

<https://doi.org/10.31891/2219-9365-2024-77-9>

УДК 621

МРАК Василь

Національний університет «Львівська політехніка»

<https://orcid.org/0009-0002-6066-5592>

vasyl.b.mrak@lpnu.ua

КЛИМАШ Михайло

Національний університет «Львівська політехніка»

<https://orcid.org/0000-0003-2867-1482>

mykhailo.m.klymash@lpnu.ua

МОДЕЛЬ СИСТЕМИ РОЗПІЗНАВАННЯ ОБЛИЧЧЯ ДЛЯ НЕРУХОМИХ СИСТЕМ ВІДЕОСПОСТЕРЕЖЕННЯ

У сучасних системах відеоспостереження використовуються методи глибокого навчання. Такі рішення створюють багато задач, особливо доступності обчислювальних ресурсів. З появою підходів глибокого навчання ці задачі стали ще складнішими, це через те, що алгоритми глибокого навчання базуються на штучних нейронних мережах і потребують багато специфічних ресурсів, таких як відеокарти (графічні обчислювальні пристрої). Для подолання цих задач пропонуються рішення, які вимагають меншого використання ресурсів. Це передбачає додавання алгоритму виявлення руху, що призводить до обмеження подальших обчислень лише до зони інтересу, тобто області, що містить людей, автомобілі тощо. У цій статті пропонується модель розпізнавання обличчя на основі алгоритму виявлення руху для нерухомих камер відеоспостереження. Дослідження спрямоване лише на людей і не охоплює всі типи об'єктів.

Ключові слова: глибоке навчання, розпізнавання обличчя, видалення фону, рухомий об'єкт, системи відеоспостереження.

MRAK VasyI, KLYMASH Mykhailo

Lviv Polytechnic National University

FACIAL RECOGNITION SYSTEM MODEL FOR FIXED VIDEO SURVEILLANCE SYSTEMS

To create intelligent video surveillance systems, it is necessary to perform a certain number of tasks related to computer vision. One of the main tasks is the detection of moving objects. This is usually done through pixel segmentation, which allows you to separate pixels of interest from others. These pixels are usually called foreground pixels, while the others are called background pixels. This has led to the development of many approaches by scientists, each of them trying to overcome the challenge in a specific scenario, such as complex background, high illumination, poor quality input image. This process is called background subtraction. The proposed approach will process a collection of images from a video, called frames, and should be fast enough to be applied in real time. Among the mentioned operations there is detection and localization in the case of processing a specific object. If we are dealing with people, the detection will involve localizing the faces in the input frame, which processes the entire frame to find the positions of the faces. In large video surveillance systems, this creates a problem of using resources, especially memory and computing power. The idea is to reduce the number of video sequences to be processed by limiting further computation of the input frame to the region of interest only. This paper proposes an approach that achieves these goals and shows the impact of such operations on resource consumption as well as the accuracy of the results.

Modern video surveillance systems utilize deep learning methods. Such solutions pose many challenges, especially regarding computational resource availability. With the emergence of deep learning approaches, these challenges have become even more complex because deep learning algorithms are based on artificial neural networks and require a lot of specific resources, such as GPUs (Graphics Processing Units). To overcome these challenges, solutions requiring less resource consumption are proposed. This involves adding a motion detection algorithm, which leads to restricting further computations only to the area of interest, i.e., the area containing humans, cars, etc. This article proposes a model for face recognition based on motion detection for stationary surveillance cameras. The study is focused solely on humans and does not cover all types of objects.

Keywords: deep learning, face recognition, background subtraction, moving object, video surveillance systems.

