

АЛГОРИТМ УТОЧНЕНИЯ ГРАНИЦ ОБЪЕКТА ПРИ ИНИЦИАЛИЗАЦИИ ПРОЦЕССА СЛЕЖЕНИЯ С БЕСПИЛОТНОГО ЛЕТАТЕЛЬНОГО АППАРАТА

Аглямутдинова Д.Б.* , Сидякин С.В.**

*Государственный научно-исследовательский институт авиационных систем,
ГосНИИАС, ул. Викторенко, 7, Москва, 125319, Россия*

** e-mail: aglyamutdinova.d@yandex.ru*

*** e-mail: sersid@bk.ru*

Настоящая работа посвящена проблеме слежения за выбранным объектом с беспилотного летательного аппарата (БПЛА). Предложен подход к уточнению граничного прямоугольника на основе алгоритма выделения объекта внимания по ключевым признакам. Проведено исследование существующих алгоритмов слежения за объектами с БПЛА на самой большой и сложной базе видеороликов UAV 123. Показано, что качество прослеживания в результате инициализации предложенным алгоритмом изменяется в допустимых пределах для успешного слежения за объектом. По результатам исследования рекомендован ряд подходов к слежению, устойчивых к начальной инициализации предложенным алгоритмом и применимых в системах слежения в режиме реального времени, установленных на БПЛА.

Ключевые слова: слежение за объектами, инициализация алгоритмов прослеживания, беспилотные летательные аппараты.

Введение

Автоматическое слежение за объектом является одной из основных задач наряду с обнаружением и распознаванием объектов, решаемых БПЛА и может быть полезно как для определения положения прослеживаемого объекта, так и для планирования и формирования оптимальной траектории полета БПЛА в трехмерном пространстве.

В задачах визуального слежения (трекинга) критически важным моментом является начальная инициализация. От нее во многом зависит качество прослеживания, так как грубо определенное положение и размер объекта интереса на первом кадре способствует скорому срыву процедуры прослеживания. В результате неточной инициализации информация о фоновой обстановке может попасть в модель объекта, в другом случае важные части объекта интереса могут быть отброшены.

Точная инициализация особенно затруднительна в режиме реального времени, когда одновременно движутся и объект интереса, и камера, с которой осуществляется видеонаблюдение. При слежении за объектом с беспилотного летательного аппарата ситуация дополнительно осложняется тем, что с момента захвата объекта оператором, находящимся на земле, до момента получения данных сервером, установленным на БПЛА, может прой-

ти немало времени и позиция объекта в кадре может существенно измениться.

Оператор вынужден действовать быстро, поэтому, как правило, выделение объекта им сводится к грубому обведению ограничивающим прямоугольником (часто, квадратом). Когда информация о размере и положении объекта доходит до трекера, он быстро уточняет реальный размер и положение объекта на том же кадре, на котором он был захвачен оператором, после чего прослеживает объект на всех кадрах, прошедших за время получения информации от оператора, нагнав реальное время. Соответственно алгоритм слежения должен быть достаточно быстрым, чтобы компенсировать разницу во времени, и потенциально устойчивым к незначительным огрехам в инициализации.

Для поиска положения и размера прослеживаемого объекта могут быть использованы алгоритмы семантической сегментации. Сегментация стремится разбить изображение на семантически независимые регионы. Но такая информация является избыточной в данной ситуации, так как нас интересует только область, содержащая прослеживаемый объект. Вдобавок алгоритмы семантической сегментации достаточно медленные.

Поскольку оператор захватывает область, уже содержащую объект интереса, требуется алгоритм,

который сможет уточнить размер и положение этого объекта внутри выделенной области изображения. В этом случае могут быть применены алгоритмы, моделирующие карту зрительного внимания человека, известные в международной литературе как «saliency detection». Термин «saliency» какого-либо объекта на изображении или одиночного пикселя — это состояние или качество, которым он отличается от своих соседей. Методы, обнаруживающие подобные состояния, используют принципы, которые позволяют человеку сосредоточить свои ограниченные перцептивные и когнитивные ресурсы на наиболее важном подмножестве имеющихся сенсорных данных. В области компьютерного зрения исследуются различные модели внимания, в частности механизм внимания «снизу-вверх» [11]. Такой процесс еще называется визуальным выявлением признаков («visual saliency detection»). В основном рассматриваются два вида моделей для имитации механизма внимания «снизу-вверх». Первый подход основан на анализе пространственного контраста. Например, выносится предположение о том, что зрительное предпочтение отдается предметам, лежащим в центральной области [15]. Другой путь сформирован на анализе частотной области [16—18]. В работе [16] использовался амплитудный спектр для назначения значимости редко встречающимся величинам, в [17] — фазовый спектр; в [18] была предложена система, использующая как амплитудный, так и фазовый спектр.

В связи с прогрессом сверточных нейронных сетей появился и третий подход. В отличие от классических методов, основанных на ключах контрастности, методы, основанные на CNN, отменяют необходимость в ручных инструментах, облегчают зависимость от знаний о смещении центра и, следовательно, принимаются многими исследователями. К сожалению, известным ограничением нейросетевых алгоритмов являются высокие требования к вычислительным мощностям и аппаратному обеспечению [6]. Оборудование, обладающее высокой вычислительной мощностью, стоит дорого, кроме того следует учитывать вес оборудования при анализе возможности его использования на БПЛА.

В данной работе предлагается быстрый алгоритм уточнения положения объекта на кадре, пригодный для использования на БПЛА. Предложенный подход не требует специального оборудования (вроде графических ускорителей) и может работать на маломощных системах на кристалле (SoC, FPGA) в режиме реального времени. Дополнительно исследуется устойчивость методов прослеживания к начальной инициализации предложенным алгорит-

мом на примере самой большой и сложной базы видеороликов, снятых с БПЛА — UAV 123 [21].

