ФЕДЕРАЛЬНОЕ АГЕНТСТВО

ПО ТЕХНИЧЕСКОМУ РЕГУЛИРОВАНИЮ И МЕТРОЛОГИИ

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

ГОСТ Р 72563— 2026

Информационные технологии. Искусственный интеллект

СИТУАЦИОННАЯ ВИДЕОАНАЛИТИКА

Методология определения значений функциональных характеристик

Издание официальное

Москва
Российский институт стандартизации
2026

gost.e-ecolog.ru

Предисловие

1 РАЗРАБОТАН Обществом с ограниченной ответственностью «ВИДЕОИНТЕЛЛЕКТ» (ООО «ВИДЕОИНТЕЛЛЕКТ»)

2 ВНЕСЕН Техническим комитетом по стандартизации ТК 164 «Искусственный интеллект»

3 УТВЕРЖДЕН И ВВЕДЕН В ДЕЙСТВИЕ Приказом Федерального агентства по техническому регулированию и метрологии от 18 февраля 2026 г. № 144-ст

4 ВВЕДЕН ВПЕРВЫЕ

*Правила применения настоящего стандарта установлены в статье 26 Федерального закона от 29 июня 2015 г. № 162-ФЗ «О стандартизации в Российской Федерации». Информация об изменениях к настоящему стандарту публикуется в ежегодном (по состоянию на 1 января текущего года) информационном указателе «Национальные стандарты», а официальный текст изменений и поправок — в ежемесячном информационном указателе «Национальные стандарты». В случае пересмотра (замены) или отмены настоящего стандарта соответствующее уведомление будет опубликовано в ближайшем выпуске ежемесячного информационного указателя «Национальные стандарты». Соответствующая информация, уведомление и тексты размещаются также в информационной системе общего пользования — на официальном сайте Федерального агентства по техническому регулированию и метрологии в сети Интернет (*gost.e-ecolog.ru

© Оформление. ФГБУ «Институт стандартизации», 2026

Настоящий стандарт не может быть полностью или частично воспроизведен, тиражирован и распространен в качестве официального издания без разрешения Федерального агентства по техническому регулированию и метрологии

Содержание

1 Область применения 1
2 Нормативные ссылки 2
3 Термины, определения, сокращения и обозначения 2
4 Основные положения 5
5 Требования к объекту испытаний 7
6 Требования к условиям проведения испытаний 8
7 Требования к исходным данным 8
8 Методология проведения испытаний 13
9 Методики определения и расчета метрик 16
10 Документирование и отчетность 18
Приложение А (справочное) Форма представления результатов испытаний 20
Библиография 21

Введение

Преобразование визуальных данных в структурированную информацию становится все более востребовано в условиях стремительной цифровизации и общего роста объемов информации. Ситуационная видеоаналитика — это технологическая область, которая позволяет решать задачи распознавания событий на изображениях или последовательностей изображений (видеопотоков) при помощи методов компьютерного зрения и различных алгоритмов автоматизированного анализа.

Современные системы ситуационной видеоаналитики играют ключевую роль в обеспечении безопасности и оперативного контроля в различных областях, таких как транспорт, здравоохранение, промышленность и охрана правопорядка. Уровни эффективности и надежности таких систем зависят от множества эксплуатационных характеристик, которые необходимо точно определять и оценивать.

В настоящее время в Российской Федерации отсутствуют стандарты, определяющие терминологию, способы классификации, эксплуатационные характеристики и методологию проведения испытаний разнородных средств ситуационной видеоаналитики. Подходы к оценке эффективности работы таких систем, включая виды и методики испытаний, носят специализированный характер и не имеют единого базиса. Это приводит к тому, что федеральные органы исполнительной власти и организации промышленности используют различные наименования эксплуатационных характеристик и методики испытаний, что вызывает разногласия и затруднения при оценке результатов разработок различных компаний, а также при формировании и реализации технических и эксплуатационных требований к средствам ситуационной видеоаналитики.

Таким образом, актуальной задачей является выработка единых подходов и разработка стандартизированных терминологии, требований к эксплуатационным характеристикам и методологии проведения испытаний разнородных средств ситуационной видеоаналитики.

Настоящий стандарт фокусируется на функциональных характеристиках систем ситуационной видеоаналитики, которые занимают одно из центральных мест в совокупности эксплуатационных характеристик. Способность системы качественно выполнять свои задачи является основным критерием ее ценности, без которого остальные характеристики теряют смысл.

Методология, описанная в настоящем стандарте, разработана с целью установления единых, обоснованных требований и методов определения значений функциональных характеристик, обеспечивающих объективную оценку и возможность сравнения различных систем ситуационной видеоаналитики. Настоящий стандарт описывает основные принципы, процедуры и методы, используемые для расчета наиболее критически важных характеристик, таких как точность и вероятность детекции требуемых сценариев ситуаций.

Введение настоящего стандарта позволит не только повысить качество предлагаемых решений на рынке, но и укрепить доверие пользователей к данным системам, обеспечив возможность проверки их соответствия установленным требованиям и нормативам.

Стандарт предназначен для использования при разработке технических заданий, программ и методик испытаний, а также при производстве продукции в данной предметной области. Он предназначен для использования коммерческими организациями, испытательными лабораториями и сертификационными органами в целях проведения объективных и воспроизводимых расчетов значений функциональных характеристик, что в свою очередь будет способствовать развитию и совершенствованию технологий в данной области.

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

**Информационные технологии.
Искусственный интеллект**

СИТУАЦИОННАЯ ВИДЕОАНАЛИТИКА

Методология определения значений функциональных характеристик

Information technology. Artificial intelligence. Situational video analytics. Methodology for determining the values of functional characteristics

Дата введения — 2026—03—01

1 Область применения

Настоящий стандарт распространяется на системы ситуационной видеоаналитики, используемые для автоматизированного анализа видеоданных с целью выявления, распознавания, отслеживания и регистрации различных объектов и событий.

Стандарт применим к системам ситуационной видеоаналитики, за исключением биометрических, в следующих областях:

- общественная безопасность и правоохранительная деятельность;
- транспортный комплекс;
- железнодорожный транспорт;
- промышленные и коммерческие объекты;
- здравоохранение;
- образование и учебные заведения;
- государственный сектор и административные учреждения;
- финансовые учреждения;
- энергетика и критическая инфраструктура;
- жилищно-коммунальное хозяйство.

Перечисленные области применения приведены в качестве примеров и не ограничивают другие возможные сценарии использования систем ситуационной видеоаналитики.

Настоящий стандарт:

- устанавливает общие требования к проведению функциональных и сравнительных испытаний систем ситуационной видеоаналитики;
- определяет наиболее существенные эксплуатационные характеристики функционирования систем ситуационной видеоаналитики;
- устанавливает требования к исходным данным, процедурам испытаний и форме представления протоколов с результатами испытаний;
- является основой для разработки программ и методик испытаний, исключающих систематические ошибки, которые обусловлены несоответствующими требованиям настоящего стандарта процедурами сбора и анализа данных с целью проверки значений характеристик, заявляемых производителями систем ситуационной видеоаналитики в технических условиях;
- определяет ключевые количественные показатели, подлежащие расчету для оценки значений функциональных характеристик систем ситуационной видеоаналитики;
- устанавливает методики и процедуры определения ключевых количественных показателей;
- устанавливает требования к условиям определения значений функциональных характеристик;

Издание официальное

- обеспечивает единство и сопоставимость результатов, получаемых различными организациями и специалистами;
- определяет формы представления результатов испытаний.

2 Нормативные ссылки

В настоящем стандарте использованы нормативные ссылки на следующие стандарты:

ГОСТ 19.501 Единая система программной документации. Формуляр. Требования к содержанию и оформлению

ГОСТ 19.505—79 Единая система программной документации. Руководство оператора. Требования к содержанию и оформлению

ГОСТ Р 58973 Оценка соответствия. Правила к оформлению протоколов испытаний

ГОСТ Р 59385 Информационные технологии. Искусственный интеллект. Ситуационная видеоаналитика. Термины и определения

ГОСТ Р ИСО/МЭК 25010—2015 Информационные технологии. Системная и программная инженерия. Требования и оценка качества систем и программного обеспечения (SQuaRE). Модели качества систем и программных продуктов

Примечание — При пользовании настоящим стандартом целесообразно проверить действие ссылочных стандартов в информационной системе общего пользования — на официальном сайте Федерального агентства по техническому регулированию и метрологии в сети Интернет или по ежегодному информационному указателю «Национальные стандарты», который опубликован по состоянию на 1 января текущего года, и по выпускам ежемесячного информационного указателя «Национальные стандарты» за текущий год. Если заменен ссылочный стандарт, на который дана недатированная ссылка, то рекомендуется использовать действующую версию этого стандарта с учетом всех внесенных в данную версию изменений. Если заменен ссылочный стандарт, на который дана датированная ссылка, то рекомендуется использовать версию этого стандарта с указанным выше годом утверждения (принятия). Если после утверждения настоящего стандарта в ссылочный стандарт, на который дана датированная ссылка, внесено изменение, затрагивающее положение, на которое дана ссылка, то это положение рекомендуется применять без учета данного изменения. Если ссылочный стандарт отменен без замены, то положение, в котором дана ссылка на него, рекомендуется применять в части, не затрагивающей эту ссылку.

