

**САНКТ-ПЕТЕРБУРГСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ТЕЛЕКОММУНИКАЦИЙ
им. проф. М. А. БОНЧ-БРУЕВИЧА**

Факультет В и ЗО

Дипломная работа

на тему

**Анализ эффективности методов сжатия
при передаче текстовых документов
в системах факсимильной связи**

Дипломник

Крюковец И. В.

Руководитель работы

Доронин Е. М.

**Санкт-Петербург
2005 г.**

Реферат

Дипломная работа посвящена анализу эффективности методов сжатия при передаче текстовых документов для факсимильных аппаратов G3.

Для анализа эффективности методов сжатия будем использовать один из трех критериев оценки сжатия документов (коэффициент сжатия, помехозащищенность и сложность алгоритмов реализации) коэффициент сжатия, как наиболее показательный.

Рассмотренные методы кодирования длин серий (КДС) и адресно-позиционное кодирование (АПК) позволяют эффективнее передавать монохромные документы.

Для метода КДС: за счет нахождения оптимальных длин серий путем последовательного перебора значений, при которых коэффициент сжатия будет максимальным.

Для метода АПК: за счет нахождения зависимости коэффициента сжатия от вероятности новых значений.

Пояснительная записка к дипломной работе содержит 61 страницу текста, 20 рисунков, 13 таблиц и 5 графиков.

Ключевые слова: факсимильный аппарат, методы сжатия, документы, коэффициент сжатия.

Содержание

Введение	4
1. Передача изображений в системах факсимильной связи	4
1.1. Стандарты на терминалы факсимильной связи	7
1.2. Характеристики и возможности факсимильных аппаратов G3	10
1.3. Рекомендации для факсимильных аппаратов G3	11
1.4. Современные факсимильные системы	16
2. Примеры методов сжатия двухградационных изображений	24
2.1. Кодирование длин серий (КДС)	25
2.2. Адресно-позиционное кодирование (АПК)	29
2.3. Алгоритм оптимального неравномерного кодирования (метод Хаффмана)	32
2.4. Двумерный код Хаффмана	37
2.5. Другие методы кодирования	38
3. Анализ эффективности метода КДС при передаче текстовых документов	39
3.1. Пример оценки избыточности для двухградационного черно-белого изображения	39
3.2. Расчет оптимальных длин кодовых комбинаций для метода КДС	40
3.3. Программная реализация статистической обработки и сжатия текстовых документов методом КДС	43
4. Оценка эффективности метода АПК для сжатия монохромных изображений документов	53
4.1. Постановка задачи	53
4.2. Расчет коэффициента сжатия. Выводы о применимости метода АПК	55
Заключение	60
Список литературы	61

Введение

Происходящее в последнее десятилетие в России и за рубежом широкомасштабное внедрение систем факсимильной связи рассматривается как ключевое звено при реализации новых информационных технологий в различных сферах управления и производства.

В общем случае системы факсимильной связи включают достаточно представительные множества передающего, сетевого и оконечного оборудования, основной отличительной особенностью которых является ориентация на более эффективную передачу различного рода факсимильных сообщений (ФС).

Для существующих и перспективных систем факсимильной связи характерно использование:

- практически полного множества линий и сетей связи в качестве основной коммуникационной среды;
- аналоговых и цифровых методов на различных этапах обработки и передачи ФС;
- широкого набора специальной оконечной аппаратуры факсимильных служб связи, осуществляющей эффективное кодирование и оперативную передачу графической информации с высоким уровнем автоматизации всех процессов ввода, передачи и регистрации документов и широким набором сервисных услуг.

Анализ особенностей организации факсимильной связи показал, что для большинства современных систем наибольшая специфика процессов передачи факсимильной информации определяется оконечной терминальной аппаратурой, реализующей сложные, в большинстве случаев автоматически выполняемые функции кодирования, обработки и передачи ФС. При этом одним из наиболее существенных факторов, влияющих на эффективность передачи факсимильной информации, является использование цифровых методов кодирования со сжатием неподвижных изображений.

Наличие различных методов кодирования для сжатия факсимильных изображений требует детального их анализа с точки зрения оценки эффективности для уменьшения объема передаваемого факсимильного сообщения.

1. Передача изображений в системах факсимильной связи

Для передачи и воспроизведения на расстоянии неподвижных изображений предназначена система факсимильной связи. Передаваемым изображением может быть рукописный, машинописный или типографский текст, фотография, чертеж, карта и др.

Факсимильное сообщение следует считать непрерывным - любой участок изображения сколь угодно малых размеров может иметь любое значение оптической плотности в некотором диапазоне плотностей. Ограниченные возможности зрения позволяют дискретизировать непрерывное факсимильное сообщение и тем самым снизить его избыточность до разумной величины. Дискретизация производится по площади изображения и по оптической плотности отдельных участков с учетом разрешающей способности и контрастной чувствительности глаза.

Суть дискретизации по площади заключается в разложении всей площади изображения на отдельные элементарные площадки. Каждая площадка характеризуется одним единственным значением оптической плотности, получаемым путем усреднения плотностей в пределах этой площадки.

В факсимильных аппаратах (ФА) общего назначения размер элементарной площадки должен составлять около 0,1 мм.

Дискретизация по оптической плотности заключается в округлении полученного значения плотности элементарной площадки оригинала до ближайшей фиксированной величины. В результате на копии будут воспроизведены только определенные, фиксированные плотности, число которых сравнительно невелико. Для качественной передачи художественной черно-белой фотографии необходимо передать и воспроизвести на копии 10-15 градаций оптической плотности, отличающихся друг от друга. При передаче двухградационных изображений дискретизацию производят двумя уровнями оптической плотности - черным и белым.

В передающей части факсимильного аппарата световой луч просматривает (сканирует) неподвижное изображение и образует на светочувствительном приемнике его электрическую копию (рис. 1.1).

Каждой точке (элементу) изображения оригинала соответствует электрический сигнал. В процессе считывания он превращается в последовательность «0» и «1» - цифровую кодовую комбинацию. Цифровые комбинации преобразуются далее в аналоговые сигналы - в последовательность импульсов, которые и поступают в канал связи. На приемной стороне процесс происходит в обратном порядке. Аналоговые сигналы демодулируются и преобразуются в оцифрованное изображение, которое распечатывается на бумаге.

Перечисленные операции в том или ином виде реализованы в факсимильном аппарате любой системы и постоянно совершенствуются с появлением новых технических решений.

Передача черно-белых изображений документов и другой графической информации и черно-белом варианте осуществляется службами факсимильной связи.

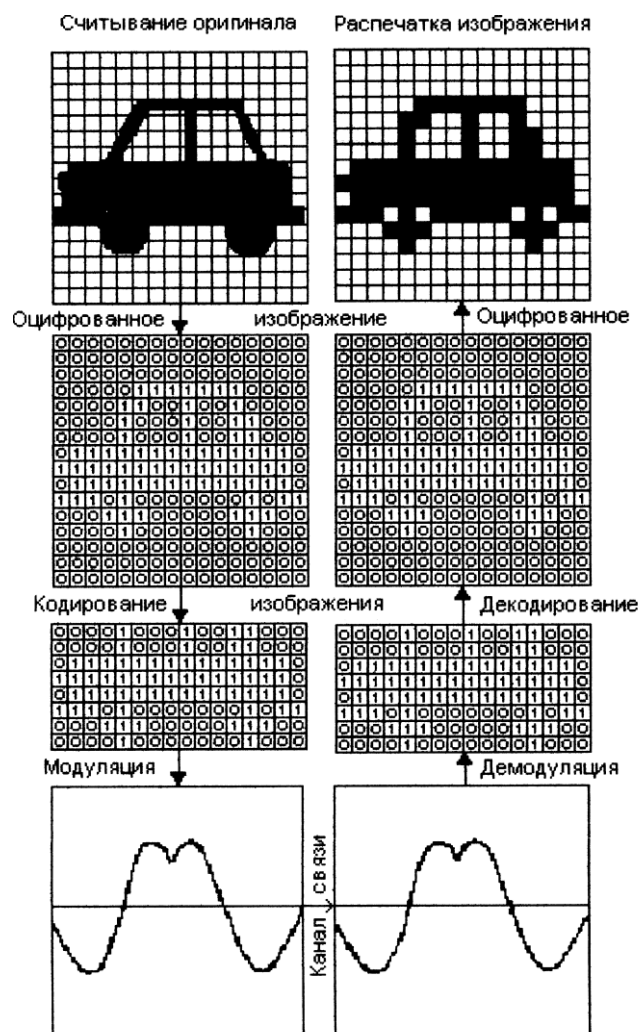


Рис. 1.1. Процесс факсимильной передачи и приема

Телефакс

Телефакс является самым распространенным в настоящее время способом передачи черно-белой графической информации. В качестве окончного оборудования выступают факсимильные аппараты, подключаемые к обычной телефонной линии и представляющие собой гибрид электронного телефонного аппарата и печатающего устройства.

Сам факсимильный аппарат снабжен электронной системой автоматического приема и передачи графической информации, закодированной в цифровом виде. В качестве службы, отвечающей за саму передачу информации, выступают автоматические телефонные станции с вытекающими отсюда недостатками и преимуществами. Абонент ставится в зависимость от качества канала, предоставляемого телефонной станцией. Канал может иметь шумы, помехи или малый уровень сигнала (тихо слышно). Для передачи речевой информации эти условия могут быть достаточны, но для передачи цифровой информации абсолютно неприемлемы. Современные цифровые АТС предоставляют такой вид услуг, как предоставление высококачественных каналов связи для

определенных абонентов, использующих цифровую аппаратуру связи (факсы и модемы).

Бюрофакс

Служба «Бюрофакс» введена для предоставления услуг по передаче факсимильных сообщений с использованием более качественной оконечной и промежуточной аппаратуры связи для более широкого круга потребителей. Аппаратура для приема и передачи факсимильных сообщений размещается в отделениях связи, где можно разместить заказ на прием или передачу факсимильного сообщения. Это так называемая клиентская служба факсимильной связи.

Комфакс

Комфакс является абонентской службой передачи факсимильных сообщений с накоплением. Служба комфакс используется в случае, если необходима пересылка факсимильных сообщений большому количеству адресатов, имеющих в своем распоряжении факсимильные аппараты, но нет полной уверенности в том, что приемная сторона будет готова к приему и связь не будет прервана по какой-либо причине. Служба комфакс состоит из автоматизированного центра по приему факсимильных сообщений, сервера для временного хранения рассылаемых сообщений и аппаратуры для передачи сохраненных факсимильных сообщений указанным адресатам. Эта служба получила широкое распространение с появлением компьютерной почты и в частности с появлением сети Интернет. Для передачи факсимильного сообщения достаточно выслать по электронной почте заказ на передачу документа. При этом нет необходимости в связи по междугородным линиям и постоянном контроле над ходом самой пересылки. Задача по пересылке сообщений абонентам ложится на сам центр службы комфакс.

1.1. Стандарты на терминалы факсимильной связи

Факсимильная связь является видом документальной связи, предназначенной для передачи не только содержания, но и внешнего вида самого документа. Допустимые форматы документов ISO A4, ISO B4 и ISO A3. При этом применяется стандартная плотность строк сканирования в вертикальном направлении 3,85 строк/мм или более высокие плотности (7,7 строк/мм, 15,4 строк/мм).

Согласно рекомендациям Международного союза электросвязи (ITU-T – International Telecommunications Union – Telecommunications) в зависимости от типа сети связи и используемого вида модуляции различают ФА 4-х групп:

1. Группа 1 (G1). Аналоговый метод передачи. Аппаратура, использующая модуляцию с двумя боковыми полосами без специальных мер по сжатию ширины полосы передаваемого сигнала и обеспечивающая передачу документов формата ISO A4 с плотностью развертки 4 строки/мм в течение приблизительно 6 минут. В настоящее время ФА G1 практически не используются;
2. Группа 2 (G2). Аналоговый метод передачи. Аппаратура, использующая методы сжатия ширины полосы с целью снижения времени передачи до 3 минут. Сжатие ширины полосы включает кодирование и/или частичное подавление боковой полосы (БП), но исключает обработку данных для снижения избыточности. В настоящее время ФА G2 практически не используются;
3. Группа 3 (G3). Цифровой метод передачи. Аппаратура, содержащая средства обработки данных для снижения избыточности до процесса модуляции и обеспечивающая передачу документа формата ISO A4 приблизительно за 1 минуту. Аппаратура может осуществлять сжатие ширины полосы сигнала. Аппаратура может подключаться к 64 Кбит/с каналам ЦСИО / ISDN;
4. Группа 4 (G4). Цифровой метод передачи. Аппаратура, содержащая средства обработки данных для снижения избыточности до процесса передачи. Аппаратура может использоваться в КТСОП, для чего будет применяться соответствующий метод модуляции.

В последние 5 лет интенсивно осуществлялась доработка ФА аппаратов 3 группы; в том числе подключение их к каналам цифровых сетей с интегрированным обслуживанием (ЦСИО) и организация взаимодействия с терминалами G4.

Это привело к расширению функциональных возможностей терминалов G3. В то же время проводились работы по расширению возможностей терминалов G4. Была поставлена задача, чтобы перспективные терминалы, в том числе универсальные (комбинированные), обеспечивали совместимость с уже существующими.

Стандарт на ФА G3 был определен рекомендацией ITU-T T.4 в 1980 году. Этот стандарт был переиздан в 1984 г. и в 1988 г. В модификации этого стандарта от 1990 г. были одобрены схемы кодирования, разработанные для факсимильных аппаратов G4, а также более высокие скорости передачи, определяемые стандартами V. 17, V. 29 и V. 33. Радикальное отличие ФА G3 от более ранних заключается в полностью цифровом методе передачи со скоростями до 14400 бит/с.

В результате сжатия данных ФА G3 передает страницу за 30-60 с. Согласно стандарту для ФА G3 возможны две степени разрешения: стандартное, обеспечивающее 1728 точек по горизонтали и 100 точек/дюйм по вертикали; и высокое, удваивающее количество точек по

вертикали, что дает разрешение 200x200 точек/дюйм и вдвое уменьшает скорость.

Развитие техники связи идет в направлении цифровизации. Это связано с тем, что цифровые системы имеют ряд существенных преимуществ по сравнению с аналоговыми: возможность регенерации сигнала на промежуточных пунктах; универсальность способа передачи, что позволяет передавать факсимильную информацию наряду с другими видами сообщений по единой цифровой системе связи; возможность сокращения времени передачи за счет оптимального кодирования факсимильного сигнала; применение хорошо развитой цифровой элементной базы при создании новых образцов цифровой факсимильной аппаратуры; возможность сопряжения цифровых факсимильных аппаратов с ЭВМ в целях обработки факсимильной информации.

Факсимильные аппараты G3 знаменуют переход от аналоговых к цифровым методам передачи факсимильных сообщений и существенно отличаются от аппаратов G1 и G2. Цифровые факсимильные аппараты рассматривают изображение как набор дискретных элементов (пикселей). Каждый пиксель может быть или черным (ему соответствует двоичная 1) или белым (двоичный 0). Последовательность двоичных цифр, вырабатываемая при оцифровке изображения, может быть сжата, передана по каналу связи и проконтролирована на наличие ошибок с использованием любых методов, применяемых в модемах. Все аппараты G3 обеспечивают стандартное вертикальное разрешение 3,85 линии на миллиметр. Возможна установка более высокого разрешения - 7,7 линии на миллиметр. В обоих случаях разрешение по горизонтали устанавливается на уровне 1728 пикселей для стандартной линии сканирования длиной 215 мм, что эквивалентно 8 пикселям на миллиметр. В модификации стандарта T.4, датированной 1992 г., добавляются новые режимы кодирования с более высоким разрешением: 15,4 линии на миллиметр по вертикали и 16 пикселей на миллиметр по горизонтали. Таким образом, каждая строка сканирования содержит 1728 элементов изображения (ЭИ) на 215 мм ширины, 2048 ЭИ на 255 мм ширины и 2432 ЭИ на 303 мм ширины соответственно для документов формата А4, В4 и А3. Это стандартные базовые длины строк.

В цифровых факсимильных аппаратах блоки электрооптического анализа и синтеза дискретные (в качестве преобразователей «свет-сигнал» используется, например, ПЗС-линейка и термопечатающая линейка).

В результате цифровых преобразований получается последовательность единичных элементов (двоичный код), поэтому скорость передачи факсимильной информации измеряют количеством бит в секунду (бит/с).

Цифровые данные, формирующие каждую линию сканирования, уплотняются перед передачей.

Поэтому время, затрачиваемое на передачу одной линии сканирования, может изменяться от линии к линии внутри некоторого

произвольно определяемого интервала. Стандарт T.4 рекомендует затрачивать на передачу каждой линии сканирования не менее 20 мс, хотя допустимо изменение этого параметра в пределах от 0 до 40 мс. Максимальное время передачи линии сканирования не должно превышать 5с.

Внутри линии сканирования группы смежных черных и белых пикселей чередуются. Для обеспечения синхронизации цветов воспроизводимого и передаваемого изображений все линии должны начинаться с группы белых пикселей. Если линия начинается с черного пикселя, передается цепочка белых пикселей нулевой длины.

Факсимильные аппараты первых трех групп ориентированы на использование аналоговых телефонных каналов КТСОП. В 1984 году ITU-T принял стандарт G4, который предусматривает разрешение до 400x400 точек/дюйм и повышение скорости при более низком разрешении. ФА G4 дают разрешение очень высокого качества. Однако они нуждаются в высокоскоростных каналах связи, которые могут предоставить сети ISDN, и не могут работать через каналы КТСОП.

Применение аппаратов группы G4 в основном концентрируется на интеграции службы факсимильной связи с другими службами связи и передачи сообщений. Служба факсимильной связи с использованием аппаратов G4 добавляет некоторые полезные усовершенствования в передачу факсимильных сообщений, не отказываясь при этом от концепции растровой графики. Определены, например, базовые величины разрешения: 200, 300 и 400 точек на дюйм. Добавлен режим передачи без сжатия данных, позволяющий осуществлять передачу изображения, не применяя обычные алгоритмы сжатия. Новый режим, называемый смешанным режимом, обеспечивает пересылку текста как данных и передачу графических изображений в растровой форме.

