

МОДЕЛЬ НАКОПЛЕНИЯ СТАТИЧЕСКОГО ФОНА С УЧЕТОМ НЕСТАБИЛЬНОСТИ СЦЕНЫ

Тарасов Кирилл Викторович

студент, Московский институт электронной техники, РФ, г. Зеленоград

Воротников Владимир Сергеевич

научный руководитель, канд. техн. наук, доцент, Московский институт электронной техники, РФ, г. Зеленоград

Модель фона

Модель фона (накопление фона) – это метод отделения элементов переднего плана от фона, который выполняется путем создания маски переднего плана. Данный метод используется для обнаружения динамически движущихся объектов со статических камер и активно используется в решении задач, связанных с охраной периметра [1].

В данной статье мы, основываясь на существующих подходах, таких как «Real-time detection of abandoned bags using CNN» [1] и «Robust Abandoned Object Detection Using Dual Foregrounds» [2], реализуем собственную модель фона, устойчивую к шумовым движениям и резким изменениям сцены. А в конце проведем сравнительный анализ полученной модели с моделями, чей подход был взят за основу, а также моделью, основанной на нейронных сетях.

Существующие подходы

Для построения модели фона нередко используется концепция скользящего среднего, в которой видеопоследовательность анализируется по определенному набору кадров. В течение этой последовательности кадров вычисляется скользящее среднее по текущему и предыдущим кадрам, в итоге получая фоновую модель, где любой новый объект, появившийся в текущей последовательности кадров, становится частью переднего плана.

Таким образом, каждый новый кадр содержит новый объект с фоном, на котором вычисляется абсолютная разница между фоновой моделью (которая является функцией времени) и самим кадром (1):

$$dst(x, y) = (1 - \alpha) \cdot dst(x, y) + \alpha \cdot src(x, y)$$

Так, dst – накопительное изображение с тем же количеством каналов, что и входное изображение; x ,

у – координаты изображения; src – новый кадр, α – вес входного изображения, который регулирует скорость обновления фона (насколько быстро накопитель фона «забывает» о более ранних изображениях) [1].

Для того, чтобы отделить временно статичные пиксели от фона сцены, одного фонового изображения не достаточно, поэтому используется модель с двумя фонами для получения как долговременного фона, так и кратковременного фонов (рис. 1). Также, исходя из сцены, для каждого накопительного изображения эмпирическим путем подбираются α : в данном случае 0.02 и 0.1 для долговременного и кратковременного, соответственно.

По кратковременному фону строится маска движения – бинаризованное по порогу накопительное изображение (рис. 2).

