

# Поиск визуально подобных изображений на основе машинного обучения

© А. Д. Варламов

Владимирский государственный университет им. А. Г. и Н. Г. Столетовых

Муром

varlamov\_aleks@mail.ru

© Р. В. Шарапов

info@vanta.ru

## Аннотация

Рассматриваются вопросы поиска изображений по визуальному образцу. Предлагается методика разработки алгоритма оценки визуальной близости пары изображений на основе машинного обучения. Анализируются признаки изображений, которые могут применяться в алгоритмах поиска. Решается проблема выбора из них ключевых.

## 1 Введение

Развитие компьютерной техники и технологий привели к бурному росту количества цифровых изображений. Источников таких изображений множество — цифровые фотоаппараты и видеокамеры, видеорегистраторы, камеры видеонаблюдения, мобильные телефоны, сканеры и т. д. Появились цифровые архивы, содержащие миллионы снимков. При таких объемах данных важное значение приобретает поиск нужных изображений.

Существует несколько направлений поиска по изображениям: поиск по содержанию (найти фотографию человека или изображение, к примеру, берёзы), поиск по визуальному образцу (найти изображения, похожие на заданное), поиск по описаниям (найти изображение, помеченное как «Совершенно секретно» или «Москва») и т. д. Каждое из направлений поиска имеет свои особенности и сферы применения.

Целью работы является рассмотрение вопросов построения системы поиска по визуальному образцу в относительно больших (десятки тысяч) коллекциях изображений.

## 2 Текущее состояние проблемы

Поиск изображений по визуальному образцу является интереснейшей задачей, решению которой посвящено множество работ. Достаточно подробные обзоры приведены в [1, 2, 6, 7].

Данный вид поиска заключается в извлечении

существенных свойств изображений и построении на их основе сигнатур, используемых в дальнейшем для сравнения пар изображений. В каждую пару всегда входит изображение из коллекции и изображение-образец, подобное которому стремится найти пользователь. Результатом сравнения является величина, называемая визуальным подобием (иногда тематической близостью) изображений. С точки зрения информационного поиска данная оценка, выполненная человеком-экспертом, будет являться содержательной релевантностью, а рассчитанная в системе — релевантностью формальной.

Свойства оцениваются алгоритмами вычисления значений важных численных величин и называются признаками. Подобных признаков можно выделить достаточно много. В общем случае они разделяются на две группы: глобальные и локальные. К глобальным признакам можно отнести основные цвета, текстуры, формы, значимые элементы всего изображения. Локальные признаки высчитываются для небольших частей (блоков) изображения [2]. Для обозначения значений признаков также употребляют термин дескриптор. Набор признаков (иногда говорят вектор признаков), описывающий изображение для определенной задачи, называется сигнатурой. Признаки, отобранные из множества альтернативных для их применения в некоторой задаче анализа изображений и вошедшие в сигнатуру, называют ключевыми.

При поиске изображений по образцу анализируются не отдельные экземпляры, а пары изображений, которые сопоставляются друг другу. Поэтому в отличие от других задач анализа в текущей проблеме от признаков изображений переходят к признакам пар, значения которых находятся как абсолютные разности значений соответствующих признаков каждого из изображений пары. Такая несложная формула позволяет моментально их рассчитать для проиндексированных изображений и реализовать на практике поиск с приемлемыми временными характеристиками.

## 3 Существующие признаки изображений

Системы, осуществляющие поиск по визуальному образцу, активно используют различные признаки изображений. Эти признаки отличаются большим разнообразием.

Вопросы использования цветовых характеристик обсуждаются в [4]. Выделяются такие признаки, как общий цвет, наиболее частотные цвета, регулярность размещения цветов, сложность образов и т.д.

Работа [8] посвящена использованию текстурных признаков. Авторы выделяют признаки контрастности, грубости, направленности, линейных образов, регулярности и шероховатости текстур.

В [9] рассматриваются различные признаки формы: Фурье признаки (FD), признаки кривизны пространства (CSSD), признаки Церник-моментов (ZMD) и сеточные признаки (GD). Анализируются слабые и сильные стороны указанных признаков изображений.