3 Термины, определения, сокращения и обозначения

3.1 Термины и определения

В настоящем стандарте применены термины по ГОСТ Р 59385, а также следующие термины с соответствующими определениями:

3.1.1

автоматизированное рабочее место; АРМ: Программно-технический комплекс АС, предназначенный для автоматизации деятельности определенной категории пользователей или определенного вида деятельности.

[ГОСТ Р 59853—2021, статья 3]

3.1.2 **анализ распределения ошибок:** Исследование равномерности распределения ошибок по группам испытаний.

3.1.3 **внутренняя структура кадровых диапазонов:** Заранее определенная организация последовательностей кадров в исходных данных, разделяющая их на логические сегменты для решения задач испытаний (прогрев, контроль, определение кадрового сдвига обнаружения).

3.1.4 **доля правильных исходов:** Доля правильно классифицированных случаев среди всех реально существующих.

3.1.5 **идентификация систематических отклонений:** Выявление и фиксация проблем или потенциальных уязвимостей в работе алгоритма объекта испытаний.

3.1.6 **индекс Жаккара:** Отношение площади пересечения областей двух прямоугольных рамок или замкнутых контуров, описывающих один и тот же объект, к площади объединения этих областей.

3.1.7 **истинно отрицательный исход:** Испытание с отсутствием сценария ситуации считается пройденным успешно, если объект испытаний не зарегистрировал требуемого сценария в соответствующей контрольной части исходных данных.

3.1.8 **истинно положительный исход:** Испытание с наличием сценария ситуации считается пройденным успешно, если объект испытаний зарегистрировал один требуемый сценарий в соответствующей контрольной части исходных данных.

3.1.9 **кадр начала:** Номер кадра в исходных данных, определенный объектом испытаний в качестве начала сценария ситуации.

3.1.10 **кадровый сдвиг обнаружения:** Разность между номером эталонного кадра начала сценария ситуации и номером кадра, по результатам обработки которого объектом испытаний зафиксировано начало сценария; выражается в кадрах и характеризует задержку обнаружения.

3.1.11 **критерий достоверности:** Значение для заданного уровня достоверности.

3.1.12 **критерий точности:** Допустимая погрешность (ошибка выборки), выраженная в десятичной форме.

3.1.13 **ложноотрицательный исход:** Испытание с наличием сценария ситуации считается пройденным неуспешно, если объект испытаний не зарегистрировал требуемый сценарий или зарегистрировал больше одного требуемого сценария в соответствующей контрольной части исходных данных.

3.1.14 **ложноположительный исход:** Испытание с отсутствием сценария ситуации считается пройденным неуспешно, если объект испытаний зарегистрировал один или более требуемых сценариев в соответствующей контрольной части исходных данных.

3.1.15

метрика: Материальная мера некоторых аспектов характеристик качества.
[ГОСТ Р 59898—2021, пункт 3.12]

3.1.16 **нулевая гипотеза:** Объект испытаний не идентифицирует наличие сценария ситуации, когда сценарий отсутствует.

3.1.17 **объект испытаний:** Система ситуационной видеоаналитики, разработанная для автоматизированного анализа видеоданных с целью выявления, классификации и регистрации предопределенных сценариев ситуаций по заданному алгоритму анализа видеоизображений или по набору алгоритмов.

3.1.18 **отклонение кадра начала:** Разница в количестве кадров между кадром начала сценария и эталонным кадром начала сценария в исходных данных.

3.1.19 **оценка стабильности объекта:** Исследование распределения ошибок по группам испытаний для оценки стабильности объекта испытаний в различных условиях (в зависимости от выбранного признака или набора признаков для группировки).

3.1.20 **ошибка второго рода (β-ошибка):** Нулевая гипотеза принимается, хотя она неверна, то есть система ситуационной видеоаналитики не выявляет сценарий ситуации, который действительно присутствует.

3.1.21 **ошибка первого рода (α-ошибка):** Нулевая гипотеза отвергается, хотя она верна, то есть система ситуационной видеоаналитики ложно идентифицирует сценарий ситуации, которого на самом деле нет.

3.1.22 **предварительное значение доли успеха:** Предполагаемая оценка чувствительности, определяемая по результатам проведения предварительного тестирования или посредством использования данных, заявленных в эксплуатационных документах на объект испытаний.

3.1.23 **размер выборки:** Статистически значимое количество испытаний в серии, которое необходимо при исследовании пропорций в генеральной совокупности.

3.1.24 **специфичность:** Доля правильно классифицированных отрицательных случаев среди всех реально отрицательных.

3.1.25 **стенд:** Специализированная программно-аппаратная среда, используемая для испытания систем ситуационной видеоаналитики в части комплексной оценки значений их функциональных характеристик.

3.1.26 **точность:** Доля правильно классифицированных случаев (как с наличием, так и с отсутствием искомых сценариев) относительно общего числа случаев, классифицированных объектом испытаний как положительные.

3.1.27

функциональная корректность: Степень обеспечения продуктом или системой необходимой степени точности корректных результатов.
[ГОСТ Р ИСО/МЭК 25010—2015, пункт 4.2.1.2]

3.1.28

функциональная полнота: Степень покрытия совокупностью функций всех определенных задач и целей пользователя.

[ГОСТ Р ИСО/МЭК 25010—2015, пункт 4.2.1.1]

3.1.29

функциональная пригодность: Степень, в которой продукт или система обеспечивают выполнение функции в соответствии с заявленными и подразумеваемыми потребностями при использовании в указанных условиях.

[ГОСТ Р ИСО/МЭК 25010—2015, пункт 4.2.1]

3.1.30

функциональная целесообразность: Степень функционального упрощения выполнения определенных задач и достижения целей.

[ГОСТ Р ИСО/МЭК 25010—2015, подпункт 4.2.1.3]

3.1.31 **чувствительность:** Доля правильно классифицированных положительных случаев среди всех реально положительных.

3.1.32 **эталонный кадр начала:** Номер кадра в исходных данных, заранее установленный в качестве начала сценария ситуации.

3.2 Сокращения

В настоящем стандарте применены следующие сокращения:

AVI — аудио-, видеочередование (Audio Video Interleave);

CSV — значения, разделенные запятыми (Comma-Separated Values);

FN — ложноотрицательный (False Negative);

FP — ложноположительный (False Positive);

HLS — потоковая передача по протоколу HTTP (HTTP Live Streaming);

HTTP — протокол передачи гипертекста (HyperText Transfer Protocol);

JPG (JPEG) — объединенная группа экспертов по фотографии (Joint Photographic Experts Group);

JSON — обозначение объектов на языке программирования JavaScript (JavaScript Object Notation);

MOV — формат файла мультимедийной платформы QuickTime;

MP4 — международный стандарт MPEG-4 часть 14 (MPEG-4 Part 14);

MPEG — экспертная группа по видеоданным (Moving Picture Experts Group);

PNG — переносимая сетевая графика (Portable Network Graphics);

RTSP — протокол потоковой передачи в реальном времени (Real-Time Streaming Protocol);

SDK — набор инструментов для разработки программного обеспечения (Software Development Kit);

TN — истинно отрицательный (True Negative);

TP — истинно положительный (True Positive);

XML — расширяемый язык разметки (eXtensible Markup Language).

3.3 Обозначения

В настоящем стандарте применены следующие обозначения:

α — ошибка первого рода;

β — ошибка второго рода;

Ac — метрика «Доля правильных исходов»;

DFS — метрика «Кадровый сдвиг обнаружения»;

E — критерий точности;

H_0 — нулевая гипотеза;

J — метрика «Индекс Жаккара»;

K — размер выборки;

Me_{dfs} — медиана кадрового сдвига обнаружения;

Me_J — медиана индекса Жаккара;

Me_{sfa} — медиана отклонения кадра начала;

p — предварительное значение доли успеха;
Pr — метрика «Точность»;
Se — метрика «Чувствительность»;
SFA — метрика «Отклонение кадра начала»;
Sp — метрика «Специфичность»;
Z — критерий достоверности.

4 Основные положения

4.1 Общие положения

4.1.1 Настоящий стандарт устанавливает единые требования и методики для определения значений функциональных характеристик систем ситуационной видеоаналитики.

