

УДК 004.932.2

О.В. Самощенко, Д.А. Савченко
Донецький національний технічний університет
samoshchenko@cs.donntu.edu.ua

Відстеження регіонів інтересу в послідовності відеокадрів за допомогою коваріаційних матриць

Розглядається використання коваріаційних матриць для рішення задачі відстеження регіонів інтересу у відеопослідовності. За допомогою запропонованого підходу є можливість кодувати частини відео із різним бітрейтом. Методика застосовується для рухомих об'єктів з постійною та непостійною формами. Розглядається вибір характеристик регіонів інтересу, які використовуються при відстеженні.

Ключові слова: регіон інтересу, коваріаційні матриці, алгоритм ієрархічного пошуку, відновлення моделі відстеження.

Вступ

Відстеження регіонів інтересу (ROI – Region of Interest) необхідне для кодування, щоб знизити обсяг даних. Основна ідея полягає у тому, щоб зберегти якість важливих об'єктів та знехтувати якістю фону.

У багатьох роботах ROI кодування застосовують для вирішення завдання передачі матеріалів відео конференцій, де регіоном інтересу є обличчя. На кожному кадрі важливість областей призначається залежно від кольору [1]. Такий підхід призводить до того, що зберігається якість неважливих частини зображення. Ідею ROI кодування можливо розвинути, якщо застосувати алгоритми відстеження. У роботі [2] використовується алгоритм, до якого необхідно внести модифікації для поліпшення його роботи.

Метою роботи є розробка алгоритму, який забезпечує відстеження регіонів інтересу, що можуть бути характеризовані різними властивостями. Для реалізації поставленої мети розв'язуються наступні задачі:

- Розробка базового алгоритму відстеження з використанням коваріаційних матриць,
- Оптимізація алгоритму за допомогою ієрархічного пошуку,
- Додавання масштабування регіонів інтересу та відновлення моделі відстеження,
- Експериментальне дослідження відстеження.

Задача відстеження регіонів інтересу

Відстеження регіонів інтересу містить у собі виявлення частини кадру за допомогою апріорної інформації, що описує об'єкти пошуку. Хибно виявлені регіони інтересу призводять до зниження якості сприйняття відео порівняно з початковим відео.

Якщо V відео, що складається з n кадрів

$F1..Fn$ і $B1$ – це контур регіону інтересу в кадрі $F1$, проблема відстеження полягає у визначенні контурів $B2..Bn$ з $B1$. Таке відстеження може бути реалізовано не тільки за допомогою моделей просторового подання, але й з використанням алгоритмів компенсації руху. Відстеження регіонів інтересу доречно застосовувати при телевізійному спостереженні (кримінальні телепрограми, спортивна статистика або репортажі), при пост опрацюванні фільмів (застосування особливих кольорових ефектів стосовно об'єктів), при передачі відео за допомогою мереж (відео конференції) та інше.

Регіон інтересу – це область у відеокадрі, яка характеризується декількома атрибутами, такими як колір, структура, текстура, яскравість та інше. Відстеження ROI містить у собі виявлення таких областей у послідовності кадрів з використанням їхнього місця розміщення на першому кадрі. Складність полягає у тому, що регіон інтересу може зазнавати незначних змін у зовнішньому вигляді, освітленні та інше. Завдання для системи відстеження регіонів інтересу полягає у застосуванні методу, який би виявляв ROI у присутності таких відхилень. Відстеження регіонів інтересу може бути реалізовано за допомогою двох шляхів:

1. Порівняння цільового блоку (ROI) з блоками кандидатами. Найбільш співпадаючий блок – це наступний цільовий блок.
2. Оцінка стану (цільового блоку) з урахуванням усіх вимірів до цього моменту, або розробка еквівалентної функції ймовірності щільності розташування об'єкту.

Відстеження можливо розглядати як рішення проблеми відповідності. Задача полягає у пошуку блоку кандидату в поточному кадрі, що відповідає цільовому блоку попереднього кадру. Міри подібності використовуються для визначення кількості відповідності. Вимірюються подібності геометричних особливостей, таких як

форма, положення, структура та негеометричних особливостей, таких як колір та освітлення. Блок кандидат, що є найбільш подібний, визначається новим регіоном інтересу.

