Федеральное государственное автономное образовательное учреждение
высшего образования
«Московский физико-технический институт
(национальный исследовательский университет)»
Физтех-школа Прикладной Математики и Информатики
Кафедра интеллектуальных систем

Направление подготовки: 03.03.01 Прикладные математика и физика (бакалавриат)
Направленность (профиль) подготовки: Компьютерные технологии и интеллектуальный анализ данных

Оценивание параметров в задаче слежения за множеством объектов в видеопотоке

(бакалаврская работа)

Студент:
Григорьев Алексей Дмитриевич

(подпись студента)

Научный руководитель:
Гнеушев Александр Николаевич,
к-т физ.-мат. наук

(подпись научного руководителя)

Москва 2020

Содержание

Данная работа посвящена задаче слежения за множеством объектов в видео потоке изображений. Предлагается метод слежения, опирающийся на ре-идентификацию с отбором объектов на основе оценки показателя "качества". Существующие методы в данной области имеют низкую вычислительную эффективность и склонны к ошибкам предсказания и разрывам траекторий объектов. Отбор объектов из траектории для задачи ре-идентификации значительно уменьшает вычислительную сложность алгоритма. Биометрический подход на основе оценки параметра "качества" позволяет исключить из ре-идентификации объекты с низкой информативностью и объекты, не являющиеся предметом наблюдения. Вычислительный эксперимент, проведенный на данных с камер видеонаблюдения с использованием детекторов различной сложности, показал вычислительную эффективность предложенных методов и значительное уменьшение числа некорректных переключений оцениваемых траекторий между различными объектами.

Ключевые слова: слежение за множеством объектов; ре-идентификация; отбор кандидатов к ре-идентификации; оценка биометрического качества.

1 Введение

Проблема сопровождения объектов является очень актуальной в области компьютерного зрения в настоящее время. Задачи слежения возникают как в системах безопасности, автоматизации контроля и учета, так и области беспилотных транспортных средств, робототехнике. Оперативная обработка поступающих данных в режиме реального времени позволяет снизить вычислительную нагрузку на систему и обрабатывать поступающие данные локально, независимо от наличия соединения с удаленным сервером, повысить отказоустойчивость и использовать вычислительные устройства с большей эффективностью.

Задача сопровождения объектов связана с их обнаружением на видео последовательности кадров, является составной частью систем идентификации и верификации объектов, оценки их пространственного положения. Современные детекторы объектов на видео изображениях не позволяют оценить координаты объектов с необходимой для рассматриваемых систем точностью. Использование моделей слежения помогают уточнить координаты объекта путем оценивания траектории движения по измерениям его положения на каждом кадре, предсказывать его перемещение в будущие моменты времени. Предсказанное положение объекта позволяет значительно уменьшить площадь областей для последующего анализа изображения, увеличить вычислительную эффективность системы в целом. Получение множества изображений одного объекта с помощью системы слежения позволяет значительно увеличить точность его идентификации на основе подходов накопления вероятности распознавания и агрегации признаков.

Задача слежения может быть решена путем восстановления траекторий объектов в видеопотоке изображений на основе обнаружения объектов на каждом кадре, при этом предполагается, что число объектов заранее неизвестно. Особый интерес представляет задача слежения, включающая дополнительное ограничение работы в онлайн режиме (online mode). В этом режиме в каждый момент времени доступна информация, полученная только из предыдущих по времени видео кадрах. Данная задача возникает в практических приложениях, работающих в режиме реального времени.

Как правило, задача сопровождения объектов в онлайн режиме допускает представление в виде нескольких подзадач: инициализация траекторий объектов, обнаружение объекта на кадре и сопоставление его положения с оцениваемой траекторией, предсказание траектории на следующий кадр, определение конца траектории. Данные подзадачи решаются последовательно для каждого кадра видео последовательности.

Методы сопровождения объектов используются во множестве приложений, на которые накладываются дополнительные ограничения. Они должны быть вычислительно эффективны, работать в режиме реального времени, то есть частота обработки кадра алгоритмом должна быть не меньше частоты кадров в видео потоке. Часто данное ограничение приводит к упрощению используемых методов в каждой подзадаче, в частности, для подзадачи предсказания траектории на следующий кадр используют линейную модель движения объектов, если это допускается геометрией сцены наблю-

дения [1].

Подзадачи обнаружения объекта на кадре и сопоставление его положения с оцениваемой траекторией являются наиболее сложными в задаче слежения. Обнаружение и локализация объекта на кадре решается с помощью различных систем детектирования, обученных на целевой класс изображений объектов. Наиболее эффективными на текущий момент являются нейросетевые детекторы, основанные на популярных архитекторах SSD [2], YOLO [3], RetinaNet [4]. Они хорошо себя зарекомендовали и показывают допустимую точность и относительно приемлемую скорость для большинства практических приложений, в которых возможно использование специализированных матричных процессоров. Однако нестабильность локализации объекта детектором, невозможность быстрой идентификации каждого вновь найденного объекта на кадре приводит к необходимости сопоставления найденного объекта с уникальной траекторией для его ре-идентификации, уточнения его положения и предсказания движения.

Подзадача сопоставления положения объекта с его траекторией сводится к линейной задаче о назначениях с некоторой функцией стоимости, которая эффективно решается существующими методами [5, 6]. Есть различные подходы к выбору функции стоимости. Метод, предполагающий линейную модель движения объекта, положение которого не претерпевает существенного изменения на последовательных кадрах, использует отношение площади пересечения к площади объединения областей предсказания и обнаружения объекта, IoU (Intersection over Union) меру сходства в качестве функции стоимости в задаче о назначениях. Данный подход показывает высокое качество для простых сценариев в задачах сопровождения [7, 8]. В то же время отмечается [10], что в этом случае возникают проблемы с разрывами траектории объекта, не наблюдавшегося в видео потоке на протяжении некоторого числа кадров. Данная проблема связывается со сложным движением подобных объектов, не соответствующим линейной модели.

Для решения проблемы с разрывами траекторий слежения предлагается подход ре-идентификации временно пропадавших из кадра объектов [10]. Данный подход использует признаковое описание изображения объекта (дескриптор), позволяющий сравнивать области объекта на изображении, производить их повторное обнаружение и привязку к существующей траектории. Для построения дескриптора объекта используется сверточная нейронная сеть, выходы ее последнего полносвязного слоя формируют признаковое векторное пространство описания изображения целевого объекта. В данном пространстве вводится косинусная мера близости векторов признаков для сопоставления области объекта, предсказанной по траектории, и новых областей обнаружения объектов. Использование нейронной сети приводит к уменьшению числа неверных реидентификаций объектов в видео потоке ценой увеличения вычислительной сложности алгоритма слежения.

В данной работе рассматривается задача сопровождения лиц людей по видео последовательности кадров и предлагается развитие подхода к решению подзадачи ре-

идентификации, сопоставления объектов с использованием процедуры предварительного отбора областей-кандидатов на основе оценки показателя их "качества". Показатель "качества" определяется мерой полезности объекта для задачи распознавания, который предлагается оценивать на основе степени уверенности детектора и специально разработанного алгоритма оценки близости изображения объекта к своему классу. Предлагаемый подход позволяет увеличить эффективность ре-идентификации объектов, уменьшить вероятность разрывов траекторий и некорректных переключений.

Цель и задачи исследования. Целью исследования является разработка системы сопровождения лиц людей по видео последовательности кадров, основанной на ре-идентификации объектов с использованием процедуры предварительного отбора найденных объектов на основе оценки показателя "качества". Для достижения этой цели поставлены следующие задачи:

- изучить существующие методы решения задачи онлайн сопровождения объектов по видео последовательности изображений;
- реализовать наиболее успешные методы, адаптировав их к задаче слежения за лицами;
- предложить метод решения задачи сопоставления областей обнаруженных объектов с их траекториям на основе ре-идентификации, сформулировать критерии выбора объектов-кандидатов;
- реализовать предложенный метод и провести вычислительные эксперименты;
- сравнить предложенный метод с существующими на тестовых выборках с помощью принятых показателей качества и оценки вычислительной эффективности;
- провести анализ полученных результатов.

Научная новизна. Сформулированная постановка задачи одновременного онлайн сопровождения объектов на видео изображении сведена к задаче о назначениях обнаруженных объектов к соответствующим траекториям с одновременной оценкой их параметров. Для проблемы ре-идентификации объектов предложен подход модификации критерия стоимости в задаче о назначениях, позволяющий отбирать для подсчета стоимости назначения только несколько наиболее информативных дескрипторов изображений объектов с помощью оценки показателя "качества".

Практическая ценность. Предложенный метод обладает лучшей по сравнению с аналогами вычислительной эффективностью и работает в режиме реального времени при сопоставимом качестве с существующими системами слежения. Применение данного метода в системах распознавания объектов позволит снизить нагрузку на систему, использовать вычислительные устройства с большей эффективностью, повысить отказоустойчивость и увеличить точность распознавания.

2 Обзор литературы

2.1 Модели слежения за множеством объектов

К настоящему времени предложено множество методов онлайн слежения, функционирующих в режиме реального времени [1, 10]. Наиболее успешный подход SORT [1] основывается на использовании линейного фильтра Калмана [18] в задаче предсказания траектории на следующий кадр. Метод ограничивается линейной моделью движения объектов в видеопотоке и существенно опирается на допущение, что частота кадров достаточна велика и объекты меняют свое пространственное положение в кадре несильно. Данное допущение позволяют применять IoU в качестве меры близости объектов. Задача соответствия обнаруженных объектов и их траекторий решается венгерским алгоритмом [5]. Инициализация траекторий осуществляется при наблюдении объекта, движение которого соответствует линейной модели на протяжении некоторого числа кадров. Удаление траектории производится при отсутствии сопоставления между новыми обнаружениями объекта и данной траекторией на протяжении некоторого числа последовательных кадров видео потока. В сценариях, соответствующим введенным допущениям, данный подход показывает приемлемое качество в задаче онлайн сопровождения объектов, существенно превосходя альтернативные подходы, работающие в реальном режиме времени. Ограничение модели линейностью движения объектов является строгим, его невыполнение приводит к многочисленным сбоям в отслеживании объектов. Данный метод имеет проблемы с ре-идентификацией объекта, не найденного на протяжении некоторого числа последовательных кадров, поскольку такие объекты имеют сложное движение, не соответствующее линейной модели. Данный недостаток существенно сужает сферу применимости метода, ограничивая его простыми сценариями с низкой плотностью объектов в видеопотоке.

