

А.В. Смирнов, Е.Е. Власов

КРИТИЧЕСКИЙ ОБЗОР СУЩЕСТВУЮЩИХ МЕТОДОВ ВЫДЕЛЕНИЯ ОБЪЕКТОВ В ПОТОКЕ ВИДЕО

Аннотация. Выделение контуров объектов в потоке видео в интеллектуальных системах видеонаблюдения является одной из основных операций при обработке изображения для дальнейшего анализа, поскольку контур содержит всю необходимую информацию для распознавания объектов по форме. Цель данной статьи – дать критический обзор существующих методов выделения объектов в потоке видео. Также статья посвящена анализу методов распознавания изображений и поиска их в видеопотоке. Проанализирована эволюция структуры сверточных нейронных сетей, используемых в области диагностики компьютерных видеопотоков. Произведен обзор существующих методов выделения объектов в потоке видео. Распознавание всего объекта лишь по его контуру позволяет не рассматривать внутренние точки изображения и существенно сократить объем обрабатываемой информации, предоставляя возможность осуществлять анализ изображений в режиме реального времени.

В данной статье даны ответы на вопросы, что представляет собой методология и методика научных исследований. Так, в статье рассмотрена проблема выделения объектов в потоке видео в задачах обнаружения тревожных событий интеллектуальными системами видеонаблюдения. В результате сделаны выводы, что перспективным направлением дальнейших исследований является разработка алгоритмов выделения контуров изображений объектов, реализующих двухмасштабную статистическую модель изображения. Практическая значимость статьи заключается в том, что с целью улучшения основных характеристик интеллектуальных систем видеонаблюдения предложены алгоритмы выделения контуров изображений объектов, необходимых для обеспечения выявления четырех типов тревожных событий: появление и нахождение объекта в зоне наблюдения, перемещение объекта в запрещенном направлении, оставление и опрокидывание предмета.

Ключевые слова: интеллектуальные системы видеонаблюдения, видеоаналитика, поток видео, цифровая обработка изображений, обнаружение объектов, отслеживание объектов, контуры изображений объектов, контурный.

A.V. Smirnov, E.E. Vlasov

A CRITICAL REVIEW OF THE EXISTING METHODS OF OBJECT SELECTION IN A VIDEO STREAM

Abstract. The selection of contours of objects in the video stream in intelligent video surveillance systems is one of the main operations in image processing for further analysis, since the contour contains all the necessary information for recognizing objects by their shape. The purpose of this article is to give a critical overview of existing methods for selecting objects in a video stream. The article is also devoted to the analysis of image recognition methods and their search in the video stream. The evolution of the structure of convolutional neural networks used in the field of diagnostics of computer video streams is analyzed. An overview of existing methods for selecting objects in the video stream is made. It is worth saying that the recognition of the entire object only by its contour allows you not to consider the internal points of the image and thus significantly reduce the amount of information processed, providing the opportunity to analyze images in real time. In this article, based on available sources and published literature, answers are given to questions about what is the methodology and methodology of scientific research. So, the

Смирнов Артем Владимирович

ассистент базовой кафедры № 536 программного обеспечения систем радиоэлектронной аппаратуры, МИРЭА – Российский технологический университет, Москва. Сфера научных интересов: математическое моделирование алгоритмов обнаружения движущихся объектов. Автор семи опубликованных научных работ.
Электронный адрес: 5725757@mail.ru

Власов Евгений Евгеньевич

кандидат технических наук, доцент базовой кафедры № 536 программного обеспечения систем радиоэлектронной аппаратуры, МИРЭА – Российский технологический университет, Москва. Сфера научных интересов: математическое моделирование алгоритмов обнаружения движущихся объектов. Автор 20 опубликованных научных работ.
Электронный адрес: vlasov@mirea.ru

article considers the problem of selecting objects in the video stream in the tasks of detecting alarming events by intelligent video surveillance systems. As a result of the work, the author came to the conclusion that a promising direction for further research is the development of algorithms for selecting the contours of images of objects that implement a two-scale statistical image model. The practical significance of the article lies in the fact that, in order to improve the basic characteristics of intelligent video surveillance systems, algorithms are proposed for selecting the contours of images of objects necessary to ensure the identification of four types of alarming events: the appearance and location of an object in the observation zone, the movement of an object in a prohibited direction, the abandonment of an object and the overturning of an object.

Keywords: intelligent video surveillance systems, video analytics, video stream, digital image processing, object detection, object tracking, contours of images of objects, contour.

Введение

Целью статьи является обзор существующих методов выделения объектов в потоке видео интеллектуальными системами видеонаблюдения. Методы распознавания образов, основанные на искусственном интеллекте, применяются в системах компьютерного зрения (далее – СКЗ) для автоматической классификации образов объектов или ситуаций по кадрам видеопотока и отнесения их к определенному классу.

Опираясь на имеющиеся источники и опубликованную литературу, даны ответы на вопросы о том, что представляет собой методология и методика научных исследований.

