**ACCV**

**FRiS-based dynamical keypoint model for tracking and matching**

1. Introduction

Введение

Задача слежения за априори неизвестными объектами (трекинг объектов) – представляет значительный интерес в области компьютерного зрения. Под априори неизвестным объектом подразумевается, что на вход алгоритму трассировки подается только область, содержащая объект в первом кадре последовательности. Задача слежения за объектом состоит в том, чтобы, зная область, содержащую объект на первом кадре, определить ее местоположение на последующих кадрах последовательности.

При трекинге объектов алгоритм сталкивается с рядом трудностей: низкая контрастность между объектом и фоном, перекрытие объекта интереса другими объектами, временное исчезновения объекта из кадра, изменения освещения. Дополнительные сложности возникают при изменении формы, угла обзора или текстуры объекта. Примером таких изменений может служить съемка футбольного матча, когда одного игрока могут снимать с разных ракурсов и в разном масштабе. Алгоритмы трекинга [LC,MS,CamS] не в состоянии определить координаты объекта в случае его исчезновения из поля зрения и появления в другом месте кадра. Данная проблема решается сочетанием отслеживания и обнаружения объекта [TLD]. В этом случае алгоритм трекинга работает, когда смещения объекта между кадрами незначительны, а детектор работает при потере объекта трекером.

Ключевую роль в задачах отслеживания играет представление (модель) объекта в алгоритме. Модели объекта в виде совокупности частей [], более устойчивы к вышеперечисленным трудностям, так как локальные изменения действуют только на определенные части объекта. При этом объект стабильно обнаруживается даже при частичном перекрытии, когда ряд фрагментов объекта отсутствуют в кадре. С другой стороны, объект может описываться не только тем набором частей, которые присутствовали на первом кадре последовательности, но другими частями, видимыми под иным ракурсом. Поэтому выглядит перспективным использование изменяющейся во времени (динамической) [] модели объекта. Построение динамической модели объекта сводится к решению задачи машинного обучения, что приводит к ряду проблем, связанных с построением решающих правил. Типичным примером может служить частичное перекрытие двух объектов, в результате которого модель начинает соответствовать перекрывающему объекту. Несмотря на это динамическая модель позволяет надежнее отслеживать объект при изменениях формы и угла обзора.

В алгоритме CMT [] предлагается использовать представление объекта в виде набора особых точек []. Особые точки, найденные алгоритмами BRISK [] и SIFT [] устойчивы к масштабированию, повороту и незначительным изменениям освещения, это делает их использование в качестве частей объекта перспективным. При этом модель в алгоритме CMT состоит из статической части, включающей в себя координаты особых точек, их взаимное расположение, дескрипторы и динамической части, описывающей объект на кадре координатами центра, его масштабом и углом поворота. При изменении формы и угла обзора набор особых точек, описывающий объект, также меняется, но данные изменения никак не отражаются в модели. В данной работе предлагается модификация алгоритма CMT с использованием динамически изменяемого набора особых точек. Алгоритм CMT был выбран за сочетание точности, скорости работы и компактного представления модели.

Основной вклад работы заключается в следующем:

1. Предложен алгоритм рейтинга особых точек в модели, позволяющий оценивать степень доверия к данной точке. Рейтинг сочетает в себе время жизни точки в последовательности, расстояние до геометрического центра объекта и степень уникальности дескриптора точки.
2. Алгоритм добавления и удаления точек в зависимости от рейтинга доверия. Удаляются случайные точки и точки фона. При изменении набора особых точек переоцениваются все параметры модели.
3. Рассмотрено построение решающих правил для добавления новых точек в модель на основе FRiS []. Применение FRiS позволяет оценивать сходство добавляемой точки с точками объекта в конкуренции с точками фона. При фиксированных весах характеристик FRiS функция измеряется сходство в абсолютной шкале, что позволяет улучшить методы анализа данных.

Было оценено качество работы предложенного алгоритма на данных CMT и на данных эталонного теста VOT []. В результате сравнения было показано, что предложенные модификации позволили улучшить качество трекинга объекта алгоритмом CMT на XX%.

