

МІНІСТЕРСТВО ОСВІТИ І НАУКИ, МОЛОДІ ТА СПОРТУ УКРАЇНИ
Тернопільський національний економічний університет
Факультет комп'ютерних інформаційних технологій

ОПОРНИЙ КОНСПЕКТ ЛЕКЦІЙ

з курсу

“КОМПРЕСІЯ ТА АРХІВУВАННЯ ДАНИХ”

для студентів спеціальності:

6.09150 “Спеціалізовані комп'ютерні системи”,

ТЕРНОПІЛЬ – 2012

Тема 1. Роль методів стиснення даних в сучасних інформаційних та автоматизованих системах.

Проблема стиснення даних з'явилася разом з появою техніки. Розвиток інформаційних мереж і систем, зокрема автоматизації систем управління різних рівнів, обчислювальної мережі, телеметричних комплексів, інформаційно-довідкових систем, довідкових комплексів, мереж зв'язку, привів до суттєвого збільшення потоків між територіально розподіленими джерелами інформації і користувачами.

Для підвищення ефективності використання комунікаційних та інформаційно-обчислювальних ресурсів вказаних системою використовують різні методи та засоби. Серед них важливу роль відіграють методи зменшення надлишковості даних, які забезпечують стиснення об'єму інформації, що дозволяє значно завантажити канали зв'язку і системи обробки та зберігання даних за рахунок виключення непотрібних повідомлень. Це еквівалентне підвищення пропускну здатності системи збою. Передавання та обробки даних або збільшення ємності запам'ятовуючих пристроїв.

Комплекс задач, вирішення яких забезпечує створення високоефективних систем зі стисненням даних включає в першу чергу синтез і аналіз методів обробки первинних повідомлень та проектування програмних засобів. Основний принцип на якому ґрунтується стиснення даних полягає в економному описі повідомлень (мовний сигнал, фото або телезображення, телеметрична інформація, будь-які числові дані) за яким можливе є відновлення початкового повідомлення.

З іншого боку опис повідомлень з затратою мінімальної кількості двійкових одиниць не є чітко сформульована задача теорії інформації. Однак шляхи реалізації конкретних процедур обробки повідомлень для усунення їх природної надлишковості за невеликим виключенням не знайдено. Особливо великі труднощі викликає нестаціонарний характер реальних повідомлень і недостатньо апріорні відмінності про фізичний процес, інформація про який цікавить споживача. Запропоновані багатьма дослідниками процедури обробки повідомлень враховують можливості сучасної техніки часто наближені за своїми

параметрами до оптимальних алгоритмів стиснення даних, які синтезуються з використанням математичних методів теорії інформації.

Теорії інформації

Чим менша імовірність події, тим більшу невизначеність знімає повідомлення про її появу і тим більшу інформацію воно несе. Хартлі запропонував кількість інформації I , що припадає на одне повідомлення, визначення логарифму загального числа можливих повідомлень: $I(N) = \log_2 N$. Якщо вся множина можливих повідомлень складається з одного повідомлення, то це відповідає повну її відсутність інформації у даному випадку $I(N) = \log_2 1 = 0$.

$$I(N) = \log_2 1 = 0.$$

При наявності незалежних джерел повідомлень з N_1 і N_2 кількістю можливих повідомлень, $I(N) = \log(N) = \log(N_1 + N_2) = \log N_1 + \log N_2$ буде дорівнювати кількості інформації сумі кількостей інформації, які були б отримані від двох незалежних джерел взятих окремо.

Якщо можливість появи будь-якого символу алфавіту рівний ймовірно тобто її ймовірність дорівнює $p = 1/m$, тоді $I = \log m = \log 1/p = -\log p$ кількість інформації на кожний рівно ймовірний сигнал дорівнює $-\log$ ймовірності (кількість інформації) окремого сигналу.

Одиниця кількості інформації, що являє собою вибір з двох рівно ймовірних подій отримала назву двійкової одиниці або біта.

Чим менша ймовірність отримання того чи іншого повідомлення тим більш воно несподіване, а отже більш інформативне.

Повідомлення ймовірність якого велика, несе небагато інформації. Формула Хартлі дозволяє визначити кількість інформації повідомлення тільки для випадку коли поява символів рівно ймовірна і вони статично незалежні.

Формула Шенона:

$$I = - \sum p_i \log_2 p_i$$

Тема 2: Інформаційні моделі повідомлень.

Повідомлення, що отримуються на приймальній стороні несе корисну інформацію лише в тому випадку, коли має місце невизначеність відносно стану джерела повідомлень. Величину, що характеризує невизначеність окремого і-го повідомлення прийнято називати частковою ентропією.

$$M(x_i) = \log_2 1/p(x_i).$$

Кількість інформації і невизначеність для всієї сукупності випадкових повідомлень можна отримати усередненням по всіх повідомленнях

$$1) I(x) = -\sum p(x_i) \log_a 1/p(x_i);$$

$$2) H(x_i) = -\sum P(x_i) \log_a p(x_i) - \text{ентропія.}$$

Незважаючи на співпадіння залежностей ентропія і кількість інформації принципово різні. Ентропія, що виражає середню невизначеність джерела повідомлень є об'єктивною характеристикою джерела повідомлення і, якщо відома статистика повідомлень може бути визначена апіорно (тобто до отримання повідомлення).

$I(x)$ є апіорною характеристикою і визначає кількість інформації, що отримується з надходження повідомлення $H(x)$ є міра нестачі інформації про стан окремої системи. З надходженням інформації про стан системи ентропія останньої зменшується. Співпадіння виразів 1 і 2 свідчить про те, що кількість отриманої інформації чисельно дорівнює ентропії, яка має місце відносно джерела повідомлення на розглянутому зв'язку кількості інформації з ентропією проявляється відомий діалектичний закон.

Одиниці вимірювання кількості інформації і ентропії залежить від вибору основи алгоритму у формулах 1 і 2.

Для десяткового логарифму – діт.

Для двійкового – біт.

Для натурального – ніт.

Середня ентропія повідомлення при однаковій кількості елементів може бути різною в залежності від статистики характеристик повідомлень. При наявності корелятивних зв'язків між елементами повідомлення ентропія зменшується. Для випадків, коли корелятивні зв'язки охоплюють два або три

елементи вираз для обчислення ентропії набуде наступного вигляду. Для 2 елементів:

$$3) H(x) = -\sum \sum p(x_i, x_j) \log_a p(x_i, x_j).$$

$$4) H(x) = -\sum \sum \sum p(x_i, x_j, x_k) \log_a p(x_i, x_j, x_k).$$

Повідомлення, ентропія яких максимально є оптимальною з точки зору найбільшої кількості інформації. Мірою того, наскільки дане реальне повідомлення за своєю ентропією відрізняється від відповідному йому оптимальному повідомлення є коефіцієнтом стиснення:

$$\mu = H(x) / H(x)_{\max},$$

$H(x)$ – реальне; $H(x)_{\max}$ – ентропія відповідному йому оптимального повідомлення.

Якщо не оптимальні і оптимальні повідомлення характеризуються однаковою законною ентропією, то справедлива наступна рівність $n^* H(x) = n' H(x)_{\max}$;

n – число елементів не оптимального повідомлення;

n' – число елементів відносно оптимального повідомлення.

Оскільки ентропія оптимального повідомлення максимальна, то число елементів не оптимального повідомлення n завжди більша від числа елементів відповідного оптимального повідомлення n' .

Коефіцієнт стиснення можна виразити через кількість елементів повідомлення $\mu = n' / n$. Таким чином реальне повідомлення при однаковій інформативності володіють певною надлишковістю в елементах порівняно з оптимальним повідомленням.

Дійсним коефіцієнтом стиснення рахується:

$$K_c = N_{\text{вх}} / N_{\text{вих}}.$$

$N_{\text{вх}}$ – кількість війкових розрядів на вході пристроїв стиснення.

$N_{\text{вих}}$ – кількість війкових розрядів на виході пристрою стиснення.

У 1989 р. група дослідників запропонували оцінювати коефіцієнт стиснення універсальних методів за допомогою набору файлів, що отримав назву Calgt.

Файл	Розмір, байт	Опис
Bib	111261	Бібліографічний список. UNIX
Book1	768771	Художня книга Томас Харді. Неформатований текст ASCII. Містить велику кількість ORC помилок.
Book2	610856	Технічна книга.
Geo	102400	Геофізичні дані, 32 бітні числа
News	370109	Набір повідомлень конференцій USNET формат ASCII.
Obj1	21504	Об'єктний файл типу VAX
Obj2	246814	Об'єктний файл для ПК Apple Macintosh
Paper1	53161	Технічна стаття Уітен, Мілк. Стаття тематична. Формат UNIX
Paper2	82794	Технічна стаття Bitten... Формат UNIX
Pic	513216	Факсимільна двоколірна картинка 1728*2726 точок. Являє собою 2 сторінки технічної книги на французькій мові від скановані з роздільною здатністю 200 точок на дюйм
ProgC	39611	Програма на мові C
Prog1	71746	Програма на мові Mops
ProgI	49379	Програма на мові Pascal ASCII коди
Trans	93695	Розширення термінальної сесії формат MMAX ASCII

Тема 3. Класифікація методів стиснення даних.

Основні причини виникнення надлишковості:

1. Ненавмисне завищення розмірності вектора відомостей, що надаються користувачу. Воно може бути викликано не повним врахуванням властивостей користувача, якому для правильного сприйняття достатньо вектора меншої розмірності. Таким чином перша причина виникнення надлишковості пов'язано з психофізичними властивостями користувача сигматичними і прагматичним аспектам сприйняття інформації.

2. Ненавмисне завищення числа компонент процесу породжуваного об'єкту, що призводить до збільшення розмірності повідомлення. Це може бути пов'язано з неоптимальністю методів отримання відомостей по параметрах процесу.

3. Не повне використання характеристик процесу не може бути викликано фізичною нереалізованістю методів, які повністю враховували апріорну інформацію. Складність реалізації і чутливість оптимальних процедур до неточного подання необхідних характеристик параметрів об'єкта. При цьому розробка оптимальних методів стиснення даних, які вичерпно враховували б відомості про процес являє теоретичний інтерес для дослідження потенційно можливої ефективності стиснення.

