

РАЗРАБОТКА АЛГОРИТМА СЕМАНТИЧЕСКОЙ СЕГМЕНТАЦИИ АЭРОФОТОСНИМКОВ РЕАЛЬНОГО ВРЕМЕНИ

Ю.Б. Блохинов¹, В.А. Горбачев¹, Ю.О. Ракутин¹, А.Д. Никитин¹

¹Государственный научно-исследовательский институт авиационных систем, Государственный научный центр Российской Федерации (ГНЦ ФГУП «ГосНИИАС»), Москва, Россия

Аннотация

Статья посвящена разработке эффективного алгоритма семантической сегментации изображений для работы в реальном времени, обладающего наилучшей точностью в своем классе. На основе сравнительного анализа методов предварительной сегментации, методов вычисления признаков по сегментам изображения, а также различных алгоритмов машинного обучения выявлены наиболее эффективные из них как по точности, так и по быстродействию. По результатам исследования построен модульный алгоритм семантической сегментации со временем выполнения, близким к реальному. Обучение и тестирование проводились на коллекции ISPRS «Vaihingen» аэрофотоснимков видимого и инфракрасного диапазонов, к которым прилагается пиксельная карта высот изображенной местности. Предложен оригинальный способ получения нормализованной матрицы высот по исходной цифровой модели рельефа.

Ключевые слова: анализ изображений, распознавание образов, детектирование, классификация, аэрофотоснимки, матрица высот, суперпиксели, набор признаков, семантическая сегментация, машинное обучение, условные случайные поля.

Цитирование: Блохинов, Ю.Б. Разработка алгоритма семантической сегментации аэрофотоснимков реального времени / Ю.Б. Блохинов, В.А. Горбачев, Ю.О. Ракутин, А.Д. Никитин // Компьютерная оптика. – 2018. – Т. 42, № 1. – С. 141-148. – DOI: 10.18287/2412-6179-2018-42-1-141-148.

Введение

Задача автоматического детектирования и классификации наземных объектов является одной из наиболее интересных задач современного компьютерного зрения. Если в задачах классификации необходимо определить только тип изображённого объекта, в задачах детектирования – построить ограничивающий прямоугольник (или определить координаты) для всех объектов заданного типа, то в задаче семантической сегментации требуется не только обнаружить и классифицировать объекты, но и определить их границы. Иными словами, для каждого пиксела изображения необходимо определить класс объекта, к которому он относится. Таким образом, задача семантической сегментации является наиболее трудной задачей обработки снимков. Трудность обработки дополняется высокой изменчивостью объектов внутри одного класса и высокой схожестью элементов объектов разных классов.

Особый интерес представляет возможность решения задачи семантической сегментации на бортовых вычислительных системах в реальном времени. Он обусловлен бурным развитием автономных беспилотных аппаратов, которое происходит в настоящее время. Семантическая информация об окружающей сцене чрезвычайно важна для систем навигации и систем принятия решений на борту этих аппаратов.

В качестве исходных данных, помимо визуального канала (видимого и ближнего ИК-диапазона), рассматривается канал глубины в виде матрицы высот (*digital surface model*, DSM). Данные о высоте могут быть получены аппаратно с помощью LiDAR-сенсора или в результате применения вычислительной процедуры трёхмерной стереорекострукции [1] или многовидовой реконструкции [2] матрицы высот по ис-

ходным изображениям. В данной работе использована матрица высот, полученная в результате стереорекострукции и нормализованная с помощью оригинального алгоритма, описанного ниже.

Целью статьи являлся поиск наиболее эффективного способа семантической сегментации изображений с точки зрения компромисса между скоростью работы и точностью результата. Вопросы скорости работы крайне важны для прикладного использования алгоритмов анализа изображений в реальном времени, например, в бортовых системах самолётов и БПЛА.

На сегодняшний день одним из самых популярных средств анализа изображений являются свёрточные искусственные нейронные сети (*Convolutional artificial neural net*, CNN), в частности, полносвёрточные сети для семантической сегментации [3]. Несмотря на эффективность и удобство их использования (отсутствие необходимости вручную строить признаковое описание объектов), применение таких сетей достаточно затратно с вычислительной точки зрения и, как правило, требует задействования графического ускорителя (GPU). Однако ресурсы бортовых вычислительных систем более ограничены, чем у настольных персональных ЭВМ. Поэтому в данной работе исследовались подходы на основе классификаторов с признаками, выбираемыми вручную, и имеющими более высокое быстродействие.

Большинство подходов к построению алгоритмов семантической сегментации включают в себя следующие этапы:

1. Предварительная обработка данных.
2. Предварительная сегментация.
3. Признаковое описание.
4. Обучение классификатора и классификация.
5. Постобработка с учётом контекста.

Можно отметить, что алгоритмы имеют модульную структуру, допускают выбор различных методов на каждом этапе и их комбинирование.

Статья включает в себя два параграфа. В первом параграфе обсуждаются возможности выбора методов на разных этапах работы алгоритма. Во втором, экспериментальном, анализируются численные характеристики, описывающие результаты работы исследуемых комбинированных алгоритмов. В заключении приведены обобщение результатов и основные выводы.

