

АЛГОРИТМ СПЕКТРАЛЬНО-ТЕКСТУРНОЙ КЛАССИФИКАЦИИ СПУТНИКОВЫХ ИЗОБРАЖЕНИЙ С ИСПОЛЬЗОВАНИЕМ ЧАСТИЧНОГО ОБУЧЕНИЯ

Сергей Александрович Рылов

Институт вычислительных технологий СО РАН, 630090, Россия, г. Новосибирск, пр. Академика Лаврентьева, 6, кандидат технических наук, научный сотрудник, e-mail: RylovS@mail.ru

При классификации спутниковых изображений обучающая выборка зачастую является непредставительной, что приводит к низкому качеству сегментации. В таких условиях целесообразно использовать методы классификации с частичным обучением, которые одновременно используют как размеченные (обучающую выборку), так и неклассифицированные данные. Вместе с тем, спутниковые изображения высокого пространственного разрешения характеризуются высокой внутриклассовой неоднородностью спектральных характеристик, что вызывает острую необходимость учитывать пространственную информацию. В работе предлагается алгоритм классификации мультиспектральных изображений, совместно учитывающий спектральные и текстурные признаки. При этом использование концепции частичного обучения позволяет повысить качество результатов классификации в условиях малого объема обучающей выборки. Приводятся результаты экспериментов на модельных и спутниковых изображениях, подтверждающие эффективность предложенного алгоритма.

Ключевые слова: частичное обучение, частично контролируемая классификация, спектрально-текстурная сегментация, спутниковые изображения, высокое пространственное разрешение.

SEMI-SUPERVISED SPECTRAL-TEXTURE IMAGE CLASSIFICATION ALGORITHM

Sergey A. Rylov

Institute of Computational Technologies SB RAS, 6, Prospect Akademik Lavrentiev St., Novosibirsk, 630090, Russia, Ph. D., Researcher, e-mail: RylovS@mail.ru

When classifying satellite images, training sample often turns out to be unrepresentative. This leads to low segmentation quality. In such conditions, it is advisable to use semi-supervised classification methods, which simultaneously utilize both training sample and unclassified data. At the same time, high resolution satellite images are characterized by high interclass heterogeneity of spectral characteristics, which demands to take spatial information into account. We propose a new semi-supervised classification algorithm for multispectral images, that utilizes both spectral and texture features. The use of the semi-supervised concept allows improving the classification quality when the amount of training sample is small. The results of experiments on model and satellite images confirming the effectiveness of the proposed algorithm are given.

Key words: semi-supervised classification, spectral-texture segmentation, satellite images, high spatial resolution.

Введение

Спутниковые данные незаменимы при решении задач, связанных с мониторингом протяженных или труднодоступных территорий. Одним из наиболее

распространенных подходов при тематической обработке спутниковых изображений является использование алгоритмов классификации с обучением. Однако, процесс получения необходимой обучающей выборки (ОВ), как правило, требует значительных материальных и временных затрат. Использование малого объема ОВ при построении классификации может приводить к неудовлетворительному качеству распознавания, особенно при анализе больших и сложных сцен.

В условиях недостаточно полной ОВ целесообразно использовать методы классификации с частичным обучением (semi-supervised learning) [1–11]. В процессе их работы для построения решающего правила используется информация, содержащаяся не только в помеченных (ОВ), но и в непомеченных (неклассифицированных) данных. Это позволяет повысить точность разделения классов и качество результатов классификации при малом объеме ОВ. Однако, в настоящее время в распространенных пакетах обработки спутниковых изображений (ERDAS Imagine, ENVI, ArcGIS, SNAP, GRASS и др.) алгоритмы на основе частичного обучения не представлены. Это во многом связано с тем, что их практическое применение к обработке спутниковых изображений большого размера затруднено ввиду большого числа настраиваемых параметров и высоких вычислительных затрат [8–11].

Кроме того, на спутниковых изображениях высокого пространственного разрешения наблюдается увеличение внутриклассовой неоднородности спектральных характеристик [12]. Вследствие чего применение классических методов, учитывающих лишь спектральные характеристики, приводит к чрезмерной раздробленности искомых классов [13]. Поэтому для получения качественных результатов классификации необходимо учитывать одновременно спектральные и текстурные признаки [14]. Однако, существующие алгоритмы классификации с частичным обучением, учитывающие спектрально-текстурные характеристики [9–11], как правило, после расширения ОВ на первом этапе применяют стандартные алгоритмы классификации, использующие только спектральные признаки. Таким образом, классы с сильно разнородными или пересекающимися спектральными характеристиками не могут быть качественно распознаны.