4. Принципова неможливість повного знання поведінки реалізації або досліджуваних характеристик процесу. Оскільки в протилежному випадку створення інформаційної системи втрачає зміст. В даному випадку йдеться про природну надлишковість повідомлень.

5. Навмисне введення надлишковості з метою підвищення достовірності оцінок цих векторів і отримуваних відомостей.

Отже, всі методи стиснення даних за першою класифікаційною ознакою – видом надлишковості, яка усувається можуть бути поділені на класи методів, що реалізують стиснення даних з врахуванням позицій 1 – 5.

Методи зменшення розмірності повідомлень розглядаються в тісному зв'язку із захопленням задачами дослідження об'єкта: класифікації, ідентифікації і т.д. і є предметом теорії розпізнавання образів прийняття рішень.

Проте, оскільки дані методи істотно визначають обсяг даних про об'єкт, що обробляється в системі стиснення даних, виходячи з міркувань спільності розгляду процесів скорочення обсягу даних. Стиснення даних в традиційній постановці пов'язаний з усуненням 3 і 4 видів усунення надлишковості.

2. Тип повідомлення. За видом методи стиснення поділяються на:

- 1) Призначені для обробки аналогових сигналів.
- 2) Аналогово-дискретних сигналів.
- 3) Дискретних сигналів.

3. Число координат повідомлень, що одночасно обробляються методом стиснення. Поділяються на методи орієнтовані на обробку кожної координати окремо (роздільна обробка) та декількох координат одночасно (сумісна обробка).

Сумісна обробка дозволяє у ряді випадків при наявності детермінованих взаємних зв'язків координат повідомлення отримати суттєвий вииграш у стисненні порівняно із роздільною обробкою. В той же час сумісна обробка реалізується більш складними підсистемами стиснення даних.

4. Предметна область на яку орієнтований метод стиснення даних, тобто на певний клас об'єкта. За цією ознакою можна виділити класи методів не пов'язаних з конкретною предметною областю (універсальні) і ті, що використовують в певній мірі специфіку класу об'єктів (спеціалізовані). Прикладом останнього є методи стиснення звуку, зображення, технологічних даних.

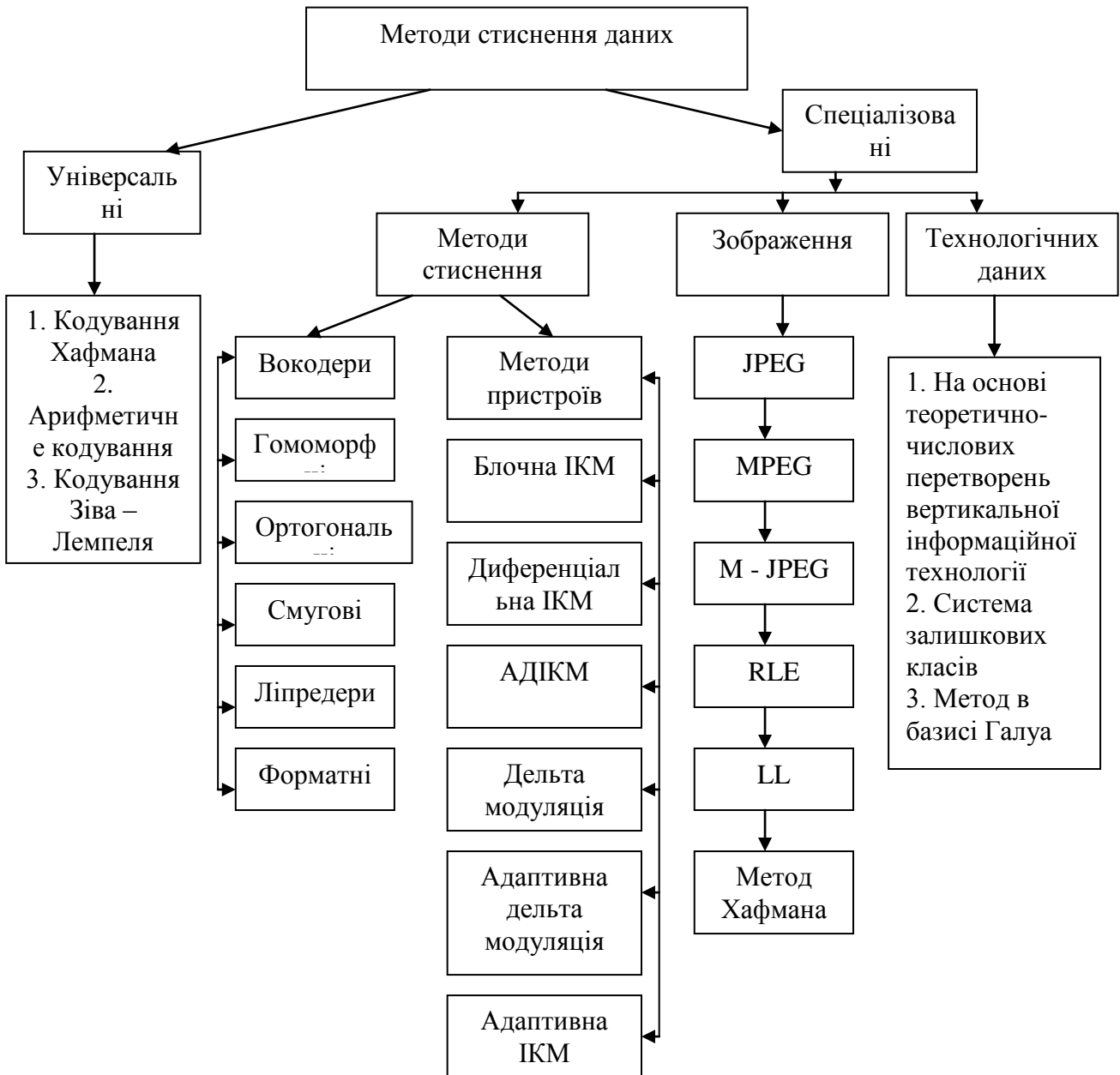
5. Число елементів (при цифровій обробці число кодових слів), що одночасно обробляються з метою стиснення.

Перший клас складають методи в яких кожен елемент (кодове слово) обробляється окремо. Ці методи визначені як не блочні. Другий клас складають групові або ж блочні методи в яких одночасно обробляються декілька вибірок (кодових слів). Блочні методи дозволяють підвищити ефективність стиснення даних джерел з пам'яттю.

Можливість відновлення початкової реалізації повідомлення з контрольованою похибкою. За цією ознакою методи стиснення поділяються на

методи стиснення без втрат і методи стиснення з втратами. У першому випадку відновлення із заданою точністю можливе у другому ні. Оскільки інформація, що надходить являє собою узагальнені параметри об'єкта.

До методів стиснення без втрат відноситься : кодування Хафмана, арифметичне кодування, кодування довжин повторів, кодування Зіва – Лемпеля, розподілу експонент і мантис.



Одним із методів стиснення без втрат є алгоритм Хафмана, але наприкінці 70-их років з'явилися дві нові ідеї, одна полягає у відкритті метода арифметичного кодування, інша у створенні метода кодування Зіва – Лемпеля. Існують два основних способи проведення стиснення: статистичний і

словниковий. Кращі статистичні використовують арифметичне кодування. Кращі словникові – Зіва – Лемпеля.

До методів з втратами належать векторне квантування, хвильовий алгоритм, фрактальний алгоритм, JPEG. Для стиснення звукових – методу імпульсно-кової модуляції, логарифмічної (тобто скомпандуванням по А та μ закону), адаптивної імпульсно-кової модуляції (АІКМ), диференційованої імпульсно-кової модуляції (ДІКМ), адаптивної диференційної імпульсно-кової модуляції (АДІКМ), адаптивної дельта модуляції (АДМ), застосування вокодерів, кліпірування.

Тема 4. Кодування Хафмана

Найбільш розповсюджений метод генерування кодів змінної довжини для символів на основі частоти їхнього використання називається кодуванням Хафмана. Цей метод запропонований Хафманом у 1952 р. Ідея алгоритму Хафмана полягає у наступному: якщо дані містять різні числа подані у вигляді бітових ланцюгів фіксованої довжини і частота з якою різні числа присутні у цих даних істотно відрізняються, то замінивши бітові ланцюги фіксованої довжини на бітові ланцюги різної довжини при чому, таким чином щоб числам, які зустрічаються часто відповідали коротші ланцюги можна отримати зменшення обсягу даних.

Процедура формування кодів Хафмана для набору значень на основі частот їхнього входження достатньо проста, вона включає створення двійкового дерева починаючи знизу до верху, при чому символи, які зустрічаються найрідше знаходяться найдалі від кореня. Спочатку створюється pool (загальний фонд), який містить або значення або вузли дерева. Спочатку цей фонд містить всі значення і не містить вузли. Приведена нижче процедура повторюється до тих пір поки pool не буде містити один вузол дерева і не буде жодного символу.

1. Знайти два значення або вузла дерева з найменшою частотою появи і видалити їх з pool-у. Якщо є декілька елементів з найменшою частотою, використовуючи зв'язки можна розірвати за допомогою довільного вибору.

2. Створіть новий вузол дерева і зробіть елементи з попереднього кроку його двома гілками.

3. Присвойте новому вузлу дерева частоту, що дорівнює сумі частот відгалужених від нього гілок.

4. Додайте новостворений вузол у pool.

Після того, як всі значення будуть об'єднані в одне дерево для кожного вузла присвоюється значення 0 – одній гілці, а 1 – другій гілці вузла. Код Хафмана для кожного кодованого значення визначається проходом по шляху від кореня до заданого значення і додаванням загального кода кожної гілки, яка визначається по мірі просування. Слід зауважити, що при роботі з кодами Хафмана оперують з бітовими стрічками. Ця стрічка змінної довжини може

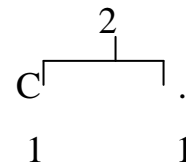
складатися з 0 і 1. Множина кодованих значень і пов'язаних з ним кодів утворює так звану таблицю Хафмана. Для того, щоб розглянути кодування Хафмана на прикладі розглянемо наступну стрічку:

A M A N A P L A N A C A N A L P A N A M A.