В результате работы были сделаны выводы, что перспективным направлением дальнейших исследований является разработка алгоритмов выделения контуров изображений объектов, реализующих двухмасштабную статистическую модель изображения.

Практическая ценность работы заключается в разработке улучшения основных характеристик интеллектуальных систем видеонаблюдения, предложены алгоритмы выделения контуров изображений объектов, необходимых для обеспечения выявления четырех типов тревожных событий: появление и нахождение объекта в зоне наблюдения, перемещение объекта в запрещенном направлении, оставление и опрокидывание предмета.

На видеокдрах потока могут быть замечены физические и технические объекты, каждый из которых обладает набором свойств. Важно учитывать тот факт, что поток описывается взаимодействием этих объектов. Тогда задачами алгоритмов, применяемых в СКЗ, является нахождение объектов на изображении и извлечение их ключевых признаков с последующей классификацией и выработкой управляющих воздействий [1; 2].

Теоретические основы исследования

Для проведения исследования использовались следующие методы исследования: общенаучные (диалектический, анализа и синтеза имеющихся литературных данных, сравнения и аналогии, аннотирование, конспектирование и реферирование информации, полученной из современных научных источников), специальные (системный, сравнительного анализа, моделирование и др.). В результате проведенного анализа проведено сравнительное изучение методов выделения объектов, распознавания изображений и поиска их в видеопотоке. Проблема исследования заключается в выделении в нем объектов в задачах обнаружения тревожных событий интеллектуальными системами видеонаблюдения. Так, абсолютное большинство современных нейросетей применяется для анализа изображений. Поскольку любое видео представляет собой набор упорядоченных изображений, его необходимо разбить на отдельные кадры или выбирать кадры через определенные интервалы, к которым будет применяться анализ нейросетями.

Задача обнаружения объектов является одновременно задачей регрессии и классификации. Прежде всего для оценки пространственной точности нужно удалить блоки с низкой степенью распознаваемости (обычно модель выводит намного больше блоков, чем реальные объекты). Затем используется участок пересечения объединений рамок (Intersection of Union – IoU) со значениями от 0 до 1. Это соответствует зоне перекрытия между прогнозируемым и реальным изображением в видеопотоке. Чем выше IoU, тем лучше прогнозируемое расположение рамок для данного объекта. Обычно сохраняются ограничительные коэффициенты кандидатов с IoU больше, чем некоторый порог.

Первые работы, в которых описывалась концепция CNN (Convolutional Neural Networks), появились еще в 60-х годах прошлого века. Однако тогда она не была успешно применена на практике из-за нехватки вычислительных мощностей и ограниченности данных для обучения. Внимание к этому типу архитектур снова возросло в 2020 году, поскольку на данный момент описанные выше проблемы легко решаемы. С тех пор разработано большое количество архитектур сверточных нейросетей и подходов к обработке данных, основные из которых проанализированы ниже.

Статья [3] считается одной из самых влиятельных в этой области. Исследователи описывают созданную ими глубокую сверточную нейросеть Alex Net, которая состояла из пяти сверточных чередующихся слоев. Сеть способна классифицировать объекты среди 100 различных категорий, а общее число параметров в сети составляло 6,5 млн. Авторы исследования [4] усовершенствовали архитектуру Alex Net, придерживаясь принципа «просто, но глубоко». В предложенной ими нейросети использовались фильтры для свертки размерности 3×3 в отличие от 7×7 в Alex Net. Эта статья оказала большое влияние на научное сообщество, поскольку подтвердила убеждение, что нейросети с увеличением глубины могут значительно улучшить результаты, хотя и требуют большего объема данных и вычислительной мощности для обучения.

Разработанная сеть GoogLe Net, описанная в статье [5], состояла из 22 слоев и существенно отходила от общепринятой в то время архитектуры с последовательным со-

единением сверточных и подвыборочных слоев, то есть модулей ввода. Авторы статьи отмечают, что применение таких модулей позволяет существенно сократить использование памяти и вычислительной мощности. В 2015 году Microsoft представила нейросеть со 152 слоями, которая была описана в работе Deep Residual Learning for Image Recognition [6; 7], установившей новые рекорды и практически не уступающей человеческому глазу в сфере обнаружения объектов на изображении.

Хиршик Р. вместе с другими исследователями создали Radial Based CNN (R-CNNs) [8]. В статье описывается механизм, в котором рамки разных размеров «скользят» по изображению, и к каждому снимку из окна применяется нейросеть. Статья [9] посвящена проблеме нахождения возможных мест расположения объектов для его распознавания. Исследователи представляют метод выборочного поиска, который сочетает в себе сильные стороны метода исчерпывающего поиска и метода сегментации объектов.

Первые модели R-CNN [10] интуитивно начинались с поиска объектов, к которым затем применялась классификация. В них используется метод выборочного поиска для локализации объектов. Модель объединяет выборочный метод поиска [11] для выявления предложений региона и глубокого обучения для выявления объекта в этих регионах [12].

Fast-R-CNN [13] является прямым потомком R-CNN и напоминает оригинал многими способами. Faster-R-CNN [14] сейчас является канонической моделью для обнаружения объектов на основе глубокого обучения.

