

Анализ совместной работы методов сопровождения объектов в видеопотоке, получаемом с летательного аппарата*

В.А. Павлов, С.В. Завьялов, С.В. Волвенко

pavit@bk.ru | zavyalov_sv@cee.spbstu.ru | volk@cee.spbstu.ru

Санкт-Петербургский государственный политехнический университет Петра Великого

Рассмотрены варианты решения задачи сопровождения объекта в видеопотоке, получаемом с камеры беспилотного летательного аппарата. Проведён анализ современных алгоритмов сопровождения объектов и принципов их работы. Рассмотрена совместная реализация алгоритмов сопровождения.

Ключевые слова: TLD, CMT, Struck, сопровождение объекта, летательный аппарат

Analysis of collaboration methods for tracking objects in a video stream received from an aircraft*

V.A. Pavlov, S.V. Zavialov, S.V. Volvenko

Peter the Great St.Petersburg Polytechnic University, St.Petersburg, Russia

The variants of solving the problem of tracking an object in the video stream received from the camera of the unmanned aerial vehicle are considered. The analysis of modern algorithms for tracking objects performed. Joint implementation of the tracking algorithm is considered.

Keywords: TLD, CMT, Struck, object's tracking, aerial vehicle

1. Введение

Одной из наиболее сложных и актуальных задач машинного зрения является проблема обнаружения и сопровождения движущихся объектов в видеопоследовательности при наличии различного вида помех [1]. Представляет интерес два основных случая движения [2]: неподвижная камера и постоянный фон, подвижная камера и изменяющийся фон. В первом случае движение объектов характеризуется изменением яркости пикселей в определённых областях изображения, поскольку фон не изменяется. Второй случай является более сложным, из-за обобщённого движения объекта и камеры.

2. Проблемы, возникающие при работе с видеопотоком, получаемым с летательного аппарата

В данной статье рассматривается расположение камеры на борту летательного аппарата, соответствующее второму случаю движения. Работа с видеопотоком, получаемым с летательного аппарата, накладывает специфические требования на алгоритмы сопровождения, из-за следующих особенностей видеопоследовательностей: потеря одного или нескольких кадров; частично или полностью повреждённые кадры; артефакты сжатия с потерями; резкие рывки камеры; выход объекта за пределы кадра и возврат в кадр через несколько секунд; размытие (motion blur) всего кадра при резком движении камеры из-за низкой скорости затвора; размер сопровождаемого объекта может меняться от минимального (около 16x16 пикселей) до максималь-

ного (704x576), что соответствует размеру кадра; движение камеры характеризуется частым сдвигом, приближением/удалением в очень широких пределах, вращением камеры; наличие данных телеметрии поверх изображения. Также следует отметить, что часто возникает ситуация, когда в кадре присутствуют нескольких одновременно движущихся объектов с близкими характерными признаками и пересекающимися траекториями.

Целью настоящего исследования является анализ, выбор и тестирование совместной работы алгоритмов сопровождения за объектами.

3. Описание выбранных алгоритмов

В ходе анализа публикаций, освещающих решения поставленной задачи, были выбраны следующие подходы: TLD [3], CMT [4], Struck [5]. Эти подходы частично удовлетворяют следующим критериям: устойчивая работа при полном или частичном исчезновении объекта из поля зрения камеры; требуемая производительность; точность сопровождения. Далее рассмотрим более подробно выбранные алгоритмы.

3.1 Принцип работы TLD

Подход, который использован в алгоритме TLD (рис. 1) называется: Сопровождение-Обучение-Обнаружение (Tracking-Learning-Detection) и состоит из трёх компонентов: краткосрочного сопровождения, обучения и обнаружения. Модуль сопровождения осуществляет сопровождение объекта от кадра к кадру. Модуль обнаружения локализует все появления объекта, которые наблюдались в ближайшее время и корректирует модуль сопровождения, если это необходимо. Модуль обучения оце-

Работа опубликована при финансовой поддержке РФФИ, грант 15-07-20347.