Значення	Частота
A	10
C	1
L	2
M	2
N	4
P	2
probil	6
.	1

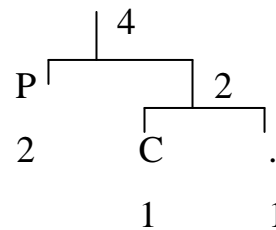
A L M N P probil

10 2 2 4 2 6



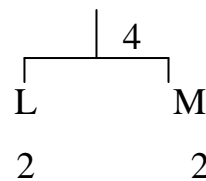
A L M N probil

10 2 2 4 6



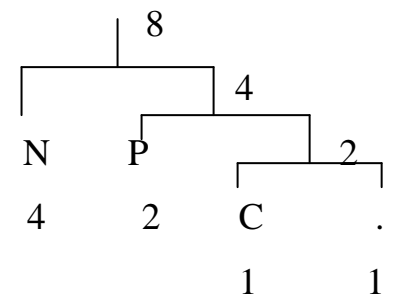
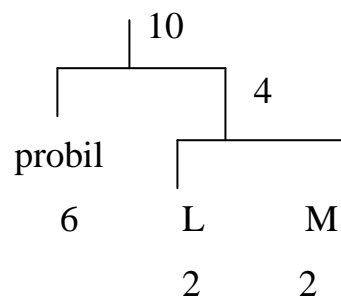
A N probil

10 4 6

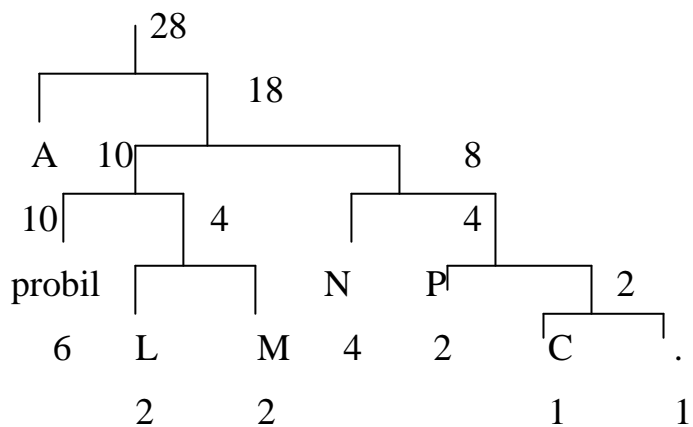
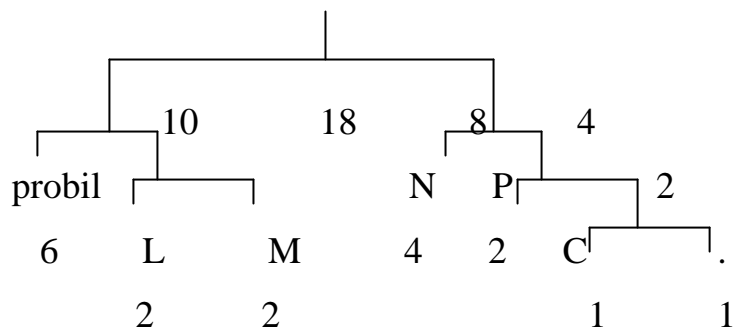


A

10

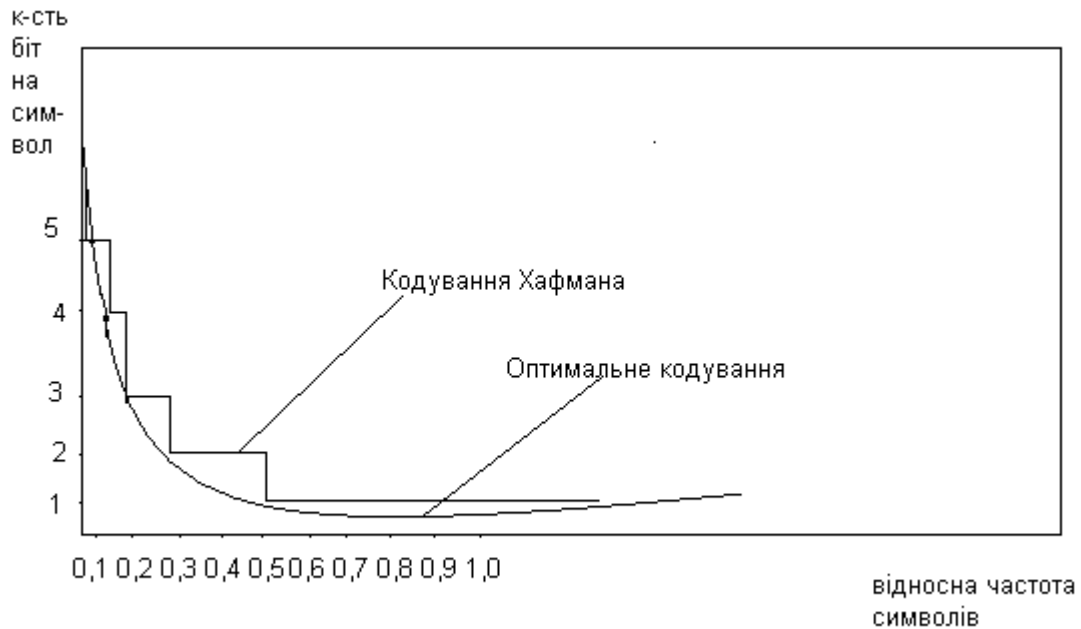


A
10



Значення	Код Хафмана	Довжина коду	Частота використання символів	Загальна кількість біт
A	0	1	10	10
C	11110	5	1	5
L	1010	4	2	8
M	1011	4	2	8
N	110	3	4	12
P	1110	4	2	8
probil	100	3	6	18
.	11111	5	1	5

Графік порівняння оптимального кодування і кодування Хафмана:



За т. Шенона найкраще стиснення в двійковій арифметиці отримаємо, якщо будемо кодувати символ з відносною частотою FC за допомогою $-\log_2(FC)$ біт.

З графіку видно, що у випадку коли відносні частоти не є степенями двійки, стиснення стає менш ефективним (ми витрачимо більше біт ніж потрібно).

Нехай ми маємо два символи A і B з відносними частотами $253/256$ і $3/256$
 $-\log_2(253/256) * 253 - \log_2(3/256) * 3 = 23,546$.

При кодуюванні методом Хафмана ми закодуємо A і B як 0 і 1 і в результаті отримаємо кодову послідовність довжиною 256 біт ($253+3$), тобто в десять разів більше ніж при оптимальному кодуюванні.

Розглянемо алгоритм, що дає близький до оптимального результат.

Арифметичне кодуювання

Основні принципи арифметичного кодуювання були розроблені наприкінці 70-их років. А арифметичне кодуювання так само, як і ймовірнісні методи використовує в якості основи технології стиснення ймовірність появи символу у файлі. Однак сам процес арифметичного кодуювання має принципові відмінності. В

результаті арифметичного кодування символну послідовність замінюється дійсним числом більшим 0 і менше 1.

Текст отриманий при стисненні розглядається як дріб, де кожна буква в алфавіті пов'язана з деяким під інтервалом відкритого з права інтервалу $[0,1]$. Текст джерела можна розглядати, як буквальноє представлення дроби, що використовує систему числення, де кожна буква в алфавіті використовується в якості числа, а інтервал значень пов'язаних з нею залежать від частоти появи цієї букви. Перша буква стиснутого тексту може бути декодована знаходженням букв, напівінтервал якої включає значення дроби, що представляє текст. Після визначення чергової букви вихідного тексту дріб перераховується для знаходження наступної. Це здійснюється відніманням від дроби основи пов'язаної з знайденою букви під області і діленням результату на ширину її пів інтервалу. Після закінчення цієї операції можна декодувати наступну букву.

Розглянемо процес арифметичного кодування

REDUNDANCE (надлишковість)

Частота появи кожної букви в цьому слові 0,1, за виключенням E, N, D, що зустрічаються двічі. І відповідно ймовірність їхньої появи – 0,2. Далі кожній букві присвоюється інтервал ймовірності довжина якого обчислюється виходячи з ймовірності її появи в слові (табл.1)

Таблиця 1. Інтервали ймовірностей для символів у слові **REDUNDANCE**

Буква	Ймовірність	Інтервал
A	0.1	0.0 – 0.1
C	0.1	0.1 – 0.2
D	0.2	0.2 – 0.4
E	0.2	0.4 – 0.6
N	0.2	0.6 – 0.8
R	0.1	0.8 – 0.9
U	0.1	0.9 – 1.0

Подальшу процедуру арифметичного стиснення пояснимо з допомогою табл.2.

Перша буква R отримує інтервал із нижньою границею 0,8 і верхньою 0,9. Нижня границя інтервалу і буде першим значенням цифрового коду. Потім

виконуємо розрахунок границь під інтервалів для кожної з наступних букв за таким алгоритмом:

Нижня границя нового інтервалу дорівнює нижній границі (для букви R це 0,8) додати добуток попереднього інтервалу (для букви E=0,9-0,8=0,1) і нижньої (а для розрахунку верхньої границі інтервалу - верхньої). Границі інтервалу для букви E (ці значення відповідно дорівнюють 0,4 і 0,6). В результаті послідовність символів REDUNDANCE – 0,8478570048.

Таблиця 2. Покрокове представлення рядка REDUNDANCE методом арифметичного кодування.

Символ	Нижня границя	Верхня границя
R	0.8	0.9
E	0.84	0.86
D	0.844	0.848
U	0.8476	0.848
N	0.84784	0.84792
D	0.847856	0.847872
A	0.847856	0.8478576
N	0.84785696	0.84785728
C	0.847856992	0.847857024
E	0.8478570048	0.8478570112

$$E = 0.9 - 0.8 = 0.1 * 0.4 = 0.8 + 0.04 = 0.84$$

$$E = 0.6 * 0.1 = 0.06 + 0.8 = 0.86$$

Таким чином кінцева довжина дорівнює добутку ймовірностей усіх символів, а його початок залежить від порядку слідування символів у потоці. Арифметичне кодування дає високу степінь стиснення даних особливо у тих випадках, коли частота появи різних символів істотно відрізняється в той же час, а сама процедура арифметичного кодування потребує потужних обчислювальних ресурсів і мало застосовується.