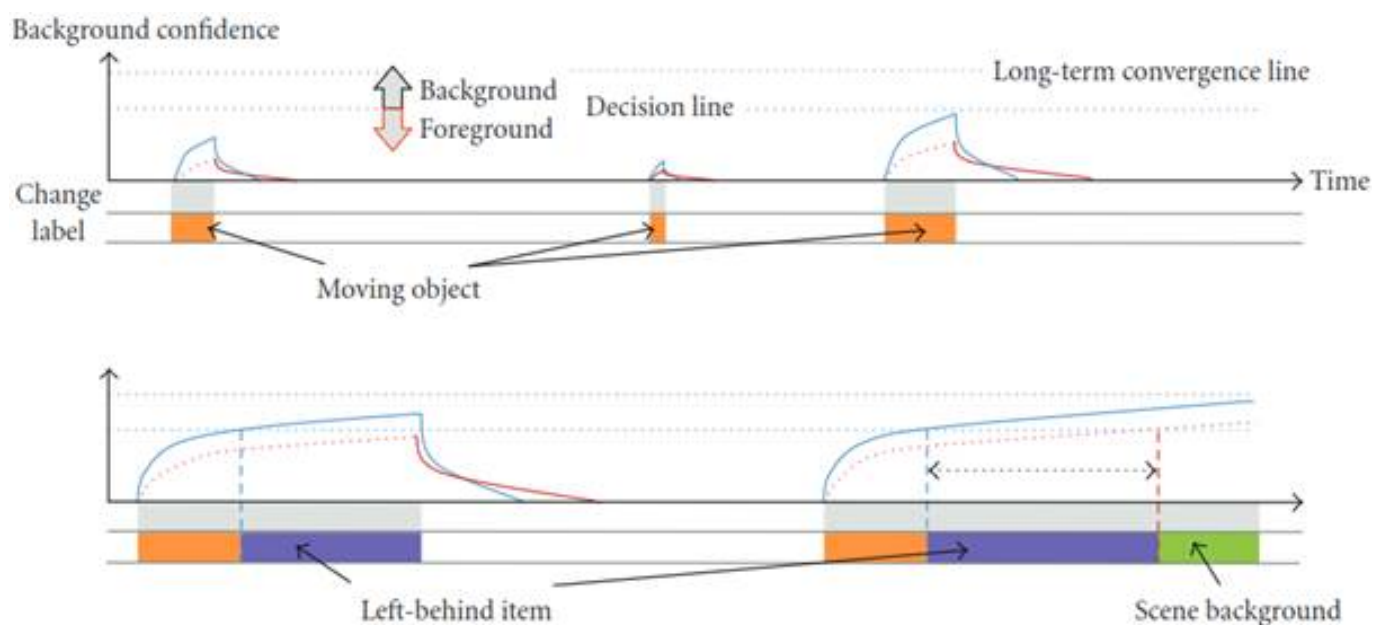


Рисунок 1. Достоверность моделей долгосрочного и краткосрочного фонов

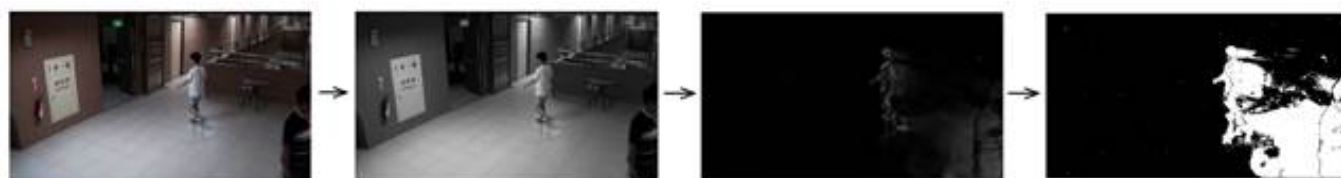


Рисунок 2. Построение маски движения и ее бинаризация

Вместе с маской движения находится долговременный фон с последующей бинаризацией (рис. 3).



Рисунок 3. Построение долговременного фона и его бинаризация

Затем полученная маска движения вычитается из долговременного фона, в итоге получая устойчивый статический фон (рис. 4) [2].



Рисунок 4. Устойчивый статический фон, полученный в результате накопления двух фонов

Альтернативное решение

Исходя из рис. 4, есть основания полагать, что данная модель не достаточно помехоустойчива. В связи с этим, попробуем реализовать собственную модель, а затем проведем сравнительный анализ, для оценки используя MSE (Mean Squared Error – среднеквадратичная ошибка) и среднее время обработки кадра.

Чтобы сделать систему более помехоустойчивой, в отличие от моделей с двумя фонами, используем три фона, где первые два будут служить для построения долговременного фона, а третий – кратковременного. Такой подход обеспечит дополнительную фильтрацию шумового движения и резких изменений сцены.

Как и в предыдущем методе, по кратковременному фону строится маска движения с α , равным 0.1, и долговременный фон с α , равным 0.02. Вместе с этим строится еще один долговременный фон с α , равным 0.016 (рис. 5). Тогда разница первых двух накопительных изображений, с учетом исключения из накопления пикселей, содержащихся в маске движения, будет являться устойчивым статическим фоном (рис. 6).



