

Выделение объектов в видеопотоке на основе разрезов графов

Антон Конушин, Никита Четвериков
Лаборатория Компьютерной Графики и Мультимедиа
Московский Государственный Университет, Москва, Россия
{ktosh,nchetverikov}@graphics.cs.msu.ru

Аннотация

Алгоритмы поиска объектов переднего плана широко используются в видеонаблюдении. Однако большинство существующих алгоритмов работает либо быстро, но неточно, либо точно, но медленно. В данной статье предлагается новый алгоритм выделения объектов в видеопотоке, основанный на «прицеливании», т.е. предварительном выборе областей интереса [5]. К выделенным областям интереса затем применяется качественный, но медленный алгоритм на основе минимизации функционала энергии с помощью разрезов графов [1]. Такой подход позволяет достичь большей скорости по сравнению с применением алгоритма разрезов графов ко всему кадру, но сохранить точность определения границ объекта интереса.

Ключевые слова: поиск объектов в видео, трекинг, разрезы графов.

1. ВВЕДЕНИЕ

В статье рассматривается задача выделения объектов переднего плана в видеопоследовательности, полученной со статичной камеры. Основным применением алгоритмов выделения переднего плана являются системы видеонаблюдения со стационарными камерами. В таких системах наблюдаемая сцена мало меняется от кадра к кадру в видеопотоке, поэтому называется «фоном». Все новые объекты, попадающие в области видимости, отличаются от фона и называются «передним планом». Формально задача выделения объектов переднего плана в видеопотоке ставится как определение для каждого пикселя каждого кадра, принадлежит ли он фону, либо объекту на переднем плане (см. рис. 1).

Для последующего анализа выделенных объектов необходимо как можно точнее определить их границы. Если в выделенные области объектов попадает фон, то анализ таких объектов, например, поиск похожих в видеоархиве, затруднён. Другой проблемой является разделение одного наблюдаемого объекта на несколько несвязанных областей. Наиболее высокую точность выделения объектов показывают методы на основе минимизации функционала энергии специально вида, заданного на всём изображении. Однако даже вычислительно эффективные методы минимизации энергии, подходящие для решения только для случая бинарных переменных, такие как методы на основе разрезов графов, оказываются недостаточно быстрыми для обработки видеопотока высокого разрешения в реальном времени.

В данной работе предлагается применить идею «прицеливания» [5] для выбора областей интереса, и применять методы на основе минимизации функционала энергии только к ним. За счёт этого удастся обеспечить высокую точность границ объектов при высокой скорости работы алгоритма.



Рис. 1: Задача выделения объектов в видеопоследовательностях. Сверху: исходный кадр. Снизу: обработанный кадр (пиксели, отнесённые к фону, затемнены, к объекту переднего плана – «подсвечены»).

2. СУЩЕСТВУЮЩИЕ МЕТОДЫ

Основную часть существующих методов выделения объектов переднего плана в видеопотоке можно отнести к трём основным категориям:

- Попиксельные методы [2] [3] [9];
- Поблочные методы [11], [13];
- Методы, основанные на минимизации функционала энергии по всему изображению [7].

Попиксельные методы обрабатывают все точки изображения независимо. Обычно вначале они строят цветовую модель фона и во время работы оценивают, насколько текущий цвет пикселя ей соответствует. В качестве модели фона может использоваться, к примеру, нормальное распределение либо смесь из статически [2] либо динамически [9] заданного числа нормальных распределений. К этой же группе можно отнести методы, осуществляющие попиксельную классификацию, но использующие при этом статистику распределения цветов по окрестности пикселя [3]. Подобные методы обычно обладают достаточно высокой скоростью работы, однако

чувствительны к шуму и не учитывают связанность разметки (среди большинства пар соседних пикселей либо оба принадлежат фону, либо переднему плану).

Методы из второй категории обрабатывают независимо не отдельные пиксели, а группы пикселей, объединенные в блоки. Несмотря на то, что использование таких методов позволило в некоторых случаях достичь высокой скорости работы [11], подобный подход даже теоретически не позволяет добиться идеальной точности, т.к. к фону либо переднему плану относится сразу весь блок пикселей. Если блок содержит границу, то в таких областях граница объекта определяется заведомо неточно.

