

Разработка и анализ методов выделения объектов изображения

М.И. Хотилин¹, Р.А. Парингер^{1,2}, И.А. Рыцарев¹, Н.С. Кравцова¹

¹Самарский национальный исследовательский университет им. академика С.П. Королева, Московское шоссе 34А, Самара, Россия, 443086

²Институт систем обработки изображений РАН – филиал ФНИЦ «Кристаллография и фотоника» РАН, Молодогвардейская 151, Самара, Россия, 443001

Аннотация. Данная работа посвящена анализу изображений и нахождению располагающихся на них объектов. Рассматриваются существующие методы и подходы, имеющие место при выполнении данной задачи. В работе исследованы и намечены к доработке алгоритмы выделения и кластеризации объектов, присутствующих на изображениях.

1. Введение

Распознавание и выделение объектов на изображениях являют собой одну из наиболее значимых задач компьютерного зрения и обработки изображений. Она находит свое решение в различных сферах деятельности человека: в медицине, например при нахождении различного рода опухолей, в системах безопасности, при слежении за объектами, при восстановлении изображений и во многих других. Основным ключевым моментом при распознавании объектов является набор различных признаков (категорий) объектов и их описание [1]. Для этого требуется набор надежных и повторяемых параметров (дескрипторов), которые получаются из тестовых изображений или из модели объекта.

Под «распознаванием» обычно подразумевается тот факт, что исследуемый объект, представленный в виде совокупности наблюдений, следует относить к тому или иному взаимоисключающему классу [2]. Таким образом, в данном контексте, распознавание образов и объектов являет собой одну из разновидностей классификации [3], и в тех случаях, когда класс содержит только один объект, классификация является идентификацией, то есть присвоением рассматриваемому объекту однозначного названия. Иными словами, получение характерных точек на изображении позволит в дальнейшем классифицировать рассматриваемый объект.

В настоящее время существуют три основных направления в области распознавания:

- 1) сопоставление изображения с эталоном;
- 2) распознавание изображения по характерным точкам (способ получения данных характерных точек может отличаться)
- 3) распознавание посредством нейронных сетей.

Количество и тип методов могут изменяться, варьируясь в зависимости от объекта распознавания (например, для распознавания лиц наряду с указанными выше методами, применяются также: метод главных компонент, анализ оптических потоков изображений, и

другие). В данной работе рассматриваются два из указанных подходов, а именно распознавание посредством нейронных сетей и распознавание изображений по характерным точкам.

2. Распознавание объектов по характерным точкам

Рассмотрим алгоритм выявления характерных точек. Зачастую, использование данного алгоритма подразумевает под собой использование следующих основных этапов [1]:

- 1) получение нормализованного полутонового изображения;
- 2) поиск исследуемых областей;
- 3) выделения краев на исследуемой области;
- 4) преобразование рассматриваемого участка в монохромное изображение;
- 5) анализ полученного монохромного и полутонового изображения в исследуемой области.

При выполнении первого этапа используется следующая формула:

$$I(C) = 0,3 \cdot R(C) + 0,59 \cdot G(C) + 0,11 \cdot B(C),$$

где I - интенсивность в точке полутонового изображения; R, G, B (значения 0..255) – красная, зеленая и синяя компонента цвета C .

Для выделения контуров изображения или исследуемой области принято применять метод Собеля, позволяющий получить массив чисел, характеризующих изменения яркости в различных точках изображения. Метод Собеля использует для вычисления градиента первого порядка функции интенсивности специальные ядра, называемые операторами Собеля, имеющие следующий вид:

$$S_x = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}, S_y = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}, I = \begin{pmatrix} I_{11} & \dots & I_{1n} \\ \dots & \dots & \dots \\ I_{n1} & \dots & I_{nn} \end{pmatrix},$$

где S_x и S_y - X и Y операторы Собеля соответственно; I - матрица исходного изображения, содержащая в качестве элементов интенсивности в точках.

Ядра применяются к каждому пикселю изображения следующим образом: пиксель помещается в центр ядра, и производится умножение значения интенсивности в соседних точках на соответствующие коэффициенты ядра, после чего полученные значения суммируются. X и Y операторы, примененные к матрице 3×3 исходного изображения, дают величину горизонтальной и вертикальной составляющих градиента интенсивности в центральной точке данной матрицы. Величина градиента определяется как квадратный корень из суммы квадратов вертикальной и горизонтальной составляющих:

$$I' = \sqrt{(S_x \otimes I)^2 + (S_y \otimes I)^2}.$$

Это и есть искомый массив чисел.

