

# Gotcha-I: A Multiview Human Videos Dataset

Paola Barra<sup>1</sup>, Carmen Bisogni<sup>1</sup>, Michele Nappi<sup>1</sup>, David Freire-Obregón<sup>2</sup>, and Modesto Castrillón-Santana<sup>2</sup>

<sup>1</sup> Università degli Studi di Salerno, Salerno, 84084, Italy  
`{pbarra, cbisogni, mnappi}@unisa.it`

<sup>2</sup> Universidad de Las Palmas de Gran Canaria (ULPGC), Gran Canaria, 35017, Spain  
`{david.freire, modesto.castrillon}@ulpgc.es`

**Abstract.** The growing need of security in large open spaces led to the need to use video capture of people in different context and illumination and with multiple biometric traits as head pose, body gait, eyes, nose, mouth, and further more. All this trait are useful for a multibiometric identification or a person re-identification in a video surveillance context. Body Worn Cameras (BWCs) are used by the police of various countries all around the word and their use is growing significantly. This raises the need to test new recognition methods that use multibiometric traits on person re-identification. The purpose of this work is to present a new video dataset called Gotcha-I. This dataset has been obtained using more mobile cameras in movement to adhere to the data of BWCs. The dataset includes videos from 62 subjects in indoor and outdoor environment to address both security and surveillance problem. Subjects have different behavior in videos including walk freely, follow a path, go upstairs, try to avoid the camera, for a total of 493 videos including a 180 degree videos for each face of the subjects in the dataset. Furthermore, there are already processed data, such as: the 3D model of the face of each subject with all the poses of the head in pitch, yaw and roll; and the body keypoint coordinates of the gait for each video frame. It's also shown an application of gender recognition performed on Gotcha-I, confirming the usefulness and innovativeness of the proposed dataset.

**Keywords:** Dataset, biometric, face, gait, head pose estimation, mobile device, body worn cameras.

## 1 Introduction

For security purposes, there are various types of videos to examine with different devices, from different perspectives and in different environments. This type of data increasingly comes from mobile devices such as cameras worn by the body by UK police officers. Body worn cameras are revolutionizing road safety. In addition to providing important information on the health of the agents, the main use of these cameras is to record videos useful in dangerous situations to perform subject and action recognition. To test algorithms suitable for data like these, we created Gotcha-I. Gotcha-I contains videos with different modes. The

subjects in the videos are both cooperative and non-cooperative, in order to simulate the user's attempts to avoid the camera. They move along a path or freely and have been captured in different lighting modes. One of the innovations made by Gotcha-I is the possibility of working on videos captured by the moving camera. The movement is manual and not forced, this means that both the subjects and the recorders have different speeds during different videos.

### 1.1 Related work

As new needs arise in the world of video surveillance and security, new datasets are created to test recognition methods. In this section we will show an overview of dataset comparable with the one proposed.

COMPACT [2] is a biometric dataset focused on less-cooperative face recognition. Images are in high resolution but acquired in a fully automated manner. This allow to have real-world degradation like expressions, occlusion, blur etc. Differently from COMPACT, UBEAR [3] is focused on ear images. Subjects in UBEAR are in movements and under varying lighting conditions. The subject can move the head freely and acquired images are in gray scale.

In the outdoor environment we can found datasets as QUIS-CAMPI [4]. In QUIS-CAMPI, subjects are on the move and at about 50m of distance. For the enroll of data, the same subjects were acquired also in indoor scenario and a 3D model of each subject is also available. There are full body images and, from them, a PTZ camera extracted face images.

In previous example images were captured by a camera, but in the surveillance purpose there are also dataset obtained by drones. In DRONEFACE [5], the authors focused on face tracking. This task became difficult to approach due to the distance between the drone and the subjects. For this reason, they built DRONEFACE, composed of facial images taken from various combinations of distances and heights for evaluating how a face recognition technique works in recognizing designated faces from the air.

Another dataset focused on faces on distance is SALSA [7]. Differently from the previous, in salsa we have a fixed camera network that record subjects in two different modalities, both in indoor environment. The first modality simulate a poster presentation, in which there is a presenter and an audience. The second modality simulate a cocktail party in which subjects are freely to move and interact.

All previous datasets are focused on one, or at most two biometric traits. However, recently, also multibiometric dataset were created.

