорной информации,

закрывающейся в том, что поглощение излучения капиллярами отличается от поглощения на других участках, как это наглядно видно на рисунке 4.4, а. Однако выделение капилляра на единственном видеокadre не обеспечивает надежного определения границ капилляра (см. рисунок 4.4, б) ввиду локальной изменчивости свойств биоткани. В отдельных точках отличие капилляра и окружающей биоткани незначительно, что приводит к утрате информации. Отметим, что свойства биоткани весьма изменчивы от участка к участку и для различных субъектов, поэтому не существует алгоритмов, позволяющих надежно выделить границы капилляров на основе критерия связности области капилляра.

Дополнительная априорная информация состоит в том, что капилляры содержат движущиеся эритроциты, совокупность которых представляет собой пространственно распределенный изменчивый динамический объект. Использование этой информации позволяет надежно выделить области капилляров, однако для этого требуется обработка набора видеокadre, зарегистрированных с достаточно высоким быстродействием, позволяющим обеспечить разрешение отдельных движущихся эритроцитов.

На рисунке 4.8, а–д показан набор отнормированных и совмещенных изображений с использованием методики, рассмотренной выше (см. рисунок 4.6) и детально описанной в [4.3].

Зарегистрированный набор изображений $I(i, j, k)$, где (i, j) – дискретные координаты пикселей, $k = 1, \dots, K$ – номер видеокadre, усредняется по нескольким десяткам кадров, предшествующих k -му видеокadre и следующим за ним, с вычислением усредненного значения для этого видеокadre

$$B(i, j, k) = \frac{1}{2l+1} \sum_{k'=k-l}^{k'+l} I(i, j, k'), \quad (4.7)$$

где $2l+1$ – количество усредненных видеокadre. Как видно на рисунке 4.8, а–д, положения областей капилляров совмещены, однако несмотря на усреднение значения $B(i, j, k)$ несколько отличаются друг от друга в некоторых точках вследствие значительной изменчивости изображений в видеопоследовательности.

Средние значения (4.7) вычитаются из каждого видеокadre. Полученные разности

$$\tilde{s}(i, j, k) = I(i, j, k) - B(i, j, k) \quad (4.8)$$

показаны на рисунке 4.8, а'–д'. Видны отдельные движущиеся эритроциты и их агрегаты в различных положениях внутри капилляра.

Светлый участок внутри капилляра представляет собой плазму крови, не содержащую эритроцитов и перемещающуюся вместе с эритроцитами.

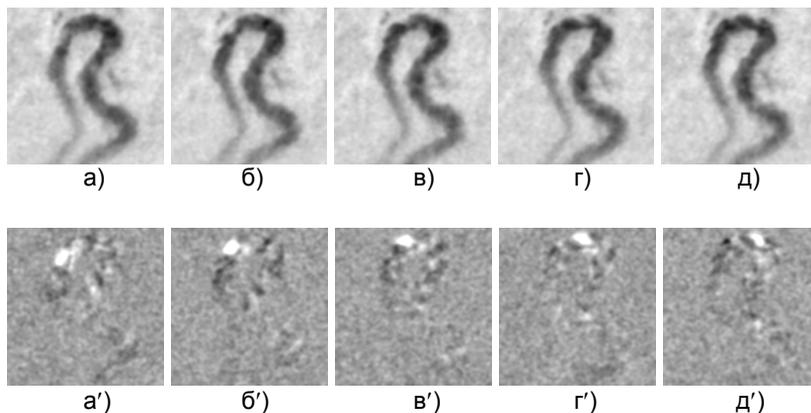


Рисунок 4.8. Нормированные усредненные изображения $B(i, j, k)$, относящиеся к различным участкам видеопоследовательности (различным значениям k) (а–д) и информативные составляющие изображений, представленные в полутоновой шкале интенсивности, после вычитания фоновой составляющей (а'-д') [4.7]

Для обнаружения областей капилляров вычисляется дисперсия $\sigma^2(i, j)$ изменения сигнала в каждом пикселе (i, j) по последовательности видеокадров в форме

$$\sigma^2(i, j) = \frac{1}{K} \sum_{k=1}^K [\tilde{s}(i, j, k) - \langle \tilde{s}(i, j, k) \rangle]^2, \quad (4.9)$$

где $\langle \tilde{s}(i, j, k) \rangle$ – усредненное изображение по всей видеопоследовательности, содержащей K видеокадров. Значение (4.9) характеризует общую динамическую изменчивость отраженного излучения в каждом пикселе (i, j) .

Полученные согласно (4.9) значения иллюстрируются на рисунке 4.9, а. Видно, что область капилляра четко выделена. Однако внутри области значения параметра (4.9) изменяются вследствие искажений и дефектов распределения интенсивности, что не позволяет надежно выделить границы капилляра.

Данный недостаток устраняется при нормировке на среднее значение каждого из видеок кадров $\tilde{s}(i, j, k)$ в (4.9) и вычитания фоновой составляющей, что позволяет получить улучшенную оценку области капилляра (см. рисунок 4.9, б).

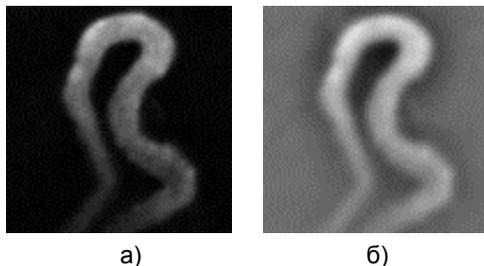


Рисунок 4.9. Оценки дисперсии отклонений интенсивности в видеопоследовательности для различных точек области капилляра (а) и улучшенные оценки при дополнительной обработке (б)

Рассмотренные выше процедуры устранения влияния дефектов и шума в изображениях позволяют применить простое пороговое правило в форме $\sigma^2(\hat{i}, \hat{j}) > \sigma_{\text{п}}^2$, где (\hat{i}, \hat{j}) – координаты пикселей, в которых дисперсия (4.9) превышает заданное фиксированное значение $\sigma_{\text{п}}^2$, для определения границ капилляров, т.е. выполнить сегментацию изображения, показанного на рисунке 4.9, б, по признаку превышения заданного порогового значения.

Определение геометрических характеристик капилляров

На рисунке 4.10, а показана выделенная область капилляра. Однако данный результат промежуточный и требует дополнительного уточнения, основанного на идентификации движущихся эритроцитов в капилляре.

Компоненты вектора скорости оцениваются как

$$v_x(i, j, k) = f\Delta i(k), \quad (4.10)$$

где f – частота видеок кадров, $\Delta i(k)$ – перемещение точки (i, j) между соседними кадрами k и $k-1$, т.е. локальная мгновенная скорость вдоль координаты X (дискретная координата i) и

$$v_y(i, j, k) = f\Delta j(k), \quad (4.11)$$

$\Delta j(k)$ – перемещение точки (i, j) по координате Y (дискретная координата j).

Перемещения $\Delta i(k)$ и $\Delta j(k)$ оцениваются согласно выражению

$$G(i, j, \Delta i, \Delta j, k) = \arg \min_{\Delta i, \Delta j} \sum_{k=1}^K [\tilde{s}_R(i, j, k) - \tilde{s}_R(i + \Delta i, j + \Delta j, k - 1)]^2, \quad (4.12)$$

где значения $\tilde{s}_R(i, j, k)$ вычисляются из соотношения (4.8) для точек $(i, j) \in R$, R – маска в форме круга малого диаметра внутри области капилляра с центром в пикселе (i, j) . Преобразование (4.12) представляет собой оценку двумерной пространственной взаимной корреляции между точками потока эритроцитов в соседних видеокдрах видеопоследовательности. В результате формируются оценки сдвигов $(\Delta i, \Delta j)$ взаимно коррелированных точек изображения, соответствующих потоку эритроцитов.

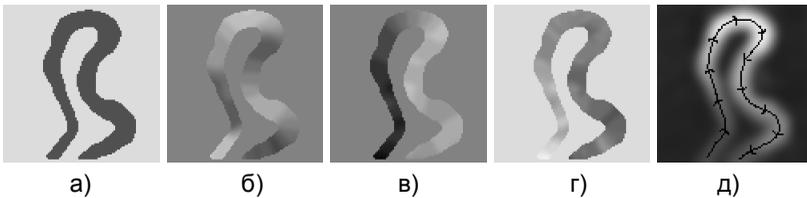


Рисунок 4.10. Оценка области капилляра (а), пространственное распределение скорости эритроцитов (в полутоновом представлении) вдоль горизонтального направления (координата i) (б), пространственное распределение скорости эритроцитов вдоль вертикального направления (координата j) (в), оценки модуля скорости (г), оценки локального направления скорости вдоль центральной линии капилляра (д) [4.7]

Полученные согласно (4.10) и (4.11) значения пространственных изменений компонентов вектора скорости $v_x(i, j, k)$ и $v_y(i, j, k)$ иллюстрируются в полутоновом представлении на рисунке 4.10, б–в (для фиксированного k -го видеокдра). На рисунке 4.10, г–д показаны соответствующие значения модуля скорости и локальные направления вектора скорости вдоль центральной линии капилляра.

Рассмотренные выше процедуры обеспечивают надежное определение области капилляра, в которой модуль скорости потока эритроцитов отличен от нуля, а также центральной линии капилляра, вдоль которой модуль скорости принимает максимальные значения в точках вдоль центральной линии, отмеченных на рисунке 4.10, д,

применительно к капиллярам, имеющим индивидуальные (изменяющиеся) геометрические характеристики.

Предварительно выделенная центральная линия капилляра (см. рисунок 4.11, а) сглаживается методом вычисления локального скользящего среднего вдоль центральной линии капилляра (рисунок 4.11, б). В результате центральная линия капилляра точно соответствует области капилляра (рисунок 4.11, в). Одновременно уточняются локальные значения диаметра капилляра, как показано на рисунке 4.11, г. Центральная линия капилляра разделяется на равные отрезки (рисунок 4.11, д).

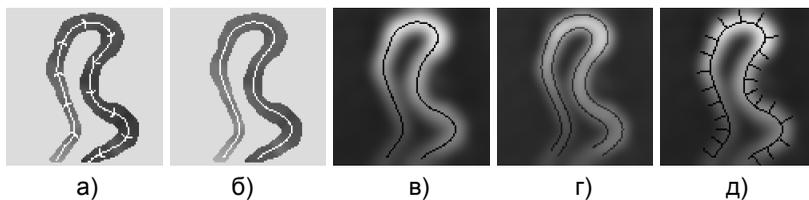


Рисунок 4.11. Предварительная оценка центральной линии капилляра (а), сглаженная центральная линия капилляра (б), уточненное положение центральной линии в области капилляра, показанной на рисунке 4.9, б (в), оценки локального диаметра капилляра (г) и равномерно распределенные отрезки вдоль центральной линии капилляра (д) [4.7]

Рассмотренные выше преобразования позволяют надежно определить область капилляра и его локальный диаметр, а также получить оценки скорости потока капилляров вдоль центральной линии капилляра. Следует отметить, что описанные выше процедуры необходимы ввиду значительной изменчивости индивидуальных свойств капилляров и их морфологии в каждом отдельном эксперименте.