В настоящем исследовании используется модель You Only Look Once (YOLO) как наиболее точная и охватывающая все преимущества рассмотренных нейросетей [15]. Она позволяет напрямую прогнозировать ограничительные рамки и вероятности класса с использованием единой сети в пределах одной оценки. Аналогично модели YOLO была разработана модель Single-Shot Detector-SSD [8], чтобы одновременно предсказывать все ограничивающие рамки и вероятности классов с помощью единой CNN.

Результатом является несколько вариантов видеопотоков, которые могут содержать объект, собранный путем объединения нескольких небольших видеопотоков. На Рисунке 1 визуализированы небольшие видеопотоки в изображении и объединены по иерархическому принципу, то есть последняя группа – это рамка, содержащая все изображение целиком.



Рисунок 1. Визуализация работы метода выборочного поиска
Рисунок создан на основе [11]

Обнаруженные области (см. Рисунок 2) объединяются по разным цветовым промежутоккам и показателям сходства. Каждое предложение видеопотока изменяется так, чтобы соответствовать входу CNN, из которого мы получаем векторные признаки размера 4096 bit. Вектор признаков поступает в несколько классификаторов для получения вероятностей, относящихся к каждому классу.

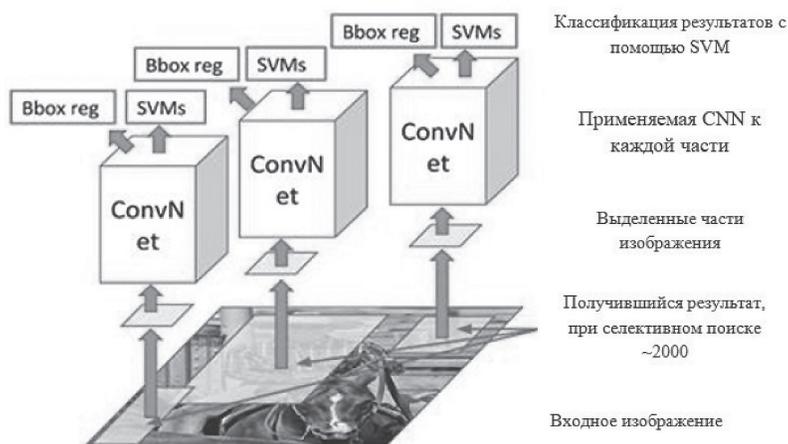


Рисунок 2. Типовая структура работы R-CNN
Рисунок создан на основе [12]

Каждый из этих классов имеет классификатор на основе SVM, обученный вывести вероятность обнаружения этого объекта для определенного вектора функций. Сформированные векторы поступают в модуль линейной регрессии для коррекции форм ограничительной рамки для предложения региона, что уменьшает ошибку локализации.

Благодаря Fast-R-CNN можно улучшить скорость обнаружения через два основных изменения: выполнение функции выделения признаков изображения перед применением алгоритма поиска регионов (таким образом, изображение проходит через CNN только один раз) и замена SVM на слой softmax, что позволяет расширить возможности нейронной сети для прогнозирования вместо создания и тренировки отдельных SVM-моделей. Остается лишь одна проблема в Fast-R-CNN – выборочный алгоритм поиска для создания регионов предложений. Однако появившаяся позже Region Proposal Network (RPN) напрямую генерировала видеопотоки предложений, предусматривала ограничивающие рамки и обнаруживала объекты. Модель CNN принимает в качестве входного параметра целое изображение и создает карты фильтров (признаков) (см. Рисунок 3).

Окно размером 3×3 перемещает все фильтры признаков и выводит вектор признаков, который передается в два полностью связанных слоя (Fully Connected – FC), – один для коррекции регрессии и один для классификации рамок. Предсказание для предложенных областей происходит FC-слоями. Faster R-CNN использует RPN, чтобы избежать выборочного метода поиска. Это ускоряет обучение и тестирование и улучшает эффективность. RPN может использовать предварительно подготовленную модель над набором данных Image Net для классификации и точно настроена на массив данных с VOC PASCAL (см. Рисунок 4).

Fast R-CNN

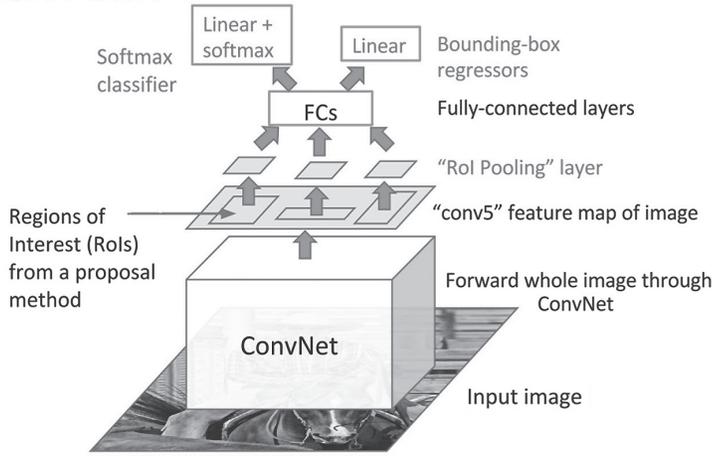


Рисунок 3. Схема работы FasterR-CNN
 Рисунок создан на основе [13; 14]

