VOL. 18, NO. 3, 2019, 49-56
www.elektrika.utm.my
ISSN 0128-4428

ELEKTRIKA
Journal of Electrical Engineering

A Review of Content-Based Video Retrieval Techniques for Person Identification

Syahmi Syahiran Ahmad Ridzuan[1], Zaid Omar [2], Usman Ullah Sheikh [2]

[1]School of Electrical Engineering, Universiti Teknologi Malaysia, Johor Bahru, Malaysia.
[2]Department of Electric and Computer Engineering, School of Electrical Engineering,
Universiti Teknologi Malaysia, Johor Bahru, Malaysia

[*]Corresponding author: syahmisyahiran.ahmadridzuan@gmail.com

Abstract: The rise of technology spurs the advancement in the surveillance field. Many commercial spaces reduced the patrol guard in favor of Closed-Circuit Television (CCTV) installation and even some countries already used surveillance drone which has greater mobility. In recent years, the CCTV Footage have also been used for crime investigation by law enforcement such as in Boston Bombing 2013 incident. However, this led us into producing huge unmanageable footage collection, the common issue of Big Data era. While there is more information to identify a potential suspect, the massive size of data needed to go over manually is a very laborious task. Therefore, some researchers proposed using Content-Based Video Retrieval (CBVR) method to enable to query a specific feature of an object or a human. Due to the limitations like visibility and quality of video footage, only certain features are selected for recognition based on Chicago Police Department guidelines. This paper presents the comprehensive reviews on CBVR techniques used for clothing, gender and ethnic recognition of the person of interest and how can it be applied in crime investigation. From the findings, the three recognition types can be combined to create a Content-Based Video Retrieval system for person identification.

Keywords: Clothing Recognition, Content-Based Video Retrieval, Ethnic Recognition, Gender Recognition, Person Identification.

Article History: received 26 September 2019; accepted 1 December 2019; published 18 December 2019.

1. INTRODUCTION

Person identification is a technique commonly used by the police to identify criminals or a missing person. A witness of a crime or the close relative of the missing person will describe the distinct characteristics of the wanted person to a police officer before a sketch artist or a computer program creates a face from the descriptions. The first person who invent the anthropometric measurements of a person is a French criminologist named Bertillon who lived in the late 1800s. He creates an arrest card for each criminal which contain the physical measurements such as head length, nose breadth and forehead width along with other details such age, nativity, complexion, fingerprint and the mugshots [1].

Nowadays with the advent of video cameras technology, it is possible to install Closed-Circuit Television (CCTV) as a preventive measure for crimes which creates a video surveillance network. The footage from CCTV can be used for identifying the criminal (refer Boston Bombing 2013) and later served as the evidence. However, the large collection of video footage creates another problem – Data Management. Previously, the police needed to search manually through a bulk of video collection to find a suspect. Then, facial recognition is utilized in this field to ease the burden on the enforcers. However, this technique only detects face images during the scene. The method of conveniently finding suspect by filtering his traits like a search engine is still not available.

As for now, video search tools such as Google and Yahoo are primarily based on textual annotation of videos, which means that the search does not actually involve the video itself, but rather is focused on the keywords within the surrounding paragraphs on video's page, the complete opposite of their image department which allows reverse image searching. Using these tools, videos are manually annotated with keywords and then retrieved using text-based search methods. However, this manual video annotation is subjective, cumbersome, inconsistent and time-consuming, especially when the video database is large and diverse. The recent Youtube demonetization fiasco is an example the failure of differentiating the video caption with video content. Due to the pressure from the advertisers, Youtube is trying to stop showing the adverts for racial or hate speech videos which eventually funds the content creator. Unfortunately, with the current algorithm Youtube not only filtered the hate speech videos but also other political satirist or critics videos [2]. These shortcomings have led to the development of Content-based Video Retrieval (CBVR) techniques to simplify and improve the video retrieval performance.