Методы на основе минимизации энергии функционала специального вида позволяют использовать информацию со всего изображения в совокупности, включая информацию о градиенте яркости между соседними пикселями. Такой подход позволяет учесть, например, тот факт, что граница между объекта чаще разделяет пиксели, сильно отличающиеся по цветам, чем похожие по цвету. За счёт учета пикселей в совокупности методы оказываются менее чувствительными к шуму в изображении, чем попиксельные методы, и дают большую точность, чем поблочные методы. Недостатком методов на основе функционала энергии является низкая скорость работы, определяемая низкой скоростью методов минимизации энергии.

Одним из наиболее известных методов, использующих для поиска объектов в видеопотоке минимизацию энергии по всем пикселям изображения, является метод BackgroundCut [7]. Он учитывает цвет пикселя и его контраст с «соседями», а также принимает во внимание возможные изменения фона: изменение освещения, движение в фоне, появление новых объектов на заднем плане и дрожание камеры. Однако скорость его работы составляет 12-15 кадров в секунду для разрешения 320x240 (при обработке на процессоре 3.2 GHz) и является недостаточной для многих задач видеонаблюдения.

Отдельно стоит отметить различные методы ускорения и улучшения качества работы алгоритмов. Одним из распространённых методов увеличения скорости является использование многомасштабных представлений [10], когда обрабатывается не само исходное изображение, а его уменьшенные в несколько раз копии. Ряд методов учитывает специфику предметной области для добавления в используемую модель дополнительных условий. Например, при решении задачи отслеживания лабораторной мыши может быть учтена форма тела животного [12]. Другим примером является использование информации о наиболее вероятном положении камеры и типичном виде отслеживаемых объектов: если камера находится на высоте 1-2 метра над землёй и направлена параллельно ей, то подходящие к ней спереди люди занимают, как правило, несколько непрерывных соседних столбцов точек на кадре [4].

3. ПРЕДЛОЖЕННЫЙ АЛГОРИТМ

Основная идея предлагаемого алгоритма заключается в том, чтобы применять метод выделения объектов на основе минимизации функционала энергии не ко всему изображению в целом, а только к некоторым областям. Для этого вначале необходимо выбрать области интереса, которые потенциально содержат объекты переднего плана, т.е. «прицелить» алгоритм. Как и другие алгоритмы выделения объектов на основе вычитания фона, предлагаемый алгоритм состоит из двух этапов: построения модели фона и выделения объектов переднего плана.

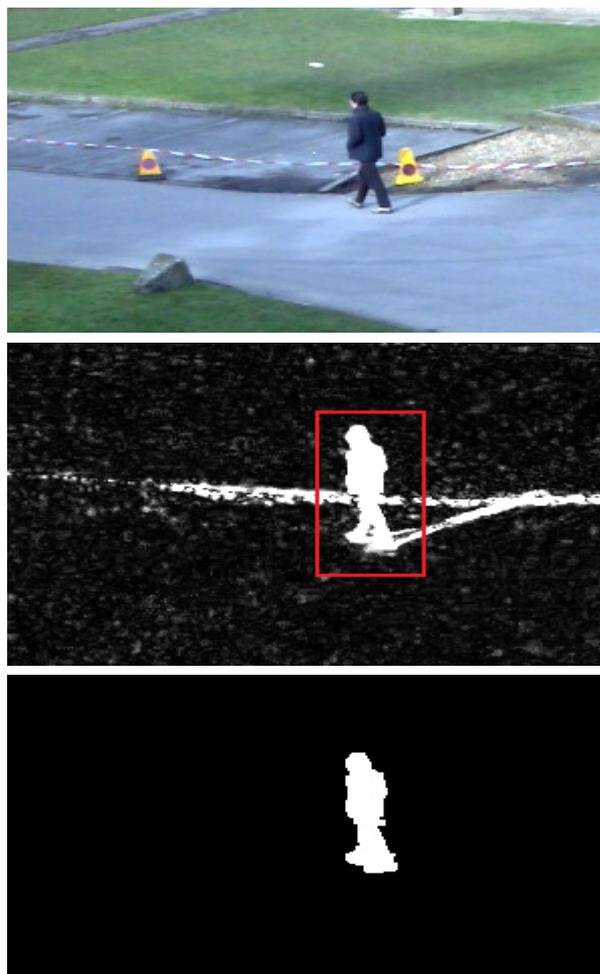


Рис. 2: Основные шаги алгоритма. Сверху вниз: исходный кадр, карта вероятностей с выбранной областью интереса, результат.

