

ЦИФРОВАЯ ОБРАБОТКА ИЗОБРАЖЕНИЙ ДЛЯ ОПРЕДЕЛЕНИЯ СТРУКТУРЫ ПОЧВ

Бутенко Анатолий Иванович,

профессор кафедры математики,

физики и информационных технологий

ФГБОУ ВО Мичуринский ГАУ, г. Мичуринск, РФ.

but_tolik@mail.ru,

Гордеев Александр Сергеевич,

профессор кафедры

агроинженерии и электроэнергетики АПК

ФГБОУ ВО Мичуринский ГАУ, г. Мичуринск, РФ.

gorde2020@gmail.com,

Придорогин Михаил Викторович,

старший научный сотрудник

ФГБОУ ВО Мичуринский ГАУ, г. Мичуринск, РФ.

rfsad2017@mail.ru,

Мишин Борис Сергеевич,

ассистент кафедры

агроинженерии и электроэнергетики АПК

ФГБОУ ВО Мичуринский ГАУ, г. Мичуринск, РФ.

boris.sergeewitch@yandex.ru,

Антонов Антон Владимирович,

магистрант кафедры

агроинженерии и электроэнергетики АПК

ФГБОУ ВО Мичуринский ГАУ, г. Мичуринск, РФ.

Егоров Андрей Михайлович,

аспирант

РГАУ-МСХА

Аннотация. Рассмотрены методические основы кластеризации почв по цвету. В работе предлагается кластеризация образцов почвы по гистограммам их цветных изображений.

Ключевые слова: почва, излучение, цвет, кластеризация, гистограмма, изображение.

Введение

Цвет – основной морфологический признак свойств почв. Это диагностический показатель, полученный путем измерения отраженного от поверхности почвы оптического излучения в красной, синей и зеленой областях спектра. Цвет характеризует ее различный минералогический и химический состав. Однако разными людьми цвет воспринимается неоднозначно. Объективные методы оценки образцов можно получить с помощью цифровой обработки изображений.

Для описания цвета в компьютерной графике существуют разные модели цветовых пространств. Наиболее простым является кодирование цвета каждого пикселя монитора тройкой чисел RGB. В этом случае R соответствует красному цвету (от английского red – красный), G отражает зеленый цвет (по-английски green – зеленый) и B соответствует синему цвету (по-английски blue – синий). Яркость каждого из цветов может меняться от 0 до 255. Если все три числа будут нулями, то пиксель будет черного цвета, а если все три будут равны 255 – белого цвета. Любому цвету пикселя однозначно соответствует определенная тройка значений RGB [4].

Как правильно сравнивать изображения образцов почвы по цвету? Ведь если перемешать частицы почвы в образце, то образец не изменится, а в изображениях первоначального и «перемешанного» образцов пиксели с одинаковыми яркостями цветов будут занимать разное положение. В этом случае сравнение изображений по каждому пикселю даст большие различия, что неприемлемо. Если сравнивать средние по каждому цвету значения яркости, то будут потеряны локальные цветовые различия образцов.

В цифровой обработке изображений глобальной информативной характеристикой изображения является гистограмма [4]. Для цветного изображения получают три гистограммы. Каждая гистограмма является графиком функции, где по оси ОХ откладываются значения яркости соответствующего цвета, а по оси ОУ – количества пикселей на изображении с заданной яркостью.

Методика исследования

В процессе исследования были использованы: результаты почвенных полевых обследований в разных регионах РФ, методы цифровой обработки изображений и кластерного анализа. Почвенные полевые исследования осуществлялись в соответствии с «Общесоюзной инструкцией по почвенному обследованию земель сельскохозяйственного назначения» [3].

Для нахождения гистограмм мы использовали библиотеку компьютерного зрения OpenCV[1]. Эта библиотека распространяется бесплатно и содержит большое количество функций и алгоритмов цифровой обработки фото- и видеоматериалов. Мы использовали версию библиотеки 2.4.10. Эта версия совместима с Visual Studio 2010, которую использовали для написания программы обработки на языке C++ функций OpenCV. Составленная нами программа выдает для каждого изображения три массива чисел, соответствующих трем гистограммам. Все изображения образцов имели одинаковый размер 287×287 пикселей, поэтому гистограммы были сопоставимы без дополнительного масштабирования.

Кластеризацию полученных данных проводили с помощью пакета STATISTICA 6 [2]. Вначале проводили кластерный анализ по каждому цвету отдельно. Объединение или метод древовидной кластеризации используется при формировании кластеров несходства или расстояния между объектами. В пакете предусмотрено несколько видов расстояния (несходства) между объектами. При иерархической кластеризации происходит последовательное объединение меньших кластеров в большие или разделение больших кластеров на меньшие.

