

УДК 004.855.3

## СИСТЕМА ОБУЧЕНИЯ ВИЗУАЛЬНЫМ ПОНЯТИЯМ НА ОСНОВЕ СООТНЕСЕНИЯ ЛЕКСЕМ И КЛЮЧЕВЫХ ТОЧЕК

В.И. Филатов<sup>а</sup>, А.С. Потапов<sup>а</sup>

<sup>а</sup> Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

Адрес для переписки: sigal89@mail.ru

### Информация о статье

Поступила в редакцию 24.05.16, принята к печати 20.06.16

doi: 10.17586/2226-1494-2016-16-4-689-696

Язык статьи – русский

**Ссылка для цитирования:** Филатов В.И., Потапов А.С. Система обучения визуальным понятиям на основе соотношения лексем и ключевых точек // Научно-технический вестник информационных технологий, механики и оптики. 2016. Т. 16. № 4. С. 689–696. doi: 10.17586/2226-1494-2016-16-4-689-696

### Аннотация

**Предмет исследования.** Исследован процесс формирования визуальных понятий на основе двух неразмеченных потоков информации (визуального и текстового). **Метод.** Формирование визуальных понятий осуществляется путем соотношения визуальных паттернов с появляющимися одновременно с ними лексемами. Процесс обучения разбивается на два этапа: раннее зрительное обучение (первичное обучение), обучение на основе лексико-семантической информации (вторичное обучение). Первый этап заключается в построении словаря визуальных слов, служащего основой для второго этапа. Вторичное обучение предполагает анализ двух потоков информации, связанных по времени. В каждом канале проводится разделение информации на отдельные единицы, которые, в свою очередь, описываются на основе векторов признаков. Визуальные понятия формируются путем выявления связи между признаками, извлеченными из двух каналов информации. **Основные результаты.** Реализована система формирования визуальных понятий, которая протестирована на видеоданных, сопровождаемых субтитрами. Результаты тестирования показали принципиальную возможность формирования визуальных понятий данной системой. **Практическая значимость.** Описанная в работе система может быть использована для задач распознавания статических и динамических объектов, поиска изображений объектов в архиве видеоданных, автоматического создания баз визуальных понятий на основе произвольных видеофайлов, содержащих текстовое описание.

### Ключевые слова

обучение понятиям, визуальные понятия, интерпретация сцен, ключевые точки, дескрипторы, машинное обучение

### Благодарности

Работа выполнена при поддержке Министерства образования и науки Российской Федерации и частично при государственной поддержке ведущих университетов Российской Федерации (субсидия 074-U01).

## VISUAL CONCEPT LEARNING SYSTEM BASED ON LEXICAL ELEMENTS AND FEATURE KEY POINTS CONJUNCTION

V.I. Filatov<sup>а</sup>, A.S. Potapov<sup>а</sup>

<sup>а</sup> ITMO University, Saint Petersburg, 197101, Russian Federation

Corresponding author: sigal89@mail.ru

### Article info

Received 24.05.16, accepted 20.06.16

doi: 10.17586/2226-1494-2016-16-4-689-696

Article in Russian

**For citation:** Filatov V.I., Potapov A.S. Visual concept learning system based on lexical elements and feature key points conjunction. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2016, vol. 16, no. 4, pp. 589–696. doi: 10.17586/2226-1494-2016-16-4-689-696

### Abstract

**Subject of Research.** The paper deals with the process of visual concept building based on two unlabeled sources of information (visual and textual). **Method.** Visual concept-based learning is carried out with image patterns and lexical elements simultaneous conjunction. Concept-based learning consists of two basic stages: early learning acquisition (primary learning) and lexical-semantic learning (secondary learning). In early learning acquisition stage the visual concept dictionary

is created providing background for the next stage. The lexical-semantic learning makes two sources timeline analysis and extracts features in both information channels. Feature vectors are formed by extraction of separated information units in both channels. Mutual information between two sources describes visual concepts building criteria. **Main Results.** Visual concept-based learning system has been developed; it uses video data with subtitles. The results of research have shown principal ability of visual concepts building by our system. **Practical Relevance.** Recommended application area of described system is an object detection, image retrieval and automatic building of visual concept-based data tasks.