Алгоритм уточнения границ объекта при инициализации процесса слежения

В качестве входных данных для алгоритма слежения предоставляется область изображения, выделенная оператором для уточнения объекта. После получения области изображения ее необходимо преобразовать с целью удаления шума, искажающего границы выделяемых объектов.

Предварительная обработка изображения

С задачей шумоподавления хорошо справляются сглаживающие фильтры. Существует множество линейных и нелинейных алгоритмов сглаживания, которые включают в себя удаление шума, стабилизацию яркости, повышение контрастности и четкости. Популярным методом сглаживания является фильтр Гаусса, благодаря которому можно уменьшить влияние пикселей друг на друга в зависимости от расстояния. Ядро фильтра Гаусса выражается формулой

$$F(i, j) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{i^2 + j^2}{2\sigma^2}\right),$$

где i, j — координаты пикселя изображения; σ — шум, находящийся на изображении. Размер ядра свертки 3×3 пикселя.

Сглаженное изображение конвертировалось в цветовое пространство CIE LAB, которое представляет собой цвета в виде набора из трех компонент: L — светлота, A — красно-зеленый оттенок, B — желто-синий. Полученные колориметрические значения удобны как для визуальной оценки положения цвета в пространстве, светлоты и насыщенности цвета, так и для численного сравнения различных цветов и определения цветоразличия. В ходе тестирования алгоритма также использовалось цветовое пространство RGB, но результаты CIE LAB оказались лучше.

С целью повышения быстродействия и робастности алгоритма перейдем от дискретных пикселей к областям (суперпикселям) путем сегментации.

Сегментация изображения

Среди методов сегментации, таких, как метод k-средних [22], метод водораздела [23], метод разреза графов [23], выделяется метод простой линейной итеративной кластеризации (Simple Linear Iterative Clustering, SLIC) [1], так как он способен разбить исходное изображение на различные, но внутри себя в некотором смысле однородные области за приемлемое количество времени.

Алгоритм сегментации SLIC:

1. Разбиение изображения на k фрагментов размера $a \times a$ с центрами C_i .

2. Коррекция координат C_i из условия $\nabla I(C_i) \rightarrow \min$.

3. Формирование локальных кластеров в окрестности.

Расстояние между фрагментом p с цветовыми компонентами l, a, b и координатами x, y и фрагментом C_i :

$$d_{Lab} = (l - C_l)^2 + (a - C_a)^2 + (b - C_b)^2,$$

$$d_{xy} = (x - C_x)^2 + (y - C_y)^2,$$

$$dist = d_{Lab} + d_{xy} \cdot inwt.$$

Пусть $sSize$ — размер суперпикселей, тогда

$$inwt = \left(\frac{compactness}{\sqrt{sSize}} \right)^2,$$

где $compactness$ — мера «компактности сегментов», т. е. то, насколько они по форме близки к квадрату.

4. Определение новых центров.

5. Повторение шагов 3 и 4 определенное количество раз.

В данном алгоритме используются следующие входные параметры: $compactness = 20$, $sSize = 500$.

Результатом выполнения процедуры сегментации является распределение пикселей по суперпикселям. В качестве контейнера для хранения информации о сегментах используется неориентированный граф G . Вершинами этого графа являются суперпиксели. В каждой вершине хранится информация о средних значениях цветовых компонент, координат и о том, является ли данный сегмент граничным. Весом ребра, соединяющего две соседние вершины, является евклидово расстояние между средними цветовыми компонентами данных вершин.

Далее необходимо определить меру принадлежности каждого суперпикселя к фону и объекту.

Вычисление меры принадлежности суперпикселя к фону

В основе определения степени принадлежности суперпикселя к фону лежит следующая идея: наиболее значимый объект находится в середине изображения, а объекты, которые имеют наиболь-

ший периметр на границе, скорее всего являются фоном [2].

Используем геодезическое расстояние, определяемое как кратчайший путь на графе между двумя суперпикселями и вычисляемое по алгоритму Джонсона.

Пусть дано множество суперпикселей S , $|S| = N$ и B — множество граничных суперпикселей, $B \in S$, $d_{app}(p, q) = dist(p, q)$ — компонента, отвечающая за цветовую контрастность, т.е. евклидово расстояние между средними цветовыми компонентами суперпикселей p и q .

Вычислим геодезическое расстояние по следующей формуле:

$$d_{geo}(p, q) = \min_{p=p_1, \dots, p_n=q} \sum_{i=1}^{N-1} d_{app}(p_i, p_{i+1}),$$

где $p, q \in S$.

Длина периметра, граничащего с краем картинки:

$$Len_{bnd}(p) = \sum_{q \in B} \exp \left(\frac{-d_{geo}(p, q)^2}{2\delta_{clr}^2} \right),$$

где $p \in S$.

Площадь области охвата суперпикселя:

$$Area(p) = \sum_{q \in S} \exp \left(\frac{-d_{geo}(p, q)^2}{2\delta_{clr}^2} \right),$$

где $p \in S$.

Тогда мера принадлежности к фону для суперпикселя p рассчитывается как

$$w_{bg}(p) = \frac{Len_{bnd}(p)}{\sqrt{Area(p)}},$$

где $p \in S$.

Вычисление меры принадлежности суперпикселя к объекту

В качестве меры принадлежности к объекту используется контрастирование, взвешенное по фону («Background weighted contrast») [2].

Мера определяется следующим выражением:

$$w_{ctr}(p) = \sum_{p_i \in S} d_{app}(p, p_i) w_{spa}(p, p_i) w_{bg}(p_i),$$

где $p \in S$.

Пространственная компонента

$$w_{spa}(p, q) = \exp\left(\frac{-d(p, q)^2}{2\delta_{spa}^2}\right),$$

где d — евклидово расстояние между средними цветовыми компонентами суперпикселей p и q .

Зная меру принадлежности, можно охарактеризовать каждый суперпиксель, а значит, и каждый обычный пиксель с помощью меры значимости. Чем больше мера значимости пикселя, тем больше вероятность, что он принадлежит к объекту.