В данной работе предлагается вычислительно эффективный алгоритм спектрально-текстурной классификации с частичным обучением для обработки спутниковых изображений высокого пространственного разрешения. Алгоритм основан на оригинальном подходе к описанию мультиспектральной текстуры [15–16], который позволяет компактно описывать многомерную гистограмму для локальных областей изображения с помощью предварительного использования непараметрической кластеризации.

Методы и материалы

Для большинства объектов природного происхождения структурные особенности, такие как пространственная ориентация, не являются важными,

т.к. пространственная композиция их составных элементов, как правило, имеет случайный характер [17]. При этом количественное содержание составных элементов в локальных областях изображения для текстур одного и того же класса примерно совпадает, а для текстур разных классов отличается. Поэтому для описания естественных текстур целесообразным является использование статистического подхода на основе анализа гистограмм спектральных яркостей [18].

Для описания спектрально-текстурных характеристик областей мультиспектрального изображения предлагается использовать подход на основе векторов частот встречаемости кластеров [15–16], полученных при непараметрической кластеризации изображения по спектральным признакам. «Спектральные» кластеры соответствуют некоторому разбиению многомерного пространства спектральных признаков, которое можно рассматривать как способ квантования, позволяющий сократить размер анализируемых характеристик. Выделенные кластеры соответствуют определенным модам плотности спектральных яркостей и могут характеризовать составные элементы объектов изображения. Основная идея используемого подхода к описанию мультиспектральной текстуры заключается в том, что в локальных областях изображения для мультиспектральных текстур одного и того же класса процентное содержание пикселей из разных «спектральных» кластеров примерно одинаково, а для текстур разных классов заметно отличается.

Пусть картосхема, полученная с помощью кластеризации по спектральным признакам, состоит из K кластеров. Тогда для произвольного фрагмента изображения f вектор спектральных частот $z(f) = (z^{(1)}, \dots, z^{(K)})$ определяется следующим образом: $z^{(i)} = N_f^{(i)} / N_f, i = 1, \dots, K$, где N_f – число всех пикселей фрагмента f , $N_f^{(i)}$ – число пикселей фрагмента f , относящихся к кластеру с номером i . Представленные векторы спектральных частот позволяют описывать многомерную гистограмму фрагментов изображения.

Расстояние между векторами спектральных частот $x = (x^{(1)}, \dots, x^{(K)})$ и $y = (y^{(1)}, \dots, y^{(K)})$ задается с помощью манхэттенской метрики (расстояние городских кварталов): $\mu(x, y) = 1/2 \cdot \sum_{i=1}^K |x^{(i)} - y^{(i)}|$. При совпадении векторов x и y $\mu(x, y) = 0$; в случае, если они не имеют общих ненулевых компонент, $\mu(x, y) = 1$.

Предлагаемый алгоритм спектрально-текстурной классификации мультиспектральных изображений с частичным обучением состоит из четырех основных шагов.

Шаг 1. Кластеризации по спектральным признакам. Производится сегментация исходного изображения с помощью непараметрического алгоритма кластеризации НЕСА [19]. Этот алгоритм обеспечивает выделение кластеров различной формы, размера и плотности в пространстве спектральных признаков. Кроме того, благодаря использованию сеточного подхода он обладает высоким быстродействием. В результате получается картосхема, состоящая из N пикселей и K кластеров.

Шаг 2. Формирование векторов спектральных частот. Картосхема разбивается на множество непересекающихся фрагментов размера $h \times h$ пикселей (где h – задаваемый параметр): $\{f_1, \dots, f_M\}$. Число фрагментов $M \approx N/h^2$. Для всех фрагментов вычисляются соответствующие векторы спектральных частот $\{z_1, \dots, z_M\}$ согласно определению. Для всех представителей обучающей выборки рассматриваются фрагменты с центром в соответствующих пикселях, и для них также вычисляются векторы спектральных частот $\{y_1, \dots, y_L\}$.

