

ОБ ОДНОМ ПОДХОДЕ К РЕШЕНИЮ ЗАДАЧИ ВИДЕОДЕТЕКТИРОВАНИЯ ТРАНСПОРТНЫХ СРЕДСТВ

В.Д. Кустикова

Нижегородский госуниверситет им. Н.И. Лобачевского

E-mail: valentina.kustikova@gmail.com

Рассматривается задача видеодетектирования транспортных средств. Приводится схема решения задачи с использованием методов поиска и сопровождения на потоке видеоданных. Описывается прототип разработанной системы для решения задачи видеодетектирования. Приводятся результаты вычислительных экспериментов.

Введение

Задача видеодетектирования транспортных средств является актуальной, ее решение позволяет оценивать пропускную способность дорожных участков, эффективно выполнять автоматическое регулирование движения, принимать обоснованные решения о необходимости строительства дорог-дублеров и развязок, проектировать дорожную сеть в масштабах города. Решение данной задачи может быть интересно не только дорожным организациям, но и рекламным службам, владельцам автостоянок и крупных торговых комплексов с точки зрения организации их работы.

1. Постановка задачи

Метод видеодетектирования автомобилей работает с потоком видеоданных. Обозначим входной видеопоток V . По существу, V можно представить в виде последовательности кадров или изображений $I_0, I_1, I_2, \dots, I_{N-1}$, где N – количество кадров.

Задача видеодетектирования транспортных средств состоит в том, чтобы каждому кадру I_k исходного видео V поставить в соответствие совокупность положений объектов BB_k . В зависимости от алгоритма под положением может пониматься расположение контуров объекта [1], особых точек или расположение трехмерной геометрической модели объекта [2]. Далее примем, что положение объекта определяется расположением прямоугольника, его окаймляющего, аналогично работам [3–6]. Тогда задача видеодетектирования сводится к построению отображения $\varphi: \{I_k, k = \overline{0, N-1}\} \rightarrow \{BB_k, k = \overline{0, N-1}\}$, где $BB_k = \{bb_l^k, l = \overline{0, s_k-1}\}$ – множество окаймляющих прямоугольников, обнаруженных на кадре I_k ; $s_k = |BB_k|$ – количество найденных прямоугольников. При этом каждый прямоугольник bb_l^k определяется набором следующих компонент: $bb_l^k = ((x_1^l, y_1^l), (x_2^l, y_2^l), [s^l, c^l])$, где $(x_1^l, y_1^l), (x_2^l, y_2^l)$ – координаты левого верхнего и правого нижнего углов прямоугольника ($x_1^l \geq 0, y_1^l \geq 0, x_2^l \geq 0, y_2^l \geq 0$), $s^l \in \mathbb{R}$ – достоверность того, что объект продетектирован правильно, а $c^l \in C = \{‘car’, ‘bus’, \dots, ‘train’\}$ – класс, которому принадлежит транспортное средство. Отметим, что в зависимости от специфики решаемой задачи и выбора метода ее решения последние две компоненты могут отсутствовать. Также положение прямоугольника может быть задано, например, координатами левого верхнего угла и значениями ширины и высоты сторон.

Результат работы выбранного алгоритма не всегда зависит только от информации, извлекаемой из текущего кадра. В общем случае отображение φ зависит от целого

набора параметров $\varphi(I_{k-l_1}, I_{k-l_1+1}, \dots, I_{k-1}, I_k, I_{k+1}, \dots, I_{k+l_2}, P^{(d)}, P^{(t)}) = BB_k$: информация, полученная на l_1 -предыдущих кадрах $I_{k-l_1}, I_{k-l_1+1}, \dots, I_{k-1}$ видео V ($l_1 \leq k$); информация, извлекаемая из l_2 -следующих кадров $I_{k+1}, \dots, I_{k+l_2}$ видео V ($l_2 \leq N - 1 - k$); совокупность параметров $P^{(d)}$ алгоритма поиска объектов; набор параметров $P^{(t)}$ алгоритма сопровождения.

В рассматриваемой постановке необходимо дополнительно выполнять построение траекторий движения объектов. Требуется ввести алгоритм сопоставления ψ окаймляющих прямоугольников на паре последовательных кадров I_k и I_{k+1} . Обратим внимание, что на некотором кадре объект исчезает из поля зрения камеры. Поэтому требуется рассматривать множество образов, расширенное специально выделенным окаймляющим прямоугольником с несуществующими координатами, например, $bb = ((-1, -1), (-1, -1)[\forall s^l, \forall c^l])$. Тогда $\psi: BB_k \rightarrow BB_{k+1} \cup \{bb\}, \forall k = \overline{0, N-2}$. Допустим, что I_k – первый кадр, на котором был обнаружен объект, содержащийся в прямоугольнике с индексом $r_0(k) \in \{0, 1, 2, \dots, s_k\}$, а q – количество последующих кадров, на которых объект виден наблюдателю. Тогда под траекторией движения $T_{r_0(k)}^k$ транспортного средства будем понимать последовательность положений соответствующих окаймляющих прямоугольников $(bb_{r_0}^k, bb_{r_1}^{k+1}, bb_{r_2}^{k+2}, \dots, bb_{r_{q-1}}^{k+q-1})$, где $r_i \in \{0, 1, 2, \dots, s_{k+i}\}, bb_{r_i}^{k+i} = \psi(bb_{r_{i-1}}^{k+i-1}), i = \overline{1, q-1}$.