Развитием данного подхода является введение признакового представления объектов, что позволяет производить ре-идентификацию временно пропавших из кадра объектов [10]. Метод Deep-SORT также использует линейный фильтр Калмана [18] для решения задачи предсказания траекторий и аналагичный методу SORT подход к решению задачи инициализации и удаления траекторий. Новизна метода заключается в альтернативном подходе к решению задачи сопоставления траекторий с новыми обнаружениями объектов. В этой работе для изображений объектов предлагается использовать признаковое представление (дескрипторы) в некотором векторном пространстве. В предложенном методе дескрипторы получают с использованием сверточной нейронной сети. Нейросеть представляет собой авторскую модификацию архитектуры ResNet [16]. Нейронная сеть обучается на размеченной части видеоряда, решая задачу мультикласовой классификации, где число классов соответствует числу объектов в видеопотоке. Выходом сети является векторный дескриптор, нормированный по евклидовой норме. На многомерной сфере вводится косинусная мера схожести дескрипторов, на основе которой производится сопоставление ранее полученных дескрипторов на траектории

объекта и новых обнаруженных объектов. Для уменьшения ошибок сопоставления производится отсеивание кандидатов по порогу максимального значения расстояния Махаланобиса [10] параметров кандидата от предсказанных характеристик траектории фильтром Калмана [18].

При составлении матрицы стоимости в задаче о назначениях производится попарное вычисление косинусных мер сходства между дескрипторами новых объектов и всеми ранее полученными вдоль всех траекторий. Вес, соответствующий некоторому назначению, определяется как наибольшая косинусная мера сходства между дескриптором данного объекта и объектов в соответствующей траектории. Таким образом, сложность вычисления матрицы стоимости зависит от числа дескрипторов для уже построенных траекторий. Данное ограничение существенно увеличивает вычислительную сложность алгоритма, что приводит к существенному уменьшению скорости обработки видеопотока по сравнению с исходным методом SORT.

Более того, данный подход при определенных условиях подвержен проблеме некорректной ре-идентификации, поскольку при составлении матрицы стоимости используются дескрипторы всех объектов, связанных с траекторией, в том числе ошибочно связанные. Поскольку метод выделения признакового описания существенно опирается на наличие в области объекта заданного класса, то при применении данного метода к некоторой области изображения, не содержащей целевого объекта, приводит к некорректной работе алгоритма. Признаковое представление, полученное по изображению, не соответствующему распределению данных, на которых была обучена нейронная сеть, может иметь произвольную косинусную схожесть с другими объектами. Данная особенность негативно влияет на ре-идентификацию с использованием всех объектов, связанных с траекторией, без их предварительного отбора и может приводить к накапливанию ошибки с течением времени.

Таким образом, использование нейронной сети приводит к уменьшению числа неверных ре-идентификаций объектов в видеопотоке, уменьшая число некорректных сопоставлений траекторий и новых обнаружений, ценой увеличения вычислительной сложности алгоритма слежения.

Для задачи слежения за несколькиими объектами предложено несколько подходов к оценке качества алгоритмов [7–9]. Число мер качества велико, поскольку разные приложения накладывают разные требования на решение задачи. Большинство показателей качества предполагает решение вспомогательной задачи сопоставления траекторий, полученных в результате работы алгоритма, с истинными траекториями [8, 9].

Основные и наиболее репрезентативные меры качества — это обобщение мер точности (precision) и полноты (recall). Важное значение имеет показатель качества, соответствующий числу некорректных переключений при продлении траекторий, данная мера качества коррелирует с числом ложных ре-идентификаций одного и того же объекта в видеопотоке [8].

За последние годы предложено множество выборок тестовых данных для оценки

качества моделей слежения [11–15]. Большинство предлагаемых подходов тестируются на стандартных для данной задачи выборках данных [11, 12], что делает доступными полную информацию о конфигурации модели и ее измеренному показателю качества. Это существенно упрощает определение наиболее успешных и эффективных подходов в области задачи слежения за множеством объектов.

2.2 Оценка качества изображения для задачи распознавания

Оценка качества изображения является стандартной задачей в биометрии и, в частности, в области распознавания лиц. Данная задача сводится к нахождению функции, отображающей изображения объекта в действительное число, являющееся оценкой качества данного объекта. Под качеством в биометрии понимается ожидаемая точность распознования по данному объекту, другими словами, величина, коррелирующая с полезностью данного объекта для задачи распознования [19].

Базовый подход [19] к решению задачи оценки качества изображения предполагает использование частных показателей, основанных на контрастности, яркости, фокусировке, резкости и освещенности. Для каждой из характеристик производится оценка плотности вероятности правдоподобия, которая строится в классе гауссовских распределений. Значения, близкие к среднему значению функции правдоподобия для каждого параметра, соответствуют хорошему качеству. Итоговая оценка качества определяется как среднее геометрическое оценок, полученных на основе каждой из указанных характеристик изображения.

Основным недостатком данного подхода является использование гауссовского приближения, что не всегда является корректным допущением на практике.

В работе [20] предложен метод оценки качества, не опирающийся на предположения нормальности распределения частных показателей. Оценка качества сводится к обучению с учителем в задаче регрессии, где ответы получены в результате асессорской разметки данных и в качестве алгоритма машинного обучения используется метод опорных векторов [21]. Предварительно по изображениям обучающей выборки строятся признаковые представления с использованием сверточной нейронной сети, предобученной для решения задачи классификации множества объектов исследуемого класса.

При большой обобщающей способности данного метода основными его недостатками являются необходимость большого количества размеченных данных для обучения модели регрессии и ручная разметка данных, с плохо контролируемым уровнем ошибок, которые приводят с смещению результата для обученной по таким данным модели.

3 Постановка задачи

Дана последовательность изображений видео потока с фиксированным шагом дискретизации по времени $\{\mathbf{I}_t\}_{t=1}^T$, где T – длина видеоряда, $\mathbf{I}_t \in \mathbb{I} \subset \mathbb{R}^{C\times H\times W}$, где C, H, W – число цветовых каналов, высота и ширина изображения соответственно. Задано множество объектов $O = \{j|j \in \overline{1,N}\}$, N – число объектов в видеоряде $\{\mathbf{I}_t\}_{t=1}^T$.

Пусть $\mathbf{y}_{j,1}, \ldots, \mathbf{y}_{j,T}$ – измеренные наблюдаемые характеристики объекта j в каждый момент времени t, $\mathbf{x}_{j,1}, \ldots, \mathbf{x}_{j,T}$ – скрытые характеристики объекта в каждый момент времени t, состояние объекта, подлежащие оцениванию. Для кадра t определим измеренные характеристики $\mathbf{z}_{1,t}, \ldots, \mathbf{z}_{M_t,t}$ неизвестных объектов, где M_t – число обнаруженных в кадре неизвестных объектов. Задача сопровождения множества объектов в онлайн режиме состоит в нахождении совместных оценок характеристик $X_t = \{\mathbf{x}_{j,t}, |j = 1, \ldots, N\}$ всех объектов N на основе всех измерений данных объектов на предыдущих кадрах $Y_t = \{\mathbf{y}_{j,1}, \ldots, \mathbf{y}_{j,t-1}, |j = 1, \ldots, N\}$ и измеренных характеристик $Z_t = \{\mathbf{z}_{1,t}, \ldots, \mathbf{z}_{M_t,t}\}$ неизвестных объектов на данном кадре t.

Предполагается, что изменение характеристик каждого объекта j обладают свойством марковского процесса, то есть величина состояния $\mathbf{x}_{j,t}$ зависит только от состояния объекта $\mathbf{x}_{j,t-1}$ в предыдущий момент времени, причем наблюдаемые характеристики $\mathbf{y}_{j,t}$ определяются исключительно скрытыми параметрами объекта $\mathbf{x}_{j,t}$. Тогда задаче фильтрации характеристик объекта j соответствует байессовская сеть. Используя допущение, что изменение характеристик объекта $\mathbf{x}_{j,t}$ подчиняется линейной зависимости, данный процесс для каждого объекта j описывается моделью линейной динамической системы (ЛДС) и все распределения в задаче задаются гауссовской моделью:

$$\begin{aligned} p(\mathbf{x}_{j,t}|\mathbf{x}_{j,t-1}) &= N(\mathbf{A}_j\mathbf{x}_{j,t-1}, \mathbf{\Gamma}_j), \\ p(\mathbf{y}_{j,t}|\mathbf{x}_{j,t}) &= N(\mathbf{B}\mathbf{x}_{j,t}, \mathbf{\Sigma}_j), \\ p(\mathbf{x}_{j,1}) &= N(\boldsymbol{\mu}_{j,0}, \mathbf{\Gamma}_{j,0}), \end{aligned} \tag{1}$$

где $p(\mathbf{x}_{j,t}|\mathbf{x}_{j,t-1})$ - описывает изменение состояние объекта и с матрицей $\mathbf{A}_j$ определяет линейный закон движения объекта и изменения его параметров с случайным шумом с ковариационной матрицей $\mathbf{\Gamma}_j$; $p(\mathbf{y}_{j,t}|\mathbf{x}_{j,t})$ - описывает закон измерения, линейную связь между состоянием объекта и его измерениями с помощью матрицы преобразования $\mathbf{B}$ с добавлением случайного шума с ковариационной матрицей $\mathbf{\Sigma}_j$; $p(\mathbf{x}_{j,1})$ - априорное распределение скрытых параметров в начальный момент времени со средними значениями $\boldsymbol{\mu}_{j,0}$ и шумом с ковариационной матрицей $\mathbf{\Gamma}_{j,0}$.

