УДК 004.89, 681.5
DOI:10.25729/ESI.2026.42.2.004

Анализ битемпоральных изображений системой управления коллаборативным роботом-манипулятором для определения вновь появившихся объектов в поле подсистемы технического зрения

Калушев Константин Александрович, Макаров Илья Андреевич

Национальный исследовательский ядерный университет «МИФИ»,
Россия, Москва, *konstantin.kalushev@gmail.com*

Аннотация. Одной из задач, относящихся к созданию интерактивного коллаборативного робота-манипулятора, является темпоральный анализ рабочей сцены, то есть определение порядка появления (выбытия) объектов в поле технического зрения. Традиционно, данный вопрос рассматривался применительно к спутниковым снимкам и недостаточно прорабатывался в литературе в отношении сцен, находящихся на расстоянии около 1 м от камеры. Вместе с тем, анализ рабочей сцены на основе битемпоральных изображений является актуальной областью исследований в контексте развития робототехники в целом и физического искусственного интеллекта в частности. Создание качественных темпоральных масок изменения рабочей сцены позволяет определить контуры и геометрические центры новых объектов для последующего захвата роботом-манипулятором. Качественная темпоральная маска не должна иметь ложно определенных областей изменений (объектов изменений, которых на самом деле не существует), но при этом позволять четко обрисовывать контуры истинных объектов изменений рабочей сцены. В статье осуществлена математическая постановка задачи темпорального анализа и на ее базе предложен метод создания темпоральных масок областей изменений на основе дифференцирования изображений «до» и «после», комбинирующий классические методы технического зрения и нейросетевую сегментационную модель SAM (Segment Anything Model). Новизна предлагаемого подхода заключается в применении к дифференцированному изображению не алгебраической обработки, а его сегментация на две области (область изменений и область без изменений) с использованием нейросетевой сегментационной модели. Предложенный подход сопоставлялся с алгебраическими методами создания темпоральных масок (Change Vector Analysis – CVA и Slow Feature Analysis – SFA) и использованием нейросетевой архитектуры многослойного перцептрона (внешний слой из 12 нейронов, внутренний слой из 512 нейронов, наружный слой из 1 нейрона). Продемонстрировано, что предложенный подход позволяет генерировать качественные маски изменений для разнообразных объектов на большом количестве фонов (включая пестрые), чего сложно добиться приведенными для сопоставления методами. Вместе с тем, предложенный подход может быть реализован «на лету», то есть в реальном времени работы оператора-робота, только при наличии графического ускорителя (Graphics Processing Unit – GPU).

Ключевые слова: коллаборативный робот, битемпоральные изображения, SAM, бинарные маски изменений

Цитирование: Калушев К.А. Анализ битемпоральных изображений системой управления коллаборативным роботом-манипулятором для определения вновь появившихся объектов в поле подсистемы технического зрения / К.А. Калушев, И.А. Макаров // Информационные и математические технологии в науке и управлении, 2026. – № 2(42). – С. 42-54. – DOI:10.25729/ESI.2026.42.2.004.

Введение. Согласно данным International Federation of Robotics, промышленные коллаборативные роботы набирают все большую популярность и увеличивают присутствие на рынке, в частности, количество новых установок таких роботов в 2024 году по сравнению с предыдущим годом выросло на 12.9%, а доля коллаборативных роботов в общем количестве установленных роботов увеличилась с 2.8% в 2017 году до 11.9% в 2024 году [1]. Традиционно, системы управления коллаборативными роботами-манипуляторами построены на принципе «от точки к точке», то есть требуется прямо программировать все движения робота, что ограничивает сферу использования, существенно расширить которую возможно за счет интеллектуализации систем управления.

Эволюция коллаборативной робототехники привела к появлению понятия «интерактивный коллаборативный робот» [2], под которым подразумевается машина, обладающая элементами искусственного интеллекта и развитой информационно-сенсорной

системой, что позволяет выступать партнером человека при выполнении самых различных задач. Ключевая особенность системы управления интерактивным коллаборативным роботом – максимально полное восприятие окружающей среды, что, в конечном итоге, позволит качественно планировать действия машины в ответ на команды оператора.

Восприятие внешней среды достигается, прежде всего, через систему технического зрения, которая решает две ключевые задачи – определение объектов интереса (англ. – object detection) [3, 4] и сегментацию (англ. – segmentation) рабочей области [5, 6] с целью планирования последующих манипуляционных операций. К дополнительным задачам может относиться определение глубины изображения для расширенного понимания рабочей сцены (например, используя SwiftDepth++ [7], представляющую собой легкую модель оценки глубины, которая обеспечивает конкурентоспособные результаты при сохранении низкого уровня вычислительных затрат).

Несмотря на общий прогресс технического зрения в робототехнике, на сегодня недостаточно проработанной остается задача темпорального анализа рабочей сцены, цель которого – достижение понимания системой управления коллаборативным роботом порядка появления (выбытия) объектов на рабочей сцене. На практике темпоральное восприятие рабочей сцены может помочь оператору формулировать команды с компонентом временной последовательности, такие, как «возьми последнее упавшее яблоко», «убери со стола все предметы, кроме последнего».

В настоящей статье предложен подход к определению новых объектов в поле технического зрения, особенностью которого выступает комбинация классических методов машинного зрения (вычитание изображений) и нейросетевых инструментов для сегментации и создания темпоральных масок. Специфика решения задачи определения изменений на изображении рабочей области коллаборативного робота-манипулятора – повышенные требования к производительности (необходимость работы «на лету», с соответствующим показателем FPS) и к качеству масок новых объектов, которые могут использоваться для выработки стратегии их захвата.

1. Обзор литературы. В решении задачи определения изменений на изображениях можно выделить два принципиальных подхода [8]: отслеживание попиксельных изменений (включая алгебраические, трансформационные и классификационные методы) и отслеживание изменений объектов на изображении (например, посредством кластеризации изображений на такие объекты).