As an example, MUBIDIUS-I [6] is a multibiometric and multienvironment dataset, acquired by drones and cameras. There are many biometric traits in this dataset as ear, face, iris and full body. Most of the modality are at less distance and with fully cooperative subjects, but there are still videos with less-cooperative subjects at distance in a outdoor environment.

Completely different from the previous ones are then the dataset captured with pose estimation purpose. In order to obtain an accurate ground truth, the most used dataset in this field are like BIWI [8]. BIWI is a face dataset captured by a

kinect, that allows faces and 3D models of the subject captured in an indoor environment and in a cooperative mode. There are only 20 subjects in the dataset and only with face informations.

The proposed dataset Gotcha-I brings together all the features of the previous ones. This dataset provide a 3D model of each subjects, despite not having any image captured by depth camera. This make us able to perform pose estimation on real faces with a very precise ground truth as BIWI. As MUBIDIUS-I, our dataset can be used for multiple biometric traits due to the different distances and modalities. As in SALSA, our subjects can move freely in different modalities and as DRONEFACE we are able to perform the tracking of a subjects through different videos. We have also outdoor environments as in QUIS-CAMPI and a 180 degree videos of the subjects that allows us to perform ear recognition as in UBEAR. Finally our modality is more than less-cooperative than COMPACT because our subjects deliberately try to avoid being filmed.

In table 1, a list of overall specification of each presented dataset is available, compared with Gotcha-I.

In table 2 a list of annotation furnished from each compared dataset is present. Very few dataset has 3D models of the subjects and Landmark annotation, by which we mean coordinates of keypoint on the faces or bodies of the subjects.

Dataset	Subjects	Biometrics	Environment	Device	C./N.C.
COMPACT	108	Face	Indoor	Camera	No
UBEAR	126	Ear	Indoor	Camera	No
QUIS-CAMPI	320	Full Body	Outdoor	Camera	Yes/No
DRONEFACE	11	Face	Outdoor	Camera	Yes/Yes
MUBIDIUS-I	80	Multi	Multi	Multi	Yes/No
SALSA	18	Full Body	Multi	Multi	No/Yes
BIWI HEADPOSE	20	Face	Indoor	Camera	Yes/No
Gotcha-I	62	Full Body/Face	Multi	Camera	Yes/Yes

**Table 1.** Reference datasets with overall specifications. C./N.C. means Cooperative/Non Cooperative Mode

## 1.2 Security purpose applications

Videos captured by mobile camera with different environments are representative of surveillance data. In the last ten years, more and more countries provided their police officers with body worn cameras [9]. Differently from fixed cameras, mobile cameras has the ability to move around big spaces and cover big areas. However, data as images and videos from mobile cameras are quite different from fixed cameras due to the different point of view during recording [10]. Our dataset is proposed as a starting point to test algorithms operating on that topic and we will introduce some algorithms that could benefit it.

Dataset	3D Models	Landmark Annotation
COMPACT	No	No
UBEAR	No	No
QUIS-CAMPI	Yes	No
DRONEFACE	No	No
MUBIDIUS-I	No	No
SALSA	No	Yes
BIWI HEADPOSE	Yes	No
Gotcha-I	Yes	Yes

**Table 2.** Reference datasets with type of annotations

**Identity recognition** One of the hot topic in surveillance is identity recognition. One of the most used biometric traits for this purpose is the face. There are a lot of algorithms in last years that use faces to recognize a subject, both using neural networks [11] [12] or mixed methods [13]. In order to do that, on this type of data may be useful use previously a head pose estimation method to select the most frontal frame. This aim can be reached with various algorithm in literature that works in real time, both with Neural Networks and without [14] [15][16]. The proposed dataset also allows to detect a subject from other biometric traits like ear or iris, as algorithms in [17][18] performs. This is possible due to the different distance of the subject from the camera during recording. Thanks to the fact that this happens in the same video sequence, we are also able to fuse biometric traits in order to perform multibiometric recognition using different frames for different biometric trait [19] [20] [21].

