

На правах рукописи



ФАВОРСКАЯ МАРГАРИТА НИКОЛАЕВНА

**МОДЕЛИ И МЕТОДЫ РАСПОЗНАВАНИЯ  
ДИНАМИЧЕСКИХ ОБРАЗОВ НА ОСНОВЕ  
ПРОСТРАНСТВЕННО-ВРЕМЕННОГО АНАЛИЗА  
ПОСЛЕДОВАТЕЛЬНОСТЕЙ ИЗОБРАЖЕНИЙ**

05.13.17 – «Теоретические основы информатики»

**АВТОРЕФЕРАТ**

диссертации на соискание ученой степени  
доктора технических наук

Красноярск – 2011

Работа выполнена в Государственном образовательном учреждении высшего профессионального образования «Сибирский государственный аэрокосмический университет имени академика М. Ф. Решетнева», г. Красноярск

Научный консультант: доктор физико-математических наук, профессор  
Попов Алексей Михайлович

Официальные оппоненты: доктор технических наук, профессор  
Кашкин Валентин Борисович

доктор технических наук, профессор  
Киричук Валерий Сергеевич

доктор физико-математических наук, профессор  
Рязанов Владимир Васильевич

Ведущая организация ФГУП «Государственный научно-исследовательский институт авиационных систем», г. Москва

Защита состоится 18 февраля 2011 года в 14 часов на заседании диссертационного совета Д 212.099.11 при Сибирском федеральном университете по адресу: 660074, г. Красноярск, ул. акад. Киренского, 26, ауд. УЛК 115.

С диссертацией можно ознакомиться в научной библиотеке Сибирского федерального университета по адресу: г. Красноярск, ул. акад. Киренского, 26, Г274.

Автореферат разослан 12 января 2011 г.

Ученый секретарь  
диссертационного совета



Л. И. Покидышева

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**Актуальность работы.** Существует класс задач, где особую важность приобретает информация о структуре и движении объектов сцены (видеонаблюдение в закрытых помещениях, в местах большого скопления людей, управление движением робототехнических комплексов, наблюдение за движением транспортных средств и т.д.). Последовательности изображений являются сложным информационным ресурсом, структурированным в пространстве и во времени и объединяющим исходную информацию в виде многомерных сигналов, форму ее представления в компьютере и физические модели динамических объектов, явлений, процессов.

Анализ пространственно-временных данных позволяет выявлять не только статические, но и динамические признаки объектов наблюдения. В этом случае задачу распознавания можно определить как классификацию совокупностей состояний или как классификацию траекторий, решение которой не может быть найдено классическими методами распознавания, т.к. временные переходы могут порождать преобразования изображений, не описываемые известными аналитическими зависимостями. Также наряду с задачей распознавания динамических объектов возникают задачи распознавания активных действий объектов и событий сцены, например, для выявления несанкционированных действий в местах скопления людей или определения жанра сцены для индексации в мультимедийных базах данных. Если рассматривать задачу распознавания объектов и событий по последовательностям изображений в виде единого процесса, то наиболее целесообразным является системный иерархический подход с элементами параллельной обработки на каждом уровне.

Первые работы исследователей из США, Швеции, Франции, Японии по обработке последовательностей изображений и распознаванию динамических объектов были опубликованы в конце 1980-х гг. Позже аналогичные работы стали появляться и в России: в Москве (МГУ, МФТИ), С-Петербурге (СПбГУ, ГУАП), Самаре (СГАУ), Рязани (РГРУ), Новосибирске (НГУ), Томске (ТГПУ) и других городах. Следует отметить особый вклад выдающихся российских ученых, занимающихся в данной области, таких, как академик РАН, д.ф.-м.н. Ю. И. Журавлев, член-корреспондент РАН, д.т.н. В. А. Сойфер, д.т.н. Н. Г. Загоруйко, д.т.н. Л. М. Местецкий, д.т.н. В. П. Пяткин, д.т.н. Б. А. Алпатов и другие. На сегодняшний день достигнуты значительные успехи при построении систем видеонаблюдения, систем аутентификации личности по изображениям и т. д. Однако существуют нерешенные проблемы при распознавании динамических образов из-за сложности и многообразия поведения объектов реального мира. Таким образом, данное направление нуждается в совершенствовании моделей, методов и алгоритмов распознавания динамических объектов и событий по последовательностям изображений в различных диапазонах электромагнитного излучения.

**Целью диссертационной работы** является повышение эффективности распознавания динамических объектов, их активных действий и событий в сложных сценах по последовательностям изображений для систем наружного и внутреннего видеонаблюдения.

Достижение цели предполагает решение следующих основных задач:

– Провести анализ методов оценки движения и нахождения признаков движения объектов по набору последовательных изображений, методов сегментации динамических объектов и семантического анализа сложных сцен, а также подходов к построению систем распознавания и слежения за динамическими объектами различного целевого назначения.

– Разработать модели распознавания статических и динамических образов, основываясь на иерархической процедуре обработки временных рядов, в частности, последовательностей изображений.

– Разработать метод оценки движения динамических структур по пространственно-временной информации, полученной в различных диапазонах электромагнитного излучения, позволяющий выбирать методы сегментации в зависимости от характера движения и, тем самым, выполнять адаптивное распознавание динамических образов.

– Создать модель многоуровневого движения динамических структур в сложной сцене, позволяющую на основе полученных одометрических данных строить траектории движения динамических структур и выдвигать гипотезы о существовании визуальных объектов на основе анализа предыстории движений.

– Разработать комплексный алгоритм сегментации, учитывающий совокупность выявленных признаков динамических структур при произвольных направлениях перемещений и перекрытий проекций объектов, основываясь на модели многоуровневого движения в сложных сценах.

– Разработать метод распознавания динамических образов, представленных в терминах формальной грамматики и видеографа сцены, на основе метода коллективного принятия решений, а также методы распознавания активных действий и событий в сложной сцене, использующие графы активных действий и событий (расширяющие видеограф сложной сцены) и байесовскую сеть.

– На основе разработанных методов и моделей спроектировать экспериментальные системы различного назначения, предназначенные для обработки последовательностей изображений объектов, характеризующихся фиксированным и произвольным набором *2D*-проекции, и распознавания динамических образов в сложных сценах.

**Методы исследований.** При выполнении диссертационной работы использовались методы теории распознавания образов, дескриптивной теории распознавания изображений, теории обработки сигналов, методы векторного анализа и тензорного исчисления, а также теория групп, теория формальных грамматик.

**Научная новизна** диссертационной работы состоит в разработке нижеприведенных теоретических положений, совокупность которых является новым научным достижением в области распознавания динамических образов на основе пространственно-временного анализа последовательностей изображений.

1. Построена новая модель преобразования динамических изображений, отличающаяся расширенными иерархическими уровнями сегментации (по локальным и глобальным векторам движения) и распознавания (объектов и их активных действий), позволяющая находить целевые признаки для статических сцен с движущимися объектами и динамических сцен на основе понятия максимального динамического инварианта.

2. Расширена дескриптивная теория распознавания изображений введением четырех новых принципов: учет цели распознавания на начальных стадиях анализа, распознавание поведения динамических объектов, оценка предыстории, переменное количество объектов наблюдения, что позволяет повысить качество распознавания движущихся объектов за счет повышения информативности исходных данных.

3. Впервые разработан адаптивный пространственно-временной метод оценки движения в синхронных последовательностях видимого и инфракрасного диапазонов электромагнитного излучения, позволяющий извлекать признаки движения на различных иерархических уровнях, сочетая достоинства обоих типов последовательностей изображений.

4. Разработана новая модель многоуровневого движения, позволяющая проводить декомпозицию сцены на отдельные уровни, не ограничивающаяся общепринятым разделением на передний план и фон, что позволяет выполнять более достоверную сегментацию изображений объектов в сложных перспективных сценах.

5. Обоснован и построен новый обобщенный алгоритм сегментации динамических объектов с применением множества признаков, включающих предыстории поведения, и позволяющий отслеживать как динамику отдельных визуальных объектов, так и взаимодействия объектов в сцене (перекрытия проекций, появление/исчезновение объектов из поля зрения видеодатчика) на основе групповых преобразований и впервые предложенном анализе общей части проекций объекта (взятых из двух соседних кадров) с применением интегральных и инвариантных оценок.

6. Модифицирован метод коллективного принятия решений, отличающийся нахождением признаков межкадровых проекций объекта и позволяющий учитывать предысторию наблюдений для распознавания активных действий и событий на основе байесовской сети, а также предложены четыре вида псевдо-расстояний для нахождения меры сходства динамических образов с эталонными динамическими образами в зависимости от представления динамических признаков.

**Практическая значимость.** Предложенные в диссертационной работе методы и алгоритмы предназначены для практического применения при мониторинге автотранспортных средств при многополосном движении в рамках государственного проекта «Безопасный город», в системах автоматизированного контроля различных технологических производственных процессов по видеопоследовательностям, в системах наружного видеонаблюдения и видеонаблюдения в закрытых помещениях, а также в системах идентификации объектов на аэрофотоснимках и распознавании ландшафтных изображений. На основе диссертационных исследований разработаны программные комплексы обработки и распознавания динамических объектов, применяемые в различных сферах деятельности.

**Реализация результатов работы.** Семь разработанных программ зарегистрированы в Российском реестре программ для ЭВМ (перечень прилагается в списке основных работ автора).

Получены акты о передаче и использовании алгоритмического и программного обеспечения для распознавания корпусов холодильников на сборочной линии (ОАО КЗХ «Бирюса», г. Красноярск), для идентификации изображений объектов на ландшафтных изображениях (Концерн радиостроения «Вега», ОАО КБ «Луч», г. Рыбинск Ярославской области), для сегментации лесной растительности по набору последовательных аэрофотоснимков (ООО «Альтекс Геоматика», г. Москва), для обнаружения пластин государственных регистрационных знаков автотранспортных средств в видеопоследовательностях при многопоточном движении и повышении качества их отображения (УГИБДД ГУВД по Красноярскому краю, г. Красноярск).

Разработанные алгоритмы и программное обеспечение используются в учебном процессе при проведении занятий по дисциплинам «Интеллектуальная обработка данных», «Компьютерные технологии в науке и образовании», «Теоретические основы цифровой обработки изображений», «Распознавание образов», «Нейронные сети», «Алгоритмы обработки изображений», «Алгоритмы обработки видеопоследовательностей», «Анализ сцен и машинное зрение» в Сибирском государственном аэрокосмическом университете имени академика М.Ф. Решетнева (СибГАУ).

**Достоверность полученных в диссертационной работе результатов** обеспечивается корректностью используемых методов исследования, математической строгостью выполненных преобразований, а также соответствием сформулированных положений и выводов результатам их экспериментальной проверки.

**Основные положения, выносимые на защиту:**

1. Модель обработки и распознавания динамических образов в сложных сценах, существенно расширенная иерархическими уровнями сегментации и распознавания не только объектов, но и их активных действий.

2. Расширение дескриптивной теории распознавания изображений для временных рядов (последовательностей изображений) за счет повышения информативности анализируемых данных не только в пространственной области, но и по временной составляющей.

3. Адаптивный пространственно-временной метод оценки движения на основе тензорных представлений локальных 3D объемов в синхронных последовательностях видимого и инфракрасного диапазонов электромагнитного излучения.

4. Модель многоуровневого движения в сложных сценах, расширяющая декомпозицию перспективных сцен на отдельные уровни для более достоверного анализа траекторий движения объектов.

5. Обобщенный алгоритм сегментации динамических объектов, позволяющий на основе групповых преобразований и предложенных интегральных и инвариантных оценок выявлять перекрытия проекций объектов, появление/исчезновение объектов из поля зрения видеодатчика.

6. Методы распознавания динамических образов на основе модифицированного метода коллективного принятия решений и нахождения псевдо-расстояний в метрических пространствах, а также активных действий и событий в сложных сценах.

