

М.Э. КУССУЛЬ, О.Г. САДОВАЯ, А.С. СЫЧЕВ

СИСТЕМА РАСПОЗНАВАНИЯ ПЕШЕХОДА С ПОМОЩЬЮ ОДНОЙ ВИДЕОКАМЕРЫ

Abstract: This paper is dedicated to investigations of pedestrian detection system based on single video camera. The purpose of the investigations is to find limitations and possibilities of subsystem, which utilizes one frame for recognition. It is shown possibility to obtain neural network classifier for pedestrian recognition with accuracy up to 99,99%. Developed classifier is very important part of pedestrian recognition system based on video camera. In spite of high classification rate one classifier is not enough to solve the problem in the whole. Pedestrian detection system can be applicable if it based on two-steps classification.

Key words: optical recognition system, intelligent safety systems.

Анотація: Робота присвячена дослідженню систем розпізнавання пішохода за допомогою однієї відеокамери. Мета дослідження – виявити обмеження і можливості підсистеми, яка виконує розпізнавання по окремому кадру. Показана можливість отримання нейронної мережі типу класифікатор з точністю розпізнавання пішоходів до 99,99%. Розроблений класифікатор є надзвичайно важливою частиною системи розпізнавання образів за допомогою однієї відеокамери. Незважаючи на високий відсоток розпізнавання, одного класифікатора недостатньо для вирішення проблеми в цілому. Система розпізнавання пішохода може бути дієздатною, якщо вона базується на двохступеневій класифікації.

Ключові слова: система оптичного розпізнавання, інтелектуальні системи безпеки.

Аннотация: Статья посвящена исследованию систем распознавания пешехода с помощью одной видеокамеры. Цель исследований – выяснить ограничения и возможности подсистемы, использующей распознавание по отдельному кадру. Показана возможность получения нейронной сети типа классификатор с точностью распознавания пешеходов до 99,99%. Разработанный классификатор является чрезвычайно важной частью систем распознавания образов с помощью видеокамеры. Несмотря на высокий процент распознавания, одного классификатора недостаточно для решения проблемы в целом. Система распознавания пешехода может быть применима, если она основана на двухступенчатой классификации.

Ключевые слова: система оптического распознавания, интеллектуальные системы безопасности.

1. Введение

В последние годы автомобильная промышленность вместе с такими организациями, как National Highway Traffic Safety Administration (NHTSA) или Европейская комиссия, уделяют очень много внимания защите весьма уязвимых пользователей автомобильных дорог. Обнаружение пешехода – одна из наиболее важных проблем, решение которой способствует созданию эффективной системы его защиты. Многие исследователи склоняются к выводу, что визуальное распознавание пешеходов должно быть частью такой системы. Подход к обнаружению пешехода на дороге, основанный на видеозрении, широко обсуждается в литературе [1–5]. Однако многие опубликованные результаты относятся к «нормальным условиям», таким, например, как дневное время суток, когда есть ограничение на вариации освещенности, или дороги, которые не покрыты снегом, или нахождение пешехода непосредственно на проезжей части дороги, а не на борке или тротуаре и т.д. Другие исследования основываются только на движущихся объектах [6, 7]. Значительное число систем распознавания пешеходов используют нейронные сети или обучаемые части систем [2–7].

Несмотря на огромное количество проведенных в этой области работ, есть несколько вопросов, которые до сих пор требуют решения для разработки эффективных систем распознавания пешеходов, основанных на видеозрении, с использованием нейронных систем. Первый – это определение ограничений на визуальное распознавание пешехода. Вторым вопросом относится к архитектуре самой

системы, сбору данных и их маркировке для обучения нейронных сетей с учителем. Эта статья посвящена описанию нашей попытки ответить на поставленные выше вопросы.

2. Постановка задачи

В экспериментах была использована обычная цветная портативная видеокамера. Разрешение обрабатываемого изображения 360x288 пикселей и получение кадров со скоростью 24 кадра в секунду. Камера была установлена за ветровым стеклом автомобиля. Рабочая дистанция была определена в пределах от ~10 до ~35 метров при скорости автомобиля от 0 до 50 км/час.

