

## ОБРАБОТКА И КЛАССИФИКАЦИЯ ГРАФИЧЕСКОГО СЛОВАРЯ СИМВОЛЬНЫХ ДАННЫХ ПРИ СЖАТИИ ИЗОБРАЖЕНИЯ ТЕКСТА

Иванов В.Г., Ломоносов Ю.В., Любарский М.Г.

*Национальный университет "Юридическая академия Украины имени Ярослава Мудрого", г. Харьков*

Основной задачей классификации при сжатии изображения текста является такое разбиение изображений символов на классы, чтобы различные изображения одного и того же символа попадали в один и тот же класс. Изображение текста, после классификации символов, можно представить в виде «графического словаря» – набора изображений каждого символа, и «карты регионов» – описанием положения каждого символа в тексте.

Целью настоящей работы является усовершенствование алгоритма сжатия текстовых изображений, основанного на дополнительном сжатии «графического словаря». Предлагаемый метод базируется на представлении всех изображений символов, входящих в состав «графического словаря», последовательностью вертикальных элементов строки и их автоматической классификации с последующим построением карты их размещения. Общий алгоритм сжатия текстовых изображений разбивается на два этапа (рис.).

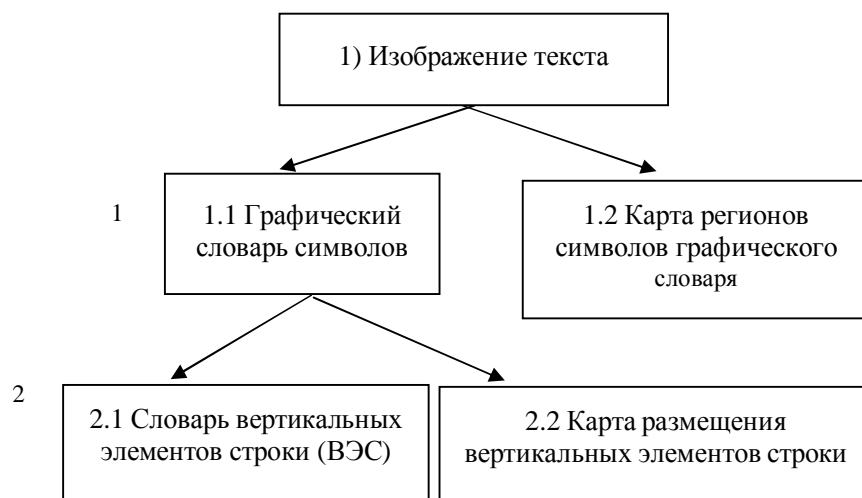


Рис. Схема двухэтапной обработки изображения текста

Практическая ценность полученных результатов заключается в том, что в сравнении с лучшим в настоящее время специальным алгоритмом сжатия изображений текста – JB2 (формат DjVu), предлагаемый двухэтапный алгоритм имеет преимущество в степени сжатия данных в среднем на 25% для часто используемых разрешений изображения текста (200-600 dpi).