Один з прикладів вирішення проблеми є застосування гістограм кольору з наступними перевагами [3]:

1. Метод надійний при обертанні об'єкту, зсуву камери, часткової оклюзії та інше.

2. Алгоритм показує гарні результати, якщо регіон інтересу специфічний стосовно кольорів. Гістограми кольору варто використовувати для відстеження облич, оскільки колір обличчя можливо описати у двох просторових компонент кольору.

Але цей метод не варто використовувати, коли гістограми незначно відрізняються. Таке трапляється, коли регіон інтересу важко характеризувати кольором.

Відстеження також можливо розглядати як проблему оцінювання. Воно також називається ймовірнісне виявлення/відстеження. У цьому методі регіон інтересу кадру Y_t моделюється як Z_t з параметрами θ_t , також застосовується модель простору станів (описується переходами станів та моделями спостережень) для розміщення декількох кадрів.

Зазначена проблема оцінки може бути реалізована за допомогою моделі спостереження (моделі зовнішнього вигляду), що є складовою методу фільтру часток. Також можливе використання алгоритму ЕМ (Expectation-Maximization).

Одна з головних переваг при вирішенні відстеження як проблеми оцінювання полягає у тому, що подальше розташування об'єкту прогнозується, оскільки використовується інформація переміщень з попередніх кадрів. Алгоритм не бажано використовувати, якщо рух регіону інтересу є хаотичним. Алгоритм дуже трудомісткий, тому що він стохастичний, що потребує аналізу випадкових явищ. Також цей алгоритм є дуже складним, тому його використання не бажане.

Регіон інтересу повинен бути добре описаний/змодельований для виконання відстеження ROI. Модель, що описана з використанням різних властивостей регіону, називається дескриптором регіону. Властивості, такі як колір, контур (градієнт), інтенсивність та інші, які є особливими для кожного регіону, формують параметри моделі. У результаті, вибір властивостей є одним з найважливіших кроків для проблем відстеження та класифікації. Гарні властивості повинні бути характерними та легко обчислюватися. Такі властивості як колір, градієнт, відклик фільтру – це найпростіший вибір. Їх використовували протягом багатьох років для вирішення завдань комп'ютерного бачення. Гістограми – це природне розширення

таких властивостей, що є ефективними при стеженні за об'єктами непостійної форми.

Однією з основних проблем є брак компетентних критеріїв подібності, що фіксує просторові та статистичні властивості, тобто більшість підходів або залежать тільки від розподілу кольору, або від розподілу структурних моделей. Є можливість використання гістограм з об'єднаними властивостями, але це занадто ускладнить обчислення. У цьому плані коваріаційні матриці є переважаючими [4]. Вони надають гарний спосіб для об'єднання властивостей та добре масштабуються, оскільки розмір дескриптора (матриці коваріацій) залежить тільки від числа властивостей d . Вони мають лише $(d^2 + d)/2$ різних значень.

Коваріація як дескриптор регіону

Нехай I – одновимірне полутонове зображення або тривимірне кольорове зображення, а F – властивість зображення, яка отримана з I . Кількість вимірів $F = W * H * d$, де W і H – розмір ROI, і d – це кількість властивостей, що характеризують регіон інтересу.

Використовується наступний вираз:

$$F(x, y) = \phi(I, x, y), \quad (1)$$

де ϕ – будь-яка властивість зображення: інтенсивність, колір, градієнти, відклики фільтрів та інше. Треба зазначити, що властивість зображення/вектору може складатися з використанням двох типів атрибутів: просторових та структурних.

$$f_k = [x, y, I(x, y), I_x(x, y) \dots], \quad (2)$$

де k – номер пікселя, x та y – координати пікселя, $I(x, y)$ – інтенсивність пікселя k , $I_x(x, y)$ – градієнт інтенсивності за напрямком x . Після отримання властивості F регіону R , створюється матриця коваріацій:

$$C_R = \frac{1}{MN} \sum_{k=1}^{MN} (f_k - \mu)(f_k - \mu)^T, \quad (3)$$

де f_k для $k=1..M*N$ – точки властивості у d вимірах всередині регіону інтересу R та μ – середнє значення точок.

