

Трекинг объектов на видео при помощи фильтра частиц

Нижибицкий Евгений Алексеевич

научный руководитель: д.ф.-м.н. Дьяконов А.Г.

Московский государственный университет имени М. В. Ломоносова
Факультет Вычислительной Математики и Кибернетики
Кафедра Математических Методов Прогнозирования

2 июня 2014 г.

Задача видеотрекинга

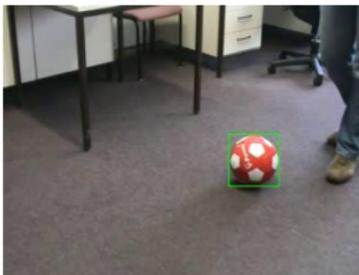
Исследуем задачу сопровождения предложенного объекта на видео — его начальное положение при этом считается заданным. Для представления местоположения цели используются координаты прямоугольника, отвечающего участку изображения, содержащему объект.



Рис. : Выделение объекта на видео.

- Многие подходы опираются на достаточно ограниченный и разный для каждой работы набор признаков, не рассматривая другие или из-за самостоятельной неэффективности или из-за вычислительной сложности для многократного использования.
- Исследовать эффективность композиций простых признаков по сравнению с (не)эффективностью признаков самих по себе.
- Исследовать способы оптимизации многократного извлечения признаков из участков изображений.

В качестве датасета для экспериментов выступает Bonn Benchmark on Tracking (BoBoT) — датасет общей направленности с различными видео, отвечающими тем или иным осложнениям при трекинге.



Фильтр частиц для видеотрекинга

Общий план

Распределение $p(x_t)$ возможных положений отслеживаемого объекта моделируется взвешенными частицами $s_t^i = (x_t^i, \pi_t^i)$.

$$\hat{p}(x_t) = \sum_{i=1}^N \pi_t^i \delta_{x_t^i}(x_t).$$

В работе состояние частицы $x = (x, y, v_x, v_y, s)^T$.

Новый набор S_t получается из предыдущего S_{t-1} , откуда s_{t-1}^i выбирается с вероятностью, пропорциональной весу π_{t-1}^i .

Новые состояния x_t^j получаются из модели движения:

$$p(X_t | X_{t-1} = x_{t-1}^i).$$

Измерение в новом кадре Z_t используется при обновлении весов π_t^i с помощью подсчета правдоподобия наблюдения:

$$\pi_t^j = p(Z_t | X_t = x_t^j, Z_0, Z_1, \dots, Z_{t-1}).$$

Фильтр частиц для видеотрекинга

Модель наблюдения (правдоподобие)

Рассматриваются признаки $\{f\}$, для которых определены меры схожести $\{\rho_f\}$ с шаблонами для реального объекта.

Общее правдоподобие наблюдения:

$$p(\mathbf{z}_t | \mathbf{x}_t) \propto \prod_f \exp\{-\rho_f^2(\hat{h}_f, h_f(\mathbf{x}_t)) / \lambda_f\},$$

\hat{h}_f — шаблон, а $h_f(\mathbf{x}_t)$ — признаки для региона изображения, соответствующего частице \mathbf{x}_t .



Рис. : Слева: исходное изображение F. По центру: интегральное изображения для изображения F. Справа: 4 региона, которые используются при подсчете суммы для региона D изображения F.

Признаки

Цветовые гистограммы (Color)

Для каждого цветового канала рассчитываем количество пикселей региона, интенсивность которых равна заданному значению или попадает в один из равных промежутков (в работе рассмотрен случай 8 «корзин»), объединяем в вектор.

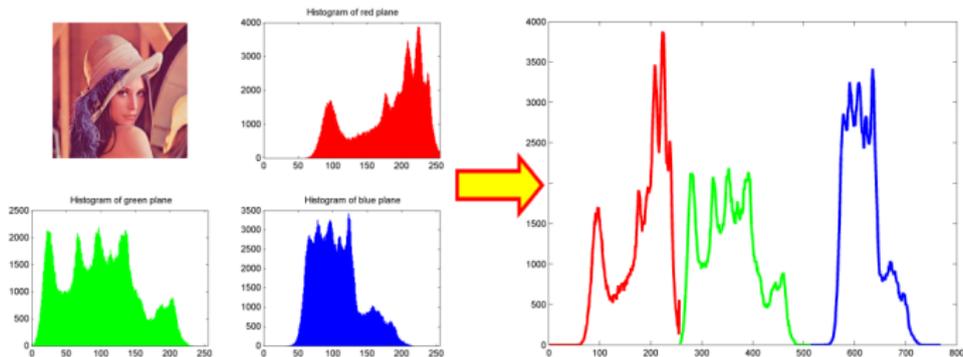


Рис. : Объединение гистограмм для каналов.

