

УДК 004.621.398

С.С. Довнар<sup>1</sup>, Ю.А. Скудняков<sup>2</sup>, Н.Н. Гурский<sup>1</sup>

## ПРОГРАММНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ РАСПОЗНАВАНИЯ И МОНИТОРИНГА МОБИЛЬНЫХ ОБЪЕКТОВ

Белорусский национальный технический университет<sup>1</sup>,  
Белорусский государственный университет информатики и радиоэлектроники<sup>2</sup>

Предложены к использованию математические модели, разработаны алгоритмы и программно-технические средства для распознавания и мониторинга мобильных объектов с использованием основных принципов и операций компьютерного зрения и нейронных сетей. Программное обеспечение с использованием аппаратуры в виде камеры видеонаблюдения (IP-камеры) позволяет следить за перемещениями мобильного объекта.

Использование полученных результатов позволяет повысить оперативность и качество распознавания и мониторинга мобильных объектов различного назначения.

*Ключевые слова:* математические модели; алгоритмы; программно-технические средства; распознавание; мониторинг; мобильные объекты; компьютерное зрение; нейронные сети; камеры видеонаблюдения.

### Введение

В настоящее время в системах обработки информации широко применяются алгоритмы из теории компьютерного зрения. Основные задачи, которые ставит перед собой теория, - анализ изображений с целью обнаружения, распознавания, отслеживания и классификации объектов.

Изучение компьютерного зрения проводится с 1960-х годов, а широкая распространенность задач, связанных с анализом изображений, и принципиальные сложности в разработке алгоритмов с течением времени только увеличивают количество публикаций и исследований в этой области.

На данный момент системы обработки информации с использованием компьютерного зрения применяются в таких сферах, как медицина, автоматизированные военные системы, автомобильная промышленность, поисковые системы, системы охраны объектов, системы контроля качества продукции и др. [1-2].

В данной работе исследуются возможности использования компьютерного зрения для распознавания и мониторинга мобильных объектов.

### Методы распознавания и мониторинга

В работе для решения задач распознавания и мониторинга мобильных объектов предложено использовать детектор Хаара, метод Виолы–Джонса, фильтр Калмана. Благодаря быстрой обработке изображения, можно с лёгкостью обрабатывать потоковое видео. Детектор Хаара используется для распознавания большинства классов объектов. К ним относятся лица и другие части тела людей, номера автомобилей, пешеходы, дорожные знаки, животные и т.д.

В настоящее время метод Виолы–Джонса является популярным методом для поиска объекта на изображении в силу своей высокой скорости и эффективности. В основу метода Виолы–Джонса положены: интегральное представление изображения по признакам Хаара, построение классификатора на основе алгоритма адаптивного бустинга и способ комбинирования классификаторов в каскадную структуру. Эти идеи позволяют осуществлять поиск объекта в режиме реального времени.

Интегральное представление изображения – это матрица, одинаковая по размерам с

исходным изображением. В каждом элементе матрицы хранится сумма интенсивностей всех пикселей, находящихся левее и выше данного элемента – правого нижнего угла прямоугольной области (0,0) до (x,y). Элементы матрицы L можно рассчитать по формуле:

$$L(x, y) = \sum_{i=0, j=0}^{i \leq x, j \leq y} I(i, j)$$

где  $I(i, j)$  – яркость пиксела исходного изображения.

Расчёт значений элементов матрицы проходит за время, пропорциональное числу пикселей в исходном изображении, поэтому интегральное изображение просчитывается за один проход.

Элементы матрицы рассчитываются по формуле:

$$L(x, y) = I(x, y) + L(x-1, y-1) + L(x, y-1) + L(x-1, y).$$

С помощью интегрального представления изображения можно быстро рассчитать суммарную яркость произвольной прямоугольной области на изображении.

Классификатор формируется на примитивах Хаара путём расчёта значений признаков. Для обучения на вход классификатора сначала подаётся набор «правильных» изображений с предварительно выделенной областью на изображении, дальше происходит перебор примитивов и расчёт значения признака.