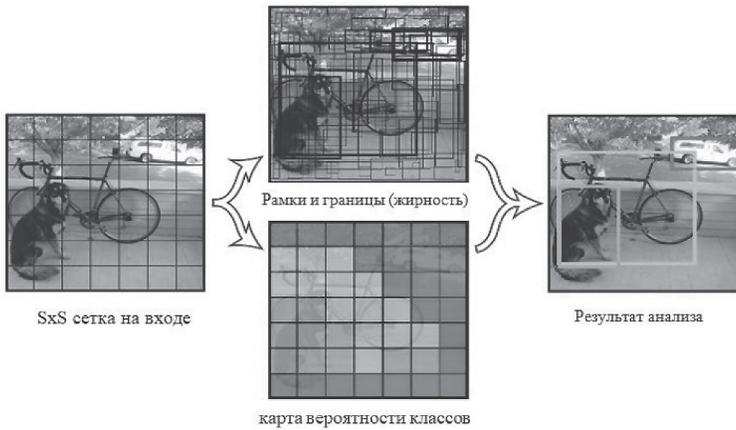


Рисунок 4. Модель визуализации работы предсказаний YOLO
 Рисунок создан на основе [15]

Простота модели YOLO позволяет создавать прогнозы в реальном времени. Первоначально модель принимает изображение как входной параметр. Она разделяет его на сетку размером $S \times S$. Каждая ячейка этой сетки прогнозируется в ограничительных рамках с оценкой уверенности для каждой ячейки. Используемая в YOLO CNN создана на основе GoogLe Net, состоящей из Inception модулей (см. Рисунок 5).

Модель SSD принимает изображение на входе, которое проходит через несколько слоев с разными размерами фильтров (10×10 , 5×5 и 3×3). Специальные карты из сверточных слоев в разных местах сети используются для прогнозирования ограничительных рамок. Они обрабатываются специфическими сверточными слоями с фильтрами размером 3×3 , называемыми дополнительными функциональными слоями, для создания набора ограничительных рамок, подобных якорным рамкам Fast-R-CNN [8; 12].

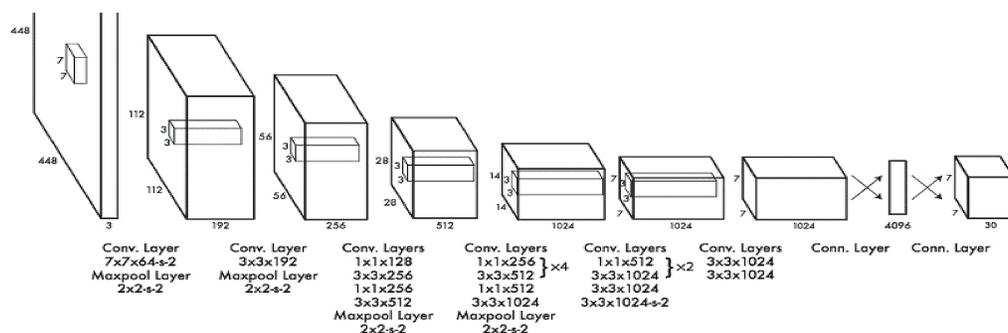


Рисунок 5. Схема CNN в YOLO

Рисунок создан на основе [15]

Каждая рамка имеет 4 параметра: координаты центра, ширина и высота. В то же время она создает вектор вероятности, соответствующий уверенности принадлежности объекта для каждого из классов (см. Рисунок 6).

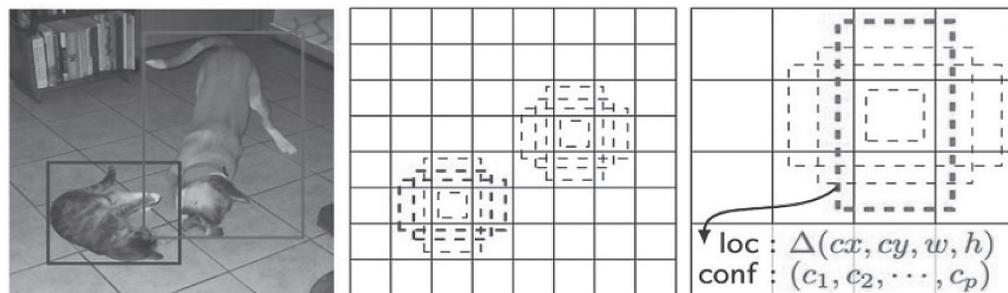


Рисунок 6. Модель построения вероятностей в SSD

Рисунок создан на основе [8]

В модели SSD используется метод немаксимального подавления (Non-Maximum Suppression) для хранения наиболее подходящих ограничительных коробок. Кроме того, используется метод жесткого удаления негативов для удаления отрицательных рамок. Лучшие коробки выбираются в зависимости от соотношения между отрицательными и положительными изображениями, что не должно превышать $1/3$.