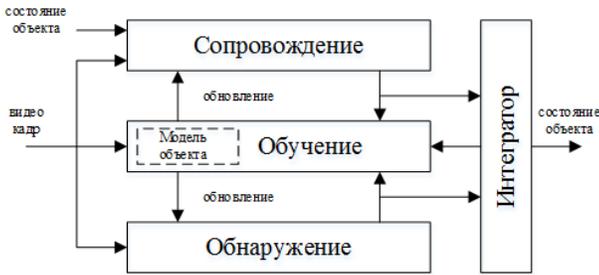


Рис. 1: Принцип работы TLD

нивает ошибки модуля обнаружения и обновляет его, чтобы избежать их в будущем. Интегратор усредняет выходные ограничительные прямоугольные рамки модулей сопровождения и обнаружения в одну выходную ограничительную прямоугольную рамку. Если на выходе ни модуля сопровождения, ни модуля обнаружения нет ограничительной рамки, объявляется что сопровождаемого объекта не видно.

Достоинства: высокая производительность; устойчивость к частичному и полному исчезновению, а также к различным изменениям внешнего вида объекта.

3.2 Принцип работы СМТ

Основная идея алгоритма СМТ (рис. 2) заключается в обнаружении ключевых точек, принадлежащих сопровождаемому объекту и их поиске в каждом кадре. Для этого используются два подхода:

- отслеживание ключевых точек от кадра к кадру с помощью оценки их оптического потока;
- сопоставление ключевых точек, путем сравнения их дескрипторов.

Поскольку оба этих метода подвержены ошибкам, осуществляется поиск согласованного решения по обоим подходам в найденных ключевых точках. По каждой ключевой точке определяется предполагаемый центр объекта. Вычисленные предполагаемые центры объекта затем интегрируются и побочные максимумы удаляются. На основе оставшихся ключевых точек вычисляется новая ограничивающая прямоугольная рамка и процесс продолжается.



Рис. 2: Принцип работы СМТ

Достоинства: высокая производительность; точное сопровождение при смещении, изменении масштаба и вращении различных классов объектов; устойчивость к частичному и полному исчезновению объекта, а также к различным изменениям внешнего вида.

3.3 Принцип работы Struck

В подходе Struck для сопровождения объектов используется метод опорных векторов с использованием функции ядра [5], который позволяет проводить обучение в режиме реального времени. Проблемой метода опорных векторов с использованием функции ядра является то, что число векторов неограниченно растет с ростом количества обучающих данных.

Для решения этой проблемы в данном подходе применяется механизм бюджетирования (или ограничения). Если бюджет уже полный и необходимо добавить новый опорный вектор, то выполняется процедура определения и удаления подходящего опорного вектора с наименьшим весовым коэффициентом и коррекция коэффициентов остальных опорных векторов. Достоинства: высокая производительность, устойчивость к шумам, устойчивость к частичному и полному исчезновению объекта.

4. Система тестирования

Для анализа выбранных алгоритмов сопровождения была разработана система тестирования (рис. 3). Реализация алгоритмов сопровождения основывалась на библиотеке OpenCV 2.4.9 [6]. В качестве тестовых видеопоследовательностей использовались записи с камер летательного аппарата. Для каждой тестовой видеопоследовательности предварительно создавалась аннотация (координаты и размеры объекта на каждом кадре) вручную в формате «XML». Результат работы алгоритмов сопровождения фиксировался, затем выполнялось сравнение аннотаций и результатов работы алгоритмов сопровождения. После сравнения формировался отчет в формате «HTML» с подробной статистикой.