На этапе обнаружения объекта в методе Виолы-Джонса используется окно определенного размера, которое движется по изображению. Для каждой области изображения, над которой проходит окно, рассчитывается признак Хаара, с помощью которого происходит поиск нужного объекта [3].

Ян ЛеКун предложил использовать для решения рассматриваемых задач так называемые сверточные нейронные сети [4].

Идея сверточных нейронных сетей заключается в чередовании сверточных слоев (C-layers), субдискретизирующих слоев (S-layers) и наличия полносвязных (F-layers) слоев на выходе.

Такая архитектура включает в себе три основные парадигмы: 1) локальное восприятие; 2) разделяемые веса; 3) субдискретизацию.

Локальное восприятие подразумевает, что на вход одного нейрона подается не все изображение (или выходы предыдущего слоя), а лишь некоторая его область.

Такой подход позволил сохранять топологию изображения от слоя к слою.

Концепция разделяемых весов предполагает, что для большого количества связей используется очень небольшой набор весов. Т.е. если у нас имеется на входе изображение размером 32x32 пикселя, то каждый из нейронов следующего слоя примет на вход только небольшой участок этого изображения размером, к примеру, 5x5, причем каждый из фрагментов будет обработан одним и тем же набором. Самых наборов весов может быть много, но каждый из них будет применен ко всему изображению. Такие наборы часто называют ядрами (kernels). Нетрудно посчитать, что даже для 10 ядер размером 5x5 для входного изображения размером 32x32 количество связей окажется равным примерно 256000 (сравниваем с 10 млн), а количество настраиваемых параметров всего 250!

На качестве распознавания это сказывается в лучшую сторону. Дело в том, что такое искусственно введенное ограничение на веса улучшает обобщающие свойства сети (generalization), что в итоге позитивно сказывается на способности сети находить инварианты в изображении и реагировать главным образом на них, не обращая внимания на прочий шум. Для решения данной задачи можно использовать другой подход с применением двумерных фильтров. Фильтр представляет собой матрицу коэффициентов, обычно заданную вручную. Эта матрица применяется к изображению с помощью математической операции, называемой *сверткой*. Суть этой операции в том, что каждый фрагмент изображения умножается на матрицу (ядро) свертки поэлементно и результат суммируется и записывается в аналогичную позицию выходного изображения. Основное свойство таких фильтров заключается в том, что значение их выхода тем больше, чем больше фрагмент изображения похож на сам фильтр.

Таким образом, изображение, свернутое с неким ядром, дает другое изображение, каждый пиксел которого будет означать степень похожести фрагмента изображения на фильтр.

Процесс распространения сигнала в С-слое показан на рис. 1.