Словникові методи стиснення даних

Вхідну послідовність символів можна розглядати як послідовність рядків, що містять довільну кількість символів. Ідея словникових методів стиснення даних полягає в заміні рядків символів на коди, які можна інтерпретувати (трактувати) як індекс рядків деякого словника.

Словник – це набір фраз, які будуть зустрічатися в послідовності, що обробляються. Індеси фраз повинні бути побудовані таким чином щоб їхнє представлення мало менший об’єм ніж рядок, який ми ними замінюємо. За рахунок цього і відбувається стиснення.

Класичні алгоритми

Алгоритм Зіва – Лемпеля

Алгоритми словникового стиснення Зіва – Лемпеля з’явився у другій половині 70 –их років. Це є алгоритми – LZ 77 і LZ 78.

LZ 77 – 78 є універсальними алгоритмами, в яких словник формується на основі вже обробленої частини вхідного потоку, тобто адаптивним. Принциповою відмінністю є лише спосіб формування фраз. Публікація Зіва -Лемпеля мали теоретичний характер. Вони займались проблемою вимірювання складності рядка і застосування їх до стиснення є лише її частковим результатом. Практично всі використовувані словникові алгоритми належать до сімейства LZ .

Алгоритм LZ 77 – є найстарішим серед методів Зіва – Лемпеля. Цей алгоритм є родоначальником цілої групи алгоритмів з ковзаючим вікном або ковзаючим словником. В LZ 77 в якості словника використовують блок вже закодованої послідовності по мірі обробки вхідного потоку, розміщені цього блоку відносно початку послідовностей постійно змінюваного словника, “ковзаючого” по вхідному потоку даних.

Ковзаюче вікно має довжину N – символів і складається з двох частин:

1. Послідовність довжин $W=N-n$ вже закодованих символів, яка є словником.
2. Буфер попереднього перегляду, що має довжину n (на порядок менше ніж N).

Ідея методу LZ 77 полягає в пошуку самого довгого співпадіння між рядком буфера і усіма фразами словника, ці фрази можуть виходити за область словника і частково захоплювати область буфера, але повинні знаходитись в межах вікна, отримана в результаті пошуку фраза кодується двома числами:

1. Зміщення відносно початку буфера і.
2. Довжина співпадіння j.

Зміщення і довжина співпадіння відіграють роль вказівника. додатково у вихідний потік записується символ S, який слідує за фразою, що співпадає з словником.

у_бобра_добра_багато

крок	Ковзаюче вікно		Фраза, що співпадає	Закодовані дані		
	Словник	Буфер		i	j	S
1	_____	у_бобра_	_____	1	0	“у”
2	у	_бобра_д	_____	1	0	“_”
3	у_	бобра_до	_____	1	0	“б”
4	у_б	обра_доб	_____	1	0	“о”
5	у-бо	бра_добр	“б”	2	1	“р”
6	у_бобр	а_добра_	_____	1	0	“а”
7	у_бобра	_добра_б	_____	6	1	“д”
8	у_бобра_д	обра_ба	“обра_”	6	5	“б”
9	у-бобра_добра-б	агато	“а”	3	1	“г”
10	у_бобра_добра_баг	ато	“а”	2	1	“г”
11	у_бобра_добра_багато	о	“о”	10	1	_____

$$5+3+8=16$$

$$16*11=176$$

Алгоритм LZ78

Алгоритм цієї групи не використовує ковзаючого вікна і в словник записує не усі рядки, які зустрічаються, на кожному кроці в словник записується нова фраза, що представляє собою об'єднання фрази із словника (що має найбільше співпадіння з рядком буфера) і наступний символ S. На відміну від LZ77 в словнику не може бути одної фрази, коди генерують послідовність кодів фраз, кожен код складається з індекса n фрази у словнику і символу S.

у_бобра_добра_багато;

n – номер початкової фрази чи символу;

S – символ.

№	Фраза, що додається у словник		Буфер	Фраза співпадіння	Закодовані дані	
	фраза	її номер			n	S
1	у	2	у_бобра_	—	1	“у”
2	_	3	_бобра_д	—	1	“_”
3	б	4	бобра_до	—	1	“б”
4	о	5	обра_доб	—	1	“о”
5	бр	6	бра_добр	“б”	4	“р”

Нульове зміщення означає кінець кодування. Одиничка задає порожню фразу словника.

РОЗДІЛ 2. СПЕЦІАЛІЗОВАНІ МЕТОДИ СТИСНЕННЯ ДАНИХ

Кодування довжин повторів

Це один із найпростіших алгоритмів стиснення без втрат. Нехай ми маємо ланцюг однакових символів, при цьому існує явна надлишковість стиснення здійснених наступним чином:

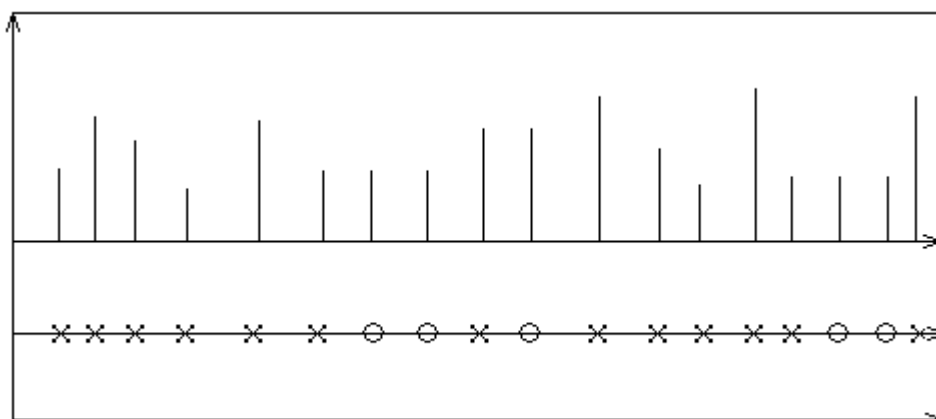
1. Обчислюється кількість однакових символів у ланцюгу;
2. Замість ланцюга символів записуємо символ і скільки разів він повторюється.

Приклад: Рядок із 40 пропусків, записуємо байт-код, який вказує на те, що йде символ, який повторюється потім число 40 (кількість повторів) і нарешті пропуск (символ, що повторюється) рядок, довжиною що стискається до трьох байт, даний алгоритм використовується для стиснення зображення, зокрема ділової графіки.

Методи стиснення технологічних даних

Адаптивні методи стиснення даних

Адаптивні методи стиснення базуються на аналізі стану об'єкта керування, процедура кодування полягає у визначенні активних і не активних відліків, що можна показати на прикладі.



Якщо виконується умова $x_i = x_{i-1}$, то даний відлік не активний. Якщо $x_i \neq x_{i-1}$, то це активний відлік і позначається x_i^*

Ефект стиснення даних досягається за рахунок кодування тільки активних відліків і їхніх номерів. розрядність коду для самих відліків визначаються за формулою $n = E[\log_2 a]$, де n розрядність відліків для представлення двійкового коду x_i , a - діапазон квантування, E – цілочисельна функція для заокруглення до більшого цілого.

В результаті такого адаптивного кодування формується наступний потік даних $\{x_i^*\}$, де $i=1 \dots m$, m – загальна кількість відліків x_i .

Для визначення розрядності коду номеру активного відліка використовують формулу $L = E[\log_2 m]$.

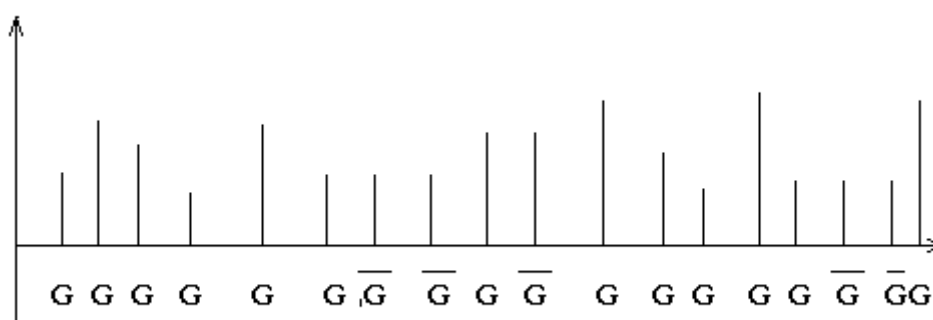
Сумарна розрядність, що реєструється визначається f . Для різних процесів число активних відліків f може змінюватися в діапазоні від 1 до m . Тоді коефіцієнт стиснення визначається за формулою:

$$k_c = m \cdot n / (1 + n) \cdot f$$

f – число активних відліків.

Можливий інший варіант адаптивного кодування з використанням рекурентних послідовностей Галуа. Суть методу полягає в наступному: для кодування номерів активного чи не активного відліку використовують послідовність Галуа.

Кожному активному відліку відповідає прямий ввід Галуа, не активному – інвертований.



$$k = m \cdot n / m(f) \cdot n,$$

де m – загальна кількість відліків, n – розрядність для представлення значення відліку, f – кількість активних відліків.

При умові, що заздалегідь відомо, що даний об'єкт контролю на протязі тривалого часу не змінює свого стану тобто, що ми маємо достатньо довгі послідовності неактивних відліків доцільно використовувати одну із модифікацій даного методу, суть якої полягає у використанні буферу для бітів Галуа на передавальній стороні з приходом кожного інвертованого біту Галуа вміст буферу змінюється (послідовність ніби зсувається вправо). У разі слідування прямого біту Галуа, що свідчить про активний відлік вміст буферу передається у канал зв'язку, а за ним змінюється саме значення відліку.

При використанні даного методу коефіцієнт стиснення визначається тільки динамікою зміни значень об'єкту контролю.

Однак при кодуванні технологічних даних досить часто постає проблема кодування стану технологічної установки в цілому, а не окремих параметрів, як розглядалось раніше.

