

Навигация по коллекциям цифровых изображений на основе методов автоматической классификации

Е. В. Мясников

Самарский Государственный Аэрокосмический Университет
mevg@smr.ru

Аннотация

В работе предлагается метод навигации по коллекциям цифровых изображений. Центральным вопросом при построении систем навигации является способ отображения коллекции в двумерное навигационное пространство. Суть предлагаемого метода состоит в реализации отображения в два этапа. На первом этапе по коллекции изображений строится система иерархически вложенных кластеров. На втором этапе производится отображение исходного пространства описания изображений в пространство навигации. Это отображение формируется с учетом отношений между изображениями и ограничений, накладываемых иерархией кластеров.

В работе дан обзор методов, используемых при построении систем навигации, рассмотрены требования к разрабатываемому методу навигации, приведены результаты исследований предложенного метода и сравнения его с известными методами. Полученные результаты позволяют сделать вывод о возможности успешного применения разработанного метода.

1. Введение

Поиск изображений по содержанию в последнее время является актуальной областью исследований. Однако поиск изображений по шаблону или в соответствии с какими-либо параметрами запроса далеко не всегда приводит пользователя к желаемому результату. Кроме того, пользователь часто желает просто посмотреть, какие изображения могут находиться в цифровой библиотеке. И в том и другом случае пользователю необходимо переходить от одного изображения к другому. Тогда мы имеем дело уже не с поиском изображений по содержанию, а с навигацией на основе содержания. Система навигации может служить прекрасным дополнением к системе поиска изображений, использоваться отдельно в коллекциях цифровых изображений домашнего пользования или в рамках создания пользовательских интерфейсов интернет-поисковых систем.

Постановка самой задачи навигации по коллекциям цифровых изображений является в значительной степени новой, оригинальной, и, с учетом постоянно растущих объемов цифровых данных, актуальной.

В настоящей работе предлагается новый метод навигации по коллекциям изображений. Метод основан на совместном использовании методов автоматической классификации и понижения размерности данных. Такое совмещение различных методов позволяет удовлетворить ряду естественных требований, налагаемых пользователями на процесс навигации.

Работа организована следующим образом. Во

втором разделе рассмотрены методы, используемые при построении систем навигации по коллекциям цифровых изображений, дается обзор известных систем навигации.

В третьем разделе работы приводятся требования, которым должна удовлетворять система навигации.

Четвертый раздел посвящен описанию предлагаемого подхода. Детально рассматриваются возможные модификации разработанного метода. Приводится описание всех исследуемых методов кластеризации и снижения размерности пространства.

Результаты экспериментального исследования предложенного метода приводятся в пятом разделе. Полученные результаты содержат как оценки качества предложенного метода, так и результаты сравнения с известными методами.

Шестой раздел посвящен вопросу выбора признаков, используемых при построении систем навигации.

В конце работы дается заключение, содержащее перечень основных ее результатов.

2. Обзор современного состояния

Центральным вопросом при построении систем навигации по коллекциям цифровых изображений является способ построения отображения коллекции изображений в двумерное пространство навигации. Существует целый ряд средств, позволяющих отображать изображения коллекции, упорядочивая изображения по времени создания или другой сопутствующей информации. Альтернативный подход к отображению состоит в извлечении из изображений каких-либо признаков и размещении изображений на экране в соответствии со значениями признаков. Так как размерность пространства признаков может в десятки и сотни раз превышать размерности пространства отображения, то для создания двумерных (трехмерных) отображений необходимо применять методы снижения размерности. Такие методы обычно подразделяют на линейные и нелинейные.

Линейные методы используют дискретный вариант разложения Карунена – Лоэва, называемый также методом главных компонент (PCA – Principal component analysis). В этом методе осуществляется поворот системы координат в исходном пространстве признаков таким образом, чтобы в проекции на новые оси – главные компоненты – дисперсия всего множества точек была максимальна. При этом дисперсия сосредоточена большей частью в первых компонентах, что позволяет рассматривать

только их, отбрасывая остальные. Метод PCA был использован при создании относительно небольшого числа систем [11, 16].

Наиболее распространенным классом методов нелинейного снижения размерности пространства является многомерное масштабирование (шкалирование) — MDS (multidimensional scaling).

Под MDS понимают техники, с помощью которых производится отображение множества векторов многомерного пространства в пространство малой размерности (как правило, двух- или трехмерное) с сохранением, по возможности, расстояний между ними.

Все MDS алгоритмы пытаются минимизировать функцию потерь, характеризующую величину рассогласования расстояний между первоначальными и полученными векторами в пространстве малой размерности. Наиболее часто функция потерь определяется как

$$E = \sum_{j < i} \frac{(d_{ij} - d_{ij}^*)^2}{d_{ij}^2}, \quad (1)$$

где d_{ij} — расстояние между объектами i и j в исходном пространстве, d_{ij}^* — расстояние между объектами i и j в пространстве малой размерности.

Как правило, алгоритмы MDS имеют итерационный характер, и в качестве начального приближения используются случайные значения. Вследствие этого, MDS алгоритмам свойственно останавливаться в локальных минимумах, порождая каждый раз различные представления в навигационном пространстве.

В работе [19] было предложено запускать MDS алгоритм для каждого набора входных данных некоторое малое число раз (5–10) и выбирать наилучшую конфигурацию в соответствии с используемым критерием. В [19] выделяется два класса алгоритмов, в зависимости от того, является ли порождаемое навигационное пространство непрерывным или дискретным. К первому типу алгоритмов относятся алгоритмы, использующие итеративную процедуру Ньютона-Рапсона, и алгоритмы, использующие процедуру имитации отжига.

