

Композиции признаков для видеотрекинга при помощи фильтра частиц*

Нижбицкий Е. А.

nizhibitsky@cs.msu.ru

Москва, Факультет ВМК МГУ имени М. В. Ломоносова

В работе рассмотрены модели правдоподобия в фильтре частиц для задачи отслеживания объектов на видео, основанные на композиции мер сходства извлекаемых из изображений признаков. Предложены способы оптимального многократного извлечения признаков из различных регионов одного кадра видео. Оптимизация при этом выполняется за счет использования интегральных изображений, впервые использованных в компьютерном зрении для признаков Хаара, для других исследуемых признаков. Экспериментально показана возможность эффективного использования композиций групп признаков при неэффективности использования каждой группы в отдельности. С помощью рассмотренных композиций признаков достигнуто качество трекинга, сравнимое с более продвинутыми методами, основанными на построении сложных ансамблей с помощью бустинга, и превышающее результаты схожей работы с использованием метода каскадов.

Ключевые слова: *трекинг, фильтр частиц, интегральное изображение, гистограмма направленных градиентов, признаки Хаара, локальные бинарные шаблоны, композиция признаков.*

Feature composition in video tracking using particle filters*

Nizhibitsky E. A.

Faculty of Computational Mathematics and Cybernetics, Moscow State University

The target of the survey is likelihood models based on similarity measures of features extracted from images for the task of object tracking in video sequences using particle filters. Methods for optimal multiple feature extraction from several regions of the same image are proposed. Optimization is performed using integral images, first prominently used in CV within the Viola–Jones object detection framework for Haar rectangles, for other studied features. Experimentally demonstrated that feature compositions can be used even in circumstances where each of the features are somehow useless on their own. The performance achieved using proposed compositions is greater than one in the similar study and sometimes comparable to the performance of more complicated models based on ensemble boosting.

Keywords: *tracking, particle filter, integral image, histogram of oriented gradients, HOG, Haar features, local binary patterns, LBP, feature composition.*

1 Введение

2 Задача трекинга объектов на видео является неотъемлемой частью многих приклад-
3 ных областей, таких как построение систем видеонаблюдения, отслеживания дорожного
4 трафика (в частности, наблюдения за определенными транспортными средствами в пото-
5 ке), создание интерфейсов человек-компьютер, программ для передачи и сжатия видео и
6 других.

7 За последние годы было предложено множество успешных подходов по решению дан-
 8 ной задачи. Тем не менее, многие из них накладывают определенные ограничения на
 9 обрабатываемые данные, как то статичный фон и фиксированный ракурс [5], знание о
 10 типе наблюдаемого объекта [7] или даже наличие множества камер [10]. Одни подходы
 11 уделяют мало внимания вычислительной сложности, другие наоборот учитывают строгие
 12 ограничения по ресурсам, примером чего являются приложения в робототехнике [2].

13 Так или иначе значительное число подходов опирается на использование фильтра ча-
 14 стиц для приближения вероятностного распределения на положения объекта на видео с
 15 помощью частиц, или сэмплов, которым отвечают регионы на видео и те или иные допол-
 16 нительные характеристики. Для каждой частицы при этом необходимо подсчитывать ее
 17 вес, который отвечает схожести данного региона с регионом для отслеживаемого объекта,
 18 а значит необходимо уметь выделять признаки из них.

19 Большинство работ при определении таких мер сходства опираются на достаточно
 20 ограниченный и разный для каждой работы набор признаков. Другие признаки при этом
 21 не рассматриваются или по причине самостоятельной неэффективности в условиях каких-
 22 то возникающих на исследуемых видео сложностей (например, изменяющееся освещение)
 23 или из-за вычислительной сложности многократного выделения признаков для каждой
 24 частицы.

25 Целью данной работы является исследование возможности эффективного использова-
 26 ния композиций признаков даже там, где каждый из них может быть неэффективен сам по
 27 себе, а также построение способов оптимального подсчета этих признаков для многократ-
 28 ного извлечения из различных регионов одного изображения, соответствующего кадру на
 29 видео.

