УДК 004.932.2 URL: http://ptsj.ru/catalog/iemim/sta/989.html

ОСОБЕННОСТИ КАСКАДНОЙ МОДЕЛИ ДИЗАЙН-ПРОЕКТИРОВАНИЯ НА ПРИМЕРЕ РАЗРАБОТКИ МНОГОФУНКЦИОНАЛЬНОГО УСТРОЙСТВА (МИНИ-ПОГРУЗЧИКА)

П.В. Малышев pasha_malyshev01@mail.ru

МГТУ им. Н.Э. Баумана, Москва, Россия

Рассмотрены алгоритмы обнаружения и отслеживания объектов на видеопоследовательности, выбраны наиболее подходящие методы для распознавания и выделения контуров объектов в видеопотоке без применения машинного обучения. Предложен алгоритм создания простой системы видеонаблюдения. Многие подходы отслеживания объектов сочетают отслеживание, обучение и обнаружение. Алгоритм отслеживания, или так называемый трекер, следует за объектом от кадра к кадру. Алгоритм обнаружения, детектор, локализует все особенности, которые наблюдались до сих пор, и при необходимости корректирует трекер. В процессе обучения ошибки детектора оцениваются и его работа корректируется во избежание этих ошибок в будущем. Отметим, что обучение обнаружению объектов обычно выполняется при условии, что все обучающие примеры помечены. В данной работе рассмотрены алгоритмы и методы, которые можно применить при создании детектора и трекера.

Ключевые слова: алгоритм, обнаружение объектов, отслеживание объектов, видеонаблюдение, системы видеонаблюдения, обработка изображений, модель фона, компьютерное зрение

Введение. Одной из ключевых задач в системах с компьютерным зрением является идентификация объектов, попавших в поле зрения камеры видеонаблюдения [1]. Задача компьютерного зрения заключается в том, чтобы научить компьютер понимать, что запечатлено на изображении. Для компьютера, так же как и для человека, зрение служит источником семантической, качественной и метрической информации о трехмерном мире.

В системах с компьютерным зрением графическая информация обычно поступает в виде оцифрованного видеопотока, представляющего собой сплошную последовательность отдельных изображений. Видеоданные могут поступать с камеры видеонаблюдения, которая ведет съемку в реальном времени, либо из заранее отснятого видеофайла.

Оценка сложности. Компьютерное зрение представляет собой сложную задачу по ряду причин. Изображения одних и тех же объектов очень сильно различаются в разных условиях. При этом создается целый ряд сложностей:

 ISSN 2541-8009

– в зависимости от точки наблюдения объекты выглядят по-разному;

– освещение сильно влияет на внешний вид объекта;

– размер объектов одного и того же класса может варьироваться;

– объекты могут претерпевать существенные деформации;

– часто объекты перекрывают друг друга, тогда видимой является только часть объекта;

– объекты могут быть замаскированы (например, маскировка живых объектов);

– в движении объекты претерпевают деформацию и смазываются [2];

– объекты одного и того же класса могут быть слишком разными;

– классы объектов могут зависеть не только от самих объектов, но и от контекста, в котором они находятся. Если по одному объекту нельзя определить, чем он является, необходимо изучить всю сцену в целом;

– локальная неоднозначность — разные объекты на большом расстоянии могут выглядеть одинаково.

При этом в изображении объекта также присутствует множество подсказок, позволяющих правильно идентифицировать его. Задача компьютерного зрения заключается в том, чтобы интерпретировать их. Основные подсказки, содержащиеся в изображении:

– цвет (зачастую по цвету можно распознать, чем является объект [3]);

– освещенность (форму объекта можно восстановить по градации серого);

– отбрасываемые тени — по теням можно определить положение объекта);

– группировка (объекты можно объединять в классы по близости друг к другу или взаимному расположению);

– линейная перспектива и уменьшение объектов с расстоянием (позволяют определить более точное расположение объекта в пространстве);

– текстурный градиент;

– упорядочивание по глубине (перекрытие объектов позволяет определить, какие из них находятся ближе, а какие — дальше);

– туман и фокусировка (по размытости в тумане можно судить о расстоянии до объекта).

