

Ю. И. Майборода, М. Ю. Синцов, А. Ю. Озерин, А. А. Кузин,
О. О. Варламов

Система автоматического тегирования изображений на основе миварных технологий

Аннотация. В данной статье рассматривается проблема создания системы автоматического тегирования изображений. В статье приведен обзор существующих задач и программных решений в области классификации объектов на изображении. Для решения задачи распознавания множества объектов на изображениях предлагается реализовать модули снятия контекстной неопределенности и управляющий модуль, производящий интеллектуальный запуск остальных модулей системы. Данные модули используют миварные технологии: миварную базу знаний и миварный метод логического вывода. В работе приведена модульная схема системы, описание работы модулей системы, а также полученные результаты.

Ключевые слова и фразы: искусственный интеллект распознавание изображений, тегирование изображений, мивар, миварные технологии.

Введение

Сегодня элементы искусственного интеллекта и, в частности, методы теории распознавания образов широко используются в автоматизированных системах во многих предметных областях. Например, решение задач из области распознавания образов востребовано в таких предметных областях, как военное дело, охрана правопорядка, медицина и т.д. Одной из таких востребованных задач является распознавание объектов на статичных изображениях.

На данный момент в области распознавания объектов на статичных изображениях можно выделить несколько основных направлений исследований. Так, ниже приведены типы задач, а также некоторые конкретные промышленные и/или исследовательские образцы с указанием разработчика данных систем:

-
- © Ю. И. Майборода, М. Ю. Синцов, А. Ю. Озерин, А. А. Кузин, О. О. Варламов, 2014
 - © Московский физико-технический институт (государственный университет), 2014
 - © Московский автомобильно-дорожный государственный технический университет, 2014
 - © Программные системы: теория и приложения, 2014

- (1) *Поиск изображений по образцу.* Например, существует проект IRMA (Image Retrieval in Medical Applications) [1] — совместный проект нескольких департаментов (Department of Diagnostic Radiology, the Department of Medical Informatics, Division of Medical Image Processing, the Chair of Computer Science VI) Ахенского технологического университета (RWTH Aachen). В рамках данного проекта разрабатывается система, позволяющая по входному статичному изображению выдавать подобные изображения. Работа ведется с медицинскими изображениями. Существуют и другие подобные проекты, которые реализуются в других областях знаний.
- (2) *Детектирование объектов на статичных изображениях.* Примеры систем, использующих методы детектирования и распознавания лиц:
 - программный продукт Google Picasa [2] компании Google;
 - программный продукт iPhoto [3] компании Apple;
 - встроенная функция детектирования распознавания лиц в социальной сети Facebook [4];
 - программный продукт Windows Live Photo gallery [5] компании Microsoft. В каждом из этих продуктов реализована технология детектирования и распознавания лиц для упрощения процедуры тегирования изображений, т.е., программа помогает находить фотографии для последующего тегирования, на которых она распознает одинаковые лица.
- (3) *Классификация объектов, присутствующих на статичном изображении.* Примерами таких систем служат следующие:
 - программный продукт Google Googles [6] компании Google;
 - исследовательский проект TextonBoost [7] компании Microsoft;
 - исследовательский проект Linkon [8] компании Microsoft.

Проекты Linkon и Google Googles направлены на классификацию (присвоение метки с названием категории) одиночных объектов на фотографиях, в то время как целью проекта Textonboost является классификация категорий всех объектов, присутствующих на статичном изображении.

Первая и вторая задачи имеют достаточно качественные решения, которые широко внедрены в автоматизированные системы. Что касается третьей задачи, то ее решения на данный момент являются весьма слабыми. Кроме чисто академического интереса, актуальность данной задачи заключается в том, что на данный момент

существует большое количество фотобанков и личных фотоархивов пользователей, которые не используют автоматическое тегирование изображений — теги проставляются вручную.

В данной статье описываются принципы построения системы для решения задачи автоматического тегирования изображений (задача классификации всех объектов, присутствующих на статичном изображении). То есть суть задачи — поступающему на вход изображению поставить в соответствие его максимально полное и точное текстовое описание, в котором будут присутствовать названия объектов, их свойства, и отношения между объектами.