CBVR is a technique which uses visual contents to search videos from large scale video databases according to the user's interest and has been an active research area since the 90's [3-6]. Information retrieval, meanwhile, is defined as the process of converting a request for information into a meaningful set of references, where early work on image retrieval can be traced back to the late 70's [7,8]. Early techniques were not generally based on visual features but on the textual annotation of images. The biggest different between CBVR and content-based image retrieval (CBIR) techniques are the latter does not consider the temporal information, which allows human activity recognition.

2. CONTENT-BASED VIDEO RETRIEVAL OVERVIEW

Content-Based Video Retrieval system is generally composed of:

- Object Detection & Tracking
- Feature Extraction
- Feature Classification
- Indexing
- Similarity Measure
- Retrieval System

Object Detection & Tracking is the step where the object of interest which in this particular case is human is detected from the video and his movement is then tracked frame by frame throughout the video. To isolate the subject from the surrounding, normally a bounding box will be used to track it. The method of detection can be divided into 3 types; appearance-based, motion-based and shape-based [9,10]. Furthermore, the tracking methods can be classified into 4 types: region-based tracking, active-contour-based tracking, feature-based tracking, and model-based tracking.

Feature Extraction is the process where some information is extracted from the video which will later be classified. It is also the same technique used in the object detection step. But instead of detecting human, this process will extract the features which allows to determine his gender, ethnicity and clothing.

Feature Classification is the step where the extracted features are assigned to relevant attributes. To achieve it, a training data will be utilized to train the classification technique to the extend where the machine is able to label the features correctly.

Indexing is the step of managing the classified features in a table. This table will later be served for similarity comparison. In Content-Based Image Retrieval, each image will be linked with its own features. Meanwhile, for the video it will take a longer time to accomplish while pursuing the same method. Therefore, to reduce processing time, all the frames that are associated to a single person will be treated as a single data and only few more salient frames undergo the feature extraction and classification.

Similarity measure is a metric to compare the features set at the input with the features in the index table which enable the machine to show the most similar or highest likelihood result.

Retrieval system is a system that decides how to show the result to the user. It can be divided into rule-based or classification-based. Rule-based is where the user set how to decide the outcome like the decision tree. It will involve many pruning where some potential results are eliminated in the early process, possibly giving inaccurate result. But this is the fastest method. For classification-based, the method will compare the similarity for each aspects of the data before deciding the most accurate output automatically. It can be Support Vector Machine which requires smaller training sample or Neural Network which is more accurate but requires much larger training sample.

The Figure 1 shows the process flow from the raw footage until the output desired by the user. This paper will focus more on feature extraction and classification which are crucial for the recognition of a person's ethnicity, gender and clothing.

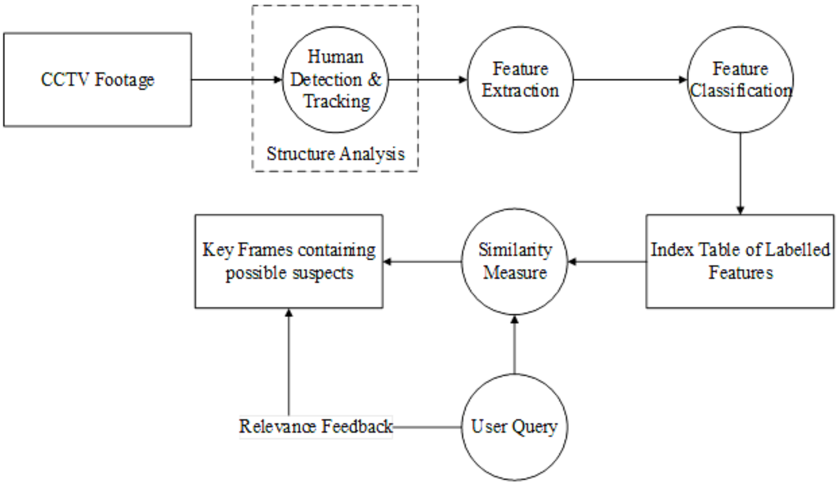

Figure 1. CBVR Flow Chart

3. REVIEW ON PERSON IDENTIFICATION TECHNIQUES

According to Chicago Police Department's 'How to describe a suspect' criminal identification guidelines [6], there are two major parts that contain a lot of discriminative features; suspect's face and clothing. There are other discriminative criminal descriptions that are important the police to narrow down the suspect such as gender, ethnicity, age, height and weight. According to Heckathorn et al. [11], combining visible physical characteristics like gender, ethnicity, eye color and body marks with anthropometric components could increase the accuracy of finding the relevant person.