**Traits Classification** Not only the identity of the subject can be useful in security purpose, often we are interested in some physical traits like the gender, the age or facial characteristics in order to classify many subjects in few time. Regarding the gender recognition, it is often performed using face [22] [23]. However, it is also possible to use videos in which subjects can move freely, also far from camera, using the gait as biometric traits [24] [25] or data collected by mobile devices [26]. It is also interesting how different results are if we consider the cooperativeness or the non cooperativeness of the subjects, labeled in this dataset, as the work in [27]. Using the face or the movement of the subject we can also extract information about their age, as in [29] [30]. This problem is often performed together with the previous one, gender recognition, to extract the most intuitive and general information about the subjects [31]. Once some characteristics are captured we may be able to follow the subjects along different paths and cameras, as our dataset allow us. This is called tracking [32] [33] and it is a very hot topic in security due to the ability to find the same people in different environments, in real time, without the need to know the exact identity of the subject [34]. For this purpose, if we focus our attention on face, there are various characteristics we can use to discriminate subjects, as in [28]. At

the same times various classification and clustering algorithms were built in this sense [35] [36] [37].

In conclusion, taking a look at identification and classification of subjects in security purpose, our dataset allow users to train, test and perform very different algorithms at the state of the art.

## 2 Gotcha-I

Gotcha-I dataset stems from the growing need to extract biometrics from video surveillance data and from the need to understand who the user is, where he is and what he is doing. Gotcha-I-I dataset allows to extract different biometrics: the face, the nose, the mouth, the eyes, the ears and the periocular area. Given the nature of the videos it is also possible to extract behavioral biometrics from gait. Gotcha-I dataset simulates the acquisition of the body worn camera in which a moving subject is acquired by a moving camera. Gotcha-I dataset is available for download at the site in [1].

### 2.1 Content of the dataset

To simulate real-world conditions, no accessories (clothes, hats or glasses) were controlled, they were left participant dependent. About the procedure followed by each participant, there were two recording procedures: (I) a cooperative mode with the camera where the subject walks and collaborates with the camera watching it during the walk, see Fig.1 (top-left), and (II) a non cooperative mode where the same subject walks trying to avoid the camera, see Fig.1 (top-right). The dataset contains a total of 493 videos with an average duration of 4 minutes, including 62 subjects, 15 women and 47 men in an average age between 18 and 20 years.

In order to be able to create robust systems, several possible scenarios were considered for the previous described procedures. The dataset is composed of 11 different video modes in different environmental and behavioral contexts.

The contents of the dataset are listed below:

- (1) indoor with artificial light - cooperative mode;
- (2) indoor with artificial light - non cooperative mode;
- (3) indoor without any lights but the camera flash - cooperative mode;
- (4) indoor without any lights but the camera flash - non cooperative mode;
- (5) outdoor with sunlight - cooperative mode;
- (6) outdoor with sunlight - non cooperative mode;
- (7) 180°head video;
- (8) stairs outdoor - cooperative mode;
- (9) stairs outdoor - non cooperative mode;
- (10) path outdoor - cooperative mode ;
- (11) path outdoor - non cooperative mode;
- (12) derived files attached, detailed in section 2.2.

All the videos have been acquired with the camera of the Samsung S9+ mobile phone; the modes (8-9-10-11) have also been acquired with an iPhone 10 and a Samsung Galaxy A5.

**Illumination differences.** Some real world problems can occur regarding the selected illumination settings. Videos in (5-6-8-9-10-11) were captured outdoors with natural sunlight, Fig. 1 (top-left). Videos in (1-2-7) have been acquired in a room with a white background with the artificial lights on, see Fig. 1 (top-right). Videos in (3-4) have been acquired in the same room with the lights off and the flash camera on, in these videos the use of camera flash can generate blur frames in some sequences Fig. 1 (down); as we can see, in this mode some frames could be blurred and this adds greater competitiveness to the dataset.