4.1.2 Стандарт направлен на обеспечение обоснованности, точности, надежности, воспроизводимости, объективности и сопоставимости результатов определения значений функциональных характеристик для различных систем ситуационной видеоаналитики и условий их применения.

4.2 Роли и участники испытаний

Роли и участники испытаний, как правило, являются общепринятыми (такие как разработчик, сертификационная лаборатория и заказчик) и определяются в зависимости от специфики и условий проведения испытаний. Состав участников, их полномочия и ответственность могут быть уточнены и согласованы сторонами, принимающими участие в испытаниях, в установленном порядке и не требуют дополнительного описания в стандарте.

4.3 Объект испытаний

4.3.1 Целевые сценарии ситуаций, подлежащие детектированию объектом испытаний, включают сценарии согласно [1] (пункт 37), но не ограничиваются только ими.

Примечание — Целевые сценарии ситуаций могут охватывать другие области применения и определяются с учетом назначения системы и нормативных требований, соответствующих области применения.

4.3.2 Эксплуатационные характеристики объекта испытаний определяются моделью качества системы/программного продукта по ГОСТ Р ИСО/МЭК 25010—2015 (рисунок 4) и представляют собой совокупность свойств и показателей, которые характеризуют различные аспекты использования системы/программного продукта в реальных условиях.

4.3.3 В модели качества продукта отражены сведения о свойствах качества продукта, которые представлены следующими характеристиками: функциональная пригодность, уровень производительности, совместимость, удобство использования, надежность, защищенность, сопровождаемость и переносимость.

4.3.4 Каждая характеристика, в свою очередь, состоит из нескольких связанных подхарактеристик, которые могут быть определены количественными показателями — метриками.

4.3.5 В рамках настоящего стандарта для объекта испытаний наиболее существенной и критически важной эксплуатационной характеристикой является функциональная пригодность.

4.3.6 Подхарактеристиками функциональной пригодности являются:

- функциональная полнота.

Примечание — В настоящем стандарте функциональная полнота достигается за счет соответствия требованиям к минимальной функциональности объекта испытаний согласно 5.1;

- функциональная целесообразность.

Примечание — В настоящем стандарте не приводятся методики определения метрик для функциональной целесообразности в силу субъективности и сложности количественной оценки того, насколько функции системы ситуационной видеоаналитики могут соответствовать ожиданиям и задачам пользователей;

- функциональная корректность.

4.3.7 В рамках настоящего стандарта для количественного определения функциональной корректности используются следующие метрики (см. рисунок 1):

- отклонение кадра начала;

- кадровый сдвиг обнаружения;
- индекс Жаккара;
- чувствительность;
- специфичность;
- точность;
- доля правильных исходов.

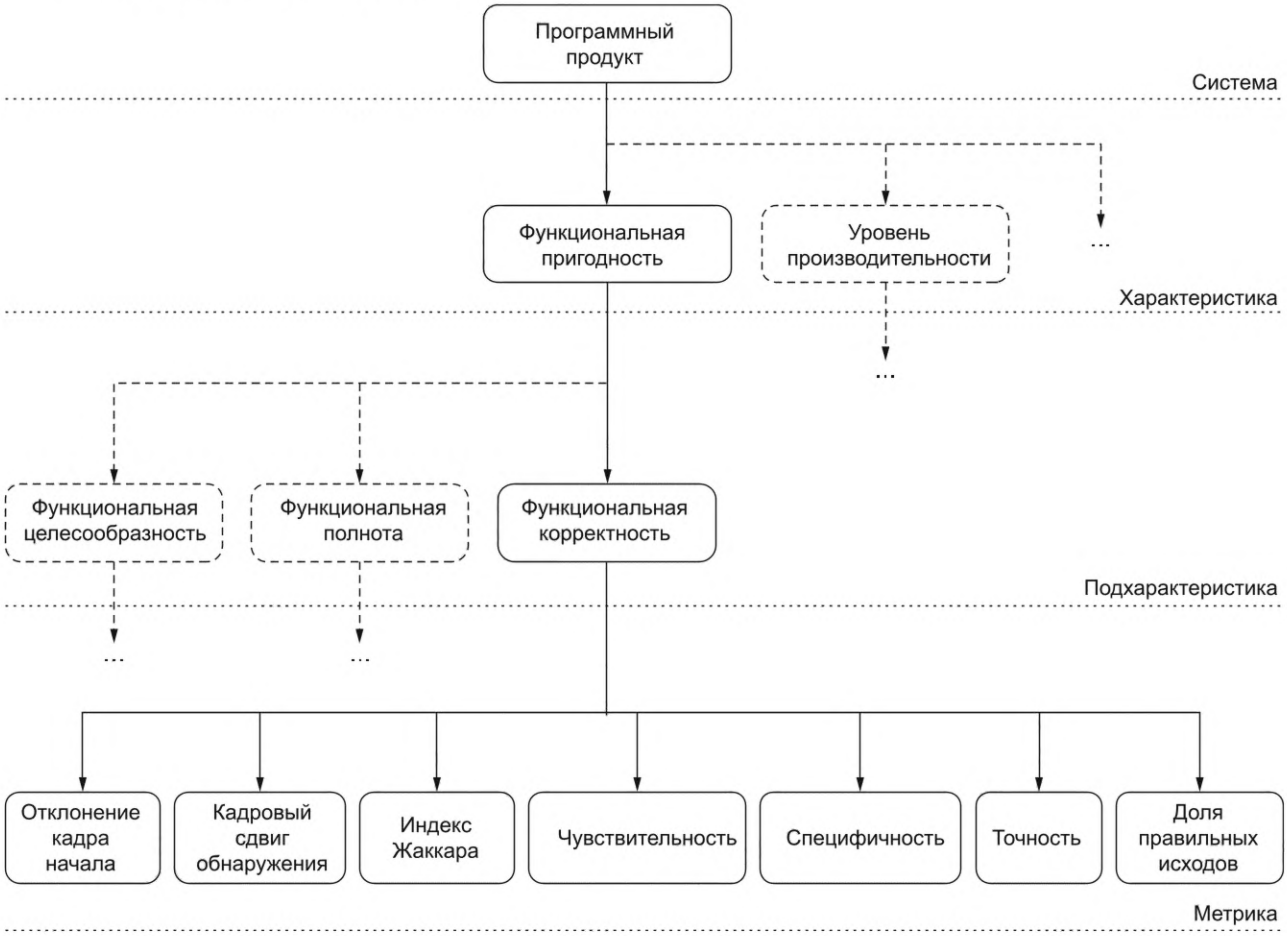


Рисунок 1 — Метрики функциональной корректности

4.4 Основные принципы и подходы к определению

4.4.1 Объективность и воспроизводимость: все методики и процедуры, описанные в стандарте, направлены на обеспечение объективности и воспроизводимости определения значений функциональных характеристик.

4.4.2 Универсальность: стандарт ориентирован на создание универсальных и применимых методик, которые могут использоваться в различных условиях и сценариях.

4.4.3 Учет внешних условий и факторов: при определении значений функциональных характеристик необходимо учитывать и обеспечивать воздействие внешних условий и факторов, которые соответствуют эксплуатационным характеристикам систем ситуационной видеоаналитики. Внешние условия и факторы должны быть задокументированы.

4.4.4 Для проведения испытаний необходимо использовать оборудование и программное обеспечение, прошедшие соответствующую поверку, сертификацию и лицензирование в установленном порядке.

4.4.5 Профессиональная компетентность: испытания могут проводить только специалисты, обладающие подтвержденной в установленном порядке квалификацией, необходимыми знаниями и опытом эксплуатации систем ситуационной видеоаналитики, включая их настройку, обслуживание и мониторинг.

5 Требования к объекту испытаний

5.1 Общие требования к функциональности

5.1.1 В общем случае объект испытаний представляет собой функциональный блок, функционирующий в условиях, отвечающих требованиям эксплуатационной документации. Структурная схема функциональной модели объекта испытаний показана на рисунке 2.

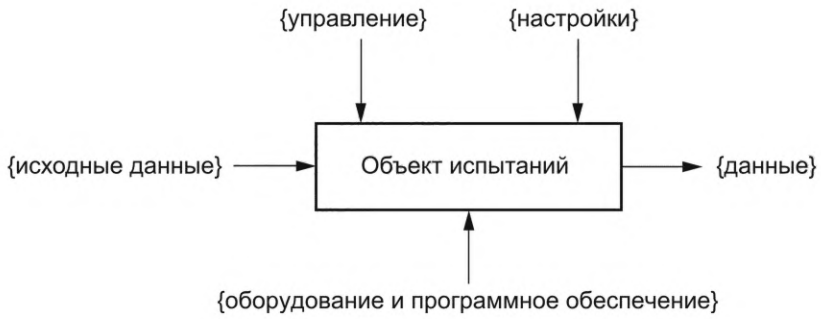


Рисунок 2 — Функциональная модель объекта испытаний

Функциональная модель предполагает, что испытания проводятся с использованием оборудования и программного обеспечения, а для анализа исходных данных с целью формирования результатов (данные) у объекта испытаний необходимо наличие интерфейса для получения исходных данных, управляющих сигналов (управление) и файлов конфигурации (настройки).

