

Решение Задачи Одновременного Сопровождения и Оконтуривания Методом Максимального Правдоподобия

Ю. В. Слынько

Кафедра Информационных Систем (ОАО «МАК «Вымпел»)

Московский Физико-Технический Институт (Государственный Университет)

UrasSI@yandex.ru

1. ВСТУПЛЕНИЕ

Сопровождение выделенных объектов в видеопоследовательности – актуальная задача для систем наблюдения, охранных систем, интеллектуальных систем управления автомобилем и робототехники. Кроме того, результаты решения этой задачи могут быть использованы как начальный этап для создания более сложной системы анализа видеопоследовательности.

Предложенный в данной работе алгоритм разрабатывался для управляемых камер с возможностью большого увеличения, что накладывает достаточно серьезные требования на его функциональность.

Алгоритм не должен иметь в своей основе представления о статичности и известности фона. Это обусловлено двумя причинами: во-первых, камера может быстро перемещаться (особенно в условиях сильного увеличения), что не дает возможности накопить информацию о фоне; во-вторых, фон может быть сильно нестационарным, например, поверхность моря.

Еще одним условием является отсутствие априорных знаний о размерах и форме объекта. Это требование связано с тем, что алгоритм разрабатывался для систем, в которых оператор, помимо выделения объектов, вынужден управлять камерой. В этом случае он не имеет возможности точно задать размеры объекта, которые могут варьироваться из-за изменения увеличения.

На данный момент задача сопровождения объектов в общем виде не имеет единственного удовлетворительного решения. Существует два различных способа ее решения. Первый – на основе отделения движущихся объектов от фона с последующей классификацией (чтобы выбрать на кадре из всех объектов только интересующий). Второй – на основе поиска на следующем кадре выделенного объекта с предыдущего кадра.

В первом случае алгоритм работает с неподвижными камерами и имеет ряд недостатков, например, невозможность сопровождать неподвижные объекты или объекты на сильно изменяющемся фоне. Такие задачи характерны для систем видеонаблюдения и мониторинга (большинство коммерческих продуктов относится именно к этой категории). Здесь сопровождение объекта основано на использовании стационарности хорошо известного «подстилающего» фона. Обзор таких методов дан в [1].

Для сопровождения объектов на более сложном фоне, как правило, используются алгоритмы с накоплением информации (фильтрацией). Базовыми являются алгоритмы фильтрации на основе фильтра Калмана [2], [3]. Среди наиболее изощренных алгоритмов следует выделить фильтрацию методом частиц [4], [5] и с использованием ядра объекта [6], [7].

Основная проблема алгоритмов с накоплением заключается в завязке сопровождения, т.е. в обработке нескольких первых кадров, пока не накоплено достаточное количество информации. Это особенно актуально, когда алгоритм должен управлять камерой, следящей за объектом. Если алгоритм не справится с задачей, объект исчезнет из поля зрения камеры за несколько кадров.

Для решения задачи сопровождения объекта с одновременным нахождением его контура, алгоритм классификации (разделения) кадра на объект и фон должен быть сильно взаимосвязан с алгоритмом сопровождения (нахождения сдвига). Один из примеров таких алгоритмов представлен в [8]. Однако здесь заложено жесткое ограничение на форму объекта – она должна быть близка к прямоугольной.

Предложенный в данной статье алгоритм позволяет решить задачу сопровождения и оконтуривания движущихся объектов при отсутствии априорной информации об их форме, размерах и характере изменения фона.

2. ТЕОРЕТИЧЕСКИЕ ОСНОВЫ МЕТОДА

Будем решать задачу одновременного определения контура объекта и его сдвига методом максимального правдоподобия. В случае отсутствия априорной информации данный подход является оптимальным.