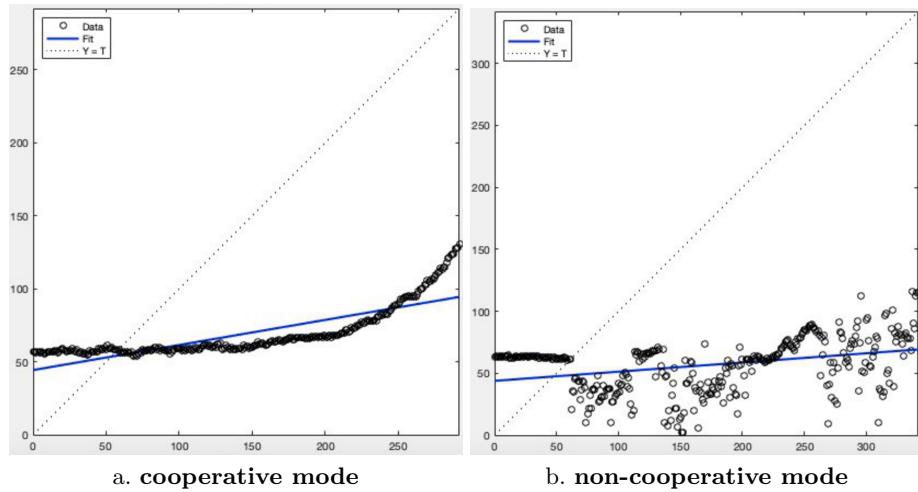


**Fig. 1.** Some Gotcha-I Dataset samples: outdoor sunlight in cooperative mode (top left), indoor with artificial light in non cooperative mode (top right) and Indoor with the camera flash in cooperative mode (down).

**Cooperative and non-cooperative mode.** In cooperative video sequences, the subjects look at the camera during the acquisition and follow the camera

lens during the motion. In non-cooperative video sequences, the subjects try to avoid the camera during the motion, as in Fig. 1 (top-right). This modality is clearly most competitive.

Fig. 2-a shows the distance from the neck to the nose in a cooperative video and Fig. 2-b shows this distance in a non-cooperative video of the same subject. We can observe that the cooperative mode exhibits a linear behaviour, while the non cooperative mode behaves irregular. The difference in the regularity of the movements of the subjects in cooperative and non-cooperative videos has led us to carry out experiments on different methods to study the results.



**Fig. 2.** The measurement of the distance from the coordinate of the nose to that of the neck in subject 01 in Indoor video with artificial light.

**Path and stairs outdoor** In these two modes the videos were acquired from different points of view simultaneously between three different cameras: Samsung S9 +, iPhone 10 and a Samsung Galaxy A5. Furthermore not all subjects are present. These videos have been created specifically to perform re-identification and action recognition algorithms. The videos simulate the acquisition from a video surveillance camera, so once the face (or the gait) is acquired it is possible to re-identify it and trace it for the whole journey. Furthermore, the action of "going up the stairs" allows to perform algorithms of action recognition in order to predict if a subject is going up the stairs or walking. Example of frame extracted from this videos are in 3.

**180° head video** The videos of the face were acquired with the most favorable lighting conditions: indoor with lights on. There are 62 videos, one for each



**Fig. 3.** Images from stairs and path videos respectively.

subject, close to less than a meter from the face by rotating the camera around the head 180 degrees: from the left ear to the right ear. The subjects were asked to sit on a chair placed in a room with a white panel behind them, the operator then made the video by turning around the subject. This mode has the purpose of acquiring the details of the face and can be used for the acquisition of the biometries of the face that require a high resolution of the image, such as iris, ear, profile, nose, mouth, and periocular area.

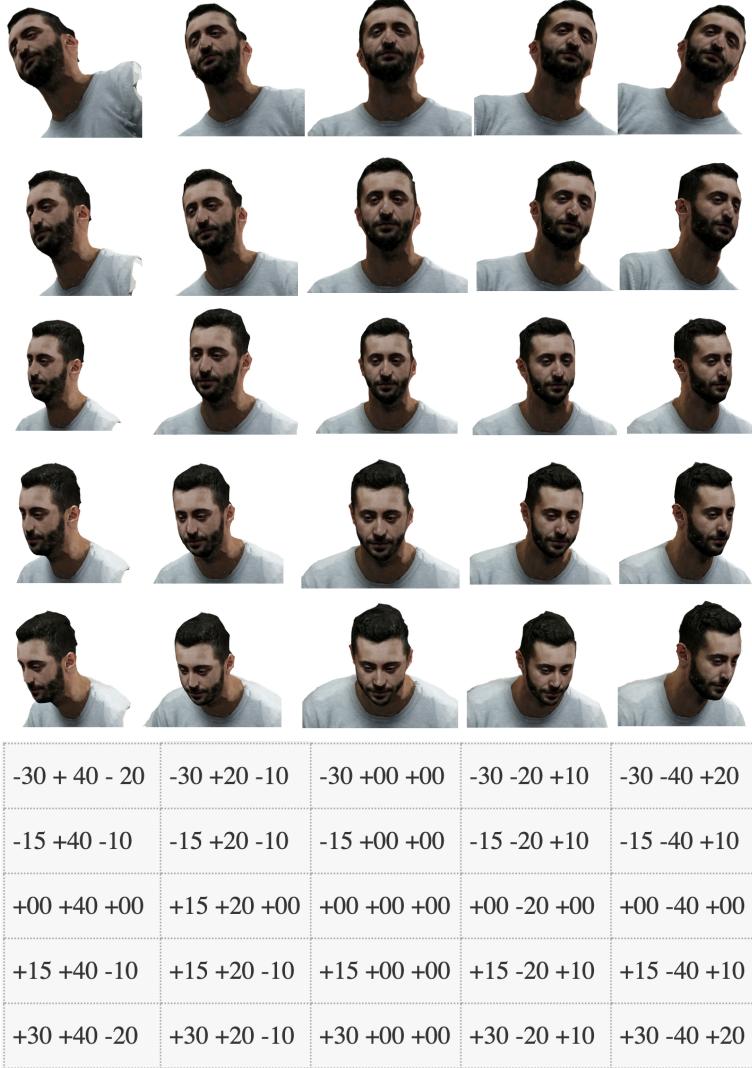
## 2.2 Derived files attached.

