

УДК 004.032.26:004.93

ЗАСТОСУВАННЯ НЕЙРОМЕРЕЖЕВОВОГО ПІДХОДУ ДЛЯ ОЖИВЛЕННЯ ЗОБРАЖЕНЬ ОБЛИЧЧЯ

Льолін Д.В.

e-mail: danylo.lolin@nure.ua

Науковий керівник – д.т.н., проф., Гороховатський В.О.

Харківський національний університет радіоелектроніки, каф. ІНФ
м. Харків, Україна

In this work, methods for animating 3D facial models using the neural network approach of the First Order Motion Model (FOMM) and the integration of animation data into Blender are discussed. Key points are used to create a motion map that allows the transmission of facial expressions and gestures to the 3D model. The obtained animation data is synchronized with Shape Keys in Blender, providing natural movement of the model without the need for complex rigging.

У сучасному світі технології машинного навчання та комп'ютерної графіки дозволяють створювати реалістичну анімацію на основі статичних зображень. Моєю метою було розробити систему, яка оживляє обличчя на фотографії, використовуючи нейромережевий підхід First Order Motion Model (FOMM), інтегрований у Blender [1, 2].

Я прагну створити метод, який дозволяє завантажити 3D-модель обличчя в Blender, застосувати до неї анімацію на основі моделі руху обличчя та отримати результат у реальному часі. Це дозволяє використовувати Blender не лише як інструмент для 3D-моделювання, але і як платформу для анімації за допомогою штучного інтелекту.

Розвиток глибокого навчання дозволив автоматизувати процес анімації обличчя. Одним із найперспективніших алгоритмів є First Order Motion Model (FOMM). First Order Motion Model (FOMM) – це метод глибокого навчання, який дозволяє анімувати статичне зображення, використовуючи відео як джерело руху. Алгоритм був розроблений дослідниками у сфері машинного навчання та комп'ютерного зору для передачі міміки та жестів на нерухомі зображення [2–4].

Для навчання ми використовуємо велику колекцію відеопослідовностей, що містять об'єкти однієї і тієї ж категорії. Ця модель навчається реконструювати навчальне відео шляхом поєднання одного кадру та вивченого прихованого представлення руху у відео. Спостерігаючи пари кадрів (початковий і рухомий), кожен із яких витягнуто з одного і того ж відео, вона вчиться кодувати рух як комбінацію зміщень ключових точок, специфічних для руху, та локальних афінних перетворень.

Під час тестування ця модель застосовується до пар кадрів, що складаються з початкового зображення та кожного кадру рухомого відео, виконуючи анімацію початкового об'єкта.

Огляд підходу представлений на рис. 1. Ця структура складається з двох основних модулів: модуля оцінки руху та модуля генерації зображення. Метою модуля оцінки руху є прогнозування щільного поля руху. Ми припускаємо, що існує абстрактна система відліку, і незалежно оцінюємо два перетворення: від початкового кадру до рухомого та від рухомого до початкового. Такий вибір дозволяє незалежно обробляти початкові та рухомі кадри. Це бажано, оскільки під час тестування модель отримує пари початкових зображень і рухомих кадрів, взятих із різних відео, які можуть значно відрізнитися візуально.