В работе [5] для поиска изображений по содержанию предлагается использовать признак однородности текстуры (Homogeneous Texture Descriptor, HTD), признак граничной гистограммы (Edge Histogram Descriptor, EHD), масштабируемый признак цвета (Scalable Color Descriptor, SCD), признак доминирующего цвета (Dominant Color Descriptor, DCD). Эти признаки подробно описаны в [3] и используются в различных системах.

Достаточно интересны работы, выполненные в рамках Российского семинара по оценке методов информационного поиска (РОМИП). Начиная с 2008 года на семинаре проводятся дорожки (секции, посвященные конкретным проектам) поиска изображений по образцу.

В работе [14] предложен алгоритм поиска изображений на основе нечетких цветовых гистограмм, вычисляемых в пространстве CIE  $L^*a^*b^*$ . Предлагается кроме главной использовать набор дополнительных гистограмм, построенных для определенных участков изображения.

В работе [11] рассматривается возможность цветовой модели HIS (Hue, Saturation, Intensity) при поиске изображений по содержанию. HSI — это цилиндрическое представление цветовой модели RGB, состоящее из следующих компонент: цветовой тон, насыщенность и интенсивность. Автор также предлагает использовать модель HLD (Hue Layout Descriptor).

В [16] используются уменьшенные копии изображений для поиска похожих снимков. В основу работы положен тот факт, что человеческое восприятие позволяет с хорошей точностью понимать содержание изображений с маленьким разрешением. Для цветных изображений человеку достаточно разрешения  $32 \times 32$ , чтобы достичь точности распознавания сцен превышающей 80%.

В работе [10] предлагается использовать цветовые признаки. Для этого строится цветовая гистограмма, учитывающая пространственное расположение цветов, и вычисляются цветовые моменты. Для оценки подобия двух изображений сравниваются параметры распределения их цветов: математическое ожидание по каждому из цветовых каналов и попарные ковариации распределений каналов.

В работе [13] используются признаки цветовых автокорреллограмм, текстурный признак Tamura, гистограммы ориентаций градиентов.

Таким образом, существует достаточно большое количество признаков, которые могли бы использоваться в задачах поиска изображений по визуальному образцу. Тем не менее, большинство работ включают лишь часть существующих признаков изображений, комбинируя их и разрабатывая на их основе сигнатуры. В связи с тем, что многие признаки могут быть взаимосвязаны между собой, важной задачей является их рациональный отбор.

#### 4 Методика проведения исследования

Схематично предлагаемая методика работ для построения алгоритма определения визуального подобия изображений на основе машинного обучения показана на рисунке 1. Ее можно представить в виде следующих этапов, которые должны быть последовательно выполнены:

Этап 1. Из коллекции изображений формируется обучающая выборка. Она состоит из множества прецедентов, каждый из которых включает в себя пару изображений и степень их близости, оцененную экспертом.

Этап 2. Программно реализуются (при необходимости и разрабатываются) алгоритмы вычисления всевозможных признаков на изображениях.

Этап 3. Для каждой пары изображений вычисляются значения всех признаков.

Этап 4. Проводится факторный анализ признаков, в результате которого определяются главные признаки, значения которых будут использоваться в обучении нейронной сети и алгоритмы вычисления которых станут частью конечного алгоритма оценки визуального подобия изображений.

Этап 5. Обучается нейронная сеть. По завершению этого процесса запоминаются данные, характеризующие величины ошибок нейронной сети относительно входной выборки. Эти показатели преобразуются в полноту и точность — целевые значения оценки качества выполняемой работы.