In addition of that, due to limitation of obtaining clear face of the suspect from a video footage, this paper is only focusing on gender, ethnicity and clothing recognitions. Furthermore, there has been many researches done on these three types of recognition for various application and most of them only use one or two of the them. By reviewing all three recognitions, it enables us to reaffirm the decision for other types as they are interconnected, especially for gender and clothing. In this section, the focal points that will be discussed are feature extraction and classification only.

3.1 Gender Recognition

Identifying a person's gender is one of the key elements in person according to Chicago Police Department [6]. Based on facial and body features alone, a person can be decided having masculine or feminine traits. For the attire features meanwhile, most of the men's clothes can be worn by women but many of women's clothes are only used uniquely by women such as gown and skirt.

To start the gender recognition, some features are extracted first from images or prelabelled images if training-based technique is used. Then, they use feature classification to determine the gender.

Most of feature-based techniques use face images, but there are also others who use full body images and silhouette images. For those that use face images, they must be clear, high definition and not occluded [12-14], else they need to be trained to adapt in uncontrolled environment [13,15,16].

Tariq et al. [17] applied shape context-based matching on 441 silhouette images and achieved an average accuracy of 71.2% with 23.41% more accurate in identifying females.

Mozaffari et al. [18] combined appearance-based features, Discrete Cosine Transform (DCT) and Local Binary Pattern (LBP) with geometric-based feature, Geometrical Distance Feature (GDF) to obtain better classification result. The overall classification accuracy increased by 15.7% where DCT-LBP fusion only obtained 80.3% compared to DCT-LBP-GDF fusion's 96% using AR and Ethnic database.

Gutta et al. [12] introduced hybrid classifiers which is composed of Radial Basis Function (RBF) network and Inductive Decision Tree (DT). Adding RBF on top of DT gives the more adaptive threshold due to clustering ability of RBF. This method allows the user to gain 96% average accuracy rate on gender classification.

Hatipoglu et al. [19] implemented Speeded Up Robust Feature (SURF) based on Bag of Visual Words (BoW) and Support Vector Machine (SVM) algorithm on 3560 samples from FERET database. The authors showed that his proposal surpasses the accuracy rate of previous works that uses Principal Component Analysis (PCA) and Genetic Algorithm (GA) method (92%), Scale Invariant Feature Transform with BoW method (89.2%), Gabor Filter, RBF and SVM combined method (95.38%) and COSFIRE Filter method (93%) by obtaining 96% accuracy rate.

Orozco et al. [20] employed Convolutional Neural Network to classify the candidate regions extracted using Haar features embedded in Adaboost. The average accuracy rates obtained using this method are 95.42% for training set and 91.48% for the test set.

Azzopardi et al. [14] used SURF to extract 51 facial landmarks and COSFIRE filter to create trainable features. The output is then going through SVM classifier to decide the person's gender. The accuracy rates from this method are 94.7% for GENDER-FERET and 99.4% for the labeled faces in the wild data sets, better than LBP, Histogram of Gradient (HOG) and Gabor filter methods. The use of SURF features also adds robustness to most face variations.

Cirne et al. [21] detected 68 fiducial points from face images and establish a geometric descriptor by calculating the Euclidean distance of each points. The geometric descriptors are then feed into SVM classifier to classify gender. Using AR, Fei Face, Adience and Feret database, the proposed method excels more than other geometric descriptors such as Fellous distances and Gupta triangulation in most cases.

Nistor et al. [22] also utilized Convolutional Neural Network but he conceptualizes the idea of 'Network-in-Network' by introducing the Inception module. It contains convolutional layers of size 1*1, 3*3, 5*5 and a max pooling layer. This network, named 'Inception-v4' is trained using 70000 images and achieved the accuracy rate of 98.2% on their own dataset, and 84% on Adience dataset.