1. Этапы реализации алгоритма

В качестве данных использована коллекция ISPRS 2D Semantic Labeling Contest [4]. Она содержит аэрофотоснимки с двумя цветными и инфракрасным (ИК) каналом (IR-R-G), карту высот (DSM) и размеченные вручную изображения (рис. 1). Разметка была произведена по 6 классам: дорога, здание, низкая растительность, деревья, автомобили, прочее. Поскольку в обучающей коллекции всего 16 снимков, 15 использовались для обучения, 1 – для теста.

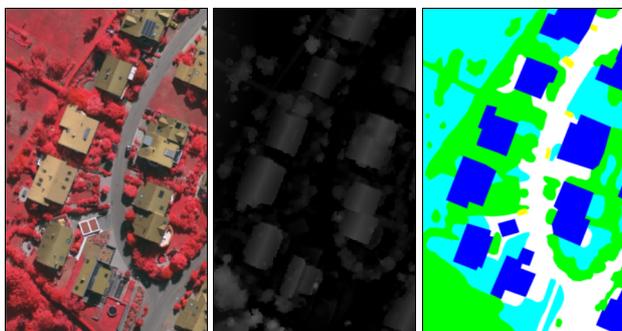


Рис. 1. Обучающие данные: IR-R-G изображение, DSM, ручная разметка изображения (ground truth)

При работе с данными было выяснено, что непосредственное использование канала высоты (DSM) не улучшает результаты классификации. Это очевидно, так как абсолютные значения высоты, например, для дороги и здания могут совпадать в случае, если здание находится в низине или дорога находится на возвышенности. Поэтому необходимо использовать т.н. нормализованную матрицу высот (nDSM), в которой записано не абсолютное значение высоты, а относительное значение высоты над поверхностью. Для автоматического построения nDSM по DSM авторами предложен следующий алгоритм на основе двойной ранговой фильтрации (рис. 2):

1. Медианная фильтрация – устранение выбросов и мелких деталей.
2. Фильтрация низкоранговым фильтром большой апертуры – устранение объектов над поверхностью.
3. Фильтрация высокоранговым фильтром – восстановление «провалов» поверхности.
4. Вычитание фона (рельефа) из исходной матрицы высот для получения нормализованной матрицы высот.

Шаг 3 алгоритма необходим для корректного получения рельефа в регионах, где имеется несколько уровней высоты поверхности или протяжённый равномерный перепад. При всех операциях используется маска фильтра круглой формы.

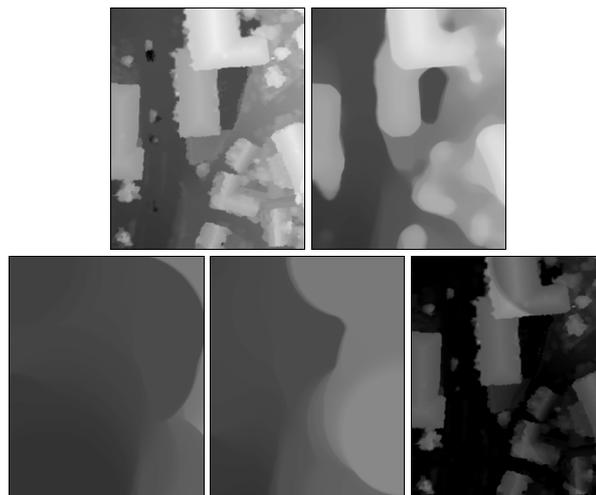


Рис. 2. Этапы получения nDSM. Исходная DSM, медианная фильтрация, низкоранговая фильтрация, высокоранговая фильтрация, разность DSM и фона – nDSM

Результатом семантической сегментации изображения является изображение-карта разметки исходного изображения, в котором каждому пикселу соответствует номер класса объекта, к которому он был отнесён. Однако можно производить классификацию не каждого отдельного пиксела, а т.н. «суперпикселей», небольших однородных сегментов изображений размером 50–100 пикселей. Это позволяет, во-первых, на несколько порядков снизить число классифицируемых в процессе семантической сегментации сущностей и, как следствие, снизить время обработки, во-вторых, обрабатывать более семантически наполненные сущности. Для области изображения можно вычислить более содержательные признаки, чем для пиксела, в первую очередь текстурные и градиентные признаки. Подобный процесс предварительной сегментации изображения на суперпиксели, то есть пересегментация, должен удовлетворять следующим требованиям: сегменты имеют приблизительно одинаковый размер, внутри каждого сегмента должны содержаться только пиксели одного класса. На практике это с хорошей точностью выполняется, если сегменты малы и однородны по цвету. Алгоритмами пересегментации могут служить SLIC, QuickShift, Turbopixel, Fast Graph-based Method. Показано, что наиболее подходящим по скорости и качеству работы является алгоритм SLIC [5] (рис. 3). Алгоритм SLIC основан на методе k-средних для кластеризации в пространстве координат-цветов пикселей XYRGB. При этом начальные центры кластеров расставляются равномерно по изображению. А пересчёт центров кластеров производится локально, то есть только между ближайшими центрами. Благодаря этому метод быстро сходится, имеет линейную по числу пикселей сложность.

В качестве признаков для суперпикселей были использованы статистические характеристики каналов (среднее, дисперсия, максимальное по сегменту значение), текстурные признаки в виде статистики откликов банка фильтров Габора [6] (рис. 4 слева) по области и статистика локальных бинарных шаблонов

(LBP) [7]. Кроме того, отдельно рассмотрены текстурные признаки, вычисляемые с помощью банка признаков Хаара [8] (являющихся разностями между суммами яркостей пикселей в соседних прямоугольниках (рис. 4 справа). Такие признаки в какой-то степени аналогичны признакам Габора, но имеют более простую (бинарную) маску. Благодаря структуре признаков Хаара они могут быть эффективно вычислены на основе интегрального изображения.