Принимая во внимание видимые размеры пешехода на рабочей дистанции, возможный уклон дороги, качание автомобиля и расположение пешеходов у края дороги, мы считаем возможным ограничиться такими размерами входных изображений: от 100 до 280 пикселей по вертикали и от 10 до 350 пикселей по горизонтали. Края горизонтальной области исключены, чтобы избежать граничных эффектов во время алгоритмической обработки образа.

Для определения ограничений к визуальной системе необходимо собрать множество данных. Схема для сбора данных должна состоять из всевозможных условий, требуемых для анализа данных. Такая схема включает следующие условия:

- времена года: лето, весна\осень, зима;
- погодные условия: солнечно, пасмурно, идет дождь, идет снег, ветрено;
- взаимное поведение: автомобиль и пешеходы неподвижны, автомобиль и пешеходы движутся (включая резкие повороты автомобиля), движется или автомобиль, или пешеход;
- окружающая среда: разные варианты дорожного покрытия, различная ширина и уклон дороги (например, холмы). Присутствие деревьев, зданий, столбов, заборов, а также наличие других движущихся объектов, в частности, автомобилей.

Анализ собранных данных показывает, что в некоторых обстоятельствах информация о цвете изображения, получаемого от видеокамеры, может практически отсутствовать. Это зависит от времени года, погоды и освещения. Поэтому мы не можем очень рассчитывать на информацию о цвете объекта, и основные алгоритмы должны использовать черно-белые изображения. Так как мы рассматриваем возможность обнаружения неподвижных объектов, то не можем полагаться на технологию, которая использует вычитание образов (контуров).

Исходя из вышеизложенного, на архитектуру подсистемы распознавания пешехода можно наложить следующие ограничения: в качестве базового она должна использовать один кадр от видеокамеры (а не два, как в случае с вычитанием изображений) и только черно-белую информацию об объекте. Алгоритмы, обрабатывающие информацию о цвете объекта, равно как и технология, которая использует вычитание образов, могут играть только дополнительную роль. Конечно, в полноценную систему защиты пешехода должны входить алгоритмы слежения за пешеходом, сенсор ночного видения (IR camera) и сенсор измерения расстояния.

3. Маркировка данных

Сбор и маркировка данных имеют особую важность для решения такой сложной задачи, как распознавание пешехода в естественных условиях. Под маркировкой данных мы понимаем подготовку данных для обучения нейронной сети с правилом обучения – с учителем. Если схема сбора данных в основном обусловлена анализом проблемы, то правильность маркировки данных становится очевидной только после сравнения работы уже обученных систем на данных, маркированных различными способами.

Маркировка данных непосредственно связана с системой обнаружения объекта. В нашем случае система обнаружения объекта имеет две стадии: первый алгоритм предназначен для нахождения так называемого региона особого интереса (РОИ) – это прямоугольник, внутри которого потенциально может находиться изображение пешехода, а второй – это классификатор, предназначенный для определения наличия пешехода в РОИ. Используемый нами алгоритм быстрого поиска для нахождения РОИ основан на высокой вертикальной симметрии тела пешехода. Он анализирует контуры изображения: если в заданном прямоугольнике количество вертикальных контуров превышает некий порог, то данный прямоугольник считается РОИ.

Видимые размеры пешехода внутри рабочей зоны изменяются более, чем в 2 раза, это порождает некоторые проблемы, поскольку многие системы кодирования оперируют с фиксированным числом элементов. Есть несколько подходов избежать этого противоречия. Некоторые системы включают настройку РОИ под предопределенные размеры, другие изменяют дискретизацию зон с различными размерами.

Наилучшим в нашем случае оказался второй подход. Размер РОИ вычисляется следующим образом: для выбранной рабочей области мы определяем две перекрывающиеся области. Будем называть их зонами инспектирования (ЗОИ). ЗОИ_1 включает дистанцию приблизительно от 7 до 18 метров, для данной зоны РОИ определяется как прямоугольник в 57x140 пикселей. ЗОИ_2 включает расстояние от 16 до 40 метров с прямоугольником РОИ в 37x85 пикселей. Во время работы алгоритм нахождения РОИ сканирует картинку, чтобы покрыть рабочую зону соответствующими прямоугольниками, в которых проводится подсчет вертикальных контуров.

