

УДК 004.627

СПОСОБ ОБРАТИМОГО СЖАТИЯ КАДРОВ ИЗМЕРИТЕЛЬНЫХ ДАННЫХ НА ОСНОВЕ «ПАРКЕТНОГО» РАЗБИЕНИЯ

И. В. Богачев, А. В. Левенец, Е. У. Чье

*Тихоокеанский государственный университет,
680035, г. Хабаровск, ул. Тихоокеанская, 136
E-mail: ilya.bogachev.pnu@mail.ru*

Рассмотрено представление измерительных данных как битовое пространство. Предложены способ и базирующийся на нём алгоритм обратимого геометрического сжатия кадров измерительных данных, в основе которого лежит их преобразование к битовому виду с последующим отображением на плоскость и разбиением на строго однородные области. Приведены экспериментальные результаты, показывающие, что предложенный алгоритм обеспечивает достаточно высокую общую эффективность сжатия.

Ключевые слова: измерительные данные, кадр данных, сжатие, коэффициент сжатия.

DOI: 10.15372/AUT20180307

Введение. Наиболее активно проблема сжатия измерительных данных обсуждалась в середине–конце XX века, но и в настоящее время ей уделяется достаточно пристальное внимание в работах как отечественных [1–5], так и зарубежных авторов [6–10]. Активное развитие и усложнение информационно-измерительных систем (ИИС) за последние десятилетия вызвало необходимость решения задачи уменьшения избыточности объёма передаваемых данных путём сжатия [6, 11] в целях достижения следующих условий: оптимизации использования каналов связи, увеличения экономической эффективности и т. д.

Для измерительных данных, полученных от разных источников одного сложного измерительного объекта, можно декларировать наличие двух видов корреляционных связей: явных и неявных [12]. Корреляционные связи явного вида определяются особенностями ИИС и измеряемых физических процессов, в то время как неявные корреляционные связи задаются неучтёнными взаимодействиями между параметрами системы, принятыми способами представления и кодирования информации в ИИС и т. д. Выявление неявных связей возможно путём введения процедур предварительной обработки и преобразования.

В классических способах сжатия принято учитывать только явные корреляции, что, в свою очередь, задаёт границу верхнего коэффициента сжатия (КС) для всех способов сжатия, применяемых к измерительным данным. В таком случае более эффективным может оказаться подход, основанный на двухступенчатом сжатии, где в качестве первой ступени предлагается способ предварительной обработки, заключающийся в использовании дельта-кодирования, который позволит проявить неявные корреляции, а в качестве второй ступени — способ сжатия, учитывающий корреляции внутри кадра. Важно отметить, что корреляции могут быть между единичными и между нулевыми битами, что также ведёт к разработке различных способов, принимающих во внимание данную особенность.

В современных измерительных системах наиболее часто используются широко известные универсальные способы сжатия данных, готовых к передаче по цифровому каналу связи [13]. Очевидно, что в ряде случаев применяются специализированные способы, информация о которых, однако, почти всегда закрыта. Таким образом, существует практическая необходимость в разработке высокоэффективных специализированных способов сжатия телеметрических данных.

Представление кадров данных. Для измерительных данных можно предположить, что в штатном режиме работы ИИС, как правило, арифметическая разность между одним и тем же отсчётом в двух соседних кадрах будет составлять не более четырёх единиц (без учёта знака) [14]. Это свойство предлагается называть «стационарностью» кадра измерительных данных и далее в работе термин используется именно в таком смысле. Из свойства стационарности следует, что в качестве способа предварительной обработки эффективным будет алгоритм дельта-кодирования, представляющий собой поэлементное вычисление операции арифметической разности между одними и теми же отсчётами в текущем и предыдущем кадрах, что позволит существенно снизить их динамический диапазон.