Нахождения искомых оценок представляет собой задачу максимизации совместной апостериорной вероятности скрытых характеристик объектов X_t при известных прошлых наблюдениях характеристик объектов Y_t и измеренных характеристиках Z_t неизвестных объектов, обнаруженных на данном кадра:

$$\max_{X_t} p(X_t|Y_t, Z_t). \tag{2}$$

При допущениях, что объекты не взаимодействуют друг с другом и двигаются независимо, будем предполагать независимость состояний, скрытых характеристик объектов друг от друга и независимость состояния объекта от измеренных ранее характеристик других объектов. Таким образом, совместную апостериорную вероятность скрытых характеристик объектов X_t можно записать в факторизованном по объектам виде:

$$p(X_t|Y_t, Z_t) = \prod_{j=1}^{N} p(\mathbf{x}_{j,t}|Y_t, Z_t) = \prod_{j=1}^{N} p(\mathbf{x}_{j,t}|\mathbf{y}_{j,1}, \ldots, \mathbf{y}_{j,t-1}, Z_t), \tag{3}$$

Введем реалистичное допущение, что состояние объекта $\mathbf{x}_{j,t}$ статистически связано только с параметрами одного обнаруженного на кадре объекта, но заранее неизвестного, причем от характеристик данного объекта не зависят состояния других объектов $\mathbf{x}_{k,t}$, $k \neq j$. Тогда, с учетом этого допущения, выражение (3) может быть переписано в следующем виде:

$$\begin{aligned} p(X_t|Y_t, Z_t) &= \prod_{j=1}^{N} \left\{ \sum_{i=1}^{M_t} a_{i,j}\, p(\mathbf{x}_{j,t}|\mathbf{y}_{j,1}, \ldots, \mathbf{y}_{j,t-1}, \mathbf{z}_{i,t}) \right\} = \\ &= \prod_{j=1}^{N} \left\{ p(\mathbf{x}_{j,t}|\mathbf{y}_{j,1}, \ldots, \mathbf{y}_{j,t-1}, \sum_{i=1}^{M_t} a_{i,j}\mathbf{z}_{i,t}) \right\} = \\ &= \prod_{j=1}^{N} \left\{ p(\mathbf{x}_{j,t}|\mathbf{y}_{j,1}, \ldots, \mathbf{y}_{j,t-1}, \mathbf{y}_{j,t}) \right\}, \end{aligned} \tag{4}$$

при условиях:

$$\begin{aligned} & a_{i,j} \in \{0, 1\},\ i \in \overline{1, M_t},\ j \in \overline{1, N} \\ & \sum_{i=1}^{M_t} a_{i,j} = 1,\ j \in \overline{1, N} \\ & \sum_{j=1}^{N} a_{i,j} = 1,\ i \in \overline{1, M_t}, \end{aligned} \tag{5}$$

где

$$\mathbf{y}_{j,t} = \sum_{i=1}^{M_t} a_{i,j}\mathbf{z}_{i,t}. \tag{6}$$

- выбранное измерение для объекта j. Условия (5) определяют соответствие одного измерения $\mathbf{z}_{i,t}$ одному состоянию $\mathbf{x}_{j,t}$, т.е. значение коэффициента $a_{i,j} = 1$, если $\mathbf{x}_{j,t}$ зависит от наблюдения $\mathbf{z}_{i,t}$, иначе $a_{i,j} = 0$. Таким образом, коэффициенты $a_{i,j}$ определяют выбор апостериорной вероятности состояния объекта, с которой связано измерение $\mathbf{z}_{i,t}$.

Для оценивания состояния объектов X_t (2) оптимальный выбор измерений $\mathbf{y}_{j,t}$ для каждого объекта j среди всех $\mathbf{z}_{i,t}$ может быть реализован путем максимизации функции правдоподобия апостериорного распределения. Пусть для текущего кадра t имеется некоторая оценка состояния $\mathbf{x}_{j,t} = \widetilde{\mathbf{x}}_{j,t}$, которая может быть получена на основе состояния $\mathbf{x}_{j,t-1}$ на предыдущем кадре, используя модель ЛДС (1). Тогда для объекта

j в соответствии с (4) определим функцию правдоподобия $L_{\widetilde{\mathbf{x}}_{j,t}}$:

$$L_{\widetilde{\mathbf{x}}_{j,t}}(\mathbf{y}_{j,1},\ldots,\mathbf{y}_{j,t-1},\mathbf{z}_{i,t}) = p(\widetilde{\mathbf{x}}_{j,t}|\mathbf{y}_{j,1},\ldots,\mathbf{y}_{j,t-1},\mathbf{z}_{i,t}) \tag{7}$$

Выражение для максимизации функции правдоподобия $L_{\widetilde{\mathbf{x}}_{j,t}}$ объекта j по измерениям $\mathbf{z}_{i,t}$ с учетом (6) имеет вид

$$\begin{aligned}\max_{\mathbf{z}_{i,t}\in Z_t} L_{\widetilde{\mathbf{x}}_{j,t}}(\mathbf{y}_{j,1},\ldots,\mathbf{y}_{j,t-1},\mathbf{z}_{i,t}) &= \max_{a_{i,j}} L_{\widetilde{\mathbf{x}}_{j,t}}(\mathbf{y}_{j,1},\ldots,\mathbf{y}_{j,t-1},\sum_{i=1}^{M_t} a_{i,j}\mathbf{z}_{i,t}) = \\ &= \max_{a_{i,j}} \sum_{i=1}^{M_t} a_{i,j} L_{\widetilde{\mathbf{x}}_{j,t}}(\mathbf{y}_{j,1},\ldots,\mathbf{y}_{j,t-1},\mathbf{z}_{i,t})\end{aligned} \tag{8}$$

при ограничениях (5).

Для нахождения совместных назначений измерений Z_t всем объектам X_t необходимо оптимизировать функцию правдоподобия совместного апостериорного распределения (4). Используя выражение (8) из (4) получаем для оценок на кадре $X_t = \widetilde{X}_t$

$$\max_{a_{i,j}} p(\widetilde{X}_t|Y_t,Z_t) = \max_{a_{i,j}} \prod_{j=1}^{N} \left\{ \sum_{i=1}^{M_t} a_{i,j}\, L_{\widetilde{\mathbf{x}}_{j,t}}(\mathbf{y}_{j,1},\ldots,\mathbf{y}_{j,t-1},\mathbf{z}_{i,t}) \right\}, \tag{9}$$

при ограничениях (5).

Преобразуем задачу (9) к следующему виду:

$$\begin{aligned}&\max_{a_{i,j}} \log p(\widetilde{X}_t|Y_t,Z_t) = \max_{a_{i,j}} \sum_{j=1}^{N} \log \left(\sum_{i=1}^{M_t} a_{i,j}\, L_{\widetilde{\mathbf{x}}_{j,t}}(\mathbf{y}_{j,1},\ldots,\mathbf{y}_{j,t-1},\mathbf{z}_{i,t}) \right) = \\ &= \max_{a_{i,j}} \sum_{j=1}^{N}\sum_{i=1}^{M_t} a_{i,j}\, \log L_{\widetilde{\mathbf{x}}_{j,t}}(\mathbf{y}_{j,1},\ldots,\mathbf{y}_{j,t-1},\mathbf{z}_{i,t}) = \min_{a_{i,j}} \sum_{j=1}^{N}\sum_{i=1}^{M_t} a_{i,j}\, \mathbf{C}_{i,j}, \\ &\text{с ограничениями (5),}\end{aligned} \tag{10}$$

где $\mathbf{C}_{i,j} = -\log L_{\widetilde{\mathbf{x}}_{j,t}}(\mathbf{y}_{j,1},\ldots,\mathbf{y}_{j,t-1},\mathbf{z}_{i,t})$ - матрица штрафов за назначение объекту j наблюдения i с измеренными параметрами $\mathbf{z}_{i,t}$, $a_{i,j}$ - матрица назначения наблюдения i объекту j.

С учетом решения задачи (10) в виде (6) и выражения (4) задача (2) примет следующий вид:

$$\begin{aligned}&\max_{X_t} \log p(X_t|Y_t,Z_t) = \max_{\mathbf{x}_{1,t},\ldots,\mathbf{x}_{N,t}} \log \prod_{j=1}^{N} \left\{ p(\mathbf{x}_{j,t}|\mathbf{y}_{j,1},\ldots,\mathbf{y}_{j,t-1},\mathbf{y}_{j,t}) \right\} = \\ &= \sum_{j=1}^{N} \max_{\mathbf{x}_{j,t}} \log p(\mathbf{x}_{j,t}|\mathbf{y}_{j,1},\ldots,\mathbf{y}_{j,t-1},\mathbf{y}_{i,t}).\end{aligned} \tag{11}$$

Отсюда следует, что на основе решения задачи о назначениях (10), задача совместного

оценивания (2) состояний множества X_t объектов, приводит к подзадачам индивидуальной оценки скрытых характеристик каждого объекта j:

$$\mathbf{x}^*_{j,t} = \arg\max_{\mathbf{x}_{j,t}} \log p(\mathbf{x}_{j,t}|\mathbf{y}_{j,1}, \ldots, \mathbf{y}_{j,t-1}, \mathbf{y}_{j,t}), \tag{12}$$

где $\mathbf{x}^*_{j,t}$ - оценка состояния и скрытых характеристик объекта j на текущем кадре t.