Вычисление меры значимости

Для вычисления меры значимости используется подход [2], комбинирующий меру фона и меру объекта. Пусть значимость суперпикселей будет равна $\{s_i\}_{i=1}^N$. Тогда функция стоимости будет определена как

$$\underbrace{\sum_{i=1}^N w_i^{bg} s_i^2}_{\text{фон}} + \underbrace{\sum_{i=1}^N w_i^{fg} (s_i - 1)^2}_{\text{объект}} + \underbrace{\sum_{i,j} w_{ij} (s_i - s_j)^2}_{\text{гладкость}},$$

где w_i^{bg} — мера принадлежности к фону $p_i \in S$; w_i^{fg} — мера принадлежности к переднему плану (объекту), в качестве которого используется контрастирование, взвешенное по фону («Background weighted contrast») [2]; w_{ij} — вес, определенный между смежными суперпикселями.

Компонента гладкости отвечает за непрерывность функции стоимости:

$$w_{ij} = \exp\left(-\frac{d_{app}^2(p_i, p_j)}{2\delta_{nei}^2}\right) + \mu.$$

Для нахождения оптимального значения $\{s_i\}_{i=1}^N$ используется метод наименьших квадратов.

Вычислим меру значимости, решив уравнение

$$Ax = w_{ctr}(i);$$

$$A = D - W + E^{bg} + E^{fg}.$$

Компонента D представляет сумму весов, соединяющих каждый суперпиксель со всеми остальными:

$$D_{ij} = \begin{cases} weight(i), & i = j; \\ 0, & i \neq j. \end{cases}$$

Компонента W определяет вес, соединяющий суперпиксели p_i и p_j :

$$W_{ij} = w(p_i, p_j).$$

Компонента E^{bg} определяет меру принадлежности к фону:

$$E_{ij}^{bg}(i, j) = \begin{cases} w_{bg}(i) \cdot \lambda_{bg}, & i = j; \\ 0, & i \neq j. \end{cases}$$

Компонента E^{fg} определяет меру переднего плана или объекта:

$$E_{ij}^{fg}(i, j) = \begin{cases} w_{ctr}(i), & i = j; \\ 0, & i \neq j; \end{cases}$$

$$weight(i) = \sum_{p_j \in S} w(p_i, p_j);$$

$$w(p, q) = \exp\left(\frac{-d_{app}(p, q)^2}{2\delta_{nei}^2}\right) + adjacent(p, q) \cdot \mu;$$

$$adjacent(p, q) = \begin{cases} 1, & p, q - adjacent; \\ 0, & else. \end{cases}$$

В ходе экспериментов были подобраны следующие значения гиперпараметров:

$$\delta_{ctr} = 7, \delta_{spa} = 0.4, \lambda_{bg} = 5, \delta_{nei} = 10, \mu = 0.1.$$

В результате вычисления меры значимости в каждом пикселе генерируется карта значимости (рис. 1), в которой хорошо просматривается объект интереса. Соответственно, его размер и положение могут быть найдены с использованием представленной карты путем последовательного применения операций бинаризации, удаления тени, морфологических операций фильтрации и алгоритма выделения связанных областей. Рассмотрим каждую из этих операций.

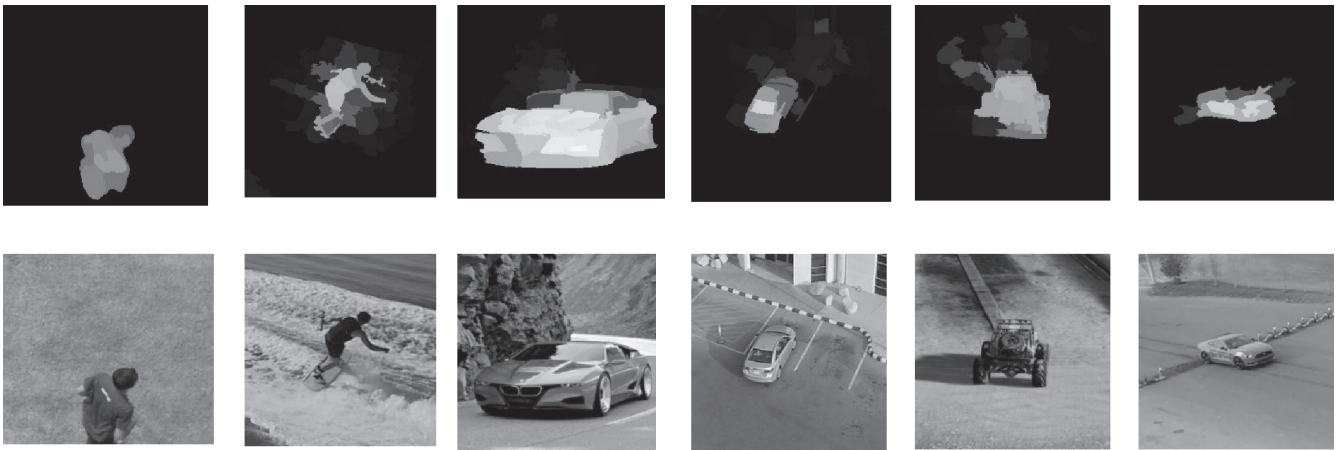


Рис. 1. В первой строке представлены карты значимости, во второй — соответствующие им изображения