Если кадр содержит отсчёты n датчиков, то разностный кадр представим в виде вектор-столбца $\Delta \mathbf{d}$ размерностью n : $\Delta \mathbf{d} = (\Delta d_1, \Delta d_2, \dots, \Delta d_i, \dots, \Delta d_n)^T$. Предполагая, что измерительные данные имеют разрядность m , каждый элемент вектор-столбца Δd_i можно преобразовать в битовую последовательность и записать в виде вектор-строки $\Delta \mathbf{b}$ размерности m : $\Delta d_i = \Delta \mathbf{b}_i = (\Delta b_{i,1}, \Delta b_{i,2}, \dots, \Delta b_{i,j}, \dots, \Delta b_{i,m})$. Таким образом, вектор-столбец $\Delta \mathbf{d}$ приводится к виду матрицы $\Delta \mathbf{B}$ размера $n \times m$: $\Delta \mathbf{B} = (\Delta \mathbf{b}_1, \Delta \mathbf{b}_2, \dots, \Delta \mathbf{b}_i, \dots, \Delta \mathbf{b}_n)^T$. Предложенный подход позволяет совместить операции уменьшения динамического диапазона данных после дельта-кодирования и обнаружения корреляций между соседними отсчётами.

Способ сжатия на основе «паркетного» разбиения. В основе предлагаемого способа лежит последовательное деление кадра измерительных данных, представленного в виде отображённой на плоскость матрицы $\Delta \mathbf{B}$, на строго однородные области произвольного размера, которые могут состоять либо только из нулевых, либо только из единичных бит. В качестве одного из способов такого разбиения можно предложить «паркетное», при котором вся матрица $\Delta \mathbf{B}$ сегментируется на множество областей произвольного размера, полностью покрывающих её поверхность, как это показано на рис. 1, где серым цветом обозначены единичные биты, а белым — нулевые.

После разбиения полученные области предлагается кодировать одной парой координат $P_i[x_i, y_i]$, которая задаёт положение правого нижнего угла области, и состоянием её левого верхнего бита, кодирующим тип области (нулевая либо единичная). Такой подход позволяет сжимать области, размер которых превышает число бит, отводимое на хранение пары координат, описывающих положение области и бита её типа.

Важно отметить, что поиск областей осуществляется по принципу наибольшей площади, т. е. сначала будет найдена область, имеющая наибольшую возможную площадь,

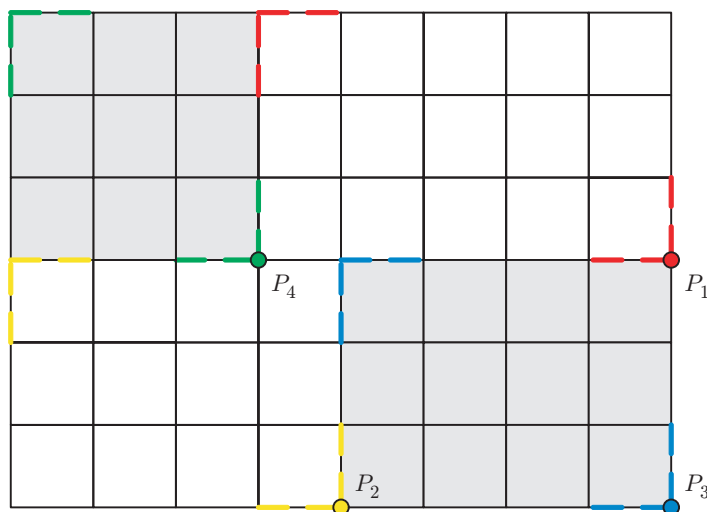


Рис. 1. Пример разбиения кадра данных

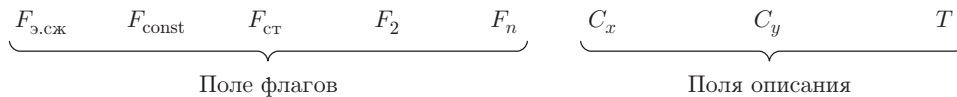


Рис. 2. Структура формата описания сжатого кадра данных

затем область, следующая за ней по величине, и т. д., пока не будет произведено разбиение всего кадра.

