Person Recognition Beyond The Visible Spectrum: Combining Body Shape and Texture from mmW Images

Ester Gonzalez-Sosa
Nokia Bell-Labs
Madrid, Spain
ester.gonzalez@nokia-bell-labs.com

Ruben Vera-Rodriguez, Julian Fierrez
BiDA-Lab
Universidad Autonoma de Madrid
Madrid, Spain
{*ruben.vera,julian.fierrez*}*@uam.es*

Vishal M. Patel
Rutgers University
Piscataway, NJ, USA
vmp93@soe.rutgers.edu

***Abstract*—Biometrics have been tradittionally based on images acquired in the visible spectrum. In this paper, we will go first into details regarding the regions beyond the visible spectrum that have been already explored in the literature for biometrics to overcome some of the limitations found in the visible region. Later, we will introduce millimeter imaging as a new region of the spectrum that has also potential in biometrics. To this aim, we first consider shape and texture information individually for person recognition. Later, we compare them and study to what extent the joint use of shape and texture can provide further improvements. Results suggest that both sources of information can complement each other, reaching verification results of 1.5% EER. This result motivates us to think that in the future, person recognition can be integrated within the millimeter screening scanners already deployed in airports, and enhance this way security.**

***Keywords*-mmW Imaging; body shape; body texture; CNN; security; multimodal fusion; multi algorithmic fusion**

I. Introduction

Biometrics is a technological area, whose aim is to discriminate automatically between subjects in a reliable way and according to some target application based on one or more signals derived from physical or behavioral traits [12]. The most widely research biometric traits have been face, fingerprint, iris and voice. However, there is also additional biometric information available in some particular application domains that has not been extensively studied. For instance, body information has been primarily studied through gait biometrics, but not much using information derived from single-shot images [18].

From the entire electromagnetic spectrum, the visible range of the spectrum comprises only the range between $400 - 750\ nm$. Despite that most biometric recognition applications use images acquired in the visible spectrum, such images are affected by, among other factors, lighting conditions and body occlusions viz. clothing, make-up or hair. To overcome these limitations, the biometric community have explored other ranges of the spectrum with interesting properties. Table I summarizes some of the regions beyond the visible spectrum that have been explored for biometrics, pointing out the associated wavelength, and their properties.

As illustrated, X-ray has a wavelength in the range of $0.01 - 10\ nm$ and enough energy to pass through cloth and human tissues. An interesting example on biometric identification through X-ray is dental biometrics [5], which are useful in forensic scenarios.

The infrared band of the electromagnetic spectrum (IR) lies between the microwave and visible regions and its wavelength is located in the range of $1\ mm$ to $750\ nm$. Some examples of usages of the IR region for biometric purposes are: i) using the NIR subregion to be more robust against illumination changes [14] or allowing identification in the darkness, ii) using the MWIR and LWIR bands to deploy liveness detection, being the later less explored than NIR. A novel approach to address MWIR-based biometrics is through polarimetric images. By acquiring the polarization-state, additional textual and geometric information can be obtained and be used to enhance the conventional thermal face images [19], [11].

A. Submillimeter and Millimeter Wave Imaging

Submillimeter (smW) and millimeter waves (mmW) fill the gap between the IR and the microwaves Specifically, mmW waves lie from $10-1\ mm$ ($30-3000$ GHz), covering the highest frequencies of the microwave region and the smW regime lies in the range of $1\ mm$ to $100\ \mu m$ (300 GHz to 3 THz), occupying the lowest frequencies of the infrared region. The smW band is within the well-known Terahertz band, which occupies the range between $0.3-300$ μm ($1000 - 1$ THz).

Although most of the radiation emitted by the human body belongs to the MWIR and LWIR bands, it emits radiation in the smW and mmW regions as well, hence allowing passive imaging. One of the more interesting properties is that clothing is highly transparent to the mmW radiation and partially transparent to the smW radiation. Natural applications of mmW and smW imaging include Concealed Weapon Detection (CWD), non-destructive inspection, low visibility navigation enhancement and medical imaging.