Одно и то же изображение может давать несколько различных интерпретаций, одного изображения для полноценного анализа, как правило, недостаточно. Даже все подсказки, присутствующие в одном изображении, не могут решить задачу компьютерного зрения однозначно. Чтобы правильно интерпретировать сцены, нужны априорные знания структуре и свойствах реального мира.

Обнаружение объектов. Видеонаблюдение включает в себя решение следующих задач:

– обнаружение объектов интереса в видео;

– отслеживание их движения в последующих кадрах.

Обнаружение объектов — это задача локализации объектов на входном изображении. Определения объекта также могут варьироваться. Объектом может быть один экземпляр или целый класс объектов, при этом сам объект определяется его местоположением и протяженностью в одном кадре. Методы обнаружения, как правило, основаны на локальных особенностях изображения или скользящем окне. Обычно задача обнаружения объекта сводится к решению следующих подзадач:

– обнаружение признаков;

– распознавание признаков;

– распознавание объектов.

Выбор объекта интереса может осуществляться несколькими способами:

– в простейшем случае — ручная инициализация объекта;

– использование детектора объектов — детектора «пешехода» и детектора «лица» [4];

– выделение движущихся объектов путем сегментации видео [5].

На выделении движущихся объектов основано большинство существующих систем видеонаблюдения. Они опираются на два следующих упрощения. Первое упрощение заключается в том, что камера является стационарной, т. е. закреплена и смотрит в определенное место. В случае, когда камера стационарна, возникает второе упрощение: фон будет мало изменяться между кадрами.

Существует два основных метода выделения объектов интереса в видео. Простой вариант представляет собой метод обнаружения с использованием ограничивающего прямоугольника. Более сложный вариант подразумевает использование пиксельной маски [6].

Современные системы компьютерного зрения могут обнаруживать множество различных объектов, попавших в поле зрения видеокамеры, в том числе различающихся по цветам. Для решения этой задачи определяют цветовые маски для каждого интересующего объекта. Затем полученные маски складывают и накладывают на исходное изображение. В результате получают цветовой фильтр, позволяющий выделять искомую группу объектов.

Отслеживание объектов. При поиске смещения объектов на видео (video tracking) часто применяют простые алгоритмы. Исходный кадр, на котором присутствует рассматриваемый объект, а также один из последующих кадров видеопоследовательности можно представить в виде матриц. Тогда смещение на изображении можно распознать путем поэлементного вычитания из второй матрицы первой. Поскольку на изображениях часто присут-

 ISSN 2541-8009

ствует шум, при учете всех пикселей, для которых такое вычитание дает некоторую величину, отличную от нуля, в качестве искомого объекта будет выделено практически все изображение. Поэтому вводят некоторое пороговое значение. Разницу для каждого пикселя сравнивают с порогом и, если разница больше порога, то пиксель принадлежит объекту. Если разница меньше порога, то пиксель принадлежит фону.

Однако данный алгоритм будет работать только для статичного фона (когда камера закреплена и не двигается) и одинакового освещения. Для изменяющегося фона строят модели изменения яркости по времени. Каждый пиксель изображения рассматривается независимо от остальных. Для каждого пикселя фона можно построить график изменения яркости (цвета) от времени. Полученную модель необходимо параметризовать. Если на новом кадре яркость пикселя не удовлетворяет модели фона, значит, этот пиксель принадлежит движущемуся объекту.

В случае статичной камеры, когда на каждом кадре выделяются пятна потенциальных объектов, задача слежения сводится к задаче сопоставления или ассоциации данных [7]. Необходимо определить, какому объекту, какое пятно соответствует на каждом кадре. Существует огромное количество как вероятностных, так и детерминированных методов, которые позволяют решать такую задачу сопоставления.

Детерминированные методы по сути представляют собой методы перебора, которые определяют соответствие гипотез некоторым ограничениям.

Для каждого пятна на кадрах возможны три варианта:

1) пятно может быть новым объектом;

2) пятно может быть новым положением старого объекта (если его можно сопоставить старому следу);

3) пятно может исчезнуть, в таком случае след необходимо завершить (пропадание объекта).

