

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ КЛАССИФИКАЦИИ СИМВОЛОВ ПРИ ОБРАБОТКЕ ИЗОБРАЖЕНИЯ ТЕКСТА

Современные методы сжатия, основанные на различных ортогональных преобразованиях, дают хороший результат при сжатии размытых изображений, но не эффективны для битональных изображений, тем более изображений текста, избыливающего множеством мелкими деталями – буквами, цифрами, знаками препинания. В настоящее время лучшие алгоритмы для сжатия битональных изображений текста основаны на выделении изображений символов и их классификации. Это – алгоритмы JB2 и JBIG2, используемые соответственно в широко распространённых форматах DjVu и PDF. Степень сжатия информации с помощью методов классификации тем выше, чем меньше классов образуется при классификации и чем больше элементов в каждом классе. В идеале при сжатии изображения страницы текста, изображения каждого символа должны находиться в одном и только одном классе. Однако ни один из известных алгоритмов этому условию не удовлетворяет. Дело в шумах (случайных искажениях), возникающих при печати страницы и ее последующем сканировании.

Новый подход к сжатию графических текстовых данных заключается в следующем. Если представить себе прямоугольник, охватывающий какую-либо строку, то вертикальным элементом этой строки будем называть пересечение прямоугольника с любой вертикальной линией шириной в один пиксель.

Шумы печати и сканирования случайным образом искажают вертикальные элементы. Так что среди них могут быть искаженные и неискаженные элементы. Однако нет необходимости разбивать совокупность вертикальных элементов, составляющих изображение страницы, на классы тождественных или почти тождественных элементов, поскольку многие из них могут быть искажениями сразу нескольких неискаженных элементов.

Используя отдельный этап представления связанных символов изображения текста, в виде вертикальных элементов строки и применив их нечеткую классификацию, получена минимальная наиболее правдоподобная совокупность неискаженных элементов строки. Приняв во внимание ограничения вероятностной модели, для каждого вертикального элемента строки получена вероятность того, что он является искажением найденного неискаженного элемента строки.

Исследовав возможности компрессии данных представленных в виде словаря неискаженных вертикальных элементов строки и их карты размещения, была проведена классификация связанных символов на основе нечеткой классификации вертикальных элементов. Формирование словаря связанных символов основывалось не на сравнительном анализе геометрических форм сравниваемых символов [1], а на вероятностной оценке соответствующих вертикальных элементов строки, которые представляют состав классифицируемых связанных символов [2, 3].

Предложенный алгоритм представления и обработки изображения текста позволил получить достаточно высокую степень сжатия при хорошем качестве восстановленного изображения. Сравнение с лучшим в настоящее время специальным алгоритмом сжатия изображений текста – JB2, входящим в формат DjVu, показало, что предлагаемый алгоритм сжатия изображения текста имеет преимущество в степени компрессии данных.

Для наиболее часто используемого разрешения изображения текста 300 dpi были получены следующие сравнительные показатели сжатия:

- в работе [1] преимущество над JB2 – 8 %;
- в работе [2] преимущество над JB2 – 25 %;
- в работе [3] преимущество над JB2 – 37 %.

Это открывает новые возможности повышения информативности электронного представления графических текстовых данных в инженерных реализациях.

Список литературы

1. 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. 111–122.
2. Иванов В. Г. Сжатие изображения текста на основе формирования и классификации вертикальных элементов строки в графическом словаре символьных данных [Текст] / В. Г. Иванов, М. Г. Любарский, Ю. В. Ломоносов // Проблемы управления и информатики. – 2011. – № 5. – с. 98–109.
3. Иванов В. Г. Сжатие изображения текста на основе статистического анализа и классификации вертикальных элементов строки [Текст] / В. Г. Иванов, Ю. В. Ломоносов, М. Г. Любарский // Восточно-Европейский журнал передовых технологий. – 2014. – № 4/2 (70). – с. 4–15.

ЗМІСТ

МихайловаЄ.О. МОЖЛИВОСТІ ІМІТАЦІЙНОГО МОДЕЛЮВАННЯ В ЕКОЛОГІЇ.....	80
Москаленко В.Б., Фонта Н.Г. СТРУКТУРА ІНФОРМАЦІОННОЇ СИСТЕМИ УПРАВЛІННЯ ЕФЕКТИВНОСТЮ ПРІДПРИЯТТЯ С УЧЕТОМ КАСКАДИРОВАНИА КРІ.....	81
Степанов В.П. ЕФЕКТИВНОСТЬ ІСПОЛЬЗОВАНИА ПРОГРАММНИХ СРЕДСТВ В EXCEL ДЛІ ОБРАБОТКИ ЕКОНОМІЧЕСКОЇ ІНФОРМАЦІ.....	82
Ткаченко В.В. ПЕРЕВАГИ ПРОЦЕСНОЇ МОДЕЛІ ПРЕДСТАВЛЕННЯ НЕЯВНИХ ЗНАНЬ.....	83

