

## МЕТОДЫ РАСПОЗНАВАНИЯ ОБРАЗОВ И ЗАДАЧИ ЛОГИЧЕСКОГО ВЫДЕЛЕНИЯ ОБЪЕКТОВ

**Иванько А.Ф., Иванько М.А., Горчакова Я.В.**

*Московский политехнический университет, Москва,*

*e-mail: alekfed@mail.ru, mihaleks@mail.ru, jane.gorchakova@gmail.com*

Данная статья состоит из двух взаимосвязанных частей. Первая часть дает основные понятия о теории распознавания образов, о машинном зрении, изображении как о двумерном массиве данных, характеризующемся тремя величинами и рассматриваемом в качестве модульной парадигмы, и описывает классические методы фильтрации изображения. Для приведения многомерных задач к одномерным и выделения ключевых характеристик изображения используются такие методы, как гистограммная обработка, бинаризация порогов полутоновых изображений, сегментация многомодальных изображений, преобразования Фурье и линейная фильтрация изображения в пространственной и частотной области. А вторая часть содержит алгоритмы логического решения задач на основе уже проведенной фильтрации изображения. Включая способ математической морфологии с алгоритмами выделения неизвестного объекта в известной среде и известного объекта в неизвестной среде, а также способ нахождения объектов посредством выделения особых точек, принадлежащих трем классам: стабильные на коротком отрезке времени, стабильные при малейших движениях и смене освещения и постоянно стабильные. И в том числе приводятся яркие примеры использования методов фильтрации изображений и выявления образов и очертаний объектов, искомых в заданных задачах.

**Ключевые слова:** теория распознавания образов, машинное обучение, гистограмма, сегментация, бинаризация, обработка изображения, машинное зрение

## METHODS OF RECOGNITION OF IMAGES AND THE TASK OF LOGICAL ISOLATION OF OBJECTS

**Ivanko A.F., Ivanko M.A., Gorchakova Ya.V.**

*Moscow Polytechnic University, Moscow,*

*e-mail: alekfed@mail.ru, mihaleks@mail.ru, jane.gorchakova@gmail.com*

This article consists of two interrelated parts. The first part gives the basic concepts of the theory of pattern recognition, of machine vision, the image as a two-dimensional data array, characterized by three values and regarded as a modular paradigm, and describes the classical methods of image filtering. Methods such as histogram processing, binarization of halftone image thresholds, segmentation of multimodal images, Fourier transforms and linear filtering of the image in the spatial and frequency domain are used to bring multidimensional problems to one-dimensional and to highlight key characteristics of the image. And the second part contains algorithms for logical problem solving based on the image filtering already performed. Including, a method of mathematical morphology with algorithms for isolating an unknown object in a known environment and a known object in an unknown environment, as well as a method for finding objects by selecting special points belonging to three classes: stable for a short period of time, stable with the slightest movement and change of lighting and constantly stable. And including vivid examples of the use of image filtering methods and the identification of images and outlines of objects that are required in the given tasks.

**Keywords:** pattern recognition theory, machine learning, histogram, segmentation, binarization, image processing, machine vision

На сегодняшний день самой актуальной темой развития в информационных технологиях является развитие нейронных сетей и роботостроение. Нельзя не заметить как роботы все чаще и чаще появляются в различных спектрах нашей повседневной жизни. Представить себе машину с автопилотом, повара-робота в ресторане, робота, создающего музыку, и многое другое, ранее представлялось нам невозможным. Люди хотят усовершенствовать свою жизнь, улучшить её с помощью внедрения рабочей силы в качестве роботов.

Поэтому теория распознавания образов пронизывает многие аспекты информационных технологий, как мы уже сказали, это

и машиностроение, и создание нейронных сетей, а также роботостроение.

Цель исследования: исследование и создание краткого описания методов распознавания образов и решения задач логического выделения объектов для использования данной информации в работах начинающих программистов и инженеров.