Стандарты для аппаратов группы G4 ввели три класса факсимильных терминалов в соответствии с их рабочими характеристиками. Терминалы, которые просто посылают и принимают факсимильные сообщения, определяются как терминалы класса 1. Терминалы класса 2 могут получать как телексы (текстовые сообщения, пересылаемые между факсимильными аппаратами), так и смешанные сообщения. Терминалы группы G3, могут создавать, передавать и принимать сообщения в любой форме.

Практически все продаваемые в настоящее время ФА основаны на стандарте группы G3.

1.2. Характеристики и возможности факсимильных аппаратов G3

Современные ФА G3 отличаются по способу нанесения изображения на бумагу и, соответственно, типу используемой бумаги. Традиционный способ заключается в выжигании изображения на специальной термочувствительной бумаге, но большая часть ФА, использует обычную

писчую бумагу типа ксероксной. Из других возможных функций ФА можно выделить следующие:

1. АВТООТВЕТЧИК. Двух типов: кассетный и электронный. Электронные, в которых отсутствует магнитная лента и головки, качественнее и надежнее, но недостатком является малая емкость записи (до 10 минут).
2. ГРОМКАЯ СВЯЗЬ. Возможность вести переговоры, не снимая трубки.
3. ПАМЯТЬ НОМЕРОВ. Возможность присваивать функциональным кнопкам номера абонентов, а также наличие встроенной электронной записной книжки, позволяющей заносить в память не только номера, но и комментарии к ним.
4. АВТООБРЕЗ. ФА отрезает лист по окончании приема.
5. ПОЛУТОНА. Наибольшее количество оттенков серого цвета, передаваемых ФА, обеспечивает более качественное изображение при передаче или приеме, если ФА абонента имеет сходные характеристики.
6. ОТЛОЖЕННАЯ ПЕРЕДАЧА. ФА фиксирует в памяти изображение и передает его автоматически из памяти в удобное время.
7. ОТЛОЖЕННЫЙ ПРИЕМ. ФА осуществляет прием сообщений при отсутствии бумаги в свою память.
8. АВТОДОЗВОН. ФА на стадии набора номера определяет, свободен абонент или занят.
9. МАСШТАБИРОВАНИЕ. Возможность уменьшать/увеличивать передаваемое изображение.
10. ДИСПЛЕЙ. ФА могут быть оборудованы жидкокристаллическим экраном, что позволяет видеть больше информации.

Технические характеристики факсимильных аппаратов группы G3 приведены в табл. 1.1.

1.3. Рекомендации для факсимильных аппаратов G3

Рассмотрим более подробно рекомендации для ФА группы 3 (табл. 1.2)

Таблица 1.2

Рекомендация	Скорость передачи (бит/с)
V.21	300
V.17	9600, 14400
V.27bis	4800
V.27ter	2400, 4800
V.29	7200, 9600

Кроме указанных протоколов факсимильный аппарат должен соответствовать рекомендациям МККТТ Т.4 и Т.30 из перечня рекомендаций серии Т «Оконечное оборудование и протоколы для телематических служб».

Протокол V.21

Дуплексный протокол с частотным разделением каналов и частотной модуляцией. Скорость передачи данных 300 бит/с. Предназначен для использования как в коммутируемых телефонных сетях общего пользования, так и на арендованных линиях. Так как модуляция двоичных данных осуществляется путем изменения частоты сигнала, то скорость модуляции равна скорости передачи данных. На нижнем канале (его обычно использует для передачи вызывающий модем) "1" передается частотой 980 Гц, а "0" - 1180 Гц. На верхнем канале (передает отвечающий модем) "1" передается частотой 1650 Гц, а "0" - 1850 Гц. Данные могут передаваться в синхронном либо асинхронном режиме.

Протокол V.21 используется факсимильными аппаратами группы 3 (для передачи управляющих команд при факсимильной связи используется только верхний канал) и факс-модемами на этапе установления связи. Сама передача факсимильных сообщений происходит уже в соответствии с другим, более эффективным протоколом.

Протокол V.17

Полудуплексный факсимильный протокол для использования в коммутируемых телефонных сетях общего пользования. В нем используется модуляция с решетчатым кодированием, или треллис-модуляция (ТСМ - Trellis Coded Modulation). ТСМ является вариантом многопозиционной амплитудно-фазовой модуляции, которую называют еще квадратурной амплитудной модуляцией (QAM - Quadrature Amplitude Modulation). Здесь изменяется и фаза, и амплитуда сигнала, однако, для повышения помехозащищенности передачи информации, вводится избыточность, что позволяет в процессе демодуляции использовать более совершенный алгоритм декодирования.

Частота несущего сигнала - 1800 Гц, модуляционная скорость 2400 Бод. Протокол имеет режимы 16-ТСМ, 32-ТСМ, 64-ТСМ и 128-ТСМ. Соответственно информационная скорость может быть 7200, 9600, 12000 и 14400 бит/с.

Протокол V.27ter

В этом протоколе применяется фазоразностная модуляция (DPSK - Differential Phase Shift Keying) с частотой несущего сигнала 1800 Гц. Могут использоваться два режима с разными информационными скоростями: 2400 и 4800 бит/с. Информационная скорость 2400 бит/с достигается модуляционной скоростью 1200 Бод и кодированием дибита (4-позиционная DPSK) а 4800 бит/с - скоростью 1600 бод и кодированием трибита (8-позиционная DPSK).

Протоколы данного семейства - V.27 и V.27bis используются редко. Они отличаются от V.27ter типом канала - выделенного четырехпроводного, для которого они предназначены.

Протокол V.29

Протокол для дуплексной работы ее скоростью передачи данных 9600 бит/с по 4-проводным арендованным каналам тональной частоты, удовлетворяющим рекомендациям M.1020 или M.1025 (рекомендации о характеристиках каналов особого качества). В протоколе применяется квадратурная амплитудная модуляция. Частота несущего сигнала - 1700 Гц, модуляционная скорость - 2400 Бод. Имеет режимы 8- и 16-позиционной QAM. Соответственно информационная скорость может быть 7200 и 9600 бит/с. Предусмотрена возможность снижения скорости до 7200 или 4800 бит/с. В состав модема, реализующего этот протокол, входит адаптивный автоматический корректор, а также скремблер. Может входить также мультиплексор для разделения общего потока (например, со скоростью 9600 бит/с) на несколько подканалов с меньшими скоростями (например, 2400 и 4800 бит/с).



Модель	Brother intelli FAX - 1270	Canon B 120	Canon L 900	Panafax DF-1100	Panasonic KX- FP363	Panasonic UF-560	Sagem Phonefax 320	Samsung MJ 4500C	Samsung SF-531P	Sharp FO- 4700	Xerox 7042	XEROX DWC450 cp	Xerox X7241
Технология печати	Лазерная	Струйная	Лазерная	Лазерная	Термо перен.	Лазерная	Лазерная	Струйная	Лазерная	Лазерная	Лазерная	Струйная	Термографическая
Формат бумаги	A4	A4	A4	Legal	A4	Legal	Legal	Legal	A4	Legal	Legal	A4	A4
Объем памяти, стр (стандартно / максим-но)	20	50	80/592	70/670	28	12/335	25	80	160	60/120	40/120	23	
Скорость модема max, Бит/сек	9600	14400	33600	14400	14400		9600	14400	33600	14400	9600	14400	9600
Скорость передачи, стр/мин	4	4,4	20	6	7,5	10	4	6	20	10	6	8,5	3
Скорость печати, стр/мин	2	1,13	8	10			2	8(ч/б), 3(цв)	8			7	

Разрешение печати, dpi	200x200	360x360	600x600	400x400		400x400		1200x1200	600x600		400x400	1200x600	
Разрешение печати в текстовом режиме, lpi (stand/fine/superfine)			98/196/392	98/196/391		98/196/392	98/196/391			100/200/400	98/196	98/196/391	98/196
Ширина сканирован., мм	208	208	208	212	208			210		213		212	210
Запас бумаги, стр	200	50	1100	250(1000 max)	50		230	130	150	250/750	200	150	
Объем автоподатчика, стр	20	15	50	30	10	80	-	30	30	50	20	20	10
Потребляем. мощность, Вт		34/4,5	380(max)	470(max)	135/4,4	460(max)				400		45(max)	35
Габариты, мм	385x387x213	400x228x177	406x460x355	459x500x412	185x355x275	430x440x280	267x234x95	190x360x325	373x325x228	691x594x462	155x380x180	305x419x444	314x233x121
Вес, кг	5,2	4,7	19	22	3,4	15	2,7	5,4	6,26	22		8,4	3,3
Автоответчик		10 мин.			18 мин.								опция
Сетевой интерфейс			да	опция		опция							

Рекомендация Т.4

Рекомендация Т.4 определяет стандарты факсимильной аппаратуры группы 3 для передачи документов: размер листа передаваемого документа, направление сканирования документа, разрешение, алгоритмы сжатия передаваемой графической информации и др. Например, для формата А4 (210x 297мм) зона гарантированного воспроизведения должна составлять 196,6x281.46 мм, направление развертки должно быть слева направо, сверху вниз; стандартная разрешающая способность - 3,85 строки/мм, минимальное время передачи всей кодированной строки развертки 20 мс, максимальное - не более 5 секунд; в качестве алгоритма сжатия определена одномерная схема кодирования длин серий (допускается использование двумерной схемы кодирования). Аппаратура группы 3, работающая по коммутируемой телефонной сети общего пользования, должна использовать сигнал модуляции, скремблера, коррекции и синхронизации, определенные в рекомендации V.27ter. При работе по арендованным каналам или коммутируемым цепям высокого качества указанные сигналы могут определяться Рекомендацией V.29.

Рекомендация Т.30

Рекомендация Т.30 «Процедуры для факсимильной передачи документов по коммутируемой телефонной сети общего пользования» регламентирует процедуру установления соединения, согласование параметров связи, передачу образа документа и завершение связи. Так, установление соединения может осуществляться вручную или автоматически; передаче сообщения предшествует процедура предсообщения, которая должна осуществлять идентификацию возможностей и контроль выбранных условий. Процедура передачи сообщения должна отвечать требованиям соответствующей Рекомендации, касающейся оборудования. В процессе передачи сообщения должно обеспечиваться управление синхронизацией, обнаружением и исправлением ошибок, а также контролем линии. Завершение связи должно включать последовательное выполнение двух этапов: процедуры пост-сообщения, обрабатывающей сигналы «конец сообщения», «многостраничная передача» и «конец факсимильной процедуры», и этапа разъединения соединения, выполняемого вручную или автоматически.

В ФА G3 используются различные методы сжатия информации.

1.4. Современные факсимильные системы

ФА представляет собой электромеханическое устройство, состоящее из сканера, модема, принтера, мотора и шестерней. Мотор и шестерни отвечают за нормальную подачу бумаги в сканер и принтер. Сканер считывает изображение документа, оцифровывает его и передает

информацию в модем. Модем преобразует цифровые сигналы в последовательность модулированных сигналов и обеспечивает их передачу на другой факсимильный аппарат через обычную телефонную линию. Модем принимающего телефакса преобразует данную последовательность обратно в цифровую и передает ее на принтер. Принтер распечатывает изображение на специальной термобумаге в соответствии с полученной информацией.

Недостатки телефаксов

- Подверженность значительному механическому износу. При частом использовании сканер телефакса забивается пылью и грязью, попадающим со считываемых документов. Пластиковые шестерни изнашиваются. Все это приводит к перекосам и неравномерной подаче как считываемых документов в сканер, так и термобумаги в принтер. Таким образом, качество передаваемых и принимаемых документов значительно ухудшается.
- Невозможность автоматизированной работы. По обычному ФА сложно отправлять документы большому числу адресатов.
- Неэффективное использование термобумаги. Большинство факсимильных аппаратов распечатывает все получаемые сообщения (в том числе и не несущие никакой полезной информации) на специальной термобумаге. Кроме высокой цены, у нее есть еще один существенный недостаток - изображение на ней выцветает со временем. Поэтому все важные сообщения необходимо копировать для хранения.

Новая аппаратура факсимильной связи

Сегодня многие пользователи заинтересованы в приобретении не простых автономных ФА, выполняющих строго определенные функции, а более совершенных систем, которые позволяют автоматизировать процесс приема, обработки и рассылки факсимильных сообщений и исключить отмеченные недостатки.

Идея создания интегрированных систем и применения в них персонального компьютера впервые была реализована в 1985 г., когда фирма GAMMA LINK выпустила первую компьютерную факсимильную плату. Это позволило подключить телефонную линию непосредственно к компьютеру и превратить его в мощный и многофункциональный ФА.

Сегодня компьютерные факсимильные платы выпускает огромное количество производителей. Их продукция, различающаяся по некоторым функциональным возможностям, служит одной цели - автоматизации процесса передачи, приема и распределения факсимильных сообщений, обмен которыми происходит по обычным телефонным линиям. Системы,

строящиеся на базе ПК с применением таких плат, обладают рядом существенных преимуществ перед обычными факсимильными аппаратами.

- Удобство использования. Интеграция ПК с телефонной сетью и наделение его возможностями ФА позволяет пользователям получать, обрабатывать и отправлять факсимильные сообщения, не отрываясь от своих компьютеров.
- Эффективное использование телефонных линий. Факсимильная система, строящаяся на базе ПК, обеспечивает эффективный обмен информацией по малому числу телефонных линий, заменяя собой множество автономных телефаксов, для каждого из которых требуется отдельная линия.
- Высокое качество передаваемого изображения. Любой документ текстового или графического редактора может быть передан в виде факсимильного сообщения высокого качества. Для этого с помощью специального программного обеспечения он преобразуется в формат, используемый факсимильной платой для передачи сообщений. Таким образом, гарантируется высокое качество изображения, поскольку документ не может быть «испорчен» ни низким качеством печати принтера, ни загрязнением сканера телефакса, ни неполадками в механизме подачи бумаги.
- Сохранение конфиденциальности принимаемых сообщений. В отличие от обычных телефаксов, распечатывающих все поступающие сообщения на едином рулоне бумаги факсимильные системы на базе ПК принимают и сохраняют их в персональных директориях пользователей, доступ к которым ограничивается паролем. Таким образом, полностью исключается просмотр важных документов посторонними людьми. Кроме того, применение ПК для управления работой факсимильных карт позволяет реализовывать множество полезных и удобных приложений. Многие из них предоставляют возможность полностью автоматизировать процесс обмена факсимильными сообщениями. К таким приложениям, получившим наиболее широкое распространение, относятся: ФАКС-СЕРВЕР, ФАКС ПО ЗАПРОСУ И ФАКС - РАССЫЛКА.

Применение ФАКС - СЕРВЕРА сводит к минимуму временные и материальные затраты при приеме и передаче факсимильных сообщений.

ФАКС ПО ЗАПРОСУ позволяет автоматизировать процесс предоставления абонентам часто запрашиваемых документов.

ФАКС - РАССЫЛКА значительно упрощает работу персонала при рассылке большого количества разных документов большому числу адресатов.

Факс-сервер

ФАКС - СЕРВЕР представляет собой компьютер, оборудованный несколькими специальными факсимильными платами (или одной многоканальной картой) и интегрированный с локальной вычислительной сетью (ЛВС). Он обладает многими преимуществами по сравнению с группой из нескольких автономных телефаксов, позволяя обмениваться факсимильными сообщениями с лучшим качеством, большими удобствами и меньшими издержками. ФАКС - СЕРВЕР наделяет каждого пользователя ЛВС возможностью передавать и принимать факсимильные сообщения с помощью своего рабочего ПК. При его использовании отпадает необходимость в термобумаге, так как все принятые сообщения сохраняются в виде файлов, которые в случае необходимости можно распечатать с помощью обычного сетевого или локального принтера. Облегчается контроль затрат на пересылку сообщений (ФАКС - СЕРВЕР регистрирует все процессы в файле-отчете), и наконец, ФАКС - СЕРВЕР является более дешевым вариантом, чем подключение каждой рабочей станции к телефонной сети с помощью модема.

ПЕРЕДАЧА.

На каждом ПК локальной сети устанавливается специальная программа. Она дает возможность пользователю отправлять документы со своего компьютера. Достаточно указать документ, подлежащий отправке, и телефонный номер адресата. Все остальное ФАКС - СЕРВЕР сделает сам, оповестив пользователя об успешной передаче документа адресату. Причем все несрочные сообщения могут быть сохранены на диске ФАКС - СЕРВЕРА и отправлены в ночное время по более низким тарифам. Некоторые факсимильные серверы также позволяют рассылать документы большому числу адресатов.

ПРИЕМ.

ФАКС - СЕРВЕР принимает каждое поступающее факсимильное сообщение и сохраняет его в общей директории либо в персональной директории пользователя, извещая об этом в первом случае секретаря, а во втором - конкретного пользователя. Права доступа к обеим директориям для каждого пользователя могут быть ограничены. Этим обеспечивается сохранение конфиденциальности принимаемой информации. Понятно, что сохранение поступающих сообщений в персональных директориях пользователей является наиболее удобным, однако для реализации такой возможности требуется применение специальных способов маршрутизации сообщений.

Факс по запросу

Системы ФАКС ПО ЗАПРОСУ (ФПЗ) позволяют автоматизировать обработку запросов абонентов с предоставлением им факсимильных сообщений. Как показывает практика, при обработке запросов вручную

обычно выполняется следующая последовательность действий. Абонент звонит по номеру, на котором не установлен ФА, и запрашивает какой-либо документ. Сотруднику нужно найти необходимый документ, дойти до аппарата, установить с абонентом новое соединение (известно, что 75% вызовов не достигают цели с первой попытки - занято, не отвечает, и т.д.) и отправить документ. Данный процесс обычно отнимает не менее пяти минут рабочего времени, а при передаче многостраничных документов временные затраты увеличиваются еще больше. Если допустить, что в день приходится обрабатывать 10 запросов, на каждый из которых уходит пять минут, то получается, что примерно одна восьмая часть рабочего времени тратится на выполнение рутинной работы. Если даже запрос поступает по линии, на которой установлен ФА, то не каждый абонент, тем более иногородний, может позволить себе ждать, пока секретарь найдет необходимый документ.

