

УДК 528.873

М.Ю. Катаев, А.В. Богомолов

## Особенности кластеризации многоспектральных изображений спутникового прибора Landsat

Исследуются особенности тематической обработки данных дистанционного зондирования Земли, полученных в различных диапазонах оптического спектра с помощью прибора Landsat-8. Кластеризация данных многоспектральной съемки, основанная на измерениях в первых каналах спутникового прибора (пространственное разрешение 30 м), не позволяет с высокой точностью выделять типы поверхности за счет неоднородности их смешивания. Изучение проблемы выделения однородных участков по различным признакам и является предметом данной статьи. Приведены результаты обработки реальных данных спутниковых измерений. Полученные результаты показывают хорошую согласованность кластеризованных данных с известным мировым аналогом GlobeLand30.

**Ключевые слова:** дистанционное зондирование, обработка спутниковых изображений, многоспектральный, кластеризация.

**doi:** 10.21293/1818-0442-2018-21-2-54-59

В последнее десятилетие мониторинговые методы дистанционного зондирования Земли (ДЗЗ) из космоса являются одними из важнейших подходов к исследованию параметров атмосферы и поверхности Земли [1, 2]. Особую значимость имеют направления, связанные с изучением поверхности Земли, позволяющие изучать природные и техногенные объекты, а также, изменения, которые связаны с объектами. Для огромных территорий России единственно возможными для контроля экологического состояния, решения производственных и научных задач и др. являются данные мониторинга ДЗЗ. Получаемые спутниковые данные проходят несколько этапов обработки: предварительную, тематическую и анализ [1].

На этапе предварительной обработки разработано множество методов, которые привязаны к каждому спутнику отдельно, так как учитывают индивидуальные характеристики орбиты и измерительного прибора. Методы тематической обработки являются более унифицированными и разбиты на группы, связанные с размерностью измерений (пространство, время, спектральный канал). Одним из наиболее развитых подходов тематической обработки является формирование попиксельных изображений, где каждое изображение связано с характеристиками спектральных каналов, и их совместный анализ. Анализ возможен независимо для каждого изображения или их функциональных преобразований (индексов) [1].

Важным направлением исследований является выделение типов поверхности на спутниковом снимке и изучение свойств выделенных классов. Для спутниковых снимков с низким пространственным разрешением (более 100 м) выделение классов проходит более успешно, чем для среднего (10–50 м). Это связано с тем, что в первом случае за счет осреднения различных типов поверхности, попадающих в пиксель, классы являются более однородными, чем для второго случая. Однако более высокое пространственное разрешение позво-

ляет получить дополнительные возможности при изучении пространственно-временных изменений типов поверхности.

### Постановка задачи

При тематической обработке данных среднего пространственного разрешения типично проводят выделение однородных участков по спектральным или иным признакам. Процедура выделения таких участков связана с кластеризацией (сегментацией) спутниковых изображений [3–8]. Проблемой выделения однородных участков по различным признакам является обнаружение неперекрывающихся классов. Однако выполнение этого условия для реальных спутниковых данных весьма проблематично, поскольку для естественных ландшафтов сказывается рельеф (наклон поверхности относительно точки освещения Солнцем и наблюдением), а также типы поверхности, попадающие в пиксель с разной площадью. Нельзя забывать и про влияние состояния атмосферы (освещенность, замутненность, наличие дымки, облачности и т.д.). Эти аспекты, а также их естественные суточные и сезонные изменения делают задачу кластеризации весьма сложной.

Кластеризация спутниковых изображений заключается в разбиении изображения на непересекающиеся пространственные области на основе близости их спектральных, индексных (например, вегетационных, водных и др.), пространственных или пространственно-временных характеристик (например, текстурных).

В настоящее время известны и широко используются на практике методы кластеризации, позволяющие разбивать спутниковые изображения на классы, соответствующие различным типам поверхности (лес, вода, город, поле и др.). Существует достаточно много подходов к кластеризации: иерархические или неиерархические методы, четкие или нечеткие, масштабируемые или немасштабируемые, с выбором числа и типа кластеров или нет, последовательные или параллельные,

отличающиеся по выбору меры расстояния (Евклидово, городские кварталы, Чебышева, степенное и др.), метрики (односвязные, полносвязные, взвешенные и др.) и др. При решении практических задач существуют особенности, которые необходимо учитывать при выполнении процедуры кластеризации. Выбор метода кластеризации из множества подходов связан с их эффективностью при решении конкретной задачи. При этом необходимо учитывать проблему выбора метода, меры расстояния, числа кластеров, формы кластеров, их пересечения и др.

