

Информационные системы и технологии

УДК 004.054: 004.931.2

Ю.В. Бухтияров, М.В. Дидковская, канд. техн. наук

Метрики качества систем распознавания государственных номерных знаков транспортных средств

В данной статье выделены классы бизнес задач, поставленных перед системами распознавания регистрационных номеров транспортных средств, проанализированы их характеристики и разработаны соответствующие метрики качества систем распознавания.

In this paper the classes of business tasks of the registration numbers of vehicles' recognition systems were identified, their characteristics were analyzed and appropriate metrics of recognition systems' quality were developed.

Ключевые слова: системы распознавания государственных номерных знаков, качество систем распознавания, метрики качества систем распознавания.

Введение

Постоянный рост объемов транспортных потоков усложняет задачу их анализа вручную и делает внедрение современных инструментов мониторинга и управления экономически оправданным. На сегодняшний день все большее распространение получают системы распознавания государственных номерных знаков (ГРЗ) транспортных средств (ТС) и возникает задача определения качества функционирования таких систем.

Принятая метрика отношения количества распознанных номеров к общему количеству проехавших транспортных средств [1-3] не является достаточно репрезентативной, так существующие на рынке системы распознавания не являются автономными алгоритмами, а зачастую являют собой программно-аппаратные комплексы, которые не могут рассматриваться в отрыве от решаемой ими задачи. Без достаточно точного измерения качества таких систем, не представляется возможным их контролировать, а в дальнейшем сертифицировать. Под измерением качества будем понимать процесс, при котором цифры или символы присваиваются атрибутам объектов реального мира таким образом, чтобы охарактеризовать их в соответствии с четко определенными правилами [4].

Несмотря на существование ряда метрик для качества, в частности, надежности программного обеспечения [5] и исследования в этой области [6], они не применимы в чистом виде в рамках рассматриваемой задачи, так как в зависимости от бизнес задачи, качество функционирования системы будет определяться разными группами признаков.

Целью данной работы является выделение классов бизнес задач, поставленных перед системами распознавания регистрационных номеров транспортных средств и разработка соответствующих метрик качества.

1. Бизнес задачи

В рамках данной работы предлагается выделить следующие классы бизнес задач и провести анализ их характеристик:

1. Подсчет транспортных средств и анализ загруженности автомобильных дорог:

- от системы требуется функционал захвата и обеспечения уникальности транспортного средства;
- распознавание номеров в таких системах не является критичным.

2. Автоматический розыск ТС:

- от системы требуется, чтобы в базу попадали все номера (в т.ч. имеющие некорректно или неполно распознанные символы);
- существенным является обнаружение ТС без ГРЗ;
- обеспечение уникальности ГРЗ ТС, т.е. единственности попадания соответствующей записи об обнаружении транспортного средства в базу данных, не является критичным;
- желательно наличие функционала нечеткого сравнения с эталоном при поиске.

При анализе систем данного класса, введем понятие «частично распознанный» номер – номер ТС, который содержит ошибки распознавания, но несет в себе полезную информацию и может быть использован для поиска объекта с применением нечетких методов сравнения.

3. Фиксация нарушений ПДД (в т.ч. контроль скоростного режима):

- от системы требуется, чтобы информация о ТС была уникальной – в противном случае будет возникать ситуация с дублированием штрафных санкций за одно и тоже нарушение;
- обнаружение ТС без ГРЗ не нужно, ввиду невозможности определить владельца и адрес отправки протокола о нарушении ПДД;
- требуется наличие критерия позволяющего ввести разделение результатов распознавания на 2 группы:
 - печать "с участием оператора" (необходима проверка оператором);
 - "автоматическая" печать протоколов (процент неправильных результатов распознавания не должен превышать результата достижимого при ручной обработке 0.5-3%).

2. Анализ систем распознавания ГРЗ ТС

Как было сказано выше, на сегодняшний день системы распознавания ГРЗ ТС являются программно-аппаратные комплексы, в которых целесообразно выделить две основные подсистемы:

1. Подсистема распознавания.
2. Подсистема детекции ТС.

Работа систем распознавания ГРЗ ТС сопряжена с наличием дублирующих систем, которые могут решать задачу детекции ТС другими техническими методами. К дублирующим системам относятся:

1. Радары
2. Детекторы ТС

- детекторы на основе эффекта Доплера
- видеодетекторы (возможность покрытия до 8 полос движения ТС одним устройством)
- индуктивные детекторы

Благодаря интеграции с дублирующими системами, можно производить дополнительную фильтрацию результатов работы основной системы и существенно улучшать качество.