Непосредственно алгоритм «паркетного» разбиения можно описать как последовательность, состоящую из шагов:

- 1) создание буфера кодов (БК), в который предлагается помещать коды, описывающие однородные области;
- 2) поиск однородной области, обладающей максимальной площадью;
- 3) занесение в БК кода, описывающего полученную однородную область, после чего она помечается как найденная и больше в процессе поиска не участвует, затем происходит переход к шагу 2, в противном случае разбиение завершается.

Для измерительного кадра данных, представленного на рис. 1 и сформированного из шести восьмиразрядных отсчётов, будет произведено разбиение на четыре однородные области со следующими координатами правых нижних углов: $P_1[8, 3]$; $P_2[4, 6]$; $P_3[8, 6]$ и $P_4[3, 3]$.

Для описания результатов работы алгоритма предлагается использовать следующий формат (рис. 2), который будет описывать сжатый кадр в целях его декодирования и дальнейшей передачи по каналу связи. Важно отметить, что использование формата необходимо лишь в том случае, когда кадр данных является стационарным, поддаётся сжатию и значения хотя бы некоторых его отсчётов изменились. В общем случае заголовок такого формата состоит из пяти бит флагов $F_{\text{э.сж}}$, F_{const} , $F_{\text{ст}}$, F_2 и F_n , после которых идут поля описания выделенных областей C_x , C_y и T .

Флаги имеют следующий смысл: $F_{\text{э.сж}}$ фиксирует эффективность сжатия кадра данных (устанавливается, если коэффициент сжатия кадра больше единицы); F_{const} фиксирует неизменность значений отсчётов кадра (устанавливается, если результат дельтакодирования над всеми отсчётами кадра равен нулю); $F_{\text{ст}}$ фиксирует стационарность отсчётов кадра (устанавливается, если кадр содержит только стационарные отсчёты); F_n фиксирует наличие отрицательных отсчётов (устанавливается, если результат дельтакодирования хотя бы одного отсчёта является отрицательным); F_2 фиксирует использование второго разряда в отсчётах (устанавливается, если второй разряд хотя бы одного отсчёта содержит единичный бит). Поля описания имеют следующий смысл: C_x описывает координаты правых нижних углов областей на оси абсцисс, C_y — координаты правых нижних углов областей на оси ординат, T — типы областей.

Процедура восстановления кадра измерительных данных после сжатия является асимметричной для предложенного алгоритма и основывается на значениях, хранящихся в полях описания, при этом выполняется последовательность шагов:

1. Считываются пара координат и тип первой восстанавливаемой области.
2. На основе считанных значений происходит восстановление области, при этом её верхний левый элемент совмещается с верхним левым элементом восстанавливаемого кадра, который ранее не был помечен как восстановленный, после чего вся область помечается как восстановленная.
3. Если все области были восстановлены, то данная процедура завершается, в противном случае считываются пара координат и тип следующей восстанавливаемой области, после чего происходит переход к шагу 2.

Результаты экспериментальных исследований. Тестирование предложенного алгоритма обратимого сжатия проводилось на данных телеизмерений и телесигнализации, полученных от ряда объектов энергетики (подстанции, генерирующие станции) с помощью системы сбора и передачи телемеханики ТМ-800. В тестовые наборы данных (НД) входили отсчёты, полученные как в штатном режиме работы объекта (условно называемые стационарными), так и в режиме перевода энергосети из одного режима в другой (нестационарные). Несмотря на то что оборудование ТМ-800 обеспечивает передачу девятиразрядных данных, при организации сбора использовались только 8 бит, что определило разрядность отсчётов. В предлагаемом исследовании приведены результаты работы с пятью наиболее характерными наборами данных, объём которых составлял от 11 до 19,5 тысяч кадров. В табл. 1 сведены основные свойства наборов, такие как число кадров (ЧК) и число датчиков (ЧД) в кадрах, при этом оценивались дисперсия (Д) и математическое ожидание (МО). Анализируя результаты, представленные в таблице, можно сделать вывод, что стационарность наиболее ярко выражена в наборах НД1, НД3 и НД5, при этом наборы НД2, НД4 обладают этим свойством в меньшей степени.