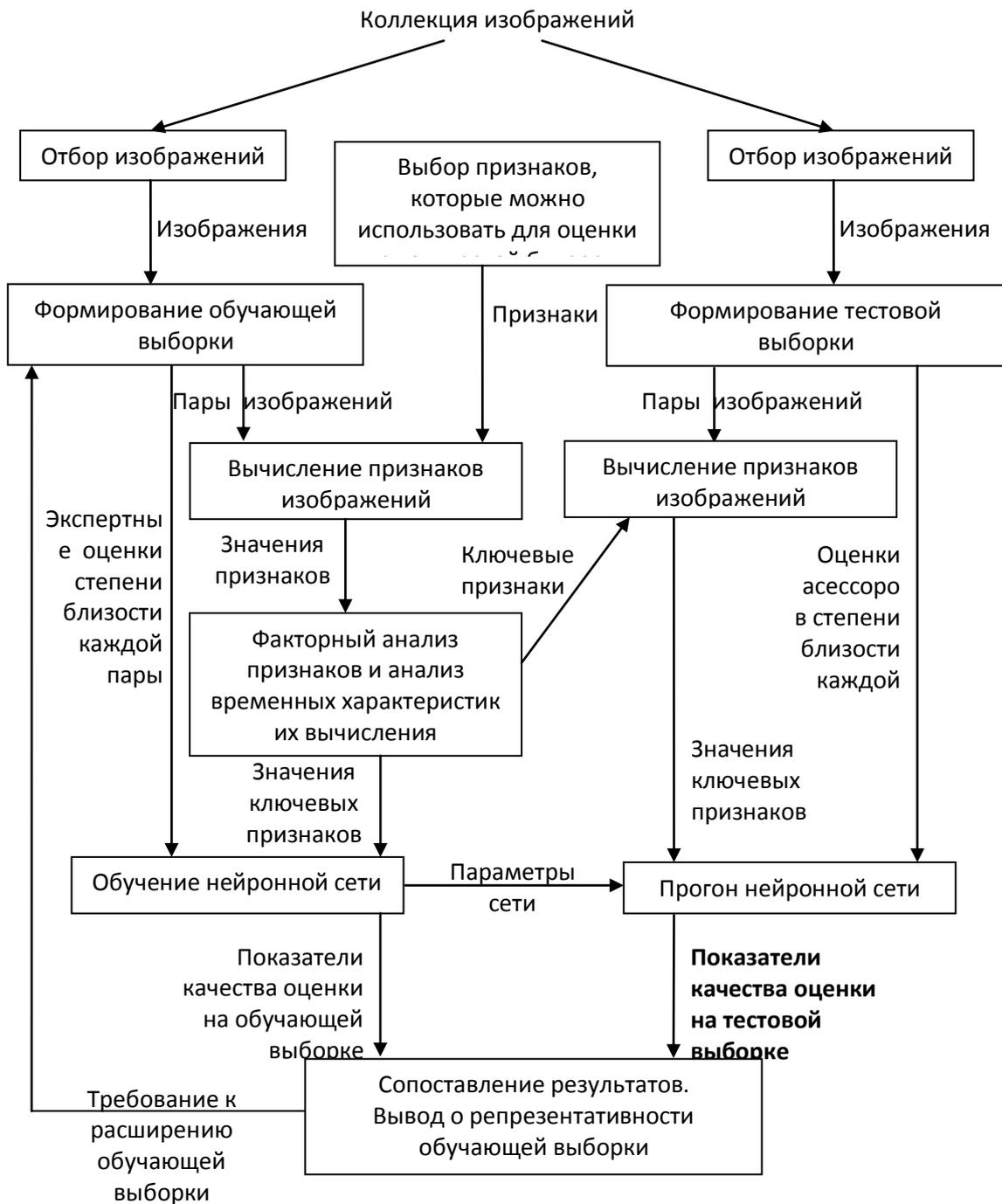


Рис. 1 Методика проведения работ для построения алгоритма определения визуального подобия изображений на основе машинного обучения

Этап 6. Из коллекции изображений формируется тестовая выборка. Ее структура и характеристики (объем, соотношение близких и небликих пар) должны соответствовать структуре и характеристикам обучающей выборки.

Этап 7. Вычисляются значения ключевых признаков всех изображений, входящих в тестовую выборку.

Этап 8. Прогоняется нейронная сеть на тестовой выборке с определением полноты и точности.

Этап 9. Сравниваются полнота и точность анализа нейронной сетью изображений обучающей выборки с полнотой и точностью анализа нейронной сетью изображений тестовой выборки. При приблизительно одинаковых значениях соответствующих показателей работа считается завершенной.

По завершению пятого этапа, на котором выполняется машинное обучение, может быть уже достигнута цель — получены все данные для их сборки в алгоритм определения подобия пары

произвольных изображений. На дальнейших этапах проводится анализ полученного результата и, при необходимости, принимается решение о возврате на предыдущие этапы работы. По завершении всего цикла становятся известными основные численные характеристики, отражающие качество полученного алгоритма: полнота и точность.

