УДК 621.3(045)

А. М. Негода, асп.

АНАЛІЗ СУЧАСНИХ АЛГОРИТМІВ СТИСНЕННЯ ВІДЕО

Інститут аерокосмічних систем управління НАУ, e-mail: negoda@nevo.com.ua

Проаналізовано алгоритми стиснення зображень для оцінювання їх можливостей стиснення потокового відео.

Ключові слова: алгоритми стиснення, дискретне косинусне перетворення, дискретне вейвлет-перетворення, цифрове відео, цифрове фото.

Вступ. Сучасні системи передавання і зберігання цифрових фотографій, цифрового відео та інші використовують алгоритми стиснення зображень.

Стиснення даних є надзвичайно важливим і актуальним практичним завданням у зв'язку з інтенсивним розвитком комп'ютерних засобів комунікації. На сьогодні це одне з надзвичайно важливих завдань для науковців усього світу, оскільки функціонування супутникових систем цифрового телебачення, цифрових фото- і відеокамер, відеотелефонів, інтернет-систем відеоконференцій неможливе без стиснення відеозображень.

Відео – це по суті тривимірний масив кольорових пікселів. Два виміри означають вертикальний і горизонтальний розміри кадру, а третій вимір – момент часу. Кадр – це масив усіх пікселів, які розміщені перед камерою в певний момент часу, або просто зображення.

Стиснення відео було б неможливе, якби кожен кадр був унікальним, а розміщення пікселів – повністю випадковим, але це не так. Тому можна стискати, по-перше, саме зображення – наприклад, фотографія блакитного неба без сонця фактично зводиться до опису граничних точок і градієнта заливки. По-друге, можна стискати схожі сусідні кадри. Зрештою алгоритми стиснення картинок і відео схожі, якщо розглядати відео як тривимірне зображення з часом як третю координату.

Аналіз останніх досліджень і публікацій. Стиснення відео ґрунтується на двох важливих принципах [1, с. 287]. Перший – це просторова надлишковість, притаманна кожному кадру відеоряду. Другий принцип заснований на тому факті, що більшу частину часу кожен кадр подібний до свого попередника. Це називають часовою збитковістю. Таким чином, типовий метод стиснення відео розпочинається з кодування першого кадру з допомогою декількох алгоритмів стиснення зображення. Потім кодується кожний наступний кадр, знаходяться розбіжності між цим кадром і його попередником і кодується ця розбіжність. Якщо новий кадр сильно відрізняється від попереднього, то його можна кодувати незалежним чином.

Стиснення відео являє собою зменшення кількості даних, які використовуються для подання відеопотоку. Стиснення відео дозволяє ефективно зменшувати потік даних, необхідний для передачі відео по каналах радіомовлення, зменшувати простір, необхідний для зберігання даних на CD, DVD або жорстких дисках. Недоліки: якщо стиснення відбувається із втратами, виникають характерні і добре помітні артефакти – наприклад, блочність (розбиття зображення на блоки 8×8 пікселів), замилювання (втрата дрібних деталей зображення) і т. ін. Існують і способи стиснення відео без втрат, але натепер вони зменшують дані недостатньо.

У разі використання стиснення без утрат за декомпресії результат точно (біт до біта) відповідатиме оригіналу. Проте стисненням без утрат неможливо досягти високих коефіцієнтів стиснення на реальному (не штучному) відео. З цієї причини майже все відео є стисненим з втратами.

Постановка завдання. Алгоритми стиснення відеоданих можна поділити на дві групи:
– алгоритми стиснення без втрат (CorePNG, GIF89a);
– алгоритми стиснення із втратами (MPEG-1, MPEG-2, MPEG-4).

Стиснення цифрових даних без утрат виконується шляхом заміни оригінальної послідовності бітів іншою послідовністю – що містить опис оригінальної послідовності. При цьому скорочення відбувається за рахунок фрагментів, що повторюються, і фрагментів, що містять закономірний ряд змін. У разі потреби отримати початковий потік даних його відновлюють за потоком-описом.

Стиснення даних із втратами виконують за двома схемами:

1. У трансформувальних кодеках фрейми зображень трансформуються в новий базисний простір і виконується квантування. Трансформація може здійснюватися або для всього фрейму цілком (як, наприклад, у схемах на основі wavelet-перетворення), або поблоково (характерний приклад – JPEG). Результат потім стискується методами ентропії.

