

УДК 630\*587

И. В. Толкач, доцент (БГТУ); А. А. Пушкин, доцент (БГТУ);  
С. В. Ковалевский, ассистент (БГТУ)

## ФОРМИРОВАНИЕ ОБУЧАЮЩИХ ВЫБОРОК ДЛЯ КЛАССИФИКАЦИИ ЛЕСНЫХ ЗЕМЕЛЬ ПО СНИМКАМ ВЫСОКОГО РАЗРЕШЕНИЯ НА ОСНОВЕ ПОВЫДЕЛЬНОЙ БАЗЫ ДАННЫХ

Приведены основные подходы к классификации изображений космических снимков высокого разрешения, описана оригинальная методика формирования обучающих выборок на основе повыделельной базы данных ГИС и цифровых планово-картографических лесоустроительных материалов путем параметрической оценки статистических показателей распределения яркостей различных спектральных каналов.

Some basic approaches of picture classification of high resolution space images are given in the article. Also a unique methodic of forming of so called learning extracts are solely described. The methodic is based on GIS standwisely collected forest database as well as digital plan and cartographic forest inventory data. The proposed methodic uses a parametric evaluation procedure of some statistical indexes of different spectral channels brightness distribution.

**Введение.** Классификация изображений – сложный и неоднозначный процесс, заключающийся в том, чтобы на основе спектральной информации из различных диапазонов проанализировать каждый пиксель и отнести его к тому или иному классу. Таким образом, целью классификации изображений является распределение пикселей снимка по определенным классам объектов. Результирующее изображение является, по существу, тематической картой.

При проведении классификации различают информационные и спектральные классы. Информационные классы – это те объекты, которые необходимо распознать на снимке: различные виды растительности, определенные геологические структуры, виды земель и т. д. Спектральный класс – это группа пикселей, обладающих приблизительно одинаковой яркостью в некотором спектральном диапазоне. Одна из основных целей классификации состоит в том, чтобы совместить спектральные классы с информационными [1].

Выделяют два основных направления классификации – контролируемую (с обучением) и неконтролируемую. При неконтролируемой классификации спектральные классы выделяются на основании только той информации, которая представлена в данных космической съемки, и лишь затем оператор пытается сопоставить их с реальными пространственными объектами. Для контролируемой классификации используются эталонные области, которые выбираются оператором в соответствии с их принадлежностью к определенному информационному классу. Используется также комбинация различных методов классификации, что дает оптимальные результаты, особенно для значительных массивов данных [2].

Основными методами, дающими приемлемые результаты, являются параметрические ме-

тоды классификации с обучением, в основе которых лежит оценка спектральных и яркостных показателей отдельных пикселей цифровых изображений снимков и определение их принадлежности к тому или иному классу. Для определения возможности разделения на классы используют как статистические, так и графические методы анализа, при этом различные комбинации спектральных диапазонов ранжируются в зависимости от возможности разделения классов. Наиболее часто для решения этой задачи используют численные статистические методы. Основными этапами классификации являются:

- геометрическая и цветовая коррекция изображения снимка, трансформирование и привязка снимка к системе координат;
- формирование эталонных обучающих выборок;
- оценка однородности эталонов и разделимости классов, формирование спектральных библиотек;
- классификация изображения;
- оценка результатов классификации;
- векторизация и формирование тематических карт [2].

Все этапы имеют важное значение при интерпретации изображений снимков, однако этап формирования эталонных обучающих выборок при контролируемой классификации имеет особое значение. Основными требованиями к данным обучающих выборок являются их однородность и разделимость. Существуют различные методы формирования обучающих выборок, которые требуют различных временных и финансовых затрат.

Одним из способов является закладка эталонно-калибровочных участков (ЭКУ). Количество ЭКУ определяется количеством классов и изменчивостью спектральных и яркостных показателей в пределах класса. Недостатком

такого метода являются большие затраты, необходимые для выполнения полевых работ по закладке ЭКУ, поэтому его более целесообразно использовать для оценки результатов классификации и в случаях, если имеются неклассифицированные участки.

Более эффективным и не требующим больших затрат способом является использование имеющихся вторичных данных. В отрасли лесного хозяйства практически на все лесохозяйственные учреждения республики имеются векторные карты и повыделльные базы данных, содержащие таксационные показатели насаждений. Основная проблема при их использовании заключается в том, что в связи с интенсивной хозяйственной деятельностью, природными катаклизмами и естественным ростом насаждений информация в базах данных устаревает и требует актуализации.

Целью исследования стала разработка методики формирования обучающих выборок на основе повыделльной базы данных и планово-картографических лесоустроительных материалов.