Примечание — Обычно для функционирования видеоаналитического детектора получение настроек является опциональным, но для обеспечения детектирования целевых сценариев ситуаций настоящего стандарта условие предопределения зон, линий и значений времени реагирования является обязательным для соответствующих сценариев ситуаций.

5.1.2 В соответствии с обобщенной функциональной моделью объект испытаний должен удовлетворять следующим минимальным функциональным требованиям:

- получению исходных данных испытаний для анализа;
- получению и выполнению управляющих сигналов;
- формированию данных с результатами анализа исходных данных;
- настройке параметров работы объекта испытаний.

5.2 Требования к документации

5.2.1 Документация на объект испытаний должна соответствовать требованиям нормативных документов (Единая система программной документации) и предоставляться вместе с объектом испытаний на электронных или бумажных носителях.

5.2.2 Комплект документов, сопровождающий объект испытаний, в обязательном порядке должен включать:

- формуляр, разработанный в соответствии с требованиями ГОСТ 19.501;
- руководство оператора, разработанное в соответствии с требованиями ГОСТ 19.505.

5.2.3 Независимо от состава комплекта документов, информация о сопровождении и эксплуатации объекта испытаний должна содержать сведения:

- о наименовании и описании объекта испытаний;
- назначении и области применения объекта испытаний;
- технических характеристиках объекта испытаний;
- настройках, параметрах и факторах, влияющих на функционирование;
- условиях и ограничениях эксплуатации;
- способах и интерфейсах взаимодействия с объектом испытаний;
- совместимости с различными операционными системами, платформами и устройствами.

6 Требования к условиям проведения испытаний

6.1 Общие требования к испытательному стенду

6.1.1 Испытания объекта необходимо проводить с использованием испытательного стенда, соответствующего и воспроизводящего условия предполагаемой реальной эксплуатации.

6.1.2 Программная и аппаратная части испытательного стенда должны удовлетворять требованиям, указанным в руководстве оператора на объект испытаний по ГОСТ 19.505—79 (пункт 2.2).

6.1.3 Стенд должен иметь системы безопасности, которые обеспечивают защиту компонентов и данных от несанкционированного доступа.

6.2 Требования к составу программной части

В общем случае программная часть стенда для испытаний объекта должна включать в себя:

- операционную систему (или набор систем), конфигурация которой позволяет испытывать объект в контролируемых условиях;
- программные инструменты, реализующие поведение устройств или сервисов, с которыми может взаимодействовать объект испытаний;
- набор данных, используемых для испытаний объекта, включая входные данные, настройки и др.;
- скрипты или программы, автоматизирующие процесс испытаний, включая запуск тестов, проверку результатов и генерацию отчетов;
- инструменты для мониторинга программного обеспечения, включая профилирование, трассировку и анализ ошибок.

6.3 Требования к составу аппаратной части

В общем случае аппаратная часть стенда для испытаний объекта должна включать в себя:

- автоматизированное рабочее место, предназначенное для конфигурации, управления испытаниями и контроля состояния объекта испытаний;
- оборудование, обеспечивающее возможность запуска объекта испытаний и обработки исходных данных;
- хранилище данных, обеспечивающее возможность хранения исходных данных, результатов испытаний, промежуточных данных, образов, файлов установки и др.;
- сетевое оборудование (линии связи, маршрутизаторы, коммутаторы и другие устройства), которое обеспечивает возможность взаимодействия частей испытательного стенда между собой.

7 Требования к исходным данным

7.1 Требования к качественному составу исходных данных

7.1.1 Для достижения статистической значимости результатов испытаний необходимо проведение серии испытаний (см. рисунок 3) для объекта по каждому отдельному сценарию ситуации.

7.1.2 Серия испытаний s должна представлять собой совокупность репрезентативных элементов исходных данных e двух типов t, объединенных в группы g по принципу единого набора признаков сценария ситуации.

7.1.3 Совокупность репрезентативных элементов исходных данных представляет собой выборку наблюдаемых событий или объектов, сформированную из всей доступной генеральной совокупности и отражающая ее структуру и разнообразие.

7.1.4 Исходными данными являются заранее подготовленные по 7.1.16 референсные данные в виде видеофайлов определенных форматов, которые после декодирования в последовательность изображений (кадров) передаются объекту испытаний для видеоанализа.

7.1.5 Характеристики исходных данных (в том числе декодированных исходных данных) должны соответствовать требованиям, указанным в руководстве оператора на объект испытаний по ГОСТ 19.505—79 (пункты 2.2, 2.3).

Примечания

1 Исходные данные могут быть представлены в виде видеофайлов в распространенных контейнерных форматах (например, AVI, MP4, MOV), отдельных изображений (например, PNG, JPEG) или потокового видео, передаваемого по стандартным сетевым протоколам (например, RTSP, HLS), включая данные от тепловизионных камер.

2 Декодирование исходных данных может осуществляться как инструментами испытательного стенда, так и набором инструментов (SDK), которые предоставляются поставщиком объектом испытаний.

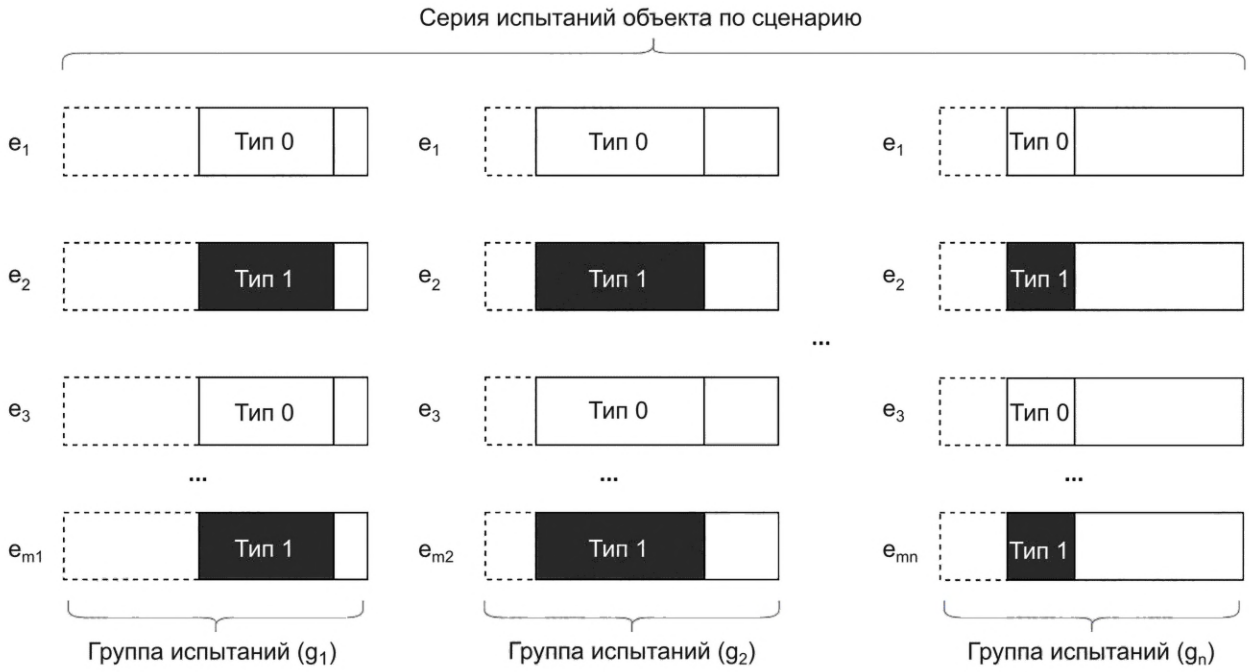


Рисунок 3 — Структура серии испытаний

7.1.6 Достоверное представление о значениях функциональных характеристик объекта испытаний можно получить, определяя метрики только в контролируемых условиях с заранее известными исходами.

7.1.7 Следующие метрики: чувствительность, специфичность, точность и доля правильных исходов — являются статистическими показателями, выводимыми из α-ошибок (ложно положительные заключения) и β-ошибок (ложно отрицательные заключения) в бинарной классификации с нулевой гипотезой H_0.

Таким образом, элементы исходных данных, входящие в группы серии испытаний, должны быть представлены двумя типами по формуле

$$type(e_i) = t_i \in \{0,1\}, \quad (1)$$

где тип 0 — это элемент исходных данных с отсутствием сценария ситуации;

тип 1 — это элемент исходных данных с наличием сценария ситуации.