Второй класс алгоритмов MDS работает на дискретной сетке, что, по мнению ряда исследователей, больше удовлетворяет потребности пользователей, чем случай непрерывного пространства отображения, приводящий к перекрытию иконок. Очевидным способом получения отображения на дискретную сетку является привязка к дискретной сетке результата работы, полученного в непрерывном пространстве. Однако в этом случае могут возникать конфликты, когда несколько изображений должны быть привязаны к одной ячейке сетки. Некоторые стратегии разрешения таких конфликтов предложены в [20]. Из работы [19] известен также «жадный» (*greedy*) алгоритм, позволяющий получить приемлемый результат за короткое время. Для получения отображения на дискретное пространство навигации был также использован генетический алгоритм. Этот алгоритм позволил исследователям получить существенно меньшие значения функции потерь, однако время его ра-

боты оказалось существенно выше, чем для «жадного» алгоритма.

Следует отметить, что на практике алгоритмы MDS применимы к небольшим множествам объектов (порядка сотни) из-за вычислительных ограничений. По той же причине рекомендуется вычислять матрицу расстояний между объектами заранее.

Существует отдельный класс алгоритмов многомерного масштабирования, использующих силовые методы укладки графов. Работа этих алгоритмов основана на математических моделях механических процессов. Наиболее известными являются модели Фрухтермана — Рейнгольда [10] и Камада-Кавайи [12].

В работе [23] метод MDS применен для размещения на плоскости результатов поиска изображений. Интересной стороной работы является использование цветовых сигнатур (*color signatures*) для описания изображений и введение экскаваторного расстояния (*Earth Mover's Distance — EMD*) между сигнатурами, трактуемого как минимальное количество работы, необходимое для преобразования одной сигнатуры в другую. К сожалению, пространство, образуемое сигнатурами, не является линейным. Это создает препятствие использованию такого метода снижения размерности, как PCA.

В работе [15] MDS подход использован для размещения на экране различных версий одного изображения, сгенерированных автоматически путем варьирования вектора параметров. Система, описанная в работе, предназначена для принятия решения о том, какие наборы параметров более всего удовлетворяют пользователя — специалиста по компьютерной графике.

При навигации по большим коллекциям изображений возникают различные проблемы. В частности, попытка снизить размерность пространства признаков для большого количества объектов занимает непомерно большое количество времени. Кроме того, отображение всех объектов на экране одновременно просто не представляется возможным. Вследствие этого, исследователями предпринимаются попытки построения иерархических структур, группирующих объекты в соответствии с их характеристиками. Наиболее очевидным способом выполнить такого рода разбиение является кластеризация.

В системе [6] реализован подход, основанный на иерархической кластеризации. На каждом шаге пользователю представляется набор изображений, каждое из которых является представителем кластера. Выбирая одно из представленных изображений, пользователь переходит к следующему уровню иерархии и процесс повторяется. Организация базы данных имеет иерархическую структуру квадродерева. В работе использован иерархический агломеративный алгоритм кластеризации.

Возможен также подход, при котором на каждом шаге просмотра пользователь выбирает несколько кластеров. Далее, все объекты, входящие в данные кластеры, кластеризуются заново, представляя пользователю новый набор кластеров, и процесс повторяется. Описанный принцип (*Scatter/*

Gather), был использован в работах [26, 7]. Изначально он был разработан для просмотра коллекций текстовых документов.

В работе [14] использованы нейронные сети Кохонена (так называемые самоорганизующиеся карты — SOM), организованные в иерархическом порядке. Полученная иерархическая структура сохраняется и используется при поиске в базе данных. Изображения располагаются в узлах дискретной сетки в соответствии с их схожестью.

В системе AutoAlbum [18] все изображения автоматически разбиваются на альбомы, образуя, таким образом, двухуровневую иерархию. В системе используется два вида кластеризации: кластеризация по времени создания фотографий и кластеризация по содержанию фотографий. Для кластеризации используется вероятностный алгоритм кластеризации (best-first model merging), в качестве вероятностной модели применяется скрытая марковская модель (НММ). Данный алгоритм позволяет работать на упорядоченных множествах данных и схож с агломеративными методами кластеризации. Отображение фотографий для пользователя внутри кластера осуществляется в соответствии со временем создания.

Система PhotoTOC [17] является дальнейшим развитием системы AutoAlbum. В PhotoTOC используются те же основные алгоритмы, улучшен пользовательский интерфейс и изменен алгоритм выбора представителя кластера (представитель выбирается на основании минимума дивергенции Кульбека-Либера (Kullback-Lieberg) между гистограммами фотографии и центра кластера).

Подавляющее большинство систем навигации (просмотра) коллекций изображений производят кластеризацию или строят двумерные отображения исключительно на основе примитивных признаков. Очевидно, что это является их серьезным недостатком, так как визуальная схожесть изображений на примитивном уровне далеко не всегда соответствует семантической схожести, оцениваемой человеком. На этот недостаток указывает ряд исследователей. Следует отметить, что данный недостаток на сегодняшний день не может быть преодолен из-за так называемого семантического провала [8].

3. Требования к системе навигации

При проектировании систем навигации и разработке метода отображения в пространство навигации следует четко определить те требования, которым должен удовлетворять разрабатываемый метод. В настоящей работе основными требованиями являются:

- представление изображений коллекции на текущем уровне отображения производится в виде двумерных векторов (например, точек/иконок на экране монитора);

- переход к подмножеству изображений коллекции, соответствующему меньшему уровню различий (по отношению к текущему уровню), производится при понижении уровня отображения или «приближении» области просмотра;

- переход к подмножеству изображений коллекции, соответствующему большему уровню различий (по отношению к текущему уровню), производится при повышении уровня отображения или «удалении» от текущей области просмотра.

