

ПІДХІД ДО ПОБУДОВИ СИСТЕМИ ПОВТОРНОЇ ІДЕНТИФІКАЦІЇ ЛЮДЕЙ В ПРИМІЩЕННІ

Остапенко М.С., Штогріна О.С.

Інститут телекомунікаційних систем КПІ ім. Ігоря Сікорського, Україна

E-mail: mak7.ostapenko@gmail.com

Approach for building an indoor person re-identification system

In this paper, we propose the whole pipeline of person re-identification system. Re-identification problem is one of the oldest problems in computer vision. The main contributions of the paper first are describing challenges during building and adjusting system and second is the approach for choosing the main parts of re-identification pipeline based on research and production experience.

В останні роки спостерігається підвищена зацікавленість до систем повторної ідентифікації в області безпеки та персоналізації послуг різними установами. Для забезпечення безпеки найбільш широко використовуються системи ідентифікації на основі електромагнітних перепусток. Їх недоліки це відсутність гарантії того що перепустка не попаде до сторонніх осіб та не буде підроблена. Для покращення надання персоналізованих послуг використовуються картки лояльності. Проблема пов'язана із картами лояльності - складність їх поширення. Тому для вирішення задачі повторної ідентифікації зростає зацікавленість в системах, які базуються на комп'ютерному зорі.

В даній роботі було взято за мету проаналізувати складові для системи комп'ютерного зору повторної ідентифікації (СКЗПІ), описати основні проблеми при їх побудові та описати розроблену СКЗПІ.

СКЗПІ приймають на вхід відеопотік з камер відеоспостереження, а видають аналітику пов'язану із людьми, які були в кадрі. Переваги таких систем перед іншими: можливість до масштабування, практично повна автономність системи, ліквідація витрат на поширення електронних перепусток та їх підтримку. Зовнішні фактори, що впливають на якість і складність СКЗПІ:

- 1) потужність обчислювальних ресурсів для аналізу відеопотоку;
- 2) положення камери в просторі по відношенню до цільових об'єктів;
- 3) роздільна здатність камери та швидкість зйомки;
- 4) наявність світло відбиваючих поверхонь;
- 5) умови освітленості в приміщенні.

Основна ідея СКЗПІ полягає в описанні унікальної людини унікальним вектором-дескриптором і прийняття рішення ідентифікації порівнюючи дескриптори. В загальному СКЗПІ складається із наступних частин: детектор, модель відстеження або кластеризації, модель отримання дескрипторів

об'єкта, модель пошуку в базі даних (БД) та модель прийняття рішення авторизації.

Базовим детектором лиць та його ключових точок є MTCNN [1]. Під час роботи із MTCNN були виявлені його недоліки: багато негативних детекцій, та недостатня швидкість для систем реального часу, тому що в його основі лежать три послідовні нейронні мережі (НМ). Детектор SSD [2] за рахунок базової НМ та ієрархічної структури прийняття рішення є швидшим та точнішим за MTCNN, але не має блоку знаходження ключових точок лиця. Архітектура детектора YOLO [3] схожа на SSD, за виключенням відсутності ієрархічного прийняття рішення. YOLO більш швидкий за SSD, але програє йому по точності детекції.

Відомим алгоритмом відслідковування об'єктів є KLT Tracker [4]. Із-за того що після ініціалізації об'єктів KLT Tracker намагається працювати паралельно із детектором, то його слід використовувати при великій кількості негативних спрацювань детектора. Більш простішим та швидким є IoU Tracker [5]. Його доцільно використовувати при швидкому та точному детекторі.

В якості базової моделі отримання дескрипторів рекомендовано обирати згорткові НМ [6] типу ResNet [7] переднавчені на наборі даних ImageNet на 1200 класів. Необхідно, щоб дескриптори, які належать до одного класу були сконцентровані із малою дисперсією в певній області гіперпростору, а дескриптори, які належать до різних класів, мали достатньо велику дисперсію в гіперпросторі, щоб їх можна було розділити гіперплощиною для класифікації. L2-constrained Softmax Loss [8] - стандартна функція помилок для класифікації, але вона не забезпечує попередньо висунутих умов. Center Loss [9] дозволяє згуртувати всі дескриптори одного класу в певній частині гіперсфери. Triplet Loss [10] дозволяє відділити між собою дескриптори різних класів.