Следующим этапом, после применения оператора Собеля, изображение подвергается инвертированию, что позволяет получить результат, указанный на рисунке 1.

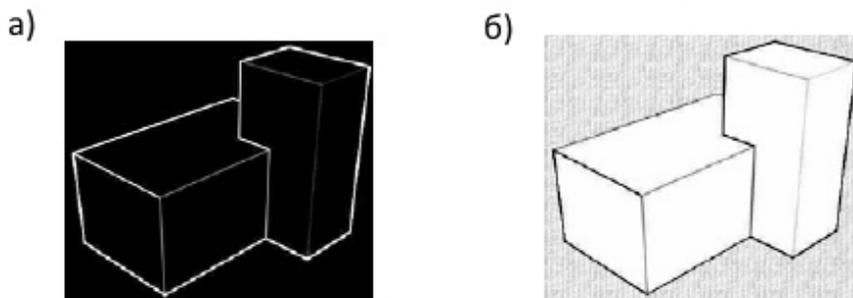


Рисунок 1. Инвертирование изображения: а) исходное изображение с применением оператора Собеля; б) инверсия.

Далее изображение преобразуется в монохромное, например, в соответствии с методом Отсу вычисления порога бинаризации для полутонового изображения [1].

$$I'' = Ot(\bar{I}') = \begin{cases} 0, \bar{I}'_{ij} \leq T^{opt}; \\ 1, \bar{I}'_{ij} > T^{opt}; \end{cases}$$

где T^{opt} - оптимальный для \bar{I}' порог бинаризации.

Заключительным этапом является анализ полученного изображения, зачастую производящийся посредством применения фильтра Габора, или путем вычисления коэффициента корреляции, что дает возможность получить решение указанной выше задачи.

3. Нейронные сети в обработке изображений

Использование нейронных сетей является иным подходом к решению данной задачи. Зачастую говоря о нейронных сетях, подразумевают полносвязные нейронные сети прямого распространения с обратным распространением ошибки. В сетях такого типа каждый нейрон связан с каждым, сигнал идет только в направлении от входного слоя к выходному, нет никаких рекурсий[4]. Пример подобной сети изображен на рисунке 2.

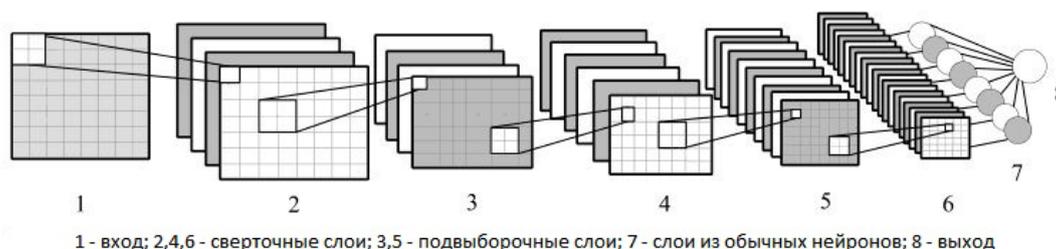


Рисунок 2. Общий вид свёрточной нейронной сети.

Однако применение нейронных сетей данного типа достаточно затруднительно, что обуславливается определенной вычислительной сложностью. При использовании полносвязной нейронной сети самым простым решением подготовки входных данных является выражение двумерной матрицы изображения в виде одномерного вектора. Например, в задаче распознавания рукописных цифр, при рассмотрении изображения 28×28 необходимо уже 784 входа. С учетом различных методик редукции сети и с учетом количества нейронов на скрытых слоях получаем итоговое количество связей порядка $\sim 10^6$. Обработка данной нейронной сети является затратной вычислительно и ресурсно. Кроме того задача распознавания подразумевает умение сети быть устойчивой к небольшим сдвигам, поворотам и изменению масштаба изображения, т.е. она должна извлекать из данных некие инварианты, не зависящие от почерка того или иного человека. Так какой же должна быть нейронная сеть, чтобы быть не очень вычислительно сложной и, в тоже время, более инвариантной к различным искажениям изображений?

Решением данной проблемы является использование, так называемых, свёрточных нейронных сетей. Идея свёрточных нейронных сетей заключается в чередовании свёрточных слоев (C-layers), субдискретизирующих слоев (S-layers) и наличии полносвязных (F-layers) слоев на выходе. Подобный подход архитектуры подразумевает три парадигмы, а именно:

- локальное восприятие;
- разделяемые веса;
- субдискретизация.

Под локальным восприятием подразумевается, что на вход одного нейрона подается лишь некоторая область изображения, а не всё изображение (или выходы предыдущего слоя) целиком.