#### Keywords

concept learning, visual concepts, scene understanding, feature key points, descriptors, machine learning

#### Acknowledgements

This work was supported by the Ministry of Education and Science of the Russian Federation and partially by the Government support of leading universities of the Russian Federation (074-U01 subsidy).

### Введение

Решение задачи интерпретации сцен [1, 2] в частном виде выполняется во многих сферах информационных технологий. Однако существующие алгоритмы и методы решения такого рода задач обладают рядом существенных ограничений.

Частным случаем задачи интерпретации сцен является задача обучения понятиям, решение которой заключается в нахождении способа представления понятий в форме лексико-семантического описания. Многие современные задачи требуют для своего решения формирования понятий, дающих необходимую информацию об изображении сцены в момент ее анализа. Под «понятием» будем понимать семантическую единицу описания объектов, действий, событий. В качестве «визуальных понятий» определим понятия, формируемые на основе анализа сцен, представленных на изображениях или видеоданных, и лексических данных.

Наиболее часто встречающимися методами построения визуальных понятий являются:

- методы, использующие данные, размеченные частично или полностью вручную [3];
- методы на основе общих признаков [4–6];
- методы на основе лексических описаний [7];
- методы на основе синтеза аудио- и видеоданных [8];
- методы на основе иерархических представлений [9].

Одним из наиболее интересных подходов к формированию визуальных понятий является соотношение визуальных паттернов с появляющимися одновременно с ними лексемами. В работах Д. Роя [10–12] описан подход, где в качестве входной информации для анализа сцен используются два неразмеченных потока. Формирование понятий при такой постановке задачи осуществляется извлечением набора признаков из лингвистического и семантического информационных каналов с последующим их сопоставлением.

В настоящей работе осуществляется дальнейшее развитие этого подхода. Основным ограничением упомянутых работ Д. Роя является использование упрощенных моделей изображений, допустимых благодаря тому, что обучение проводится по простым видеоданным, содержащим единственный объект, хорошо отделяемый от фона. Мы используем современные методы анализа изображений естественных сцен, в частности, основанные на обнаружении и описании ключевых точек изображений. В разработанной нами системе на основе привлечения современных методов обработки изображений возможен анализ сложных сцен, содержащих произвольное количество объектов интереса и обладающих динамическим неоднородным фоном.

### Описание системы обучения визуальным понятиям

Входные данные разработанной системы формирования визуальных понятий – информация с сенсоров двух модальностей (визуальной и речевой), связанных по времени и в совокупности являющихся описанием наблюдаемой сцены. В качестве таких данных рассматриваются видеозаписи со звуковым сопровождением (в упрощенном случае аудиоканал может быть заменен текстовым описанием).

Для построения системы обучения визуальным понятиям необходимо обладать информацией об объектах интереса наблюдаемой сцены, поэтому первичным является определение количества объектов наблюдения в кадре. Количество таких объектов зависит от того, что принимать за объект интереса. В данной работе под объектом интереса будем понимать связанную область пикселей на изображении, представляющую некоторый объект. Объект может быть статическим или динамическим (в зависимости от того, предполагается ли наличие информации о движении для исследуемого объекта интереса).

Для детектирования динамических объектов используется алгоритм, осуществляющий вычисление поля оптического потока. Случай установления количества и положения статических объектов интереса менее тривиален и не может быть рассмотрен без задания дополнительных ограничений. Статическим объектом интереса на изображении является не произвольная область этого изображения, а область, для которой можно построить описание (например, в форме вектора признаков), по которому можно отличить

наблюдаемый объект от прочих. Для статических объектов интереса применяется разработанный алгоритм сегментации.

Одним из общих подходов к анализу изображений, который может быть, в частности, применен к описанию объектов интереса, является подход на основе дескрипторов (векторов локальных признаков) ключевых точек. Существует большое разнообразие методов детектирования ключевых точек, которые в основном сводятся к поиску экстремумов некоторой функции интереса. Среди наиболее часто используемых методов выделяют следующие:

- методы максимизации частных производных первого порядка (фильтры Собеля, Канни);
- методы максимизации вторых частных производных [13];
- корреляционные методы и аппроксимации (детектор Харриса [14], аппроксимация кривой);
- методы оценки оптического потока [15, 16].