2. Схема решения

Типичная схема работы систем видеодетектирования посредством обнаружения и сопровождения объектов не зависит от типа детектируемых объектов (человек, автомобиль и др.) и состоит из следующей последовательности действий:

1. Извлечение кадра из потока видеоданных.
2. Предобработка полученного статического изображения – кадра (фильтрация, подавление теней, сегментация, выделение областей интереса).
3. Поиск положения объекта на изображении и оценка степени достоверности нахождения объекта в данной области.
4. Сопровождение (трекинг) обнаруженных объектов. На данном этапе выполняется построение траекторий движения транспортных средств, выделенных на предыдущем шаге, от момента их обнаружения до их выхода из области обзора камеры.
5. Анализ результатов поиска и сопровождения (например, оценка направления движения объекта на основании траектории, подсчет числа объектов и т.п.). Функциональность блока анализа результатов в основном зависит от конкретной задачи, которая поставлена перед разработчиками системы видеодетектирования.

3. Программная реализация

Множество всех модулей системы можно разделить на 5 групп (рис. 1):

1. Модули подготовки тестовых данных:
 - модуль разметки окаймляющих прямоугольников для транспортных средств разных классов;
 - модуль формирования траекторий движения транспортных единиц на основании разметки с последующим их сохранением.
2. Модули детектирования и сопровождения:
 - модуль детектирования транспортных средств разных классов на отдельном кадре видео как самостоятельном изображении с помощью параллельной реализации алгоритма Latent SVM [3, 4];

- модуль сопровождения протестированных транспортных средств на последовательности кадров видеопотока;
 - приложения, содержащие реализации различных схем видеодетектирования транспортных средств.
3. Модули апробации результатов видеодетектирования транспортных средств:
 - модуль вычисления некоторых количественных показателей качества детектирования и сопровождения транспортных средств;
 - модуль VOCdevkit [7] для определения средней точности предсказания положения объектов – модуль третьих разработчиков.
 4. Служебные модули визуализации окаймляющих прямоугольников и траекторий.
 5. Служебные модули для автоматизации сбора численных показателей качества видеодетектирования.



Рис. 1. Архитектура системы видеодетектирования транспортных средств

Модули подготовки тестовых данных, детектирования и сопровождения разработаны с использованием средств библиотеки OpenCV [8].

4. Критерии оценки качества поиска объектов

Выделяются два принципиально разных подхода к оцениванию качества работы систем видеодетектирования. Один из подходов предполагает определение числа правильно и неправильно протестированных объектов на всем видеопотоке и вычисление различных относительных и абсолютных показателей. Согласно второму подходу, видео рассматривается как последовательность изображений, следовательно, качество поиска оценивается по аналогии с качеством алгоритмов детектирования объектов на изображениях.

В работах [9–11] используются количественные показатели, которые вычисляются на основании информации о числе правильно и неправильно протестированных объектов:

1. Истинно положительный показатель (the true positive rate [9] или detection rate [10]) – отношение количества правильно протектированных транспортных средств к общему числу транспортных средств $TPR = \frac{\text{detected vehicles}}{\text{total number of vehicles}}$.
2. Показатель числа ложных срабатываний (the false detection rate) – отношение количества ложных срабатываний к общему числу срабатываний детектора $FDR = \frac{\text{false positives}}{\text{detected vehicles} + \text{false positives}}$. При этом считается, что объект протектирован правильно, если перекрытие протектированного $\text{bbox}_{\text{detect}}$ и размеченного $\text{bbox}_{\text{marked}}$ окаймляющего прямоугольника превышает некоторое пороговое значение threshold (в [10] в качестве порога выбирается 80%) $\frac{S(\text{bbox}_{\text{detect}} \cap \text{bbox}_{\text{marked}})}{S(\text{bbox}_{\text{detect}} \cup \text{bbox}_{\text{marked}})} > \text{threshold}$.