2. Related Works

Обзор литературы

На данный момент существует большое количество алгоритмов трекинга объектов по последовательности изображений. Последние сравнительные данные по существующим алгоритмам трекинга приведены в обзорной работе [2] и в результатах соревнования VOT (Visual Object Tracking 2015) []. Существующие алгоритмы оцениваются по качеству работы, при этом зачастую игнорируется вычислительная сложность и размер модели объекта в памяти. Существуют два подхода к построению алгоритма трекинга: построение обобщенной модели изображения объекта интереса [] и построение разделяющего правила между объектом и фоном []. В работе [TLD] предложено обучение в реальном времени разделяющего классификатора с использованием схемы P-N обучения [] и пространственных ограничений. Объект интереса представлен набором бинарных шаблонов [], максимально разделяющим фрагменты объекта и фрагменты фона. В данном алгоритме сочетается детектирование фрагментов объекта с оценкой оптического потока. В работе [STRuck] предложено использовать структурную SVM (SSVM) [] для построения разделяющего правила между объектом и фоном. Для каждого кадра решаются задачи проверки соответствия признаков SSVM внутри окон интереса и обучение SSVM на обновленном положении объекта. В работе [KCF] предложено обучать SVM в пространстве Фурье. Функция правдоподобия использует гистограммы цветов и ориентированных градиентов.

В последнее время популярность набирает использование сверточных искусственных нейронных сетей для трекинга объектов []. Основным достоинством нейронных сетей является то, что они используют автоматически выбранные признаки для разделения объекта и фона. В работе [FCNT] предлагается использование значений активаций со слоев предобученной сверточной нейронной сети VGG [] в качестве вектора признаков и тренировка дополнительной нейронной сети для поиска объекта. В работе [DeepSRDCF] предложено использовать значения активаций нейронной сети в качестве пространства признаков для разделительного корреляционного фильтра DCF []. В работе [MDNet] предложено обучать нейронную сеть на последовательности изображений объекта для построения решающего правила.

В работах [] используется совместно обобщающая и разделяющая модель объекта для достижения большей точности трекинга. В качестве основы нашей работы был выбран алгоритм CMT, использующий представление объекта в виде набора особых точек BRISK. Модель в алгоритме состоит из центра масс, координат особых точек относительно центра масс, данных об их взаимном расположении и дескрипторов. Данный алгоритм использует совместную работу трекера и детектора. Трекер определяет положение точек объекта на новом кадре, используя вычисление прямого и обратного оптического потока[], что позволяет более точно определить новое положение точек объекта. На основании нового местоположения точек вычисляется смещение объекта, изменение его размера и угла поворота. Для вычисления центра объекта и удаления выбросов применяется кластеризация. Затем детектор сравнивает дескрипторы кадра с дескрипторами в модели объекта и уточняет результат работы трекера. На основании полученных данных об изменении модели, вычисляется новое положение области объекта на кадре.

Функция конкурентного сходства (FRiS) [] представляет собой удобный инструмент для решения ряда задач машинного обучения, таких как построение решающего правила [FRiS -Stolp], кластеризация [FRiS Tax] и другие [FRiS UC].

1. **FRiS-based dynamical keypoint model**

Даны последовательность кадров входного видео и область ,ограничивающая объект на первом кадре последовательности. Задача алгоритма состоит в оценке местоположения этой области на каждом последующем кадре или определении того, что объект не видим на текущем кадре. За основу данной работы взят алгоритм СМТ, в котором для оценки местоположения объекта используются такие параметры, как его центр масс , масштаб и угол поворота .

Объект наблюдения представлен центром масс , набором особых точек объекта , где – координата каждой его особой точки относительно центра масс. Для оценки изменения масштаба и поворота вычисляются вектора разности координат для каждой пары точек объекта, в модели они представлены матрицей длин этих векторов и матрицей углов в полярных координатах. Каждая особая точка описывается дескриптором, дескрипторы точек объекта составляют множество . Дескрипторы особых точек первого кадра последовательности, не принадлежащих области , образуют множество дескрипторов точек фона . Для быстроты вычислений были использованы двоичные дескрипторы BRISK размерности 512.

* 1. Рейтинги.