7.1.8 Структура обоих типов элементов исходных данных (см. рисунок 4) для испытаний должна содержать:

- начальное количество данных (диапазон кадров) для прогрева, то есть вхождения объекта испытаний в рабочее состояние.

Примечание — Количество начальных данных для прогрева может быть равным нулю, фиксированным или вычисленным (подобранным путем постепенного увеличения до вырождения существенного влияния на значения функциональных характеристик);

- контрольные данные (диапазон кадров), которые характеризуются наличием или отсутствием изменений, подлежащих обнаружению объектом испытаний в качестве сценария ситуации;
- данные для определения кадрового сдвига обнаружения (диапазон кадров), которые необходимы для определения скорости обнаружения сценария ситуации.

7.1.9 Все события, зарегистрированные на начальных данных прогрева и данных для определения кадрового сдвига обнаружения, не учитываются в качестве результатов испытаний.

7.1.10 Объем данных, необходимых для определения кадрового сдвига обнаружения, должен соответствовать заявленным техническим характеристикам на объект испытаний — то есть быть не менее эквивалента количества кадров, которое производитель указывает как максимальное время реакции системы на обнаружение требуемого сценария ситуации.

Примечание — Для объектов испытаний, у которых формирование событий в режиме реального времени не предусмотрено, допускается количество данных для определения кадрового сдвига обнаружения приравнивать к нулю.

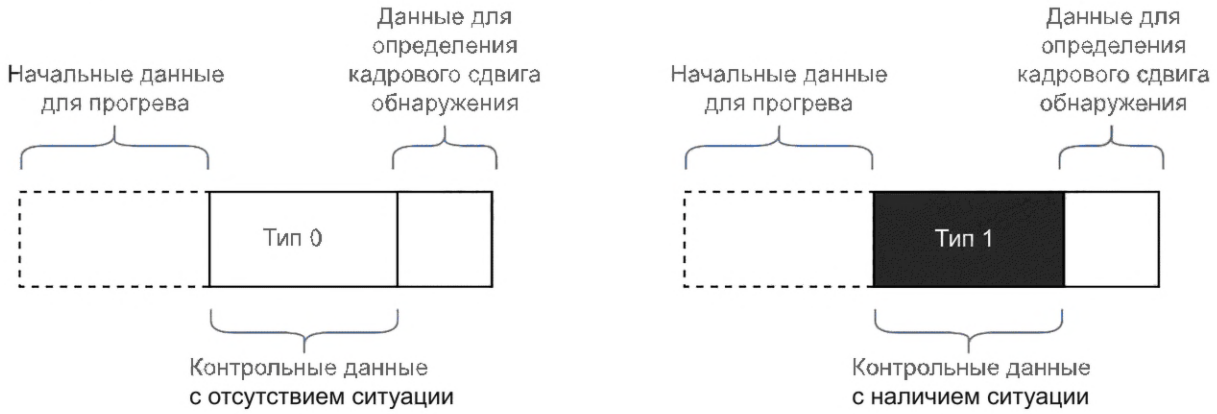


Рисунок 4 — Структура типов элементов исходных данных для испытаний

7.1.11 Суммарное количество данных в структуре каждого типа определяется при подготовке программы испытаний.

7.1.12 Разница в количестве начальных данных, контрольных данных и данных для определения кадрового сдвига обнаружения должна быть минимальна у всех элементов исходных данных в рамках одной группы в серии.

7.1.13 В серии испытаний элементы исходных данных должны быть сгруппированы по определенным признакам или набору признаков, которые определяются при подготовке программы испытаний.

В качестве объединяющего признака может выступать как сама сцена видеонаблюдения, так и объекты детектирования в сцене видеонаблюдения, условия освещенности или любые другие значимые для сценария ситуации признаки.

7.1.14 В каждой группе должны быть представлены оба типа данных, причем каждому элементу одного типа должен соответствовать элемент другого типа с теми же признаками, совпадающими по структуре и количеству.

7.1.15 Вероятность получения истинного значения статистической метрики в серии испытаний, состоящей из одной группы, вычисляется по схеме Бернулли, т. к. выполняются условия по 8.3.1.

Примечание — Вероятность P того, что из n испытаний будет точно k успешных исходов (истинных значений метрики), может быть описана биномиальной формулой

$$P(X = k) = C_n^k \cdot p^k \cdot (1 - p)^{n-k}, \quad (2)$$

где X — количество успешных исходов;

$C_n^k = \frac{n!}{k! \cdot (n-k)!}$ — число сочетаний (биномиальный коэффициент);

p — вероятность успеха;

n — общее количество испытаний.

7.1.16 В целях обеспечения точности, воспроизводимости и понятности испытаний элементы исходных данных должны быть сопровождены референсными данными (аннотированы), содержащими следующую информацию:

- наименование серии испытания (например, наименование или шифр сценария ситуации);
- принадлежность к группе испытаний;
- тип элемента исходных данных;

- общую продолжительность исходных данных в кадрах (и/или длительность в секундах);
- внутреннюю структуру кадровых диапазонов;
- эталонные кадры начала сценария ситуации;
- координаты зон детектирования (относительные);
- описание сцены видеонаблюдения;
- описание объектов и их действий;
- описание сценария ситуации (для типа 1 исходных данных);
- координаты границ объектов (bounding boxes/contour);
- метаданные (частота кадров, разрешение видео, условия съемки и другие технические параметры).

7.1.17 Указатель на конкретный кадр или диапазон кадров в видеофайле может быть представлен в произвольном виде, но должен быть однозначным для обеспечения точного сравнения с результатами испытаний.

Примечание — В случае использования временных меток рекомендуется указывать точные временные метки в виде кадра и времени (например, в формате hh:mm:ss:ff, где ff — номер кадра).

7.1.18 Серия испытаний также должна сопровождаться отдельной аннотацией, объясняющей структуру наименования файлов элементов исходных данных (например, s-g-e-t.format, где s — шифр серии, g — номер группы, e — номер элемента в группе, t — тип элемента, .format — формат элемента), описывающей шифры, группы и т. п.

Аннотации могут быть представлены в виде структурированных текстовых файлов форматов XML, JSON, CSV или др.

7.1.19 Формирование тестовой выборки (элементов исходных данных для серии испытаний) может быть произведено из архивных данных фактической работы систем видеонаблюдения и ситуационной видеоаналитики на реальных объектах при выполнении всех требований к качественному и количественному составу исходных данных. В общем случае процесс признаковой сегментации таких данных должен включать:

- сбор видеоданных с реальных объектов в качестве исходных данных;
- подготовку данных к использованию в тестовой выборке (разметка входных данных);
- анализ степени репрезентативности и признакового разнообразия сцен видеонаблюдения исходных данных;
- определение интересуемых признаковых групп для серии испытаний;
- определение структуры типов элементов исходных данных по группам;
- формирование элементов исходных данных серии испытаний;
- определение референсных данных.

7.1.20 Итоговое количество элементов исходных данных в серии, полученной способом признаковой сегментации, должно быть максимально возможным при условии сохранения высокой степени репрезентативности.

7.2 Требования к количественному составу исходных данных

7.2.1 Серия испытаний состоит из n групп (g_i), имеющих m_i элементов в каждой группе (см. рисунок 3), то есть $|g_i| = m_i$, где m, $n \in N$. Таким образом, общее количество элементов исходных данных в серии — это $M = \sum_{i=1}^{n} m_i$, которое должно удовлетворять соотношению

$$M \geq K, \tag{3}$$

где M — общее количество элементов серии;

K — минимально необходимый размер выборки для серии испытаний, обеспечивающий необходимую статистическую значимость результатов испытаний.

7.2.2 Для обеспечения точности и объективности результатов определения значений функциональных характеристик объекта испытаний необходимо производить равнозначное воздействие на оценку α- и β-ошибок как во всей серии испытаний, так и в отдельных группах, то есть элементы исходных данных в группах (как следствие и во всей серии) должны удовлетворять соотношению

$$|i{:}\{type(e_i) = 0\}| = |i{:}\{type(e_i) = 1\}|. \tag{4}$$

Примечание — Если испытания не будут равномерно распределены между наличием (t_i = 1) и отсутствием (t_i = 0) сценария ситуации, то одна из ошибок может быть искусственно занижена или завышена, что приведет к неправильной оценке значений функциональных характеристик.

Равенство количества элементов исходных данных (равномощность) в группах серии ($|g_1| = |g_2| = \ldots = |g_n|$) важно для предотвращения смещения в оценке ошибок.

Допускается формирование групп разного объема (разномощных) в рамках одной серии испытаний при соблюдении принципа пропорционального включения их элементов в общее количество испытаний серии для обеспечения наилучшей репрезентативности всей выборки.