В случае кластерного анализа спутниковых изображений отсутствуют априорные сведения о количестве классов (типов поверхности), спектральных и пространственных характеристик (возможного наличия в пикселе нескольких типов поверхности). Известные программные продукты, такие как Multispec [<https://engineering.purdue.edu/~biehl/MultiSpec/>], ERDAS Imagine [[www.mapinfo.ru/product/erdas](http://www.mapinfo.ru/product/erdas)], ENVI [HarrisGeospatial.com], ArcGIS [<https://www.arcgis.com/>], решают задачи кластеризации для спутниковых приборов различного пространственного разрешения, учитывают особенности для исследуемой территории Земли (определяются типы поверхности, углы освещения Солнцем, состояние атмосферы и др.). Однако для получения качественного результата требуется значительный вклад ручного труда и знаний пользователя. Наиболее распространенный подход к кластеризации мультиспектральных спутниковых изображений в отсутствие обучающей выборки основан на применении алгоритмов, использующих пространство спектральных признаков изображения (спектральных каналов).

Изучая опыт практической реализации различных авторов [2–16], нами выбран алгоритм кластеризации k-средних. Число кластеров нами было ограничено и задано 12, учитывая опыт построения баз данных типов поверхности. Целью является изучение различных аспектов кластеризации для понимания процессов, необходимых для разработки автоматизированной процедуры кластеризации.

#### **Спутниковый прибор Landsat-8**

Использование измеренных изображений прибором Landsat [<https://landsat.usgs.gov/>] в научных и производственных целях начинается с 1970 г. Этому способствует несколько причин: 1) близкие по числу и спектральным свойствам характеристики спектральных каналов приборов, 2) достаточно высокое пространственное разрешение (30 м в видимом и ближнем инфракрасном спектральном диапазоне) и 3) наличие панхроматического канала (15 м). Спектральные каналы приборов Landsat позволяют выделять разнообразные типы поверхности, в том числе и наземную растительность. Однако результаты анализа литературных источников говорят, что имеются некоторые отличия между различными ис-

следованиями одной и той же территории. Поэтому важным элементом для исследований остается вопрос точности выделения типов поверхности.

#### **Кластеризация спутниковых изображений**

Особенностью многоспектральной спутниковой измерительной техники Landsat является возможность наблюдать поверхность Земли в любое время суток, используя каналы видимого, ближнего ИК или ИК диапазона спектра. Основой измерений является спектральная яркость типов поверхности, попадающих в пиксель, которая определяется коэффициентами отражения типов поверхности и их площадью в пикселе. Автоматическое распознавание типов поверхности на спутниковом изображении необходимо для составления разнообразных карт, включающих в себя информацию о растительности, почвах, водных объектах, городах и др.

Сравнивая карты за различные промежутки времени, удаётся выявлять изменения формы, площади и свойств типов поверхности за счёт естественных или антропогенных процессов. На первом этапе тематической обработки спутниковых изображений выполняется кластеризация изображения, т.е. выделение отдельных участков, характеризующихся однородностью внутри участка и существенными отличиями между участками. На втором этапе обработки выделенные участки подлежат классификации или определению типа поверхности (лес, вода, город или др.).

Для решения задач кластеризации необходимо выбрать множество признаков, характеризующих изображение, что определяет набор методов исследования. Одним из наборов признаков могут быть измеренные значения спектральной яркости в каждом выбранном канале. Эти значения могут быть усреднены на некоторой выбранной области (например, области, состоящей из 3×3 пикселей, что ведет к понижению пространственного разрешения). Если используется попиксельный подход, то каждый кластер состоит из пикселей, представленных на изображении. На основе спектральных каналов можно рассчитать тот или иной индекс, например вегетационный, который и будет являться изображением, подлежащим анализу. Можно использовать не только спектральные, но и пространственные характеристики, используя, например, текстурные или морфологические характеристики, и др.

Важным направлением является переход из пространства спектральных признаков в пространство цветовых признаков, например RGB. В этом случае можно использовать классические подходы обработки изображений. Поэтому, поиск условий для автоматизации процесса кластеризации является важным.

#### **Анализ результатов кластеризации**

Нами выполняются работы по анализу пространственно-временных данных спутникового прибора MODIS, пространственное разрешение

пикселя которого составляет 250, 500 м [17, 18]. В некоторых научных и практических задачах возникает необходимость более детального рассмотрения процессов изменения типов поверхности, и для этих целей применяют данные прибора Landsat.

