International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 04 | Apr 2022 www.irjet.net p-ISSN: 2395-0072

Person Acquisition and Identification Tool

Swastik Pattanaik [1], Sachin Mudaliyar [2], Pushpak Pachpande [3], Balasaheb Balkhande [4]

[1,2,3] UG Student, Dept. of Computer Engineering, Bharati Vidyapeeth College of Engineering,Mumbai,India
[4] Professor, Dept. of Computer Engineering, Bharati Vidyapeeth College of Engineering,Mumbai,India

---

Abstract - *Closed Circuit Television (CCTV) has been used in everyday life for various needs. In its development, the use of CCTV has evolved from a simple, passive surveillance system into an integrated intelligent surveillance system. In this article we propose a facial recognition on CCTV video system which can generate timestamp based data on the presence of target individuals, augmented for specific usage and purposes of modern day surveillance scenarios. This is proposed to be done with a three step approach of detection, super resolution and recognition. We also intend to explore the possibility and various outcomes which come from implementation of a Siamese network as a part of face recognition component for recognizing unbounded face identity and subsequently doing one shot learning to record newly recognized identity.*

Key Words: **Surveillance, CCTV, Timestamp, Facial detection, Identity recognition, Super Resolution**

1. INTRODUCTION

Facial expressions and features are one of the most distinct traits that everyone has. Face recognition can be used to identify ownership or to ensure that two face images belong to the same person or not. Today, face recognition is already in use in various fields such as the military, surveillance, mobile security system, etc. The emergence of facial recognition techniques is enhanced as it uses seep learning as the backbone. In 2014 DeepFace [34] was introduced as the first face recognition method using deep learning with really good performance around 97.35% accuracy. This practice is further developed over advanced models such as FaceNet [33], VGGFace[32], and VGGFace2 [24] with precision more than 99%. Most advanced face recognition systems are not designed and trained for low resolution recognition. Although in reality very few systems are capable of capturing high resolution images. Computer performance is another barrier that needs to be considered. Not all face recognition systems operate on supercomputers having multiple GPUs. We need to develop a system that can run on a slow computer and even a cell phone.

In this paper, we will focus on building a comprehensive low-resolution facial recognition system. The complete system contains at least three main components which are face detection, resolution adjustment or image enhancement, and face recognition. We will also compare between various approaches and techniques for each component to decide upon which is the better for facial recognition activity and is the most lightweight model that can work on cheaper devices. The Siamese network will be used as part of the face recognition feature so that our system can detect unbounded identity and perform one-shot learning to record the feature of the newly identified identity. This will be done with the end goal of providing laymen to use the tool with ease from mobile devices.

So to summarise it all, Person Identification and Acquisition tool is a proposed solution for cutting down hours of manpower invested in video footage scouring during several day to day law enforcement and defence scenarios. It can be adapted and modified for various other purposes with foremost example of such an application being semi-tracking and tracing of a target individual based on video based data.

2. LITERATURE REVIEW

A. Hinori Hattori (2018)[1] worked on a pedestrian detector and pose estimator system for static video surveillance which in reference to the work we intend to do proposes a solution for scenarios where there are zero instances of real pedestrian data (e.g., a newly installed surveillance system in a novel location in which no labeled real data or unsupervised real data exists yet) and a pedestrian detector must be developed prior to any observations of pedestrians.

B. Kamta Nath Mishra (2019)[2] this study proposes a solution for human identification based on the human face recognition in images extracted from conventional cameras at a low resolution and quality through optimal super resolution techniques and proposes pipelines which can help in the process.

C. P. Satyagama (2020)[8] seemed to explore the concepts of using Siamese network as a part of face recognition component for recognizing unbounded face identity and subsequently doing one shot learning to record newly recognized identity while addressing the issue of low resolution recognition scenarios.

D. Anurag M. (2002)[18] presented a method that uses multiple synchronized cameras to track all the people in a cluttered scene while segmenting, detecting and detecting their movement at the same time. It introduces an algorithm based on region

data that can be used to search for 3D points inside an object if we are aware of the regions within the object from two different viewpoints. People were constrained to move in only a small region..

E. Koen Buys (2014)[10] approach relies on an underlying kinematic model. This approach uses an additional iteration of the algorithm that segments the body from the background. It presents a method for RDF training, including data generation and cluster-based learning that enables classifier retraining for different body models, kinematic models, poses or scenes.