Применительно к решаемой задаче полнота (recall) определяется как отношение числа правильно идентифицированных близких пар изображений, к общему числу близких пар в выборке; точность (precision) определяется как отношение количества найденных близких пар изображений к числу пар, которые система посчитала тематически близкими.

Если на последнем этапе выявлено расхождение значений полноты или точности в выборках, то исходная выборка является нерепрезентативной. В данном случае у обучающей выборки будут наблюдаться более лучшие показатели, чем у тестовой. Но за счет неучастия тестовой выборки в процессе машинного обучения она дает более объективные показатели, на которые следует ориентироваться.

Репрезентативность определяет соответствие характеристик выборки генеральной совокупности. В нашем случае под генеральной совокупностью понимается множество изображений, доступных в сети интернет (или в коллекции, по которой ведется поиск). Проще говоря, если выборка репрезентативна, то ее целесообразно использовать при обучении нейронной сети, полагая, что обобщенные численные показатели результатов обучения будут схожи с результатами работы с генеральной совокупностью.

Сделать выборку более репрезентативной можно путём увеличения ее объема, то есть с возвратом на первый этап и, соответственно, повторением последующих этапов работы.

## 5 Выбор ключевых признаков

Очевидно, что можно использовать очень большое число признаков для анализа пар изображений на визуальное подобие. Но такая численность является крайне нежелательной. Во-первых, время работы результирующего алгоритма напрямую зависит от количества признаков, так как в данном случае велико количество величин, которые требуется рассчитать. Во-вторых, большее число признаков усложняет архитектуру нейронной сети, что приводит к увеличению требований к ресурсам, задействованным в процессе машинного обучения, и времени самого обучения. И, самое главное, — это избыточность признаков из-за их взаимной корреляции и потенциальное наличие признаков, которые могут не влиять на результат.

Поэтому необходимо сокращение числа признаков, но проводимое не вслепую в ущерб качества результата, а с обеспечением наибольшего снижения трудоемкости их вычисления при наименьшей потере точности результата обучения. Поставленная задача является оптимизационной, и

эвристически решить ее человеку не под силу, поэтому требуется формальный подход. В основе метода сокращения количества признаков предлагается использовать факторный анализ, широко применяемый при многомерном статистическом анализе для понижения размерности данных. За фактор будем принимать абсолютную разность значений одного и того же признака, рассчитанного для обоих изображений пары. Сокращая факторы, мы будем сокращать признаки, уменьшая сложность вычислений.

Факторный анализ предполагает анализ корреляционной матрицы с построением графика под названием “Каменистая осыпь”, на котором отражается значимость каждого фактора и на основании которого принимается решение о количестве и составе выделяемых (главных) факторов (в нашем случае — признаков). В классическом варианте факторный анализ учитывает только взаимную корреляцию факторов, исключая те, которые в большей степени коррелируют с другими. Для решения нашей оптимизационной задачи необходимо учитывать также корреляцию значений факторов со степенью близости пар и временные сложности алгоритмов вычисления признаков, так как они различны.

Этап выделения ключевых признаков является одним из самых важных и составляет ту соль, которая делает ее, с одной стороны, аккумулятором лучших предложений от специалистов по вычислению свойств изображений; с другой стороны, отфильтровывает избыточные, трудоемкие и потому ненужные алгоритмические решения вычисления признаков изображения.

Следует отметить, что с введением этого этапа в метод, стремление оптимизировать полноту-точность дополняется оптимизацией временных характеристик. Опишем математически данную задачу.

Пусть  $s_{ij}$  — значение  $j$ -го признака, вычисленное для  $i$ -й пары изображений,  $i=1, \dots, I$ ;  $j=1, \dots, J$ ;

$I$  — количество пар изображений;

$J$  — количество признаков;

$t_j$  — среднее время вычисления  $j$ -го признака на изображении;

$a_i$  — степень близости изображений в паре  $i$ . С учетом использования двузначной шкалы принимает значение 1, если изображения пары являются тематически близкими и 0 в противном случае.