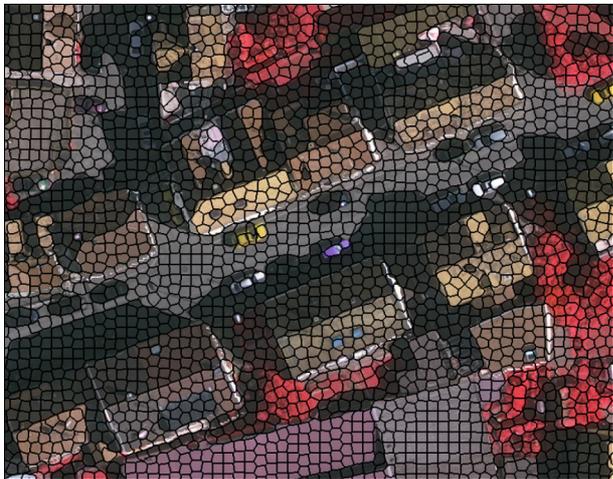


Рис. 3. Пересегментация с помощью алгоритма SLIC

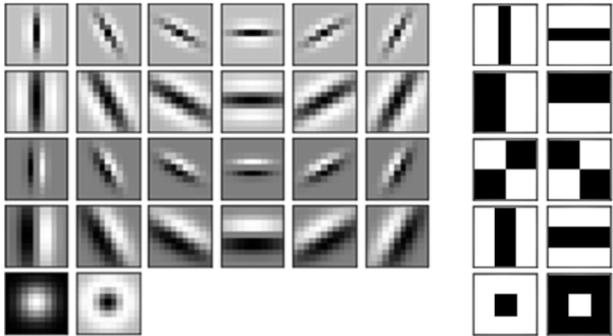


Рис. 4. Ядра фильтров Габора (слева), фильтров Хаара (справа)

Для выбора классификатора было произведено сравнение четырёх алгоритмов: метода опорных векторов (SVM) с ядром RBF (radial-based function), градиентного бустинга над решающими деревьями (GBT) и «случайного леса» (RF), состоящего из 500 и 100 деревьев [9].

Классификация сегментов на основе только их признаков не учитывает контекст, который является существенным для распознавания изображений. Поэтому следующим шагом после проведения классификации сегментов может являться повторная классификация на основе учёта их соседства. Она выполняется для устранения выбросов и в определённом смысле сглаживания результата сегментации. Учёт контекста заключается в гипотезе о том, что соседний сегмент для данного сегмента имеет тот же класс (то есть границы между классами относительно редки), а вероятность встретить по соседству сегменты различных классов соответствует вероятности на обучающей выборке. Математически такая гипотеза формулируется в виде

модели условного случайного поля (CRF). Вероятность того, что изображению I соответствует разметка Y , можно описать следующим образом:

$$P(Y | I) = \frac{1}{Z(I)} \prod_i \Phi(y_i | I) \prod_i \prod_{j \in N(i)} \Psi(y_i, y_j | I),$$

где $\Phi(y_i | I)$ – фактор, выражающий вероятность того, что сегмент i изображения будет иметь класс y_i , $\Psi(y_i, y_j | I)$ – фактор, выражающий вероятность того, что сегмент i и его сосед j из окрестности $N(i)$ будут иметь классы y_i и y_j одновременно. Перемножение производится по всем сегментам i изображения I . $Z(i)$ – константа нормализации, равная сумме по всем возможным разметкам произведений факторов. Максимизацию вероятности можно заменить минимизацией логарифма вероятности, т.н. «энергии» карты разметки:

$$E(Y | I) = \sum_i \phi(y_i | I) + \sum_i \sum_{j \in N(i)} \psi(y_i, y_j | I),$$

где $\phi = \ln \Phi$ – «унарный потенциал», $\psi = \ln \Psi$ – «бинарный потенциал». При этом вычислять константу нормализации не требуется.

Для реализации модели CRF можно использовать метод SSVM [10], при котором и унарные потенциалы, и парные потенциалы обучаются одновременно по данным. Недостатком метода SSVM является очень длительное обучение. Другой вариант использования модели CRF состоит в том, чтобы сначала обучить классификатор сегментов и использовать вычисленные им вероятности классов для унарных потенциалов, а парные потенциалы задать на основе модели Поттса или статистики взаимной встречаемости, а затем применить алгоритм вывода, определяющий карту разметки с минимальной энергией. В проведённых экспериментах метод SSVM не показал преимущества по точности, поэтому применялся второй подход. Функционал энергии задавался в следующем виде:

$$E(Y) = \sum_{i \in I} \ln P(y_i | x_i) + \sum_{i \in I} \sum_{j \in N(i)} K \cdot [y_i \neq y_j],$$

где $P(y_i | x_i)$ – вычисляемая классификатором оценка вероятности принадлежности сегмента с признаками x_i к классу y_i , [true]=1, [false]=0, K – коэффициент. Для вывода использован метод QPBO [11], реализованный в библиотеке PyStruct [12].