3.1 Модель фона

Здесь и далее предполагается, что цвет каждой точки изображения задаётся в формате RGB, то есть вектором из трёх значений (по одному для каждой составляющей цвета).

Предлагаемый алгоритм использует достаточно простую модель фона. Предполагается, что цвет c_{xy} в каждой отдельно взятой точке фона (x, y) определяется нормально распределённой случайной величиной $C_{xy} \sim N(\mu_{xy}, \Sigma)$. Отметим, что векторы математических ожиданий задаются в каждой точке отдельно, в то время как матрица ковариации считается одинаковой для всех точек. Такой выбор обусловлен предположением о том, что цвет каждой точки фона достаточно стабилен, и его изменения на разных кадрах вызваны в основном шумом, одинаковым для всех точек изображения.

3.2 Построение модели фона

На этапе обучения алгоритм получает на вход неупорядоченный набор кадров, содержащих изображение фона без каких-либо посторонних объектов. Допускается различное освещение сцены на разных кадрах, а также незначительные изменения фона (например, колебания лёгких предметов из-за ветра).

По полученным данным производится расчёт средних значений цвета в каждой точке, а также матрицы ковариации, единой для всех точек. Кроме того, для каждого пикселя сохраняется последовательность значений

цветового вектора, которая впоследствии используется при динамическом обновлении модели фона.

3.3 Поиск объектов в видеопотоке

На данном этапе алгоритм последовательно получает и обрабатывает кадры видеопотока, которые могут содержать либо не содержать объекты переднего плана. Работа с кадром состоит из следующих основных шагов (рис. 2):

1. Для кадра по модели фона рассчитывается карта вероятностей, то есть матрица, размеры которой совпадают с разрешением кадра, и значение каждого элемента равно вероятности того, что соответствующая точка кадра принадлежит фону;
2. По полученной карте производится поиск областей, содержащих достаточно большое число точек, не принадлежащих фону, т.е. проводится «прицеливание».
3. Задача поиска объекта в каждой из найденных областей сводится к задаче минимизации энергии определённого функционала, которая, в свою очередь, сводится к поиску минимального разреза некоторого графа;
4. Производится обновление модели фона с использованием информации о том, какие части кадра были отнесены к переднему плану, а какие – к фону.

Так как шаг 1 достаточно очевиден, рассмотрим более подробно лишь шаги 2-4.



Рис 3: Регион с объектом, найденный алгоритмом «прицеливания» (слева), и объект после уточнения границ (справа).

3.3.1 Поиск областей интереса

Для быстрого поиска регионов изображения, содержащих объекты переднего плана, используется метод «скользящего окна». В ходе его работы по карте вероятностей P с некоторым шагом перемещается окно переменного размера, и для каждого его положения и размера вычисляется функция

$$f(x, y, w, h) = \frac{\sum_{i=x}^{x+w} \sum_{j=y}^{y+h} P_{ij}}{w * h}$$

где

- (x, y) – координаты верхнего левого угла окна;
- w и h – ширина и высота окна.

Если значение f для какого-либо положения и размера окна ниже заданного порога, это окно считается содержащим объект переднего плана. Если два или более таких окон пересекаются либо расстояние между ними меньше заданного порога, для дальнейшей работы выбирается ограничивающий прямоугольник их совокупности.

Для ускорения работы этой части алгоритма используется представление карты вероятностей в виде интегрального изображения (такой подход применяется, например, в алгоритме Виолы-Джонса [8]).

3.3.2 Уточнение границ объектов

Для уточнения границ объектов, найденных на предыдущем этапе, используются метод минимизации функционала энергии специального вида.

