

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И ПРИНЯТИЕ РЕШЕНИЙ
ARTIFICIAL INTELLIGENCE AND DECISION MAKING

УДК 004.82

DOI: 10.18413/2518-1092-2025-10-2-0-7

Басов О.О.¹
Офицеров А.И.²

АВТОМАТИЧЕСКИЙ РАЗБОР ДОРОЖНО-ТРАНСПОРТНЫХ
ПРОИСШЕСТВИЙ НА ОСНОВЕ МУЛЬТИМОДАЛЬНОГО
АНАЛИЗА ДАННЫХ С ИСПОЛЬЗОВАНИЕМ ТЕХНОЛОГИЙ
ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

¹) Акционерное общество «АСТ», пер. Капанова, 3, стр. 2, г. Москва, 123242, Россия

²) Федеральное государственное казённое военное образовательное учреждение высшего образования «Академия Федеральной службы охраны Российской Федерации», ул. Приборостроительная, 35, г. Орёл, 302034, Россия

e-mail: o.basov@acti.ru

Аннотация

В работе предложено задачу разбора дорожно-транспортных происшествий автоматизировать и интеллектуализировать. Для этого важные для разбора данные, представленные разными модальностями, автоматически поступают в интеллектуальную систему, где происходит их интерпретация, синхронизация по времени и пространству и объединение для принятия итогового решения. Реализация каждой из процедур автоматического разбора дорожно-транспортного происшествия требует однозначного определения вида, объема и семантики информации, поступающей от объектов наблюдения. Полученные в работе результаты свидетельствуют о перспективности выбранного подхода к решению задачи автоматического разбора дорожно-транспортных происшествий на основе мультимодального анализа данных с использованием технологий искусственного интеллекта.

Ключевые слова: дорожно-транспортное происшествие; интерпретация; мультимодальность; автоматический разбор; интеллектуальная поддержка принятия решения

Для цитирования: Басов О.О., Офицеров А.И. Автоматический разбор дорожно-транспортных происшествий на основе мультимодального анализа данных с использованием технологий искусственного интеллекта // Научный результат. Информационные технологии. – Т. 10, №2, 2025. – С. 69-83. DOI: 10.18413/2518-1092-2025-10-2-0-7

Basov O.O.¹
Ofitserov A.I.²

AUTOMATIC PARSING TRAFFIC ACCIDENTS BASED
ON MULTIMODAL DATA ANALYSIS USING ARTIFICIAL
INTELLIGENCE TECHNOLOGIES

¹) Joint Stock Company «AST», trans. Kapranova, 3, building 2, Moscow, 123242, Russia

²) Federal state military educational institution of higher professional education "Academy of the Federal security service of the Russian Federation", 35 Priborostroitel'naya St., Orel, 302034, Russia

e-mail: o.basov@acti.ru

Abstract

The paper proposes to automate and intellectualize the task of analyzing traffic accidents. To do this, the data presented in different modalities, which is important for analysis, is automatically fed into an intelligent system, where it is interpreted, synchronized in time and space, and combined to make a final decision. The implementation of each of the procedures for the automatic analysis of

a traffic accident requires an unambiguous definition of the type, volume and semantics of the information coming from the objects of supervision. The results received in work testify to prospects of the chosen approach to the solution of a problem of automatic analysis of road accidents on the basis of the multimodal analysis of data with use of technologies of artificial intelligence.

Keywords: traffic accident; interpretation; multimodality; automatic parsing; intellectual support of decision-making

For citation: Basov O.O., Ofitserov A.I. Automatic parsing traffic accidents based on multimodal data analysis using artificial intelligence technologies // Research result. Information technologies. – Т. 10, №2, 2025. – P. 69-83. DOI: 10.18413/2518-1092-2025-10-2-0-7

ВВЕДЕНИЕ

В современной практике наблюдается значительное расширение спектра задач, в которых для лиц, принимающих решения, критически важен детальный анализ структуры и динамики объектов сцены. Это охватывает широкий спектр приложений, включая анализ ситуаций в закрытых пространствах, мониторинг в местах массового скопления людей, управление движением робототехнических систем и разбор дорожно-транспортных происшествий (ДТП). Ключевыми аспектами этих задач являются распознавание как статических, так и динамических объектов сцены, а также выявление активных действий и событий, что особенно важно для обнаружения несанкционированных действий в общественных местах, разбора ДТП, а в перспективе – определения событий, предшествующих реальной сцене, и прогнозирование последующих событий. Однако на современном этапе развития техники и технологий отсутствует эффективный научно-методический аппарат решения данного класса задач из-за сложности и многообразия поведения объектов реального мира.

Разработка интеллектуальных методов и систем интерпретации сцен и событий требует создания эффективных подходов к формализации различных модальностей данных, таких как изображения, видеопоследовательности, аудиосигналы и текстовые данные. Целью данной работы является извлечение семантического описания сцены, которое адекватно соответствует поставленным перед системой целевым управляющим воздействиям. При этом наличие других модальностей, помимо визуального канала, является важнейшим информативным условием полной и достоверной интерпретации.

Возрастание спроса на автомобили и, соответственно, значительный рост их числа на дорогах обусловило увеличение количества ДТП, в том числе и достаточно сложных для разбора. В то же время число сотрудников групп разбора (в первую очередь высококвалифицированных) не увеличилось (а в некоторых субъектах даже уменьшилось). Все это способствует значительному повышению нагрузки на сотрудников, что в конечном итоге может привести к снижению эффективности их работы. В связи с этим остро встала проблема автоматизации отдельных процедур (а в идеале всего цикла) разбора ДТП на основе интеллектуальных методов и систем, связанных с интерпретацией сцен и событий.

ПОСТАНОВКА ПРОБЛЕМЫ АВТОМАТИЧЕСКОГО РАЗБОРА ДОРОЖНО-ТРАНСПОРТНОГО ПРОИСШЕСТВИЯ

В области анализа ДТП в настоящее время отсутствует систематическая методологическая база и стандартизированные подходы к их разбору. Основным методом, применяемым на практике, остается автотехническая экспертиза. Однако ограниченные возможности существующих методов и недостаточное их применение в расследовании ДТП часто приводят к ошибочным выводам, что подчеркивает необходимость разработки более совершенных аналитических инструментов. Полная реконструкция механизма ДТП является критически важным условием для обеспечения обоснованности и объективности выводов по всем связанным с происшествием вопросам.

Традиционный подход к проведению автотехнической экспертизы основывается на фундаментальных принципах кинематики и трасологии. Методологические основы данного подхода

включают использование кинематических расчетов, направленных на определение параметров движения идеализированной материальной точки в пространстве и времени [1].

В рамках данного подхода анализируются такие ключевые параметры движения, как путь, перемещение, скорость и следы, образовавшиеся на поверхности в результате динамического взаимодействия. Однако данный метод характеризуется высокой степенью абстракции и требует применения значительного числа допущений и упрощений. Для минимизации этих условностей и обеспечения более точной оценки обстоятельств ДТП используются поправочные коэффициенты и методы усреднения значений параметров движения. Эти коэффициенты, как правило, базируются на эмпирических данных и включают такие параметры, как время срабатывания тормозного привода, коэффициенты сцепления, сопротивления перемещению, эффективности торможения и другие [2].