Вопросы анализа изменений на изображениях получили наиболее глубокое рассмотрение применительно к анализу изображений удаленных объектов (remote sense detection), прежде всего анализу спутниковых снимков. Классическим алгебраическим методом выступает анализ вектора изменений (англ. – Change Vector Analysis – CVA), суть которого сводится к определению какой-либо метрики расстояния (например, евклидового расстояния) между двумя корреспондирующими пикселями изображений «до» и «после» с последующем вычислением маски на ее основе для изображения в целом (например, как это сделано в работах [9, 10]). Применение нашла также адаптация подхода анализа наименее изменяемых компонентов (англ. - Slow Feature Analysis – SFA) для целей анализа удаленных объектов (например, как в [11]), суть которого сводится к определению набора функций, которые извлекают наименее изменяющиеся признаки многомерного входного сигнала, такого, как фотографическое изображение. Существует множество реализаций SFA для различных применений [12], в том числе на основании фреймворка Python MDP [13].

Анализ изменений на изображениях нашел свое развитие в применении нейросетевых методов. Например, в работе [14] предложена архитектура IDJANet, интегрирующая механизмы деформируемого внутреннего внимания (англ. – deformable self-attention) и кросс-

внимания (англ. – cross-attention), а в работе [15] – модель HFNet для извлечения признаков темпоральных изменений, использующая двухкомпонентный энкодер для связанных и дифференцирующих признаков.

Реже в литературе рассматриваются методы определения изменений на изображениях «открытого мира». Например, в статье [16] предлагается метод определения изменений в изображениях, полученных с камеры мобильного робота-патрульного. В статье [17] рассматривался метод определения изменений при одновременном осуществлении семантической сегментации. Метод базируется на архитектуре сверточной нейронной сети U-net, используя один энкодер и два декодера для осуществления одновременной сегментации снимков «до» и «после» и генерации маски для выявления изменений на изображениях.

Вопросы определения изменений на 3D-сцене, сопоставимой с рабочей областью коллаборативного робота-манипулятора, рассматривались в статье [18], в которой предложен метод SemanticDifference, основанный на 4D гауссовом сплэтинге для 3D-представления сцен «до» и «после» на основе множества снимков рабочей сцены. Предложено осуществлять рендеринг изображения «после» из точки, где располагалась камера «до», и после этого сопоставлять изображения на базе выделения потенциальных сегментов, в которых могли произойти изменения. Принципиальным отличием от предлагаемого в настоящей статье подхода выступает построение полной 3D-модели рабочей области на основании нескольких снимков, что требует соответствующих вычислительных ресурсов и не подходит для работы «на лету».

Методы темпорального анализа рабочей сцены робота-манипулятора к настоящему моменту широкого отражения в литературе не нашли.

2. Постановка задачи. Пусть имеются два битемпоральных изображения $I_1 \subseteq R^{3 \cdot H \cdot W}$ и $I_2 \subseteq R^{3 \cdot H \cdot W}$, где 3 представляет собой количество каналов, H – высоту изображения в пикселях, W – ширину изображения в пикселях. При этом пусть $C_1 = \{c_1, c_2, \dots c_N\}$ представляет собой семантическое множество объектов на I_1, а $C_2 = \{c_1, c_2, \dots c_M\}$ – на I_2.

Под совокупностью новых объектов на I_1 будет пониматься подмножество $\mathrm{C}' = \mathrm{C}_2/\mathrm{C}_1$, , то есть подмножество, состоящее из элементов:

$$C' = \{c_{N+1}, \dots c_M\} \quad (1)$$

Задача определения маски новых объектов на I_2 состоит в определении:

$$M(x, y) \in \{0, 1\}^{1 \cdot H \cdot W}, \quad (2)$$

где значение 1 присваивается всем пикселям (x, y), таким, что

$$\mathrm{I}_2(\mathrm{x,y}) \in \mathrm{C}'. \quad (3)$$

Пикселям, не удовлетворяющим данному условию, присваивается значение 0.

Ключевыми особенностями задачи определения новых объектов в поле технического зрения робота-манипулятора выступают:

1. Необходимость максимально геометрически точного определения масок объектов для целей последующего формирования стратегии захвата манипулятором (gripping strategy planning).

2. Расположение камеры технического зрения стационарно (вибрации также отсутствуют). Камера технического зрения может располагаться в фиксированном (над рабочей сценой) либо в квази-фиксированном положении (непосредственно рядом со схватом робота на конечном звене, в позиции eye-in-hand).

3. Расстояние от объектива камеры до основной рабочей поверхности фона составляет около 1 метра.

4. Отсутствие резких перепадов освещения между кадрами (например, сценариев «день-ночь», включение-выключение точечного или фонового источника освещения).

5. Возможность использовать только коммерчески доступные ЭВМ, целесообразные с экономической точки зрения (например, если речь идет об одноплатном компьютере, то не производительнее / не дороже NVIDIA Jetson Orin Nano).

3. Методология. Предлагается совместить классические методы компьютерного зрения (для предобработки, формирования дифференциального изображения алгебраическим методом дифференцирования изображений, и постобработки) с последующей сегментацией с использованием нейросетевых моделей.

Подход, предлагаемый к реализации автором, описан на рис. 1 и состоит из следующих последовательных шагов. Прежде всего, осуществляется предобработка I_1 и I_2 с целью выравнивания их взаимной яркости и контрастности изображений. Предобработка осуществляется с помощью тривиальных методов, в связи с чем в настоящей статье не описывается.

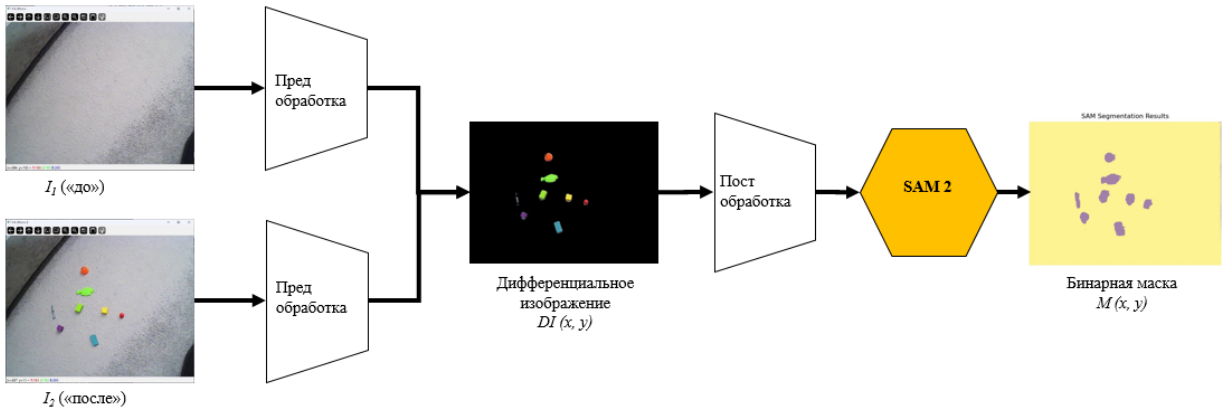


Рис. 1. Структурная схема предложенного подхода к определению изменений на изображениях