В качестве признаков для регионов считаем разности между суммарными интенсивностями светлых и тёмных субрегионов:



Рис. : Рассматриваемые виды прямоугольников Хаара.

Получаем локальную структуру пикселя, сравнивая его с соседями — для каждого соседа получаем 1, если его интенсивность больше, и 0 иначе.

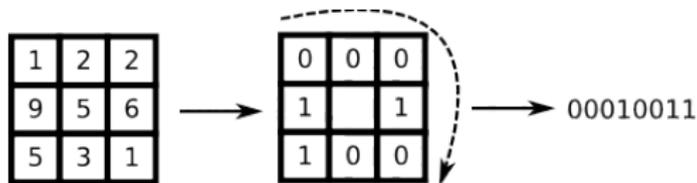


Рис. : Получение локального бинарного шаблона.

Силы резкости границ распределяются по K «корзинам» с направлениями, после чего значения суммируются:

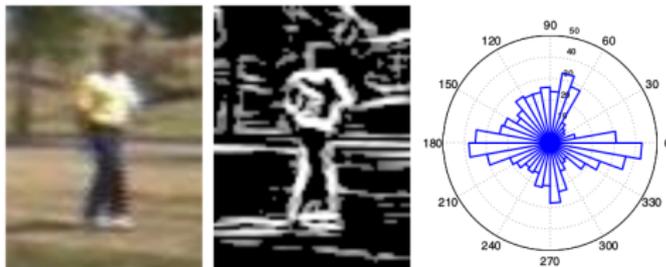


Рис. : Получение HOG для изображения.

$$G_x(x, y) = K_x * I(x, y),$$

$$G_y(x, y) = K_y * I(x, y),$$

$$S(x, y) = \sqrt{G_x^2(x, y) + G_y^2(x, y)},$$

$$\theta = \arctg(G_y(x, y)/G_x(x, y)).$$

Программная реализация

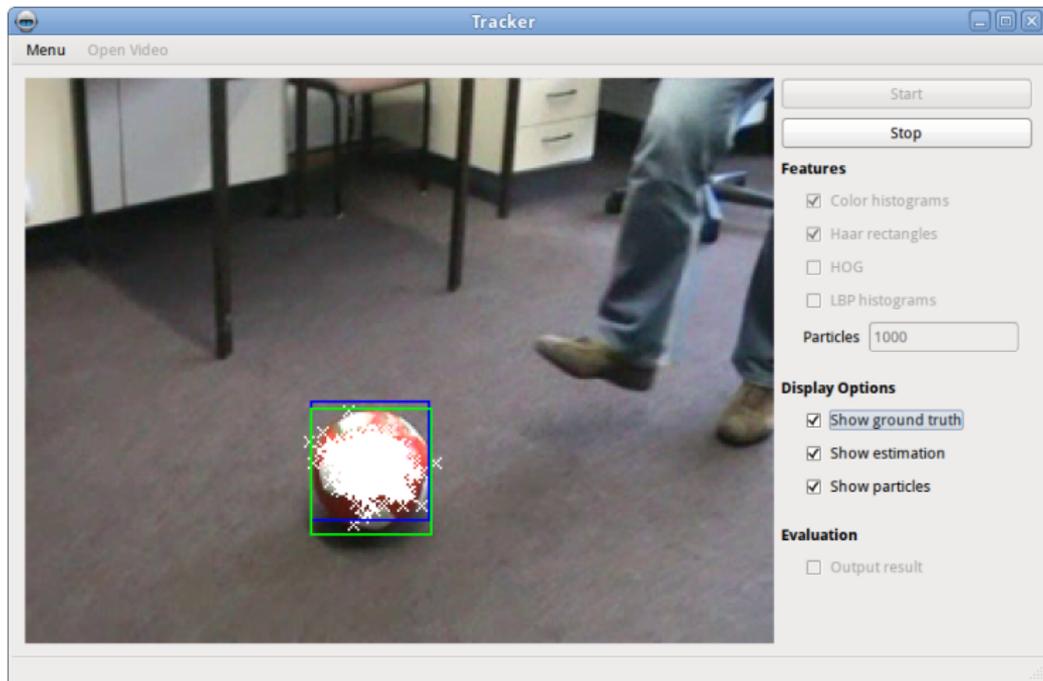


Рис. : Окно работающей программы.

Мера качества позиционирования объекта — доля пересечения фактической и предсказанной областей в их объединении.

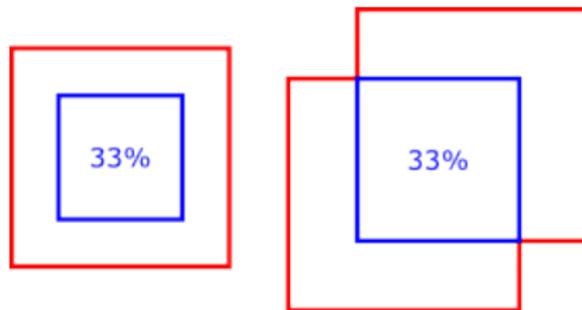


Рис. : Пример для значения качества, равного 33%.

Качеству выше 33% соответствует правильное определение положения объекта.

Результаты экспериментов

Первая группа признаков

Сначала рассматривались алгоритмы только на основе одного из признаков для поиска «самодостаточных» признаков.

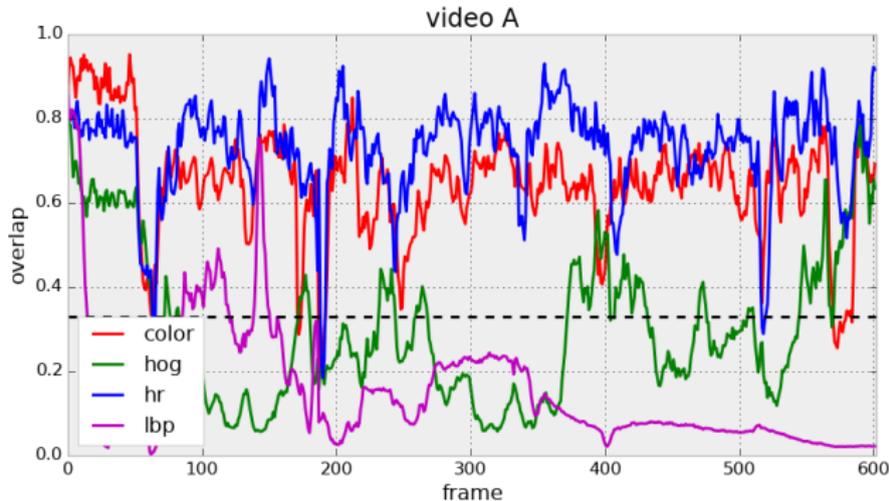


Рис. : Видео А: распределение качества по кадрам для одиночных признаков.

Результаты экспериментов

Вторая группа признаков

Затем исследовался эффект от объединения в композицию обоих самостоятельных «цветовых» признаков.

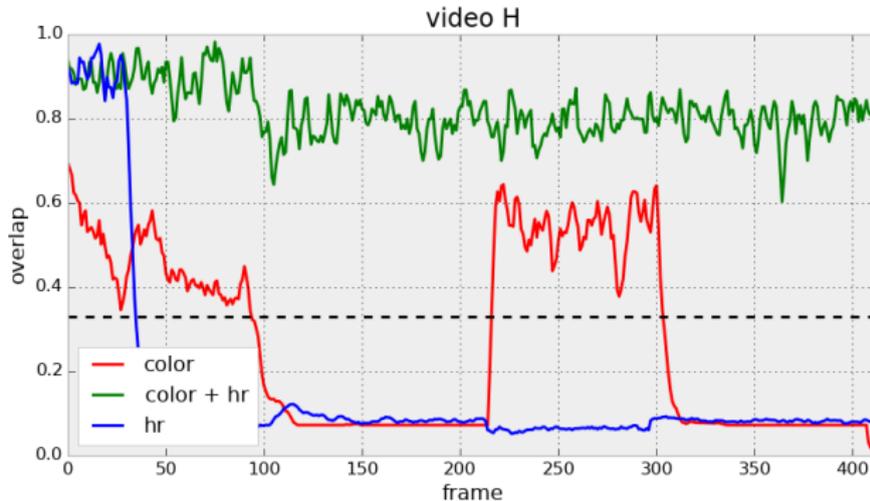


Рис. : Видео H: распределение качества по кадрам для цветowych признаков.

Результаты экспериментов

Третья группа признаков

В конце исследовалось то, какое улучшение даёт добавление признаков с информацией о структуре/текстуре объекта.

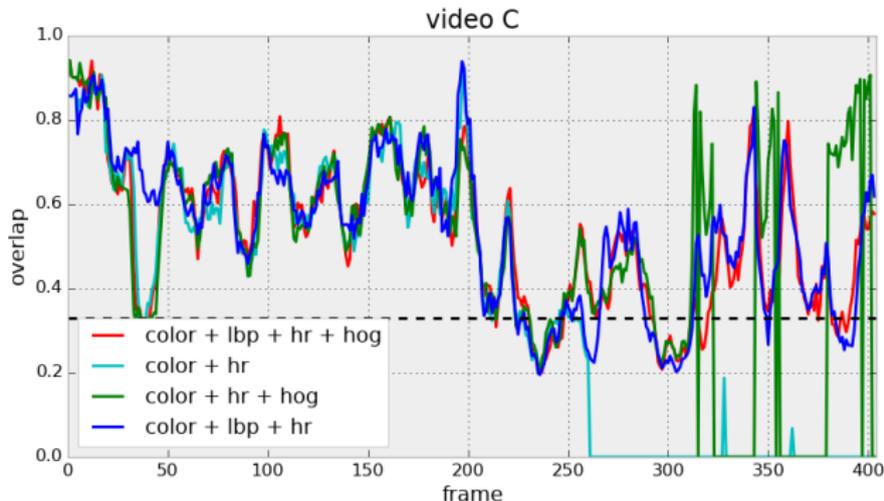


Рис. : Видео С: распределение качества по кадрам для структурных признаков.

Результаты экспериментов

Сводная таблица

В качестве результата рассматривалось отношение количества кадров с успешным нахождением объекта (рассмотренная мера качества при этом $\geq 33\%$) к общей длине видео.

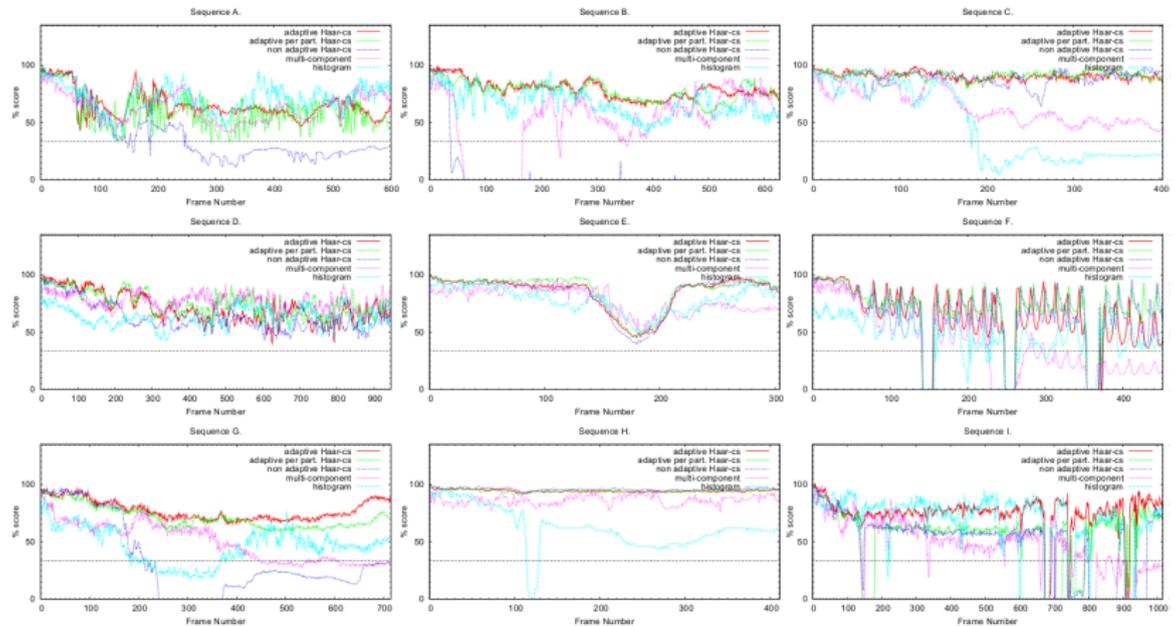
Группа признаков	A	B	C	D	E	F	G	H	I
color	0.966	0.732	0.252	0.947	0.931	0.556	1.000	0.441	0.719
hr	0.986	0.804	0.475	0.977	0.855	0.439	1.000	0.084	0.649
hog	0.367	0.184	0.178	0.714	0.200	0.569	0.117	0.415	0.074
lbp	0.102	0.081	0.091	0.114	0.347	0.256	0.146	0.038	0.066
hr+color	0.993	1.000	0.564	0.991	1.000	0.602	0.997	1.000	0.779
hr+color+lbp	0.998	1.000	0.863	0.991	1.000	0.613	1.000	1.000	0.782
hr+color+hog	0.996	0.992	0.764	0.991	1.000	0.613	0.853	1.000	0.770
hr+color+lbp+hog	0.996	0.992	0.863	0.990	0.996	0.613	0.863	1.000	0.772

Рис. : Сравнительная таблица всех рассмотренных алгоритмов.

Результаты экспериментов

Сравнение с другими реализациями

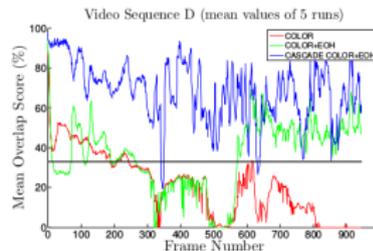
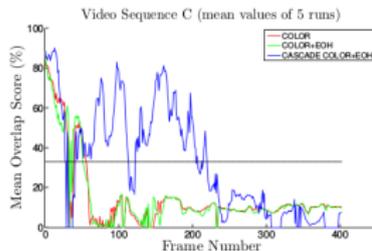
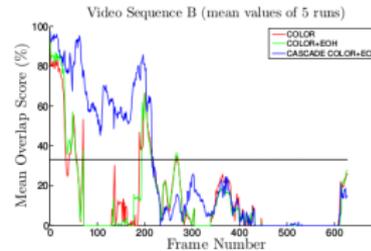
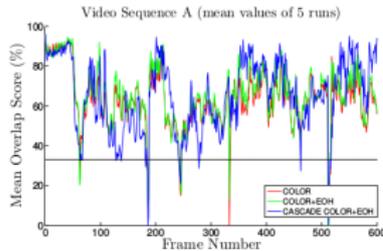
Результаты в оригинальной работе, в которой впервые был предложен для анализа изученный датасет VoVoT.



Результаты экспериментов

Сравнение с другими реализациями

Результаты в работе схожей тематики на основе метода каскадов. Из цветовых использовались только прямоугольники Хаара, гистограммы не использовались под предлогом меньшей устойчивости к изменениям по сравнению с ними!



- Разработан метод, обобщающий использование интегральных изображений для оптимизации многократного подсчета рассмотренных признаков.
- Экспериментально показано, что рассмотрение композиций признаков осмысленно даже для неработоспособных самих по себе признаков.
- С помощью рассмотренных композиций признаков достигнуто качество трекинга, сравнимое с более продвинутыми методами, основанными на построении сложных ансамблей с помощью бустинга, а также превышающее результаты для схожей методики.
- Создана программная реализация с графическим интерфейсом для демонстрации работы всех изученных алгоритмов с возможностью создания произвольных композиций на основе рассмотренных признаков.