Матриця коваріацій – це ефективний шлях для об'єднання декількох властивостей. Діагональні значення відповідають за розбіжність кожної властивості (дисперсії), не діагональні значення представляють коваріації. Оскільки при коваріації береться середнє значення та обчислюється його коваріація, то шум, який є присутній на зображенні, у більшості випадків є відфільтрованим.

Є декілька переваг при використанні матриць коваріацій як дескрипторів регіону. Деякі з них:

- Це досить зрозумілий та простий шлях для об'єднання структурних та статистичних властивостей.

- Матриці є низьковимірними при порівнянні з іншими дескрипторами. Вони мають тільки $(d_2+d)/2$ різних значень, оскільки розмір матриці коваріацій не залежить від розмірів регіону, а залежить лише від кількості використаних властивостей.
- Можливе досягнення інваріантності при масштабуванні або обертанні в залежності від того, як описані властивості.
- Коваріація не залежить від змін середніх значень, як це відбувається при зсуві значень кольору, оскільки при коваріації різниця середніх значень береться з дійсних значень (можливо відстежувати об'єкти при освітленні, що змінюється).

Для обчислення відстані між коваріаційними матрицями необхідно використовувати спеціальний підхід. Просте віднімання матриць не спрацює. Використовується інша метрика, що включає об'єднані характеристичні значення матриць [5]. Відстань між значеннями обчислюється за допомогою інваріантної ріманової метрики для роботи з симетричними додатними матрицями:

$$\rho(C_1, C_2) = \sqrt{\sum_{i=1}^d [\ln^2 \lambda_i(C_1, C_2)]}, \quad (4)$$

де $\lambda_i(C_1, C_2)$, для $i=1..d$ це узагальнені характеристичні значення, які обчислюються з $\lambda_i C_1 x_i - C_2 x_i = 0$ за допомогою чисельних методів, наприклад, розклад Холецького, QR та інші.

Алгоритм відстеження регіонів інтересу

Для початку роботи алгоритму відстеження необхідна апріорна інформація щодо зовнішнього вигляду об'єкту. Найбільш простим способом, який не потребує глобального пошуку регіону інтересу на першому кадрі, є такий, що використовує координати розташування регіону інтересу на першому кадрі відеопослідовності.

Відстеження ROI у подальших кадрах враховує параметр відстані локального пошуку – кількість пікселів, на яку може зміститися об'єкт. При встановленні цього параметру треба враховувати не тільки рух об'єктів, але й рух відеокамери. Це значення встановлюється користувачем.

Наприклад, встановлено відстань за горизонтальним напрямом (віссю x) $m=4$ пікселів, а вертикальним напрямом (віссю y) $n=6$ пікселів (див. рис. 1). Клітині – це пікселі, сірим кольором позначено місце розташування регіону інтересу у попередньому кадрі.

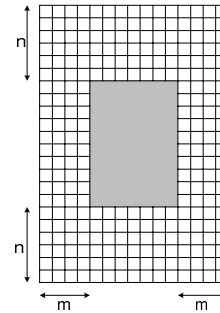


Рисунок 1 – Область локального пошуку

Кількість блоків кандидатів для перевірки:

$$n_{cand} = (2m+1)*(2n+1). \quad (5)$$

Тобто для даного прикладу необхідно обчислити 117 матриць коваріацій та порівняти їх з еталонною матрицею.

Кроки для реалізації базового алгоритму:

1. Отримати місце розташування регіону інтересу на першому кадрі. Обчислити матрицю коваріацій та назначити її еталонною.
2. Перейти на наступний кадр. Обчислити матриці коваріацій для позиції розташування ROI у попередньому кадрі та для усіх сусідніх позицій.
3. Обчислити відстань усіх матриць коваріацій регіонів кандидатів до еталонної матриці.
4. Обрати матрицю з мінімальною відстанню. Відповідний регіон встановлюється новим ROI. Запам'ятати позицію та матрицю коваріацій.
5. Повторити кроки з 2 по 4 для усіх кадрів відеопослідовності.

Ієрархічний пошук

Застосування ієрархічного пошуку зменшує кількість перевірок кандидатів. На першому кадрі обчислюється матриця коваріацій не тільки ROI, але й частини зображення (глобальний регіон), що більша за регіон розташування в усіх напрямках. Кількість додаткових пікселів глобального регіону визначається користувачем та залежить від характеру відеопослідовності.