7.2.3 Равное количество испытаний для каждого типа исходных данных позволяет использовать статистические методы для оценки параметров модели, таких как доверительные интервалы для α- и β-ошибок.

Метод доверительных интервалов для α- и β-ошибок позволяет оценить диапазон значений, в котором находится истинное значение параметра с определенной вероятностью.

Для получения достоверного значения статистических показателей необходимо провести статистически значимое количество испытаний. Этот процесс включает в себя несколько этапов:

- определение критерия точности E (например, ±2 %);
- определение критерия достоверности Z (например, для уровня достоверности:
- 90 %: Z = 1,645;
- 95 %: Z = 1,96;
- 98 %: Z = 2,326;
- 99 %: Z = 2,576);
- определение предварительного значения доли успеха p (например, p = 0,5 для случаев, когда точное значение p неизвестно);
- расчет размера выборки K рассчитывается по формуле Кокрана

$$K = \frac{Z^2 \cdot p \cdot (1-p)}{E^2}. \quad (5)$$

Формула (5) применяется для оценки объема выборки при анализе бинарных признаков в случае достаточно большой выборки, без учета размера генеральной совокупности, основана на нормальном приближении биномиального распределения и применима при выполнении условий:

$$n \cdot p \geq 5, \quad (6)$$

$$n \cdot (1 - p) \geq 5, \quad (7)$$

где n — размер выборки;

p — значение доли успехов.

Примечание — На практике под большой выборкой для формулы (5) обычно понимается объем выборки не менее 100 наблюдений.

7.2.4 Предполагают, что по предварительным данным чувствительность объекта испытаний составляет 95 % (0,95). Желаемая точность оценки — ±2 % (0,02) при уровне доверия 95 % (Z = 1,96). Размер выборки (количества испытаний) K рассчитывают по формуле

$$K = \frac{(1{,}96)^2 \cdot 0{,}95 \cdot (1-0{,}95)}{(0{,}02)^2} = 456{,}19. \quad (8)$$

Расчет показывает, что необходимо провести минимум 457 испытаний, чтобы достигнуть желаемой точности в этом примере.

7.2.5 Когда предварительная оценка чувствительности объекта испытаний может быть неизвестна (p = 0,5), рассчитаны размеры выборок для некоторых значений критериев точности и достоверности. Результаты расчетов приведены в таблице 1.

Таблица 1 — Размеры выборок при $p = 0,5$

Уровень достоверности	Значение Z	Критерий точности E			
		±1 %	±2 %	±5 %	±10 %
99 %	2,576	16 589	4147	664	166
98 %	2,326	13 526	3381	541	135
95 %	1,960	9604	2401	384	96
90 %	1,645	6765	1691	271	68

Если предварительная оценка доли успехов объекта испытаний оценивается на уровне 95 % ($p = 0,95$), то результаты расчетов необходимых размеров выборок для некоторых сочетаний критериев точности и достоверности приведены в таблице 2.

Таблица 2 — Размеры выборок при $p = 0,95$

Уровень достоверности	Значение Z	Критерий точности E			
		±1 %	±2 %	±5 %	±10 %
99 %	2,576	3152	788	126	32
98 %	2,326	2570	642	103	26
95 %	1,960	1825	456	73	18
90 %	1,645	1285	321	51	13

7.2.6 При ограниченной мощности генеральной совокупности рекомендуется использовать коррекцию Кокрана, рассчитываемую по формуле

$$n = \frac{K}{1 + \frac{K - 1}{N}}, \tag{9}$$

где K — расчетный размер выборки по формуле (5);

N — объем генеральной совокупности.

8 Методология проведения испытаний

8.1 Этапы испытаний

8.1.1 Проведение испытаний систем ситуационной видеоаналитики требует тщательного и поэтапного подхода, обусловленного необходимостью объективной и комплексной оценки значений функциональных характеристик.

Разделение испытаний на этапы позволяет:

- структурировать процесс проведения испытаний;
- минимизировать риски ошибок;
- обеспечить систематическую проверку;
- выявлять и устранять проблемы на ранних стадиях;
- предотвратить накопление ошибок.

8.1.2 Поэтапное проведение испытаний является неотъемлемой частью методологии тестирования систем ситуационной видеоаналитики, обеспечивающей высокую степень достоверности результатов.

Общая этапность проведения испытаний должна быть строго соблюдена и включать:

- подготовительный этап;

- проведение серии испытаний;
- анализ и интерпретацию результатов;
- оформление результатов испытаний.

Состав и объем конкретных работ, выполняемых на каждом этапе, могут варьироваться в зависимости от специфики объекта, условий испытаний и целей исследования.

8.1.3 Целью подготовительного этапа является подготовка объекта испытаний и условий проведения испытаний.

К задачам относятся определение типа объекта испытаний, его назначения и условий испытания, установление вида испытаний и перечня функциональных характеристик, подготовка исходных данных, подготовка программы испытаний, настройка оборудования и программного обеспечения, проверка их работоспособности, сбор необходимой документации.

8.1.4 Целью этапа проведения серии испытаний является оценка значений функциональных характеристик объекта испытаний.

К задачам относятся выполнение испытаний по программе испытаний, регистрация результатов, выявление и устранение возможных отклонений в работе объекта испытаний.

8.1.5 Целью этапа анализа и интерпретации результатов являются анализ данных, полученных в ходе испытаний, и оценка их соответствия заявленным значениям функциональных характеристик.

К задачам относятся сравнение результатов испытаний с эталонными значениями, расчет метрик, формирование заключений по группам.

8.1.6 Целью этапа оформления результатов испытаний является подготовка итогового отчета с результатами испытаний.

К задачам относятся оформление протоколов испытаний и подготовка аналитического отчета.

8.2 Интерпретация результатов испытаний

8.2.1 Для расчета ключевых статистических метрик (чувствительность, специфичность, точность, доля правильных исходов и другие статистические показатели) важно однозначно определить критерии интерпретации исходов испытаний:

- исход TN;
- исход FP;
- исход TP;
- исход FN.

8.2.2 Реидентификация сценария ситуации (по аналогии с реидентификацией объекта, когда один и тот же сценарий ситуации зарегистрирован несколько раз), должна интерпретироваться как истинно положительный исход для статистических метрик согласно таблице 3.

При вычислении нестатистических метрик [«Отклонение кадра начала» (SFA) и «Кадровый сдвиг обнаружения» (DFS)] — учитываются только факт и результаты первого обнаружения сценария ситуации.

8.2.3 Результаты испытаний в зависимости от того, присутствует ли сценарий ситуации в контрольной части данных, с классификацией типов исходов приведены в таблице 3.

Таблица 3 — Результаты испытаний

Тип элемента исходных данных	Сценарий ситуации в контрольной части исходных данных	Действие объекта испытаний	Тип исхода	Результат
Тип 0	Отсутствие	Не зарегистрировал требуемый сценарий ситуации	Исход TN	Испытание пройдено успешно
Тип 0	Отсутствие	Зарегистрировал один или более требуемых сценариев один или более раз	Исход FP	Испытание не пройдено
Тип 1	Наличие	Зарегистрировал требуемый сценарий один или более раз	Исход TP	Испытание пройдено успешно

Окончание таблицы 3

Тип элемента исходных данных	Сценарий ситуации в контрольной части исходных данных	Действие объекта испытаний	Тип исхода	Результат
Тип 1	Наличие	Не зарегистрировал требуемый сценарий или зарегистрировал более одного требуемого сценария	Исход FN	Испытание не пройдено

8.3 Биномиальное распределение в группах

8.3.1 При выполнении требований настоящего стандарта для описания распределения результатов испытаний в серии по группам допускается использовать модель биномиального распределения, т. к. выполняются условия:

- количество испытаний в серии фиксировано;
- для каждого испытания существуют только два возможных результата (см. таблицу 3);
- все испытания независимы;
- вероятность успеха постоянная.

8.3.2 Биномиальная модель описывает распределение статистических метрик в группах серии испытаний. Метрика, например чувствительность, может рассматриваться как вероятность успеха (выявления истинно положительного исхода) среди всех положительных случаев.

На рисунке 5 представлен пример распределения результатов определения доли успехов (100 %, 95 %, 90 % и 80 %) в 11 группах.

