# Вступление

Визуальное отслеживание заключается в последовательном определении местоположения целевого объекта на каждом кадре видеопотока. Эта задача находит практическое применение во многих областях, так или иначе связанных с обработкой видеозаписей для получения из них некоторой информации. Примерами систем, в которых применяются технологии отслеживания, являются системы человеко-машинного взаимодействия, видеонаблюдения, анализа спортивных матчей, появившиеся в недавнее время системы «умного дома», системы дополненной реальности и многие-многие другие.

Несмотря на значительное количество научных работ и исследований, проводимых в данной области, точное и устойчивое к ошибкам отслеживание объектов на видео остается сложной проблемой. Небольшие размеры объектов, меняющийся фон, резкие ускорения и смены траектории движения объектов, частичные или полные перекрытия, изменение масштаба отслеживаемой цели при ее приближении или удалении от камеры являются основными трудностями на пути к построению точной траектории перемещения объекта на кадрах видеозаписи.

Целью магистерской работы является разработка метода отслеживания объектов в видеопотоке, позволяющего справляться с указанными проблемами. Для достижения поставленной цели необходимо в первую очередь проанализировать существующие методы отслеживания, определить их сильные и слабые стороны и выбрать из них тот, который будет положен в основу разрабатываемого метода. Также необходимо сформулировать математическое описание предлагаемого метода и разработать на его основе алгоритм отслеживания. Для проверки работоспособности предложенного алгоритма требуется разработать реализующее его программное обеспечение и сформировать наборы данных для проведения исследования. По результатам проведенного исследования необходимо сделать вывод о работоспособности метода и накладываемых на него ограничениях.

Работа построена следующим образом. В аналитическом разделе приводится обзор существующих методов отслеживания с указанием их достоинств и недостатков, принимается решение о выборе метода, на основе которого будет вестись дальнейшая разработка. Здесь же приводится математическое описание выбранного метода. В конструкторском разделе описывается разрабатываемый алгоритм отслеживания. Технологический раздел посвящен описанию разработки программного обеспечения, реализующего данный алгоритм, выбору языка и среды программирования, используемых библиотек. В исследовательском разделе описываются условия, при которых проверяется разработанный метод, приводятся метрики оценки его качества. По результатам исследований делается вывод о применимости разработанного метода в тех или иных задачах отслеживания, его достоинствах и накладываемых на него ограничениях, предлагаются направления дальнейшего исследования.

# Аналитический раздел

В данном разделе приводится классификация существующих методов отслеживания, описываются их достоинства и недостатки, обосновывается выбор метода для дальнейшей разработки и дается его формальное описание.

## Обзор подходов к отслеживанию объектов в видеопотоке

Целью визуального отслеживания является построение траектории движения объекта, то есть последовательности его положений на каждом кадре видеопотока. В качестве дополнительной информации алгоритм отслеживания (трекер, англ. tracker) может предоставлять информацию о занимаемой объектом области на кадре. Задачи распознавания области объекта на кадре и сопоставления данных областей на соседних кадрах могут решаться как раздельно, так и совместно. В связи с этим можно выделить два основных подхода к отслеживанию объектов.

В первом подходе ключевую роль играют алгоритмы распознавания . С их помощью определяются новые объекты, появляющиеся в сцене, и инициализируются трекеры. В этом случае процесс отслеживания сводится к задаче сопоставления объектов, распознанных на текущем кадре, с объектами, выявленными на предыдущем кадре. Такой подход реализует отслеживание на основе распознавания.

В качестве примера алгоритма, реализующего данный подход, можно привести алгоритм AdaBoost, описанный в работе [4]. Для представления объекта авторы работы используют признаки Хаара (англ. Haar-like features), гистограммы направленных градиентов (англ. Histogram of Oriented Gradients, HOG) и локальные бинарные шаблоны (англ. Local Binary Patterns, LBP). Задача соотнесения объектов на соседних кадрах решается с помощью процесса сопоставления с эталоном (англ. template matching) и бинарного классификатора.

Использование алгоритмов распознавания в качестве основы для процесса отслеживания позволяет повысить качество получаемых результатов за счет предоставления трекеру достаточно точного описания занимаемой объектом области. С другой стороны, для успешного определения области объекта на кадре необходимо, чтобы объект имел отличительные внешние характеристики (форма, цвет, текстура и т. п.), а для соотнесения выявленных областей требуется, чтобы объект не претерпевал значительных изменений на соседних кадрах. Таким образом, алгоритмы, реализующие данный подход, неэффективны в тех случаях, когда нельзя заранее предсказать визуальные характеристики цели, например, при быстрых перемещениях объекта, когда его внешний вид также меняется достаточно быстро. Данный класс алгоритмов не подходит также для отслеживания целей небольших размеров и целей, не имеющих достаточных отличительных признаков.