При проведении исследования применялись алгоритмы сжатия данных RLE (Run-Length Encoding) (АС1), LZW (Lempel-Ziv-Welch) (АС2), на основе поиска нестрого однородных областей [15] (АС3), на основе «паркетного» разбиения (АС4). Важно отметить, что АС1, АС2 и АС3 представляют собой алгоритмы общего и специального назначения соответственно и используются для сравнения их параметров с предложенным в данной работе алгоритмом АС4.

В целях определения характеристик алгоритмов сжатия было проведено исследование, в котором над каждым НД осуществлялось сжатие с помощью всех предложенных алгоритмов и при этом фиксировались такие параметры, как средний коэффициент сжатия (СКС) и среднее время сжатия (СВС). Результаты тестирования алгоритмов приведены в табл. 2.

Таблица 1

Свойства наборов данных

Параметры	Набор данных				
	НД1	НД2	НД3	НД4	НД5
ЧК	17292	11440	19598	19283	19153
ЧД	32	48	40	56	48
Д	1,07059	45,15341	1,11350	19,07263	4,23198
МО	-0,00149	-0,02899	-0,00235	0,00001	-0,00222

Таблица 2

Усреднённые параметры алгоритмов

Параметры	АС	Набор данных				
		НД1	НД2	НД3	НД4	НД5
СКС, ед.	АС1	1,99	1,51	2,00	1,68	2,33
	АС2	1,76	1,63	1,89	1,88	2,25
	АС3	4,35	2,69	3,89	3,20	5,42
	АС4	3,32	1,29	2,81	1,68	4,01
СВС, мс	АС1	0,01	0,02	0,01	0,01	0,01
	АС2	0,03	0,04	0,03	0,04	0,03
	АС3	0,36	3,47	0,61	3,64	0,43
	АС4	0,07	0,23	0,10	0,21	0,09

Полученные результаты показывают, что в сравнении с алгоритмами общего назначения алгоритм АС4 затрачивает незначительно большее среднее время сжатия, обеспечивая наибольший средний коэффициент сжатия для стационарных данных, при этом для нестационарных данных этот показатель несколько ниже, чем для алгоритмов АС1 и АС2, а временные затраты более значительны. Важно отметить, что, в свою очередь, алгоритм АС3 хоть и показывает наибольший средний коэффициент сжатия среди рассмотренных алгоритмов, но при этом имеет достаточно низкую среднюю скорость сжатия как стационарных, так и нестационарных наборов данных, что связано с большими временными затратами, возникающими при поиске нестрогих однородных областей.

Поведение КС при использовании алгоритма на основе «паркетного» разбиения для наборов данных, различающихся по своим стационарным свойствам, представлено на рис. 3.