Положим  $C(k, l)$  — коэффициент корреляции между последовательностями чисел  $k$  и  $l$ . Тогда возможно построить корреляционную матрицу

$$Cs = (cs_{j1, j2}) = \left( C(s_{j1}, s_{j2}) \right)_{j1=1, j2=1}^{JJ},$$

где  $s_{jx}$  — последовательность значений  $s_{1, jx}, s_{2, jx}, \dots, s_{I, jx}$ , и вектор

$$Ca = (ca_j) \left( C(s_j, a) \right)_{j=1}^J.$$

где  $a$  – последовательность значений  $a_1, a_2, \dots, a_l$ .

То есть  $cs_{k,l}$  – корреляция между  $k$ -м и  $l$ -м признаками, а  $ca_j$  – корреляция между  $j$ -м признаком и степенью подобию изображений в паре.

Анализируя матрицу  $Cs$ , можно найти признаки, которые имеют сильную статистическую взаимосвязь с другими. Эти признаки будем называть избыточными. Анализируя вектор  $Ca$ , можно найти признаки, которые статистически никак не связаны со степенью близости пары и, соответственно, не могут применяться для ее оценки. Эти признаки будем называть бесполезными. Признаки, требующие большое количество времени на их вычисление  $t_j$ , будем называть трудоемкими.

Чтобы оценить степень избыточности каждого признака, определим вектор собственных значений матрицы  $Cs$ . В теории анализа данных эта процедура используется в факторном анализе, являющимся разделом многомерного статистического анализа и позволяющего сократить размерность данных. Пусть  $X=(x_j)$  — вектор собственных значений корреляционной матрицы. Значение  $x_j$  определяет степень статистической независимости  $j$ -го признака от других. В факторном анализе пороговой обработкой этой величины определяются главные признаки. Авторами предлагается в качестве подобной меры использовать выражение

$$h_j = \frac{x_j ca_j}{t_j}.$$

В этой формуле агрегируются три важных показателя, которые необходимо учитывать при отборе ключевых признаков: избыточность по отношению к другим признакам, корреляция с результатом и время вычисления. Если ранжированные значения  $h_j$  отложить на графике, получаем аналог графика “Каменистая осыпь” из теории факторного анализа. Отметим, что он имеет подобный ему вид (рисунок 2).

Типовой процедурой по критерию Кайзера или визуально на графике отсеиваются не ключевые признаки, оставляя те, которые будут использоваться в дальнейшей работе.

## 6 Результаты исследований

В проведенном исследовании мы остановились на работе с 25 признаками, наиболее часто используемыми в задачах анализа изображений.

В связи с тем, что поиск может осуществляться в больших коллекциях изображений, в качестве признаков мы рассматривали только те, которые можно просчитать для изображений заранее и занести в некое подобие поискового индекса. По этой причине, нами не рассматривались те признаки, которые необходимо рассчитывать по парам изображений непосредственно при осуществлении поиска. Из 25 признаков на этапе выбора главных

часть была отсеяна. Таким образом, к ключевым признакам были отнесены следующие:

- Средние значения компонент R, G, B изображения в цветовой модели RGB;
- Медиана яркости;
- Детализованность;
- Признак симметрии изображения;
- Признак наличия текстур;
- Пропорции сторон;
- Наличие лиц на изображении;
- Нормированные 16-ти уровневые гистограммы по компонентам Y (яркости), U, V изображения в цветовой модели YUV;
- Среднеквадратичное отклонение яркостей уменьшенных копий (размером 32x32 пикселя) образца и рассматриваемого изображения.

В качестве набора исходных данных использовалась коллекция Flickr семинара РОМИП [12]. В ней содержится 20000 фотографий разного качества и без единой темы; имеются фотоснимки людей, пейзажи, городские сцены и т.д. Фотографии сделаны при разном освещении, в помещениях, на улице и т.д. Из данной коллекции в обучающую выборку были отобраны 500 пар, среди которых 125 являются тематически близкими, а 375 нет. Тестовой выборкой являлось специальное задание дорожки семинара РОМИП с результатами работы группы ассессоров.

Для формирования выборок и обучения нейронной сети применялся нейроимитатор Сигнейро [15], специализирующийся на задачах обработки и анализа изображений. Была сконструирована двухслойная нейронная сеть, структура которой приведена на рисунке 3. Количество входов сети соответствует количеству ключевых признаков, выход сети — один.

Для осуществления поиска изображений по визуальному подобию на основе полученной нейронной сети была разработана программная оболочка Аxioma. В процессе поиска она извлекает данные сигнатур (значения признаков) из поискового индекса, осуществляет вычисление меры близости образца к изображениям коллекции и ранжирует результаты поиска.