Такими образом, исходная задача (2) совместного оценивания скрытых характеристик множества объектов свелась к трем последовательным задачам: задаче получения первоначальной оценки $\mathbf{x}_{j,t} = \widetilde{\mathbf{x}}_{j,t}$ для текущего кадра t по предыдущему на основе модели ЛДС (1), задаче (10) о назначениях между множеством наблюдений Z_t на данном кадре и множеством объектов с характеристиками $\widetilde{X}_t$ и задаче (12) уточнения оценки состояния объекта $\mathbf{x}^*_{j,t}$ по сопоставленному измерению (6), найденному на предыдущем этапе (10).

Для определения модели объекта зададим его скрытые характеристики в виде $\mathbf{x} = (u, v, s, r, \dot{u}, \dot{v}, \dot{s})^T$, где u и v соответствуют горизонтальной и вертикальной координате пикселя центра объекта, s и r представляют площадь и соотношение сторон прямоугольника, ограничивающего объект, соответственно, $\dot{u}, \dot{v}, \dot{s}$ - скорости изменения соответствующих параметров.

В работе ставится задача сопровождения множества объектов в онлайн режиме, нахождении оценок характеристик X_t для всех объектов N путем решения задач (10) и (12).

4 Оценивание параметров в задаче слежения

4.1 Алгоритм предсказания траекторий. Фильтр Калмана.

Задача сопровождения может быть описана в рамках задачи фильтрации сигнала [18, 22]. Используя обозначения для измерений характеристик объекта $\mathbf{y}_1, \ldots, \mathbf{y}_t$ и скрытого состояния объекта $\mathbf{x}_1, \ldots, \mathbf{x}_t$ Задача оценивания для объекта j решается путем максимизации апостериорного распределения (12). Далее в рамках данной главы индекс j опускается для лаконичности, подразумевается, что все излагается для заданного объекта j.

Исходя из постановки задачи используются гауссовские линейные модели измерения и изменения состояния объекта, совместное апостериорное распределение в модели ЛДС так же является нормальным:

$$p(\mathbf{x}_t|\mathbf{y}_1, \ldots, \mathbf{y}_t) = N(\mathbf{x}_t|\boldsymbol{\mu}_t, \mathbf{V}_t). \tag{13}$$

Пусть известно распределение $p(\mathbf{x}_{t-1}|\mathbf{y}_1, \ldots, \mathbf{y}_{t-1})$ в момент времени $t-1$. Тогда прогноз значения $\mathbf{x}_t$ вычисляется следующим образом:

$$\begin{aligned} p(\mathbf{x}_t|\mathbf{y}_1, \ldots, \mathbf{y}_{t-1}) &= \int p(\mathbf{x}_t, \mathbf{x}_{t-1}|\mathbf{y}_1, \ldots, \mathbf{y}_{t-1}) d\mathbf{x}_{t-1} = \\ &= \int p(\mathbf{x}_t|\mathbf{x}_{t-1}) p(\mathbf{x}_{t-1}|\mathbf{y}_1, \ldots, \mathbf{y}_{t-1}) d\mathbf{x}_{t-1} = \\ &= \int N(\mathbf{x}_t|\mathbf{A}\mathbf{x}_{t-1}, \boldsymbol{\Gamma}) N(\mathbf{x}_{t-1}|\boldsymbol{\mu}_{t-1}, \mathbf{V}_{t-1}) d\mathbf{x}_{t-1} = \\ &= N(\mathbf{x}_t|\mathbf{A}\boldsymbol{\mu}_{t-1}, \boldsymbol{\Gamma} + \mathbf{A}\mathbf{V}_{t-1}\mathbf{A}^T), \end{aligned} \tag{14}$$

Таким образом, прогноз значения $\mathbf{x}_t$ задается формулой:

$$\begin{aligned} &p(\mathbf{x}_t|\mathbf{y}_1, \ldots, \mathbf{y}_{t-1}) = N(\mathbf{x}_t|\widetilde{\boldsymbol{\mu}}_t, \widetilde{\mathbf{V}}_t) \\ &\widetilde{\boldsymbol{\mu}}_t = \mathbf{A}\boldsymbol{\mu}_{t-1}, \\ &\widetilde{\mathbf{V}}_t = \boldsymbol{\Gamma} + \mathbf{A}\mathbf{V}_{t-1}\mathbf{A}^T. \end{aligned} \tag{15}$$

Когда значение $\mathbf{y}_t$ становится известным, прогноз значения $\mathbf{x}_t$ может быть уточнен:

$$\begin{aligned} p(\mathbf{x}_t|\mathbf{y}_1, \ldots, \mathbf{y}_t) &= \frac{p(\mathbf{x}_t, \mathbf{y}_1, \ldots, \mathbf{y}_t)}{p(\mathbf{y}_1, \ldots, \mathbf{y}_t)} = \frac{p(\mathbf{y}_t|\mathbf{x}_t) p(\mathbf{x}_t|\mathbf{y}_1, \ldots, \mathbf{y}_{t-1})}{p(\mathbf{y}_t|\mathbf{y}_1, \ldots, \mathbf{y}_{t-1})} \propto \\ &\propto N(\mathbf{x}_t|\boldsymbol{\mu}_t, \mathbf{V}_t) = N(\mathbf{y}_t|\mathbf{B}\mathbf{x}_t, \boldsymbol{\Sigma}) N(\mathbf{x}_t|\widetilde{\boldsymbol{\mu}}_t, \widetilde{\mathbf{V}}_t). \end{aligned} \tag{16}$$

Решая задачу (12) с использованием (16), находим решение фильтра Калмана в виде:

$$
\begin{aligned}
\boldsymbol{\mu}_t &= \widetilde{\boldsymbol{\mu}}_t + \mathbf{K}_t(\mathbf{y}_t - \mathbf{B}\widetilde{\boldsymbol{\mu}}_t), \\
\mathbf{V}_t &= (\mathbf{I} - \mathbf{K}_t\mathbf{B})\widetilde{\mathbf{V}}_t, \\
\mathbf{K}_t &= \widetilde{\mathbf{V}}_t\mathbf{B}^T(\mathbf{B}\widetilde{\mathbf{V}}_t\mathbf{B}^T + \boldsymbol{\Sigma})^{-1}.
\end{aligned}
\tag{17}
$$

Таким образом, фильтр Калмана [18] задается формулами (15), (17) и состоит из двух шагов. Пусть имеется априорное распределение $p(\mathbf{x}_{t-1}|\mathbf{y}_1, \ldots, \mathbf{y}_{t-1})$. На первом шаге осуществляется прогноз значения $\mathbf{x}_t$ по формулам (15). На втором шаге, происходит коррекция прогноза для $\mathbf{x}_t$ с учетом новой информации по формулам (17). В случае если рассматриваему объекту не было назначено ни одно наблюдение на текущем кадре, коррекция предсказания характеристик объекта не производится и сразу происходит переход к следующему кадру.

Фильтра Калмана является вероятностной моделью со скрытыми переменными, оценка параметров модели производится методом максимизации неполного правдоподобия, например, с использованием EM алгоритма [23].

4.2 Задача о назначениях. Матрица стоимости.

Задача о назначениях (10) возникает в связи с необходимостью уточнения параметров объектов (12) и прогнозов их траекторий с помощью фильтра Калмана (шаг 1) (15) в задаче слежения за множеством объектов. После обнаружения и локализации объектов на кадре требуется их сопоставить с ранее найденными объектами вдоль соответствующих траекторий, чтобы для каждой из траектории произвести уточнение прогноза (шаг 2 фильтра Калмана (17) на основе сопоставленного наблюдения. Уточненные прогнозы сохраняются в состоянии объекта на траектории, после чего происходит переход на следующий кадр.

Общая схема решения задачи сопровождения множества объектов на кадре t представлена на рисунке 1.

Задача о назначениях может быть эффективно решена с использованием существующих алгоритмов [5, 6]. Сложность заключается в выборе матрицы стоимости $\mathbf{C}$, которая бы приводила к нахождению корректного соответствия между траекторией объекта и локализацией данного объекта в следующем кадре видеопотока. В соответствии с (10) элементы матрицы стоимости $\mathbf{C}_{i,j}$ определяются на основе логарифма функции правдоподобия апостериорной функции вероятности (16). В качестве оценки состояния $\mathbf{x}_{j,t}$ объекта на текущем кадре для функции правдоподобия будем использовать прогноз фильтра Калмана (15) с предыдущего кадра $\mathbf{x}_{j,t} = \widetilde{\boldsymbol{\mu}}_{j,t}$. В таком случае из (16) для логарифма функции правдоподобия апостериорной функции вероятности выражением для элемента матрицы стоимости $\mathbf{C}_{i,j}$ имеет вид:

$$
\begin{aligned}
\mathbf{C}_{i,j} &= \log N(\mathbf{z}_{i,t}|\mathbf{B}\widetilde{\boldsymbol{\mu}}_{j,t}, \boldsymbol{\Sigma}) = \\
&= \frac{1}{2}(\mathbf{z}_{i,t} - \mathbf{B}\widetilde{\boldsymbol{\mu}}_{j,t})^T\boldsymbol{\Sigma}^{-1}(\mathbf{z}_{i,t} - \mathbf{B}\widetilde{\boldsymbol{\mu}}_{j,t}) - \frac{m}{2}\log 2\pi - \frac{1}{2}\log|\boldsymbol{\Sigma}|,
\end{aligned}
\tag{18}
$$

