

ПОСТРОЕНИЕ СЛОВАРЯ ПРИ КЛАССИФИКАЦИИ СИМВОЛЬНЫХ ДАННЫХ

Методы классификации являются достаточно перспективными и активно используются в теории и практике сжатия изображений [1 – 5]. Наибольший интерес и значение эти методы приобретают при сжатии изображений текста (символьных изображений), которые используются при переводе печатной продукции в электронную форму.

Приводится новый подход к созданию общего словаря символов путем классификации символов изображения короткими словарями, которые последовательно формируются на участках изображения текста. Сам метод сжатия изображения текста на основе выделения символов и их классификации подробно изложен в работах авторов [6,7,8]. Эффективность первичного словаря (K) оценивалась как отношение количества центров (классов) вошедших в словарь (N dic) к количеству символов на котором формировался данный первичный словарь (N symbols), выражение (1)

$$K = \frac{N_dic}{N_symbols}. \quad (1)$$

Максимум отношения определяет участок изображения текста, где сформированный первичный словарь будет наиболее эффективным. Найденные центры используются для классификации на всем множестве символов. Количество итераций обработки изображения текста определяется величиной – среднее количество классифицированных символов центром первичного словаря (выражение 2).

$$K1 = \frac{N_{classific_symbols}}{N_{classes}}. \quad (2)$$

На Рис.1 представлено среднее количество символов в классе на множестве необработанных символов – сплошная линия, а приращение среднего количества символов в классе после классификации символами центрами первичного словаря – пунктирная кривая.

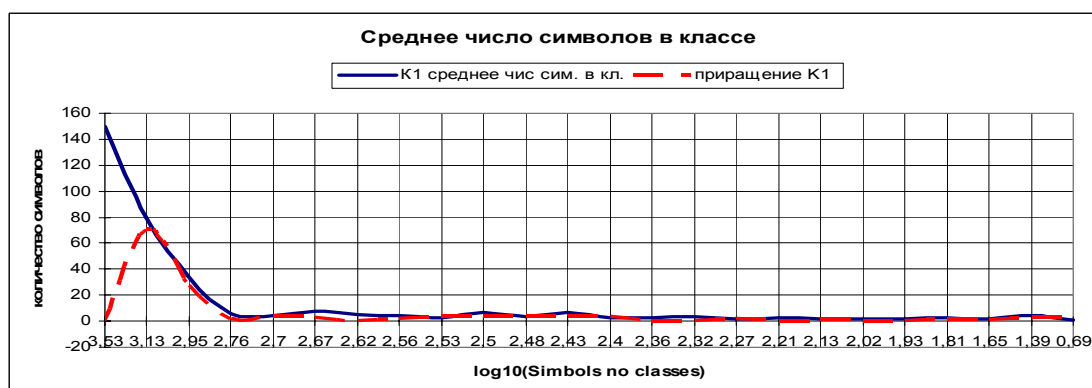


Рис.1 Среднее число символов в классе и его приращение.

Максимум приращения среднего числа символов в классе на множестве необработанных символов определяет число итераций.

Таким образом, классификация символов центрами первичных словарей наиболее эффективна при нескольких (двух) итерациях, а оставшееся множество символов можно классифицировать методом прямого перебора. Использование описаного способа формирования общего словаря символов позволяет снизить на 20-25% время обработки всего изображения по сравнению с методом прямого перебора символов.

Список литературы:

[1] Земсков В.Н. Сжатие изображений на основе автоматической классификации [Текст] / В.Н. Земсков, И.С. Ким // Известия вузов. Электроника. – 2003. – № 2. – С. 50-56.

[2] Gupta Maya R., Stroilov A. Segmenting for wavelet compression [Электронный ресурс]: Data Compression Conference, 2005. Proceedings. DCC 2005, 29-31 March 2005, USA, Utah, Snowbird. – 462 p. - Режим доступа: <http://www.computer.org/portal/web/csdl/proceedings/> - 10.04.2010 г.

[3] Иванов В.Г. Сокращение содержательной избыточности изображений на основе классификации объектов и фона [Текст] / В.Г. Иванов, М.Г. Любарский, Ю.В. Ломоносов // Проблемы управления и информатики. – 2007. – № 3. – С. 93-102.

[4] Иванов В.Г. Сжатие изображений на основе автоматической и нечеткой классификации фрагментов [Текст] / В.Г. Иванов, Ю.В. Ломоносов, М.Г. Любарский // Проблемы управления и информатики. – 2009. – №1 – С. 52-63.

[5] Прикладная статистика: Классификация и снижение размерности [Текст]: справочник / С.А. Айвазян, В.М. Бухштабер, И.С. Енюков и др.; под общ. ред. С.А. Айвазяна.– М.: Финансы и статистика, 1989. – 607 с.

[6] Иванов В.Г. Сжатие изображения текста на основе выделения символов и их классификации [Текст] / В.Г. Иванов, М.Г. Любарский, Ю.В. Ломоносов // Проблемы управления и информатики. – 2010. – № 6. – С. 74-84.

[7] Иванов В.Г. Сжатие символьных изображений на основе новой классифицирующей метрики. [Текст] / В.Г. Иванов, М.Г. Любарский, Ю.В. Ломоносов, С.В. Деркач // 17 міжнародна конференція з автоматичного управління “Автоматика -2010”. Тези доповідей. Том 2.- Харків: ХНУРЕ, 2010.- с.162-164. 306 с.

[8] Иванов В.Г. Компресія зображень тексту на основі класифікуючої метрики з подавленням шумів друку та сканування. [Текст] / В.Г. Иванов, М.Г. Любарський, Ю.В. Ломоносов, С.В. Котляр // Праці 10-ї всеукраїнської міжнародної конференції “Оброблення сигналів і зображень та розпізнавання образів” (УкрОБРАЗ’2010) – Київ, 2010. – с.161-165.