Обработанные I_1 и I_2 используются для формирования дифференциального изображения $DI(x,y)$ следующим образом. Определяются расстояния между корреспондирующими пикселями I_1 и I_2. В общем виде матрицу расстояний можно представить, как:

$$D(x,y) = \|I_1 - I_2\|, \tag{4}$$

где $\|I_1 - I_2\|$ представляет собой оценку расстояния между корреспондирующими пикселями двух изображений с помощью одной из метрик, таких, как евклидово расстояние, манхэттенское расстояние или косинусная близость.

Далее непосредственно формируется дифференциальное изображение рабочей области на основании подхода с применением порогового значения τ

$$DI(x,y) = \begin{cases} I_2(x,y), \text{если } D(x,y) > \tau \\ 0, \text{в иных случаях} \end{cases} \tag{5}$$

В связи с наличием на изображении дефектов, снижающих качество будущей бинарной маски, проводится постобработка. В частности, устраняются «внутренние» пустые пиксели в областях расположения семантических объектов из C' (см. рис. 2). Для этого используется одна из реализаций метода «наводнения» (англ. - flood fill).

На рисунке 2 слева представлено изображение до исключения внутренних областей, справа – после. Полученное после постобработки улучшенное дифференциальное изображение $DI'(x,y)$ используется в качестве входящего для нейросетевой сегментационной модели. Сегментационная модель (Segment Anything Model – SAM) [19] представляет собой базовую модель (англ. – foundation model), которая впервые была опубликована в 2023 году, а также соответствующий набор данных из более чем 1 млрд. масок для более чем 11 млн. изображений.

Рис. 2. Результат исключения внутренних областей

Как и другие базовые модели в области ИИ, SAM способна выходить за рамки тренировочного набора данных и осуществлять сегментацию произвольных изображений без предварительного обучения (англ. – zero-shot). В 2024 году была представлена модель SAM-2 [20], отличающаяся как повышенной производительностью при сегментации изображений (заявляется увеличение FPS в 6 раз), так и возможностью сегментации видео. В настоящей работе мы ссылаемся на обе модели, как на SAM.В задачах определения новых объектов в поле технического зрения SAM формирует бинарную маску $M(x, y)$ для областей изменения изображений $DI(x, y)$ (то есть областей, в которых значения пикселей на I_1 и I_2 отличаются от порогового значения).

4. Эксперимент. Далее представлено описание экспериментов, проведенных с использованием предложенного подхода к определению новых объектов в поле технического зрения коллаборативного робота-манипулятора. Основной эксперимент сводился к экспертной оценке качественных результатов формирования масок расположения новых объектов.

4.1. Технология экспериментов.

Оценка эффективности предложенного подхода осуществлялась на ЭВМ в трех конфигурациях, описанных в таблице 1. Использование нескольких конфигураций обусловлено спецификой задачи – определением подхода, который может практически использоваться для системы управления роботом-манипулятором, вычислительные ресурсы которой ограничены, а определение масок новых объектов должно осуществляться «на лету».

Таблица 1. Конфигурации оборудования для тестирования

Конфигурация	CPU	RAM	GPU
1	11th Gen Intel Core i3-1115G4 (3.00 GHz)	8,0 Гб	Отсутствует (встроенная видеокарта Intel UHD Graphics)
2	13th Gen Intel Core i7-13700H (2.40 GHz)	16,0 Гб	Отключен (встроенная Intel Iris Xe Graphics)
3	13th Gen Intel Core i7-13700H (2.40 GHz)	16,0 Гб	NVIDIA GeForce RTX 4070 Laptop GPU (8,0 Гб)

Для осуществления экспериментов был написан код на языке Python. Код, относящийся к сегментационной модели SAM, использовал соответствующую библиотеку от компании Ultralytics. Использовались веса модели от SAM-2 [20] различных типов.

Нами не найдено опубликованных наборов данных с битемпоральными масками изменения рабочей сцены робота-манипулятора (то есть для закрытой стационарной сцены с расположением камеры на высоте около 1 м над ней). Из ближайших аналогов следует

отметить набор данных Panoramic Change Detection (PCD) [21] для выявления изменений в открытой среде (ландшафты после цунами в Японии). В связи с этим тестирование проводилось на качественном уровне. Предлагаем под качественной битемпоральной маской понимать маску, которая (а) имеет четкие контуры, (б) не имеет внутренних артефактов (областей, отмеченных как области без изменений, в которых изменения фактически были) во внутренней области и (в) не имеет внешних артефактов (областей отмеченных ложных изменений).

4.2. Результаты экспериментов. Осуществлены (1) количественный тест, направленный на определение производительности предложенного подхода по определению новых объектов в поле технического зрения робота на различных конфигурациях оборудования, и (2) качественный тест, нацеленный на подтверждение способности предложенного подхода создавать адекватные битемпоральные маски для различных объектов и фонов.

Для теста 1 использовались конфигурации оборудования, описанные в таблице 1 (конфигурации 1 и 2 фактически представляют собой работу на CPU, конфигурация 3 – на GPU). Результаты Теста 1 представлены в таблице 2.

Таблица 2. Результаты теста 1 – время на сегментацию кадра с помощью SAM (в миллисекундах для различных весов)

Модель SAM	Конфигурация 1	Конфигурация 2	Конфигурация 3
SAM 2.1 tiny	3020	950 (нет распознавания)	67 (нет распознавания)
SAM 2.1 small	3850	1030	75
SAM 2.1 base	6500	1820	116
SAM 2.1 large	15850	4640	268

Результаты Теста 1 демонстрируют, что реализовать систему технического зрения робота-манипулятора, работающую «на лету» фактически (то есть с временем на сегментацию менее 1000 мс) возможно только с использованием GPU. Необходимо обратить внимание, что на практике при использовании только CPU даже по сравнению с представленными в таблице 1 результатами время обработки может резко и существенно возрастать (как показали эксперименты, в два раза) при увеличении загрузки CPU общесистемными задачами.