Важно отметить, что некоторые аспекты, такие как изменение кинетической энергии транспортных средств вследствие деформации при столкновении, часто не учитываются в рамках данного подхода. Интерпретация ДТП осуществляется методом мысленного реконструирования события, при котором эксперт, основываясь на исходной информации и логических построениях, воссоздает структуру и динамику происшествия для оценки его непротиворечивости и выявления ключевых факторов.

В профессиональной среде наблюдается значительная вариативность в оценке ДТП среди экспертов, обладающих различным уровнем квалификации. Достоверность интерпретации таких инцидентов напрямую коррелирует с глубиной научных знаний и практическим опытом специалиста. В рамках типовых сценариев, включающих определение тормозного и остановочного пути, времени торможения и дистанции между транспортными средствами на момент столкновения, высокий уровень достоверности достигается за счет применения стандартных расчетных методик. Однако с усложнением ситуации возрастает роль субъективных факторов, влияющих на интерпретацию обстоятельств ДТП. Это требует разработки новых методологических подходов, основанных на интеллектуализации процессов принятия решений и мультимодальном анализе данных о происшествиях.

Для того, чтобы результаты разбора ДТП обладали максимальной достоверностью и полнотой, не зависели от сложности происшествия и квалификации эксперта, процедуры разбора целесообразно автоматизировать и интеллектуализировать, а входная информация, на основе которой будут получены эти результаты, должна содержать по возможности данные различных модальностей. В качестве таковых может выступать визуальная, акустическая и текстовая информация, а также навигационные данные о геолокации объектов интереса в конкретные моменты времени. Информация визуальной модальности может поступать с видеорегистраторов самого автомобиля и автомобилей – свидетелей ситуации, со стационарных камер системы «Безопасный город», видеозаписывающих устройств пешеходов, а также видеокамер группы разбора ДТП (как ручных, так и установленных на беспилотных воздушных судах (БВС)). Информацию акустической модальности, полезную для последующей интерпретации ДТП, могут предоставлять телефонные звонки участников ДТП, а также свидетелей происшествия, важной для последующего анализа причинно-следственных факторов является аудиоинформация с принятых видеопоследовательностей (визг тормозов, звук удара и т. п.). В качестве текстовой информации, которая после семантического разбора будет представлять определенную ценность для интерпретации ДТП, могут выступать объяснительные участников и свидетелей ДТП с пояснением ситуации, а также текстовая информация с принятых видеопоследовательностей (государственные регистрационные номера автомобилей, названия улиц, номера домов, текст баннеров и т. п.). Источниками навигационной информации могут являться спутниковые навигационные системы автомобилей (например, «ЭРА-ГЛОНАСС»), а также геоданные при включенной функции геолокации в сотовых телефонах участников ситуации. Причем эти данные не должны ограничиваться сугубо местом совершения ДТП. Для получения полной картины происшествий, выявления факторов и причин, способствовавших их совершению, зачастую бывает полезно «отмотать назад» течение событий и проанализировать, что им предшествовало, начиная с начала движения автомобиля – участника ДТП, заканчивая его эвакуацией с места происшествия.

Возможна и постановка обратной задачи, связанной с прогнозированием ДТП. В данном случае при фиксации, например, факта совершения участником дорожного движения опасного (нестандартного) действия, возникновения конфликтной ситуации на дороге и т. п. система должна запустить алгоритм «сопровождения опасного водителя», передавая информацию о нем элементам системы. Так как не все инициаторы опасных ситуаций на дороге в последующем становятся участниками ДТП, то для предотвращения перегрузки системы информацией алгоритм сопровождения (и, соответственно, сбор данных) целесообразно остановить через три-пять участков сопровождения (например, зон действий камер системы «Безопасный город») или через определенное установленное время. В прямой постановке задачу прогнозирования в настоящее время решить достаточно проблематично, ввиду отсутствия правовых и организационных механизмов ее реализации. Однако определенные процедуры обратной задачи возможно использовать в задаче разбора ДТП, например, для выявления факторов и условий, способствовавших или спровоцировавших его совершение.

ИНТЕЛЛЕКТУАЛЬНАЯ ПОДДЕРЖКА ПРИНЯТИЯ РЕШЕНИЯ НА ОСНОВЕ МУЛЬТИМОДАЛЬНОГО АНАЛИЗА ДАННЫХ О ДОРОЖНО-ТРАНСПОРТНОМ ПРОИСШЕСТВИИ

Для реализации возможности автоматического разбора ДТП на основе мультимодального анализа данных с использованием технологий искусственного интеллекта интеллектуальные системы мониторинга дорожной ситуации должны строиться по принципу объединения в единую сеть различных приемников модальностей (ПрМ), регистрирующих обстановку в зоне наблюдения. Такими ПрМ могут являться видеоприборы: стационарные камеры и камеры с возможностью управления положением оптической оси и величиной оптической силы объектива (например, стационарные камеры видеofиксации, камеры видеонаблюдения организаций и учреждений), а также камеры, размещенные на подвижных объектах (например, видеорегистраторы автомобилей, смартфоны пешеходов, камеры БВС), изменяющие свое положение как в пространстве, так и относительно друг друга. Некоторые из них, такие как стационарные камеры, установленные на автодорогах и зданиях городской инфраструктуры, уже объединены в рамках системы «Безопасный город» и будут представлять основные источники мультимодальной информации для последующего анализа, интеграция других в единую автоматизированную систему интеллектуального мониторинга требует реализации дополнительных механизмов. Кроме того, ПрМ могут являться приемники аудиоинформации (микрофоны видеокамер, накопители «Системы обеспечения вызова экстренных оперативных служб по единому номеру 112»), также для привязки к конкретным координатам необходимы приемники навигационных сигналов (ГЛОНАСС / GPS).

Создание интеллектуальных систем, базирующихся на мультимодальном анализе данных, представляет собой сложную задачу, требующую разработки надежных методов распознавания сигналов из различных каналов передачи информации, а также эффективных алгоритмов интеграции данных различных модальностей. Интеграция данных может осуществляться на разных уровнях их представления и с применением разнообразных фундаментальных стратегий, включая методы машинного обучения, статистический анализ и когнитивные модели [3]:

1) на уровне признакового описания («раннее объединение») информативные признаки f_1, \dots, f_N извлекаются из сигналов, поступающих от различных модальностей. Эти признаки интегрируются в единый вектор $F_{1,N}$, который затем поступает в блок обработки для формирования итогового решения D ;

2) на семантическом уровне принятия решений («позднее объединение») отдельные блоки обработки формируют локальные решения d_1, \dots, d_N на основе соответствующих признаков f_1, \dots, f_N . Эти локальные решения объединяются в вектор $D_{1,N}$, который служит основой для принятия итогового решения D относительно решаемой задачи или выдвинутой гипотезы распознавания;

3) гибридный подход объединяет преимущества обоих описанных методов и позволяет эффективно решать задачи распознавания мультимодальной информации.