Мы использовали иерархические агломеративные методы, которые характеризуется последовательным объединением исходных элементов и соответствующим уменьшением числа кластеров. В начале работы такого алгоритма, когда каждый объект представляет собой отдельный кластер, расстояния между объектами определяются выбранной мерой. Однако, когда связываются вместе несколько объектов, возникает вопрос, как следует определить расстояния между кластерами? Другими словами, необходимо правило объединения или связи для двух кластеров. В пакете есть несколько методов объединения кластеров. На наших данных мы сравнили все меры определения расстояния и все методы объединения кластеров в Statistica 6. Кластеры, полученные разными методами, не очень сильно различаются. Наиболее четкое разделение образцов получили при использовании евклидова расстояния с объединением кластеров по методу Варда.

Евклидово расстояние между гистограммами двух изображений находится по формуле

$$= \left(\sum_{i=0}^{255} (x_i - y_i)^2 \right)^{\frac{1}{2}},$$

где X_i, y_i – i -е значения гистограммы у двух сравниваемых объектов.

В методе Варда применяют внутригрупповую сумму квадратов отклонений, которая есть не что иное, как сумма квадратов расстояний между каждой точкой (объектом) и средней по кластеру, содержащему этот объект. На каждом шаге объединяются такие два кластера, которые приводят к минимальному увеличению этой целевой функции, т. е. внутригрупповой суммы квадратов. Этот метод направлен на объединение близко расположенных кластеров.

Для разных цветовых каналов получили разные кластеры, поэтому возникла проблема объединения данных. При проведении кластерного анализа попутно находится также матрица парных расстояний. Можно провести объединение трех полученных матриц. Обозначим через D_r, D_g, D_b матрицы парных расстояний по красному, зеленому и синему цветам

соответственно. Мы использовали следующие формулы объединения: среднее арифметическое $D = \frac{1}{3}(D_r + D_g + D_b)$, среднее геометрическое $D = \sqrt[3]{D_r \cdot D_g \cdot D_b}$, среднее векторное $D = \sqrt{D_r^2 + D_g^2 + D_b^2}$.

В этих формулах все операции над матрицами осуществляли поэлементно.

Проводили объединение также по гистограммам. Обозначим через r , g , b соответствующие значения гистограмм красного, зеленого и синего цветов. Мы использовали следующие формулы объединения:

$h = 0,299r + 0,587g + 0,114b$ (такая комбинация используется при конвертировании цветного изображения в изображение в оттенках серого),

$$h = \max(r, g, b).$$

Результаты исследования

По каждой из рассмотренных формул кластерный анализ объединил 21 образец в три кластера. Состав образцов в кластерах, полученных при разных объединениях, немного различался. Чтобы выбрать лучший метод мы брали среднее значение гистограмм одного цвета, для образцов одного кластера и смотрели, как визуально отличаются средние гистограммы трех кластеров. Наиболее различающиеся кривые получены при объединении по последней формуле. Эти кривые приведены на рис. 2, а дендрограмма кластеров приведена на рис. 1. Как видно из этой дендрограммы, в первый кластер входят образцы p_{11} , p_{41} , p_{12} , p_{61} , p_{31} , p_{42} , p_{72} . Вторым кластером составляют образцы p_{13} , p_{21} , p_{33} , p_{51} , p_{62} . И в третий кластер входят образцы p_{22} , p_{52} , p_{43} , p_{53} , p_{23} , p_{63} , p_{32} , p_{71} , p_{73} .

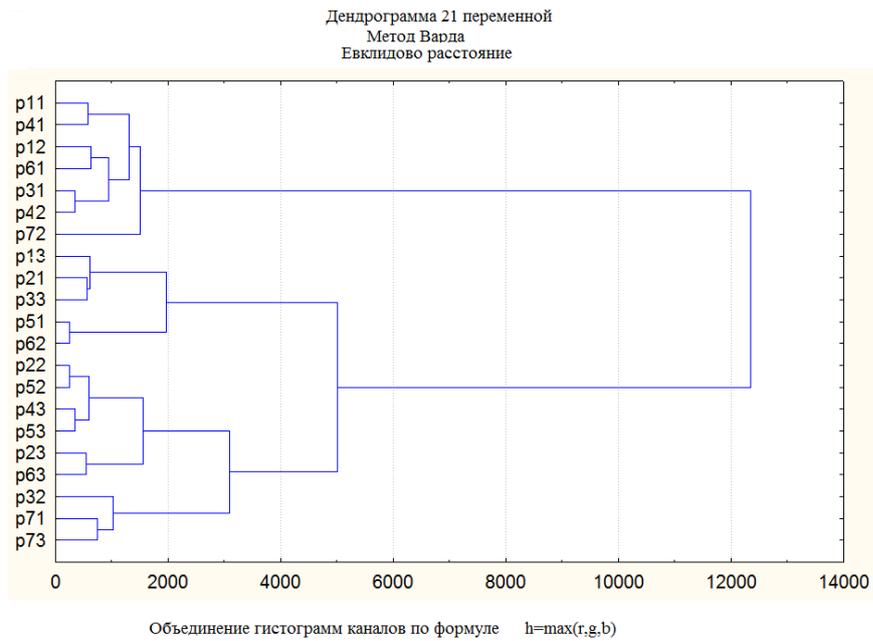


Рис. 1.