Постановка проблеми у загальному вигляді

та її зв'язок із важливими науковими чи практичними завданнями

Для створення інтелектуальних систем відеоспостереження необхідно виконати певну кількість задач, що стосуються комп'ютерного зору. Одна з основних задач - це виявлення рухомих об'єктів. Зазвичай це відбувається через сегментацію пікселів, що дозволяє відокремлювати пікселі інтересу від інших. Ці пікселі зазвичай називаються переднім планом, тоді як інші - пікселями фону. Це призвело до розробки багатьох підходів вченими, кожен з них намагається подолати виклик у конкретному сценарії, такому як складний фон, високе освітлення, погана якість вхідного зображення. Цей процес називається відніманням або усуненням фону. Запропонований підхід оброблятиме колекцію зображень з відео, які називаються кадрами, і повинен бути виконаний достатньо швидко, щоб його можна було застосувати реальному часі. Серед згаданих операцій існує виявлення та локалізація у випадку обробки конкретного об'єкта. Якщо маємо справу з людьми, виявлення буде включати локалізацію обличчя у вхідному кадрі, який обробляє весь

кадр, щоб знайти положення облич. У великих системах відеоспостереження це створює проблему використання ресурсів, особливо пам'яті та обчислювальної потужності. Ідея полягає в тому, щоб зменшити кількість відеопослідовностей, що підлягають обробці, обмеживши подальші обчислення вхідного кадру лише до області інтересу. У цій статті пропонується підхід, який досягає цих цілей і показує вплив таких операцій на споживання ресурсів, а також точність результатів.

Аналіз досліджень та публікацій

Алгоритми віднімання фону можна розділити на параметричні та непараметричні. Параметричні (також відомі як ймовірнісні підходи) стосуються методів, які моделюють фон за допомогою нормального (гаусівського) розподілу значень інтенсивності пікселів зображення [1], тоді як непараметричні методи моделюють пікселі за значеннями вибірки та середнім значенням обчислень. Піксель класифікується як належний до фону або до переднього плану. Деякі ймовірнісні підходи включають GMM [2], TLGMM [3], STGMM [4], а непараметричні - [5], байєсівський підхід [6], підхід оцінки щільності ядра (KDE) [23]. Детальнішу інформацію про алгоритми віднімання фону можна знайти в [1], [7].

Виявлення рухливих об'єктів передусе виявленню або локалізації обличчя. Ці підходи можна класифікувати як методи Viola John [8], які були одними з перших запропонованих. Тут використовуються певні шаблони Adaboost[9], для перевірки наявності або відсутності обличчя на вхідному кадрі. Інший метод на основі кольору шкіри, де [10] для пошуку облич, використовується метод, заснований на кольорі, порівнюючи колір обличчя з фоном, використовуючи той факт, що у людських облич значна колірна розподіленість, яка відрізняється від фонових об'єктів. Інша класифікація - це використання методу опорних векторів SVM [11] з альтернативою, що називається підходом на основі вигляду [12].

Алгоритмів розпізнавання обличчя настільки багато, що їх можна розглядати як методи на основі вигляду, на основі ознак та штучних нейронних мереж. У методах на основі вигляду зображення представлено з математичним розподілом, щоб здійснити зменшення розмірності. У цьому контексті можна виділити такі методи, як метод Фішера [13], аналіз головних компонент [14] та метод опорних векторів [11]. Підхід на основі ознак, намагається знайти вектор, який зазвичай називається вектором ознак для кожного зразка в базі знань. Деякі підходи, пов'язані з цим, - це модель прихованих марківських моделей (HMM), активна модель вигляду (AAM) - двовимірна змінна та тривимірна змінна модель [15]. Існують підходи на основі штучних нейронних мереж для знаходження інтуїтивного представлення обличчя. Деякі з найвідоміших - це модель Facenet [16], модель VGG Google [17], модель OpenFace [18] та модель MobileNet [19]. Детальніша інформація про розпізнавання обличчя доступна в [15], [17].

Формулювання цілей статті

Метою даної роботи - представити та дослідити модель розпізнавання обличчя, засновану на виявленні руху. Ця модель використовує алгоритми виявлення руху для локалізації та відстеження об'єктів в кадрі, а потім застосовує методи розпізнавання обличчя для ідентифікації цих об'єктів.