Кадр f_t момента времени t состоит из конечного количества частей – пикселей (для ускорения алгоритма можно использовать части, состоящие из нескольких пикселей; далее будет использоваться термин «часть»). Будем считать, что каждая часть полностью принадлежит либо объекту, либо фону. Первоначальный выбор объекта осуществляется оператором и заключается в выборе части i_0 , которая заведомо принадлежит объекту.

Запишем функцию правдоподобия для данной задачи, т.е. принадлежности всех частей кадра к объекту или фону:

$$J(I) = \prod_{i \in I} p(i \in object) \prod_{i \notin I} p(i \notin object), \quad (1)$$

где i – номер части в кадре, I – множество частей, принадлежащих объекту, $p(i \in object)$ – вероятность того, что часть i принадлежит объекту.

Таким образом, для нахождения сдвига необходимо найти максимум правдоподобия, т.е.

$$I = \arg \max_I J(I) \quad (2)$$

Основная проблема заключается в нахождении вероятности того, что часть принадлежит объекту. Т.к., по условиям задачи, объект отличается от фона только вектором сдвига, такая вероятность должна зависеть от сдвига объекта и фона.

Вычислим функцию невязки (как функцию сдвига $(\delta x, \delta y)$) для каждой части и всего объекта (подробности вычисления этой функции могут быть найдены в [9]):

$$\begin{aligned} \tilde{C}_i(\delta x, \delta y) &= \sum_{l, k \in P_i} (f_t(l, k) - f_{t-1}(l - \delta x, k - \delta y))^2, \\ \tilde{F}(\delta x, \delta y) &= \sum_{i \in I} \tilde{C}_i(\delta x, \delta y), \end{aligned} \quad (3)$$

где P_i – пиксели, принадлежащие части i .

Будем рассматривать вычисленные функции невязки как оценки истинных C_i и F . Минимум этих функций будет оценкой сдвига каждой части или всего объекта. Можно было бы выбрать части объекта, сдвиг которых близок к сдвигу части i_0 . Но как показали эксперименты, этот алгоритм чрезвычайно неустойчив из-за больших ошибок определения сдвига отдельной части.

Более точным будет вычисление вероятности $p(i \in object)$ как вероятности того, что минимум C_i и F совпадают. Тогда

$$p(i \in object) = \sum_{\delta x, \delta y} p_i(\delta x, \delta y) p_F(\delta x, \delta y) \text{ или приближенно} \quad (4)$$

$$p(i \in object) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_i(\delta x, \delta y) p_F(\delta x, \delta y) d\delta x d\delta y \quad (5)$$

где $p_i(\delta x, \delta y)$ и $p_F(\delta x, \delta y)$ - вероятности того, что значение сдвига части или объекта равны $(\delta x, \delta y)$.

Эти вероятности могут быть оценены из оценок функций невязок. Для этого заменим распределение оценок сдвига каждой части нормальным распределением:

$$p_i(\delta x, \delta y) = \frac{1}{2\pi\sqrt{\det(\Gamma_i)}} e^{-\frac{1}{2}(\bar{x}-\bar{x}_i)^T \Gamma_i^{-1}(\bar{x}-\bar{x}_i)}, \text{ где} \quad (6)$$

$$\bar{x} = (\delta x, \delta y), \quad \Gamma_i = \begin{pmatrix} \sigma_{ix}^2 & \rho_i \sigma_{ix} \sigma_{iy} \\ \rho_i \sigma_{ix} \sigma_{iy} & \sigma_{iy}^2 \end{pmatrix}$$

и аналогично для F .