Тест 2 заключался в исследовании распознавания различных объектов на фонах различного типа (монотонный, цветной и пестрый) с использованием конфигурации 3 (см. выше). Выбор конкретных объектов был продиктован их цветовой палитрой, с целью необходимости убедиться в работоспособности метода в различных потенциальных окружениях.

Практические результаты экспериментов представлены на таблице 3 ниже. Для иллюстративных целей в таблице 3 представлены не бинарные маски областей изменений (см. образец на рисунке 1), а обратные бинарные маски $Mrev(x, y) = \sim M(x, y)$, закрывающие черным цветом области I_2, в которых изменений нет.

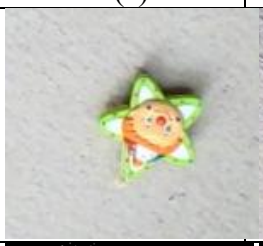
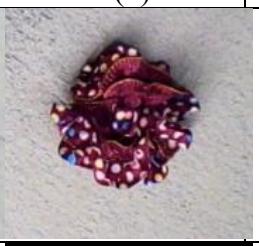
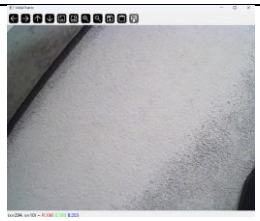
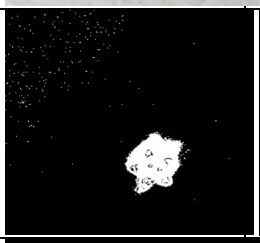
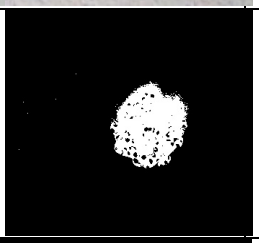
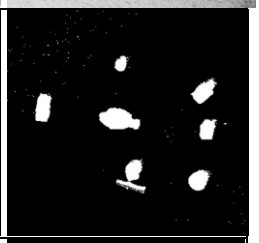
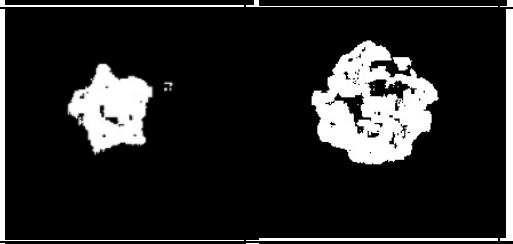
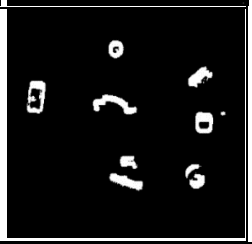
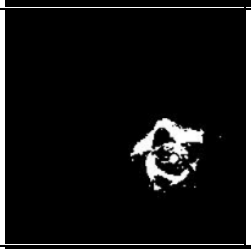
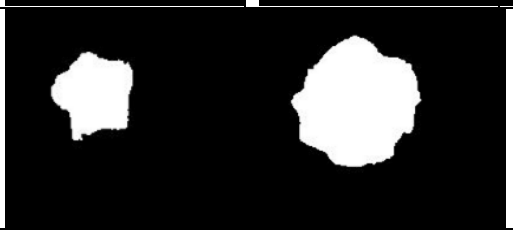
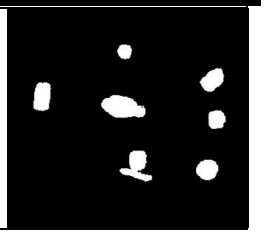
Результаты теста 2, полученные с использованием предложенного метода, сопоставлялись с результатами, полученными при реализации методов Change Vector Analysis (CVA) по аналогии с тем, как это было сделано в [9, 10], и одной из возможных реализаций метода Slow Feature Analysis (SFA) на основе фреймворка Python MDP, как это было сделано в [11].

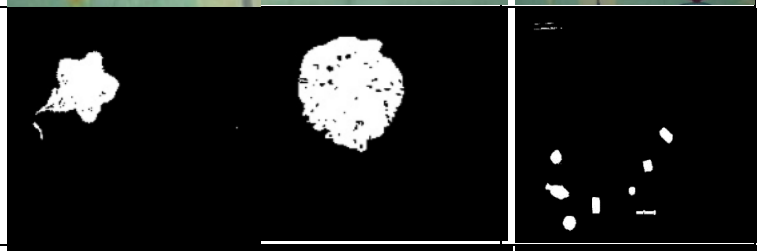
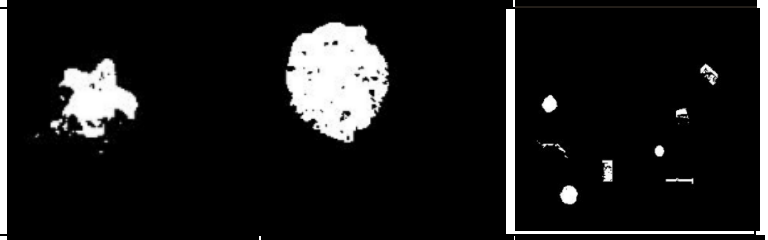
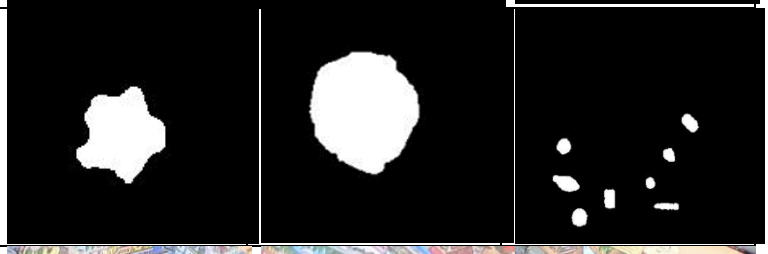
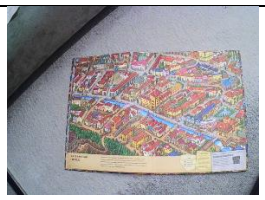
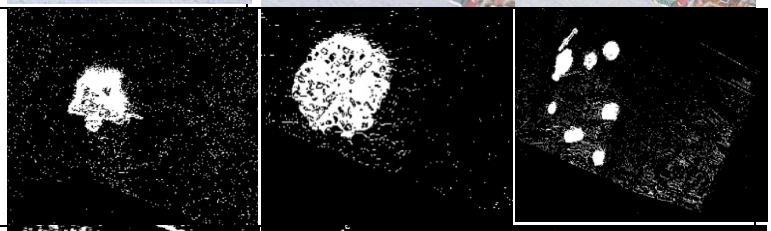
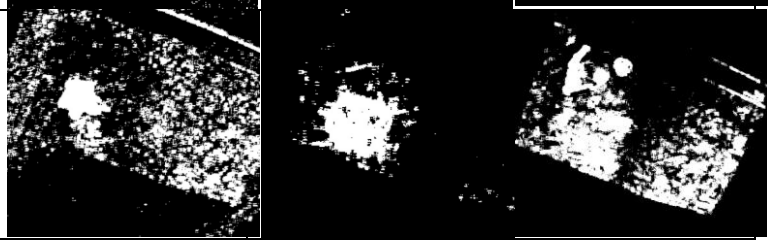
Дополнительные результаты получены и использованием нейронной сети прямого распространения, построенной по архитектуре многослойного перцептрона (Multi-layer Perceptron – MLP): внешний слой из 12 нейронов, внутренний слой из 512 нейронов, наружный