Виклад основного матеріалу

Модель розпізнавання обличчя на основі видалення фону базується в основному на шести кроках, на відміну від чотирьох кроків у звичайному підході. У звичайному підході перший крок - це попередня обробка вхідного зображення, яка може включати фільтрацію, зміну розміру тощо, за якою слідує алгоритм виявлення обличчя, крок вилучення ознак і завершується розпізнаванням обличчя або ідентифікацією особи рис. 1. Крок виявлення служить для виявлення об'єкта, який потрібно аналізувати або ідентифікувати. Це може включати особу, автомобіль, тварину. У випадку цього дослідження основний акцент робиться на виявленні обличчя, оскільки модель використовується у відеоспостереженні. Таким чином, другий крок буде полягати у виявленні обличчя за допомогою одного з підходів, описаних у [21] [22]. Заключний крок полягає у ідентифікації виявлених обличчя. Цей крок також відомий як крок класифікації. У випадку відеоспостереження за особами крок ідентифікації буде полягати у визначенні осіб. Іншими словами, він буде полягати у наданні імені виявленій особі.

Для обробки відеопотоків, безумовно, потрібно більше ресурсів, ніж у звичайних випадках. Відповідно, розумним буде пропонувати рішення, які зменшать споживання зазначених ресурсів. У випадку відеоспостереження вважається розумним спочатку виявляти наявність рухомих об'єктів, перед запуском обробки на обрізаному вхідному кадрі, отриманому після сегментації пікселів. Якщо виявлені рухомі об'єкти, відбудеться обмеження зони інтересу, яка в майбутньому сформує нове вхідне зображення. Для цього використовується набір передніх пікселів, виявлених сегментацією, що впливає з ViBE [20]. На цьому етапі є бінарна матриця розміром вхідного кадру, де передні пікселі у білому (255), а фонові пікселі - у чорному (0). Потім обчислюється площа, зайнята цими передніми пікселями, після чого виконується виявлення країв цієї області. Як тільки ця зона буде обмежена, будь-який піксель всередині вважається пікселем інтересу. Таким чином виконується вилучення цієї зони інтересу за допомогою механізму обрізки, який матиме той самий розмір, що і вхідний кадр.



Рис. 1. Етапи виявлення та розпізнавання обличчя

Архітектура моделі розпізнавання обличчя представлена на рисунку 2 і складається з наступних кроків:

- Крок 1 Попередня обробка: Як у звичайному сценарії, вона включає всі механізми попередньої обробки, які можуть включати отримання зображення, шумопониження тощо.
- Крок 2 Виявлення руху: Це полягає в виявленні рухомих об'єктів, оскільки дослідження проводиться в контексті відеоспостереження, ідея полягає в тому, щоб сфокусуватися тільки на рухомому об'єкті та проігнорувати решту зображення. Використовується підхід віднімання фону, який був запропонований в [20].
- Крок 3 Прийняття рішення: Це полягає в зупинці всього процесу, якщо об'єктів для аналізу немає, і поверненні до кроку 1, що допомагає заощадити обчислювальні ресурси. У випадку виявлення рухомих об'єктів слідує сегментація та вилучення, виконується вилучення зони зображення, що цікавить.
- Крок 4 Виявлення обличчя: Тепер, коли отримано зображення (кадр), алгоритм продовжується, виконуючи виявлення обличчя для пошуку людського обличчя на зображенні. Тут використаний алгоритм CenterFace [22], який складається з каскаду нейронних мереж для ефективного виявлення обличчя.
- Крок 5 Вилучення ознак: З нормалізованого обличчя необхідно за допомогою глибокої нейронної мережі вилучити вектор ознак, які будуть достовірно його описувати для подальшого розпізнавання. Для цього була обрана модель ResNet [23]. Вона була навчена за допомогою функцій втрат ArcFace [24].
- Крок 6 Нарешті, виконується розпізнавання виявленої особи. Проаналізувавши переваги та недоліки методів, іоцільно використовувати для невеликої кількості класів SVM, через те, що він описується математично і витрачає менше ресурсів і пам'яті [25].