Самая простая стратегия отслеживания объекта — сопоставление с ближайшим. Пусть имеется точка, которую необходимо отследить. На следующем кадре имеется несколько пятен. Из них со следом сопоставляется тот, которых находится к следу ближе всего. Такая система хорошо работает во многих случаях, но оказывается малопригодной при большом количестве объектов.

Часто при использовании систем видеонаблюдения наблюдаемых объектов не так много, и решение сводится к простому перебору вариантов. В подобных случаях выдвигают гипотезы, каким объектам первого кадра какие пятна последующего кадра соответствуют. Затем перебирают все ограничения и фильтруют соответствие. Иногда эти операции не выполняют для двух со-

седних кадров, вместо этого количество кадров увеличивают и находят ограничения на нескольких кадрах сразу, т. е. выдвигают сразу несколько гипотез, а затем перебирают все варианты.

На практике в простых системах видеонаблюдения бывает достаточно учитывать несколько основных признаков — то, что объект недалеко сдвинулся и то, что скорость его движения не очень высока.

Базовый метод, реализованный в большинстве систем, выглядит следующим образом:

– обучается модель фона;

– для каждого нового кадра происходит вычитание фона;

– обрабатывается маска (фильтрация, морфология);

– выделяются связанные компоненты;

– вычисляется соответствие объектов предыдущим кадрам по близости и скорости;

– если какие-то отслеживаемые пятна не соответствуют старым следам, инициализируются новые следы;

– обновляется фон всюду, где не были обнаружены объекты.

Данный метод работает только тогда, когда фон статичен. В остальных случаях (например, в системах автопилотирования автомобилей) необходимо строить модель объекта и искать на следующем кадре похожий объект.

Резюмируя вышесказанное, можно выделить следующую схему отслеживания некоторой области:

– инициализируются модели;

– выбирается пространство признаков;

– модель представляется в выбранном пространстве;

– определяется начальное положение объекта;

– в последующем кадре в окрестности предыдущего положения происходит поиск нового положения объекта, такого, чтобы старая модель была наиболее похожей на новую;

– положение, в котором модели максимально точно совпадают, считается новым положением объекта.

Таким образом, имеется некоторая функция качества, которую необходимо максимизировать. Данные операции повторяют для каждого последующего кадра.

Алгоритм построения простой системы видеонаблюдения. ***Вычитание фона.*** Пусть имеется изображение чистого фона, не содержащего объекты интереса. Берут изображение фона, на котором находится интересующий объект, и из него вычитают первое изображение. Полученная разница будет являться предполагаемым искомым объектом.

 ISSN 2541-8009

Поскольку на изображениях часто присутствуем шум, при учете всех пикселей, для которых такое вычитание дает некоторую величину, отличную от нуля, в качестве объекта интереса будет выделено практически все изображение. Поэтому вводят некоторое пороговое значение. Разницу для каждого пикселя сравнивают с порогом, и, если разница больше порога, то пиксель принадлежит переднему плану. Если разница меньше порога, то пиксель принадлежит фону. Результатом такого сравнения является маска переднего плана, которая содержит искомый объект.

После этого на полученном бинарном изображении выделяют объекты переднего плана как связанные компоненты. В системах видеонаблюдения такие связанные компоненты представляют собой пятна. Если присутствует шум, то маску переднего шума необходимо очистить с использованием одного из методов фильтрации (например, математической морфологии или медианной фильтрации).

Данный система видеонаблюдения будет хорошо работать только в искусственных условиях, когда освещение одинаково, нет подвижных теней и фон не изменяется. При этом тень объекта будет также выделяться, как и сам отслеживаемый объект.

В реальных условиях фон существенно меняется с течением времени. В этом случае строят модели изменения яркости по времени. Каждый пиксель изображения i рассматривают независимо от остальных. Для каждого пикселя фона можно построить график изменения яркости (цвета) x_i от времени t. Полученную модель необходимо параметризовать. Если на новом кадре яркость пикселя не удовлетворяет модели фона, значит, этот пиксель принадлежит движущемуся объекту.