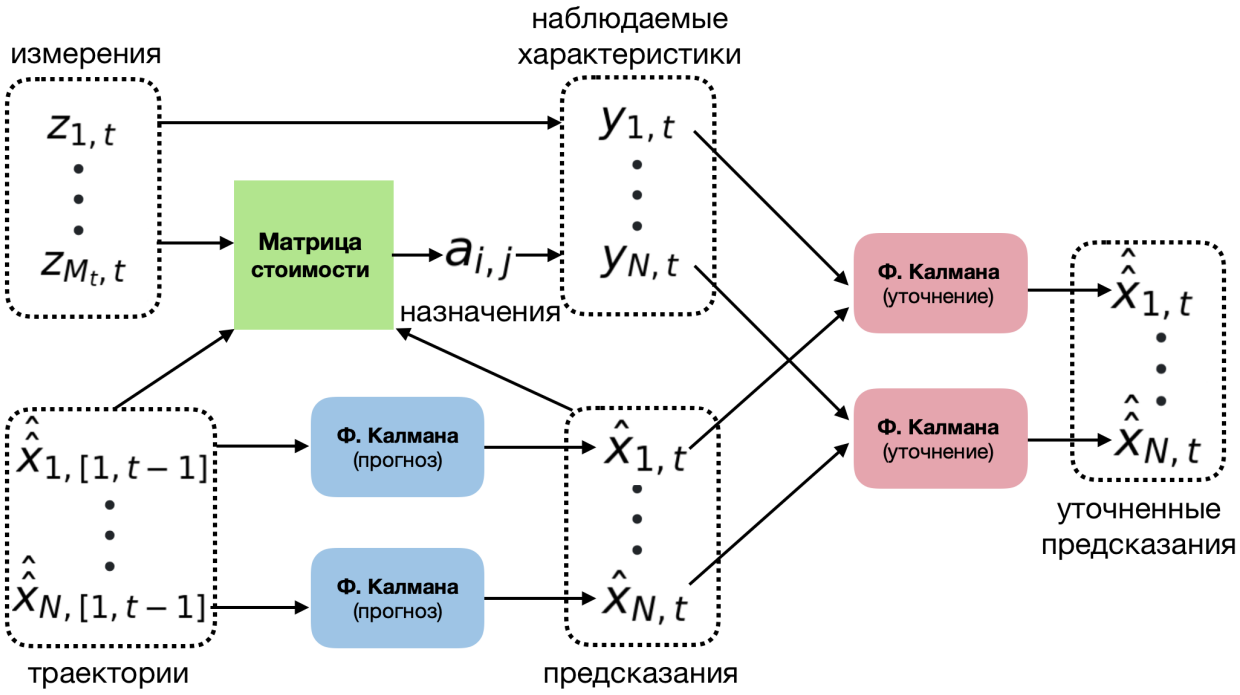

Рис. 1: Алгоритм слежения за множеством объектов для кадра t

где m размерность вектора $\mathbf{x}$. Таким образом, матрица стоимости назначения определяется близостью характеристик измерения неизвестного объекта и скрытых параметров объекта на траектории. Однако в сложных условиях наблюдения, таких как пересечения траекторий разных объектов или срыв слежения из-за пропусков измерений критерий (18) не достаточно хорошо работает.

В методе [5, 6] предлагается использовать дескрипторы изображения объекта, основанные на выделении характерных признаков целевых объектов. Определяется отображение пространства изображений, содержащих объект исследуемого класса, в пространство дескрипторов — векторное пространство фиксированной размерности, на котором определена операция скалярного произведения. Таким образом, изображению $\mathbf{I}$ ставится в соответствие дескриптор $\mathbf{d} = h(\mathbf{I}|\boldsymbol{\theta})$. Используя дескрипторы можно производить сравнение разных изображений объекта в разные моменты времени вдоль траектории. Дескрипторы объекта от кадра к кадру подвержены существенно нелинейным искажениям, модель которых неизвестна, поэтому в качестве модели оценок векторов дескрипторов будем использовать все множество дескрипторов, полученных по предыдущим оценкам положений объекта вдоль траектории. Величину отклонения измерения дескриптора объекта на текущем кадре от дескрипторов, построенных вдоль траектории объекта ранее будем определять по порядковой статистике обратных мер схожести измерений на текущем кадре с множеством предыдущих измерений для данного объекта.

Добавим в выражение для элемента матрицы стоимости $\mathbf{C}_{i,j}$ соответствующую

добавку:

$$\tilde{\mathbf{C}}_{i,j} = \mathbf{C}_{i,j} + \|\boldsymbol{\ell}_{i,j}\|^2, \tag{19}$$

где $\boldsymbol{\ell}_{i,j}$ – вектор размерности $t-1$, компоненты которого определяются как $\boldsymbol{\ell}_{i,j,\tau} = \ell(\mathbf{d}_{i,t}, \mathbf{d}_{j,\tau})$ где ℓ – нормированное расстояние между дескрипторами (обратная мера схожести дескрипторов), $\mathbf{d}_{i,t}, \mathbf{d}_{j,\tau}$ – дескрипторы наблюдений $\mathbf{z}_{j,t}$ и $\mathbf{y}_{j,\tau}$ соответственно.

Добавление критерия отнесения неизвестного объекта к объекту на траектории по дескриптору задает модель ре-идентификации. Она позволяет минимизировать ошибку назначения в сложных условиях наблюдения, таких как пересечения траекторий разных объектов или срыв слежения из за пропусков измерений и не назначений измерения объекту из-за плохой работы детектора объектов на промежуточных кадрах.

В качестве отображения $h(\cdot|\boldsymbol{\theta})$ выступает сверточная нейронная сеть архитектуры ResNet18 [16] с L_2-нормализацией выхода сети. Блочная схема архитектуры нейронной сети изображена на рисунке 2.

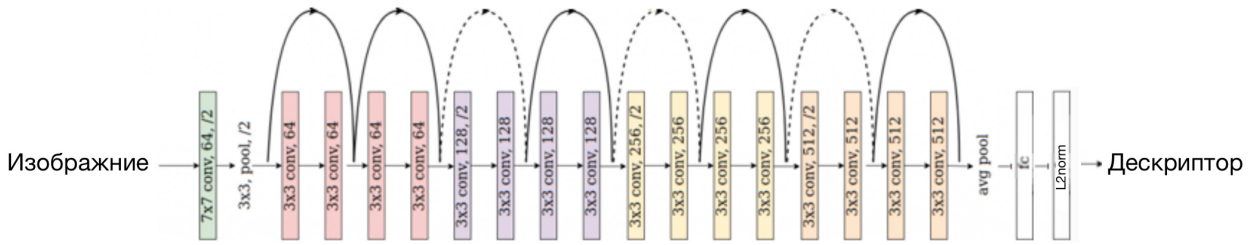

Рис. 2: Схема слоев архитектуры Resnet18 с L_2-нормализацией выхода

Обучение параметров $\boldsymbol{\theta}$ нейронной сети производится в рамках задачи классификации объектов исследуемого класса. Для обучения используется Cosine Softmax Cross-Entropy функции потерь [17]:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{\exp\left(\kappa \cdot \tilde{\mathbf{W}}_{c_i}\mathbf{d}_i\right)}{\sum\limits_{j=1}^{N}\exp\left(\kappa \cdot \tilde{\mathbf{W}}_{j}\mathbf{d}_j\right)}, \;\; \tilde{\mathbf{W}} = \frac{\mathbf{W}}{\|\mathbf{W}\|_2}, \;\; \|\mathbf{d}_i\|_2 = 1, \tag{20}$$

где $\mathbf{W}$ – матрица весов, $\mathbf{d}_i = h(\mathbf{I}_i|\boldsymbol{\theta})$ – декриптор i-ого объекта выборки, c_i – метка класса для i-ого объекта выборки, N – число объектов в выборке, $\tilde{\mathbf{W}}_j$ и $\tilde{\mathbf{W}}_{c_i}$ – j-ый и c_i-ый столбцы $\tilde{\mathbf{W}}$ соответственно, κ – параметр масштабирования (обратная температура). Таким образом, нейронная сеть задает отображение пространства изображений объектов в пространство признаковых представлений (дескрипторов), представляющее собой многомерную единичную сферу в силу наличия нормировки выхода сети. Данная функция потерь позволяет при обучении добиться большой внутриклассовых косинусной схожести и малой межклассовой соответственно, что позволяет производить ре-идентификацию объектов на основе косинусной схожести.

Определение добавки к элементу матрицы стоимости $\mathbf{C}$ сводится к задаче реидентификации между объектом, требующим назначения, и объектами траектории-

кандидата. В базовом методе Deep-SORT [10] элемент матрицы стоимости, отвечающий назначению наблюдения i некоторого объекта, присутствующего на изображении $\mathbf{I}_i$, траектории объекта j, включающей изображения $\mathbf{I}_{j,1}, \ldots, \mathbf{I}_{j,\tau}$ определяется следующим образом. Всем изображениям ставятся в соответствие дескрипторы, выделенные нейронной сетью $h(\cdot|\boldsymbol{\theta})$:

$$\mathbf{d}_i = h(\mathbf{I}_i|\boldsymbol{\theta});\ \mathbf{d}_{j,t} = h(\mathbf{I}_{j,t}|\boldsymbol{\theta}),\ t \in \overline{1,\tau}. \tag{21}$$

По выделенным дескрипторам вычисляются косинусные сходства между наблюдением i неизвестного объекта и всеми объектами траектории j, наибольшее из определенных косинусных сходств домножается на -1 и вычитается из единицы. Полученное значение принимается за значение добавки (19) к элементу матрицы стоимости, которая принимает вид:

$$\tilde{\mathbf{C}}_{i,j} = \mathbf{C}_{i,j} + r\left[1 - \max\left\{\frac{\langle \mathbf{d}_i, \mathbf{d}_{j,1}\rangle}{\|\mathbf{d}_i\|\,\|\mathbf{d}_{j,1}\|}, \ldots, \frac{\langle \mathbf{d}_i, \mathbf{d}_{j,\tau}\rangle}{\|\mathbf{d}_i\|\,\|\mathbf{d}_{j,\tau}\|}\right\}\right]^2, \tag{22}$$

где r - масштабный множитель. Таким образом, в составлении элемента матрицы участвуют все объекты из соответствующей траектории.

