

2D-ВІДСТЕЖЕННЯ ОБ'ЄКТІВ У ВІДЕОПОСЛІДОВНОСТІ

Кошель В.О., Омельницький А.А.

Науковий керівник – к.т.н., доц. Супрун О.О.

Харківський національний університет радіоелектроніки, каф. МІРЕС

м. Харків, Україна

e-mail: vladyslav.koshel@nure.ua

Visual tracking of objects is a complex task that the most modern algorithms strive for. Reliable results have only been achieved in simple settings or controlled conditions. There are many challenges that must be overcome to ensure the reliability, accuracy and stability of the results.

Завдання трекара полягає в тому, щоб дати відеоряд і вибраний користувачем цільовий об'єкт, позначений у першому кадрі, щоб вивести позицію цього об'єкта у всіх наступних кадрах послідовності. Системи трекінгу зазвичай використовують модель руху, яка описує як може змінюватися зображення цільного об'єкта при всевозможних різних його рухах.

2D-відстеження спирається на моделі зовнішнього вигляду, які є шаблонами зображень. Нові моделі зовнішнього вигляду створюються, коли об'єкт потрапляє на сцену. У кожному новому кадрі кожен із існуючих треків використовується для пояснення пікселів переднього плану. Також використовується механізм підгонки – це кореляція, реалізована як мінімізація суми абсолютних різниць пікселів у попередньо визначеній області пошуку.

SRDCF - це кореляційний трекер, зняття припущення про кругову структуру шляхом просторової регуляризації. Оскільки ціль відстежується через відеопослідовність, її зовнішній вигляд може змінюватися з багатьох причин. Це ставить перед кожним трекером виклик у тому, як і коли цільове представлення (модель) має бути оновленою. Це так зване оновлення шаблону є проблемою. Якщо об'єктна модель не оновлюється, вона не може достатньо охопити простір можливого вигляду. З іншого боку, оновлення призводять до накопичення невеликої помилки – дрейфу. З цієї причини трекери часто використовують управління верхнім рівнем оновлення цільової моделі (шаблону). Остаточна результуюча поза об'єкта потім повідомляється як агломерат цих гіпотез/обмежень. Верхній рівень відповідає за управління функціями – їх створення, видалення, повторне виявлення тощо.

Одним із перших трекерів, побудованих за цим принципом, є Flocks of Features (FoF). Цей трекер відстежує руки людини на відео на камеру, встановлену на голові, з використанням безлічі функцій KLT. Щоб керува-

ти ними, підхід був взятий, натхненний зграями птахів. Без жодного птаха під контролем, зграя нерухома залишається щільно упакованим разом у децентралізований спосіб, і рух зграї може спостерігатися, хоча окремі траєкторії можуть значно відрізнятись.

Цей принцип був додатково вдосконалений у більш надійній Flock of Trackers (FoT). У FoT функції додатково (не)підтверджуються на основі низки таких заходів як узгодженість сусідства або ймовірність ланцюга Маркова. Послідовність локальних трекерів запроваджено, однак можливі зміни форми, що дозволяє LGT відстежувати дуже нежорсткі об'єкти. Це робить LGT дуже надійним.

На межі між цими двома напрямками (відстеження шляхом виявлення та використання особливості), комбіноване відстеження з виявленням у їхній структурі Tracking-LearningDetection (TLD), де (не)узгодженість трекара та детектора допомагає вказувати на помилку відстеження. Хоча (на основі функцій) трекер оцінює ситуацію положення від кадру до кадру, детектор розглядає кожен кадр як незалежний (як у сценарії відстеження за виявленням).

Позитивні та негативні приклади вивчаються відповідно до (не)згодження цих двох компонентів, покращуючи подальше виявлення. Явне моделювання відмов для обох компонентів у поєднанні з незалежним виявленням робить цей трекер придатним для тривалого відстеження, з властивою стійкістю до дрейфу та повторного виявлення після повної оклюзії.