Moghaddam et al. [23] proposed using Nonlinear SVM with Gaussian RBF kernel where the authors achieved lowest overall error rate of 3.4% compared to SVM with cubic polynomial kernel, RBF, Bayesian, Fisher, Nearest Neighbor and linear classifier. In addition of that, when the method is compared between low and higher definition images, the difference is only at 1%. Thus, it is more robust to scale and degree of the face.

Liu et al. [24] programmed a gender recognition system from a full body image. The authors begin the system by segmenting the person into head, upper torso and lower torso. Then, features are extracted by HoG and classified into male or female using SVM classifier. It was done by detecting facial hair and type of clothing. From 400 images, this technique achieves 98.5% accuracy rate and uses multiple attribute to confirm someone's gender rather than face only.

From a video processing point of view, framework introduced by Liu et al. [24] is more compatible due to not relying only on face images but also someone's clothing to determine the gender. Due to video low definition, it is much harder to extract fine details such as facial landmarks from a person's face.

3.2 Clothing Recognition

Clothing recognition has been used for fashion purposes [25,26], pedestrian profiling [24,29] and also to reaffirm the decision for gender recognition [24]. The clothing also has been used for criminal identification [6]. Most of the techniques started by processing a full-body image. It was later separated into different body part location since the clothing type can be determined by a human physical.

Weng et al. [25] proposed using Cascade Color Moment to categorize a person's attire. The clothing image is divided into many blocks of 16*16 to obtain regional color information and to add spatial information. Thus, why it is called Cascade Color Moment or Cascade HSV Histogram depending on which type was used. Evaluation using 2000 images shows that Cascade Color Moment retrieved better than Cascade HSV Histogram, HSV Histogram or Color Moment.

Hidayati et al. [26] introduced style elements to identify types of cloth worn by a person. Firstly, the authors divided a full-body image into different parts; head, arms, waist, ankle and torso. Then five upper wear and seven lower wear features are extracted. These features which is called style elements are then utilized by a trained classifier to categorize them accordingly.   proposed method achieves overall precision of 88.7%. much better than end-to-end deep learning (60.13%) and deep-learned features with SVM (62.58%).

Chao et al. [27] utilized low-level features to recognize the clothing styles. The initial step is to detect faces using Viola-Jones detector and the position of upper body is estimated relative to the detected face. LBP and HOF are then used to extract texture features and local color histogram to detect the clothing's color.

Yang et al. [28] suggested a method combining color histogram with 3 different texture descriptors. The authors started with face detection and tracking from a surveillance video. Then the detected person's clothing was segmented using Voronoi image. 3 texture descriptors; HoG, BoW, DCT and combination of the three are then utilized for clothing recognition. Using 25441 cloth images, it is concluded that the combination of the 3 descriptors are better than in [27].

Kurnianggoro et al. [29] used deep networks of three convolutional layers and ten task specific classifier; gender, top-clothing, boots, hat, backpack, bag, handbag, shoes, upper-body color, and lower-body color. To train the network, the authors utilized the images for 702 persons and the rest as testing dataset. The proposed method achieved the accuracy rate of 80.5% compared which is better than ResNet and InceptionV3 methods.

Li Sun et al [30] uses BSP (B-Spline Patch) and TSD (Topology Spatial Distances) for features extraction and obtained 83.2% accuracy rate.

Huang et al. [31] focused on extracting the clothing landmarks using deep networks with prior of key point associations. By accurately locating the landmarks, the accuracy rate of clothing recognition can be increased even further. His method achieved 86% for collar, 95.2% for sleeve and 81.2% clothing using CCAP database, which are better than fast RCNN and YOLO method.

By comparing the performance of clothing recognition, most of the techniques didn't achieved more than 90% like gender and ethnic recognition methods. It is due the complexity of defining a cloth type or genre. However, in [26] the authors managed to outline the elements which defines the clothing type and achieved one of the highest accuracy rates. In paper [31] also offered another solution by extracting clothing landmarks which can later be used for classification.