Параметры этого распределения $\bar{x}_i, \sigma_{ix}, \sigma_{iy}, \rho_i$ выберем исходя из оценки функции невязки (подробности см. в Приложении 1). Тогда, вычисляя интеграл (5), находим:

$$p(i \in object) = \frac{\exp(\frac{B}{2A})}{2\pi\sqrt{-A}}, \text{ где} \quad (7)$$

$$A = 2\rho_i \rho_F \sigma_{Fx} \sigma_{Fy} \sigma_{ix} \sigma_{iy} +$$

$$\sigma_{Fx}^2 ((\rho_F^2 - 1) \sigma_{Fy}^2 - \sigma_{iy}^2) +$$

$$\sigma_{Fx}^2 ((\rho_i^2 - 1) \sigma_{iy}^2 - \sigma_{Fy}^2)$$

$$B = (\sigma_{Fy}^2 + \sigma_{iy}^2)(x_F - x_i)^2 -$$

$$2(\rho_F \sigma_{Fx} \sigma_{Fy} + \rho_i \sigma_{ix} \sigma_{iy})(x_F - x_i)(y_F - y_i) +$$

$$(\sigma_{Fx}^2 + \sigma_{ix}^2)(y_F - y_i)^2$$

3. ДЕТАЛИ РЕАЛИЗАЦИИ

Для решения поставленной задачи – нахождения контура объекта с одновременным определением его сдвига – необходимо максимизировать функцию правдоподобия (1). Основная

проблема заключается в том, что функция невязки F для объекта зависит от контура объекта I . Это затрудняет поиск максимума. Задачу можно решить методом перебора, что чрезвычайно трудоемко и исключает практическое применение. Для сокращения объема вычисления и повышения устойчивости, части выбираются как квадраты размером 8x8 пикселей.

Поэтому предлагается следующая схема: контур объекта вычисляется одним из стандартных алгоритмов кластеризации, а затем рекуррентно уточняется, используя функцию правдоподобия.

Задачей алгоритма кластеризации является нахождение первого приближения контура объекта. Кластеризация ведется по сдвигам отдельных частей – т.е. вычисляется множество сдвигов всех частей и делится на две группы, одна из которых соответствует объекту, а вторая – фону.

В качестве алгоритма кластеризации были использованы два хорошо известных алгоритма: на основе матрицы сходства и метод К-средних. Описание обоих методов дано в [10].

Задав начальное разбиение кадра на объект и фон можно найти локальный максимум функции правдоподобия методом последовательных приближений.

Для этого вычислим функцию невязки $\tilde{F}(\delta x, \delta y)$ для объекта по контуру I . Затем для каждой части найдем вероятности $p(i \in object)$ и $p(i \notin object)$. Если $p(i \in object) > p(i \notin object)$, то часть принадлежит объекту, иначе – фону (это следует из критерия максимизации функции правдоподобия). Таким образом, строится новый контур объекта. Итерация повторяется (т.е. для нового контура вычисляется $\tilde{F}(\delta x, \delta y)$ и т.д.). Когда контур I перестанет изменяться – это будет означать, что достигнут локальный максимум, и итерации следует прекратить.

На практике такой подход позволяет достаточно быстро (за 3-4 итерации) находить локальный максимум, который достаточно близок или вовсе совпадает с глобальным.

4. РЕЗУЛЬТАТЫ ТЕСТИРОВАНИЯ И ВЫВОДЫ

Работа алгоритма была проверена на большом количестве видеопоследовательностей разного характера. Особенно внимательно исследовались случаи существенно нестационарного фона и быстрых перемещений камеры (а значит и фона).

Помимо реальных видеопоследовательностей, качество работы алгоритма проверялось и на модельных, где на нестационарный фон накладывалось изображение движущегося объекта. Такой метод позволяет получить оценки точности алгоритма определения сдвига. На исследованных видеопоследовательностях СКО определения сдвига составила 0.08 пикселей.