3. GAP ANALYSIS

Facial expressions and features are one of the most unique traits that everyone has. Face recognition can be used to identify ownership or to ensure that two face images belong to the same person or not.

Today, face recognition is already in use in various fields such as the military, surveillance, mobile security system, etc. The emergence of facial recognition techniques is enhanced as it uses in-depth learning as the backbone.

A. In 2014 DeepFace was introduced as the first face recognition method using in-depth learning with really good performance around 97.35% accuracy. This practice is further developed over advanced models such as FaceNet, VGGFace, and VGGFace2 with precision performance affecting more than 99%.

B. Most advanced face recognition systems are not designed and trained for low vision correction. Although in the case of actual use, not all face recognition systems can achieve a high-resolution facial image. Computer performance is another barrier that needs to be considered. Not all face recognition systems operating on supercomputers have multiple GPUs.

C. We need to develop a system that can run on a slow computer and even a cell phone and/or on distributed systems.

D. In this proposal, we will focus on building a comprehensive low-resolution facial recognition system. The complete system contains at least three main components which are face detection, resolution adjustment or image enhancement, and face recognition.

E. We will also compare between strategies for which each component collects information on which is the best face-to-face facial recognition activity and which is the most lightweight model that can work on cheaper devices.

F. The Siamese network will be used as part of the face recognition feature so that our system can detect unlimited identity and perform one-on-one reading to record the feature of the newly identified identity. This will be done with the end goal of providing laymen to use the tool with ease from mobile devices

4. OBJECTIVES

A. Given still(s) & video images of a scene, identify or verify one or more target individuals of whom the still(s) have been provided for. The solution should look for optimization for implementation on CCTV footage (enhancing recognition). Certain conditions must be satisfied in the output in the post recognition stage:-

- The system needs to report back the decided identity from the input of target (known) individuals.
- Timestamps indicating the presence of the suspected match of target individual must be intimidated to the user.

B. The primary objective of the system is to create a solution which can provide timestamp based data about the presence of the target individual when provided with a facial sample of the target individual and the footage to be scoured.

C. The secondary objective is to cut down time and effort in several Law enforcement scenarios which arise in due course of any case/situation in major metropolitan cities in India by empowering the foot-soldiers with accurate and easy to use tools.

5. PROPOSED METHODOLOGY

Our approach is based on experimental research methods, where we experimentally decide on the best possible technique for each stage of the proposal based on a certain dataset. We will then evaluate the result of the whole system with different configuration using accuracy and execution time metrics.

The system can be broken down into three major functional components:

A. Face detection

B. Super resolution

C. Face recognition

The basic approach for such a system, with every component connected into a pipeline so that a full system of Low resolution face recognition is as shown in the flowchart below

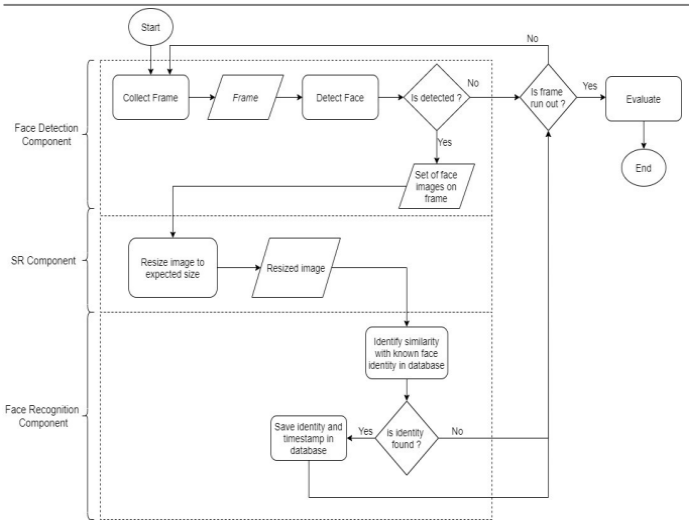

Fig -1: Proposed Pipeline

A. First stage of system pipeline is a face detection component which serves the roles of collecting the frames from the video at specified rate based on system capabilities or form a direct camera feed and detect every face on each frame and yield a set of cropped face images that are detected on frame

B. In the Second stage, the face images are needed to be uniformed to one size by using Super resolution component. It serves the role to resize all detected face images, by the use of two main operations: Upscale and downscale. Upscaling is executed with a deep-learning based technique with purpose to collect better low resolution image and Downscaling is done with standard bicubic interpolation technique.