Шаг 3. Классификация фрагментов изображения. Векторы спектральных частот $\{z_1, \dots, z_M\}$ классифицируются с помощью модифицированного алгоритма кластеризации CLUSTER [20], в котором в качестве начальных центров классов выступают представители ОВ $\{y_1, \dots, y_L\}$. Кроме того, процедура сдвига центров выполняется несколько раз (мы ограничивались 10 итерациями). Данный алгоритм схож с известным алгоритмом К-средних, отличие состоит лишь в том, что центры сдвигаются при последовательном рассмотрении каждого элемента данных, а не после рассмотрения их всех. Таким образом имеющиеся представители ОВ корректируются в пространстве спектрально-текстурных признаков с учетом неклассифицированных данных. Пиксели фрагментов изображения относятся к тем классам, к которым были отнесены соответствующие им векторы спектральных частот.

Шаг 4. Попиксельная коррекция границ классов. Последовательно рассматриваются границы смежных фрагментов, отнесенных к разным классам. Каждый пиксель, попадающий на границу, относится к тому из соседних классов, к центру (или центрам) которого окажется ближе вектор спектральных частот фрагмента размера $h \times h$ с центром в рассматриваемом пикселе. При этом могут появиться новые граничные пиксели. Процесс продолжается до тех пор, пока все граничные пиксели не будут рассмотрены.

Вычислительная сложность предложенного алгоритма классификации с частичным обучением линейно зависит от числа пикселей изображения N . Параметр h определяет размер фрагментов, описывающих текстуру (характерный размер текстуры).

На основе представленного подхода к описанию спектрально-текстурных характеристик был также разработан алгоритм классификации [14] (без использования частичного обучения), в котором векторы спектральных частот $\{z_1, \dots, z_M\}$ классифицируются по ОВ путем отнесения к ближайшему элементу без учета структуры неклассифицированных данных.

Результаты

Предложенный алгоритм классификации исследовался на модельных и спутниковых изображениях высокого пространственного разрешения. Обработка проводилась на ПЭВМ с процессором Intel Core i5 (4 ядра, 3.5 ГГц).

На рис. 1 представлено цветное модельное изображение, состоящее из двух текстурных классов. Оба класса содержат одну общую цветовую компоненту, составляющую 50% текстуры в каждом из них. Для левого класса харак-

терное содержание темно-зеленой цветовой компоненты составляет 40 % и светло-зеленой – 10%. Для правого класса наоборот: 10% – для темно-зеленой и 40% – для светло-зеленой. Такое содержание цветовых компонент соблюдается в полной мере на правом и левом краях изображения, а ближе к середине (границе классов) характеристики плавно меняются, моделируя неоднородность и смешение естественных текстур.

При удачном выборе ОВ (показана крестиками на рис. 1), представляющей наиболее характерные «чистые» области классов, точность классификации с использованием и без использования частичного обучения составляет 99.77% и 99.09% соответственно. Однако, в случае неудачного выбора ОВ (показана кружочками на рис. 1) в результате классификации без применения частичного обучения граница классов значительно смещается (показана черной линией), и точность составляет 85.52%. В то время как точность классификации предложенным алгоритмом с частичным обучением составляет 99.22% (граница показана белой линией).

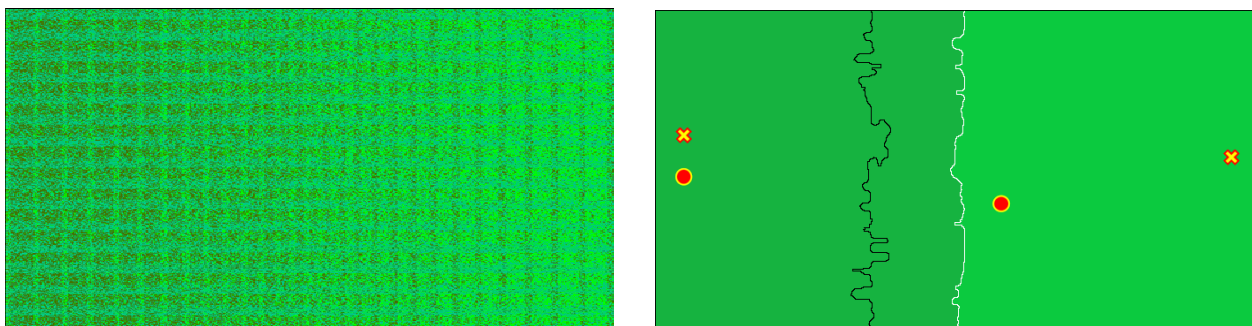


Рис. 1. Модельное изображение из двух текстурных классов и результат спектрально-текстурной классификации с применением частичного обучения (белая линия границы) и без него (черная линия) с нанесенными метками ОВ