В данной работе для целей кластеризации рассматривается алгоритм *k*-средних (*k*-means), суть которого связана с разбиением всех пикселей на кластеры таким образом, чтобы минимизировать сумму расстояний от каждого пикселя до соответствующих им центров кластеров. Параметрами кластера могут быть: значения RGB-каналов синтезированного изображения, значения спектральных каналов или индексов (например, вегетационный индекс). Множество пикселей изображения делится на *k* кластеров, сама кластеризация выполняется за счёт смещения центров для поиска из устойчивого положения. В качестве спутниковых данных нами были выбраны спутниковые изображения Landsat-8. Разбиение на кластеры проводится на основе расстояний между пикселями изображения, на которое потом накладывается определенное ограничение, разделяющее кластеры между собой. Примерами функций, определяющими расстояние, могут быть евклидово расстояние (1), или Чебышевская мера (2), представленные формулами:

$$D(i,k) = \sqrt{\sum_{j=1}^N (I(i,j) - I(j,k))^2}, \quad (1)$$

$$D(i,k) = \max_{1 \leq j \leq N} |I(i,j) - I(j,k)|, \quad (2)$$

где *i* – индекс текущего наблюдения, *k* – индекс кластера, *N* – количество признаков цветового пространства (*N* = 3) и **I** = {*R*, *G*, *B*} – вектор в выбранном цветовом пространстве.

На рис. 1 показан результат кластеризации синтезированного спутникового изображения

(*R* = 4, канал Landsat-8, *G* = 3, *B* = 2) для территории, расположенной рядом с г. Томском, полученный в июне 2016 г., на основе выражения (1). Число обнаруженных классов равно 12. Из рис. 1 видно, что имеются однородные области, связанные с водой – река Томь, участками города, сельскохозяйственными и заливными полями, лесом. В большей своей части пиксели того или иного участка соответствуют определенным типам поверхности, однако имеется много артефактов. Например, на поверхности воды обнаруживаются пиксели, характеристики которых совпадают с пикселями, расположенными на поверхности Земли. Это всё приводит к выводу, что помимо цветовых признаков необходимо использовать спектральные.



Рис. 1. Пример кластеризации синтезированного спутникового изображения для спектральных каналов 4-3-2

На рис. 2 представлены спектральные кривые коэффициентов отражения в первых семи каналах прибора Landsat-8.

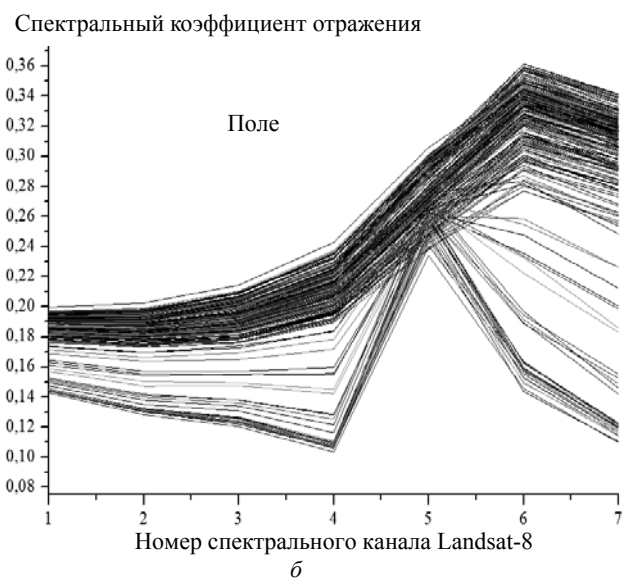
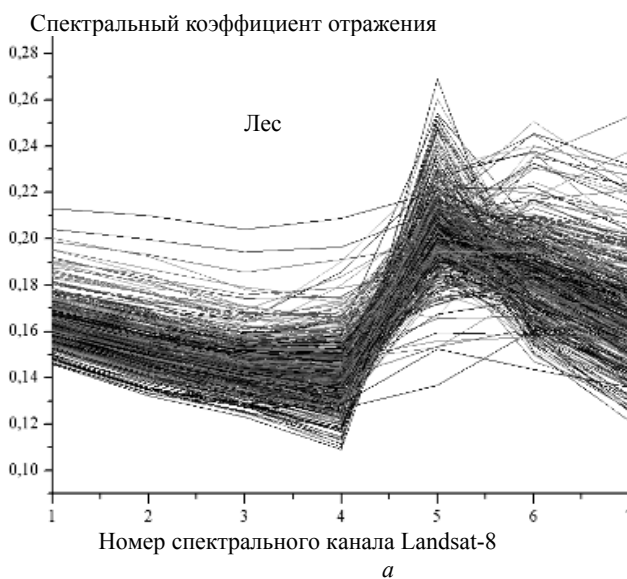