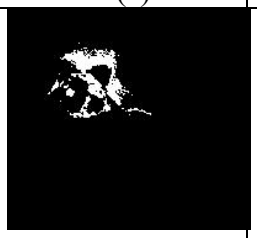
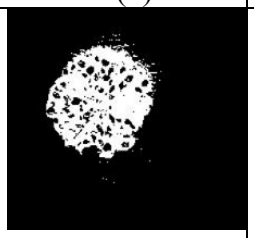
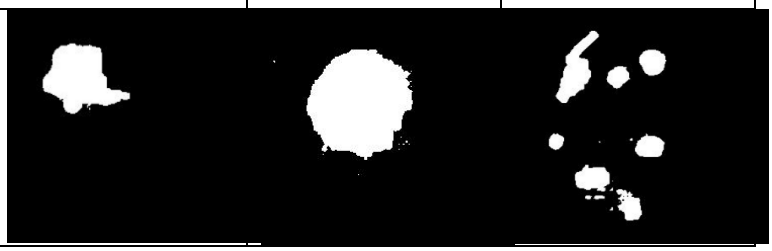
слой из 1 нейрона (использовался как классификатор – были ли изменения в данном пикселе). На внешний слой обучения модели попиксельно подавались векторы, состоящие из нормированных значений пикселей изображения «до», нормированных значений пикселей изображения «после», разницы в значениях нормированных пикселей и квадрата такой разницы. Метки классов (0 – изменения в пикселе не было, 1 – изменение в пикселе было) брались из битемпоральной маски, созданной с применением изложенного в настоящей статье метода, качество которой было признанно наилучшим.

Обращаем внимание, что мы не проводили количественный тест затрат машинного времени для методов сопоставления, аналогичный тесту 1 для предложенного подхода с использованием SAM, в связи с тем, что их реализация в целом требует значительно меньшей производительности оборудования (для любого из приведённых в таблице 3 изображений менее 1 000 мс) по сравнению с использованием нейросети, для работы «на лету» которой необходимо наличие GPU.

Таблица 3. Качественные результаты тестирования

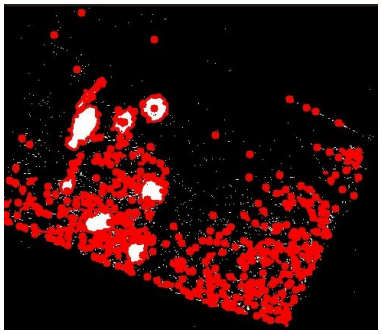
Фон	Метод	Объект (а)	Объект (б)	Объект (в)
Оригинальное изображение				
Монотонный 	CVA			
	SFA			
	MLP			
	Наш			

Фон		Метод	Объект (а)	Объект (б)	Объект (в)
Оригинальное изображение					
Цветной		CVA			
		SFA			
		MLP			
		Наш			
Оригинальное изображение					
Пестрый		CVA			
		SFA			

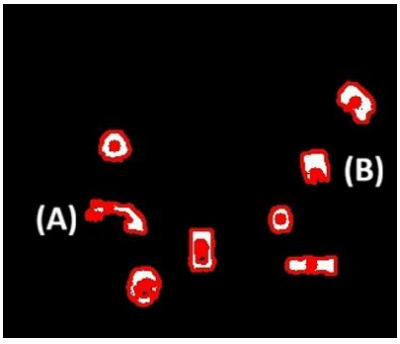
Фон		Метод	Объект (а)	Объект (б)	Объект (в)
		MLP			
		Наш			

Проведенным экспериментом продемонстрировано, что на монотонном фоне возможно созданием маски битемпоральных изменений с использованием любого метода (маски, созданные с помощью использованной реализации метода SFA хуже для маленьких объектов). Однако при усложнении фона (пестрый) альтернативные методы показывают резкое визуально очевидное ухудшение качества масок. Таким образом, продемонстрированно, что при соответствующем выборе порогового значения (параметр τ) предложенным в настоящей статье методом удается создать высокого качества битемпоральные маски на сложных типах фона, даже на пестром (с ухудшением качества масок по сравнению с более простыми типами фонов).

Отметим, что некачественные битемпоральные маски не позволяют выделять контур конкретного объекта интереса (красная линия на рис. 3а) и делают невозможным определить геометрический центр объекта (красная точка на рис. 3б) с целью последующего захвата роботом.