Таким образом, с учетом поставленных бизнес задач и анализа их требований предлагается выделить ряд формальных параметров, которые позволят сформировать метрики для оценки качества систем.

Сделаем ряд допущений про исходные условия работы систем распознавания ТС ГРЗ:

- Тестирование проходит в условиях отсутствия дублирующих систем.
- Введено понятие изображения ТС (или его части) необходимой для обнаружения.
- Введено понятие "частично распознанных" номеров - результат подсистемы распозна-

вания пригодный для применения в задаче "Розыск".

- Введен допустимый процент ошибок для группы результатов распознавания предназначенных для "автоматической" печати протоколов.

Введем следующие обозначения:

Q_{all} - все ТС, которые содержатся в выборке (включая автомобили без ГРЗ)

Q_{rec} - все ТС в выборке, ГРЗ которых поддаются распознаванию (список ТС, входящих в данное множество, формируется экспертной комиссией)

Th_{max} - процент ошибок при распознавании оператором

3. Метрики качества функционирования систем распознавания ГРЗ ТС

Проанализируем задачу подсчета транспортных средств и анализа загруженности автомобильных дорог. Тестирование подобных систем осуществляется посредством набора эталонных видеороликов, количество транспортных средств на которых известно. Производится сравнение количества ТС, обнаруженных системой, с эталонным. При выборе тестового образца следует обращать внимание на то, чтобы на нем присутствовали ТС без ГРЗ.

В случае корректной работы, при решении данной задачи сигнал должен поступать от системы распознавания, если номер присутствует на ТС и от подсистемы детекции, если номер отсутствует, причем сама система не обязательно должна содержать подсистему детекции, последняя может быть внешней.

Введем понятие коэффициентов качества безусловной и условной детекции.

Коэффициент качества безусловной детекции

$$D = \frac{M}{Q_{all}}, \quad (1)$$

где $M = K_r + K_d$,

K_r - количество уникальных номеров, для которых было осуществлено распознавание (как полное, так и ошибочное), но не было сигнала срабатывания подсистемы детектора.

K_d - количество уникальных срабатываний подсистемы детектора, для которых не было получено информации от подсистемы распознавания.

Коэффициент качества условной детекции

$$D_c = \frac{\sum_{i=1}^{Q_{all}} F(x_i)}{Q_{all}},$$

где x_i - объект, выявленный в системе (запротоколированный экспертной комиссией объект, м.б. как с номером, так и без него);

$F(x_i)$ - функция детекции.

$$F(x_i) = \begin{cases} 1, & \text{если получен сигнал от подсистемы распознавания,} \\ & \text{либо от подсистемы детектора} \\ 0, & \text{в противном случае} \end{cases}$$

Коэффициент качества условной детекции учитывает уникальность анализируемого объекта, в то время как с помощью коэффициента безусловной детекции можно понять сколько в среднем ложных срабатываний было допущено для объекта.

В связи с тем, что объект может характеризоваться как записью от системы распознавания так и от системы детекции, введем относительный коэффициент избыточности системы:

$$S_{red} = \frac{Q_{sys}}{M},$$

Q_{sys} - количество записей, идентифицирующих объект, в системе.

Коэффициент, показывающий внутреннюю избыточность системы, т.е. сколько информации фиксируется относительно количества выявленных объектов.

Для простоты анализа ошибочности системы введем коэффициент пропусков системы:

$$S_{ms} = 1 - D = 1 - \frac{M}{Q_{all}},$$

отношение количества объектов, которые не были выявлены системой, к общему количеству запротоколированных объектов.

Проведем анализ задачи автоматического розыска ТС.

Для данного класса задач существенным является степень распознанности номера, которая может быть полной либо частичной. При тестировании подобных систем для эталонного ролика фиксируется ТС с его ГРЗ, наличием ТС без ГРЗ можно пренебречь.

Введем понятие коэффициентов качества безусловного и условного распознавания.

Коэффициент качества безусловного распознавания

$$R = \frac{N}{Q_{rec}}, \quad (2)$$

где N - количество распознанных номеров, для которых осуществилось полное совпадение

цифробуквенной последовательности распознанного системой номера с номером из эталонного протокола, когда ему соответствует только один номер журнала системы.