Применение систем ФАКС ПО ЗАПРОСУ является одним из подходов к решению подобного рода проблем. Системы ФАКС ПО ЗАПРОСУ позволяют в цепочке абонент-сотрудник-документ полностью исключить звено "сотрудник". Это достигается за счет широких возможностей взаимодействия с абонентами. Именно по этому признаку системы ФАКС ПО ЗАПРОСУ можно разделить на три вида - ПРОСТОГО ОТВЕТА, ВЕЩАТЕЛЬНЫЕ и ИНТЕРАКТИВНЫЕ.

Работа систем ПРОСТОГО ОТВЕТА выглядит следующим образом. Организация, имеющая в своем распоряжении систему ФАКС ПО ЗАПРОСУ, указывает в рекламе рядом с каждым коммерческим предложением его индекс. Клиент, позвонив с телефакса по номеру системы ФАКС ПО ЗАПРОСУ, вводит один из этих индексов, используя функцию тонального набора, доступную на любом современном телефаксе.

В ответ система автоматически высылает соответствующий индексу документ, в котором содержится подробная информация о коммерческом предложении. Все документы хранятся в системе в виде файлов специального формата. Их можно создать двумя путями: позвонить на номер системы с обычного телефакса и передать документ, после чего он будет автоматически преобразован в используемый формат, преобразовать документ текстового или графического редактора в используемый формат с помощью специального программного обеспечения.

ВЕЩАТЕЛЬНЫЕ системы ФАКС ПО ЗАПРОСУ, в отличие от систем ПРОСТОГО ОТВЕТА, позволяют абоненту выбирать документы, следуя кратким речевым инструкциям. Система "поднимает трубку", приветствует абонента и проигрывает речевое меню. Разные пункты меню содержат информацию, соответствующую разным видам продукции или услуг, предлагаемых организацией. В ответ абонент может сделать следующее: как в предыдущем случае, ввести известный индекс предложения и сразу получить необходимый документ, нажать клавишу соответствующую одному из пунктов меню, переключиться на оператора. При выборе

абонентом пункта меню система проигрывает речевое сообщение, которое содержит информацию о конкретных товарах или услугах и соответствующих им индексах. Абонент вводит заинтересовавший его индекс (или несколько индексов) и получает соответственно один или несколько документов.

ВЕЩАТЕЛЬНЫЕ системы позволяют организовывать древовидные речевые меню с разным числом ветвей и уровней, информация в которых может быть представлена как в факсимильном, так и в речевом виде. Техническое обслуживание **ВЕЩАТЕЛЬНЫХ** систем заключается не только в пополнении библиотеки документов, но и в строгом контроле за структурой меню. Организация таких меню требует аккуратного и продуманного подхода. Стоимость вещательных систем, по сравнению с системами простого ответа, выше минимум в полтора раза. Это объясняется применением дополнительного аппаратного обеспечения, для управления работой которого также требуются более сложные программы.

ИНТЕРАКТИВНЫЕ системы **ФАКС ПО ЗАПРОСУ**, так же как и **ВЕЩАТЕЛЬНЫЕ**, позволяют абонентам получать доступ к информации с помощью индексов и древовидных речевых меню. Разница между этими системами состоит в том, что **ИНТЕРАКТИВНЫЕ** системы **ФАКС ПО ЗАПРОСУ** могут предоставлять каждому абоненту не только однотипные документы, но и индивидуальные. Они создаются системой в процессе общения, в зависимости от набираемой абонентом на телефонном аппарате последовательности команд. Выглядит это следующим образом. Абонент звонит на номер системы **ФАКС ПО ЗАПРОСУ**, вводит свой персональный номер, код, и следуя речевым инструкциям, получает доступ к персональной информации. Другой вариант - абонент отвечает на вопросы системы, которая формирует сообщение в зависимости от ответов, включая/не включая те или иные данные в документ. Таким образом, все сообщения, высылаемые интерактивной системой **ФАКС ПО ЗАПРОСУ**, индивидуальны для каждого абонента.

ИНТЕРАКТИВНЫЕ системы **ФАКС ПО ЗАПРОСУ** предоставляют абонентам самый высокий уровень сервиса. Они сложнее в установке и эксплуатации, так как для создания высылаемых документов требуется оперативное обращение к различным базам данных а, следовательно, интеграция с ЛВС (локальной вычислительной сетью). Соответственно, цена таких систем является самой высокой среди систем **ФАКС ПО ЗАПРОСУ** и определяется как стоимостью аппаратного и программного обеспечения, так и стоимостью услуг специалистов по ее наладке и запуску.

Каждый из описанных видов систем **ФАКС ПО ЗАПРОСУ** может обеспечивать один или два режима доставки документов. Первый - режим **ОДНОГО ВЫЗОВА**. Абонент звонит со своего телефакса, выбирает нужные документы и инициирует их немедленное принятие нажатием на клавишу «Старт». Все расходы по оплате телефонного вызова оплачивает абонент. Такой способ доставки документов применяется на всех

системах ФАКС ПО ЗАПРОСУ. Для систем ПРОСТОГО ОТВЕТА он является единственно возможным.

Другой режим - ДВУХ ВЫЗОВОВ - предоставляет абонентам более высокий уровень сервиса. Абонент звонит на номер системы ФАКС ПО ЗАПРОСУ, выбирает необходимые документы и вводит номер телефакса, на который следует выслать затребованную информацию. Сразу после окончания разговора система установит соединение и вышлет затребованные документы. Соответственно время, затраченное на передачу документов, оплачивает владелец системы ФАКС ПО ЗАПРОСУ. Многие современные вещательные и интерактивные системы ФАКС ПО ЗАПРОСУ предоставляют такой способ доставки документов наряду с режимом ОДНОГО вызова.

Применение систем ФАКС ПО ЗАПРОСУ выручает сотрудников компаний в тех случаях, когда им приходится обрабатывать большое число запросов абонентов. Экономится время сотрудников, не требуется постоянного технического обслуживания, неизбежного при частом использовании обычных телефаксов. Все отправляемые документы имеют идеальное качество, так как большинство систем позволяет создавать библиотеку документов с помощью специального программного обеспечения, которое преобразует любой текстовый или графический документ в формат, используемый системой для передачи документальной информации. Время, требуемое на обработку запроса, сводится к минимуму, что на руку и организации (уменьшаются счета за использование телефонных линий) и абонентам (они быстро получают необходимую информацию).

Факс-рассылка

ФАКС - РАССЫЛКА является популярным приложением в организациях, которым по роду своей деятельности приходится рассылать большие объемы факсимильных сообщений большому числу адресатов. Системы ФАКС – РАССЫЛКИ позволяют одновременно рассылать по разным линиям различные по содержанию документы разным группам адресатов.

Для рассылки документов абонентам достаточно объединить их телефонные номера во временную группу и соотнести ее с рассылаемыми документами. Системы ФАКС - РАССЫЛКИ обычно взаимодействуют с любой широко распространенной базой данных, используя хранящуюся в ней информацию о номерах абонентов. После этого, если необходимо, указывается интервал времени, в течение которого следует производить рассылку. Все остальное система делает автоматически.

Системы ФАКС - РАССЫЛКИ не имеют каких-либо общепринятых стандартов. Их функции и возможности целиком зависят от производителя. Тем не менее, существует некий перечень требований, на наличие которых следует обращать внимание при выборе системы.

Только система, удовлетворяющая этим требованиям, может гарантировать действительно качественную доставку документов.

«Интеллектуальная» досылка. Нередко при передаче факсимильных сообщений происходят сбои, в результате которых документ передается не полностью. Для минимизации временных затрат следует приобретать систему, которая осуществляет досылку только переданных страниц, а не всего документа. Это важно при передаче документов большого размера.

Выбор маршрута наименьшей стоимости. Хорошие системы должны обладать возможностью выбора маршрута для передачи документа с наименьшими тарифными ставками. В данных целях могут использоваться разные префиксы - для выбора разных поставщиков телефонных услуг в зависимости от времени суток и телефонного номера вызываемой стороны.

Настройка «шапки». Система должна предоставлять возможность создания «шапки», данные в которую вносятся в процессе передачи. Это может быть имя, адрес, персональное сообщение и т.д.

Передача нескольких документов. Важной является возможность передачи нескольких документов в течение одного соединения - один длинный вызов дешевле, чем несколько коротких. Сегодня предлагается много систем, реализующих данную возможность.

Задержанная рассылка. Система должна обладать возможностью рассылки несрочных сообщений в ночное время для экономии затрат по оплате междугородных (международных) разговоров, с приостановкой процесса до следующего вечера в случае невозможности рассылки за одну ночь.

Обновление оригиналов документов в процессе работы. Подразумевается внесение изменений в документ или алгоритм работы системы без ее отключения. Некоторые системы реализуют эту возможность, другие требуют освобождения всех линий - это может стать существенной проблемой для загруженной системы ФАКС - РАССЫЛКИ.

В России, где пока наиболее развитой телекоммуникационной инфраструктурой является телефонная, а доступ к ресурсам глобальных компьютерных сетей ограничен, использование для передачи информации в документальном виде факсимильной связи является наиболее рациональным подходом. Проблема высоких тарифов на услуги российских поставщиков телефонной связи остро ставит вопрос о применении интеллектуальных факсимильных систем, способных не только освободить персонал от выполнения рутинной работы, но и повысить эффективность использования телефонных линий, а также сократить затраты, связанные с их эксплуатацией.

2. Примеры методов сжатия двухградационных изображений

При переходе к цифровому факсимильному сигналу увеличивается полоса частот сигнала, а, следовательно, требуется более широкий канал связи для передачи сигнала в цифровом виде по сравнению с передачей аналогового сигнала. Поэтому весьма остро ставится вопрос о сокращении полосы частот факсимильного сигнала. Методы сокращения полосы частот факсимильного сигнала основаны на учете избыточности изображения. Как полутонные, так и двухградационные черно-белые изображения имеют большую статистическую избыточность.

Имеется ряд статистических моделей двухградационного изображения. Так, при электрооптическом анализе двухградационного изображения после соответствующей обработки сигнала, образующегося на выходе фотоэлектрического преобразователя, получается сигнал в виде случайной последовательности единиц и нулей, причем единицы и нули группируются. Если принять сигнал, соответствующий «черному» за «1», а сигнал, соответствующий «белому» за «0», то получатся пачки (серии) единиц и нулей.

Статистика появления серий единиц и нулей обуславливается правилами формирования знаков (букв, цифр). При разбиении изображения на элементарные площадки (элементы изображения) число их достаточно велико. Если же элементами изображения считать серии черно-белых отрезков изображения вдоль строки, то объем передаваемых сообщений уменьшится. Модель двухградационного изображения в виде серий является основой для разработки методов эффективного кодирования (методов сжатия).

Еще одним критерием оценки различных методов сжатия является сложность построения их алгоритмов, которая в свою очередь определяет максимальные затраты времени на обработку сигнала процессором и объем вычислительных ресурсов. Благодаря постоянному совершенствованию цифровой элементной базы, относительная стоимость обработки сигналов постоянно снижается по сравнению с общей стоимостью аппаратуры факсимильной связи. Кроме того, тактовые частоты современных процессоров в среднем на пять порядков выше верхней частоты факсимильного сигнала, поэтому время передачи в целом в меньшей степени зависит от сложности алгоритмов программ и, соответственно, от скорости обработки цифрового сигнала изображения. В значительной степени скорость передачи факсимильной информации определяется скоростью развертки и временем прохождения трафика в сети. Таким образом, сложность алгоритмов сжатия-восстановления не является определяющим критерием при обосновании и выборе различных методов.

2.1. Кодирование длин серий (КДС)

При кодировании по этому методу двухуровневый факсимильный сигнал от каждой строки развертки разбивается на отдельные элементы - дискретизируется. Элементам белого приписывается значение логического нуля, а черного единицы. При этом учитывается, что, как правило, в среднем длины черных участков изображения много короче белых. Длины белых участков кодируются, например, шестиэлементными кодовыми комбинациями (до 63 элементов сигнала), а отрезки черного - трехэлементными (до семи элементов сигнала). Кодирование строки всегда начинается с белого; для определения начала строки передается служебная комбинация из девяти нулей. При кодировании белого различают отрезки; меньше 63 элементов изображения; больше, но не кратные 63 элементам; равные или кратные 63 элементам. В первом случае отрезки кодируются шестиэлементными кодовыми комбинациями. Во втором случае длина отрезка кодируется несколькими шестиэлементными комбинациями, причем все они, кроме последней, состоят из одних единиц. В третьем случае длина отрезка кодируется одной или несколькими шестиэлементными комбинациями, состоящими из одних единиц, и для определения окончания белого передается служебная кодовая комбинация из шести нулей.

При кодировании черного различают отрезки: меньше 7 элементов изображения; больше, но не кратные 7 элементам; равные или кратные 7 элементам. Они кодируются так же, как и отрезки белого, но при этом используются не шести, а трехэлементные кодовые комбинации.

Для оценки методов сжатия используется коэффициент сжатия, который определяется по формуле:

$$K = N / N_{сж},$$

где N - число элементов на изображении, $N_{сж}$ - число двоичных разрядов для представления сжатого изображения (с учетом служебных кодовых комбинаций).

Метод КДС обладает низкой помехоустойчивостью. Неверно принятый код длины серии приводит к так называемому «треку ошибок», поскольку декодер воспроизведет серию не той длины. Таким образом, даже одиночная ошибка за счет помехи в канале приводит к «разрушению» строки (к сдвигу всех последующих серий до конца строки).

Блок схема алгоритма кодирования двухградационного сигнала по методу КДС показана на рис. 2.1.

Пример использования метода КДС

Пусть двухградационное тестовое изображение состоит из двух строк, структура которых показана в таблице, приведенной ниже.

Номер строки	Серия						
	белая	черная	белая	черная	белая	черная	белая
1	129	11	128	12	127	13	0
2	0	14	126	15	125	16	124

Пусть $m_b = 7$ и $m_c = 4$ – длины равномерных кодов, используемых для кодирования длин серий белого и черного соответственно.

При кодировании белого различают отрезки:

1) меньше 127. В этом случае отрезки кодируются одним 7-элементным кодом;

2) больше, но не кратные 127. В этом случае длина отрезка кодируется несколькими 7-элементными кодами, причем первые коды состоят из единиц, а последняя кодовая комбинация из остатка;

3) кратные 127. В этом случае длина отрезка кодируется одним или несколькими 7-элементными кодами, состоящими из всех единиц, и для определения окончания белой серии передается кодовая комбинация из семи нулей.

При кодировании черной серии происходит аналогичное разделение на отрезки, но с учетом, что черные серии кодируем 4-элементными кодами:

При кодировании черного различают отрезки:

1) меньше 15;

2) больше, но не кратные 15;

3) кратные 15.

Для определения начала строки передается служебная комбинация из 11 нулей.

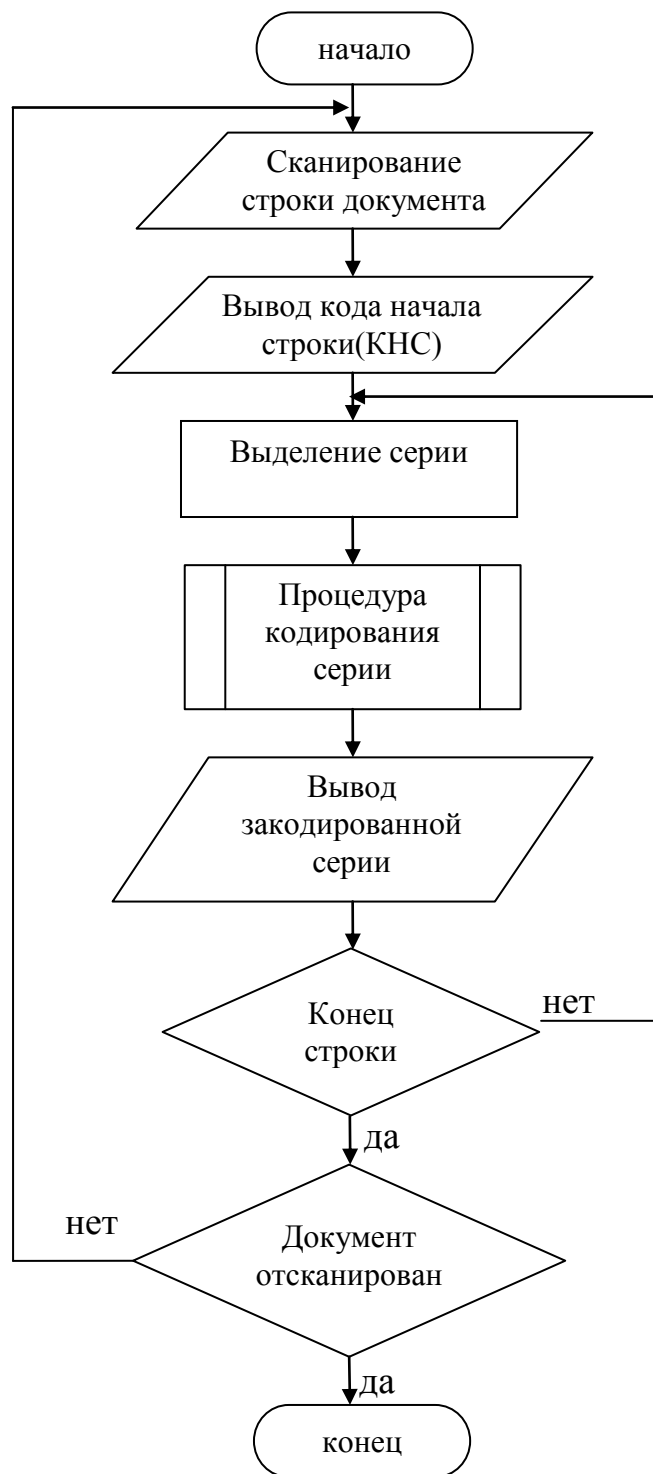


Рис. 2.1. Блок-схема алгоритма кодирования двухградационного сигнала по методу КДС

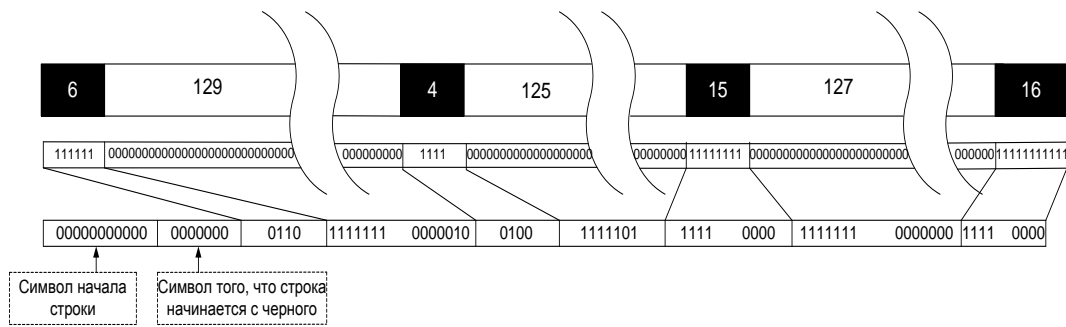


Рис. 2.2. Кодирование двухградационного факсимильного сигнала по методу КДС

Алгоритм сжатия (кодирования) по методу КДС заключается в проведении следующих операций:

1) каждая строка начинается с признака начала новой строки. Этому признаку должна соответствовать кодовая комбинация, которая никогда не встречается и не страдает избыточностью. Для рассматриваемого примера кодовая комбинация для обозначения начала каждой сжатой строки - 000000000000;

2) первая кодовая комбинация в строке обязательно должна соответствовать закодированной белой серии, если строка начинается с чёрной серии, то первое кодовое слово соответствует нулевой белой серии (0000000);

3) далее поочередно белые (или черные) серии заменяются на кодовые комбинации: 7-разрядные (или 4-разрядные) как было описано выше.