30 **Фильтр частиц для видеотрекинга**

31 Распределение $p(x_t)$ возможных положений отслеживаемого объекта на i -м кадре видео
 32 приближается набором $S_t = \{s_t^i\}_{i=1}^N$ (далее $N=1000$) взвешенных частиц $s_t^i = (x_t^i, \pi_t^i)$:

$$33 \quad \hat{p}(x_t) = \sum_{i=1}^N \pi_t^i \delta_{x_t^i}(x_t).$$

34 В работе в качестве состояния частицы рассматривается вектор $(x, y, v_x, v_y, s)^\top$, учиты-
 35 вающий положение и скорость перемещения рамки, содержащей отслеживаемый объект, а
 36 так же масштаб относительно ее первоначальных размеров при инициализации (считаем,
 37 что пропорции объекта не изменяются).

38 При инициализации начальное положение вместе с рамкой (x, y, w, h) передается ал-
 39 горитму, веса частиц при инициализации принимаются равными между собой — $\pi_0^j = \frac{1}{N}$,
 40 начальные скорости (v_x, v_y) получаются из нормального распределения.

41 Для каждого нового кадра получается новый набор на основе алгоритма фильтра ча-
 42 стиц с промежуточным ресэмплингом с учетом важности (веса) каждой частицы (Sampling
 43 Importance Resampling Particle Filter):

44 Таким образом, на каждом шаге на основе имеющихся частиц получают новые с
 45 помощью моделирования их перемещения, затем оценивается правдоподобие нового со-
 46 стояния каждой частицы. В качестве промежуточного этапа перед новым шагом вместо
 47 частиц с различными весами мы будем получать частицы с одинаковыми весами, где их
 48 состояния будут выборкой нужного размера из дискретного распределения на предыду-
 49 щих состояниях с вероятностями, пропорциональными их весам.

Алгоритм 1 Фильтр частиц для видеотрекинга (SIR PF)**Вход:** S_{t-1} **Выход:** S_t для $i = 1$ to N получить $x_t^i \sim p(x_k | x_{t-1}^i)$ // модель движения вычислить $\pi_t^i = p(Z_t | X_t = x_t^i, Z_0, Z_1, \dots, Z_{t-1})$ // модель правдоподобиявычислить $w = \sum_{i=1}^N \pi_t^i$ для $i = 1$ to N вычислить $\pi_t^i = w^{-1} \pi_t^i$ // нормализация весовдля $i = 1$ to N получить $\hat{x}_t^i \sim \{p(x_k = x_t^j) = \pi_t^j, j = 1, \dots, N\}$ // ресэмплинг**return** $\{(\hat{x}_t^i, N^{-1})\}_{i=1}^N$

50 Модель движения в рассматриваемых экспериментах учитывает имеющиеся скорости
 51 частиц для определения новых координат. Сами же скорости, как и масштаб рамки для
 52 каждой частицы, получаются на основе зашумления предыдущих значений нормальным
 53 распределением:

$$v_{x,t} = v_{x,t-1} + N(0, \sigma_x^2),$$

$$v_{y,t} = v_{y,t-1} + N(0, \sigma_y^2),$$

$$54 \quad x_t = x_{t-1} + v_{x,t},$$

$$y_t = y_{t-1} + v_{y,t},$$

$$s_t = s_{t-1} + N(0, \sigma_s^2).$$

55 Модель правдоподобия, построение которой и является целью исследования, рассмот-
 56 рена отдельно в следующем разделе.

Модель правдоподобия на основе композиции признаков

57 Чтобы вычислить вес каждой частицы, нужно оценить правдоподобие наблюдения,
 58 отвечающего ей, т.е. оценить схожесть региона, отвечающего частице, с шаблоном — таким
 59 же регионом для отслеживаемого объекта. Для этого мы будем извлекать признаки из
 60 регионов изображения и сравнивать их между собой. Далее под шаблоном также будут
 61 пониматься значения признаков для целевого объекта.

62 В данной работе рассматривались признаки, зарекомендовавшие себя в задаче отсле-
 63 живания объектов, моделирующие представления объекта с разной стороны, а значит
 64 подходящие для различных сложностей в исследуемых видео — одни модели используют
 65 цветовые признаки, другие же моделируют текстуру, контур или иные характеристики
 66 объекта. Для набора рассматриваемых признаков $\{f\}$ определялись меры схожести $\{\rho_f\}$
 67 с шаблонами для реального объекта.