1. Описание работы модулей системы

Интересующее пользователя изображение попадает на конвейер и проходит через модули предобработки, сегментации, детектирования, классификации, определения признаков, модуль построения графа вещь-свойство-отношение (ВСО) и модуль снятия контекстной неопределенности. При этом каждый из модулей может включать в себя один или несколько динамически подключаемых методов обработки изображения, что позволяет получить наиболее полную информацию об объектах предметной области (область знаний, ограниченная понятиями, контекстно связанными друг с другом) пользователя. По результатам математической обработки изображения строится граф вещь-свойство-отношение, который может быть использован для большого класса задач, связанных с интеллектуальной обработкой информации. Граф Вещь-Свойство-Отношение представляет собой универсальную форму представления человеческих знаний, причем, каждый из элементов графа может выступать в трех состояниях: как вещь, как свойство и как отношение.

В подразделах ниже представлено описание основных модулей системы (рис. 1).

1.1. Модуль предобработки и управляющий модуль

В рамках работы было исследовано большое количество методов сегментации, детектирования, классификации, а также определения признаков объектов. В зависимости от поставленной задачи, типа поступающей информации (изображение, видео, другие данные) и особенностей самого изображения для качественного тегирования изображения необходимо использовать различные методы обработки изображений. Для этой цели разрабатывается управляющий модуль,

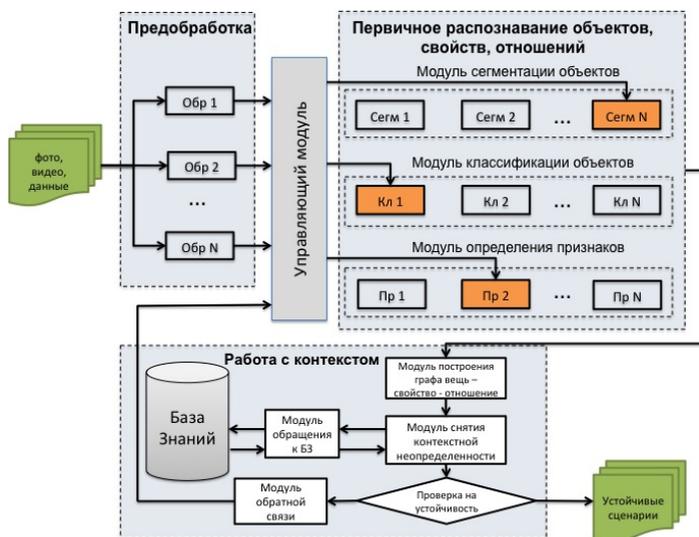


Рис. 1. Модульная схема системы

который действует по принципу производственной системы принятия решений: первоначально ряд предобработчиков анализируют входной поток информации, извлекая из нее характерные признаки, после чего на основании значений этих признаков принимается решение о запуске того или иного модуля (метода). Принцип работы производственной системы основан на миварном методе логического вывода, подробно описанный, например, в [9–16].

1.2. Модуль сегментации

В модуле сегментации происходит выделение масок объектов, т.е. из физически непрерывного набора визуальной информации в изображении строится дискретный набор сегментов — областей пикселей, связанных друг с другом по какому-либо критерию: общность цвета и/или текстуры, разделение общими границами, нахождение в общем фокусе и др. В дальнейшем, полученные сегменты ложатся в основу создания графа ВСО в качестве объектов. Используемые методы в данном модуле: Watershed, Grabcut, SWA, Graph Segmentation, SLIC.

1.3. Модуль классификации объектов

После модуля сегментации запускается модуль классификации объектов, где происходит непосредственно распознавание категорий объектов (наделение объектов именами категорий), таких как стул, стол и т.д. В зависимости от предметной области классификатор распределяет полученные сегменты к той или иной категории понятий. В бета-версии сервиса классификатор настроен на объекты из контекста «комната». Для классификации объектов используется концепция Bag of Words.

1.4. Модуль детектирования

Связка из сегментатора и классификатора представляет собой общий метод поиска объектов. В то же время, возможен поиск конкретных объектов на изображении (например, лиц, номеров машин, круглых объектов), что достигается применением модуля детектирования.

1.5. Модуль определения признаков

На данный момент в модуле происходит определение цвета, текстуры, формы и пространственных отношений (например, «стул находится слева от стола») между объектами. Реализовано определение пола и эмоций (возможные значения: «позитивный», «нейтральный», «негативный») для объекта «лицо человека».