Таким образом, можно говорить о том, что метод навигации подразумевает создание двумерной навигационной «карты», на которой должны быть определены *операции перемещения*, а также *операции приближения и удаления*.

Учитывая стохастический характер большинства методов кластеризации и снижения размерности, следует сказать о неочевидном, на первый взгляд, дополнительном требовании, которое в дальнейшем именуется *требованием обратимости*. Оно заключается в следующем. Пусть в некоторый момент времени пользователь имеет определенную картину отображения коллекции. Пусть, далее, он проделал ряд операций перемещения, приближения или удаления. Тогда, проделывая обратные операции, пользователь должен получить ту же картину отображения коллекции, которую он имел первоначально.

4. Описание предлагаемого подхода

Как показал обзор литературы, существующие системы навигации по коллекциям изображений строятся на основе двух различных групп методов. К первой группе относятся методы кластеризации, позволяющие разбить всю коллекцию изображений на небольшие группы, внутри которых разница между изображениями невелика. Ко второй группе относятся методы снижения размерности пространства признаков, позволяющие расположить элементы коллекции на плоскости.

Для построения более эффективных средств навигации представляется целесообразным разработка и исследование средств совмещения методов, относящихся к различным группам. Однако именно этому вопросу исследователями пока не было уделено достаточно внимания.

В данной работе проведено исследование ряда комбинированных алгоритмов, функционирующих на основе широко известного метода снижения размерности Сэммона и методов кластеризации. Кроме того, в работе предлагается метод снижения размерности, позволяющий сохранять линейную разделимость кластеров при построении отображения в двумерное пространство.

4.1. Алгоритмы кластеризации

Для реализации этапа кластеризации коллекции изображений и построения иерархии вложенных кластеров в процессе работы были рассмотрены как алгоритмы иерархической агломеративной кластеризации, так и нейросетевой алгоритм WTA.

В качестве *алгоритмов иерархической кластеризации* рассматривались алгоритмы Single Link и Complete Link [25].

Нейросетевой кластеризационный алгоритм WTA функционирует на базе нейронной сети Кохонена [2], представляющей собой однослойную сеть, в которой каждый нейрон соединен со всеми ком-

понентами входного вектора. При подаче на вход сети в момент времени t вектора $\mathbf{x}(t)$ в конкурентной борьбе побеждает тот нейрон из общего числа K , веса которого в наименьшей степени отличаются от соответствующих компонент входного вектора. Для нейрона-победителя p выполняется соотношение

$$d(\mathbf{x}(t), \mathbf{w}_v(t)) = \min_{1 \leq i \leq K} d(\mathbf{x}(t), \mathbf{w}_i(t)).$$

Здесь $d(\mathbf{x}(t), \mathbf{w}(t))$ – расстояние между входным вектором $\mathbf{x}(t)$ и текущим состоянием синаптических весов $\mathbf{w}_i(t)$ нейрона i . Выходной сигнал v_p выигравшего нейрона устанавливается в единицу, выходы остальных нейронов v_i становятся равными нулю. Нейрон-победитель подвергается адаптации, в ходе которой векторы весов изменяются в направлении вектора $\mathbf{x}(t)$ по правилу коррекции:

$$\mathbf{w}_v(t+1) = \mathbf{w}_v(t) + \eta(t)[\mathbf{x}(t) - \mathbf{w}_v(t)].$$

Здесь $\eta(t)$ – коэффициент обучения в момент времени t , значение которого со временем уменьшается. Основные параметры алгоритма подробно описаны в работе [1].

В работе было проведено исследование всех приведенных алгоритмов кластеризации. Результаты исследования приводятся в разделе 5.1.

4.2. Линейный и нелинейный методы снижения размерности пространства

Для *линейного преобразования* пространства в работе используется дискретный вариант преобразования Карунена-Лоэва, реализуемый нейронной сетью РСА [2] с применением правила Сенгера. При обучении сети используется адаптивный алгоритм изменения коэффициента обучения. В процессе обучения входная выборка предъявляется на вход сети многократно, вплоть до стабилизации весов.

Для *нелинейного преобразования пространства* в предлагаемом методе взят за основу алгоритм двумерного отображения, описанный, в частности, в [3] и впервые примененный Сэммоном в 1969 г. Данный алгоритм позволяет минимизировать ошибку представления многомерных данных, выражаемую в виде:

$$\varepsilon = \frac{1}{\sum_{i < j} d_{ij}} \cdot \sum_{i < j} \frac{(d_{ij} - d_{ij}^*)^2}{d_{ij}}. \quad (2)$$

Здесь d_{ij} и d_{ij}^* – расстояние между объектами i и j , соответственно, в многомерном и двумерном пространстве.

Работа алгоритма имеет итеративный характер, связанный со следующим рекуррентным соотношением для координат в двумерном пространстве y_{jk} :

$$y_{ik}(t+1) = y_{ik}(t) + \frac{2 \cdot \alpha}{\sum_{i < j} d_{ij}} \cdot \sum_{\substack{j=1 \\ j \neq i}}^N \frac{d_{ij} - d_{ij}^*}{d_{ij} \cdot d_{ij}^*} \cdot (y_{ik}(t) - y_{jk}(t)).$$

Настраиваемый параметр α влияет на скорость работы и сходимость алгоритма.