Второй поход к визуальному отслеживанию основан на совмещении процессов выделения областей объекта на кадре и их соотнесения между соседними кадрами. Основная идея заключается в итеративном обновлении положения объекта и занимаемой им области на каждом кадре, то есть область объекта вычисляется на основе результатов отслеживания, полученных на предыдущих кадрах. Данный подход реализует распознавание на основе отслеживания. Его отличительной особенностью является описание множества возможных состояний объекта с помощью некоторой функции, анализ которой позволяет получить информацию о его текущем положении.

Для дальнейшего изучения был выбран второй подход, так как он позволяет учитывать динамику объекта за счет использования всей информации о положении и области объекта, полученной с начала отслеживания, а не только с предыдущего кадра. Также данный подход является менее затратным с вычислительной точки зрения, поскольку в большинстве своем алгоритмы, реализующие распознавание на основе отслеживания, используют более слабую модель представления объекта, позволяя избавиться от затрат на распознавание объекта и снимая требование наличия у объекта отличительных внешних характеристик.

Алгоритмы, реализующие распознавание на основе отслеживания, делятся на две группы: детерминированные и стохастические (вероятностные) алгоритмы. Наиболее известными представителями класса детерминированных алгоритмов отслеживания являются сдвиг среднего (англ. Mean Shift) и непрерывно адаптивный сдвиг среднего (англ. Continuously Adaptive Mean-shift, CAM-shift). Их основная идея заключается в задании для каждого кадра весовой функции, такой, чтобы искомое положение объекта соотносилось с ее минимум или максимумом, и сведении задачи отслеживания к оптимизации данной функции. Алгоритмы сдвига среднего основаны на следующем принципе: на каждом кадре ищется область фиксированной формы с переменным размером, цветовое содержание которой наиболее близко к эталонной цветовой модели. Поиск начинается с положения, определенного на предыдущем кадре, и продолжается до тех пор, пока не будет найден минимум функции, задающей расстояние до эталонной цветовой гистограммы. Данный метод показывает достаточно высокие результаты, кроме тех случаев, когда объект находится на фоне, близком к нему по цвету, а также если он на некоторое время полностью перекрыт другими элементами сцены. Общий недостаток детерминированных подходов заключается в низкой способности к восстановлению после потери объекта, поскольку поиск нового положения всегда начинается с последней точки, которая может быть определена некорректно.

Стохастические (вероятностные) алгоритмы отслеживания основаны на теории байесовского оценивания. В них задача отслеживания сводится к итеративной оценке функции плотности распределения вероятности вектора состояния объекта, производимой на основе измерения его значений на последовательных кадрах. Под измерением будем понимать процесс получения оценки возможного состояния объекта на основе информации, извлекаемой из текущего кадра. Результат процесса измерения является наблюдением текущего состояния объекта. Вектор состояния объекта может описывать его положение, размер, цвет, скорость, ускорение и другие характеристики. Вероятностные методы отслеживания позволяют учитывать погрешности, получаемые в ходе измерения значения вектора состояния, а также шумовые воздействия, которым подвержен данный процесс. Помимо этого учитываются стохастические ошибки модели динамики объекта, позволяющие моделировать движение с резкими изменениями направления и скорости. Однако для получения фактической оценки состояния объекта из получаемых наблюдений необходимо избавиться от данных погрешностей. Для этого часто применяют процесс фильтрации.

