Human height estimation using AI-assisted computer vision for intelligent video surveillance system

K. Iyshwarya Ratthi [a], B. Yogameena [a],*, S. Saravana Perumaal [b]

a *Department of Electronics and Communication Engineering, Thiagarajar College of Engineering, Madurai 625015, Tamil Nadu, India*
b *Department of Mechanical Engineering, National Institute of Technical Teachers' Training and Research, Chennai 600113, Tamil Nadu, India*

ABSTRACT

In urban areas, technological advancements have led to an increased focus on height as a critical human characteristic for surveillance purposes. Face recognition often encounters challenges due to occlusion and masks, necessitating the use of height, build, and torso. Accurately estimating human height in surveillance scenarios is complex due to camera calibration, posture variations, and movement patterns. This research introduces a novel human height estimation method for surveillance systems, along with a dedicated dataset. The process begins with camera calibration to rectify lens distortions. A deep learning-based YOLOv7-Occlusion Aware (YOLOv7- OA) target detection technique is employed to precisely locate individuals within the frame. The study assesses the impact of camera height and deflection angle on height estimation across different areas of the field of vision (FOV). The proposed method yields a mean absolute error of 0.02 cm to 0.8 cm across various FOV zones, surpassing the previous 1.39 cm benchmark findings.

CCS CONCEPTS

• Computing methodologies • Artificial Intelligence • Computer vision • Computer vision tasks • Biometrics

KEYWORDS

Human height estimation, Computer Vision, Artificial Intelligence, Human height dataset, Video Surveillance, YOLOv7, Camera Calibration

ACM Reference format:

2024. Human height estimation using AI-assisted computer vision for intelligent video surveillance system. In *Proceedings of 15th Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP '24).Bangalore,INDIA,2pages.*https://doi.org/10.1145/1234567890

1 Human height estimation using AI-assisted Computer Vision

The increasing number of missing children has led to a significant increase in the use of CCTV systems. Computer vision algorithms and deep learning methodologies have made it an indispensable tool in person retrieval, enabling efficient identification and retrieval of individuals. Height, a non-intrusive characteristic, can be passively acquired and used for

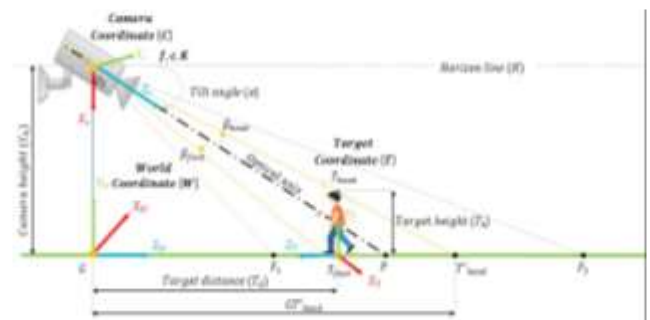

Figure 1: Two-dimensional view of the target distance and height estimation environment depicted following the Right-handed coordinate system that represents various coordinate frames.

Figure 2: The 'Height-Sense' surveillance dataset, gathered from various publicly accessible locations, across various age groups, genders, camera heights, object distances, and ground levels.

identification in situations where facial recognition may not be successful.

However, estimating human height from surveillance footage can be challenging due to camera position, angle, and illumination conditions. This research introduces a novel human height estimation method for surveillance systems based on a mathematical model combined with a monocular surveillance camera. Please refer to the final published version at https://doi.org/10.1016/j.measurement.2024.115133

1.1 Contribution

1. A novel human height estimation method based on a mathematical model combined with a monocular surveillance camera is proposed, which eliminates the need for depth information. The proposed method is depicted in **Figure 1**.

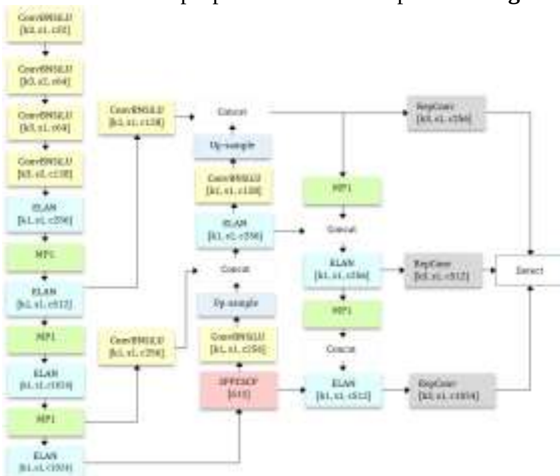

Figure 3: Illustration of the proposed YOLOv7-OA architecture.