Поскольку используется библиотека Open CV, алгоритм для детектирования изображений может быть заменен на использование сверточной нейросети. Модуль оптимизации принимает на входе видеопотоки с изображениями, обнаруженные на фазе детектирования, и выполняет оптимизацию этих изображений средствами Open CV: выравнивание, поворот, увеличение контрастности, освещение и др. Модуль оптического распознавания текста (OCR) использует библиотеку современной системы для распознавания текста на основе рекуррентных нейросетей типа LSTM (longshort-term memory). Выбранная система Tesseract 4.00. Tesseract 4.00 имеет открытый исходный код и относится к лучшим системам оптического распознавания текста. Для применения в прикладной программе распознавания номерных знаков была выбрана среда Open ALPR через специализацию и возможность модификации с языком программирования C#.

Применение

Важными задачами, которые должны решать интеллектуальные системы видеонаблюдения, является идентификация объектов и определение траекторий их движения; изменение скорости движения объектов; выявление тревожных событий в задачах объектно-территориальной защиты в режиме реального времени.

Весь спектр задач, которые приходится решать при распознавании объектов на изображениях, можно разделить на две группы:

1. Распознавание или классификация изображений.
2. Поиск и распознавание объектов (специфических локальных участков) на изображениях.

Такое разделение связано с особенностями реализации процесса распознавания.

В первой группе задач распознавание или классификация осуществляется для всего изображения в целом, то есть все изображения в процессе распознавания относят к одному из нескольких классов. Таким образом, решение задачи распознавания в данной группе – реализация отображения, где изображение – условный номер класса.

В задачах второй группы процесс распознавания относится к технологии обработки изображения, связанной с поиском геометрических объектов по всему участку наблюдения. Объектами в данной ситуации являются относительно маленькие локальные участки, возникающие в любой точке изображения. Причем информация о наличии объектов на изображении, их количестве, ориентации, размерах чаще всего отсутствует. Результатом решения задачи распознавания в этой ситуации является не только класс найденного объекта, но и его характеристики: состояние, размер, цвет, ориентация объекта в плоскости изображения и др. Примером задач второй группы является задача объектно-территориальной защиты, которые заключаются в выявлении в реальном времени четырех типов тревожных событий: движение в стерильной зоне, оставление предмета и опрокидывание предмета, отслеживание перемещения объекта совмещенными видеокамерами [3]. Следует отметить, что неопределенность в ряде характеристик объектов делает задачу их поиска и распознавания на изображении в математическом и вычислительном плане более сложной по сравнению с задачами первой группы.

Построение описания изображения на основе его представления с использованием признаков – одна из самых сложных задач в процессе построения любой системы распознавания визуальной информации. И если в пределах некоторых математических моделей удалось формализовать процесс классификации, то процесс выбора признаков до сих пор остался эвристической процедурой, зависимой как от предметной сферы, так и от разработчика. В то же время определенный опыт, накопленный за годы использования средств распознавания образов и обработки изображений для решения практических задач, позволяет выделить ряд основных групп признаков, успешно применяемых для описания и распознавания изображений.

Достаточно часто в качестве классификационных признаков берутся признаки, основанные на использовании геометрических характеристик (геометрические размеры изображенного объекта по вертикали или горизонтали; расстояние между наиболее удаленными точками на изображенном объекте; периметр и площадь) изображенного объекта, компактность объекта как отношение между его площадью и периметром, числовые характеристики описанных или вписанных в изображение объекта геометрических фигур: круги, многоугольники и др.).

Одной из основных операций при обработке изображения для его дальнейшего анализа является выделение контуров, поскольку контур несет основную информацию об изображении, и именно по контурам осуществляется обнаружение объектов или их составных элементов [4].

Контурный анализ

Контурный анализ является совокупностью методов выделения, описания и обработки контуров изображений, позволяющих описывать, хранить, сравнивать и отыскивать объекты, представленные в виде своих внешних очертаний-контуров, а также эффективно решать основные проблемы распознавания образов – переноса, поворота, изменение масштаба изображения объекта [4]. При этом под контуром понимается пространственно-протяженный разрыв, перепад или скачкообразное изменение значений яркости.

Следует отметить, что контур полностью определяет форму изображения и содержит всю необходимую информацию для распознавания изображений. Такой подход позволяет не рассматривать внутренние точки изображения, существенно сокращая объем информации, обрабатываемой при анализе изображения, следовательно, дает возможность осуществлять анализ изображений в режиме реального времени.

Существует ряд проблем при выделении контуров изображения:

- разрывы контура в местах, где яркость меняется не слишком быстро;
- наличие ошибочных контуров в результате шума на изображении;
- широкие контурные линии из-за размытости или шума.

В настоящее время используются разные способы контурного анализа [5; 6], и количество новых решений продолжает расти. К наиболее распространенным методам описания контура, которые используются также для решения задач распознавания, относятся цепные коды Фримена (Freeman Chain Code) [7].