Следует отметить, что возможно использова-

ние *двухэтапного метода* снижения размерности пространства. В таком методе на первом этапе с помощью нейронной сети РСА находится грубое приближение решения. На втором этапе полученное приближение уточняется описанным выше нелинейным методом снижения размерности пространства. Двухэтапный метод обладает потенциально более высокой точностью и более быстрой сходимостью. Результаты экспериментального исследования такого метода и входящих в его состав алгоритмов представлены в разделе 5.2.

4.3. Комбинированный метод снижения размерности пространства

Приведенный в предыдущем разделе алгоритм нелинейного снижения размерности хорошо работает на небольших объемах данных, однако на больших объемах данных возможность применения алгоритма ограничивает вычислительная сложность. При объеме выборки N и числе итераций, сравнимом с объемом выборки объем вычислений для получения отображения равен $O[N^3]$.

Решением данной проблемы может быть алгоритм, использующий аппроксимации приращений координат точек на каждой итерации. При этом, если вычислительная сложность для построения аппроксимационной оценки на каждой итерации $O[k]$, где $k \ll N$, то вычислительная сложность всего алгоритма может быть снижена до $O[N^2]$.

Среди существующих решений в качестве такой аппроксимации может быть использован метод, подобный предложенному Чалмерсом в работе [5]. В этом подходе на каждой итерации для каждого корректируемого элемента формируется 2 множества. В первом из множеств содержатся элементы, наиболее близкие рассматриваемому в многомерном пространстве. Во втором множестве содержатся элементы, отбираемые на каждой итерации случайным образом. Такой подход был использован для минимизации ошибки, выражаемой в виде (1), однако он может быть применен и при минимизации ошибки Сэммона (2) (далее будем называть этот метод *методом ЧС*).

В настоящей работе предлагается новый *комбинированный метод (КМ)*. Он состоит в том, чтобы в качестве аппроксимации использовать результаты иерархической кластеризации в рамках двухэтапной процедуры следующего вида. На первом этапе для всех k кластеров самого верхнего уровня строится двумерное отображение центров этих кластеров. На втором этапе строится k отображений для подкластеров и объектов второго уровня. При этом для построения отображения каждого подкластера расположение координат центров суперкластеров фиксируется, а производится оптимизация только объектов, находящихся в рассматриваемом подкластере. Процесс повторяется для третьего уровня иерархии и так далее, пока не будет построено отображение всех объектов.

В случае, когда мощность кластера равна k , вычислительная сложность двумерного отображения элементов кластера равна $O[k^3]$. Далее, число кластеров при иерархической кластеризации в слу-

чае сбалансированного дерева кластеров составляет:

$$M = \frac{N}{k} + \frac{N}{k^2} + \dots + \frac{N}{k^{\log_k N}} = N \cdot \sum_{i=1}^{\log_k N} \frac{1}{k^i} = \frac{N-1}{k-1}.$$

Таким образом, *вычислительная сложность* в случае сбалансированного дерева кластеров равна

$$\frac{N-1}{k-1} O[k^3] \sim O[k^2 N].$$

В случае сбалансированного дерева двухуровневой иерархии $k = \sqrt{N}$. Тогда сложность алгоритма, как и в случае предыдущего алгоритма, равна $O[N^2]$. Однако с ростом количества уровней иерархии сбалансированного дерева сложность алгоритма уменьшается. Для случая сбалансированного дерева высотой $L = \log_k N$, когда в каждом кластере оказывается $k = N^{1/L}$ элементов, выражение сложности принимает вид $O\left[N^{1+\frac{2}{L}}\right]$. Очевидно, по мере роста числа уровней L это выражение стремится к $O[N]$.

В работе также рассматривается *модификация комбинированного метода (МКМ)*, в которой при построении отображения для любого кластера (кроме кластера верхнего уровня) в качестве фиксированных центров берутся центры кластеров верхнего уровня. Очевидно, *вычислительная сложность* комбинированного метода с модификацией остается той же.

Результаты экспериментального исследования представленных в этом разделе методов приведены в разделе 5.2.

4.4. Комбинированный метод снижения размерности пространства с ограничениями

Рассмотренные методы снижения размерности пространства имеют существенный недостаток. А именно, их применение не дает гарантии того, что в двумерном пространстве будет сохранена та разделимость кластеров, которая имела место в многомерном пространстве. Для пользователя это выражается в проявлении нежелательного эффекта, заключающегося в том, что непохожие изображения оказываются рядом друг с другом. Чтобы избежать подобного эффекта потребуем сохранения разделимости между кластерами при отображении. Существенно, что предложенный в предыдущем разделе комбинированный метод естественным образом может быть доработан для удовлетворения указанному требованию сохранения разделимости. Назовем разрабатываемый метод *комбинированным методом снижения размерности пространства с ограничениями (КМО)*.

Суть метода КМО состоит в следующем. Алгоритм кластеризации разбивает все множество объектов в многомерном пространстве на непересекающиеся подмножества – кластеры. Потребуем, чтобы образы кластеров в двумерном пространстве навигации были разделимы, причем каждый кластер мог быть ограничен прямыми, образующими выпуклую область. Естественным способом удовлетворения этого требования является использо-

вание областей Вороного. Построение областей Воронова, то есть получение набора ограничивающих каждую область прямых, производится на основании информации о расположении центров кластеров на плоскости. Имея описание прямых – *границ кластеров* для каждого уровня – можно корректировать результат отображения в двумерное пространство.

Собственно предлагаемый метод КМО представим следующим образом.