Пример процедуры сжатия и превращения длинных серий в кодовые комбинации приведен на рис. 2.2. На входе кодера мы имеем строку изображения, состоящую из нулей (белое) и единиц (черное), на выходе выдается закодированная строка, которая должна передаваться в канал связи.

Алгоритм восстановления (декодирования) строки, которая была закодирована по методу КДС, основывается на следующих принципах:

1) определение начала строки по кодовой комбинации начала строки (000000000000);

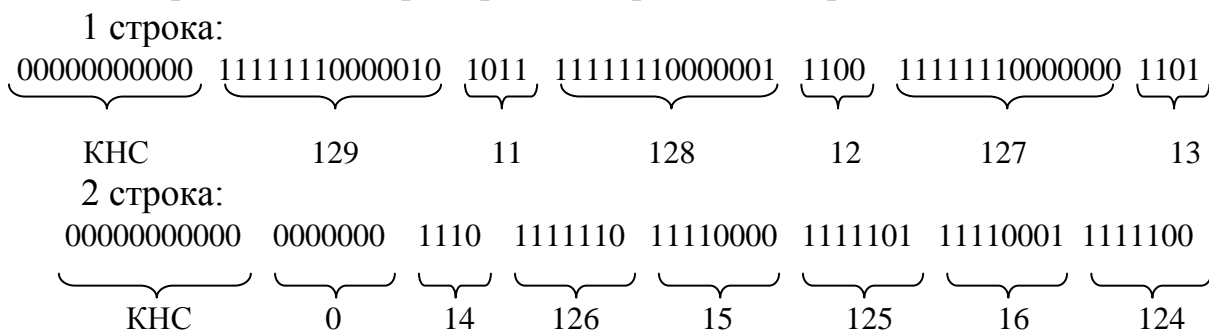
2) проверка на наличие «нулевой белой серии» - кодовой комбинации (0000000), если она есть, то первой декодируется сжатая черная серия, иначе декодер первую сжатую серию восстанавливает как белую;

3) последовательное считывание кодовых комбинаций согласно установленной размерности (по 7 бит - белое, по 4 бита - чёрное) и заполнение регистра, длина которого равна строке изображения. Суммирование значений кодовых комбинаций, если необходимо (закодирована белая серия, длина которой больше 127, или закодирована черная серия, длина которой больше 15);

4) изменение цвета происходит в трёх случаях:

- при обнаружении «нулевой белой серии» в начале строки,
- распознавании числа меньше 127 (для белого) и меньше 15 (для черного),
- при распознавании кодовой комбинации «0000000» или «0000», что свидетельствует о том, что было закодировано число кратное 127 (для белой серии) или кратное 15 (для черной серии).

Для приведенного примера закодированные строки имеют вид:



2.2. Адресно-позиционное кодирование (АПК)

При кодировании штриховых изображений возможны новые значения двух типов: переход от белого к чёрному и переход от чёрного к белому. Координата элемента с новым значением яркости (КНЗ) отсчитывается относительно начала строки изображения. Для безошибочного восстановления изображения по сжатым данным необходимо в начале кодограммы строки указать признак (код) её начала и значение яркости первого элемента строки:

[код начала строки] [код первого элемента] [КНЗ] [КНЗ] [КНЗ].

При равномерном кодировании длина кода будет определяться из условия, что элемент с новым значением яркости находится в конце строки. Таким образом, длина двоичного кода $k = \log_2 N$, где N - число элементов изображения на строке.

Данный метод эффективен при передаче метеокарт, графиков, чертежей, укрупненных контурных рисунков и т. п. Избыточность сокращается примерно в 5-6 раз.

Метод АПК менее чувствителен к ошибкам, чем КДС, т.е. обладает большей помехоустойчивостью. Помехоустойчивость АПК заключается в том, что ошибка координаты нового значения яркости приводит к искажению небольшой группы элементов строки до следующего нового значения яркости.

Повышенная помехозащищенность метода АПК позволяет вести передачу в условиях помех, несколько худших, чем это допускают многие другие методы сжатия, например КДС.

Блок-схема алгоритма кодирования двухградационного сигнала по методу АПК показана на рис. 2.3.

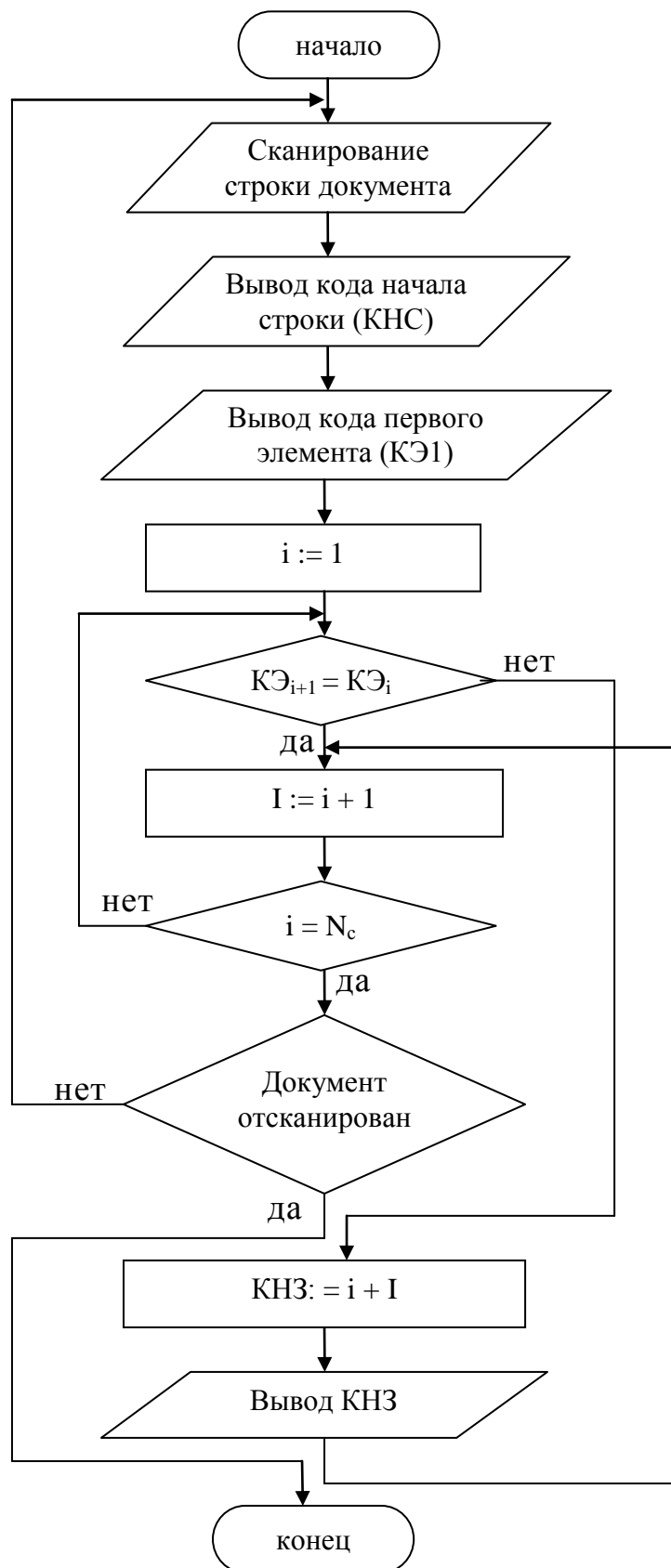


Рис. 2.3. Блок-схема алгоритма кодирования двухградационного сигнала по методу АПК

Пример использования метода АПК

Двухградационное тестовое изображение, представленное в цифровом виде, состоит из двух строк и имеет следующий вид:

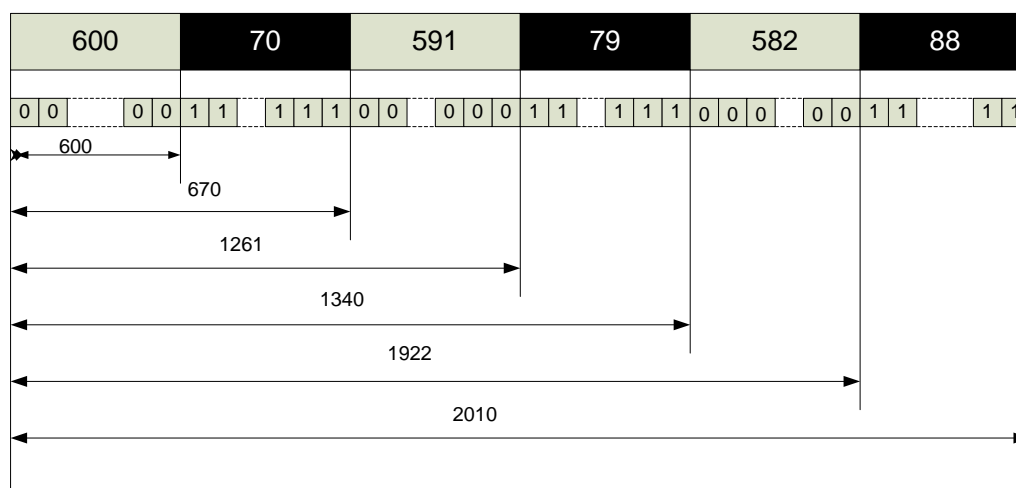
Номер строки	Серия						
	белая	черная	белая	черная	белая	черная	белая
1	600	70	591	79	582	88	0
2	0	72	598	81	589	90	580

N – число элементов изображения на строке;

m – длина двоичного кода.

$$N = 2010; m = \lceil \log_2 N \rceil = \lceil \log_2 2010 \rceil = 11.$$

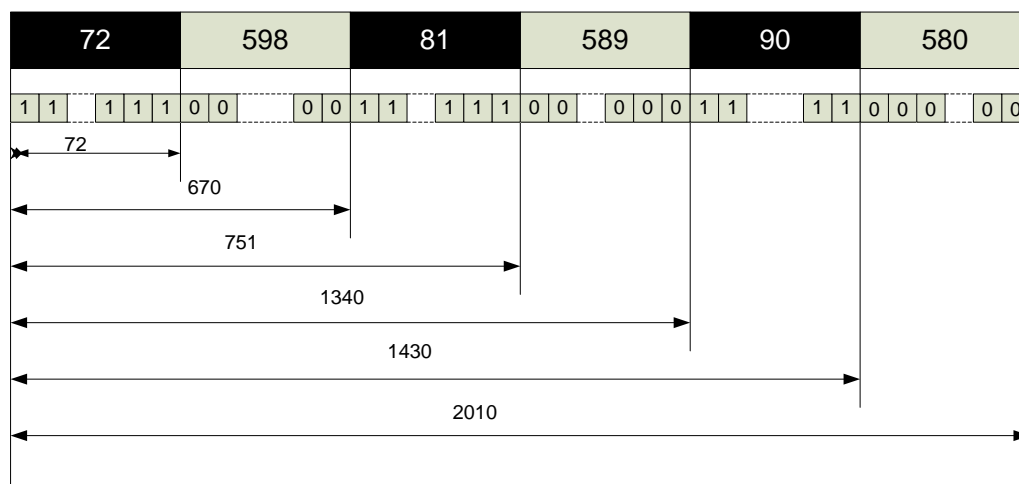
Исходный факсимильный сигнал для 1-й тестовой строки:



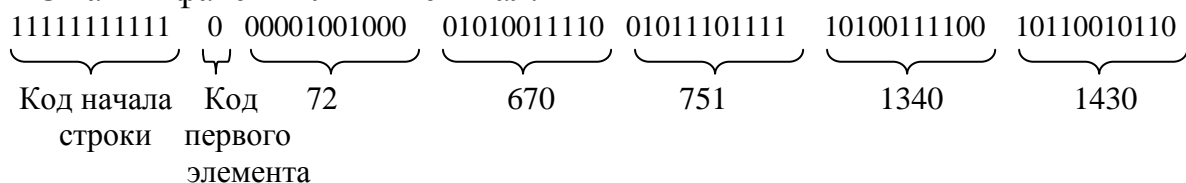
Сжатый факсимильный сигнал:

1111111111 0 01001011000 01010011110 10011101101 10100111100 11110000010
 Код начала строки Код первого элемента 600 670 1261 1340 1922

Исходный факсимильный сигнал для 2-й тестовой строки:



Сжатый факсимильный сигнал:



2.3. Алгоритм оптимального неравномерного кодирования (метод Хаффмана)

Алгоритм оптимального неравномерного кодирования учитывает статистику исходного двоичного изображения: наиболее вероятные длины серий кодируются наиболее короткими кодовыми комбинациями, а наименее вероятные - наиболее длинными. Примером такого кода является одномерный код Хаффмана. При использовании этого кода серии черных и белых элементов кодируются в соответствии с кодовой таблицей. Размер кодовой комбинации в таблице обратно пропорционален вероятности появления соответствующей длины серии.

Характерная особенность этих кодовых комбинаций заключается в том, что они не содержат повторяющихся последовательностей единичных и нулевых элементов. Этим обеспечивается однозначное декодирование и, кроме того, не требуется вводить дополнительные элементы для обозначения начала или конца кодовых комбинаций. Иначе говоря, в любой последовательности кодовых комбинаций всегда однозначно определяются начало и конец каждой из них, т.е. код является самосинхронизирующимся.

Код Хаффмана имеет ряд модификаций. Модифицированные коды Хаффмана (МКФ) широко применяются в факсимильной аппаратуре третьей группы.

Также используются способы двумерного кодирования. Для них характерно, что позиция каждого меняющегося элемента кодируемой

строки кодируется с учетом позиции соответствующего меняющегося элемента опорной строки, расположенной непосредственно над кодируемой строкой. При этом используются корреляционные свойства изображения соседних строк (их «похожесть»). После кодирования опорной строки следующая строка изображения вычитается из опорной, а затем кодируется разностный сигнал. Поскольку при вычитании фактически осуществляется декорреляция изображения (разрушаются статистические связи), то устраняется избыточность. Энтропия преобразованного сигнала возрастает, а, следовательно, возрастает эффективность кодирования.

Двумерное кодирование особенно чувствительно к ошибкам, поскольку ошибки приводят не только к «трекам ошибок», но и распространяются на соседние строки. Чтобы предотвратить вертикальное распространение искажений от строки к строке, двумерным способом кодируется не более двух - четырех соседних строк, после чего строка кодируется одномерным способом.

Начало каждой сжатой строки документа обозначается кодовой комбинацией EOL.

Конец передачи документа обозначается путем передачи шести кодовых комбинаций EOL.

На рис. 2.4 показаны несколько строк развертки данных, начиная с начала переданной страницы. На рис. 2.5 показана последняя кодируемая строка развертки страницы.

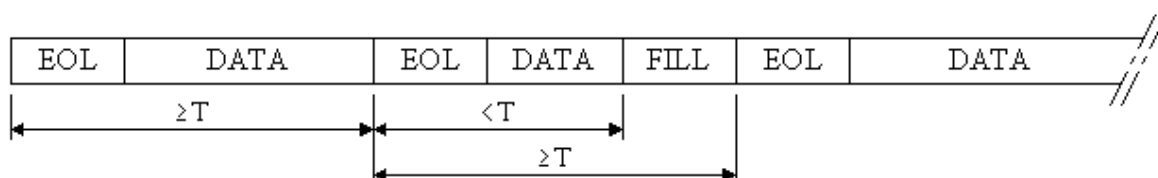


Рис. 2.4. Первые n кодируемых строк документа

T - минимальное время передачи полной кодируемой строки развертки.

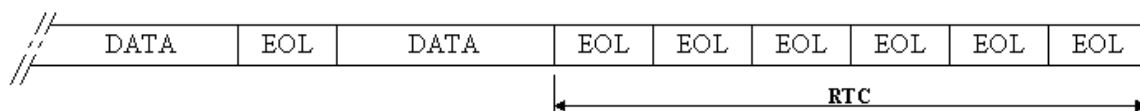


Рис. 2.5. Последняя кодируемая строка документа

Код Хаффмана может быть двухпроходным (двухмерным) и однопроходным (одномерным). Первый строится по результатам подсчета частот (вероятностей) появления различных символов в данном сообщении. Второй использует готовую таблицу кодирования, построенную на основе вероятностей символов в сообщениях похожего типа. Например, кодирование текста на русском языке в первом случае включает его предварительный анализ, подсчет вероятностей символов,

построение дерева кода и таблицы кодирования индивидуально для данного сообщения. Во втором случае будет работать готовая таблица, построенная по результатам анализа множества русскоязычных текстов. Двухпроходный код более полно использует возможности сжатия. Однако при этом вместе с сообщением нужно передавать и кодовую таблицу. Однопроходный код не оптимален, однако прост в использовании, поэтому на практике обычно применяют именно его.

