

*В.В. Мороз, кандидат технічних наук, О.С. Чубач, аспірант,
(Одеський національний університет імені І.І. Мечникова, Україна)*

МЕТОД ЧАСОВОЇ ІНТЕРПОЛЯЦІЇ ВІДЕО НА ОСНОВІ ОБ'ЄКТНОЇ МОДЕЛІ КАДРУ

В статті пропонується метод інтерполяції відео послідовності з метою підвищення частоти кадрів на стороні декодера. Оцінка та компенсація руху виконується на основі обчислення вейвлетних апроксимацій оптичного потоку. Для покращення якості інтерпольованих кадрів пропонується компенсація руху з застосуванням об'єктної моделі кадру.

Більшість сучасних технологій, що використовують обробку відеосигналів, стикаються з проблемою перетворення частоти кадрів [1]. Дана проблема виникає при передачі та збереженні відеоданих і зачіпає дискретизацію кадрів (зображень) не тільки в просторі, але і в часі. Зміна частоти кадрів виконується також при переході від одного стандарту відео до іншого (PAL/SECAM та NTSC), при відновленні кадрів, що були зіпсовані під час неякісної передачі. Тому необхідно передбачати інформацію в пропущених/зіпсованих кадрах ґрунтуючись на інформації, що є в сусідніх кадрах.

Поставлена проблема використовує часову модель послідовності кадрів і є типовою задачею інтерполяції, в якій відома таблиця значень функцій при заданих значеннях аргументу, а треба обчислити значення функції в проміжних точках. Але, при використанні алгоритмів лінійної інтерполяції, помітні спотворення зображень при відтворенні об'єктів, що швидко рухаються: розмиття меж, тремтіння, уривчастість. Причина цих спотворень полягає у тому, що застосування прямої інтерполяції не завжди допустиме. Тому замість прямої інтерполяції застосовується інтерполяція з компенсацією руху.

В задачах обробки відео час може розглядатися як додатковий вимір. В зв'язку з цим, в більшості досліджень відеосигнал розглядається як тривимірна функція $f(x,t)$, де x – двовимірний вектор піксельних координат, а t – час. З фізичної точки зору $f(x,t)$ – інтенсивність зображення, як функція простору і часу.

Таким чином, відеосигнал можна розкласти в ортонормований базис для тривимірних сигналів. Але, виходячи з того, що найбільші зміни зображення з часом є результатом руху сцени по відношенню до відеокамери, таке представлення сигналу малоефективне. Це в першу чергу пов'язано із зміщенням рівнів яскравості в кадрах відеопослідовності.

Тому, найбільш поширеною технікою при кодуванні відео, є техніка на основі компенсації руху. Інтерполяція з компенсацією руху ґрунтується на аналізі зображення, виявленні об'єктів, що рухаються, і оцінці швидкості їх переміщення в кадрі. Задача розв'язується шляхом оцінки руху об'єктів. Для цього вимірюється абсолютна величина швидкості і напрямок переміщення об'єкту, тобто швидкість визначається як векторна величина. На практиці визначають горизонтальне та вертикальне зміщення об'єкту за час між двома кадрами. Отримані значення є відповідно горизонтальною та вертикальною координатами вектора, пропорційного швидкості, який є вектором руху.

В найпростішому випадку усі елементи кадру можуть переміщатися в одному напрямку з однією швидкістю. В загальному випадку необхідно визначити об'єкти, або множини елементів зображення, що характеризуються одним і тим же рухом. Кожному такому об'єкту призначається власний вектор руху.

Для кращої компенсації руху в зображенні необхідна більша кількість таких незалежних областей з призначеними векторами руху. В граничному випадку кількість незалежних векторів руху може дорівнювати числу елементів зображення. Коли відомі вектори руху для всіх елементів зображення, то можливо перемістити елементи в розрахункові точки і створити проміжне зображення.

Переміщення об'єктів розглядається з позиції вимірювання оптичного потоку, який є зміщенням рівнів яскравості в зображенні. Для обчислення зміщення рівнів яскравості з часом, інтенсивність зображення $f(x,t)$ має бути гладкою по x . Тоді оптичний потік може розглядатися, як поле траєкторій пікселів між послідовними кадрами. Маючи оптичний потік, можна побудувати точний прогноз для більшості пікселів між послідовними кадрами переміщуючи кожен піксел першого кадру вздовж його вектора оптичного потоку.

Обчислення оптичного потоку може бути виконане як за допомогою технічних прийомів, так і за допомогою диференційного потоку [2].

Але точне обчислення оптичного потоку призводить до великих обчислювальних затрат. Тому на практиці частіше застосовують не піксел, як одиницю переміщення, а блок пікселів. Неправильне визначення векторів руху окремих блоків призводить до небажаних візуальних ефектів, які проявляються у вигляді тремтіння або деформацій елементів зображення. Для зменшення подібних небажаних артефактів обчислимо кратномасштабні вейвлетні апроксимації оптичного потоку з подальшим узгодженням векторів руху на об'єктній моделі кадрів.