68 Распределение правдоподобия на основе одной метрики $\rho(\cdot, \cdot)$ можно получить по фор-
 69 муле:

$$71 \quad p(Z_t | x_t) \propto \exp\{-\rho^2(\hat{h}_f, h_f(x_t))/\lambda\},$$

72 где \hat{h}_f — шаблон, $h_f(x_t)$ — признаки для региона изображения, соответствующего состоя-
 73 нию x_t частицы s_t , а λ — параметр, подбираемый отдельно для каждой пары признака и
 74 соответствующей ему метрики.

75 Общее правдоподобие наблюдения при можно посчитать как произведение правдопо-
 76 добий по каждому признаку:

$$77 \quad p(Z_t | x_t) \propto \prod_f \exp\{-\rho_f^2(\hat{h}_f, h_f(x_t))/\lambda_f\}.$$

78 Далее мы сначала рассмотрим интегральные изображения, потом опишем используе-
79 мые признаки, и в конце покажем, как можно оптимизировать их многократный подсчет
80 для различных регионов одного изображения.

81 **Интегральные изображения.** Чтобы получить интегральное изображение I на ос-
82 нове исходного F , для каждого пикселя необходимо вычислить значение по формуле

$$83 \quad I(x, y) = F(x, y) + I(x - 1, y) + I(x, y - 1) - I(x - 1, y - 1),$$

84 где $I(x, -1) = I(-1, y) = 0$. А это, очевидно, можно сделать за один проход по результи-
85 рующему изображению с помощью динамического программирования.

86 После того, как было получено интегральное изображение, чтобы посчитать сумму
87 интенсивностей для прямоугольника с верхним левым углом (x_1, y_1) и нижним правым
88 углом (x_2, y_2) , нужно воспользоваться формулой:

$$89 \quad \sum_{x=x_1}^{x_2} \sum_{y=y_1}^{y_2} F(x, y) = I(x_2, y_2) - I(x_2, y_1 - 1) - I(x_1 - 1, y_2) + I(x_1 - 1, y_1 - 1).$$

90 Данное выражение эквивалентно подсчету суммы для региона D изображения F с помо-
91 щью вычисления $(A + B + C + D) - (A + B) - (A + C) + A$ для изображения I .

92 Интегральные изображения нашли широкое применение в работе Виолы-Джонса [9],
93 где использовались для многократного выделения признаков Хаара, рассматриваемых да-
94 лее, из различных регионов одного изображения в их быстром алгоритме детектирования
95 объектов.

96 **Цветовые гистограммы.** В качестве первой группы признаков рассматривались
97 простые поканальные гистограммы для каждого из трех каналов RGB-изображения. Для
98 оптимизации в дальнейшем эти значения в каждом канале объединялись в 8 *корзин* из 32-х
99 значений интенсивности. Три группы корзин, объединенные между собой и затем норма-
100 лизованные, образуют вектор-признак из 24 значений. Метрика сходства двух гистограмм
101 при этом определялась на основе Евклидовой метрики:

$$102 \quad \rho_{hist}(\hat{h}, h(\mathbf{x}_t)) = \sqrt{\sum_{i=1}^{24} (\hat{h}_i - h_i)^2},$$

103 где \hat{h} — шаблон, а $h(x_t)$ — гистограмма для региона x_t , отвечающего частице s_t .

104 **Признаки Хаара.** Признаки Хаара впервые были описаны в [9], где они исполь-
105 зовались в алгоритме для распознавания лиц, и на сегодняшний день применяются во
106 многих алгоритмах классификации, т.к. обладают большей дискриминативной способно-
107 стью, чем значения пикселей сами по себе. Для получения признаков каждый подрегион
108 разбивается дополнительно на условно светлые и темные области, состоящие из одного
109 или нескольких прямоугольников (см. рис. 1), для каждой из которых затем вычисляется
110 среднее значение по каналам.

111 Чтобы вычислить разницу между признаком для региона, полученным для частицы, и
112 шаблоном, можно также воспользоваться Евклидовой метрикой для векторов из разностей
113 значений в светлых и темных областях:

$$114 \quad \rho_{hr}(\hat{c}, c(x_t)) = \sqrt{\sum_{v=1}^4 ((red_{\hat{c}_v} - red_{(x_t,v)})^2 + (green_{\hat{c}_v} - green_{(x_t,v)})^2 + (blue_{\hat{c}_v} - blue_{(x_t,v)})^2)},$$

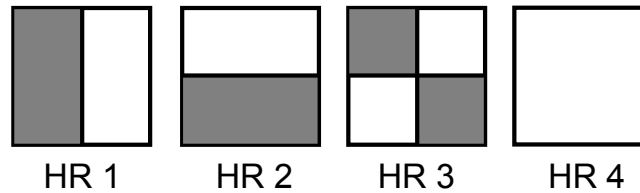


Рис. 1. Четыре признака Хаара для извлечения средних значений по каналам.