1.6. Модуль построения графа вещь-свойство-отношение

После определения объектов, их свойств и связей между объектами строится граф вещь-свойство-отношение (описание миварного пространства дается, например, в [17–25]), где вещь — это категория объекта; свойства — цвет, текстура, форма (для объекта «лицо человека» свойствами также являются пол и эмоции); отношения — пространственные отношения.

1.7. Модуль снятия контекстной неопределенности и модуль обращения к миварной базе знаний

Модуль снятия контекстной неопределенности позволяет уточнять полученные теги. После построения графа ВСО, он может быть верифицирован с миварной базой знаний (описание миварного подхода к построению БЗ описывается в [17, 24]).

В результате применения алгоритмов машинного зрения может получиться информация, не являющаяся математически противоречивой, однако не имеющая смысла для человека. Для этого заведомо ложная информация из разбора удаляется в модуле верификации ВСО следующим образом: данный модуль через модуль обращения к миварной базе знаний получает по каждому объекту из графа вещь-свойство-отношение его семантическую сеть, которую сопоставляет с этим же графом. В результате этого сопоставления система принимает решение об удалении из тега свойств объектов и отношений между объектами. В дальнейшем, граф ВСО может быть сохранен, как отдельная модель, как часть другой модели.

1.8. Модуль обратной связи

На данный момент еще не реализован. С его помощью планируется бороться с оверсегментацией изображения.

2. Миварная база знаний

Важной частью системы является миварная база знаний. В базе знаний хранятся понятия и системные отношения между понятиями, которые показывают связь между ними. Понятиями могут быть существительные («стол»), прилагательные («зеленый»), глаголы («бегать») и другие части речи. Системные отношения представляют из себя словосочетания и бывают следующих типов:

- (1) Общее – частное: «машина – легковая»,
«млекопитающее – человек»;
- (2) Часть – целое: «туловище – рука», «часы – стрелка»;
- (3) Словосочетания имеющие смысл: «круглый мяч», «серый стул»;
- (4) Пространственные отношения: «кружка находится на столе»,
«стул находится под столом».

Системой автоматического тегирования изображений на данный момент используются системные соотношения типов: словосочетания имеющие смысл и пространственные отношения. Как было сказано в подразделе 1.7, в случае, если того или иного понятия или системного отношения из получаемого графа вещь-свойство-отношение нет в базе знаний, это понятие или отношение удаляется из графа.

ТАБЛИЦА 1. Текущие и планируемые технические характеристики системы

Параметр	Возможности прототипа	Развитие прототипа
Ввод данных	Загрузка фото через веб-интерфейс; веб-камеру, камеру Microsoft Kinect	Использование других типов камер, датчиков
Классификация объектов	Классификация по 10 произвольным категориям	Добавление детекторов произвольных объектов; распознавание категорий вплоть до количества объектов «текстового» словаря системы мивар-текст [22, 23, 25], повышение точности распознавания, создание иерархии образов
Определение признаков	Цвет, текстура, пространственные отношения, форма, определение пола и эмоций человека, определение сценария (например, «улица», «офис» и т.д.)	Определение следующих признаков: позы людей и животных, движения объектов
Сегментация	Реализованные методы: Water-shed, Grabcut, SWA, Graph Segmentation, SLIC	Повышение точности сегментации отдельных сегментаторов, создание модуля обратной связи

3. Полученные результаты

Реализован программный прототип на основе технологий, перечисленных в предыдущем разделе. Разработка велась на языке Python с использованием библиотеки OpenCV.

В табл. 1 представлены основные характеристики системы, а также характеристики системы, которые планируется достичь в будущем:

Заключение

В данной статье был представлен принцип построения системы автоматического тегирования изображений. Система кроме определения категорий объектов также позволяет определять и другие признаки объектов на изображении, такие как: цвет, текстура, форма, пространственные отношения, пол и эмоции человека. Также реализовано определение сценария изображения. Кроме стандартных методов детектирования, сегментации, классификации в данной системе используются миварные технологии (миварная база знаний и миварный метод логического вывода).

Миварные технологии должны позволить реализовать интеллектуальный запуск модулей в зависимости от параметров входящего изображения, а также модуль снятия контекстной неопределенности. Кроме того дальнейшими шагами в разработке системы являются борьба с оверсегментацией изображения и ложными срабатываниями классификатора. Также предполагается наращивание типов определяемых признаков: поз людей и животных, движений объектов.