Таким образом, обновленный алгоритм видеонаблюдения выглядит следующим образом:

– вначале инициализируется модель фона;

– на каждом новом кадре вычисляется разница между моделью и текущим кадром;

– полученная разница обрабатывается путем сравнения с пороговым значением;

– удаляется шум;

– выделяются связанные компоненты;

– обновляется модель фона везде, где не присутствуют объекты переднего плана.

Построение модели фона. Самый простой вариант построения модели фона — попиксельно усреднить некоторое количество кадров по интенсивности цвета:

$$I_0(x,t) = \frac{1}{T}\sum_{t=1}^{T} I(x,t),$$

где I — значение интенсивности; t — момент времени; x — номер пикселя; T — количество временный отсчетов для усреднения.

Данный метод не будет работать, если в кадре есть движущиеся объекты или случайные и резкие изменения яркости (блики, засветка). Поэтому используется более продвинутый вариант — вместо среднего берут медиану для каждого пикселя. Медианный фон также вычитают из кадра, на котором необходимо обнаружить движущийся объект.

Медианный фон также может работать не всегда, например, в случаях, когда наблюдаемый объект долгое время (более 50 % всех кадров тестового видеоролика) находится на одном месте.

Существует целая группа методов построения модели фона на основе оптимизации целевой функции. Все изображения можно разбить на сегменты и затем выбирать сегменты из разных кадров для получения наиболее плавной и стабильной картинки.

Еще один метод — построение модели фона при постепенном изменении освещенности [8]. В этом случае предлагается сделать плавное постепенное обновление фона, например, брать движущееся среднее. Берут взвешенную сумму текущего и предыдущих N кадров, которая затем усредняют:

$$I_0(x,t) = \frac{\omega_a I(x,t) + \sum_{i=1}^{N} \omega_i I(x,t-i)}{\omega_c},$$

где ω_a — вес текущего кадра; ω_i — вес i-го кадра; ω_c — константа для нормирования итоговой суммы взвешенных кадров, представляющая собой сумму весов всех кадров; N — количество кадров.

Каждый новый кадр добавляется в буфер, получается новый фон, последний кадр отбрасывается. Данный процесс потребляет много памяти, поскольку чем больше кадров будет храниться в памяти, тем лучше будет работать построенная модель. При этом данный метод не будет работать при возникновении многих распространенных помех, таких как изменение положения объекта фона, плавные или резкие изменения в освещении, колебания объектов фона и т. д.

Объединение трекеров. Поскольку все методы отслеживания очень просты и очень часто не срабатывают, наилучший алгоритм должен объединять несколько методов [9], т. е. в нем должна использоваться комбинация из последовательно применяемых наилучших методов. Для этого ряд базовых тре-

 ISSN 2541-8009

керов тестируют на наборе выборок, оценивают их качество и надежность. Среди трекеров ищут наиболее хорошо дополняющие друг друга группы методов, которые используют для отслеживания объекта.

Для разных комбинаций исследуют зависимость вероятности того, что на последующем кадре объект будет отслежен правильно, от параметра доверия, специфичного для каждого метода. Комбинации методов объединяют в каскады, при падении доверия к первому методу в группе (когда ошибка падает ниже заданного порога) происходит переключение на второй метод и т. д. Если последний метод также не сработал, детектор необходимо переинициализировать, и запустить каскад снова.

Оценка точности. Для оценки качества работы методов видеонаблюдения вводят специальные метрики. При этом размеченные данные должны иметь вид длительных видеороликов, в каждом кадре которых (или в каждом ключевом кадре которых) отмечено положение объекта интереса. Задача разметки данных является непростой, поскольку даже 1 минута видео при частоте 30 кадров в секунду включает в себя 1800 кадров. Поэтому чаще всего размечают не все кадры, а только ключевые.

В случае, когда требуется отслеживать обстановку и распознавать объекты, нужно будет находить не только сами объекты, но и их контуры. Соответственно, задача разметки кадров еще больше усложнится. Контур представляет собой видимый край объекта, который разграничивает искомый объект интереса и область фона. Контуры используются при анализе изображений вместо пикселей. Выделение и фильтрация контуров также позволяет отсеивать из рассмотрения контуры ложных объектов, которые часто присутствуют на обрабатываемых кадрах [10].