0-7695-6403-8/18/31.00 ©2018 IEEE
DOI 10.1109/ICB2018.2018.00044

Table I
REGIONS BEYOND THE VISIBLE SPECTRUM COMMONLY USED FOR BIOMETRIC APPLICATIONS. ABBREVIATIONS USED: ARCHITECTURE (ARCH.); ACTIVE (A); PASSIVE (P); CONCEALED WEAPON DETECTION (CWD); INFRARED (IR); NEAR INFRARED (NIR); MEDIUM WAVE INFRARED (MWIR); LONG WAVE INFRARED (LWIR); SUBMILLIMETER WAVE (SMW); MICROWAVE (MW); MILLIMETER WAVE (MMW).

Spectrum Region		Wavelength	Properties
X-ray		$0.01 - 10nm$	Penetration through clothes and human tissues
IR	NIR	$750nm - 0.9um$	Robust to illumination visibility in the darkness
	MWIR	$3um - 5um$	Invariant to illumination, acquires temperature
	LWIR	$8um - 14um$	Invariant to illumination, acquires temperature
	smW	$100um - 1mm$	Non-polar dielectric material, partial transparency
MW	mmW	$1mm - 10mm$	Non-polar dielectric material, High transparency, visibility in adverse conditions

CWD is deployed on mmW imaging due to the high degree of contrast achieved between the human body and weapons or other metallic objects in that region [3], [16]. Indeed, there are already mmW wave scanners deployed in several airports which acquire the full body signature with that purpose, exploiting the different responses (due to difference of temperatures) between metallic objects and the human body skin [17].

As these waves can penetrate through clothing, in mmW images, we are able to see things that can not be seen in a visible image, such as the torso information. As a result, information collected in a mmW image may be useful for person recognition. Concretely, shape information retrieved from the mmW images may be more robust to clothing variations than visible images. For the same reason, mmW imaging can also be exploited to retrieve texture information, which could potentially be used as discriminatory information. Furthermore, mmWs are also able to pass through facial accessories such as balaclavas, caps or artificial beards. This property makes mmW images less susceptible to spoofing attacks when compared to images in the visible spectrum [10] as people modifying their body constitution or using sophisticated facial masks may be easily detected using mmW images.

Contrary to CWD, few research works have been carried out for biometric recognition applications using mmW images. This has been mainly due to privacy concerns and lack of suitable databases for person recognition. The pioneers of this line of work were Alefs *et al.*, who put forward a person recognition system using real mmW passive images acquired in outdoors scenarios from the mmW TNO database [2]. They exploited the texture information contained in the torso region of the image through multilinear eigenspaces techniques, reaching very promising verification results (86% of Verification rate at FAR=0.001). However, the experimental protocol followed in [2] was very optimistic (distances were computed comparing pairs of images under the same head pose and point of view conditions, which was not a realistic situation). In [8], the utility of mmW images for person recognition under realistic scenarios was explored with a more realistic experimental protocol. They analysed shape information using real mmW images from the same database as [2], finding out that shape information could provide sufficient evidence for both verification and identification tasks, with EERs of 10.0% while using contour-based and image-based shape approaches.

Later, in [9], they explored the use of texture information extracted from real mmW images. The results achieved concluded that the torso was the best mmW body part, with the best EER around 5.00%.

In this paper, we aim to gain more insight through the joint use of shape and texture information from mmW images. With this goal in mind, we consider first biometric systems based on mmW texture and mmW shape standalone. Later, we will study different fusion schemes regarding combination of different texture approaches, and a joint fusion between shape and texture. Fig. 1 shows the overall pipeline of the system.

This paper is structured as follows. A brief review of the selected shape-based and texture-based features approaches are given in Section II and III, respectively. The mmW TNO database and experimental protocol used in this paper is described in Section IV. Results of these methods are presented in Section V. Finally, Section VI concludes the paper with a brief summary and discussion.

II. TEXTURE-BASED APPROACHES

For the transparency property of millimeter waves, mmW imaging for person recognition can be exploited through texture information. Concretely, we have considered three mmW body parts: face, torso and wholebody information. To address mmW texture information here, we consider various hand-crafted and deep learning approaches. Among the wide variety of hand-crafted features presented in the literature for biometrics, we select two of the most widely used ones: i) Local Binary Patterns (LBP), and ii) Histogram of Oriented Gradients (HOG). In what concerns deep learning approaches, several experiments have been conducted through fine tuning two different pre-trained models: Alexnet [13] and VGG-face [15]. In what follows we give more details regarding each of them.