Определение скорости потока эритроцитов

Определение центральной линии капилляра позволяет выполнить преобразование координат и представить поток эритроцитов как движущийся вдоль прямой линии, как это иллюстрируется на рисунке 4.12, т.е. привести двумерную задачу к одномерной.

Из рисунка 4.12 видно, что пространственное распределение отраженного излучения, соответствующее микроструктуре потока крови и определяемое эритроцитами и их агрегатами, смещается слева

направо при одновременных изменениях деталей этой структуры: некоторые пики появляются или исчезают, становятся уже или шире и т.п.

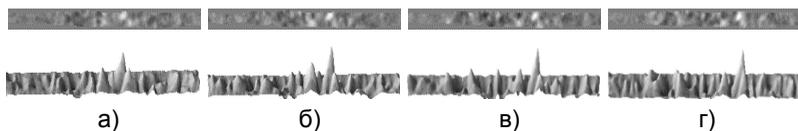


Рисунок 4.12. Представление смещений во времени потока эритроцитов и их агрегатов вдоль прямой линии (а)–(г) в форме полутонового отображения (верхний ряд) и трехмерной визуализации интенсивности отраженного излучения (нижний ряд)

Совместная обработка данных, показанных в верхней части рисунка, позволяет провести прямые измерения скорости потока крови при определении смещения в течение заданного промежутка времени. Для решения этой задачи используется корреляционный метод, позволяющий определять взаимный сдвиг сигналов. Поскольку сигналы регистрируются при помощи видеокамеры с заданной частотой кадров, известен промежуток времени, в течение которого произошел пространственный сдвиг. Скорость определяется отношением этого сдвига к соответствующему ему промежутку времени.

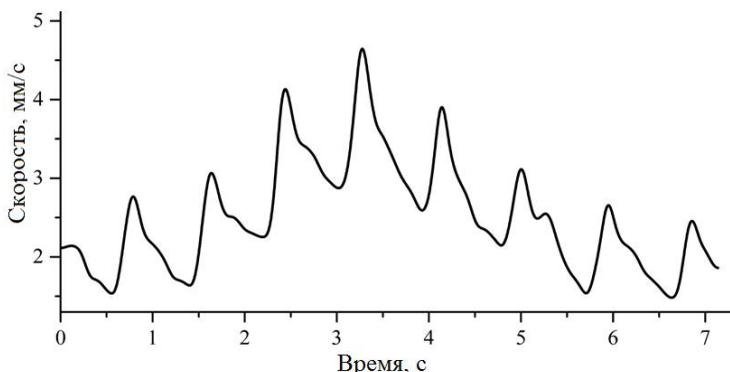


Рисунок 4.13. Пример результатов прямого измерения скорости эритроцитов в капилляре

Рассмотренные выше процедуры обработки данных позволяют определить изменения скорости движения эритроцитов в капиллярах. Пример полученных экспериментально оценок скорости

иллюстрируется на рисунке 4.13. На рисунке видны циклы изменения скорости, соответствующие биениям сердца, а также низкочастотная составляющая, соответствующая респираторному циклу.

Определение свойств потока крови в капиллярной сети важно для решения проблем диагностики состояния кровеносной системы. Поэтому рассмотренная выше методика была применена для оценивания скорости движения эритроцитов одновременно в нескольких капиллярах на наблюдаемом участке. Результаты представлены на рисунке 4.14.

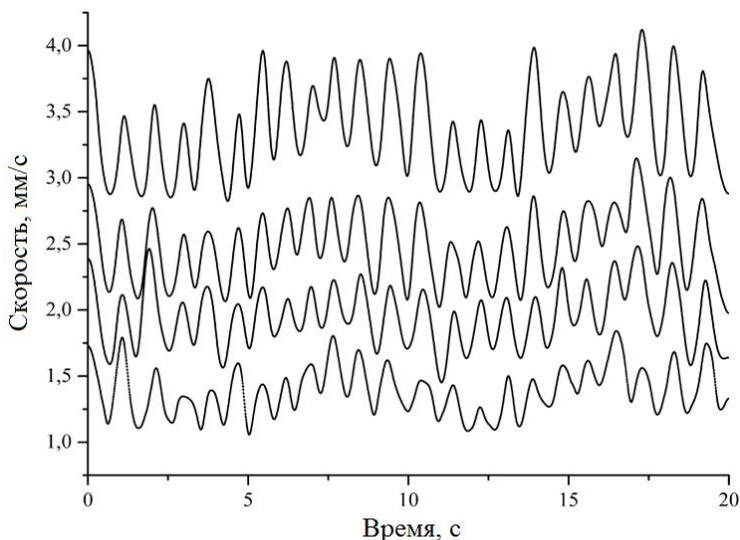


Рисунок 4.14. Пример результатов прямого измерения скорости эритроцитов в капиллярной сети [4.7]

Из рисунка видно, что скорость движения эритроцитов существенно различна в разных капиллярах, однако периоды изменения скорости во всех капиллярах синхронны поскольку обусловлены биениями сердца.

Отметим, что в зависимости, показанные на рисунке 4.14, получены с использованием процедур усреднения, рассмотренных выше, поэтому кривые выглядят сглаженными. Однако при использовании рассмотренного метода возможно получение информации о скорости эритроцитов внутри цикла сердечного сокращения с более высоким разрешением, как показано на рисунке

4.13. С другой стороны, кривые на рисунке 4.14 содержат также низкочастотные составляющие, которые также синхронизированы.

Таким образом, метод позволяет получать данные с высоким пространственным разрешением и регулируемой степенью разрешения во времени, которое устанавливается в зависимости от конкретной решаемой задачи диагностики кровеносной системы.

4.3 Системы оптической когерентной микроскопии

Методы оптической когерентной томографии позволяют получать информацию о внутренней микроструктуре частично прозрачных объектов, в том числе неоднородных и слоистых сред с высоким разрешением.

При исследованиях микрообъектов используются методы интерференционной микроскопии. Начиная с публикации [4.8] по проблематике отображения сечений внутренней микроструктуры биологических объектов данный подход носит название *оптической когерентной томографии* (ОКТ). В случае использования объективов с большой числовой апертурой, что обеспечивает повышенное разрешение в боковом направлении, данный метод принято называть *оптической когерентной микроскопией* (ОКМ). Ввиду того, что не существует четкой границы между ОКТ и ОКМ далее используются оба термина.

Формирование когерентных световых полей в нескольких длинах волн

Основные особенности формирования сигналов в ОКТ кратко изложены в учебном пособии [4.9]. Обратимся к рассмотрению методов ОКТ с позиций системного подхода.

При освещении объекта монохроматическим излучением с комплексной амплитудой $E_1(x, y, z)$ в некоторой точке наблюдения $P_0(x, y, z)$ формируется поле с комплексной амплитудой

$$A(x, y, z) = \sum_{n=1}^N \frac{1}{\sqrt{N}} a_n(x, y, z) = \frac{1}{\sqrt{N}} \sum_{n=1}^N |a_n| \exp(j\psi_n), \quad (4.13)$$

где N – количество точек объекта, в которых происходит отражение. В этом случае имеем структурную схему со многими входами и многими выходами, показанную на рисунке 3.5. Очевидно, что по единственному наблюдаемому значению $A(x, y, z)$ невозможно однозначно определить параметры a_n и ψ_n . Поэтому для того, чтобы

найти a_n и ψ_n нужно иметь как минимум $2N$ независимых уравнений, для получения которых можно использовать набор длин волн источника излучения, освещающего объект.

Применительно к системам ОКТ обычно предполагается однократное отражение от каждой точки микроструктуры объекта вдоль линии A -скана [4.9]. При этом структурная схема системы имеет вид в соответствии с рисунком 3.4, т.е. представляет собой многоканальную систему с пространственно разнесенными каналами.

Обратимся к рассмотрению систем ОКТ с учетом использования для освещения объекта источника с набором различных длин волн.

Как отмечалось ранее в разделе 3, электрическое поле измерительной и опорной волн в интерферометре (см. рисунок 3.1) выражается в форме

$$E_1(v, t) = \sqrt{\alpha} E_0(v, t), \quad (4.14)$$

$$E_2(v, t, \tau) = \sqrt{\beta} E_0(v, t) \exp(-j2\pi\nu\tau), \quad (4.15)$$

где α и β – коэффициенты отражения / пропускания в светоделителе по интенсивности, $E_0(v, t) = A(v) \exp(-j2\pi\nu t)$ представляет оптические колебания источника излучения с амплитудным спектром $|A(v)|$, $\tau = \Delta/c$ – время задержки опорной волны по отношению к измерительной волне при оптической разности хода (ОРХ) $\Delta = 2n_s z$, n_s – показатель преломления среды, c – скорость света.

Интенсивность интерференционных полос согласно (3.1) определяется в форме

$$I(v, \tau) = \langle EE^* \rangle = \langle E_1^2 \rangle + \langle E_2^2 \rangle + 2 \operatorname{Re} \langle E_1 E_2^* \rangle, \quad (4.16)$$

где $E = E_1 + E_2$ и угловые скобки обозначают усреднение во времени. Ввиду использования практически доступных фотодетекторов с ограниченным быстродействием время усреднения в (4.16) значительно превышает период оптических колебаний $1/\nu$, при этом из (4.14)–(4.16) получаем выражение

$$I(v, \tau) = I_1(v) + I_2(v) + 2\sqrt{I_1(v)I_2(v)} \cos 2\pi\nu\tau, \quad (4.17)$$

где

$$I_1(v) = \langle A_1(v)A_1^*(v) \rangle, \quad I_2(v) = \langle A_2(v)A_2^*(v) \rangle, \quad (4.18)$$

$A_1(v)$ и $A_2(v)$ – комплексные амплитуды интерферирующих волн на выходе интерферометра. Уравнение (4.17) выражает известный закон

спектральной интерференции (см., например, [4.10]) в упрощенной форме.

Первое и второе слагаемые в (4.17) соответствуют спектру источника излучения, третье слагаемое – автокорреляционная функция гармонического колебания с частотой ν . При таком рассмотрении излучение источника интерпретируется как состоящее из гармонических колебаний с известными амплитудами $|A(\nu)|$.