Однако нередко используются специфические методы детектирования ключевых точек, разработанные совместно с методами их описания (т.е. построения дескрипторов). Примером может служить метод SURF [17], включающий как детектор, так и дескриптор ключевых точек.

Описание объекта интереса как совокупности дескрипторов ключевых точек является неустойчивым: от кадра к кадру будут меняться как сами дескрипторы, так и их количество. В разработанной нами системе вводится дополнительный уровень представления объектов интереса – уровень визуальных слов.

Визуальное слово – это кластер в пространстве дескрипторов с соответствующим центром. Определение визуальных слов – задача кластеризации, которая может быть решена заданием модели группирования данных. Такая модель осуществляет кластеризацию совокупности дескрипторов ключевых точек на основе их близости в пространстве признаков и формирует совокупность кластеров – визуальных слов. Центры кластеров, каждый из которых соответствует одному визуальному слову, будем называть дескрипторами визуальных слов.

Объект интереса тогда описывается не совокупностью дескрипторов, а совокупностью визуальных слов (кластеров в пространстве признаков), к которым эти дескрипторы относятся. Такая совокупность также называется «мешком визуальных слов».

Процесс анализа информации речевого канала выполняется схожим с визуальным каналом образом. Сначала неразмеченный поток информации разбивается на набор отдельных лексических единиц (в случае текстового потока лексической единицей будет отдельное слово). Частотная характеристика появления лексем в речевом канале будет являться оценкой априорного (безусловного) распределения вероятностей появления отдельных слов для текстовых данных. В случае привлечения аудиальной информации необходимо ввести описание лексических единиц на основе вектора признаков, извлекаемых из речевого канала.

Для формирования отдельных наборов визуальных и аудиальных слов необходимо совместное рассмотрение наборов слов и определение лексико-семантических пар (визуальных понятий), описывающих объекты интереса. В качестве критерия выделения визуального понятия служит взаимная информация аудиального и визуального слова, оцениваемая на основе вероятности одновременного (в рамках некоторого временного интервала) появления аудиальных и визуальных слов в наблюдаемой видеопоследовательности.

В предлагаемом подходе процесс обучения разделен на два основных этапа: раннее зрительное обучение (первичное), обучение на основе лексико-семантической информации (вторичное). Первый этап характеризуется извлечением набора признаков из визуального канала для построения словаря визуальных слов, служащего основой для второго этапа. Словарь визуальных слов используется в качестве промежуточного уровня представления для описания наблюдаемых сцен. Совокупность визуальных слов, построенных на основе извлеченных из изображения ключевых точек и их дескрипторов, является промежуточным описанием интерпретируемой сцены.

Этап обучения на основе лексико-семантической информации предполагает нахождение способа объединения информации связанных потоков информации (визуального и речевого). Для этой цели сначала выполняется извлечение набора признаков в каждом из каналов по отдельности. Исследование связи между извлеченными наборами признаков сводится к оценке совместной плотности распределения вероятностей появления визуальных слов и лексем.

### Основные алгоритмы системы обучения визуальным понятиям

**Алгоритм формирования словаря визуальных слов.** Процесс раннего зрительного обучения служит для формирования словаря визуальных слов, являющегося базовым представлением при интерпретации сцен в разрабатываемой системе. В качестве входных данных алгоритма используются анализируемая видеопоследовательность, формируемое количество визуальных слов и шаг выбора кадров из видеопоследовательности.

В каждом обрабатываемом кадре осуществляется выделение набора ключевых точек на основе детектора ключевых точек SURF [17]. Каждая ключевая точка описывается 64-компонентным вектором признаков (дескриптором) SURF. После формирования наборов дескрипторов по кадрам сцены выполня-

ется процесс их группирования (по алгоритму кластеризации) на заданное количество кластеров (которое соответствует количеству визуальных слов, являющемуся входным параметром алгоритма) с помощью алгоритма  $k$ -средних. Сформированный словарь визуальных слов описывается и хранится как набор дескрипторов визуальных слов.

**Алгоритм разметки сцен кадра.** Текстовое описание в представленной нами системе задается на основе предоставляемого на вход системы файла субтитров в формате \*.srt, содержащего информацию для лексического анализа. Структура файла субтитров содержит данные о времени начала и конца субтитра, номер записи субтитра и текст субтитра. В нашей системе временные интервалы субтитров переводятся в номера кадров; проводится формирование вектора меток, где каждый элемент обозначает индекс субтитра, используемый в системе в качестве индекса для номера сцены. После этого осуществляется разметка исходного текстового потока на набор сцен кадров. По завершении процесса разметки выполняется выделение информативной составляющей текста субтитров и проводится ее анализ с целью установления совокупности отдельных лексем (отдельных текстовых слов).

**Алгоритм выделения ключевых точек на основе словаря визуальных слов.** Основная задача предлагаемого алгоритма заключается в формировании набора ключевых точек на основе имеющейся информации в словаре визуальных слов и входного кадра изображения. В стандартных методах (SURF, SIFT) выделение ключевых точек осуществляется отдельно от их описания на основе функции интереса, не учитывающей дальнейшее использование ключевых точек. Цель предлагаемого алгоритма – найти такие ключевые точки, которые обладают наибольшей информативностью для описания объекта интереса в виде «мешка визуальных слов».

Однако, чтобы можно было использовать разработанный алгоритм выделения ключевых точек, необходимо наличие словаря визуальных слов. В связи с этим в системе имеется два режима детектирования ключевых точек и построения визуальных слов на их основе – SURFWORD и VISWORD.

Режим работы SURFWORD не является оригинальным. Его основная идея заключается в привлечении словаря визуальных слов к детектору ключевых точек SURF с последующей маркировкой ключевых точек индексом ближайшего (по евклидовой метрике) визуального слова в пространстве признаков. Для дескриптора каждой ключевой точки определяется расстояние в пространстве признаков до каждого дескриптора визуального слова. Индекс ближайшего визуального слова заносится в карту индексов, а значение расстояния – в карту расстояний. После формирования карты индексов и карты расстояний каждая ключевая точка маркируется индексом визуального слова с учетом информации, хранящейся в карте индексов.

В отличие от алгоритма SURFWORD, в алгоритме VISWORD выделение ключевых точек в кадре осуществляется не на основе полученных в процессе работы SURF ключевых точек, а путем нахождения таких пикселей на изображении, которые лучше всего соответствуют имеющимся визуальным словам. С этой целью сначала для каждого пикселя вычисляются характеристики ориентации ключевой точки детектора SURF и выполняется построение дескриптора. Затем для каждого дескриптора определяется индекс ближайшего визуального слова и расстояние до него. В качестве ключевых точек выделяются точки, которые соответствуют локальным минимумам на построенной карте расстояний. При этом для выделенных ключевых точек сразу оказывается известным, каким визуальным словам они принадлежат. Результат работы алгоритма – набор ключевых точек и их дескрипторов с меткой визуального слова из словаря визуальных слов.

**Алгоритм уточнения ключевых точек на основе поля оптического потока.** Для алгоритма уточнения ключевых точек на основе поля оптического потока входными и выходными данными является набор ключевых точек. Данный алгоритм предназначен для уточнения необходимой информации о движении объектов интереса динамического характера. В качестве базового для построения поля оптического потока используется алгоритм Гунера–Фернебаха [18, 19]. Эксперименты в работе [7] показали возможность использования пары алгоритмов: алгоритм построения поля оптического потока и алгоритм выделения ключевых точек, что позволяет повысить инвариантность по отношению к изменению освещения.

**Алгоритм сегментации на основе словаря визуальных слов.** Данный алгоритм может быть использован как в качестве средства для уточнения ключевых точек кадра (выделение объектов интереса на основе принципа объект-фон), так и для маркировки пикселей изображения для установления количества объектов.

Входными данными алгоритма является набор пикселей изображения, каждый из которых представляется на основе вектора признаков  $\mathbf{x} = \{r, g, b, x, y, \mathbf{d}\}$ , где  $r, g, b$  – значения RGB-компонент цвета пикселя,  $x, y$  – координаты пикселя,  $\mathbf{d} = \{d_1, \dots, d_N\}$  – вектор, содержащий расстояния от дескриптора ключевой точки текущего пикселя до всех дескрипторов визуальных слов в словаре размера  $N$ .

Так как вектор расстояний содержит значения расстояний в пространстве признаков между дескриптором текущего пикселя и дескриптором каждого визуального слова в словаре, качество сегментации определяется словарем визуальных слов, сформированным в процессе раннего зрительного обучения.

После построения соответствующего вектора признаков  $\mathbf{x}$  для каждого пикселя осуществляется группировка на основе вектора признаков. Для этого в алгоритме применяется метод  $k$ -средних. Полученные центры кластеров задают индексы для меток объектов интереса на кадре. Критерий, задающий предполагаемое количество объектов интереса, устанавливается на основе принципа минимальной длины описания.

**Алгоритм лексико-семантического обучения.** Данный алгоритм относится к процессу вторичного обучения и заключается в оценивании совместной встречаемости дескрипторов визуальных слов и лексем. На основе работы алгоритма устанавливаются оценки маргинальных и совместного распределения вероятностей появления дескрипторов визуальных слов и лексем.

#### Обработка экспериментальных данных и анализ результатов

В ходе работы было проведено экспериментальное исследование системы на основе набора видеопоследовательностей с субтитрами. Примеры полученных результатов представлены ниже. На рис. 1, 2 отображены примеры результатов вычислений программы по двум различным сценам наблюдения, полученных в процессе обучения системы. В качестве режима детектора ключевых точек на представленной серии изображений был использован алгоритм VISWORD. В одном из примеров объекты интереса предполагались динамическими (рис. 1), а в другом статическими (рис. 2).

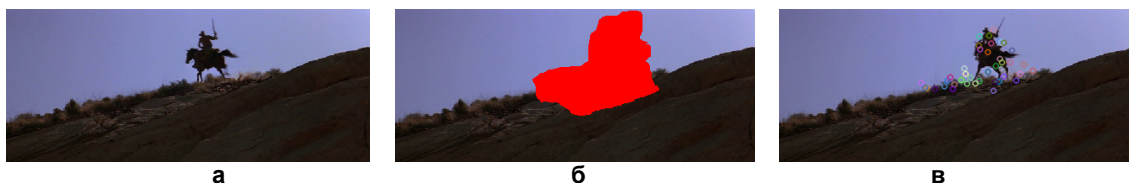


Рис. 1. Исходное изображение (а), поле оптического потока (б) и изображение с выделенными ключевыми точками (в) на основе алгоритма VISWORD

Альтернативой использованию оптического потока (динамические объекты интереса) является применение алгоритма сегментации для нахождения областей предполагаемых объектов интереса без учета информации о движении. Результат полученных данных алгоритма сегментации визуальных слов представлен на рис. 2.



Рис. 2. Исходное изображение (а), кадр с выделенными ключевыми точками (б) и результат разработанного алгоритма сегментации (в)

Для алгоритма сегментации также указаны информационные параметры (рис. 3): карта индексов, карта расстояний (отображены путем нормировки значений в диапазоне от 0 до 255) и гистограмма визуальных слов, встретившихся в текущем кадре. Значения для карты индексов заданы в пределах от 0 (белый цвет) до 64 (черный цвет), что соответствует диапазону индексов слов в словаре. Значения карты расстояний отображаются переходом цветов от белого к черному по мере увеличения хранимых расстояний. Гистограмма визуальных слов хранит относительные частоты встречаемости визуальных слов в кадре.

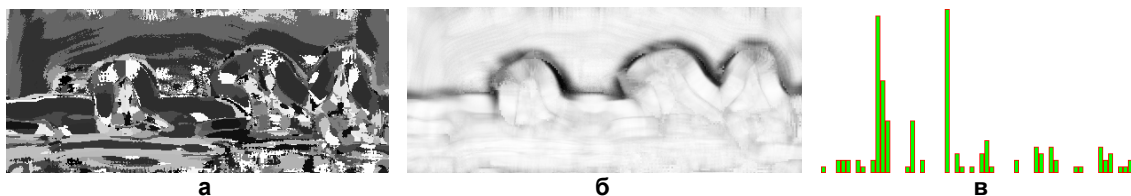


Рис. 3. Карта индексов визуальных слов (а), карта расстояний до ближайшего визуального слова (б), гистограмма визуальных слов в кадре (в)

В процессе выполнения лексико-семантического обучения (вторичного обучения) устанавливается вид совместного распределения вероятностей появления дескрипторов визуальных слов и лексем. Следуя алгоритму разметки сцен кадров, в каждой сцене определяются вероятности встречаемости лексем и дескрипторов визуальных слов (например, для изображения на рис. 2 получены соответствующие лексемы и значения функции правдоподобия: horse –  $2.498e-37$ , hunter –  $3.024e-39$ , man –  $1.623e-60$ ). Полученные значения маргинальных распределений лексем и дескрипторов визуальных слов, а также их совместного распределения вероятностей сохраняются в виде бинарных файлов. Пример получившихся в результате обучения графиков для вероятностей представлен на рис. 4–7.

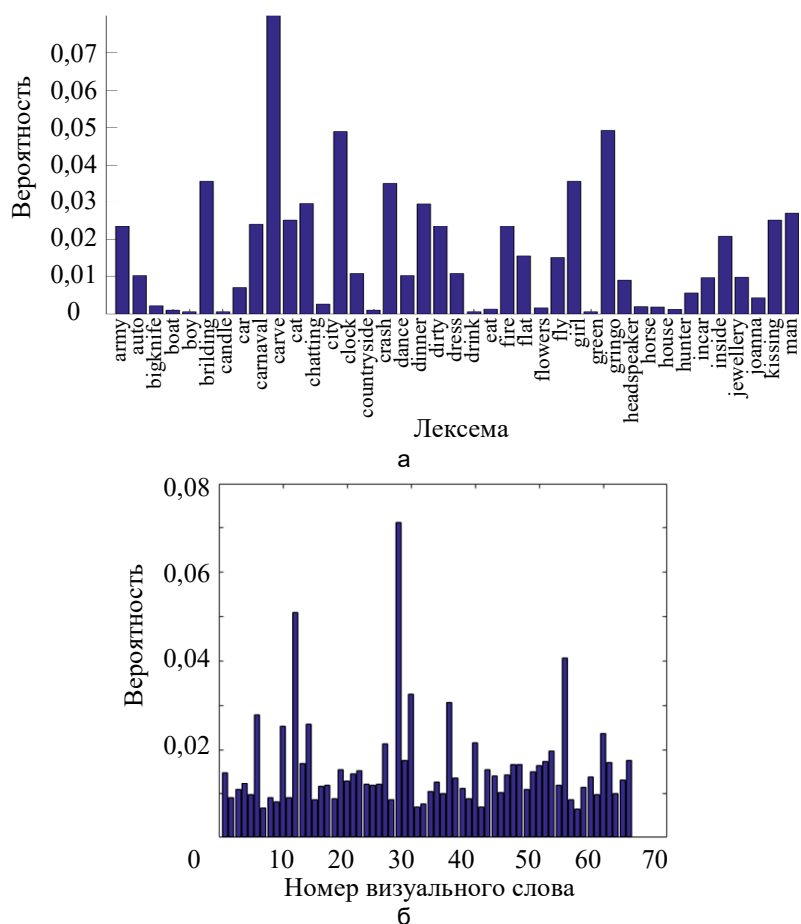


Рис. 4. Маргинальные распределения лексем (а) и визуальных слов (б)

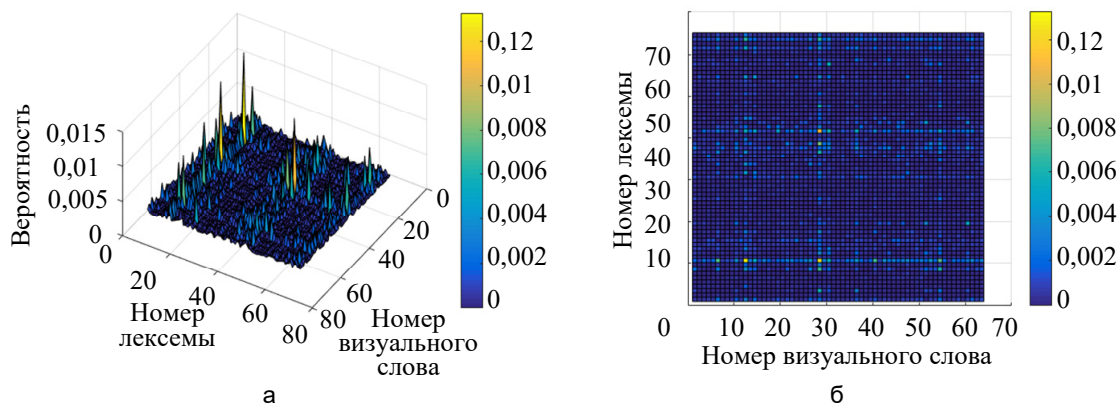


Рис. 5. Совместное распределение вероятности появления визуальных слов и лексем (а, б)

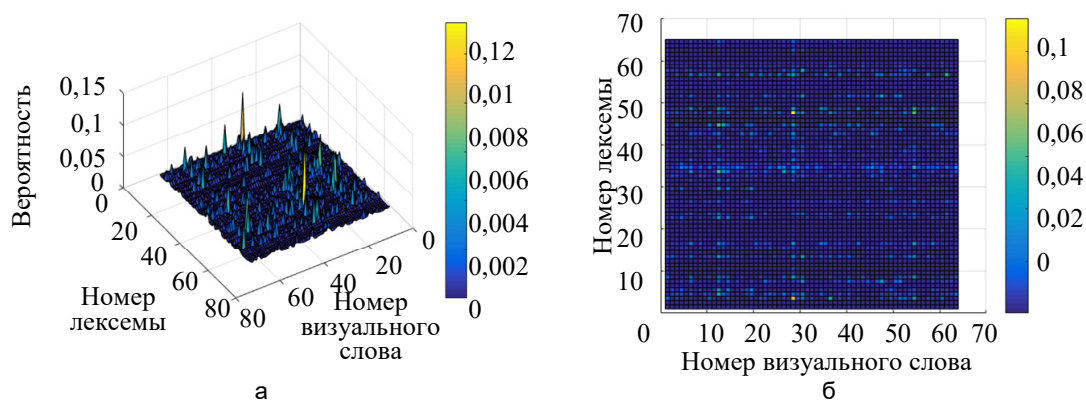


Рис. 6. Условное распределение вероятности появления визуальных слов и лексем (а, б)

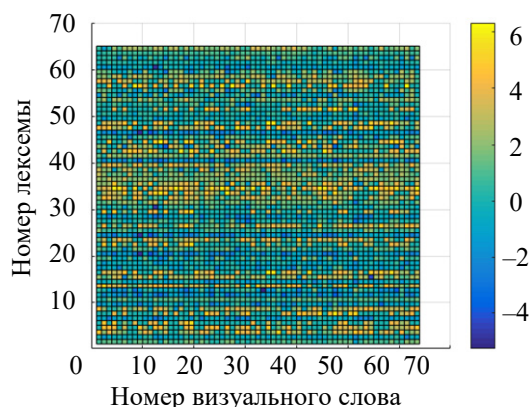


Рис. 7. Количество взаимной информации между парами лексем и визуальных слов

Как видно из представленных графиков, условное распределение вероятностей появления визуальных слов и лексем не является равномерным, а взаимная информация между ними может быть существенно ненулевой, что свидетельствует о возможности установления связи между наблюдаемыми на сцене объектами и соответствующими им лексемами.

### Заключение

Анализ полученных результатов показывает возможность применения алгоритмов обучения визуальных понятий для анализа видеоданных. Полученное условное распределение вероятностей появления дескрипторов визуальных слов при заданной лексеме подтверждает наличие зависимости между данными. Установлена возможность обучения системы на основе словаря визуальных слов. Система позволяет проводить автоматизированное обучение объектам сцен, содержащим нескольких объектов интереса и обладающим сложным динамическим фоном.

### Литература

1. Ito S., Yoshioka M., Omatu S., Kita K., Kugo K. An image recognition method by rough classification for a scene image // *Artificial Life and Robotics*. 2006. V. 10. N 2. P. 120–125. doi: 10.1007/s10015-005-0353-9
2. Ko B., Peng J., Byun H. Region-based image retrieval using probabilistic feature relevance learning // *Pattern Analysis and Application*. 2001. V. 4. N 2–3. P. 174–184. doi: 10.1007/s100440170015
3. Li Z. Learning Visual Concepts from Social Images: Master Thesis. Leiden Institute of Advanced Computer Science, Netherlands, 2011. 18 p.
4. Fergus R., Fei-Fei L., Perona P., Zisserman A. Learning object categories from Google's image search // *Proc. 10<sup>th</sup> IEEE Int. Conf. on Computer Vision*. Beijing, China, 2005. V. II. P. 1816–1823. doi: 10.1109/ICCV.2005.142
5. Fei-Fei L., Fergus R., Perona P. One-shot learning of object categories // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2006. V. 28. N 4. P. 594–611. doi: 10.1109/TPAMI.2006.79
6. Grauman K., Darrell T. The pyramid match kernel: discriminative classification with sets of image features // *Proc. 10<sup>th</sup> IEEE Int. Conf. on Computer Vision*. Beijing, China, 2005. V. II. P. 1458–1465. doi: 10.1109/ICCV.2005.239
7. Filatov V.I. Concept-based lexical-semantic unsupervised learning system // *Journal of Physics: Conference Series*. 2014. V. 536. N 1. Art. 012016. doi: 10.1088/1742-6596/536/1/012016.
8. Naphade M., Huang T. A probabilistic framework for semantic video indexing, filtering and retrieval // *IEEE Transactions on Multimedia*. 2001. V. 3. N 1. P. 141–151. doi: 10.1109/6046.909601
9. Ranzato M. Unsupervised Learning of Feature Hierarchies: PhD thesis. New York University, 2009. 167 p.
10. Roy D. Learning from Sights and Sounds: a Computational Model: PhD thesis. MIT, 1999. 176 p.
11. Roy D.K., Pentland A.P. Learning words from sights and sounds: a computational model // *Cognitive Science*. 2002. V. 26. N 1. P. 113–146. doi: 10.1016/S0364-0213(01)00061-1
12. Roy D. Learning visually grounded words and syntax of natural spoken language // *Evolution of Communication*. 2002. V. 4. N 1. P. 33–56.
13. Lowe D.G. Object recognition from local scale-invariant features // *International Conference of Computer Vision*. Kerkyra, Greece, 1999. V. 2. P. 1150–1157.
14. Lowe D.G. Distinctive image features from scale-invariant keypoints // *International Journal of Computer Vision*. 2004. V. 60. N 2. P. 91–110. doi: 10.1023/B:VISI.0000029664.99615.94
15. Lucas B.D., Kanade T. An iterative image registration technique with an application to stereo vision // *Proc. 7<sup>th</sup> Int. Joint Conference on Artificial Intelligence*. Vancouver, Canada, 1981. V. 2. P. 674–679.
16. Barron J., Fleet D., Beauchemin S. Performance of optical flow techniques // *International Journal of Computer Vision*. 1994. V. 12. N 1. P. 43–77. doi: 10.1007/BF01420984

17. Bay H., Ess A. Speeded-up robust features // Computer Vision and Image Understanding. 2008. V. 110. N 3. P. 346–359. doi: 10.1016/j.cviu.2007.09.014
18. Farneback G. Two-frame motion estimation based on polynomial expansion // Lecture Notes in Computer Science. 2003. V. 2749. P. 363–370.
19. Farneback G. Disparity estimation from local polynomial expansion // Proc. SSAD Symposium on Image Analysis, 2001.

- |                                   |  |
|-----------------------------------|--|
| <b>Филатов Владислав Игоревич</b> | – аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, sigal89@mail.ru                           |
| <b>Потанов Алексей Сергеевич</b>  | – доктор технических наук, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, pas.aicv@gmail.com |
| <b>Vladislav I. Filatov</b>       | – postgraduate, ITMO University, Saint Petersburg, 197101, Russian Federation, sigal89@mail.ru                         |
| <b>Alexey S. Potapov</b>          | – D.Sc., Associate professor, ITMO University, Saint Petersburg, 197101, Russian Federation, pas.aicv@gmail.com        |