115 где \hat{c} — цветовая информация из шаблона, $c(x_t)$ — цветовая информация для частицы с
116 состоянием x_t , а v — один из типов признаков, изображенных на рис. 1.

117 **Гистограммы направленных градиентов.** Признаки на основе цветов подходят
118 для многих задач трекинга, даже когда происходят частичные наложения. Тем не менее,
119 они плохо себя показывают в ситуации, когда на фоне присутствуют похожие цвета. Было
120 предложено множество других типов признаков для использования вместе с цветовыми.
121 В [3] показали, что комбинация цветовой модели вместе с моделью контуров позволяет
122 получить более быстрое и стабильное отслеживание объекта.

123 Для получения информации о контурах мы будем использовать гистограммы направ-
124 ленных градиентов (Histogram of Oriented Gradients, HOG). По своей природе они устой-
125 чиво себя ведут в условиях изменения освещенности и в случае схожести фона и объекта
126 по цветовой модели. Для нахождения границ мы сначала переводим RGB-изображение в
127 оттенки серого, а затем используем операторы Собеля K_x и K_y [8]:

$$128 \quad G_x(x, y) = K_x * I(x, y), G_y(x, y) = K_y * I(x, y) \text{ (под } * \text{ понимается свертка).}$$

129 Тогда сила (резкость перехода) и ориентация границы вычисляются по формулам:

$$130 \quad \begin{aligned} S(x, y) &= \sqrt{G_x^2(x, y) + G_y^2(x, y)}, \\ \theta &= \arctan(G_y(x, y)/G_x(x, y)). \end{aligned}$$

131 Чтобы избавиться от шума, можно применить порог T к значению $S(x, y)$ (используем
132 $T = 100$). Затем границы распределяются по K корзинам в соответствии с их направле-
133 ниями, после чего значения их сил суммируются, и получается нужная нам гистограмма.
134 В оригинальной работе [1] лучше всего показало себя $K = 9$. Схожесть между шаблоном
135 и гистограммой для каждой частицы вычисляются аналогично предыдущим признакам с
136 помощью Евклидовой метрики для векторов из корзины.

137 **Локальные бинарные шаблоны.** Другим примером получения информации о
138 структуре региона изображения является извлечение локальных бинарных шаблонов
139 (Local Binary Patterns, LBP), которые в некотором смысле характеризуют текстуру изоб-
140 ражения в каждой конкретной точке, для чего и были описаны впервые в [6].

141 Главная идея данного метода состоит в извлечении локальной структуры путем срав-
142 нения интенсивности каждого пикселя с его соседями — для каждого соседа мы получим
143 число, которое будет равно 1, если интенсивность соседа больше рассматриваемого цен-
144 трального пикселя, и 0 в противном случае. В итоге, если объединить полученные зна-
145 чения, например по часовой стрелке, то текстуру в окрестности каждого пикселя будет
146 описывать вектор из 0 и 1 вроде 00010011, который и называется локальным бинарным
147 шаблоном.

148 Полученные вектора можно просуммировать по всем пикселям региона и, пронормиро-
149 вав, рассматривать их как гистограмму, определяющую текстурную характеристику всего

150 региона. Схожесть между шаблоном и гистограммой для каждой частицы вычисляются
151 аналогично предыдущим признакам.

152 **Оптимизация многократного выделения признаков.** В этом параграфе нами бу-
153 дут предложены способы оптимизации многократного выделения рассматриваемых при-
154 знаков из различных регионов одного изображения.

155 1) Для оптимизации подсчета цветových гистограмм для регионов мы на основе значения
156 каждого канала всего текущего кадра видео создадим 8 бинарных изображений, в каж-
157 дом из которых значения будут характеризовать попадание интенсивности в нужную
158 корзину гистограммы. Таким образом каждое значение интенсивности цветового кана-
159 ла «оставит след» ровно на одном из 8 изображений, которые мы и назовем исходными
160 для получения интегральных изображений. Затем уже на основе этих интегральных
161 изображений можно производить оптимальный расчет гистограмм для любых регио-
162 нов в кадре — каждый элемент вектора гистограммы получается на основе соответ-
163 ствующих интегральных изображений за несколько простейших операций.