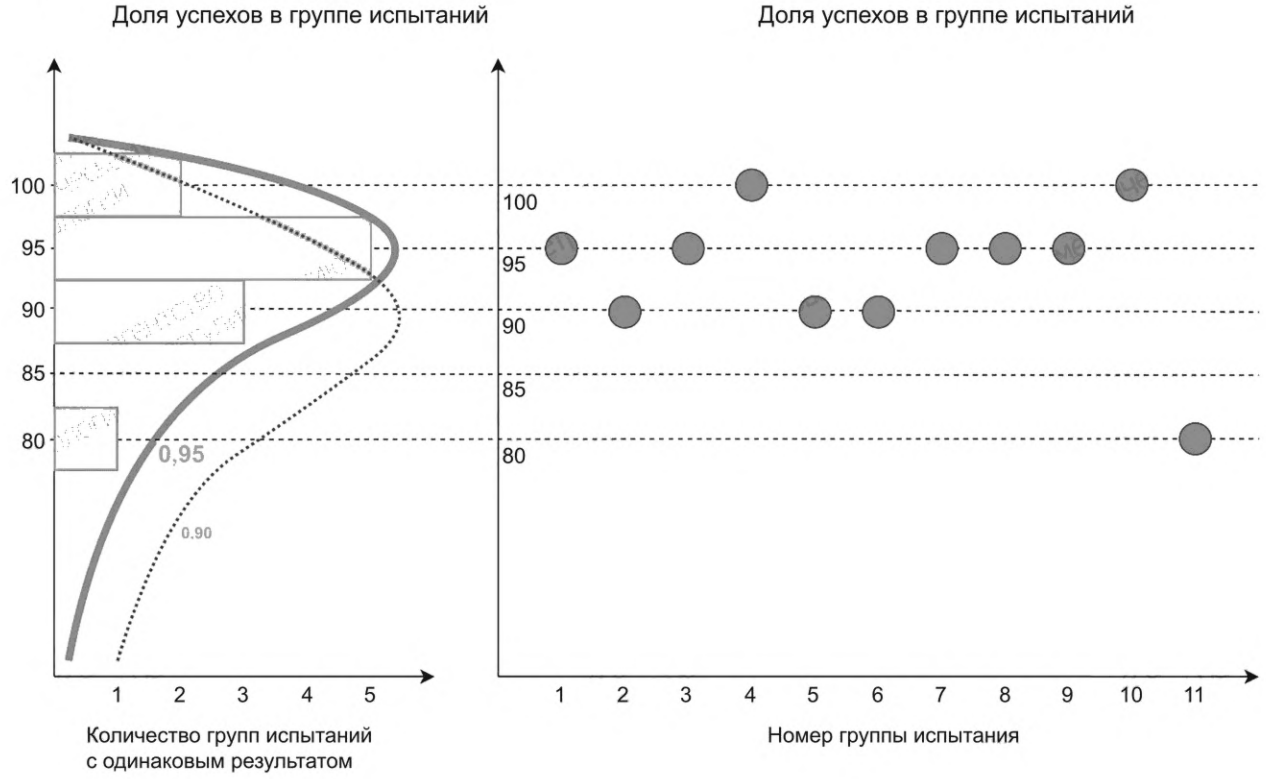


Рисунок 5 — Пример биномиального распределения метрики

8.3.3 Рассмотрение биномиального распределения результатов испытаний серии показывает оценку распределения ошибок по группам испытаний и отражает стабильность объекта испытаний в различных контекстах.

В зависимости от требуемой глубины анализа в протоколе испытаний испытатель может отразить:

- анализ распределения ошибок.

Примечание — Если ошибки распределяются неравномерно, это может указывать на систематические проблемы или потенциальные уязвимости в работе алгоритма объекта испытаний, которые не будут видны при анализе только итоговых (интегральных) значений метрик;

- оценку стабильности объекта.

Примечание — Если результаты значительно варьируются между группами, это может свидетельствовать о том, что объект испытаний чувствителен к определенным признакам, которые не учитываются в общих статистических показателях;

- идентификацию систематических отклонений.

Примечание — Использование биномиального распределения помогает валидировать методику испытаний. Если результаты соответствуют ожидаемому распределению, это подтверждает корректность методологии и отсутствие систематических ошибок.

9 Методики определения и расчета метрик

9.1 Нестатистические метрики

9.1.1 Нестатистические метрики рассчитываются только для исходных данных типа 1, то есть для тех, где присутствует сценарий ситуации, и только при условии исхода TP.

В целях уменьшения влияния выбросов в результатах испытаний в качестве среднего значения для серии и групп испытаний следует рассчитывать и приводить в протоколах медианные значения Me нестатистических метрик:

- Me_{sfa} — медиана отклонения кадра начала;
- Me_{dfs} — медиана кадрового сдвига обнаружения;
- Me_J — медиана индекса Жаккара.

Для нечетного количества отсортированных по возрастанию значений медианой считается среднее значение в отсортированном ряду, а для четного — среднее арифметическое двух центральных значений.

9.1.2 Метрика SFA «Отклонение кадра начала»:

- определяется в кадрах;
- показывает, на сколько кадров детектируемый объектом кадр начала сценария ситуации отличается от эталонного кадра начала;
- вычисляется по формуле

$$SFA = |D - E|, \quad (10)$$

где E — номер эталонного кадра начала сценария ситуации, определенный в референсных данных;

D — номер детектируемого кадра начала сценария ситуации, определенного объектом испытаний.

Для одного испытания типа 1 значение метрики «Отклонение кадра начала» вычисляется по модулю, т. к. важна только величина отклонения. Чем меньше разница между детектируемым и эталонным кадрами, тем точнее работает объект испытаний.

Примечание — Например, если эталонный кадр начала сценария ситуации равен 150, а детектируемый — 155, то ошибка будет 5 кадров, что означает отклонение в 5 кадров относительно эталона.

9.1.3 Метрика DFS «Кадровый сдвиг обнаружения»:

- определяется в кадрах;
- показывает, какое количество кадров, начиная с эталонного кадра E, требуется объекту испытаний для детектирования начала сценария ситуации;
- вычисляется по формуле

$$DFS = |S - E|, \quad (11)$$

где E — номер эталонного кадра начала сценария ситуации, определенный в референсных данных;

S — номер кадра, переданного на обработку объекту испытаний, после обработки которого объектом было зафиксировано наличие сценария ситуации в поданных на обработку кадрах.

Для одного испытания типа 1 значение метрики «Кадровый сдвиг обнаружения» вычисляется по модулю, т. к. важна только величина отклонения и не учитывается возможность предиктивного обнаружения события. Чем меньше разница между номером эталонного кадра начала сценария и номером кадра, по результатам обработки которого объектом испытаний зафиксировано начало сценария, тем с меньшей задержкой реагирует объект испытаний.

Примечание — Видеоанализ — процесс, осуществляемый на уровне отдельных кадров и логично определяемый именно в кадрах, а не в миллисекундах или других временных единицах, а само определение значения в кадрах дает более точное и универсальное представление о задержке, так как интервал между кадрами фиксирован, известен и позволяет легко адаптировать метрику к видео с разной частотой кадров.

9.1.4 Метрика J «Индекс Жаккара» — это коэффициент перекрытия рамок (контуров), оценивающий точность выделения группы пикселей в кадре, который рассчитывают как отношение площади пересечения эталонной и детектируемой рамок (контуров) к их общей площади по формуле

$$J = \frac{|A \cap B|}{|A \cup B|}, \quad (12)$$

где A и B — два множества (детектируемая область и эталонная область);

$A \cap B$ — размер пересечения множеств A и B, общая часть, которая принадлежит обоим множествам;

$A \cup B$ — размер объединения множеств A и B (все уникальные элементы из обоих множеств).

Для одного испытания типа 1 значение метрики J «Индекс Жаккара» вычисляется как отношение двух наборов пикселей в кадре, то есть доля пересекающихся пикселей относительно общего числа пикселей в двух областях. Индекс Жаккара принимает значения от 0 до 1,

- где 0 — это случай, когда множества не пересекаются;
- 1 — это случай, когда множества полностью совпадают (см. рисунок 6).

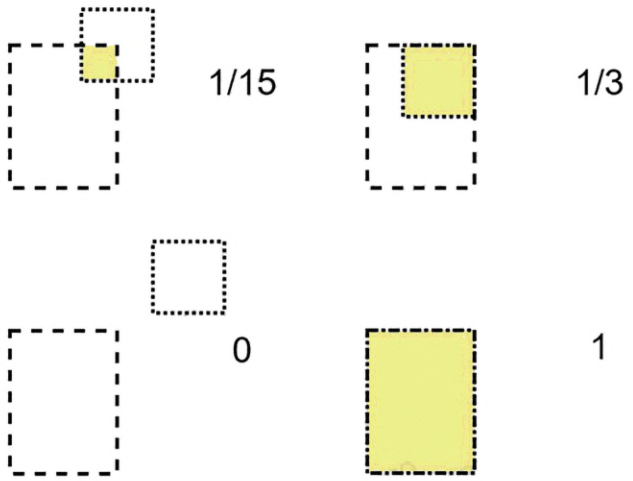


Рисунок 6 — Пример пересечения рамок

9.2 Статистические метрики

9.2.1 На основании выполнения требований настоящего стандарта, в том числе интерпретации результатов испытаний по 8.2, может быть рассчитан набор статистических метрик для оценки модели, который позволяет всесторонне оценить производительность объекта испытаний.