Нехай рівень яскравості i -го кадру $-f_i(x)$, а Δ – час, за який демонструється кадр. Якщо існує вектор зміщення $m_i(x)$ за проміжок часу з $(i \cdot \Delta)$ по $(i+1) \cdot \Delta$, то $f_{i+1}(x) = f_i(x + m_i(x))$. Відповідний вектор швидкості визначається як $v_i = m_i(x)/\Delta$. Для невеликих змін швидкості в деякому околі δ та достатньо малому z припустимо, що $f_{i+1}(x) \approx f_i(x + m_i(\delta))$, коли $|x - \delta| \leq z$. Виходячи з цього, узгодження зводиться до оцінки вектора зміщення $m_i(x)$ шляхом мінімізації норми різниці рівнів інтенсивності на квадратному блоці розміру z :

$$\varepsilon(\delta, m) = \iint_{|x-\delta| \leq z} |f_{i+1}(x_1, x_2) - f_i(x_1 + m_1, x_2 + m_2)|^2 dx_1 dx_2.$$

Таким чином, використовуючи зсув m_δ , який мінімізує похибку $\varepsilon(\delta, m_\delta) = \min_m \varepsilon(\delta, m)$, можна побудувати апроксимацію оптичного потоку $m_i(\delta)$.

При дуже малих значеннях z , $\varepsilon(\delta, m)$ має кілька локальних мінімумів для різних m . Так як локальні мінімуми по амплітуді мало відрізняються, то вибір глобального мінімуму може призвести до неправильної оцінки вектора руху $m_i(\delta)$ при наявності шуму в зображенні. При дуже великих розмірах зони узгодження, швидкість $v_i(x)$ може бути не постійною в її межах, а тому мінімізація похибки $\varepsilon(\delta, m)$ також призводить до помилкової оцінки руху. До того ж, обчислювальна складність в даному випадку для зображень в кадрі розміром $N \times N$ пікселів складає $O(z^2 N^4)$. Знизити обчислювальну складність алгоритмів компенсації руху, що базуються на узгодженні блоків, дозволяють кратномасштабні апроксимації зображення. Алгоритми, що ґрунтуються на використанні КАЗ, спочатку обчислюють грубу оцінку поля векторів руху. Для цього обчислюється узгодження кожного пікселя двох послідовних кадрів \tilde{f}_i^j та \tilde{f}_{i+1}^j шляхом мінімізації відстані, обчисленої по блоках із z^2 пікселів, де \tilde{f}_i^j – апроксимація i -го кадру в масштабі 2^j . Потім обчислюється уточнене узгодження для зменшеного масштабу 2^{j-1} .

Таким чином кратномасштабна апроксимація використовує спочатку обчислення похибки узгодження в зонах великих розмірів, а потім розмір зони зменшується при уточненні масштабу. При цьому складність таких алгоритмів складає $O(K^2 z^2 N^2)$, де K^2 – кількість цілочисельних векторів зміщення m , які задовольняють вимозі узгодження на двох послідовних масштабах при пороговому значенні K .

Обчислимо оптичний потік за допомогою повного приросту. Виходячи з того, що рівень яскравості $f(x(t), t)$ залишається постійним в межах кількох кадрів, то повний

приріст може бути записаний як

$$\frac{df(x(t), t)}{dt} = \frac{\partial f(x, t)}{\partial x_1} x'_1(t) + \frac{\partial f(x, t)}{\partial x_2} x'_2(t) + \frac{\partial f(x, t)}{\partial t} = 0.$$

Тоді рівняння оптичного потоку, що зв'язує вектор швидкості та рівень яскравості в будь-якій точці зображення з часом t , буде мати вид:

$$\bar{\nabla}f \cdot v = \frac{\partial f(x, t)}{\partial x_1} v_1 + \frac{\partial f(x, t)}{\partial x_2} v_2 = -\frac{\partial f(x, t)}{\partial t}, \quad (1)$$

де:

$v = (v_1, v_2) = (x'_1, x'_2)$ — вектор швидкості

$\bar{\nabla}f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2} \right)$ — градієнт f .

Нехай $v(x, t)$ — гладка вектор-функція x і, відповідно, її можна апроксимувати константою в досить малій області. Точна оцінка вектора руху може бути обчислена проекцією рівняння оптичного потоку на багатомаштабні вейвлети [3], а в [4], [5] запропонований швидкий алгоритм обчислення оптичного потоку у вейвлетному каркасі.

Головними недоліками алгоритмів, що ґрунтуються на подібному обчисленні вейвлетного потоку, є те, що малі об'єкти та об'єкти, що швидко рухаються, можуть бути втрачені вже на етапі, коли використовуються блоки великого розміру в поєднанні з вейвлетною фільтрацією.