Для автоматического разбора ДТП на основе мультимодального анализа данных целесообразным является применение стратегии гибридного объединения. При этом решение задачи поиска объектов интереса на изображениях и видеопоследовательностях (например, для определения номеров автомобилей, попавших во временной интервал совершения ДТП в зону действия камеры наблюдения, для последующего запроса информации с их видеорегистраторов) предусматривает применение первого подхода – стратегии «раннего объединения». Второй подход, известный как «позднее объединение» (*late fusion*), демонстрирует свою эффективность при интеграции методов анализа сигналов, полученных из различных модальностей и временных интервалов, а также из удаленных пространственных точек. В качестве примера можно привести использование скрытых марковских моделей (СММ) для анализа аудиосигналов и метода опорных векторов (*SVM*) для обработки изображений. Этот метод обеспечивает более высокую гибкость и адаптивность по сравнению с ранним объединением (*early fusion*), позволяя более эффективно интегрировать разнородные данные на поздних этапах обработки.

Таким образом, интеллектуальная поддержка принятия решения на основе мультимодального анализа данных о ДТП может быть сведена к последовательности процедур, реализуемых системой в автоматическом режиме и приводящих в итоге к оптимальному результату разбора:

- 1) выявление факта совершения ДТП;
- 2) сбор и обработка мультимодальной информации от всех возможных источников с места ДТП;
- 3) по результатам обработки – определение источников дополнительной мультимодальной информации, запрос и получение информации, обработка информации;
- 4) синхронизация и объединение входных модальностей от всех источников;
- 5) интерпретация ДТП и принятие решения по результатам разбора.

Для реализации каждой из процедур, ведущих к достижению поставленной цели – принятию решения по результатам разбора ДТП – необходимо однозначно определить вид, объем и семантику информации, поступающей от объектов наблюдения (рис. 1).

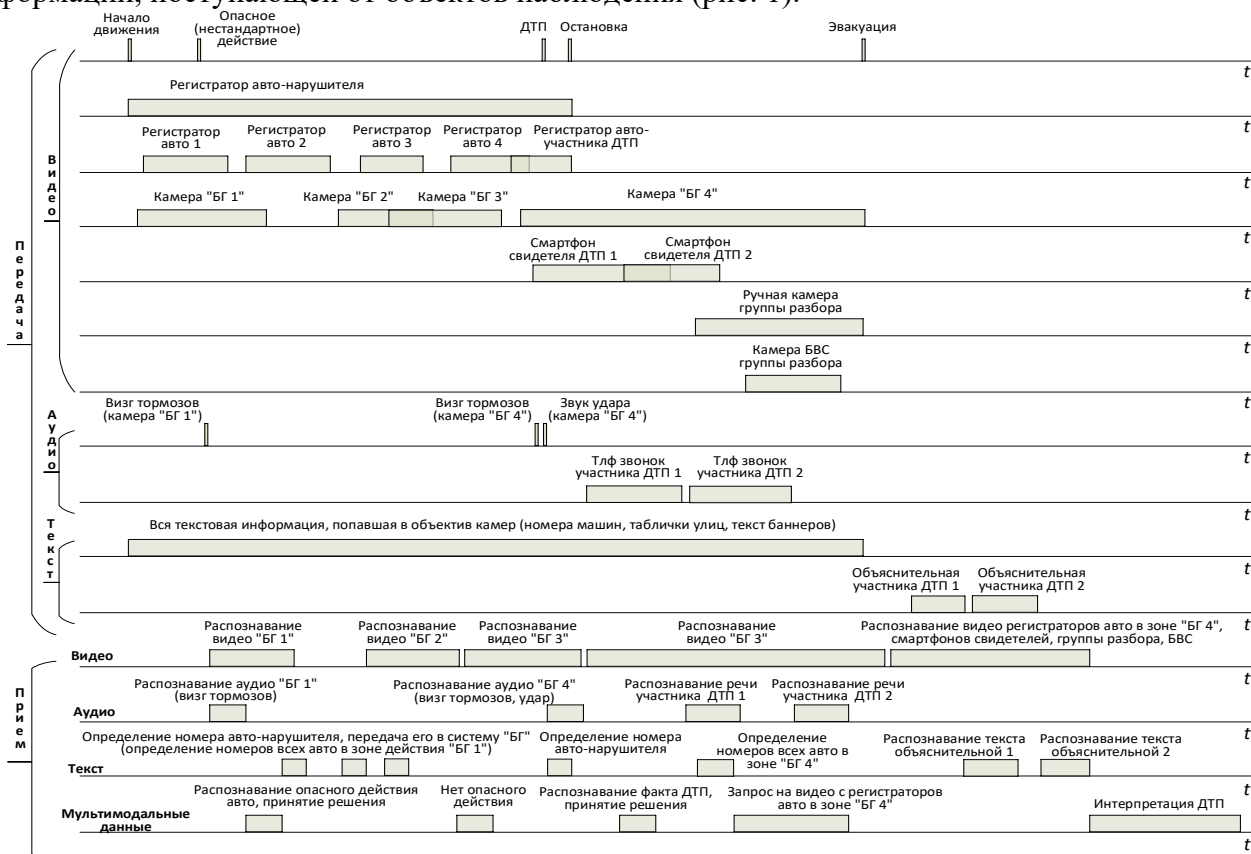


Рис. 1. Временные диаграммы процедур автоматического разбора ДТП
Fig. 1. Time diagrams of automatic accident analysis procedures

1. Выявление факта совершения ДТП

Для получения качественных результатов разбора ДТП выявление факта его совершения при интеллектуальной поддержке принятия решения должно осуществляться в режиме времени, близком к реальному [4], и в качестве входных данных для автоматического распознавания события ДТП могут выступать визуальная информация с видеокамер (например, стоящие вместе поперек дороги автомобили, упавший мотоцикл, лежащий на дороге человек и др.), текстуальная информация – заявления участников ДТП, а также аудиоинформация в режиме *on-line* (например, визг тормозов, звук удара при столкновении и др.) и в режиме *off-line* (например, с телефонных звонков участников или свидетелей ДТП).

Выявление факта совершения ДТП на основе информации визуальной модальности основано на достаточно изученных и широко распространенных методах распознавания статических и динамических объектов, составляющих основу большинства типичных задач видеонаблюдения. Некоторые методики решения задач данного типа основываются на применении СММ или искусственных нейронных сетей (ИНС), которые используют двухэтапный подход к обучению. На первом этапе СММ проходит обучение на большом наборе стандартных событий, что позволяет ей формировать обобщенные вероятностные модели этих событий. На втором этапе для обработки нестандартных событий применяется стохастическая ИНС, которая способна адаптироваться к новым условиям и распознавать редкие или неизвестные паттерны. Однако данный подход сталкивается с определенными проблемами достоверности результатов, связанными с тем, что в зоне действия первичных датчиков может присутствовать неизвестное количество объектов, подлежащих распознаванию. Эти объекты могут находиться в различных пространственных положениях и частично перекрываться, что затрудняет их точное определение и классификацию.

