

Министерство образования Республики Беларусь

Учреждение образования

Белорусский государственный университет

информатики и радиоэлектроники

УДК 004.89

Ветров

Юрий Владимирович

Система распознавания объектов видеопотока

АВТОРЕФЕРАТ

на соискание степени магистра технических наук

по специальности 1-40 81 01 - Информатика и технологии разработки  
программного обеспечения

---

Научный руководитель

Одинец Д. Н.

кандидат технических наук

---

Минск 2017

## ВВЕДЕНИЕ

Искусственный интеллект с каждым днём меняет наш мир. Появление методов интеллектуального анализа данных позволило решить широкий круг задач классификации, прогнозирования и динамического управления, которые до какого-то момента считались частично или полностью не решаемыми с помощью классических методов.

В противовес математическим подходам, технологии, основанные на машинном обучении, моделируют способность человека к узнаванию и диагностике информации, получаемой непрерывным потоком из внешнего мира. Это позволяет приблизить уровень обработки информации к формату человеческого восприятия.

Актуальность проблемы распознавания образов и принятия решений постоянно возрастает. Такой интерес обусловлен необходимостью переработки больших потоков неструктурированной информации, расширением круга нечётко формализуемых интеллектуальных задач, а также постепенным переходом от пользовательского контроля технического средства к автономному и самоорганизующемуся управлению.

Особое место в интеллектуальном анализе данных занимает распознавание визуальной информации. Алгоритмы, позволяющие отслеживать и классифицировать объекты на фото или видео, в данном случае являются «глазами» более глобальной интеллектуальной системы. Программные модули, инкапсулирующие функциональность распознавания визуальных объектов – это фундамент для таких комплексов, как системы автопилота, охранные компоненты контроля территорий, приложения для идентификации личности и многих других решений.

Для того, чтобы реализованные библиотеки алгоритмов были готовы для работы с подобными задачами, необходимо, чтобы они отвечали ряду требований.

Первое необходимое условие, без которого само существование подобных алгоритмов теряет свой смысл, – это их инвариантность к контенту и представлению обрабатываемых данных. Спектр применения подобных систем достаточно обширный, и данные, которым необходимо обучить модуль распознавания, могут принимать самые различные формы: от относительно формализованных образов наподобие номерных знаков автомобилей (в случае систем контроля оплаты парковочных мест [1]) до объектов с достаточно высоким уровнем динамики – таких как люди (охранные системы защиты от несанкционированного проникновения либо системы идентификации личности

[2]), визуальные очаги внештатных ситуаций (программные комплексы отслеживания возгораний [3]) и даже домашние животные (автоматическая кормушка [4]).

Второе требование к исполнению подобных алгоритмов – их быстродействие. В большинстве случаев работа системы актуальна только при режиме функционирования в режиме реального времени – и в данном случае программной реализации придётся столкнуться с числом обрабатываемых кадров в одну секунду, измеряемых десятками. Помимо очевидных инженерных решений, таких как распараллеливание вычислений между отдельными потоками либо передача расчётов на видеокарту, а также реализация приложения в виде распределённой системы, могут применяться различные подходы, позволяющие упростить вычислительную сложность на логическом уровне обработки данных.

К третьему, достаточно важному требованию следует отнести переносимость: система должна одинаково хорошо функционировать на всех популярных платформах, как мобильных (Windows Phone, Android, iOS), так и десктопных (Windows, Linux, Mac OS). Библиотека с алгоритмами должна быть готова к встраиванию в любой программный комплекс, и поэтому все программные решения должны быть реализованы самостоятельно, без использования каких-либо сторонних библиотек, платформозависимых фреймворков и т. д.

Разрабатываемая система представляет собой программный комплекс, позволяющий отслеживать объекты на видеопотоке в режиме реального времени и классифицировать их в соответствии с набором знаний, сформированным на этапе обучения.

Обучение системы производится единожды на этапе задания её конфигурации. Ключевой особенностью системы является использование сложных и достаточно тяжеловесных вычислительных модулей, обладающих высоким уровнем точности распознавания, совместно с адаптивными логическими компонентами увеличения производительности, что в результате позволяет получить достаточную эффективность и необходимое быстродействие. Адаптивность в данном случае достигается за счёт динамической подстройки модулей, направленных на увеличение производительности системы, в режиме реального времени и в зависимости от особенностей текущей выполняемой задачи.

Разрабатываемая система может рассматриваться как набор базовой функциональности для более сложных приложений, требующих инструменты для сканирования и распознавания объектов видеопотока в режиме реального

времени, так и как самостоятельное приложение – например, для видеофиксации и логгирования различных логических объектов.

Целевой направленности на определённые платформы разрабатываемая система не имеет: программные реализации доступны как для мощных высокопроизводительных десктопных систем, так и для низкопроизводительных мобильных платформ. Итоговая эффективность разрабатываемой системы и области её применения будут определяться на этапе конечного пользования в зависимости от объёма и сложности (нелинейности) знаний, которые система получила в ходе её обучения.

Научная значимость разработанного решения находится на стыке дисциплин распознавания образов и машинного обучения (Pattern Recognition and Machine Learning) с одной стороны, и инженерных подходов к увеличению производительности – с другой.

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Разрабатываемая система представляет собой инструмент, позволяющий получать данные из внешнего мира в формате видеопоследовательности и в режиме реального времени производить сканирование объектов на этих видеоданных и их классификацию. Ключевым аспектом системы является модуль распознавателя, позволяющий предсказать координаты объектов на следующем шаге, значительно сэкономя вычислительные ресурсы как на сканирование, так и на классификацию.