При опрацюванні подальших кадрів відбувається обчислення матриці коваріацій для глобальних регіонів. Користувач визначає кількість позицій для пошуку глобального регіону. Матриці коваріацій порівнюються з еталонною матрицею глобального регіону. Координати регіону з мінімальною відстанню запам'ятовуються. Пошук регіону інтересу проходить тільки у межах одного з глобальних регіонів. Після того, як обрано новий регіон інтересу, необхідно обчислити та запам'ятати матрицю коваріацій глобального регіону. Алгоритм повторюється для усіх кадрів.

Вікно пошуку глобального регіону

зміщується не на 1 піксель, як це відбувається при відстеженні ROI, а на кількість пікселів, на які глобальний регіон перевищує регіон інтересу. Використання більшого зсуву вікна прискорить роботу алгоритму. Необхідно враховувати, щоб вікна глобального пошуку частково перекривали один одного. Це усуне ситуацію, якщо об'єкт буде розташовано на межах вікна. Пошук ROI відбувається у тому глобальному регіоні, що є найбільш схожим з попереднім.

Наприклад, кількість можливих позицій розташування регіонів інтересу без використання ієрархічного пошуку при обраній відстані локального пошуку $m=n=8$ пікселів дорівнює 289. При використанні ієрархічного підходу є можливість знизити зону локального пошуку до 4 пікселів. Тоді глобальний регіон буде перевищувати ROI на 4 пікселі за усіма напрямками. Кількість позицій пошуку дорівнює 1. У такому випадку кількість обчислень відстаней між матрицями дорівнює: глобальний регіон $(2*1+1)*(2*1+1) = 9$, регіон інтересу $(2*4+1)*(2*4+1) = 81$. Разом 90 обчислень відстаней між матрицями.

При використанні ієрархічного підходу потрібно додатково для кожного кадру обчислювати матрицю коваріацій для глобального регіону та зберігати її (що не становить проблему у зв'язку з невеликими розміром матриці). На рисунку 2 пояснено відпрацювання за глобальним регіоном.

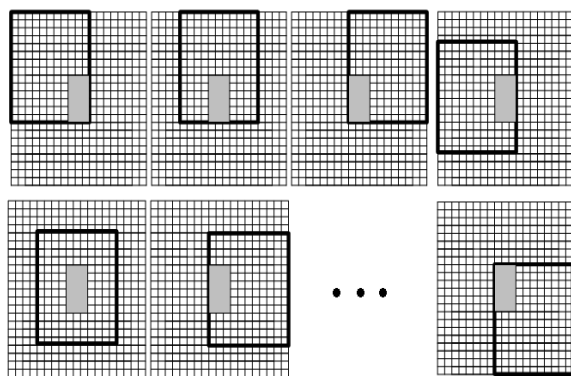


Рисунок 2 – Відпрацювання за глобальним регіоном

Використання алгоритму ієрархічного пошуку призведе до того, що на матрицю коваріацій буде мати вплив частина фону. Тому не варто використовувати занадто великий глобальний регіон, бо це може призвести до втрати об'єкту. Також якщо глобальний регіон є більшим за регіон інтересу на незначну кількість пікселів, то не буде досягнуто значного прискорення роботи. Бажано обрати кількість пікселів, на яку глобальний регіон є більшим за ROI, у відсотковому відношенні (наприклад 5%).

Якщо глобальний регіон перевищує ROI на m і n пікселів та вікно відстеження зміщується

на таку ж саму кількість позицій, то кількість перевірок кандидатів обчислюється з:

$$n_{cand} = (2m/(m)+1)*(2n/(n)+1) + (m+1)*(n+1) = 9 + (m+1)*(n+1). \quad (6)$$

Масштабування вікна

Для вирішення проблеми зменшення/збільшення регіонів інтересу необхідно використовувати масштабування вікна. Якщо масштабування не робити, то, наприклад, при збільшенні об'єкту буде відстежуватися лише деяка його частина, а не об'єкт в цілому. Коли об'єкт відстеження зменшується, то обчислена матриця коваріацій буде містити також інформацію про значну частину фону, що має негативний вплив на вирішення задачі відстеження. Також можлива ситуація, коли об'єкт поступово змінює свою форму на відміну від постійного фону. Це може призвести до переміщення вікна відстеження у бік постійного фону.