C. The third and the final stage of the pipeline runs the Face recognition component. The role of this stage is to identify the face image with known face identities that are already recorded in database or provided by the user, and then save the recognition log to a database or throw output immediately intimidating the presence of target individual in the frame ,with a timestamp. Here, the resized images given to this component will be extracted to face feature vectors by a deep-learning based face feature extraction model and then the resulted vector and every existing face feature vector in the database or input will be fed to a Siamese classifier. The classifier will yield the confidence rate to determine whether the two face features belong to same identity or not

6. CONCLUSIONS

In conclusion the proposed project wishes to address and positively solve the lack of modern investigatory tools which use technologies which define today's world the way we know it. Here we address the specific lack of a truly malleable tool which can assist "non tech savvy" or computer illiterate personnel in dynamic video based evidence gathering or tracking-tracing scenarios which surface on a day to day basis in any law enforcement organization in the world which is tasked to a major metropolitan city.

REFERENCES

[1] Hironori Hattori, Namhoon Lee, Vishnu Naresh Boddeti, Fares Beainy,Kris M Kitani, Takeo Kanade, "Synthesizing a Scene-Specific Pedestrian Detector and Pose Estimator for Static Video Surveillance",International Journal of Computer Vision, pp. 1-18, 2018.

[2] Kamta Nath Mishra, "An Efficient Technique for Online Iris Image Compression and Personal Identification," in Proceedings of International Conference on Recent Advancement on Computer and Communication, pp. 335-343, 2018.

[3] A.Robert Singh, A. Suganya, "Efficient Tool For Face Detection And Face Recognition In Color Group Photos", In IEEE proceedings of third International Conference on Electronics Computer Technology(ICECT), Kanyakumari, India, pp. 263-265, 2011.

[4] B.Balkhande, D.Dhadve, P.Shirsat, M.Waghmare, "A Smart Surveillance System," International Journal of Recent Technology and Engineering, vol-9, no. 1, pp. 1135-38,May 2020, ISSN: 2277-3878.

[5] R.Paunikar, S.Thakare, B.Balkhande, U.Anuse, "Literature Survey On Smart Surveillance System," International Journal of Engineering Applied Sciences and Technology, vol. 4, no. 12, pp. 494-496, April 2020, ISSN No. 2455-2143

[6] P. Kakumanu, S. Makrogiannis, N. Bourbakis, "A survey of skin-color modeling and detection methods", In Journal of Pattern Recognition,Elsevier, pp. 1106-1122, 2007.

[7] O. Manyam, N. Kumar, P. Belhumeur, D. Kriegman, "Two faces are better than one: Face recognition in group photographs", In IEEE proceedings of International Joint Conference on Biometrics (IJCB),Washington, USA, pp. 1-8, 2011.

[8] P. Satyagama and D. H. Widyantoro, "Low-Resolution Face Recognition System Using Siamese Network," 2020 7th International Conference on Advance Informatics: Concepts, Theory and Applications (ICAICTA), 2020, pp. 1-6, doi: 10.1109/ICAICTA49861.2020.9428885

[9] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[10] K. Buys, C. Cagniart, A. Baksheev, T.-D. Laet, J. D. Schutter and C. Pantofaru, "An adaptable system for RGB-D based human body detection and pose estimation," Journal of

visual communication and image representation, vol. 25, pp. 39-52, Jan 2014.

[11] A. Jalal, Y.-H. Kim, Y.-J. Kim, S. Kamal and D. Kim, "Robust human activity recognition from depth video using spatiotemporal multi-fused feature," Pattern recognition, vol. 61, pp. 295-308, 2017.

[12] B. Enyedi, L. Konyha and K. Fazekas, "Threshold procedures and image segmentation," in proc. of the IEEE International symposium ELMAR, pp. 119-124, 2005.

[13] A. Jalal, and S. Kamal, "Real-time life logging via a depth silhouettebased human activity recognition system for smart home services," in Proceedings of AVSS, Korea, pp. 74-80, Aug 2014.

[14] A. Sony, K. Ajith, K. Thomas, T. Thomas, and P. L. Deepa, "Video summarization by clustering using euclidean distance," in proc. of the SCCNT, 2011.