Применение условных случайных полей позволяет получить более «гладкую» карту сегментации, снизить количество выбросов и в целом повысить точность сегментации (рис. 5). Однако данная процедура достаточно вычислительно затратная, вывод оптимальной карты разметки занимает время, превышающее время классификации суперпикселей.

2. Сравнительный анализ элементов алгоритма

На первом этапе было замерено время вычисления значений признаков сегментов изображения (табл. 1).

Необходимо заметить, что все цифры по времени вычисления признаков являются относительными.

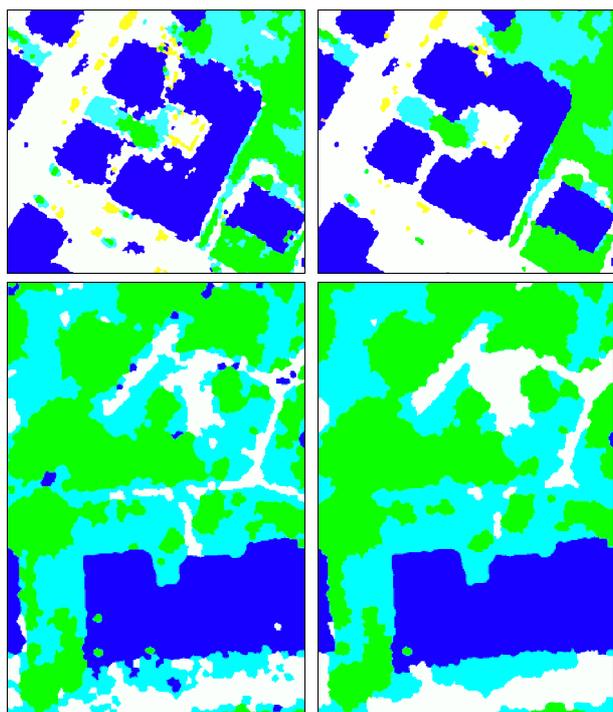


Рис. 5. Результаты сегментации с использованием модели CRF (справа) и без неё (слева)

Табл. 1. Время подготовки суперпикселей и вычисления их признаков

Тип сегментации	SLIC	SLIC ускоренная	Прямоугольная сетка 10×10	Прямоугольная сетка 5×5
Сегментация	337,11	50,72	0	0
Простые признаки	61,25	61,69	60,72	59,72
Текстурные признаки	81,62	84,41	79,55	79,55
Упрощённые текстурные признаки	42,52	43,39	42,38	42,38
Граф соседства	20,56	19,88	0	0

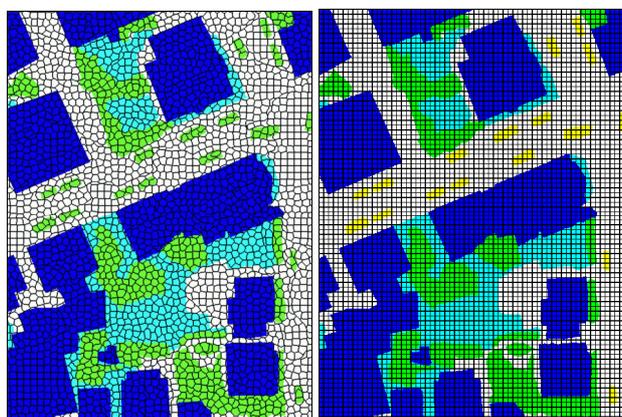


Рис. 6. Сегменты изображения, наложенные на карту разметки. SLIC слева, равномерная сетка 10×10 справа

Затем вычислялась точность такой разметки, подсчитывалась доля пикселей полученной карты, метка класса которых совпала с эталонной. Результаты оценки точности приведены в табл. 2. Видно, что три основных метода близки по точности.

В ходе экспериментов на этапе распознавания необходимо было выяснить:

1. Какой набор признаков является предпочтительным: только простые статистические признаки по

Оценка времени производилась при вычислении на Python, оптимизация не производилась. Для оценки более реалистичного времени вычисления признаков было произведено вычисление простых признаков на равномерной сетке 10×10 с использованием библиотеки OpenCV на языке C++ с применением небольших оптимизаций (в частности, интегрального изображения). Время вычисления в таком случае составило менее 0,4 секунды. Изображение имело размер 2000×2500 пикселей (5 мпикс). Вычисления производились на компьютере с процессором Intel(R) Core(TM) i5-3470, 3,20 GHz, оперативной памятью 16 Gb, под операционной системой Windows 8.1 64bit.

На втором этапе была вычислена ошибка, вносимая алгоритмами пересегментации ещё до этапа распознавания. Ошибки возникают тогда, когда границы сегментов не совпадают с границами объектов, один сегмент содержит пиксели различных классов (рис. 6). Точность пересегментации замерялась следующим образом. Строилась карта семантической сегментации, на которой всем пикселям каждого сегмента присваивался тот класс, который преобладал для пикселей этого сегмента в эталонной разметке (наилучшая разметка для данных сегментов).

каналам, простые с добавлением текстурных на основе признаков Хаара, простые с добавлением текстурных на основе фильтров Габора и LBP.

2. Какой метод пересегментации является предпочтительным: SLIC, упрощённый SLIC, или достаточно разбивать изображение на небольшие квадраты.
3. Какой метод обучения является предпочтительным: SVM, GBT, RF или RF с уменьшенным числом деревьев.

