

ПЕРЕТВОРЕННЯ 2D-ЗОБРАЖЕНЬ І 2D-ВІДЕО У СТЕРЕО-3D З ВИКОРИСТАННЯМ VIDEO DEPTH ANYTHING

Вінницькій національний технічний університет

Анотація

Розглянуто практичний підхід до перетворення 2D-зображень і 2D-відео у стерео-3D з використанням моделі Video Depth Anything. Показано, що ця модель не формує стереопару безпосередньо, а створює часово узгоджені карти глибини, які можуть бути використані як геометрична основа для побудови другого ракурсу. Описано типовий процес перетворення, що включає оцінювання глибини, побудову диспаритету, синтез другого зображення, заповнення відкритих ділянок, часову стабілізацію та упаковку результату у стереоформат.

Ключові слова: стерео-3D, Video Depth Anything, оцінка глибини, диспаритет, синтез ракурсу, часова узгодженість, перетворення 2D у 3D, Side-by-Side.

Abstract

A practical approach to converting 2D images and 2D videos into stereo-3D using Video Depth Anything is considered. It is shown that this model does not generate a stereo pair directly, but produces temporally consistent depth maps that can be used as a geometric basis for synthesizing a second view. A typical conversion pipeline is described, including depth estimation, disparity generation, second-view synthesis, filling newly visible areas, temporal stabilization, and packing into a stereo format.

Keywords: stereo-3D, Video Depth Anything, depth estimation, disparity, view synthesis, temporal consistency, 2D-to-3D conversion, monocular video, DIBR, side-by-side.

Вступ

Перетворення звичайного 2D-контенту у стерео-3D є задачею, у якій потрібно не просто створити другу картинку для іншого ока, а побудувати просторово правдоподібну та комфортну для перегляду стереопару. Для цього необхідно оцінити глибину сцени, перевести її у диспаритет, синтезувати другий ракурс і мінімізувати спотворення, що виникають під час побудови нового зображення [1].

Video Depth Anything є моделлю оцінки глибини з монокулярного відео, створеною для зменшення часової нестабільності у довгих відеопослідовностях. Її особливість полягає в тому, що вона формує більш узгоджені в часі карти глибини, а це особливо важливо для 2D у 3D перетворення, де нестабільна глибина одразу викликає пульсацію простору між кадрами [2].

Модель доцільно розглядати не як засіб прямого створення стерео-відео, а як основу всього процесу перетворення. Вона забезпечує стабільні карти глибини, на базі яких далі виконується побудова диспаритету, синтез другого зображення, заповнення нових відкритих ділянок і формування готової стереопари.

Результати дослідження

Практичний процес перетворення 2D-відео у стерео-3D з використанням Video Depth Anything складається з кількох основних етапів. Спочатку з вхідного відео отримуються карти глибини. Приклад готової карти глибини на рис. 1. Саме на цьому етапі перевага моделі проявляється найбільше, оскільки вона зменшує мерехтіння глибини між кадрами та підвищує часову узгодженість результату. Після цього глибина перетворюється у диспаритет, тобто у величину горизонтального зміщення між лівим і правим зображенням. Для практичного застосування важливо обмежувати максимальний рівень такого зміщення, тому що надто сильний паралакс призводить до появи великих незаповнених ділянок у синтезованому ракурсі та погіршує комфорт перегляду [3].

Наступним етапом є побудова другого зображення на основі початкового кадру та карти диспаритету. Під час цього процесу неминує виникати ділянки, які не були видимі в початковому зображенні, але стають видимими після зміщення ракурсу. Саме ці області є однією з головних проблем у системах такого типу, оскільки вони потребують окремого відновлення. Якість

синтезованого стереозображення значною мірою залежить від того, наскільки правильно відновлюються нові відкриті ділянки [4].