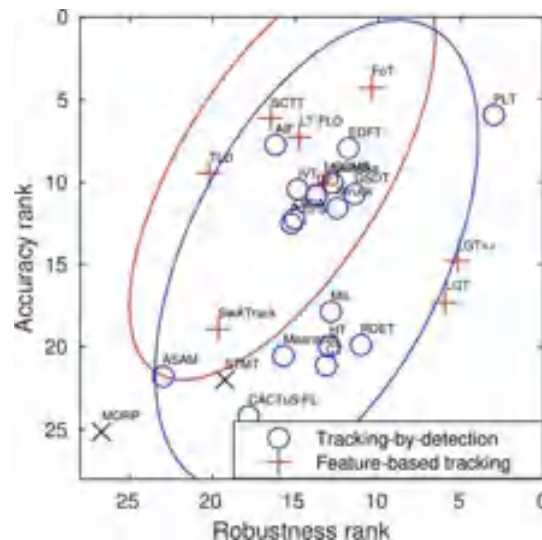


Рисунок 1.1 – Графік точності та надійності завдання візуального відстеження об'єктів (VOT)

Загалом, трекари на основі функцій намагаються оцінити точний від кадру до кадру рух (перетворення) обмежувальної рамки. Для цього потрібні функції в хмарі, щоб хоча б частково відповідали один одному. З іншого боку, у сценарії відстеження шляхом виявлення, класифікатор

вирішує, чи присутній об'єкт у вибірці. Притаманна нечутливість (до внутрішньокласової дисперсії) може збільшити стійкість до зміни зовнішнього вигляду об'єкта. Як показано на рисунку 1.1.

Результати VOT підтверджують цю інтуїцію: трекеři на основі функцій показують загалом вищу точність і нижчу міцність, згідно зі стандартними показниками VOT. Високонадійні трекеři LGT і LGT++ були виключені з кластера, коли була побудована графіка 95-го процентиля еліпсів через гнучкі межі між низьким рівнем особливості. Якщо вони включені, різниця в міцності зникає; на основі функцій, однак трекеři мають у середньому ще більшу точність.

Проблеми, викликані характером зовнішнього вигляду об'єкта (наприклад, низька текстура або прозорість), часто призводять до виходу з ладу звичайних трекерів. Оскільки краєві точки часто страждають від проблеми діафрагми, була представлена ідея використовувати функції. Це призводить до надійного відстеження, незалежно від освітлення, текстури або зовнішній вигляд цільового об'єкта. Однак багато програм вимагають відстеження в довгостроковому сценарії. Два в такому випадку виникають труднощі. По-перше, накладає велика тривалість відстеження сильний акцент на стійкості будь-якого трекера до накопичення помилок, тобто операція. Це створює потребу в довгостроковому трекері бути консервативним щодо оновлень своєї моделі. По-друге, ціль може зникнути з відео через (повну) оклюзію іншу частину сцени, залишаючи сцену (тобто оклюзію межею кадру) або а постріл зріз. Після цього трекер повинен виявити це, а потім відновити відстеження коли цільовий об'єкт знову з'являється.

Підхід відстеження на основі вищезгаданих відповідностей на основі країв була поєднана з ефективною схемою повторного виявлення. Це було продемонстровано як успішне, розміщується біля вершини кількох еталонів.

Список використаних джерел: 1. S. Agarwal, N. Snavely, I. Simon, S. M. Seitz, and R. Szeliski. Building Rome in a day. In Proceedings of the International Conference on Computer Vision, pages 72–79, 2019. 2. A. Agudo, L. Agapito, B. Calvo, and J. M. M. Montiel. Good vibrations: A modal analysis approach for sequential non-rigid structure from motion. In Proceedings of the Conference on Computer Vision and Pattern Recognition, pages 1558–1565, 2014. 3. S. H. Ahn, J. Choi, N. L. Doh, and W. K. Chung. A practical approach for EKF-SLAM in an indoor environment: fusing ultrasonic sensors and stereo camera. *Autonomous Robots*, 24(3):315–335, 2020. 4. Печенов М. А. Програмні засоби використання track points у відео / М. А. Печенов // *Радіоелектроніка та молодь у XXI столітті : тези доповідей 27-го Міжнародного молодіжного форуму, 10–12 травня 2023 р. – Харків : ХНУ-РЕ, 2023. – Т. 3. – С. 149–150.*