164 2) Для признаков Хаара все гораздо проще — нам требуется создать только три инте-
165 гральных изображения на основе цветových каналов изображения из текущего кадра
166 видео, с помощью которых для каждого прямоугольника Хаара каждого региона, от-
167 вечающего частице, подсчет нужных характеристик будет производиться за несколько
168 простейших операций. Именно для этого интегральные изображения вводились в [9].

169 3) Для эффективного вычисления гистограммы направленных градиентов подобно слу-
170 чаю «обычных» гистограмм для всего изображения можно построить K интегральных
171 изображений на основе K исходных, каждое из которых вместо значений интенсивно-
172 стей будет содержать либо значение силы границы, если направление совпадает с од-
173 ним из девяти имеющихся, либо 0 в противном случае — т.о. все значения мощностей
174 границ разойдутся по 9-ти исходным изображениям. Элементы вектора гистограм-
175 мы региона затем получаются на основе сумм значений интенсивности в регионе для
176 каждого из полученных исходных изображений, которые вычисляются за константное
177 время на основе интегральных изображений.

178 4) Так же как и в случае с гистограммами направленных градиентов, значения для ло-
179 кальных бинарных шаблонов можно разделить на 8 отдельных корзин, для каждой из
180 которых построить сначала исходное изображение, а затем интегральное, и с помощью
181 рассмотренных выше приемов для каждого региона получать нужные гистограммы за
182 константное время.

183 С помощью описанных выше оптимизаций удастся избежать затратного многократно-
184 го извлечения признаков из регионов для каждой частицы. Так, например, для регионов,
185 занимающих 10% площади кадра мы рассчитываем один раз признаки для всего изоб-
186 ражения, которое в 10 раз больше региона, но без оптимизаций для 1000 частиц мы бы
187 производили 1000 таких расчетов для самого региона, т.е. в среднем в 100 раз больше
188 операций.

189 **Вычислительные эксперименты**

190 Все рассмотренные алгоритмы с учетом различных признаков были реализованы на
191 языке C++. Для высокоуровневой работы с изображениями и видео использовалась биб-
192 лиотека `OpenCV` — в частности, с помощью нее происходила загрузка видео и разбор по
193 кадрам на отдельные изображения, для которых в свою очередь применялись встроен-
194 ные функции для операторов Собеля (`cv::Sobel`), подсчета интегральных изображений
195 (`cv::integral`) и проверки на попадание в нужный интервал (`cv::inRange`).

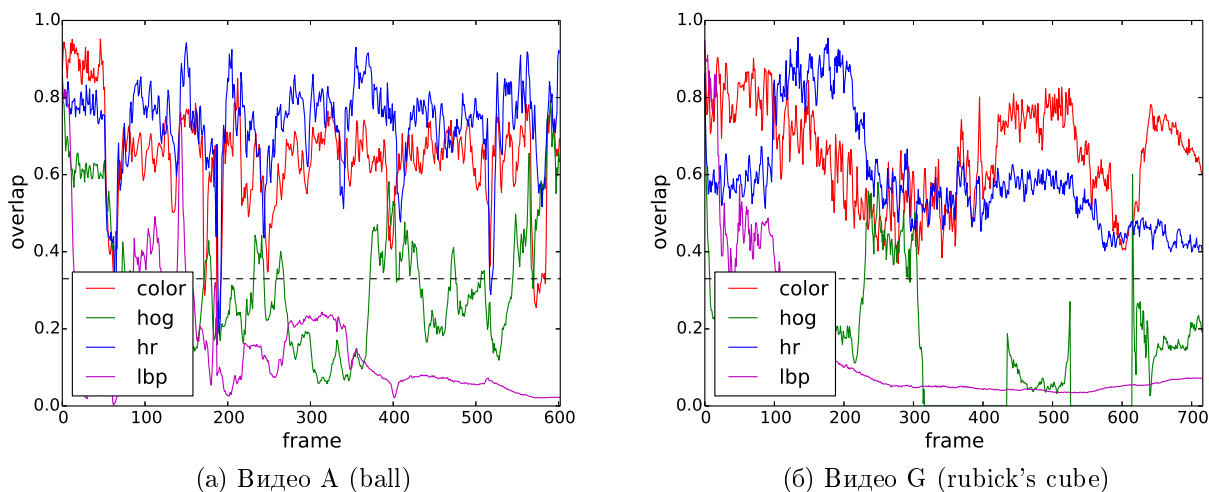
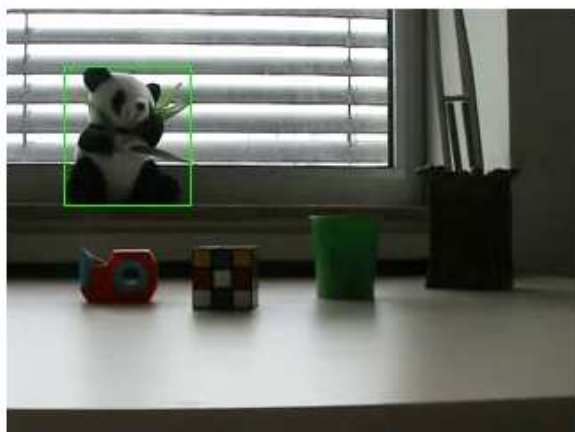