Рис. 2. Значения спектрального коэффициента отражения для каждого из семи спектральных каналов Landsat-8 для области 15×15 пикселей (450×450 м) для леса – *a* и поля – *б*

Нами были выбраны однородные участки по типам поверхности размером 15×15 пикселей и для каждого из них выбраны значения спектральных коэффициентов отражения. Хорошо видно, что вариации значений спектральных кривых для всех пикселей весьма значительны, а также явно просматриваются разные типы поверхности в выбранной области. Это говорит о том, что в части пикселей присутствует несколько типов поверхностей с разной площадью, что изменяет типичное для определенного типа поверхности распределение значений (см. рис. 2). Например, на рис. 2, а в области класса поля в нескольких пикселях присутствуют деревья (нижняя часть рисунка), хотя большая часть пикселей принадлежит классу поля.

Учитывая специфику распределения значений коэффициентов отражения, нами выделены средние значения коэффициентов отражения для нескольких классов: поле, вода, город, лес и др. Далее проведена кластеризация согласно заданному количеству и типу классов.



Рис. 3. Сравнение полученных результатов кластеризации – а и данных глобальной базы данных типов поверхности GlobeLand30 – б, по данным Landsat-8

В целом полученные на рис. 3 данные показывают хорошее согласие по основным типам поверхности, которые выделены нами и которые присутствуют в известной глобальной базе данных поверхности GlobeLand30.

#### Заключение

В работе рассмотрены основные особенности, возникающие в задаче многоспектральной кластеризации спутниковых изображений. Выполнен обзор методов кластеризации. Указаны условия, по которым спутниковые изображения среднего пространственного разрешения (10–50 м) можно кластеризовать. Выбран широко известный метод кластеризации – k-средних с евклидовой метрикой, который хорошо работает при обработке данных большого объема. Полученные результаты показывают на хорошую согласованность кластеризованных данных в работе с известным мировым аналогом GlobeLand30.

Работа выполнена в центре космического мониторинга ТУСУРа и в рамках государственного

Результат приведен на рис. 3, а. Сравнение полученных нами результатов кластеризации и глобальной базы данных типов поверхности GlobeLand30 [http://www.global-landcover.com] по данным Landsat-8 показано на рис. 3. Из рис. 3 видно, что в глобальной базе данных проводилось существенное увеличение пространственного разрешения для достижения однородности областей (увеличение площади одного из типов поверхности по отношению к другому). В выполненных нами расчетах такого осреднения не сделано, что приводит к появлению множества мелких областей с одним из выделенных типов поверхности. Обнаружено, что в отражении воды р. Томи присутствуют участки с разной глубиной, и это приводит к появлению не одного, а нескольких типов водной поверхности. Поэтому наличие точных значений участков глубины, полученных, например, с помощью эхолота, позволит провести калибровку спутниковых данных водной поверхности.

задания Министерства образования и науки РФ, проект № 8.8184.2017/8.9 «Методология создания систем энергогенерирующих и энергопреобразующих устройств для наземных и бортовых комплексов наземного, космического и подводного базирования».

#### Литература

1. Бондур В.Г. Основы аэрокосмического мониторинга окружающей среды: курс лекций. – М.: МИИГАиК, 2008. – 546 с.
2. Бондур В.Г. Современные подходы к обработке больших потоков гиперспектральной и многоспектральной аэрокосмической информации // Исследование Земли из космоса. – 2014. – № 1. – С. 4–16.
3. Burnett C. Multi-scale segmentation/object relationship modelling methodology for landscape analysis / C. Burnett, T.A. Dlaschke // Ecological Modelling. – 2003. – Vol. 168, No. 3. – PP. 233–249.
4. Chen G. Object-based change detection / G. Chen, G.J. Hay, L.M.T. Carvalho, M.A. Wulder // International Journal of Remote Sensing. – 2012. – Vol. 33, No. 14. – PP. 4434–4457.

5. Haralick R.M. Textural features for image classification / R.M. Haralick, K. Shanmugam, I. Dinstein // *IEEE Transactions on Systems, Man and Cybernetics*. – 1973. – Vol. 3, No. 6. – PP. 610–621.