Контур объекта подается в дискретном поле (решетке) посредством последовательности отрезков прямых линий определенной длины и направления. В основе этого представления лежит четырех- или восьмисвязная модель (см. Рисунок 7). Длина каждого отрезка определяется разрешением решетки, а направления задаются выбранным кодом. Для построения кода фиксируется начальная точка отсчета, затем осуществляется обход контура в выбранном направлении с последующим описанием последовательности цифр. С помощью относительно простых многочисленных методов можно выполнить масштабирование, повороты или другие преобразования, необходимые для распознавания.

Практически все подходы к выделению контуров изображений объектов можно разделить на две категории [9]:

- методы, основанные на отыскании максимумов;
- методы, основанные на отыскании нулей.

Большинство алгоритмов, по которым осуществляется выделение контуров изображений объектов, базируются на вычислении градиента изображения. Для функции f градиент принимает вид

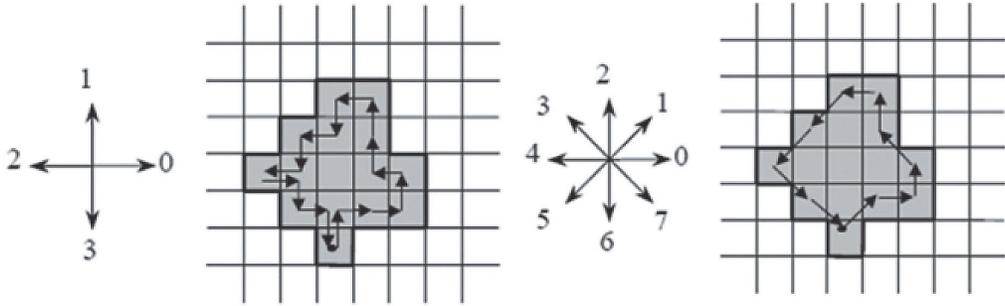
$$\nabla = \begin{bmatrix} \frac{\partial}{\partial x} f \\ \frac{\partial}{\partial y} f \end{bmatrix}. \quad (1)$$

Вычисление градиента изображения состоит в получении частных производных $\partial f / \partial x$ и $\partial f / \partial y$ для каждой точки.

Один из самых ранних алгоритмов выделения контуров изображения объекта принадлежит Лоуренсу Робертсу. Этот алгоритм базируется на дифференцировании амплитуды сигнала, что равнозначно вычислению дискретных разниц амплитуд отсчетов [9]:

$$\frac{\partial f(x, y)}{\partial x} \approx \Delta f_x(x, y) - f(x-1, y), \tag{2}$$

$$-f\left(x, y \frac{\partial f(x, y)}{\partial y} \approx \Delta f_x(x, y) - 1\right), \tag{3}$$



Цепной код 10012112323203036

Цепной код 1023245577

Рисунок 7. Примеры построения цепного кода:
a – четырехсвязная модель; *b* – восьмисвязная модель

Градиент изображения рассчитывается с помощью матриц свертки:

$$G_x = \begin{bmatrix} +1 & 0 \\ 0 & -1 \end{bmatrix} * A, G_y = \begin{bmatrix} 0 & +1 \\ -1 & 0 \end{bmatrix} * A, G = \sqrt{G_x^2 + G_y^2}, \tag{4}$$

где *A* – изображение; * – оператор свертки; *G* – градиент изображения.

Для повышения скорости вычисления градиента с пониженной точностью оператор Робертса упрощается:

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, A' = |a_{11} - a_{22}| + |a_{12} - a_{21}|, \tag{5}$$

где *A'* – обработанное изображение.

Наиболее простым и быстрым методом выделения контуров является перекрестный градиент Робертса

$$R(i, j) = \sqrt{(B_{i,j} - B_{i-1,j-1})^2 + (B_{i-1,j} - B_{i,j-1})^2}, \tag{6}$$

или

$$R_1(i, j) = |B_{i,j} - B_{i-1,j-1}| + |B_{i-1,j} - B_{i,j-1}|, \tag{7}$$

и попарно разностный оператор

$$G_1(i, j) = |B_{i-1,j} - B_{i+1,j-1}| + |B_{i,j-1} - B_{i,j+1}|, \tag{8}$$

где *B_{i,j}* – яркость элемента.

Для повышения эффективности подчеркивания контуров произвольного направления используется объединенный оператор

$$G_2(i, j) = [R_1(i, j) \geq e] V[G_1(i, j) \geq e]. \quad (9)$$

Доктор Джудит Превитт (Judith Prewitt) для обнаружения контуров медицинских изображений применила оператор, маски которого получили ее имя (Prewitt operator). Оператор Prewitt основан на понятии центральной разницы [10]:

$$f(x, y)/x = (f(x + 1, y) - f(x - 1, y))/2; \quad (10)$$

$$f(x, y)/y = (f(x, y + 1) - f(x, y - 1))/2. \quad (11)$$

Оператор вычисляет градиент интенсивности изображения в каждой точке, задавая направление максимально возможного роста от светлого до темного и скорость изменения в этом направлении. Результат показывает, насколько резко или плавно изменяется изображение в этой точке, и, следовательно, насколько вероятно, что часть изображения является краем, а также то, как этот край будет ориентирован.