После того как положение объекта на каждом ключевом кадре отмечено, можно вычислить ошибку отслеживания, отражающую то, насколько измеренное положение объекта отличается от его действительного положения:

$$e_t^k = d(\hat{x}_t^k, x_t^{gt}),$$

где e_t^k — ошибка; k — порядковый номер объекта; t — момент времени; d — функция расчета расстояния между измеренным и действительным положениями объекта; $\hat{x}_t^k$ — предсказанное положение объекта; x_t^{gt} — действительное положение объекта на ключевом кадре.

Усреднив все эти данные по видео, можно определить ожидаемую ошибку $E(e^k)$ на видеопоследовательности:

$$E(e^k) = \frac{1}{T}\sum_t e_t^k, \;\; k = 1,\ldots,K.$$

Ожидаемая ошибка представляет собой не очень полезный показатель, гораздо большее значение имеет показатель, называемый точностью (*precision*). Он отображает долю кадров, на которых объект нашелся с ошибкой, не превышающей заданный порог:

$$1 - E(e^k \mid e^k < \tau),$$

где τ — заданное пороговое значение.

Надежность отслеживания измеряется как вероятность p того, что слежение на последующем кадре не будет утеряно:

$$p(e_t^k < \tau \mid e_{t-1}^k < \tau).$$

Фактически надежность показывает количество ошибок системы. Если система часто ошибается, то необходимо распознать ошибку и переинициализировать систему. Сама переинициализация системы представляет собой довольно сложный процесс.

Заключение. В данной статье были рассмотрены основные методы отслеживания объектов на видеопоследовательности.

Основой всех современных систем видеонаблюдения со стационарным фоном является вычитание фона. Если фон динамический, то нужно отдельно выделять объекты и затем применять методы отслеживания объектов. При этом сам движущийся объект может быть представлен в разных видах: в виде точек, сочлененных моделей, контуров или оптического потока. В отдельных случаях хорошо работают достаточно простые базовые методы слежения:

– сопоставление с ближайшими объектами;

– сопоставление точек объектов (пятен);

– вычитание матриц для поиска сдвигов.

Наилучший результат достигается при использовании комбинации этих методов, когда одновременно запускается каскад из нескольких методов и происходит переключение между ними, но иногда это существенно замедляет скорость работы. Стоит отметить, что все методы видеонаблюдения, которые были придуманы в последние 10 лет, как правило, работают хуже, чем известные методы, скомбинированные между собой правильным образом.

Заметим, что все существующие методы отслеживания не дают идеальной точности. Существует ряд проблем, которые необходимо решить, чтобы получить более надежную и общую систему, основанную на комбинирован-

 ISSN 2541-8009

ном подходе. Многие методы не работают в случае полного поворота объекта вне плоскости. Кроме того, из-за несовершенства алгоритмов отслеживания некоторые характерные точки могут быть потеряны.

Литература

[1] Darrell T., Pentland A.P. Pfinder: Real-Time Tracking of the Human Body. *IEEE PAMI*, 1997, vol. 19, iss. 7. http://dx.doi.org/10.1109/34.598236

[2] Verschae R., Ruiz-del-Solar J. Object Detection: Current and Future Directions. *Front. Robot. AI*, 2015, vol. 2. https://doi.org/10.3389/frobt.2015.00029

[3] Ramakant C., Rohit R., Rohit M., Upasana Si., Alok Kumar S.K., Hiral R. Enhanced the moving object detection and object tracking for traffic surveillance using RBF-FDLNN and CBF algorithm. *Expert Systems with Applications*, 2021, vol. 191. https://doi.org/10.1016/j.eswa.2021.116306

[4] Bartels A., Zeki S. The architecture of the colour centre in the human visual brain: new results and a review. *European Journal of Neuroscience*, 2000, vol. 12. https://doi.org/10.1046/j.1460-9568.2000.00905.x