Результаты экспериментов сведены в табл. 3. Пропуски в табл. 3 соответствуют ситуациям, когда обучение требовало слишком большого расхода оперативной памяти или времени и не было завершено.

Точность сегментации рассчитывалась как доля пикселей изображения, вычисленная метка которых совпала с эталонной.

Видно, что самый точный результат 88,67% достигается на самом полном наборе признаков, включающем текстурные, с применением леса из 500 деревьев (RF-500) и сегментации с помощью SLIC. На его получение требуется больше всего времени как на этапе классификации (2,65 с), так и на этапе предварительной сегментации и вычисления признаков. Однако видно, что можно получить почти такие же высокие показатели точности, заменив сложные текстур-

ные признаки на упрощённые (Хаара), и использовать разбиение изображения на прямоугольники вместо суперпикселей. Это позволяет избежать процедуры SLIC и почти вдвое сэкономить на времени вычисления текстурных признаков, при этом точность составит 88,27%. Как ни странно, можно вообще отказаться от текстурных признаков и сложной сегментации и

сохранить достаточно высокую точность: 88,09%. Это можно объяснить тем, что, во-первых, разбиение сеткой вносит в классификацию ошибку незначительно большую, чем SLIC (табл. 2), во-вторых, тем, что значения каналов высоты и ИК являются достаточно информативными, чтобы различить рассматриваемые классы.

Табл. 2. Ошибка, вносимая предварительной сегментацией изображения

Тип сегментации	SLIC	SLIC ускоренная	Прямоугольная сетка 10×10	Прямоугольная сетка 5×5
Ошибка, вносимая сегментацией	4,15 %	4,39 %	4,63 %	2,34 %

Табл. 3. Результаты классификации

Тип сегментации		SLIC		SLIC ускоренная		Сетка 10×10		Сетка 5×5	
		Точность	Время,с	Точность	Время,с	Точность	Время,с	Точность	Время,с
Статистич. признаки	GBT	87,56	1,53	87,81	1,44	87,71	1,55	87,47	7,00
	RF 500	88,23	2,53	88,35	2,23	88,09	2,77	–	–
	RF 100	88,09	0,48	88,11	0,51	88,05	0,58	87,94	2,86
	SVM	78,39	339,25	78,13	352,16	78,14	516,52	–	–
Статистич. упрощённые текстурн. признаки	GBT	87,70	1,44	87,93	1,43	87,84	1,57	87,61	6,73
	RF 500	88,24	2,29	88,47	2,35	88,27	2,70	–	–
	RF 100	88,15	0,48	88,24	0,50	88,14	0,59	87,96	2,78
	SVM	76,69	377,13	76,61	383,11	77,27	561,14	–	–
Статистич. текстурн. признаки	GBT	88,25	1,78	88,20	1,79	88,06	2,04	88,23	9,16
	RF 500	88,67	2,65	88,62	2,61	88,55	3,15	–	–
	RF 100	88,38	0,47	88,47	0,47	88,33	0,63	88,33	3,12
	SVM	80,03	525,01	80,13	519,51	79,93	760,34	–	–

Более мелкое дробление сетки (5×5 пикселей) изображения приводит к очень существенному росту времени обучения и расходов памяти. В ряде случаев завершить процесс обучения за разумное время (сутки) не удалось. Время классификации ожидаемо увеличивается более чем в 4 раза (почти пропорционально числу сегментов). При этом ощутимого роста точности классификации замечено не было.

Существенно понизить время классификации (0,47 с) можно, используя RF с меньшим количеством более простых деревьев (100 вместо 500). В этом случае точность составляет 88,38%. Кроме того, видно, что если вообще избежать сложной пересегментации, ограничиться вычислением только простейших признаков (статистик каналов по сегментам) и использовать упрощённый классификатор, то точность составит 88,05% при времени 0,58 с. Это всего на 0,62% процентных пункта меньше, чем максимальная точность. Как было упомянуто в предыдущем пункте, для вычисления таких признаков на обычном персональном компьютере при использовании библиотеки OpenCV достаточно 0,4 секунды (при вычислении без использования графического ускорителя). На рис. 7 представлен график точность-время для основных результатов классификации.

Таким образом, эксперименты показали, что можно достичь точности 88,05% при времени обработки снимка 5 мпикс всего за 1 секунду. Это говорит о возможности применения методов семантической сегментации с достаточно высокой точностью в реальном времени. Однако последующее повышение точности даже на пару процентов требует уже заметно больших

затрат времени как на предварительную сегментацию, так и на вычисление более сложных признаков и классификацию. Достижимыми показателями точности рассмотренной группы алгоритмов является около 88,7% (на коллекции ISPRS benchmark on Urban Classification and Semantic Labeling [4]).

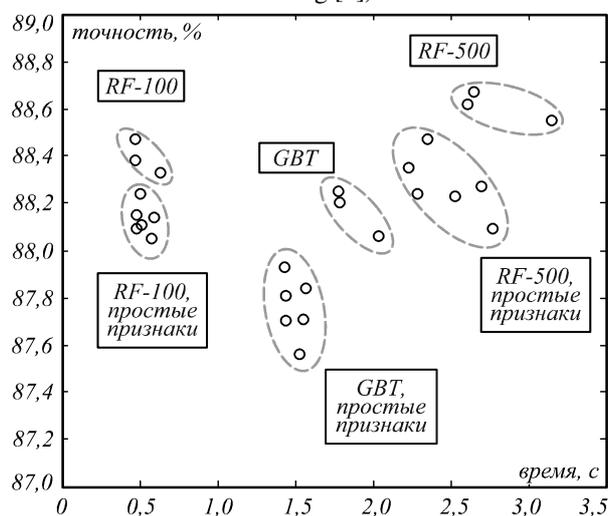


Рис. 7. График результатов работы классификаторов. Простые признаки – только статистические и упрощённые текстурные

Результаты применения модели CRF после классификации приведены в табл. 4. Указано только время вывода без учёта вычисления вероятностей классификатором и создания графа соседства суперпикселей. Метод SVM и сетка 5×5 не тестировались ввиду очень большого времени на обучение и недостаточной точности.