Hand-crafted features: we consider LBP and HOG. In both cases, first images are resized to 100×150 ($width \times height$ format). The image is then divided into non

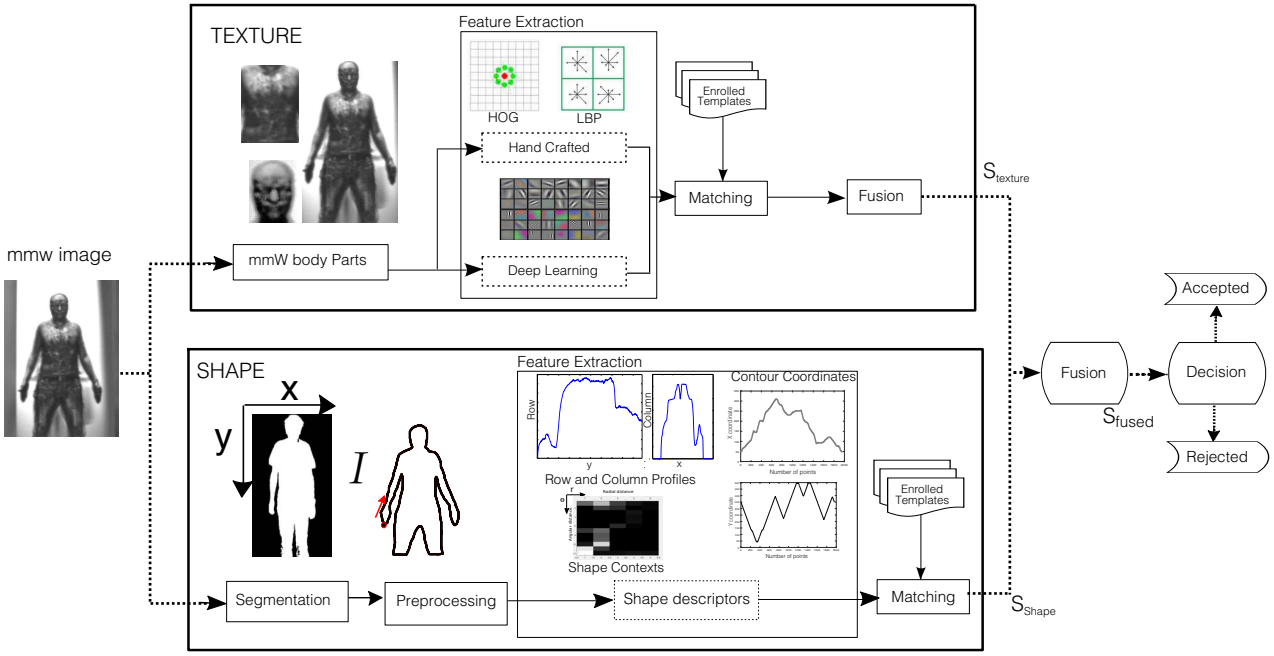

Figure 1. **An overview of the proposed fusion scheme for person recognition from mmW images**. The upper branch depicts the approach that address mmW texture while the lower branch sketches the approach that deals with mmW shape.

overlapping 10×10 blocks. In what concerns LBP, for each block we compute the LBP descriptor with radius 1, 8 neighbours and uniform patterns, resulting in a 59-length vector. The final feature vector of each image is the concatenation of all the histograms from all blocks. LBP features are extracted based on the implementation provided by [1] Regarding HOG, each block is described by a histogram of gradients with 8 number of orientations, with each gradient quantized by its angle and weighted by its magnitude. Then, four different normalizations are computed using adjacent histograms, resulting in 8×4-length feature vector for each block. The final feature vector of a given image is the vectorization of the HOG features from all blocks. HOG features are extracted using the implementation provided by [6]. In both cases, the distance-matcher employed has been the cosine distance.