3.3 Ethnic Recognition

Ethnicity is what separates the people in the world and let them form their own unique group. Being the same ethnic also means having many similar sets of appearance features such as skin colour, build, head shape, hair, face shape, and blood type. Most of the ethnicity recognition used face images to classify the race.

Guo et al. [32] proposed using biologically-inspired features (BIF) to classify ethnicity. To create BIF, the authors used Gabor filter which produces layers of simple (S) and complex (C) cell units. The extracted BIF is then combined with manifold learning techniques to reduce the dimensionality and SVM is used to finally classify the ethnic.

Gutta et al. [12] introduced hybrid classifiers which is composed of RBF network and Inductive DT. This method allows the user to gain 94% average accuracy rate on ethnic classification.

Lin et al. [33] used combination of Gabor filter banks and Adaboost learning to extract the facial features and SVM classifier to classify those features as gender. For ethnic recognition, the authors defined 3 classes; yellow (Mongoloid), black (African) and white (Caucasian) The authors also added pre-processing to help increase the accuracy rate. He achieved the accuracy rate of 94.58% for yellow against white, 95.59% yellow against black and 96.21% white against black.

Mohammad et al. [34] began with fusion of features extracted by HOG and LBP methods. Then, four classifiers namely, SVM, Multi-Layer Perceptron (MLP), Linear Discriminant Analysis (LDA), and Quadratic Discriminant Analysis (QDA) are compared before deciding which classifiers are best in classification. The authors showed that overall accuracy rate of 98.5% was achieved using SVM with polynomial kernel which excelled other classifiers.

Based on performance alone, [34] is better than [33] but [32] also achieved a good result in recognizing black and white people with more database needed to accurately recognize other ethnics.

Table 1. Techniques of gender, clothing and ethnicity recognition

Types	Year	Techniques	Database used	Comments
Gender	2009	Shape context-based matching and training using KNN [7]	3D face models collected by Hu et al.	- Average accuracy for gender was 71.20% and for ethnicity 71.66%. - Accuracy was significantly higher for some classes; 83.41% for females and 80.37% for East and South East Asians
Gender	2010	Discrete Cosine Transform (DCT), and Local Binary Pattern (LBP), and	AR and Ethnic	- Novel geometric feature improves the gender classification accuracy by 13%.

		geometrical distance feature (GDF) [18]		
Gender & Ethnic	1999	RBF networks ad inductive decision trees (DT)[12]	FERET	- Average accuracy rate of 96% on the gender classification task - 94% on the ethnic classification task
Gender	2017	SURF based BoW and SVM Methods [19]	FERET	better gender recognition performance on FERET database and the accuracy level of on left and right face images is a bit lower than the average accuracy level of frontal ones
Gender	2017	Deep convolutional network architecture to classify as male or female person the candidate regions previously detected using Haar features embedded in an AdaBoost [20]	Labeled Faces in the Wild and Gallagher	The method reached 95.42% and 91.48% average accuracies for the training set and for the test set respectively
Gender	2018	SURF descriptors extracted from 51 facial landmarks related to eyes, nose, and mouth as domain-dependent features, and the COSFIRE filters as trainable features [14]	GENDER-FERET, labeled faces in the wild, UNISA-Public	The method achieved 94.7% on GENDER-FERET and 99.4% on the labeled faces in the wild data sets
Gender	2017	Geometric descriptors [21]	Fei Face, AR, FERET, Adience	The proposed method excels more than other geometric descriptors such as Fellous distances and Gupta triangulation in most cases.
Gender	2017	CNN [22]	Adience	
Gender	2002	Nonlinear Support Vector Machines [23]	FERET	
Gender and Clothing	2017	Multiple-attributes (MA) recognition method [24]	Random images from web	Achieved 98.5% accuracy rate
Clothing	2013	Cascade Colour Moment [25]	Randomly downloaded images from the dangdang.com	The method reached a retrieval precision of 0.618
Clothing	2018	- Face Detection - Body Parts Identification - Multiclass supervised learning algorithm - Set of style elements [26]	Self-built data from various popular E-commerce websites	The proposed method achieves overall precision of 88.7%.
Clothing		low-level features [27]		
Clothing	2011	colour histograms with histogram of oriented gradient (HOG), Bag-of-Words (BOW)	Self-collected images which includes 937	It has better accuracy rate than [27]