Наряду с указанными метриками авторы [9] вводят несколько относительных показателей качества:

1. Количество ложных срабатываний, которое в среднем приходится на один кадр (the average false positives per frame) $FPperFrame = \frac{\text{false positives}}{\text{total number of processed frames}}$.
2. Среднее количество ложных срабатываний, приходящееся на объект (the average false positives per object) $FPperObject = \frac{\text{false positives}}{\text{true vehicles}}$.

Отметим, что авторы [9] не рассматривают задачу классификации транспортных средств, тем не менее, приведенные показатели можно распространить на случай многоклассового детектирования транспортных средств, если рассматривать эти метрики для каждого класса в отдельности.

Авторы [12] предлагают рассматривать видео как последовательность кадров видеопотока и использовать для оценки качества среднюю точность предсказания (average precision). В состав VOC Development Kit [7] входит Matlab-модуль, позволяющий вычислить данный показатель на основании результатов детектирования, записанных в специальном формате. Метрика определяется как математическое ожидание точностей:

$$AP = \frac{1}{11} \sum_{r \in \{0; 0.1; 0.2; \dots; 1\}} p(r), p(r) = \max_{\bar{r}, \bar{r} \geq r} p(\bar{r}),$$

где $p(\bar{r}) = \frac{1}{a+c}$ – точность, \bar{r} – процент перекрытия детектированного окаймляющего прямоугольника и прямоугольника, который был размечен на исходном изображении как окаймляющий $\bar{r} \in \{0; 0.1; 0.2; \dots; 1\}$, a – количество объектов, для которых процент перекрытия не меньше, чем \bar{r} (т. е. считается, что объект детектирован правильно), c – количество объектов с процентом перекрытия, меньшим, чем \bar{r} (объект найден ошибочно).

5. Тестовые данные

Для проведения экспериментов по оцениванию качества видеодетектирования транспортных средств было снято несколько треков (25FPS, 720x405 пикселей), из которых выбраны отдельные последовательности кадров:

- *track_09_0-2000* (2000 кадров = 80 секунд) содержит большое количество мелких транспортных средств трех классов (CAR, TRAIN, BUS), движущихся в двух противоположных направлениях (рис. 2, слева). Отметим, что значительную площадь кадра занимает фон.
- *track_10_5000-7000* (2000 кадров = 80 секунд) – видео с транспортными средствами единственного класса CAR, которые двигаются в 4 полосы одного направления (рис. 2, справа). Отличительной чертой является наличие большого количества кадров, не содержащих объектов.

- *track_10_7000-8000* (1000 кадров = 40 секунд) – видео, снятое при тех же условиях, что и *track_10_5000-7000*. Содержит объекты двух классов CAR и BUS.



Рис. 2. Кадры треков *track_09_0-2000*, *track_10_5000-7000*

6. Результаты экспериментов

Результаты видеодетектирования с помощью разработанных схем ниже сведены в общей таблице (таблица 1):

- 1 – схема полного покадрового детектирования;
- 2 – схема полного детектирования с отсечением прямоугольников, находящихся вне областей интереса;
- 3 – схема детектирования с одношаговым трекингом и отсечением. Идея применения трекинга на один кадр вперед состоит в том, чтобы обеспечить дополнительный уровень проверки корректности работы детектора с целью отсечения ложных срабатываний;
- 4 – схема детектирования с одношаговым трекингом без отсечения;
- 5 – схема детектирования с трекингом на 5 кадров вперед и отсечением;
- 6 – схема детектирования с трекингом на 5 кадров вперед без отсечения;
- 7 – схема детектирования с трекингом на 3 кадра вперед и отсечением.

Таблица 1. Сводная таблица результатов видеодетектирования

№ схемы	Название тестового видео	Средняя точность предсказания (average precision)		Значение показателя <i>TPR</i>		Значение показателя <i>FDR</i>		Значение показателя <i>FPperFrame</i>		Значение показателя <i>FPperObject</i>	
		BUS	CAR	BUS	CAR	BUS	CAR	BUS	CAR	BUS	CAR
1	<i>track_09_0-2000</i>	0.213	0.202	0.291	0.201	0.924	0.786	1.627	5.898	3.530	0.738
	<i>track_10_5000-7000</i>	-	0.684	-	0.728	-	0.735	-	2.942	-	2.015
	<i>track_10_7000-8000</i>	0.514	0.674	0.657	0.745	0.910	0.813	0.660	3.567	6.677	3.243
2	<i>track_09_0-2000</i>	0.217	0.149	0.287	0.114	0.802	0.663	0.532	1.792	1.168	0.224
	<i>track_10_5000-7000</i>	-	0.684	-	0.726	-	0.676	-	2.209	-	1.513
	<i>track_10_7000-8000</i>	0.514	0.674	0.657	0.745	0.905	0.784	0.619	2.980	6.263	2.709
3	<i>track_10_5000-7000</i>	-	0.624	-	0.680	-	0.437	-	0.772	-	0.529
	<i>track_10_7000-8000</i>	0.494	0.617	0.596	0.697	0.879	0.631	0.427	1.313	4.313	1.193
4	<i>track_10_5000-7000</i>	-	0.625	-	0.683	-	0.518	-	1.074	-	0.735
	<i>track_10_7000-8000</i>	0.494	0.617	0.596	0.697	0.884	0.674	0.448	1.583	4.525	1.440
5	<i>track_10_5000-7000</i>	-	0.534	-	0.586	-	0.508	-	0.885	-	0.606
	<i>track_10_7000-8000</i>	0.520	0.526	0.606	0.596	0.869	0.667	0.340	1.134	4.04	1.194
6	<i>track_10_5000-7000</i>	-	0.617	-	0.622	-	0.605	-	1.390	-	0.952
	<i>track_10_7000-8000</i>	0.520	0.602	0.606	0.621	0.879	0.721	0.436	1.770	4.404	1.609
7	<i>track_10_5000-7000</i>	-	0.619	-	0.644	-	0.521	-	1.023	-	0.700
	<i>track_10_7000-8000</i>	0.516	0.610	0.606	0.667	0.875	0.659	0.419	1.416	4.242	1.287