Код Хаффмана при декодировании наиболее устойчив к возникающим ошибкам в кодовых последовательностях серий. С возникновением ошибки данная строка может быть пропущена или заменена на предыдущую, что не влияет на дальнейшее декодирование и информационное значение всего сообщения.

В таблицах, приведенных ниже, заданы два вида кодов:

- Коды завершения серий — заданы с 0 до 63 с шагом 1 (табл. 2.1);
- Начальные (дополнительные) коды — заданы с 64 до 2560 с шагом 64, они используются, если длина серии превышает 63 (табл. 2.2).

Таблицы 2.1 и 2.2 были составлены на основе анализа многочисленных факсимильных изображений на предмет определения вероятностей, с которыми следуют серии разной длины и разных цветов.

Серии длиной от 0 до 63 пикселей кодируются с помощью определенных завершающих кодовых слов. Для черных и белых серий существуют разные списки кодовых слов. Серии длиной от 64 до 2623 (2560+63) пикселей, кодируются сначала начальным кодовым словом, которое описывает ближайшую по длине, но не более длинную серию, чем требуемая. За ним следует завершающее кодовое слово, представляющее разность между требуемой серией и серией, описанной начальным кодовым словом.

Таблица кодов завершения

Таблица 2.1

Длина серии	Код белой подстроки	Код черной подстроки	Длина серии	Код белой подстроки	Код черной подстроки
0	00110101	0000110111	32	00011011	000001101010
1	00111	010	33	00010010	000001101011
2	0111	11	34	00010011	000011010010
3	1000	10	35	00010100	000011010011
4	1011	011	36	00010101	000011010100
5	1100	0011	37	00010110	000011010101
6	1110	0010	38	00010111	000011010110
7	1111	00011	39	00101000	000011010111
8	10011	000101	40	00101001	000001101100
9	10100	000100	41	00101010	000001101101
10	00111	0000100	42	00101011	000011011010
11	01000	0000101	43	00101100	000011011011
12	001000	0000111	44	00101101	000001010100
13	000011	00000100	45	00000100	000001010101
14	110100	00000111	46	00000101	000001010110
15	110101	000011000	47	00001010	000001010111
16	101010	0000010111	48	00001011	000001100100
17	101011	0000011000	49	01010010	000001100101
18	0100111	0000001000	50	01010011	000001010010
19	0001100	00001100111	51	01010100	000001010011
20	0001000	00001101000	52	01010101	000000100100
21	0010111	00001101100	53	00100100	000000110111
22	0000011	00000110111	54	00100101	000000111000
23	0000100	00000101000	55	01011000	000000100111
24	0101000	00000010111	56	01011001	000000101000
25	0101011	00000011000	57	01011010	000001011000
26	0010011	000011001010	58	01011011	000001011001
27	0100100	000011001011	59	01001010	000000101011
28	0011000	000011001100	60	01001011	000000101100
29	00000010	000011001101	61	00110010	000001011010
30	00000011	000001101000	62	00110011	000001100110
31	00011010	000001101001	63	00110100	000001100111

Таблица начальных кодов

Таблица 2.2

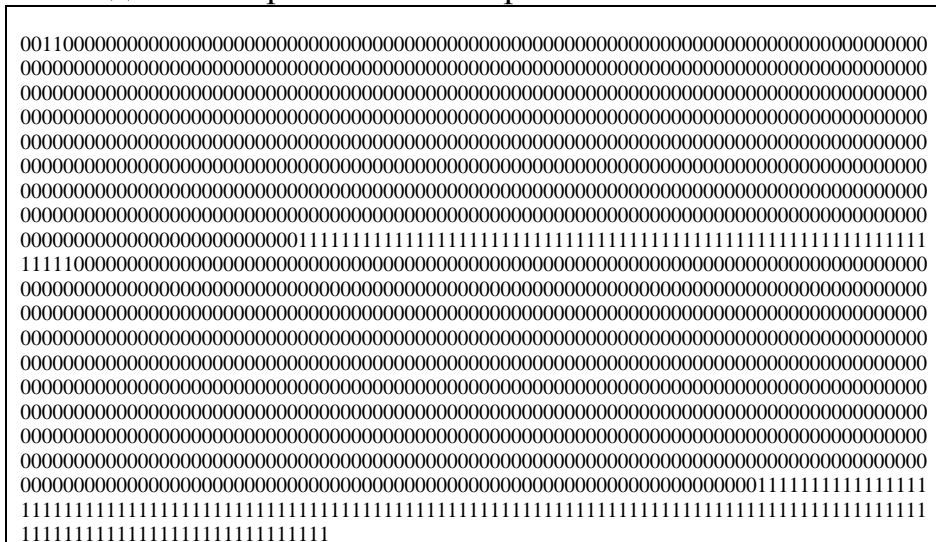
Длина серии	Код белой подстроки	Код черной подстроки	Длина серии	Код белой подстроки	Код черной подстроки
64	11011	0000001111	1344	011011010	0000001010011
128	10010	000011001000	1408	011011011	0000001010100
192	01011	000011001001	1472	010011000	0000001010101
256	0110111	000001011011	1536	010011001	0000001011010
320	00110110	000000110011	1600	010011010	0000001011011
384	00110111	000000110100	1664	011000	0000001100100
448	01100100	000000110101	1728	010011011	0000001100101
512	01100101	0000001101100	1792	00000001000	совп. с белой
576	01101000	0000001101101	1856	00000001100	— // —
640	01100111	0000001001010	1920	00000001101	— // —
704	011001100	0000001001011	1984	000000010010	— // —
768	011001101	0000001001100	2048	000000010011	— // —
832	011010010	0000001001101	2112	000000010100	— // —
896	011010011	0000001110010	2176	000000010101	— // —
960	011010100	0000001110011	2240	000000010110	— // —
1024	011010101	0000001110100	2304	000000010111	— // —
1088	011010110	0000001110101	2368	000000011100	— // —
1152	011010111	0000001110110	2432	000000011101	— // —
1216	011011000	0000001110111	2496	000000011110	— // —
1280	011011001	0000001010010	2560	000000011111	— // —

Пример использования кода Хаффмана

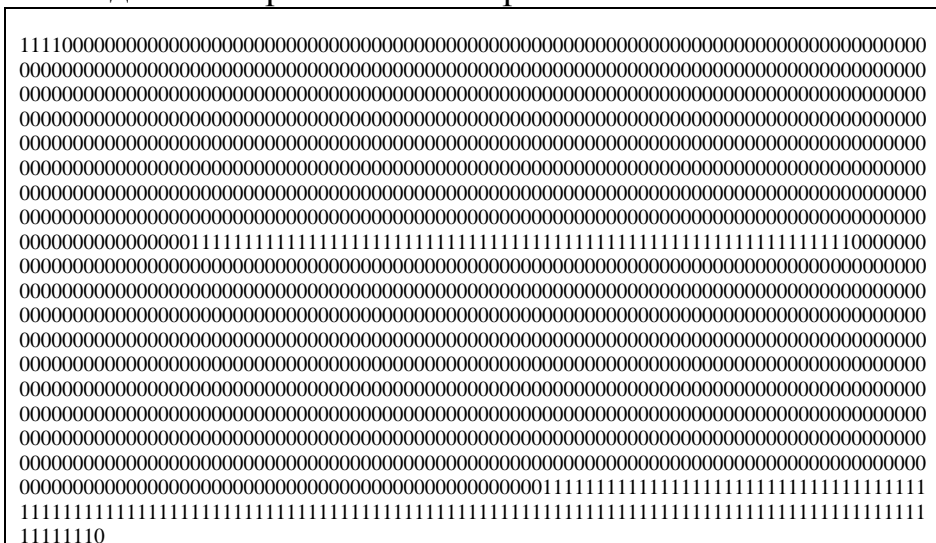
Тестовое двухградационное изображение, представленное в цифровом виде:

Номер строки	Серия						
	белая	черная	белая	черная	белая	черная	белая
1	2	2	702	64	829	129	0
2	0	5	699	63	830	130	1

Исходное изображение 1-й строки:



Исходное изображение 2-й строки:



Сжатая последовательность строки 1:

011111011001110011001100000111100011011101100110100110010000011001000010

Сжатая последовательность строки 2:

0011010100110110011101001010000011001110110011010011001100001100100011000111

Коэффициент сжатия $\approx 19,53$.

2.4. Двумерный код Хаффмана

Двумерная схема кодирования имеет менее избыточный код, чем одномерная схема. В сравнении с кодом Хаффмана код READ более уязвим при возникновении ошибок, так как последующие строки связаны с предыдущими. Возникновение ошибок в опорной строке влечет за собой сбой в последующих $K-1$ двумерных строках, что приводит к нечитабельности K строк документа. Данный сбой восстанавливается только после безошибочной опорной строки.

Двумерная схема кодирования является расширением одномерной схемы кодирования. Двумерная схема или код READ - это также метод построчного кодирования, при котором положение каждого “изменяющегося” элемента изображения в текущей или кодируемой строке кодируется по отношению к положению соответствующего отсчетного элемента, расположенного на кодируемой строке, либо на отсчетной строке, которая располагается непосредственно над кодируемой строкой. После кодирования кодируемая строка становится отсчетной строкой по отношению к следующей кодируемой строке.

Для того, чтобы ограничить область поражения в случае наличия ошибок передачи в двумерной схеме после каждой строки, закодированной по одномерному методу, по крайней мере, $K-1$ последовательных строк должны кодироваться по двумерному методу. Строка, закодированная по

одномерной схеме, может передаваться чаще, чем через каждые K линий. После того, как передается строка, закодированная по одномерному методу, начинается следующая серия строк $K-1$, кодируемых по двумерной схеме. Значение K устанавливается при стандартной вертикальной разрешающей способности, равной 2 или 4 при более высокой разрешающей способности.

Используя уникальность кода Хаффмана и кодовых слов при вертикальном кодировании, параметры развертки изображения, а также размер, контрольную сумму и номер сбойного блока, можно разработать алгоритм восстановления искаженных при передаче данных, не прибегая к повторному перезапросу.

2.5. Другие методы кодирования

Две схемы сжатия, кодирование Хаффмана и LZW-кодирование (по начальным буквам фамилий Лемпел, Зив и Велч), формируют основу для многих других систем сжатия.

И кодирование Хаффмана, и LZW-кодирование являются алгоритмами сжатия без потерь. Они подходят для сжатия любого типа данных, потому что развернутое представление идентично исходным входным данным системы сжатия.

Кодирование Хаффмана уменьшает количество битов, используемых для представления часто встречающихся символов, и увеличивает количество бит, используемых для редких символов. Метод LZW кодирует строки символов, используя входной поток для построения расширенного алфавита, основанного на строках, которые он обрабатывает. Оба эти подхода работают путем уменьшения лишней информации во входных данных.

Код LZW

В отличие от схемы Хаффмана, схеме LZW не требуется перед кодированием создавать таблицу кодов. Начиная с более простой таблицы кодов, алгоритм формирует более эффективную таблицу по мере своего продвижения. При реализации LZW обычно не используются более короткие коды для часто встречающихся элементов. В обнаруженных последовательностях схема LZW использует избыточность.

Коэффициент сжатия, обеспечиваемый, при кодировании методом LZW находится между 1:1 и 3:1.

Алгоритмы сжатия этой группы наиболее эффективны для текстовых данных больших объемов и малоэффективны для файлов маленьких размеров (за счет необходимости сохранения словаря).

Основным достоинством метода LZW является высокая скорость декодирования.

Арифметическое кодирование

Арифметическое сжатие (АС), подобно кодированию по алгоритму Хаффмана, использует более короткие коды для часто появляющихся участков и более длинные коды для редко появляющихся. Это более эффективная схема, хотя она (подобно LZW) сжимает последовательности величин, не сами величины. Также эта схема более подходит для работы с файлами, в которых большинство данных состоят из повторений одной и той же последовательности. Фактически, арифметическое кодирование изображений близко подходит к теоретическим пределам для сжатия.

Для АС существует несколько алгоритмов. АС отображает каждую отличительную последовательность величин пикселей диапазоном цифр, в пределах воображаемого числа между 0 и 1. Эта область затем представляется как двоичная дробь переменной точности (количество бит). Последовательности, встречающиеся, реже требуют более высокой точности (больше бит).

Отличительной особенностью алгоритма от метода Хаффмана является его недостаток - значительная трудоемкость и достоинство - эффективное уменьшение размера файла в зависимости от исходного изображения и точности используемой статистической модели (максимальное сжатие 100:1).

Наличие различных методов кодирования для сжатия факсимильных изображений требует детального их анализа. С точки зрения оценки эффективности для уменьшения объема передаваемого факсимильного сообщения.

3. Анализ эффективности метода КДС при передаче текстовых документов

3.1. Пример оценки избыточности для двухградационного черно-белого изображения

Двухградационное черно-белое изображение обладает большой избыточностью. Так для машинописного текста справедливо следующее рассуждение. Текстовый документ содержит много белых полей, не несущих полезной информации, и около 700-1000 знаков. Пусть параметры разложения: 3,85 строк/мм по вертикали и 8 пикселей/мм вдоль строки (по горизонтали). Количество элементов разложения на бланке документа формата А4: $210 \cdot 297 \cdot 8 \cdot 3,85 = 1,92 \cdot 10^6$ элементов.

Изображение буквы размером $3 \cdot 2 \text{ мм}^2$ содержит $8 \cdot 2 \cdot 3,85 \cdot 3 = 185$ элементов.

Поэтому $1,92 \cdot 10^6 - (1000 \cdot 185) = 1,7 \cdot 10^6$ элементов являются «лишними».

Рассмотрим пример для конкретного документа.

Допустим, требуется передать с помощью факсимильного аппарата документ “Расписание занятий” (рис. 3.1).

Если на передающем ФА установлен режим Fine (плотность сканирования по вертикали 7,7 линий на мм), то количество элементов разложения на бланке документа (рис. 3.1)

$$N=210 \cdot 297 \cdot 8 \cdot 7,7=3,84 \cdot 10^6.$$

На документе содержится примерно 3500 знаков размером $1 \cdot 1 \text{ мм}^2$ и 20 горизонтальных линий длиной 180 мм. Таким образом, количество элементов разложения для смысловой («полезной») информации составляет приблизительно

$$N_{\text{и}} = (8 \cdot 1 \cdot 7,7 \cdot 1) \cdot 3500 + (8 \cdot 180) \cdot 20 = 0,24 \cdot 10^6.$$

Таким образом, $3,6 \cdot 10^6$ элементов соответствуют белой градации и несут информационную нагрузку только об оформлении документа.

Если ФА поддерживает протокол V.17 со скоростью $V = 9600$ бит/с, то для передачи документа потребуется

$$T_{\text{пер}} = \frac{N}{V} = \frac{3,84 \cdot 10^6}{9600} = 400 \text{ с} \approx 7 \text{ мин.}$$

Рекомендация Т.4 для ФА G3 рекомендует время передачи для документа формата А4 приблизительно 1 мин. Следовательно, обязательно необходимо использовать после сканирования документа процедуру сжатия. Коэффициент сжатия для данного документа должен быть не менее семи.

Оценим возможность использования для сжатия данного документа (рис. 3.1) метода КДС.

3.2. Расчет оптимальных длин кодовых комбинаций для метода КДС

Для оценки эффективности методов сжатия, а также метода КДС используется коэффициент сжатия, который определяется по формуле:

$$K=N/N_{\text{сж}},$$

где N – число элементов изображения,

$N_{\text{сж}}$ – число двоичных разрядов для представления сжатого изображения (с учетом служебных кодовых комбинаций).

Очевидно, что, чем меньше $N_{\text{сж}}$, тем больше коэффициент сжатия.

Рассмотрим, как определяется $N_{\text{сж}}$ для метода КДС.

Для метода КДС должны быть выбраны $m_{\text{б}}$ и $m_{\text{ч}}$ - длины равномерных кодов, используемые для кодирования длин серий белого и черного соответственно. При этом учитывается, что, как правило, в среднем длины черных участков изображения много короче белых. Поэтому $m_{\text{б}} > m_{\text{ч}}$.