Базисні функції Галуа нульового порядку

Назва	Кодон	Послідовність Галуа
G_0	0000	000010110010111
G_1	1000	100001010011011
G_2	1100	110001010011011
G_3	1110	111001010011011

Методи стиснення звуку

Оцифрований звук (мова) погано стискається без втрат. Очевидно, що є потреба застосовувати алгоритми стиснення з втратами. Найбільшого поширення набули застосування :

1. Замість лінійної ІКМ логарифмічної ІКМ (з компандуванням по А- і μ -законам), адаптивної ІКМ, диференціальної ІКМ, адаптивно-диференціальної ІКМ, блочної ІКМ і адаптивної дельта модуляції.

2. Кліпірування мовного сигналу

3. Статистичне ущільнення

4. Застосування вокодерів

Також використовуються алгоритми стиснення без втрат:

1. Стиснення за рахунок пауз в сигналі (паузи видаляються з вхідного сигналу, а їхня тривалість запам'ятовується).

2. При зберіганні і передачі музики замість оцифрованого звука зберігається інформація про те, які ноти коли і на якому інструменті виконуються. Всі вище згадані алгоритми можна реалізувати як апаратно так і програмно. Апаратно реалізуються варіанти перетворення аналогового сигналу в цифровий, застосовуються ІКМ і спорідненні з нею методи: АІКМ, ДІКМ, АДІКМ, БІКМ.

ІКМ і пов'язані з нею методи стиснення інформації

Опис схеми:

Відліки n – розрядного АЦП розбиваються на блоки по N – відліків. Кожному блоку знаходять відлік з максимальним для даного блоку рівнем. Даному рівню відповідає певний номер старшого значущого розряду і всі старші розряди цього блоку будуть нульовими. Записаний у двійковому коді номер цього розряду утворює масштабну інформацію, котра є дуже важливою і як правило захищається заводозахисним кодом. В результаті масштабна інформація утворює разом з перевірочними символами m – значну комбінацію, яку додають до основної інформації.

Серед різноманітних систем ІКМ найбільш поширеною є система БІКМ.

n										j									
0	0	0	0	0	0	0	0	1	1	0	1	0	x	x	x	x	x	x
0	0	0	0	0	1	1	1	0	1	0	1	1	x	x	x	x	x	x
0	0	0	0	0	1	0	1	1	0	1	0	1	x	x	x	x	x	x
.....																			
0	0	0	0	0	0	0	1	0	0	1	0	1	x	x	x	x	x	x
0	0	0	0	0	0	1	0	0	0	1	1	0	x	x	x	x	x	x

k- розрядів відібраних для передавання

Основна інформація формується вибором k – розрядів з n- початкових. При цьому старшим розрядом є розряд з номером описаним у масштабі інформації. Основна інформація для кожного із блоків об’єднується з масштабами в єдиний цифровий потік. Результиуюча швидкість цифрового потоку на виході системи БІКМ дорівнює:

$$R = f_g(k+m/N), \text{ де } f_g \text{ – частота дискретизації.}$$

Використовуються наступні параметри:

$$f_g=8 \text{ кГц, } n=10 - 13, \text{ } k=6 - 8, \text{ } N=8 - 16, \text{ } m=6-8.$$

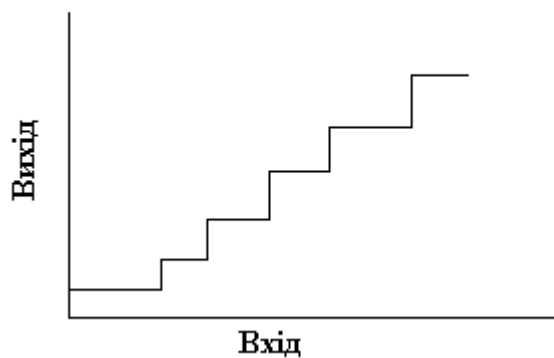
При однакових умовах передачі блочна ІКМ дає кращу якість ніж ІКМ, тому можна знизити швидкість передачі до 32 – 56 кбіт/с. При ДІКМ замість відліків кодуються різниці, які як правило менші ніж самі відліки. Швидкість передачі цифрового потоку знижується до 48 – 56 кбіт/с.

В системах з ЛІКМ використовуються А- і μ - закони компандування для реалізації нерівномірного квантування. Людська мова характеризується унікальними статистичними властивостями.

Для більшості каналів мовного зв’язку домінують дуже низькі тони: 50% часу напруга, що характеризує енергію виявленої мови складає менше 1/4 середньо - квадратичного значення. Значення з великими амплітудами зустрічаються

відносно рідко. Тільки 15% часу напруга перевищує середньоквадратичне значення. В системах, що використовують рівномірне квантування. При такому кодуванні відношення сигнал – шум буде гіршим для сигналів низьких рівнів ніж для сигналів високих рівнів. Нерівномірне квантування слабких сигналів може забезпечити краще квантування слабких сигналів і гірше сильних сигналів. В такому випадку шум квантування пропорційний сигналу. Результатом є збільшення загального відношення сигнал – шум – зменшення шуму для переважаючих слабких сигналів за рахунок збільшення шуму для сильних сигналів, які зустрічаються рідко.

В стандартному телефонному зв'язку для обробки великого діапазону можливих вхідних рівнів сигналу використовується не звичайний пристрій з рівномірним квантуванням, а пристрій з логарифмічним стисненням. При цьому відношення сигнал – шум на виході не залежить від розподілу рівнів сигналу на вході. Одним із шляхів нерівномірного квантування є використання пристрою, який має характеристику зображену на рис.1.



Частіше нерівномірне квантування реалізується наступним чином:

- 1) Спочатку вхідний сигнал деформується за допомогою пристрою, що має логарифмічну характеристику стиснення (Рис.2) , а потім використовується пристрій квантування з рівномірним кроком.



Характеристика стиснення ефективно змінює розподіл амплітуд вхідного сигналу таким чином, щоб на виході системи стиснення вже не переважали сигнали малих амплітуд. Після прийому сигнал пропускається через пристрій з характеристикою оберненою до зображеної на Рис.2, яка називається розширенням. Таким чином загальна передача не є деформованою. Описана пара станів обробки сигналу (стиснення і розширення) називається компандуванням. В ранніх системах ІКМ функції стиснення були легкими логарифмічними. Більшість використовували кусково – лінійну апроксимацію функції логарифму стиснення. В Північній Америці характеристика пристрою описується наступним законом:

$$y = y_{\max} \ln[1 + \mu|x|/x_{\max}] / \ln[1 + \mu] * \text{sgn } x$$

$$\text{sgn } x = +1, \text{ при } x > 0 \text{ та } -1, \text{ при } x < 0$$

μ – додатня константа

x, y – напруги на вході, виході

x_{\max}, y_{\max} - максимальні додатні амплітуди напруг на вході і на виході.

В Північній Америці стандартним значенням для μ є 255. Слід зазначити, що $\mu=0$ відповідає лінійному підсиленню (рівномірне квантування). В Європі для опису характеристики пристрою стиснення використовують інший закон.

$$y = y_{\max} A(|x|/x_{\max}) / (1 + \ln A) * \text{sgn } x \quad 0 < |x|/x_{\max} < 1/A$$

$$y = y_{\max} (1 + \ln[A(|x|/x_{\max})]) / (1 + \ln A) * \text{sgn } x \quad 1/A < |x|/x_{\max} < 1$$

де A – додатна константа.

Однією важливою відмінністю між характеристиками стиснення A - і μ – законів є те, що стандарт A – закону має характеристику з 0 на границі кроку

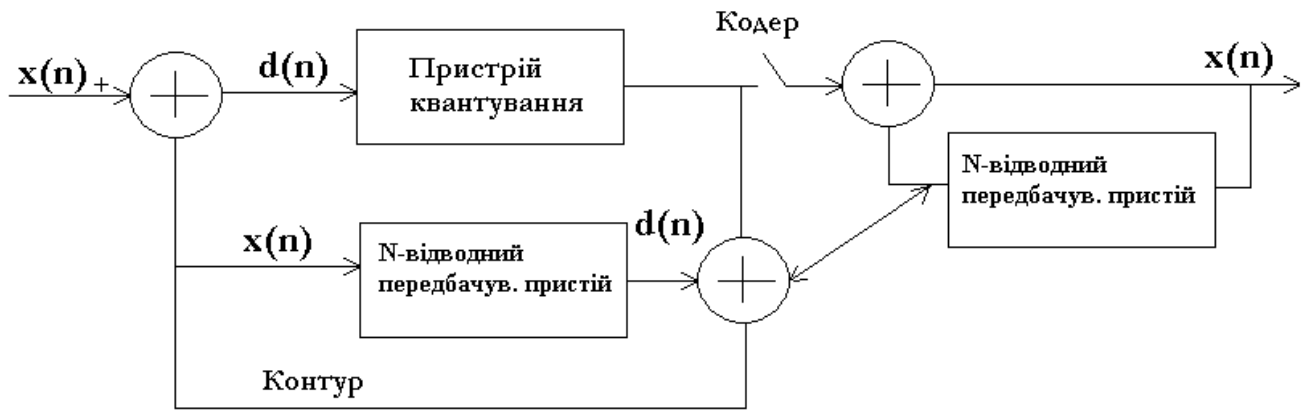
квантування. В той час, як μ – закон характеристику з 0 в центрі кроку квантування.

Диференціальна ІКМ

При ДІКМ передбачення наступного значення вибірки формуються на основі попередніх значень. Для квантуючого пристрою передбачення попереднє значення є інструкцією по пошуку наступних значень вибірки у конкретному інтервалі. Якщо для передбачення використовують надлишковість сигналу, то область невизначеності скорочується і квантування можна проводити із зменшеним числом біт для заданого рівня квантування.

Зменшення надлишковості реалізується шляхом віднімання передбачуваного значення від наступного вибіркового значення. Ця різниця називається помилкою передбачення. Кореляційні характеристики джерела можна подати у часовій області з допомогою вибірки його автокореляційної функції і в частотній області – його спектром потужності.

