



# Основы сжатия данных

Лекция 7

# Избыточность информации

- степень избыточности данных зависит от принятой системы кодирования
- избыточность русского языка выше, чем английского
- избыточность графических данных выше, чем текстовых
- избыточность видеоданных выше, чем графических

# Способы сжатия данных

- изменение содержания данных
  - методы сжатия с регулируемой потерей информации
    - *.JPG для графических данных*
    - *.MPG для видеоданных*
    - *.MP3 для звуковых данных*
- изменение структуры данных
  - обратимые методы сжатия
    - *.GIF, .TIF, .PCX для графических данных*
    - *.AVI для видеоданных*
    - *.ZIP, .ARJ, .RAR, .LZH, .LH, .CAB*

# Обратимые методы

- Для любой последовательности данных существует теоретический предел сжатия, который не может быть превышен без потери части информации
- Для любого алгоритма сжатия можно указать такую последовательность данных, для которой он обеспечит лучшую степень сжатия, чем другие методы
- Для любого алгоритма сжатия можно указать такую последовательность данных, для которой этот алгоритм вообще не позволит получить сжатия

# Алгоритмы обратимых методов

<b>Алгоритм</b>	<b>Выходная структура</b>	<b>Сфера применения</b>	<b>Примечание</b>
RLE (Run-Length Encoding)	Список (вектор данных)	Графические данные	Эффективность алгоритма не зависит от объема данных
KWE (Keyword Encoding)	Таблица данных (словарь)	Текстовые данные	Эффективен для массивов большого объема
Алгоритм Хаффмана	Иерархическая структура (дерево кодировки)	Любые данные	Эффективен для массивов большого объема