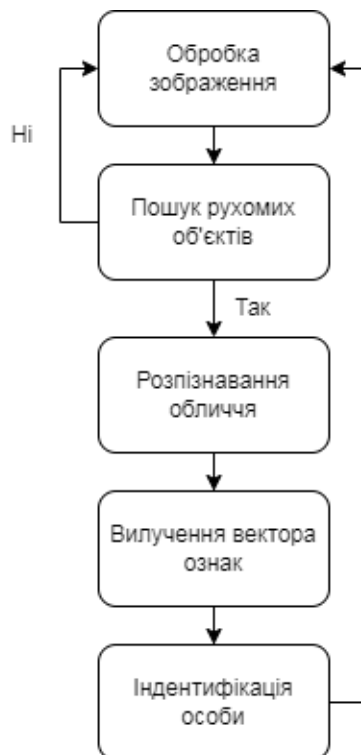


Рис. 2. Загальна архітектура запропонованого підходу

Дослідження впливу віднімання фону

Для тестування використовувалось відео тривалістю декілька хвилин, до якого застосовувався алгоритм сегментації ViBE [20]. Цей алгоритм виділяє зону інтересу в кадрі, яка є частиною, що містить рухомий об'єкт. Тоді було помічено, що з точки зору використання системної пам'яті використання підходу фонового віднімання не впливає суттєво на пам'ять, яка використовується для всього процесу, оскільки цей підхід є легким. Це показано на малюнку 3, де можна побачити зміну споживання пам'яті з відніманням фону та без нього. З іншого боку, було зроблено запис розмірів зображень, які обробляються під час операції. Це дозволило помітити, що при застосуванні алгоритму віднімання фону, розмір обробленого зображення менший, ніж розмір зображення, яке використовується без підходу віднімання фону, завдяки обробці лише частина зображення.

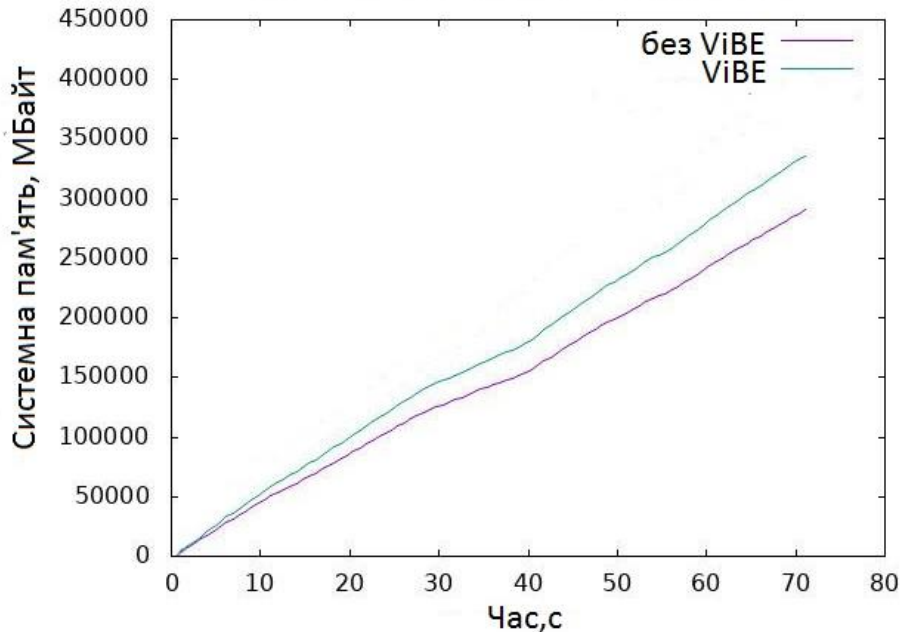


Рис. 3. Використання системної пам'яті в залежності від часу

На цьому малюнку видно, що зображення, обрізане за допомогою ViBE, займає меншу кількість системної пам'яті. Також на малюнку 4 представлено зміну розміру зображення, і можна побачити значний ефект використання підходів віднімання фону, тому що без них розглядається весь кадр (зображення).