При масштабуванні вікна необхідно визначити кількість пікселів, на яку може змінюватися об'єкт за шириною та висотою. Оскільки об'єкт може постійно змінюватися у відео, то використання фіксованої кількості пікселів для перевірки масштабування може призвести до того, що алгоритм масштабування не впорається із своїм завданням. Доречно використовувати відсоткову кількість пікселів.

Є два підходи до реалізації масштабування. Це повна чи часткова перевірка. При повній перевірці використовується кожне з можливих вікон у області локального пошуку. Такий варіант є більш надійним, але його застосування занадто знизить швидкість відстеження.

При другому варіанті відстеження спочатку виконується звичайний пошук. У результаті отримано матрицю з мінімальною відстанню до еталонної. Відштовхуючись від отриманого місця розташування змінюють ширину вікна. Серед отриманих матриць обирається та, що більш співпадає з еталонною матрицею коваріацій. Далі звужену або розширену матрицю змінюють за висотою. Матриця з мінімальною відстанню буде призначено новим регіоном інтересу.

На рисунку 3 зображено приклад відстеження регіону інтересу з масштабуванням. Спочатку знаходиться найбільш схожий регіон з використанням вікна попереднього кадру. Далі відбуваються кроки масштабування. По черзі змінюються права, ліва, нижня, верхня границі.

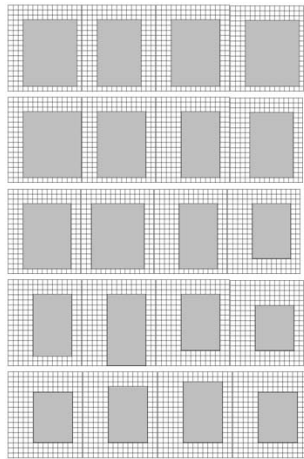


Рисунок 3 – Масштабування вікна пошуку

Для реалізації масштабування не достатньо лише обчислити матрицю коваріацій зменшеного/збільшеного регіону. При використанні такого підходу буде присутня суттєва відстань між поточною та еталонною матрицями. Перед етапом обчислення матриці коваріацій необхідно привести обраний регіон до розміру попереднього регіону. Для цього використовують афінні перетворення.

Відновлення моделі відстеження

Об'єкти, що знаходяться у регіонах інтересу, можуть змінювати свою форму або бути частково загордженими. Відновлення моделі відстеження знижує ймовірність хибного виявлення ROI.

Для відновлення моделі запам'ятовується T попередніх матриць коваріацій $[C_1 \dots C_T]$, де C_1 позначає поточну матрицю коваріацій. З цього набору обчислюється усереднена коваріаційна матриця, що змішує усі попередні матриці. У такому випадку усі регіони інтересу, що визначені раніше, та відповідні властивості зберігаються. З'єднану матрицю можна отримати за допомогою наступного рівняння:

$$\tilde{C} = \begin{pmatrix} \sigma_{1,1}^2 & \sigma_{1,2}^2 & \dots \\ \sigma_{2,1}^2 & \sigma_{2,2}^2 & \dots \\ \vdots & \vdots & \ddots \end{pmatrix} \quad (7)$$

Елементи матриці обчислюються за наступним рівнянням:

$$\sigma_{u,v}^2 = \frac{1}{MNT} \sum_{t=1}^T \sum_{k=1}^{MN} [f_k^t(u) - \mu(u)][f_k^t(v) - \mu(v)], \quad (8)$$

де $f_k^t \in R_f$. Середнє значення μ обчислюється серед усіх регіонів R_1, \dots, R_T . Хоча це формулювання достатньо просте, його використання передбачає той факт, що розмір вікон відстеження регіонів інтересу повинен бути однаковим. Також усі відеокадри мають однаковий пріоритет. Крім того, відновлення моделі потребує багато обчислювальних ресурсів та великої кількості

пам'яті для зберігання ROI попередніх кадрів.