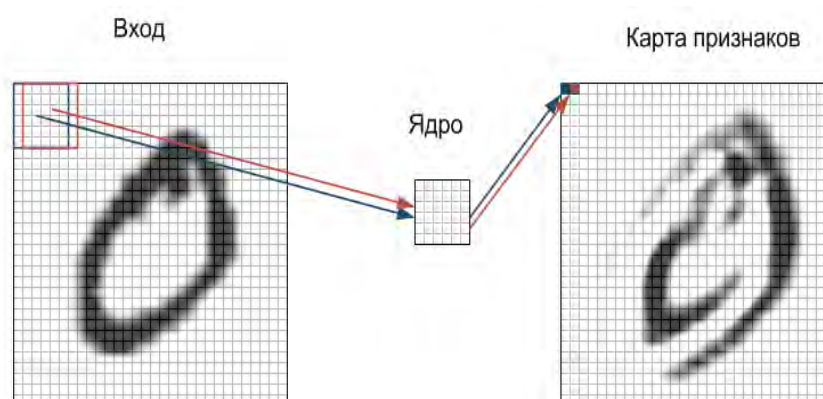


Рис. 1. Процесс распространения сигнала в С-слое

Каждый фрагмент изображения поэлементно умножается на небольшую матрицу весов (ядро), результат суммируется. Эта сумма является пикселом выходного изображения, которое называется картой признаков. Здесь опустим тот факт, что взвешенная сумма входов еще пропускается через функцию активации (как в любой другой нейросети). На самом деле это может происходить и в S-слое, принципиальной разницы нет. Следует отметить, что в идеале не разные фрагменты проходят последовательно через ядро, а параллельно все изображение проходит через идентичные ядра. Кроме того, количество ядер (наборов весов) определяется разработчиком и зависит от того, какое количество признаков необходимо выделить. Еще одна особенность сверточного слоя в том, что он немного уменьшает изображение за счет краевых эффектов.

Суть субдискретизации S-слоев заключается в уменьшении пространственной размерности изображения, т.е. входное изображение грубо (усреднением) уменьшается в заданное количество раз. Чаще всего в 2 раза, хотя может быть и не равномерное изменение, например, 2 по вертикали и 3 по горизонтали. Субдискретизация нужна для обеспечения инвариантности к масштабу.

Итак, в результате работы алгоритма выделения объектов может быть получен список обнаруженных на данном кадре сегментов, однако, чтобы решить задачу слежения за объектами, необходимо сопоставить каждому из этих сегментов объект, известный на предыдущем кадре, либо принять решение об обнаружении нового объекта. Иногда ошибочно обнаруженные сегменты являются достаточно устойчивыми во времени и существуют в близких точках пространства на протяжении нескольких кадров. С другой стороны, иногда правильно обнаруженные сегменты кратковременно исчезают, поэтому должен быть механизм, который, *во-первых*, принимает решение об обнаружении нового объекта, *во-вторых*, фильтрует неустойчивые во времени ложные сегменты, *в-третьих*, устанавливает соответствие между известными ранее объектами и новыми сегментами, *в-четвёртых*, прогнозирует координаты объектов при кратковременном пропадании, *в-пятых*, принимает решение о потере объекта.

Для придания алгоритму слежения устойчивости к временному закрытию объекта участками фона для каждого объекта вводится персональный траекторный фильтр, главная задача которого – прогноз координат объекта в следующем кадре на основе анализа траектории движения объекта. Обычно в качестве такого фильтра используется фильтр Калмана.

При временном закрытии объекта участками фона или другими объектами уточнение параметров траектории не выполняется, и траекторный фильтр работает в режиме прогнозирования координат.

При наличии в последовательности изображений множества объектов слежения воз-

никает вероятность перепутывания объектов в ситуациях пересечения траекторий их движения. После пересечения траекторий объекты должны получать те же идентификаторы, какие были у них до пересечения.

### Программно-техническая реализация

В процессе выполнения работы была создана программа, которая реализует алгоритм распознавания объекта на сцене и последующего его мониторинга.

Для получения видеоизображения с камеры используется протокол http. После получения видеоизображения, оно раскадровывается и далее работа происходит с каждым кадром по отдельности.

Основная часть программы работает в два (три) потока, количество потоков зависит от того, работает в данный момент мониторинг или нет. В одном потоке к изображению применяется детектор Хаара для поиска искомым примитивов (например: глаза, нос, рот или лицо целиком) на изображении, информация о найденном совпадении записывается в память компьютера, далее запускается в отдельном потоке мониторинг найденного объекта или нескольких объектов, запуск мониторинга может быть автоматическим или с подтверждением запуска у пользователя, в случае, когда на изображении детектор Хаара нашел, например, глаза, рот и лицо целиком. В ходе мониторинга записываются все изменения координат объекта, также имеется возможность вручную задать объект, за которым производится мониторинг. В случае ручного выбора объекта можно выбрать опцию поворота камеры вслед за объектом.