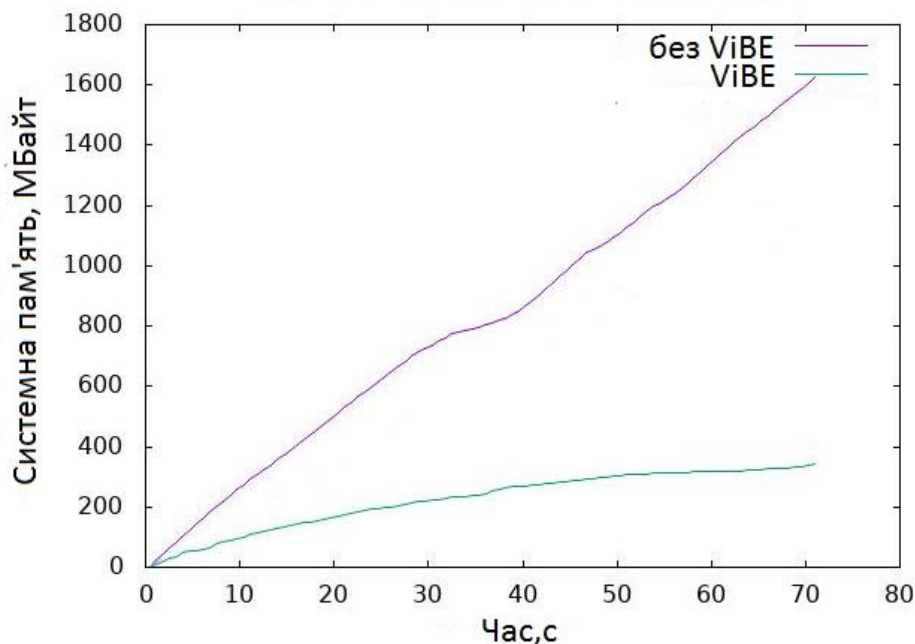


Рис. 4. Зміна розміру використовуваного зображення в залежності від часу

Висновки з даного дослідження і перспективи подальших розвідок у даному напрямі

У роботі представлено модель розпізнавання обличчя, заснована на виявленні руху. Цей метод зосереджується не лише на етапах виявлення та розпізнавання облич, але включає виявлення рухомих об'єктів, що зменшує обсяг даних, які потрібно обробити. Цей підхід має ряд переваг, таких як: зменшення обсягу даних, підвищення швидкості, зниження споживання енергії, універсальність, покращення продуктивності. Виявлення рухомих об'єктів дозволяє обробляти лише ті частини зображення, де відбуваються зміни, що значно зменшує обсяг даних, які потребують обробки. Система може обробляти зображення значно швидше, що робить її більш придатною для задач реального часу. Обробка меншого обсягу даних потребує менше енергії, що робить цю модель більш енергоефективною. Модель може бути адаптована до різних середовищ та умов освітлення. Завдяки поєднанню виявлення руху з методами розпізнавання облич, модель може підвищити точність за допомогою використання додаткової інформації про рух може допомогти усунути помилки, пов'язані з шумом та складними фоновими зображеннями. Зменшити кількість помилкових спрацювань за рахунок реагування лише на ті об'єкти, які дійсно рухаються, що мінімізує ймовірність помилкового розпізнавання облич.