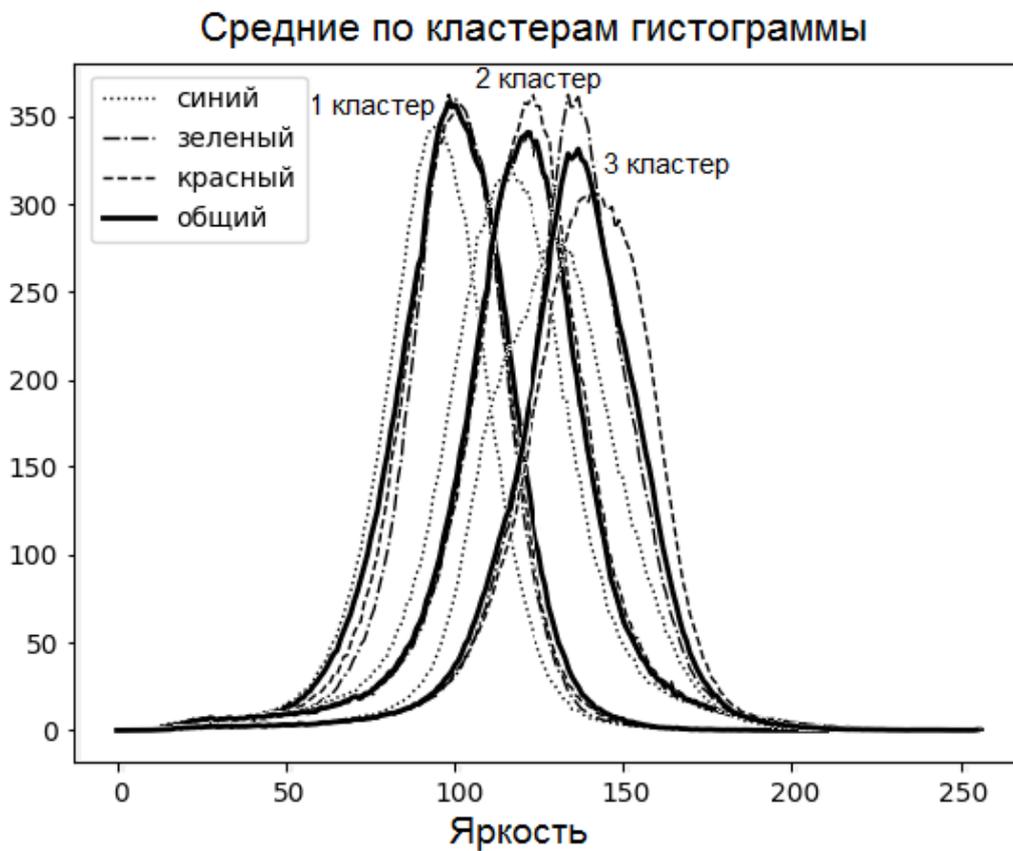


Рис. 2.

Выводы

Эту статью мы рассматриваем как предварительную на пути получения объективных экспресс-оценок качеств почв и их классификации. На этом пути необходимо выявить наиболее информативные признаки изображений и

связать их с известными показателями, получаемыми в лабораторных условиях.

Список использованной литературы

1. Бовырин А.В., Разработка мультимедийных приложений с использованием библиотек OPENCVи IPP / А.В. Бовырин, П.Н. Дружков, В.Л. Ерухимов, В.Д. Кустикова, И.Д. Лысенков, И.Б. Мееров, В.Н. Писаревский, А.Н. Половинкин, А.В. Сысоев. – М.: ИНТУИТ, 2016. – 515 с.

2. Боровиков В.П., Популярное введение в современный анализ данных в системе STATISTICA. Учебное пособие для вузов/ В.П. Боровиков. – М.: Горячая линия – Телеком, 2013. – 288 с.

3. Общесоюзная инструкция по почвенным обследованиям и составлению крупномасштабных почвенных карт землепользования. – М.: Колос, 1973. – 80 с.

4. Яне Б. Цифровая обработка изображений. – М.: Техносфера, 2007. – 584 с.

DIGITAL IMAGE PROCESSING FOR DETERMINING THE SOIL STRUCTURE

Anatoly Butenko,

Professor

Department of Mathematics, Physics and Information Technology

but_tolik@mail.ru,

Alexander Gordeev,

Professor

Department of Agroengineering and Electric Power Industry of the AIC

gorde2020@gmail.com,

Mikhail Pridorogin,

Senior Researcher,

rfsad2017@mail.ru,

Boris Mishin,

Assistant

Department of Agroengineering and Electric Power Industry of the AIC

boris.sergeewitch@yandex.ru,

Antonov Anton Vladimirovich,

master student

Department of Agroengineering and Electric Power Industry of the AIC

Michurinsk State Agrarian University

Michurinsk, Russia

Egorov Andrei Mikhailovich,

graduate student

RGAU-ICCA

Annotation. The methodological foundations of soil clustering by color are considered. The paper proposes clustering of soil samples according to histograms of their color images.

Key words: soil, radiation, color, clustering, histogram, image.