Для оценки качества работы алгоритмов сопровождения на всей видеопоследовательности использовалась мера «точность сопровождения», которая для одной видеопоследовательности определяется по формуле:

$$p = \sum_{i=1}^{N(n)} \left(S_i / N(n) \right),$$

$$S_i = \begin{cases} 1, & \left[\begin{array}{l} A(R_1 \cap R) / A(R) > e \\ A(R_1 \cap R) / A(R_1) > e \end{array} \right] \\ 0, & \text{иначе} \end{cases},$$

где $N(n)$ — число кадров в видеопоследовательности, на которых присутствует объект; n — число кадров, на которых было произведено выделение сопровождаемого объекта; S_i — мера перекрытия для покадровой оценки производительности алгоритмов сопровождения (результат обнаружения на очередном кадре); R_1 — прямоугольник найденного алгоритмом на кадре объекта; R — прямоугольник объекта из файла аннотаций; $A(R)$, $A(R_1)$, $A(R_1 \cap R)$ — площади соответствующих прямоугольников и площадь пересечения этих прямоугольников; e — точность, устанавливается равной 0.3 (значение обусловлено плохим качеством видеопоследовательностей).

То есть объект считается найденным верно в кадре, если площадь пересечения найденного прямоугольника и прямоугольника из файла аннотаций не меньше, чем 30% площади каждого из них соответственно. Если объект сопровождения на очередном кадре вышел за его пределы, а алгоритм сопровождения данный объект на кадре нашёл, то S_i считается равным 0.

5. Результаты работы TLD, CMT, Struck на тестовых видеопоследовательностях, полученных с камеры летательного аппарата

Тестирование проводилось на 77 видеопоследовательностях, с различными сценариями поведения объекта, освещением, временем года, фоном, типом объекта, типом камеры и качеством съёмки. Следующие факторы осложняли работу алгоритмов: наличие посторонних объектов (данные телеметрии); плохое качество съёмки (объект трудно различить и отследить даже человеческому глазу); на некоторых видеопоследовательностях удавалось определить объект лишь в динамике; на видеопоследовательностях, снятых на тепловизор, зачастую, трудно различить очертания объектов и выделить их из общей сцены; наличие объектов близких по характеристикам с наблюдаемым объектом; составная структура некоторых объектов; значительное увеличение и уменьшение размеров сопровождаемого объекта; частичный и полный поворот камеры или объекта.

В таблицах 1, 2, 3 представлены результаты работы алгоритмов сопровождения для произвольных видеопоследовательностей.

Был проведён анализ целесообразности применения предобработки кадров перед подачей на вход алгоритма. Путём тестирования было показано, что предобработка кадров имеет смысл только для CMT, поскольку средняя точность сопровождения по всем тестовым последовательностям возросла примерно на 10%. Причем для CMT ключевые точки будут детектироваться с большей вероятностью

Таблица 1: Результаты тестирования TLD для шести произвольных видеопоследовательностей.

Номер видеопоследовательности	TLD	
	Точность, %	Ср. время, с
1	12.6	0.02
2	81.2	0.03
3	86.3	0.02
4	71	0.02
5	56.4	0.05
6	100	0.02
Среднее по 77 видео	51.7	0.04

Таблица 2: Результаты тестирования CMT для шести произвольных видеопоследовательностей.

Номер видеопоследовательности	CMT	
	Точность, %	Ср. время, с
1	97.9	0.04
2	*	-
3	98.6	0.03
4	27.1	0.03
5	20.8	0.04
6	0	0.01
Среднее по 77 видео	21.5	0.03

Таблица 3: Результаты тестирования Struck для шести произвольных видеопоследовательностей.

Номер видеопоследовательности	Struck	
	Точность, %	Ср. время, с
1	21.8	0.04
2	99.1	0.05
3	100	0.02
4	12.6	0.02
5	100	0.04
6	100	0.02
Среднее по 77 видео	32.8	0.05

Таблица 4: Результаты тестирования для двух произвольных видеопоследовательностей с применением выделения границ оператором Scharf

Номер видеопоследовательности	CMT	
	Без обработки	Выделение границ
7	48%	76.9%(+28.9%)
8	*	60.3%
Среднее по 77 видео	21.5%	31.5%(+10%)

за счёт повышения детализации объекта сопровождения. В таблице 4 приведены результаты тестирования для двух произвольных видео с применением выделения границ оператором Scharf.