Градиент изображения вычисляется по матрицам свертки:

$$G_x = \begin{bmatrix} -1 & 0 & +1 \\ -1 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} \cdot A, G_y = \begin{bmatrix} -1 & 0 & -1 \\ 0 & 0 & 0 \\ +1 & 0 & +1 \end{bmatrix} \cdot A, G = \sqrt{G_x^2 + G_y^2}.$$

Как недостаток оператора Prewitt можно отметить его чувствительность к шуму изображения.

Применяемый чаще оператор Собела (Sobel) также основывается на понятии центральной разницы. Однако вес центральных пикселей растет вдвое:

$$G_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} \cdot A, G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix} \cdot A, G = \sqrt{G_x^2 + G_y^2}.$$

Это позволяет снизить влияние шума угловых элементов, что существенно при работе с производными.

Также наравне с градиентными методами широко используются методы, основанные на лапласиане изображения, важной особенностью которого является его инвариантность к вращению. Математическое их описание в настоящей статье не приводится, отметим лишь, что с помощью этого и вышеописанного метода определяются условия, при которых возможно обнаружение изображений объектов. Это алгоритмы обнаружения, учитывающие основные и дополнительные признаки отличия объекта от окружающего фона.

Заключение

Таким образом, обоснован выбор метода и архитектуры сверточных нейросетей для выделения объектов в потоке видео. Прослежена эволюция моделей нейросетей и методов, позволяющих ускорить их работу и точность распознавания за счет SSD.

Выделение контуров изображений объектов является одной из основных операций предварительного анализа изображения в интеллектуальных системах видеонаблюдения.

Настоящая статья посвящена анализу методов распознавания изображений и поиска их в видеопотоке. Проанализирована эволюция структуры сверточных нейронных сетей, используемых в области диагностики компьютерных видеопотоков.

Уделено внимание созданию комбинированной системы, сочетающей в себе технологию искусственного интеллекта и компьютерного зрения на основе нечеткой логики.

Перспективным направлением дальнейших исследований является разработка алгоритмов выделения контуров изображений объектов, реализующих двухмасштабную статистическую модель изображения.

Литература

1. Ainsworth T. (2002) Buyer Beware. *Security Oz*, 2002. № 19, Pp. 18–26.
2. Сальников И.И. Критерии отнесения устройств и систем обработки информации к интеллектуальным. XXI век: результаты прошлого и проблемы настоящего плюса. Пенза: Изд-во Пензенской технологической академии, 2012. С. 11–15.
3. Крючкова Л.П., Кременский М.С. Методы выявления тревожных событий в интеллектуальных системах видеонаблюдения // Современная защита информации. 2019. № 3. С. 64–69.
4. Heikkila M., Pietikainen M. (2006) A Texture-based Method for Detecting Moving Objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, No. 28(4), Pp. 657–662. DOI: 10.5244/C.18.21
5. Сакович И.О., Белов Ю.С. Обзор основных методов контурного анализа для выделения контуров движущихся объектов // Инженерный журнал: наука и инновации. 2014. № 12. URL: <http://engjournal.ru/catalog/it/hidden/1280.html> (дата обращения: 25.06.2023)
6. Сирота А.А., Соломатин А.И. Статистические алгоритмы обнаружения границ объектов на изображениях // Вестник ВГУ. Серия: Системный анализ и информационные технологии. 2008. № 1. С. 58–64.
7. Хачумов М.В. Сжатие, передача и распознавание контуров ригидных объектов, описанных цепными кодами // Современные наукоемкие технологии. 2020. № 8. С. 79–85.
8. Girshick R. (2015) Fast R-CNN. 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, 2015:1440–1448. DOI: 10.1109/ICCV.2015.169
9. Shih F.Y. (2010) Image processing and pattern recognition: fundamentals and techniques. IEEE Press, 2010, 537 p.
10. Гонсалес Р.С., Вудс Р.Э. Цифровая обработка изображений. 3-е изд., испр. и доп. М.: Техносфера, 2012. 1104 с.
11. Ren S., He K., Girshick R., Sun J. (2017) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, No. 39(6), Pp. 1137–1149. DOI: 10.1109/TPAMI.2016.2577031
12. Redmon J., Divvala S., Girshick R., Farhadi A. (2016) You Only Look Once: Unified, Real-Time Object Detection. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, Pp. 779–788. DOI: 10.1109/CVPR.2016.91
13. Liu W. (2016) SSD: Single Shot MultiBox Detector: *Computer Vision – ECCV 2016*. *ECCV 2016. Lecture Notes in Computer Science*, Vol. 9905. Springer, Cham. DOI: 10.1007/978-3-319-46448-0_2
14. Girshick R., Donahue J., Darrell T., Malik J. (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, 2014, Pp. 580–587. DOI: 10.1109/CVPR.2014.81
15. Chen C., Chen Q., Huaqi Q., Giacomo T., Jimming D., Wenjia B., Daniel R. (2020) Deep Learning for Cardiac Image Segmentation: A Review. *Frontiers in Cardiovascular Medicine*, 2020, 7:25. DOI: 10.3389/fcvm.2020.00025
16. Эриксен К. и Хоффман Дж. Временные и пространственные характеристики выборочного кодирования с визуальных дисплеев // Восприятие и психофизика. 1972. № 12 (2В). С. 201–204.