В якості моделі пошуку схожих дескрипторів найбільш простою є k-Nearest Neighbors [11]. Його основний недолік - швидкість пошуку у великих БД. Для прискорення пошуку можна використовувати кластеризацію або індексацію дескрипторів. Один із найефективніших наближених алгоритмів пошуку є Faiss [12]. Важлива перевагою Faiss є можливість паралельного пошуку та виконання на графічних процесорах.

Для порівняння дескрипторів та прийняття рішення ідентифікації зазвичай використовують порогове порівняння дескрипторів, моделі бінарної класифікації та поєднання різних моделей в ансамблі.

В запропонованому підході СКЗПІ на основі проведеного аналізу були вибрані основні частини. Для детекції лиць обраний SSD. Використовуючи техніки одночасного навчання НМ [13], до SSD був доданий блок детекції ключових точок лиця. Для відслідковування об'єктів між кадрами був обраний IoU Tracker, оскільки детектор є досить точним.

В якості базової моделі отримання дескрипторів була обрана НМ ResNet18. Кількість нейронів в передостанньому повнозв'язному шарі НМ була вибрана 512, оскільки 128 і 256 під час експериментів виявились не

достатніми, а більше 512 нейронів не призвели до значного зростання якості. Далі НМ була натренована на класифікацію людських лиць використовуючи набір даних WIDER FACE [14]. Під час навчання в якості метрики помилок використовувалась зважена сума L2-constrained Softmax Loss, Center Loss та Triplet Loss.

Для швидкого пошуку схожих дескрипторів в БД був обраний Faiss. Далі для прийняття рішення авторизації був розроблений ансамбль методів. Методи, що входять в ансамбль прийняття рішень: попарне порогове порівняння дескрипторів, CatBoost [15] поверх категоризованих дескрипторів.

Запропонований підхід дозволяє побудувати СКЗПІ з можливістю подальшого використання в реальному світі. Подальша робота над описаною СКЗПІ буде полягати в продовженні експериментів з моделями розпізнавання лиць та пошуку по БД підвищенню їх якості та швидкості роботи.

Література

1. Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, Yu Qiao "Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks". arXiv:1604.02878, 2016.
2. Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg "SSD: Single Shot MultiBox Detector". arXiv:1512.02325, 2016.
3. Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi "You Only Look Once: Unified, Real-Time Object Detection". arXiv:1506.02640, 2015.
4. Carlo Tomasi and Takeo Kanade. Detection and Tracking of Point Features. Carnegie Mellon University Technical Report CMU-CS-91-132, April 1991.
5. E. Bochinski, V. Eiselein, T. Sikora. High-Speed Tracking-by-Detection Without Using Image Information. In International Workshop on Traffic and Street Surveillance for Safety and Security at IEEE AVSS 2017, 2017.
6. A. Krizhevsky, I. Sutskever, and G. Hinton "Imagenet classification with deep convolutional neural networks". In NIPS, 2012.
7. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun "Deep Residual Learning for Image Recognition". arXiv:1512.03385, 2015.
8. Rajeev Ranjan, Carlos D. Castillo, Rama Chellappa "L2-constrained Softmax Loss for Discriminative Face Verification". arXiv:1703.09507, 2017.
9. Yandong Wen, Kaipeng Zhang, Zhifeng Li and Yu Qiao. A Discriminative Feature Learning Approach for Deep Face Recognition. The Chinese University of Hong Kong, Sha Tin, Hong Kong, 2016.
10. Florian Schroff, Dmitry Kalenichenko, James Philbin "FaceNet: A Unified Embedding for Face Recognition and Clustering". arXiv:1503.03832, 2015.
11. Cover TM, Hart PE (1967). Nearest neighbor pattern classification. IEEE Transactions on Information Theory. 13 (1): 21–27.
12. Johnson, Jeff and Douze, Matthijs and J'egou, Herv'e "Billion-scale similarity search with GPUs". arXiv:1702.08734, 2017.
13. Sebastian Ruder "An Overview of Multi-Task Learning in Deep Neural Networks". arXiv:1706.05098, 2017.
14. Yang, Shuo and Luo, Ping and Loy, Chen Change and Tang, Xiaoou. WIDER FACE: A Face Detection Benchmark. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
15. Anna Veronika Dorogush, Vasily Ershov, Andrey Gulin "CatBoost: gradient boosting with categorical features support". Workshop on ML Systems at NIPS 2017.