Каждому пикселю изображения с координатами (i, j) сопоставляется случайная переменная x_{ij} . Кроме того, вводится множество меток $L = \{0, 1\}$; метка 0 отождествляется с фоном, 1 – с объектом переднего плана. Задача заключается в поиске разметки, минимизирующей энергию вида

$$E(x) = \sum_{i=0}^W \sum_{j=0}^H \varphi_{ij}(x_{ij}) + \sum_{i=0}^W \sum_{j=0}^H \sum_{(i', j') \in N(i, j)} \psi_{ij i' j'}(x_{ij}, x_{i' j'})$$

где

- W, H - ширина и высота изображения;
- $N(i, j)$ - множество таких координат (i', j') , что точки (i, j) и (i', j') являются соседними;
- $\varphi_{ij}(x_{ij})$ – унарные потенциалы, отвечающие за «стоимость» назначения переменной определённой метки;
- $\psi_{ij i' j'}(x_{ij}, x_{i' j'})$ – парные потенциалы, соответствующие тому факту, что у соседних пикселей, скорее всего, одинаковые или похожие метки.

В предлагаемом алгоритме унарные потенциалы получаются из соответствующих значений карты вероятности вычитанием последних из 1 (так как каждая точка обязательно принадлежит либо объекту, либо фону, то сумма соответствующих вероятностей равна 1).

Для задания парных потенциалов используется модель Поттса:

$$\psi_{ij i' j'} = \begin{cases} 0 & \text{если } x_{ij} = x_{i' j'} \\ w * \text{abs}(p_{ij} - p_{i' j'}) & \text{если } x_{ij} \neq x_{i' j'} \end{cases}$$

где w – параметр.

Для поиска минимума энергии используется алгоритм Бойкова-Колмогорова на основе поиска разреза минимальной стоимости [12].

На рис. 3 показан пример результата работы алгоритма уточнения границ объекта.

3.3.3 Обновление модели фона

После окончательного определения границ всех найденных объектов производится обновление модели фона. При этом для каждого пикселя, отнесённого к фону, удаляется самое «старое» значение из истории его значений, в конец её добавляется значение из текущего кадра, после чего производится перерасчёт среднего значения цвета в данной точке. Такой подход позволяет динамически адаптировать модель к достаточно плавным изменениям фона, например, изменению уровня освещения.

4. ЭКСПЕРИМЕНТЫ И СРАВНЕНИЕ

Предложенный алгоритм был протестирован на нескольких видеопоследовательностях, взятых из базы PETS 2010 [6], содержащей набор городских сцен с пешеходами. Для проведения численной оценки одна из видеопоследовательностей была попиксельно размечена вручную для оценки точности работы алгоритма.

В ходе тестирования производилось сравнение описанного алгоритма с каждым из базовых для него методов: отдельно взятый алгоритмом поиска области интереса (найденным объектом считались точки (x, y) из области интереса, для

которых $P_{xy} < \frac{1}{2}$) и алгоритмом, использующим минимизацию функционала энергии по всему изображению. Оценивалась скорость работы, а также точность и полнота нахождения объектов переднего плана. Результаты для одной из последовательностей можно увидеть в таблице 1.

	Точность	Полнота	F-оценка	Время (мсек/кадр)
Предложенный алгоритм	0.488228	0.835179	0.965023	112
Марковские поля для всего изображения	0.275857	0.838707	0.863122	281
Алгоритм поиска области интереса	0.329894	0.77707	0.914774	110

Таблица. 1: Результаты тестирования

Нужно сразу отметить, что основной целью тестирования было не получение как можно более высокой точности результатов или скорости работы, а сравнение предложенного алгоритма с каждым из двух алгоритмов, лежащих в его основе. Как следствие, результаты могут быть улучшены как по качеству (более тонкой настройкой параметров алгоритмов или использованием другой модели фона), так и по скорости (например, распараллеливанием отдельных частей алгоритма).

Основным достижением предложенного алгоритма, как можно увидеть из результатов тестирования, является скорость работы, эквивалентная скорости работы алгоритма поиска областей интереса, при качестве результатов, достигающем и даже превосходящем качество результатов метода, использующего случайные Марковские поля.