Learned features: in order to extract deep learning features for the problem of mmW person recognition with only small databases available, we need to apply transfer learning techniques such as fine-tuning or use some pre-trained models as feature extractor. After performing several experiments, we realized that for this particular problem, the fine-tuning strategy worked better. This might be due to the dissimilarity between images from the source and target datasets. Concretely, we have fine-tuned the pre-trained models from the Alexnet and VGG-face networks, originally designed for the ISLVRC competition in 2012 and 2014, respectively. We have adapted the very last layers of the CNN architecture, changing the classification task from 1000 different objects or 2662 facial images to 50 different subjects. In total, we have 6 fine-tuned models, one per each mmW body part and pre-trained model. For the training, we have used the stochastic gradient descent to learn the parameters, with a momentum of 0.9, a number of epochs of 100, and a batch size of 32 samples. The learning rate and regularization parameter for the mmW person recognition task were respectively set to 10^{-5} and 10^{-3} for Alexnet and to 10^{-6} and 10^{-4} for VGG-face. After fine-tuning the AlexNet or VGG-face on the target dataset, we extract the deep feature as the output of the $fc7$ layer, which is a 4096 dimension vector. This feature vector will be later compared to other feature vector from a different subject to obtain a similiarity score through cosine distance.

A. Texture-based Fusion Schemes

We propose two different fusion schemes to further reach additional improvements over the individual texture approaches.

Multialgorithmic Fusion: we consider here fusing discriminative features extracted from a particular mmW body part. Concretely, we fuse at score level, information from the two hand-crafted (LBP and HOG) and two deep learning approaches (Alexnet and VGG-face), which use cosine similarity as the matcher.

Multimodal CNN-Level Fusion Torso and Wholebody: this multimodal fusion implies combining information from

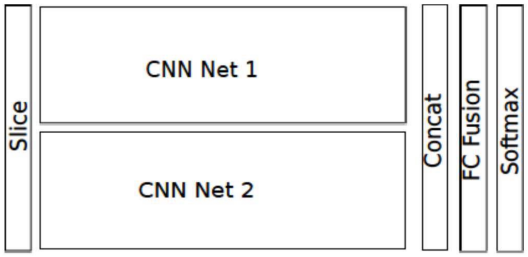

Figure 2. **CNN-level fusion**. This architecture allows you to fuse information from two different biometric traits.

different mmW body parts: face, body and wholebody. We studied different multimodal fusion schemes: feature, score and CNN-level fusion, finding out that the CNN-level outperformed the tradditional multimodal fusion schemes. The proposed CNN-level fusion was inspired by previous works in the area of object recognition [7]. Concretely, this multimodal network consists of two CNN branches, which then are combined in a late fusion approach (see Fig.2).

The training of this multimodal CNN is performed stage-wise, that is, first each branch is trained individually and then the overall architecture is trained using the individual trained models from branch 1 and branch 2. Branch 1 and branch 2 are merged by concatenating features from the *fc7* layer, resulting in a concat vector of size $2 \times 4096 = 8192$. After that, an additional fully connected layer (*fc8*) is introduced followed by a *Softmax* classification layer. To assure that mmW body parts come from the same identity, a *slice* layer is introduced at the beginning of the network. Due to the lack of a sufficient number of samples, both branches are fine-tuned using a particular pre-trained model. After that, the overall CNN architecture is trained to learn the parameters associated to the new introduced layer $fc8$, using a batch size of 50, learning rate of 0.0001 and 100 epochs.

III. Shape-based Approaches

As a consequence of the clothing transparency of millimeter waves, shape information retrieved from the mmW images may be more robust to clothing variations than visible images. Among the different shape-based approaches studied in a previous work [8], here we focus in those shape approaches with better verification performance, namely: i) Contour-Coordinates (CC) ii) Shape Contexts (SC), and iii) Row and Column Profiles (RCP). CC and SC are related to the contour coordinates themselves or to their relative position, respectively. RCP retains shape information, extracted from binarized images. The matchers employed for these shape approaches are Dynamic Time Warping (DTW) and Modified Hausdorff Distance (MHD), which both are distance-based matchers able to provide dissimilarities between sequences of different points.