1. Отображаем центры x_u^l кластеров $C_u^l \in C^0$ верхнего уровня в вектора y_u^l на плоскости с использованием метода снижения размерности пространства (нелинейного метода Сэммона или двухэтапного метода). Устанавливаем границы всей области отображения $\Gamma^0 = \emptyset$.

2. Для каждого кластера $C_u^k \in C_v^{k-1}$ текущего уровня k выполняем пункты 3–6.

3. Строим границы Γ_u^k кластера C_u^k в двумерном пространстве с учетом кластеров $C_m^k, m=1..|C_v^{k-1}|$ текущего уровня k , используя для этого центры соответствующих кластеров y_m^k на плоскости.

4. Дополняем границы кластера Γ_u^k с учетом границ Γ_v^{k-1} родительского кластера C_v^{k-1} предыдущего уровня: $\Gamma_u^k = \Gamma_u^k \cup \Gamma_v^{k-1}$.

5. Отображаем центры x_i^{k+1} всех подкластеров $C_i^{k+1} \in C_u^k$ (или координаты самих изображений O_i) уровня $k+1$ в вектора y_i^{k+1} на плоскости, учитывая границы Γ_u^k кластера, с использованием следующего рекуррентного соотношения:

$$y_i^{k+1}(t+1) = y_i^{k+1}(t) + \frac{2 \cdot \alpha}{\sum_{j>i} d_{ij}} \cdot \sum_{\substack{j=1 \\ j \neq i}}^N \frac{d_{ij} - d_{ij}^*}{d_{ij} \cdot d_{ij}^*} \cdot (y_i^{k+1}(t) - y_j(t)) - \varphi(y_i^{k+1}(t), \Gamma_u^k)$$

где $d_{ij}^* = \|y_i^{k+1}, y_j\|$, $d_{ij} = \|y_i^{k+1}, x_j\|$,

$y_j \in \{y_i^{k+1}, l=1..|C_u^k|\} \cup \{y_l^l, l=1..|C_u^l|\}$,

$x_j \in \{x_i^{k+1}, l=1..|C_u^k|\} \cup \{x_l^l, l=1..|C_u^l|\}$,

$N = |C_u^k| + |C_u^l|$.

6. Применяем описанную в пп. 3–5 процедуру для отображения дочерних кластеров C_i^{k+1} в рекурсивном порядке.

Представленный метод дан с точностью до определения функции штрафа за приближение или выход точки за границы соответствующего кластера. Функция $\varphi(y, \Gamma)$ может быть выбрана исходя из минимизации общего функционала, состоящего из ошибки Сэммона и функции границы (градиентное правило).

В настоящей работе рассматривались два эвристических правила определения функции $\varphi(y, \Gamma)$.

Правило полной коррекции (КМО-1) – величина $\varphi(y, \Gamma)$ выбиралась таким образом, чтобы при выходе точки за границу кластера на следующем шаге образ точки y_i попадал на границу;

Кусочно-линейное правило коррекции (КМО-2) – величина $\varphi(y, \Gamma)$ выбиралась таким образом, чтобы обеспечить «притяжение» точки y_i к центру кластера, пропорциональное расстоянию до центра кластера или границы.

Результаты экспериментального исследования

представленного в этом разделе метода приведены в разделе 5.2.

4.5. Метод навигации по коллекции цифровых изображений

Используя теоретические основы, изложенные в данном разделе, можно предложить следующий метод навигации по коллекциям цифровых изображений.

На первом этапе для всей коллекции цифровых изображений производится расчет признаков (вопрос выбора системы признаков рассматривается в 6 разделе). На втором этапе производится иерархическая кластеризация изображений в многомерном пространстве признаков с использованием алгоритма WTA, описанного в разделе 4.1. Далее применяется двухэтапный метод отображения, предложенный в разделе 4.4, который для каждого изображения коллекции рассчитывает координаты его расположения на двумерной плоскости отображения (навигации).

Результатом этих трех этапов являются рассчитанные координаты всех изображений коллекции на двумерной плоскости навигации. На основе этой базовой информации производится процесс навигации, интерактивно управляемый пользователем.

Простейшими операциями навигации являются операции перемещения вверх, вниз, вправо, влево, определяемые на двумерном пространстве естественным образом. Кроме того, реализуются операции увеличения и уменьшения масштаба. Операция увеличения масштаба (приближение) соответствует переходу к подмножеству изображений коллекции, соответствующему меньшему уровню различий. Операция уменьшения масштаба (удаление) соответствует переходу к подмножеству изображений коллекции, соответствующему большему уровню различий. При переходе к отображению с большим или меньшим уровнем различий для показа используются представленные кластеры соответствующего уровня иерархии.

Помимо указанных базовых операций навигации разработанный метод позволяет определить операцию перехода к конкретному кластеру в дереве иерархии.

Для ускорения операций навигации данные с координатами в двумерном пространстве могут быть пространственно проиндексированы.

Таким образом, по построению, предложенный метод автоматически удовлетворяет естественным требованиям раздела 3. Заметим, что поскольку координаты расположения изображений коллекции на плоскости отображения можно рассчитать заранее и сохранить, то проблема, связанная со стохастическим характером алгоритмов (требования обратимости), снимается.

Дополнительным преимуществом предлагаемого метода является то, что в нем возможно явное определение границ кластеров в двумерном навигационном пространстве и сохранение их характеристик. Это открывает широкие возможности для порционной обработки данных и пространственной индексации коллекций изображений значительного объема.

Следует также отметить, что использование предложенного в п. 4.4 метода не создает проблем

при добавлении или удалении единичных изображений из коллекции.