Вследствие того, что спектр излучения источника известен априори, оказывается возможным найти из (4.17) полезное третье слагаемое $\tilde{I}(\nu, \tau)$ в (4.17) в форме

$$\tilde{I}(\nu, \tau) \propto |A(\nu)|^2 \cos 2\pi\nu\tau. \quad (4.19)$$

В случае N отражений это выражение принимает вид

$$\tilde{I}(\nu, \tau_n) \propto |A(\nu)|^2 \cos 2\pi\nu\tau_n. \quad (4.20)$$

При этом возможно определить задержку $\tau_n = \Delta_n / c$ измерительной волны по отношению к опорной волне (см. рисунок 4.15), соответствующую ОРХ Δ_n , при получении оценок интенсивности $\tilde{I}(\nu_k, \tau_n)$ на наборе различных длин волн $\lambda_k = c/\nu_k$, $k = 1, \dots, K$, $K > 2N$.

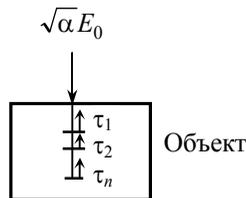


Рисунок 4.15. Отражения измерительной волны от границ слоев объекта с задержками $\tau_1, \tau_2, \dots, \tau_n$

При взаимодействии с объектом измерительная волна (4.14) изменяется и определяется в форме

$$E_1(\nu, t) = \sqrt{\alpha} E_0(\nu, t) H(\nu), \quad (4.21)$$

где $H(\nu)$ – комплексная частотная характеристика исследуемого образца. При этом выражение (4.17) принимает вид

$$I(\nu, \tau) = I_1(\nu) |H(\nu)|^2 + I_2(\nu) + 2\sqrt{I_1(\nu)I_2(\nu)} \operatorname{Re}\{H(\nu) \exp(j2\pi\nu\tau)\}, \quad (4.22)$$

откуда следует, что первое слагаемое зависит от свойств исследуемого объекта, и для наблюдения доступна только действительная часть комплексной частотной характеристики объекта.

Системы ОКТ во временной области

Интенсивность излучения (4.22) преобразуется в фотоэлектрический сигнал $s = \mu I$, где μ – коэффициент преобразования. При освещении объекта одновременно на всех K длинах волн и фотодетектировании сигнал интерференционных полос выражается как

$$s(\tau) = \mu \sum_{k=1}^K I(v_k, \tau) = B + 2\mu\sqrt{\alpha\beta} \operatorname{Re} V(\tau), \quad (4.23)$$

где

$$B = \mu \sum_{k=1}^K I_1(v_k) |H(v_k)|^2 + \mu \sum_{k=1}^K I_2(v_k) \quad (4.24)$$

– фоновая составляющая и

$$V(\tau) = \sum_{k=1}^K |A(v_k)|^2 H(v_k) \exp(j2\pi v_k \tau) \quad (4.25)$$

– параметр, характеризующий видность интерференционных полос.

Соотношение (4.25) представляет собой обратное преобразование Фурье от произведения двух функций. Согласно известной теореме о свертке это выражение может быть представлено во временной области в форме

$$V(\tau) = R(\tau) \otimes h(\tau), \quad (4.26)$$

где $R(\tau)$ – автокорреляционная функция излучения источника,

$$R(\tau) = \sum_{k=1}^K |A(v_k)|^2 \cos 2\pi v_k \tau, \quad (4.27)$$

$$h(\tau) = \sum_{k=1}^K H(v_k) \exp(j2\pi v_k \tau) \quad (4.28)$$

– импульсная характеристика исследуемого образца. При использовании источника излучения с широким спектром, содержащим много длин волн и имеющим узкую корреляционную функцию (4.27), свертка (4.26) приводит к соотношению $V(\tau) \cong h(\tau)$, и

видность интерференционных полос представляет импульсную характеристику объекта $h(\tau)$.

Для описания свойств объекта при помощи импульсной характеристики (4.28) требуется определить видность интерференционных полос (4.25), изменяя во времени шаг за шагом положение горизонтального сечения объекта по всей глубине объекта. Толщина слоя в сечении объекта определяет ОРХ в интерферометре в пределах длины когерентности l_c , где существуют интерференционные полосы. Системы этого вида носят название *ОКТ во временной области*.

Преимущество систем ОКТ во временной области состоит в возможности фокусировки излучения в пределах длины когерентности, что позволяет использовать микрообъективы с увеличенной апертурой и обеспечить высокую разрешающую способность в плоскости наблюдения в системах класса ОКМ.

Системы ОКТ в частотной области

При регистрации интенсивности на каждой из длин волн в отдельности интерферометрический сигнал определяется выражением

$$s(v_k, \tau_n) = B(v_k) + 2\mu\sqrt{\alpha\beta}|A(v_k)|^2 \sum_{n=1}^N \operatorname{Re}\{ H(v_k) \exp(j2\pi v_k \tau_n) \}, \quad (4.29)$$

где

$$B(v_k) = \mu I_1(v_k) |H(v_k)|^2 + \mu I_2(v_k) \quad (4.30)$$

– фоновая составляющая, зависящая от частоты колебаний v_k во временной области, т.е. от длины волны λ_k . Операция суммирования в (4.29) обусловлена N одновременными отражениями внутри микроструктуры исследуемого объекта и может быть выражена в форме

$$\sum_{n=1}^N \operatorname{Re}\{ H(v_k) \exp(j2\pi v_k \tau_n) \} = \sum_{n=1}^N |H(v_k)| \cos[(j2\pi v_k \tau_n + \delta_k)], \quad (4.31)$$

где $H(v_k) = |H(v_k)| \exp(j\delta_k)$, δ_k – сдвиги фазы, вносимые объектом на длинах волн λ_k . Если эти сдвиги малы, то (4.31) можно интерпретировать как косинусное преобразование импульсной характеристики объекта $h(\tau_n)$.

Ввиду регистрации сигналов на различных частотах (длинах волн) системы рассматриваемого вида носят название *ОКТ в частотной области*.

Сравнение характеристик систем ОКТ во временной и частотной области показывает, что системы второго вида имеют преимущество в быстродействии благодаря одновременной регистрации отраженного излучения на всей глубине объекта, тогда как в ОКТ во временной области оценивается отражение по глубине последовательно слой за слоем, что обуславливает потери информации вне слоя и занимает большее время.

Проведенное рассмотрение позволяет выделить следующие важные особенности систем ОКТ.

Влияние количества длин волн K проявляется в ОКТ-системах во временной и в частотной области различным образом. В системах ОКТ во временной области это количество связано с шириной спектра и, следовательно, снижением ширины автокорреляционной функции (4.27) с увеличением K и уменьшением длины когерентности, где существуют интерференционные полосы. Время задержки τ рассматривается только в пределах дины когерентности.

В системах ОКТ в частотной области интерференционные полосы формируются на каждой длине волны при отражении предметной волны в пределах всей глубины проникновения излучения. Количество длин волн должно удовлетворять условию $K > 2N$, где N – количество отражений (на равных шагах по глубине), соответствующих задержкам τ_n .

В системах ОКТ во временной области фоновая составляющая (4.24) постоянна и не зависит от времени задержки τ в пределах длины когерентности излучения (однако обычно зависит от координаты z по глубине вследствие изменения степени отражения предметной волны). В системах ОКТ в частотной области фоновая составляющая (4.30) существенно изменяется в результате влияния частотной характеристики объекта и неравномерности спектра излучения источника. Эти особенности учитываются при обработке сигналов в ОКТ-системах.

Оптическая когерентная эластография

Оптическая когерентная эластография (ОКЭ, англ. ОСЕ) представляет собой современный метод визуализации степени эластичности биологических тканей и отдельных слоев и позволяет

оценивать механические свойства биоткани в микрометровом масштабе в трехмерном представлении, что не может быть достигнуто при помощи традиционных эластографических методов. Ввиду преимуществ ОКТ по разрешающей способности, а также повышенных требований к чувствительности при неразрушающем контроле биомеханических характеристик, эта новая технология получила активное развитие.

Область масштабов исследуемых объектов методом ОКЭ среди других методов эластографии представлена на рис. 4.16 [4.11]. Благодаря высокому пространственному разрешению в ОКТ, разрешение в ОКЭ лежит в диапазоне от нескольких единиц до сотен микрометров.

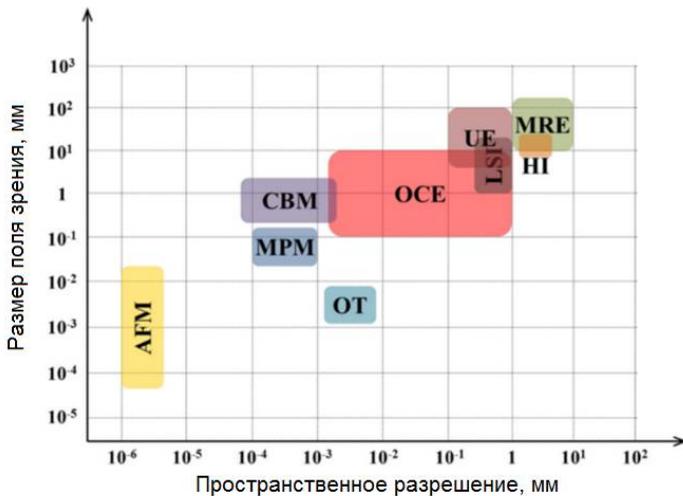


Рисунок 4.16. Область масштабов ОКЭ (OCE) и других методов неразрушающего контроля эластичности биоткани. AFM: атомно-силовая микроскопия, OT: оптический пинцет, MPM: многофотонная микроскопия, CBM: конфокальная микроскопия Бриллюэна, LSI: измерения на основе лазерных спеклов, UE: ультразвуковая эластография, HI: голографическая визуализация, MRE: магнитно-резонансная эластография

Динамическая ОКЭ первоначально основывалась на использовании систем ОКТ во временной области и для исследований кожи человека в естественных условиях с использованием пьезоэлектрического кольца [4.12], осуществляющего возбуждение механического напряжения в исследуемом образце. Метод фазовой

ОКЭ предложен в 2007 году [4.13], когда получили широкое распространение системы ОКТ в частотной области, в которых оценка фазы комплексных данных ОКТ обладает большей устойчивостью, чем в системах ОКТ во временной области. Деформация, скорость сжатия и растяжения мягких тканей были измерены в реальном времени в нанометровом масштабе. Данный метод использует информацию о фазе, доступную в комплексных ОКТ-изображениях, и позволяет оценить изменения фазы между последовательными *B*-сканами. При этом регистрируются мгновенные деформации биоткани.

Исследования продемонстрировали возможности использования ОКЭ для диагностики заболеваний без прямого терапевтического вмешательства, а также визуализации артериальных патологий в организме человека (см., например, [4.14]). Используя преимущества высокой скорости ОКТ-визуализации, система ОКЭ позволяет исследовать живые биоткани. Благодаря высокому разрешению, позволяющему различить отдельные клетки, при глубине проникновения в рассеивающие ткани до нескольких миллиметров ОКЭ обеспечивает уникальные возможности неинвазивного оценивания биомеханических свойств тканей в естественных условиях. Возможные клинические применения ОКЭ включают дифференциацию твердых и мягких масс во время биопсии, визуализацию состава артериальных бляшек и оценку процесса заживления ран.