6. Сидорова В.С. Оценка качества классификации многоспектральных изображений гистограммным методом // *Автометрия*. – 2007. – Т. 43, № 1. – С. 37–43.

7. Halkidi M. On clustering validation techniques / M. Halkidi, Y. Batistakis, M. Vazirgiannis // *Journal of Intelligent Information Systems*. – 2001. – No. 17. – PP. 107–132.

8. Chen C.H. The Handbook of Pattern Recognition and Computer Vision / C.H. Chen, L.F. Pau, P.S.P. Wang. – Singapore: World Scientific Publishing Co, 1998. – 1004 p.

9. Magnussen S. Contextual classification of Landsat TM images to forest inventory cover types / S. Magnussen, P. Boudewyn, M. Wulder // *Int. J. Remote Sens.* – 2004. – No. 25. – PP. 2421–2440.

10. Duda T. Unsupervised classification of satellite imagery: choosing a good algorithm / T. Duda, M.J. Canty // *Int. J. Remote Sens.* – 2002. – No. 23. – PP. 2193–2212.

11. Li C. Comparison of classification algorithms and training sample sizes in urban land classification with Landsat Thematic Mapper imagery / C. Li, J. Wang, L. Wang, L. Hu, P. Gong // *Remote Sens.* – 2014. – No. 6. – PP. 964–983.

12. Li M. A review of remote sensing image classification techniques: The role of spatio-contextual information / M. Li, S.Y. Zang, B. Zhang, S.S. Li, C.S. Wu // *Eur. J. Remote Sens.* – 2014. – No. 47. – PP. 389–411

13. Асмус В.В. Контролируемая классификация данных дистанционного зондирования Земли / В.В. Асмус, А.А. Бучнев, В.П. Пяткин // *Автометрия*. – 2008. – № 4. – С. 60–67.

14. Книжников Ю.Ф. Аэрокосмические методы географических исследований / Ю.Ф. Книжников, В.И. Кравцова, О.В. Тутубалина. – М.: Академия, 2004. – 336 с.

15. Лурье И.К. Теория и практика цифровой обработки изображений. Дистанционное зондирование и географические информационные системы / И.К. Лурье, А.Г. Косиков. – М.: Научный мир, 2003. – 168 с.

16. Yonggang L. PHA: A fast potential-based hierarchical agglomerative clustering method / L. Yonggang // *Patt. Recogn.* – 2013. – Vol. 46, No. 5. – PP. 1227–1239.

17. Xu R., Wunsch D. I. Survey of clustering algorithms // *IEEE Trans. Neural Networks*. – 2005. – Vol. 16, No. 3. – PP. 645–678.

18. Катаев М.Ю. Обнаружение экологических изменений природной среды по данным спутниковых измерений / М.Ю. Катаев, А.А. Бекеров // *Оптика атмосферы и океана*. – 2014. – Т. 27, № 7. – С. 652–656.

19. Катаев М.Ю. Интернет-информационная система накопления, обработки и анализа спутниковых данных MODIS / М.Ю. Катаев, А.А. Бекеров, А.К. Лукьянов // *Доклады ТУСУР*. – 2015. – Т. 35, № 1. – С. 93–99.

#### Катаев Михаил Юрьевич

Д-р техн. наук, профессор каф. автоматизированных систем управления систем (АСУ)

Томского государственного университета систем управления и радиоэлектроники (ТУСУР)

Ленина пр-т, д. 40, г. Томск, Россия, 634050

Тел.: (382-2) 70-15-36

Эл. почта: kmy@asu.tusur.ru

#### Богомолов Александр Владимирович

Магистрант 2-го курса каф. АСУ ТУСУР

Ленина пр-т, д. 40, г. Томск, Россия, 634050

Тел.: +7-951-587-04-08

Эл. почта: alex1.bogomolov@gmail.com

Kataev M.Yu., Bogomolov A.V.

#### Features of clustering the multispectral satellite Landsat images

In the article, features of thematic processing of Earth remote sensing data obtained in different ranges of the optical spectrum using the Landsat-8 instrument are studied. The clustering of multispectral survey data, based on measurements in the first channels of a satellite instrument, with a resolution of 30 m, does not allow the high-precision determination of surface types due to heterogeneity of their mixing. The study of these aspects is the subject of this article. The results of real data processing of satellite measurements are presented.