Далее (согласно 9.2.2—9.2.5) приведены формулы и описания основных статистических метрик, которые широко используются в машинном обучении для оценки качества классификаторов. Они являются базовыми, так как дают представление о том, насколько хорошо модель справляется с предсказаниями, и используются для построения более сложных показателей, таких как F1-мера и др.

Для объективной оценки качества работы объекта испытаний необходимо рассчитать как минимум статистические показатели «Чувствительности» (recall) и «Точности» (precision), гарантирующие в совокупности сбалансированную оценку между ложными обнаружениями событий и пропущенными событиями.

9.2.2 Метрика Se «Чувствительность» (полнота, истинно положительная пропорция) показывает оценку вероятности того, что сценарий ситуации в элементе исходных данных испытаний с наличием этого сценария (тип 1) будет детектирован, то есть насколько хорошо объект испытаний находит истинные положительные случаи среди всех реальных положительных. Вычисляется по формуле

$$Se = \frac{TP}{TP + FN}. \quad (13)$$

9.2.3 Метрика Sp «Специфичность» (избирательность, истинно отрицательная пропорция) показывает оценку вероятности того, что сценарий ситуации в элементе исходных данных испытаний с отсутствием этого сценария (тип 0) не будет детектирован, то есть способность объекта испытаний правильно идентифицировать отрицательные случаи среди всех реальных отрицательных. Вычисляется по формуле

$$Sp = \frac{TN}{TN + FP}. \quad (14)$$

Примечание — В различных нормативных документах могут встречаться различные определения эксплуатационных характеристик алгоритмов и аппаратно-программных средств. Например, в [1] термин «специфичность» определен как характеристика, отражающая долю истинно положительных обнаружений событий среди всех обнаружений. Однако, согласно общепринятым международным стандартам и терминам, данное определение скорее соответствует термину «Точность», который показывает, насколько велика доля истинно положительных результатов среди всех предсказанных положительных результатов.

9.2.4 Метрика Pr «Точность» показывает оценку вероятности того, что сценарий ситуации для испытаний с наличием этого сценария (тип 1) и для испытаний с отсутствием этого сценария (тип 0) будет детектирован, то есть как много из обнаруженных положительных исходов действительно положительные или насколько точно объект испытаний предсказывает положительные результаты. Вычисляется по формуле

$$Pr = \frac{TP}{TP + FP}. \quad (15)$$

9.2.5 Метрика Ac «Доля правильных исходов» (accuracy) показывает оценку вероятности того, что сценарий ситуации или отсутствие этого сценария для испытаний с наличием этого сценария (тип 1) и для испытаний с отсутствием этого сценария (тип 2) будет детектирован или не детектирован соответственно, то есть оценивает общую долю правильно классифицированных случаев (как положительных, так и отрицательных) относительно общего числа исходов. Вычисляется по формуле

$$Ac = \frac{TP + TN}{TP + TN + FP + FN}. \quad (16)$$

10 Документирование и отчетность

10.1 Общие требования

10.1.1 Документирование результатов испытаний должно обеспечивать запись и сохранение всех необходимых данных для воспроизведения и оценки проведенных испытаний.

10.1.2 Оформление протоколов должно выполняться в соответствии с требованиями ГОСТ Р 58973.

10.1.3 Условия испытаний, включая используемое оборудование и программное обеспечение, должны быть задокументированы в протоколе результатов испытаний.

10.2 Структура и содержание отчета

В отчет о проведении испытаний обычно рекомендуется включать следующие разделы:

- титульный лист — с указанием наименования испытательной лаборатории, наименования объекта испытаний, наименования организации-разработчика, даты проведения испытаний и регистрационного номера отчета;
- предисловие — с описанием цели и основания для проведения испытаний, ссылки на договор, заявку или поручение, на сведения о заказчике, разработчике и исполнителе испытаний;
- содержание — с перечнем всех разделов, подразделов и приложений отчета с указанием соответствующих страниц;
- введение — с кратким обоснованием необходимости проведения испытаний, характеристикой объекта и общим описанием условий и порядка выполнения работ;
- наименование — с полным наименованием отчета, включающим вид испытаний, наименование объекта и иные уточняющие сведения (например, модификации или версии);
- область применения — с указанием, в каком контексте, условиях или сферах могут использоваться результаты испытаний, включая ограничения применимости;
- нормативные ссылки — с перечнем нормативных документов, методик и стандартов, применявшихся при проведении испытаний;
- термины и определения — с определениями терминов, используемых в отчете, для обеспечения единообразного понимания результатов испытаний;
- обозначения и сокращения — с расшифровкой всех используемых сокращений, символов, переменных и иных обозначений, применяемых в тексте отчета;
- основные нормативные положения — с изложением примененных подходов, методик, критериев оценки, требований к испытаниям и методологических основ проведения работ;
- приложения — с подробными данными об испытаниях, расчетных метриках, их значениях в табличном виде (см. приложение А) и с дополнительными материалами (например, графики, диаграммы, копии актов, протоколы, фотоматериалы и иные подтверждающие данные);
- библиография — с перечнем использованных нормативных документов, научных источников, методических материалов, на которые даны ссылки в тексте отчета;
- библиографические данные — с указанием сведений об отчете: полного наименования, номера, даты составления, наименования — и реквизитов исполнителя, фамилий и подписей ответственных лиц.

Приложение А
(справочное)

Форма представления результатов испытаний

Таблица А.1

Параметр	Значение
Наименование серии	«Нетипичные изменения в сцене»
Шифр серии	S-XXXXXX
Количество испытаний в серии	1826
Количество групп в серии	18
Уровень доверия	95,00 %
Погрешность	±1,00 %
Предварительная оценка (p = 0,95)	95,00 %

Таблица А.2

Группа	Количество	Me_{sfa}	Me_{dfs}	Me_j	TP	FN	TN	FP	Se	Sp	Pr	Ac
g1	100	10	25	1	48	2	45	5	0,9600	0,9000	0,9057	0,9300
g2	100	11	25	1	50	0	50	0	1,0000	1,0000	1,0000	1,0000
g3	100	12	25	1	49	1	48	2	0,9800	0,9600	0,9608	0,9700
…	…	…	…	…	…	…	…	…	…	…	…	…
g8	100	10	25	1	46	4	49	1	0,9200	0,9800	0,9787	0,9500
g9	100	9	25	1	47	3	48	2	0,9400	0,9600	0,9592	0,9500
g10	100	9	25	1	49	1	48	2	0,9800	0,9600	0,9608	0,9700
…	…	…	…	…	…	…	…	…	…	…	…	…
g18	126	13	25	1	62	1	62	1	0,9841	0,9841	0,9841	0,9841
Итого	1826	11	25	1	889	24	883	30	0,9737	0,9671	0,9674	0,9704

Таблица А.3

Испытываемая группа	Наименование элемента исходных данных	Общее количество кадров	Длительность (при 25 к/с)	Исход
0001 g1	s1-g1-001-0.avi	2250	00:01:30	TP
0002 g1	s1-g1-001-1.avi	2250	00:01:30	TP
0003 g1	s1-g1-002-0.avi	1975	00:01:19	TP
…	…	…	…	…
0894 g8	s1-g8-094-1.avi	2550	00:01:42	FP
…	…	…	…	…
1826 g18	s1-g18-126-1.avi	1950	00:01:18	TP

Библиография

[1] Постановление Правительства Российской Федерации от 26 сентября 2016 г. № 969 «Об утверждении требований к функциональным свойствам технических средств обеспечения транспортной безопасности и Правил обязательной сертификации технических средств обеспечения транспортной безопасности»

УДК 004.89:006.354 ОКС 35.240.01

Ключевые слова: информационная технология, искусственный интеллект, ситуационная видеоаналитика, методология, функциональные характеристики

Редактор *М.В. Митрофанова*
Технический редактор *В.Н. Прусакова*
Корректор *И.А. Королева*
Компьютерная верстка *А.Н. Золотаревой*

Сдано в набор 19.02.2026. Подписано в печать 05.03.2026. Формат 60×84⅛. Гарнитура Ариал.
Усл. печ. л. 3,26. Уч.-изд. л. 2,64.

Подготовлено на основе электронной версии, предоставленной разработчиком стандарта

Создано в единичном исполнении в ФГБУ «Институт стандартизации»
для комплектования Федерального информационного фонда стандартов, 117418 Москва, Нахимовский пр-т, д. 31, к. 2.
gost.e-ecolog.ru gost.e-ecolog.ru