5. Результаты экспериментальных исследований

В разделе приводятся результаты экспериментальных исследований методов, изложенных в разделе 4. При проведении исследования использовались изображения из набора данных «База изображений Яндекса».

5.1. Экспериментальные исследования алгоритмов кластеризации

Экспериментальные исследования алгоритмов кластеризации, носящие вспомогательный характер, проводились с целью сравнения качества кластеризации, получаемого различными алгоритмами, и выбора алгоритма для дальнейшего использования. При проведении исследований качество кластеризации оценивалось по погрешности, выражаемой в виде:

$$\varepsilon_q = \frac{1}{U} \cdot \sum_{j=1}^U \|\mathbf{x}_j - \mathbf{w}_{v(j)}\|^2,$$

где \mathbf{x}_j – входной образ, $\mathbf{w}_{v(j)}$ – центр соответствующего кластера, рассчитанный как выборочное среднее входящих в него образов, U – количество входных векторов. Выбор данной формулы для оценки качества кластеризации, помимо простоты вычисления, был обусловлен спецификой выбранного в работе пространства признаков. Оно основано на цветовом пространстве CIE L*a*b*, разработанном специально таким образом, чтобы степень рассогласования цветов в данном пространстве, рассчитанная как евклидово расстояние, соответствовала человеческому восприятию.

При исследовании использовались наборы по 1000 изображений. Такое относительно небольшое количество изображений представляется для вспомогательного исследования достаточным. В табл. 1 приведены средние значения ошибки квантования, получаемые в результате кластеризации на верхнем уровне иерархии для каждого из рассматриваемых алгоритмов. Были проведены эксперименты, в которых максимальное количество кластеров для верхнего уровня иерархии (а также для любой ветви иерархии) составляло 25, 50 и 100 кластеров.

Таблица 1

Средние значения ошибки квантования

Количество кластеров	Алгоритм		
	Single link	Complete link	WTA
25	0,387	0,193	0,169
50	0,359	0,150	0,139
100	0,293	0,116	0,112

Из приведенных результатов видно, что наилучшие значения погрешности обеспечивает алгоритм WTA. Следует отметить, что алгоритм Complete Link так же обеспечивает хорошие значения погрешности, худшие значения – у алгоритма Single Link.

Опираясь на результаты проведенного исследования, при проведении описанных в следующих разделах экспериментов для кластеризации всюду выбирался алгоритм WTA.

Следует отметить, что максимальное число кластеров верхнего уровня (а также любой ветви иерархии) является настройкой алгоритма. При выборе данного значения следует учитывать как максимально возможное число объектов, участвующих в двумерном отображении, так и порождаемую глубину иерархии. При больших значениях этого параметра работа методов двумерного отображения может занять неприемлемо большое время. При малых значениях возможно получение больших значений погрешности при применении комбинированных методов двумерного отображения вследствие недостаточного количества фиксируемых центров суперкластеров. Практически рассматриваемое значение следует выбирать из диапазона от нескольких десятков до сотни.

5.2. Экспериментальные исследования методов двумерного отображения

В первой серии экспериментов исследования линейного, нелинейного и двухэтапного методов снижения размерности проводились следующим образом. Из набора, содержащего порядка 10 тысяч различных цветных изображений, случайным образом производились выборки по 100 изображений. Далее для каждой такой выборки применялись линейный, нелинейный, а также двухэтапный методы снижения размерности, описанные в разделе 4.2. Малый объем выборки (100 шт.) обусловлен, прежде всего, высокой вычислительной сложностью алгоритма Сэммона снижения размерности. Целью экспериментов было определение целесообразности использования двухэтапного метода отображения, описанного в разделе 4.2.

Для оценки качества работы алгоритмов значение ошибки рассчитывалось в соответствии с критерием (2). При применении нелинейного метода снижения размерности Сэммона проводилось по 10 запусков, для каждого из которых начальное приближение выбиралось случайным образом. При этом фиксировалось также минимальное значение ошибки по 10 запускам.

В табл. 2 приведены средние значения ошибок по всем сгенерированным наборам изображений

Таблица 2

Сравнение линейного, нелинейного и двухэтапного методов

	Количество итераций		
	100	200	300
Среднее значение ошибки для метода Сэммона	0,126	0,078	0,056
Среднее из минимальных значений ошибки для метода Сэммона	0,093	0,051	0,035
Среднее значение ошибки для нейросетевого алгоритма РСА		0,139	
Среднее значение ошибки для двухэтапного метода	0,044	0,041	0,039

для каждого из трех рассмотренных методов. Для нелинейного метода снижения размерности Сэммона также указано среднее из полученных минимальных значений ошибки.

Как видно из таблицы, значение ошибки (2) для алгоритма РСА является наихудшим. Однако использование результатов применения алгоритма РСА в качестве начального приближения дает возможность получить в двухэтапном методе достаточно малые значения ошибки уже при 100 итерациях. Использование нелинейного метода даже при достаточно большом количестве итераций в среднем не позволяет достичь результатов, полученных при использовании двухэтапного метода.

С другой стороны, при использовании в нелинейном методе разумного количества попыток (до 10) и достаточного числа итераций (порядка 300 для наборов до 100 элементов) можно добиться меньших значений погрешности, чем при использовании двухэтапного метода. Однако с ростом количества попыток пропорционально растет и время решения задачи нелинейным методом.

В следующей серии экспериментов было проведено исследование методов, описанных в пп. 4.3, 4.4. Было проведено 200 экспериментов, в которых каждый из методов применялся к набору из 1000 и 5000 изображений. Для оценки качества работы алгоритмов значение ошибки, как и в предыдущем случае, рассчитывалось в соответствии с критерием (2). Работа алгоритмов снижения размерности останавливалась, когда наибольшая относительная коррекция положения точек за одну итерацию не превышала заданной величины, равной 10^{-5} . Результаты исследований приведены в табл. 3, 4.