В работе [4.14] представлена общая схема эластографического исследования, показанная на рисунке 4.17. Основными шагами являются:

- Определение структурных характеристик биоткани (изотропной или анизотропной) и создание соответствующей теоретической базы, учитывающей ограничения на возможные соотношения приложенных напряжений или сил сжатия или растяжения.
- Разработка соответствующих экспериментальных методов, в том числе обнаружение естественной деформации биоткани из-за влияния пульсаций потока крови и / или внешнего возбуждения на основе теоретических положений. Эластограммы смещений или деформаций затем могут быть получены с помощью методов медицинской визуализации, в том числе МРТ, ультразвука и ОКТ.
- Получение надежного соотношения между механическим напряжением и деформацией, а также карт эластичности. Этот процесс является итерационным и позволяет свести к минимуму различия между механическими напряжениями, оцененными при

помощи предложенной модели, и экспериментальными результатами.

Эластограммы смещения или напряжения получают, как правило, в результате непосредственной обработки изображений исследуемой структуры, полученных с помощью методов биомедицинской визуализации, таких как ультразвук или ОКТ, в то время как карты эластичности восстанавливаются при помощи дальнейших расчетов. Подобно ультразвуковой эластографии для деформации мягких тканей, как правило, применяется внешнее воздействие, например, акустические волны для стимуляции мягких тканей.

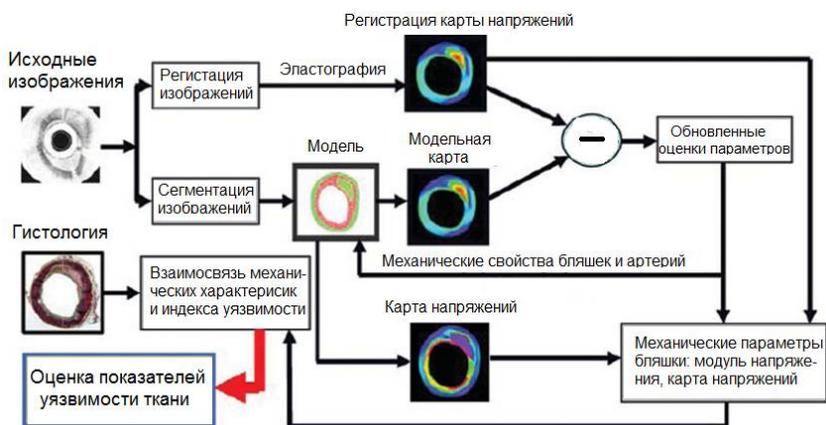


Рисунок 4.17. Схема ОКЭ и процедуры оценки параметров при исследовании артериальных бляшек [4.14]

Система ОКЭ включает в себя две основные подсистемы: систему ОКТ и систему возбуждения, используемую для деформации биологических тканей. Используются два типа стимулов: статические (квазистатические) и динамические.

Возможность определения характеристик биоткани, в частности, обнаружение патологических и физиологических изменений, определили развитие ОКЭ для многих приложений. Методы ОКЭ применяются для диагностики состояния семи основных видов биотканей, включая роговицу и хрусталик глаза, опухоли мягких тканей, стенки артерий, мышцы, кожу, одиночные клетки, а также нескольких других видов биотканей, таких как сухожилия и стенки дыхательных путей. Механические характеристики этих тканей имеют большое значение для доклинических исследований и

соответствующего клинического диагноза. Метод ОКЭ весьма продуктивен благодаря уникальным возможностям визуализации биомеханических характеристик биологических тканей.

Контрольные вопросы

1. Какие особенности объектов следует учитывать при построении оптической системы биомедицинской диагностики?
2. Приведите схему и объясните структуру системы «вход – выход» в условиях возмущений.
3. Каким образом учитываются в математической модели самопроизвольные смещения биологического объекта в процессе наблюдения?
4. На каких априорных предположениях основываются алгоритмы устранения шума наблюдения и фоновой составляющей в изображениях биологических объектов?
5. Какие процедуры анализа данных проводятся при диагностике состояния капилляров по набору зарегистрированных изображений?
6. Приведите схему и объясните структуру системы идентификации движущихся эритроцитов в капиллярах.
7. Назовите основные виды систем оптической когерентной томографии. В чем состоит различие методов оптической когерентной томографии и оптической когерентной микроскопии?
8. Приведите сравнение характеристик основных методов оптической когерентной томографии.
9. В чем состоит назначение систем оптической когерентной эластографии?
10. Приведите схему системы оптической когерентной эластографии и объясните процедуры оценивания параметров при исследовании характеристик биологических тканей.

Список литературы

- 4.1. Оптическая биомедицинская диагностика / Под ред. В.В. Тучина. – М.: Физматлит, 2007.
- 4.2. Allen J., Howell K. Microvascular imaging: techniques and opportunities for clinical physiological measurements // *Physiol. Meas.* 2014. V. 35. P. R91–R141.
- 4.3. Гуров И. П., Волков М. В., Маргарянц Н. Б., Потемкин А. В. Метод совмещения локально изменяющихся изображений в

- видеокапилляроскопии //Оптический журнал. 2019. Т. 86. № 12. С. 35-42.
- 4.4. Zijlstra W. G., Buursma, A., Meeuwssen-van der Roest W. P. Absorption spectra of human fetal and adult oxyhemoglobin, deoxyhemoglobin, carboxyhemoglobin, and methemoglobin // Clinical Chemistry. 1991. V. 37. P. 1633–1638.
- 4.5. Chochia P. A. Detection of capillaries in the images formed by a capillaroscope // J. Commun. Technol. Electron. 2013. V. 58. P. 1314–1323.
- 4.6. Гуров И. П. Формирование и анализ сигналов в системах компьютерной фотоники. – СПб: Университет ИТМО, 2018.
- 4.7. Gurov I., Volkov M., Margaryants N., Pimenov A., Potemkin A. High-speed video capillaroscopy method for imaging and evaluation of moving red blood cells // Opt. Lasers Eng. 2018. V. 104. P. 244–251.
- 4.8. Huang D., Swanson E. A., Lin C. P., *et al.* Optical coherence tomography //Science. 1991. V. 254. P. 1178–1181.
- 4.9. Вольнский М.А., Гуров И.П. Анализ изображений в оптической когерентной томографии. Учебно-методическое пособие по лабораторным работам. – СПб: Университет ИТМО, 2014.
- 4.10. Мандель Л., Вольф Э. Оптическая когерентность и квантовая оптика. – М.: Физматлит, 2000.
- 4.11. Wang S., Larin K. Optical coherence elastography for tissue characterization: a review // J. Biophotonics. 2015. V. 8. P. 279–302.
- 4.12. Kennedy B., Hillman T., McLaughlin R., Quirk B., Sampson D. In vivo dynamic optical coherence elastography using a ring actuator // Opt. Express. 2009. V.17. P. 21762–21772.
- 4.13. Wang R., Kirkpatrick S., Hinds M. Phase-sensitive optical coherence elastography for mapping tissue microstrains in real time // Appl. Phys. Lett. 2007. V. 90. P. 164105–164105-3.
- 4.14. Karimi R., Zhu T., Bouma B., Mofrad M. Estimation of nonlinear mechanical properties of vascular tissues via elastography // Cardiovasc. Eng. 2008. V.8. P. 191–202.

5 Информационные характеристики систем в оптике

5.1 Представление характеристик оптических полей на основе теоретико-информационного подхода

При проведении исследований требуется получить сведения об изучаемом явлении, процессе или объекте, т.е. *полезную информацию*. В оптических системах информация представляется в закодированном виде в форме электромагнитного поля.

Предположим, что источник информации – это неизвестный источник излучения, и требуется определить его спектральные характеристики (которые представляют полезную информацию). Для решения задачи можно использовать спектрометр, показывающий спектральное распределение интенсивности излучения по длинам волн.

В зависимости от степени априорной неопределенности совокупность возможных кривых спектрального распределения интенсивности $I(\lambda)$, либо отклонений от известной номинальной кривой, можно рассматривать как случайный процесс [5.1]. При этом получаемые в конкретном опыте значения интенсивности $I_k = I(\lambda_k)$ в спектре для каждой длины волны λ_k , $k = 1, \dots, K$ являются случайными величинами и образуют K -мерный случайный вектор $\mathbf{I} = (I_1, I_2, \dots, I_K)^T$, статистически полностью определяемый плотностью вероятности $p(\mathbf{I})$.

Произведение $p(\mathbf{I}) \Delta I_1 \Delta I_2 \dots \Delta I_K$, в свою очередь, определяет вероятность $P(\mathbf{I})$ того, что значения $I(\lambda)$ в точках λ_k будут попадать в узкие интервалы значений (см. рисунок 5.1, а) $I_1 + \Delta I$, $I_2 + \Delta I, \dots, I_K + \Delta I$, где ΔI – малое отклонение интенсивности, которое возможно зарегистрировать. Следовательно, при достаточно большом числе сечений K случайного процесса многомерная плотность вероятности $p(\mathbf{I})$ позволяет найти вероятность получения отдельной случайной функции – реализации случайного процесса. Заметим, что после проведения наблюдений (апостериорно) функция $I(\lambda)$, очевидно, является детерминированной.

Обозначим $P(\mathbf{I}; \lambda_1, \lambda_2, \dots, \lambda_K)$ совместную вероятность дискретных значений интенсивности на длинах волн $\lambda_1, \lambda_2, \dots, \lambda_K$.

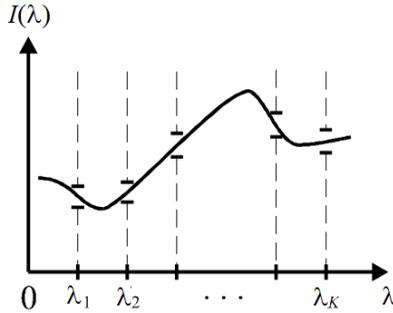


Рисунок 5.1. Спектральное распределение интенсивности как реализация случайного процесса

Если количество возможных (разрешаемых) значений интенсивности равно m , тогда, по определению, энтропия источника сообщений равна

$$H(\mathbf{I}) = - \underbrace{\sum_{i=1}^m \sum_{j=1}^m \dots \sum_{K}}_K P(\mathbf{I}; \lambda_1, \lambda_2, \dots, \lambda_K) \log P(\mathbf{I}; \lambda_1, \lambda_2, \dots, \lambda_K). \quad (5.1)$$

Из (5.1) можно вычислить среднее количество информации, содержащееся в полученном спектральном распределении интенсивности излучения исследуемого источника.