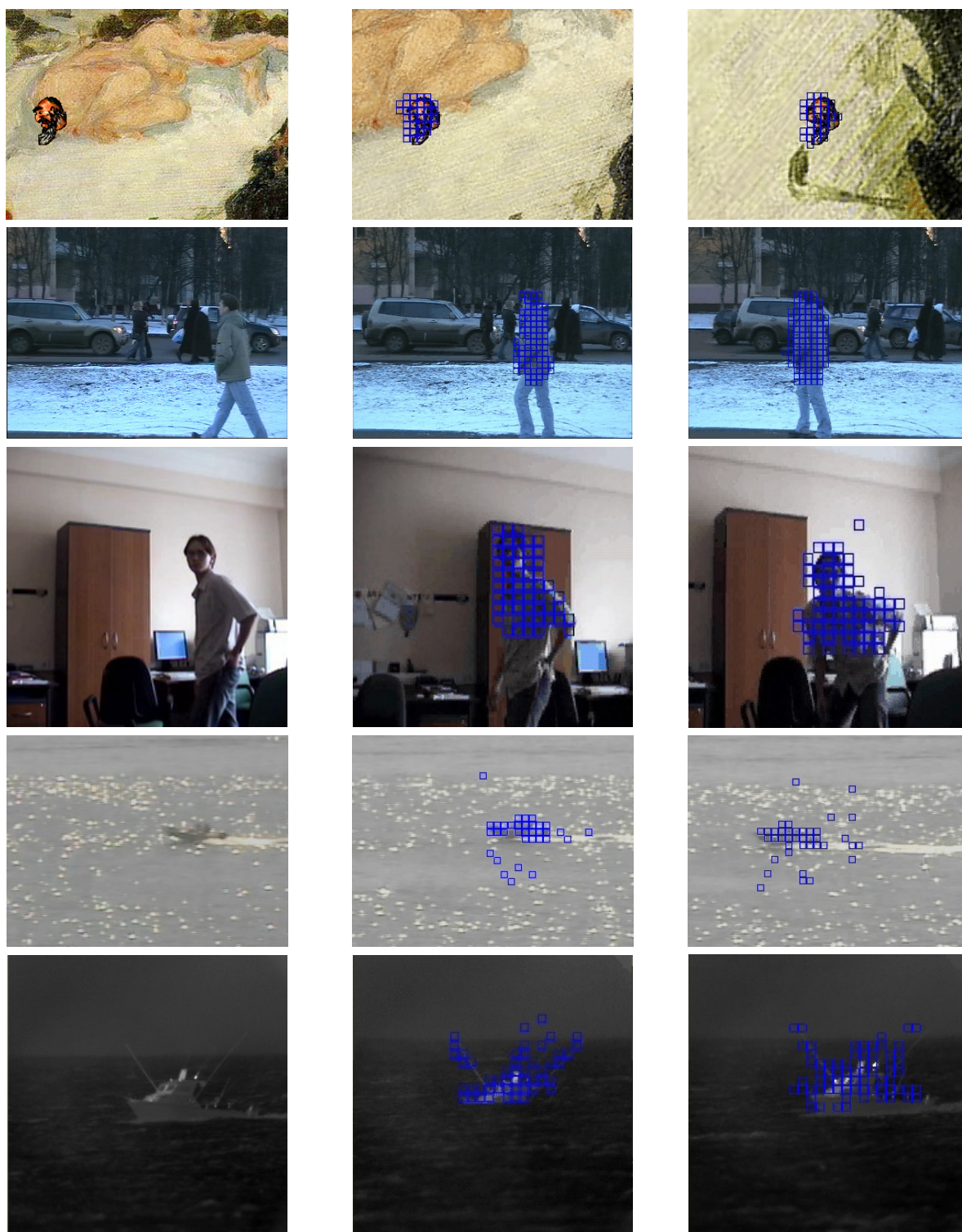


Рис. 1. Результаты работы алгоритма. Синими квадратами обозначены части, определенные как принадлежащие объекту.

На рис. 1 приведены результаты работы алгоритма с различными видеопоследовательностями. В верхнем ряду представлены результаты работы алгоритма на синтетических кадрах. Два средних ряда – сопровождение человека на нестационарном фоне. В двух нижних рядах – сопровождение

лодки на водной поверхности с яркими информативными бликами при перемещающейся камере. Видно, что некоторые части фона ошибочно определяются как принадлежащие объекту. Они в дальнейшем отсеиваются методами морфологической обработки. На изображениях нижнего ряда видно, что определились как части объекта даже удочки и антенны.

