

УДК 004.932

ИССЛЕДОВАНИЕ АЛГОРИТМОВ КЛАССИФИКАЦИИ С ПРИМЕНЕНИЕМ МЕТОДОВ ОБРАБОТКИ BIG DATA

Ситникова Н. В., Парингер Р. А., Куприянов А. В.

Самарский государственный аэрокосмический университет имени академика
С. П. Королёва (национальный исследовательский университет), г. Самара

Существует класс изображений кристаллических микро- и наноструктур, получаемых с помощью оптической и электронной микроскопии, для которых модели, методы и соответствующие алгоритмы текстурного анализа недостаточно проработаны и исследованы [1]. Человек достаточно легко справляется с выделением текстурной информации и разделением текстурных изображений, в то время как автоматизированный текстурный анализ, то есть анализ текстур на изображении, является сложной задачей. В связи с многообразием разновидностей текстур [2, 3] для расчёта признаков по изображениям предлагается использовать технологию Big Data.

Когда говорят о термине Big Data, то используют определение трёх «V», что означает Volume – объём данных, Velocity – необходимость обрабатывать информацию с большой скоростью и Variety – многообразие и часто недостаточная структурированность данных.

В настоящее время на рынке информационных систем и программного обеспечения синонимом Big Data является технология Hadoop, которая представляет собой программный фреймворк, позволяющий хранить и обрабатывать данные с помощью компьютерных кластеров, используя парадигму MapReduce. В соответствии с подходом MapReduce обработка данных состоит из двух шагов: Map и Reduce. На шаге Map выполняется предварительная обработка данных, которая осуществляется параллельно на различных узлах кластера. На шаге Reduce происходит сведение предварительно обработанных данных в единый результат.

Для проведения вычислительных экспериментов используются текстурные изображения из базы MeasTex [4], которые разделены на 12 классов похожих текстур. Было исследовано 84 больших образца размерами 512×512 по семь из каждого класса, которые затем были разбиты на более мелкие информативные изображения размером 256×256 точек. Полученные текстурные изображения были достаточно велики, чтобы быть информативными представителями различных типов изображений. Таким образом, было доступно 336 образца для представления набора из 12 классов, число представителей каждого класса равнялось 28.

При исследовании качества классификации было сформировано две выборки: обучающая и тестовая. На основе обучающейся выборки был построен классификатор по k ближайшим соседям, с помощью которого классифицировалась тестовая выборка.

В результате работы формировалась матрица ошибок классификации, значениями которой являлись оценки вероятностей попадания объекта из класса по строке в класс по столбцу (таблица 1). На диагонали этой матрицы располагаются вероятности правильной классификации объектов.

Таблица 1. Матрица ошибок, вычисленная с использованием двумерных признаков

№ класса	1	2	3	4	5	6	7	8	9	10	11	12
1	1	0	0	0	0	0	0	0	0	0	0	0
2	0	0,61	0	0	0	0	0	0,38	0	0	0	0
3	0	0	1	0	0	0	0	0	0	0	0	0
4	0	0	0	1	0	0	0	0	0	0	0	0
5	0	0	0	0,17	0,82	0	0	0	0	0	0	0
6	0	0,26	0	0	0	0,54	0	0	0,19	0	0	0
7	0	0	0	0	0	0	1	0	0	0	0	0
8	0	0,17	0	0	0	0	0	0,82	0	0	0	0
9	0	0	0	0	0	0	0,24	0	0,75	0	0	0
10	0	0,40	0	0	0	0	0	0	0	0,59	0	0
11	0	0,17	0	0	0	0	0,38	0	0	0	0,54	0
12	0	0	0	0	0	0	0	0	0	0	0	1

Процент верной классификации равен 80,9 %.

Библиографический список

1. Куприянов, А. В. Анализ текстур и определение типа кристаллической решётки на наномасштабных изображениях [Текст]/ А. В. Куприянов // Компьютерная оптика – 2011. – Т. 35.-2. – С. 145-152.
2. Kupriyanov, A. V. Statistical Features of Image Texture for Crystallogram Classification / A. V. Kupriyanov, A. G. Khramov, N. Yu. Ilyasova // Pattern Recognition and Image Analysis. – 2001. – Vol. 11, N 1. – P. 180-183.
3. Ильясова, Н. Ю. Классификация кристаллограмм с использованием методов статистического анализа текстурных изображений [Текст]/ Н. Ю. Ильясова, А. В. Куприянов, А. Г. Храмов // Компьютерная оптика. – 2000. – № 20. – С. 122-127.
4. База данных изображений MeasTex Image Texture Database [Электронный ресурс]. – Режим доступа: <http://www.texturesynthesis.com/meastex/meastex.html>.