

потока поступающих – планка минимального проходного бала при поступлении как на бюджетные, так и на платные места, будет подниматься.

Обобщая, можно сделать следующие выводы. Только рассчитывая на свои силы и возможности при поступлении на ту или иную специальность, учитывая современные проблемы на рынке труда и нынешнюю переменчивую ситуацию в сфере высшего профессионального образования, можно сделать правильный выбор учебного заведения. Для привлечения школьников на технические специальности и направления необходимо повышать престиж профессии в целом. Кроме того, в современных условиях необходима перестройка системы подготовки специалистов высшей технической квалификации, что отмечается большинством специалистов, и необходима грамотная стратегия продвижения и рекламы данных специальностей среди молодежи.

### Литература

1. Логунова О.С. Потенциальная емкость рынка услуг высшего образования в условиях экономического кризиса// Социология образования. №6. С. 42-50
2. Лоранж П. Новый взгляд на управление образованием: задачи руководства. М., 2004.
3. Увайсов С.У., Бушмелева К.И., Кривицкая М. Выбор критериев оптимальности при разработке рабочего учебного плана// Качество. Инновации. Образование. 2013. №1. С. 68-71
4. Авдеюк О.А. Социальные основания популярности инженерного образования / Соловьев А.А., Авдеюк О.А., Шведов Е.Г., Авдеюк Д.Н. // Молодой учёный. - 2013. - № 5 (ч. 4). - С. 766-768.
5. Авдеюк О.А. К проблеме адаптации в вузе студентов заочной формы обучения / Авдеюк О.А., Асеева Е.Н., Крохалев А.В., Приходьков К.В., Савкин А.Н. // Социосфера. - 2011. - № 2. - С. 65.
6. Увайсов С. У., Аютова И. Информационные технологии защиты персональных данных в ВУЗе // В кн.: Информационные технологии в гуманитарном образовании. Материалы V Международной научно-практической конференции, 20-24 июня 2012 года / Отв. ред.: Г. Воробьев. Пятигорск: Пятигорский государственный лингвистический университет, 2013. С. 271-277.
7. Увайсов С. У., Шайторова И. А., Бушмелева К. И. Методика адаптации молодых преподавателей в вузе // В кн.: Информационные технологии в гуманитарном образовании. Материалы V Международной научно-практической конференции, 20-24 июня 2012 года / Отв. ред.: Г. Воробьев. Пятигорск: Пятигорский государственный лингвистический университет, 2013. С. 262-265.
8. Увайсов С.У., Увайсова С.С., Увайсова А.С. Электронный ассистент UVAYSOV докладчика. Инновации на основе информационных и коммуникационных технологий. 2011. № 1. С. 61-62
9. Увайсов С. У., Аютова И. Модифицированная графическая модель классификации информационных систем обработки персональных данных // В кн.: Инновационные информационные технологии. Материалы международной научно-практической конференции / Отв. ред.: И. Иванов, Л. Агеева, Д. Дубоделова, В. Еремина; под общ. ред.: С. У. Увайсов. М.: МИЭМ, 2012. С. 370-373.
10. Увайсов С. У., Шайторова И. А., Бушмелева К. И. Автоматизированная информационная система «Индивидуальная траектория движения ППС в ВУЗе»// В кн.: Информационные технологии в гуманитарном образовании. Материалы V Международной научно-практической конференции, 20-24 июня 2012 года / Отв. ред.: Г. Воробьев. Пятигорск: Пятигорский государственный лингвистический университет, 2013. С. 269-271

### ПОСТРОЕНИЕ ЭФФЕКТИВНОГО СЛОВАРЯ ПРИ КЛАССИФИКАЦИИ СИМВОЛЬНЫХ ДАННЫХ

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

*Национальный университет «Юридическая академия Украины имени Ярослава Мудрого»*

Показано, что использование коротких первичных словарей в двухэтапном алгоритме сжатия символьных данных дает возможность уменьшить время кодирования на 20-25%. Представлены способы и критерии формирования первичных словарей символов, а так же показатель их итерационного использования.

**Construction of effective dictionary during classification of character dates. Ivanov V., Lomonosov U., Lyubarsky M.**

It is shown, that at usage in double-step algorithm of compression of the character given short primary dictionaries the general processing time can be reduced by 20-25 %. Modes and measure of creation of primary dictionaries of classification of characters, and as an index of their iterative usage are presented.

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.

Сам метод сжатия изображения текста на основе выделения символов и их классификации подробно изложен в работах авторов [6,7,8]. Установлено, что степень сжатия изображений текста является очень высокой при качестве восстановленного текста существенно лучшем (благодаря операциям усреднения), чем у исходного текста. Однако минимизация вычислительных затрат предлагаемых алгоритмов в этих работах не рассматривалась.