«УТВЕРЖДАЮ»
 Декан ФППК
 профессор **М.А. СИБЕРС**
 " " апреля 2005 года

РАСПИСАНИЕ ЗАНЯТИЙ
 слушателей факультета повышения квалификации и переподготовки
 инженерно-педагогических кадров
 по теме: "Сеть Internet и предоставляемые услуги"
 с 12 апреля по 22 апреля 2005 года

Дата	Время	Вид зан.	Ауд.	Тематика	Преподаватель
12 апреля вторник	09 ⁰⁰ -16 ¹⁵	ЛК	ВО ФППК	Тенденции развития вычислительных сетей. Общие сведения об Internet. Структура сети Internet. Стек протоколов TCP/IP. Сравнительная оценка сетевых архитектур OSI и TCP/IP.	Когновицкий О.С.
13 апреля среда	09 ⁰⁰ -10 ³⁵	ПР	263	Принципы построения вычислительных сетей. Практическая работа в LAN.	Доронин Е.М.
	10 ⁴⁵ -16 ¹⁵	ЛК	267	Протокол межсетевое взаимодействие IP. Основные функции, характеристики и механизмы протокола IP. Структура IP-пакета. Фрагментация IP-пакетов в сети Internet. Новая версия протокола межсетевых взаимодействий (IPv6).	Когновицкий О.С.
14 апреля четверг	09 ⁰⁰ -10 ³⁵	ЛК	267	Адресация в IP-сетях. Представление и структура сетевого IP-адреса (версия IPv4).	Доронин Е.М.
	10 ⁴⁵ -12 ²⁰	ЛК	267	Маски сетей. Подсети: адреса и маски. Доменная система имен – DNS. Назначение протокола ARP.	Доронин Е.М.
	12 ³⁰ -14 ⁴⁵	ПР	263	Использование протокола ARP для определения MAC-адреса по IP-адресу.	Доронин Е.М.
	14 ³⁵ -16 ²⁰	ЛК	267	Транспортный уровень архитектуры TCP/IP. Протоколы UDP и TCP.	Когновицкий О.С.
15 апреля пятница	09 ⁰⁰ -16 ²⁰	ПР	263	Практическая работа пользователей в сети Internet: мониторинговые исследования, электронная почта, поиск информации в Internet и др.	Доронин Е.М.
18 апреля понедельник	09 ⁰⁰ -10 ³⁵	ЛК	267	Принципы и алгоритмы маршрутизации в сети Internet. Внутренние и внешние протоколы маршрутизации. Функциональная модель маршрутизатора.	Когновицкий О.С.
	10 ⁴⁵ -14 ²⁵	ПР	263	Анализ работы протоколов в сети Internet.	Доронин Е.М.
	14 ³⁵ -16 ²⁰	ЛК	267	Системы абонентского доступа к сети Internet.	Когновицкий О.С.
19 апреля вторник	09 ⁰⁰ -16 ²⁰	ЛК	267	Современные сетевые технологии Frame Relay, ATM. Цифровые сети интегрального обслуживания, их роль в построении современных компьютерных сетей.	Харитонов В.Х.
20 апреля среда	09 ⁰⁰ -16 ²⁰			Технические и программные средства Internet-провайдера, предоставляемые услуги (на примере компании).	Когновицкий О.С.
21 апреля четверг	09 ⁰⁰ -16 ²⁰			Опыт создания и эксплуатация узла Internet компании, предоставляемые услуги.	Когновицкий О.С.
22 апреля пятница	09 ⁰⁰ -10 ³⁵	ЛК	267	Оценка качества предоставляемых услуг.	Когновицкий О.С.
	10 ⁴⁵ -14 ²⁵	ПР	263	Практическая работа пользователей в сети Internet: мультимедиа и видеоконференцсвязь в сети Internet, основы web-хостинга, создание web-страниц.	Доронин Е.М.
	14 ³⁵ -16 ²⁰		267	ЗАЧЕТ	Когновицкий О.С.

Примечание:
 ауд. 263 и 267 находятся по адресу – наб. р. Мойки, д. 61, СПбГУТ (станция метро «Невский проспект», выход на канал Грибоедова).

Заведующий кафедрой ОПДС
 профессор

Когновицкий О.С.

Рис. 3.1. Пример текстового документа «Расписание занятий»

Можно записать, что $N_{сж} = N_{сж}^0 + N_{сж}^1$, где $N_{сж}^0$ - количество двоичных разрядов, которые требуются для всех кодовых комбинаций, кодирующих белые серии на изображении; $N_{сж}^1$ - количество двоичных

разрядов, которые требуются для всех кодовых комбинаций, кодирующих черные серии на изображении.

Как было показано в п. 2.1, максимальная длина белой серии, которую можно закодировать одной кодовой комбинацией длиной m_b , равна $2^{m_b} - 1$, а максимальная длина черной серии, которую можно закодировать одной кодовой комбинацией длиной m_c , равна $2^{m_c} - 1$. Следовательно, количество кодовых комбинаций для представления всех белых и черных серий на изображении документа зависит от его структуры. Для каждого документа могут быть определены оптимальные m_b и m_c , обеспечивающие наименьшие значения $N_{сж}^b$ и $N_{сж}^c$ соответственно.

Сформулируем задачу по расчету оптимальных m_b и m_c . В общем случае после сканирования документа можно получить следующие таблицы (табл. 3.1 и табл. 3.2).

Таблица 3.1

$N_{п/п}$	Длина белой серии	Кол-во серий
1	1	n_1
2	2	n_2
3	3	n_3
⋮	⋮	⋮
i	li	n_i
⋮	⋮	⋮
C_b	L	n_L

Таблица 3.2

$N_{п/п}$	Длина черной серии	Кол-во серий
1	1	n_1
2	2	n_2
3	3	n_3
⋮	⋮	⋮
i	li	n_i
⋮	⋮	⋮
C_c	M	n_M

Здесь L – максимальная длина белой серии для конкретного отсканированного документа, а M – максимальная длина черной серии. Тогда для $N_{сж}^b$ может быть составлена следующая зависимость

$$N_{сж}^b = \sum_{i=1}^{C_b} \left[\left(\left[\frac{li}{2^{m_b} - 1} \right] + 1 \right) \cdot m_b \cdot n_i \right] + m_b \cdot b, \quad (3.1)$$

где C_b – количество различных длин белых серий,

C_c – количество различных длин черных серий.

Здесь выражение в круглых скобках (...) есть количество кодовых комбинаций для представления белой серии длиной li . Обратные квадратные скобки означают целую часть от деления li на $2^{m_b} - 1$.

Выражение в квадратных скобках [...] определяет количество двоичных разрядов для представления всех серий длиной li . Сумма от $i=1$

до S определяет количество двоичных разрядов для представления всех белых серий на изображении документа.

Выражение $m_b \cdot b$ определяет количество двоичных разрядов для представления части кодовых комбинаций начала строки (b – количество отсканированных линий на документе, например, для режима Fine $b = 8 \cdot 297 = 2376$).

Аналогичная зависимость может быть составлена и для расчета $N_{сж}^4$.

Последовательно вычисляя $N_{сж}^6$ для различных $m_b = 1, 2, 3, \dots$, можно построить функцию $N_{сж}^6 = f(m_b)$.

Также можно построить функцию $N_{сж}^4 = f(m_c)$.

Минимальное значение этих функций и будет определять наилучшие значения m_b^* и m_c^* для конкретного документа. Эти значения m_b^* и m_c^* позволяют получить максимальный коэффициент сжатия для конкретного документа, обработанного методом КДС.

3.3. Программная реализация статистической обработки и сжатия текстовых документов методом КДС

Была составлена программа для статистической обработки отсканированных текстовых документов, расчета оптимальных длин m_b^* и m_c^* равномерных кодов для кодирования длин белых и черных серий методом КДС и вычисления коэффициента сжатия.

Отсканированный документ для программной обработки должен быть представлен в формате .bmp.

Порядок работы с программой.

1. Запуск программы. На экран выводится окно «Преобразование двухцветного BMP в текст» (рис. 3.2).
2. Просмотр отсканированного файла в формате .bmp. Для просмотра нажать кнопку «Открыть файл».
3. Преобразование файла в формат .txt. Для включения процесса преобразования нажать кнопку «Преобразовать».

Процесс преобразования занимает несколько минут (чем сложнее документ по структуре, тем дольше идет процесс преобразования).

4. Преобразованный документ отображается в окне в виде 0 и 1. Документ можно просмотреть с помощью значка прокрутки по вертикали (справа) и горизонтали (внизу). А также, набрав координаты X и Y и нажав кнопку «Показать», можно просмотреть любую часть документа (рис. 3.3).

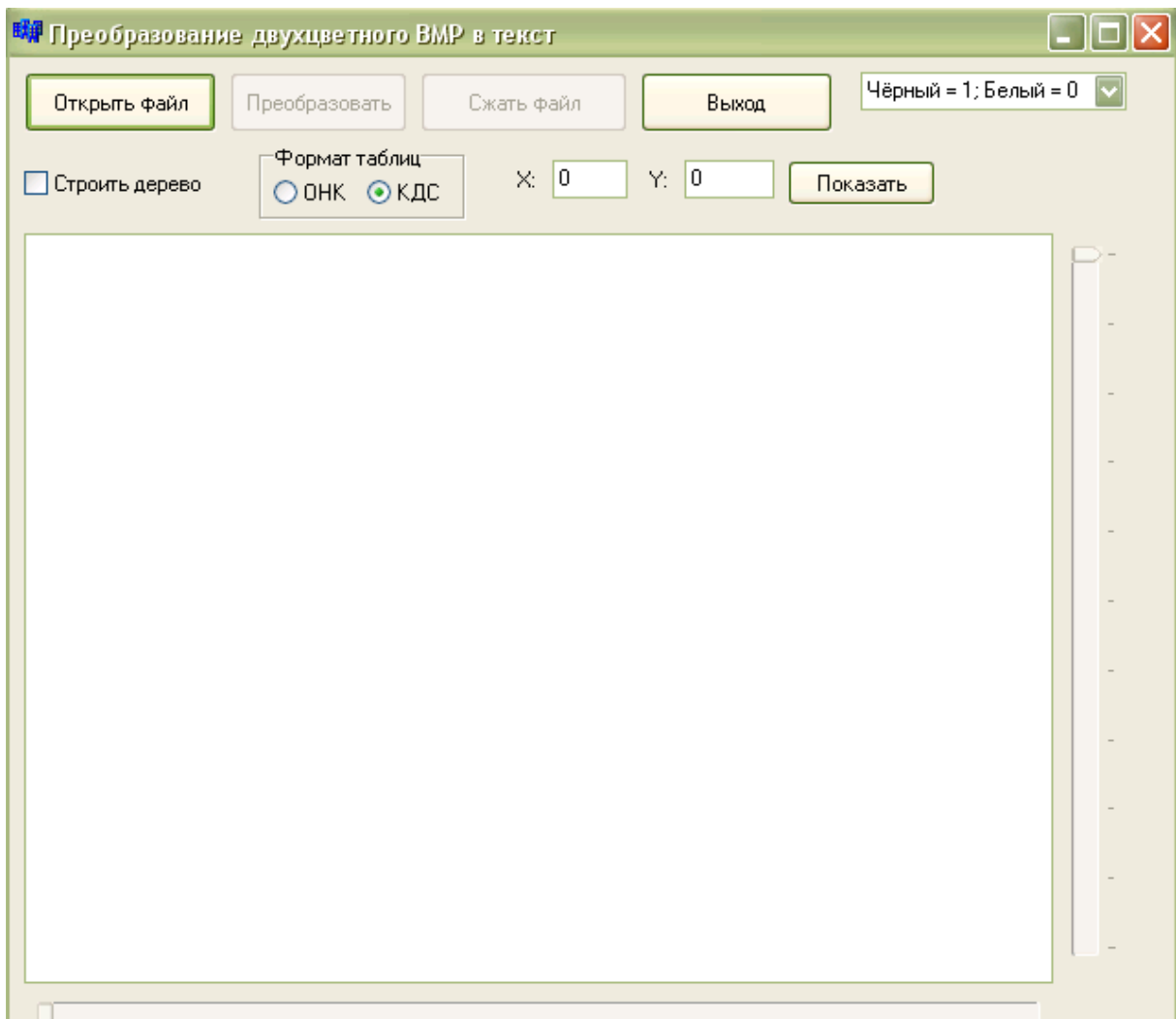


Рис. 3.2. Графический интерфейс для преобразования файла .bmp в текстовый формат

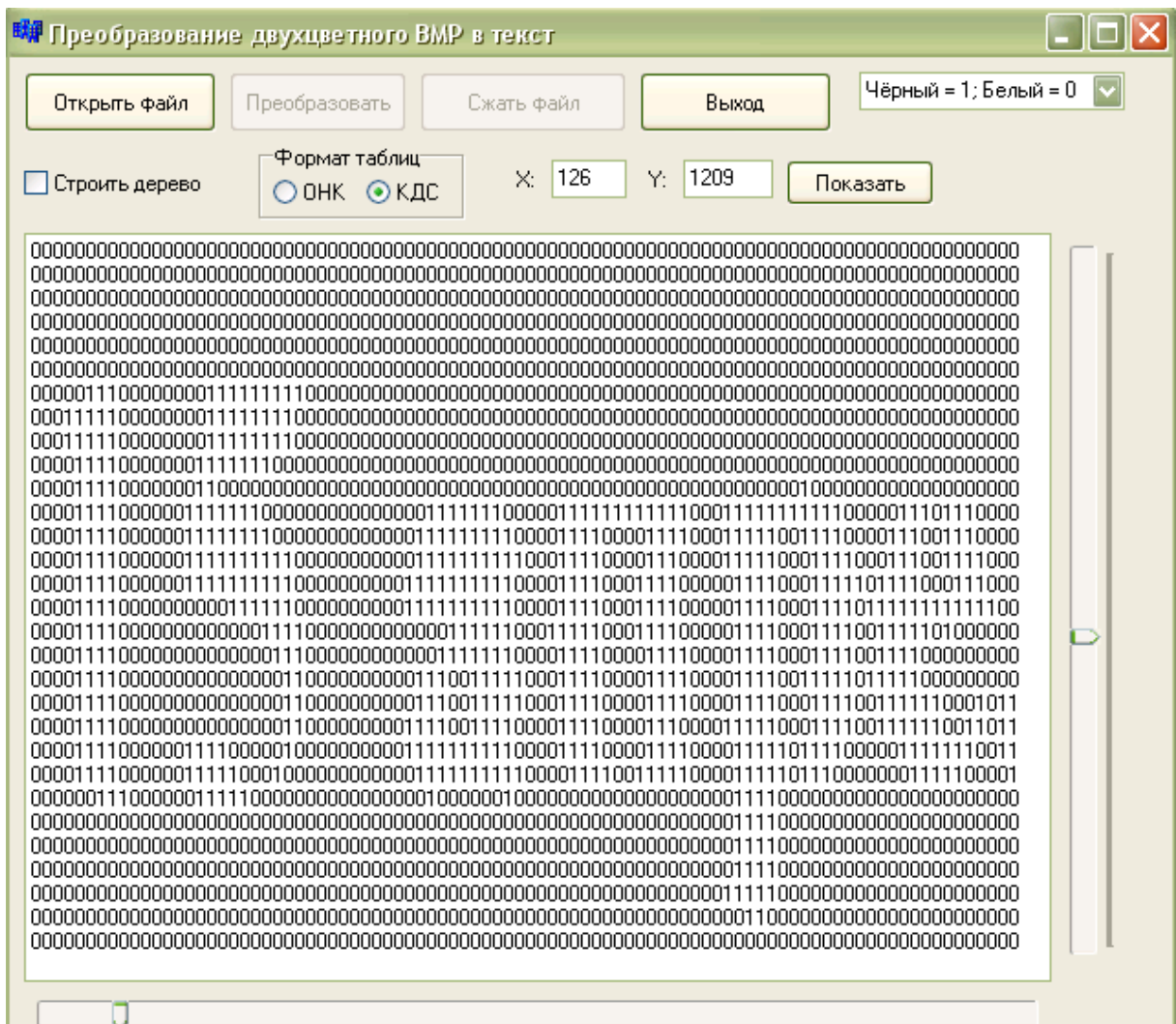


Рис. 3.3. Просмотр документа в формате .txt

Таким образом, документ представлен в двоичном виде.

Во время преобразования файла программа генерирует таблицы со следующим содержанием.

Таблица «рис.One» содержит данные о длине черной серии (первый столбец) и их количестве (второй столбец). С помощью прокрутки можно посмотреть все значения (рис. 3.4).

Таблица «рис.Zero» содержит данные о длине белой серии (первый столбец) и их количестве (второй столбец). С помощью прокрутки можно посмотреть все значения до конца (рис. 3.5).

Таблица «Result» содержит результат формулы (3.1), где первый столбец - это m от 1 до 15, второй – $N_{сж}^6$, третий – $N_{сж}^4$ (рис. 3.6).

Можно посмотреть графики зависимости $N_{сж}^6$ и $N_{сж}^4$ от m . На графиках можно визуально определить, при каком значении m $N_{сж}^6$ ($m = 7$) и $N_{сж}^4$ ($m = 3$) – оптимально для данного документа; можно сравнить эти значения с табличными (рис. 3.7).

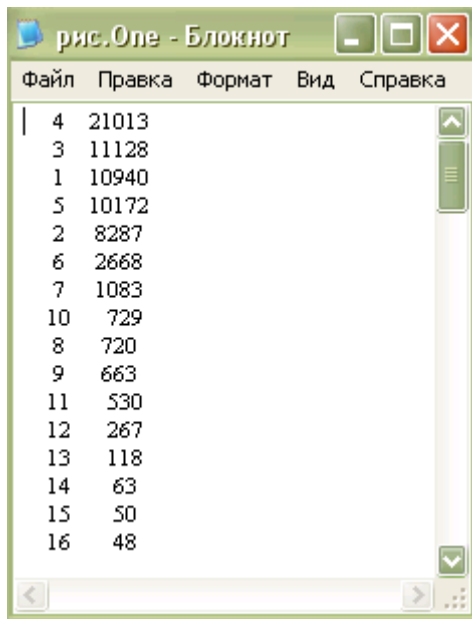


Рис 3.4

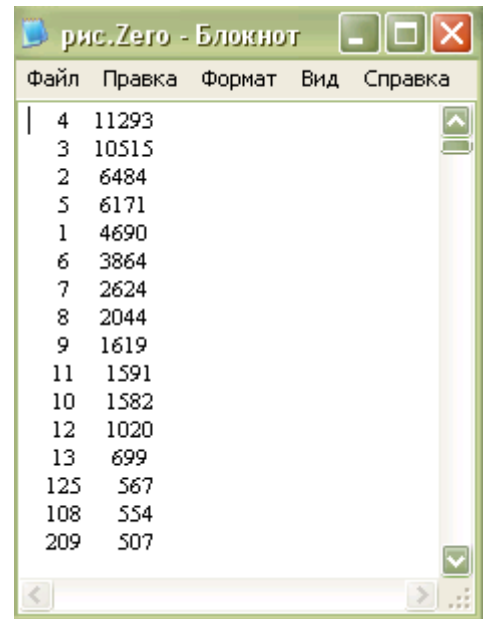


Рис 3.5

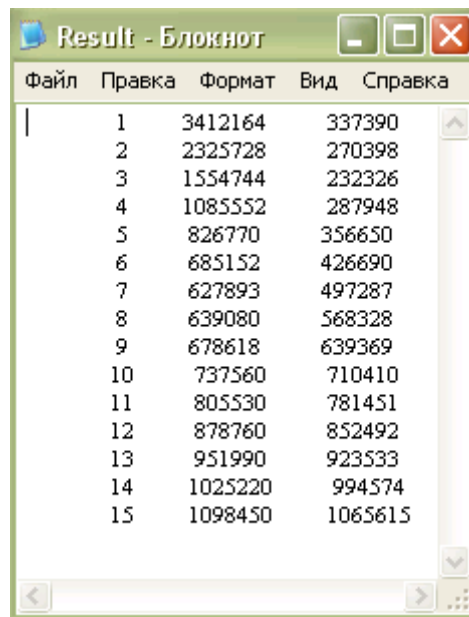


Рис. 3.6

На рис. 3.7 показаны зависимости $N_{сж}^{\delta} = f(m_{\delta})$ и $N_{сж}^{\chi} = f(m_{\chi})$. $N_{сж}^{\delta}$ обозначено как SZ_m , а $N_{сж}^{\chi}$ обозначено как SO_m .