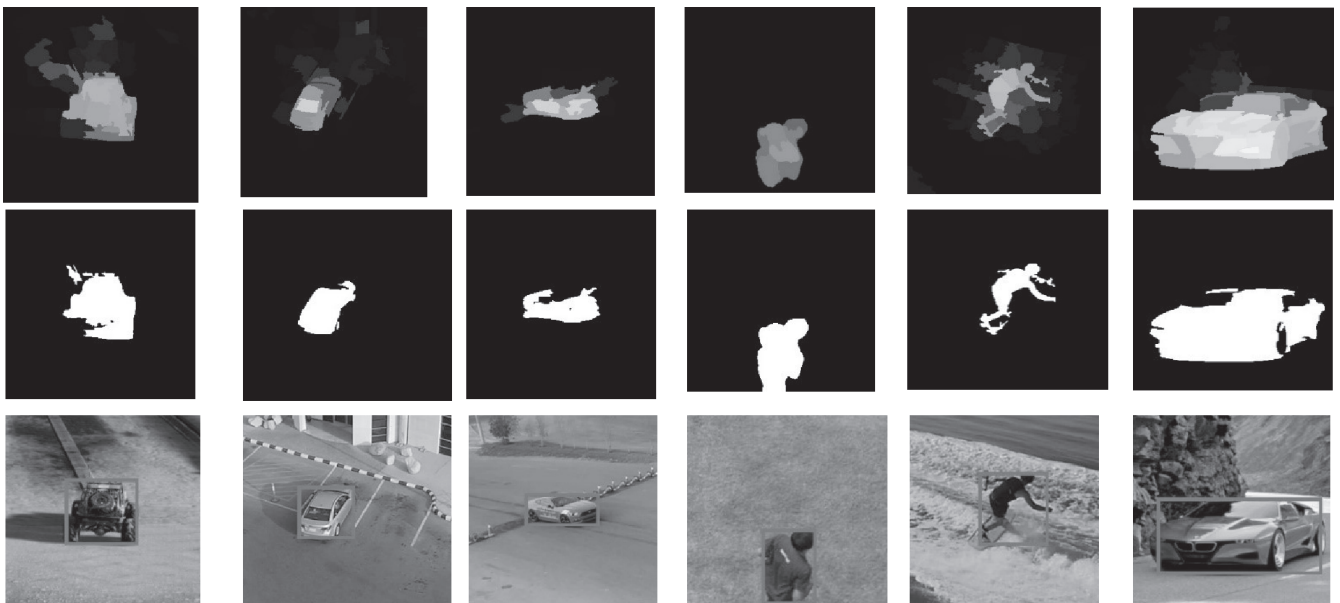


Рис. 2. В первой строке представлены карты значимости, во второй — соответствующие бинарные изображения, в последней — ограничивающие прямоугольники найденных объектов

Бинаризация результата

На предыдущем шаге была создана карта значимости исходного изображения. Каждый суперпиксель был оценен мерой принадлежности к объекту. В результате получен набор значений меры значимости, принадлежащих интервалу [0, 1], где значения, близкие к нулю, обозначают наиболее вероятную принадлежность текущего пикселя к фону, а близкие к единице — характеризуют суперпиксель как часть объекта. Для нахождения пикселей объекта необходимо бинаризовать полутоновое изображение карты значимости, получив два класса: фон и объект.

Для бинаризации (рис. 2) воспользуемся пороговой обработкой. Среди распространенных методов, применяющих пороговую обработку, часто используется бинаризация с верхним порогом, с ниж-

ним порогом и с двойным ограничением. В нашей задаче большим значениям соответствует большая вероятность принадлежности к объекту, следовательно, используем верхнюю пороговую обработку:

$$S(x, y) = \begin{cases} 1, S(x, y) \geq t; \\ 0, S(x, y) < t. \end{cases}$$

Качество разделения зависит от значения данного порога. В данном алгоритме используется метод Отсу [24].

Удаление тени с изображения

Тень, отбрасываемая объектом, нередко имеет четкие очертания и выделяется на окружающем фоне. То есть она вписывается в представление главного объекта на изображении и, будучи соеди-

ненной с объектом, может быть распознана как объект, а не часть фона, какой она и является. При определенных ракурсах тень от объекта занимает пространство на изображении, равное самому объекту, и значительно расширяет определяемую алгоритмом область. В обработке изображений существуют два принципиально разных подхода к детектированию тени:

- используются свойства тени с учетом модели освещения, формы объекта отражения и т. д.;
- учитываются свойства цвета: яркость, плотность градиента.

В данной работе для удаления тени за основу был взят метод [4], который использует яркостную статистику. Подход [4] опирается на следующие соображения:

- пиксель принадлежит к тени, если его яркость меньше 60% средней яркости всего изображения;
- пиксель принадлежит к тени, если его яркость меньше 70% от средней яркости в окне переменного размера.

Алгоритм [4] был модифицирован, т.е. вместо скользящего окна переменных размеров использовалась сегментация пикселей по суперпикселям, дополнительно учитывались данные от шага бинаризации (анализировались только суперпиксели объекта).

Один из недостатков алгоритма определения тени — ложное определение тени на изображении, где наиболее темной областью является сам объект. Во избежание ложного детектирования в этих ситуациях выдвигаются следующие условия для определения пикселей тени:

$$\begin{aligned} FullAverage \cdot \gamma_f > Y_i, \quad FgAverage \cdot \gamma_{fg} > Y_i, \\ FullAverage \cdot \gamma_{f2} < Y_i, \quad FgAverage \cdot \gamma_{fg2} < Y_i, \end{aligned}$$

где $FullAverage$ — средняя яркость для изображения; $FgAverage$ — средняя яркость суперпикселей, принадлежащих объекту; Y — яркость пикселя; $\gamma_f = 0.8$, $\gamma_{fg} = 0.8$, $\gamma_{f2} = 0.3$, $\gamma_{fg2} = 0.4$.

На рис. 3 представлены поэтапные результаты работы алгоритма, из которых можно сделать вывод об эффективности метода удаления тени. Область объекта в некоторых случаях уменьшается вдвое, увеличивая точность обнаружения цели.