Работа выполнялась сотрудниками кафедры лесоустройства (БГТУ) совместно с сотрудниками отдела приема и обработки космической информации (РУП «Белгослес») в рамках задания по теме «Разработать технологию комплексной обработки данных дистанционного зондирования Земли для решения задач лесного хозяйства», подпрограмма «Применение космической информации в интересах лесного хозяйства». Национальной программы исследования и использования космического пространства в мирных целях. Для формирования обучающих выборок использовались материалы дистанционного зондирования высокого разрешения: *ALOS AVNIR* – 10 м (мультиспектральная съемка), *ALOS PRISM* – 2,5 (панхроматическая съемка). В качестве основного программного продукта использовался программный комплекс *ENVI*. При работе с векторными данными использовались ГИС *MapInfo* и *ArcView*.

**Основная часть.** В представленной работе на примере части территории ГЛХУ «Березинский лесхоз» была выполнена контролируемая классификация с использованием обучающих выборок, сформированных на основе картографической и повыделльной баз данных ГИС «Лесные ресурсы». Для этого была разработана методика формирования обучающих выборок, в соответствии с которой предусматривалось решение следующих основных задач:

- привязка векторных слоев выделов и снимков;
- формирование повыделльной базы данных и ее конвертация в формат поддерживаемый программным комплексом *ENVI*;

- оценка спектральных и яркостных показателей изображений выделов, статистический анализ их изменчивости и формирование базы данных;

- группировка выделов по классам, оценка однородности сформированных классов, удаление выделов, ошибочно попавших в класс;

- оценка спектральных и яркостных показателей классов, формирование обучающих выборок и спектральных библиотек;

- классификация изображения, формирование векторных слоев;

- закладка и таксация насаждений на эталонно-калибровочных участках;

- оценка точности классификации на основе данных выборочной таксации эталонно-калибровочных участков.

Отбор выделов для формирования обучающих выборок выполнялся в программном комплексе *ENVI* на основе векторного слоя выделов и повыделльной базы данных насаждений на участке изображения ГЛХУ «Березинский лесхоз». Из базы данных выбирались все выделы, относящиеся к данному классу. Выборки формировались по преобладающим породам с различной долей участия породы и возрастным группам. Всего отобрано пять наиболее представленных на данной территории пород деревьев (ель, сосна, береза, осина и ольха черная), которые были разделены на четыре возрастные группы: молодняки, средневозрастные, приспевающие и спелые. Выборки выполнены также и для категорий непокрытых лесом и нелесных земель – сельскохозяйственных угодий, гарей, вырубок, прогалин.

Совместное использование базы данных и векторного слоя ГИС позволило осуществить выбор эталонных участков для определения спектральных сигнатур различных классов лесных и нелесных объектов, однако, как показал анализ, использовать участки, отобранные по результатам запросов из базы данных как эталонные, нельзя. В результате ошибок совмещения векторного слоя карты и спутникового изображения, неоднородности атмосферных условий, образования смешанных пикселей на границах различных классов, интенсивной хозяйственной деятельности и естественного роста насаждений в один класс попадали объекты из других классов. Это потребовало проведения дальнейшего анализа с использованием статистических и графических методов.

Статистический метод заключался в вычислении статистических показателей для каждой выборки – минимального, максимального и среднего значения, стандартного отклонения и ковариационной матрицы.

Результаты статистической обработки сформированных выборок показали, что спектральная яркость по каналам ранжирует в широких

пределах, в некоторых случаях от 0 до 255, что указывало на неоднородность информационных классов. Средние значения по первым трем каналам (R, G, B) практически совпадали, заметные различия наблюдались лишь в четвертом (NIR) канале.

Графический метод состоял в построении для каждой выборки частотной гистограммы для всех спектральных диапазонов. При наличии у гистограммы нескольких пиков, необходимо идентифицировать эталонные области, которые вызывают этот эффект и исключить их из рассмотрения. Гистограммы распределения значений яркостей RGB и NIR каналов для информационных классов «болота» для отобранных участков на основе записей повыделочной базы данных представлены на рис. 1.

Графический анализ показал, что средние значения спектральных яркостей для сосновок и ельников отличаются в 4-м канале, спектральная яркость по каналам ранжирует в широких пределах, график спектральной яркости 4-го канала (NIR) для ели имеет несколько пиков. Все это указывает, что эталонные области не является репрезентативными для классов.

Среднее значение спектральной яркости NIR канала березняков отличается от чернольшаников и осинников, спектральная яркость по каналам ранжирует в широких пределах, что указывает на содержание в эталонных областях пикселей разных типов объектов.

Многомодальность, присущая на гистограммах спектральной яркости классов нелесных земель, указывает на неоднородность данных классов и необходимость более тщательного их подбора.