Derived files attached to the dataset are data extracted from videos and are useful for data manipulation. What exists in the 3D model of the face from the videos (9) extracted through photogrammetry. Later it was possible to extract from each 3D model the pitch yaw and roll rotation of the face for the Head Pose Estimation.

**3D model and Head Pose Estimation data.** From the videos in "(7) 180° head video" we have reconstructed the 3D model of the head in .obj format available within the derived files attached. From the 3D model of the head adequately elaborated through the Blender software, the images of the head were extracted in all the pitch, yaw and roll poses with 5° deviations in the following ranges of values:

- Pitch (-30°; +30°);
- Yaw (-40°; +40°);
- Roll (-20°; +20°).

For each subject of the dataset 2.223 related to as many different poses of the head were extracted, for 62 subjects therefore 137.826 images were extracted. In figure 4 there is a subset of 25 images of the head pose estimation of the subject 62.



**Fig. 4.** Above some examples of image extracted from 3D model of the subject 62. At the bottom the degrees in pitch, yaw and roll of the head pose estimation corresponding to the position in the table.

**Landmark extraction data.** For each video frame, except for the videos in ”(7) 180° head video”, the landmarks of the 2D pose estimation of the body and the 68 landmarks of the face were extracted using the OpenPose software [38]. These data are useful for gait analysis and performance action recognition.

### 3 Conclusion

In conclusion, Gotcha-I is a multiview dataset built to meet the needs of surveillance data from Body Warn Cameras. With a total of 62 subjects in 11 different modalities, Gotcha-I results particularly suitable for tasks such as people tracking and recognition. The high definition and full bodies in video allow to perform different type of biometric traits, both physical and behavioral. Compared to the other Datasets in literature, Gotcha-I presents a remarkable difference between cooperative and non cooperative modalities, allowing to test the response of the same algorithms on more or less complex data. Additional contents as the 3D model of each subjects, face and body coordinates and annotated head pose images, make the dataset very versatile in terms of possible testable applications.