### *Методы исследования*

Мы использовали метод анализа вариантов для создания удобной и понятной классификации методов фильтрации изображений. В процессе исследования мы с помощью синтезирования уже существующих алгоритмов выявления образов наш-

ли самые удобные и интересные для решения практических задач.

### Результаты исследования и их обсуждение

Теория распознавания образа представляет собой раздел машинного обучения, который позволяет осуществлять классификацию и идентификацию предметов, явлений, процессов, сигналов и ситуаций. Наш мир состоит из сигналов, которые и обрабатывает человек. А представьте теперь, что сигналы обрабатывает не человек, а программа, которая видит, как человек, и может принимать «человекоподобные» решения.

Термин «машинное зрение» можно связывать с задачей интерпретации сцены наблюдения по ее двумерным проекциям, а также практическое использование этих результатов.

Мы не замечаем, как ежедневно имеем дело с самыми распространёнными компьютерами, с «компьютерным зрением»: считывальщики штрихкодов, которые используются для автономного чтения закодированной информации; анализ автомобильных номеров в камерах, анализирующих скорость на автомобильных дорогах, и многое другое.

Для того чтобы начать решать задачи распознавания объектов в реальной среде, нужно учитывать биологические и психологические аспекты человеческого зрения, сформулировать этапы наращивания опыта запоминания и классификации объектов. Представим человека который впервые увидел объект, например самолёт. У человека в голове возникает описательный образ предмета и, возможно, его характеристики и свойства. Впоследствии, повторной встрече, человеческий мозг, опираясь на опыт и воспоминания, воссоздаёт располагающийся в памяти образ и классифицирует объект. А как все-таки происходит распознавание у компьютера? Компьютер не может опереться на опыт, но в его программу закладываются определенные алгоритмы-правила, с помощью которых программа и выводит ответ. Поэтому изначально определяются массивы данных, описывающих факторы (факторы среды, условий, света и так далее), которые впоследствии и задают примеры, а после чего анализируются и выводятся в формулу или математический закон.

На сегодняшний день изображение рассматривается в согласии с *модульной парадигмой*, предложенной Д. Марром. В связи с этим обработка изображения осуществляется на основе нескольких уровней восходящей информационной линии, включая от

«иконического» до символического представления объектов [1].

Стоит акцентировать внимание на самом изображении, данное в видимом поле, которое на самом деле представлено в виде функции распределения яркости или цвета на двумерной плоскости,  $f(x, y)$ , где  $x$  и  $y$  – декартовы координаты, описывающие плоскость изображения. И при этом изображение состоит из дискретных цветных прямоугольников в виде регулярной прямоугольной матрицы. А вот с точки зрения математики цифровое изображение – это двумерная матрица  $Im[x, y]$  размера  $DimX \times DimY$ , где  $x$  – целое число от 0 до  $DimX-1$ , которое описывает номер элемента в строке матрицы,  $y$  – целое число от 0 до  $DimY-1$ , которое описывает номер строки матрицы, в которой расположен данный элемент. Элемент цифрового изображения, который имеет скалярное целочисленное значение, равное значению функции распределения яркости, называется пикселем [2].

Пиксель изображения кодируется тремя величинами  $\langle x, y, I \rangle$ . Первые две отвечают за геометрическое положение в плоскости, и третья –  $I$ , показывает яркость и интенсивность пикселя. Тем самым такое представление позволяет разработчику легче понимать задачу, каждый раз понижая порядок, и, соответственно, существенно экономить время. И еще нужно отметить, что яркостная составляющая  $I$  описывает одномерный массив гистограммы, а именно частоты встречаемости одинаковых по яркости пикселей. Но другие величины описываются двумерным массивом информации. Поэтому и возникает сведение двумерных задач к одномерным.

Так же используется рассмотрение изображения с точки зрения его цветовых характеристик. Зачастую используются изображения, закодированные в пространстве RGB. По осям располагаются цвета: красный, зелёный, синий ( $x, y, z$  соответственно), при этом интенсивность варьируется в диапазоне от 0 до 255. При этом каждый сигнал кодируется 8 битами информации.