Література

1. Thierry Bouwmans, Recent Advanced Statistical Background Modeling for Foreground Detection: A Systematic Survey, Computer Science Review, vol. 4 pp 147-176 (2011) [13] Chris Stauffer, Eric Grimson, Adaptive background mixture models for real-time tracking, Computer Vision and Pattern Recognition, pp 246-252 (1999)
2. Bo Han, Xinggang Lin, Update the GMMs via adaptive Kalman filtering, International Society for Optical Engineering, pp 1506-1515 (2005)
3. Yang Hong, Yihua Tan, Jinwen Tian, Jian Liu, Accurate dynamic scene model for moving object detection, International Conference on Image Processing (ICIP), pp 157-160 (2007)
4. Wei Zhang, Xiangzhong Fang, Xiaokang Yang, Jonathan Wu, Spatio-temporal Gaussian mixture model to detect moving objects in dynamic scenes, Journal of Electronic Imaging, (2007)
5. O. Barnich, M. V. Droogenbroeck, Vibe: A powerful random technique to estimate the background in video sequences, International Conference on Acoustics, Speech, and Signal Processing (ICASSP)(2009) 19–24
6. Jong Geun Park, Chulhee Lee, Bayesian rule-based complex background modeling and foreground detection Optical Engineering, Optical Engineering, (2010)
7. BowmaZhihao Wang, Jian Chen, Steven C.H. Hoi, "Deep Learning for Image Super-resolution:A Survey", IEEE Transactions on Pattern Analysis and Machine Intelligence PP(99):1-1 March 2020.
8. Cui-huan DU, Hong ZHU, Li-ming LUO, Jie LIU, Xiang-yang HUANG, Face detection in video based on AdaBoost algorithm and skin model, The Journal of China Universities of Posts andTelecommunications, Vol. 20 pp 6-9 (2013)
9. Yi-Qing Wang, An Analysis of Viola-Jones Face Detection Algorithm, Published in Image Processing On Line, pp:128-148, August 31, (2013)
10. YuseokBan, Sang-KiKim, SooyeonKim, Kar-AnnToh, Face detection based on skin color likelihood, Pattern Recognition, vol. 47, pp: 1573-1585 (2014)
11. Mohammed Javed, Bhaskar Gupta, Performance Comparison of Various Face Detection Techniques, International Journal of Scientific Research Engineering Technology (IJSRET), pp 019-0027, Vol. 2 April (2013)
12. Henry A. Rowley, ShumeetBaluja, Takeo Kanade, Rowley Neural Network-Based Face Detection, IEEE PAMI, (1998)
13. Peter N.Belhumeur, JoaoP.Hespanha, David Kreigman, Eigenfaces vs. Fisherfaces Recognition using class specific Linear Projection, IEEE Trans. PAMI, (1997)
14. M. Turk and A. Pentland, Face recognition using eigenfaces, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (1991)
15. T. Kanade, Computer Recognition of Human Faces, Basel and Stuttgart Birkhauser, (1977)
16. Florian Schroff, Dmitry Kalenichenko, James Philbin, FaceNet: A Unifed Embedding for Face Recognition and Clustering, arXiv (2015)
17. O. M. Parkhi, A. Vedaldi, A. Zisserman, Deep Face Recognition, British Machine Vision Conference, (2015)
18. B. Amos, B. Ludwiczuk, M. Satyanarayanan, Openface: A general-purpose face recognition library with mobile applications, CMU-CS-16-118, CMU School of Computer Science, Tech. Rep., (2016).
19. A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, Mobilenets: Efficient convolutional neural networks for mobile vision, applications. arXiv preprint (2017)
20. Xavier Bou, Thibaud Ehret, Gabriele Facciolo, Jean-Michel Morel, and Rafael Grompone von Gioi, Reviewing ViBe, a Popular Background Subtraction Algorithm for Real-Time Applications, Image Processing On Line, 12 (2022), pp. 527–549. <https://doi.org/10.5201/ipol.2022.434>

21. Методи розпізнавання обличчя у системах відеоспостереження з використанням машинного навчання В. Мрак Інфокомунікаційні технології та електронна інженерія, Вип. 2, № 3, С. 33–42 (2023) <https://doi.org/10.23939/ict2023.02.033>
22. Xu, Y., Yan, W., Yang, G., Luo, J., Li, T., & He, J. (2020). CenterFace: joint face detection and alignment using face as point. *Scientific Programming*, 2020
23. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
24. Deng, J., Guo, J., Xue, N., & Zafeiriou, S. (2019). Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 4690-4699).
25. Yadav, A. (2018, October 22). Support Vector Machines(SVM). Medium. URL: <https://towardsdatascience.com/support-vector-machines-svmc9ef22815589>.