[5] Ren Li, Xian Weifu, Tang Hao, Jiang Yadong, Jia Haitao, Li Jing. Pedestrian and Face Detection with Low Resolution Based on Improved MTCNN. *ICCPR 2020: 2020 9th International Conference on Computing and Pattern Recognition*, 2020. https://doi.org/10.1145/3436369.3436492

[6] Ochs P., Malik J., Brox T. Segmentation of moving objects by long term video analysis. *IEEE transactions on pattern analysis and machine intelligence*, 2014, vol. 36. https://doi.org/ 10.1109/TPAMI.2013.242

[7] Yu Zhang, Hongjie Wang, Yongkai Yin, Wenjie Jiang, and Baoqing Sun. Mask-based single-pixel tracking and imaging for moving objects. *Opt. Express*, 2023, vol. 31, pp. 32554–32564. https://doi.org/10.1364/OE.501531

[8] Lionel Rakai, Huansheng Song, ShiJie Sun, Wentao Zhang, Yanni Yang. Data association in multiple object tracking: A survey of recent techniques. *Expert Systems with Applications*, 2021, vol. 192.
https://doi.org/10.1016/j.eswa.2021.116300

[9] Shanq-Jang Ruan. Illumination-Sensitive Background Modeling Approach for Accurate Moving Object Detection. *IEEE Transactions on Broadcasting*, 2011, vol. 57 (4), pp. 794–801. https://doi.org/ 10.1109/TBC.2011.2160106

[10] Myung-Cheol Roh, Tae-Yong Kim, Jihun Park, Seong-Whan Lee. Accurate object contour tracking based on boundary edge selection. *Pattern Recognition*, 2007, vol. 40, iss. 3.

Поступила в редакцию 07.05.2024

Малышев Павел Викторович — студент кафедры «Информационные системы и телекоммуникации», МГТУ им. Н.Э. Баумана, Москва, Россия.

Научный руководитель — Локтев Даниил Алексеевич, д-р техн. наук, доцент кафедры «Информационные системы и телекоммуникации», МГТУ им. Н.Э. Баумана, Москва, Россия.

Ссылку на эту статью просим оформлять следующим образом:
Малышев П.В. Алгоритмы обнаружения и отслеживания объектов на видеопоследовательности. *Политехнический молодежный журнал*, 2024, № 04 (93). URL: https://ptsj.bmstu.ru/catalog/icec/inf_tech/989.html

 ISSN 2541-8009

VIDEO SEQUENCE OBJECT DETECTION AND TRACKING ALGORITHMS

P.V. Malyshev pasha_malyshev01@mail.ru

Bauman Moscow State Technical University, Moscow, Russian Federation

The paper considers the video sequence object detection and tracking algorithms, and selects the most suitable methods in identifying and highlighting the object contours in a video stream without using the machine learning. It proposes an algorithm for creating a simple video surveillance system. Many approaches to object tracking combine tracking, learning, and detection. The tracking algorithm, or the so-called tracker, follows an object from one frame to another. The detection algorithm, or the detector, localizes all the features observed so far and, if necessary, adjusts the tracker. During learning, errors of the detector are assessed and its operation is corrected to avoid these errors in future. Let us note that learning in object detection is usually performed under the condition that all the learning examples are labeled. This paper analyzes algorithms and methods that could be introduced to create detectors and trackers.

Keywords: algorithm, object detection, object tracking, video surveillance, video surveillance systems, image processing, background model, computer vision

Received 07.05.2024

Malyshev P.V. — Student, Department of Information Systems and Telecommunications, Bauman Moscow State Technical University, Moscow, Russia.

Scientific advisor — Loktev D.A., Dr. Sc. (Eng.), Associate Professor, Department of Information Systems and Telecommunications, Bauman Moscow State Technical University, Moscow, Russia.

Please cite this article in English as:
Malyshev P.V. Video sequence object detection and tracking algorithms. *Politekhnicheskiy molodezhnyy zhurnal*, 2024, no. 04 (93). (In Russ.). URL: https://ptsj.bmstu.ru/catalog/icec/inf_tech/989.html