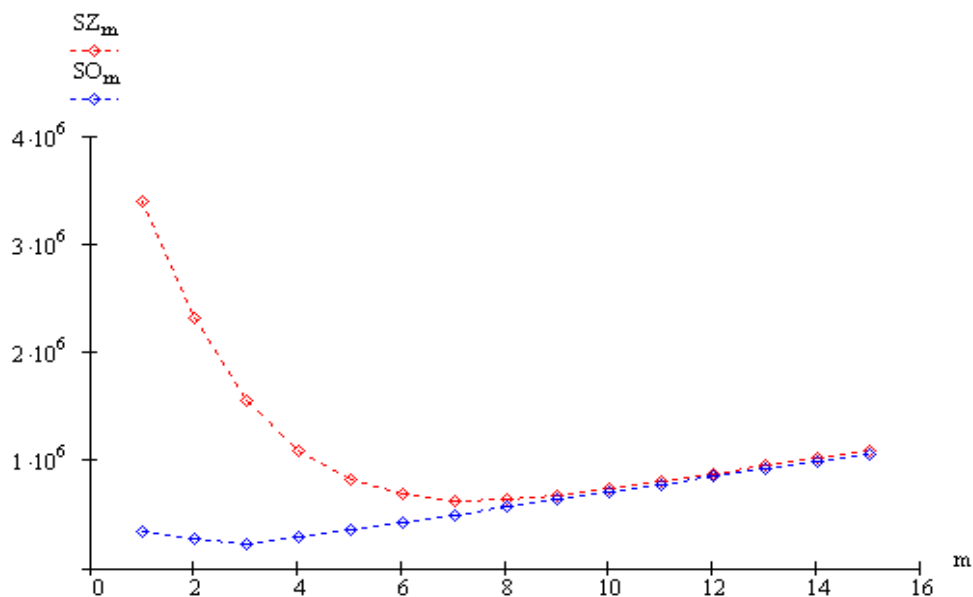


Рис. 3.7. Графики для выбора m_{δ}^* и m_{χ}^*

Из графиков видно, что минимальное значение $N_{сж}^{\delta}$ для анализируемого документа равно $6,3 \cdot 10^5$. При этом $m_{\delta}^* = 7$. Минимальное значение $N_{сж}^{\chi}$ для анализируемого документа равно $2,3 \cdot 10^5$.

При этом $m_{\chi}^* = 3$. Коэффициент сжатия:

$$K_{сж} = \frac{N}{N_{сж}^{\delta} + N_{сж}^{\chi}} = 4,19.$$

На рис. 3.8 представлена таблица «Коэффициент сжатия», которая содержит зависимости $K_{сж}$ от m .

m	$N_{сж}^{\delta}$	$N_{сж}^{\chi}$	N	$K_{сж}^{\delta}$	$K_{сж}^{\chi}$	$K_{сж}$	X	Y	$K_{сж\text{ опт}}$
1	3412164	337390	3605283	0,978538546	0,789439521	0,961523157	3338934	266349	4,191122261
2	2325728	270398		1,435651117	0,985025777	1,388716495			
3	1554744	232326		2,147577994	1,146445081	2,01742685			
4	1085552	287948		3,075793698	0,924989929	2,624887514			
5	826770	356650		4,038528248	0,746807795	3,046494905			
6	685152	426690		4,873274835	0,624221332	3,242621703			
7	627893	497287		5,317679923	0,535604188	3,204183331			
8	639080	568328		5,22459473	0,468653665	2,985969117			
9	678618	639369		4,920196635	0,416581035	2,735446556			
10	737560	710410		4,526999837	0,374922932	2,489887912			
11	805530	781451		4,145015083	0,340839029	2,271787123			
12	878760	852492		3,79959716	0,312435777	2,082471529			
13	951990	923533		3,507320455	0,288402255	1,922281412			
14	1025220	994574		3,256797565	0,267802094	1,784975597			
15	1098450	1065615		3,039677728	0,249948621	1,665977223			

Рис. 3.8

Рассчитаем коэффициент сжатия, который определяется по формуле $K_{сж} = \frac{N}{N_{сж}}$, где N – общее количество элементов на изображении данного документа. Так как изображение документа состоит из 2189 строк, а каждая строка содержит 1647 элемента, то $N=2189 \cdot 1647=3,6 \cdot 10^6$ элементов.

$N_{сж}$ – число двоичных разрядов для представления сжатого изображения документа, т.е. $N_{сж} = N_{сж}^б + N_{сж}^ч$. Были найдены оптимальные значения m_b^* и m_c^* , для которых $N_{сж}^б$ и $N_{сж}^ч$ минимальные. Таким образом, используя значения $N_{сж}^б$ при $m=7$ и $N_{сж}^ч$ при $m=3$, находим оптимальное $N_{сж}$ для данного документа. $N_{сж} = 8,6 \cdot 10^5$ элементов.

Для того, чтобы найти наилучший коэффициент сжатия для белых ($K_{сж}^б$) и черных ($K_{сж}^ч$) элементов, используем оптимальные значения $N_{сж}^б$ и $N_{сж}^ч$, а также долю белых элементов (нулевых двоичных разрядов) от всего документа и долю черных элементов (единичных двоичных разрядов) соответственно. Долю нулевых двоичных разрядов обозначим X , а долю единичных двоичных разрядов обозначим Y .

$K_{сж}^б$ находим по формуле $K_{сж}^б = \frac{X}{N_{сж}^б} \approx 5,3$, аналогично находим

$$K_{сж}^ч = \frac{Y}{N_{сж}^ч} \approx 1,14.$$

Графики зависимости K , $K_{сж}^б$ и $K_{сж}^ч$ от m представлены на рис. 3.9.

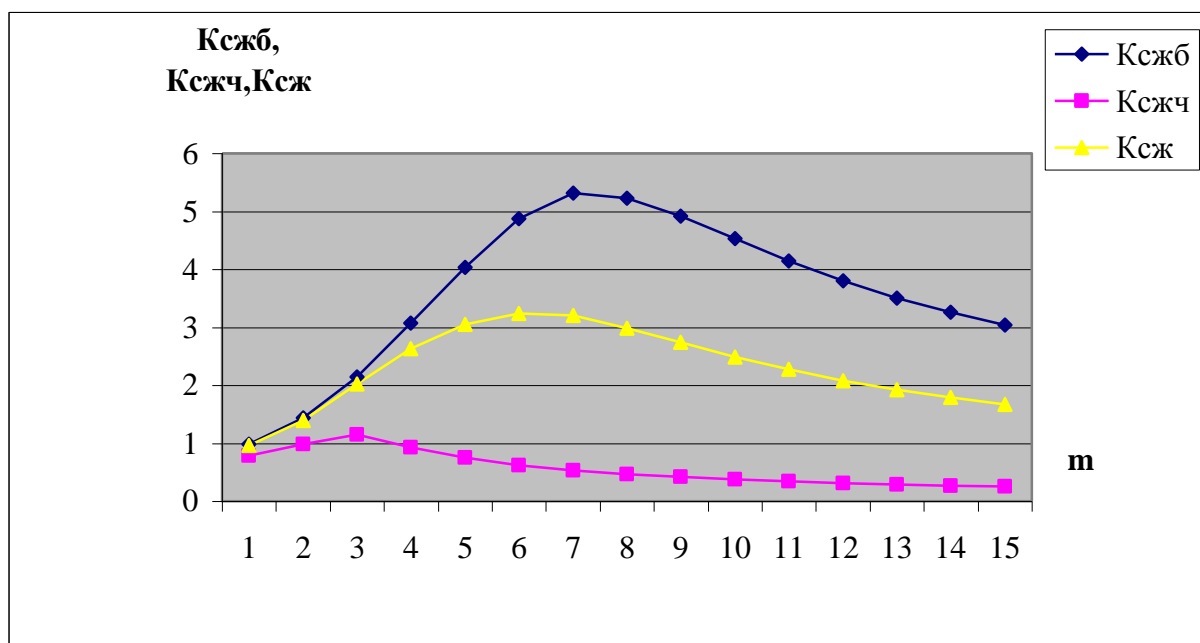


Рис. 3.9. Графики K , $K_{сж}^б$, $K_{сж}^ч$ в зависимости от m

Вывод.

Для данного текстового документа (рис. 3.1) оптимальные значения длин серий m_b^* и m_c^* обеспечивают коэффициент сжатия $K_{сж} \approx 5$.

В п. 3.1 было показано, что коэффициент сжатия для рассматриваемого в качестве примера текстового документа должен быть больше или равен 7.

Таким образом, данный документ не подходит для сжатия по методу КДС, так как не позволяет выполнить требования Рекомендации Т.4.

Рассмотрим пример преобразования по методу КДС документа «Письмо» (рис. 3.10). Найдем $K_{сж}$ для этого документа по схеме аналогичной предыдущей. Документ представлен в двоичном виде.

Во время преобразования файла программа генерирует таблицы со следующим содержанием.

Таблица «рис 1.Zero» содержит данные о длине белой серии (первый столбец) и их количестве (второй столбец). С помощью прокрутки можно просмотреть все значения до конца (рис. 3.11).

Таблица «рис 1.One» содержит данные о длине черной серии (первый столбец) и их количестве (второй столбец). С помощью прокрутки можно просмотреть все значения (рис. 3.12).

Таблица «Result 1» содержит результат формулы (3.1), где первый столбец - это m от 1 до 15, второй – $N_{сж}^b$, третий – $N_{сж}^c$ (рис. 3.13).

Можно посмотреть графики зависимости $N_{сж}^b$ и $N_{сж}^c$ от m . На графиках можно визуальным образом определить, при каком значении m $N_{сж}^b$ ($m = 7$) и $N_{сж}^c$ ($m = 3$) – оптимально для данного документа; можно сравнить эти значения с табличными (рис. 3.14).

7	2628
1	2347
5	2197
6	1916
1646	1691
8	1440
4	1221
2	1104
3	905
9	872
10	819
11	400
12	271
16	269
13	251

Рис 3.11

4	4475
1	4164
2	4157
3	3289
5	2256
6	591
7	278
8	121
9	78
10	58
11	31
12	19
14	16
13	13
15	12

Рис 3.12

Начальникам управлений
Государственного комитета
Российской Федерации
по телекоммуникациям

Руководителям организаций

18.11.99 № 40-у

О Министре Российской Федерации по
связи и информатизации

Указом Президента Российской Федерации от 12 ноября 1999 года № 1488 (в соответствии с пунктом «д» статьи 83 Конституции Российской Федерации) Министром Российской Федерации по связи и информатизации назначен Рейман Леонид Дододжонович.

ПРЕДЛАГАЮ:

Принять изложенный Указ Президента Российской Федерации к сведению.

Заместитель
председателя комитета

Н.Ф.ПОЖИТКОВ

Рис. 3.10. Пример текстового документа «Письмо»

Файл	Правка	Формат	Вид	Справка
1	871838	82573		
2	596830	69364		
3	407790	67548		
4	285372	87416		
5	224220	109110		
6	197154	130932		
7	190183	152754		
8	195320	174576		
9	207684	196398		
10	225940	218220		
11	246136	240042		
12	268512	261864		
13	290888	283686		
14	313264	305508		
15	335640	327330		

Стр 15

Рис. 3.13

На рис. 3.14 показаны зависимости $N_{сж}^6 = f(m_6)$ и $N_{сж}^4 = f(m_4)$. $N_{сж}^6$ обозначено как SZ_m , а $N_{сж}^4$ обозначено как SO_m .

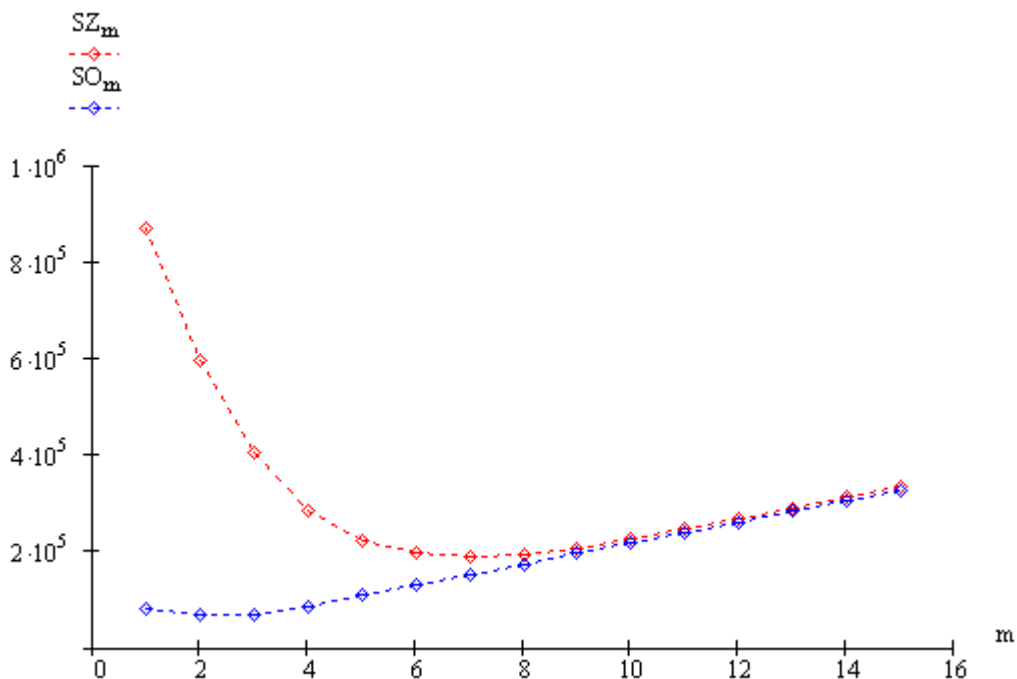


Рис. 3.14. Графики для выбора m_6^* и m_4^*

Из графиков видно, что минимальное значение $N_{сж}^6$ для анализируемого документа равно $1,9 \cdot 10^5$. При этом $m_6^* = 7$. Минимальное значение $N_{сж}^4$ для анализируемого документа равно $6,7 \cdot 10^4$. При этом $m_4^* = 3$.

Коэффициент сжатия:

$$K_{сж} = \frac{N}{N_{сж}^6 + N_{сж}^4} = 14,3.$$

На рис. 3.15 представлена таблица «Коэффициент сжатия», которая содержит зависимости $K_{сж}$ от m .

m	Nсжб	Nсжч	N	Kсжб	Kсжч	Kсж	X	Y	Kсж опт
1	871838	82573	3693624	4,166884215	0,736027515	3,870055982	3632848	60776	14,3
2	596830	69364		6,086905819	0,876189378	5,544366956			
3	407790	67548		8,908624537	0,899745366	7,770521187			
4	285372	87416		12,7302188	0,695250297	9,90810863			
5	224220	109110		16,20215859	0,557015856	11,08098281			
6	197154	130932		18,42644836	0,46417988	11,25809696			
7	190183	152754		19,10185453	0,397868468	10,77056136			
8	195320	174576		18,59946754	0,34813491	9,985574324			
9	207684	196398		17,49219006	0,309453253	9,140778357			
10	225940	218220		16,07881739	0,278507928	8,315976225			
11	246136	240042		14,75951506	0,253189025	7,597266845			
12	268512	261864		13,52955548	0,23208994	6,964161274			
13	290888	283686		12,48882044	0,214236868	6,428456561			
14	313264	305508		11,59676184	0,198934234	5,969281092			
15	335640	327330		10,82364438	0,185671952	5,571329019			

Рис. 3.15

Рассчитаем коэффициент сжатия, который определяется по формуле $K_{сж} = \frac{N}{N_{сж}}$, где N – общее количество элементов на изображении данного документа. Так как изображение документа состоит из 2244 строк, а каждая строка содержит 1646 элемента, то $N=2244 \cdot 1646 = 3,7 \cdot 10^6$ элементов.

$N_{сж}$ – число двоичных разрядов для представления сжатого изображения документа, т.е. $N_{сж} = N_{сж}^6 + N_{сж}^4$. Были найдены оптимальные значения m_b^* и m_c^* , для которых $N_{сж}^6$ и $N_{сж}^4$ минимальные. Таким образом, используя значения $N_{сж}^6$ при $m=7$ и $N_{сж}^4$ при $m=3$, находим оптимальное $N_{сж}$ для данного документа. $N_{сж} = 2,6 \cdot 10^5$ элементов.

Для того, чтобы найти наилучший коэффициент сжатия для белых ($K_{сж}^6$) и черных ($K_{сж}^4$) элементов, используем оптимальные значения $N_{сж}^6$ и $N_{сж}^4$, а также долю белых элементов (нулевых двоичных разрядов) от всего документа и долю черных элементов (единичных двоичных разрядов) соответственно. Долю нулевых двоичных разрядов обозначим X , а долю единичных двоичных разрядов обозначим Y .

$K_{сж}^6$ находим по формуле $K_{сж}^6 = \frac{X}{N_{сж}^6} \approx 19,1$, аналогично находим

$$K_{сж}^u = \frac{Y}{N_{сж}^u} \approx 0,9.$$

Графики зависимости $K_{сж}$, $K_{сж}^b$ и $K_{сж}^ч$ от m представлены на рис. 3.16.