При моделюванні обиралися такі властивості: позиція (x,y) , атрибути зовнішнього вигляду (кольорові компоненти RGB), структура області зображення (градієнт).

На рисунку 4 наведено приклад відстеження рухомого об'єкта постійної форми.

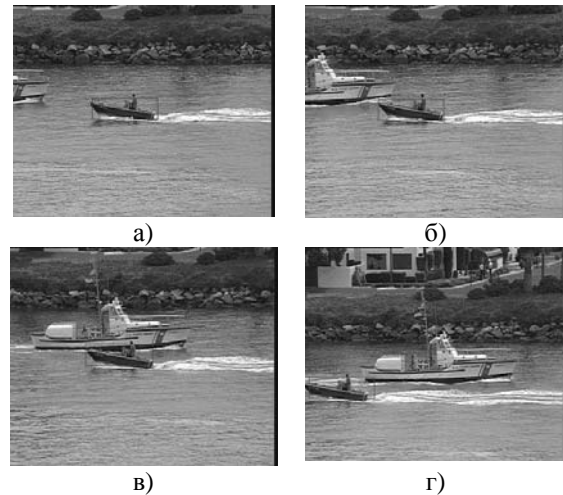


Рисунок 4 – Відстеження човна, який рухається, на послідовності відеокадрів

Найбільш проблематичним є відстеження об'єктів непостійної форми, що швидко рухаються. Постійні зміни можуть привести до того, що буде втрачено об'єкт пошуку. На рисунку 5 зображено приклад відстеження об'єкта непостійної форми (регіоном інтересу обрано тенісиста).

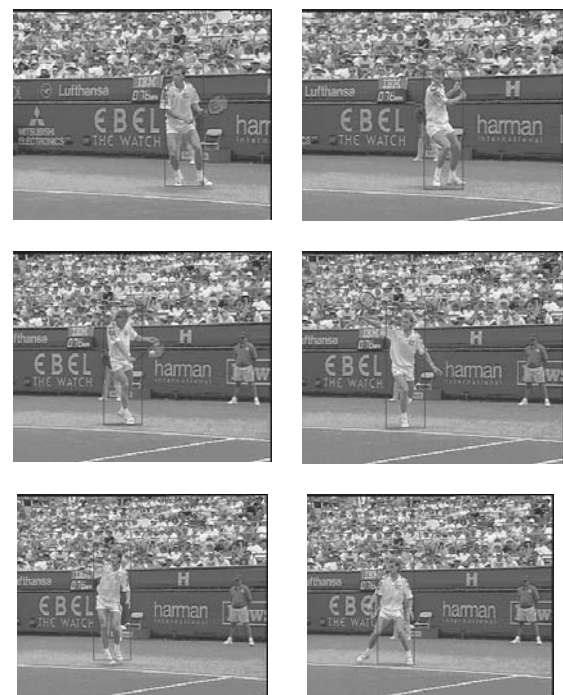


Рисунок 5 – Відстеження об'єкта з непостійною формою

Попереднє опрацювання регіонів інтересу

Один з методів кодування ROI – це використання попереднього опрацювання відео (ROI preprocessing). Воно опрацьовує відео таким чином, що регіону інтересу призначається більше біт ніж фону. Попереднє опрацювання ROI вносить модифікації до відеопослідовності перед етапом кодування.

Для оцінки попереднього опрацювання використовуються два критерії: коефіцієнт стиснення та пікове відношення сигналу до шуму (PSNR). Чим є більший коефіцієнт стиснення, тим менший бітрейт, що є ціллю ROI кодування. Значення PSNR визначає кількісну міру якості відеопослідовності. Метрика визначається для регіону інтересу і для фону. Попереднє опрацювання вказано для прикладів, що наведені вище (див. рис.6).