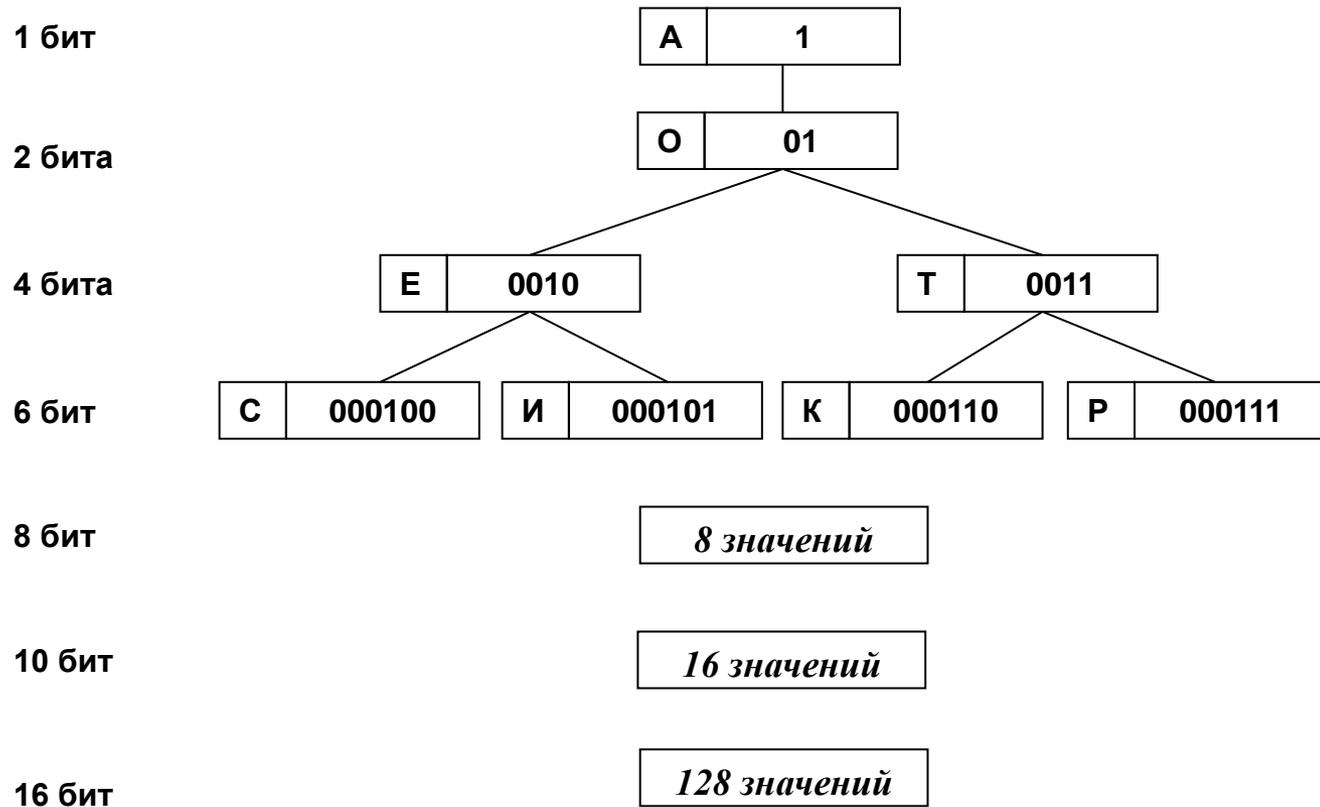
# RLE – групповое кодирование

- использует принцип выявления повторяющихся последовательностей данных и замены их простой структурой, в которой указывается код данных и коэффициент повтора
  - пример – для последовательности: 0; 0; 0; 127; 127; 0; 255; 255; 255; 255 (всего 10 байтов) образуется следующий вектор: 0; 3; 127; 2; 0; 1; 255; 4 (всего 8 байтов), что дает коэффициент сжатия равный 80%

# KWE – сжатие со словарем

- использует кодирование лексических единиц исходного документа группами байтов фиксированной длины
- результат кодирования сводится в таблицу, которая прикладывается к результирующему коду и представляет собой словарь
  - *токены* - двухбайтная кодировка слов английского языка

# Код переменной длины (Хаффмана)



# Синтетические алгоритмы

- сложные алгоритмы, основанные на комбинации нескольких теоретических методов
- используют предварительный просмотр и анализ исходных данных для индивидуальной настройки алгоритма

# Диспетчер архивов

- извлечение файлов из архивов
- создание новых архивов
- создание самораспаковывающихся архивов
- создание распределенных архивов на носителях малой емкости
- тестирование целостности структуры архивов
- полное или частичное восстановление поврежденных архивов
- защита архивов от просмотра и несанкционированной модификации

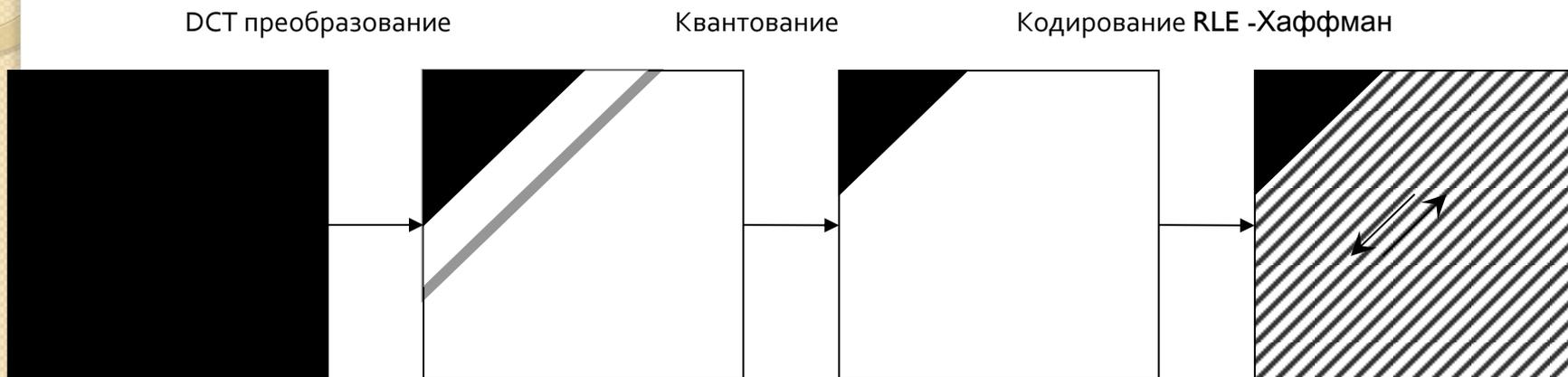
## Самораспаковывающиеся архивы

- готовится на базе обычного архива с присоединением к нему модуля программы распаковки
- файл архива имеет расширение *.EXE* и запускается как программа
- защита архивов выполняется с помощью пароля

# Алгоритм JPEG

- преобразование в цветовую модель с яркостным каналом
- преобразование участков 8x8 в числовой массив данных с помощью DCT
- разные уровни квантования частот
  - нули – групповое кодирование
  - ненулевые – алгоритм Хаффмана

# Сжатие JPEG



- декомпрессия файла JPEG происходит в обратном порядке
- данные, потерянные при квантовании, не могут быть восстановлены

# Форматы файлов JPEG

<b>Тип</b>	<b>Описание файла</b>
<b>JFIF</b>	<b>JPEG File Interchange Format</b>
<b>Raw JPEG</b>	<b>как JFIF</b>
<b>HSI</b>	<b>Handmade Software</b>
<b>Storm</b>	<b>Storm Technology</b>
<b>PICT-JPEG</b>	<b>Macintosh PICT</b>
<b>PhotoCD</b>	<b>Kodak Photo CD</b>

# Варианты JPEG

- чересстрочный режим использует Progressive JPEG
- технология вейвлетного сжатия в формате JPEG2000
- метод сжатия без потерь – **JBIG Joint Bi-level Image Group** позволяет управлять разрешением при декодировании изображений



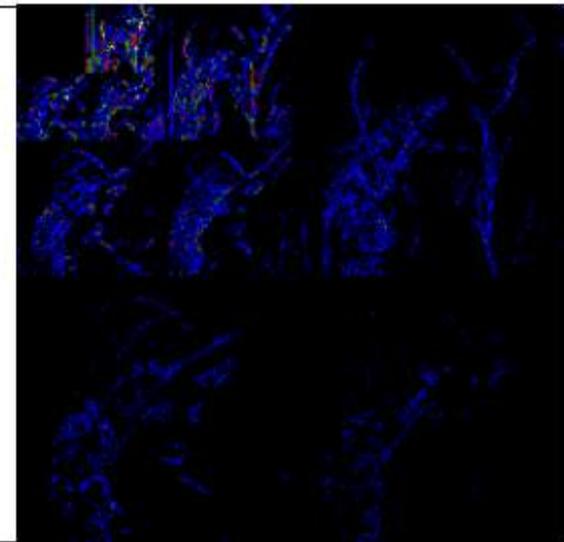
# Применение JPEG 2000

- мультимедийные устройства
- клиент/серверные взаимодействия
- военное применение
- медицинские изображения
- хранение видео
- сенсорные устройства
- цифровые устройства/архивы



# Сравнение JPEG

	JPEG	JPEG-2000
Структура разбиения	пространственная	частотная
Проход по коэффициентам	частотный	пространственный
Кодирование коэффициентов	групповое кодирование не нулевых	проходы по соответствующим битам коэффициентов