Морфологические методы обработки бинарного изображения

В результате бинаризации и удаления тени могут появиться мелкие объекты, не имеющие отношения к основному объекту, или сам объект может распасться на части. Для решения данной проблемы используются операции морфологической фильтрации [20]. Мелкие фрагменты удаляются с помощью операции эрозии со структурирующим элементом в форме круга с радиусом в 2 пикселя, в результате чего все найденные объекты будут уменьшены в размере. Для восстановления форм объектов далее применяется операция дилатации с тем же структурирующим элементом, затем для соединения мелких частей в один объект используется операция дилатации с кругом радиусом в 3 пикселя.

Для определения положения и размера фильтрованного и бинаризованного объекта используется алгоритм выделения связных областей [20], который и дает финальный ограничивающий прямоугольник.

Практическая часть

Исследование предложенного алгоритма произведено на коллекции видеозаписей UAV 123 [21]. Этот DataSet содержит набор видеозаписей, на которых присутствует объект интереса, чье местоположение необходимо проследить. Разрешение изображений составляет 1920×1080 пикселей. Линейные размеры прослеживаемых объектов варьируются в пределах от 29 до 102 пикселей.

При исследовании эффективности предложенного алгоритма инициализации объект, положение и размер которого необходимо уточнить, занимал примерно 40% области, заданной оператором, и был расположен максимально близко к центру кадра.

Качество полученной инициализации оценивалось ее влиянием на процесс прослеживания путем сравнения с результатами прослеживания при эталонной инициализации. Для оценки качества слежения использовались две метрики: точность и успех.

Точность измеряется как расстояние между центром ограничивающего прямоугольника, выдаваемого процедурой слежения, и центром ограничивающего прямоугольника из заранее известной эталонной разметки. График точности показывает процент ограничивающих прямоугольников, полученных алгоритмом слежения и удаленных от действительного положения объекта не более чем на значения порога, измеряемого в пикселях кадра. Для ранжирования алгоритмов прослеживания по

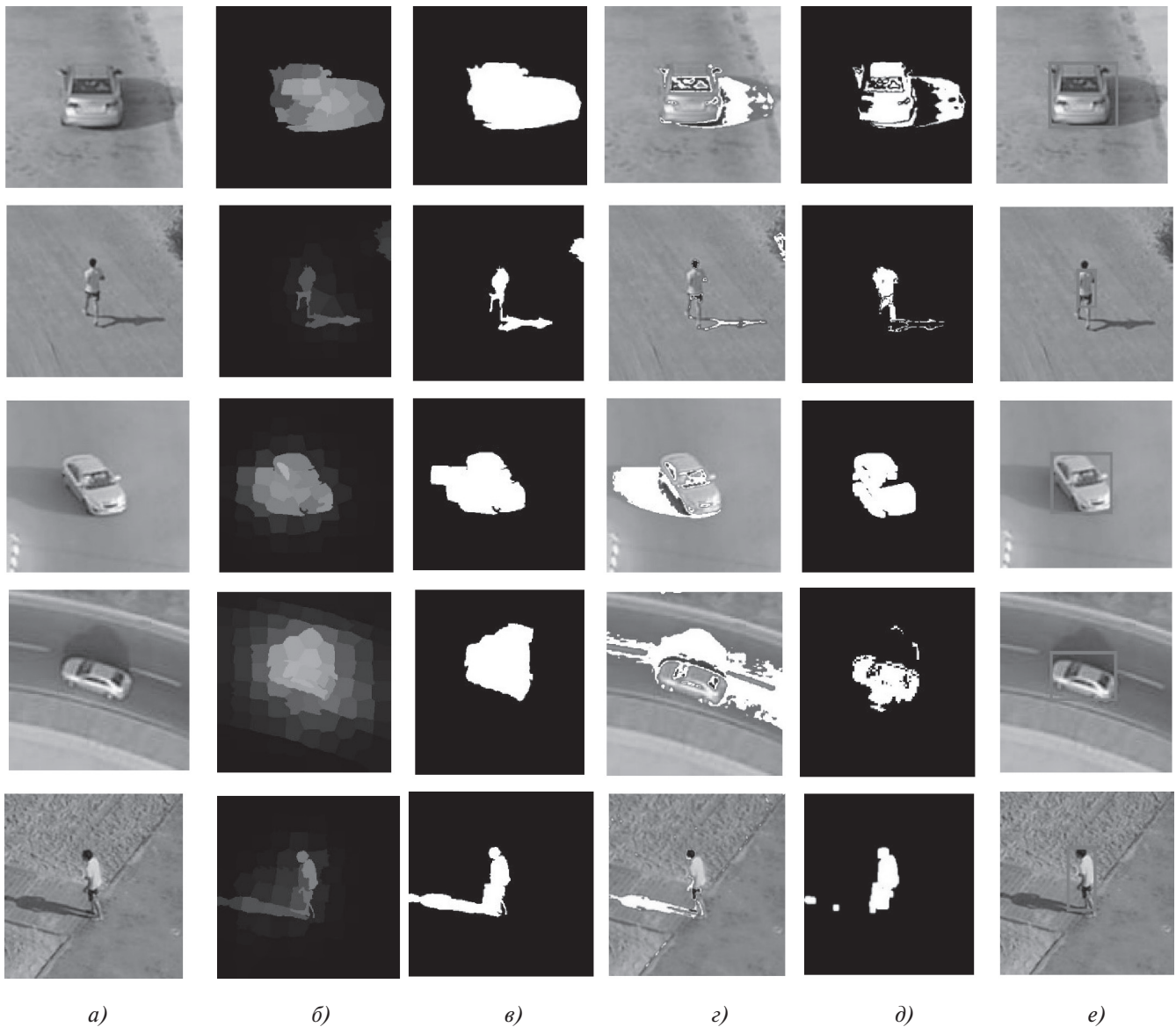


Рис. 3. Этапы выделения объекта на изображении: *a* — исходное изображение; *b* — карта значимости изображения; *в* — бинаризованное исходное изображение; *г* — изображение с выделением тени белым цветом; *д* — бинаризованное изображение после удаления тени; *е* — изображение с выделением объекта, обнаруженного алгоритмом уточнения

Таблица 1

Графики успешности и точности различных алгоритмов слежения с инициализацией предложенным алгоритмом уточнения и эталонной разметкой

Алгоритм слежения	График успеха с инициализацией предложенным алгоритмом уточнения	График точности с инициализацией предложенным алгоритмом уточнения	График успеха с эталонной инициализацией	График точности с эталонной инициализацией
DCF_CA				

показателю точности используется порог в 20 пикселей.