Рассмотренный выше общий подход можно распространить на другие задачи, например, для определения количества информации, содержащейся в изображении.

При формировании изображений оптической системой с числовой апертурой α количество точек объекта, в которых оценивается комплексная амплитуда отраженного излучения, определяется разрешающей способностью $\lambda/(2\alpha)$ оптической системы, где λ – длина волны. В наблюдаемой области площадью S количество точек составит $4\alpha^2(S/\lambda^2)$.

В физике наименьшее количество действительных параметров, которые необходимы, чтобы полностью описать систему, называется *числом степеней свободы* системы. Поскольку комплексная амплитуда задается двумя действительными числами изображение имеет $N = 8\alpha^2(S/\lambda^2)$ степеней свободы. Это наибольшее число *независимых данных*, которые содержатся в изображении.

Пусть двумерное изображение состоит из $N_I = P \times Q$ пикселей. Вероятность выбора одного пикселя из N_I возможных равна $P_1 = 1/N_I$ при равновероятном выборе. Среднее количество информации в одном пикселе изображения согласно формуле Шеннона составит

$$H_{\text{микс}} = - \sum_{i=1}^{N_I} P_i \log P_i, \quad (5.2)$$

где P_i – вероятность выбора i -го пикселя.

Если значение в пикселе квантуется по m уровням, то среднее количество информации в фиксированном пикселе изображения равно

$$H_{\text{ур}} = - \sum_{j=1}^m P_j \log P_j, \quad (5.3)$$

где P_j – вероятность получения значения сигнала на j -ом уровне квантования, $j = 1, \dots, m$.

Совместная энтропия определяется суммой (5.2) и (5.3) при статистической независимости выбора пикселя и имеющегося в нем дискретного уровня сигнала.

Отметим, что число степеней свободы, соответствующее количеству N_I пикселей изображения (и количеству уровней квантования) конечно, а исследуемого объекта, вообще говоря, бесконечно. Поэтому данному изображению могут соответствовать многие объекты. Для разрешения неоднозначности требуется априорная информация, на основе которой отбирается правильное представление объекта.

5.2 Информационная оптимизация оптико-цифровых систем

При бесконтактном контроле объект как источник информации характеризуется формой поверхности $z(x, y)$ в трехмерном пространстве. Определение формы осуществляется при освещении объекта, при этом наиболее важна пространственно-временная структура освещения и степень когерентности. После взаимодействия излучения с объектом получают представление объекта в форме отраженного (рассеянного) излучения.

Излучение направляется в канал передачи информации (см. рисунок 1.2 в разделе 1.1), который включает оптическую часть (распространение излучения от объекта до фотодетектора) и электронную часть (фотодетектор, формирующий и передающий зарегистрированные сигналы). Основное назначение оптической части

состоит в сборе максимального объема информации от объекта. При этом наиболее важна апертура наблюдательной системы.

Максимальное количество передаваемой информации в единицу времени носит название *пропускной способности* канала. На пропускную способность оказывает влияние шум в канале, присущий физической системе: шум источника излучения, шум спеклов, дробовой шум фотодетектирования, шум электронных компонентов и шум квантования сигнала по уровню. Особенности основных видов шума и их характеристики рассмотрены в учебном пособии [5.2].

На выходе канала формируются зарегистрированные реальные данные, которые декодируются при помощи соответствующего алгоритма с получением оценки $\hat{z}(x, y)$ трехмерной формы объекта. Эти данные направляются получателю информации.

Канал передачи информации содержит три части (см. рисунок 5.2): оптический канал распространения отраженного излучения I в свободном пространстве, оптико-электронный канал с формированием фотоэлектрического сигнала $s = \mu I$, где μ – коэффициент преобразования, и преобразователь сигналов в цифровую форму.

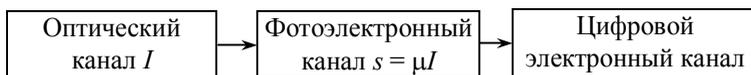


Рисунок 5.2. Структура информационного канала

При создании системы требуется обеспечить необходимую пропускную способность, чтобы передать именно ту информацию, которой достаточно для адекватного представления свойств объекта.

Существует максимальный объем информации, которая может быть передана в условиях воздействия шума. При передаче информации со скоростью, не превышающей пропускную способность канала, передача информации не сопровождается ошибками.

Пропускная способность канала с аддитивным гауссовским шумом определяется известной формулой Шеннона (см., например, [5.3], с. 250)

$$C_T = B \log_2 \left(1 + \frac{S_{in}}{N} \right) = B \log_2 \left(\frac{S}{N} \right), \quad (5.4)$$

где B – односторонняя ширина спектра во временной области, S_{in} – мощность сигнала на входе системы, N – средняя мощность шума, $S = S_{in} + N$ – регистрируемая мощность на выходе системы. Согласно

известной теореме Шеннона для канала без помех требуется обеспечить выполнение условия $C_T \geq H'$, где H' – производительность источника (бит/с). Объем информации, передаваемой в течение интервала времени ΔT составляет

$$C = \Delta T B \log_2(S/N). \quad (5.5)$$

Следовательно, количество переданной информации пропорционально произведению длительности на ширину полосы сигнала, а также логарифму отношения сигнал-шум.

Пропускная способность оптического канала

При оценивании пропускной способности оптического канала требуется рассматривать двумерную область в боковом направлении. Пропускная способность оптического канала определяется выражением [5.4]

$$C_O = 2\Delta X \Delta Y B_x B_y \log_2(S_O / N_O) = -4\Delta X \Delta Y B_x B_y \log_2 \mu, \quad (5.6)$$

где $\Delta X \times \Delta Y$ – размеры фотодетектора,

$$B_x B_y = \frac{\pi D^2}{16 \langle \lambda \rangle^2 f^2} \quad (5.7)$$

– ширина полосы в области пространственных частот, определяемая дифракционно-ограниченной системой в виде линзы диаметром D с фокусным расстоянием f при среднем значении длины волны $\langle \lambda \rangle$, и отношение сигнал-шум

$$\frac{S_O}{N_O} = \frac{\langle I^2 \rangle}{\sigma_I^2} \approx \frac{\langle I \rangle^2}{\sigma_I^2} = \frac{1}{\mu^2} \quad (5.8)$$

определяется главным образом шумом спеклов, присущим оптической системе с когерентным освещением (см., например, [5.5]). В (5.8), σ_I^2 – дисперсия значений интенсивности спеклов, μ – контраст спеклов [5.5].

Принимая в расчет пропускание оптического излучения через систему с площадью зрачка Σ_p и площадью фотодетектора Σ_d , пропускную способность оптического канала можно выразить в форме [5.4]

$$C_O = -\frac{\Sigma_p \Sigma_d}{\langle \lambda \rangle^2 f^2} \log_2 \mu, \quad (5.9)$$

Следовательно, для передачи максимального объема информации об объекте требуются широкий пространственно не когерентный источник (при использовании которого $\mu \ll 1$), большая апертура и большая площадь фотодетектора.

Пропускная способность электронного канала

Оптическое излучение регистрируется фоточувствительной матрицей форматом $P \times Q$ элементов. Ширина полосы пропускания в области пространственных частот при этом ограничена критерием Найквиста и определяется для одной пространственной координаты выражением

$$B_x = \frac{P}{2\Delta x}, \quad (5.10)$$

где Δx – шаг дискретизации. Используя соотношение, аналогичное (5.10) для второй пространственной координаты, пропускную способность электронного канала с учетом (5.4) можно выразить в форме

$$C_E = \frac{1}{2} PQ \log_2 \left(\frac{S_E}{N_E} \right). \quad (5.11)$$

Энергия оптического излучения, приходящаяся на один светочувствительный элемент, весьма мала, поэтому составляющая шума N_E в электронном канале возникает главным образом в форме дробового шума фотонов со статистикой Пуассона. В этом случае дисперсия шума σ_n^2 , как известно [5.2], равна среднему количеству регистрируемых фотонов $\langle n \rangle$, что определяет отношение сигнал-шум в (5.11) в виде

$$\frac{S_E}{N_E} = \frac{\langle n^2 \rangle}{\sigma_n^2} \approx \frac{\langle n \rangle^2}{\sigma_n^2} = \langle n \rangle. \quad (5.12)$$

При этом выражение (5.11) принимает форму

$$C_E = \frac{1}{2} PQ \log_2 \langle n \rangle. \quad (5.13)$$

Следовательно, объем информации, принятой фотодетектором, линейно возрастает с увеличением количества фоточувствительных элементов и возрастает по логарифмическому закону в зависимости от количества фотонов, зарегистрированных в среднем каждым элементом.

Пропускная способность при квантовании сигналов по уровню

Зарегистрированные сигналы преобразуются в цифровую форму при квантовании по уровню. При количестве уровней квантования m пропускная способность определяется выражением

$$C_{E_m} = \frac{1}{2} PQ \log_2 m. \quad (5.14)$$

в предположении о равномерном распределении сигнала по уровням квантования.

Результирующая пропускная способность канала

Результирующая пропускная способность не может превышать значения для каждого из рассмотренных выше компонентов информационного канала. Часть канала с минимальной пропускной способностью определяет верхнюю границу для канала в целом. К примеру, при когерентном освещении основной источник шума представляет шум спеклов в оптическом канале. Для снижения этого вида шума, как отмечено выше, требуется широкий источник с низкой пространственной когерентностью. При высокой интенсивности излучения становится доминирующим дробовой шум фотодетектирования. Количество уровней квантования сигнала должно быть достаточным, чтобы не ограничивать пропускную способность по сравнению с ограничениями от влияния шума.

Оптическая система способна обеспечить весьма высокую пропускную способность, поэтому ограничение накладывает электронный канал, особенно при использовании видеокамеры с ограниченным быстродействием. Следовательно, при создании системы целесообразно по возможности не вносить повышенных требований к пропускной способности электронного канала.

Корреляционные и спектральные характеристики объектов

Для обеспечения информационной эффективности системы требуется устранить избыточность источника информации (исходных данных), что особенно важно для трехмерной реконструкции объектов и при решении других задач, когда регистрируется большой набор видеоданных.

Форму поверхности объекта на участке $z(x, y)$ можно интерпретировать как реализацию случайного процесса, включающего ансамбль случайных функций (определяемых вариантами формы поверхности). Плотность вероятности значений $z(x, y)$, как правило, априорно не известна. Поэтому обратимся к статистическим характеристикам второго порядка, а именно, автокорреляционной функции $R_z(x, y)$ рассматриваемого случайного процесса и спектральной плотности мощности (СПМ) $G(f_x, f_y)$, определяемой в области пространственных частот (f_x, f_y) . Большинство объектов могут быть представлены гладкой поверхностью (на низких частотах) в сочетании с деталями поверхности (на высоких частотах) с меньшей амплитудой. Поэтому СПМ принимает большие значения на низких частотах со снижением при возрастании частоты (см. рисунок 5.3).