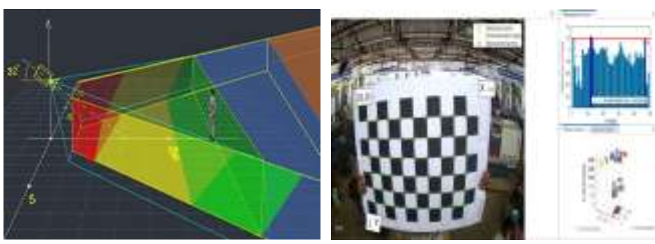

Figure 4: The scene environment generated with IP video design tool and Zhang's calibration method.

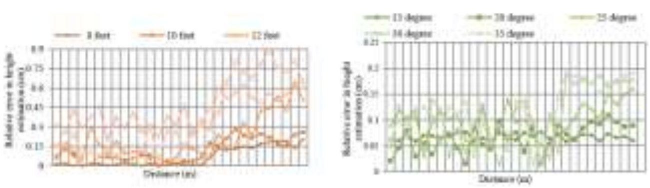

Figure 5: Influence of camera height and tilt angle on human height estimation.

2. A human height surveillance dataset containing camera characteristics such as tilt angle, camera height, and ground truth height of the subjects required for height estimate that matches the complexity of the real-world environment is introduced. This is the first dataset to include a wide range of collections of surveillance video with height information, covering both children and adults. The proposed dataset is depicted in **Figure 2**.
3. Recognizing the importance of occlusion management, the YOLOv7- Occlusion Aware is developed as an upgraded version of YOLOv7 designed specifically to address occlusion difficulties. To assist the model in focusing on relevant regions and contextual information, a hybrid attention mechanism (HAM) is introduced into the feature extraction section of the original YOLOv7. It is shown in **Figure 3**.
4. A detailed experimental examination of various height estimate relevant parameters such as camera height, tilt angle, horizontal and vertical fields, age, and gender is performed. Influence of camera height and tilt angle on human height estimation is depicted in **Figure 4**.
5. To ensure target detection across the full FOV, CCTV's FOV is separated into five zones termed Monitoring (M), Detection (D), Observation (O), and Recognition (R), Identification (I). Zhang's calibration is used to remove any distortion in the video frames so that YOLOv7-OA can recognize and localize pedestrians in complex backgrounds. The scene environment generated with IP video design tool is given in **Figure 5**.
6. A comparison of several human height estimation methodologies and YOLO-based pedestrian detection algorithms is provided, including YOLOv5, YOLOv6, YOLOv7, and YOLOv8 and shown in **Figure 6**.

2 Future Works

The proposed system achieved a mean absolute error (MAE) of 0.02 cm to 0.8 cm across various zones, significantly improving upon the previous benchmark of 1.39 cm. A sample frame is depicted in **Figure 7**. Future work should incorporate posture recognition techniques into height estimation, adjusting algorithms based on upper body measurements and anthropometric ratios. This approach considers human proportions and compensates for seated positions, while also considering surrounding environment cues.

Figure 6: Detection of 'Person' class by YOLOv5, v6, v7 & v8.

Figure 7: The height estimation results obtained inn various zones of the FOV. Row 1 represents Zone 1, and the corresponding rows represent Zone 3, 2 and 1.

References

[1] C. Galiano L´opez, J. Hunter, T. Davies, A. Sidebottom, Further evidence on the extent and time course of repeat missing incidents involving children: a research note, The Police J: Theory, Practice and Principles 96 (1) (2023), https://doi.org/10.1177/0032258x211052900.

[2] L. Boulton, J. Phoenix, E. Halford, A. Sidebottom, Return home interviews with children who have been missing: an exploratory analysis, Police Pract. Res. 24 (1) (2023), https://doi.org/10.1080/15614263.2022.2092480.

[3] M. Shorfuzzaman, M.S. Hossain, M.F. Alhamid, Towards the sustainable development of smart cities through mass video surveillance: a response to the COVID-19 pandemic, Sustain. Cities Soc. 64 (2021).

[4] B. Hassan, E. Izquierdo, T. Piatrik, Soft biometrics: a survey, Multimed Tools Appl. (2021), https://doi.org/10.1007/s11042-021-10622-8.