В таблице 5 представлены результаты тестирования

Таблица 5: Результаты тестирования для двух произвольных видеопоследовательностей с применением фильтра увеличивающего резкость

Номер видеопоследовательности	СМТ	
	Без обработки	Резкость (Sharpen filter)
9	20.5%	99.2% (+78.7%)
10	*	60.3%
Среднее по 77 видео	21.5%	0.05

ния для двух произвольных видеопоследовательностей с применением фильтра увеличивающего резкость.

Тестирование показало, что применять к кадрам выделение границ оператором Scharr перед подачей на вход СМТ нецелесообразно, из-за значительного повышения времени работы СМТ.

6. Реализация совместной работы алгоритмов

Для повышения точности сопровождения целесообразно использовать совместную работу алгоритмов TLD, СМТ и Struck с учётом коэффициента уверенности и весовых коэффициентов для каждого алгоритма. Весовой коэффициент определяет вклад выходного результата алгоритма при формировании итогового результата. Коэффициент уверенности необходим для определения момента потери объекта при сопровождении или для переключения с одного алгоритма на другой при совместной работе алгоритмов. В оригинальной реализации TLD — это «conf», для СМТ в качестве коэффициента уверенности можно использовать отношение найденных ключевых точек при инициализации к количеству ключевых точек на текущем кадре, а в реализации Struck — параметр «bestScore».

На рис. 3, 4, 5 представлены примеры зависимости коэффициентов уверенности TLD, СМТ и Struck от номера кадра, а также процент перекрытия, нормированный к единице от номера кадра для произвольной видеопоследовательности. Анализ коэффициента корреляции между зависимостями коэффициента уверенности от номера кадра и коэффициентом точности сопровождения от номера кадра позволяет определить целесообразность использования коэффициентов уверенности алгоритмов сопровождения при совместной реализации.

Средний коэффициент корреляции между коэффициентом уверенности и коэффициентом точности сопровождения для 77 видеопоследовательностей для TLD составляет 0.84, для СМТ 0.83, для Struck — 0.08. Таким образом для совместной работы алгоритмов можно использовать коэффициент

* не удалось детектировать ключевые точки



Рис. 3: Зависимости коэффициента уверенности и точности сопровождения от кадра для произвольной видеопоследовательности для TLD (коэфф. корреляции 0.99)



Рис. 4: Зависимости коэффициента уверенности и точности сопровождения от кадра для произвольной видеопоследовательности для СМТ (коэфф. корреляции 0.98)



Рис. 5: Зависимости коэффициента уверенности и точности сопровождения от кадра для произвольной видеопоследовательности для Struck (коэфф. корреляции 0.84)

ты уверенности TLD и СМТ, которые в должной мере описывают момент исчезновения объекта из поля зрения наблюдателя.

Результаты экспериментов на 77 последовательностях показали, что при исчезновении объекта из

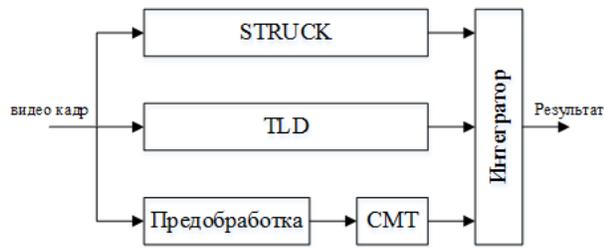


Рис. 6: Принцип совместной работы TLD, CMT, Struck

поля зрения, значение коэффициента «bestScore» алгоритма Struck некорректно по отношению к выходному результату и его использование при сопровождении нецелесообразно. Поэтому при совместной реализации при значении «bestScore» меньшим или равном 0 на более чем N кадрах (N подбирается опытным путём и равно 3-6), алгоритм Struck переинициализировался выходной ограничительной рамкой алгоритма TLD, которая подавалась на вход функции инициализации Struck при значении коэффициента уверенности TLD больше или равным 0.5. При неудачной инициализации CMT, на последующих кадрах осуществляется инициализация алгоритма выходной ограничительной рамкой алгоритма TLD, аналогично описанному выше случаю переинициализации Struck.