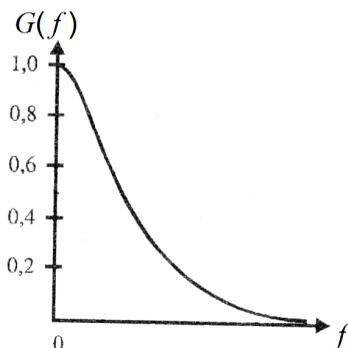


Рисунок 5.3. Нормированная зависимость спектральной плотности мощности от частоты

Более важный факт состоит в том, что расстояние до объекта обычно гораздо больше диапазона контролируемых отклонений поверхности. Очевидно, что передача информации о расстоянии до объекта (которая при контроле *формы* объекта не важна) представляется избыточной. Иными словами, информация от источника $z(x, y)$ сосредоточена при этом в окрестности нулевой частоты, тогда как для повышения информационной эффективности желательно, чтобы СПМ полезного сигнала приближалась к СПМ «белого» шума, $G(f_x, f_y) = \text{const}$, когда автокорреляционная функция близка к дельта-функции, т. е. значения сигнала взаимно не коррелированы, и нет избыточности в передаваемых данных.

«Отбеливание» спектра можно частично обеспечить при дифференцировании функции $z(x, y)$ поскольку согласно свойствам преобразования Фурье спектр производной

$$G'(f_x, f_y) = j2\pi(f_x + f_y) G(f_x, f_y),$$

т.е. возрастает пропорционально частоте и равен нулю на нулевой частоте. Однако дифференцирование необходимо выполнить оптическими методами до фотоэлектрической регистрации сигнала. В противном случае операция дифференцирования смеси сигнала и шума приведет к значительному возрастанию влияния шума.

Одним из возможных методов является восстановление формы объекта при его освещении под различными углами при неизменном направлении наблюдения, при этом оказывается возможным определение локального наклона поверхности (который имеет смысл производной) с последующим восстановлением формы поверхности. Данный метод успешно использован при восстановлении микропрофиля поверхности (см. [5.6]).

Информационная эффективность систем оптической когерентной томографии

В разделе 4.3 рассмотрены два основных вида систем оптической когерентной томографии (ОКТ): во временной области (ОКТВ) и в частотной области (ОКТЧ). Проведем сравнение пропускной способности этих двух видов систем ОКТ.

Для сравнения пропускной способности оптического канала в ОКТВ и ОКТЧ определим отношение [5.7]

$$\frac{C_{OB}}{C_{OC}} = \gamma \left(\frac{f_C}{f_B} \right)^2 \approx \gamma \left(\frac{2L}{l_c} \right) \gg 1, \quad (5.15)$$

где

$$\gamma = \Sigma_{pB} / \Sigma_{pC} \approx \frac{\pi D}{4 \delta Y} > 1 \quad (5.16)$$

– параметр, учитывающий использование площади зрачка в оптической системе, δY разрешение в боковом направлении в ОКТЧ (см. рисунок 5.4, б). Глубина фокусировки d_z объектива по координате z рассчитывается с использованием соотношения [5.8] $d_z = 1,8\langle\lambda\rangle / (NA)^2 = 1,8\langle\lambda\rangle f^2 / D^2$, где NA – числовая апертура объектива, при этом

$$\frac{d_{zB}}{d_{zC}} = \left(\frac{f_B}{f_C} \right)^2 \approx \frac{l_c}{2L}. \quad (5.17)$$

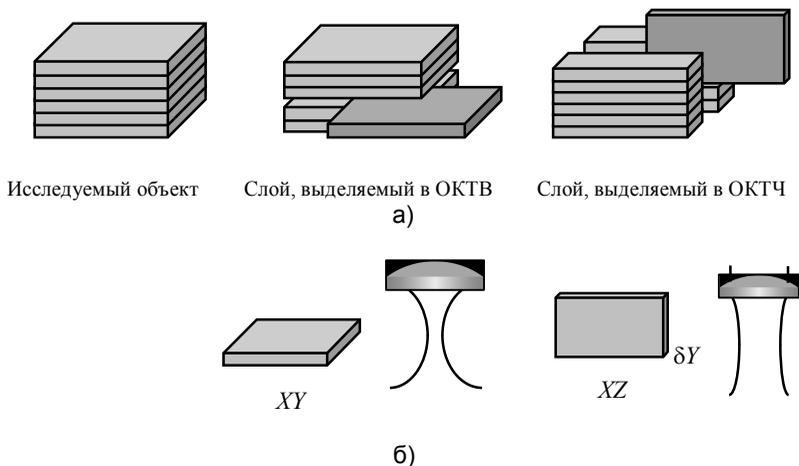


Рисунок 5.4. Сравнение геометрического расположения исследуемых слоев (а) и фокусировки зондирующего пучка в ОКТВ и ОКТЧ (б) [5.9]

Выражение (5.15) показывает, что пропускная способность оптического канала в ОКТВ существенно выше, чем в ОКТЧ. Это объясняется двумя ограничениями в ОКТЧ: необходимостью использования низкоапертурных объективов для обеспечения фокусировки на всей глубине объекта и использованием только части полезной площади зрачка при отображении одной линии в поле наблюдения вдоль оси X на входную щель спектрального прибора.

Второе ограничение устраняется в системах ОКТ с перестраиваемым по длине волны источником излучения в режиме, когда источник освещает все поле зрения. Однако такой режим не нашел широкого применения вследствие формирования B -сканов в сечении объекта при боковом сканировании зондирующего пучка.

Обратимся к рассмотрению пропускной способности электронного канала.

Оптическое излучение регистрируется фоточувствительной матрицей, содержащей $P \times Q$ элементов. Количество элементов примем одинаковым в системах ОКТВ и ОКТЧ. При этом справедливы выражения (5.10) – (5.14).

Важно отметить, что в ОКТВ полезную информацию несут только $\langle n_c \rangle = \langle n \rangle l_c / 2L$ фотонов в пределах длины когерентности излучения l_c , тогда как в ОКТЧ эту информацию переносят фотоны $\langle n \rangle$, отраженные по всей глубине объекта. При этом отношение

$$\frac{C_{BE}}{C_{CE}} = \frac{\log_2 \langle n_c \rangle}{\log_2 \langle n \rangle} < 1. \quad (5.18)$$

Выражение (5.18) показывает, что системы ОКТЧ обеспечивают выигрыш в пропускной способности электронного канала в логарифмической зависимости от разрешающей способности по глубине. С другой стороны, согласно (5.15) пропускная способность оптического канала в ОКТВ возрастает линейно по отношению к ОКТЧ при увеличении разрешающей способности по глубине. Следовательно, ОКТВ имеет потенциальное преимущество в пропускной способности при исследованиях трехмерной микроструктуры объектов с высоким разрешением, если электронный канал в ОКТВ не ограничивает общую пропускную способность системы (например, при использовании быстродействующей видеокамеры и скоростной шины ввода данных в вычислительную систему).

Двумерное и трехмерное представление информации в системах оптической когерентной томографии

Очевидно, что трехмерное отображение внутренней микроструктуры объекта более информативно по сравнению с двумерным сечением. Однако важны два вопроса: как передавать и обрабатывать большой объем информации при исследованиях трехмерной микроструктуры и как использовать эту информацию.

Восприятие человека ориентировано на наблюдение и интерпретацию главным образом двумерных изображений. По этой причине большинство ОКТ-систем создаются для визуализации двумерных изображений в форме *B*-сканов. При этом снижается объем информации, регистрируемой и обрабатываемой одновременно, и обеспечивается высокое быстродействие при визуализации.

Тем не менее трехмерная информация очень важна, когда сложно оценить объект с не вполне ясными свойствами на основе наблюдения единственного *B*-скана. В этом случае исследователь рассматривает несколько параллельных *B*-сканов, чтобы разрешить неопределенность, используя дополнительную информацию.

Трехмерная информация важна и ее использование перспективно для автоматических систем диагностики, которые способны обеспечить оперативное получение трехмерной информации и принятие решения о свойствах исследуемого объекта. При этом целесообразно, в частности, использование ОКТ-систем полного поля. Исследования направлены на повышение быстродействия таких систем при формировании, регистрации и обработке трехмерных данных.

5.3 Информационный критерий качества систем бесконтактного контроля объектов

При бесконтактном контроле различных объектов на основе формирования и анализа изображений основное значение имеют показатели, которые включают три основные группы широко используемых критериев, не учитывающих пространственную структуру изображения (см., например, [5.10], с. 151–154; [5.11], с. 612–614): отношение сигнал-шум (ОСШ, англ. SNR, signal-to-noise ratio), в том числе пиковое отношение (ПОСШ, англ. PSNR), среднее квадратичное отклонение (СКО, англ. RMSE, root mean square error) и вероятностные критерии.

Отношение сигнал-шум определяется выражением

$$\text{SNR} = \frac{\sigma^2}{\sigma_n^2}, \quad (5.19)$$

где σ^2 – дисперсия исходных значений изображения, σ_n^2 – дисперсия обусловленных влиянием шума отклонений от истинных значений. Отметим, что известны другие определения ОСШ, в частности, отношение максимального значения сигнала к СКО шума.

Пиковое отношение сигнал-шум [5.12]

$$\text{PSNR} = 10 \lg \frac{(2^n - 1)^2}{\sigma_n^2} \quad (5.20)$$

вычисляется в логарифмической шкале в форме отношения квадрата максимального значения пиксела изображения в n -битном представлении к дисперсии ошибки (обычно обусловленной влиянием шума).

СКО вычисляется согласно выражению [5.11]

$$\sigma_I = \sqrt{\frac{1}{LW} \sum_{i=1}^L \sum_{j=1}^W (\hat{I}(i, j) - I(i, j))^2}, \quad (5.21)$$

где $\hat{I}(i, j)$ – оценка изображения с координатами пикселей (i, j) , $I(i, j)$ – истинное (идеальное) изображение, $L \times W$ – размер изображения в пикселях.

Критерии (5.19) – (5.21) удобны ввиду их простоты, но не всегда адекватны решаемым задачам (см., в частности, примеры в [5.12]) вследствие того, что недостаточно полно учитывают вероятностные характеристики изображения.

Этот недостаток частично устраняется при использовании вероятностного критерия, который определяется выражением [5.10]

$$\int_{\varepsilon_{\min}}^{\varepsilon_{\max}} p(\varepsilon) d\varepsilon = P, \quad (5.22)$$

где $\varepsilon(i, j) = \hat{I}(i, j) - I(i, j)$, ε_{\min} и ε_{\max} – минимальное и максимальное значение ошибки, $p(\varepsilon)$ – плотность вероятности ошибки. Значение критерия (5.22) можно получить экспериментально из гистограммы распределения ошибки, однако для его применения требуется знание «эталонного» изображения $I(i, j)$.