Концепция разделяемых весов предполагает, что для большого количества связей используется очень небольшой набор весов, который называют ядром. Например, если на вход подается изображение 32×32 пикселя, то каждый из нейронов следующего слоя примет на

вход лишь небольшой участок данного изображения заданного размером, (например 6×6), причем каждый из фрагментов будет обработан одним и тем же набором.

Задача субдискретизации S-слоев заключается в уменьшении пространственной размерности изображения. Иными словами, входное изображение уменьшается в заданное количество раз путем усреднения. Зачастую используют уменьшение вдвое, однако изменение может быть неравномерным, например, втрое по вертикали и вдвое по горизонтали.

После прохождения сигнала через свёрточные слои, получается так называемая карта признаков. То есть, каждый фрагмент исходного изображения умножается поэлементно на ядро (матрицу весов) и результат суммируется. Эта сумма является пикселем выходного изображения, которое и называют картой признаков (рисунок 3).

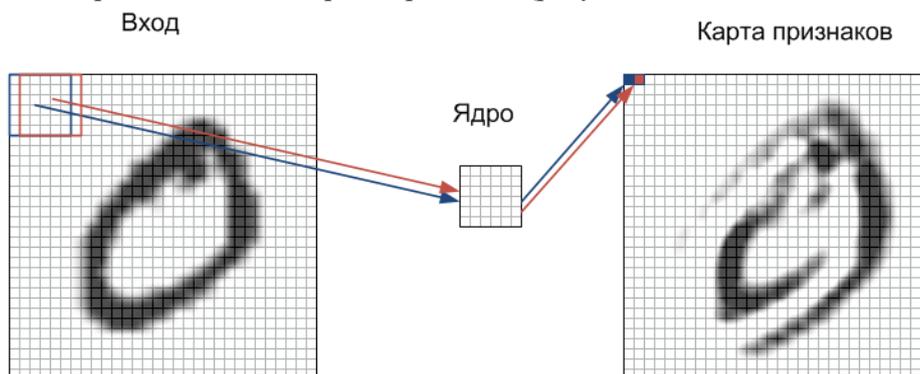


Рисунок 3. Получение карты признаков.

Чередование слоев позволяет составлять карты признаков из карт признаков, что на практике означает способность распознавания сложных иерархий признаков.

Зачастую, после прохождения нескольких слоев, карта признаков вырождается в вектор или скаляр, однако число подобных карт признаков может исчисляться сотнями. В таком виде они подаются на один-два слоя полносвязной сети. Выходной слой такой сети может иметь различные функции активации.

В настоящее время существуют и активно развиваются инструменты, направленных на создание нейронных сетей, обработку изображений и распознавание объектов на них.

4. Результаты исследования метода выделения объектов на изображении

Каждое изображение можно рассматривать как сцену, представляющую собой множество объектов различных цветов и размеров. Рассмотрим следующий пример.

Пусть дано исходное изображение, показанное на рисунке 4. На этом изображении каждый объект имеет определённый цвет, а рельеф объекта выделяется за счёт изменения интенсивности цвета. Такое изображение позволяет наглядно показать все этапы работы алгоритма.



Рисунок 4. Исходное изображение.

Необходимо выделить каждый из объектов, указанных на данном изображении. Предположим, что каждый объект однозначно определяется своими цветами – основным признаком, по которому можно его обозначить.

С задачей распознавания объектов на изображении помогает справляться такой подход, как кластеризация [6]. Кластеризация в данной задаче является исключительно полезной, поскольку:

- разбиение выборки на группы схожих объектов позволяет упростить дальнейшую обработку данных и принятия решений;
- если исходная выборка избыточно большая, то можно сократить её, оставив по одному наиболее типичному представителю от каждого кластера;
- в процессе кластеризации выделяются нетипичные объекты, которые не удаётся присоединить ни к одному из кластеров.

В данном случае может применяться иерархическая кластеризация, и задачи подобного типа называются задачами таксономии. Результатом таксономии является древообразная иерархическая структура, при этом каждый объект характеризуется перечислением всех кластеров, которым он принадлежит, обычно от крупного к мелкому. В качестве алгоритма кластеризации был выбран алгоритм k-means ввиду его простоты и быстроты, а также понятности. В качестве программной среды использовался Matlab компании Mathworks.

Применяя k-means к исходному изображению, получаем следующее сегментированное изображение (рисунок 5).



Рисунок 5. Сегментированное изображение.

Отсюда, даже невооруженным взглядом, можно видеть, что большинство кластеров не имеют четких границ, частично накладываются друг на друга. Также можно заметить, что появились нетипичные объекты, такие как разводы на заднем плане. Это связано с тем, что алгоритм слишком чувствителен к выбросам, которые могут исказить среднее и, необходимо выполнить предварительную обработку изображения.