Повільно змінна автокореляційна функція свідчить про те, що при переході від вибірки до вибірки відбувається тільки не велика зміна, а для повної зміни амплітуди потрібен часовий інтервал, що перевищує інтервал кореляції. Зокрема значення кореляції для типового одиничного вибіркового запізнення лежить в діапазоні від 0,79 до 0,87. А радіус кореляції має порядок від 4 до 6 вибіркових інтервалів. Оскільки різниця між частотними сусідніми вибірками для мови мала, то використовуваний метод базується на передачі від вибірки до вибірки різниць, а не дійсних вибіркових значень. Процес обробки мовного сигналу за вказаною схемою здійснюється N – відведених передбачаючими пристроями, яких називають кодерами з передбаченням і поправками. Структурна схема приведена на рисунку.



n – відвідний диференціальний імпульсний модулятор з передбаченням.

У даному перетворенні передавальний і приймальний пристрій мають однакову модель передбачення отриману з автокореляційної характеристики сигналу. Математичний контур передбачення описується наступним чином:

$$d(n) = x(n) - \hat{x}(n)$$

$x(n)$ – n – вхідна вибірка.

$\hat{x}(n)$ – передбачене значення вибірки.

$d(n)$ – помилка передбачення.

Ця операція відбувається у контурі передбачення порівнянь. Кодер коректує свої передбачення складаючи суму передбачення значення і помилки передбачення. Математичний контур корекції описується так

$$d(n) = \text{quand}(n) \quad \tilde{x}(n) = \hat{x}(n) + \tilde{d}(n).$$

quand – операція квантування.

- $\tilde{x}(n)$ - квантування значень помилки передбачення.

- \hat{x} - квантування інвертності від вибірки.

Ці операції виконуються у контурі передбачення і поправок у нижньому циклі кодера і в єдиному циклі декодера. Декодер повторяє обернений цикл кодера. Перетворювачі з передбаченням повинні мати короткочасну пам'ять, яка підтримує операції в реальному часі необхідні для алгоритму передбачення. Крім того необхідна довготривала пам'ять, яка підтримує автоматичне регулювання підсилення.

Адаптивне передбачення

Ефективність використання кодера з передбаченням обмежується з можливими розбіжностями між сигналами джерел і передбачуваним фільтром, які в першу чергу пов'язані із не стаціонарністю розподілення амплітуди і спектральних або кореляційних властивостей сигналу. Адаптивні кодери містять допоміжні схеми для оцінки параметрів, які необхідно для одержання локальної продуктивності. Ці допоміжні ланки періодично програмують, модифікації для передбачення параметрів і таким чином уникають розбіжності передбачень. Міжнародний консультативний комітет з питань телеграфії і телефонії (МККТТ) обрав адаптивну диференціальну імпульсно-кодову модуляцію зі швидкістю 32 кбіт/с.

Пряма адаптація

В алгоритмах прямої адаптації вхідні дані, які повинні бути закодовані буферизуються і обробляються з метою отримання реальних характеристик, таких як перші N – вибіркового значень автокореляційної функції. Кореляційне значення $R_x(0)$ з нульовим запізненням є короткочасною оцінкою дисперсії. Ця оцінка використовується для узгодження автоматичного регулювання підсилення з метою отримання оптимального узгодження. Масштабування вхідного сигналу з динамічною областю пристрою квантування.

Цей процес позначають:

AQFC – adapter quantization forward control (контроль прямим адаптивним квантуванням).

$(N-1)$ кореляційних оцінок, що залишилось використовувати для отримання нових коефіцієнтів для фільтра з передбаченням. Цей процес називається контролем прямим адаптивним передбаченням APF (adapter protection forward).

На рисунку зображена форма адаптивного алгоритму.

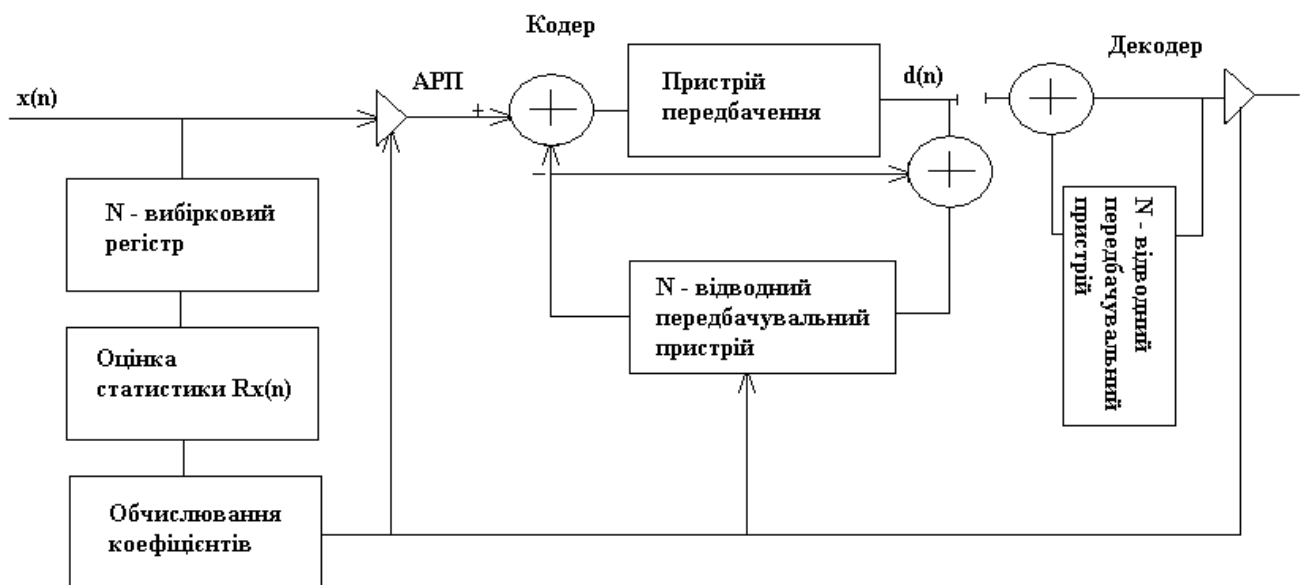


Рис.4. Пряме адаптивне передбачення і кодування – квантування.

АРП – автоматичний регулюючий пристрій передбачає коефіцієнт виводу з вхідних даних, що є побічною ініціалізацією.

Вони повинні передаватися з кодера на декодер разом з помилками передбачення. Швидкість зміни адаптивних коефіцієнтів пов'язана з часом на протязі якого сигнал вважається локально – стаціонарним.

Приклад: Мовний сигнал не може змінювати свої характеристики швидше ніж 10 – 20 раз/с. Це дає інтервал оновлення від 50 до 100 мс. Використовуючи арифметику простих, але субоптимальних алгоритмів для обчислення локальних параметрів вимагає більш швидких змін, для обчислення параметрів 10 – відводного фільтру прийнятий інтервал зміни 20 мс.

Адаптивно – диференціальна імпульсно кодова модуляція (АДІКМ)

МККТТ обрав стандартом якісного телефонного зв'язку АДІКМ G726. Цей стандарт кодує послідовно вибірки, передбачаючи значення кожної вибірки з відновленим значенням мовного сигналу попередньої вибірки з використанням адаптивного пристрою зі зворотнім зв'язком. Даний пристрій приймає якісний мовний сигнал, претворений з допомогою 8 – бітового лінійного перетворення з використанням А або μ закону зі швидкістю 64 кбіт/с. Видає стиснутий мовний сигнал зі швидкістю 16, 24, 32, 40 кбіт/с.

Кодер використовує декодер у контурі зворотнього зв'язку для аналізу і модифікації параметрів алгоритмів з метою мінімізації помилки відновлення. Передбачаючий пристрій використовує фільтр 6 –порядку для моделювання нулів. Блочна діаграма кодера зображена на рис.5

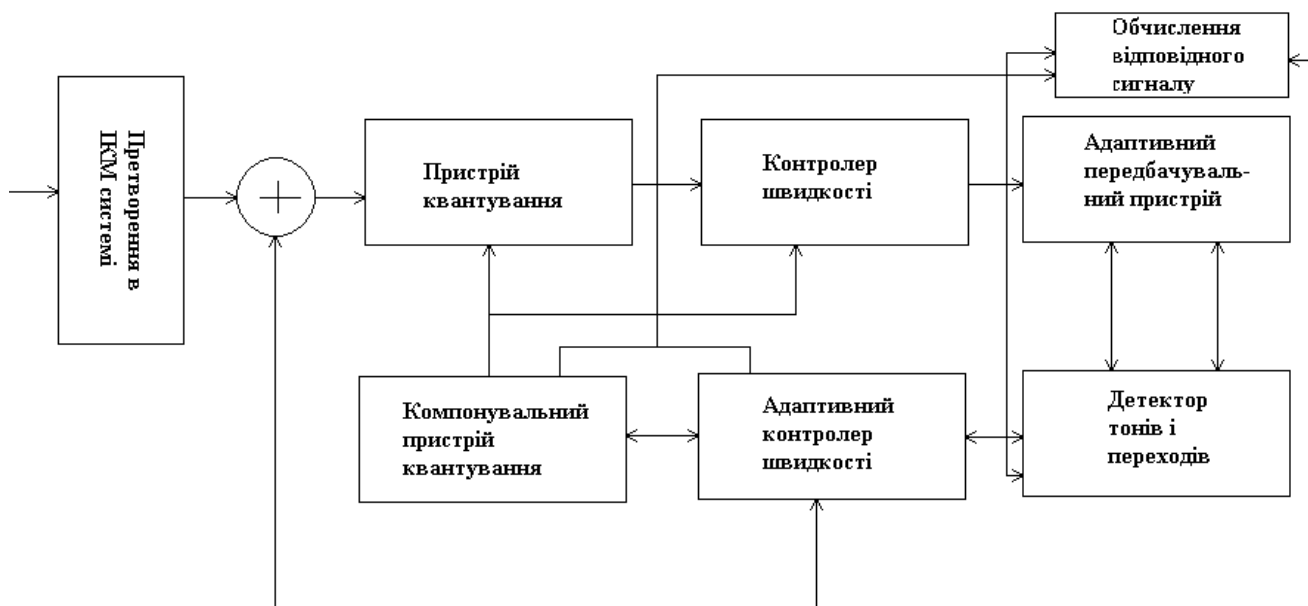


Рис.5. Мовний кодер АДІКМ (стандарт G726)

Аудіостиснення

Аудіостиснення широко застосовують у звичайних та професійних аудіопродуктах, зокрема у компактах CD, цифрових аудіолентах DAT, міні дисках MD, цифрових компакт касетах DCC, універсальних цифрових дисках DVD, цифровому аудіомовленню DAB, аудіопродукції в форматі MP3 від експертої групи з питань рухомого зображення MPEG. Крім того в стисненні телефонних, зокрема в стільникових, необхідно для економії смуги частот і продовження дії батареї, що і заклало дії основи великої кількості стандартів стиснення даних. Аудіо і мовні схеми стиснення умовно можна поділити за такими параметрами, як діапазон частот, частота дискретизації, швидкість передачі.