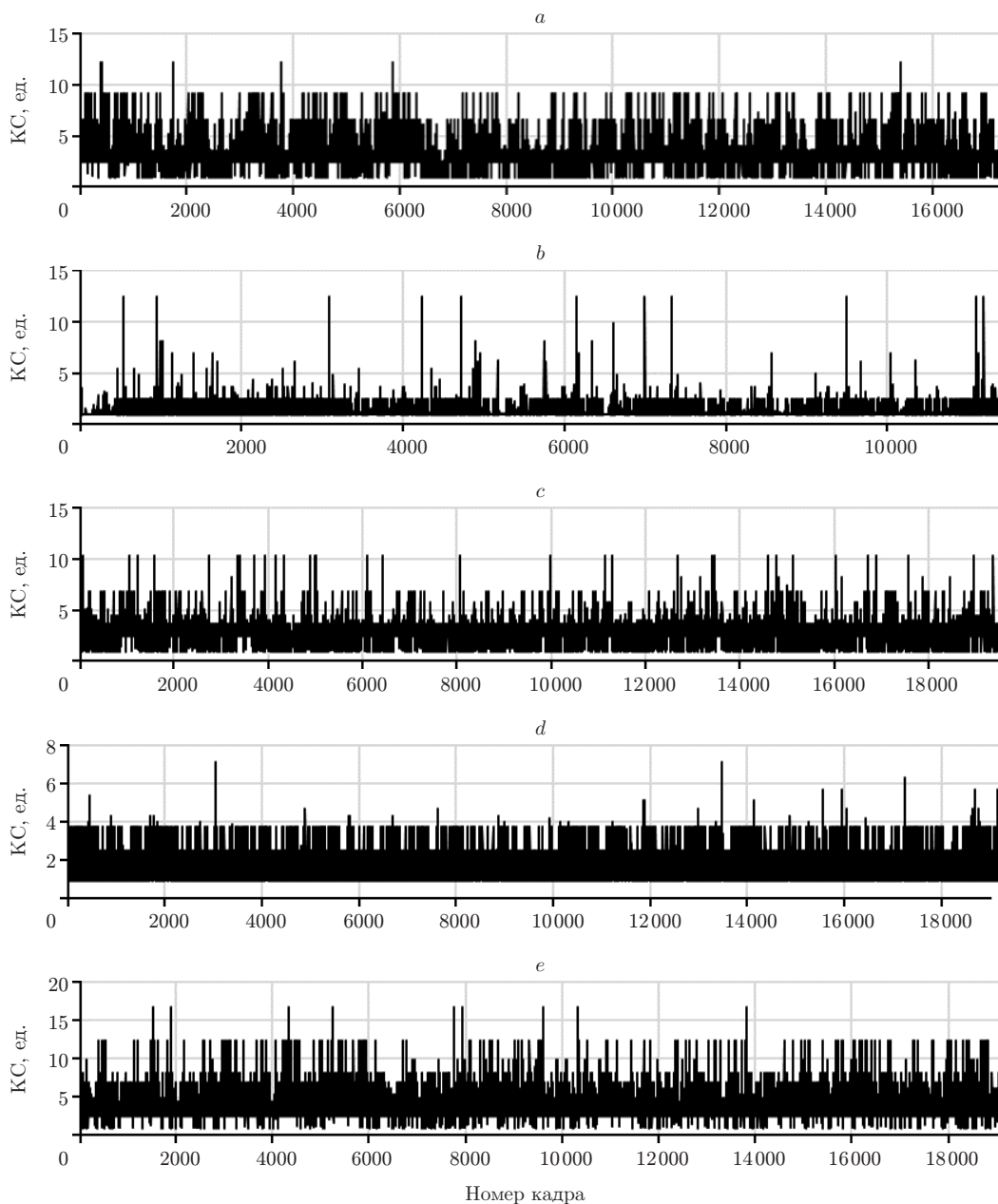


Рис. 3. Поведение коэффициента сжатия при работе алгоритма АС4 с наборами: *a* — НД1, *b* — НД2, *c* — НД3, *d* — НД4, *e* — НД5

Следует отметить, что для НД2 выход на рабочее значение среднего коэффициента сжатия происходит с некоторой задержкой, при этом возможны переходные периоды, в течение которых происходит спад, а затем достаточно плавный выход среднего значения коэффициента сжатия на некоторый постоянный уровень. Этот факт можно связать с внезапными существенными изменениями свойств потока сжимаемых данных. Анализ результатов сжатия стационарных потоков кадров показывает, что фиксируется большое число кадров, для которых обеспечивается высокий по отношению к среднему коэффициент сжатия.

Также стоит заметить, что для обоих типов потоков кадров характерен существенный разброс значений коэффициентов сжатия по отношению к среднему значению, что связано с нелинейным характером влияния числа выделенных однородных областей внутри кадра на поведение коэффициента сжатия.