**Keywords:** remote sensing, satellite image processing, multispectral, clustering.

**doi:** 10.21293/1818-0442-2018-21-2-54-59

#### References

1. Bondur V.G. *Osnovy aerokosmicheskogo monitoringa okruzhayushchey sredy. Kurslektsiy* [Fundamentals of airspace environment monitoring. Lecture course]. Moscow, MII GAIK Publ., 2008. 546 p. (In Russ.)

2. Burnett C., Dlaschke T.A Multi-scale segmentation/object relationship modelling methodology for landscape analysis. *Ecological Modelling*, 2003, vol. 168, no. 3, pp. 233–249.

3. Chen G., Hay G.J., Carvalho L.M.T., Wulder M.A. Object-based change detection. *International Journal of Remote Sensing*, 2012, vol. 33, no. 14, pp. 4434–4457.

4. Bondur V.G. Modern approaches to the processing of huge hyperspectral and multispectral airspace data flow. *Issledovanie Zemli iz kosmosa*, 2014, no. 1, pp. 4–16 (In Russ.).

5. Haralick R.M., Shanmugam K., Dinstein I. Textural features for image classification. *IEEE Trans. Syst. Man and Cybernetics*, 1973, vol. 3, no. 6, pp. 610–621.

6. Sidorova V.S. Evaluation of the quality of classification of multispectral images by the histogram method. *Avtometriya*. 2007, vol. 43, no. 1, pp. 37–43 (In Russ.).

7. Halkidi M., Batistakis Y. and Vazirgiannis M. On clustering validation techniques. *Journal of Intelligent Information Systems*, 2001, no. 17, pp. 107–132.

8. Chen C.H., Pau L.F., Wang P.S.P. The Handbook of Pattern Recognition and Computer Vision. World Scientific Publishing Co, 1998. 1004 p.

9. Magnussen S., Boudewyn P., Wulder M. Contextual classification of Landsat TM images to forest inventory cover types. *Int. J. Remote Sens.*, 2004, no. 25, pp. 2421–2440.

10. Duda T., Canty M.J. Unsupervised classification of satellite imagery: choosing a good algorithm. *Int. J. Remote Sens.*, 2002, no. 23, pp. 2193–2212.

11. Li C., Wang J., Wang L., Hu L., Gong P. Comparison of classification algorithms and training sample sizes in urban land classification with Landsat Thematic Mapper imagery. *Remote Sens.*, 2014, no. 6, pp. 964–983.

12. Li M., Zang S.Y., Zhang B., Li S.S., Wu C.S. A review of remote sensing image classification techniques: The

role of spatio-contextual information. *Eur. J. Remote Sens.*, 2014, no. 47, pp. 389–411.

13. Asmus V.V., Buchnev A.A., Pyatkin V.P. Controlled classification of remote sensing data of the Earth, *Avtometriya*, 2008, no. 4, pp. 60–67.

14. Knizhnikov Yu.F., Kravtsova V.I., Tutubalin O.V. Aerospace methods of geographical research. Moscow, Academy, 2004. 336 c.

15. Lurie I.K. *Theory and practice of digital image processing. Remote sensing and geographic information systems* / I.K. Lurie, A. G. Kosikov. Moscow, The scientific world, 2003. 168 p.

16. Yonggang L. PHA: A fast potential-based hierarchical agglomerative clustering method. *Patt. Recogn.*, 2013, vol. 46, no. 5, pp. 1227–1239.

17. Xu R., Wunsch D.I. Survey of clustering algorithms. *IEEE Trans. Neural Networks*, 2005, vol. 16, no. 3, pp. 645–678.

18. Kataev M.Yu. Obnaruzhenie ekologicheskikh izmeneniy prirodnoi sredy po dannym sputnikovyh izmerenij. *Optika atmosfery i okeana*, 2014, vol. 27, no. 7, pp. 652–656 (In Russ.)

19. Kataev M.Yu., Bekerov A.A., Lukyanov A.K. Internet information system for the accumulation, processing and analysis of satellite data MODIS. *Proceedings of TUSUR University*, 2015, vol. 35, no. 1, pp. 93–99 (In Russ.)

---

**Mikhail Yu. Kataev**

Doctor of Engineering Sciences, professor,  
Department of Control Systems, Tomsk State University  
of Control Systems and Radioelectronics (TUSUR)  
40, Lenina pr., Tomsk, Russia, 634050  
Phone: +7 (382-2) 70-15-36  
Email: kmy@asu.tusur.ru

**Alexander V. Bogomolov**

Master student, Department of Control Systems, TUSUR  
40, Lenina pr., Tomsk, Russia, 634050  
Phone: +7-951-587-04-08  
Email: alex1.bogomolov@gmail.com