Для предварительной обработки подходит медианная фильтрация, которая используется для эффективного удаления шума при сохранении важных для последующего распознавания деталей изображения. В результате применения медианного фильтра наклонные участки и резкие перепады значений яркости на изображениях не изменяются. Это очень полезное свойство именно для изображений, на которых контуры несут основную информацию. При медианной фильтрации зашумленных изображений степень сглаживания контуров объектов напрямую зависит от размеров апертуры фильтра и формы маски. При малых размерах апертуры лучше сохраняются контрастные детали изображения, но в меньшей степени подавляются импульсные шумы. При больших размерах апертуры наблюдается обратная картина. Оптимальный выбор формы сглаживающей апертуры зависит от специфики решаемой задачи и формы объектов [6]. Результат медианной фильтрации представлен на рисунке 6.

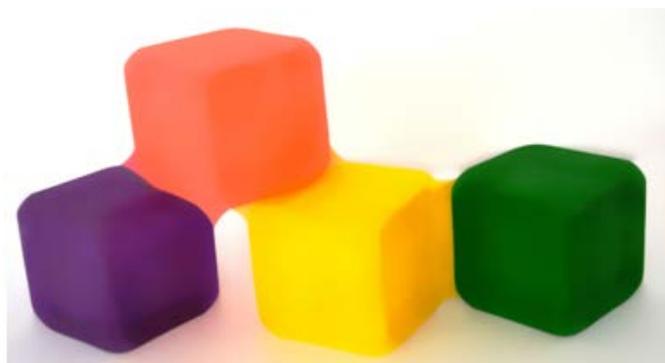


Рисунок 6. Результат медианной фильтрации.

Следующим этапом было принято решение использования нерезкого маскирования (unsharpmask). Данный прием используется в обработке фотографий и позволяет добиться эффекта ощущения большей их резкости за счет усиления контраста тональных переходов. Нерезкое маскирование усиливает локальный контраст изображения на тех участках, где изначально присутствовали резкие изменения градиентов цвета. Благодаря этому изображение визуально воспринимается как более резкое. Результат медианной фильтрации и нерезкого маскирования представлен на рисунке 7.



Рисунок 7. Результат обработки с помощью “unsharpmask”.

Из данного рисунка, путем применения нерезкого маскирования получаем более четкие границы областей с резким изменением цвета. На этом этапе задача, предобработки изображения может считаться решенной

В результате применения кластеризатора на основе метода k-means к предобработанному изображению, получаем изображение, разделённое на 5 кластеров белого, желтого, красного, зелёного и фиолетового цвета (рисунок 8).

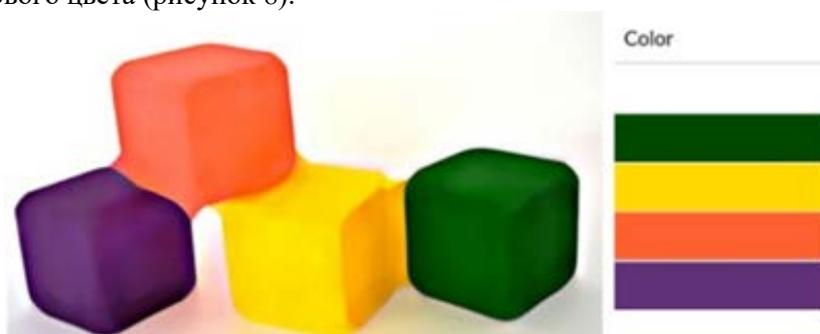


Рисунок 8. Результат сегментации предобработанного изображения.

Сравнивая рисунки 5 и 8 можно сделать выводы, что кластеризация после предобработки позволяет каждому объекту на изображении присвоить кластер с его собственным основным цветом, не порождая лишние кластеры.

Для выделения объекта на изображении применим следующий метод. После выполнения кластерного анализа, можно определить координаты целевого объекта на изображении. Для этого кластеры можно представить бинарными изображениями (масками), на которых значения каждого пикселя условно кодируются (0 — задний план или фон, 1 — объект). Иными словами, происходит сегментация с выделением границ (рисунок 9).