Заключение. Результаты проведённых исследований продемонстрировали, что предложенный способ обратимого сжатия измерительных данных обеспечивает достаточно высокий средний коэффициент сжатия на данных с различными стационарными свойствами. Так, показано, что алгоритм АС4 эффективно работает с данными стационарной группы, обеспечивая для них средний коэффициент сжатия порядка 3,38, что больше, чем алгоритмы АС1 и АС2, для которых это значение составляет 2,11 и 1,97 соответственно, но ниже, чем значение для АС3 (4,55). Алгоритм АС4 создаёт условия для работы с нестационарными данными, показывая для них приемлемый средний коэффициент сжатия 1,49, что несколько ниже значений, полученных для алгоритмов АС1 (1,60), АС2 (1,76) и АС3 (2,95). При этом для алгоритма АС4 среднее время сжатия положительно коррелирует с эффективностью сжатия стационарных наборов, чего нельзя сказать при работе с наборами, обладающими нестационарными свойствами.

Дальнейшее развитие предложенного способа сжатия можно связать с оптимизацией описания полученных в результате разбиения областей, в частности, с уменьшением объёма данных, хранимых в полях формата описания сжатого кадра данных, но даже в этом виде его применение может оказаться эффективным в системах передачи телеметрических данных объектов энергетики, космических и беспилотных летательных аппаратов и т. п.

СПИСОК ЛИТЕРАТУРЫ

1. **Чье Ен Ун, Левенец А. В., Токарев Р. Е.** Применение популярных алгоритмов компрессии изображений для сжатия измерительных данных // Вестн. ТОГУ. 2012. **27**, № 4. С. 125–132.
2. **Ломтев Е. А., Мясникова М. Г., Мясникова Н. В., Цыпин Б. В.** Совершенствование алгоритмов сжатия-восстановления сигналов для систем телеизмерений // Измерительная техника. 2015. № 3. С. 11–15.
3. **Победоносцев В. А.** Определение количества информации о непрерывных сигналах. Элементарная теория. М.: Радиотехника, 2017. 168 с.
4. **Бевецкий А. В., Левенец А. В.** Алгоритм блочного сжатия измерительных данных // Ученые заметки ТОГУ. 2013. 4, № 4. С. 811–818.
5. **Трофимов В. К., Храмова Т. В.** Оптимальное равномерное по выходу кодирование для объединения различных множеств источников // Автометрия. 2017. **53**, № 1. С. 53–62.
6. **Bose R.** Combined data encryption and compression using chaos functions // Proc. SPIE. 2004. **5561**. P. 164–175.
7. **Candes E. J., Wakin M. B.** An introduction to compressive sampling // IEEE Signal Process. Mag. 2008. **25**, N 2. P. 21–30.

8. **Duarte M. F., Shen G., Ortega A., Baraniuk R. G.** Signal compression in wireless sensor networks // Philos. Trans. Royal Soc. A. 2012. **370**, Is. 1958. P. 118–135.
9. **Bormin H., Huang H.-L., Chen H. et al.** Data compression studies for NOAA Hyperspectral Environmental Suite (HES) using 3D integer wavelet transforms with 3D set partitioning in hierarchical trees // Proc. SPIE. 2004. **5238**. P. 255–265.
10. **Salomon D.** Data Compression. N. Y.: Springer, 2004. 900 p.
11. **Виттих В. А., Заездный А. М.** Постановка задачи сжатия измерительной информации и характеристики сжимателей информации // Автометрия. 1968. № 1. С. 13–18.
12. **Богачев И. В., Чье Ен Ун, Левенец А. В.** Статистический анализ телеметрических данных с точки зрения задачи сжатия // Информационно-управляющие системы. 2017. **86**, № 1. С. 11–16.
13. **Назаров А. В., Козырев Г. И., Шитов И. В. и др.** Современная телеметрия в теории и на практике. С.-Пб.: Наука и техника, 2007. 672 с.
14. **Левенец А. В.** Классификация телемеханических данных и их разностных рядов с точки зрения задачи сжатия // Вестн. ТОГУ. 2009. **15**, № 4. С. 71–80.
15. **Богачев И. В., Чье Ен Ун, Левенец А. В.** Геометрический подход к сжатию данных телеметрических систем // Информатика и системы управления. 2015. **46**, № 4. С. 16–22.

Поступила в редакцию 1 декабря 2017 г.