Таблица 3

Анализ комбинированных методов на наборах по 1000 изображений

Метод	Среднее значение ошибки	СКО ошибки	Среднее время работы, с
РСА	0,1171	0,01449	2
ЧС	0,02880	0,005668	62
КМ	0,03407	0,002638	17
МКМ	0,02767	0,002047	19
КМО-1	0,02972	0,002218	13
КМО-2	0,03494	0,002590	60

Таблица 4

Анализ комбинированных методов на наборах по 5000 изображений

Метод	Среднее значение ошибки	СКО ошибки	Среднее время работы, с
РСА	0,1148	0,009690	11
ЧС	0,02592	0,001657	1880
КМ	0,06220	0,028756	31
МКМ	0,02840	0,001962	67
КМО-1	0,03143	0,001890	31
КМО-2	0,03723	0,002076	276

Из табл. 3 и 4 видно, что наилучшими методами по качеству являются метод ЧС и метод МКМ. Судя по результатам, можно говорить о приблизительно равном качестве построения нелинейного отображения коллекции изображений. Однако, очевидно, метод МКМ требует существенно меньше времени, причем выигрыш по времени растет с ростом размеров коллекции.

Отметим также, что существенным преимуществом методов группы КМ является относительно малый размер требуемой оперативной памяти. При применении методов этой группы требуется хранение лишь $k \cdot k/2$ расстояний.

Учитывая указанные факторы, можно *рекомендовать предложенный метод МКМ к практическому использованию*.

Методы КМО, позволяющие сохранить разделимость между кластерами, имеют значения погрешности, отличающиеся от наилучших на 20-40%. Временные характеристики этих методов являются приемлемыми. Поэтому в тех случаях, когда *требования сохранения разделимости кластеров являются существенными, следует ориентироваться на методы КМО*. Визуальный анализ результатов отображения этих двух методов позволяет выделить *КМО-2 как более предпочтительный*, так как КМО-1 приводит в ряде случаев к сосредоточению образов на границах кластеров.

Пример работы методов МКМ и КМО-2 для набора из 10000 изображений показан на рисунке.

6. Выбор системы признаков

При построении систем распознавания и поиска изображений выбор системы информативных признаков является сложной задачей, решение которой во многом определяется личным опытом и предпочтениями разработчика. Следует отметить, что в настоящее время задача поиска может быть удовлетворительно решена лишь для узкоспециализированных областей. При построении систем поиска общего назначения, тем не менее, можно отметить склонность к использованию сложных систем признаков, построенных на основе цветовых, текстурных при-

знаков, признаков формы и пространственных отношений между областями [4].

При решении задачи навигации одним из важнейших факторов, влияющих на выбор системы признаков, является понятность пользователю формируемого порядка расположения изображений. Даже довольно изощренная система признаков может оказаться неэффективной, если пользователь системы не понимает, каким образом распределяются изображения на экране, или по каким критериям формируются кластеры. Причиной такого эффекта в большой степени является значительная потеря информации при проецировании в пространстве малой размерности [21].

В системах поиска и навигации цветовые признаки применяются наиболее часто, так как получаемые результаты наглядны и хорошо согласуются с ожиданиями пользователей. Поэтому в настоящей работе используются хорошо зарекомендовавшие себя гистограммы цветов, в качестве метрики взято евклидово расстояние между гистограммами.

При построении цветовых гистограмм было использовано цветовое пространство CIE L*a*b*. Основным преимуществом данного цветового пространства является тот факт, что рассогласование цветов (расчитанное как Евклидово расстояние между цветами) соответствует человеческому восприятию.

В работе также проводились исследования для следующих систем признаков:

- спектральные признаки, формируемые как коэффициенты низкочастотной части двумерного спектра Фурье;
- отсчеты яркости изображений, приведенных к малым размерам;
- геометрические моментные характеристики, рассчитываемые отдельно по цветовым плоскостям пространства RGB;
- цветовые гистограммы, рассчитанные отдельно по областям переднего плана и фона.

Следует отметить, что визуальные результаты отображения коллекций изображений на плоскость навигации для приведенных систем признаков хуже отвечают критерию визуального восприятия, чем цветовые гистограммы в пространстве CIE L*a*b*.



Пример работы комбинированных методов (а - метод МКМ, б – метод КМО-2)

7. Заключение

В работе разработан метод навигации по коллекциям цифровых изображений. В его основе лежит предложенный комбинированный метод отображения исходного описания изображений на двумерную плоскость навигации. Работа комбинированного метода основана на применении алгоритма кластеризации, используемого для иерархического представления коллекции изображений, и метода снижения размерности пространства, используемого для отображения элементов иерархии на плоскости. Предложены и реализованы несколько модификаций комбинированного метода.

Проведены экспериментальные исследования предложенных методов, дано сравнение предложенных методов с известными методами. Полученные результаты позволяют сделать вывод о возможности успешного применения разработанного метода навигации и предложенных методов снижения размерности, входящих в его состав.

8. Благодарности

Автор выражает благодарность ООО «ЯНДЕКС» (www.yandex.ru) за поддержку данной работы и предоставленный набор данных для проведения исследования.