В процессе слежения за объектом набор особых точек, описывающих объект может измениться достаточно сильно – старые точки могут перестать присутствовать на объекте, могут появиться новые. В соответствии с изменением объекта необходимо изменить и его модель. Для определения того, какие точки нужно добавить, а какие удалить применяется система рейтинга точек, определяющая степень доверия к каждой точке.

При начальной инициализации каждой точке объекта присваивается некоторый вес , где – множество весов всех точек объекта. С каждого кадра извлекаются особые точки, их дескрипторы сопоставляются с дескрипторами модели объекта. Если установлено присутствие точки объекта на новом кадре последовательности, то считается что такая точка более удачна для трекинга, и ее вес увеличивается. По достижении максимального веса вес точки более не меняется, так как считается, что такая точка достаточно хорошо описывает объект. Если точку не удалось найти на новом кадре, то ее вес уменьшается. Таким образом удается избавиться от «случайных» точек, вызванных шумом или неточностью выделения области объекта, и точек, переставших быть видимыми. Выбор начального веса и критерий добавления новой точки в модель определяются с использованием FRiS.

* 1. FRiS решающее правило

Система рейтинга точек должна учитывать степень доверия к каждой из точек объекта. В качестве меры степени доверия к точке используется функция, оценивающая степень сходства данной точки с точками фона и объекта в абсолютной шкале. Данная функция конкурентного сходства (FRiS) обладает следующими свойствами:

1. Свойство локальности: мера сходства зависит не от характера распределения всех точек, а от особенностей распределения точек в окрестности заданной точки.
2. Свойство нормированности: функция принимает максимальное значение 1 при совпадении точки с некоторой точкой объекта, минус 1 при совпадении с некоторой точкой фона. В остальных случаях функция принимает значения от минус 1 до 1.
3. Свойство антисимметричности: значения сходства точки с фоном и объектом связаны свойством антисимметричности, при одинаковых расстояниях до ближайшей точки фона и объекта, точка будет в равной степени похожа на фон и объект, а функция будет принимать значение 0.
4. Свойство инвариантности: значения меры сходства сохраняются при афинных преобразованиях.

Таким образом, FRiS позволяет определить степень схожести точки с ближайшей точкой фона в конкуренции с ближайшей точкой объекта. Данная мера сходства воспроизводит механизм оценки сходства, используемый человеком, и инвариантна к распределению особых точек.

Начальный вес точек объекта рассчитывается следующим образом: для каждой точки вычисляется расстояние между дескриптором этой точки и ближайшим дескриптором фона и расстояние между дескриптором точки и ближайшим дескриптором объекта . Затем вычисляется значение функции конкурентного сходства для каждой точки . В результате, вес , где F – линейная функция, со следующими свойствами: , где – максимально возможный вес точки, а – некоторый натуральный коэффициент.

Процесс добавления точек в модель также использует FRiS. Для начала для каждой точки кадра определяется, принадлежит ли она области, определенной как объект. Если это так, то вычисляется значение FRiS-функции Если , где – заданный пороговые значения, и расстояние между этой точкой и точкой в пространстве больше некоторого порога , то считается, что эта точка кадра новая и ее следует добавить в модель с некоторым начальным весом .

* 1. Обновление модели

После удаления и добавления точек необходимо пересчитать модель объекта. Исходя из нового набора точек заново вычисляется центр масс объекта и расположение точек относительно него. Заменяются соответствующие строки и столбцы в матрицах A и L.

1. **Эксперименты**
2. **Методология оценки []**

*Тестовый набор, критерий сравнения (можно сдуть с смт). Тут думаю логично подкрепить описание всего этого примерами(картинками)*

1. Влияние параметров на качество трекинга

*Здесь про прогон алгоритма на разных значениях порогов? Возникает вопрос какой вариант алгоритма считать финальным, что прогонять и на какой базе? Здесь видимо стоит рассмотреть, когда алгоритм дает лучшие результаты. Картинки-графики всех этих прогонов сюда еще нужно(тут по-честному все или самые красивые?))*

1. Сравнение с CMT на их БД
2. Сравнение VOT

Заключение