[15] A. Jalal and S. Kim, "The mechanism of edge detection using the block matching criteria for the motion estimation," in Proceedings of HCI Conference, Korea, pp. 484-489, Jan 2005

[16] Nilam Prakash Sonawale, B.W.Balkhande, "CISRI - Crime Investigation System Using Relative Importance: A Survey," International Journal of Innovative Research in Computer and Communication Engineering, vol 4, no 2, pp-2279-2285, Feb – 2016 ISSN 2320-9798

[17] L. Kaelon, P. Rosin, D. Marshall and S. Moore, "Detecting violent and abnormal crowd activity using temporal analysis of grey level cooccurrence matrix (GLCM)-based texture measures," MVA, vol. 28, no. 3, pp. 361-371, 2017

[18] Anurag Mittal and Larry S. Davis, M2Tracker: A Multi-View Approach to Segmenting and Tracking People in a Cluttered Scene. International Journal of Computer Vision. Vol. 51 (3), Feb/March 2003.

[19] J. Redmon and A. Farhad, "YOLOv3: An Incremental Improvement," Retrieved from: https://pjreddie.com/media/files/papers/YOLOv3.pdf, 2018.

[20] H. Tao, H.S. Sawhney and R. Kumar, Dynamic Layer Representation with Applications to Tracking, Proc. of the IEEE Computer Vision & Pattern Recognition, Hilton Head, SC, 2000.

[21] S. Kamal and A. Jalal, "A hybrid feature extraction approach for human detection, tracking and activity recognition using depth sensors," Arabian Journal for science and engineering, 2016.

[22] Ahn, N., Kang, B., & Sohn, K. A. (2018). Fast, accurate, and lightweight super-resolution with cascading residual network. In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 252-268).

[23] Bevilacqua, M., Roumy, A., Guillemot, C., & Alberi-Morel, M. L. (2012). Low-complexity single-image super-resolution based on nonnegative neighbor embedding.

[24] Cao, Q., Shen, L., Xie, W., Parkhi, Omkar M., & Zisserman, A. (2018). VGGFace2: A dataset for recognising faces across pose and age. arXiv:1710.08092v2.

[25] Cheng, Z., Zhu, X., & Gong, S. (2018, December). Low-resolution face recognition. In Asian Conference on Computer Vision (pp. 605-621). Springer, Cham.

[26] Dong, C., Loy, C. C., He, K., & Tang, X. (2014, September). Learning a deep convolutional network for image super-resolution. In European conference on computer vision (pp. 184-199). Springer, Cham.

[27] Dong, C., Loy, C. C., & Tang, X. (2016, October). Accelerating the super-resolution convolutional neural network. In European conference on computer vision (pp. 391-407). Springer, Cham.

[28] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.

[29] Huang, J. B., Singh, A., & Ahuja, N. (2015). Single image super-resolution from transformed self-exemplars. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 5197-5206).

[30] Huang, G. B., Mattar, M., Berg, T., & Learned-Miller, E. (2008, October). Labeled faces in the wild: A database forstudying face recognition in unconstrained environments.

[31] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg A. C. (2016). Ssd: Single shot multibox detector. ECCV.

[32] Parkhi, Omkar M., Vedaldi, A., & Zisserman, A. (2015). Deep face recognition. bmvc. Vol. 1. No. 3. [12] Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A Unified Embedding for Face Recognition and Clustering. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 815-823.

[33] Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A Unified Embedding for Face Recognition and Clustering. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 815-823.

[34] Taigman, Y., Yang, M., Ranzato, M., & Wolf, L. (2014). Deepface: Closing the gap to human-level performance in face verification. CVPR, page 1701–1708.

[35] Viola, P., & Jones, M. (2001, December). Rapid object detection using a boosted cascade of simple features. In Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001 (Vol. 1, pp. I-I). IEEE.

[36] Yang, S., Luo, P., Loy, C. C., & Tang, X. (2016). Wider face: A face detection benchmark. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 5525-5533).

[37] Yeo, J. (2019). PyTorch implementation of Accelerating the Super-Resolution Convolutional Neural Network. https://github.com/yjn870/FSRCNN-pytorch.

[38] Yeo, J. (2019). PyTorch implementation of Image Super-Resolution Using Deep Convolutional Networks. https://github.com/yjn870/SRCNN-pytorch.

[39] Yixuan, H. (2018). Tensorflow Face Detector. https://github.com/yeephycho/tensorflow-face-detection.

[40] Zangeneh, E., Rahmati, M., & Mohsenzadeh, Y. (2017). Low resolution face recognition using a two-branch deep convolutional neural network architecture. arXiv:1706.06247v1.

[41] Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters, 23(10), 1499-1503.