2. У кодеках, які використовують метод передбачення, попередні й наступні дані використовуються для того, щоб передбачити поточний семпл зображення. Помилка між передбаченими даними і реальними разом з додатковою інформацією потім квантується і кодується.

У деяких системах обидві наведені техніки комбінуються через використання трансформуючих кодеків для стискання помилкових сигналів, згенерованих на стадії передбачення.

Дослідження основних алгоритмів стиснення відеозображень. На сьогодні майже всі алгоритми стиснення відео (наприклад, стандарти, прийняті ITU-T або ISO) використовують дискретне косинусне перетворення (ДКП) або його модифікації для усунення просторової надмірності. Інші методи, такі як фрактальне стискання і дискретне вейвлет-перетворення (ДВП), також були об'єктами досліджень, але тепер їх зазвичай використовують тільки для компресії нерухомих зображень.

Дискретне косинусне перетворення широко використовують для стиснення зображень. Стандарт стискання статичної графіки JPEG, використовуваний у відеоконференціях стандарт H.263, цифрові відеостандарти MPEG (Mpeg-1, Mpeg-2 і Mpeg-4) – всі вони використовують ДКП. У цих стандартах застосовують, зокрема, двовимірне ДКП послідовно до блоків зображення розмірністю 8×8 пікселів. Дискретне косинусне перетворення обчислює 64 (8×8 = 64) коефіцієнти, які потім квантуються, забезпечуючи тим самим реальне стиснення. У більшості зображень значна частина коефіцієнтів ДКП через свою незначну вагу після квантування обнуляються. Цю властивість ДКП і покладено в основу безлічі алгоритмів стиснення, які використовують ДКП.

Дискретне косинусне перетворення виконується множенням вектора на матрицю перетворення. При цьому матриця оберненого перетворення з точністю до множника дорівнює транспонованій матриці. Матриці вибирають таким чином, щоб перетворення було ортонормованим, а постійний множник дорівнював одиниці. В комп'ютерних програмах це не завжди так.

Різні періодичні довжини сигналу приводять до необхідності використовувати різні типи ДКП. Матриці для перших чотирьох типів:

$$DCT-1_n = \left[\cos kl\frac{\pi}{n-1}\right]_{0\le k,l<n},$$

$$DCT-2_n = \left[\cos k\left(l+\frac{1}{2}\right)\frac{\pi}{n}\right]_{0\le k,l<n},$$

$$DCT-3_n = \left[\cos\left(k+\frac{1}{2}\right)l\frac{\pi}{n}\right]_{0\le k,l<n},$$

$$DCT-4_n = \left[\cos\left(k+\frac{1}{2}\right)\left(l+\frac{1}{2}\right)\frac{\pi}{n}\right]_{0\le k,l<n},$$

де $k, l = 1,...,n$; n = кількість вибірок в блоці.

Саме *DCT-2* (двовимірне ДКП) частіше за інші матриці перетворення використовують на практиці.

Звісно, використання ДКП на блоці з n вибірок потребує $n \cdot n$ операцій множення і підсумовування. Проте завдяки рекурсивній структурі матриці ДКП реально буде потрібна набагато менша кількість математичних операцій, а саме $n \log_2(n)$ [2]. Ця властивість робить ДКП реально застосовним на сучасних математичних процесорах персональних ЕОМ.

Алгоритми кодування з ДКП мають такі недоліки:

1. "Блочність" за високої компресії.

2. Закруглення гострих кутів зображення. Випадкове "розмивання" гострих країв зображень.

3. Кодування дуже трудомістке. Тільки останнім часом вдалося виконати процес кодування програмно, а не апаратно.

Дискретне вейвлет-перетворення – алгоритм, що ґрунтується на передаванні сигналу, наприклад зображення через пару фільтрів: низькочастотний і високочастотний. Низькочастотний фільтр видає грубу форму вихідного сигналу. Високочастотний фільтр видає сигнал різниці або додаткової деталізації. У свою чергу, результат на виході високочастотного фільтра (додатковий сигнал деталізації) може бути підданий тій же процедурі й так далі.

Простим прикладом ДВП є ДВП Хаара.

