

ВЫДЕЛЕНИЕ ЗОНЫ ДЕТЕКЦИИ С ИСПОЛЬЗОВАНИЕМ МОДУЛЯ ESP32-CAM

Русакович Е.С., магистрант

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Ярмолик В.Н. – д-р техн. наук, профессор

Описывается подход к определению зоны детекции в видеопотоке, получаемом с модуля ESP32-CAM. Сравниваются алгоритмы пороговой фильтрации, иллюстрируется типовая проблема алгоритмов удаления фона.

Компьютерное зрение активно развивается с середины 20-го века и успешно применяется для решения задач распознавания образов, отслеживания перемещений, воссоздания цифровых моделей объектов, детекции аномалий и др. Современное развитие встроенных систем позволяет применять различные комплексы мониторинга, включающие модули компьютерного зрения, в труднодоступных местах, сохраняя при этом небольшие габариты и высокую энергоэффективность устройства. Встроенное компьютерное зрение может быть использовано для контроля технологических параметров, детекции сбоев и нарушений работы оборудования, контроля движения в специальных зонах и соблюдения техники безопасности.

При разработке таких программно-аппаратных комплексов необходимо иметь механизм классификации событий по кадрам входного видеопотока. Значительного улучшения производительности системы можно добиться, выполняя классификацию не для каждого кадра видеопотока, а только для зон детекции (ROI) [1], определение которых сводится к вычислению маски попавших в кадр объектов, для чего применяется алгоритм удаления фона с последующей фильтрацией шума.

Одним из самых эффективных с точки зрения потребления ресурсов и метрик (точность, полнота, F-мера и др.) является вычитание фона с адаптивным обновлением фоновой модели [2]. Этот метод формирует фоновую модель на основе начальных кадров. Затем для каждого пикселя фоновой модели и пикселей последующих кадров вычисляется разность по модулю, формируя маску зафиксированных объектов. Отличительной особенностью алгоритма является то, что приходящие кадры, умноженные на константный коэффициент (скорость обучения), прибавляются к фоновой модели, тем самым адаптируя её к постепенным изменениям окружающей среды, таким как изменение освещения, формирование теней и движение фона. Причём суммируется только та часть кадра, которая не попадает в итоговую зону детекции.

Получившаяся маска объектов содержит шумовые значения, вызванные в том числе и физическим устройством сенсора камеры. Для их удаления широко применяются алгоритмы пороговой обработки: пиксели меньше или равные выбранному значению устанавливаются в 0. При установке глобального порога для всего кадра фильтрация может давать некорректные результаты при нестабильном освещении, тенях объектов и других усложняющих факторах. Для решения этой проблемы может применяться алгоритм адаптивной пороговой фильтрации, который устанавливает порог для каждого пикселя в зависимости от значений в его локальной окрестности.

Непосредственно пороговое значение может задаваться эмпирически. Однако более стабильные результаты получаются при автоматическом выборе порога на основании пикселей кадра, например, в алгоритме Отсу [3]. Он основывается на статистических свойствах маски объектов, предполагая, что гистограмма для получившихся пикселей стремится к бимодальной, что позволяет разделить все пиксели изображения на два класса: фоновый и объекта. Пороговым выбирается значение пикселя, которому будет соответствовать максимальная межклассовая дисперсия. Межклассовая дисперсия вычисляется по формуле:

$$\sigma_b^2(t) = w_0(t) \cdot w_1(t) \cdot (\mu_0(t) - \mu_1(t))^2, \quad (1)$$

где w_0 и μ_0 – вероятность и математическое ожидание пикселей класса фона; w_1 и μ_1 – вероятность и математическое ожидание пикселей класса объекта; t – пороговое значение, разделяющее классы.

Существуют различные варианты алгоритма, более устойчивые к мультимодальным гистограммам и зашумлённым кадрам. Но все они значительно повышают вычислительную сложность, учитывая, что классическая реализация линейно зависит от количества пикселей и их уникальных значений. В статье [4] приводится способ сокращения асимптотики алгоритма с линейной до логарифмической. Метод основывается на вычислении межклассовых дисперсий не для всей гистограммы значений, а только для значений, получаемых применением численного метода бисекции. Выбор следующего отрезка определяется тем, в каком направлении возрастает дисперсия. Алгоритм предполагает, что распределение межклассовой дисперсии унимодально для изображений масок,

полученных после вычитания фона. Проведены экспериментальные исследования с целью сравнения качества классификации алгоритмов. Тестовый объект и фоновая модель представлены на рисунке 1, результаты фильтрации фоновых пикселей – на рисунке 2.