Ввиду сложности и разнообразия задач бесконтактного контроля объектов целесообразно обратиться к теоретико-информационному подходу [5.14] и обобщенной структуре видеoinформационной системы (см. рисунок 5.5), обозначив вход оптико-цифровой системы как X и выход как Y , и использовать информационный критерий, характеризующий качество системы в целом, в форме

$$Q = \frac{H(X) - H(X|Y)}{H(X)} = 1 - \frac{H(X|Y)}{H(X)}, \quad (5.23)$$

где $H(X)$ – энтропия источника сообщений (в данном рассмотрении – исследуемого объекта), $H(X|Y)$ – потери информации вследствие влияния искажений и помех. Качество системы тем выше, чем большие значения принимает критерий (5.23). Очевидно, что максимальное значение $Q_{\max} = 1$.

Информационный критерий известен в теории связи, согласно (5.23) он выражается через взаимную информацию X и Y (см., например, [5.3], с. 227).



Рисунок 5.5. Блок-схема оптико-цифровой системы

Рассмотрим особенности применения информационного критерия качества при решении прикладных задач бесконтактного контроля объектов на основе материалов публикаций [5.14] и [5.15].

Согласно теории информации исследуемый объект можно интерпретировать как источник сообщений [5.14], от которого информация передается через оптический и электронный каналы (соответственно посредством отраженного оптического излучения и зарегистрированных камерой видеокадров). В частности, в случае контроля характеристик поверхности объекта под сообщениями понимается информация о степени отражения r_p на расстоянии до объекта z_q (пространственные координаты в боковом направлении опускаем для упрощения записи).

Вероятность получить коэффициент отражения r_p на расстоянии z_q можно выразить как

$$P(r_p, z_q) = P(r_p)P(z_q) \quad (5.24)$$

ввиду взаимной статистической независимости коэффициента отражения и положения точки поверхности. При этом, следуя теории информации по Шеннону, энтропию источника сообщений можно вычислить как

$$H(R, Z) = H(R) + H(Z) = -\sum_{p=1}^{M_r} P(r_p) \log_2 P(r_p) - \sum_{q=1}^{M_z} P(z_q) \log_2 P(z_q), \quad (5.25)$$

где $r_p \in R$, $z_q \in Z$, R и Z – случайные величины. Если коэффициент отражения представлен, например, 256-ю уровнями (M_r) и разрешение по расстоянию до поверхности объекта соответствует 250-ти интервалам Δz ($M_z = 250$), энтропия (5.25) равна примерно двум байтам на одну точку рельефа. Эта информация должна быть передана через оптический и электронный субканалы с максимальной скоростью и минимальными потерями информации. Канал передачи информации должен обеспечивать требуемую пропускную способность.

Критерий (5.23), очевидно, можно записать в форме

$$Q = \frac{H(Y) - H(Y|X)}{H(X)}, \quad (5.26)$$

где энтропия наблюдаемых значений (по выходу) $H(Y)$ вычисляется по формуле Шеннона при подстановке значений вероятности для наблюдаемых оценок в условиях влияния помех. Условная энтропия $H(Y|X)$ представляет собой энтропию помехи. Зная $H(Y)$ и $H(Y|X)$, можно найти оценку критерия Q согласно (5.26).

Рассмотрим в качестве примера применение информационного критерия к интерферометрическим системам. Регистрируемые изображения в этом случае представляют собой картины интерференционных полос, определяемые хорошо известной моделью для пространственного распределения интенсивности света

$$I(x, y) = I_a(x, y) \cos \varphi(x, y) + I_b(x, y), \quad (5.27)$$

где $I_a(x, y)$, $\varphi(x, y)$, $I_b(x, y)$ – амплитуда, фаза и фоновая составляющая картины интерференционных полос, соответственно. Критерий качества в рассматриваемом случае обычно выражается погрешностью $\delta\varphi$ определения разности фаз опорной и измерительной волн в интерферометре. Представляет интерес сравнение информационного и традиционных критериев качества интерферометрических систем.

Обратимся вначале к рассмотрению статистических характеристик интерферометрических сигналов, формируемых в результате преобразования интенсивности света в фотоэлектрический сигнал, $s = \mu I$, где μ – коэффициент преобразования,

$$s(i, j) = a(i, j) \cos \varphi(i, j) + b(i, j), \quad (5.28)$$

где $a(i, j) = \mu I_a(i, j)$, $b(i, j) = \mu I_b(i, j)$, (i, j) – дискретные координаты пикселей видеокамеры, $x_i = i\Delta x$, $y_j = j\Delta y$, с шагом Δx и Δy , соответственно, $i = 1, \dots, L$, $j = 1, \dots, W$.

Для идеализированного сигнала (5.28), когда $a = b = 1$, и значения фазы равновероятны, т.е. плотность вероятности $p_\varphi(\varphi) = 1/(2\pi)$, значения интерферометрического сигнала не являются равновероятными и определяются известным выражением для плотности вероятности значений гармонического процесса

$$p_s(s) = \begin{cases} 1 / \left(\pi \sqrt{a^2 - s^2} \right), & |s| \leq a, \\ 0, & |s| > a. \end{cases} \quad (5.29)$$

Соответствующая функция распределения вероятностей определяется выражением [5.13]

$$P_s(s) = \begin{cases} 0, & s < -a, \\ \frac{1}{\pi} \left(\frac{\pi}{2} + \arcsin \frac{s}{a} \right), & |s| \leq a, \\ 1, & s > a. \end{cases} \quad (5.30)$$

Пусть имеются M уровней s_m квантования сигнала, $m = 0, \dots, M - 1$ при шаге квантования Δs . Полагая, что s_m представляют собой набор значений случайной величины S , можно использовать значения

вероятностей из (5.30) и согласно формуле Шеннона вычислить энтропию

$$H(S) = - \sum_{m=0}^{M-1} P(s_m) \log_2 P(s_m). \quad (5.31)$$

Оценку энтропии на практике можно получить методом построения гистограммы значений сигнала (5.28) для набора значений фазы $\varphi(i, j)$. При достаточно большом количестве значений фазы в наборе, которое превышает количество уровней квантования M , энтропия (5.31) будет близка к величине $\log_2 M$, равной максимальной энтропии. Экспериментальная проверка показала, что при 10^4 заданных значений фазы, распределенных по равновероятному закону, оценки энтропии (5.31) составили $H(S) = 7,68$ бит/отсчет, $9,61$ бит/отсчет и $11,34$ бит/отсчет соответственно для $M = 256, 1024, 4096$. Отметим, что эти значения ниже максимальных H_{\max} , достигаемых при равновероятном законе распределения по уровням квантования и равных $8, 10$ и 12 бит/отсчет соответственно. Данное отличие объясняется погрешностью квантования сигнала, обусловленной округлением значений сигнала, которую принято называть шумом квантования. Кроме этого, проявляется так называемое явление дефицита алфавита ([5.16], с. 234), когда несмотря на большое количество задаваемых значений фазы некоторые окна квантования не содержат ни одного отсчета сигнала. В рассмотренном выше примере доля незаполненных окон квантования составила около 2% при 11-ти двоичных разрядах и более 10% при 12-ти разрядах. Следовательно, при увеличении разрядности представления сигналов в видеокамере снижается шум квантования, однако возрастает влияние дефицита алфавита, приводящее к некоторым потерям информации.

Обратимся к взаимосвязи разрешающей способности интерферометра по фазе и информационного критерия (5.26) с учетом влияния квантования интерферометрического сигнала по уровню.

Энтропия источника сообщений $H(X)$ определяется вероятностью значений фазы интерференционных полос. Обозначим $\Delta\varphi$ разрешающую способность интерферометра по фазе. Полагая, что значения фазы как случайной величины Φ имеют равновероятное распределение, получим оценку количества разрешаемых значений фазы в пределах фазового цикла равную $2\pi/\Delta\varphi$. Пусть $\Delta\varphi = 2\pi/1000$ (значение, достижимое в системах прецизионного оптического контроля). Тогда энтропия источника (значений фазы) $H(X) = H(\Phi)$ составит около 10 бит/отсчет. Энтропия «правильных» значений сигнала есть $H(Y) - H(Y|X) = H(S) - H(S|\Phi)$, где второе слагаемое

(условная энтропия) определяется потерями информации из-за влияния шума квантования. Это слагаемое отражает тот факт, что при фиксированном значении фазы значения сигнала имеют отклонения, вызванные случайным шумом квантования и выражаемые условной вероятностью $P(S|\Phi)$. При равновероятном попадании значений сигнала в окно квантования Δs СКО шума квантования, очевидно, составляет $\Delta s/(2\sqrt{3})$ и при большом количестве уровней квантования (т.е. малых значениях Δs) пренебрежимо мало. При этом $H(S) \approx H(Y)$, и информационный критерий (5.26) выражается как

$$Q = \frac{H(S)}{H(\Phi)}. \quad (5.32)$$

Из вышеизложенного и выражения (5.32) следует, в частности, что при использовании 8-битной видеокамеры информационный критерий имеет величину менее 0,8. Видеокамера с 10-ю разрядами обеспечивает повышение информационного критерия до значений, близких к единице. Однако дальнейшее повышение разрядности не меняет величину критерия: энтропия значений квантованного полезного сигнала не может превышать энтропию фазы.

Таким образом, информационный критерий позволяет количественно охарактеризовать процедуру квантования сигналов по уровню и сделать обоснованный выбор видеокамеры с необходимой разрядностью представления сигналов с учетом ожидаемой разрешающей способности интерферометра по фазе.

Реальные регистрируемые сигналы подвержены также влиянию шума наблюдения. Обозначим n_m отсчеты нормально распределённого аддитивного шума с нулевым средним значением, статистически не зависимо от значений сигнала s_m . При этом значения регистрируемого сигнала, искажённого шумом, определяются как

$$\xi_m = s_m + n_m. \quad (5.33)$$

Энтропии значений принятого сигнала ξ и шума n обозначим соответственно как $H(\Xi)$ и $H(N)$.

При обработке рассматриваются матрица дискретных отсчетов сигнала (5.28) с учетом влияния шума в виде

$$\xi(i, j) = a \cos \varphi(i, j) + b + n(i, j) \quad (5.34)$$

и потери информации ввиду вносимых шумом искажений сигнала.

В качестве примера зададим пространственное распределение фазы

$$\varphi(i, j) = 2\pi l [(i - i_0)^2 - (j - j_0)^2] [(W - i_0)^2 + (L - j_0)^2]^{-1}, \quad (5.35)$$

где, как и ранее, $i = 1, \dots, L, j = 1, \dots, W, (i_0, j_0)$ – координаты пиксела с нулевым значением фазы, l – количество фазовых циклов (наблюдаемых интерференционных полос). Распределение фазы (5.35) при $l = 5, L = W = 100, i_0 = j_0 = 35$ представлено на рисунке 5.6, а. Соответствующая картина интерференционных полос (5.28) иллюстрируются на рисунке 5.6, б.