В основном, маркировка данных – это трудоемкий процесс, который занимает много времени и производится вручную, поэтому вопрос о ее автоматизации, хотя бы частичной, является достаточно актуальным. Нами был предложен алгоритм автоматизации, в основу которого легла наша же система распознавания пешехода. Суть его заключается в следующем: оператор вручную размечает небольшое количество первичных данных. На их основе проводится обучение нейронной сети, и настраиваются все необходимые алгоритмы для нахождения пешехода. Когда количество правильных ответов системы становится около 60–70% (как правило, для этого необходимо маркировать небольшое количество данных – около 10–15% от всего количества), систему распознавания пешехода используют в качестве помощника оператора при маркировке последующих изображений. Естественно, оператор производит контроль и исправляет ошибки системы. После маркировки следующей части данных система переучивается, качество ее работы повышается, что соответственно увеличивает продуктивность работы оператора.

Данный метод позволяет довольно сильно ускорить процесс маркирования данных, однако он имеет и существенные недостатки. Главным недостатком является неточность позиционирования прямоугольника относительно пешехода. Часто пешеход не попадает в центр данного прямоугольника либо частично выходит за его пределы. Если оператор будет исправлять такие ошибки, то ускорение процесса по времени не происходит. Если же не производить центрирование пешехода вовсе, то качество работы системы резко ухудшается. Как показали проведенные нами эксперименты, при использовании данных, маркированных полностью вручную, качество работы системы возрастет более, чем в 10 раз.

Однако, невзирая на отрицательный опыт, на наш взгляд, не следует полностью отказываться от данной идеи автоматизации процесса маркирования. В дальнейшем, если удастся дополнить данный алгоритм простыми и надежными алгоритмами центрирования объекта, он поможет резко сократить время работы оператора без существенных потерь качества.

Поскольку количество РОИ с пешеходами значительно меньше, чем РОИ без пешеходов, то для увеличения числа «пешеходов» и для стабильности обучения нейронной сети мы применили метод, известный в литературе по распознаванию образов, как «подвижка» или «дисторсия» [8]. Метод заключается в небольших перемещениях РОИ вокруг места, которое маркировано оператором как «пешеход». Каждый предварительно маркированный прямоугольник смещается относительно начальной позиции на $\pm 1-2$ пикселя по вертикали и ± 1 пиксель по горизонтали, включая взаимные (диагональные) сдвиги. Так, число «пешеходов» увеличивается в 15 раз. При таком увеличении размеров обучающей выборки качество работы системы увеличивается в 2-3 раза, что было проверено экспериментально.

Примеры собранных данных и результаты обнаружения пешехода показаны на рис.1.

4. Кодировка данных

Пешеход на картинках характеризуется пропорциями тела и наличием таких отличительных частей тела, как голова и ноги (черты лица и руки часто скрыты). Эти характеристики можно получить путем анализа контуров. Поэтому мы должны извлечь контуры пешехода настолько точно, насколько это возможно, так как это главные признаки для целей распознавания.

Перед кодированием мы применяем алгоритм предобработки, похожий на хорошо известный алгоритм обнаружения границ Саппу, и создаем бинарный образ, используя порог. Окончательно мы имеем несколько признаков, чтобы из входного образа закодировать объект:

- сглаженная яркость;
- градиенты яркости (границы);
- части контуров, полученные после бинаризации.

Полярность, контрастность цвета и другие параметры, настраиваемые на пиксель, которые обычно используются для текстурного анализа, могли бы быть полезны как дополнительные признаки. К сожалению, алгоритмы сегментации образа, основанные на текстурах, дают контуры с шумом, поэтому эти признаки не были использованы в обсуждаемой системе.

Наша кодировка должна отражать форму объекта и все те признаки, которые используются для определения контуров прямо или косвенно. Мы попробовали несколько подходов кодировки, которые могут быть условно поделены на две категории: “Кодировка на основе контуров” и “Кодировка на основе контрастности”.

А. Кодировка на основе контуров

Контуры тела пешехода и относительные размеры его частей определяют пешехода полностью.



Рис.1. Примеры собранных данных и результаты обнаружения пешехода

Существует две проблемы, связанные с кодировкой на основе контуров.

Первая проблема – исчезновение некоторых частей контуров на входном образе. Иногда такие отсутствующие куски составляют достаточно большую часть контура. Однако мы ожидаем, что нейронные сети будут способны распознать пешехода, используя только частичную информацию.