График полнота-точность полученного нейросетевого алгоритма приведен на рисунке 4. На этом графике виден выступ, крайняя точка которого соответствует решению, полученному нейронной сетью. Однако, более объективными можно считать данные работы ассессоров, отраженные на рисунке 5 и представленные в сравнении с результатами других исследователей (в рамках семинара РОМИП в дорожке поиска изображений по визуальному подобию).

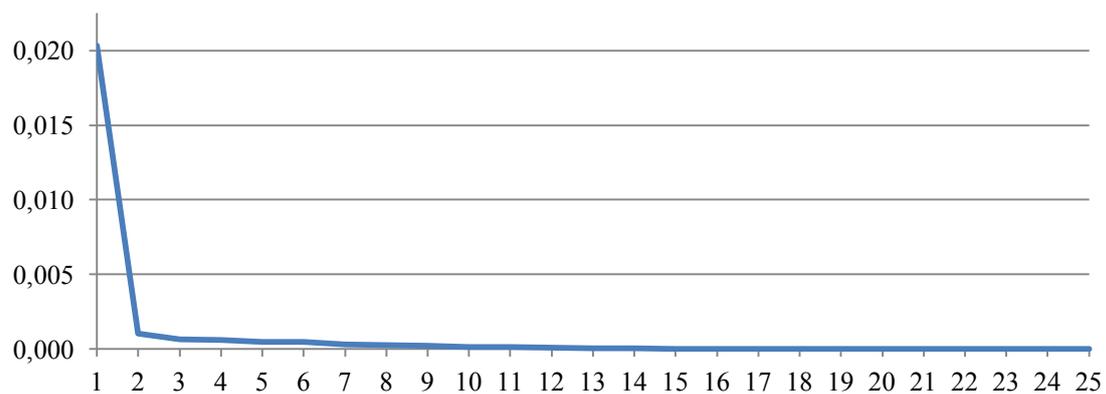


Рис. 2 Ранжирование признаков по значимости

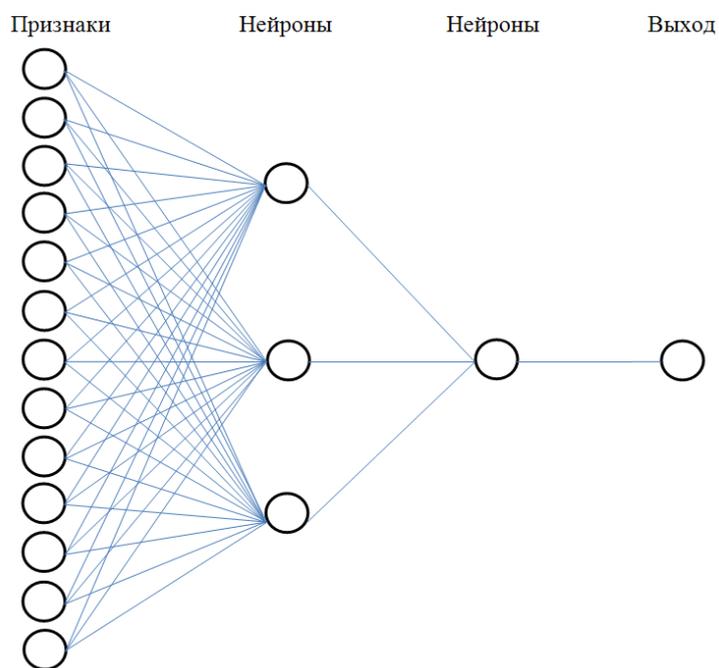


Рис. 3. Структура нейронной сети

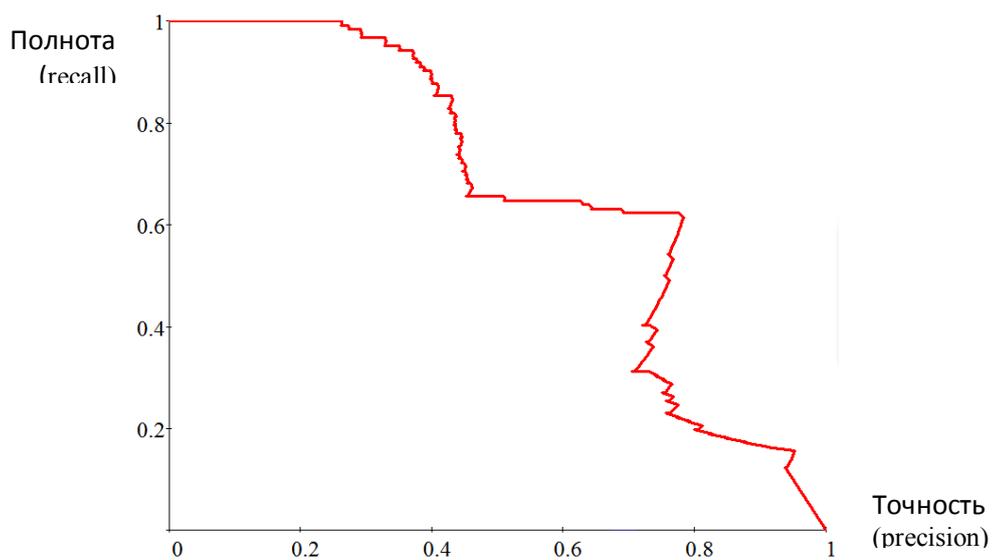


Рис. 4. График полнота-точность

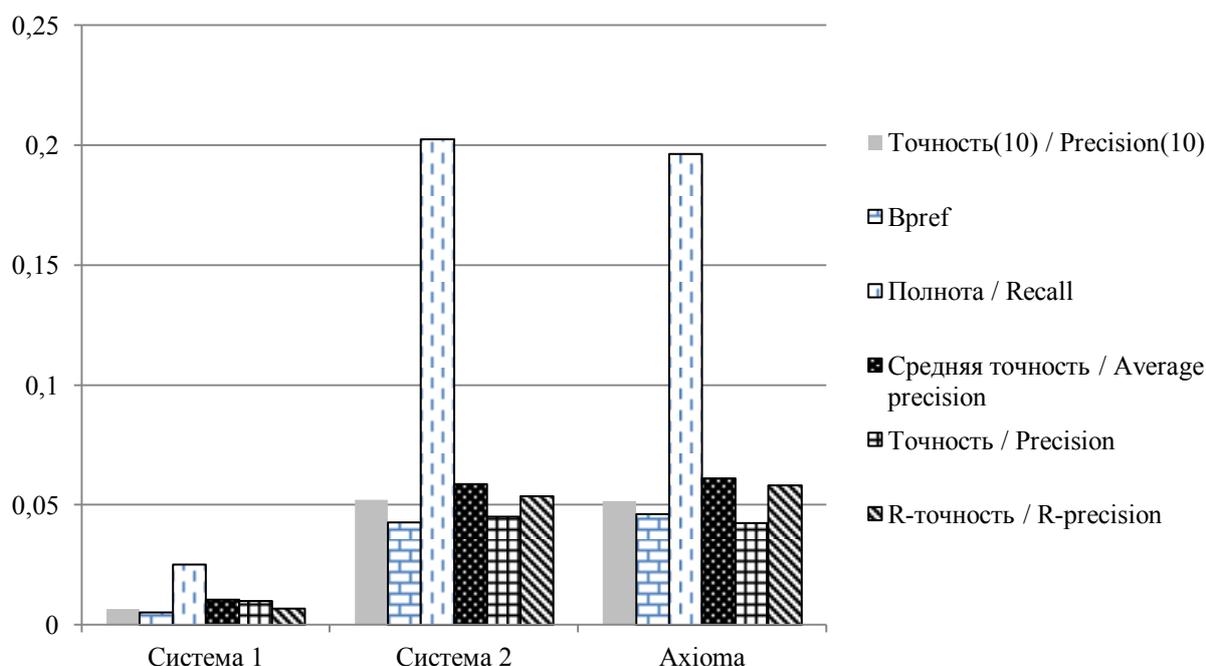


Рис. 5. Сравнение результатов участников дорожки поиска изображений по визуальному подобию семинара РОМИП (оценки «vote weak»)

Анализируя их и в целом проведенную работу, можно сделать следующие выводы:

1. Полученные качественные характеристики построенного алгоритма соответствуют современному уровню достижений науки в данной области.

2. Вошедшая в оптимизационный критерий временная характеристика позволила минимизировать трудоемкость итогового алгоритма, который обладает линейным порядком временной сложности.

3. В дальнейшем, при увеличении объемов обучающей и тестовой выборок, качественный показатель может быть улучшен и даже превзойден значения аналогов, так как на данный момент имеются расхождения в показателях обучающей и тестовой выборок.

## Литература

- [1] Aigrain P., Zhang H., Petkovic D. Content-based representation and retrieval of visual media: A review of the state-of-the-art. *Multimedia Tools and Applications*, 3(3):179-202, 1996.
- [2] Datta R., Joshi D., Li J., Wang J. Image Retrieval: Ideas, Influences, and Trends of the New Age // *ACM Computing Surveys*, Vol. 40, No. 2, pp. 1-60.
- [3] Manjunath B. S., Salembier P., Sikora T. Introduction to MPEG7: Multimedia Content Description Language. 1st edition, 2002.
- [4] Mojsilović R., Kovačević J., Hu J., Safranek R. J., Ganapathy S. K. Matching and retrieval based on the vocabulary and grammar of color patterns // *IEEE Trans. Image Processing*, 2000, volume 9, pp. 38-54.
- [5] Quack T., Monich U., Thiele L., Manjunath B.S. Cortina: A System for Largescale, Contentbased Web Image Retrieval // *MM'04*, October 1016, 2004, New York, USA.
- [6] Smeulders A. W., Worring M., Santini S., Gupta A., Jain R. Contentbased image retrieval at the end of the early years. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(12):1349-1380, 2000.
- [7] Snoek C. G. M., Worring M. Multimodal video indexing: A review of the state-of-the-art. *Multimedia Tools and Applications*, 25(1):5-35, 2005.
- [8] Tamura H., Mori S., Yamawaki T. Texture features corresponding to visual perception // *IEEE Transactions on System, Man and Cybernetic*. 1978, volume 8(6), pp. 460-473.
- [9] Zhang D., Lu G. Content-Based Shape Retrieval Using Different Shape Descriptors: A Comparative Study // In *IEEE International Conference on Multimedia and Expo*, 2001, pp. 289-293.
- [10] Васильева Н., Гладышева Ю. Взвешенный CombMNZ для комбинирования результатов поиска изображений по цветовым признакам // *Российский семинар по Оценке Методов Информационного Поиска. Труды РОМИП 2010.* (Казань, 15 октября 2010г.) - Казань, 2010. С. 98-108.
- [11] Волосных Д.Ф. Использование визуальных особенностей восприятия компонент цветовой модели HSI при поиске изображений по содержанию // *Российский семинар по Оценке Методов Информационного Поиска. Труды*

- РОМИП 2010. (Казань, 15 октября 2010г.) - Казань, 2010. С.154-163.
- [12] Коллекция изображений Flickr. 2008. <http://gomip.ru/ru/collections/flickr.html>
- [13] Мельниченко А., Гончаров А. ЛММИИ на РОМИП-2009: Методы поиска изображений по визуальному подобию и детекции нечетких дубликатов изображений с.108-121
- [14] Пименов В.Ю. Простые методы поиска изображений по содержанию // Российский семинар по Оценке Методов Информационного Поиска. Труды РОМИП 2010. (Казань, 15 октября 2010г.) - Казань, 2010. С.69-79.
- [15] Сигнейро. Нейронные сети в обработке и анализе изображений. <http://www.signneuro.ru>
- [16] Слесарев А.В., Мучник И.Б., Михалев Д.К., Крайнов А.Г., Котляров Д.И., Беляев Д.В. Яндекс на РОМИП 2010: Поиск похожих

изображений и дубликатов // Российский семинар по Оценке Методов Информационного Поиска. Труды РОМИП 2010. (Казань, 15 октября 2010г.) - Казань, 2010. С. 148-153.

### **Machine learning of visually similar images search**

Alexey Varlamov, Ruslan Sharapov

In this paper we consider the problem of image retrieval by visual similarity. We offer a method of developing the algorithm for determining similar visual images based on machine learning. We analyze the key features of images that can be used in the search. We solve the problem of selecting features that are used for machine learning search of visually similar images.