Таким образом, был создан алгоритм сопровождения и оконтуривания выделенных объектов при нестационарном фоне и отсутствии априорной информации о размерах и контуре объекта.

5. ЛИТЕРАТУРА

- [1] Fatih Porikli. *Achieving Real-Time Object Detection and Tracking Under Extreme Conditions*.
- [2] C. Wren, A. Azarbayejani, T. Darell, and A. Pentland, "Pfinder: Real-time tracking of the human body," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, pp. 780–785, 1997.
- [3] Y. Boykov and D. Huttenlocher, "Adaptive bayesian recognition in tracking rigid objects," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Hilton Head, SC*, vol. II, 2000, pp. 697–704.
- [4] M. Isard and I. Blake, "Condensation – conditional density propagation for visual tracking," in *Intl. J. of Computer Vision*, vol. 29, 1998, pp. 5–28.
- [5] N. Bouaynaya, W. Qu, and D. Schonfeld, "An online motion-based particle filter for head tracking applications," in *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Philadelphia*, 2005.
- [6] D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Hilton Head, SC*, vol. 1, 2000, pp. 142–149.
- [7] F. Porikli and O. Tuzel, "Object tracking in low-frame-rate video," in *Proc. of PIE/EI - Image and Video Communication and Processing, San Jose, CA*, 2005.
- [8] S. Avidan, "Ensemble tracking," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition, San Diego, CA*, 2005.
- [9] Yu. V. Slyn'ko, V. N. Lagutkin, and A. P. Luk'yanov. *Development and Investigation of Real-Time Robust Algorithms for Estimating the Parameters of Geometric Transformations of Video-Sequence Frames. Journal of Communications Technology and Electronics*, 2007, Vol. 52, No. 3, pp. 332–338.
- [10] Д. Форсайт, Ж. Понс. *Компьютерное зрение. Современный подход*. М: Вильямс, 2004.

ПРИЛОЖЕНИЕ 1

Для оценки разброса значений функции невязки C_i предлагается следующая схема.

Исходя из предположения о том, что шумы в пикселях распределены нормально и некоррелированы, находится доверительная область нахождения минимума функции невязки (подробности см. в [9])

$$\tilde{D} = \{\delta x, \delta y : \tilde{C}_i(\delta x, \delta y) < m_n + B \cdot \sigma_n\}, \quad (8)$$

где m_n - значение функции невязки в минимуме, $\sigma_n = \frac{m_n \sqrt{2}}{\sqrt{N_n}}$ - дисперсия значений невязки, N_n - количество точек в области, по которой подсчитана функция C_i (для части размеров 8 на 8 $N_n = 64$), B - порог, определяющий доверительную вероятность ошибки.

По этой области рассчитываются параметры распределения (6)

$$\bar{x}_i = \begin{pmatrix} x_i \\ y_i \end{pmatrix} = \frac{1}{M(\tilde{D})} \sum_{\delta x, \delta y \in \tilde{D}} \begin{pmatrix} \delta x \\ \delta y \end{pmatrix}, \quad (9)$$

$$\begin{pmatrix} \sigma_{ix} \\ \sigma_{iy} \end{pmatrix} = \frac{1}{M(\tilde{D})} \sum_{\delta x, \delta y \in \tilde{D}} \begin{pmatrix} (\delta x - x_i)^2 \\ (\delta y - y_i)^2 \end{pmatrix},$$

$$\rho_i = \frac{1}{M(\tilde{D})} \frac{1}{\sigma_{ix} \sigma_{iy}} \sum_{\delta x, \delta y \in \tilde{D}} (\delta x - x_i)(\delta y - y_i),$$

Вычисление параметров для F производится аналогично.