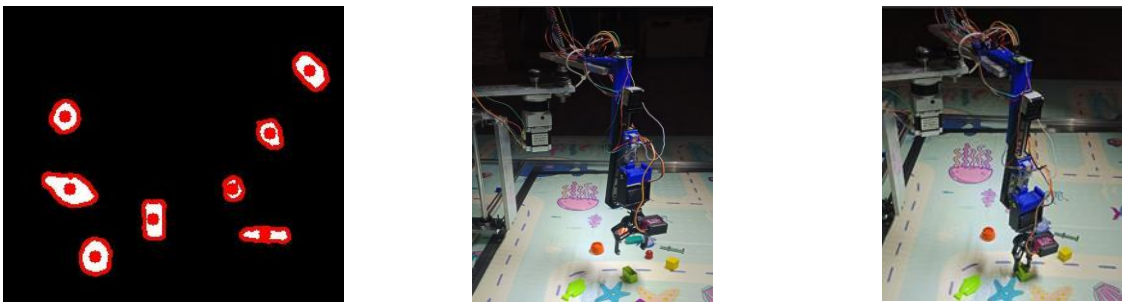
(а)



(б)

Рис. 3. Некачественные битемпоральные маски: (а) Внешнее зашумление не позволяет определить конкретный объект интереса и (б) Некорректные контуры объекта искажают положение геометрического центра фигуры (например, A и B)

Нами проведено практическое тестирование возможности захвата объекта интереса с помощью представленного метода. Тестирование осуществлялось с использованием демонстративного робота SCARA, построенного одним из авторов и описанного в статье [22]. Использование качественных битемпоральных масок изменений (без внешних и внутренних шумов на изображениях), позволяет легко определять контуры объектов интереса и находить их геометрические центры (см. рисунок 4).



(а) (б) (в)

Рис. 4. Качественные битемпоральные маски (а) позволяют осуществить (б) ориентирование схвата робота-манипулятора и (в) захват «нового» объекта на рабочей сцене

Описанный демонстративный результат достигнут с учетом следующих принципиальных особенностей (хоть и практически приемлемых для среды функционирования коллаборативного робота-манипулятора):

1. Наличие теней от самих объектов из-за внешних источников освещения (специальное освещение рабочей сцены не использовалось) может привести к искажению контуров объектов на дифференциальном изображении и, как следствие, невозможности создания точной бинарной маски,

2. Внешние тени (например, от фигуры оператора, находящегося рядом с роботом) могут признаваться в качестве объектов изменения, полностью деструктурируя картину распознавания,

3. Изменение общей освещенности рабочей сцены (например, из-за включения освещения в помещении, в котором функционирует робот), также может привести к искажению контуров объектов, хотя и не всегда критическому.

Заключение. В настоящей статье предложен и экспериментально апробирован подход к определению новых объектов в поле технического зрения коллаборативного робота-манипулятора на основании битемпоральных изображений с минимальной задержкой по времени, то есть для работы в реальных условиях «на лету». Особенностью предлагаемого подхода выступает совмещение методов классического технического зрения для создания дифференциального изображения и последующей нейросетевой сегментации для генерации бинарной маски областей изменения битемпоральных изображений. С учетом ограничений и особенностей, присущих техническому зрению роботов-манипуляторов, подход показал свою практическую применимость для различных типов фонов рабочей области робота, позволяя четко определять контуры и геометрические центры «новых» объектов с целью последующего захвата манипулятором. Вместе с тем, при его использовании необходимо учитывать, что по принципу «на лету» он может быть реализован только на ЭВМ с GPU.

Список источников

1. World Robotics 2025 Report, International Federation of Robotics. Available at: https://ifr.org/downloads/press_docs/PressConference2025_presentation.pdf (accessed: 10/20/2025).
2. Ющенко А.С. Коллаборативная робототехника и человеческий фактор / А.С. Ющенко // Актуальные проблемы психологии труда, инженерной психологии и эргономики. – Москва: Институт психологии РАН, 2020. – С. 83-103.
3. Бадика Е.М. Модель инициализации промышленных роботов с помощью обнаружения объектов на основе глубокого обучения / Е.М. Бадика, В.П. Кузьменко // Флагман науки, 2023. – № 9(9). – С. 377-380.
4. Dong Y.J., Cheng J., Meng L. Object Recognition and Grasping for Baxter Dual-Arm Robot Based on YOLOv8_OBB. Chinese Control Conference (CCC), Chongqing, China, 2025, pp. 4662-4667, DOI:10.23919/CCC64809.2025.11179705.
5. Матвеев В.Д. Разработка модели семантической сегментации RTC-SAM для определения препятствий на пути мобильного робота / В.Д. Матвеев, А.Е. Архипов, И.С. Фомин // Известия ЮФУ. Технические науки, 2025. – № 2(244). – С. 212-220.