Типові значення параметрів для трьох класів аудіосигналів приведені у таблиці.

Клас сигналу	Діапазон частот	Частота дискретизації	Біт ІКМ на вибір	Швидкість передачі
1. телефонна розмова	300-3400 Гц	8 кГц	8	64 кбіт/с
2. широко-смугова мова	60-7000 Гц	10 кГц	14	224 кбіт/с
3. широкосмугове аудіо	10000-20000 Гц	48 кГц	16	778 кбіт/с

Лінійне кодування з передбаченням

Адаптивні передбачувальні пристрої були створені для передбачення або створення оцінок вхідного сигналу. В адаптивній формі коефіцієнт обчислюється як побічна інформація на основі періодичного вивчення вхідних даних. Потім різниця між вхідними даними і передбаченням передається на приймальну сторону. Лінійні кодери з передбаченням (LPC) є розширенням N – відводних кодерів з передбаченням.

Якщо коефіцієнт фільтра періодично обчислюється за оптимальним алгоритмом то передбачення є настільки точним, що інформація про помилку, яку необхідно передавати приймачу не існує. Замість того, щоб передавати ці помилки передбачувальна система LPC передає коефіцієнт фільтра. Отже, даними, якими передаються в LPC є високочастотна побічна інформація класичного адаптивного алгоритму. Модель LPC для синтезу голосу зображена на Рис. 1.

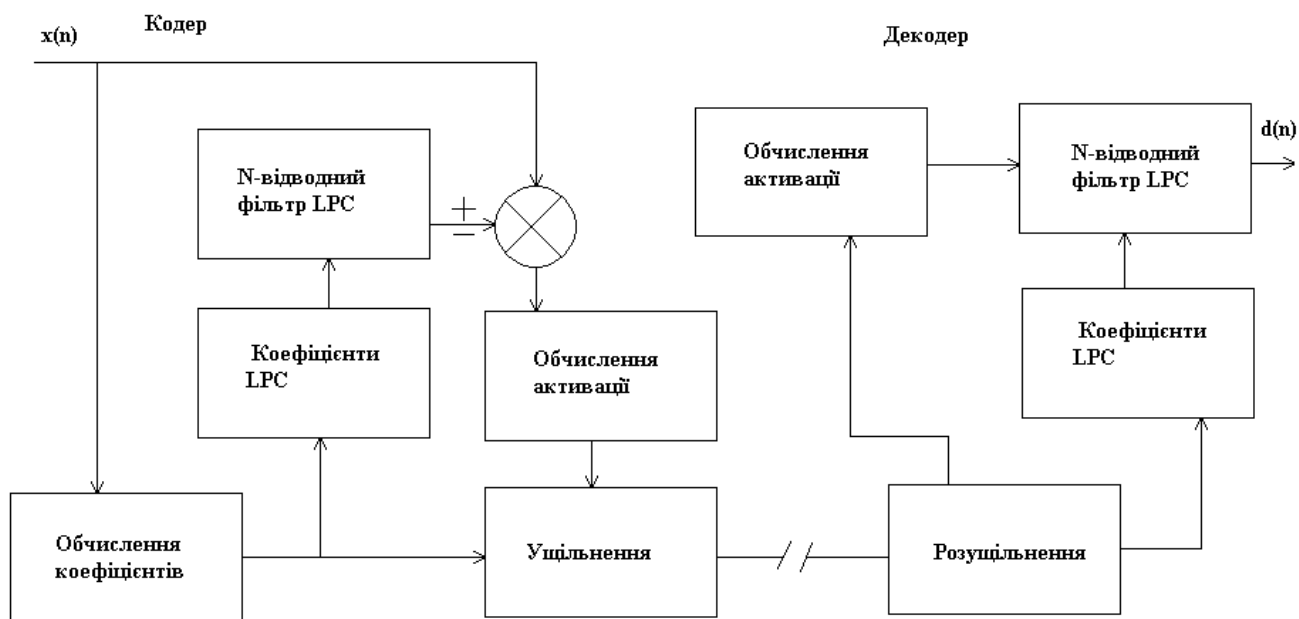


Рис. 1. Блочна діаграма: моделювання мови за допомогою кодера лінійного передбачення.

Побічна інформація: коефіцієнт LPC і параметри активації кожних 20 мс.

Кодери LPC являють собою ядро із змішаних кодерів, яке містить кодер і керуючий генератор в контурі аналізу через синтез, призначений для мінімізації різниці між вхідними і синтезованими сигналами. У сотових телефонах для отримання якісного зв'язку зі швидкістю передачі даних 9,6 кбіт/с використовують кодери PRE (активація регулярними імпульсами). Використовуються кодери CELP (лінійне передбачення, що активується кодовою книгою).

Система GSM (глобальна система мобільного зв'язку) використовує стиснення RPE в той час, як для мобільних систем створюється згідно стандартів IS-95 відносно багаточисельного доступу з кодовим розділенням каналів (CDMA) використовується варіант CELP.

Схема CELP

Мовні кодери, що використовують лінійні фільтри з передбаченням можуть дати високу якість мови закодованої зі швидкістю більше 16 кбіт/с, але із зменшенням швидкості якість погіршується.

Кодери LPC можуть бути модифіковані з метою отримання високоякісного стиснення мови зі швидкістю від 4,8-9,6 кбіт/с. За рахунок приведення задачі синтезу до двох етапної процедури, яка називається синтез через аналіз. На першому етапі утворюється модель LPC 10 порядку. Для сигналу, дійсного на протязі короткого інтервалу, наприклад, кожні 20 мс.

На другому етапі знаходиться хвильовий сигнал, який при застосуванні до моделі LPC утворює вихідний сигнал по можливості близький до вхідного синтезованого сигналу. Закінчується ця задача за допомогою послідовності застосувань підходящого сигналу активації до моделі і порівнянні кожної синтезованої форми сигналу з вхідним сигналом. З подальшим вибором того, який мінімізує помилку між вхідним сигналом і виходом керованої моделі.

З теорії процесу формування мови відомо, що активація мови часто складається з періодичних імпульсів (утворюється за рахунок вібрації голосових зв'язків). Період періодичних імпульсів P пов'язаний з голосом людини, одновідвідний рекурсивний фільтр описується двома параметрами:

1. P – число інтервалу запізнення в контурі зворотного зв'язку.

2. G – коефіцієнт зворотного зв'язку.

Імпульсна характеристика цього фільтра – це затухаюча послідовність з P рівними вибірками між послідовностями не нульових вибірок. Вихід цього фільтра генерує періодичний сигнал активації, який подається на вхід моделі ALPC.

Алгоритм синтезу повинен перевіряти можливі значення P з переліку тих, що підходять. Два параметри голосу оцінюються кожних 5 мс. Вхід в умовний фільтр вибирається з таблиці підходящих послідовностей активацій. Вихід фільтра в свою чергу керує моделлю LPC.

Таблиця, яка як правило містить 1024 позиції називається кодовою книгою. Кодову книгу використовують кожні 2,5 мс. Коли найкраща комбінація позицій кодової книги і період голосу визначається за допомогою пошуку,

формується група, яка містить послідовні параметри голосу, послідовний адрес кодової книги і інформацію про коефіцієнт LPC.

Кодер повинен подати параметри, що описують модель LPC на декодер. Спектральна характеристика фільтра LPC дуже чутлива до квантування коефіцієнта і повинна подаватись за допомогою не прийняття великої кількості біт. Тому коефіцієнт LPC перетворюється на іншу множину параметрів, які називаються лінійними спектральними параметрами, що є не чутливими до квантування. Системи створюються згідно стандарту IS – 95 використовують наступний формат кадру LPC. Кодер, що потрібен для опису 2 мс даних містить 192 байти, що присвоюються закодованим параметрам.

Загальна швидкість передачі для цієї системи складає 192 біти за 20 мс або 9600 біт/с. Швидкість передачі може бути зменшена, якщо кодер виявляє мовну паузу.

Інші формати стиснення звуку CODEC ADPCM. Даний формат використовує простий алгоритм стиснення, який забезпечує високу швидкість кодування на слабких комп'ютерах, але не придатний для зберігання музики.

MPEG

Даний алгоритм за принципом подібний до MP2, типовий бітрейд лежить в діапазоні 160...180 кбіт/с, швидкість роботи кодера досить висока.

Якість файлів при створенні МС3 перевищує якість аналогічних файлів MP3. Одним із серйозних недоліків МС3 є обмеження на формат файлу: 44 кГц, стерео і це робить його не придатним для стиснення звукових доріжок до фільмів на DVD, а також відсутність апаратної підтримки.

AAC – цей формат має ширші можливості ніж MP3. Це можливість кодування декількох звукових каналів з частотою дискретизації до 96 кГц. Краще співвідношення якості і розміру.

Використання AAC призводить до недоліків:

1. алгоритми кодування, які використовуються в даному форматі досить складні.

2. значний час, системні ресурси.

WMA9 поступова підтримка до 166 кбіт/с, якість звучання на рівні MP3

Формат WMA9, що підтримує кодування зі змінним бітрейдом і також стиснення без втрат.

OGG Vorbis файлів використовує кодування в режимі змінної бітової швидкості, що дозволяє досягти кращого співвідношення, якості, розміру за рахунок оптимального вибору бітрейду в залежності від характеру музики, хоча можливе використання постійного бітрейду. Інформація про виконавця, назва композиції зберігається в унікодї, що дозволяє усунути проблеми із символами. За якістю звучання OGG мають перевагу над MP3.