Задача идентификации факта ДТП на основе его семантических характеристик представляет собой сложную проблему, требующую применения современных методов анализа данных и искусственного интеллекта. В данном контексте поиск осуществляется на основании текстового или акустического описания, извлеченного из объяснительных записок, телефонных переговоров участников ДТП или других источников информации. Эти описания должны удовлетворять критериям, подтверждающим факт совершения ДТП, и обладать достаточной семантической насыщенностью для корректной интерпретации. Примером семантического описания может служить формулировка: «поврежденный автомобиль (или упавший мотоцикл), находящийся на пересечении соответствующих улиц». Для реализации семантического поиска необходимо выполнить операцию составления семантического индекса, включающую выделение семантических признаков из изображений и текстов, а также последующее построение сбалансированного дерева поиска, основанного на этих характеристиках. Важно отметить, что объем информации, полученной в результате выделения семантических признаков, значительно меньше объема исходных данных, что позволяет оптимизировать вычислительные ресурсы и повысить эффективность поиска.

2. Сбор и обработка мультимодальной информации от всех возможных источников с места ДТП

После выявления факта совершения ДТП система дает команду на сбор мультимодальной информации от всех ПрМ, находившихся в момент времени совершения ДТП в зоне интереса, ограниченной площадью, необходимой и достаточной для полного и достоверного разбора ДТП (например, перекресток / перекресток с прилегающими участками дорог / перекресток с прилегающими участками дорог и пешеходными дорожками и т. д.). При этом визуальная и аудиоинформация, полученная со стационарных камер системы «Безопасный город» в режиме времени, близком к реальному, является основной для предварительной интерпретации сцены ДТП и последующего определения источников дополнительной мультимодальной информации, необходимой для полной интерпретации ДТП.

Помимо стационарных камер системы «Безопасный город» источниками визуальной и акустической модальностей с места ДТП могут являться:

– видеорегистраторы автомобилей – участников ДТП (информация появляется в системе после предоставления ее автовладельцами сотрудникам группы разбора);

– видеорегистраторы автомобилей – свидетелей ДТП (информация может появиться в системе в случае, если после предварительной интерпретации сцены ДТП с помощью камер системы «Безопасный город» будут определены номера автомобилей, находящихся в зоне ДТП, автовладельцам будет послан запрос, после чего они предоставят информацию с видеорегистратора сотрудникам группы разбора);

– смартфоны пешеходов – свидетелей ДТП (информация может появиться в системе в случае, если после прибытия на место ДТП группой разбора обнаружены свидетели, находившиеся в зоне ДТП и заснявшие событие, которые предоставят информацию со смартфона сотрудникам группы разбора);

– камеры группы разбора – ручные и установленные на БВС (съемка производится в полном соответствии с требованиями автотехнической экспертизы).

Для динамических объектов в сцене ДТП применяется более комплексный подход, включающий анализ дополнительных параметров, таких как предыстория движения объекта, временные события и процедуры ассоциации. Предыстория движения позволяет классифицировать движения и интерпретировать их на понятийном уровне, что способствует более точной реконструкции событий. Временные события, накопленные в виде предыстории, играют ключевую роль в синтаксическом анализе сцены, позволяя идентифицировать последовательность действий и их взаимосвязь. (табл. 1).

Таблица 1

Классы движения

Table 1

Classes of movement

| Название | Краткое описание | Области применения |
|----------------------|--|--|
| Статические объекты | Неподвижные во времени и пространстве ситуации | Распознавание однотипных объектов сцены |
| Динамические объекты | Периодически изменяемые во времени ситуации | Сопровождение движущихся объектов сцены, определение их роли и места |
| События | Взаимодействие статических и динамических объектов | Анализ взаимодействий объектов сцены, отслеживание происшествий и нестандартных ситуаций различного рода |
| Сцены | Многоуровневые движения | Интерпретация активных действий объектов и событий сцены, определение количественных или качественных показателей наблюдаемого объекта или группы объектов |

Траектория движения объекта, представленная в виде функциональной зависимости координат от времени, может быть подвергнута упрощенной аппроксимации. Основная цель данного подхода заключается в определении общего характера движения, а не в детализированном анализе его специфических параметров [5].

В рамках анализа дорожной сцены интерпретация информации осуществляется как процесс перемещения объектов на концептуальном уровне. Критически значимым аспектом данного анализа является исследование временных корреляций между объектами, включая такие феномены, как столкновение транспортных средств, падение мотоцикла на дорожное покрытие, проявления агрессивного или неадекватного поведения водителя или пешехода, а также генерацию звукового сигнала торможения и другие релевантные события.

Для проведения комплексного анализа ДТП рекомендуется применять процедуру ассоциации, которая характеризуется двумя фундаментальными параметрами [6]:

– мера сходства, учитывающая близость векторов признаков объектов, относительных преобразований и степень значимости объектов;

– подмножество объектов, подверженных глобальному движению.

Для формирования ассоциаций каждому элементу базы знаний, помимо его непосредственного описания, присваивается набор дополнительных характеристик. Значения этих характеристик рассчитываются на основе общей динамики сцены и включают следующие параметры:

– расстояние до объекта интереса, определяющее принадлежность к траектории движения;

– степень близости отношений с соседними объектами, связанными с рассматриваемым элементом;

– величина ассоциации, характеризующая значимость элемента, которая используется для определения приоритетности внимания к его траектории.

Формирование понятий основывается на пропорциональном увеличении значимости объектов, ассоциированных с определенными семантическими следами. Понятия конструируются как часто встречающиеся подструктуры, включающие наиболее релевантные объекты. На рисунке 2 представлен пример семантического анализа исходного изображения, который может быть использован для построения графа семантического разбора.