Успешность измеряется как мера сходства (Intersection over union, IoU):

$$S = \frac{|bb_{tr} \cap bb_{gt}|}{|bb_{tr} \cup bb_{gt}|},$$

где bb_{tr} — ограничивающий прямоугольник трекара; bb_{gt} — ограничивающий прямоугольник эталонной разметки.

График успешности показывает процент ограничивающих прямоугольников, выдаваемых процедурой слежения, мера сходства которой больше заданного порога.

Согласно результатам, представленным в табл. 1 и 2, наилучшую точность с инициализацией, выполненной программой уточнения, показывают алгоритмы слежения (трекеры) «SRDCF» и «MOSSE_CA». Наиболее успешными оказываются трекары «SRDCF», «SAMF», «DCF_CA».

Таблица 2

Графики успешности и точности различных алгоритмов слежения с инициализацией предложенным алгоритмом уточнения и с эталонной разметкой

Алгоритм слежения	График успеха с инициализацией предложенным алгоритмом уточнения	График точности с инициализацией предложенным алгоритмом уточнения	График успеха с эталонной инициализацией	График точности с эталонной инициализацией
MOSSE_CA				
DSST				
MOSSE				
SRDCF				
SAMF				

Таблица 3

Среднее количество кадров в секунду, разница между графиками точности и успешности в контрольных точках с инициализацией алгоритмом уточнения и с известной эталонной инициализацией

Название трекера	Среднее количество кадров в секунду	Разница между значениями графиков успеха при совпадении больше порогового значения 0.5	Разница между значениями графиков точности при пороговом расстоянии 20 пикселей
DCF_CA	182,2102	0,176	0,176
MOSSE_CA	271,7864	0,064	0,065
SAMF	8,5000	0,216	0,182
DCF	238,9312	0,160	0,154
DSST	130,7184	0,155	0,165
MOSSE	253,6889	0,136	0,124
SRDCF	8,9679	0,234	0,175

По данным табл. 3 можно заключить, что наиболее устойчивы к начальной инициализации алгоритмы слежения за объектом «MOSSE_CA» и «MOSSE». Наилучшее быстродействие показывают трекары «MOSSE_CA», «MOSSE» и «DCF».

Согласно полученным результатам наиболее подходящим алгоритмом слежения для задачи трекинга с БПЛА является «MOSSE_CA», который наименее чувствителен к начальной инициализации и показывает высокую скорость.

Выводы

Предложен подход к уточнению граничного прямоугольника на основе алгоритма выделения объекта внимания по ключевым признакам. Данный алгоритм не требует специального аппаратного обеспечения и может работать в условиях реального времени. Алгоритм был реализован на языке C++; среднее время, требуемое для уточнения объекта, занимающего 40% изображения размером 256×256 пикселей, равно 60 миллисекунд на процессоре Intel® Core™ i5-3470 CPU @ 3.20GHz 3.60GHz.

Проведено исследование существующих алгоритмов слежения за объектами с БПЛА на самой большой и сложной базе видеороликов UAV 123. Показано, что наилучшее качество прослеживания в результате инициализации предложенным алгоритмом достигается методами слежения «SRDCF» и «MOSSE_CA». При оценке быстродействия алгоритм слежения «MOSSE_CA» заметно выигрывает, и можно сделать вывод, что он является наиболее подходящим алгоритмом для слежения за объектами с БПЛА, как наименее чувствительный к точности начальной инициализации и более быстродействующий по сравнению с конкурирующими методами.

Работа поддержана грантом РФФ № 16-11-00082 и грантом РФФИ № 16-57-52042 МНТ_а.

Библиографический список

1. Achanta R., Shaji A., Smith K., Lucchi A., Fua P., Susstrunk S. SLIC superpixels compared to state-of-the-art superpixel methods // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2012. Vol. 34. No. 11, pp. 2274-2282. DOI: 10.1109/TPAMI.2012.120
2. Zhu W., Liang S., Wei Y., Sun J. Saliency Optimization from Robust Background Detection // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2014, pp. 2814-2821. DOI: 10.1109/CVPR.2014.360
3. Hornung A., Pritch Y., Krahenbuhl P., Perazzi F. Saliency filters: Contrast based filtering for salient region detection // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2012, pp. 733-740. DOI: 10.1109/CVPR.2012.6247743
4. Blajovici C., Kiss P.J., Bonus Z., Varga L. Shadow detection and removal from a single image // 19th Summer School on Image Processing (SSIP). 2011. Szeged, Hungary, 6 p.
5. Borji A., Cheng M.-M., Jiang H., Hou Q., Jiang H., Li J. Salient object detection: A survey // Computer Vision and Pattern Recognition. 2014, 21 p. DOI: arXiv:1411.5878
6. Кульчин Ю.Н., Ноткин Б.С., Седов В.А. Нейроитерационный алгоритм томографической реконструкции распределенных физических полей в волоконно-оптических измерительных системах // Компьютерная оптика. 2009. Т. 33. № 4. С. 446-455.
7. Довженко А.Ю., Крашаков С.А. Параллельная нейронная сеть с удаленным доступом на базе распределенного кластера ЭВМ // Тезисы докл. II Международного симпозиума «Компьютерное обеспечение химических исследований» (Москва, 22-23 мая 2001) и III Всерос. школы-конф. по квантовой и вычисл. химии им. В.А. Фока. — М., 2001. С. 52-53.