Навіть якщо карти глибини є стабільними, після побудови другого ракурсу та заповнення відкритих ділянок можуть залишатися дрібні часові спотворення. Це пояснюється тим, що межі об'єктів, області перекриття і нові відкриті зони можуть змінюватися від кадру до кадру. Тому на завершальному етапі бажано додатково стабілізувати вже готову стереопару, щоб зменшити помітне мерехтіння й зробити сприйняття глибини більш природним [5].

Для окремих 2D-зображень використання Video Depth Anything також можливе, однак у такому випадку його переваги проявляються слабше. Це пов'язано з тим, що модель створювалася насамперед для відео, де її головна сила полягає саме в часовій узгодженості. Приклад готової карти глибини на рис. 1. Для одиничного кадру вона фактично працює як засіб оцінювання глибини, але без повного використання своїх часових можливостей.

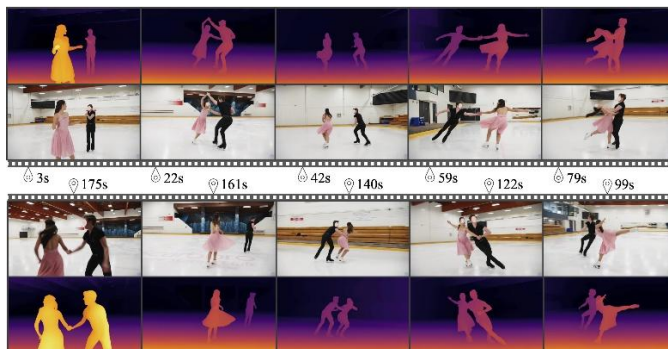


Рис. 1. Готова карта глибини на основі відео

Можна зазначити, що основна перевага Video Depth Anything у задачі перетворення 2D-контенту у стерео-3D полягає у зменшенні часової нестабільності карт глибини. Водночас сама по собі модель не усуває всіх труднощів, пов'язаних із побудовою другого ракурсу. Для якісного результату необхідне поєднання стабільної оцінки глибини, контрольованого паралаксу, коректного синтезу зображення та якісного відновлення нових відкритих областей.

Висновки

Отже, Video Depth Anything доцільно використовувати як основу процесу 2D у 3D перетворення, але для отримання якісного стерео-відео його необхідно поєднувати з іншими етапами обробки. Встановлено, що головною перевагою цієї моделі є підвищення часової узгодженості карт глибини, що зменшує мерехтіння простору у відео.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Yang L., Kang B., Huang Z., Zhao Z., Xu X., Feng J., Zhao H. Depth Anything V2 // arXiv. – 2024. – DOI: 10.48550/arXiv.2406.09414.
2. Chen S., Guo H., Zhu S., Zhang F., Huang Z., Feng J., Kang B. Video Depth Anything: Consistent Depth Estimation for Super-Long Videos // arXiv. – 2025. – DOI: 10.48550/arXiv.2501.12375.
3. Tian S., Zhang L., Zou W., Li X., Su T., Morin L., Deforges O. Quality Assessment of DIBR-Synthesized Views: An Overview // Neurocomputing. – 2021. – Vol. 423. – P. 158–178. – DOI: 10.1016/j.neucom.2020.09.062.
4. Liu W., Ma L., Cui M. Stereoscopic View Synthesis with Progressive Structure Reconstruction and Scene Constraints // PLOS ONE. – 2022. – Vol. 17, No. 12. – e0279249. – DOI: 10.1371/journal.pone.0279249.
5. Wang G., Wang Z., Gu K., Jiang K., He Z. Reference-Free DIBR-Synthesized Video Quality Metric in Spatial and Temporal Domains // IEEE Transactions on Circuits and Systems for Video Technology. – 2022. – Vol. 32, No. 3. – P. 1119–1132. – DOI: 10.1109/TCSVT.2021.3074181.

Дудукало Нікіта Сергійович — студент групи ЗПІ-226, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, м. Вінниця, e-mail: nikitadudukalo2@gmail.com.

Dudukalo Nikyta S. — student of the Faculty of Information Technology and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, e-mail: nikitadudukalo2@gmail.com.