[5] P. Drozdowski, C. Rathgeb, A. Dantcheva, N. Damer, C. Busch, Demographic bias in biometrics: a survey on an emerging challenge, IEEE Trans on Technol. Soc. 1 (2) (2020) 89‑103.

[6] Günel, S., Rhodin, H., Fua, P. (2018). What Face and Body Shapes Can Tell About Height. *CoRR*, *abs/1805.1*.

[7] Di Bieler, S.G. Gunel, P. Fua, H. Rhodin, Gravity as a reference for estimating a person' s height from video, Proce. IEEE Int. Conference on Comp. Vision (2019), https://doi.org/10.1109/ICCV.2019.00866.

[8] F. Yin, S. Zhou, Accurate estimation of body height from a single depth image via a four-stage developing network, Proce. IEEE Computer Society Conference on Comp. Vision and Pattern Recognition (2020), https://doi.org/10.1109/CVPR42600.2020.00829.

[9] O. Kainz, M. Dopiriak, M. Michalko, F. Jakab, P. Fecil' ak, Estimating the height of a person from a video sequence, 2021 19th Int. Conference on Emerging eLearning Technol. Applications (ICETA) (2021) 150 ‑ 156, https://doi.org/10.1109/ICETA54173.2021.9726680.

[10] Zhu, R., Yang, X., Hold-Geoffroy, Y., Perazzi, F., Eisenmann, J., Sunkavalli, K., Chandraker, M. (2020). Single View Metrology in the Wild. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12356 LNCS*. Doi: 10.1007/978-3-030-58621-8_19.

[11] D.S. Lee, J.S. Kim, S.C. Jeong, S.K. Kwon, Human height estimation by color deep learning and depth 3D conversion, Appl. Sci. (Switzerland) 10 (16) (2020), https://doi.org/10.3390/app10165531.

[12] A.M. Olver, H. Guryn, E. Liscio, The effects of camera resolution and distance on suspect height analysis using PhotoModeler, Forensic Sci. Int. 318 (2021), https://doi.org/10.1016/j.forsciint.2020.110601.

[13] N. Thakkar, H. Farid, On the feasibility of 3D model-based forensic height and weight estimation, IEEE Comp. Society Conference on Comp. Vision and Pattern Recognition Workshops (2021), https://doi.org/10.1109/CVPRW53098.2021.00106.

[14] N. Thakkar, G. Pavlakos, H. Farid, The reliability of forensic body-shape identification, IEEE Comp. Society Conference on Comp. Vision and Pattern Recognition Workshops (2022), https://doi.org/10.1109/CVPRW56347.2022.00014.

[15] J. Wei, J. Jiang, A. Yilmaz, MOHE-Net: monocular object height estimation network using deep learning and scene geometry, Int. Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences – ISPRS Archives 43 (B2-2021) (2021), https://doi.org/10.5194/isprs-archives-XLIII-B2-2021-557-2021.

[16] I.S. Kim, H. Kim, S. Lee, S.K. Jung, HeightNet: monocular object height estimation, Electronics (Switzerland) 12 (2) (2023), https://doi.org/10.3390/electronics12020350.

[17] A. Criminisi, I. Reid, A. Zisserman, Single view metrology, Int. J. Comput. Vis. 40 (2) (2000) 123‑148.

[18] F. Tosti, C. Nardinocchi, W. Wahbeh, C. Ciampini, M. Marsella, P. Lopes, S. Giuliani, Human height estimation from highly distorted surveillance image, J. Forensic Sci. 67 (1) (2022) 332‑344.

[19] Q. Li, L. Mou, Y. Hua, Y. Shi, S. Chen, Y. Sun, X.X. Zhu, 3DCentripetalNet: Building height retrieval from monocular remote sensing imagery, Int. J. Appl. Earth Obs. Geoinf. 120 (2023) 103311.

[20] A Déak, O Kainz, M Michalko, F Jakab. (2017) Estimation of human body height from uncalibrated image. In 2017 15th International Conference on Emerging eLearning Technologies and Applications (ICETA), pages 1‑4. IEEE.

[21] J. Shao, S.K. Zhou, R. Chellappa, Robust height estimation of moving objects from uncalibrated videos, IEEE Trans. Image Process. 19 (8) (2010) 2221‑2232.

[22] Yu. Chai, X. Cao, A real-time human height measurement algorithm based on monocular vision, in: In 2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), IEEE, 2018, pp. 293‑297.

[23] J. Jung, I. Yoon, S. Lee, J. Paik, Object detection and tracking-based camera calibration for normalized human height estimation, J. Sensors 2016 (2016).