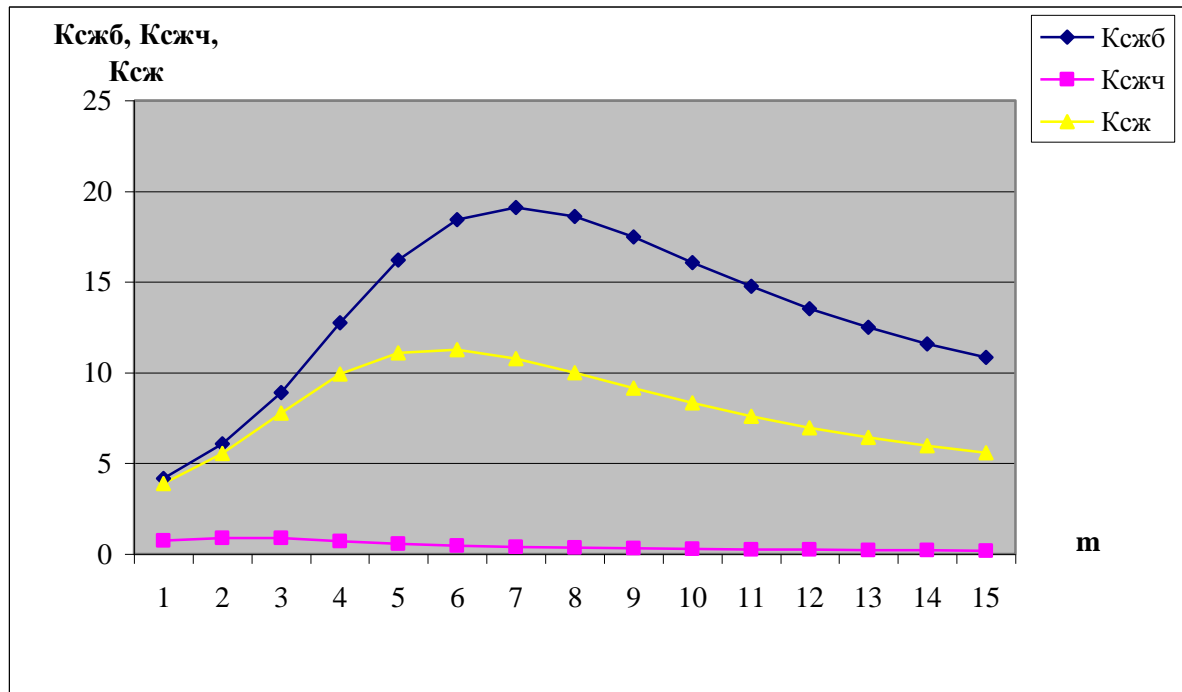


Рис. 3.16. Графики $K_{сж}$, $K_{сж}^b$, $K_{сж}^ч$ в зависимости от m

Вывод.

Для данного текстового документа (рис. 3.9) оптимальные значения длин серий m_b^* и m_c^* обеспечивают коэффициент сжатия $K_{сж} \approx 14,3$.

В п. 3.1 было показано, что коэффициент сжатия для документа на бланке формата А4 должен быть больше или равен 7.

Таким образом, данный документ подходит для сжатия по методу КДС, так как позволяет выполнить требования Рекомендации Т.4.

4. Оценка эффективности метода АПК для сжатия монохромных изображений документов

4.1. Постановка задачи

В качестве критерия для оценки метода сжатия АПК выберем коэффициент сжатия.

Как было показано в п. 2.2 сжатая строка в методе АПК имеет следующую структуру:

[код начала строки] [код первого элемента] [КНЗ] [КНЗ] [КНЗ].

Длина кода для представления координат элементов с новым значением яркости (КНЗ) при равномерном кодировании будет

определяться из условия, что элемент с новым значением яркости находится в конце строки. Т. е. $k = \log_2 N_c$, где N_c - число элементов изображения на строке.

Коэффициент сжатия $K_{сж}$ одной строки при равномерном кодировании новых значений можно вычислить, зная вероятность $P_{нз}$ появления таких элементов:

$$K_{сж} = \frac{N_c}{m_n + m_y + N_c \cdot P_{нз} \cdot \log_2 N_c},$$

где m_n – длина кода начала строки;

m_y – длина кода яркости первого элемента.

Обычно длина кода начала строки m_n равна длине кода для представления координат элементов с новым значением яркости, т.е.

$$m_n = k = \log_2 N_c = m.$$

Длину кода яркости первого элемента можно выбрать равной одному двоичному разряду, т. е. $m_y = 1$.

Если значение этого разряда равно нулю, то первый элемент белый. Если значение равно единице, то первый элемент черный.

Вероятность $P_{нз}$ можно представить так:

$$P_{нз} = \frac{n_c - 1}{N_c},$$

где n_c – количество белых и черных серий на строке. Действительно, если строка полностью белая или черная, то $n_c = 1$ и $P_{нз} = 0$. Это означает, что новых элементов нет.

$$\text{Если } P_{нз} = 0, \text{ то } K_{сж} = \frac{N_c}{m+1} > 1.$$

Максимальное значение $P_{нз}$ будет тогда, когда черные и белые элементы чередуются, т. е. $n_c = N_c$.

$$\text{Таким образом, } P_{нз \max} = \frac{N_c - 1}{N_c}, \text{ а}$$

$$K_{сж} = \frac{N_c}{m+1 + N_c \cdot \frac{N_c - 1}{N_c} \cdot m} = \frac{N_c}{m+1 + (N_c - 1) \cdot m} < 1.$$

Подобные рассуждения можно провести и для всего изображения документа.

Пусть b - число строк линий на изображении. Тогда выражение для коэффициента сжатия будет иметь вид:

$$K_{сж} = \frac{N}{b \cdot m + b + N \cdot P_{нз} \cdot m},$$

где $N = b \cdot N_c$ – число элементов на изображении.

Вероятность появления новых элементов (серий) для полного изображения определяется по формуле:

$$P_{нз} = \frac{S_n - b}{N},$$

где $S_n = \sum_{i=1}^b n_{ci}$, т.е. необходимо знать количество белых и черных серий на всем изображении.

4.2. Расчет коэффициента сжатия. Выводы о применимости метода АПК

Рассчитаем Ксж как функцию от вероятности появления новых серий на изображении $P_{нз}$ для документов формата А4, плотности сканирования по горизонтали $S_r = 8$ линий/мм и плотности сканирования по вертикали S_b :
3,85 линий/мм (режим STANDART),
7.7 линий/мм (режим FINE),
15,4 линий/мм (режим SUPER FINE).

Число строк на изображении формата А4

Режим a,b	STANDART (S)	FINE (F)	SUPER FINE (SF)
$a=S_r \cdot 210$	1680	1680	1680
$b=S_b \cdot 297$	1143	2287	4574

Число элементов на изображении формата А4

Режим N	STANDART (S)	FINE (F)	SUPER FINE (SF)
$N=a \cdot b$	$1.92 \cdot 10^6$	$3.84 \cdot 10^6$	$7.68 \cdot 10^6$

Длина кода для представления координат элементов с новым значением яркости равна $m = \log_2 a = \log_2 1680 \approx 11$.

Коэффициент сжатия:

$$K_{сж} = f(N, b, m, P_{нз}) = \frac{N}{b \cdot m + b + N \cdot P_{нз} \cdot m} =$$

$$= \frac{N}{11 \cdot b + b + 11 \cdot N \cdot P_{нз}} = \frac{N}{12 \cdot b + 11 \cdot N \cdot P_{нз}}.$$

Построим таблицы и графики Ксж от Рнз для разных режимов сканирования по вертикали.

Рнз		0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	0,1
Ксж	S	139,98	8,5	4,4	2,9	2,2	1,7	1,4	1,2	1,1	1	0,9
	F	139,922	8,5	4,4	2,9	2,2	1,7	1,4	1,2	1,1	1	0,9
	SF	139,921	8,5	4,4	2,9	2,2	1,7	1,4	1,2	1,1	1	0,9

Рнз		0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Ксж	S	0,4	0,3	0,2	0,1	0,1	0,1	0,1	0,1	0
	F	0,4	0,3	0,2	0,1	0,1	0,1	0,1	0,1	0
	SF	0,4	0,3	0,2	0,1	0,1	0,1	0,1	0,1	0

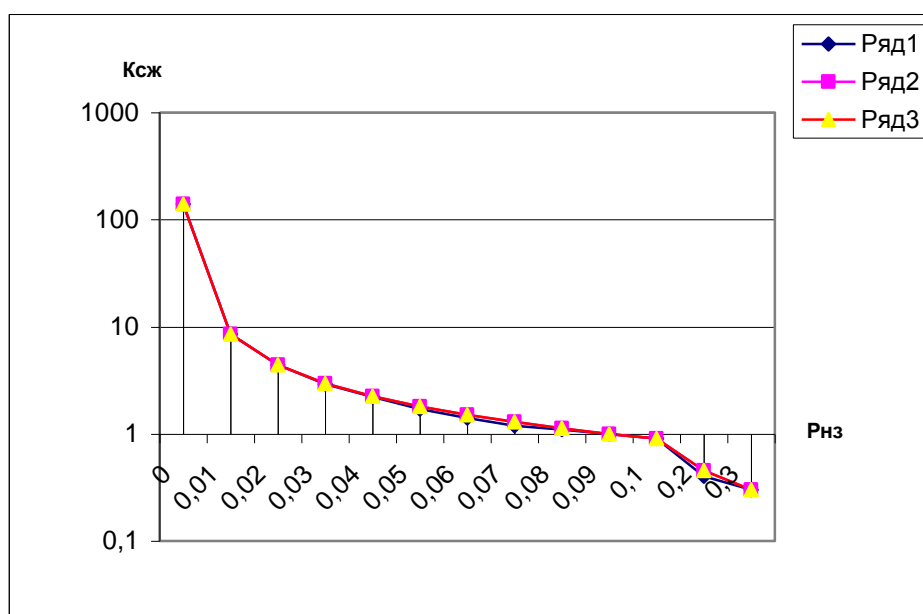


Рис. 4.1. График зависимости Рнз от Ксж

Если $P_{нз} \geq 0,09$, то сжатие по методу АПК использовать нельзя, т.к. $K_{сж} \leq 1$.

Рассмотрим сжатие по методу АПК для документов, которые были обработаны по методу КДС.

Рассчитаем коэффициент сжатия для документа «Расписание занятий» (рис. 3.1).

$$K_{сжс} = \frac{N}{12 \cdot b + 11 \cdot N \cdot P_{нз}}.$$

Для данного документа:

$N = 1647 \cdot 2189 = 3605293$ элементов;

$b = 2189$ элементов;

$$S_n = \sum_{i=1}^b n_{ci} = \sum_{i=1}^{2189} n_{ci} = 139893$$

$$P_{нз} = \frac{S_n - b}{N} = \frac{139893 - 2189}{3605293} = 0,04;$$

$$K_{сжс} = \frac{N}{12 \cdot b + 11 \cdot N \cdot P_{нз}} = \frac{3605293}{12 \cdot 2189 + 11 \cdot 3605293 \cdot 0,04} = 2,4.$$

Рассчитаем коэффициент сжатия для документа «Письмо» (рис. 3.10).

$$K_{сжс} = \frac{N}{12 \cdot b + 11 \cdot N \cdot P_{нз}}.$$

Для данного документа:

$N = 1646 \cdot 2244 = 3693624$ элементов;

$b = 2244$ элементов;

$$S_n = \sum_{i=1}^b n_{ci} = \sum_{i=1}^{2244} n_{ci} = 41402$$

$$P_{нз} = \frac{S_n - b}{N} = \frac{41402 - 2244}{3693624} = 0,01;$$

$$K_{сжс} = \frac{N}{12 \cdot b + 11 \cdot N \cdot P_{нз}} = \frac{3693624}{12 \cdot 2244 + 11 \cdot 3693624 \cdot 0,01} = 8,1.$$

Рассчитаем коэффициент сжатия для документа «Графики» (рис. 4.2).

$$K_{сжс} = \frac{N}{12 \cdot b + 11 \cdot N \cdot P_{нз}}.$$

Для данного документа:

$N = 1652 \cdot 2271 = 3751692$ элементов;

$b = 2271$ элементов;

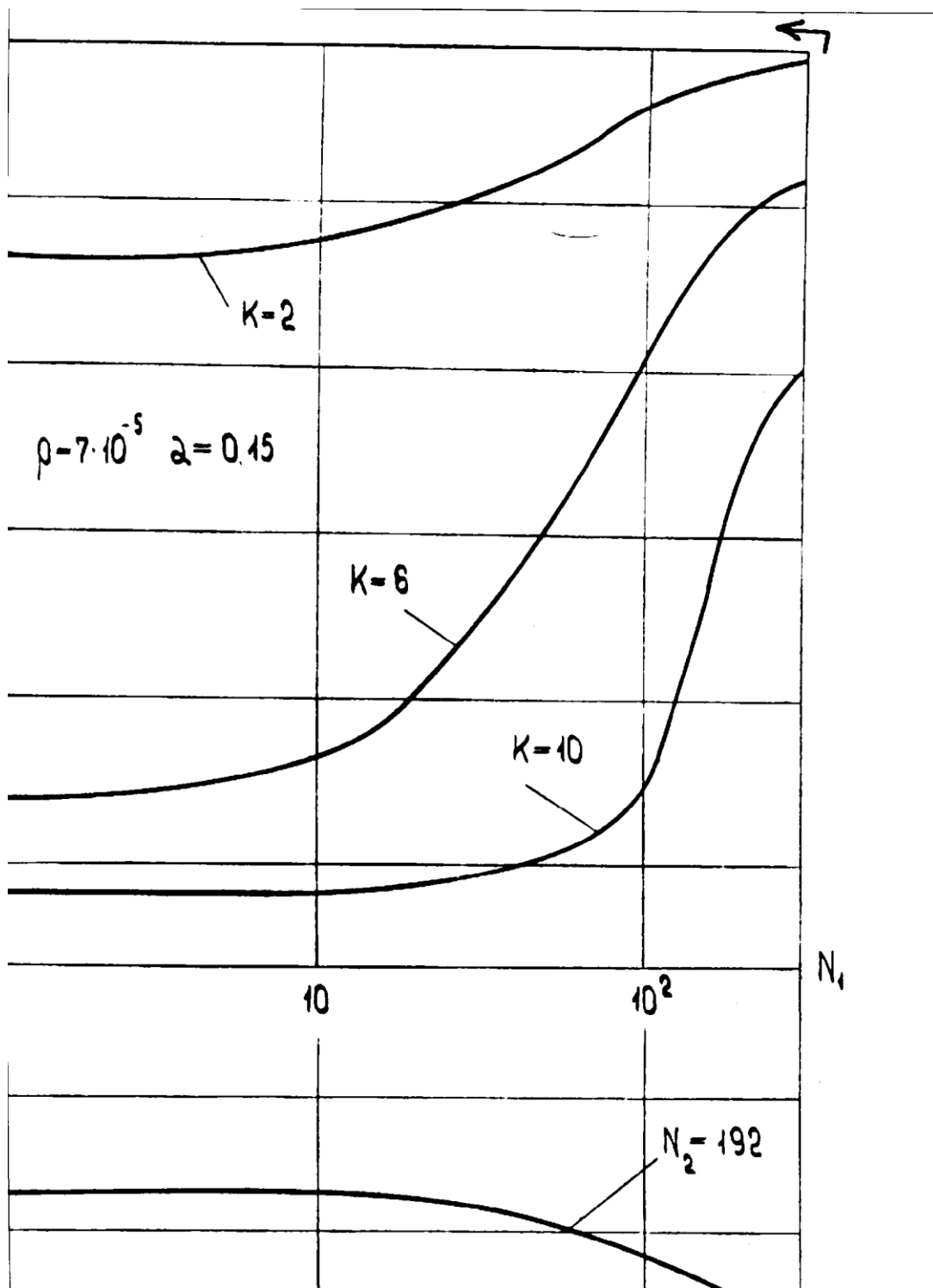


Рис. 4.2. Пример текстового документа «Графики»

$$S_n = \sum_{i=1}^b n_{ci} = \sum_{i=1}^{2271} n_{ci} = 27326$$

$$P_{нз} = \frac{S_n - b}{N} = \frac{27326 - 2271}{3751692} = 0,007;$$

$$K_{сж} = \frac{N}{12 \cdot b + 11 \cdot N \cdot P_{нз}} = \frac{3751692}{12 \cdot 2271 + 11 \cdot 3751692 \cdot 0,007} = 11,9.$$

Таким образом, можно сделать вывод, что Ксж при использовании метода АПК существенно зависит от вероятности новых значений $P_{нз}$ и для определенных документов (например, типа документа «Графики») может быть успешно использован.

Заключение

В дипломной работе проведен анализ эффективности двух методов сжатия текстовых документов для факсимильных аппаратов G3. Были проанализированы алгоритмы сжатия для факсимильных аппаратов G3.

Были рассмотрены методы кодирования длин серий (КДС) и адресно-позиционное кодирование (АПК).

Из 3-х критериев оценки сжатия документов (коэффициент сжатия, помехозащищенность и сложность алгоритмов реализации) был выбран коэффициент сжатия, как наиболее показательный.

Сделаны выводы о возможности использования метода КДС для монохромных текстовых документов, в результате нахождения оптимальных длин серий путем последовательного перебора значений, при которых коэффициент сжатия был максимальным (например, документ «Письмо»); метода АПК для монохромных текстовых документов, так как коэффициент сжатия существенно зависит от вероятности новых значений, чем меньше вероятность, тем коэффициент сжатия больше (например, документ «Графики»), а также было найдено пороговое значение вероятности новых значений.

Проведенный анализ показал, что рассмотренные методы сжатия монохромных текстовых документов КДС и АПК могут использоваться в современных факсимильных аппаратах при разработке факсимильных программ.

Список литературы

1. Щелованов Л.Н. Системы факсимильной связи / Уч. пособие. - Л.: ЛЭИС, 1991.
2. Зелигер Н. Б. Основы передачи данных. Учебное пособие для Вузов. М., «Связь», 1974.
3. Медведев Г. Н., Старков Ф. А. Технологии сетей связи. Особенности кодирования цифровых факсимильных сообщений. Учебное пособие. Курский государственный технический университет, 2002.
4. Передача дискретных сообщений. - М.: Радио и связь, 1990.
5. Питерсон У., Уэлдон Э. Коды, исправляющие ошибки.— М.: Мир, 1976.
6. Олифер В.Г., Олифер Н.А. Компьютерные сети. Принципы, технологии, протоколы: Учебник для вузов.— СПб.: Питер, 2003.
7. Руководящий документ отрасли РД.45.129-2000. Телематические службы.