Другая проблема – выделение объекта. Важно определить контуры, что принадлежат к одному и тому же объекту. Без такого выделения в кодировку, основанную на контурах, может вноситься большое количество шума. В данном случае имеет место противоречие: чтобы распознать объект, мы должны найти его контуры, а чтобы найти контуры, мы должны знать, какой это объект.

В простейшем случае кодировка на основе контуров представляет собой некоторое множество гистограмм расстояний между контурами внутри РОИ. При построении гистограмм используются два направления: по вертикали и горизонтали. РОИ делится на три перекрывающиеся части вдоль одного из направлений, поэтому код состоит из 6 частей, каждый из которых представляет гистограмму из 19 элементов. Как показали проведенные нами эксперименты, увеличение числа направлений не улучшает качества распознавания.

В. Кодировка, основанная на яркости

При построении кодов, основанных на гистограммах, мы использовали информацию о контурах



Рис. 2. Примеры маркированных пешеходов

изображения, а для построения контуров мы используем информацию о градиентах яркости и контрастности. Логично было бы построить коды, непосредственно основанные на градиентах яркости, и таким образом избавиться от шума, вносимого при бинаризации изображения. Для сохранения рельефа градиентов мы разбиваем РОИ на достаточно малые прямоугольники (8 по вертикале и 4 по горизонтали, всего 32 прямоугольника (см. рис. 2)) и определяем наиболее представительные (максимальные) градиенты яркости вдоль выбранных направлений для каждого прямоугольника.

В дополнение к градиентам мы так же используем в коде средние яркости указанных прямоугольников.

Для того, чтобы малые колебания яркости не влияли на создаваемый код, они отсеиваются при помощи порога. По нашему мнению, взаимные позиции максимальных градиентов (позиции прямоугольников) несут ту же информацию, что и гистограммы расстояний.

5. Экспериментальные результаты

В системе использовались такие нейронные сети: многослойный перцептрон с Extended Delta-Bar-Delta [10, 11] правилом обучения. Мы использовали отдельные сети Net_1 и Net_2 для работы с РОИ_1 и РОИ_2 соответственно. При этом результаты распознавания имели приоритетность. Это значит, что если пешеход был найден в некотором РОИ_1, то тогда нет необходимости анализировать сеть РОИ_2.

Пример, как работает система, показан на рис. 1. Из примера видно, что система находит пешехода достаточно хорошо. Система обнаруживает пешехода во всех рассмотренных ситуациях: движущийся пешеход, неподвижный пешеход, частично невидимый пешеход (небольшая часть скрыта), группу пешеходов и резкий поворот автомобиля. Подсистема робастна для разных погодных условий.

Результаты распознавания нейронной сетью показаны в табл. 1. Нейронные сети были обучены

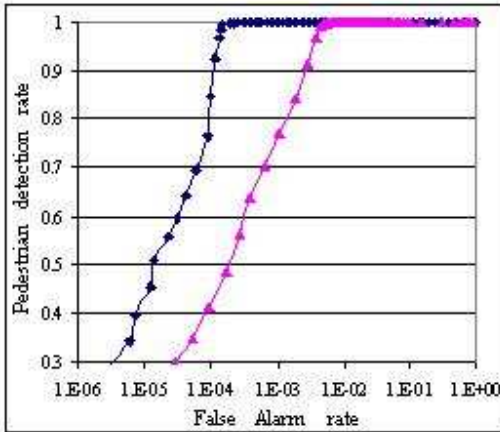


Рис. 3. ROC кривая: левая кривая - ROI_1, правая кривая - ROI_2

на нескольких группах данных, закодированных разными кодировками, описанными выше. Следующие обозначения используются в таблице для наименования разных кодировок: Н – гистограммы, G – максимум градиентов вдоль выбранного направления, В – средняя яркость.

Многие авторы указывают, что очень трудно дать количественные результаты обсуждаемых систем. Очевидно, что сравниваться могут системы, работающие в одинаковых условиях. Однако на данный момент принято использовать для сравнения таких систем так называемые ROC кривые [4]. ROC кривые для нашей системы, полученные для ROI_1 и ROI_2, показаны на рис. 3.

Как видно из представленных материалов, лучшие результаты были получены, когда использовались комбинации всех кодировок. При этом уровень распознавания варьировался на уровне 99,88% – 99,99%. Это значит, что на каждые 10000 проверяемых РОИ система делала 1–5 ошибок.