На рис. 2, а представлено мультиспектральное спутниковое изображение высокого пространственного разрешения WorldView-2 с нанесенными метками ОВ (7 меток, по одной на каждый класс). На снимке присутствуют два спектрально-неоднородных класса: лиственный лес (светло-зеленый) и хвойный лес (темно-зеленый). На рис. 2, б и в представлены результаты спектрально-текстурной классификации без и с применением частичного обучения соответственно. Применение частичного обучения позволило заметно повысить качество классификации в условиях непредставительной ОВ. Размер изображения – 2048×2048 пикселей. При обработке использовались каналы: 1, 4, 6, 7. Время работы алгоритма составило 0.53 с.

Результаты экспериментов на модельных и спутниковых изображениях высокого пространственного разрешения показали, что предложенный метод классификации с частичным обучением позволяет сократить зависимость от качества представителей обучающей выборки и повысить качество результатов спектрально-текстурной классификации в условиях ограниченной ОВ.

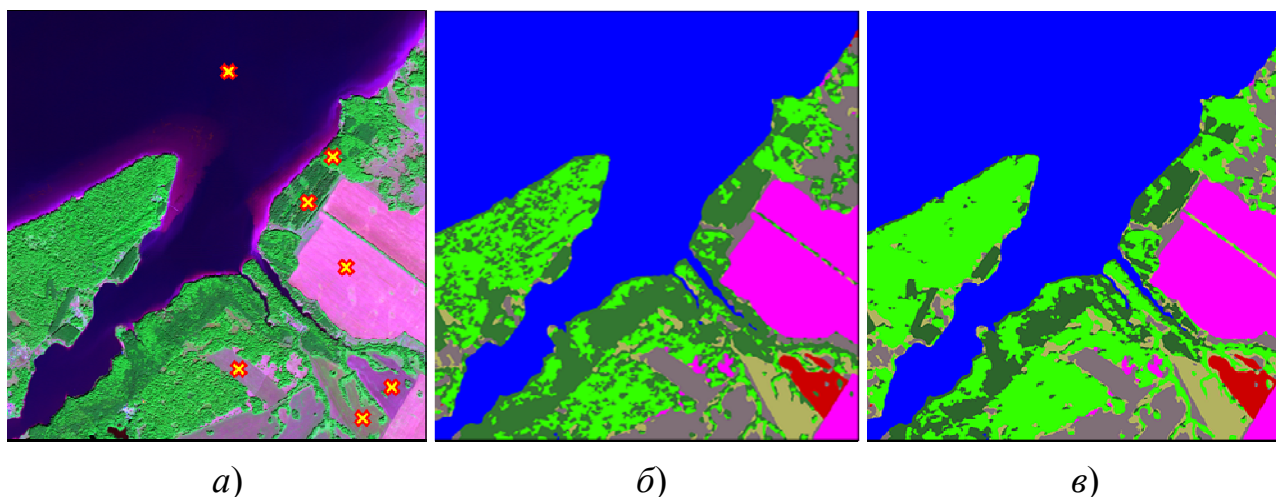


Рис. 2. RGB-композит снимка WorldView-2 (каналы: 4, 7, 1) с нанесенными метками ОБ (а); результат спектрально-текстурной классификации без применения частичного обучения (б) и с частичным обучением (в)

Заключение

В работе предложен новый алгоритм спектрально-текстурной классификации изображений с частичным обучением. Применение концепции частичного обучения дает возможность эффективно выделять спектрально неоднородные классы, даже если имеющаяся выборка мала и слабо отражает разнородную структуру классов. Вычислительная эффективность предложенного алгоритма обеспечивает его применимость к мульти-спектральным спутниковым изображениям большого размера. Все это позволяет облегчить работу пользователя и максимально автоматизировать процесс тематической обработки спутниковых снимков высокого пространственного разрешения.

Благодарности

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (грант № 18-37-00492-мол_а).