Рис. 2. Графики качества для алгоритмов на основе одной группы признаков.

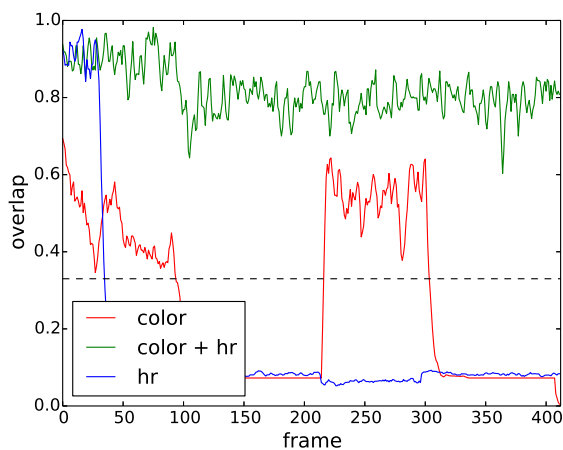
196 **Данные для экспериментов.** Для вычислительных экспериментов использовался
 197 датасет ВоВоТ, содержащий около десятка видео размера 320×240 , каждое из которых
 198 отвечает тем или иным сложностям, возникающим при трекинге объектов, как то слож-
 199 ный неоднородный фон, изменяющееся освещение, перекрытие объектов, сильные пере-
 200 мещения объекта и/или камеры. Он доступен для скачивания на сайте одного из авторов
 201 работы [4] (<http://www.iai.uni-bonn.de/~kleind/tracking/>), где также предоставлен
 202 исходный код программы на Java для оценки качества на основе файлов истинной раз-
 203 метки и разметки, получаемой алгоритмами.

204 **Используемая мера качества.** Для того, чтобы определять, насколько выделен-
 205 ный нами регион похож на реальный регион, соответствующий отслеживаемому объекту,
 206 нужно учитывать не только то, насколько близок центр рассчитанного нами выделения
 207 к реальному центру отслеживаемого объекта, но и то, насколько велика разница между
 208 реальными и вычисленными размерами объекта. Для экспериментов в данной работе ис-
 209 пользовалась мера качества, предложенная в той же работе [4], в которой был представлен
 210 используемый датасет, определяющая долю пересечения двух областей в их объединении.
 211 Так же как и в упомянутой работе, будем считать, что качеству выше 33% соответствует
 212 правильное определение положения объекта.

213 **Результаты экспериментов.** В первой части экспериментов были рассмотрены ал-
 214 горитмы, использующие только одну характерную группу признаков. На рис. 2 изображе-
 215 ны примеры графиков с распределением качества по кадрам для алгоритмов, который для
 216 подсчета правдоподобия используют гистограммы цветов (`color`), гистограммы направ-
 217 ленных градиентов (`hog`), прямоугольники Хаара (`hr`) или локальные бинарные шаблоны
 218 (`lbp`). Пунктирными линиями отображен тот порог, по которому мы определяем успеш-
 219 ность отслеживания объекта в данный момент времени. Из приведенных графикам можно
 220 сделать выводы, что самодостаточными для трекинга признаками являются цветные ги-
 221 стограммы и прямоугольники Хаара. Лишь в некоторых видео значимый результат также
 222 показывали гистограммы градиентов. Использование локальных бинарных шаблонов ви-
 223 дится осмысленным только в композиции с другими признаками.