Отдельно стоит пояснить относительно невысокую точность результатов. Одна из причин заключается в том, что алгоритм часто не отделяет движущегося человека от его тени и отмечает их как один объект. Так как тень занимает на изображении площадь, сравнимую с площадью самого человека, то точность существенно падает. Другой распространённой ошибкой является попадание в область интереса посторонних предметов, например ленты, колеблющейся на ветру (см. рис. 4). Этой же причиной обусловлена ещё более низкая точность работы алгоритма, использующего минимизацию энергии для всего изображения: так как область интереса фактически является всё изображение, то к переднему плану относятся больше посторонних объектов.

Алгоритм был реализован на C++; тестирование производилось на компьютере с процессором Core 2 Duo 2.0 GHz. Разрешение всех кадров, использованных для тестирования – 768x576. Построение модели фона по 50 кадрам занимает около 3 секунд.

5. ЗАКЛЮЧЕНИЕ

В статье предложен новый алгоритм выделения объектов переднего плана, основанный на вычислении попиксельной карты вероятности принадлежности пиксела фону, выборе областей интереса с помощью скользящего окна по этой карте, и применению метода на основе минимизации функционала энергии специального вида к выбранным областям.



Рис. 4: Примеры ошибок алгоритма: тень, отнесённая к переднему плану (слева); часть ленты, попавшая в область интереса и также отнесённая к найденному объекту.

6. БЛАГОДАРНОСТИ

Работа выполнена при поддержке гранта Президента Российской Федерации для молодых ученых - кандидатов наук МК-4644.2012.9 и проекта МРП-2010-050 с компанией Microsoft Research.

7. ССЫЛКИ

- [1] Y. Boykov, V. Kolmogorov, "An Experimental Comparison of Min-Cut/Max-Flow Algorithms for EnergyBK Minimization in Vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004.
- [2] P. KaewTraKulPong, R. Bowden, "An Improved Adaptive Background Mixture Model for Real-time Tracking with Shadow Detection," *Proc. 2nd European Workshop on Advanced Video Based Surveillance Systems*, 2001.
- [3] T. Ko, S. Soatto, D. Estrin, "Background Subtraction on Distributions," *ECCV '08 Proceedings of the 10th European Conference on Computer Vision: Part III*, 2008, pp. 276-289..
- [4] V. Konushin, A. Konushin, "Improvement of background subtraction by mask constraints," *Proc. GraphiCon*, 2010, pp. 96-99.
- [5] I. Laptev, P. Pérez, "Retrieving actions in movies," in *Proc. ICCV'07*, 2007..
- [6] PETS 2010. <http://www.cvg.rdg.ac.uk/PETS2010/a.html#s0>.
- [7] J. Sun, W. Zhang, X. Tang, H.-Y. Shum, "Background Cut," *ECCV Proceedings*, Berlin, 2006, pp. 628-641.
- [8] P. Viola, M. Jones, "Rapid object detection using a boosted cascade of simple features," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001, pp. I-511-I-518.
- [9] Z. Zivkovic, "Improved Adaptive Gaussian Mixture Model for Background Subtraction," *ICPR '04 Proceedings of the Pattern Recognition*, 2004, pp. 28-31
- [10] Н.В. Птицын, "Встроенная видеоаналитика для детектирования и сопровождения объектов при помощи многомасштабных признаков," *Proc. GraphiCon*, 2010, pp. 200-205.
- [11] Р. Салимзибаров, А. Разумков, А. Коробков, "Адаптивный метод выделения движущихся объектов," *Proc. GraphiCon*, 2010, pp. 333-334.
- [12] М. Синдеев, А. Конушин, "Отслеживание контура лабораторной мыши в видеопоследовательности на основе метода ветвей и границ," *GraphiCon*, 2010, pp. 214-217.
- [13] Е. Шальнов, В. Кононов, В. Конушин, "Алгоритм вычитания фона, основанный на поблочных классификаторах," *Proc. GraphiCon*, 2011, pp. 227-230