Коэффициент качества условного распознавания

$$R_c = \frac{\sum_{i=1}^{Q_{all}} R_{TC}(y_i)}{Q_{rec}},$$

где y_i - ГРЗ ТС, поддающегося распознаванию;

$R_{TC}(y_i)$ - функция распознавания.

$$R_{TC}(y_i) = \begin{cases} 1, & \text{журнал регистрации содержит хотя бы один} \\ & \text{"условно распознанный" ГРЗ ТС} \\ 0, & \text{в противном случае} \end{cases},$$

где «условно распознанный» ГРЗ ТС – номер ТС, для которого осуществилось полное либо частичное (не менее 80% символов) совпадение цифробуквенной последовательности распознанного системой номера с номером из эталонного протокола.

Для простоты анализа ошибочности системы введем коэффициент пропусков подсистемы распознавания:

$$R_{lost} = 1 - R_c$$

Полезным также является коэффициент безусловной детекции, указанный в (1).

Проведем анализ задачи фиксации нарушений ПДД.

Для систем данного класса существенным является распознавание номера, поэтому необходимо использовать коэффициент качества безусловного распознавания (2). Специфично для рассматриваемого класса задач предлагается выделить коэффициент автоматической печати протоколов нарушений:

$$K_{auto} = \frac{R_{auto}}{Q_{recs}},$$

где R_{auto} - количество ГРЗ ТС, выделенное в подгруппу автоматической печати (согласно критерия производителя), Q_{recs} - общее количество ГРЗ ТС распознанных системой.

Для анализа качества работы системы предлагается использовать коэффициент процента ошибок, допускаемых при автоматической печати протоколов нарушений:

$$K_{bad} = \frac{R_{bad}}{R_{auto}},$$

где R_{bad} - количество номеров ГРЗ ТС цифробуквенная последовательность, которых не совпадает с номером из эталонного протокола.

Заметим, что если процент ошибок допускаемых системой превышает T_{max} коэффициент автоматической печати протоколов нарушений принимается равным 0.

Если система не осуществила автоматическую печать протокола, то его печать осуществляется вручную, что отображается коэффициентом ручной печати протоколов нарушений:

$$K_{hand} = 1 - K_{auto}.$$

Выводы

В работе было проведено исследование систем распознавания государственных регистрационных знаков транспортных средств. Были выделены классы бизнес задач, ставящихся перед такими системами и проанализированы их характеристики.

Предложены метрики, позволяющие оценить качество систем распознавания в зависимости от поставленной бизнес задачи.

Представленные метрики могут быть положены в основу методики тестирования систем распознавания государственных номерных знаков транспортных средств.

Литература

1. Optical recognition of motor vehicle license plates [Text]/ P. Comelli, P. Ferragina, M.N. Granieri, F. Stabile // IEEE Trans. On Vehicular Technology/ — 1995. — vol. 44, No. 4. — pp. 790-799.
2. Wang T.-H. Robust License Plate Recognition based on Dynamic Projection Warping [Text] / Tsang-Hong Wang, Feng-Chou Ni, Keh-Tsong Li, Yon-Ping Chen // proc. IEEE Int. Conf. on Networking, Sensing & Control — 2004. — pp.784-788.
3. Wisam Al Faqheri. A Real-Time Malaysian Automatic License Plate Recognition (M-ALPR) using Hybrid Fuzzy [Text] / Wisam Al Faqheri and Syamsiah Mashohor//IJCSNS International Journal of Computer Science and Network Security — Vol.9 No.2. — 2009.— pp. 333-340
4. Fenton N. E. Software Metrics: A Rigorous and Practical Approach, 2nd ed. [Text] / N. E. Fenton, S. L. Pfleeger. — 2nd Edition Revised ed. — Boston, MA: International Thomson Computer Press, 1998. — 656 p.— ISBN 0534954251
5. Основы инженерии качества программных систем [Текст] : монография / Ф.И. Андон, Г.И. Коваль, Т.М. Коротун, В.Ю. Суслов ; НАН Украины. Ин-т прогр. систем. — К. : Академперіодика, 2002. — 502 с. — ISBN 966-8002-41-5
6. Дідковська М.В. Аналіз моделей оцінювання надійності програмного забезпечення [Текст] / М.В. Дідковська // Вісник НТУУ "КПІ". Інформатика, управління та обчислювальна техніка. — 2004. — №41. — С. 103-120.