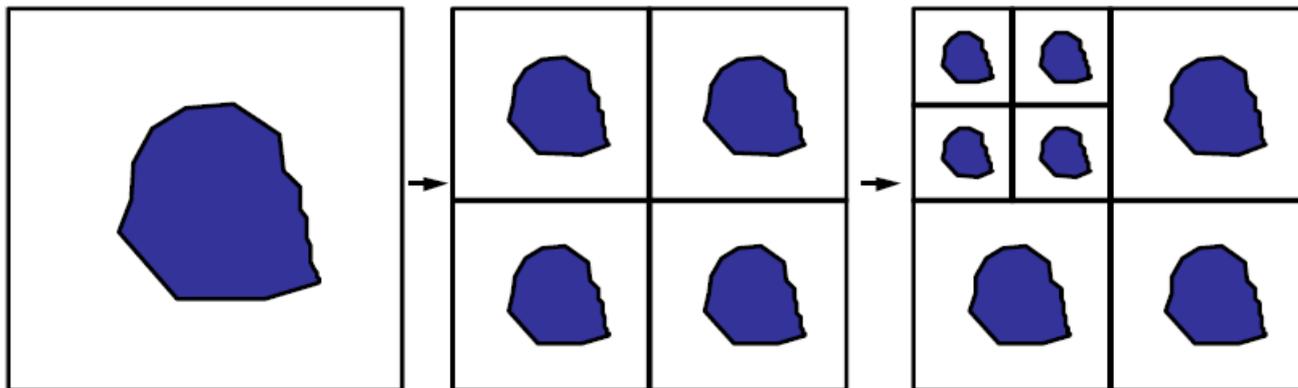
# JPEG/JPEG-2000 сжатие 130:1



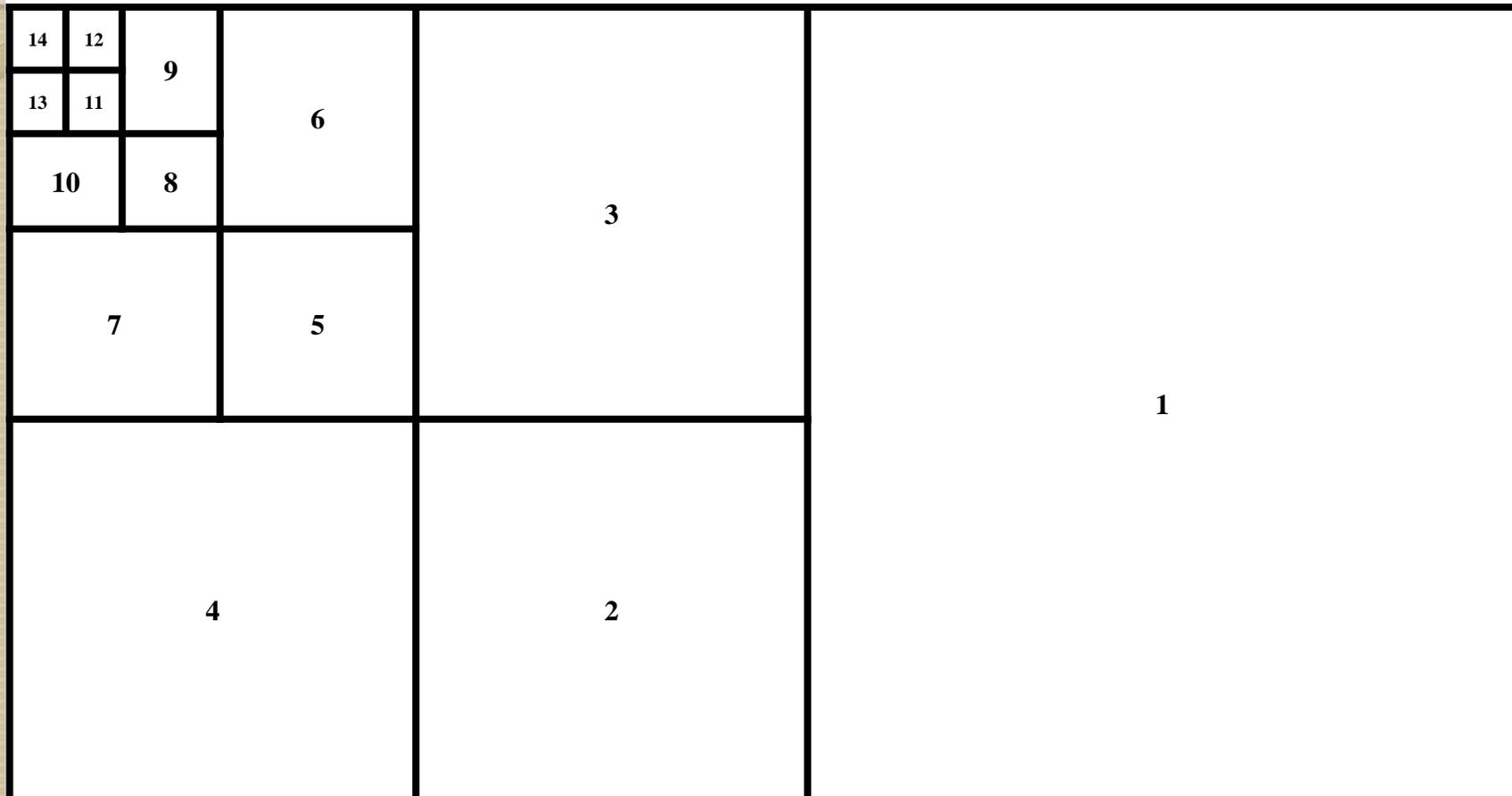
JPEG: сохранено больше деталей JPEG-2000: отсутствие блочных артефактов