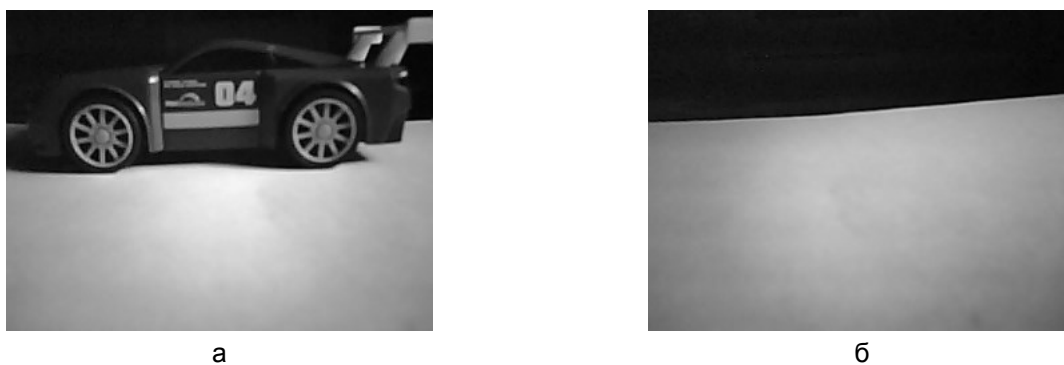


Рисунок 1 – Объект в зоне детекции (а) и фоновая модель (б)



Рисунок 2 – Пороговая фильтрация кадра классическим алгоритмом (а) и с бисекцией (б)

Пороговое значение пикселей для маски, полученной на рисунке 2а, составило 60, в то время как для маски, полученной методом Оtsу с бисекцией, – 143. Как следствие, зона детекции, выбранная после пороговой фильтрации классическим методом, точнее захватывает появившийся в кадре объект, поэтому классический метод предпочтительнее для последующей классификации.

На рисунках 2а и 2б проявляется типовая проблема алгоритмов вычитания фона: зафиксированный объект отбрасывает тень, которая включается в результирующую маску. Согласно исследованию [2] наиболее близкая к реальной маска объекта может быть получена при помощи алгоритма вычитания фона, учитывающего цвет и текстуру пикселей [5], однако объём необходимой для работы оперативной памяти делают его непригодным для реализации в рамках встроенной системы на базе ESP32. В плате предусмотрено 520 КБ встроенной оперативной памяти с возможностью адресации дополнительных 8 МБ псевдостатической оперативной памяти, встроенной в общее виртуальное адресное пространство и значительно уступающей в скорости обращения [6].

В результате проведенных экспериментов рассмотрены различные варианты получения маски объектов непосредственно на модуле ESP32-CAM. Проведено сравнение классического алгоритма Оtsу и его модификации на основе метода бисекции, показано, что несмотря на заявленное в статье [4] уменьшение вычислительной сложности алгоритма, снижается точность классификации пикселей.

Список использованных источников:

1. Noman, M. An optimized and fast scheme for real-time human detection using Raspberry Pi / M. Noman, M. H. Yousaf, S. A. Velastin // 2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA): proc., Gold Coast, Australia, 2015 / IEEE. – [Gold Coast], 2016. – P. 1-7.
2. Sobral, A. A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos / A. Sobral, A. Vacavant // Computer Vision and Image Understanding. – 2014. – Vol. 122. – P. 4-21.
3. Otsu, N. A Threshold Selection Method from Gray-Level Histograms / N. Otsu // IEEE Transactions on Systems, Man, and Cybernetics. – 1979. – Vol. 9, No. 1. – P. 62-66.
4. Kodathala, V. Fast OTSU Thresholding Using Bisection Method [Электронный ресурс]. – Режим доступа: <https://arxiv.org/abs/2509.16179v1> – Дата доступа: 23.03.2026.
5. Yao, J. Multi-Layer Background Subtraction Based on Color and Texture / J. Yao, J. -M. Odobez // 2007 IEEE Conference on Computer Vision and Pattern Recognition: proc., Minneapolis, USA, 17-22 June 2007 / IEEE. – [Minneapolis], 2007. – P. 1-8.
6. ESP32 Technical Reference Manual [Электронный ресурс]. – Режим доступа: https://documentation.espressif.com/esp32_technical_reference_manual_en.pdf – Дата доступа: 23.03.2026.