8. *Fatica M.* CUDA for High Performance Computing // Materials of HPC-NA Workshop 3 (January 2009), 40 p.
9. Belgian researchers develop desktop supercomputer // FASTRA. URL: <http://fastra.ua.ac.be/en/index.html>
10. *Wu Y., Lim J., Yang M.-H.* Online Object Tracking: A Benchmark // IEEE Conference on Computer Vision and Pattern Recognition. 2013, pp. 2411-2418. DOI: 10.1109/CVPR.2013.312
11. *Harel J., Koch C., Perona P.* Graph-based visual saliency // 19th International Conference on Neural Information Processing Systems. 2006, pp. 545-552.
12. *Achanta R., Shaji A., Smith K., Lucchi A., Fua P., Susstrunk S.* Slic superpixels compared to state-of-the-art superpixel methods // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2012. Vol. 34. No. 11, pp. 2274-2281. DOI: 10.1109/TPAMI.2012.120
13. *Achanta R., Hemami S., Estrada F., Susstrunk S.* Frequency-tuned salient region detection // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2009, 8 p. DOI: 10.1109/CVPR.2009.5206596
14. *Frintrop S., Rome E., Henrik I.C.* Computational visual attention systems and their cognitive foundations: A survey // ACM Transactions on Applied Perception. 2010. Vol. 7. No. 1, 46 p. DOI: 10.1145/1658349.1658355
15. *Itti L., Koch C., Niebur E.* A model of saliency-based visual attention for rapid scene analysis // IEEE Transactions on Pattern Analysis and Machine Intelligence. 1998. Vol. 20. No. 11, pp. 1254—1259. DOI: 10.1109/34.730558
16. *Hou X., Zhang L.* Saliency Detection: A Spectral Residual Approach // IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, MN, USA. 2007. DOI: 10.1109/CVPR.2007.383267
17. *Guo C., Ma Q., Zhang L.* Spatio-temporal Saliency detection using phase spectrum of quaternion fourier transform // IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, AK, USA. 2008, 8 p. DOI: 10.1109/CVPR.2008.4587715
18. *Li J., Levine M.D., An X., Xu X., He H.* Visual Saliency Based on Scale-Space Analysis in the Frequency Domain // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2013. Vol. 35. No. 4, pp. 996-1010. DOI: 10.1109/TPAMI.2012.147
19. *Пытьев Ю.П.* Задачи морфологического анализа изображений // Математические методы исследования природных ресурсов Земли из космоса / Под ред. В.Г. Золотухина. — М.: Наука, 1984. С. 41-83.
20. *Визильтер Ю.В., Желтов С.Ю., Бондаренко А.В., Ососков М.В., Моржин А.В.* Обработка и анализ изображений в задачах машинного зрения: Курс лекций и практических занятий. — М.: Физматкнига, 2010. — 672 с.
21. *Mueller M., Smith N., Ghanem B.* A Benchmark and Simulator for UAV Tracking // European Conference on Computer Vision (ECCV 2016), pp.445-461. DOI: 10.1007/978-3-319-46448-0_27
22. *Mirkes E.M.* K-means and K-medoids applet // University of Leicester, 2011.
23. *Grady L.* Random Walks for Image Segmentation // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2006, pp. 1768-1783. DOI: 10.1109/TPAMI.2006.233
24. *Otsu N.* A threshold selection method from gray-level histograms // IEEE Transactions on Systems, Man, and Cybernetics. 1979. Vol. 9. No. 1, pp. 62-66. DOI: 10.1109/TSMC.1979.4310076
25. *Сидякин С.В., Егоров А.И., Малин И.К.* Распознавание движущихся человеческих фигур на видеопоследовательности с использованием дисковой карты толщин // Вестник Московского авиационного института. 2014. Т. 21. № 5. С. 157-166.
26. *Мирзоян А.С., Малышев О.В., Хмаров И.М., Канивец В.Ю.* Распознавание летательных аппаратов оптической системой в реальном масштабе времени // Вестник Московского авиационного института. 2014. Т. 21. № 5. С. 145-156.

AN OBJECT BOUNDING BOX REFINEMENT ALGORITHM WHILE THE TRACKING PROCESS INITIALIZATION FROM THE UAV

Aglyamutdinova D.B.* , Sidyakin S.V.**

*State Institute of Aviation Systems,
7, Viktorenko str., Moscow, 125319, Russia*

** e-mail: aglyamutdinova.d@yandex.ru*

*** e-mail: sersid@bk.ru*

Abstract

The presented article deals with the problem of semi-automatic initialization of the selected object tracking by unmanned aerial vehicles (UAVs) or drones. Here, we proposed an algorithm of the position and sizes refining of the boundary rectangle of the tracked object

at the start time (on the first frame) based on saliency detection algorithm, which simulates the map of human attention. The advantage of the proposed approach is that it applies the principles used by the human visual system: the color contrast, the main attention is centered on the central objects. The first stage of the proposed

approach consists in preliminary image processing (noise removal) by the Gaussian filter and converting the image into the CIE LAB color space. The next stage is segmenting the image into homogeneous areas (superpixels) by simple linear iterative clustering (SLIC) algorithm. Undirected graph is employed as a container for information on segments storage. Based on information from the resulting graph, measures of identity, which assign superpixels to the background or an object, are computed. The resulting saliency measure is computed for each superpixel by optimizing the target cost function, which combines the measures of identity to the background, an the object and the smoothing component. The obtained saliency map of the image superpixels is binarized by the Otsu method. After that, the pixels belonging to the shadow can be determined. At the final stage, the operations of morphological filtering were applied to reduce fragmentation of objects and an algorithm for allocating coherent components, assigning the final dimensions and position of the object of interest for tracking initialization.

The algorithm was used to initialize a number of fast and effective methods of object tracking: DCF_CA, MOSSE_CA, SAMF, DCF, DSST, MOSSE, SRDCF. At the same time, the quality of the tracking was tested on the largest and most complex database of video clips, shot from an unmanned aerial vehicle - UAV 123.