Отличительной особенностью предложенного подхода является метод составления матрицы стоимости. Для того, чтобы избежать недостатков базового подхода Deep-SORT, предлагается ограничить число объектов-кандидатов из траектории, используемых для ре-идентификации. Это позволит ограничить алгоритмическую сложность составления матрицы и исключить зависимость сложности от числа объектов в траектории. Данное ограничение позволит существенно увеличить вычислительную эффективность алгоритма по сравнению с базовым, где исходный подход являлся основной причиной низкой скорости обработки видеопотока по сравнению с методом SORT.

Введение ограничения на число объектов из траектории, используемых при реидентификации, потенциально приводит к уменьшению обучающей способности алгоритма и соответственно к ухудшению качества работы. Для решения данной проблемы предлагается производить тщательный отбор объектов-кандидатов для ре-идентификации.

Поскольку, не все объекты в траектории равноправны, предлагается производить отбор или взвешивание объектов из траектории для подсчета общей стоимости назначения. Взвешивание производится в два этапа. Первоначально каждому изображению $\mathbf{I}_{j,t}$, $t = 1, \ldots, \tau$, соответсвующему наблюдению $\mathbf{y}_{j,t}$ из траектории j длины τ, ставится в соответствие показатель "качества" $q_{j,t} \in [0,1]$, данное отображение $q : \mathbf{I}_{j,t} \mapsto q_{j,t}$ задается методом оценки качества. После чего производится сортировка элементов в порядке убывания показателя "качества", то есть ищется перестановка $p(1)p(2)\ldots p(\tau)$ с индексами $\{1, \ldots, \tau\}$ такая, что показатели "качества" расположились бы в порядке невозрастания:

$$q_{j,p(1)} \geq q_{j,p(2)} \geq \cdots \geq q_{j,p(\tau)}. \tag{23}$$

Отбор осуществляется выбором K кандидатов $\mathbf{y}_{j,p(1)}, \ldots, \mathbf{y}_{j,p(K)}$ с наибольшим значением показателя "качества" и исключением остальных объектов. Таким образом, вес $w_{j,t}$ каждого из наблюдений $\mathbf{y}_{j,t}$, $t = 1, \ldots, \tau$ в траектории определяется как:

$$w_{j,t} = \begin{cases} 1 & \text{если } q_{j,t} \leq Q_{(K)} \\ 0 & \text{иначе} \end{cases}, \tag{24}$$

где $Q_{(K)}$ − реализация $K-$ой порядковой статистики для выборки $\{q_{j,1}, \ldots, q_{j,\tau}\}$.

В работе предлагается применить к данной задаче метод оценки биометрического показателя "качества" для выбора наиболее валидных объектов из траектории для реидентификации. Оценка качества является новым подходом в задаче сопровождения множества объектов, в литературе по задачам сопровождения объектов не встречался подход подсчета матрицы стоимости, использующий предварительный отбор объектов на основе меры "качества".

Подход основан на идее отбора объектов из траектории, которые бы были наиболее близки к распредлению данных, на котором обучалась нейронная сеть, используемая для выделения векторного представления объектов. То есть производится отбор объектов-кандидатов наиболее близких к своему классу. Таким образом, это решение позволит избежать некорректного построения векторных представлений, исключив объекты, не соответсвующие распределению данных.

Стратегия выбора фиксированного числа объектов с наибольшим качеством позволяет применить различные подходы к оценке качества.

Наиболее продуктивным является подход к оценке показателя "качества", прелоложенный в [20]. Оценка качества сводится к обучению с учителем в задаче регрессии, где ответы получены в результате асессорской разметки данных, а объектами обучения являются векторные представления, выделенные из изображений сверточной нейронной сетью.

Таким образом, заданной выборке изображений $\{I_i\}_{i=1}^{\tilde{T}}$ ставится в соотвествие выборка $G = \{\mathbf{g}_i\}_{i=1}^{\tilde{T}}$, где $\mathbf{g}_i = g(I_i|\tilde{\boldsymbol{\theta}})$, $g : \mathbb{I} \to \mathbb{R}^n$ − сверточная нейронная сеть VGGFace [24], обученная для решения задачи верификации.

По выборке G и ответам $\hat{\mathbf{y}}$, полученным на основе асессорской разметки, обучается метод опорных векторов. Алгоритм в общем виде имеет вид $a(\mathbf{g}) = \langle \mathbf{g}, \mathbf{w} \rangle - w_0$, обучение соответсвует задаче оптимизации:

$$Q_\epsilon(a, G) = \sum_{i=1}^{\tilde{T}} |\langle \mathbf{g}_i, \mathbf{w} \rangle - w_0 - \hat{\mathbf{y}}_i|_\epsilon + \tau \langle \mathbf{w}, \mathbf{w} \rangle^2 \to \min_{\mathbf{w}, w_0}, \tag{25}$$

где $|z|_\epsilon = \max\{0, |z| - \epsilon\}$, $\epsilon > 0$, τ − коэффициент регуляризации.

Метод опорных векторов в данной задаче регрессии признаковых описаний объектов на асессорскую оценку качества позволяет производить оценку качества произвольного объекта с последующим успешным обором объектов для задачи распознования.

Наряду со стандартным подходом к оценки биометрического качества предложим альтернативу, не требующую дополнительных вычислений.

Мера соответствие объекта распредлению данных, на которых обучался алгоритм выделения дескрипторов, коррелирует с вероятностью наличия объекта исследуемого класса в данной области. Таким образом, в качестве простой оценки меры соответствия объекта распределению может быть использвана уверенность (confidence) детектора для соответсвующей объекту области. Допущение наличия между данными величинами порядковой связи позволяет отбирать для ре-идентификации объекты, соответсвующие локализациям с наибольшим значением уверенности детектора.

Таким образом, добавка к элемент матрицы стоимости определяется через наибольшее косинусне сходство между соответствующей характеристикой, требующей назначения, и отобранными объектами из соответсвующей траектории. Общая схема алгоритма вычисления добавки к элементу $\mathbf{C}_{i,j}$ матрицы стоимости представлена на рисунке 3.

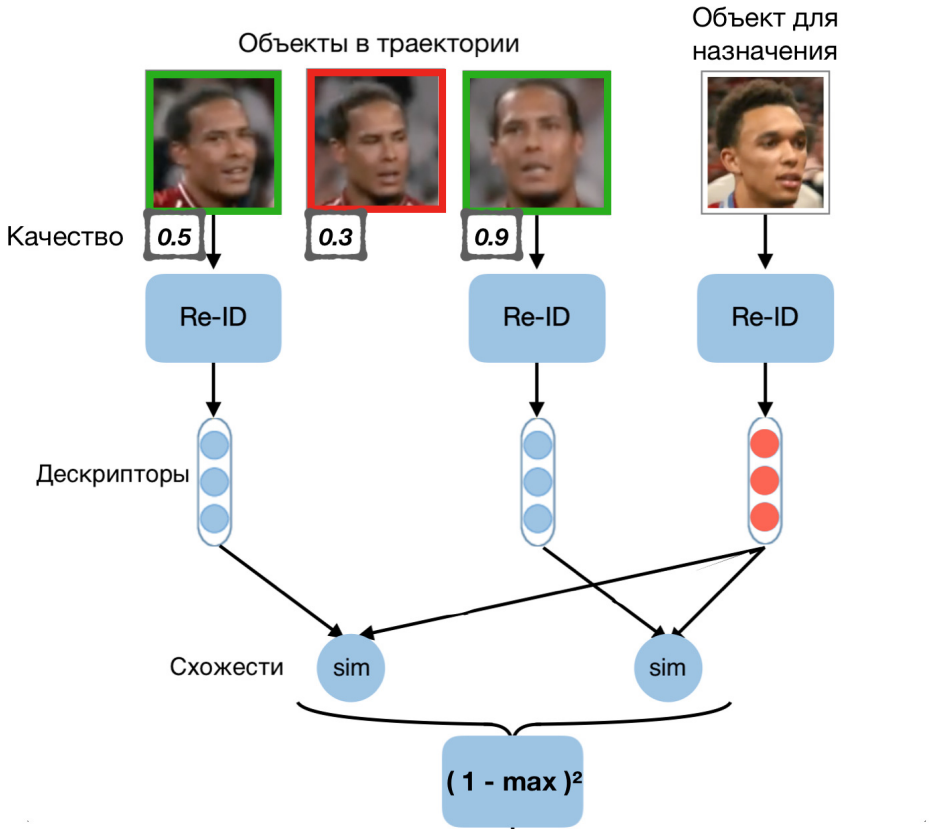

Рис. 3: Алгоритм вычисления добавки $\|\boldsymbol{\ell}_{i,j}\|^2$ к элементу $\mathbf{C}_{i,j}$ матрицы стоимости ($K = 2$)

Для решения задачи о назначениях (10) с модифицированной матрицей стоимости (22) используется алгоритм Джонкера-Волгенанта [6]. Для контроля корректности назначений вводятся пороги на компоненты $\mathbf{C}_{i,j}$ и $\|\boldsymbol{\ell}_{i,j}\|^2$ матрицы стоимости (22). Превышение порога исключает соответсвующее назначение.