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Schwenker F., Trentin E. Pattern classification and clustering: A review of partially supervised learning approaches // Pattern Recognition Letters. – 2014. – Vol. 37. – P. 4-14.
2. Hady M.F.A., Schwenker F. Semi-supervised learning // Handbook on Neural Information Processing. – Springer Berlin Heidelberg, 2013. – P. 215-239.
3. Адаскина Ю. В., Попов А. М., Реброва П. В. Метод полуавтоматической классификации для данных с несбалансированными классами // Компьютерная лингвистика и вычислительные онтологии. – 2015. – С. 15-24.
4. Травкин О. И. Подходы к агрегации данных и извлечению факторов в задаче поиска мошенничества в банковских транзакциях // Сборник трудов межд. конференции «Аналити-

ка и управление данными в областях с интенсивным использованием данных». – 2016. – С. 361-369.

5. Anand S. et al. Semi-supervised kernel mean shift clustering // IEEE transactions on pattern analysis and machine intelligence. – 2014. – Vol. 36. – N. 6. – P. 1201-1215.

6. Jing X., Chen S.Y., Fan L.L. Semi-supervised classification of multi-spectral images based on density: selected samples // Ninth International Conference on Digital Image Processing (ICDIP 2017). – International Society for Optics and Photonics, 2017. – Vol. 10420. – P. 1042030.

7. Tan K. et al. An efficient semi-supervised classification approach for hyperspectral imagery // ISPRS Journal of Photogrammetry and Remote Sensing. – 2014. – Vol. 97. – P. 36-45.

8. Banerjee B., Buddhiraju K.M. A Novel Semi-Supervised Land Cover Classification Technique of Remotely Sensed Images // Journal of the Indian Society of Remote Sensing. – 2015. – Vol. 43. – N. 4. – P. 719-728.

9. Wang L. et al. Semi-supervised classification for hyperspectral imagery based on spatial-spectral label propagation // ISPRS Journal of Photogrammetry and Remote Sensing. – 2014. – Vol. 97. – P. 123-137.

10. Luo R. et al. Spectral-spatial classification of hyperspectral images with semi-supervised graph learning // SPIE Remote Sensing. – International Society for Optics and Photonics, 2016. – P. 100040T-100040T-6.

11. Yang L. et al. Semi-supervised hyperspectral image classification using spatio-spectral Laplacian support vector machine // IEEE Geoscience and Remote Sensing Letters. – 2014. – Vol. 11. – N. 3. – P. 651-655.

12. Schiwe J. Segmentation of high-resolution remotely sensed data-concepts, applications and problems // International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences. – 2002. – Vol. 34. – N. 4. – P. 380-385.

13. Рылов С. А., Мельников П. В., Пестунов И. А. Спектрально-текстурная классификация гиперспектральных изображений высокого пространственного разрешения // Интерэкспо ГЕО-Сибирь-2016. XII Междунар. науч. конгр. : Междунар. науч. конф. «Дистанционные методы зондирования Земли и фотограмметрия, мониторинг окружающей среды, геоэкология»: сб. материалов в 2 т. (Новосибирск, 18–22 апреля 2016 г.). – Новосибирск : СГУГиТ, 2016. Т. 1. – С. 78–84.

14. Wang A., Wang S., Lucieer A. Segmentation of multispectral high-resolution satellite imagery based on integrated feature distributions // International Journal of Remote Sensing. – 2010. – Vol. 31. – N. 6. – P. 1471-1483.

15. Пестунов И. А., Рылов С. А. Алгоритмы спектрально-текстурной сегментации спутниковых изображений высокого пространственного разрешения // Вестник КемГУ. – 2012. – № 4/2 (52). – С. 104-110.

16. Рылов С. А. Методы и алгоритмы сегментации мультиспектральных спутниковых изображений высокого пространственного разрешения: автореферат дис. ... кандидата технических наук: 05.13.18. – 2016. – Новосибирск, 2016. – 20 с.

17. Junior J. J. M., Backes A. R., Cortez P. C. Color texture classification based on gravitational collapse // Pattern Recognition. – 2013. – Vol. 46. – N. 6. – P. 1628-1637.

18. Yuan J., Wang D. L., Li R. Remote sensing image segmentation by combining spectral and texture features // IEEE Transactions on Geoscience and Remote Sensing. – 2014. – Vol. 52. – N. 1. – P. 16-24.

19. Пестунов И. А., Рылов С. А., Бериков В. Б. Иерархические алгоритмы кластеризации для сегментации мультиспектральных изображений // Автометрия. – 2015. – Т. 51. – № 4. – С. 12-22.

20. Jain A. K., Murty M. N., Flynn P. J. Data clustering: a review // ACM computing surveys (CSUR). – 1999. – Vol. 31. – N. 3. – P. 264-323.

© С. А. Рылов, 2019