# Вейвлетное сжатие

- любой сигнал может быть получен разложением на базовые функции
  - преобразование Фурье
  - дискретное косинус преобразование
  - би-ортогональное преобразование



# Дерево фильтрации

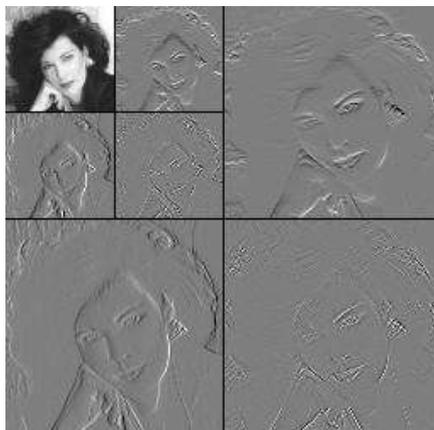
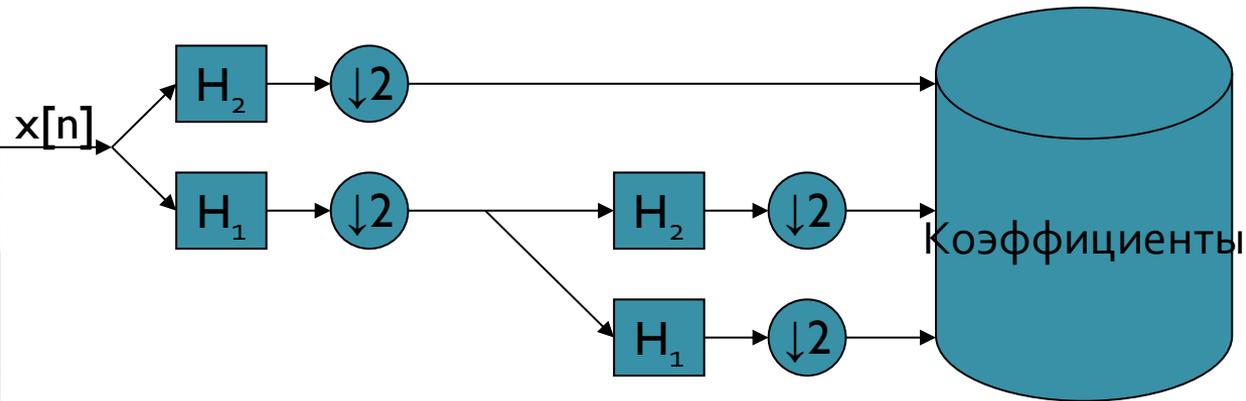


# Вейвлетное сжатие

Окончательное сжатие так же использует групповое кодирование и алгоритм Хаффмана



# Вейвлет-преобразование



## Двумерное вейвлет-преобразование

на каждом шаге получаем  
4 набора коэффициентов:  
НЧ («основные»)  
и ВЧ («детализирующие»)

# Фрактальное сжатие

- фрактал описывает структуру, состоящую из множества одинаковых форм и повторяющихся структур
- при увеличении фрактального изображения становятся различимы новые детали изображения.