Также существует модель HSV (Hue, Saturation, Value). Hue – угловая характеристика, показывающая интенсивность цвета, иначе цветовой тон. Saturation – насыщенность цвета, этот параметр показывает «чистоту цвета». Value – ось, определяющая количественную характеристику светового потока. Данная модель является нелинейной моделью RGB и близка к человеческому зрению.

Этапы обработки включают: преобразование изображений, сегментацию, выделе-

ние геометрической структуры и определе-ние структуры и семантики.

В настоящее время существует несколько алгоритмов, методов и математических формализмов, которые используются при анализе изображений.

1. Гистограммные преобразования.
2. Анализ проекций.
3. Линейная и нелинейная фильтрация.
4. Яркостная и текстурная сегментация.
- И другие.

Для начала вернёмся к описанию признаков. Смотри на идущего человека, наш мозг выделяет важные и второстепенные объекты (в данном примере назовём их признаками). Человек в нашем случае важен, так как именно его мы хотим идентифицировать, а сцена вокруг него является второстепенным признаком. Так и для компьютера существуют локальные и глобальные признаки. Но в понимании компьютера большую часть будет занимать окружающая человека среда, к примеру дорога в лесу, но не человек. Поэтому существует множество способов, нахождения объектов в реальной среде по типу человеческого зрения.

1. Гистограммные преобразования.
2. Анализ проекций.
3. Линейная и нелинейная фильтрация.
4. Яркостная и текстурная сегментация.
- И другие.

Рассмотрим гистограммную обработку изображения. Далее будем называть гистограммой частоту встречаемости пикселей одинаковой яркости. Яркостным преобразованием изображения называется преобразование двумерных функций яркости, которые описываются формулой

$$I'(x, y) = f(I(x, y)),$$

где  $f$  является функцией отображения яркости. Примером яркостных преобразований может служить линейное преобразование яркости:

$$f(I) = aI + b.$$

Коэффициент  $a$  определяет изменение контраста изображения, а  $b$  определяет изменение средней яркости изображения. Эти преобразования можно назвать фотографическими, так как с помощью них можно настраивать характеристики выдержки и диафрагмы. Функция отображения яркости задается таблицей отображения LUT (*Look-Up-Table, просмотровой таблицы*).

Также часто используют инвертирование яркости при обработке изображения,  $LUT[i] = 255 - i$ . Результатом инвертирования получается негатив.

Часто используемый метод бинаризации порогов, при котором устанавливается по-

рог и сравниваются две области изображения ниже и выше порога. Пороговая сегментация позволяет приблизить гистограмму изображения к кривой с весовыми суммами больше не менее двух вероятностей интенсивности с нормальным распределением. При этом порогом можно считать множество рядом расположенных уровней яркости с минимумом вероятности.

Представим изображение в виде двумерного массива  $Im$  с размером  $X*Y$  и целочисленными значениями от 0 до 255 включительно. Получаем представление гистограммы в виде одномерного массива  $Hist[0...255]$ , который содержит число пикселей со значением  $i$ , равным порядку нашего элемента. После чего рассматривается «подгистограмма»  $Hist[k...l]$ ,  $0 \leq k \leq l \leq 255$ . Проводим оценку дисперсии яркости  $DISP(k, l)$ . Пусть задан порог  $t$ :  $0 \leq t \leq 255$ , для которого вычисляется «критерий разделимости»

$$SC(t) = 1 - \frac{DISP(0, t) + DISP(t+1, 255)}{DISP(0, 255)}.$$

Алгоритм Отсу, при  $0 \leq SC(t) \leq 1$ , позволяет эффективно и устойчиво определить адаптивный порог бинаризации изображений

$$T = \arg \max SC(t), \text{ где } t \in \{0, \dots, 255\}.$$

Еще один метод – сегментация много-модальных изображений, с помощью которого можно оценивать количество и выраженность мод на гистограмме, который, в частности, основывается на оптимизации глобального критерия разделимости на количество мод больше одной. Пусть дан  $n + 1$ -мерный вектор  $t = \langle t_0, \dots, t_n \rangle$  от 0 до 255 [3].