(а) Кадр из видео Н



(б) График качества для видео Н

Рис. 3. Пример эффекта от объединения двух признаков в композицию.

224 При использовании двух групп признаков `color` и `hr` значимые отличия в результатах
 225 проявились в анализе двух видео. На видео В с кружкой на очень пестром и разнообразном
 226 по текстуре фоне оба алгоритма на основе этих признаков теряют на некоторое время
 227 цель, но в разное время — один алгоритм находит похожую на кружку синюю сплошную
 228 часть доски, другой видит похожие с шаблонными перепады в цветах на графиках и
 229 фотографиях на доске. На видео Н признаки Хаара быстро находят схожую по перепадам
 230 область и там же и остаются, тогда как цветовые гистограммы находят объект только при
 231 фиксированном типе освещения.

232 Т.к. оба типа признаков имеют схожую вычислительную сложность, было решено срав-
 233 нить их качество по отдельности с композицией, а затем исследовать, как улучшает каче-
 234 ство композиции добавление двух оставшихся «несамодостаточных» признаков.

235 Использование композиции позволяет нам брать лучшее от обеих групп признаков, что
 236 подтверждается экспериментами — на видео В при использовании композиции мы нико-
 237 гда не уходим от отслеживаемой кружки к схожим частям фона и добиваемся 100%-го
 238 качества в смысле отношения длины удачного трекинга к общей длине видео. На видео Е
 239 композиция лучше справляется с перекрытием объекта, и не захватывает перекрывающий
 240 объект в качестве предполагаемого. Самый сильный же эффект от использования ком-
 241 позиции достигается на видео Н — мы получаем 100%-е качество трекинга с постоянным
 242 пересечением с реальной областью на уровне 80% при том, что каждый признак сам по
 243 себе совсем не справлялся с трекингом на данном видео (см. рис. 3).

244 После добавления к изученной композиции «структурных» признаков, учитывающих
 245 структуру и текстуру объекта, на видео С стабильно лучше себя показали алгоритмы с
 246 добавлением LBP-признаков, что привело к очень высокому качеству трекинга для такой
 247 сложной задачи с значительными движением камеры и изменением масштаба объекта. На
 248 видео G добавление HOG уменьшило пересечения в конце где-то на четверть, что привело
 249 к вылету из минимальной зоны 33%-го качества (см. рис. 4).

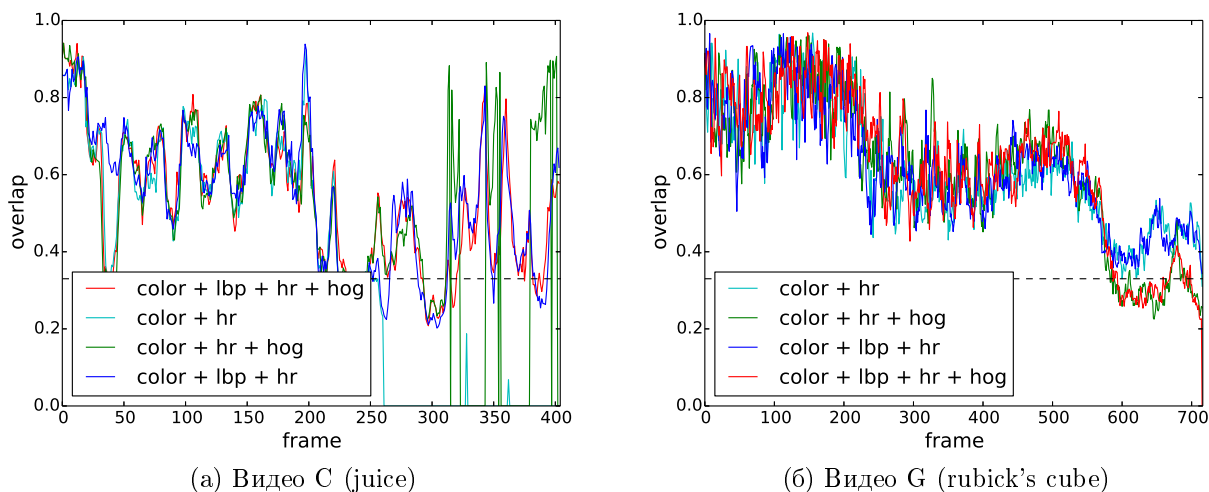


Рис. 4. Графики качества после добавления «структурных» признаков.