**Апробация работы.** Основные положения и результаты диссертационных исследований докладывались и обсуждались на 10 международной конференции «Pattern Recognition and Image Analysis: Modern Information Technologies», (S.-Petersburg, 2010), международном конгрессе «Ultra Modern Telecommunications and Control Systems ICUMT2010» (Moscow, 2010), XII международном симпозиуме по непараметрическим методам в кибернетике и системному анализу (Красноярск, 2010), II международном симпозиуме «Intelligent Decision Technologies – IDT 2010» (Baltimore, 2010), III международной конференции «Automation, Control and Information Technology – ACIT–ICT’2010» (Novosibirsk, 2010), 10-й, 11-й и 12-й международных конференциях и выставках «Цифровая обработка сигналов и ее применение» (Москва, 2008 – 2010 гг.), X международной научно-технической конференции «Теоретические и прикладные вопросы современных информационных технологий» (Улан-Удэ, 2009 г.), IX международной научно-технической конференции «Кибернетика и высокие технологии XXI века» (Воронеж, 2008), всероссийской конференции «Модели и методы обработки изображений» (Красноярск, 2007 г.), на X, XI и XIII международных научных конференциях «Решетневские чтения» (Красноярск, 2006, 2007, 2009 гг.), а также на научных семинарах Государственного университета аэрокосмического приборостроения (С-Петербург, 2009 г.), Института вычислительного моделирования СО РАН (Красноярск, 2009 г.), Института систем обработки изображений РАН (Самара, 2010).

**Публикации.** По результатам диссертационного исследования опубликовано 53 печатных работы, из них 1 монография, 26 статей (из них 14

статей – в изданиях, включенных в список ВАК, 2 статьи – в изданиях, перечисленных в «Thomson Reuters: Science Citation Index Expanded / Conference Proceedings Citation Index»), 19 тезисов докладов, 7 свидетельств, зарегистрированных в Российском реестре программ для ЭВМ, а также 3 отчета по НИР.

**Личный вклад.** Все основные результаты, изложенные в диссертации, включая постановку задач и их математические и алгоритмические решения, получены автором лично, или выполнены под его научным руководством и при непосредственном участии. Были защищены две диссертации на соискание ученой степени кандидата технических наук, при выполнении которых автор был официальным научным руководителем.

**Структура работы.** Работа состоит из введения, шести глав, заключения, библиографического списка. Основной текст диссертации содержит 326 страниц, изложение иллюстрируется 63 рисунками и 23 таблицами. Библиографический список включает 232 наименования.

## **СОДЕРЖАНИЕ РАБОТЫ**

Во **введении** обоснована актуальность диссертационной работы, сформулирована цель и поставлены задачи исследования, показана научная новизна и практическая ценность выполненных исследований, представлены основные положения, выносимые на защиту.

В **первой главе** представлена классификация задач обработки и распознавания изображений, методов оценки движения, алгоритмов сегментации движущихся объектов, методов интерпретации событий, а также рассматриваются существующие коммерческие аппаратно-программные комплексы и исследовательские разработки для систем видеонаблюдения.

Существуют три класса задач обработки и распознавания визуальной информации, классифицируемой как статические изображения, статические сцены с элементами движения и временные последовательности изображений. Последний случай является самым сложным, т. к. обладает большей информативной структурой, а динамические свойства объектов расширяют классическую постановку задач обработки и распознавания изображений, что делает непригодным использование ряда разработанных и хорошо зарекомендовавших себя классических методов распознавания.

Методы оценки движения в последовательностях изображений разделяются на сопоставительные и градиентно-ориентированные методы для различных групп физических процессов, объектов и ситуаций (табл. 1). Большинство методов основано на предположении, что локализация шаблонов движения проведена априори до распознавания и требуется установить соответствие шаблонов для оценки расположения, скорости движения и масштаба шаблона на изображении. Реализации других методов требует определения точных границы объектов и расположение их частей, как правило, для сцен с простым фоном.



Возможность автоматического формирования классов движения объектов по тестовым выборкам является сложной проблемой. Движения в последовательностях изображений с учетом их повторяемости во времени и пространстве можно разделить на временные текстуры, активные действия, события и сложные движения (табл. 2).

Таблица 1

Классификация методов оценки движения

Сопоставительные методы	Градиентно-ориентированные методы
	Динамические текстуры
	Метод анализа пространственно-временных фракталов Анализ на основе авторегрессионных функций
	Объекты с конечным набором проекций
Метод вычитания фона* Метод сопоставления блоков* Анализ функций плотности движения* Шаблоны движений	Метод отслеживания краевых точек Метод слежения за точечными особенностями Метод построения траекторий движения объектов
	Объекты с произвольным набором проекций
Метод вычитания фона* Метод сопоставления блоков*	Метод отслеживания краевых точек Метод слежения за точечными особенностями Метод оптического потока Метод построения глобальных и локальных траекторий движения объектов и их частей
	Активные действия и события
	Метод выявления относительных движений Прогнозирование траекторий движения объектов Метод построения графов событий сцены

Символом «\*» помечены быстродействующие, но менее точные методы оценки движения.

Таблица 2

Классы движения

Название	Краткое описание	Области применения
Временные текстуры	Статистические регулярности в пространстве и времени	Анализ турбулентности жидкостей и газов, распознавание ландшафтных изображений, анализ движения небольших однотипных объектов
Активные действия	Повторяемые во времени структуры	Сопровождение объектов, системы интерактивного взаимодействия человек–компьютер (при наличии стационарной видеокамеры)
События	Простые движения, не повторяющиеся во времени и пространстве	Анализ действий людей, поиск в цифровых библиотеках, анализ спортивных матчей, отслеживание чрезвычайных ситуаций, происшествий различного рода
Сложные движения	Динамические многоуровневые движения	Анализ сложных сцен при сопровождении объектов, навигации роботов, отслеживании чрезвычайных ситуаций (подвижная видеокамера)

Известные методы сегментации движущихся объектов условно можно разделить на три группы: сегментация с учетом априорных сведений об ожидаемых размерах и скоростях движения объектов, сегментация объектов, изображения которых описываются аффинной (реже проективной) группой преобразований, вероятностная сегментация объектов как задача с недостающими данными. Наиболее сложными проблемами являются появление и исчезновение объектов из поля зрения видеодатчика, перекрытия изображений объектов, «быстрые» движения объектов, многоуровневые движения в сцене.

Семантический анализ сцен развивается в трех основных направлениях: обнаружение необычных действий, категоризация действий человека и распознавание событий в сцене. Распознавание событий является ключевым моментом для семантического описания сцены. Одним из наиболее широко применяемых подходов является категоризация объектов и сцен с помощью «портфеля слов» (*Bag-of-Words*). В ряде работ события моделируются последовательностями гистограмм, построенных на основе визуальных признаков. Такие последовательности интерпретируются как строки (фразы), в которых отдельная гистограмма является символом. Классификация строк выполняется машиной опорных векторов. Словарь формируется путем кластеризации точек в пространстве признаков по методу  $k$ -средних. Размер словаря обычно определяется жанром сцены и определяется эмпирическим путем. Признаки вычисляются в отдельных кадрах, что уменьшает вычислительные затраты и делает их более простыми по структуре, однако такой подход не учитывает временные связи между различными признаками, что может привести к потере значимой информации.

Наиболее часто задача распознавание динамических объектов встречается в системах безопасности при распознавании лиц, несанкционированного проникновения людей на охраняемую территорию, в системах распознавания жестов с целью управления техническими средствами, в системах мониторинга транспортных средств различного назначения. Сравнительная характеристика некоторых коммерческих систем, предназначенных для распознавания объектов техногенного происхождения и использующих программно-аппаратные решения, приведена в табл. 3.

Завершается первая глава постановкой задачи пространственно-временной обработки последовательностей изображений. Низший уровень (I этап) характеризуется значительной изменчивостью формы регионов, по временным признакам регионов выявляется многоуровневое движение в сцене. Отдельные регионы изображения  $\mathbf{VR}$  объединяются в пространственно-временные структуры  $\mathbf{SR}_{LL}$ , низкого уровня:

$$\mathbf{SR}_{LL} = \{ \mathbf{VR} \{ \mathbf{SF}_R, \mathbf{TF}_R \}, \mathbf{IT}_R \} \quad ,$$

где  $\mathbf{SF}_R$  – множество пространственных признаков регионов;  $\mathbf{TF}_R$  – множество временных признаков регионов;  $\mathbf{IT}_R$  – множество отношений соседства регионов.

Средний уровень обработки пространственно-временной структуры  $SR_{ML}$  целесообразно представить в виде двух этапов: этап  $SR_{ML}^P$  выявление проекций визуальных объектов  $VO_P$  (II этап) и этап  $SR_{ML}^N$  нормализация проекций визуальных объектов  $VO_N$  (III этап):

$SR_{ML}^P = \{VO_P\{SF_P, TF_P\}, IT_{OE}\}$  ,  $SR_{ML}^N = \{VO_N\{SF_N, TF_N\}, IT_{OE}\}$  ,  
 где  $SF_P$  – множество пространственных признаков произвольных проекций объектов;  $TF_P$  – множество временных признаков произвольных проекций объектов;  $IT_{OE}$  – множество активных действий объектов;  $SF_N$  – множество пространственных признаков нормализованных проекций объектов;  $TF_N$  – множество временных признаков нормализованных проекций объектов.

Таблица 3

Характеристика систем распознавания динамических объектов

Название системы	Основные функциональные характеристики
Программно-аппаратная система «Вокорд» (РФ)	Мониторинг дорожного движения на городских улицах и автомагистралях, ситуация на парковках, идентификация лиц в толпе
Программно-аппаратный комплекс «ForeView» (РФ)	Предназначен для организации безопасности на территориально-распределённых объектах
Программно-аппаратный комплекс «VideoNet» (РФ)	Видеонаблюдение и аудиорегистрация, возможна интеграция с системами контроля доступа, определения автомобильных номеров, охранно-пожарной сигнализации, системой «Чек-ТВ» и другими системами
Аппаратно-программный комплекс «Ураган» (РФ)	Предназначен для регистрации номерных знаков автотранспортных средств при 4 полосах движения и архивации данных на сервере
Аппаратно-программный комплекс «Поток» (РФ)	Предназначен для считывания государственных регистрационных знаков движущихся транспортных средств и их автоматической проверки по базам данных (обслуживание платных парковок, терминалы оплаты проезда, пограничный контроль)
Распределенная интерактивная система «DIVA» (США)	Мониторинг дорожного движения на городских улицах и автомагистралях, сопровождение людей на улицах и в закрытых помещениях, идентификация лиц
Система «Tennis analyzer» (США)	Мониторинг теннисных матчей с возможностью семантического анализа игровых ситуаций

Следует различать объекты техногенного (с конечным количеством проекций) и антропогенного (с произвольным количеством проекций) происхождения. В последнем случае процедура выявления пространственно-временных признаков значительно усложняется. Учет пространственной составляющей дает возможность выявить множество активных действий объектов (см. табл. 2). Приведение проекций объектов к нормализованному виду позволяет упростить процедуру распознавания, а обнаруже-

ние активных действий и событий способствует повышению достоверности распознавания.

Высший уровень обработки предполагает отслеживание сложных событий, которые инициируют сигналы о внештатных ситуациях и рекомендации их устранения. Высший уровень обработки пространственно-временной структуры  $\mathbf{SR}_{HL}$  сцены включает этап  $\mathbf{SR}_{HL}^P$  распознавания динамических и статических образов  $\mathbf{VP}$  (IV этап) и этап  $\mathbf{SR}_{HL}^S$  распознавания жанра сцены, выявление сложных событий  $\mathbf{VE}$  при длительном видеонаблюдении (V этап):

$$\mathbf{SR}_{HL}^P = \{\mathbf{VP}\{\mathbf{SF}_O, \mathbf{TF}_O\}, \mathbf{IT}_E\}, \quad \mathbf{SR}_{HL}^S = \{\mathbf{VE}\{\mathbf{VP}, \mathbf{IT}_E\}\},$$

где  $\mathbf{SF}_O$  – множество пространственных признаков объектов;  $\mathbf{TF}_O$  – множество временных признаков объектов;  $\mathbf{IT}_E$  – множество событий сцены.

Представленная постановка задачи пространственно-временной обработки сложной сцены является системным многоуровневым подходом с идентичной формой описания пространственно-временных структур, но с принципиально различным содержанием на каждом уровне обработки.

Во **второй главе** диссертации представлены разработанные формальные модели обработки и распознавания объектов по их статическим изображениям и последовательностям изображений, а также расширена дескриптивная теория распознавания изображений.

В табл. 4 и 5 представлен порядок отображений в соответствующих пространствах, где символом « $\mathbf{I}$ » обозначено множество изображений, символом « $\mathbf{IS}$ » – множество последовательностей изображений, символом « $\mathbf{X}$ » – множество признаков, индексами  $R, F, P, O, Z$  – реальные, отфильтрованные (от шумов), проекционные, нормализованные и эталонные изображения или пространственные признаки, а индексами  $R_s, F_s, P_s, O_s, Z_s$  – реальные, отфильтрованные (от шумов), проекционные, нормализованные и эталонные последовательности изображений или пространственно-временные признаки.

Примем, что на множестве  $\mathbf{X}^R$  существует некоторое разбиение классов изображений  $\mathbf{V}^r_1, \mathbf{V}^r_2, \dots, \mathbf{V}^r_m$  такое, что

$$\mathbf{X}^R = \mathbf{V}^r_1 \cup \mathbf{V}^r_2 \cup \dots \cup \mathbf{V}^r_m, \quad \mathbf{V}^r_i \cap \mathbf{V}^r_j = \emptyset \quad \text{при } i \neq j.$$

Тогда формальную модель статического изображения можно представить как

$$\mathbf{I}^r_{ij} = S_p^{FR}(\mathbf{I}^F_{ij}, h) \quad ;$$

$$\mathbf{I}^f_{ij} = S_p^{PF}(S_p^{OP}(S_p^{ZO}(\mathbf{I}^z_{ij})))_{\beta} = S_p^{PF}(S_p^{OP}(\mathbf{I}^o_{ij}))_{\beta} = S_p^{PF}(\mathbf{I}^p_{ij})_{\beta} \quad , \quad (1)$$

где  $\mathbf{I}^f_{ij} \in \mathbf{I}^F$  – отфильтрованное (без учета шума) изображение  $i$ -го объекта  $j$ -го образа на входном поле изображения;  $i \in \mathbf{N}_O$  – переменный параметр, каждое значение которого определяет единственный объект наблюдения ( $\mathbf{N}_O$  – множество объектов наблюдения);  $S_p^{ZO}$  – функция описания нормализованных объектов;  $S_p^{OP}$  – функция описания объектов в различных проекциях;  $S_p^{PF}(\cdot)_{\beta}$  – функция описания изображений;  $\beta \in \mathbf{B}$  – параметр отображе-

ния, характеризующий неопределенность при отображении объектов наблюдения ( $\mathbf{B}$  – множество возможных значений  $\beta$ );  $h \in \mathbf{H}$  – шум, возникающий в процессе отображения ( $\mathbf{H}$  – множество возможных значений  $h$ );  $S_p^{FR}$  – функция, характеризующая воздействие шума. Физический смысл параметра  $\beta$  различен на каждом этапе описания изображений.

Таблица 4

Допустимые отображения при распознавании статических объектов

Отображения в пространстве изображений	Отображения в пространстве признаков	Краткая характеристика
$S_p^{RF}: \mathbf{I}^R \rightarrow \mathbf{I}^F$	$S_f^{RF}: \mathbf{X}^R \rightarrow \mathbf{X}^F$	Низший уровень обработки. I этап – предварительная обработка, выявление и анализ глобальных признаков изображения в целом
$S_p^{FP}: \mathbf{I}^F \rightarrow \mathbf{I}^P$	$S_f^{FP}: \mathbf{X}^F \rightarrow \mathbf{X}^P$	Средний уровень обработки. II этап – вычисление глобальных и локальных признаков изображений проекций объектов, сегментация
$S_p^{PO}: \mathbf{I}^P \rightarrow \mathbf{I}^O$	$S_f^{PO}: \mathbf{X}^P \rightarrow \mathbf{X}^O$	Средний уровень обработки. III этап – вычисление глобальных и локальных признаков изображений объектов, сегментация
$S_p^{OZ}: \mathbf{I}^O \rightarrow \mathbf{I}^Z$	$S_f^{OZ}: \mathbf{X}^O \rightarrow \mathbf{X}^Z$	Высший уровень обработки. IV этап – анализ глобальных и локальных признаков изображений объектов и распознавание объектов

Таблица 5

Допустимые отображения при распознавании динамических объектов

Отображения в пространстве последовательностей	Отображения в пространстве признаков	Краткая характеристика
$IS_p^{RsFs}: \mathbf{IS}^{Rs} \rightarrow \mathbf{IS}^{Fs}$	$S_f^{RsFs}: \mathbf{X}^{Rs} \rightarrow \mathbf{X}^{Fs}$	Низший уровень обработки. I этап – пространственная и временная фильтрация, выявление зон интереса (в частности, зон движения)
$IS_p^{FsPs}: \mathbf{IS}^{Fs} \rightarrow \mathbf{IS}^{Ps}$	$S_f^{FsPs}: \mathbf{X}^{Fs} \rightarrow \mathbf{X}^{Ps}$	Средний уровень обработки. II этап – вычисление глобальных и локальных признаков изображений проекций объектов, сегментация
$IS_p^{PsOs}: \mathbf{IS}^{Ps} \rightarrow \mathbf{IS}^{Os}$	$S_f^{PsOs}: \mathbf{X}^{Ps} \rightarrow \mathbf{X}^{Os}$	Средний уровень обработки. III этап – вычисление глобальных и локальных признаков изображений объектов с учетом предыстории, сегментация
$IS_p^{OsZs}: \mathbf{IS}^{Os} \rightarrow \mathbf{IS}^{Zs}$	$S_f^{OsZs}: \mathbf{X}^{Os} \rightarrow \mathbf{X}^{Zs}$	Высший уровень обработки. IV этап – анализ глобальных и локальных признаков изображений объектов с учетом предыстории, распознавание
$IS_p^{ZsEs}: \mathbf{IS}^{Zs} \rightarrow \mathbf{IS}^{Es}$	$S_f^{ZsEs}: \mathbf{X}^{Zs} \rightarrow \mathbf{X}^{Es}$	Высший уровень обработки. V этап – анализ глобальных и локальных событий с учетом предыстории, распознавание событий и жанра сцены

При  $i=\text{var}$  и  $j=\text{const}$  функция  $S_p^{ZO}$  описывает все возможные проекции объектов  $j$ -го образа  $\mathbf{I}_{ij}^p = S_p^{ZO}(\mathbf{I}_j)_i$ . Если же принять, что и  $i=\text{const}$ , то  $S_p^{ZO}(\mathbf{I}_j) = \mathbf{I}_j^p$ , т.е. в этом случае результатом воздействия функции является сам эталон  $j$ -го образа. Тогда модель процесса описания изображений существенно упрощается и принимает вид (при этом  $\beta \in \mathbf{B} \cup \mathbf{N}_0$ , т.е. множество возможных значений параметра  $\beta$  существенно расширяется):

$$\mathbf{I}_{ij}^p = S_p^{FR}(\mathbf{I}_j, h) ; \quad \mathbf{I}_j^p = S_p^{PF}(\mathbf{I}_j)_\beta . \quad (2)$$

Построим аналогичную формальную модель для последовательностей изображений. Особенность рассматриваемого класса задач состоит в том, что в них отображение  $IS_p^{PsFs} : \mathbf{IS}^{Ps} \rightarrow \mathbf{IS}^{Fs}$  и соответствующее ему обратное отображение  $IS_p^{FsPs} : \mathbf{IS}^{Fs} \rightarrow \mathbf{IS}^{Ps}$  зависят не только от мешающего параметра  $\beta$ , но и от времени  $t$ . Обозначим функцию отображения  $IS_p^{PsFs} : \mathbf{IS}^{Ps} \rightarrow \mathbf{IS}^{Fs}$ , изменяющуюся во времени как  $IS_p^{PsFs}(\cdot)_{\beta, t}$ , а изображение

$$\mathbf{IS}_j^{rs} = IS_p^{FsRs}(\mathbf{IS}_j^{fs}, h) , \quad \mathbf{IS}_j^{fs} = IS_p^{PsFs}(IS_p^{OsPs}(IS_p^{ZsOs}(\mathbf{IS}_j^{zs})_\gamma))_{\beta, t} , \quad (3)$$

назовем динамическим изображением (последовательностью изображений) объекта наблюдения, принадлежащего образу  $\mathbf{V}_j^{IS}$ , где  $\gamma$  – параметр, учитывающий предысторию поведения объекта. Можно рассмотреть упрощенную модель при  $\gamma = \text{const}$ . (Однако, следует помнить, что преобразование  $IS_p^{OsPs}$  становится намного сложнее из-за концепции многоуровневого движения в последовательности изображений.) Тогда

$$\mathbf{IS}_j^{rs} = IS_p^{FsRs}(\mathbf{IS}_j^{fs}, h) , \quad \mathbf{IS}_j^{fs} = IS_p^{PsFs}(IS_p^{OsPs}(\mathbf{IS}_j^{zs}))_{\beta, t} . \quad (4)$$

Такая модель, отображающая динамические характеристики объектов, существенно отличается от модели статического изображения (2). Обобщенная модель (4) отличается тем, что любые переходы  $(\beta_k \rightarrow \beta_v) \in \mathbf{B}^c$  обладают свойством гомоморфных преобразований, но переходы  $t_n \rightarrow t_{n+1}$  этим свойством не обладают. Именно эта особенность последовательностей изображений не позволяет применять обычные методы распознавания для их классификации, хотя постановка задачи формально не изменяет своего содержания.

Введем понятие динамического максимального инварианта для динамических изображений. Пусть  $\mathbf{IS}^{Rs}$  – последовательность кадров  $\{\mathbf{IS}_t^{rs}, \mathbf{IS}_{t+1}^{rs}, \dots, \mathbf{IS}_{t+N}^{rs}\}$ , где  $t$  – начальный, а  $t+1, t+2, \dots, t+N$  – последующие моменты наблюдения, причем  $t$  может быть выбрано в любой точке гиперповерхности. При этом любые две точки  $\mathbf{X}_{k,t}^{fs}$  и  $\mathbf{X}_{v,t}^{fs}$ , соответствующие одному и тому же образу, должны лежать на гиперповерхности, соответствующей заданному значению  $t$ . Тогда динамической инвариантной функцией будем называть любую функцию от  $\mathbf{IS}^{Rs}$ , если она не изменяется при изменении параметров  $\beta$  и  $t$ . Предположим, что для динамической инвариантной функции  $E$  при достаточно длинных последовательностях изображений  $\mathbf{IS}^{Rs}$  и  $(\mathbf{IS}^{Rs})^c$  равенство

$$E\{\mathbf{IS}^{Rs}\} = E\{\mathbf{IS}^{Rs}\}^c \quad (5)$$

выполняется только тогда, когда для каждой пары из серии наблюдений  $\{\mathbf{IS}^{Rs}\}$  и  $\{\mathbf{IS}^{Rs}\}^c$  существует такое преобразование  $g \in \mathbf{G}$  и всегда найдутся такие значения  $t$  и  $t^c$ , что

$$\begin{cases} \mathbf{IS}^{rs}_{k,t} = g\mathbf{IS}^{rs}_{v,t^c}, \\ \dots \\ \mathbf{IS}^{rs}_{k,t+N} = g\mathbf{IS}^{rs}_{v,t^c+N}. \end{cases} \quad (6)$$

Функцию  $E\{\mathbf{IS}^{Rs}\}$  назовем максимальным динамическим инвариантом, причем  $g \in \mathbf{G}$  – преобразования, порожденные всевозможными переходами из  $\mathbf{B}^c$ . Условия (5) и (6) выполняются только тогда, когда все наблюдения  $\{\mathbf{IS}^{Rs}\}$  и  $\{\mathbf{IS}^{Rs}\}^c$  принадлежат гиперплоскости при отсутствии шума или они сосредоточены в окрестности гиперплоскости, определяемой распределением шума  $h$ , и всегда существует возможность сдвинуть наблюдения во времени так, чтобы выполнялись равенства (6). При обучении распознаванию динамических образов задача определения полного множества преобразований  $\mathbf{G}$  существенно усложняется и практически не может быть решена без дополнительных ограничений.

Одним из решений является дескриптивный подход к распознаванию изображений, разработанный коллективом авторов под руководством академика Ю.И. Журавлева в 70-90-е гг. XX в. Подход предусматривает выбор основных эвристических процедур, построение соответствующей модели, параметризацию модели и переход к постановке и решению на эвристической модели оптимизационных задач. Однако дескриптивная теория распознавания изображений имеет дело со статическими изображениями. Распознавание временных рядов, в частности, последовательностей изображений, требует формулирования дополнительных принципов, расширяющих основные положения дескриптивного подхода. Сформулируем такие принципы: учет цели распознавания на начальных стадиях обработки последовательности изображений; распознавание поведенческих ситуаций динамических объектов; оценка предыстории динамических объектов; переменное количество объектов наблюдения в сложных сценах. Отметим, что динамические изображения являются, как правило, многоплановыми и характеризуются неоднозначностью отображаемых траекторий движения.

В **третьей главе** представлены условия и ограничения методов пространственно-временной обработки последовательностей изображений, классифицированы допустимые преобразования изображений, обоснован новый метод оценки движения в пространственно-временном наборе данных видимого и инфракрасного диапазонов излучения на основе тензорного подхода (I этап). В табл. 6 представлены основные ограничения пространственно-временной обработки и распознавания динамических объектов и событий в сложных сценах.

Введем ряд определений, необходимых для классификации движущихся объектов в последовательностях изображений.

*Определение 1.* Регионом с фиксированным набором проекций  $R_f$  называется компактная область изображения, обладающая постоянными цветовыми  $F_c$  и текстурными  $F_t$  характеристиками в определенных условиях освещения и имеющая ограниченный набор проекций  $P_s$ ,  $|P_s| \rightarrow \text{const}$ , во фронтальной плоскости, контур которых может изменяться в соответствии с аффинной или проективной группой преобразований.

Таблица 6

Условия и ограничения метода пространственно-временной обработки последовательностей изображений

Условия и ограничения	Критерий
Количество видеокамер	Одна видеокамера
Калибровка видеокамеры	Отсутствует
Подвижность видеокамеры	Допустимо перемещение видеокамеры относительно сцены
Диапазон электромагнитного излучения	Видимый, инфракрасный
Режим обработки	Режим <i>off-line</i> или режим <i>on-line</i> с заранее оговоренной задержкой
Количество целей наблюдения	Теоретически не ограничено
Количество фиксированных проекций объекта наблюдения	Не ограничено
Динамика движения	Равномерное и равноускоренное движение
Сложность фона	Фон может быть сложным с элементами движения
Сложность сцены	Движение в сцене может быть многоуровневым

*Определение 2.* Регионом с произвольным набором проекций  $R_c$  называется компактная область изображения, обладающая постоянными цветовыми  $F_c$  и текстурными  $F_t$  характеристиками в определенных условиях освещения и имеющая произвольный набор проекций  $P_d$ ,  $|P_d| \rightarrow \infty$ , во фронтальной плоскости с произвольно изменяющимся контуром.

Если видеодатчик установлен неподвижно, то сложная сцена представляет собой совокупность статических структур или статических и динамических структур. Если видеодатчик установлен на перемещающейся платформе, то сцена представляет набор динамических структур с признаками многоуровневого движения. В обоих случаях важно найти признаки движения, проследить динамику изменения структур и на основе накопленной информации выдвинуть гипотезу о принадлежности совокупности структур (как единого объекта) некоторому классу. Примем, что размеры динамических структур значительно меньше размеров изображения:

$$\int_{\Omega_{MS}} f(x, y, t) dx dy dt \ll \int_{\Omega} f(x, y, t) dx dy dt, \quad (7)$$

где  $x, y$  – пространственные координаты;  $t$  – время;  $f(x, y, t)$  – функция, описывающая пространственно-временной набор данных;  $\Omega_{MS}$  – множество



точек, принадлежащих динамическим регионам;  $\Omega$  – множество точек наблюдаемого изображения.

Новый метод оценки локальных признаков движения относится к методам оптического потока, но отличается тем, что оцениваются локальные движения регионов с применением  $3D$ -структурного тензора для последовательностей видимого диапазона и/или  $3D$  тензора потока для последовательностей инфракрасного диапазона. Оценка локальных параметров движения производится с использованием геометрических примитивов и особенных точек локального региона, что позволяет существенно снизить вычислительные затраты, исключив из вычислений геометрические примитивы и особенные точки фона. Новый метод оценки глобальных признаков движения визуальных объектов основан на выявлении периодического движения динамических структур и построении графа соседства динамических структур с целью последующей сегментации визуальных объектов с дополнительным привлечением текстурных, цветовых, геометрических, топологических и других признаков динамических регионов.

На стадии инициации требуется выявить локальные динамические структуры, претендующие на роль визуальных объектов. Примем, что остальная информация относится к фону сцены и в дальнейших измерениях не участвует. При этом процедуру выявления объектов интереса следует периодически повторять, поскольку они могут появляться и исчезать из поля зрения камер, и метеоусловия также могут изменяться. К данному этапу предъявляются требования по возможности быстрой, но грубой оценки движения. Особенностью метода является обобщенный анализ последовательностей изображений в различных диапазонах излучения.

Усовершенствованная модель фона изображения, содержащего объекты интереса на переднем плане, связана с одновременной выборкой  $N$  кадров и построении усредненных изображений  $I_{med}$  для обоих видов последовательностей. Для каждого пиксела с яркостью  $I_t(x,y)$ , где  $x, y$  – координаты пиксела, вычисляются взвешенные средние значения  $\mu(x,y)$  и отклонения  $\sigma^2(x,y)$  по  $N$  кадрам:

$$\mu(x,y) = \frac{\sum_{t=1}^N w_t(x,y) \cdot I_t(x,y)}{\sum_{t=1}^N w_t(x,y)}, \quad (8)$$

$$\sigma^2(x,y) = \frac{\sum_{t=1}^N w_t(x,y) \cdot (I_t(x,y) - \mu(x,y))^2}{1 - \frac{1}{N} \sum_{t=1}^N w_t(x,y)}, \quad (9)$$

где  $w_t(x,y)$  – весовые коэффициенты, используются для минимизации выбросов, максимально удаленных от среднего  $I_{med}$  и вычисляются по нормальному распределению, центрированному относительно величины  $I_{med}$ .

$$w_t(x, y) = \frac{1}{\sqrt{2\pi\sigma_{ex}^2}} \exp\left(-\frac{(I_t(x, y) - I_{med}(x, y))^2}{2\sigma_{ex}^2}\right). \quad (10)$$

Стандартное отклонение  $\sigma_{ex}^2$  можно рассчитать по  $N$  соседним кадрам. Использование взвешенных значений в статистической модели фона позволяет получить устойчивую модель фона без обучающих последовательностей. На основе статистической модели фона, полученной из инфракрасной последовательности (выражения (8) и (9)), можно получить маску регионов интереса  $D^{MS}$  по критерию квадрата расстояния Махаланобиса.

Зададим дискретную функцию  $\mathbf{I}^\Delta$ , определенную в точках трехмерной решетки  $\{(x_i, y_i, z_k) = \mathbf{p}_{ijk}\}$ , где  $x_i, y_i, z_k$  – координаты решетки по осям  $OX$ ,  $OY$ ,  $OZ$  соответственно,  $x_i \in [a_{OX}, b_{OX}]$ ,  $y_j \in [a_{OY}, b_{OY}]$ ,  $z_k \in [a_{OZ}, b_{OZ}]$ ,  $1 \leq i \leq i_0$ ,  $1 \leq j \leq j_0$ ,  $1 \leq k \leq k_0$ ,  $\mathbf{I}^\Delta(\mathbf{p}_{ijk}) = \mathbf{I}^\Delta_{ijk}$ . Построим математическую модель оценки движения регионов в последовательностях изображений. Найдем дифференцируемую функцию  $\mathbf{I}(x, y, z)$  такую, что в узлах решетки  $\mathbf{I}^\Delta(\mathbf{p}_{ijk}) = \mathbf{I}^\Delta_{ijk}$ . Обозначим произвольную точку из допустимого множества как  $\mathbf{p} = [x, y, t]$ , где  $x$  и  $y$  – пространственные координаты пикселей кадра по осям  $OX$  и  $OY$  соответственно, а  $t$  – временная координата, учитывающая последовательность появления кадров вдоль оси  $OZ$ . Тогда выражение для оценки пространственно-временного набора данных видеопоследовательности  $\mathbf{I}(\mathbf{p})$  (при условии постоянного освещения сцены) относительно положения некоторой локальной точки  $\mathbf{p}$  имеет вид:

$$\frac{\partial \mathbf{I}(\mathbf{p})}{\partial t} = \frac{\partial \mathbf{I}(\mathbf{p})}{\partial x} v_x + \frac{\partial \mathbf{I}(\mathbf{p})}{\partial y} v_y + \frac{\partial \mathbf{I}(\mathbf{p})}{\partial t} v_t = \nabla \mathbf{I}^T(\mathbf{p}) \mathbf{v}(\mathbf{p}), \quad (11)$$

где  $\mathbf{v}(\mathbf{p}) = [v_x, v_y, v_t]$  – вектор скорости потока относительно точки  $\mathbf{p}$ . Модуль вектора скорости  $\mathbf{v}(\mathbf{p})$  определяется из условия нахождения минимума функции (11) по локальному 3D объему  $\Omega(\mathbf{p}, \mathbf{q})$ , центрированному относительно вектора  $\mathbf{p}$ , где  $\mathbf{q}$  – локальная точка с координатами  $\mathbf{q} = [x_n, y_n, t_n]$ . Найдена оценка  $e_{ls}^s(\mathbf{p})$  скорости перемещения точки по методу наименьших квадратов в локальном объеме  $\Omega(\mathbf{p}, \mathbf{q})$  при условии, что  $\|\mathbf{v}(\mathbf{p})\| = 1$  и  $\|\nabla \mathbf{I}(\mathbf{q})\| = 1$ . Тогда

$$e_{ls}^s(\mathbf{p}) = \mathbf{v}^T(\mathbf{p}) \left( \int_{\Omega(\mathbf{p}, \mathbf{q})} (\nabla \mathbf{I}(\mathbf{q}) \nabla \mathbf{I}^T(\mathbf{q})) W(\mathbf{p}, \mathbf{q}) d\mathbf{q} \right) \mathbf{v}(\mathbf{p}) + \sum_{i=1}^N \alpha_i [1 - \mathbf{v}(\mathbf{p})^T \mathbf{v}(\mathbf{p})], \quad (12)$$

где  $\alpha_i$  – множители Лагранжа, функция  $W(\mathbf{p}, \mathbf{q})$  – это пространственно-инвариантная функция, имеющая распределение Гаусса и характеризующая изменения градиентов.

Предположим, что вектор скорости  $\mathbf{v}(\mathbf{p})$  является постоянным внутри пространственного объема  $\Omega(\mathbf{p}, \mathbf{q})$  и продифференцируем выражение (12). Подынтегральное выражение  $(\nabla \mathbf{I}(\mathbf{q}) \nabla \mathbf{I}^T(\mathbf{q}))$  можно представить как симметричное ядро скалярного произведения. В результате получим приближенную оценку вектора  $\hat{\mathbf{v}}(\mathbf{p})$

$$\mathbf{J}_s(\mathbf{p}, \mathbf{W}(\mathbf{p}))\hat{\mathbf{v}}(\mathbf{p}) = \sum_{i=1}^N \alpha_i \hat{\mathbf{v}}(\mathbf{p}) \quad ,$$

где матрица  $\mathbf{J}_s(\mathbf{p}, \mathbf{W})$  – 3D структурный тензор пространственно-временного набора данных, центрированного относительно вектора  $\mathbf{p}$ :

$$\mathbf{J}_s(\mathbf{p}, \mathbf{W}(\mathbf{p})) = \int_{\Omega(\mathbf{p}, \mathbf{q})} (\nabla \mathbf{I}(\mathbf{q}) \nabla \mathbf{I}^T(\mathbf{q})) \mathcal{W}(\mathbf{p}, \mathbf{q}) d\mathbf{q} \quad .$$

Проведено дополнительное исследование, показавшее, что при наличии больших шумовых выбросов квадратичная мера расхождения не обеспечивает устойчивость метода измерения ошибок. Для данного случая выведено выражение устойчивого структурного тензора Джомана-МакКлурра (*Geman-McClure*):

$$\mathbf{J}_s(\mathbf{p}, \mathbf{v}(\mathbf{p}), \mathbf{W}(\mathbf{p})) =$$

$$\int_{\Omega(\mathbf{p}, \mathbf{q})} \frac{m^2}{(\nabla \mathbf{I}^T(\mathbf{q}) \nabla \mathbf{I}(\mathbf{q}) - \mathbf{v}^T(\mathbf{p}) (\nabla \mathbf{I}(\mathbf{q}) \nabla \mathbf{I}^T(\mathbf{q})) \mathbf{v}(\mathbf{p}) + m^2)} (\nabla \mathbf{I}(\mathbf{q}) \nabla \mathbf{I}^T(\mathbf{q})) \mathcal{W}(\mathbf{p}, \mathbf{q}) d\mathbf{q} \quad ,$$

где  $m$  – параметр, определяющий штраф за большие выбросы ошибок.

Также рассмотрен вопрос о применении пространственно изменяемого ядра  $\mathcal{W}(\mathbf{p}, \mathbf{q})$ , адаптивного к размерам и ориентации окружения  $\Omega(\mathbf{p}, \mathbf{q})$ . Разработана процедура адаптации окружения  $\Omega(\mathbf{p}, \mathbf{q})$ , вначале имеющего форму круга, а затем в течение 2-3 итераций превращающегося в ориентированный эллипс для улучшения оценки градиентных структур при движении регионов.

Нахождение собственных значений тензора позволяет оценить градиент в окрестности точки  $\mathbf{p}$ , вычисляя след матрицы  $\mathbf{J}_s(\mathbf{p})$  и сравнивая его с пороговым значением. Далее для классификации динамических структур проведем анализ собственных значений матрицы  $\mathbf{J}_s(\mathbf{p})$ . Собственные значения  $\Lambda = \{\lambda_k\}$ , ( $k=1,2,3$ ) симметричной ковариационной матрицы  $\mathbf{J}_s(\mathbf{p})$  можно определить по локальным смещениям яркостей изображений соседних кадров и использовать для оценки локальных ориентаций динамических структур. Причем, в силу особенностей видеонаблюдения собственные значения  $\lambda_k$  указывают на локальные отклонения яркости по трем направлениям, и могут быть отсортированы в следующем порядке  $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$ . Выражение (11) можно рассматривать как матрицу, составленную из векторов градиентов в пространственно-временном наборе. В соответствии с методом главных компонент собственные значения этой матрицы сортируются в порядке убывания. Наибольшее собственное значение указывает направление наибольшего изменения данных. Отношение каждого собственного значения к сумме трех собственных значений характеризует концентрацию энергии по соответствующему направлению. Таким образом, собственные значения локального 3D структурного тензора можно использовать для обнаружения локальных изменений в последовательности кадров. Наименьшее собственное значение можно использовать для определения различий в кадрах, оно является более устойчивым к шу-

му и низко контрастным объектам фона по сравнению с простейшим методом яркостной разницы кадров. Далее на основе собственных значений  $\lambda_1(x,y,t)$ ,  $\lambda_2(x,y,t)$ ,  $\lambda_3(x,y,t)$  строятся карты  $\lambda_1(I)$ ,  $\lambda_2(I)$ ,  $\lambda_3(I)$  локального 3D структурного тензора. При этом карта собственных значений  $\lambda_1(I)$  фиксирует как динамические структуры, так и некоторые изолированные текстурные регионы фона. Карта собственных значений  $\lambda_2(I)$  является менее информативной, а карта собственных значений  $\lambda_3(I)$  генерирует небольшие разрывы внутри масок визуальных объектов. Для обнаружения движения основное внимание следует уделять собственному вектору  $\lambda_1(I)$ . Для нахождения динамических структур целесообразно вычислять коэффициент корреляции  $R$  между кадрами по критерию Пирсона, используя наименьшие собственные значения  $\lambda_3(I_t)$  и  $\lambda_3(I_{t+1})$  кадров  $Fr_t$  и  $Fr_{t+1}$ . Разброс коэффициентов корреляции кадров сцены позволяет оценить степень изменчивости формы динамических структур. Так, для регионов с фиксированным набором проекций он будет значительно меньше, чем для регионов с произвольным набором проекций.

Однако нахождение собственных значений для всех пикселей изображения сопровождается большими вычислительными затратами. Возможны два способа оценки локальных движений регионов. Первый способ основан на нахождении особых точек сцены (в отличие от существующих алгоритмов) только в найденных динамических структурах. Отметим, что между особенностями сцены и особенностями изображения есть соответствие: точечной особенностью сцены должна соответствовать точечная особенность изображений. Большинство детекторов точечных особенностей вычисляют некоторую функцию от окрестности точки и анализируют локальные максимумы. Существует целый набор функций, которые можно использовать для обнаружения точечных особенностей. Чаще всего для задач отслеживания точек сцены применяются функции, находящиеся в изображении структуры, похожие на углы.

Второй способ заключается в использовании инфракрасной последовательности данных, содержащей распределение тепловой энергии объектов. При этом воспользуемся тензором потока, который позволяет оценить движение объектов без вычисления собственных значений векторов 3D структурного тензора. Найдем вторую производную выражения (11) относительно переменной  $t$ . Тогда

$$\frac{\partial}{\partial t} \left( \frac{\partial \mathbf{I}(\mathbf{p})}{\partial t} \right) = \frac{\partial^2 I(\mathbf{p})}{\partial x \partial t} v_x + \frac{\partial^2 I(\mathbf{p})}{\partial y \partial t} v_y + \frac{\partial^2 I(\mathbf{p})}{\partial t^2} v_t + \frac{\partial I(\mathbf{p})}{\partial x} a_x + \frac{\partial I(\mathbf{p})}{\partial y} a_y + \frac{\partial I(\mathbf{p})}{\partial t} a_t$$

или в векторной форме

$$\frac{\partial}{\partial t} (\nabla \mathbf{I}^T(\mathbf{p}) \mathbf{v}(\mathbf{p})) = \frac{\partial \nabla \mathbf{I}^T(\mathbf{p})}{\partial t} \mathbf{v}(\mathbf{p}) + \nabla \mathbf{I}^T(\mathbf{p}) \mathbf{a}(\mathbf{p}) .$$

Используем аналогичный подход для нахождения ошибки  $e_{ls}^F(\mathbf{p})$ , полагая, что скорость является постоянной, а  $\|\mathbf{v}(\mathbf{p})\|=1$ . Тогда

$$e_{ls}^F(\mathbf{p}) = \mathbf{v}^T(\mathbf{p}) \left( \int_{\Omega(\mathbf{p}, \mathbf{q})} \left( \frac{\partial \nabla \mathbf{I}(\mathbf{p})}{\partial t} \cdot \frac{\partial \nabla \mathbf{I}^T(\mathbf{p})}{\partial t} \right) W(\mathbf{p}, \mathbf{q}) d\mathbf{q} \right) \mathbf{v}(\mathbf{p}) + \sum_{i=1}^N \alpha_i [1 - \mathbf{v}(\mathbf{p})^T \mathbf{v}(\mathbf{p})]. \quad (13)$$

При условии постоянной скорости в пространственном объеме  $\Omega(\mathbf{p}, \mathbf{q})$  выражение для  $3D$  тензора потока  $\mathbf{J}_F$  примет вид

$$\mathbf{J}_F(\mathbf{p}, \mathbf{W}(\mathbf{p})) = \int_{\Omega(\mathbf{p}, \mathbf{q})} W(\mathbf{p}, \mathbf{q}) \frac{\partial \nabla \mathbf{I}(\mathbf{p})}{\partial t} \cdot \frac{\partial \nabla \mathbf{I}^T(\mathbf{p})}{\partial t} d\mathbf{q}.$$

Элементы тензора потока содержат информацию об изменениях градиентов, что позволяет эффективно разделять статические и динамические структуры. След матрицы тензора потока имеет вид:

$$\text{trace}(\mathbf{J}_F(\mathbf{p})) = \int_{\Omega(\mathbf{p}, \mathbf{q})} \left\| \frac{\partial \nabla \mathbf{I}(\mathbf{p})}{\partial t} \right\|^2 d\mathbf{q}$$

и может непосредственно использоваться для классификации регионов по признаку движения без вычисления собственных значений статических регионов. При необходимости выражение (13) можно минимизировать для получения оценки вектора  $\hat{\mathbf{v}}(\mathbf{p})$ , используя равенство

$$\mathbf{J}_F(\mathbf{p}, \mathbf{W}(\mathbf{p})) \hat{\mathbf{v}}(\mathbf{p}) = \sum_{i=1}^N \alpha_i \hat{\mathbf{v}}(\mathbf{p}).$$

Для оценки движения следует вычислить производные

$$I_{xt}(\mathbf{p}) = \frac{\partial^2 I(\mathbf{p})}{\partial x \partial t}, \quad I_{yt}(\mathbf{p}) = \frac{\partial^2 I(\mathbf{p})}{\partial y \partial t}, \quad I_{tt}(\mathbf{p}) = \frac{\partial^2 I(\mathbf{p})}{\partial t \partial t}$$

и провести интегрирование по области  $\Omega(\mathbf{p}, \mathbf{q})$ . Эти производные вычисляются как свертки изображений инфракрасной последовательности с ядрами сепарабельных фильтров, которые можно представить в виде каскадных  $1D$  сверток. При этом сглаживающие фильтры позволяют снизить влияние шумов. Для вычисления производных  $I_{xt}$ ,  $I_{yt}$  и  $I_{tt}$  вычисляются пространственные свертки  $I_{xs}$ ,  $I_{ys}$  и  $I_{ss}$ , где  $s$  – сглаживающий фильтр. Для каждого входного кадра вычисляются и запоминаются наборы  $I_{xs}$ ,  $I_{ys}$  и  $I_{ss}$ . Затем при накоплении достаточного количества данных по нескольким входным кадрам находится сумма производных  $I_{xt}^2 + I_{yt}^2 + I_{tt}^2$ , по которой и строится маска движения  $MF_M$ . Маску  $MF_M$ , полученную с применением  $3D$  тензора потока, можно использовать для построения маски движения регионов  $MS_M$  видеопоследовательности кадров (т.е. найти и проанализировать собственные значения векторов матрицы  $\mathbf{J}_s(\mathbf{p})$  в локальной области кадра видеопоследовательности в соответствии с маской  $MF_M$ ).

В четвертой главе построена многоуровневая модель движения на основе структур движения, разработан обобщенный алгоритм сегментации объектов в сложных сценах (II и III этапы), учитывающий сложные случаи сегментации (перекрытия изображений, появление и исчезновение объектов из поля зрения камеры, движение на камеру).

Последовательности изображений часто состоят из структур движения, имеющих сходные параметры скоростей и ускорений. Требуется оп-

ределить набор уровней движения, которые сопоставляются с объектами, расположенными на различных расстояниях от движущейся камеры. Пусть для каждой структуры движения текущего кадра имеется локальный вектор движения, соединяющий его со структурой движения следующего кадра. Известен набор различных параметрических уровней движения сложной сцены. Вектор глобального движения определяется по локальным векторам связанных структур движения.

*Определение 3.* Статической структурой  $SS$  называется неподвижный регион изображения  $R_f, R_c$  или связанная совокупность регионов изображения в пространственно-временном объеме.

*Определение 4.* Динамической структурой  $DS$  называется регион изображения  $R_f, R_c$  или связанная совокупность регионов изображения в пространственно-временном объеме, находящаяся в одном из трех возможных состояний:

1. Состояние движения характеризуется локальным вектором скорости поступательного, вращательного или периодического движения во фронтальной плоскости наблюдения и описывается изоморфными преобразованиями проекций.

2. Состояние пересечения регионов возникает, когда состояние одного региона (по п. 1) не изменяется, а состояние других регионов описывается локальными гомоморфными преобразованиями проекций до тех пор, пока состояние пересечения не завершится.

3. Состояние пересечения с границей поля наблюдения возникает, когда состояние региона описывается локальными гомоморфными преобразованиями проекций до тех пор, пока регион не появится в поле наблюдения целиком или полностью не исчезнет из поля наблюдения.

В данной модели определяется набор различных, но внутренне подобных уровней движения, которые отвечают твердым телам, расположенным на различном расстоянии от движущейся камеры и  $g$  сегментам на изображении. Предположим, что уровень движения задан в параметрической форме, и имеется  $h$  уровней движения. Для текущего и последующего кадров требуется определить: а) уровень движения, к которому относится каждая структура движения; б) значения параметров каждого уровня. В такой постановке имеем задачу с недостающими данными. Отсутствующими данными являются значение уровня, к которому принадлежит структура движения, или параметры уровней.

Пусть некая структура движения в точке с параметрами  $(u, v)$   $i$ -го текущего кадра  $DS_i$  принадлежит  $l$ -му уровню движения, имеющему параметры  $\theta_l$ . Это означает, что на последующем  $(i+1)$ -м кадре структура движения  $DS_{i+1}$  переместится в точку с параметрами  $(u, v) + \mathbf{m}(u, v, \theta_l)$  с точностью до ошибки измерения  $C$ . Обозначим уровень движения переменной  $V_{uv,l}$ , которая принимает значение 1, если структура движения относится к  $l$ -му уровню движения, и 0 в противном случае. Предполагается, что на

значения яркостей пикселей изображения влияет гауссов шум с дисперсией  $\sigma$ . Тогда логарифм функции правдоподобия запишется как

$$L(V, \theta) = -\sum_{uv,l} V_{uv,l} \frac{(DS_i(u, v) - DS_{i+1}(u + m_u(u, v, \theta_l), v + m_v(u, v, \theta_l)))^2}{2\sigma^2} + C, \quad ,$$

где  $\theta = (\theta_1, \dots, \theta_g)$ . Далее определим вероятности уровней движения

$$P(V_{uv,l} = 1 | DS_i, DS_{i+1}, \theta), \quad ,$$

которые представляют собой карты яркостного представления максимально вероятного уровня для каждой структуры движения. Карта получается в процессе кластеризации локальных оценок движения на изображении. Каждый уровень яркости определяет уровень движения, а каждый уровень движения связан со своей моделью движения, в качестве которой обычно выступает аффинная модель

$$\begin{pmatrix} m_1 \\ m_2 \end{pmatrix} (u, v, \theta_l) = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} + \begin{pmatrix} a_{13} \\ a_{23} \end{pmatrix}, \quad ,$$

где  $\theta_l = (a_{11}, \dots, a_{23})$ . Карту можно упростить, оставив только выбранный уровень движения, для того, чтобы проверить, насколько движение соседних пикселей согласуется с движением пикселей на предыдущих и последующих кадрах. Модель многоуровневого движения связывает пиксели, «движущиеся» в одном направлении, имеется возможность определения границ движения, также последовательности кадров можно восстанавливать по имеющимся уровням движения, когда некоторые уровни отсутствуют.

*Определение 5.* Под динамическим визуальным объектом  $VO_D$  понимается совокупность динамических структур  $DS$ , обладающих постоянными на некотором временном интервале относительными направлениями и скоростями движения.

*Определение 6.* Под статическим визуальным объектом  $VO_S$  понимается совокупность статических структур  $SS$ , обладающих несущественными признаками движения или отсутствием движения.

Рассмотрим сложные случаи сегментации динамических и статических структур с использованием групповых преобразований, используя полученные в 3 главе признаки движения регионов. Зададим топологическое отображение  $\varphi$  некоторой окрестности  $\Omega_U$  единицы группы  $G$  на область  $\Omega_V$  координатного евклидова пространства  $\mathcal{R}^E$ , при котором единица переходит в начало координат. Тогда каждой точке  $p \in \Omega_U$  локальной группы  $G$  соответствует система действительных чисел  $p^1, \dots, p^r$ , являющихся координатами точки  $\varphi(p) \in \mathcal{R}^E$ . Назовем такие числа координатами точки  $p \in \Omega_U$  (при этом единица группы получает координаты, равные нулю). И наоборот, каждой системе чисел  $p^1, \dots, p^r$  (если числа достаточно малы по модулю) соответствует определенная точка  $p \in \Omega_U$ , имеющая своими координатами эти числа. Пусть  $\Omega_W$  – столь малая окрестность единицы группы

$G$ , что для любых элементов  $p$  и  $q$  из окрестности  $\Omega_W$  определено произведение  $pq \in \Omega_U$ . Тогда

$$s^i = f^i(p, q) = f^i(p^1, \dots, p^r; q^1, \dots, q^r) \quad , \quad (14)$$

где функции  $f^i$  в правых частях – непрерывные однозначные функции, определенные для достаточно малых значений аргументов. Координаты  $p^1, \dots, p^r$  локальной группы  $G$  называются дифференцируемыми, если функции (14) трижды непрерывно дифференцируемы.

Следствием теоремы об инвариантности преобразований координат группы Ли, доказанной Л. С. Понтрягиным, является следующее утверждение: всякий локальный автоморфизм  $\phi$  локальной группы  $G$  описывается в системе координат уравнениями

$$p'^j = \phi^j(p^1, \dots, p^r) \quad , \quad i=1, \dots, s$$

с трижды непрерывными дифференцируемыми (аналитическими) функциями  $\phi^j$  и с отличным от нуля функциональным определителем

$\left| \frac{\partial \phi^j}{\partial p^i} \right|_{p^1 = \dots = p^r = 0}$ . Если функции  $D$  и  $D'$  являются каноническими первого рода,

то функции  $\phi^j$  линейны. Внутренний автоморфизм группы запишется как  $\phi_a(p) = a^{-1}pa$ , где  $a$  – фиксированный элемент группы  $G$ . Функции яркости многоградационных изображений попадают под определения трижды непрерывных дифференцируемых функций (хотя на практике достаточно иметь дважды дифференцируемые функции бинарных масок), а метод оценки движения межкадровой разницы, представленный в 3 главе, позволяет установить области смещения объекта и найти «неподвижную» область, являющуюся внутренним автоморфизмом групповых координат, описывающих данный объект.

Легко показать, что евклидово пространство  $\mathfrak{R}^E$  является топологическим пространством, на котором замыкание  $\overline{\mathbf{M}}$  множества  $\mathbf{M}$  определяется как совокупность всех точек, принадлежащих  $\mathbf{M}$  или являющихся предельными для  $\mathbf{M}$ . Пусть  $G'$  и  $G$  – две линейно связанные топологические группы пространства  $\mathfrak{R}^E$ , соответствующие проекциям одного и того же объекта на двух соседних кадрах, причем группа  $G$  односвязна и локально связна. Пусть  $f$  – некоторый локальный гомоморфизм группы  $G$  в группу  $G'$ . Тогда можно продолжить локальный гомоморфизм  $f$  в гомоморфизм  $\phi$  всей группы  $G$  в группу  $G'$  единственным образом. Продолжение гомоморфизма  $f$  понимается в том смысле, что  $f$  и  $\phi$  совпадают на некоторой окрестности  $W$  единицы группы  $G$ . Действительно, в окрестности  $W$  функции  $f$  и  $\phi$  совпадают, а т. к. функция  $f$  непрерывна, то функция  $\phi$  непрерывна всюду. Если группа  $G'$  односвязна и локально связна, а функция  $f$  есть некоторый локальный изоморфизм, то гомоморфизм  $\phi$  является изоморфизмом. Таким образом, подгруппы  $G, G', \dots$ , представляющие собой проекции одного и того же объекта в ортогональной плоскости, близ-



кие в силу малого межкадрового смещения, являются, либо изоморфными при постоянном направлении и скорости относительного движения, либо гомоморфными при смене направления движения. Более того, выявленные зоны смещения проекции также являются компактными подгруппами, по степени изменения которых можно судить о характере движения отдельных частей объектов (регионов).

Далее, при перекрытиях проекций нескольких объектов во временном ряде можно проследить порождение локальных гомоморфизмов проекций объектов заднего плана в глобальный гомоморфизм вплоть до полной потери проекции. Очевидно, что аналогичная ситуация наблюдается при появлении или исчезновении проекции объекта из ограниченного поля зрения датчика. Тогда обобщенный алгоритм сегментации имеет вид:

*Предсегментация пространственно-временного набора данных*

Шаг 1.1. Построение гауссовой пирамиды и выбор слоя с малым разрешением для набора кадров  $\mathbf{IS}^{Fs}$ .

Шаг 1.2. Грубая оценка движения методом сопоставления блоков в пространственно-временном наборе данных.

Шаг 1.3. Оценка фона по модели (8,9).

Шаг 1.4. Вычисление признаков цветности, текстуры, формы регионов изображения в пространственном наборе.

Шаг 1.5. Учет статических структур  $\{SS\}$  и структур движения  $\{DS\}$ .

*Сегментация пространственно-временного набора данных*

Шаг 2.1. Для статических структур  $\{SS\}$  пространственная сегментация на основе графов, маркировка значимых регионов.

Шаг 2.2. Для динамических структур  $\{DS\}$  вычисление локальных 3D структурных тензоров и 3D тензоров потока, нахождение собственных значений матриц  $\mathbf{J}_S(\mathbf{p})$  и  $\mathbf{J}_F(\mathbf{p})$ , вычисление коэффициента корреляции  $R$  между кадрами и разброса коэффициентов корреляции.

Шаг 2.3. Нахождение регионов с фиксированным и произвольным наборами проекций  $\{R_f\}$  и  $\{R_c\}$  (определения 1, 2).

Шаг 2.4. Для динамического объекта  $VO_D$  при наличии регионов  $\{R_f\}$  анализ изоморфизма проекций объекта и локального гомоморфизма при пересечениях проекций нескольких объектов.

Шаг 2.5. Для динамического объекта  $VO_D$  при наличии регионов  $\{R_c\}$  анализ локального гомоморфизма проекций объекта и глобального гомоморфизма при резкой смене направления движения объекта или при пересечениях проекций нескольких объектов.

Шаг 2.6. Повторение шагов 2.1–2.5 для выявления нового движения в сцене. Хранение предыстории движения объектов.

*Пост-сегментация пространственно-временного набора данных*

Шаг 3.1. Окончательная оценка размеров и формы динамических объектов по пространственным признакам.

Шаг 3.2. Оценка признаков движения по пороговым значениям и классификация динамических объектов по уровням движения.

Шаг 3.3. Выбор проекций  $\mathbf{IS}^{Ps}$  объектов для нормализации и получения  $\mathbf{IS}^{Os}$ . Использование метода шаблонов при пересечениях проекций нескольких объектов или прекращение сегментации на временном интервале. Построение траекторий движения (при необходимости с предсказанием).

В пятой главе рассматривается процесс распознавания динамических образов (IV этап), использующий формальную грамматику, видеограф сцены и основанный на модифицированном методе коллективного принятия решений, а также методы распознавания активных действий и событий в сложной сцене (V этап) на основе графов активных действий, графов событий и байесовской сети.

Признаки динамических объектов, определяемые на этапе сегментации, условно можно разделить на статические (цвет, текстура сегментов изображений) и динамические (форма, траектория движения) характеристики. На этапе обучения объектов целесообразно представлять наборы признаков как эталонные графы или деревья грамматического разбора в рамках контекстной грамматики. При этом каждый признак – это множество значений признака и множество временных отсчетов, в течение которых он существовал. Для каждого объекта составляется свое дерево вывода в терминах контекстной грамматики  $\mathbf{GR}_D$ :

$$\mathbf{GR}_D = \langle \mathbf{V}_{O,E,G}, \mathbf{S}_{S,LM,GM}, \mathbf{T}_V, \mathbf{T}_S, \mathbf{T}_E, R_E, \{ \{ \mathbf{E}, \mathbf{E}(e,e'); (e,e') \in \mathbf{DSCR}_e \}, R_O, \{ R_R, \{ \mathbf{Z}, \mathbf{Z}(t,t'); (t,t') \in \mathbf{DSCR}_t \} \} \rangle .$$

где  $\mathbf{V}_{O,E,G}$  – основной словарь объектов, временных событий, жанров сцены;  $\mathbf{S}_{S,LM,GM}$  – вспомогательный словарь структурных элементов, локальных признаков движения и глобальных признаков движения;  $\mathbf{T}_V$  – множество производных элементов;  $\mathbf{T}_S$  – множество описаний;  $R_R$  – предикат построения регионов;  $R_O$  – предикат построения объектов;  $R_E$  – предикат временных событий. Элемент множества  $\mathbf{E} = \mathbf{V}_{O,E,G} \cup \mathbf{S}_{S,LM,GM}$  называется событием. Множество  $\mathbf{T}_E$  описывает набор событий. Множество  $\mathbf{T} = \mathbf{T}_V \cup \mathbf{T}_S \cup \mathbf{T}_E$  в данном случае представляет собой объединение события и описания. Множество  $\mathbf{E}(e,e')$  – множество временных событий;  $(e,e') \in \mathbf{DSCR}_e$  – описание временного события. Множество  $\mathbf{Z}(t,t')$  – множество временных регионов изображения;  $(t,t') \in \mathbf{DSCR}_t$  – описание временного региона.