The results of experimental testing allow conclude that the best tracking quality as a result of initialization by the proposed algorithm is achieved by tracking algorithms “SRDCF” and “MOSSE_CA”. In assessing the performance, you can notice that “MOSSE_CA” tracking algorithm is noticeably superior to the other algorithms. In this way, the most suitable algorithm for tracking objects by UAV, along with the proposed initialization algorithm, is “MOSSE_CA”, due to its least sensitivity to the of initial initialization accuracy and fastness among competitors.

The proposed algorithm does not require special hardware and can work in real-time. It is implemented in C++. The average time required refining the object, occupying 40% of the image size of 256×256 pixels, equals 60 milliseconds on the Intel® Core™ i5-3470 CPU @ 3.20GHz.

Keywords: objects tracking, tracking algorithms initialization, unmanned aerial vehicle (UAV)

References

1. Achanta R., Shaji A., Smith K., Lucchi A., Fua P., Sussstrunk S. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, vol. 34, no. 11, pp. 2274-2282. DOI: 10.1109/TPAMI.2012.120

2. Zhu W., Liang S., Wei Y., Sun J. Saliency Optimization from Robust Background Detection. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 2814-2821. DOI: 10.1109/CVPR.2014.360

3. Hornung A., Pritch Y., Krahenbuhl P., Perazzi F. Saliency filters: Contrast based filtering for salient region detection. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2012, pp. 733-740. DOI: 10.1109/CVPR.2012.6247743

4. Blajovici C., Kiss P.J., Bonus Z., Varga L. Shadow detection and removal from a single image. *19th Summer School on Image Processing (SSIP)*, 2011, Szeged, Hungary, 6 p.

5. Borji A., Cheng M.-M., Jiang H., Hou Q., Jiang H., Li J. Salient object detection: A survey. *Computer Vision and Pattern Recognition*, 2014, 21 p. DOI: arXiv:1411.5878

6. Kul'chin Yu.N., Notkin B.S., Sedov V.A. *Komp'yuternaya optika*, 2009, vol. 33, no. 4, pp. 446-455.

7. Dovzhenko A.Yu., Krashakov S.A. *II Mezhdunarodnyi simpozium "Komp'yuternoe obespechenie khimicheskikh issledovaniy" (Moskva, 22-23 May 2001) i III Vseros. shkoly-konferensii po kvantovoi i vychisl. khimii im. V.A. Foka*, pp. 52-53.

8. Fatica M. CUDA for High Performance Computing. *Materials of HPC-NA Workshop 3* (January 2009), 40 p.

9. *Belgian researchers develop desktop supercomputer*, <http://fastra.ua.ac.be/en/index.html>

10. Wu Y., Lim J., Yang M.-H. Online Object Tracking: A Benchmark. *IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 2411-2418. DOI: 10.1109/CVPR.2013.312

11. Harel J., Koch C., Perona P. Graph-based visual saliency. *19th International Conference on Neural Information Processing Systems*, 2006, pp. 545-552.

12. Achanta R., Shaji A., Smith K., Lucchi A., Fua P., Sussstrunk S. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, vol. 34, no. 11, pp. 2274-2281. DOI: 10.1109/TPAMI.2012.120

13. Achanta R., Hemami S., Estrada F., Sussstrunk S. Frequency-tuned salient region detection. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009, 8 p. DOI: 10.1109/CVPR.2009.5206596

14. Frintrop S., Rome E., Henrik I.C. Computational visual attention systems and their cognitive foundations: A survey. *ACM Transactions on Applied Perception*, 2010, vol. 7, no. 1, 46 p. DOI: 10.1145/1658349.1658355

15. Itti L., Koch C., Niebur E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, vol. 20, no. 11, pp. 1254-1259. DOI: 10.1109/34.730558

16. Hou X., Zhang L. Saliency Detection: A Spectral Residual Approach. *IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, MN, USA*, 2007. DOI: 10.1109/CVPR.2007.383267

17. Guo C., Ma Q., Zhang L. Spatio-temporal Saliency detection using phase spectrum of quaternion fourier

- transform. *IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, AK, USA*, 2008, 8 p. DOI: 10.1109/CVPR.2008.4587715
18. Li J., Levine M.D., An X., Xu X., He H. Visual Saliency Based on Scale-Space Analysis in the Frequency Domain. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, vol. 35, no. 4, pp. 996-1010. DOI: 10.1109/TPAMI.2012.147
 19. Pyt'ev Yu.P. *Matematicheskie metody issledovaniya prirodnikh resursov Zemli iz kosmosa. Sbornik statei*. Moscow, Nauka, 1984, pp. 41–83.
 20. Vizil'ter Yu.V., Zheltov S.Yu., Bondarenko A.V., Ososkov M.V., Morzhin A.V. *Obrabotka i analiz izobrazhenii v zadachakh mashinnogo zreniya* (Processing and analysis of images in computer vision), Moscow, Fizmatkniga, 2010, 672 p.
 21. Mueller M., Smith N., Ghanem B. A Benchmark and Simulator for UAV Tracking. *European Conference on Computer Vision (ECCV 2016)*, pp. 445-461. DOI: 10.1007/978-3-319-46448-0_27
 22. Mirkes E.M. K-means and K-medoids applet. *University of Leicester*, 2011.
 23. Grady L. Random Walks for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, pp. 1768-1783. DOI: 10.1109/TPAMI.2006.233
 24. Otsu N. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 1979, vol. 9, no. 1, pp. 62-66. DOI: 10.1109/TSMC.1979.4310076
 25. Sidyakin S.V., Egorov A.I., Malin I.K. *Vestnik Moskovskogo aviatsionnogo instituta*, 2014, vol. 21, no. 5, pp. 157-166.
 26. Mirzoyan A.S., Malyshev O.V., Khmarov I.M., Kanivets V.Yu. *Vestnik Moskovskogo aviatsionnogo instituta*, 2014, vol. 21, no. 5, pp. 145-156.