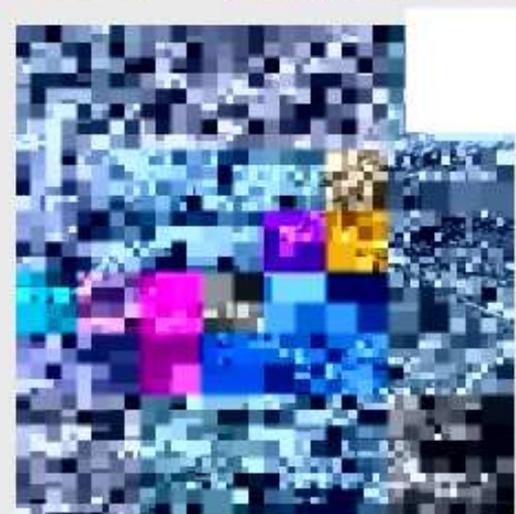
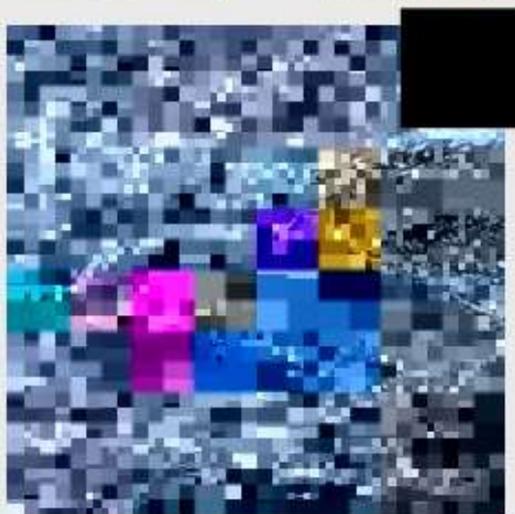


# Фрактальный алгоритм

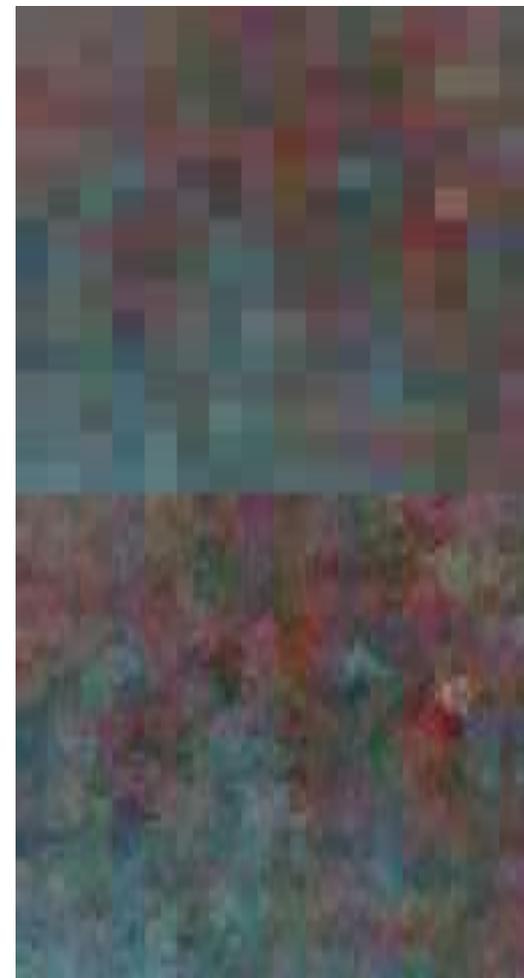
- разбиение на ряд базовых областей
- выявление последовательности зональных областей



# Фрактальная декомпрессия



# Фрактально-каскадный метод



# Тестовое изображение «Барбара»

- Много полосок (высоких частот) в разных направлениях и разной толщины



# Тестовое изображение «Boat»

- Много тонких деталей и наклонных границ в разном направлении



# Сравнение алгоритмов

Алгоритм	К-ты сжатия	Симметричн. по времени	На что ориентирован	Потери	Размерность
Групповое кодирование	1/32 1/2 2/1	1	3,4 битные	Нет	1D
<b>LZW</b>	1/100 1/4 7/5	1.2-3	1-8 битные	Нет	1D
Хаффмана	1/8 2/3 1/1	1-1.5	1-битные	Нет	1D
<b>JBIG</b>	1.5 раза	~1	1-битные	Нет	2D
<b>Lossless JPEG</b>	2 раза	~1	24-битн. сер.	Нет	2D
Рекурс. сжатие	2-20 раз	1.5	серые	Да	2D
<b>JPEG</b>	2-200 раз	~1	24-битн. сер.	Да	2D
<b>Фрактальный</b>	2-2000 раз	1000-10000	24-битн. сер.	Да	2D

**Цикл лекций подготовлен в 2011/2012 уч. году  
Кузнецовым Игорем Ростиславовичем,  
доцентом кафедры радиоэлектронных средств  
Санкт-Петербургского  
Государственного электротехнического  
университета «ЛЭТИ»**

Прочитан в дисциплине  
«Информационные технологии»