Широко известным представителем данного класса методов отслеживания является фильтр Кальмана (англ. Kalman Filter). Он основан на предположении, что функция плотности распределения вероятности вектора состояний является Гауссовой, что позволяет получить аналитическое решение ее оценки. К сожалению, данная гипотеза верна далеко не во многих случаях. Для решения этой проблемы были разработаны модификации данного метода: расширенный фильтр Кальмана (англ., Extended Kalman Filter, EKF), применяющий процесс линеаризации, и ансцентный фильтр Кальмана[[1]](#footnote-1) (англ. Unscented Kalman Filter, UKF), использующий детерминированные выборки. Они относятся к категории параметрических методов, работающих с нелинейными функциями плотности распределения [7]. Однако в большинстве своем эти решения показывают удовлетворительные результаты только при задании строгих ограничений на область применения алгоритма. Обычно они требуют наличия большого количества параметров, настройка которых поможет адаптировать алгоритм для решения конкретной задачи. Несмотря на возможность обработки функций плотности распределения вероятности, не имеющих аналитической оценки, параметрические методы вероятностного отслеживания всегда пытаются свести их к Гауссовому распределению. Но если реальная плотность распределения вероятности состояния объекта имеет, например, бимодальный или сильно ассиметричный вид, она не может быть точно описана нормальным распределением. Для таких случаев больше подходят непараметрические методы отслеживания, позволяющие провести аппроксимацию искомой плотности распределения [1].

Если пространство состояний объекта может быть разбито на ячеек, для аппроксимации функции плотности распределения вероятности вектора состояния объекта может быть использован сеточный метод (англ. approximate grid-based method), суть которого заключается в вычислении весового коэффициента для центра каждой ячейки. При этом сетка должна быть достаточно плотной для получения хорошего приближения к непрерывному пространству состояний. Поэтому при увеличении размерности пространства состояний резко возрастает и вычислительная стоимость сеточной аппроксимации. Также данный метод требует заранее определенную сетку, поэтому пространство состояний не может иметь неравномерное разбиение для получения более высокого разрешения в областях с высокой плотностью распределения вероятности, если только эти области не известны заранее [1].

К классу непараметрических методов отслеживания относятся также методы Монте-Карло, позволяющие работать с несколькими гипотезами одновременно, что обуславливает их естественную способность адаптироваться к изменениям, происходящим с отслеживаемым объектом. Среди данной группы методов можно выделить фильтр частиц (англ. particle filter), аппроксимирующий функцию плотности распределения вероятности объекта набором взвешенных частиц.

Иллюстрация изложенной классификации методов визуального отслеживания представлена в приложении А. Проведенный анализ показал, что распознавание на основе отслеживания позволяет получить более точную оценку динамики объекта, поскольку методы данного класса используют всю историю отслеживания для вычисления следующего состояния объекта, а не только данные с предыдущего шага. Вероятностные методы отслеживания, в свою очередь, повышают точность этой оценки за счет принятия во внимание случайных погрешностей модели динамики объекта и шумовых воздействий на процесс измерения его состояния. Поскольку в большинстве реальных задач отслеживания плотность распределения вектора состояния объекта не имеет аналитического представления, необходимо применять методы, способные работать с распределениями, которые не могут быть сведены к Гауссовому без значительной потери точности. Такие методы используют различные способы аппроксимации искомой функции плотности распределения. В данной работе предпочтение отдано фильтру частиц, поскольку он имеет ряд значительных преимуществ. Во-первых, он не использует никаких предположений о характере функции плотности распределения вероятности состояния объекта. Во-вторых, он позволяет анализировать только часть кадра, в отличие от сеточной аппроксимации, при которой на ячейки разбивается вся область полученного с камеры изображения, что позволяет снизить накладные расходы на процесс измерения состояния объекта. В-третьих, увеличение размерности пространства состояний, то есть добавление новых характеристик в вектор состояния объекта, не влечет за собой резкого повышения вычислительной стоимости. Еще одним его достоинством по сравнению с сеточной аппроксимацией является способность повышать концентрацию частиц в областях с высокой плотностью распределения вероятности без использования каких-либо априорных предположений о ее характере. В заключение следует отметить, что фильтр частиц достаточно прост с точки зрения реализации, что позволяет снизить затраты на разработку программного обеспечения.

Среди недостатков фильтра частиц следует выделить необходимость его настройки под конкретную задачу отслеживания. С одной стороны, это повышает гибкость и универсальность метода. Но при этом задача выбора используемых распределений вероятности и их параметров является далеко не тривиальной и требует в некоторых случаях проведения отдельного исследования. Низкая вычислительная стоимость фильтра частиц обусловлена во многом использованием слабого низкоуровневого представления объекта, чаще всего, с помощью цветового распределения, что предполагает наложение ряда ограничений, особенно в тех случаях, когда целевой объект не имеет характерных отличительных внешних признаков.

Несмотря на указанные недостатки, в основу дальнейшей разработки положен аппроксимация функции плотности распределения вероятности состояния объекта с помощью фильтра частиц. Для более глубоко изучения данного метода требуется вначале описать общий принцип вероятностного отслеживания, на котором базируется фильтр частиц.