Табл. 4. Результаты применения CRF

Тип сегментации		SLIC		SLIC ускоренная		Сетка 10×10	
		Точность	Время, с	Точность	Время, с	Точность	Время, с
Статистические признаки	GBT	88,62	3,65	88,77	3,59	88,68	4,31
	RF 500	89,16	3,89	88,95	3,86	88,76	4,36
	RF 100	88,74	4,56	88,40	4,63	88,44	4,44
Статистические + упрощённые текстурные признаки	GBT	88,37	3,59	89,06	3,63	88,88	4,32
	RF 500	89,19	3,77	89,19	3,74	88,84	4,31
	RF 100	88,84	4,46	88,53	4,55	88,62	4,38
Статистические + текстурные признаки	GBT	88,90	3,85	89,69	3,72	88,93	4,31
	RF 500	89,74	3,76	89,41	3,72	89,40	4,33
	RF 100	89,64	4,49	89,36	3,63	89,16	4,49

Учёт контекста с применением модели CRF позволяет поднять точность сегментации приблизительно на процент. Для наилучшего решения точность достигла 89,74 %. Помимо этого, видно, что подобная обработка «выравнивает» результаты различных классификаторов. Точность различных подходов становится близкой, не опускается ниже 88,4% для алгоритмов, использующих RF-классификатор. Кроме того, оказалось, что чем менее точна основная классификация, тем больше времени требует вывод CRF. Однако повышение точности требует существенных затрат времени: около 4 с на вывод оптимальной карты из модели, плюс время на создание графа соседств суперпикселей (если он не тривиальный, как в случае равномерной сетки). В силу этого применение модели CRF для уточнения результатов классификации в реальном времени не представляется возможным.

Заключение

В работе произведено сравнение основных подходов к построению алгоритмов семантической сегментации, замерены относительные показатели скорости и точности для различных архитектур алгоритмов. Полученные экспериментальные данные говорят о том, что наиболее точные результаты достигаются при использовании пересегментации по методу SLIC, использовании как статистических, так и текстурных признаков суперпикселей. Среди методов машинного обучения лучшие результаты по точности классификации даёт случайный лес (RF). Достижимая точность семантической сегментации составляет около 88,7%. С использованием контекстной информации (через модель CRF) можно достичь точности около 89,8%. Для повышения скорости работы необходимо отказаться от разбиения изображения на суперпиксели в пользу разбиения на квадратные ячейки, использовать только статистические признаки изображений и не использовать CRF. Кроме того, можно снизить количество решающих деревьев в классификаторе без существенного снижения его точности. Это позволяет снизить время классификации в пять раз. При таком подходе можно достичь точности в 88 %. Время полного цикла обработки изображения размером

5 мегапикселей в таком случае составляет менее 1 секунды. Таким образом, с одной стороны, достигаются достаточно высокие показатели точности, с другой стороны, становится возможным применение методов семантической сегментации в реальном времени. Это говорит о высокой перспективности дальнейшего развития подобных методов для применения в бортовых системах.

Благодарности

Работа выполнена при финансовой поддержке РФФИ (проект № 17-08-00191 а).

Набор данных «Vaihingen» предоставлен немецким обществом фотограмметрии, дистанционного зондирования и геоинформации (DGPF) [13, 14]: <http://www.ifp.uni-stuttgart.de/dgpf/DKEP-Allg.html>, которому авторы выражают свою искреннюю благодарность.

Литература

1. **Горбачев, В.А.** Плотная реконструкция рельефа местности на основе модифицированного алгоритма полуглобального стереоотождествления / В.А. Горбачев // Известия РАН. Теория и системы управления. – 2014. – № 2. – С. 68-79. – DOI: 10.7868/S0002338814020103.
2. **Веркеенко, М.С.** Построение плотных моделей поверхности земли на основе метода полуглобального отождествления для случая множественного перекрытия снимков / М.С. Веркеенко, В.А. Горбачев // Известия РАН. Теория и системы управления. – 2017. – № 6. – С. 100-111. – DOI: 10.7868/S0002338817060087.
3. **Long, J.** Fully convolutional networks for semantic segmentation / J. Long, E. Shelhamer, T. Darrell // IEEE Conference on Computer Vision and Pattern Recognition. – 2015. – P. 3431-3440. – DOI: 10.1109/CVPR.2015.7298965.
4. 2D Semantic labeling contest. – URL: <http://www2.isprs.org/commissions/comm3/wg4/semantic-labeling.html> (дата обращения 8.09.2017).
5. **Achanta, R.** SLIC superpixels compared to state-of-the-art superpixel methods / R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Susstrunk // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2012. – Vol. 34, Issue 11. – P. 2274-2281. – DOI: 10.1109/TPAMI.2012.120.
6. **Malik, J.** Contour and texture analysis for image segmentation / J. Malik, S. Belongie, T. Leung, J. Shi // International Journal of Computer Vision. – 2001. – Vol. 43(1). – P. 7-27. – DOI: 10.1023/A:1011174803800.