### References

1. Gotcha-I dataset : <https://gotchaproject.github.io/>
2. Włodarczyk, M., Kacperski, D., Sankowski, W., Grabowski, K. (2018). COMPACT: biometric dataset of face images acquired in uncontrolled indoor environment. Computer Science, 20(1). doi:<https://doi.org/10.7494/csci.2019.20.1.3020>
3. Rui Raposo, Edmundo Hoyle, Adolfo Peixinho, and Hugo Proen  a. UBLEAR: A dataset of ear images captured on-the-move in uncontrolled conditions, in 2011 IEEE Workshop on Computational Intelligence in Biometrics and Identity Management (SSCI 2011 CIBIM), Paris, France, April 11 - April 15, pag. 84-90, 2011.
4. Jo  o Neves, Juan Moreno, Hugo Proen  a. QUIS-CAMPI: an annotated multi-biometrics data feed from surveillance scenarios. IET Biometrics Volume: 7 , Issue: 4 , 7, 2018. DOI: 10.1049/iet-bmt.2016.0178.
5. Hwai-Jung Hsu and Kuan-Ta Chen. 2017. DroneFace: An Open Dataset for Drone Research. In Proceedings of the 8th ACM on Multimedia Systems Conference (MMSys'17). ACM, New York, NY, USA, 187-192. DOI: <https://doi.org/10.1145/3083187.3083214>
6. Luigi Di Maio, Riccardo Distasi, Michele Nappi. MUBIDUS-I: A Multibiometric and Multipurpose Dataset. SITIS 2019 - The 15h International Conference on SIGNAL IMAGE TECHNOLOGY and INTERNET BASED SYSTEMS 26-29 November 2019 - Sorrento, Italy.
7. X. Alameda-Pineda, J. Staiano, R. Subramanian, L. Batrinca, E. Ricci, B. Lepri, O. Lanz and N. Sebe: SALSA: A Novel Dataset for Multimodal Group Behavior Analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015.
8. G. Fanelli, M. Dantone, J. Gall, A. Fossati, L. Van Gool, Random forests for real time 3d face analysis, Int. J. Comput. Vision 101 (2013) 437–458.
9. Daniel E. Bromberg,   tienne Charbonneau, Andrew Smith. Public support for facial recognition via police body-worn cameras: Findings from a list experiment. Government Information Quarterly, 2 November 2019.

10. Ola Younis, Waleed Al-Nuaimy, Fiona Rowe, Mohammad H. Alomari. Real-time Detection of Wearable Camera Motion Using Optical Flow. 2018 IEEE Congress on Evolutionary Computation (CEC). DOI: 10.1109/CEC.2018.8477783.
11. Changxing Ding and Dacheng Tao. Trunk-Branch Ensemble Convolutional Neural Networks for Video-Based Face Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume: 40 , Issue: 4 , April 1 2018. DOI: 10.1109/TPAMI.2017.2700390.
12. Guodong Guo and Na Zhang. A survey on deep learning based face recognition. *Computer Vision and Image Understanding*, Volume 189, December 2019, 102805.
13. Gaili Yue and Lei Lu. Face Recognition Based on Histogram Equalization and Convolution Neural Network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume: 40 , Issue: 4 , April 1 2018. DOI: 10.1109/TPAMI.2017.2700390.
14. R. Ranjan, V. M. Patel, R. Chellappa, Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41 (2019) 121–135.
15. Paola Barra, Carmen Bisogni, Michele Nappi, Stefano Ricciardi. A survey on deep learning based face recognition. *Computer Vision and Image Understanding* Volume 189, December 2019, 102805.
16. Andrea F. Abate, Paola Barra, Carmen Bisogni, Michele Nappi and Stefano Ricciardi. Near Real-Time Three Axis Head Pose Estimation Without Training. *IEEE Access* Year: 2019, Volume: 7, Journal Article, Publisher: IEEE.
17. Debbrota Paul Chowdhury, Sambit Bakshi, Pankaj Kumar Sa and Banshidhar Majhi. Wavelet energy feature based source camera identification for ear biometric images. *Pattern Recognition Letters*. October 2018.
18. Eduardo Garea Llano, Mireya Saraí García Vázquez, Juan M. Colores Vargas, Luis M. Zamudio Fuentes, Alejandro A. Ramírez Acosta. Optimized robust multi-sensor scheme for simultaneous video and image iris recognition. *Pattern Recognition Letters* Volume 101, 1 January 2018, Pages 44-51.
19. Sonal and Ajit Singh. Review on Multibiometrics: Classifications, Normalization and Fusion Levels. 2018 International Conference on Advances in Computing and Communication Engineering (ICACCE). 22-23 June 2018. DOI: 10.1109/ICACCE.2018.8441727. IEEE
20. Carmen Bisogni, Michele Nappi. Multibiometric score-level fusion through optimization and training. 2019 3rd International Conference on Bio-engineering for Smart Technologies (BioSMART). 24-26 April 2019. DOI: 10.1109/BIOSMART.2019.8734162. IEEE
21. Andrea F. Abate, Carmen Bisogni, Aniello Castiglione, Riccardo Distasi and Alfredo Petrosino. Optimization of Score-Level biometric Data Fusion by constraint construction training.7th International Conference, iSCI 2019, Guangzhou, China, November 12–15, 2019.
22. Amit Dhomne, Ranjit Kumar, Vijay Bhan. Gender Recognition Through Face Using Deep Learning. *Procedia Computer Science* Volume 132, 2018, Pages 2-10.
23. Llukman Cerkezi, Cihan Topal. Gender recognition with uniform local binary patterns. 2018 26th Signal Processing and Communications Applications Conference (SIU). DOI: 10.1109/SIU.2018.8404587.
24. Ebenezer R.H.P. Isaac, Susan Elias, Srinivasan Rajagopalan, K.S. Easwarakumar. Multiview gait-based gender classification through pose-based voting. *Pattern Recognition Letters* Volume 126, 1 September 2019, Pages 41-50.