Таблица 1. Сравнительная таблица качества всех рассмотренных алгоритмов.

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	0.836	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

250 Заключение

251 В работе изучены способы выделения признаков из изображений для задачи трекинга
 252 объектов на видео, а также приведены способы их оптимального многократного подсчета,
 253 построены алгоритмы с использованием композиций на их основе.

254 Экспериментально показано, что использование композиций признаков позволяет по-
 255 лучить идеальные результаты с точки зрения доли времени удачного определения поло-
 256 жения объекта на видео даже там, где каждый из признаков в отдельности с этой задачей
 257 справиться не может (см. рис. 3).

258 Из проведенных экспериментов можно также заключить, что наиболее универсаль-
 259 ным в рамках рассматриваемых видео оказался композиционный алгоритм, основанный
 260 на использовании цветовых интегральных признаков, цветовых гистограмм и локальных
 261 бинарных шаблонов. Тем не менее, простая комбинация из двух цветовых групп признаков
 262 также дает высокие результаты, компенсируя недостатки каждой группы в отдельности.

263 В таблице 1 приведено сравнение результатов экспериментов для всех рассмотренных
 264 алгоритмов с указанием их качества в смысле отношения количества моментов удачного
 265 отслеживания объектов к общему количеству кадров в видео.

266 С помощью рассмотренных композиций признаков достигнуто качество трекинга, срав-
 267 нимое с более продвинутыми методами, основанными на построении сложных ансамблей
 268 с помощью бустинга [4], и превышающее результаты схожей работы [7] с использованием
 269 метода каскадов.

Литература

- 270
- 271 [1] *Dalal N., Triggs B.* Histograms of oriented gradients for human detection // *Conference on*
272 *Computer Vision and Pattern Recognition*, 2005. Pp. 886–893.
- 273 [2] *Fox D., Thrun S., Dellaert F., Burgard W.* Particle filters for mobile robot localization //
274 *Sequential Monte Carlo Methods in Practice*, New York: Springer Verlag, 2000.
- 275 [3] *Isard M., Blake A.* Icondensation: Unifying low-level and high-level tracking in a stochastic
276 framework // *Proceedings of the 5th European Conference on Computer Vision, Volume I.*, 1998.
277 Pp. 893–908.
- 278 [4] *Klein D.A., Schulz D., Frintrop S., Cremers A. B.* Adaptive real-time video-tracking for arbitrary
279 objects // *IEEE International Conference on Intelligent Robots and Systems (IROS)*, Oct 2010.
280 Pp. 772–777.
- 281 [5] *Li H., Xiong S., Duan P., Kong X.* Multitarget tracking of pedestrians in video sequences based
282 on particle filters *Advanced MultiMedia, Vol. 2012*, Jan 2012.
- 283 [6] *Pietikäinen M., Ojala T., Xu Z.* Performance evaluation of texture measures with classification
284 based on kullback discrimination of distributions // *12th IAPR International Conference on*
285 *Pattern Recognition*, 1994. Pp. 582–585.
- 286 [7] *Samuelsson O.* Video Tracking Algorithm for Unmanned Aerial Vehicle Surveillance // *Master's*
287 *Degree Project at KTH Electrical Engineering*, Stockholm, June 2012.
- 288 [8] *Sobel I., Feldman G.* A 3x3 Isotropic Gradient Operator for Image Processing // *Never published*
289 *but presented at a talk at the Stanford Artificial Project*, 1968.
- 290 [9] *Viola P., Jones M.* Rapid object detection using a boosted cascade of simple features //
291 *Conference on Computer Vision and Pattern Recognition*, 2001. Pp. 511–518.
- 292 [10] *Xu M., Orwell J., Jones G.* Tracking football players with multiple cameras // *IEEE International*
293 *Conference on Image Processing*, Los Alamitos: IEEE Computer Society Press, 2004. Pp. 2909–
294 2912