17. Трейсман А., Геладе Г. Теория интеграции функций внимания // Когнитивная психология. 1980. № 12. С. 97–136.
18. Десимоне Р. и Дункан Дж. Нейронные механизмы избирательного зрительного внимания // Ежегодник. Rev. Neurosci. 1995. № 18. С. 193–222.
19. Унгерляйдер А., Мишкин М. Две кортикальные зрительные системы. Анализ визуального поведения. Кембридж, ИнМИТ Пресс, 2010. С. 549–586.

Literature

1. Ainsworth T. (2002) Buyer Beware. *Security Oz*, 2002, No. 19, Pp. 18–26.
2. Salnikov I.I. (2012) *Kriterii otnesenija ustrojstv i sistem obrabotki informacii k intellektual'nyh. XXI vek: rezul'taty proshlogo i problemy nastojashhego pljusa* [Criteria for classifying devices and information processing systems as intelligent. XXI century: results of the past and problems of the present plus. Penza: Penz Publishing House State Technol. Acad., 2012, Pp. 11–15 (in Russian)].
3. Kryuchkova L.P., Kremensky M.S. (2019) [Methods for detecting alarm events in intelligent video surveillance systems]. *Modern information security*, 2019, No. 3, Pp. 64–69 (in Russian).
4. Heikkila M., Pietikainen M. (2006) A Texture-based Method for Detecting Moving Objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, No. 28(4), Pp. 657–662. DOI: 10.5244/C.18.21
5. Sakovich I.O., Belov Yu.S. (2014) [Overview of the main methods of contour analysis for the detection of the contours of moving objects]. *Engineering Journal: Science and Innovations*, 2014, No. 12 (in Russian). Available at: <http://engjournal.ru/catalog/it/hidden/1280.html> (accessed: 06.25.2023).
6. Sirota A.A., Solomatin A.I. (2008) [Statistical algorithms for detecting the boundaries of objects in images]. *Bulletin of VSU. Seriya: System analysis and information technologies*, 2008, No. 1, Pp. 58–64 (in Russian).
7. Khachumov M.V. (2020) [Compression, transmission and recognition of the contours of rigid objects described by chain codes]. *Modern high technologies*, 2020, No. 8, Pp. 79–85 (in Russian).
8. Girshick R. (2015) Fast R-CNN. 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, 2015:1440–1448. DOI: 10.1109/ICCV.2015.169
9. Shih F.Y. (2010) Image processing and pattern recognition: fundamentals and techniques. IEEE Press, 2010, 537 p.
10. Gonzalez R.S., Woods R.E. (2012) *Cifrovaja obrabotka izobrazhenij* [Digital image processing. 3rd ed., rev. and additional]. Moscow: Technosfera Publishing, 2012, 1104 p. (in Russian).
11. Ren S., He K., Girshick R., Sun J. (2017) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, No. 39(6), Pp. 1137–1149. DOI: 10.1109/TPAMI.2016.2577031
12. Redmon J., Divvala S., Girshick R., Farhadi A. (2016) You Only Look Once: Unified, Real-Time Object Detection. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, Pp. 779–788. DOI: 10.1109/CVPR.2016.91
13. Liu W. (2016) SSD: Single Shot MultiBox Detector: *Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science*, Vol. 9905. Springer, Cham. DOI: 10.1007/978-3-319-46448-0_2
14. Girshick R., Donahue J., Darrell T., Malik J. (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, 2014, Pp. 580–587. DOI: 10.1109/CVPR.2014.81
15. Chen C., Chen Q., Huaqi Q., Giacomo T., Jinming D., Wenjia B., Daniel R. (2020) Deep Learning for Cardiac Image Segmentation: A Review. *Frontiers in Cardiovascular Medicine*, 2020, 7:25. DOI: 10.3389/fcvm.2020.00025

16. Eriksen K. & Hoffman J. (1972) [Temporal and spatial characteristics of selective encoding from visual displays]. *Vosprijatie i psihofizika*, 1972, No. 12, Pp. 201–204 (in Russian).
17. Treisman A. and Gelade G. (1980) [The theory of integration of attention functions]. *Kognitivnaja psihologija*, 1980, No. 12, Pp. 97–136 (in Russian).
18. Desimone R. and Duncan J. (1995) [Neural mechanisms of selective visual attention]. *Ezhegodnik. Rev. Neurosci*, 1995, No. 18, Pp. 193–222 (in Russian).
19. Ungerleider L., Mishkin M. (2010) [Two cortical visual systems. Visual Behavior Analysis]. Cambridge, In MIT Press, 2010, Pp. 549–586 (in Russian).