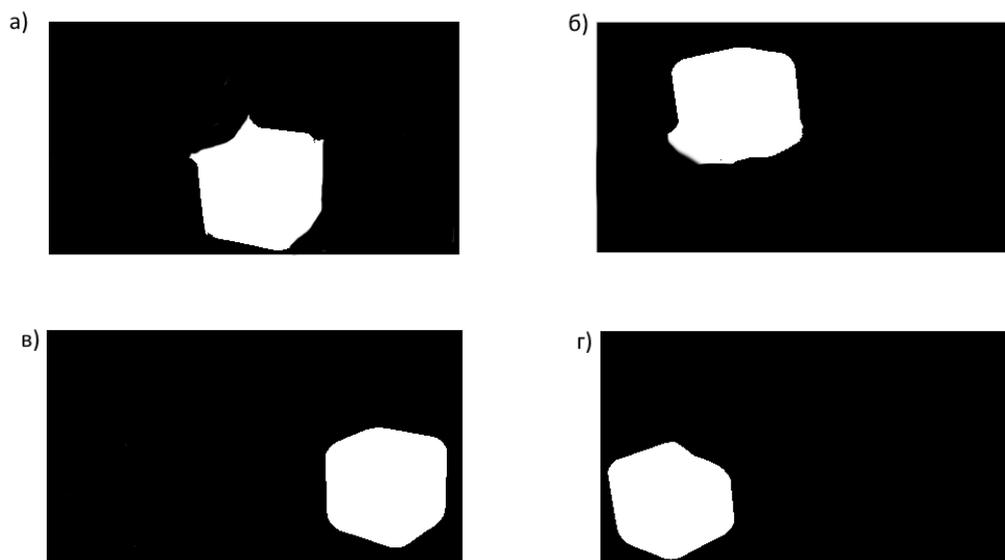


Рисунок 9. Результат выделения объектов на изображении: а – желтый объект, б – оранжевый объект, в – зелёный объект, г – фиолетовый объект.

5. Заключение

Распознавание и выделение объектов на изображениях являют собой одну из наиболее значимых задач компьютерного зрения и обработки изображений. Она находит свое решение в различных сферах деятельности человека. Зачастую, существующие методы обработки изображений и выделения объектов на них отлично работают с условно небольшими объемами исходных данных. Обработка же больших массивов исходных изображений занимает значительное время, что в ряде задач является абсолютно неприемлемым.

В настоящее время авторами данной статьи ведется работа над адаптацией указанных выше методов для работы с данными большого объема или BigData, а также ведется разработка программного средства, позволяющего осуществлять соответствующую обработку массивов изображений и распознаванию объектов на них.

6. Благодарности

Работа выполнена при поддержке Федерального агентства научных организаций (соглашение № 007-ГЗ/Ч3363/26) и Министерства образования и науки РФ в рамках реализации мероприятий Программы повышения конкурентоспособности Самарского Университета среди ведущих мировых научно-образовательных центров на 2013–2020 годы; грантов РФФИ № 15-29-03823, № 16-41-630761, № 17-01-00972.

7. Литература

- [1] Хайкин, С. Нейронные сети: полный курс, 2-е издание / С. Хайкин. – Издательский дом Вильямс, 2008. – 1131 с.
- [2] Сергеев, В.В. Применение методологии распознавания образов в задачах цифровой обработки изображений / В.В. Сергеев // Автометрия. – 1998. – Т 2. – С. 63-76.
- [3] Верхаген, К. Распознавание образов: состояние и перспективы / К. Верхаген, Р. Дёйн, Ф. Грун. – М.: РиС, 1985. – 104 с.
- [4] Применение нейросетей в распознавании изображений [Электронный ресурс]. – Режим доступа: <https://geektimes.ru/post/74326> (15.10.2017).

- [5] Самаль, Д.И. Алгоритмы идентификации человека по фотопортрету на основе геометрических преобразований / Д.И. Самаль // Диссертация на соискание ученой степени кандидата наук. Ин-т техн. киберн. НАН Беларуси. – Минск, 2002. – 170 с.
- [6] Графические фильтры на основе матрицы скручивания [Электронный ресурс]. – Режим доступа: <https://habrahabr.ru/post/43895> (01.11.2017).

Development and analysis of methods for selecting objects in an image

M.I. Khotilin¹, R.A. Paringer^{1,2}, I.A. Rytsarev¹, N.S. Kravtsova¹

¹Samara National Research University, Moskovskoye shosse, 34, Samara, Russia, 443086

²Image Processing Systems Institute of RAS - Branch of the FSRC "Crystallography and Photonics" RAS, Molodogvardejskaya street 151, Samara, Russia, 443001

Annotation. This work is devoted to the analysis of images and finding objects located on them. Existing methods and approaches that take place in the performance of this task are considered. The algorithms of selection and clustering of objects, which are presented on the images, are investigated and planned for finalization.

Keywords: large volume data, data analysis, image processing, object recognition, image recognition, k-means, median filter, neural networks.