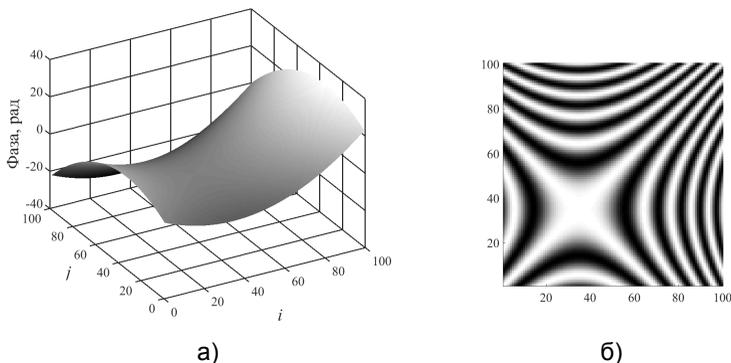


Рисунок 5.6. Пространственное распределение фазы (а) и картина интерференционных полос (б).

Построим гистограмму сигнала (5.34) при различных значениях отношения сигнал/шум

$$\text{SNR} = \frac{a}{\sigma_n}, \quad (5.36)$$

где σ_n – СКО шума. Примеры гистограмм представлены на рисунке 5.7. Гистограммы соответствуют плотности вероятности (5.29) гармонического случайного процесса и ее модификации при наличии аддитивного шума.

Оценки энтропии (5.31), полученные при 8-битном представлении регистрируемого сигнала и ОСШ = 10, 50, 100, составили $H(\Xi) = 7,88, 7,77, 7,72$ бит/отсчет соответственно. Оценки энтропии шума определяются нормальным распределением вероятностей с параметром $\sigma_n = a / \text{SNR}$ и получены равными $H(N) = 3,7, 2,0, 1,3$ соответственно.

В рассматриваемом случае формула информационного критерия (5.26) принимает вид

$$Q = \frac{H(Y) - H(Y|X)}{H(X)} = \frac{H(\Xi) - H(N)}{H(\Phi)}. \quad (5.37)$$

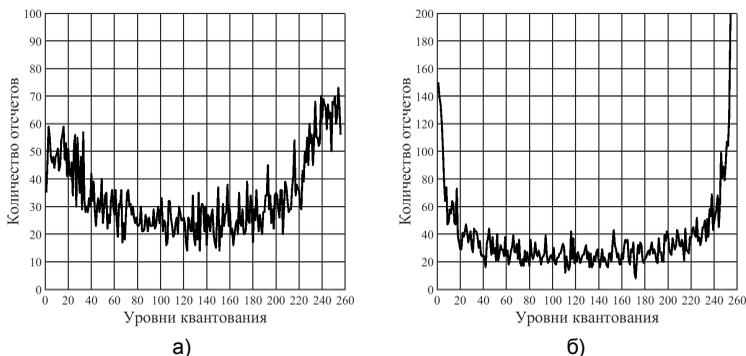


Рисунок 5.7. Гистограммы значений интерферометрического сигнала при ОСШ = 10 (а) и ОСШ = 100 (б)

Примем разрешающую способность интерферометра по фазе $2\pi/250$, что примерно соответствует 8-битному представлению в пределах одного фазового цикла 2π .

На рисунке 5.8 приведена кривая, полученная в результате моделирования с использованием выражения (5.37) и показывающая взаимосвязь информационного критерия и отношения сигнал-шум в определении (5.36). Из рисунка видно, что информационный критерий более показателен при малых и умеренных значениях ОСШ: сравнительно малым отклонениям ОСШ соответствуют существенные изменения информационного критерия.

Отметим, что распределение фазы может включать значения в диапазоне, большем 2π (например, в (5.35) при $l > 1$). Тогда энтропия значений фазы $H(\Phi)$ увеличится на величину $\log_2 l$ (при равновероятных значениях фазы). В рассматриваемом примере (см. рисунок 5.6) при $l = 5$ энтропия фазы составила $H(\Phi) \approx 10,3$ бит/отсчет. При этом значения критерия (5.37) снижаются в соответствии с кривой, представленной пунктирной линией на рисунке 5.8. Уменьшение величины критерия отражает факт потерь информации при получении оценок приведенной фазы (в интервале $(0, 2\pi)$) вместо исходной развернутой фазы (5.35) при $l = 5$.

Из вышеизложенного следует, что информационный критерий позволяет оценить важные характеристики процессов формирования и обработки видеoinформации при бесконтактном контроле объектов, определяющие требования к пропускной способности и, следовательно, к быстродействию и связанной с ним помехоустойчивости системы при регистрации, передаче и обработке

данных. Это особенно важно в случае использования получивших широкое распространение методов регистрации и обработки последовательности видеок кадров для многокурсного наблюдения объектов, исследования динамических объектов, гиперспектральных наблюдений, трехмерного отображения объектов и др., когда объем регистрируемой и обрабатываемой видеoinформации имеет существенное значение.

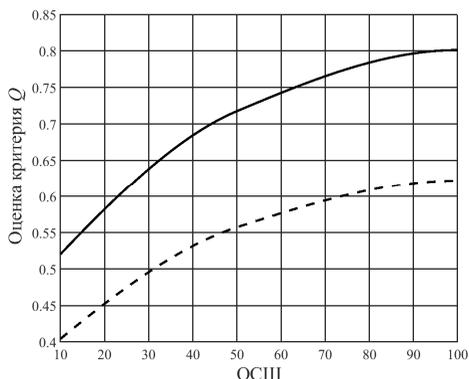


Рисунок 5.8. Взаимосвязь информационного критерия и отношения сигнал-шум в условиях воздействия аддитивного шума для развернутой фазы (сплошная линия) и для приведенной фазы (пунктир).

Информационный критерий носит универсальный характер ввиду того, что произвольное изображение можно описать моделью вида (5.33) и применить методику определения информационного критерия в форме (5.26) подобно тому, как описано выше применительно к интерферометрическим системам. Значения энтропии $H(Y)$ и условной энтропии $H(Y|X)$ в числителе выражения (5.26) определяются соответственно статистическими характеристиками зарегистрированного изображения и шума, поэтому их вычисление не составляет сложности по отношению к традиционному определению ОСШ в (5.19). Вместе с тем знаменатель $H(X)$ в (5.26) позволяет учесть априорную информацию об исследуемом объекте, которая может быть получена, в частности, методами машинного обучения на выборке ранее исследованных объектов заданного класса. Информационный критерий (5.26) есть отношение взаимной информации между входом и выходом системы к информации, потенциально содержащейся в свойствах объекта, т.е.

информационный критерий выражает долю извлеченной полезной информации.

Контрольные вопросы

1. Какое понятие обозначает наибольшее число независимых данных, которые содержатся в изображении?
2. Приведите схему и объясните структуру оптико-цифровой системы в форме информационного канала.
3. Какая величина определяет объем информации, передаваемой в течение заданного интервала времени?
4. Какие параметры оптической системы определяют ширину полосы пропускания в области пространственных частот?
5. Чем определяется объем информации, принятой фотодетектором?
6. Какие характеристики объектов учитываются при информационной оптимизации оптических систем?
7. Приведите сравнение информационной эффективности систем оптической когерентной томографии во временной области и в частотной области.
8. В чем заключаются особенности двумерного и трехмерного представления информации в системах оптической когерентной томографии?
9. Приведите и объясните формулу, определяющую информационный критерий качества системы.

Список литературы

- 5.1. Вентцель Е. С., Овчаров Л. А. Теория случайных процессов и ее инженерные приложения. – М.: Наука, 1991.
- 5.2. Гуров И. П. Формирование и анализ сигналов в системах компьютерной фотоники. – СПб: Университет ИТМО, 2018.
- 5.3. Теория электрической связи /Под ред. Д.Д. Кловского. – М.: Радио и связь, 1999.
- 5.4. Wagner, C., Häusler, G. Information theoretical optimization for optical image sensors // Appl. Opt. 2003. V. 42. P. 5418–5426.
- 5.5. Гудмен Дж. Статистическая оптика. – М.: Мир, 1988.
- 5.6. Gurov I., Volkov M., Zhukova E., et al. Evaluation of laser ablation crater relief by white light micro interferometer // Proc. SPIE. 2017. V. 10329. P. 103294I.
- 5.7. Gurov, I., Karpets A., Margariants N., Vorobeva E. Full-field high-speed optical coherence tomography system for evaluating multilayer and random tissues // Proc. SPIE. 2007. V. 6618. P. 661807.

- 5.8. Handbook of biological confocal microscopy / J.B. Pawley, ed. – NY: Plenum Press, 1995.
- 5.9. Gurov I. Signal processing methods in full-field optical coherence microscopy. In: Handbook of full-field optical coherence microscopy: Technology and applications /A. Dubois, ed.. NY: Pan Stanford Publishing Pte. Ltd, 2016. Chapt. 5. P. 183–222.
- 5.10. Методы компьютерной обработки изображений / Под ред. Сойфера В. А. М.: Физматлит, 2003.
- 5.11. Гонсалес Р., Вудс Р. Цифровая обработка изображений. М.: Техносфера, 2005.
- 5.12. Ричардсон Я. Видеокодирование. H.264 и MPEG-4 – стандарты нового поколения. М.: Техносфера, 2005.
- 5.13. Бендат Дж., Пирсол А. Прикладной анализ случайных данных. М.: Мир, 1989.
- 5.14. Васильев В. Н., Гуров И. П., Потапов А. С. Современная видеоинформатика: проблемы и перспективы // Оптический журнал. 2012. Т. 79. №11. С. 5–15.
- 5.15. Гуров И. П., Дайнеко М. В. Информационный критерий качества формирования и анализа изображений при бесконтактном контроле объектов // Оптический журнал. 2020. №7 (в печати).
- 5.16. Голдман С. Теория информации. М.: ИИЛ, 1957.



Миссия университета – открывать возможности для гармоничного развития конкурентоспособной личности и вдохновлять на решение глобальных задач.

Гуров Игорь Петрович

**ОСНОВЫ ТЕОРИИ СИСТЕМ И СИСТЕМНОГО АНАЛИЗА
В ОПТИКЕ**

Учебно-методическое пособие

В авторской редакции

Редакционно-издательский отдел Университета ИТМО

Зав. РИО

Н.Ф. Гусарова

Подписано к печати

Заказ №

Тираж

Отпечатано на ризографе

Редакционно-издательский отдел
Университета ИТМО
197101, Санкт-Петербург, Кронверкский пр., 49