СЕКЦІЯ 4. КОМП'ЮТЕРИЗОВАНИ ТЕХНОЛОГІЇ ТА СИСТЕМИ ВИДАВНИЧО-ПОЛІГРАФІЧНИХ ВИРОБНИЦТВ ТА ЕЛЕКТРОННИХ МУЛЬТИМЕДІЙНИХ ВИДАНЬ

Kazimierz Waćkowski, Tomasz Kowalczyk EFFECTIVE IT PROJECT MANAGEMENT IN ORGANIZATIONS.....	84
Бережна О. Б. РЕАЛІЗАЦІЯ ФУНКЦІЙ ПЕДАГОГІЧНОГО ДИЗАЙНУ НА ПРИКЛАДІ ДИСЦИПЛІНИ «ІНФОРМАТИКА І КОМП'ЮТЕРНА ТЕХНІКА».....	85
Бондар І.О. ЗАВДАННЯ АРХІТЕКТУРНОЇ ПОБУДОВИ ЕЛЕКТРОННИХ ЖУРНАЛІВ ІГРОВОГО СПРІАМУВАННЯ.....	86
Браткевич В.В. МОДЕЛЬ ВИБОРА ІНСТРУМЕНТАЛЬНИХ СРЕДСТВ ПРОЕКТИРОВАНИА ОБУЧАЮЩИХ СИСТЕМ.....	87
Грабовський Є.М. ЗАГАЛЬНІ ОСОБЛИВОСТІ УПРАВЛІННЯ КОНТРОЛЕМ ЯКОСТІ ПОЛІГРАФІЧНОЇ ПРОДУКЦІЇ.....	88
Завгородня О. С. РЕКОМЕНДАЦІЇ З ОПТИМІЗАЦІЇ ІГРОВИХ ЛОКАЦІЙ.....	89
Ізонін І. В., ЛотошинськаН.Д. ДОСЛІДЖЕННЯ ТЕХНОЛОГІЇ ПОБУДОВИ ТРИВИМІРНИХ МОДЕЛЕЙ.....	90
Ломоносов Ю.В. СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ КЛАССИФИКАЦИИ СИМВОЛОВ ПРИ ОБРАБОТКЕ ИЗОБРАЖЕНИЯ ТЕКСТА.....	91
Молчанов В. П. ВЗАЄМОДІЯ ПРОЦЕСІВ В WEB-ДОДАТКУ НАВЧАЛЬНОГО СЕРЕДОВИЩА.....	92
Пандорін О.К. ПЕРСПЕКТИВИ РОЗВИТКУ МОВНИХ ЗАСОБІВ У WWW СТАНДАРТАХ.....	93
Потрашкова Л. В. ІНФОРМАЦІЙНА ПІДТРИМКАРЕАЛІЗАЦІЇ СОЦІАЛЬНОЇ ВІДПОВІДАЛЬНОСТІ СПОЖИВАЧІВ ТА ВИРОБНИКІВ.....	94



Тези доповідей

Міжнародної науково-практичної
конференції

“Проблеми і перспективи розвитку ІТ-індустрії”

20–21 квітня 2017 р.

Матеріали Міжнародної науково-практичної конференції“ Проблеми і перспективи розвитку ІТ-індустрії ”: тези доповідей, 20–21 квітня 2017 р. – Х. : ХНЕУ імені Семена Кузнеця, 2017. – 105 с.

Наведені тези пленарних та секційних доповідей за теоретичними та практичними результатами наукових досліджень і розробок. Представлені результати теоретичних досліджень в галузях проектування інформаційних систем, технологій захисту інформації, використання сучасних інформаційних технологій в управлінні системами, моделювання бізнес-процесів, застосування геоінформаційних технологій, дистанційній освіті, інформаційних технологій в видавничо-поліграфічній галузі.

Матеріали публікуються в авторській редакції.

За достовірність викладених фактів, цитат та інших відомостей відповідальність несе автор.

**ТЕЗИ ДОПОВІДЕЙ
міжнародної науково-практичної конференції
“Проблеми та перспективи розвитку ІТ-індустрії”
20-21 квітня 2017 р.**

Відповідальний за випуск: *С.П. Євсєєв*

Комп'ютерна верстка: *Є.М. Грабовський*

Підписано до друку 29.03.2017. Формат 60×84/8. Папір офсетний.
Гарнітура «TimesNewRoman». Друк ризографічний. Ум.-друк. арк. – 12,2. Ціна договірна.
Наклад 250 прим.Зам. № 0329/3-17.

Видавництво «Цифрова друкарня №1»
Свідоцтво суб'єкта видавничої справи: серія ДК № 4354 від 06.07.2012 р.
61001, м. Харків, пл. Повстання, 7/8
e-mail: zebra-zakaz@mail.ru

Віддруковано з готових оригінал-макетів у друкарні ФОП Петров В.В.
Єдиний державний реєстр юридичних осіб та фізичних осіб-підприємців.
Запис № 2480000000106167 від 08.01.2009.
61144, м. Харків, вул. Гв. Широнінців, 79в, к. 137,
тел. (057) 78-17-137, e-mail: bookfabrik@mail.ua