Рисунок 5. Построение второго долговременного фона и его бинаризация

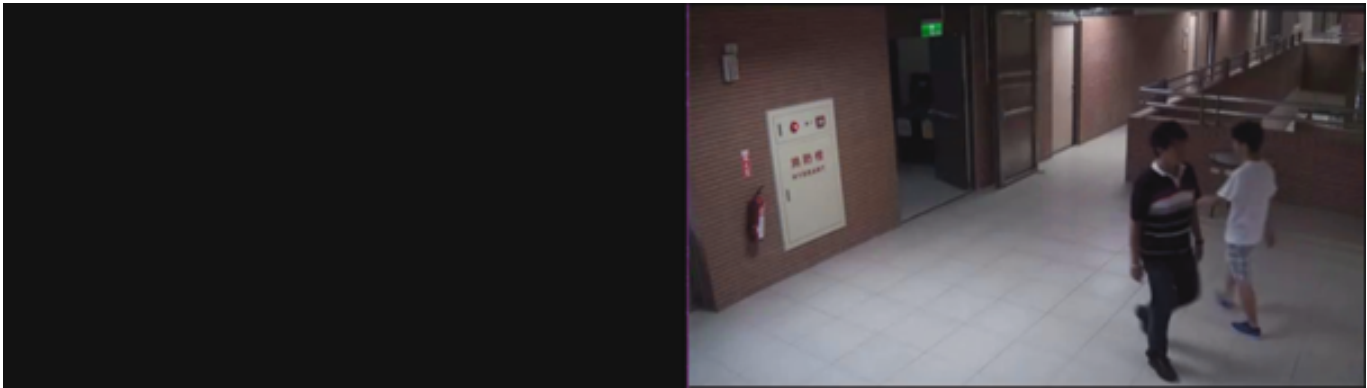


Рисунок 6. Устойчивый статический фон, полученный в результате накопления трех фонов

Теперь перейдем к сравнению. Оно проводилось на CPU на тестовом видео с разрешением 480×640 и средним FPS – 30 к/с. Используемая операционная система: Manjaro Linux, GNOME 43.1. CPU: Intel® Core™ i5-8300H × 8:

Таблица 1.

Сравнение моделей с двумя и тремя фонами

Модель фона	Среднее MSE	Среднее время к
С 2 накопительными изображениями	8.2e-3	
С 3 накопительными изображениями	1.6e-5	

Как видно из табл.1, среднее время обработки кадра имеет один порядок, чего нельзя сказать об MSE – тут данные разнятся на два порядка. А значит, модель с двумя фонами уступает модели с тремя в точности, при этом почти не уступая в скорости обработки кадров.

Также стоит сказать и про модели, принцип работы которых основан на нейронных сетях. Так, модель «A hybrid framework combining background subtraction and deep neural networks for rapid person detection» [3], используя U-Net [4] для сегментации, классифицирует объекты в кадре и отделяет фон от переднего плана.

Однако, решения такого рода имеют существенный недостаток – тяжеловесность. Продемонстрируем это, проведя сравнительный анализ полученной нами модели и модели, озвученной выше, с теми же входными параметрами, что и при сравнении моделей с разным количеством фонов:

Таблица 2.

Сравнение моделей с использованием нейронной сети и без

Модель фона	Среднее MSE	Среднее время к
С использованием нейронной сети	1.4e-5	
Без использования нейронной сети	1.6e-5	

Как видно из табл.2, теперь мы имеем обратную ситуацию: среднее MSE имеет один порядок, а среднее время обработки кадра разнится на два порядка. А значит, модель, использующая нейронную сеть, хоть и немного опережает в точности, сильно уступает в скорости обработки кадров.

Выводы

Таким образом, была получена модель с лучшей помехоустойчивостью к шумовому движению и незначительным нестабильностям сцены, не уступающая по быстродействию более легковесным моделям, и более тяжеловесным по точности.

Возможные улучшения: блочный адаптивный подбор параметров α .

Список литературы:

1. S.V.Sidyakin, B.V.Vishnyakov. Real-time detection of abandoned bags using CNN // The Federal State Unitary Enterprise "State Research Institute of Aviation System". 2017.

2. Fatih Porikli, Yuri Ivanov, and Tetsuji Haga. Robust Abandoned Object Detection Using Dual

Foregrounds // EURASIP Journal on Advances in Signal Processing. 2017.

3. Chulyeon Kim, Jiyoung Lee ,Taekjin Han and Young-Min Kim. A hybrid framework combining background subtraction and deep neural networks for rapid person detection // Journal of Big Data. 2018.

4. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation // University of Freiburg, Germany. 2015.