7. **Ojala, T.** Comparative study of texture measures with classification based on feature distributions / T. Ojala, M. Pietikäinen, D.A. Harwood // Pattern Recognition. – 1996. Vol. 29, Issue 1. – P. 51-59. – DOI: 10.1016/0031-3203(95)00067-4.
8. **Papageorgiou, C.P.** A general framework for object detection / C.P. Papageorgiou, M. Oren, T. Poggio // Proceedings of International Conference on Computer Vision. – 1998. – P. 555-562. – DOI: 10.1109/ICCV.1998.710772.
9. Машинное обучение (курс лекций, К.В. Воронцов). – URL: http://www.machinelearning.ru/wiki/index.php?title=Машинное_обучение_%28курс_лекций%2C_К.В.Воронцов%29 (дата обращения 8.09.2017).
10. **Finley, T.** Training structural SVMs when exact inference is intractable / T. Finley, T. Joachims // Proceedings of the 25th International Conference on Machine Learning (ICML '08). – 2008. – P. 304-311. – DOI: 10.1145/1390156.1390195.
11. **Kolmogorov, V.** Minimizing non-submodular functions with graph cuts – A review / V. Kolmogorov, C. Rother // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2007. – Vol. 29(7). – P. 1274-1279. – DOI: 10.1109/TPAMI.2007.1031.
12. **Müller, A.C.** PyStruct Structured prediction in Python / A.C. Müller, S. Behnke // Journal of Machine Learning Research. – 2014. – Vol. 15, Issue 1. – P. 2055-2060.
13. **Cramer, M.** The DGPF – Test on digital aerial camera evaluation overview and test design / M. Cramer // Photogrammetrie – Fernerkundung – Geoinformation. – 2010. – Vol. 2. – P. 73-82. – DOI: 10.1127/1432-8364/2010/0041.
14. Evaluierung digitaler photogrammetrischer luftbildkamerasysteme. – URL: <http://www.ifp.uni-stuttgart.de/dgpf/DKEP-Allg.html> (дата обращения 8.09.2017).

Сведения об авторах

Блохинов Юрий Борисович, 1956 года рождения, в 1980 году окончил Московский физико-технический институт по специальности «Динамика полета и управление», в 1990 году защитил кандидатскую диссертацию во ВНИИ Системных исследований, в 2012 году защитил докторскую диссертацию в Московском государственном университете геодезии и картографии (МИИГАиК). Работает начальником лаборатории в ГНЦ ФГУП «ГосНИИАС». Область научных интересов: цифровая фотограмметрия, компьютерное зрение, анализ изображений, распознавание образов. E-mail: yury.blokhinov@gosniias.ru.

Горбачев Вадим Александрович, 1988 года рождения, в 2011 году окончил Московский физико-технический институт (МФТИ (ГУ)) по специальности «Системный анализ, управление и обработка информации», в 2014 году защитил кандидатскую диссертацию в МФТИ (ГУ). Работает начальником сектора в ГНЦ ФГУП «ГосНИИАС». Область научных интересов: компьютерное зрение, машинное обучение, распознавание образов, анализ изображений. E-mail: yadim.gorbachev@gosniias.ru.

Ракутин Юрий Олегович, 1994 года рождения, в 2015 году окончил бакалавриат Московского физико-технического института по направлению 03.03.01 «Прикладные математика и физика». Учится в магистратуре МФТИ, работает инженером в ГНЦ ФГУП «ГосНИИАС». Область научных интересов: компьютерное зрение, машинное обучение, анализ данных. E-mail: rakutin@phystech.edu.

Никитин Андрей Дмитриевич, 1994 года рождения, в 2016 году окончил бакалавриат Московского физико-технического института по направлению 03.03.01 «Прикладные математика и физика». Учится в магистратуре МФТИ, работает инженером в ГНЦ ФГУП «ГосНИИАС». Область научных интересов: компьютерное зрение, машинное обучение, анализ данных. E-mail: unlostdaimos@gmail.com.

ГРНТИ: 28.23.15

Поступила в редакцию 8 сентября 2017 г. Окончательный вариант – 3 ноября 2017 г.

A REAL-TIME SEMANTIC SEGMENTATION ALGORITHM FOR AERIAL IMAGERY

Y.B. Blokhinov¹, V.A. Gorbachev¹, Y.O. Rakutin¹, A.D. Nikitin¹
¹State Research Institute of Aviation Systems, Moscow, Russia

Abstract

We propose a novel effective algorithm for real-time semantic segmentation of images that has the best accuracy in its class. Based on a comparative analysis of preliminary segmentation methods, methods for calculating attributes from image segments, as well as various algorithms of machine learning, the most effective methods in terms of their accuracy and performance are identified. Based on the research results, a modular near real-time algorithm of semantic segmentation is constructed. Training and testing is performed on the ISPRS Vaihingen collection of aerial photos of the visible and IR ranges, to which a pixel map of the terrain heights is attached. An original method for obtaining a normalized nDSM for the original DSM is proposed.