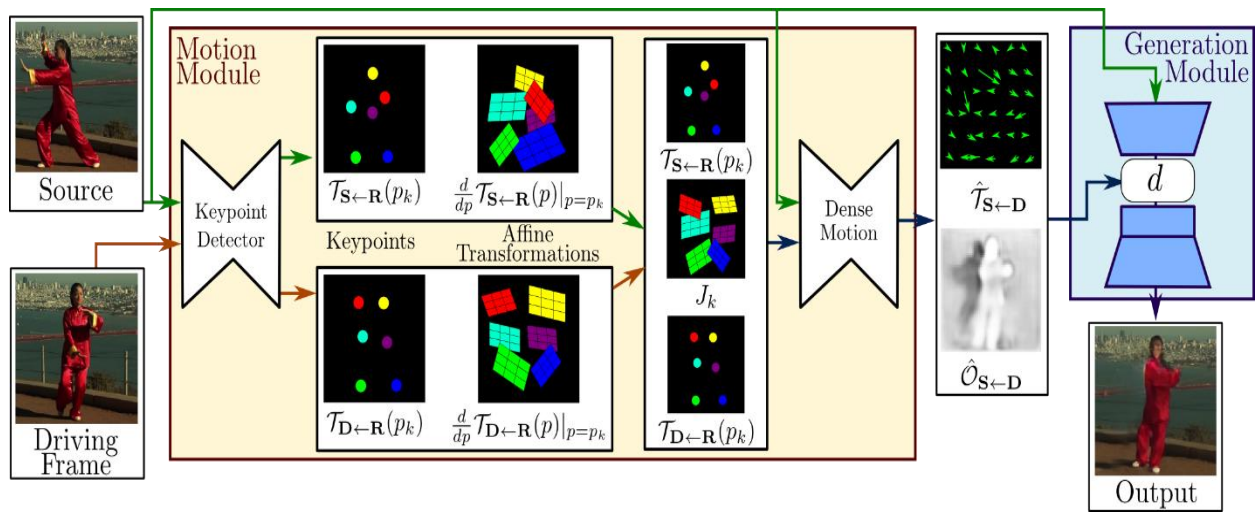


Рисунок 1 – Процес навчання FOMM

На першому етапі ця модель апроксимує обидва перетворення з набору розріджених траєкторій, отриманих шляхом використання ключових точок, вивчених самоконтрольованим методом. Вона моделює рух у околі кожної ключової точки за допомогою локальних афінних перетворень. Порівняно з використанням лише зміщень ключових точок, локальні афінні перетворення дозволяють моделювати більше сімейство перетворень. На другому етапі щільна мережа оцінки руху поєднує локальні апроксимації та отримує кінцеве щільне поле руху.

Ключові точки в моделі First Order Motion Model (FOMM) відіграють важливу роль у процесі анімації. Ці точки визначають значущі анатомічні місця на обличчі, такі як куточки очей, брови, кінчик носа, та куточки рота. Під час тренування моделі система автоматично визначає оптимальні позиції цих ключових точок на статичному зображенні, що дозволяє створювати точну мапу руху для анімації. Ця мапа руху використовується для трансформації кожної ключової точки із статичного стану в динамічний, відображаючи реалістичні вирази облич та жести, які були зняті з реальних відеозаписів.

Крім того, разом із полем руху ця мережа обчислює маску оклюзії, яка визначає, які частини початкового зображення можуть бути реконструйовані шляхом деформації, а які області слід замінити (вивести з контексту). Нарешті, модуль генерації візуалізує зображення початкового об'єкта, спроектованого у відповідний момент часу в рухомому відео.

У цьому випадку нейромережа генератора деформує початкове зображення відповідно до щільного руху та заповнює частини зображення, які були замасковані як оклюзія.

Використання цих технологій дозволило оживити 3D-моделі обличчя в реальному часі, забезпечивши високу реалістичність анімаційних процесів. Отримані результати можуть бути використані для подальших досліджень у галузі комп'ютерного зору та машинного навчання, відкриваючи нові можливості для розробки більш складних та функціональних анімаційних систем [4, 5].

Список використаних джерел:

1. Blender 4.3 Python API Documentation: вебсайт. URL: <https://docs.blender.org/api/current/index.html> (дата звернення: 28.01.2025).
2. Aliaksandr Siarohin, Stéphane Lathuilière, Sergey Tulyakov, Elisa Ricci and Nicu Sebe in NeurIPS 2019: вебсайт. URL: <https://aliaksandrsiarohin.github.io/first-order-model-website> (дата звернення: 28.01.2025)
3. Daradkeh Y.I., Gorokhovatskyi V., Tvoroshenko I., and Zeghid M. (2024) Improving the effectiveness of image classification structural methods by compressing the description according to the information content criterion, *Computers, Materials & Continua*, vol. 80, no. 2, pp. 3085-3106.
4. Gorokhovatsky, V.A., Putyatin, Y.P. (2008) Structural recognition of images on the basis of voting models of attributes of typical points, *Data recording, storage and processing*, 10(4), 75-85.
5. Gadetska, S.V., Gorokhovatskyi, V. O., Stiahlyk, N. I., Vlasenko, N.V. Statistical data analysis tools in image classification methods based on the description as a set of binary descriptors of key points. *Radio Electronics, Computer Science, Control*, 2021, №4, pp. 58-68.