Основным вычислительным ядром разрабатываемой системы являются нейронные сети. В распознавателе системы задействованы три сети:

- свёрточная нейронная сеть для сканирования объектов;
- свёрточная нейронная сеть для классификации объектов;
- рекуррентная нейронная сеть для оценки наиболее вероятных координат на следующем кадре.

Каждая нейронная сеть в данном случае является сетью глубокого обучения. Нейронная сеть классификатора является наиболее большой сетью из всех, что есть в системе, и, помимо основной оценки эффективности, по ней произведена оценка точности в сравнении с другими известными нейросетевыми топологиями.

Особое внимание уделено повышению быстродействия для возможности работы системы в режиме реального времени даже на низкопроизводительных устройствах. Программная реализация системы содержит, помимо очевидных логических и математических подходов к ускорению вычислений, структурные компоненты, позволяющие повысить скорость работы на системном уровне (CUDA).

Научная суть работы находится на стыке направлений распознавания образов и машинного обучения с одной стороны и инженерных подходов к увеличению производительности и организации вычислений – с другой.

По теме работы изучены литературные источники и проведены сравнения с имеющимися аналогами. Произведена оценка полученных результатов и тестирование итоговой программной реализации.

## КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Магистерская диссертация содержит две главы: теоретическая часть и экспериментальная часть.

В главе с теоретической частью приводится обзор литературы, описываются основные методы, которые используются для распознавания образов и машинного обучения, производится выбор подхода для реализации задачи, поставленной для магистерского исследования – в данном случае этим подходом является использование нейронных сетей глубинного обучения, приводится обоснование выбранного подхода и описывается система в целом.

Описание системы включает в себя логическую архитектуру разрабатываемой системы, формат топологий нейронных сетей, которые используются в распознавателе, обоснование этого формата, а также основные сведения, касающиеся использованных подходов к минимизации вычислений, увеличению производительности, организации процессов в распознавателе системы.

В главе с экспериментальной частью приводятся следующие блоки: описание программной реализации системы и её тестирование.

Описание программной реализации содержит перечень основных архитектурных моделей, задействованных в ходе проектирования и программирования системы. В этом подразделе описывается подход к разработке – в данном случае это объектно-ориентированное программирование, а также приводятся сведения о платформе, на которой велась разработка, используемом языке программирования, задействованных технологиях.

Блок тестирования содержит информацию о проведённых тестах итоговой программной реализации. Тестирование разбито на две части – часть, где проверялась точность системы в целом и различных её компонентов, и часть, где производилась оценка быстродействия.

## ЗАКЛЮЧЕНИЕ

Система, разработанная в данной магистерской диссертации, предназначена как для конкретных задач сканирования и классификации объектов, в режиме реального времени полученных из кадров видеопоследовательности, так и может являться основой для более комплексных систем в сфере охранных систем, программных комплексов автоматического управления транспортным средством и т. д.

В ходе разработки основной упор делался на корректный анализ данных даже в условиях отсутствия высокопроизводительных ресурсов на текущей платформе. Поэтому каждый модуль распознавателя системы – начиная от относительно простых рекуррентных сетей предиктора и оканчивая глубинной свёрточной сетью классификатора – работает в полной мере быстро и производительно. Для системы были проведены специальные нагрузочные тесты, которые позволили объективно оценить каждый из аспектов производительности.

Для центрального ядра распознавателя системы – классификатора – были произведены тесты на эталонных обучающих выборках, которые использовались для оценки и других нейронных сетей. Так, при обучении классификатора тренировочному набору CIFAR-10 и последующему тестированию на этом наборе, был получен результат 86,92%, что позволяет говорить о высоком уровне точности классификации (тест на собственных тренировочных данных, более приближенных к тем, с которыми будет работать система, показал результат 96%).

Программная реализация системы выполнена с помощью кроссплатформенных инструментов платформы Microsoft .NET Framework и поэтому является готовой к встраиванию в любой другой интеллектуальный комплекс, запрашивающий анализ поступающих видеоданных.

В качестве усовершенствования системы в дальнейшем планируется:

- увеличение мощности нейронной сети классификатора до разделения выборки на 1000 и более классов;
- организация распределённых вычислений в облаке;
- формирование собственной базы знаний за счёт подключения и обмена данными каждой реализации системы с сетью Интернет.

## СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

1. CERes Journal, Volume 1, Issue 1, 2015:

*«Handwriting recognition method in real-time mode based on distributed neural networks»*

Ключевые слова: рукописный текст, распознавание образов, нейронные сети.

Авторы: Ю. В. Ветров (магистрант каф. ЭВМ), Д. Н. Одинец (доцент каф. ЭВМ).

Доступ по ссылке: [www.ceres-journal.eu/iss150101](http://www.ceres-journal.eu/iss150101)

2. XII Международная научная конференция молодых учёных «Молодёжь в науке-2015»

*«Классификация объектов медицинских изображений на основе нейронных сетей»*

Ключевые слова: медицина, распознавание образов, нейронные сети, глубинное обучение.

Авторы: Ю. В. Ветров (магистрант каф. ЭВМ), Д. Н. Одинец (доцент каф. ЭВМ).

Доступ по ссылке: [nasb.gov.by/rus/news/add/yes2015/programm\\_2015.pdf](http://nasb.gov.by/rus/news/add/yes2015/programm_2015.pdf)