Порог на компоненту $\mathbf{C}_{i,j}$ определяется через максимального значение расстояния Махаланобиса между наблюдением $\mathbf{z}_{i,t}$ и параметрами распределения наблюдаемых

характеристик $\mathbf{y}_{j,t}$ при оценке скрытых характерестик $\mathbf{x}_{j,t} = \tilde{\boldsymbol{\mu}}_{j,t}$, заданной предсказанием фильтра Калмана для траектории j.

$$\mathcal{D}(\mathbf{z}_{i,t}) = \sqrt{(\mathbf{z}_{i,t} - \mathbf{B}\tilde{\boldsymbol{\mu}}_{j,t})^T \, \boldsymbol{\Sigma}^{-1} (\mathbf{z}_{i,t} - \mathbf{B}\tilde{\boldsymbol{\mu}}_{j,t})}, \tag{26}$$

где $\mathbf{z}_{i,t}$ – наблюдаемая характеристика, $\mathbf{B}\tilde{\boldsymbol{\mu}}_{j,t}$ и $\boldsymbol{\Sigma}$ – среднее и матрица ковариации распределения на $\mathbf{y}_{j,t}$ при оценке $\mathbf{x}_{j,t} = \tilde{\boldsymbol{\mu}}_{j,t}$ через предсказние фильтра Калмана (15). Распределение лежит в классе нормальных, поэтому квадрат расстояния Махаланобиса имеет распредление хи-кварат с числом степеней свободы, равным размерности характеристики. Тогда, рассматривая расстояние Махаланобиса в качестве статистики для проверки гипотезы о принадлежности обнаруженного объекта сопоставленной траектории, можно успешно проверять данную гипотезу для уровня значимости α, отменяя некорректнные сопоставления, отвечающие маловероятным продлениям траектории. Таким образом, порог максимального значения расстояния Махаланобиса определяется через значение квантильной функции распредления хи-квадрат в точке $1 - \alpha_0$, α_0 – установленный уровень значимости.

Порог на добавку $\|\boldsymbol{\ell}_{i,j}\|^2$ определяется в соответствии с процедурой ре-идентификации и задается через порог минимального значения наибольшей косинусной схожести между дескриптором $\mathbf{d}_i$ наблюдения и дескрипторами $\mathbf{d}_{j,p(1)}, \ldots, \mathbf{d}_{j,p(K)}$ отобранных объектов траектории j в (22). Значение ниже порогового соответствует ситуации, когда все отобранные для ре-идентификации объекты из траектории имеют схожесть с наблюдением текущего кадра ниже пороговой, то есть ни один из них не проходит успешно процедуру ре-идентификации с данным наблюдением.

4.3 Задача инициализации и удаления траекторий.

При появлении объекта в кадре и окончательном покидании кадра должна создаваться новая траектория и заканчиваться существующая соответственно. Для создания новой траектории достаточно наблюдения произвольной области, включающей объект и имеющей пересечение с локализациями объектов существующих траекторий ниже пороговой. Вектор скрытых переменных инициализируется параметрами прямоугольника локализации с нулевой скоростью. Поскольку на данном этапе отсутствует информация о скорости движения объекта в кадре, дисперсия скорости задается некоторым большим значением, определяя существующую неопределенность. Для окончательного подтверждения существования траектории предполагаемый объект должен быть наблюдаем на протяжении некоторого числа кадров и подчиняться заданной модели движения.

Удаление траекторий происходит, если соответствующий объект не наблюдается на протяжении некоторого числа последовательных кадров. Процедура удаления не позволяет числу траекторий неограниченно расти, уменьшая нагрузку на систему, и исключает ошибки, связанные с ре-идентификацией объектов чрезмерно долго отсутствующих в кадре.

5 Вычислительный эксперимент

В качестве данных использовались две выборки изображений видеоряда MOT20-01 и MOT20-02, данные представляют собой последовательность кадров, снятых камерой видеонаблюдения в людном месте с двух разных ракурсов. Большинство объектов двигалось по направлению к камере либо поперек кадра. Данные описаны в таблице 1.

Выборка	FPS	Плотность объектов на кадр	Длина	Число траекторий
MOT20-01	25	42.1	429	90
MOT20-02	25	72.7	2782	296

Таблица 1: Описание выборок

На основе данных выборок произведено обнаружение лиц с использованием двух различных фиксированный моделей локализации SSD и RetinaNet. Модели существенно отличаются по сложности, число обучаемых параметров приведено в таблице . Каждая из моделей обучена в задаче локализации лиц на базе изображений, собранной сотрудниками кафедры. Примеры кадров из выборок, с проделонной локализацией лиц, приведены на рисунке 4. Таким образом, вычислительный эксперимент производится на четырех наборах данных. Истинные траектории извлекались из исходной разметки выборок данных путем сопоставления построенных локализаций с локализациями, предоставленными к выборкам.

Архитектура	Число параметров
SSD	4 M
RetinaNet	45 M

Таблица 2: Число обучаемых параметров детекторов

В качестве нейронной сети для задачи ре-идентификации использовалась архитектура ResNet18, обученная в рамках задачи многоклассовой классификации с использованием Cosine Softmax Cross-Entropy функции потерь по выборке изображений, собранной сотрудниками кафедры. Данная модель фиксировалась и применялась для ре-идентификации как в предложенном подходе, так и в базовом. Аналогичным образом фиксировались максимальное время отсутвия объекта в кадре, которое полагалось равным 30 кадрам.

Произвелено сравнение предложенного метода слежения за множеством объектов с использованием отбора объектов для ре-идентификации с базовым подходом Deep-SORT. Рассматривались два метода отбора признаков на основе оценки качества. Первый метод использовал в качестве оценки качества уверенность (confidence) модели локализации, такой подход назван Deep-Conf-SORT. Второй метод соответствовал алгоритму оценки качества, предложенному в [20], на данный подход определен как Deep-QA-SORT. Стратегия отбора объектов-кандидатов для ре-идентифкации из траектории заключалась в выборе $K = 5$ лучших по качеству объектов.

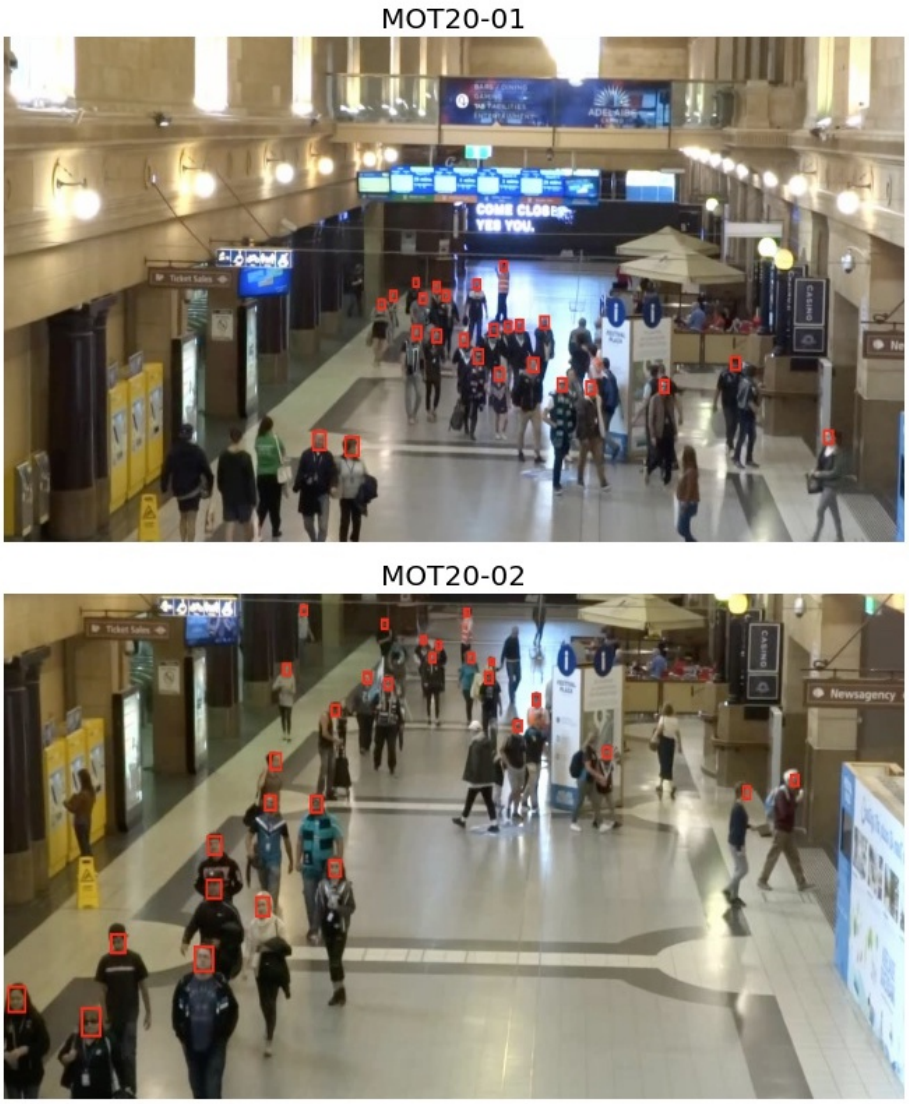

Рис. 4: Примеры локализаций лиц на кадрах из выборок MOT20-01, MOT20-02

В качестве показателей качества использовался набор стандартных мер качества в задаче слежения за множеством объектов:

- Precision $= \frac{\mathrm{TP}}{\mathrm{TP+FP}}$ – точность, где FP – число ошибок I рода (ложная тревога), TP – число верно наблюдаемых объектов;
- Recall $= \frac{\mathrm{TP}}{\mathrm{TP+FN}}$ – полнота, где FN – число ошибок II рода (пропуск цели), TP – число верно наблюдаемых объектов;
- IDS – суммарное число некорректных переключений при продлении траекторий;
- Hz – частота работы (в кадрах в секунду).