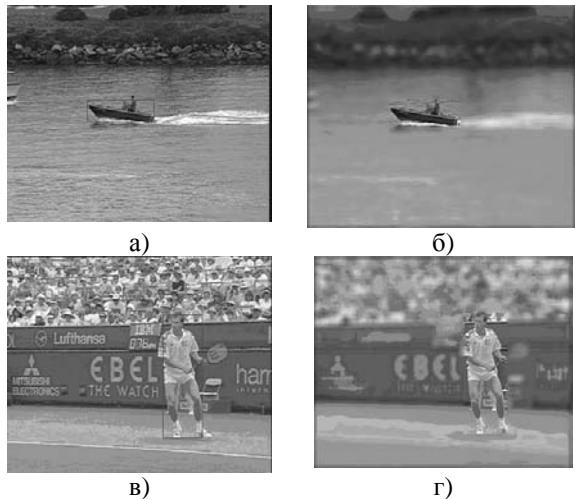


Рисунок 6 – Результати попереднього опрацювання ROI

Відеопослідовності, що наведені, проходили етап кодування після просторової фільтрації фону. При відеокодуванні поставлено мету досягти схожої якості регіону інтересу. На рисунках 7 і 8 зображено значення бітрейтів і PSNR відео із застосуванням попереднього опрацювання та без нього.

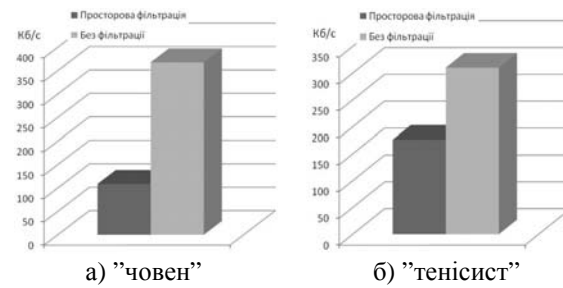


Рисунок 7 – Бітрейт закодованих відеопослідовностей

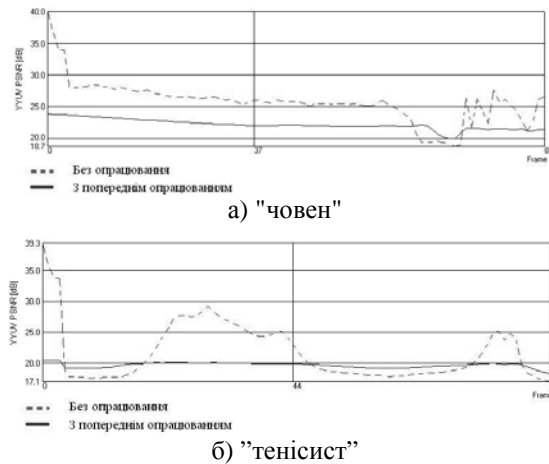


Рисунок 8 – Значення PSNR закодованих відеопослідовностей

Попереднє опрацювання може значно знизити якість фону. Тому ROI кодування варто використовувати тільки при вирішальному значенні обсягу даних.

Висновки

Матриці коваріацій – це зручний інструмент для відстеження регіонів інтересу. Головні переваги – простота реалізації, можливість об'єднувати властивості зображення, кількість яких обирається у залежності від поставленого завдання. Це робить алгоритм відстеження гнучким. Алгоритм впорався із завданням відстеження об'єктів як постійної, так і змінної форм.

Література

1. Karlsson L.S. Spatio-temporal pre-processing methods for region-of-interest video coding / Linda S. Karlsson – Department of Information Technology and Media Mid Sweden University, 2007. – 133 с.
2. Exploiting Region Of Interest For Improved Video Coding [Електронний ресурс] / Ramya Gopalan, 2009. – Режим доступу до статті: <http://etd.ohiolink.edu/send-pdf.cgi/Gopalan%20Ramya.pdf?osu1250622014>
3. Comaniciu D. Real-Time Tracking of Non-Rigid Objects using Mean Shift / D. Comaniciu, V. Ramesh, P. Meer – IEEE Transactions on Computer Vision and Pattern Recognition, 2009.
4. Covariance Tracking using Model Update Based on Lie Algebra [Електронний ресурс] / Fatih Porikli, Oncel Tuzel, 2006. – Режим доступу до статті: <http://www.merl.com/papers/docs/TR2005-127.pdf>
5. A Metric For Covariance Matrices [Електронний ресурс] / Wolfgang Forstner, Boudewijn Moonen – Режим доступу до статті: http://www.uni-stuttgart.de/gi/research/schriftenreihe/quo_vadis/pdf/foerstner.pdf

Надійшла до редакції 01.02.2011