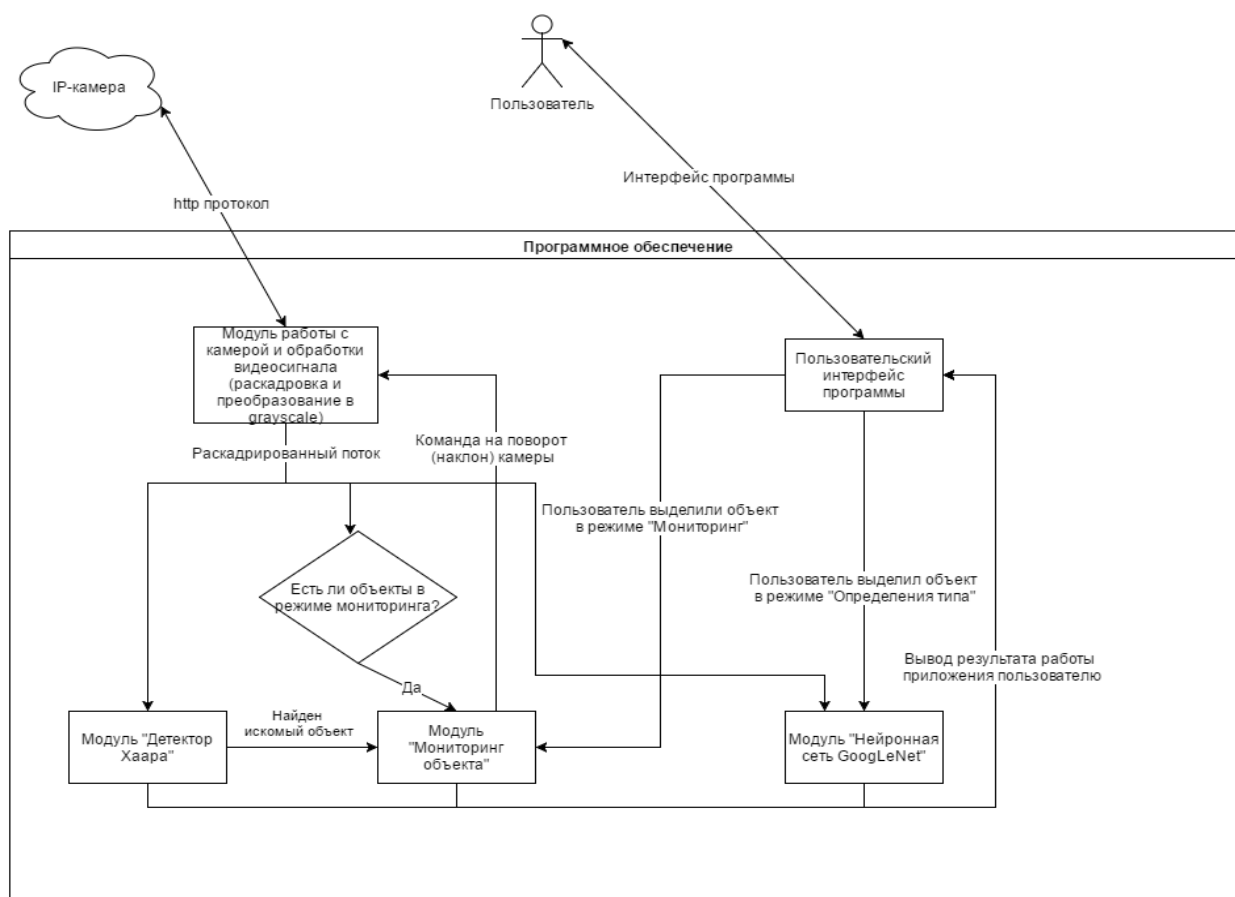


Рис. 2. Схема взаимодействия модулей программного обеспечения

Во отдельном потоке работает распознавание типа объекта на изображении с помощью нейронной сети GoogLeNet. Данный алгоритм запускается только в ручном режиме, т.е.

необходимо при помощи устройств ввода, обычно это компьютерная мышь, выделить искомый объект и дождаться результата полученного в ходе работы сверточной нейронной сети. Нейронная сеть GoogLeNet, точнее ее модель, реализована при помощи фреймворка глубокого обучения Caffe.