Вхідний сигнал $x[n]$ є множиною вибірок з індексом n. Низькочастотний фільтр Хаара (Haar Low Pass Filter) являє собою арифметичне середнє двох вибірок вхідного сигналу $x[n]$:

$$g[n] = \frac{1}{2}\left(x[n] + x[n+1]\right).$$

Високочастотний фільтр Хаара (Haar High Pass Filter) є середньою різницею двох вибірок:

$$h[n] = \frac{1}{2}\left(x[n+1] - x[n]\right).$$

Зазначимо, що

$$x[n] = g[n] - g[n]x[n+1] = g[n] + h[n].$$

Вихідні послідовності $g[n]$ і $h[n]$ містять надлишкову інформацію. Таким чином, зрозуміло, що для відтворення вхідного сигналу $x[n]$ досить узяти лише парні або лише непарні його вибірки. Як правило, беруться парні вибірки. Таким чином, вхідний сигнал $x[n]$ включає в себе лише елементи з вибірок, наприклад парних:

$$g[0], g[2], g[4]....$$

$$h[0], h[2], h[4]....$$

$$x[0] = g[0] - h[0]; \quad x[1] = g[0] + h[0];$$

$$x[2] = g[2] - h[2]; \quad x[3] = g[2] + h[2] \text{ і т. д.}$$

Вихід низькочастотного фільтра є грубою аналогією вихідного сигналу. Якщо вихідним сигналом є зображення, то на виході низькочастотного фільтра буде розпливчате, розмите зображення з низькою роздільною здатністю. Вихід високочастотного сигналу додає деталі до зображення. У поєднанні з виходом низькочастотного фільтра може бути відтворено таким чином вхідне зображення. Грубу форму вихідного сигналу (сигнал на виході низькочастотного фільтра) інколи називають основним рівнем (base layer), а додатковий сигнал деталізації – рівнем поліпшення (enhancement layer). Сигнал на виході високочастотного фільтра $h[n]$ може бути пропущений знову через пару фільтрів, і процес

таким чином може бути повторений, поки не буде досягнута достатня міра деталізації вихідного сигналу $x[n]$. Проте зрозуміло, що жодного стиснення тут не досягається. Перетворення просто відтворює таку саму кількість бітів, яка була у вихідному сигналі. Вихідні значення називають коефіцієнтами перетворення, або коефіцієнтами wavelet-перетворення.

Перетворення Хаара використовують здебільшого для стиснення зображень. Для інших цілей застосовують складніші фільтри перетворень. Стиснення ж досягається передусім застосуванням деякої форми квантування (скалярної або векторної) до додаткового сигналу деталізації. Далі для отриманих коефіцієнтів перетворення використовують техніку ймовірнісного кодування (ентропії).

Припустімо, що в наведеному вище прикладі вхідним сигналом $x[n]$ є послідовність восьмибітових вибірок растра півтонового зображення. Для виходу низькочастотного фільтра $g[n]$ тепер можна використовувати ті ж 8 біт, а для високочастотного $h[n]$ – уже менше, наприклад 4. Це, по суті, скалярне квантування. Далі вихід високочастотного фільтра прагнутиме до нуля, оскільки коефіцієнти перетворення зменшуються у міру вживання алгоритму. Таким чином, можна застосувати ймовірнісне кодування до сигналу деталізації $h[n]$.

Насправді для більшості реальних зображень сигнал $g[n]$ на виході низькочастотного фільтра буде подібний до попередніх $g[n\text{-}1]$ вибірок за винятком граней контурів; $g[n]$ прагнутиме до $g[n\text{-}1]$ унаслідок того, що реальні об’єкти мають відносно постійний коефіцієнт віддзеркалення поверхонь.

Алгоритми кодування з ДВП, як і алгоритми кодування з ДКП, мають такі переваги і недоліки:

1) більшість як статичних, так і динамічних зображень, стиснених за допомогою алгоритму ДВП, не мають характерної для алгоритму ДКП блокової структури;

2) відносна якість зображень, стиснених з використанням ДВП, перевищує якість зображень, стиснених за допомогою ДКП, за одних і тих самих коефіцієнтів стиснення;

3) дискретне вейвлет-перетворення дещо “розмазує”, закругляє гострі контури зображення, так званий контурний шум або ефект Гіббса.