Список литературы

- [1] http://ganymed.imib.rwth-aachen.de/irma/index_en.php ↑ 160.
- [2] <http://picasa.google.com> ↑ 160.
- [3] <https://www.apple.com/ru/mac/iphoto/> ↑ 160.
- [4] <https://www.facebook.com/help/463455293673370/> ↑ 160.
- [5] <http://windows.microsoft.com/ru-ru/windows-live/photo-gallery> ↑ 160.
- [6] <https://support.google.com/websearch/answer/166331> ↑ 160.
- [7] Jamie Shotton, John Winn, Carsten Rother, Antonio Criminisi, «TextonBoost for Image Understanding: Multi-Class Object Recognition and Segmentation by Jointly Modeling Texture, Layout, and Context», *International Journal of Computer Vision*, **81**:1 (2009), с. 2–23 ↑ 160.
- [8] <http://research.microsoft.com/en-us/um/people/larryz/objectrecognition.htm> ↑ 160.
- [9] Варламов О. О., «Разработка линейного матричного метода определения маршрута логического вывода на адаптивной сети правил», *Известия вузов. Электроника*, 2002, № 6, с. 43–51 ↑ 162.
- [10] Варламов О. О., «Разработка адаптивного механизма логического вывода на эволюционной интерактивной сети гиперправил с мультиактивизаторами, управляемой потоком данных», *Искусственный интеллект*, 2002, № 3, с. 363–370 ↑ 162.

- [11] Владимиров А. Н., Варламов О. О., Носов А. В., Потапова Т. С., «Программный комплекс «УДАВ»: практическая реализация активного обучаемого логического вывода с линейной вычислительной сложностью на основе миварной сети правил», *Труды НИИР*, 1 (2010), с. 108–116 ↑ 162.
- [12] Варламов О. О., «Практическая реализация линейной вычислительной сложности логического вывода на правилах «ЕСЛИ-ТО» в миварных сетях и обработка более трех миллионов правил», *Автоматизация и управление в технических системах: электрон. научн. журн.*, 2013, № 1(3), URL <http://auts.esrae.ru/3-66> ↑ 162.
- [13] Варламов О. О., Чибирова М. О., Сергушин Г. С., Елисеев Д. В., «"Облачная" реализация миварного универсального решателя задач на основе адаптивного активного логического вывода с линейной сложностью относительно правил «Если-То-Иначе»», *Автоматизация и управление в технических системах: электрон. научн. журн.*, 2013, № 2(4), с. 7–23, URL <http://auts.esrae.ru/4-78> ↑ 162.
- [14] Чибирова М. О., Сергушин Г. С., Варламов О. О., Елисеев Д. В., Хадиев А. М. и др., «Реализация общедоступного миварного универсального решателя задач на основе адаптивного активного логического вывода с линейной сложностью и облачных технологий», *Искусственный интеллект*, 2013, № 3, с. 512–523 ↑ 162.
- [15] Варламов О. О., Чибирова М. О., Сергушин Г. С., Елисеев Д. В., «Практическая реализация универсального решателя задач «УДАВ» с линейной сложностью логического вывода на основе миварного подхода и «облачных» технологий», *Приборы и системы. Управление, контроль, диагностика*, 2013, № 11, с. 45–55 ↑ 162.
- [16] Варламов О. О., *Основы миварного подхода к созданию логического искусственного интеллекта*, Учебное пособие, МАДИ, М., 2013 ↑ 162.
- [17] Варламов О. О., *Эволюционные базы данных и знаний для адаптивного синтеза интеллектуальных систем. Миварное информационное пространство*, Радио и связь, М., 2002, 288 с. ↑ 163.
- [18] Варламов О. О., «Основы многомерного информационного развивающегося (миварного) пространства представления данных и правил», *Информационные технологии*, 2003, № 5, с. 42–47 ↑ 163.
- [19] Варламов О. О., «Системы обработки информации и взаимодействие групп мобильных роботов на основе миварного информационного пространства», *Искусственный интеллект*, 2004, № 4, с. 695–700 ↑ 163.
- [20] Варламов О. О., «Создание интеллектуальных систем на основе взаимодействия миварного информационного пространства и сервисно-ориентированной архитектуры», *Искусственный интеллект*, 2005, № 3, с. 13–17 ↑ 163.