## Основные принципы вероятностного отслеживания

Как было отмечено ранее, вероятностные методы визуального отслеживания основаны на байесовском подходе, который заключается в задании состояния объекта с помощью функции плотности распределения его вероятности. Данная функция характеризует текущую степень знания о состоянии объекта и строится на основе всей полученной в ходе отслеживания информации.

Задачу отслеживания можно сформулировать с помощью следующего математического представления пространства состояний объекта:

(1.1)

(1.2)

Уравнение  описывает изменение состояния объекта при переходе от кадра к кадру и задает модель динамики объекта. Фактически состояние зависит от предыдущего состояния объекта и случайной ошибки , представляющей собой погрешность, получаемую при обновлении вектора состояния объекта. Поскольку ошибка является случайной величиной с известной статистикой, уравнение  неявно задает функцию плотности распределения вероятности . Уравнение  описывает процесс измерения, результатом которого является некоторая величина , называемая наблюдением. Наблюдение зависит от текущего состояния и случайной ошибки , описывающей шумовое воздействие, оказываемое на процесс измерения состояния объекта. Аналогично ), поскольку является случайной величиной, уравнение  неявно задает функцию правдоподобия .

С точки зрения байесовского подхода проблема отслеживания заключается в рекурсивном вычислении некоторой степени доверия к состоянию объекта на кадре с учетом всех имеющихся к текущему моменту наблюдений. Таким образом, для каждого кадра необходимо вычислить функцию плотности распределения вероятности состояния объекта ), называемую постериорной (англ. posterior). Для ее вычисления необходимы результаты, полученные на предыдущем шаге (плотность распределения вероятности ), а также априорная плотность распределения вероятности смены состояния (англ. state transition prior) и функция правдоподобия (рис.Рис. 1.1).

Вычисление плотности распределения вероятности состояния объекта состоит из двух этапов: прогноза и коррекции (рис. Рис. 1.2). Полагая известным значение искомой функции на предыдущем кадре, с помощью уравнения Чепмена-Колмогорова  можно вычислить априорную функцию плотности распределения вероятности состояния объекта на кадре :

(1.3)

|  |
| --- |
| idef0.jpg |
| Рис. 1.1. Вычисление плотности распределения вероятности состояния объекта на одном кадре |

|  |
| --- |
| idef0_1.jpg |
| Рис. 1.2. Этапы вычисления плотности распределения вероятности состояния объекта на кадре : предсказание и коррекция. |

В этом заключается этап предсказания. На кадре становится известным наблюдение , которое может быть использовано для обновления априорной функции по правилу Байеса :

(1.4)

где — нормирующий множитель, зависящий от функции правдоподобия.

Рекурсивные соотношения  и  формируют базис для нахождения оптимального байесовского решения. Однако данный подход является концептуальным с той точки зрения, что в общем случае данное решение не может быть получено аналитическим путем. Примером частного случая, когда аналитическое решение поставленной задачи все-таки существует, является упомянутый в предыдущей главе фильтр Кальмана. В более сложных случаях искомое решение можно получить с помощью аппроксимации. Одним из возможных способов для этого является фильтр частиц [1].

## Фильтр частиц

В основе метода лежит аппроксимация функции плотности распределения вероятности состояния объекта набором частиц, поведение которых регулируется с помощью их весов. Фактически, частица является представлением некоторого возможного состояния объекта. Множество   задает приближенную функцию плотности распределения вероятности для состояния при заданном наборе наблюдений .

(1.5)

где – количество частиц, — вес частицы , задаваемый с помощью функции правдоподобия. Вес частицы показывает вероятность, с которой объект примет состояние, описываемое данной частицей. Веса частиц нормализованы: . Тогда искомая постериорная плотность распределения вероятности состояния объекта может быть представлена как (1.6):

(1.6)

где — дельта функция Дирака.

Теперь задача вычисления плотности распределения вероятности состояния объекта на кадре сводится к построению множества на основе полученного на предыдущем шаге множества . Для этого необходимо сформировать новый набор частиц, производя выборку из согласно некоторому «прогнозному» распределению (англ. proposal distribution) , зависящему в общем случае от предыдущего состояния объекта и получаемого на текущем кадре наблюдения, т.е:

(1.7)

Выбор подходящего распределения зависит от особенностей решаемой задачи и способен оказать значительное влияние на получаемые результаты. Наиболее часто встречается случай, когда совпадает с плотностью распределения вероятности смены состояния , но известны и другие варианты, как, например, в работе [2].