Программа реализована на языке C++, который был выбран из-за своей достаточно высокой скорости работы, при помощи библиотек libcurl для работы с http протоколом, библиотеки OpenCV [5-6], используемой для обработки изображений и фреймворка Caffe для реализации нейронной сети. Общая схема модулей приложения приведена на рис. 2.

В аппаратной части комплекса использовалась камера D-Link DCS-5030L, имеющая ¼ дюймовый мегапиксельный CMOS-сенсор с технологией прогрессивного сканирования, фокусным расстоянием 2,38 мм и углами обзора 94,36° (по горизонтали), 59,3° (по вертикали). Данная камера поддерживает технологию PTZ (Pan-Tilt-Zoom), т.е. она поддерживает дистанционное изменение зума и поворота и наклона камеры по вертикали и горизонтали, с углом поворота от -170° до +170°, углом наклона от -20° до +90°, скоростью вращения 16° в секунду и 4-кратным цифровым зумом. Также использовалась веб-камера A4Tech PK-635E, имеющая 1/6 дюймовый 0.3 мегапиксельный CMOS-сенсор с фокусным расстоянием 3,5 мм и углом обзора 54°.

На рис. 3 приводится принцип функционирования системы IP-камер.



Рис. 3. Принцип работы системы IP-камер

### Заключение

Таким образом, в работе были получены следующие результаты:

- изучены и предложены к использованию основные принципы и операции компьютерного зрения, методы, алгоритмы и нейронные сети для распознавания и мониторинга мобильных объектов на изображении;
- разработано программно – техническое обеспечение распознавания и мониторинга мобильных объектов.

Программное обеспечение, разработанное в данной публикации, позволяет осуществлять распознавание и мониторинг мобильных объектов и следить за их передвижением, оно напрямую взаимодействует с аппаратурой, в данном случае камерой видеонаблюдения (IP-камера), и позволяет в случае движения объектов двигать соответствующую камеру.

Предложенный в работе подход не дает абсолютно точного результата распознавания и мониторинга мобильных объектов и требует своего дальнейшего развития.

#### Библиографический список

1. **Форсайт, Дэвид.** Компьютерное зрение. Современный подход издательство /Дэвид Форсайт, Жан Понс. – М.: Изд-во Вильямс, 2004. – 928 с.
2. **Претт, У.** Цифровая обработка изображений / У. Претт. – М.: Мир, 1982.
3. **Васильев, К.К.** Статистический анализ многомерных изображений / К.К. Васильев. – Ульяновск: УлГТУ, 2002. – 156 с.
4. LeCun, Y. Convolutional Networks for Images, Speech, and Time-Series, in Arbib, M. A. (Eds), The Handbook of Brain / Y. LeCun, Y. Bengio // Theory and Neural Networks, MIT Press, 1995.
5. <http://docs.opencv.org/doc/tutorials/imgproc/> - документация по применению основных алгоритмов компьютерного зрения.
6. <http://www.lektorium.tv/course/22847> — курс «Введение в компьютерное зрение» факультета вычислительной математики и кибернетики МГУ: М., 2012.

*Дата поступления  
в редакцию 26.01.2017*

S.S. Dovnar<sup>1</sup>, Y.A. Skudnyakov<sup>2</sup>, N.N. Hurski<sup>1</sup>

#### SOFTWARE AND TECHNICAL SUPPORT FOR THE RECOGNITION AND MONITORING OF MOBILE OBJECTS

<sup>1</sup>Belarusian national technical University,

<sup>2</sup>Belarusian state University of Informatics and Radioelectronics

The proposed mathematical model, developed the algorithms and software and hardware tools for recognition and monitoring of mobile objects using basic concepts and operations of computer vision and neural networks. Software using hardware in the form of surveillance cameras (IP camera) allows you to monitor the movement of mobile object.

The obtained results allows to improve the efficiency and quality of recognition and monitoring of mobile objects for various purposes.

*Key words:* mathematical models; algorithms; software and hardware; detection; monitoring; mobile objects; computer vision;neural networks; surveillance cameras.