Для оценки разделимости информационных классов в спектральном пространстве были использованы коэффициенты трансформированной дивергенции (Transformed Divergence) и Джейфриса – Матусита (Jeffries – Matusita), которые рассчитывались с использованием имеющихся опций инструментария ENVI.

Результаты расчета коэффициентов показывают, что для некоторых категорий земель и насаждений значения коэффициентов Джейфриса – Матусита и дивергенции не превышали 0,2. Это потребовало более тщательного отбора используемых для классификации виделов, для чего необходимо было выполнить редактирование либо объединение схожих эталонных классов.

Для дальнейшего формирования и оценки качества обучающих выборок был использован следующий способ. По значениям яркости пикселей отдельных выборок рассчитывались параметры распределений: средние значения и стандартные отклонения. Эти параметры использовались для представления выборок в многомерном пространстве, размерность которого равнялась числу каналов. Если выборки имели значительные перекрытия, то пиксели, находящиеся в зоне перекрытия, не могли быть разделены. Чтобы добиться разделяемости, изменения границы (величину среднеквадратического отклонения) и положение (среднее значение) выборки, добивались ее минимального перекрытия с соседними.

Границные значения были подобраны таким образом, чтобы охватывать 95% пикселей. Те же действия были выполнены и для других эталонных областей.

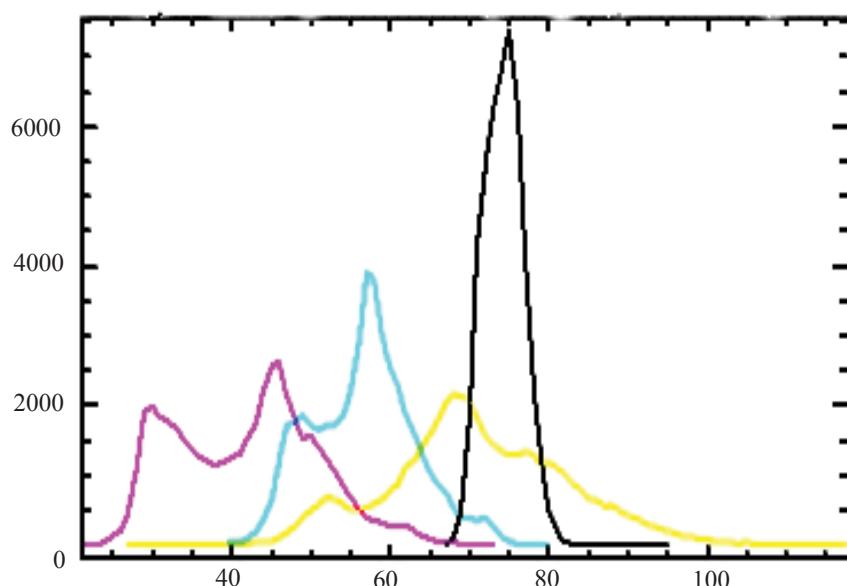


Рис. 1. Гистограммы спектральной яркости для категории нелесных земель болота после выборки из повыделочной базы данных

Результаты свидетельствуют о том, что ранжирование спектральной яркости по каналам уменьшилось. Однако средние значения по первым трем каналам у некоторых классов имели близкие значения, поэтому такие классы объединялись в один класс (например, сосновки и ельники в класс хвойные, черноольшанники и осинники).

Информационные классы, где наблюдались многовершинные гистограммы распределения яркостей пикселей, подвергались разделению на спектральные классы в пределах одного информационного. Так, например, было выделено шесть спектральных классов, определяющих информационный класс болота, два класса, определяющие класс воды.

В пределах хвойных, черноольхово-осиновых и березовых насаждений проверялась возможность выделения классов по составу, однако анализ полученных результатов показал, что информационные классы по составу отличаются весьма незначительно.

В дальнейшем была предпринята попытка разделения на классы по возрасту каждого информационного класса. Результаты говорят о том, что информационные классы по возрасту также практически не различимы. Коэффициенты разделимости данных информационных классов по возрасту оказались значительно меньше порогового значения 1,7.

Графический анализ распределения яркостей пикселей для каждого класса был выполн

ен с использованием  $N$ -мерного визуализатора ENVI. Результаты графического анализа показали, что классы достаточно хорошо разделимы между собой (рис. 2), что подтверждается достаточно высокими коэффициентами Джеффриса – Матусита (не ниже 1,5) и дивергенции (не ниже 1,8).