После формирования нового набора частиц необходимо пересчитать их весовые коэффициенты согласно выражению (1.8) и нормализовать их.

(1.8)

При новые вычисленные весовые коэффициенты пропорциональны функции правдоподобия .

Таким образом, выражение (1.8) описывает процесс коррекции (1.4) в терминах фильтра частиц.

## Выводы

В данном разделе был проведена классификация существующих подходов к визуальному отслеживанию объектов, проанализированы их достоинства и недостатки, обоснован выбор фильтра частиц в качестве основы для дальнейшей разработки. Кроме этого, было дано аналитическое описание основных принципов вероятностного отслеживания и их представление в фильтре частиц.

# Конструкторский раздел

В данном разделе приводится алгоритмическая основа разрабатываемого метода, описывается способ представления объекта на кадре и процесс измерения его состояния, задается модель динамики объекта. Также описывается предлагаемая модификация стандартного алгоритма.

## Алгоритм воспроизведения условной плотности

Основные принципы работы фильтра частиц можно рассмотреть на примере алгоритма воспроизведения условной плотности (англ. Conditional Density Propagation, Condensation). Он является базовым алгоритмом фильтра частиц, на основе которого разрабатываются различные модификации для адаптации под конкретные условия отслеживания. Основная идея алгоритма заключается в построении множества частиц на основе множества с предыдущего шага .

На каждой итерации алгоритма полагаем множество пустым. Алгоритм включает в себя 3 этапа:

* перевыборка (англ. resampling);
* предсказание;
* коррекция (обновление весов).

Этап перевыборки задается выражением (1.7). Алгоритм воспроизведения условной плотности полагает, что плотность распределения вероятности совпадает с плотностью распределения вероятности смены состояния . В этом случае первый этап алгоритма заключается в выборе некоторой частицы из множества и помещении ее в множество . Вероятность, с которой частица попадет в новое множество, равна ее весу . Данную операцию необходимо провести раз, чтобы мощность множества совпадала с мощностью . В результате перевыборки получим множество , состоящее из частиц с наибольшими весами из предыдущего множества.

На этапе предсказания каждая частица полученного множества обновляется согласно уравнению динамики :

(2.1)

Этот шаг алгоритма соответствует вычислению априорной функции плотности распределения вероятности . В качестве уравнения динамики могут быть использованы:

* модель случайных блужданий (англ. Random Walk Model) [5];
* модель динамики первого порядка [6], [3];
* авторегрессионная модель второго порядка [2] и др.

Этап коррекции заключается в обновлении весов частиц(1.8). При условии, что , новые веса вычисляются пропорционально выбранной функции правдоподобия. После обновления необходимо провести нормализацию весов, так, чтобы их общая сумма была равна единице .

(2.2)

В результате всех вышеописанных операций на каждой итерации получаем искомое множество частиц . Для оценки состояния объекта на текущем шаге необходимо взять «среднюю» частицу полученного множества :

(2.3)

/\* Конкретные реализации алгоритма (то, что я делала). Описание простого алгоритма, его недостатки, описание алгоритма с адаптацией, какие у него проблемы и что было внесено своего.\*/

# Технологический раздел

/\* выбор языка программирования, структуры данных, наверно, надо еще сказать здесь что-нибудь про тестирование \*/

# Исследовательский (Экспериментальный) раздел

/\* Оценка качества работы (описать, какие бывают, или это лучше в аналитический раздел вынести?). Какие результаты были получены, графики и пр.\*/

# Заключение

// выводы….

# Список литературы

. Arulampalam M. S., Maskell S., Gordon N., Clapp T. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking // IEEE Transactions on Signal Processing, 2002, vol. 50, no. 2, pp. 174-188. DOI: 10.1109/78.978374.

. Cai Y., de Freitas N., Little J. Robust visual tracking for multiple targets. // Computer Vision – ECCV, 2006, vol.3954, pp. 107-118. DOI: 10.1007/11744085\_9.

. Del Bimbo A., Dini F. Particle filter-based visual tracking with a first order dynamic model and uncertainty adaptation // Computer Vision and Image Understanding. 2011. Vol. 115. No. 6. P. 771-786. DOI: 10.1016/j.cviu.2011.01.004.