References

1. Thierry Bouwmans, Recent Advanced Statistical Background Modeling for Foreground Detection: A Systematic Survey, *Computer Science Review*, vol. 4 pp 147-176 (2011) [13] Chris Stauffer, Eric Grimson, Adaptive background mixture models for real-time tracking, *Computer Vision and Pattern Recognition*, pp 246-252 (1999)
2. Bo Han, Xinggang Lin, Update the GMMs via adaptive Kalman filtering, *International Society for Optical Engineering*, pp 1506-1515 (2005)
3. Yang Hong, Yihua Tan, Jinwen Tian, Jian Liu, Accurate dynamic scene model for moving object detection, *International Conference on Image Processing (ICIP)*, pp 157-160 (2007)
4. Wei Zhang, Xiangzhong Fang, Xiaokang Yang, Jonathan Wu, Spatio-temporal Gaussian mixture model to detect moving objects in dynamic scenes, *Journal of Electronic Imaging*, (2007)
5. O. Barnich, M. V. Droogenbroeck, Vibe: A powerful random technique to estimate the background in video sequences, *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)(2009)* 19–24
6. Jong Geun Park, Chulhee Lee, Bayesian rule-based complex background modeling and foreground detection *Optical Engineering*, (2010)
7. BowmaZhihao Wang, Jian Chen, Steven C.H. Hoi, "Deep Learning for Image Super-resolution:A Survey", *IEEE Transactions on Pattern Analysis and Machine Intelligence* PP(99):1-1 March 2020.
8. Cui-huan DU, Hong ZHU, Li-ming LUO, Jie LIU, Xiang-yang HUANG, Face detection in video based on AdaBoost algorithm and skin model, *The Journal of China Universities of Posts and Telecommunications*, Vol. 20 pp 6-9 (2013)
9. Yi-Qing Wang, An Analysis of Viola-Jones Face Detection Algorithm, Published in *Image Processing On Line*, pp:128-148, August 31, (2013)
10. YuseokBan, Sang-KiKim, SooyeonKim, Kar-AnnToh, Face detection based on skin color likelihood, *Pattern Recognition*, vol. 47, pp: 1573-1585 (2014)
11. Mohammed Javed, Bhaskar Gupta, Performance Comparison of Various Face Detection Techniques, *International Journal of Scientific Research Engineering Technology (IJSRET)*, pp 019-0027, Vol. 2 April (2013)
12. Henry A. Rowley, ShumeetBaluja, Takeo Kanade, Rowley Neural Network-Based Face Detection, *IEEE PAMI*, (1998)
13. Peter N.Belhumeur, JoaoP.Hespanha, David Kreigman, Eigenfaces vs. Fisherfaces Recognition using class specific Linear Projection, *IEEE Trans. PAMI*, (1997)
14. M. Turk and A. Pentland, Face recognition using eigenfaces, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (1991)
15. T. Kanade, *Computer Recognition of Human Faces*, Basel and Stuttgart Birkhauser, (1977)
16. Florian Schroff, Dmitry Kalenichenko, James Philbin, FaceNet: A Unified Embedding for Face Recognition and Clustering, *arXiv* (2015)
17. O. M. Parkhi, A. Vedaldi, A. Zisserman, Deep Face Recognition, *British Machine Vision Conference*, (2015)
18. B. Amos, B. Ludwiczuk, M. Satyanarayanan, Openface: A general-purpose face recognition library with mobile applications, *CMU-CS-16-118*, CMU School of Computer Science, Tech. Rep., (2016).
19. A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, Mobilenets: Efficient convolutional neural networks for mobile vision, applications. *arXiv preprint* (2017)
20. Xavier Bou, Thibaud Ehret, Gabriele Facciolo, Jean-Michel Morel, and Rafael Grompone von Gioi, Reviewing ViBe, a Popular Background Subtraction Algorithm for Real-Time Applications, *Image Processing On Line*, 12 (2022), pp. 527–549. <https://doi.org/10.5201/ipol.2022.434>
21. Metody rozpoznavannia oblychchia u systemakh videosposterezhennia z vykorystanniam mashynnoho navchannia V. Mrak Infokomunikatsiini tekhnolohii ta elektronna inzheneriia, Vyp. 2, № 3, S. 33–42 (2023) <https://doi.org/10.23939/ict2023.02.033>
22. Xu, Y., Yan, W., Yang, G., Luo, J., Li, T., & He, J. (2020). CenterFace: joint face detection and alignment using face as point. *Scientific Programming*, 2020
23. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
24. Deng, J., Guo, J., Xue, N., & Zafeiriou, S. (2019). Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 4690-4699).
25. Yadav, A. (2018, October 22). Support Vector Machines(SVM). Medium. URL: <https://towardsdatascience.com/support-vector-machines-svmc9ef22815589>.