Основным недостатком двухэтапной классификации [6-8] является то, что на первом этапе классификации участвуют все символы, в том числе и те, которые образуют классы состоящие из одного представителя и являются уникальными. Это приводит к неоправданным временным затратам, когда подобный символ изображения текста сравнивается с остальными и в результате не находится ни одного подобного символа, образуя класс, состоящий из одного представителя. На рис 1. приведены примеры символов, которые являются одинаковыми, но не попали в один класс. Это целое семейство символов “точка” и символа “г”. В первом случае все символы при практически равных геометрических размерах (высота, ширина) значительно разнятся по периметру (отклонение, которого допускается не более 10%, что соответствует несовпадению всего двух точек в изображении данного символа). Во втором случае представленные символы не были классифицированы в один класс в ходе плоскопараллельного переноса и вычисления симметрической разности с совмещенными центрами тяжести при процедуре “просеивания”.

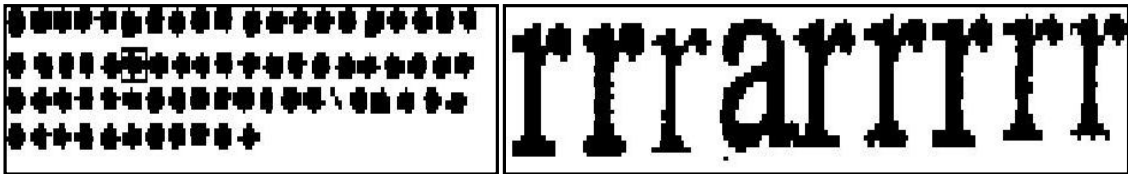


Рис. 1. Примеры классов изображений символов с одним представителем.

В данной работе предлагается - на первом этапе классификации собрать в графический словарь сначала все символы, которые формируют классы с большим числом представителей, исключив их таким образом из дальнейшей классификации при формировании следующих классов. Когда придет очередь до классификации уникальных символов, то число сравниваемых с ними символов будет гораздо меньше, что позволит сократить общее время обработки всего символьного изображения.

Необходимо напомнить, что классификация символов на первом этапе проводится методом «просеивания» [6,7], который состоит в следующем. Выбирается произвольный элемент из классифицируемого множества и в один класс с ним помещаются все элементы близкие к нему. Далее рассматриваются только элементы, не вошедшие в первый класс. Из их числа произвольно выбирается какой-либо элемент и аналогичным образом строится второй класс. Этот процесс повторяется до тех пор, пока не будут исчерпаны все элементы исходного множества.

Второй этап классификации реализует алгоритм «наращивания областей», который заключается в том, что на первом шаге, начиная с произвольно выбранного элемента классифицируемого множества, к его классу присоединяются все достаточно близкие элементы. На втором шаге к вновь присоединенным элементам добавляются все элементы, близкие к ним. Процесс «наращивания» повторяется до тех пор, пока на каком-то шаге не окажется новых элементов, которые можно было бы присоединить. Далее все элементы «выращенного» класса исключаются из классифицируемого множества и «выращивается» следующий класс. Алгоритм заканчивает работу, когда в классифицируемом множестве не остается ни одного элемента.

В представленной работе приводится иной подход к созданию общего словаря символов путем классификации символов изображения короткими словарями, которые последовательно формируются на участках изображения текста. Составление первичных словарей осуществляется на основе оценки их эффективности. Количество первичных словарей определяется такой условной характеристикой, как среднее число классифицированных символов первичного словаря.

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 средним значением классифицированных символов центром первичного словаря (выражение 2).

$$K1 = \frac{N_{\text{classific\_symbols}}}{N_{\text{classes}}} \quad (2)$$

На рис. 2 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. Максимум приращения среднего числа символов в классе определяет число итераций. Таким образом, на данном изображении классификация символов центрами первичных словарей наиболее эффективна при двух итерациях. Оставшееся множество символов можно классифицировать методом “просеивания” и далее на втором этапе методом «наращивания областей».



Рис. 2 Среднее число символов в классе и его приращение.

**Выводы.** Использование первичных словарей на первом этапе классификации методом “просеивания” (прямым перебором) позволило исключить из классифицируемого множества те символы, которые формируют классы с большим количеством представителей. Это позволило снизить общее время классификации на 20-25% по сравнению с последовательным применением метода “просеивания” и метода “наращивания областей” ко всему множеству изображений символов.

### Литература

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](http://www.computer.org/portal/web/csdl/proceedings/), 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.