Методи стиснення зображень

Зображення – це своєрідний тип даних, який характеризується трьома особливостями:

1. Зображення вимагає для зберігання значних ресурсів пам'яті.
2. Людський зір при аналізі зображення оперує контурами, загальним переходом контурів і є порівняно не чутливим до малих змін зображення.
3. Зображення є надлишковим у двох вимірах.

Алгоритми стиснення зображень практично не можливо застосувати до архівації інших видів даних.

Кожному пікселю відповідає значення, яке задає його колір. У більшості кольорових моніторів використовуються три кольори: червоний, зелений, синій.

Найбільш поширена кольорова модель RGB. Але в форматі JPEG використовується модель YcbCr.

Y- яскравість;

Cb – колірність (синій);

Cr – складова червоного кольору.

Взаємозв'язок між колірними моделями RGB та YcbCr

$$Y=0.299R+0.58G+0.114B;$$

$$Cb=-0.1687R-0.3313G+0.5B+2^{\text{точність дискретизації}/2};$$

$$Cr=0.5R-0.4187G-0.0813B+2^{\text{точність дискретизації}/2};$$

$$R=Y+1.402Cr;$$

$$G=Y-0.344+4(Cb-2^{\text{точність дискретизації}/2})-0.71414(Cr-2^{\text{точність дискретизації}/2});$$

$$B=Y+1.722(Cb-2^{\text{точність дискретизації}/2});$$

Ще одна колірна модель CMYK (блакитний, пурпуровий, жовтий, чорний) часто використовується в кольоровому блоці.

Взаємозв'язки між CMYK:

$$K=(2^{\text{точність дискретизації}/2} - 1) - \text{MAX}(R,G,B);$$

$$C=(2^{\text{точність дискретизації}/2} - 1)-R-K;$$

$$Y=(2^{\text{точність дискретизації}/2} - 1)-G-K;$$

$$M=(2^{\text{точність дискретизації}/2} - 1)-B-K.$$

Зворотне:

$$R=(2^{\text{точність дискретизації}/2} - 1) \cdot K \cdot C;$$

$$G=(2^{\text{точність дискретизації}/2} - 1) \cdot K \cdot Y;$$

$$B=(2^{\text{точність дискретизації}/2} - 1) \cdot K \cdot M.$$

Клас зображень – це сукупність зображень застосування до яких алгоритму архівації дає якісно однакові результати.

1. Зображення з невеликою кількістю кольорів (4 –16) і великими областями заповнення одним кольором.
2. Зображення із плавними переходами кольорів.
3. Контури реалістичного зображення (фото).
4. Фотореалістичні зображення з ділової графіки (реклама).

Методи стиснення, що використовуються різними форматами

Методи	BMP	GIF	PNG	JPEG
RLC	X			X
LZ		X	X	
Хафман			X	X
DCT				X

Windows BMP розробник Microsoft є вбудованим форматом зображення в ОС Microsoft Windows. Підтримує зображення з 1, 4, 8, 16, 24, 32 біт/піксел. Стиснення форматів BMP дає ефект тільки при наявності зображень областей однакового кольору. BMP рідко знаходиться у стиснутому вигляді. Структура файлів: <заголовок файлу><заголовок зображення><таблиця кольорів><піксельні дані>.

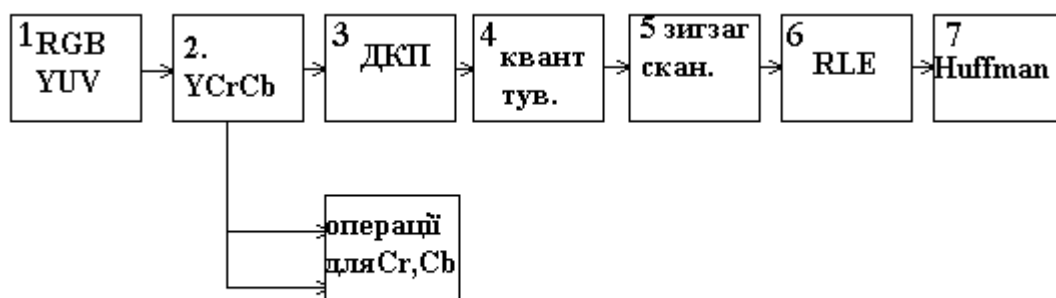
GIF (засновник COMPUSERE) – формат графічного обміну. Формат GIF і був найпоширенішим форматом, основні характеристики: підтримка 256 кольорів для опису яких використовується від 1 до 8 біт/піксел. Зберігає в одному файлі велику кількість зображення. Для фотографічних зображень на зміну GIF прийшов JPEG завдяки кращому стисненню і глибині кольорів. Формат GIF продовжив би використання інших класів зображень, але цьому заважали юридичні проблеми.

PNG (група розробників PNG) – найновіший формат (1996 р.). Формат PNG застосовує процес стиснення без втрат і підтримує наступні можливості:

1. Для 48 біт/піксел для кольорового зображення.
2. 1, 2, 4, 8, 16 бітова точність дискретизації.
3. α – канал для повного керування прозорістю.
4. Досконала колірنا відповідність.

У випадках, коли треба забезпечити стиснення без втрат зображення з 24 бітовою глибиною кольору

JPEG



1. Переводиться зображення з колірного простору RGB у колірний простір YUV.

2. Дискретизація матриці розбиває вхідні зображення на матриці 8*8. Формуємо три по матриці дискретизації косинки перетворення по 8 біт для кожного компонента. При великих степенях стиснення цей крок виконується складніше. Зображення по компоненті Y ділиться як і в першому випадку, а для компоненти Cr і Cb матриці набираються через рядок і через стовпчик. При цьому ми втрачаємо частину корисної інформації про кольорові складові.

3. ДКП в спрощеному вигляді при $n=8$ можна подати таким чином:
 $Y[u;x]*1/2$

Застосовується дискретна косинка перетворення для кожної матриці при цьому отримаємо матрицю, в якій коефіцієнти у верхньому лівому кутку відповідають низькочастотній складовій зображення, а в правому нижньому – високочастотні.

4. Квантування – це ділення робочої матриці на матриці квантування поелементно. Для кожної компоненти Y, Cr, Cb задається своя матриця квантування $q(u,v)_Y$

На даному етапі відбувається управління ступенем стиснення. Якщо задати матрицю квантування з великим коефіцієнтом, то в результаті отримаємо більше нулів і відповідно більший ступінь стиснення. Стандарт JPEG включає рекомендовані матриці квантування, які побудовані дослідним шляхом. Матриця для більшого чи меншого ступеня стиснення шляхом множення вихідної матриці на деяке число γ . З квантуванням пов'язано специфічні ефекти алгоритму:

а) При великих значеннях γ втрати в низьких частот можуть бути настільки великі, що зображення розпадається на квадрати 8×8 .

б) Втрати в високих частотах проявляються в ефектах Гібса, коли навколо контурів з різким переходом кольору утворюється своєрідний ореол.

5. Перетворивши матрицю 8×8 в 64 елементний вектор за допомогою зигзаг сканування. В результаті отримаємо спочатку низькі частоти, а в кінці високі.

6. Застосування кодувань довжин повторів до отриманих послідовностей після чого отримаємо пари <пропустити, число> де <пропустити> лічильник нулів, що пропускаються, а <число> наступне значення.

7. Згортаємо отримані пари кодування методом Хафмана з фіксованою таблицею. Переваги:

1. Задається ступінь стиснення;
2. Вихідне кольорове зображення має 24 біт/піксел.

Недоліки:

1. При підвищенні ступеня стиснення зображення розпадається на квадрати 8×8 .

3. Проявляється ефект Гібса.

ОСНОВНА ЛІТЕРАТУРА

1. Ватолин Д., Ратушняк А., Смирнов М., Юкин В. Методи сжатия данных. Устройство архиваторов, сжатие изображений и видео. – М.: Диалог-Мифи, 2002. –384 с.
2. Дж. Миано. Форматы и алгоритмы сжатия изображений в действии. Учеб. Пособ. – М.: Издательство Триумф, 2003 – 336 с.: ил.
3. Орищенко В.И. Санников. В.А., Свириденко В.А. Сжатие данных в системах сбора и передачи информации / Под ред. В.А. Свириденко. – М.: Радио и связь, 1985. –184 с.
4. Тутевич В.Н. Телемеханика: Учеб. пособие для студентов вузов спец. “Автоматика и телемеханика” - 2-е изд.; перераб. и доп.- М.: Высшая школа, 1985.- 423 с.
5. Сапожков М.А., Михайлов В.Г. Вокодерная связь. – М: Радио и связь, - 1983. – 256 с.
6. Яцків Н.Г. Методи стиснення даних в інформаційно-керуючих системах // Розвідка і розробка нафтових і газових родовищ. Серія: Технічна кібернетика та електрифікація об’єктів паливно-енергетичного комплексу. – Івано-Франківськ: ІФДТУНГ. – 2001. – №37 (том 6). – С.183-186.
7. Николайчук Я.М., Яцків Н.Г. Методи стиснення даних в багатоканальних системах на основі кодів Галуа // Вісник національного університету “Львівська політехніка”. Радіoeлектроніка та телекомунікації. – Львів. – 2002. – №443.– С.135-138.
8. Яцків Н.Г., Николайчук Я.М. Системні характеристики джерел інформації та оцінка їх ентропії // Вісник Технологічного університету Поділля. Технічні науки. – Хмельницький. – 2002. – №3 (том 1).– 197-200.

ДОДАТКОВА ЛІТЕРАТУРА

1. Вариченко Л.В., Лабунец В. Г., Раков М. А. / Абстрактные алгебраические системы и цифровая обработка сигналов – К.: Наук. думка, 1986. – 248 с.
2. Акушский И.Я., Юдицкий Д.И. Машинная арифметика в остаточных классах. - М.: Сов. радио. - 1968. – 460 с.
3. Торгашев В.А. Система остаточных классов и надёжность ЦВМ. – М.: Сов. радио. - 1973.
4. Вокодерная телефония. Методы и проблемы. /Под ред. А.А. Пирогова. - М: Связь, 1974.
5. Яцків Н.Г. Дослідження системних характеристик методів формування даних в різних теоретико-числових базисах. Вимірювальна та обчислювальна техніка в технологічних процесах: Збірник наукових праць. – Хмельницький: ТУП. – 2002. – №9 (том 2). – С. 132-136.