Розглянемо материнський вейвлет:

$$\psi_{\delta, z} = \frac{1}{z} \psi \left(\frac{x_1 - \delta_1}{z}, \frac{x_2 - \delta_2}{z} \right).$$

Припустимо, що зміщення зображення за час Δ мале в порівнянні з носієм вейвлета, розмір якого пропорційний z . Тоді величина z буде визначатися за умови, що похибка ε_z мала, коли $|v| \ll z/\Delta$ і $v(x, t) \approx v(\delta, t)$ — майже константа на носії $\psi_{\delta, z}$. Обчислюючи скалярний добуток рівняння оптичного потоку (1) на вейвлет $\psi_{\delta, z}$ та інтегруючи по частинах отримуємо:

$$\left\langle f, \frac{\partial \psi_{\delta, z}}{\partial x_1} \right\rangle v_1(\delta, t) + \left\langle f, \frac{\partial \psi_{\delta, z}}{\partial x_2} \right\rangle v_2(\delta, t) = \frac{d}{dt} \langle f, \psi_{\delta, z} \rangle + \varepsilon_z(\delta, t). \quad (2)$$

Відеопослідовність складається з кадрів, що показуються з інтервалом. Якщо частота кадрів підвищується в n_f разів, то необхідно визначити значення похідної в правій частині в моменти часу $t = (i + 1/n_f)\Delta$.

Припустимо, що $n_f = 2$. Рівняння вейвлетного оптичного потоку обчислюється за умови, що похідна $\frac{d}{dt} \langle f, \psi_{\delta, z} \rangle$ в рівнянні (2) у момент часу $t = (i + 1/2)\Delta$ може бути визначена з похибкою другого порядку за допомогою кінцевої різниці:

$$\frac{d}{dt}\langle f, \psi_{\delta,z} \rangle = \frac{1}{\Delta}\langle f_{i+1} - f_i, \psi_{\delta,z} \rangle + \varepsilon_\alpha(\delta, t).$$

Тоді рівняння (2) буде мати вигляд:

$$\left\langle \frac{f_i + f_{i+1}}{2}, \frac{\partial \psi_{\delta,z}}{\partial x_1} \right\rangle v_1(\delta, t) + \left\langle \frac{f_i + f_{i+1}}{2}, \frac{\partial \psi_{\delta,z}}{\partial x_2} \right\rangle v_2(\delta, t) = \frac{1}{\Delta}\langle f_{i+1} - f_i, \psi_{\delta,z} \rangle + \varepsilon_z(\delta, t) + \varepsilon_\alpha(\delta, t).$$

Похибкою $\varepsilon_z(\delta, t) + \varepsilon_\alpha(\delta, t)$ можна знехтувати у випадку, коли швидкість є приблизно постійною на носії вейвлета, або її амплітуда не дуже велика. Таким чином, застосування властивості кратномасштабності вейвлетів для реалізації погодження блоків, реалізує зведення оцінки векторів швидкості на великих масштабах до погодженого пошуку в вузьких областях поступово зменшуваних масштабів.

Висновки

Застосування кратномасштабних вейвлетних апроксимацій кадрів при обчисленні векторів руху дозволило значно зменшити обчислювальну складність методу завдяки тому, що оптичний потік будується на основі низькочастотних доменів першого та другого рівнів вейвлетного перетворення. При уточненні векторів руху по високочастотних доменах, проводиться їх узгодження на рівні об'єктів. Якщо вектори руху суміжних елементів кадру не співпадають, то перевіряється їх належність до одного об'єкту кадру. Даний метод дозволив значно покращити візуальну якість відновленого відео. Для перевірки роботи даного методу оригінальна послідовність кадрів була проріджена в два рази. За допомогою запропонованого методу відкинуті кадри були інтерпольовані. Результат порівняння їх з оригінальними кадрами за метрикою PSNR підтвердив кількісно суб'єктивний висновок про візуальну якість: розходження складало не більше 3,2 децибелі на більшості кадрів. Більша похибка була лише на високо динамічних сценах з перекриттям об'єктів. Тому подальша робота буде полягати в пошуку більш точного рішення для проблеми перекриття.

Список літератури

1. *J.L. Barron, D.J. Fleet and S.S. Beauchemin.* Performance of optical flow techniques. *International Jour. on Computer Vision*, 12(1): 43-77, 1994.
2. *Shou-Yi Tseng.* Motion estimation using a frame-based adaptive thresholding approach, *Real-Time Imaging*, vol. 10, Issue 1, February 2004, pp. 1-7.
3. *J. Weber and J. Malik.* Robust computation of optical flow in a multi-scale differential framework. *International Jour. of Computer Vision*, 14(1):5-19, 1995.
4. *Yu Liu, King Ngi Ngan.* Fast multiresolution motion estimation algorithms for wavelet-based scalable video coding, *Signal Processing: Image Communication*, Volume 22, Issue 5, June 2007, pp. 448-46.
5. *H.-W. Park and H.-S. Kim.* Motion estimation using low-band-shift method for wavelet based moving-picture coding, *IEEE Trans. Image Process.*, vol.9, no.4, April 2000, pp. 577-587.