6. Giacchetti M., Guerra E., García F.C. et al. Perception for Collaborative Robots in Pruning Operations. International Conference on Emerging Technologies and Factory Automation (ETFA), Padova, Italy, 2024, pp. 01-04.
7. Дайюб Я. SwiftDepth++: эффективная и легковесная модель для точной оценки глубины / Я. Дайюб, И.А. Макаров // Доклады Российской академии наук. Математика, информатика, процессы управления, 2024. – Т. 520. – № 2. – С. 182-192.
8. Fang H., Guo S., Wang X. et al. Automatic Urban Scene-Level Binary Change Detection Based on a Novel Sample Selection Approach and Advanced Triplet Neural Network. IEEE Transactions on Geoscience and Remote Sensing, 2018, vol. 61, pp. 1-18, DOI: 10.1109/TGRS.2023.3235917.
9. Wen X., Yang X. Change Detection from Remote Sensing Imageries Using Spectral Change Vector Analysis. Asia-Pacific Conference on Information Processing, Shenzhen, China, 2009, pp. 189-192, DOI:10.1109/APCIP.2009.183.
10. Xiaolu S., Bo C. Change Detection Using Change Vector Analysis from Landsat TM Images in Wuhan. Procedia Environmental Sciences, 2011, vol. 11, pp. 238-244 , DOI: 10.1016/j.proenv.2011.12.037.
11. Wu C., Du B., Zhang L. Slow Feature Analysis for Change Detection in Multispectral Imagery. IEEE Transactions on Geoscience and Remote Sensing, 2014, vol. 52, pp. 2858-2874, DOI: 10.1109/TGRS.2013.2266673.
12. Song P., Zhao C. Slow Down to Go Better: A Survey on Slow Feature Analysis. IEEE Transactions on Neural Networks and Learning Systems, 2024, vol. 35, pp. 3416-3436, DOI: 10.1109/TNNLS.2022.3201621.
13. Kamal S., Supriya M.H., Pillai P.R.S. Blind source separation of nonlinearly mixed ocean acoustic signals using Slow Feature Analysis. OCEANS 2011 IEEE, 2011, pp. 1-7, DOI: 10.1109/Oceans-Spain.2011.6003620.
14. Ma Q., Jia Y., Gong H. et al. A novel iterative deformable joint attention network for remote sensing image change detection. Multimedia Systems, 2025, DOI: 10.1007/s00530-025-01981-5.
15. Han Y., Li J., Qu Y. et al. HFNet: Semantic and Differential Heterogenous Fusion Network for Remote Sensing Image Change Detection. Journal of Geovisualization and Spatial Analysis, 2024, DOI: 10.1007/s41651-024-00202-3.
16. Choe C., Lee S., Sung N.S. Change Detection for Robotic Patrol System. Eighth IEEE International Conference on Robotic Computing (IRC), 2024, pp. 114-115, DOI: 10.1109/IRC63610.2024.00029.
17. Tsutsui S., Hirakawa T., Yamashita T. et al. Semantic Segmentation and Change Detection By Multi-Task U-Net. IEEE International Conference on Image Processing (ICIP), 2021, pp. 619-623, DOI: 10.1109/ICIP42928.2021.9506560.
18. Huang R. Li. P., Tao H., et al. SemanticDifference: Change Detection with Multi-scale Vision-Language Representation Difference, 2025, DOI: 10.1007/978-981-96-9866-0_13.
19. Kirillov A., Mintun Eric, Ravi N., et al. Segment Anything. IEEE/CVF International Conference on Computer Vision (ICCV), 2023, pp. 3992-4003, DOI: 10.1109/ICCV51070.2023.00371.
20. Ravi N., Gabeur V., Hu Y.-T. SAM 2: Segment Anything in Images and Videos, 2024, DOI: 10.48550/arXiv.2408.00714.
21. Sakurada K., Okatani T. Change detection from a street image pair using cnn features and superpixel segmentation, available at: https://www.researchgate.net/publication/301452621_Change_Detection_from_a_Street_Image_Pair_using_CNN_Features_and_Superpixel_Segmentation.
22. Калушев К.А. Разработка математической модели управления роботом SCARA на базе шаговых двигателей / К.А. Калушев, Л.И. Воронова // Робототехника и техническая кибернетика, 2025. – Т. 13. – № 2. – С. 104-114.

Калушев Константин Александрович. *Национальный исследовательский ядерный университет «МИФИ», аспирант кафедры №22 «Кибернетика». Научные интересы: системы управления роботами и компьютерное зрение. AuthorID: 1225216, SPIN: 2759-0169, ORCID: 0009-0005-7065-0716, konstantin.kalushev@gmail.com, 115409, Россия, Москва, Каширское шоссе, д. 31*

Макаров Илья Андреевич. *к.т.н., Национальный исследовательский ядерный университет «МИФИ», доцент Центра образовательных программ топ-уровня в области ИИ Института Интеллектуальных Кибернетических Систем. Научные интересы: машинное обучение и компьютерное зрение. AuthorID: 826008, SPIN: 3151-9176, ORCID: 0000-0002-3308-8825, iamakarov@hse.ru, 115409, Россия, Москва, Каширское шоссе, д. 31*

UDC 004.89, 681.5
DOI:10.25729/ESI.2026.42.2.004

The analysis of bi-temporal images by a collaborative robot control system to determine new objects in the field of view of the technical vision subunit

Konstantin A. Kalushev, Ilya A. Makarov

National Research Nuclear University "MEPhI", Russia, Moscow, *konstantin.kalushev@gmail.com*

Abstract: One of the tasks associated with developing an interactive collaborative robotic manipulator is temporal analysis of the work scene, i.e., determining the order in which objects appear (and disappear) within the field of view of the vision system. Traditionally, this issue has been considered in the context of satellite imagery and has not been sufficiently addressed in the literature with regard to scenes located approximately 1 m from the camera. At the same time, work scene analysis based on bi-temporal images is a relevant area of research in the context of the development of robotics in general and physical artificial intelligence in particular. Creating high-quality temporal change masks of the work scene makes it possible to determine the contours and geometric centers of new objects for subsequent grasping by the robotic manipulator. A high-quality temporal mask should not contain falsely detected change regions (change objects that do not actually exist), yet should clearly outline the contours of genuine change objects in the work scene. The paper presents a mathematical formulation of the temporal analysis problem and, on its basis, proposes a method for generating temporal change region masks by differentiating "before" and "after" images, combining classical computer vision techniques with the neural network segmentation model SAM (Segment Anything Model). The novelty of the proposed approach lies in applying to the difference image not algebraic processing, but rather its segmentation into two regions (a change region and a no-change region) using a neural network segmentation model. The proposed approach was compared with algebraic methods for creating temporal masks (Change Vector Analysis – CVA and Slow Feature Analysis – SFA) and with the use of a multilayer perceptron neural network architecture (input layer of 12 neurons, hidden layer of 512 neurons, output layer of 1 neuron). It is demonstrated that the proposed approach enables the generation of high-quality change masks for diverse objects against a large number of backgrounds (including cluttered ones), a result that is difficult to achieve with the methods brought for comparison. At the same time, the proposed approach can be implemented "on the fly," i.e., in real time during robot operator work, only if a Graphics Processing Unit (GPU) is available.