. Grabner H., Grabner M., Bischof H. Real-time tracking via on-line boosting // Proceedings on British Machine Vision Conference (BMVC), 2006, vol. 1, pp. 47-56.

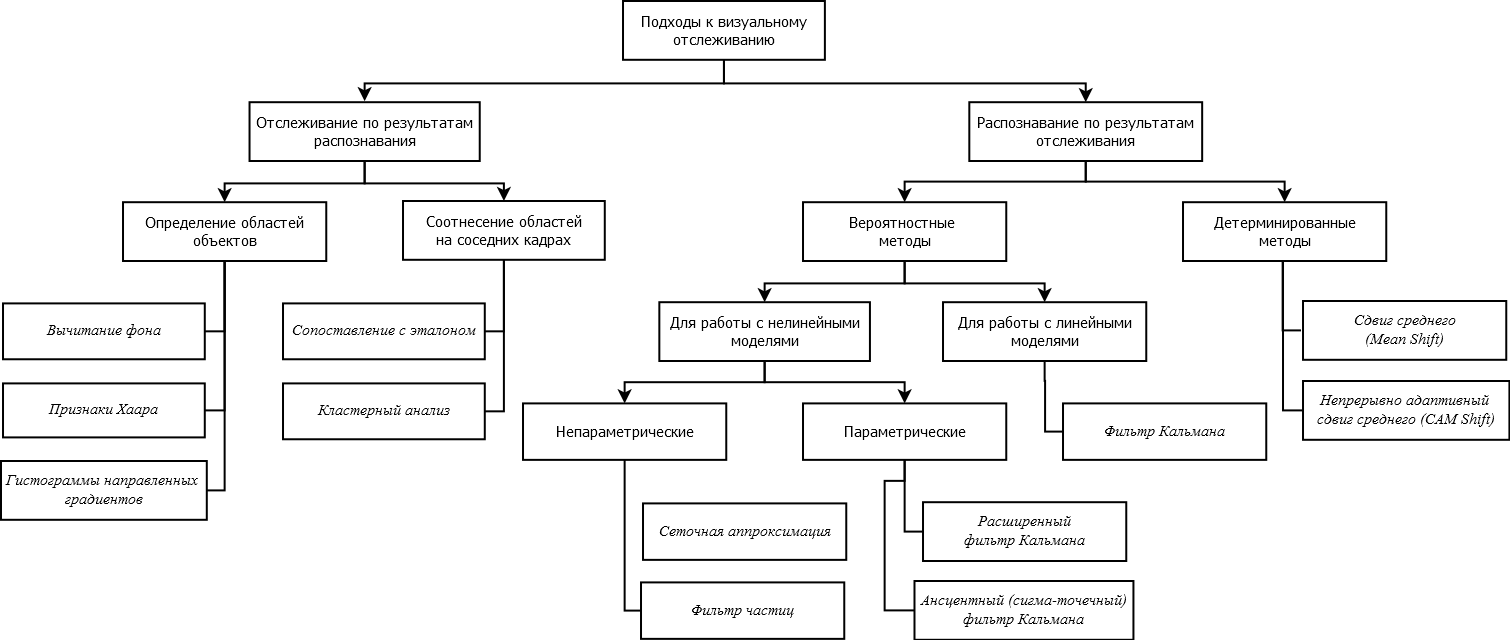
. Jaward M., Mihaylova L., Canagarajah N., Bull D. Multiple object tracking using particle filters // Aerospace Conference. IEEE, 2006, pp. 8–. DOI: 10.1109/AERO.2006.1655926.

. Nummiaro K., Koller-Meier E., Van Gool L. An adaptive color-based particle filter // Image and Vision Computing, 2003. vol. 21, no. 1, pp. 99–110. DOI: 10.1016/S0262-8856(02)00129-4.

. Rui. Y., Chen Y. Better proposal distributions: object tracking using unscented particle filter // Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (CVPR 2001), 2001, vol. 2, pp.II-786 – II-793. DOI: 10.1109/CVPR.2001.991045.

# Приложение А

# Классификация подходов к визуальному отслеживанию объектов



1. В названии алгоритма приведена транслитерация оригинального названия, поскольку устоявшегося термина в русском языке пока нет, а дословный перевод не имеет смысла (unscented – «без запаха»). Также данный алгоритм иногда называют сигма-точечным фильтром Кальмана (англ. Sigma-Point Kalman Filter) [↑](#footnote-ref-1)