Таблица 1. Качество работы классификаторов, (%)

Coding	ZOI_1		ZOI_2	
	Train	Test	Train	Test
H+G+B	99.99	99.98	99.90	99.88
G+B	99.98	99.97	99.65	99.86
G	99.84	99.82	99.53	99.42

Тем не менее, даже такие высокие результаты не могут быть приемлемы для нашей задачи распознавания. Нетрудно посчитать, что если мы имеем ≈ 120

рассматриваемых РОИ на кадр с частотой 24 кадра в секунду и уровень распознавания составляет $\approx 99,97\%$, то мы будем иметь ≈ 1 ошибку в секунду.

6. Выводы

Как видно, невзирая на хорошие результаты классификации, разрабатываемая система еще очень далека от совершенства. На наш взгляд, дальнейшее улучшение системы должно идти по пути создания таких подсистем, как трекинг, выделение «частей тела» и т.д. Применение трекинга (принятие решения на основе нескольких последовательных кадров) позволит уменьшить количество ошибок.

Еще один путь – это введение дополнительных классов. Наибольшее число неправильных ответов получается, когда в РОИ попадают объекты или часть объекта, форма которого дает контуры,

которые похожи на форму тела пешехода. Это столбы, части ограждений, деревья, контуры автомобилей, лужи и элементы конструкций. Для уменьшения ошибок такого рода при тестировании системы был проведен следующий эксперимент. Был маркирован дополнительный класс «столбы», и обучена нейронная сеть, которая предназначалась для разделения столбов от пешеходов в том случае, когда основная подсистема распознавала РОИ как содержащую пешехода. При этом качество работы данной сети было около 100%. Этот результат убеждает, что введение таких дополнительных подсистем для двусмысленных объектов позволит резко сократить ошибки данного рода. Мы надеемся, что при помощи таких подсистем нам удастся получить приемлемое качество работы всей системы в целом.

В заключении отметим, что разработанная технология может иметь более широкий круг применения и потенциально может быть использована в любых системах, где есть необходимость производить обнаружение и классификацию нестандартных объектов на видеоизображении.

СПИСОК ЛИТЕРАТУРЫ

1. Gavrila D.M. Sensor-based pedestrian protection. *IEEE Intelligent Systems*. – 2001. – Vol. 16(6). – P. 77–81.
2. Gavrila D.M., Giebel J., Munder S. Vision-Based Pedestrian Detection: The PROTECTOR System // *Intelligent Vehicles Symposium*, IEEE. – 2004. – P. 16–18.
3. Gat I., Benady M., Shashua A. A Monocular Vision Advance Warning System for the Automotive Aftermarket. *SAE World Congress & Exhibition*. – 2005.
4. Shashua A., Gdalyahu Y., Hayun G. Pedestrian detection for driving assistance systems: single-frame classification and system level performance. *IEEE Intelligent Vehicles Symposium*. – 2004. – P. 1–6.
5. Куссуль М.Э., Садовая О.Г., Сычев А.С. Обзор алгоритмов и решений задачи оптического распознавания пешехода // *Математичні машини і системи*. – 2005. – № 4. – С. 110–115.
6. Tascillo A.L., DiMeo D.M., Macneille P.R., Miller R.H. Predicting a vehicle or pedestrian's next move with neural networks // *Proceedings of the 2002 International Joint Conference on Neural Networks, IJCNN '02*, 3:2310-2314. – 2002.
7. Heisele B., Wöhler C. Motion-Based Recognition of Pedestrians // *Int. Conf. on Pattern Recognition*. – Brisbane. – 1998. – P. 1325–1330.
8. LeCun Y., Bottou L., Orr G., Muller K. Efficient BackProp // *Neural Networks: Tricks of the trade / G. Orr and K. Muller, eds.* – 1998.
9. Haykin S. *Neural Networks: A Comprehensive Foundation*. – 2nd ed. – Englewood Cliffs, NJ, Prentice-Hall. – 1999.
10. Wöhler C., Anlauf J.K., Pörtner T., Franke U. A Time Delay Neural Network Algorithm for Real-Time Pedestrian Recognition // *IEEE Int. Conf. on Intelligent Vehicles*. – Stuttgart. – 1998. – P. 247–252.