		features, and DCT responses [28]	persons and 25441 cloth instances	
Gender & Clothing	2017	Deep networks of three convolutional layers and ten task specific classifier [29]	DukeMTMC-attribute dataset	Proposed method achieves the accuracy rate of 80.5%
Clothing	2017	uses BSP (B-Spline Patch) and TSD (Topology Spatial Distances) [30]	RGBD clothing dataset	The technique obtained 83.2% accuracy rate
Clothing	2018	deep networks with prior of key point associations [31]	CCAP and DeepFashion dataset	The method achieved 86% for collar, 95.2% for sleeve and 81.2% clothing using CCAP database
Ethnic	2010	biologically-inspired features (S2 & C2) - combine the BIF and manifold learning or subspace analysis techniques, such as PCA or OLPP for face representation and linear SVM for ethnicity classification [32]	MORPH-II	The ethnicity prediction for the Black and White races achieved 98.3% and 97.1%, respectively. However, for the other three races (Hispanic, Asian and Indian), the success rates are pretty low, e.g., 74.2%, 59.5%, and 6.9%, respectively due to insufficient data
Ethnic		Gabor filter banks and Adaboost learning to extract the facial features and SVM classifier to classify genders [33]	FERET database	The accuracy rate of 94.58% for yellow against white, 95.59% yellow against black and 96.21% white against black
Ethnic		fusion of features extracted by HOG and LBP methods. SVM classifier to classify genders with polynomial kernel [34]	FERET database	The overall accuracy rate is 98.5%

4. CONCLUSION AND FUTURE WORKS

In this study, the research findings on recognition of gender, clothing and ethnicity were compared based on their performances. Most of the listed techniques are designed for still images and particularly, face images. Therefore, every technique needs to be meticulously selected and considered so that it's suitable to create a CBVR system for person identification.

For gender recognition, the technique employed by Liu et al. [24] is compatible with CBVR due to the usage of full-body images and other attributes to reaffirm the gender classification decision such as clothing. It detects the facial features such as hair types and beard presence. However, the clothing type in the paper is quite limited, therefore the style elements in [26] can be used to expand the clothing types.

Hidayati et al. [26] created style elements for clothing identification. The authors extracted five upper wears (collar, print style, shoulder skin, front button, and sleeve types) and seven lower wear (leg gap, length, print style, side, pleat, wrinkle, and width) features. These features will aid in classifying upper wears (t-shirt, formal shirts, Henley shirts, informal shirts, polo shirt, long-sleeve shirt, spaghetti shirt and tank top) and lower wears (A-line short skirts, A-line long skirts, hot pants, shorts, skinny, straight long skirts, straight short skirts and trousers).

Meanwhile, most of ethnicity recognition methods require face images for feature extraction [33,34] but in video footage cases, it's difficult to extract any meaningful features. Therefore, there are two solutions that can be implemented:

1.Use skin color extraction and decide the ethnicity based on skin variations

2.Train a machine learning technique using annotated ethnicity database and test it on video footage.

For the first solution, it is much simpler since every ethnic is assigned with their color. However, the bigger question is which color to assign which justifies the association with the race? In this link [35] and this paper [36], the authors defined several skin tones which defines the ethnicity. There are also other authors who used other metrics to define the skin variation such as in papers [37] which used Lancer Ethnicity Scale and Fitzpatrick Skin Types. To extract skin color, the method used in [25] can simply be used to extract the color.

The second solution however needed to have a vast amount of different ethnic images so that the machine learning technique can be properly trained. From all the classifiers used the other papers, SVM classifier can be used for smaller database and CNN can be used for massive database.

ACKNOWLEDGMENT

This research is fully supported by UTM Research University Grant Tier 1 vote number 19H61. The authors fully acknowledged Ministry of Higher Education

(MOHE) and Universiti Teknologi Malaysia for providing the facilities and funding for this research.