Исходя из полученных результатов анализа спектральной разделимости основных категорий лесных и нелесных земель, предлагается следующая структура классификации изображений:

1) земли, покрытые лесом:

хвойные:

сосна;

ель;

лиственные:

береза;

ольха черная, осина;

2) земли, покрытые травянистой растительностью:

поля с травой;

прогалины;

несомкнувшиеся лесные культуры;

3) земли без растительности (с поверхностной минерализацией):

свежие вырубки;

почвы;

почвы с увлажнением;

4) воды;

5) облака;

6) тени облаков;

7) болота и заболоченные участки.

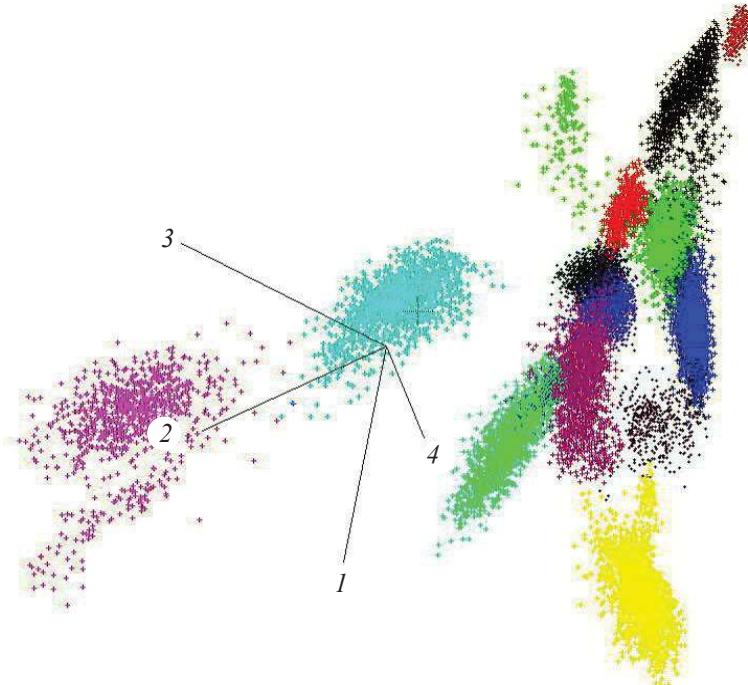


Рис. 2. Распределение яркостей пикселей отдельных классов в спектральном пространстве после второго этапа формирования:  
1–4 – спектральные каналы

На первом уровне детализации рассматриваются классы, которые можно выделить на основе визуального анализа прямых дешифровочных признаков: земли, покрытые лесом; земли, покрытые травянистой растительностью; земли без растительности; воды; облака; тени; болота и заболоченные участки.

Второй уровень детализации позволяет в пределах вышеназванных классов выделить статистически представленные более мелкие категории земель. Исходя из результатов спектрального анализа сигнатур древесных пород, преобладающих в лесах Березинского лесхоза, были выделены два основных типа лесной растительности: хвойные и лиственные леса, в пределах которых – классы сосняки, ельники, березняки и черноольхово-осиновые насаждения. Для проведения классификации было выделено 7 классов болот, 2 класса вод, 2 класса минерализованных почв. На последующих уровнях при использовании данной методики возможна и более детальная классификация.

**Заключение.** Анализ вычисленных коэффициентов, характеризующих разделяемость классов, показал, что полученные по вышеописанной методике эталонно-калибровочные участки могут использоваться для классификации изображений космических снимков и формирования спектральных библиотек.

Для формирования обучающих выборок целесообразно использовать повидельную базу данных лесохозяйственных учреждений и цифровые планово-картографические лесо-

строительные материалы, что позволит минимизировать работы по закладке эталонно-калибровочных участков и тем самым снизить затраты.

Формирование выборок необходимо выполнять в несколько этапов. На первом этапе формирования отбираются информационные классы: по категориям земель, породам, возрасту и т. д. Второй этап заключается в анализе статистических показателей распределений яркостей, выборе параметров для каждого класса и отборе участков с заданными параметрами. На заключительном этапе выполняется графический анализ и оценка разделимости классов.

### Литература

1. Чандря, А. М. Дистанционное зондирование и географические информационные системы / А. М. Чандря, С. К. Гош; пер. с англ. А. В. Кириюшина. – М.: Техносфера, 2008. – 307 с.
2. Лурье, И. К. Теория и практика цифровой обработки изображений / И. К. Лурье, А. Г. Косиков; под ред. А. М. Берлянта // Дистанционное зондирование и географические информационные системы. – М.: Научный мир, 2003. – 168 с.
3. Книжников, Ю. Ф. Аэрокосмические методы географических исследований: учебник для студ. высш. учеб. заведений / Ю. Ф. Книжников, В. И. Кравцова, О. В. Тутубалина. – М.: Академия, 2004. – 336 с.

*Поступила 14.04.2010*