На рис. 6 представлена блок-схема совместной работы TLD, CMT и Struck. При инициализации каждому алгоритму присваивается свой весовой коэффициент, лежащий в диапазоне $[0, 1]$, сумма всех весовых коэффициентов не должна превышать 1. На вход каждого алгоритма подаётся кадр видеопоследовательности, перед подачей кадра на вход CMT осуществляется предобработка фильтром, повышающим резкость. Все алгоритмы работают параллельно и независимо друг от друга. В интеграторе осуществляется усреднение выходных ограничительных рамок алгоритмов с учётом весовых коэффициентов и формирование итоговой выходной ограничительной рамки с предполагаемым расположением объекта.

При равных весовых коэффициентах средняя точность сопровождения по 77 тестовым видеопоследовательностям составила 54.8%, что всего лишь на 3% выше средней точности TLD, но при оптимальных (по критерию максимизации результирующей точности сопровождения) значениях весовых коэффициентов средняя точность сопровождения возросла до 69.7%. Количество видеопоследовательностей, на которых алгоритмы показывают точность более 70% при оптимальных значениях весовых коэффициентов возросло до 44. Предложенная схема переинициализации позволила повысить точность сопровождения Struck с 32.8% до 50.5%. Точность

CMT при использовании данной схемы переинициализации и предобработки (повышение резкости) возросла с 21.5% до 57%.

Выводы

Совместное использование TLD, CMT и Struck позволяет незначительно повысить точность сопровождения при равных значениях весовых коэффициентов до 54.8%, количество видеопоследовательностей, на которых алгоритмы показывают точность сопровождения более 70% при оптимальных значениях весовых коэффициентов возросло с 33 до 44 (из 77).

Для повышения точности сопровождения целесообразно применять: предобработку изображений (повышение резкости) перед подачей кадра на вход CMT, проверку корректности полученных кадров, а также использовать данные телеметрии для уточнения положения объекта.

Предложенная схема переинициализации позволила повысить точность сопровождения Struck с 32.8% до 50.5%. Точность CMT при использовании предложенной схемы переинициализации и предобработки (повышение резкости) возросла с 21.5% до 57%.

Целью дальнейших исследований является разработка адаптивной схемы переключения и присваивания оптимальных весовых коэффициентов алгоритмам сопровождения; повышение точности сопровождения и производительности; рассмотрение новых алгоритмов принятия решения при совместной реализации; использование модели движения камеры для дополнительного уточнения местоположения объекта на текущем кадре.

Литература

- [1] А.А. Лукьяница, А.Г. Шихин Цифровая обработка видеоизображений. – Москва: «Ай-Эс-Эс Пресс», 2009. – 518с.
- [2] Л. Шапиро, Дж. Стокман Компьютерное зрение. – Москва: БИНОМ. Лаборатория знаний, 2006. – 752с.
- [3] Zdenek Kalal, Krystian Mikolajczyk, Jiri Matas Tracking-Learning-Detection // IEEE Transactions on pattern analysis and machine intelligence, 2010. — Vol. 6, no. 1.
- [4] Georg Nebehay, Roman Pflugfelder Consensus-based Matching and Tracking of Keypoints for Object Tracking // Safety and Security Department AIT Austrian Institute of Technology 1220 Vienna, Austria, 2014.
- [5] Sam Hare, Amir Saffari, Philip H. S. Torr Struck: Structured Output Tracking with Kernels // Oxford Brookes University, Oxford, UK Sony Computer Entertainment Europe, London, UK 2012.
- [6] opencv.org – OpenCV – 2015.