25. Paola Barra, Carmen Bisogni, Michele Nappi, David Freire Obregon and Modesto Castrillon- Santana. Gender Classification on 2D human skeleton. 3rd International Conference on Bio- Engineering for Smart Technologies (BioSMART 2019), DOI: 10.1109/BIOSMART.2019.8734198.
26. Ankita Jain, Vivek Kanhangad. Gender classification in smartphones using gait information. Expert Systems with Applications Volume 93, 1 March 2018, Pages 257-266.
27. Paola Barra, Carmen Bisogni, Michele Nappi, David Freire Obregon and Modesto Castrillon- Santana. Gait Analysis for Gender Classification in Forensics. Dependability in Sensor, Cloud, and Big Data Systems and Applications, 5th International Conference, DependSys 2019, Guangzhou, China, November 12–15, 2019.
28. Liu, Z., Luo, P., Wang, X., Tang, X.: Large-scale celebfaces attributes (CelebA) dataset (Multimedia Laboratory, The Chinese University of Hong Kong, 2015)
29. D. V. Pakulich, S. A. YakimovS, S. A. Alyamkin. Age Recognition from Facial Images using Convolutional Neural Networks. Optoelectronics, Instrumentation and Data Processing, May 2019, Volume 55, Issue 3, pp 255–262.
30. Md Tauhid Bin Iqbal, Mohammad Shoyaib, Byungyong Ryu, M. Abdullah-Al-Wadud, Oksam Chae. Directional Age-Primitive Pattern (DAPP) for Human Age Group Recognition and Age Estimation. IEEE Transactions on Information Forensics and Security. Volume: 12 , Issue: 11 , Nov. 2017.
31. Pau Rodríguez, Guillem Cucurull, Josep M. Gonfaus, Josep M. Gonfaus, F. Xavier Roca, Jordi Gonzàlez. Age and gender recognition in the wild with deep attention. Pattern Recognition Volume 72, December 2017, Pages 563-571.
32. Shin-YiWen, Yu Yen, Albert Y. Chen. Human Tracking for Facility Surveillance. Science and Information Conference CVC 2019: Advances in Computer Vision pp 329-338.
33. Harihara Santosh Dadi, Gopala Krishna Mohan Pillutla, Madhavi Latha Makkenna. Face Recognition and Human Tracking Using GMM, HOG and SVM in Surveillance Videos. Annals of Data Science, June 2018, Volume 5, Issue 2, pp 157–179.
34. Young-Gun Lee, Shen-Chi Chen, Jenq-Neng Hwang, Yi-Ping Hung. An Ensemble of Invariant Features for Person Reidentification. IEEE Transactions on Circuits and Systems for Video Technology, Volume: 27 , Issue: 3 , March 2017.
35. Anzalone L., Barra P., Barra S., Narducci F., Nappi M. (2019) Transfer Learning for Facial Attributes Prediction and Clustering. In: Wang G., El Saddik A., Lai X., Martinez Perez G., Choo KK. (eds) Smart City and Informatization. iSCI 2019. Communications in Computer and Information Science, vol 1122. Springer, Singapore.
36. Pingyu Wang, Fei Su, Zhicheng Zhao. Joint multi-feature fusion and attribute relationships for facial attribute prediction. 2017 IEEE Visual Communications and Image Processing (VCIP). DOI: 10.1109/VCIP.2017.8305036.
37. Ni Zhuang, Yan Yan, Si Chen, Hanzi Wang, Chunhua Shen. Multi-label learning based deep transfer neural network for facial attribute classification. Pattern Recognition Volume 80, August 2018, Pages 225-240.
38. Cao, Zhe Hidalgo, Gines Simon, Tomas Wei, Shih-En Sheikh, Yaser. (2018). OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields.