

Метод поиска автомобиля по номеру с учетом ошибок распознавания

К.Л. Тассов¹, Д.А. Якунин¹

¹ МГТУ им. Н.Э. Баумана, Москва, 105005, Россия

Рассмотрена проблема поиска автомобиля по номеру в базе данных распознанных государственных регистрационных знаков (ГРЗ). Проанализированы причины ошибок распознавания и предложены алгоритмы поиска ГРЗ с учетом таких ошибок. Представлены выводы о работе данных алгоритмов на основе их программной реализации.

E-mail: tassov@mail.ru

Ключевые слова: ошибки распознавания, карты Кохонена, самоорганизующиеся карты Кохонена, мера схожести, перцептрон Розенблатта, государственный регистрационный знак.

Существующие системы распознавания ГРЗ дают большое число ошибок. Вероятность безусловного распознавания колеблется от 50 до 90 %. Ошибки сильно усложняют поиск номера по базе данных. Иногда операторам вручную приходится просматривать десятки тысяч записей для поиска нужного номера. В связи с этим становится актуальной задача сокращения выборки для ручного просмотра. Актуальность решения этой задачи также подтверждена отсутствием какой-либо информации о подобных системах в свободном доступе.

Целью данной работы является повышение достоверности поиска автомобиля по номеру в базе данных. Для достижения этой цели необходимо решить следующие основные задачи:

- проанализировать причины ошибок распознавания ГРЗ;
- разработать алгоритм поиска ГРЗ с учетом ошибок распознавания;
- разработать программное обеспечение для поиска ГРЗ на основе этого алгоритма.

Анализ ошибок. При распознавании могут быть допущены следующие ошибки [1].

1. Автомобиль пропущен. В этом случае отсутствует возможность восстановления информации о номере автомобиля.

2. Ошибочное выделение номера (например, вместо номерного знака может быть выделен текст на стекле автомобиля). В результате номер практически невозможно найти в базе.

3. Горизонтальное смещение рамки при выделении номера. В этом случае должна быть распознана часть номера. Существует возможность сокращения полной выборки для обнаружения автомобиля с данным номером.

4. Выбор неверного шаблона. Существующие системы распознавания ГРЗ прежде чем приступить к непосредственному распознава-

нию символов номера сначала проводят подбор шаблона номера (или типа ГРЗ). В случае ошибочного выбора типа номера последующее распознавание символов проходит с большей погрешностью. Тогда практически бесполезно использовать программные механизмы поиска.

5. Ошибка распознавания символов. Такая ошибка возникает при неверном распознавании символов ГРЗ в связи с зашумленностью изображения, засвеченным или испачканным номером. Данный тип ошибок можно использовать для повышения достоверности поиска.

Входные и выходные данные. Номера, распознанные системой распознавания ГРЗ, хранятся в базе данных. При поиске пользователем формируется запрос, в котором указываются искомый номер в виде строки и дополнительные данные о событии распознавания (время, место и пр.).

Запрос поступает в систему поиска, в которой генерируется список номеров: по ним в дальнейшем осуществляется поиск в базе данных. Номера должны генерироваться с учетом возможных ошибок распознавания. Иными словами, в этот список попадают те номера, которые потенциально могли быть распознаны вместо искомого номера.

На выходе для каждого номера из списка формируется таблица с обнаруженными в базе данных номерами, для каждой записи которой имеется фото номера, распознанный номер и другая информация. При удачной работе поисковой системы среди полученных таблиц должна быть обнаружена фотография искомого номера.

Алгоритмы, предлагаемые для поиска автомобиля по номеру.

Алгоритмы поиска с учетом горизонтального смещения рамки при выделении номера ГРЗ. Для того чтобы учесть возможное горизонтальное смещение рамки при выделении искомого номера для его поиска в базе данных, требуется:

1) определить шаблоны для поиска путем удаления символов с одного края искомого номера и добавления любого символа с другого края;

2) осуществить поиск номеров по шаблонам.

Алгоритм поиска с учетом ошибок распознавания символов.

Для того чтобы учесть ошибку распознавания для каждого символа искомого номера, требуется определить, какие символы могли быть распознаны вместо него. В связи с различной природой ошибок распознавания символов следует сформулировать различные критерии схожести символов, т. е. выделить меру схожести.

На основе выделения меры схожести основан следующий алгоритм:

1) выделение меры схожести символов;

2) подбор наиболее похожих для каждого символа искомого номера;

3) определение номеров, наиболее похожих на искомый номер, путем замены символов искомого номера на схожие с ними символы;

4) поиск по похожим номерам.

Для выделения меры схожести предлагаются такие варианты.

1. *Использование имеющейся статистики.* В процессе работы с системой распознавания оператор имеет возможность просматривать результаты распознавания и исправлять обнаруженные ошибки. Частота исправления одного символа на другой может свидетельствовать о частоте распознавания одного символа вместо другого. Информация об исправлениях собирается и может быть использована для выделения меры схожести.

2. *Выделение меры схожести с использованием самоорганизующихся карт Кохонена [2].* Карта обучается по эталонным изображениям символов. После обучения карты Кохонена мера схожести символов определяется как расстояние между ними на карте.

3. *Формирование меры схожести с использованием модели распознавания.* В данном алгоритме используется природа ошибок распознавания — сама система распознавания. При реализации данного алгоритма требуется обращаться к перцептрону Розенблатта для распознавания эталонных изображений символов алфавита, из которого состоят ГРЗ [3]. Подавая на вход вектор, соответствующий эталонному изображению символа, на выходе из сети получаем оценки возможностей распознавания данного символа вместо других символов алфавита. Такие оценки используются для формирования меры схожести.

Итак, в результате работы были проанализированы возможные ошибки при распознавании номеров, разработаны и реализованы алгоритмы поиска автомобиля по номеру в хранилище данных с учетом возможных ошибок распознавания.

Разработан программный продукт, позволяющий создавать и обучать нейронную сеть, распознавать с помощью нее символы ГРЗ и проводить поиск номера по хранилищу данных с учетом возможных ошибок распознавания. При поиске можно учитывать ошибки распознавания символов, используя меры схожести, полученные на основе самоорганизующихся карт Кохонена и с помощью системы распознавания.

В ходе проверки работоспособности программы была выявлена слабая корреляция двух реализованных методов.

Дальнейшее развитие проекта может включать:

- изучение и использование статистики, собираемой при работе оператора с системой распознавания;
- изучение явления подмены шаблонов ГРЗ при распознавании;
- совместное использование нескольких методов поиска с учетом ошибок.

СПИСОК ЛИТЕРАТУРЫ

1. Гонсалес Р., Вудс Р. Цифровая обработка изображений / пер с англ. под ред. П.А. Чочиа. М.: Техносфера, 2005. 1070 с.
2. Кохонен Т. Самоорганизующиеся карты. М.: БИНОМ. Лаборатория знаний, 2008. 656 с.
3. Медведев В.С. Нейронные сети / под общ. ред. В.Г. Потемкина. М.: Диалог-МИФИ, 2002. 496 с.

Статья поступила в редакцию 25.10.2012