REFERENCES

[1] Bertillon, A., 1896. The Bertillon system of identification. McClaughry, Ed., Chicago, IL.

[2] Julia Alexander, 2018. What is YouTube demonetization? An ongoing, comprehensive history - Polygon. Available at: https://www.polygon.com/2018/5/10/17268102/youtube-demonetization-pewdiepie-logan-paul-casey-neistat-philip-defranco. [Accessed 18 September 2018].

[3] Zhang, H. J., Wu, J., Zhong, D. and Smoliar, S. W., 1997. An integrated system for content-based video retrieval and browsing. Pattern recognition, 30(4), pp.643-658.

[4] Hu, W., Xie, N., Li, L., Zeng, X. and Maybank, S., 2011. A survey on visual content-based video indexing and retrieval. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 41(6), pp.797-819.

[5] Juan, K. and Cuiying, H., 2010, July. Content-based video retrieval system research. In Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on (Vol. 4, pp. 701-704). IEEE.

[6] Chicago Police Department, 2013. How to describe a suspect. Available at: https://portal.chicagopolice.org/portal/page/portaI/ClearPath/Get%20InvoIved/Hotlines%20and%20CPD%20Contacts/How%20to%20Describe %20a%20Suspect

[7] Goldstein, A.J., Harmon, L.D. and Lesk, A.B., 1971. Identification of human faces. Proceedings of the IEEE, 59(5), pp.748-760.

[8] Kanade, T., 1977. Computer recognition of human faces

[9] Paul, M., Haque, S. M. and Chakraborty, S., 2013. Human detection in surveillance videos and its applications-a review. EURASIP Journal on Advances in Signal Processing, 2013(1), p.176.

[10] Nguyen, D. T., Li, W. and Ogunbona, P.O., 2016. Human detection from images and videos: a survey. Pattern Recognition, 51, pp.148-175.

[11] Heckathorn, D. D., Broadhead, R. S. and Sergeyev, B., 2001. A methodology for reducing respondent duplication and impersonation in samples of hidden populations. Journal of Drug Issues, 31(2), pp.543-564.

[12] Gutta, S., Huang, J. R., Jonathon, P. and Wechsler, H., 2000. Mixture of experts for classification of gender, ethnic origin, and pose of human faces. IEEE Transactions on neural networks, 11(4), pp.948-960.

[13] González-Briones, A., Villarrubia, G., De Paz, J.F. and Corchado, J.M., 2018. A multi-agent system for the classification of gender and age from images. Computer Vision and Image Understanding.

[14] Azzopardi, G., Greco, A., Saggese, A. and Vento, M., 2018. Fusion of domain-specific and trainable features for gender recognition from face images. IEEE Access, 6, pp.24171-24183.

[15] Rodríguez, P., Cucurull, G., Gonfaus, J. M., Roca, F. X. and Gonzalez, J., 2017. Age and gender recognition in the wild with deep attention. Pattern Recognition, 72, pp.563-571.

[16] Raza, M., Sharif, M., Yasmin, M., Khan, M. A., Saba, T. and Fernandes, S. L., 2018. Appearance based pedestrians' gender recognition by employing stacked auto encoders in deep learning. Future Generation Computer Systems, 88, pp.28-39.

[17] Tariq, U., Hu, Y. and Huang, T. S., 2009, November. Gender and ethnicity identification from silhouetted face profiles. In Image Processing (ICIP), 2009 16th IEEE International Conference on (pp. 2441-2444). IEEE.

[18] Mozaffari, S., Behravan, H. and Akbari, R., 2010, August. Gender classification using single frontal image per person: combination of appearance and geometric based features. In Pattern Recognition (ICPR), 2010 20th International Conference on (pp. 1192-1195). IEEE.

[19] Hatipoglu, B. and Kose, C., 2017, October. A gender recognition system from facial images using SURF based BoW method. In Computer Science and Engineering (UBMK), 2017 International Conference on (pp. 989-993). IEEE.

[20] Orozco, C. I., Iglesias, F., Buemi, M.E. and Berlles, J. J., 2017. Real-time gender recognition from face images using deep convolutional neural network.