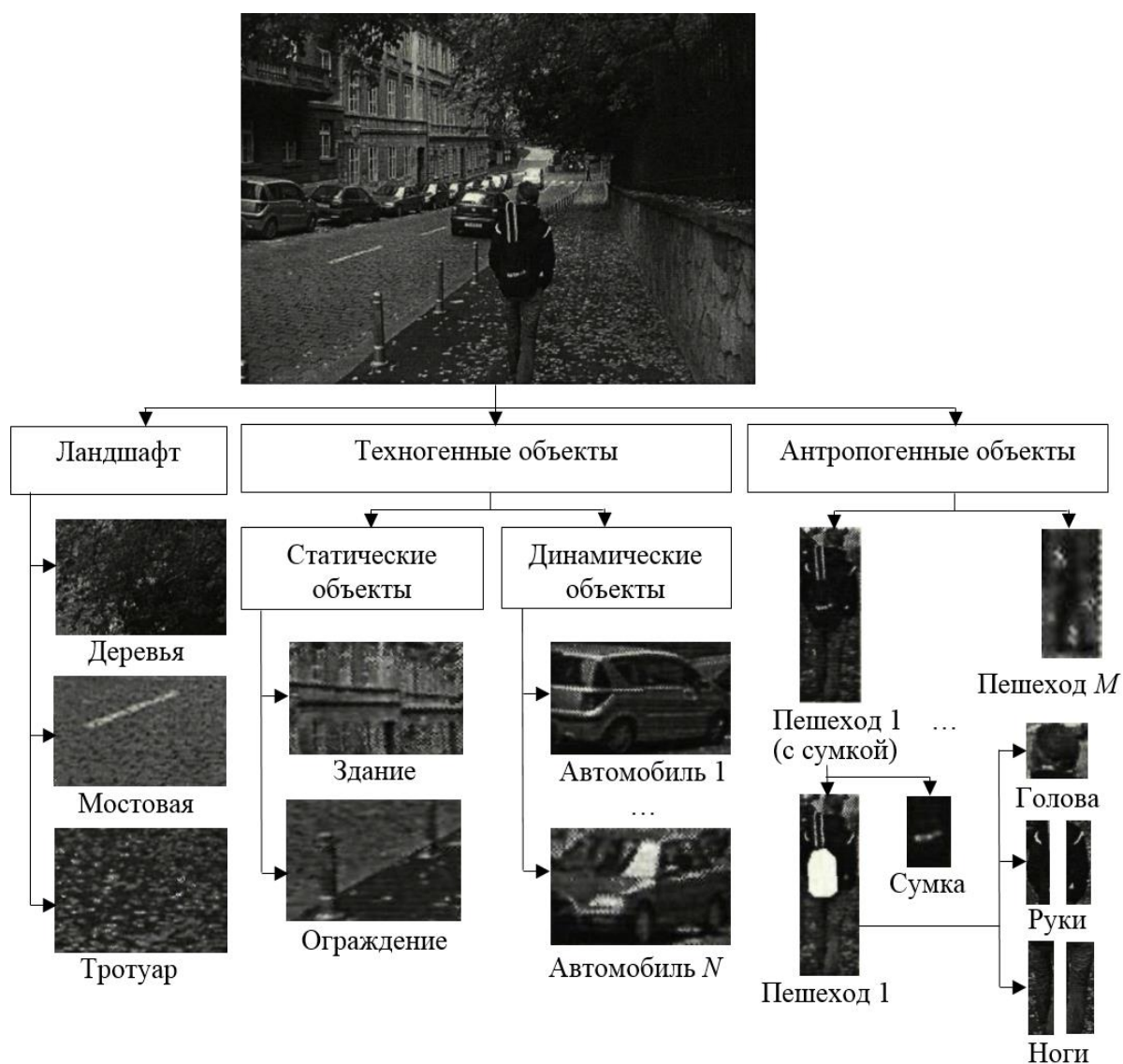


Рис. 2. Пример семантического разбора сложной сцены
Fig. 2. An example of semantic analysis of a complex scene

Создание семантических описаний участников ДТП представляет собой сложную задачу, обусловленную необходимостью интеграции низкоуровневых и среднеуровневых признаков. Эти

признаки не всегда оптимизированы для последующего семантического моделирования всей сцены ДТП. Существующие подходы опираются на упрощенные структуры сцены, где элементы могут быть представлены простейшими трехмерными геометрическими примитивами [6]. В данном контексте признаки представляют собой базовые характеристики. Определенную сложность представляет процесс идентификации признаков движения.

Сегментация сцены представляет собой высокоэффективный процесс, результатом которого является создание видеографа, где вершины содержат мультимодальный контент, включающий как видео-, так и аудиоданные, а связи классифицируются на основе пространственных и временных отношений между объектами. В связи с этим, семантические модели, предназначенные для представления видео- и аудиопоследовательностей, должны обладать способностью поддерживать два ключевых типа событий: статические и динамические, которые характеризуют эволюцию сцены.

3. Определение источников дополнительной мультимодальной информации, запрос и получение информации, обработка информации

Дополнительная мультимодальная информация в процессе автоматического разбора ДТП может понадобиться для получения более достоверных результатов разбора, учитывающих все возможные факторы и причины его совершения (предыстория ДТП), а также уточнения особенностей, не выявленных на основе мультимодальной информации с места ДТП. В качестве источников дополнительной мультимодальной информации, получаемой от ПрМ в процессе автоматического разбора ДТП, могут являться телефонные звонки участников и/или свидетелей ДТП на номер 112 и их последующие объяснительные, написанные сотрудникам группы разбора [7]. Для эффективного использования данных акустической и текстуальной модальностей необходимо их распознавание с высокой степенью достоверности, проведение семантического разбора для выделения важных составляющих, необходимых для интерпретации ДТП, синхронизация их по времени и пространству с другими модальностями для последующего объединения при принятии решения.

Для выявления дополнительной информации о факторах и причинах, способствовавших (или спровоцировавших) совершение ДТП, необходима визуальная и акустическая информация с ПрМ, не входящих в зону ДТП и находящихся по пути следования автомобилей – участников ДТП от момента времени начала их движения (по возможности) до момента начала их нахождения в зоне ДТП. Для этого данные о государственных регистрационных номерах передаются всем элементам системы (в первую очередь «Безопасный город») для распознавания указанных автомобилей и передачи видеопоследовательностей с их участием в систему автоматического разбора. Для адекватного определения причинно-следственных связей при этом необходима четкая синхронизация событий по времени, а для повышения достоверности принимаемых решений целесообразно подтверждение факта нахождения конкретного автомобиля в данный момент времени в данной точке пространства по его геопозиции, установленной с помощью приемников навигационных сигналов.

Оперативность передачи мультимодальной информации можно значительно повысить, если использовать процедуры «прогнозирования ДТП», описанные в начале статьи. В данном случае видеопоследовательности с участником ДТП, а также государственные регистрационные номера автомобилей – свидетелей опасного (нестандартного) действия – уже находятся в базе системы и могут быть оперативно представлены для разбора. Требуется лишь осуществить запрос визуальной и акустической информации с видеорегистраторов свидетелей и при ее получении группой разбора распознать и использовать видеопоследовательности для разбора ДТП.

Интерпретация понятия «событие» в контексте анализа ДТП существенно варьируется в зависимости от уровня обработки видеопоследовательности. На уровне предварительной обработки изображений осуществляется общее улучшение их характеристик (при этом за событие принимается присутствие смещений в регионах изображения). На втором уровне непосредственно вводится понятие «событие движения», основанное на данных предварительного уровня (например, фиксация телодвижений пешехода, проезд автомобилем перекрестка и другие динамические изменения в сцене). Третий уровень включает непосредственно интерпретацию сцены, при этом

события движения, накопленные в предыстории, позволяют классифицировать их по типам движений и интерпретировать на понятийном уровне. Это обеспечивает более глубокое понимание динамики сцены и ее семантического содержания, что является критически важным для анализа сложных событий, таких как ДТП (рис. 3).

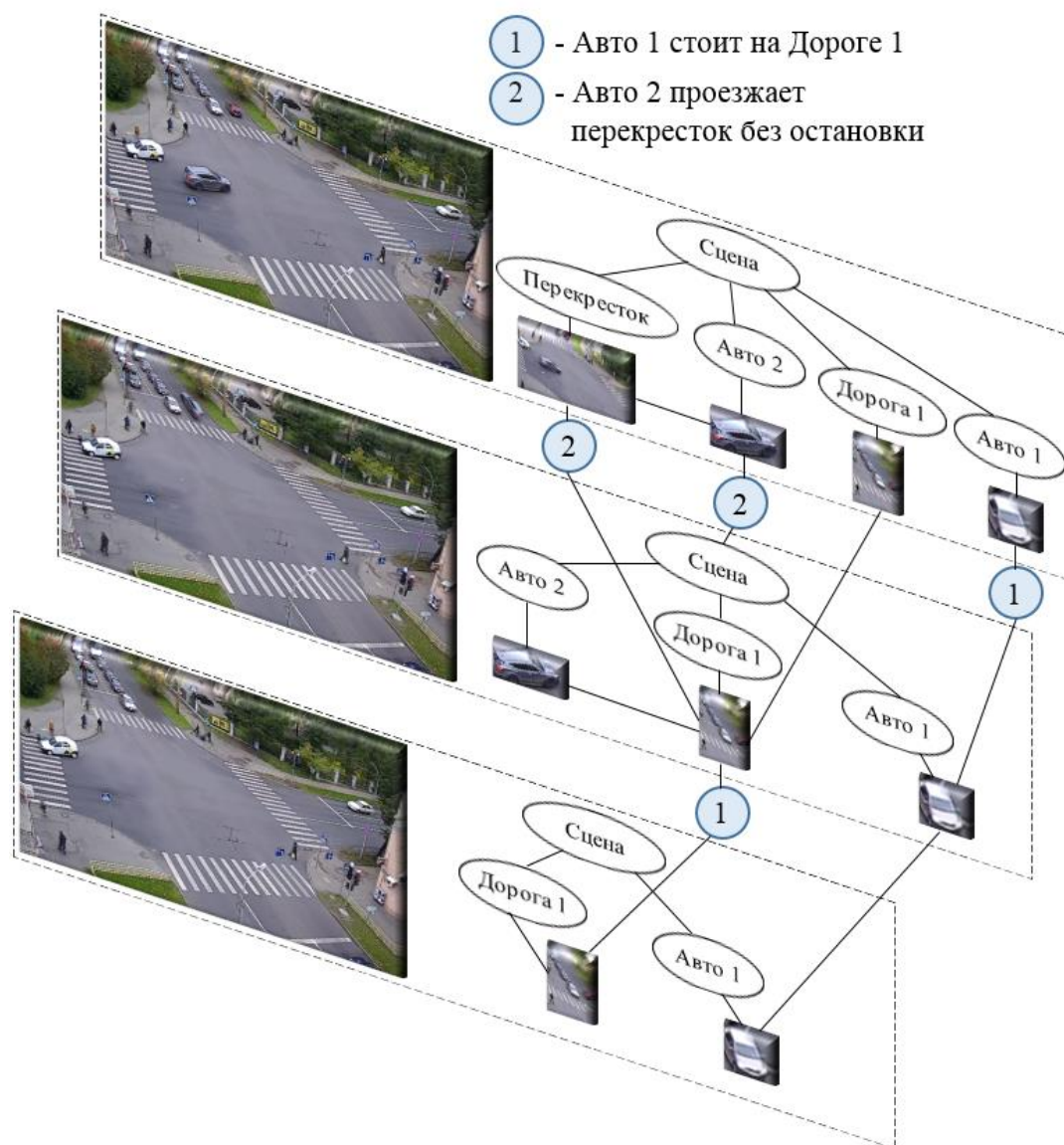


Рис. 3. Пример построения видеографа сложной динамической сцены
Fig. 3. An example of building a video grapher of a complex dynamic scene

Эффективность анализа сцен с применением предложенного метода построения видеографа в значительной степени определяется качеством локальных признаков, выявленных на этапе формирования локальных пространственных структур движения. Эти локальные признаки движения представляют собой более сложные информационные единицы по сравнению с традиционными статистическими характеристиками цветности, текстуры и геометрическими параметрами локальных областей изображений. Их использование позволяет более точно моделировать динамические процессы в сценах, учитывая нелинейные и многофакторные аспекты движения. В отличие от статистических подходов, которые часто основаны на упрощенных моделях и могут не учитывать сложные взаимодействия между элементами сцены, локальные признаки движения предоставляют более детализированную информацию о пространственно-временных характеристиках изменений в изображении. Это делает их особенно ценными для задач, требующих

высокой степени точности и чувствительности к изменениям в динамике сцены, таких как распознавание действий, анализ поведения и отслеживание объектов.

В контексте системы мониторинга дорожной обстановки, структура динамического видеографа, представляющего сложную сцену, эффективно сохраняется и индексируется для обеспечения принятия высокоуровневых решений в чрезвычайных ситуациях. Обновления видеографа происходят в режиме реального времени, отражая возникновение и исчезновение объектов наблюдения. Динамический видеограф служит основополагающим элементом для анализа ситуации и формирования обоснованных решений, что критически важно для оперативного реагирования на инциденты.

Технология распознавания динамических объектов значительно расширяет возможности интерпретации статических данных, применяя морфизмы для описания изменений проекций движущихся объектов в серии последовательных моментов времени относительно эталонных признаков. Это позволяет учитывать не только динамику объектов, но и их взаимодействие в пространственно-временном континууме.

Наблюдение за динамическими объектами открывает возможность использования множественных проекций для распознавания, которые формируют предысторию объекта при условии их принадлежности к одной и той же физической сущности. В таких случаях целесообразно применять модифицированные коллективные методы анализа и принятия решений, обеспечивающие повышенную точность и надежность результатов. Эти методы позволяют интегрировать данные из различных проекций, что способствует более полному и всестороннему пониманию ситуации.

4. Синхронизация и объединение входных модальностей от всех источников

В процессе интеграции различных сенсорных модальностей особое значение приобретает анализ корреляционных взаимосвязей между ними, который может быть реализован на различных уровнях интеграции с применением соответствующих аналитических методов. В рамках стратегии ранней интеграции, основанной на вычислении коэффициента корреляции Пирсона, были успешно решены задачи классификации видеокадров, распознавания речевой информации, отслеживания динамических объектов, идентификации источников звука и стохастического кодирования мультимедийных данных. В свою очередь, стратегия поздней интеграции, основанная на применении коэффициента конкордации Кендалла и анализе причинно-следственных связей, позволила эффективно решать задачи распознавания событий на основе наблюдательных данных, обеспечивая высокую точность и надежность результатов [8].

В процессе анализа мультимодальных данных ключевым аспектом является не только выявление корреляционных связей между отдельными сенсорными каналами, но и оценка их независимости, что в ряде случаев позволяет принимать более обоснованные и взвешенные решения. Сигналы, поступающие с различных сенсоров, как правило, регистрируются в разнообразных форматах и с различной частотой дискретизации, что обуславливает необходимость их синхронизации.

Синхронизация на уровне признаков заключается в интеграции данных, полученных в течение заданного временного интервала из разнородных, но коррелированных каналов. Это позволяет обеспечить согласованность анализа и интерпретации данных. В свою очередь, синхронизация на уровне принятия решений требует установления временных меток для интеграции выработанных решений. Таким образом, проблема синхронизации проявляется на различных уровнях интеграции данных из разнородных каналов. Для автоматизации анализа ДТП с использованием мультимодальных данных целесообразно применять методы объединения, основанные на классификации, и включающие соответствующие технологии для соотнесения наблюдаемой мультимодальной информации с заранее определенными классами.

Метод опорных векторов (*SVM*) при поздней интеграции данных из текстового, акустического и визуального каналов представляет собой эффективный инструмент для определения семантики сообщений, особенно в контексте объяснительных и телефонных звонков участников ДТП. Применение *SVM* для интеграции текстовой и визуальной информации позволяет осуществлять биометрическую аутентификацию участников события. Использование данного метода при

гибридной интеграции акустического и визуального каналов, включая такие параметры, как цвет, размер, яркость и контраст, позволяет системе анализировать мультимедийные данные, классифицировать изображения и выполнять семантическую индексацию видео на основе текстовой и визуальной информации [9].

Байесовский вывод является фундаментальным инструментом в области распознавания речи, особенно при использовании методов раннего и позднего комбинирования аудио- и визуальных признаков. Гибридные подходы, интегрирующие текстовые, акустические и визуальные каналы коммуникации, также находят широкое применение. В процессе наблюдения данные, поступающие из акустического канала (например, коэффициенты линейного предсказания) и визуального канала (такие как местоположение и площадь объектов), подвергаются комплексному анализу и интеграции для последующего распознавания событий. Этот процесс требует тщательного моделирования вероятностных распределений и применения алгоритмов оптимизации для достижения высокой точности и надежности распознавания [10].

Применение теории Демпстера–Шафера в контексте раннего и гибридного слияния визуальных характеристик обеспечивает высокоэффективную сегментацию изображений. В условиях позднего объединения данных из акустического и визуального каналов коммуникации становится возможным выполнение классификации видеоконтента, что, в свою очередь, способствует развитию систем человеко-машинного взаимодействия [11].

Динамические байесовские сети представляют собой мощный инструмент для интеграции акустических и визуальных данных в задачах классификации кадров видео, автоматического распознавания речи, биометрической аутентификации участников событий и мониторинга объектов интереса. При использовании на поздних этапах обработки визуальных данных и параметров сенсоров (камер) динамические байесовские сети обеспечивают аннотирование фотографий и отслеживание лиц на изображениях. В случае гибридного объединения информации из текстового, акустического и визуального каналов коммуникации, данные сети позволяют классифицировать видеоматериалы по тематическим категориям [12].

Интеграция данных акустического и визуального каналов коммуникации посредством мультимодальных нейронных сетей открывает новые перспективы в точной локализации источников звука и мониторинге объектов интереса, таких как транспортные средства и пешеходные потоки. Гибридные методы объединения визуальной информации, основанные на применении глубоких нейронных сетей, демонстрируют высокую эффективность в задачах компьютерного зрения, включая распознавание дорожных инцидентов и автоматизированное управление транспортными системами. Эти технологии находят применение в разработке интеллектуальных систем безопасности, улучшении транспортной инфраструктуры и оптимизации процессов экстренного реагирования [13].

Методы интеграции данных, основанные на количественных оценках, играют ключевую роль в мониторинге положения подвижных объектов, включая автомобили и пешеходов, с использованием мультимодальной информации. В частности, применение фильтра Калмана и его модификаций, а также фракционного фильтра, основанного на раннем и позднем объединении акустических и визуальных данных, демонстрирует высокую эффективность в решении задач определения местоположения, отслеживания траекторий движения транспортных средств и пешеходов, а также локализации источников звука, таких как визг тормозов или шум столкновения [14].

5. Интерпретация ДТП и принятие решения по результатам разбора

Общую схему интерпретации ДТП можно представить в следующем виде. Входной информацией являются мультимодальные данные, поступающие в систему от различных источников в различные временные интервалы: изображения и видеопоследовательности (включая звуковую дорожку), мета-данные текстового и речевого описания событий, связанных с ДТП, а также навигационные данные. Полученный ансамбль мультимодальных данных, принятый от разнородных источников, должен быть синхронизирован по времени и пространству и объединен в соответствии с оптимальной стратегией и методом объединения. Для корректной интерпретации

ДТП необходимо интегрировать базы данных слов и фраз, видеоконтент, онтологические модели, контекстную грамматику, а также словари визуальных и акустических паттернов. Эти ресурсы должны поддерживать динамическую систему обновлений и взаимодействовать с географическими информационными системами (ГИС). Анализ мультимодальных данных предполагает комплексное прохождение всех этапов обработки и распознавания, осуществляемое либо автоматически при использовании простых модальностей и высоко обученной системы распознавания, либо интерактивно с участием эксперта по базам данных и знаниям.

Таким образом, на выходе системы должно быть сгенерировано, например, текстовое описание результатов анализа входной мультимодальной информации, используемой в дальнейшем для поддержки принятия решения. Конечный пользователь сможет не только интерпретировать текстовые описания сцен и событий, но и формировать контентные запросы [15]. При этом интерпретация событий ДТП должна учитывать контекст, задаваемый пространственными отношениями объектов (например, «перекресток») или конкретными задачами (например, «обнаружение столкновения транспортных средств»), и включать распознавание множества активных действий. Применение при решении задач интерпретации ДТП моделей глубокого обучения, обеспечивающих комплексное восприятие и контекстное понимание сцены и позволяющих не только определять геометрические характеристики окружающих объектов, но и их семантическое значение, а также разработка программного обеспечения для анализа данных с ПрМ обеспечит интеллектуальную поддержку принятия решений в системах мониторинга дорожной ситуации [16].

ВЫВОДЫ

ДТП как правило является результатом сложной совокупности обстоятельств, формирующих причинно-следственные связи. Установление истинных причин нарушения правил безопасности, приведших к аварии, и факторов, способствовавших их возникновению, является ключевой задачей при расследовании правонарушений и неотъемлемой частью обеспечения безопасности дорожного движения и эксплуатации транспорта.

Полученные в работе результаты свидетельствуют о перспективности выбранного подхода к решению задачи автоматического разбора ДТП на основе мультимодального анализа данных с использованием технологий искусственного интеллекта. Однако отсутствие научно-методического аппарата в данной области требует его разработки и адаптации под конкретные практические задачи.

В качестве направлений дальнейшего исследования, обеспечивающих повышение полноты и достоверности принятия решения при автоматическом разборе ДТП, рассматриваются:

- 1) нахождение способов эффективной формализации различных модальностей (изображений, видеопоследовательностей, аудио- и текстовой информации) в соответствии с поставленными целевыми управляющими воздействиями;
- 2) комбинирование подхода к интерпретации сцен и событий на основе классических методов и технологий искусственного интеллекта;
- 3) разработка научно-методического аппарата для построения автоматической системы разбора ДТП.