Keywords: collaborative robot, bitemporal images, SAM, binary change masks

References

1. World Robotics 2025 Report, International Federation of Robotics. Available at: https://ifr.org/downloads/press_docs/PressConference2025_presentation.pdf (accessed: 10/20/2025).
2. Yushchenko A.S. Kollaborativnaya robototekhnika i chelovecheskiy factor [Collaborative robotics and the human factor]. Aktual'nyye problemy psikhologii truda, inzhenernoy psikhologii i ergonomiki [Current problems of labor psychology, engineering psychology and ergonomics]. Moscow, Institute of Psychology RAS Publ., 2020, pp. 83-103.
3. Badika E.M., Kuzmenko V.P. Model initsializatsii promyshlennykh robotov s pomoshch'yu obnaruzheniya ob"yektov na osnove glubokogo obucheniya [Initialization model for industrial robots using deep learning-based object detection]. Flagman nauki [Flagman sciences], 2023, no. 9(9), pp. 377-380.
4. Dong Y.J., Cheng J., Meng L. Object Recognition and Grasping for Baxter Dual-Arm Robot Based on YOLOv8_OBB. Chinese Control Conference (CCC), Chongqing, China, 2025, pp. 4662-4667, DOI:10.23919/CCC64809.2025.11179705.
5. Matveev V.D., Arkhipov A.E., Fomin I.S. Razrabotka modeli semanticheskoy segmentatsii RTC-SAM dlya opredeleniya prepyatstviy na puti mobil'nogo robot [Development of the RTC-SAM semantic segmentation model for obstacle detection along the path of a mobile robot]. Izvestiya YuFU. Tekhnicheskiye nauki [Proceedings of SFU. Technical sciences], 2025, no. 2(244), pp. 212-220.
6. Giacchetti M., Guerra E., García F.C. et al. Perception for Collaborative Robots in Pruning Operations. International Conference on Emerging Technologies and Factory Automation (ETFA), Padova, Italy, 2024, pp. 01-04.
7. Dayyub Y., Makarov I.A. SwiftDepth++: effektivnaya i legkovesnaya model' dlya tochnoy otsenki glubiny [SwiftDepth++: an efficient and lightweight model for accurate depth estimation]. Doklady Rossiyskoy akademii nauk. Matematika, informatika, protsessy upravleniya [Reports of the Russian academy of sciences. Mathematics, computer science, control processes], 2024, vol. 520, no. 2, pp. 182-192.

8. Fang H., Guo S., Wang X. et al. Automatic Urban Scene-Level Binary Change Detection Based on a Novel Sample Selection Approach and Advanced Triplet Neural Network. IEEE Transactions on Geoscience and Remote Sensing, 2018, vol. 61, pp. 1-18, DOI: 10.1109/TGRS.2023.3235917.
9. Wen X., Yang X. Change Detection from Remote Sensing Imageries Using Spectral Change Vector Analysis. Asia-Pacific Conference on Information Processing, Shenzhen, China, 2009, pp. 189-192, DOI:10.1109/APCIP.2009.183.
10. Xiaolu S., Bo C. Change Detection Using Change Vector Analysis from Landsat TM Images in Wuhan. Procedia Environmental Sciences, 2011, vol. 11, pp. 238-244 , DOI: 10.1016/j.proenv.2011.12.037.
11. Wu C., Du B., Zhang L. Slow Feature Analysis for Change Detection in Multispectral Imagery. IEEE Transactions on Geoscience and Remote Sensing, 2014, vol. 52, pp. 2858-2874, DOI: 10.1109/TGRS.2013.2266673.
12. Song P., Zhao C. Slow Down to Go Better: A Survey on Slow Feature Analysis. IEEE Transactions on Neural Networks and Learning Systems, 2024, vol. 35, pp. 3416-3436, DOI: 10.1109/TNNLS.2022.3201621.
13. Kamal S., Supriya M.H., Pillai P.R.S. Blind source separation of nonlinearly mixed ocean acoustic signals using Slow Feature Analysis. OCEANS 2011 IEEE, 2011, pp. 1-7, DOI: 10.1109/Oceans-Spain.2011.6003620.
14. Ma Q., Jia Y., Gong H. et al. A novel iterative deformable joint attention network for remote sensing image change detection. Multimedia Systems, 2025, DOI: 10.1007/s00530-025-01981-5.
15. Han Y., Li J., Qu Y. et al. HFNet: Semantic and Differential Heterogenous Fusion Network for Remote Sensing Image Change Detection. Journal of Geovisualization and Spatial Analysis, 2024, DOI: 10.1007/s41651-024-00202-3.
16. Choe C., Lee S., Sung N.S. Change Detection for Robotic Patrol System. Eighth IEEE International Conference on Robotic Computing (IRC), 2024, pp. 114-115, DOI: 10.1109/IRC63610.2024.00029.
17. Tsutsui S., Hirakawa T., Yamashita T. et al. Semantic Segmentation and Change Detection By Multi-Task U-Net. IEEE International Conference on Image Processing (ICIP), 2021, pp. 619-623, DOI: 10.1109/ICIP42928.2021.9506560.
18. Huang R. Li. P., Tao H., et al. SemanticDifference: Change Detection with Multi-scale Vision-Language Representation Difference, 2025, DOI: 10.1007/978-981-96-9866-0_13.
19. Kirillov A., Mintun Eric, Ravi N., et al. Segment Anything. IEEE/CVF International Conference on Computer Vision (ICCV), 2023, pp. 3992-4003, DOI: 10.1109/ICCV51070.2023.00371.
20. Ravi N., Gabeur V., Hu Y.-T., SAM 2: Segment Anything in Images and Videos, 2024, DOI: 10.48550/arXiv.2408.00714.
21. Sakurada K., Okatani T. Change detection from a street image pair using cnn features and superpixel segmentation, available at: https://www.researchgate.net/publication/301452621_Change_Detection_from_a_Street_Image_Pair_using_CNN_Features_and_Superpixel_Segmentation.
22. Kalushev K.A., Voronova L.I. Razrabotka matematicheskoy modeli upravleniya robotom SCARA na baze shagovykh dvigateley [Development of a mathematical model for controlling an SCARA robot based on stepper motors]. Robototekhnika i tekhnicheskaya kibernetika [Robotics and technical cybernetics], 2025, vol. 13, no. 2, pp. 104-114.

Kalushev Konstantin Aleksandrovich. *National Research Nuclear University «MEPhI», postgraduate student of Department No. 22 «Cybernetics», Research interests: robot control systems and computer vision. AuthorID: 1225216, SPIN: 2759-0169, ORCID: 0009-0005-7065-0716, konstantin.kalushev@gmail.com. 31, Kashirskoe Highway, Moscow, 115409 Russia.*

Makarov Ilya Andreevich. *Ph.D., National Research Nuclear University «MEPhI», associate professor of the Center of top-level educational programs in AI of the Institute of Intellectual Cybernetic Systems. Research interests: machine learning and computer vision. AuthorID: 826008, SPIN: 3151-9176, ORCID: 0000-0002-3308-8825, iamakarov@hse.ru. 31, Kashirskoe Highway, Moscow, 115409 Russia*.

Статья поступила в редакцию 15.11.2025; одобрена после рецензирования 10.12.2025; принята к публикации 12.05.2026.

The article was submitted 11/15/2025; approved after reviewing 12/10/2025; accepted for publication 05/12/2026.