Keywords: image analysis, pattern recognition, detection, classification, aerial images, DSM, superpixels, feature vector, semantic segmentation, machine learning, conditional random fields.

Citation: Blokhinov YB, Gorbachev VA, Rakutin YO, Nikitin AD. A real-time semantic segmentation algorithm for aerial imagery. *Computer Optics* 2018; 42(1): 141-148. DOI: 10.18287/2412-6179-2018-42-1-141-148.

Acknowledgements: The work was partially funded by the Russian Foundation of Basic Research, grant No. 17-08-00191 a. The Vaihingen data set was provided by the German Society for Photogrammetry, Remote Sensing and Geoinformation (DGPF) [13]: <http://www.ifp.uni-stuttgart.de/dgpf/DKEP-Allg.html>.

References

- [1] Gorbachev VA. Dense terrain stereoreconstruction using modified SGM. *Journal of Computer and Systems Sciences International* 2014; 53(2): 212-223. DOI: 10.1134/S1064230714020105.
- [2] Verkeenko MS, Gorbachev VA. Development of dense models of the Earth's surface based on the semiglobal matching method for the case of multiple overlapping images. *Journal of Computer and Systems Sciences International* 2017; 56(6): 997-1007. DOI: 10.1134/S1064230717060132.
- [3] Long J., Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Conference on Computer Vision and Pattern Recognition* 2015: 3431-3440. DOI: 10.1109/CVPR.2015.7298965.
- [4] 2D Semantic labeling contest. – Source: (<http://www2.isprs.org/commissions/comm3/wg4/semantic-labeling.html>)
- [5] 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; 34(11): 2274-2281. DOI: 10.1109/TPAMI.2012.120.
- [6] Malik J, Belongie S, Leung T, Shi J. Contour and texture analysis for image segmentation. *International Journal of Computer Vision* 2001; 43(1): 7-27. DOI: 10.1023/A:1011174803800.
- [7] Ojala T, Pietikäinen M, Harwood DA. Comparative study of texture measures with classification based on feature distributions. *Pattern Recognition* 1996; 29(1): 51-59. DOI: 10.1016/0031-3203(95)00067-4.
- [8] Papageorgiou CP, Oren M, Poggio T. A general framework for object detection. *Proceedings of International Conference on Computer Vision* 1998: 555-562. DOI: 10.1109/ICCV.1998.710772.
- [9] Machine learning (K.V. Voroncov's lecture course). – Source: (http://www.machinelearning.ru/wiki/index.php?title=Машинное_обучение_%28курс_лекций%2C_К.В.Воронцов%29).
- [10] Finley T, Joachims T. Training structural SVMs when exact inference is intractable. *Proceedings of the 25th international conference on Machine learning* 2008: 304-311. DOI: 10.1145/1390156.1390195.
- [11] Kolmogorov V, Roth C. Minimizing non-submodular functions with graph cuts – A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2007; 29(7): 1274-1279. DOI: 10.1109/TPAMI.2007.1031.
- [12] Müller AC, Behnke S. PyStruct Structured prediction in Python. *Journal of Machine Learning Research* 2014; 15(1): 2055-2060.
- [13] Cramer M. The DGPF – Test on digital aerial camera evaluation overview and test design. *Photogrammetrie – Fernerkundung – Geoinformation* 2010; 2: 73-82. DOI: 10.1127/1432-8364/2010/0041.
- [14] Evaluierung digitaler photogrammetrischer luftbildkamarasysteme. Source: (<http://www.ifp.uni-stuttgart.de/dgpf/DKEP-Allg.html>).

Author's information

Yury Borisovitch Blokhinov (b. 1956) graduated from the Moscow Institute of Physics and Technology in 1980 in the speciality "Space Flight Dynamics and Control". In 1990 he defended his PhD thesis in the State Research Institute of System Analysis, in 2012 he defended his DE thesis in the Moscow State University of Geodesy and Cartography. He works as head of the laboratory in the State Research Institute of Aviation Systems. His research interests currently are digital photogrammetry, computer vision, image analysis, pattern recognition. E-mail: yury.blokhinov@gosniias.ru.

Vadim Aleksandrovich Gorbachev, born in 1988, graduated from the Moscow Institute of Physics and Technology (MIPT (SU)) in 2011, in the specialty "System Analysis, Control and Information Processing". In 2014 he defended his PhD thesis at MIPT. He works as head of sector at the FSUE State Research Institute of Aviation Systems (SSC of RF). Research interests: computer vision, machine learning, pattern recognition, image analysis. E-mail: vadim.gorbachev@gosniias.ru.

Yury Olegovich Rakutin (b. 1994) got the Bachelor's degree in "Applied Mathematics and Physics" in the Moscow Institute of Physics and Technology in 2015. He receives the Master's degree in the Moscow Institute of Physics and Technology and works as an engineer in the State Research Institute of Aviation Systems. His research interests currently are computer vision, machine learning, and data analysis. E-mail: rakutin@phystech.edu.

Andrey Dmitrievich Nikitin (b. 1994) got the Bachelor's degree in "Applied Mathematics and Physics" in the Moscow Institute of Physics and Technology in 2016. He receives the Master's degree in Moscow Institute of Physics and Technology and works as an engineer in the State Research Institute of Aviation Systems. His research interests currently are computer vision, machine learning, and data analysis. E-mail: unlostdaimos@gmail.com.

Received September 8, 2017. The final version – November 3, 2017.