[21] Cirne, M. V. M. and Pedrini, H., 2017, October. Gender recognition from face images using a geometric descriptor. In Systems, Man, and Cybernetics (SMC), 2017 IEEE International Conference on (pp. 2006-2011). IEEE.

[22] Nistor, S. C., Marina, A. C., Darabant, A. S. and Borza, D., 2017, September. Automatic gender recognition for "in the wild" facial images using convolutional neural networks. In Intelligent Computer Communication and Processing (ICCP), 2017 13th IEEE International Conference on (pp. 287-291). IEEE.

[23] Moghaddam, B. and Yang, M.H., 2002. Learning gender with support faces. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(5), pp.707-711.

[24] Liu, H. H., Xu, S. S. D., Chiu, C. C. and Chiu, S. Y., 2017, June. Gender recognition technology of whole-body image. In Consumer Electronics-Taiwan (ICCE-TW), 2017 IEEE International Conference on (pp. 263-264). IEEE.

[25] Weng, T., Yuan, Y., Shen, L. and Zhao, Y., 2013, October. Clothing image retrieval using color moment. In Computer Science and Network Technology (ICCSNT), 2013 3rd International Conference on (pp. 1016-1020). IEEE.

[26] Hidayati, S. C., You, C. W., Cheng, W. H. and Hua, K. L., 2018. Learning and recognition of clothing genres from full-body images. IEEE transactions on cybernetics, 48(5), pp.1647-1659.

[27] Chao, X., Huiskes, M. J., Gritti, T. and Ciuhu, C., 2009, October. A framework for robust feature selection for real-time fashion style recommendation. In Proceedings of the 1st international workshop on Interactive multimedia for consumer electronics (pp. 35-42). ACM.

[28] Yang, M. and Yu, K., 2011, September. Real-time clothing recognition in surveillance videos. In Image

Processing (ICIP), 2011 18th IEEE International Conference on (pp. 2937-2940). IEEE.

[29] Kurnianggoro, L. and Jo, K. H., 2017, October. Identification of pedestrian attributes using deep network. In Industrial Electronics Society, IECON 2017-43rd Annual Conference of the IEEE (pp. 8503-8507). IEEE.

[30] Sun, L., Aragon-Camarasa, G., Rogers, S., Stolkin, R. and Siebert, J. P., 2017, September. Single-shot clothing category recognition in free-configurations with application to autonomous clothes sorting. In Intelligent Robots and Systems (IROS), 2017 IEEE/RSJ International Conference on (pp. 6699-6706). IEEE.

[31] Huang, C. Q., Chen, J. K., Pan, Y., Lai, H. J., Yin, J. and Huang, Q. H., 2018. Clothing landmark detection using deep networks with prior of key point associations. IEEE transactions on cybernetics, (99), pp.1-11.

[32] Guo, G. and Mu, G., 2010, June. A study of large-scale ethnicity estimation with gender and age variations. In Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on (pp. 79-86). IEEE.

[33] Lin, H., Lu, H. and Zhang, L., 2006. A new automatic recognition system of gender, age and ethnicity. In Proceedings of WCICA (Vol. 2, pp. 9988-9991).

[34] Mohammad, A. S. and Al-Ani, J. A., 2017, September. Towards ethnicity detection using learning-based classifiers. In Computer Science and Electronic Engineering (CEEC), 2017 (pp. 219-224). IEEE.

[35] Chrispy, 2004. Ethnic skintones. Available at: http://www.coolminiornot.com/articles/1310-ethnic-skintones. [Accessed 18 September 2018].

[36] Xiao, K., Yates, J. M., Zardawi, F., Sueeprasan, S., Liao, N., Gill, L., Li, C. and Wuerger, S., 2017. Characterizing the variations in ethnic skin colours: a new calibrated data base for human skin. Skin Research and Technology, 23(1), pp.21-29.

[37] Torres, V., Herane, M. I., Costa, A., Martin, J. P. and Troielli, P., 2017. Refining the ideas of "ethnic" skin. Anais brasileiros de dermatologia, 92(2), pp.221-225