При распознавании объектов с ограниченным допустимым количеством проекций можно предложить следующую формальную схему. Предположим, что каждый образ представлен только одним изображением (эталон образа)  $v^j$ . Пусть задано множество допустимых преобразований  $G_\beta$  исходного эталона, параметризованное по мешающему параметру  $\beta$ . Результатом применения преобразования  $G_\beta$  к эталону  $v^j$  является преобразованный эталон

$$V(j,\beta) = G_\beta v^j .$$

Соответственно для распознавания объектов с произвольным количеством проекций используется рекуррентная процедура нахождения эталонного изображения  $V_i^*(X)$  на  $i$ -м шаге, порождаемого грамматикой  $GR_D$ , которая максимизирует функцию сходства следующего вида:

$$V_i^*(X) = V_{i-1}^*(X) + \gamma_i f_V(X_{i-1}, \Delta_i) ,$$

где  $i, i-1$  – шаги аппроксимации;  $\gamma_i$  – некоторая функция, зависящая от шага аппроксимации (например, последовательность положительных чисел);  $\Delta_i$  – изменяемая в процессе анализа последовательности изображений величина;  $V_i^*(X) = \arg \max_j P(X/V(j))$  – функция сходства на  $i$ -м шаге;  $P(X/V(j, \beta))$  – распределение вероятностей многомерной случайной величины, описывающей наблюдаемые изображения.

На этапе распознавания строится видеограф по методу иерархического группирования, когда на основе признаков низшего уровня формируются локальные пространственные структуры, устойчивые во времени, локальные пространственные объекты, совокупность присущих им действий, а также пространственно-временные связи между ними. Модификация коллективного метода принятия решений основана на допущении, что существует не  $L$  различных алгоритмов решения задачи, а один алгоритм, но рассчитываются  $L$  решений в дискретные моменты времени наблюдения за динамическим объектом. Такой подход подразумевает двухуровневую процедуру распознавания: распознавание принадлежности изображения той или иной области компетентности и выбор решающего правила, компетентность которого максимальна в заданной области. Построены четыре вида псевдо-расстояний (на основе расстояний Хаусдорффа, Громова-Хаусдорффа, Фреше и метрического расстояния) для нахождения меры сходства входных динамических образов с эталонными динамическими образами в зависимости от представления динамического признака – множества числовых характеристик, множества векторов, множества функций.

При распознавании событий видеограф сложной сцены расширяется до видеографа событий. При этом событие представляет собой совокупность активных действий, и любое событие можно рассматривать как временную последовательность зависимых активных действий. Построена объектно-зависимая модель динамического объекта. В качестве функции соответствия используются простейшие классификаторы в пространстве признаков (например, по методу  $k$ -средних), т. к. сопоставление осуществляется не по всему набору шаблонов, а по ограниченному множеству шаблонов, ассоциированных с ранее опознанным объектом. Рассмотрены способы формирования шаблонов проекций визуальных объектов.

Видеограф событий строится на основе сетей Маркова. Подробно рассмотрены способы выявления активных действий агентов, порядок построения и разрезания видеографа событий для распознавания событий. Обнаружение событий в поступающих видеопоследовательностях сводит-

ся к кластеризации последовательно выполняемых активных действий на основе байесовского подхода. Выполняется рекурсивное разрезание матрицы весовых коэффициентов во входной видеопоследовательности и сравнение с эталонными событиями, полученными на этапе обучения. Данная информация является исходной для определения жанра сцены и при необходимости индексирования видеопоследовательности в базе данных. Разработана схема понимания и интерпретации изображений и видео материалов для индексирования в мультимедийных Интернет-базах.

В шестой главе приведены схемы экспериментального программного комплекса «*SPOER*», v.1.02 по обработке последовательностей изображений и распознаванию движущихся объектов и событий, показаны результаты его тестирования, а также полученные экспериментальные оценки на различных этапах обработки информации. На базе экспериментального комплекса «*SPOER*», v.1.02 были разработаны системы обработки видеoinформации различного целевого назначения, описания которых также приведены в главе. В табл. 7 приведено краткое описание основных модулей системы.

Таблица 7

Модули программного комплекса «*SPOER*» и их назначение

Название модуля	Функциональная характеристика
1. Модуль организации интерфейса с пользователем	Обеспечивает взаимодействие с пользователем и координирует работу всех модулей
2. Модуль предварительной обработки	Осуществляет фильтрацию последовательностей изображений, включает градиентные методы обработки
3. Модуль оценки движения	Находит динамические структуры на наборе изображений, рассчитывает и выдает результаты грубой и точной оценки движения, определяет степень изменчивости динамических структур
4. Модуль сегментации	Осуществляет сегментацию статических и динамических структур в пространственно-временном наборе данных, проводит нормализацию проекций, строит видеограф сцены
5. Модуль распознавания визуальных объектов	Выполняет распознавание объектов на основе структурного подхода, модифицированного метода коллективного принятия решений
6. Модуль распознавания активных действий	Осуществляет распознавание активных действий и событий на основе сетей (в частности, сетей Маркова)
7. Конфигурационный модуль	Осуществляет сохранение и загрузку параметров системы в конфигурационный файл

Экспериментальные результаты оценки движения, сегментации и распознавания объектов, полученные на тестовых последовательностях изображений («*Hamburg taxi*», «*Rubik cube*», «*Silent*», видеопоследовательности и инфракрасные последовательности из тестовой базы «*OTCBVS*») и собственных видеоматериалах, приведены в таблице 8. В таблице 9 пред-

ставлены сведения о точности распознавания некоторых активных действий людей на примерах из тестовых баз «PETS», «CAVIAR», «VACE».

Как видно из табл. 8, характер последовательности влияет на показатели. Хуже распознаются объекты, осуществляющие вращательное движение («Rubik cube»), лучше – техногенные объекты небольших размеров («Hamburg taxi», «Видео 1»). Наилучшие результаты показывает распознавание по обоим типам последовательностей, когда для сегментации локального движения применяется инфракрасная последовательность, а для оценки движения – видеопоследовательность. Экспериментальные результаты, представленные в табл. 9, демонстрируют лучшую точность распознавания периодических активных действий людей, не находящихся в группах (хождение, бег, поднятие рук). Ложные срабатывания обусловлены засветкой и наличием теней в ряде мест сцены. Приведенные в таблицах 8 и 9 цифры характеризуют обобщенные результаты применения разработанных алгоритмов ко всем кадрам тестовых последовательностей.

Таблица 8

Обобщенные результаты экспериментальных исследований

Последовательность \ Характеристика	Оценка движения		Сегментация динамических структур		Распознавание объектов	
	Точность, %	Ложное срабатывание, %	Точность, %	Ложное срабатывание, %	Точность, %	Ложное срабатывание, %
«Hamburg taxi»	93,82	2,76	91,46	3,28	94,63	2,51
«Rubik cube»	82,61	3,81	80,52	4,16	87,91	3,37
«Silent»	92,87	2,54	84,29	3,91	91,56	3,19
Thermal «Sequence 6a» из «OTCBVS»	98,15	2,18	70,21	9,52	75,39	6,83
Color «Sequence 6b» из «OTCBVS»	90,29	2,63	89,82	3,64	93,18	2,94
«Sequence 6a» и «Sequence 6b»	94,94	2,51	92,66	3,08	96,13	2,22
«Видео 1»*	90,12	2,94	90,82	2,89	93,32	2,18
«Видео 2»**	91,28	3,29	89,92	3,32	92,16	2,67
«Видео 3»***	73,04	2,12	70,25	3,94	79,62	8,82

\* – Видео с малым количеством динамических структур, до 10% от площади кадра

\*\* – Видео со средним количеством динамических структур, до 40% от площади кадра

\*\*\* – Видео с большим количеством динамических структур, до 100% от площади кадра

Также были созданы прикладные проекты с применением разработанных методов. Проект «Визуальная регистрация государственных номерных знаков автотранспортных средств при многопоточном движении» предназначен для обнаружения автотранспортных средств и локализации местоположения номерных пластин по данным из видеопоследовательности. Его основными функциями являются: считывание и воспроизведение

данных с регулируемым периодом, выбор цветовой модели, формирование карты движения на основе контурных и цветовых представлений данных, сегментация объектов движения, анализ характеристических особенностей автотранспортного средства и выделение номерной пластины на основе матрицы спектральных характеристик, улучшения качества изображения номерной пластины.

Таблица 9

Точность распознавания активных действий людей в сложных сценах

Последовательность Активные действия	«Meeting»		«Survelence1»		«Survelence2»	
	Точность, %	Ложное срабатывание, %	Точность, %	Ложное срабатывание, %	Точность, %	Ложное срабатывание, %
Поднятие рук	90,15	2,16	86,94	3,43	90,37	2,07
Рукопожатие	87,42	3,41	–	–	–	–
Ходьба	91,67	1,84	90,79	2,11	91,15	1,94
Остановка	92,04	0,67	89,12	1,04	90,66	0,85
Обмен	76,81	2,83	69,85	4,02	73,49	3,48
Бег	–	0,92	87,73	1,93	90,32	1,46

Проект «Система идентификации моделей корпусов холодильников по изображениям» позволяет как по отдельным изображениям корпусов холодильников, так и по видеопоследовательности идентифицировать модель холодильника. Программный продукт производит следующую последовательность действий: предварительная обработка изображения, улучшение яркостных характеристик изображения; поиск объекта на изображении или наборе соседних кадров видеопоследовательности (с целью выбора «наилучшего» кадра); обработка изображения, включающая бинаризацию, выделение контуров, устранение нежелательных визуальных эффектов, нормализацию изображения (сдвиги, повороты и изменения масштаба); поиск признаков модельного ряда (наличие камер, ребер и технологических отверстий, определение их количества и взаимного положения); принятие решения о модели корпуса холодильника.

В программной системе «Алгоритмы обработки и сегментации ландшафтных изображений. Идентификация объектов» ландшафтные изображения интерпретируются как совокупность текстурных фрагментов естественного происхождения и изображений антропогенных объектов. Разработана комплексная система оценки цветовых, статистических и фрактальных признаков ландшафтных изображений. Причем, используется оригинальный метод выявления техногенных объектов на ландшафтных изображениях по признаку отсутствия показателя фрактальной размерности, вычисленной в локальных областях границ техногенных объектов. Программная система имеет модификацию для сегментации и распознавания лесной растительности по последовательности аэрофотоснимков, в ос-

нову которой положена двухуровневая процедура сегментации и распознавание текстуры трехуровневой нейронной сетью прямого распространения.

**В заключении** сформулированы основные результаты и выводы, полученные в диссертационной работе.

## **ОСНОВНЫЕ РЕЗУЛЬТАТЫ И ВЫВОДЫ**

1. Построены формальные модели обработки и распознавания пространственно-временных структур на основе иерархической процедуры обработки последовательностей изображений, отличающиеся тем, что в них учтены изоморфные и гомоморфные преобразования и выведены обобщенные функции статических и динамических инвариантов. Также построены модели поиска статических и динамических признаков объектов для четырех задач анализа последовательностей изображений в зависимости от наличия движущегося видеодатчика и объектов в сцене.

2. Расширены основные положения дескриптивного подхода к распознаванию последовательностей изображений, позволяющие учитывать цели распознавания на начальных стадиях обработки последовательности изображений с последующей сегментацией областей интереса, строить траектории движения и распознавать поведение динамических объектов, учитывать предысторию движения объектов при пересечении их проекций, сопровождать переменное количество объектов наблюдения.

3. Разработан иерархический метод обработки и распознавания пространственно-временных структур, состоящий из трех уровней и пяти этапов и предполагающий нормализацию проекций объектов, что позволяет сократить количество эталонов для одного класса при распознавании сложных динамических объектов.

4. Разработан метод оценки движения для последовательностей изображений из видимого и инфракрасного диапазонов электромагнитного излучения, отличающийся тем, что используются пространственно-временные наборы данных в виде  $3D$  структурных тензоров и  $3D$  тензоров потока соответственно. Полученная оценка движения позволяет выбрать наиболее эффективный метод сегментации динамических визуальных объектов, отличающихся количеством допустимых проекций.

5. Построена модель многоуровневого движения регионов изображения на основе локальных векторов скорости, отличающаяся тем, что позволяет разделять сцену не только на объекты переднего плана и фон, но и на уровни движения объектов, удаленных от наблюдателя. Это особенно актуально для сложных сцен, регистрируемых подвижным видеодатчиком, когда все объекты сцены находятся в относительном движении.

6. Разработан адаптивный алгоритм сегментации динамических объектов: а) для объектов с ограниченным количеством проекций на основе анализа предыстории движения локальных динамических регионов, отли-

чающийся тем, что при перекрытиях изображений достраивается форма региона по текущему шаблону и при условии применения фильтра Калмана прогнозируется текущая траектория; б) для объектов с произвольным количеством проекций на основе комплексного анализа цветовых, текстурных, статистических, топологических признаков и признаков движения, отличающийся тем, что при перекрытиях изображений форма региона достраивается с использованием метода активных контуров.