9. Литература

- [1] Мясников Е. В. Нейросетевые алгоритмы кластеризации отсчетов цветных изображений. Перспективные информационные технологии в научных исследованиях, проектировании и обучении (ПИТ-2006) // Тр. науч.-техн. конф. с междунар. участием. Т. 2. – Самара, 2006. – С. 118–123.
- [2] Осовский С. Нейронные сети для обработки информации. – М.: Финансы и статистика, 2002.
- [3] Фукунага К. Введение в статистическую теорию распознавания образов. – М.: Наука, 1979. – 368 с.
- [4] Castelli V., Bergman L. D. Image Databases: Search and Retrieval of Digital Imagery. 2002.
- [5] Chalmers M. «A Linear Iteration Time Layout Algorithm for Visualising High-Dimensional Data», Proc. IEEE Visualization'96, San Francisco. P. 127–132 (1996).
- [6] Chen J.-Y., Bouman C. A., Dalton J. C. Similarity Pyramids For Browsing And Organization Of Large Image Databases. In Human Vision and Electronic Imaging III, Vol. 3299 of Proceedings of SPIE. P. 563–575, 1998.
- [7] Chen F., Gargi U., Niles L., Schutze H. Multi-modal browsing of images in web documents // Document Recognition and Retrieval VI. Vol. 3651 of Proceedings of SPIE. P. 122–133, 1999.
- [8] Eakins J. P., Graham M. E. Content-based Image Retrieval. A Report to the JISC Technology Applications Programme. Institute for Data Research, University of Northumbria at Newcastle, 1999.
- [9] Everton S. F. A Guide For The Visually Perplexed: Visually Representing Social Networks. Stanford University. Jan. 2004.
- [10] Fruchterman T., Reingold E. Graph Drawing by Force-Directed Placement. Software-Practice and Experience 21: 1129–1164, 1991.
- [11] Hiroike A., Musha Y., Sugimoto A., Mori Y. Visualization of information spaces to retrieve and browse image data. In Proceedings of the Third International Conference on Visual Information and Information Systems (VISUAL'99), volume 1614 of Lecture Notes in Computer Science, P. 155–162. Springer, 1999.
- [12] Kamada T., Kawai S. An Algorithm for Drawing General Undirected Graphs. Information Processing Letters 31: 7–15, 1989.
- [13] Barnard K., Duygulu P., Forsyth D. Clustering Art.
- [14] Koskela M., Laaksonen J., Laakso S., Oja E. The PicSOM retrieval system: Description and evaluations. In The Challenge of Image Retrieval. Electronic Workshops in Computing, 2000.
- [15] Marks J., Andalman B., Beardsley P. A., Freeman W. et al. Design Galleries: A general approach to setting parameters for computer graphics and animation. In Proceedings of SIGGRAPH '97. P. 389–400. ACM, 1997.
- [16] Peccenovic Z., Do M., Vetterli M., Pu P. Integrated browsing and searching of large image collections // Proceedings of the Fourth International Conference on Advances in Visual Information Systems (VISUAL 2000). Vol. 1929 of Lecture Notes in Computer Science. P. 279–289. Springer, 2000.
- [17] Platt J. C., Czerwinski M., Field B. A. PhotoTOC: Automatic Clustering for Browsing Personal Photographs. Technical Report. Microsoft Research. February 2002.
- [18] Platt J. C. AutoAlbum: Clustering Digital Photographs using Probabilistic Model Merging. Proc. IEEE Workshop on Content-Based Access of Image and Video Libraries. P. 96–100, 2000.
- [19] Rodden K. Evaluating Similarity-Based Visualisations As Interfaces For Image Browsing. Technical Report. University of Cambridge Computer Laboratory, 2002.
- [20] Rodden K., Basalaj W., Sinclair D., Wood K. Evaluating A Visualisation Of Image Similarity As A Tool For Image Browsing // Proceedings of the IEEE Symposium on Information Visualization (InfoVis'99). IEEE, 1999.
- [21] Rodden K., Basalaj W., Sinclair D., Wood K. A comparison of measures for visualising image similarity // The Challenge of Image Retrieval. British Computer Society Electronic Workshops in Computing, 2000.
- [22] Rose T., Elworthy D., Kotcheff A., Clare A., Tsonis P. ANVIL: a system for the retrieval of captioned images using NLP techniques // The Challenge of Image Retrieval. Electronic Workshops in Computing, 2000.
- [23] Rubner Y., Tomasi C., Guibas L. J. Adaptive color-image embeddings for database navigation // Proceedings of the Asian Conference on Computer Vision. P. 104–111. IEEE, 1998.
- [24] Santini S., Jain R. Integrated browsing and querying for image databases. IEEE Multimedia, 7(3): 26–39, 2000.
- [25] Van Rijsbergen C. J. Information Retrieval. L., 1979.
- [26] Vendrig J., Worring M., Smeulders A. W. M. Filter image browsing: Exploiting interaction in image retrieval // Proceedings of the Third International Conference on Visual Information and Information Systems (VISUAL'99), volume 1614 of Lecture Notes in Computer Science, P. 147–154. Springer, 1999.

Digital image collection navigation based on automatic classification methods

E. V. Myasnikov

The method for digital images collection navigation is proposed in this work. The central question of image navigation system construction is to build the projection of the collection into the two-dimensional space of navigation. The kernel of the method proposed is to build the projection in two steps. At the first step the hierarchical system of clusters is constructed. At the second step the initial space of image descriptions is projected into two-dimensional navigation space. This projection is constructed using the information about inter-image relationships and restrictions imposed by cluster hierarchy.

A survey of methods used for navigation system construction is given. The requirements to the method are considered. The results of experimental analysis are present. The proposed method is compared to known methods. The results of this work allow to draw a conclusion of ability to successfully apply the method developed.