Список литературы

1. Горюшинский В.С., Бичахчян А.В., Шацких Е.В. Общие принципы и особенности расследования ДТП // Вестник современных исследований. – № 12.1 (27), 2018. – С. 106-111.
2. Пинчук Л.В. Некоторые тактические аспекты детального осмотра места ДТП // Вестник Рязанского филиала Московского университета МВД России. – № 11, 2017. – С. 55-62.
3. Atrey P.K., Hossain M.A., Kankanhalli M.S. Multimodal Fusion for Multimedia Analysis: a Survey. *Multimedia Systems*. – № 16 (6), 2010. – P. 345-379.
4. Муленков Д.В. Горшков М.М. Фиксация факта, хода, содержания и результатов осмотра места происшествия // Законодательство и практика. – № 2 (31), 2013. – С. 10-13.

5. Фаворская М.Н., Шилов А.С. Алгоритмы реализации оценки движения в системах видеонаблюдения // Системы управления и информационные технологии. Перспективные исследования. – № 3.3 (33), 2008. – С. 408-112.
6. Фаворская М.Н. Сегментация визуальных движущихся объектов с применением групповых преобразований // В материалах XII междунар. симпозиума по непараметрическим методам в кибернетике и системному анализу, Красноярск. – 2010. – С. 156-159.
7. Фролов В.В. Организация выявления свидетелей и их допрос при расследовании ДТП // Гуманитарные, социально-экономические общественные науки. – № 8-9, 2017. – С. 261-264.
8. Басов О.О., Карпов А.А. Анализ стратегий и методов объединения мультимодальной информации // Информационно-управляющие системы. – №2, 2015. – С. 7-14.
9. Bredin H., Chollet G. Audio-visual Speech Synchrony Measure for Talking-Face Identity Verification // Proc. IEEE Intern. Conf. on Acoustics, Speech and Signal Processing, Paris, France. – № 2, 2007. – P. 233-236.
10. Chetty G., Wagner M. Audio-visual Multimodal Fusion for Biometric Person Authentication and Liveness Verification // Proc. NICTA-HCSNet Multimodal User Interaction Workshop, Sydney, Australia. – 2006. – P. 17-24.
11. Singh R., Vatsa M., Noore A., Singh S.K. Dempster-Shafer Theory Based Finger Print Classifier Fusion with Update Rule to Minimize Training Time // IEICE Electronics Express. – № 3 (20), 2006. – P. 429-435.
12. Nefian A.V., et al. Dynamic Bayesian Networks for Audio-visual Speech Recognition // EURASIP Journal on Applied Signal Processing. – № 11, 2002. – P. 1-15.
13. Ni J., Ma X., Xu L., Wang J. An Image Recognition Method Based on Multiple BP Neural Networks Fusion // Proc. IEEE Intern. Conf. on Information Acquisition. – 2007. – P. 429-435.
14. Gehrig T., et al. Kalman Filters for Audio-video Source Localization // Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, Germany. – 2005. – P. 118-121.
15. Макарова Е.А., Лагерев Д.Г. Модель обработки слабоструктурированных текстовых данных на русском языке для интеллектуальной поддержки информационного управления в динамических организационных системах // Модели, системы, сети в экономике, технике, природе и обществе. – № 3, 2022. – С. 104-125.
16. Офицеров А.И., Сафонов Д.А. Использование искусственного интеллекта в системах обеспечения комплексной безопасности охраняемого объекта // Экономика. Информатика. – № 50 (1), 2023. – С. 203-210.

References

1. Goryushinsky V.S., Bichakhchyan A.V., Shatskikh E.V. General principles and features of accident investigation // Bulletin of Modern Research. – № 12.1 (27), 2018. – P. 106-111.
2. Pinchuk L.V. Some tactical aspects of a detailed inspection of the accident site // Bulletin of the Ryazan Branch of the Moscow University of the Ministry of Internal Affairs of Russia. – № 11, 2017. – P. 55-62.
3. Atrey P.K., Hossain M.A., Kankanhalli M.S. Multimodal Fusion for Multimedia Analysis: a Survey. Multimedia Systems. – № 16 (6), 2010. – P. 345-379.
4. Mulenkov D.V. Gorshkov M.M. Recording the fact, course, content and results of the inspection of the scene // Legislation and practice. – № 2 (31), 2013. – P. 10-13.
5. Favorskaya M.N., Shilov A.S. Algorithms for the implementation of motion assessment in video surveillance systems // Management systems and information technologies. Promising research. – № 3.3 (33), 2008. – P. 408-112
6. Favorskaya M.N. Segmentation of visual moving objects using three transformations // In the proceedings of the XII International Symposium on Nonparametric Methods in Cybernetics and Systems Analysis, Krasnoyarsk. – 2010. – P. 156-159.
7. Frolov V.V. Organization of the identification of witnesses and their interrogation during the investigation of an accident // Humanities, socio-economic social sciences. – № 8-9, 2017 – P. 261-264.
8. Basov O.O., Karpov A.A. Analysis of strategies and methods of combining multimodal information // Information management systems. – № 2, 2015. – P. 7-14.
9. Bredin H., Chollet G. Audio-visual Speech Synchrony Measure for Talking-Face Identity Verification // Proc. IEEE Intern. Conf. on Acoustics, Speech and Signal Processing, Paris, France. – № 2, 2007. – P. 233-236.
10. Chetty G., Wagner M. Audio-visual Multimodal Fusion for Biometric Person Authentication and Liveness Verification // Proc. NICTA-HCSNet Multimodal User Interaction Workshop, Sydney, Australia. – 2006. – P. 17-24.

11. Singh R., Vatsa M., Noore A., Singh S. K. Dempster–Shafer Theory Based Finger Print Classifier Fusion with Update Rule to Minimize Training Time // IEICE Electronics Express. – № 3 (20), 2006. – P. 429-435.
12. Nefian A.V., et al. Dynamic Bayesian Networks for Audio-visual Speech Recognition // EURASIP Journal on Applied Signal Processing. – № 11, 2002. – P. 1-15.
13. Ni J., Ma X., Xu L., Wang J. An Image Recognition Method Based on Multiple BP Neural Networks Fusion // Proc. IEEE Intern. Conf. on Information Acquisition. – 2007. – P. 429-435.
14. Gehrig T., et al. Kalman Filters for Audio-video Source Localization //Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, Germany. – 2005. – P. 118-121.
15. Makarova E.A., Lagerev D.G. Model for processing weakly structured text data in Russian for intelligent support of information management in dynamic organizational systems // Models, systems, networks in economics, technology, nature and society. – № 3, 2022. – P. 104-125.
16. Ofitserov A.I., Safonov D.A. 2023. The Use of Artificial Intelligence in Systems for Ensuring the Integrated Security of a Protected Object // Economy. Informatics. – № 50 (1), 2023. – P. 203-210.

Басов Олег Олегович, доктор технических наук, доцент, руководитель центра искусственного интеллекта, Акционерное общество «АСТ», г. Москва, Россия

Офицеров Александр Иванович, кандидат технических наук, сотрудник, Академия Федеральной службы охраны Российской Федерации, г. Орёл, Россия

Basov Oleg Olegovich, Doctor of Technical Sciences, Associate Professor, Head of the Artificial Intelligence Center, Joint Stock Company «AST», Moscow, 123242, Russia

Ofitserov Alexandr Ivanovich, Candidate of Technical Sciences, Academy of the Federal security service of the Russian Federation, Orel, Russia