Получается, среднеквадратичный критерий выбора порогов сегментации равен

$$\sum_{i=0}^{n-1} DISP(t_i, t_i + 1) \rightarrow \min(t_1, \dots, t_n - 1).$$

Вследствие того, что гистограмма является одномерным массивом, то задача сводится к динамическому программированию, в результате чего набор порогов сегментации обеспечивает минимальное среднеквадратическое отклонение на  $n$  уровней изображения.

$$\sum_{i=0}^{n-1} DISP(t_i, t_i + 1) + \alpha n \rightarrow \min(t_1, \dots, t_n - 1).$$

Далее мы рассмотрим линейную фильтрацию изображения в пространственной и частотной области, которая заключается

в том, что вычисляется линейная комбинация значений яркости пикселя с коэффициентами матрицы весов фильтра – называется ядром или маской линейного фильтра.

Маска фильтра представляется в виде матрицы

Mask[-1,-1]	Mask[0,-1]	Mask[1,-1]
Mask[-1,0]	Mask[0,0]	Mask[1,0]
Mask[-1,1]	Mask[0,1]	Mask[1,1]

А фрагмент с центральным пикселом

Im[x-1,y-1]	Im[x,y-1]	Im[x+1,y-1]
Im[x-1,y]	Im[x,y]	Im[x+1,y]
Im[x-1,y+1]	Im[x,y+1]	Im[x+1,y+1]

Тогда результатом фильтрации будет

$$Im'[x,y] = \sum_{i=-hWinX}^{hWinX} \sum_{j=-hWinY}^{hWinY} Im[x+i,y+j] \times Mask[x+i,y+j],$$

$hWinX$  и  $hWinY$  – это полуширина и полувысота соответственно.

Самым простым видом оконной фильтрации является скользящее среднее в окне, с результатом значения математического ожидания, элементы маски которой равны  $1/n$ .

Классические методы фильтрации можно также применять для решения данного вида задач. С помощью преобразований Фурье можно представить любую функцию. Для двумерного массива размера  $M \times N$  для дискретного преобразования вычисляется по формуле

$$Guw = \frac{1}{N} \sum_{n=1}^{N-1} \left[ \frac{1}{M} \sum_{m=0}^{M-1} xmne^{\frac{-2\pi j m \omega}{M}} \right] e^{\frac{-2\pi j n u}{N}}.$$

Так как компоненты преобразования Фурье называют пространственными частотами.

Рассмотренные методы фильтрации изображений позволяют решать разнородные логические задачи. Теперь нужно пояснить, что результатом фильтрации являются пригодные для обработки данные, с помощью которых можно выделить объекты и их свойства [4].

Одним из логических переходов является способ математической морфологии, это можно назвать операциями наращивания и эрозии бинарных изображений. С помощью морфологии можно выделить неизвестные объекты на изображениях известной сцены. Идея решения заключается в том, что существуют два изображения одной сцены (А и В). При том эти два сним-

ка сняты в условиях разной освещенности и разных характеристик видимого поля, и на изображении В появился какой-то объект. Но при сравнении этих двух изображений мы получаем разностное изображение, с помощью которого все же невозможно понять различия исходников. Поэтому приходится создавать вспомогательное изображение С с формой изображения А, но яркость элементов будет оцениваться по В. В данном случае новое изображение С называется проекцией В на форму А. А далее существует алгоритм распознавания:

1. Выделение связных областей на изображении А.
2. Вычисление средней яркости по областям А и В.
3. Формулировка С по форме А с яркостями из В.
4. Нахождение разности С и В.
5. Выделение области с существенной разностью интенсивности пикселей.

Рассмотрим пример на двух фотографиях зимнего и летнего периода. Теперь нам нужно найти объект, отсутствующий на «летнем» фото, но находящийся на «зимнем». А внизу представлены два изображения с разностью яркостей сцены [5].