Количественные показатели качества (столбцы 4, 5, 6, 7) приведены для порогового значения процента пересечения размеченных и продетектированных окаймляющих прямоугольников, равного 0.5.

Переход от схемы 1 к схеме 2 – введение отсечения – позволил незначительно снизить количество ложных срабатываний детектора, о чем говорят значения показателей *FDR*, *FPperFrame* и *FPperObject*. Применение алгоритмов сопровождения обеспечило еще большее улучшение указанных показателей. При этом очевидно прослеживается снижение средней точности детектирования. Если для реализаций, использующих одношаговый трекинг (схемы 3 и 4), это не так заметно, то с многошаговым трекингом (схемы 5, 6, 7) падение точности наблюдается с увеличением шага для выполнения очередного детектирования. Причина состоит в том, что при увеличении шага увеличивается вероятность перекрытия транспортных средств другими объектами или частичный выход из кадра, как следствие, алгоритм детектирования может не найти данный объект, и ошибка распространится на столько предыдущих кадров, чему равен шаг.

В целом наиболее перспективной является схема с одношаговым сопровождением и отсечением. Повысив качество детектирования транспортных средств на каждом кадре за счет использования других алгоритмов и методов, вполне возможно улучшить полученные показатели качества.

Работа выполнена в рамках программы «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007-2013 годы», государственный контракт № 11.519.11.4015.

Литература

1. Shotton J., Blake A., Cipolla R. Contour-based learning for object detection // 10th IEEE International Conference on Computer Vision. 2005. Vol. 1. P. 503–510.
2. Buch N., Cracknell M., Orwell J., Velastin S. A. Vehicle localisation and classification in urban CCTV streams // 16th World Congress and Exhibition on Intelligent Transport Systems and Services. 2009.
3. Felzenszwalb P. F., Girshick R. B., McAllester D., Ramanan D. Object detection with discriminatively trained part based models // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2010. Vol. 32, No.9. P. 1627–1645.
4. Druzhkov P.N., Eruhimov V.L., Kozinov E.A. et al. On some new object detection features in OpenCV library // Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications. 2011. V. 21. No.3. P. 384–386.
5. Felzenszwalb P.F., Girshick R.B., McAllester D., Ramanan D. Cascade object detection with deformable path model // Proceedings of the IEEE Computer Vision and Pattern Recognition (CVPR'10). 2010. P. 2241–2248.
6. Sun Z., Bebis G., Miller R. On-road vehicle detection using Gabor filters and support vector machines // Digital Signal Processing. 2002. Vol. 2. P. 1019–1022.
7. PASCAL Visual Object Challenge Development Kit – [<http://pascallin.ecs.soton.ac.uk/challenges/VOC/voc2007/#devkit>].
8. Официальная страница библиотеки OpenCV – [<http://opencv.org/>].
9. Sivaraman S., Trivedi M.M. A General active-learning framework for on-road vehicle recognition and tracking // IEEE Transactions on intelligent transportation systems. 2010. Vol. 11, No. 2. P. 267–276.
10. Song X., Netavia R. A Model-based Vehicle Segmentation Method for Tracking // ICCV'05. 2005. Vol. 2. P. 1124–1131.
11. Kim Z.W., Malik J. Fast Vehicle Detection with Probabilistic Feature Grouping and its Application to Vehicle Tracking // ICCV'03. 2003. Vol. 1. P. 524–531.
12. Dollrар P., Wojek Ch., Schiele B., Perona P. Pedestrian Detection: A Benchmark // CVPR'09. 2009. P. 304–311.