Векторне квантування (Vector Quantization,VQ). Технологію VQ використовують компресорами Indeo 3.2 і Cinepak. Обидва вони застосовують колірну схему YUV.

Суть алгоритму, що ґрунтується на векторному квантуванні, полягає в розбитті зображення на блоки (розміром 4×4 пікселів у колірній схемі YUV для компресорів Indeo і Cinepak). Зазвичай деякі блоки виявляються схожими один на одного. В цьому випадку компресор ідентифікує клас подібних блоків і замінює їх одним загальним блоком. Крім того, генерується двійкова таблиця (карта) таких загальних блоків з найкоротших кодових слів. При декомпресії VQ-декодер, використовуючи таблицю, збирає зображення поблоково із загальних блоків. Це спосіб кодування із втратами якості, оскільки подібність блоків дуже відносна. Тут допускається апроксимація реальних блоків зображення до загального блока, що їх об’єднує. Процес кодування тривалий і трудомісткий, оскільки кодеру необхідно виявляти належність кожного блока зображення до якого-небудь загального блока. Проте завдання декодування в цьому випадку зводиться до побудови зображення за заданою картою із загальних блоків і не займає багато апаратних і тимчасових ресурсів. Таблицю або карту називають ще й кодовою книгою, а двійкові коди, що входять у неї, – кодовими відповідно словами. Найбільший ступінь стиснення відеозображень з використанням алгоритму VQ досягається зменшенням кількості класів загальних блоків, тобто припущенням про подібність відносно більшої кількості блоків зображення, і як наслідок зменшенням кодової книги. У міру зменшення розмірів кодової книги якість відтвореного відео погіршується. В результаті на зображенні з’являється штучна “блоковість”.

Для прикладу порівняємо три такі блоки 4×4.

Блок 1:	Блок 2:	Блок 3:
128 128 128 128	128 127 128 128	128 127 126 128
128 128 128 128	128 128 128 128	128 128 128 128
128 128 128 128	128 128 127 128	127 128 128 128
128 128 128 128	128 128 128 128	128 128 128 128

Відмінності в наведених блоках для людського ока непомітні. Таким чином, блоки 2 і 3 можна замінити першим. Тоді кодова книга матиме такий вигляд:

Кодова книга [1] = 128 128 128 128
128 128 128 128
128 128 128 128
128 128 128 128

Важливою особливістю технології VQ є те, що для стиснення відео одна й та ж кодова книга може використовуватися для декількох кадрів зображення.

Переваги і недоліки векторного квантування:

1) процес кодування дуже трудомісткий і майже неможливий без спеціального додаткового обладнання;

2) процес декодування дуже швидкий;

3) блокові спотворення за високих коефіцієнтів стиснення;

4) технології, що використовують алгоритми ДКП (БДКП), ДВП можуть досягати вищих рівнів стиснення інформації.

Висновки та перспективи вдосконалення алгоритмів стиснення відеозображень. Перспективним засобом уникнення наведених недоліків є вдосконалення алгоритмів компресії-декомпресії у напрямі використання сплайна як базисної функції. Оскільки сплайн являє собою кусково-поліноміальну функцію, то її легко можна використати під час обчислень. Алгоритми для графічного зображення кривих за допомогою сплайнів та для обчислення їх поліноміальних складових надзвичайно ефективні [3]. Більш того, оскільки сплайни мають найменший можливий носії, то можна застосовувати схеми локальної інтерполяції для апроксимації функцій за допомогою будь-якого сплайн-підпростору. Усі такі алгоритми можна використовувати у режимі реального часу.

Список літератури

1. *Селомон Д.* Сжатие данных, изображений и звука / Д. Селомон – М.: Техносфера, 2004, – 368 с.
2. *Чичева М. А.* Эффективный алгоритм дискретного косинусного преобразования четной длины / М. А. Чичева // Компьютерная оптика. – 1998. – № 18 – С. 147 – 149.
3. *Чуи К.* Введение в вельветы / Чуи К. – М.: Мир, 2001. – 412 с.

А. М. Негода

Анализ современных алгоритмов сжатия видео

Проанализированы алгоритмы сжатия изображений с целью оценки их возможностей для сжатия потокового видео.

A. N. Negoda

Analysis of modern algorithms of video compression

In the article the analysis of algorithms of compression of images is conducted with the purpose of estimation of their possibilities for the compression of video.