7. Предложен способ построения динамического видеографа сложной сцены по методу иерархического группирования комплексных признаков низшего уровня в локальные пространственные структуры устойчивые во времени, и далее в локальные пространственные объекты. Сформированный видеограф устанавливает временные отношения между объектами и сохраняет все обобщенные признаки для распознавания событий в сцене. Расширена двумерная грамматика М.И. Шлезингера в рамках структурного метода распознавания до трехуровневой контекстной грамматики.

8. Для распознавания динамических объектов модифицирован коллективный метод принятия решений, вначале осуществляющий распознавание принадлежности изображения области компетентности, а затем выбирающий то решающее правило, компетентность которого максимальна в заданной области. Построены четыре вида псевдо-расстояний для нахождения меры сходства входных динамических образов с эталонами в зависимости от представления динамических признаков.

9. Разработан метод распознавания событий на основе байесовской сети, выполняющий рекурсивное разрезание матрицы весовых коэффициентов во входной видеопоследовательности и сравнение с эталонными событиями, полученными на этапе обучения. Данная информация является исходной для определения жанра сцены и индексирования видеопоследовательностей в мультимедийных Интернет-базах.

10. Практические задачи обработки и распознавания последовательностей изображений решены с помощью адаптивно-иерархического метода пространственно-временной обработки, показана работоспособность метода, продемонстрирована эффективность применения системы иерархических методов обработки и распознавания визуальной информации с возможностью адаптивного выбора признаков в процессе решения задачи. Полученные результаты в виде спроектированных экспериментальных систем переданы заинтересованным организациям.

Таким образом, в данной диссертационной работе решена важная научно-техническая проблема информационного обеспечения систем видеонаблюдения и разработано новое направление в области распознавания динамических образов на основе пространственно-временного анализа последовательностей изображений.



**Основные положения и результаты диссертационной работы представлены в следующих работах автора.** Списки статей и материалы конференций приведены в хронологическом порядке.

Статьи в рецензируемых научных изданиях и в изданиях, входящих в перечень ВАК:

1. Фаворская, М.Н. Комплексный расчет характеристик ландшафтных изображений / М.Н. Фаворская, Н.Ю.Петухов // Оптический журнал, 77, 8, 2010. с. 54–60.

2. Фаворская, М.Н. Модель улучшения цветных изображений на основе адаптивного выравнивания спектральных диапазонов / М.Н. Фаворская, А.И. Пахирка // Системы управления и информационные технологии / ИПУ РАН, ВГТУ, № 1 (39), М.-Воронеж, 2010. с. 23–26.

3. Фаворская, М.Н. Оценка движения объектов в сложных сценах на основе тензорного подхода / М.Н. Фаворская // Цифровая обработка сигналов, М.:– Рязань, № 1, 2010. с. 2–9.

4. Favorskaya M., Zotin A., Danilin I., Smolentcheva S. Realistic 3D-modeling of Forest Growth with Natural Effect // Proceedings of the Second KES International Symposium IDT 2010, Baltimore. USA. Springer-Verlag, Berlin, Heidelberg. 2010. pp. 191–199.

5. Favorskaya M., Zotin A., Damov M. Intelligent Inpainting System for Texture Reconstruction in Videos with Text Removal // Proceedings of International Congress on Ultra Modern Telecommunications and Control Systems ICUMT2010, Moscow. Russia. 2010. Article number 1569331721.

6. Фаворская, М.Н. Формирование контекстной грамматики для распознавания сложных сцен с многоуровневым движением объектов / М.Н. Фаворская, А.М. Попов // Вестник Сибирского государственного аэрокосмического университета имени академика М.Ф. Решетнева. – Вып. 1 (27), Красноярск, 2010. с. 64–68.

7. Фаворская, М.Н. Нахождение движущихся видео объектов с применением локальных 3D структурных тензоров / М.Н. Фаворская // Вестник Сибирского государственного аэрокосмического университета имени академика М.Ф. Решетнева. – Вып. 2 (23), Красноярск, 2009. с. 141–146.

8. Фаворская, М.Н. Методы поиска движения в видеопоследовательностях / М.Н. Фаворская, А.И. Пахирка, А.С. Шилов, М.В. Дамов // Вестник Сибирского государственного аэрокосмического университета имени академика М.Ф. Решетнева. – Вып. 1 (22) в 2 частях, Ч. 2, Красноярск, 2009. с. 69–74.

9. Фаворская, М.Н. Распознавание динамических образов на основе предсказывающих фильтров / М.Н. Фаворская // Вестник Сибирского государственного аэрокосмического университета имени академика М.Ф. Решетнева. – Вып. 1(22) в 2 частях, Ч. 1, Красноярск, 2009. с. 64–68.

10. Фаворская, М.Н. Алгоритмы реализации оценки движения в системах видеонаблюдения / М.Н. Фаворская, А.С. Шилов // Системы управления и информационные технологии / ИПУ РАН, ВГТУ, № 3.3(33), М.-Воронеж, 2008. с. 408–412.

11. Фаворская, М.Н. Модели локализации лиц на изображениях / М.Н. Фаворская, А.И. Пахирка // Системы управления и информационные технологии / ИПУ РАН, ВГТУ, № 3.3 (33), М.-Воронеж, 2008. с. 404–408.

12. Фаворская, М.Н. Модель распознавания изображений рукописного текста / М.Н. Фаворская, А.Н. Горошкин // Вестник Сибирского государственного аэрокосмического университета имени академика М.Ф. Решетнева. – Вып. 2 (19), Красноярск, 2008. с. 52–58.

13. Фаворская, М.Н. Вероятностные методы сегментации видеопотока как задача с недостающими данными / М.Н. Фаворская // Вестник Сибирского государственного аэрокосмического университета имени академика М.Ф. Решетнева. – Вып. 3(16), Красноярск, 2007. с. 4–8.

14. Фаворская, М.Н. Морфологическая обработка контурных изображений в системах распознавания текстовых символов / М.Н. Фаворская, А. Г. Зотин, А. Н. Горошкин // Вестник Сибирского государственного аэрокосмического университета имени академика М.Ф. Решетнева. – Вып. 1 (14), Красноярск, 2007. с. 70–75.

15. Фаворская, М.Н. Инвариантные решающие функции в задачах распознавания статических изображений / М.Н. Фаворская // Вестник Сибирского государственного аэрокосмического университета имени академика М.Ф. Решетнева. – Вып. 1 (14), Красноярск, 2007. с. 65–70.

16. Фаворская, М.Н. Прогнозирование в системах распознавания образов на основе скрытых марковских моделей / М.Н. Фаворская, Н.Д. Торгашин, А.Г. Зотин // Вестник Сибирского государственного аэрокосмического университета имени академика М.Ф. Решетнева. – Вып. 1(8), Красноярск, 2006. с. 59–63.

Материалы конференций, статьи в сборниках:

17. Favorskaya M. N., Petukhov N. Y., Danilin I. M., Danilin A. I. Recognition of forest textures on airphotos // Proceedings of the IASTED International Conferences on Automation, Control, and Information Technology (ACIT 2010). Vol.1, Novosibirsk, Russia. ACTA Press, Anaheim, Calgary, Zurich. 2010. pp. 9–14.

18. Фаворская, М.Н. Построение видео графа сложной динамической сцены / М.Н. Фаворская // В материалах 12-й международной конференции и выставки «Цифровая обработка сигналов и ее применение», М. В 2 т. Т. 2, 2010. с. 185–189.

19. Фаворская, М.Н. Локальные пространственно-временные признаки событий в видеопоследовательностях / М.Н. Фаворская // В материалах X междунар. науч.-техн. конф. «Теоретические и прикладные вопросы со-

временных информационных технологий», ч. II, Улан-Удэ, 2009. с. 461–466.

20. Фаворская, М.Н. Пространственно-временная сегментация видеопоследовательностей на основе 3D структурного тензора / М.Н. Фаворская // В материалах 11-й междунар/ конф. и выставки «Цифровая обработка сигналов и ее применение», Москва, 2009. с. 385–388.

21. Фаворская, М.Н. Методы и модели поиска целевых информативных признаков в видеопоследовательностях / М.Н. Фаворская // В материалах IX международной научно-технической конференции «Кибернетика и высокие технологии XXI века», в 2 т., Т.1, Воронеж, 2008. с. 171–182.

22. Фаворская, М.Н. Стратегии сегментации двумерных изображений / М.Н. Фаворская // В материалах всероссийской научной конференции «Модели и методы обработки изображений ММОИ-2007», Красноярск, 2007. с. 136–140.

23. Фаворская, М.Н. Определение аффинной структуры объекта по движению / М.Н. Фаворская // Вестник Сибирского государственного аэрокосмического университета имени академика М.Ф. Решетнева. – Вып. 6, Красноярск, 2005. с. 86-89.

24. Фаворская, М.Н. Идентификация ограниченного набора образов на основе ассоциативных нейронных сетей / М.Н. Фаворская, А.Г. Зотин, М.В. Дамов, Н.Д. Торгашин // Вестник Сибирского государственного аэрокосмического университета имени академика М.Ф. Решетнева. – Вып. 5, Красноярск, 2004. с. 109-116.

25. Фаворская, М.Н. Цифровой анализ искаженных изображений по плоским проекциям / М.Н. Фаворская // В кн. «ЭВМ в проектировании и производстве», вып. 2 / Под общ. ред. Г.В.Орловского, Л., Машиностроение, 1985, с. 149-153.

26. Луцив, В.Р. Принципы унификации оптических систем роботов / В.Р. Луцив, М.Н. Фаворская // В кн. «Унификация и стандартизация промышленных роботов», Ташкент, 1984, с. 93-94.

Свидетельства о регистрации программных продуктов:

27. Пахирка, А.И. Нелинейное улучшение изображений (Nonlinear image enhancement) / А.И. Пахирка, М.Н. Фаворская // Свидетельство № 2010610658. Зарегистрировано в Реестре программ для ЭВМ г. Москва, 31 марта 2010 г.

28. Зотин, А.Г. Программа визуальной регистрации государственных номерных знаков автотранспортных средств при многопоточном движении (FNX STRAnalyzer). Версия 1.27 / А.Г.Зотин, М.Н. Фаворская, Е.А. Вейсов, С.В. Шульц // Свидетельство № 2010610570. Зарегистрировано в Реестре программ для ЭВМ г. Москва, 25 марта 2010 г.

29. Левтин, К.Э. Визуальное детектирование дыма (SmokeDetection) / К.Э. Левтин, М.Н. Фаворская // Свидетельство № 2009612795. Зарегистрировано в Реестре программ для ЭВМ г. Москва, 30 июля 2009 г.

30. Смоленцева, С.С. Система наложения визуальных природных эффектов на статическое изображение (Natural effects imitation) / С.С. Смоленцева, М.Н. Фаворская // Свидетельство № 2009612794. Зарегистрировано в Реестре программ для ЭВМ г. Москва, 30 июля 2009 г.

31. Шилов, А.С. Определение движения (MotionEstimation) / А.С. Шилов, М.Н. Фаворская // Свидетельство № 2009611014. Зарегистрировано в Реестре программ для ЭВМ г. Москва, 16 февраля 2009 г.

32. Пахирка, А.И. Локализация лица (FaceDetection) / А.И. Пахирка, М.Н. Фаворская // Свидетельство № 2009611010. Зарегистрировано в Реестре программ для ЭВМ г. Москва, 16 февраля 2009 г.

33. Горошкин, А.Н. Сегментация изображений рукописного текста (SegPic) / А.Н. Горошкин, М.Н. Фаворская // Свидетельство № 2008614243. Зарегистрировано в Реестре программ для ЭВМ г. Москва, 5 сентября 2008 г.

Монография:

34. Фаворская, М.Н. Методы распознавания изображений и видеопоследовательностей: монография / М. Н. Фаворская / Сиб. гос. аэрокосмич. ун-т. Красноярск, 2010. – 176 с.

Фаворская Маргарита Николаевна

Модели и методы распознавания динамических образов на основе пространственно-временного анализа последовательностей изображений

Автореферат

Подписано к печати 28.12.2010

Формат 60x84/16. Бумага писчая. Печ. л. 1.5

Тираж 100 экз. Заказ № \_\_\_\_\_

Отпечатано в отделе копировальной и множительной техники СибГАУ  
660014 г. Красноярск, пр. им. газеты «Красноярский рабочий», 31