Еще существует класс задач, который называется поиск известных объектов на неизвестной сцене. К примеру, у нас есть два изображения: снимок сцены в видимом диапазоне видимого излучения и автомобиль на изображении ИК-диапазона. Далее выстраивается график зависимости близости по форме объектов сравнения, где точка максимума соответствует наибольшему сходству изображений по форме. Яркости участков, которые являются автомобилем или ему принадлежат, очень различаются, но сохраняют очертание формы, что позволяет находить координаты объекта из ИК-изображения на снимке с видимым спектром.

Еще одними характеристиками объекта можно считать особые точки, которые являются уникальными характеристиками, с помощью которых можно сравнивать объекты. Они находятся с помощью большинства алгоритмов, например, выделение особых точек в соседних кадрах при смене освещения или через промежуток времени и так далее [6]. Выделяют несколько классов особых точек.

Первый из которых содержит особые точки, которые являются стабильными на коротком отрезке времени (в течение нескольких секунд). Данный вид точек служит, чтобы можно было вести объект между соседними кадрами видео или сведения изображения с двух ракурсов и камер. К ним относятся локальные максимумы, макси-

мумы дисперсии, определенные градиенты и так далее.

Второй класс включает особые точки, которые стабильны при малейших движениях и смене освещения и служат для того, чтобы построить машинное обучение классификации другого вида объектов. Так же сюда можно отнести точки, которые были найдены методом гистограмм направленных градиентов.

К третьему классу можно отнести стабильные точки, которые находятся методами SHIFT и SURF [7].

Представленные в исследовании методы логического и технического анализа на основе редактирования и фильтрации изображений позволили выделить и представить в виде классификации наиболее простые и понятные для практического применения.

### Заключение

И в заключение можно сказать, что на сегодняшний день не существует единого метода или способа создания алгоритмов анализа изображения с выделением образов объектов. Так как машинное зрение находится в процессе развития и находит разнообразные ключи к пониманию данного этапа машинного обучения. Если посмотреть в прошлое, то мы заметим, как наука сделала огромный шаг к автоматизации чело-

веческой деятельности. Поэтому в скором времени, возможно, произойдет научный прорыв и появятся новейшие методы и теории распознавания образов. А сейчас нужно трудиться и улучшать уже существующие для точной детализации и усовершенствования методов для определения объектов из динамичной сцены в реальном времени для обработки информации в компьютере работа.

### Список литературы

1. Техническое зрение Wiki ресурс «Техническое зрение» Код доступа. [Электронный ресурс]. URL: [http://wiki.technicalvision.ru/index.php/Заглавная\\_страница](http://wiki.technicalvision.ru/index.php/Заглавная_страница) (дата обращения: 16.05.2019).
2. Визильтер Ю.В., Желтов С.Ю., Бондаренко А.В., Ососков М.В., Моржин А.В. Обработка и анализ изображений в задачах машинного зрения: Курс лекций и практических занятий. М.: Физматкнига, 2010. 672 с.
3. Журавель И.М. Краткий курс теории обработки изображений (MatLab Image Processing Toolbox). [Электронный ресурс]. URL: <http://matlab.exponenta.ru/imageprocess/book2/index.php> (дата обращения: 01.06.2019).
4. Гонсалес Р., Вудс Р. Цифровая обработка изображений. 3-е издание, исправленное и дополненное. М.: Техносфера, 2012. 1104 с.
5. Richard Szeliski. ComputerVision: Algorithmsand Applications. [Электронный ресурс]. URL: <http://szeliski.org/Book/> (дата обращения: 01.06.2019).
6. Шапиро Л., Стокман Дж. Компьютерное зрение. М.: Бинум, 2009. 763 с.
7. Пару слов о распознавании образов habr. [Электронный ресурс]. URL: <https://habr.com/ru/post/208090/> (дата обращения: 16.05.2019).