Решение о наличии ошибки I или II рода принималось на основе порога отношения пересечения к объединению (IoU) предсказанной области и истинной областью локализации объекта. Области заданы прямоугольниками, ограничивающими объекты. Сопоставление между истинной разметкой и результатом работы алгоритма слежения

получено следующим образом: если истинная траектория и предсказанная траектория были сопоставлены на предыдущем кадре, то они были сопоставлены и на такущем кадре, если значение IoU между данными прямоугольниками выше порогового, равного 0.5. Такое сопоставление производилось даже в случае существования другой предсказнной траектории, мера IoU между прямоугольником которой и истинным прямоугольником выше. После сопоставлений для предсказанных траекторий с предыдущего шага производилось назначение оставшихся истинных траекторий с оставшимися предсказанными с таким же порогом IoU, равным 0.5. Окончательно области истинных траекторий, которые не были сопоставлены с предсказанными областями, определялись как ошибки II рода (FN), области предсказнных траекторий, которые не были поставлены в соответствие областям истинных траекторий, принимались за ошибки I рода (FP). Оставшиеся успешно сопоставленные области соответствовали верно наблюдаемым объектам (TP).

Результаты работы методов на выборках описаны в таблице 3 и таблице 3 соответственно. Исходя из полученных результатов, утверждается, что оба предложенных метода имеют большую вычислительную эффективность по сравнению с базовым подходом, причем преложенные методы не приводят к ухудшению качества в терминах показателей качества Precision и Recall и позволяют уменьшить число некорректных переключений при продлении траекторий (IDS). Относительный прирост качества в терминах IDS выше в случае детектора SSD, чем RetinaNet. Данное наблюдение может быть объяснено тем, что модель RetinaNet имеет большую сложность с точки зрения числа параметров, что приводит к меньшей подверженности к совершению ошибок I рода (обнаружению ложных объектов), на отсеивание которых нацелен преложенный метод оценки качества в задаче ре-идентификации.

Детектор	Метод слежения	Precision	Recall	IDS	Hz
SSD	Deep-SORT	0.851	**0.910**	828	22.6
	Deep-Conf-SORT	**0.860**	0.904	778	**31.8**
	Deep-QA-SORT	0.855	0.907	**751**	31.8*
RetinaNet	Deep-SORT	0.887	**0.945**	543	22.4
	Deep-Conf-SORT	0.891	0.941	533	**31.4**
	Deep-QA-SORT	**0.896**	0.939	**526**	31.4*

Таблица 3: Результаты для выборки MOT20-02

* Время работы алгоритма оценки качества не включено в суммарное время, поскольку оценка качества объектов применяется в системах распознавания лиц на постоянной основе.

Детектор	Метод слежения	Precision	Recall	IDS	Hz
SSD	Deep-SORT	0.847	**0.895**	203	39.2
	Deep-Conf-SORT	**0.850**	0.890	202	**44.7**
	Deep-QA-SORT	0.848	0.892	**195**	44.7*
RetinaNet	Deep-SORT	0.871	**0.922**	168	40.9
	Deep-Conf-SORT	**0.875**	0.919	166	**46.0**
	Deep-QA-SORT	0.873	0.921	**162**	46.0*

Таблица 4: Результаты для выборки MOT20-01

* Время работы алгоритма оценки качества не включено в суммарное время, поскольку оценка качества объектов применяется в системах распознавания лиц на постоянной основе.

Так же производится вычислительный эксперимент, направленный на исследование зависимости качества в теминах IDS меры качества, определяющей число некорректных переключений при продлении траектории, от числа отобранных для ре-идентифкации объектов-кандидатов. Результаты эксперимента представлены на рисунке 5. Исследованные функциональные зависимости обладают минимумом, оптимальное число отбираемых объектов K в данном эксперименте варируется в диапазоне $5-6$. Наличие точки оптимума для числа отбираемых объектов свидетельствует о обоснованной применимости такого отбора и определяет число отбираемых объектов как гипермапараметр, требующий оптимизации.

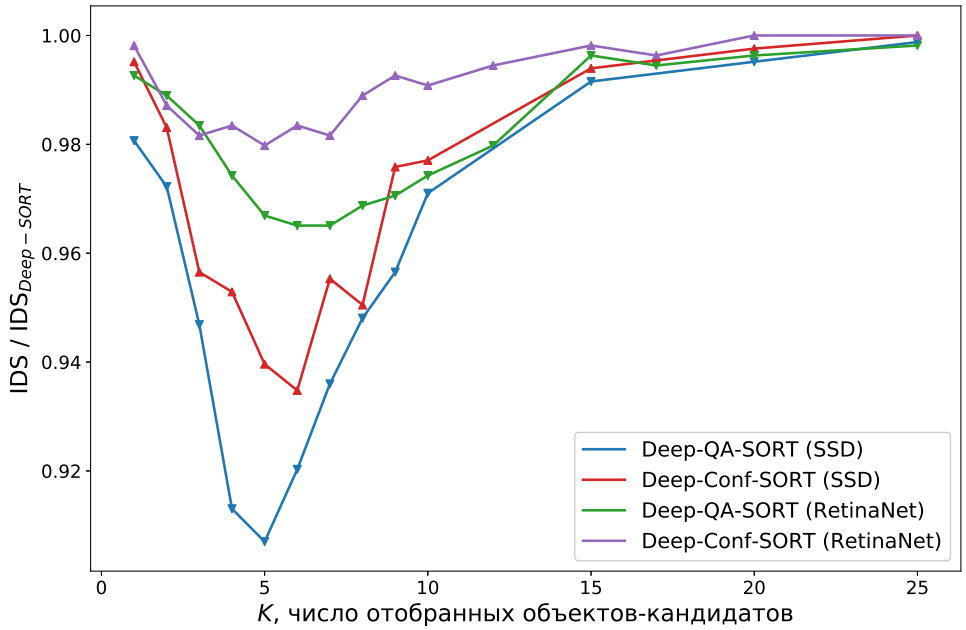

Рис. 5: Зависимость отношения IDS метода к IDS базового метода от числа отобранных кандидатов K

6 Заключение

В работе была поставлена задача слежения за множеством объектов в видеопотоке, были изученные существующие подходы и выявлены их недостатки. На основе анализа базового метода Deep-SORT был разработан метод слежения с использованием процедуры отбора объектов из траектории для ре-идентификации. В качестве стратегии отбора признаков рассматривался подход выбора K лучших объектов на основе оценки качества объекта. В дополнении к существующим методам оценки качества предложен альтернативный подход, использующий в качестве оценки уверенность детектора. Проведен вычислительный эксперимент по оценке разработанных методов в сравнении с базовыми. Проанализировано влияние предложенных подходов к отбору объектов для ре-идентификации на число некорректных переключений при продлении траекторий для моделей детекторов различной сложности. Показана высокая вычислительная эффективность и наличие прироста качества преложенных методов по сравнению с базовым методом Deep-SORT.

Список литературы

[1] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, Simple online and realtime tracking, 2016 IEEE Int. Conf. on Image Proces., 2016, pp. 3464–3468.

[2] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, and S. Reed. Ssd: Single shot multibox detector. ECCV, 2016.

[3] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You only look once: Unified, real-time object detection. CVPR, 2016.

[4] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar. Focal loss for dense object detection. ICCV, 2017.

[5] H. W. Kuhn and Bryn Yaw. The Hungarian method for the assignment problem, Naval Res. Logist. Quart, 1955. pp 83-97.

[6] R. Jonker and A. Volgenant. A Shortest Augmenting Path Algorithm for Dense and Sparse Linear Assignment Problems, Computing, 1987. vol. 38, pp. 325-340 IEEE.

[7] Bo Wu and Ram Nevatia. Tracking of multiple, partially occluded humans based on static body part detection. CVPR, 2006.

[8] Keni Bernardin and Rainer Stiefelhagen. Evaluating multiple object tracking performance: the clear mot metrics. Journal on Image and Video Processing, 2008.

[9] Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi. Performance measures and a data set for multi-target, multi-camera tracking. European Conference on Computer Vision, 2016. pp 17-35, Springer.

[10] Nicolai Wojke, Alex Bewley, and Dietrich Paulus. Simple online and realtime tracking with a deep association metric. ICIP, 2017. pp 3645–3649, IEEE.

[11] Laura Leal-Taixe, Anton Milan, Ian Reid, Stefan Roth, and Konrad Schindler. Motchallenge 2015: Towards a benchmark for multi-target tracking. arXiv preprint arXiv:1504.01942, 2015.

[12] Anton Milan, Laura Leal-Taixe, Ian Reid, Stefan Roth, and Konrad Schindler. Mot16: A benchmark for multi-object tracking. arXiv preprint arXiv:1603.00831, 2016.

[13] Patrick Dendorfer, Hamid Rezatofighi, Anton Milan, Javen Shi, Daniel Cremers, Ian Reid, Stefan Roth, Konrad Schindler, and Laura Leal-Taixe. Cvpr19 tracking and detection challenge: How crowded can it get?, 2019.

[14] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. CVPR, 2012. pp 3354–3361, IEEE.

[15] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The kitti dataset. The International Journal of Robotics Research, 32(11): pp. 1231–1237, 2013.

[16] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. CVPR, 2016.

[17] N. Wojke and A. Bewley, Deep Cosine Metric Learning for Person Re-identification. WACV, 2018.

[18] R.E. Kalman, A New Approach to Linear Filtering and Prediction Problems, Transactions of the ASME–Journal of Basic Engineering. vol. 82, Series D, pp. 35-45, 1960.

[19] A. Abaza, M. A. Harrison, and T. Bourlai. Quality metrics for practical face recognition. IAPR ICPR, 2012.

[20] L. Best-Rowden and A. K. Jain. Learning Face Image Qual- ity From Human Assessments. IEEE Transactions on Infor- mation Forensics and Security, 2018.

[21] J. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor and J. Vandewalle, Least Squares Support Vector Machines. Singapore:World Scientific, 2002.

[22] A. L. Barker, D. E. Brown and W. N. Martin, Bayesian estimation and the Kalman filter. Comput. Math. Appl., 1995. vol. 30, no. 10, pp. 55-77.

[23] A. Dempster, N. M. Laird and D. Rubin, Maximum-likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, 1977. vol. B39, pp. 1-38.

[24] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. BMVC, 2015.