- [21] Варламов О. О., «О возможности создания интеллектуальных систем на основе GRID, систем адаптивного синтеза ИВК, сервисно-ориентированной архитектуры и миварного информационного пространства», *Известия Таганрогского государственного радиотехнического университета*, **54**:10 (2005), с. 130–140 ↑ [163](#).
- [22] Варламов О. О., Сергушин Г. С., Елисеев Д. В., Адамова Л. Е., Майборода Ю. И., Антонов П. Д., Чибирова М. О., «О миварном подходе к моделированию процессов понимания компьютерами смысла текстов, речи и образов. Новые возможности расширения границ автоматизации умственной деятельности человека», *Автоматизация и управление в технических системах: электрон. научн. журн.*, 2013, № 2(4), с. 30–45, URL <http://auts.esrae.ru/4-80> ↑ [163](#), [165](#).
- [23] Варламов О. О., Адамова Л. Е., Елисеев Д. В., Майборода Ю. И., Антонов П. Д., Сергушин Г. С., Чибирова М. О., «Комплексное моделирование процессов понимания компьютерами смысла текстов, речи и образов на основе миварных технологий», *Искусственный интеллект*, 2013, № 4, с. 15–27 ↑ [163](#), [165](#).
- [24] Варламов О. О., «Эволюционные базы данных и знаний. Миварное информационное пространство», *Известия Южного федерального университета. Технические науки*, **77**:2 (2007), с. 77–81 ↑ [163](#).
- [25] Varlamov O. O., Adamova L. E. E., Eliseev D. V., Mayboroda Yu. I., Antonov P. D., Sergushin G. S., Chibirova M. O., «Mivar Technologies in Mathematical Modeling of Natural Language, Images and Human Speech Understanding», *International Journal of Advanced Studies*, **3**:3 (2013), с. 17–23 ↑ [163](#), [165](#).

Рекомендовал к публикации

Программный комитет

Третьего национального суперкомпьютерного форума *НСКФ-2014*

Об авторах:



Юрий Игоревич Майборода

Аспирант Московского физико-технического института (государственный университет). Старший научный сотрудник НИИ Мивар, Москва, Россия.

e-mail:

y.majboroda@mivar.ru



Михаил Юрьевич Синцов

Аспирант Московского физико-технического института (государственный университет). Старший научный сотрудник НИИ Мивар, Москва, Россия.

e-mail: m.sintsov@mivar.ru



Алексей Юрьевич Озерин

Аспирант Московского физико-технического института (государственный университет). Младший научный сотрудник НИИ Мивар, Москва, Россия.

e-mail: a.ozerin@mivar.ru



Артур Азатович Кузин

Аспирант Московского физико-технического института (государственный университет). Младший научный сотрудник НИИ Мивар, Москва, Россия.

e-mail: a.kuzin@mivar.ru



Олег Олегович Варламов

Доктор технических наук. Профессор кафедры «Прикладная математика» ФГБОУ ВПО Московский автомобильно-дорожный государственный технический университет (МАДИ), Россия, Москва. Генеральный директор НИИ МИВАР, Москва, Россия.

e-mail: O.Varlamov@mivar.ru

Образец ссылки на эту публикацию:

Ю. И. Майборода, М. Ю. Синцов, А. Ю. Озерин, А. А. Кузин, О. О. Варламов. Система автоматического тегирования изображений на основе миварных технологий // Программные системы: теория и приложения: электрон. научн. журн. 2014. Т. 5, № 4(22), с. 159–170.

URL http://psta.pstiras.ru/read/psta2014_4_159-170.pdf

Yurii Mayboroda, Mikhail Syntsov, Aleksei Ozerin, Artur Kuzin, Oleg Varlamov.
Automatic tagging system based on mivar technologies.

ABSTRACT. This article considers the problem of development of an automatic tagging system. The review of the existing problems and software products in the sphere of image recognition is presented. For solving the problem of classification of all objects which exist in an image it is proposed to develop the modul of context ambiguity elimination and the control modul which controls intelligent launching of the other system moduls. These moduls use mivar technologies: mivar knowledge base and mivar method of logical inference. The article presents the modul scheme of the system and also the description of how the moduls of the system work. (*in Russian*)

Key Words and Phrases: artificial intelligence, image recognition, automatic tagging, mivar, mivar technologies.