Contour Coordinates: are used as the baseline feature approach, defined as $\mathbf{CC}(n) = (x_n, y_n), n = 1, \ldots, n_{cc}-1$, being n_{cc} the number of pixels that compose the contour that the silhouette edge describes with the background (approximately 2000 points), and (x_n, y_n) the coordinates of each one of those pixels.

Row and Column Profiles: given the binarized image I, whose pixels belonging to the foreground (fg) are set to 1 ($I(\mathbf{x}_{fg}, \mathbf{y}_{fg}) = 1$) and pixels belonging to the background (bg) are set to 0 ($I(\mathbf{x}_{bg}, \mathbf{y}_{bg}) = 0$), row and column profiles are computed as the number of pixels per row and per column which belongs to the foreground, respectively.

Shape Contexts: were first introduced by Belongie *et al.* [4]. This technique describes a specific point considering the relative distance and angle of the rest of the points within a shape. This method considers the set of vectors originating from a point to all other sample points on a shape. The number of radial bins (r_bins) and theta bins (θ_bins) are the main parameters of this descriptor. As a result, the shape contexts of a shape with N points forms a vector of size $(N \times r\_bins \times \theta\_bins)$.

IV. mmW TNO Database and Experimental Protocol

The mmW TNO database is comprised of images belonging to 50 different male subjects in 4 different scenarios. These 4 different scenarios derive from the combination of 2 different head poses (*frontal head pose* and *lateral head pose*) and 2 different facial occlusions (balaclava and beard). Images were recorded using a passive stereo radiometer scanner in an outdoor scenario. Each full scanning is a set of two single images with slightly different points of view of size 696×499 ($width \times height$). By dividing this set into single images of 348×499, the TNO database is comprised of 50 subjects $\times 2$ head pose configurations $\times 2$ facial clutter configurations $\times 2$ images per set, making a total of 400 images in the whole mmW TNO database. Even if it does not contain a larger number of subjects, the mmW TNO database is the only dataset available for research purposes.

V. Results

Table II presents the results obtained with the individual shape and texture approaches and the fusion of both approaches.

Texture: after having explored the three mmW body parts with all texture approaches, we found that the torso was the most discriminative one. Surprisingly, hand crafted approaches work slightly better. Considering the three mmW body parts, the second best body part was the wholebody, with 14.00% EER. Regarding the two different texture fusion approaches, we improved the individual results to 2.00% EER when considering all texture torso features and to 2.50% when combining information from the two best mmW body parts: torso and wholebody, using alexnet as feature descriptor.

Table II
SHAPE AND TEXTURE FUSION FOR VERIFICATION. SCORE LEVEL FUSION BETWEEN THE BEST SHAPE AND TEXTURE APPROACHES. ABBREVIATIONS USED: HISTOGRAM OF ORIENTED GRADIENTS (HOG); CONTOUR COORDINATES (CC); ROW COLUMN PROFILES (RCP); SHAPE CONTEXTS (SC); DYNAMIC TIME WARPING (DTW); MODIFIED HAUSDORFF DISTANCE (MHD). VERIFICATION PERFORMANCE IS REPORTED IN TERMS OF EER (%). FUSION RESULTS, WHICH MANAGES TO OUTPERFORM SHAPE OR TEXTURE STANDALONE ARE HIGHLIGHTED IN BOLD. THE BEST OVERALL VERIFICATION RATE IS MARKED WITH *.

Texture App.	Perf.	Texture Fusion App.	Perf.	Shape App.	Perf.	Shape & Texture App.
Torso Alexnet	6.00%	Multimodal CNN-Level Fusion	2.50%	CC-DTW	11.50%	3.00%
				RCP-DTW	10.75%	**1.50%***
Wholebody Alexnet	14.00%			CC-MHD	9.25%	2.50%
				SC-MHD	15.25%	2.50%
Torso HOG	4.50%	Multialgorithmic Fusion	2.00%	CC-DTW	11.50%	2.00%
Torso LBP	6.00%			RCP-DTW	10.75%	1.86%
Torso Alexnet	6.00%			CC-MHD	9.25%	**1.50%***
Torso VGG	8.72%			SC-MHD	15.25%	2.00%

Shape: among the different combinations of shape descriptors and matchers, the best shape-based approaches are Contour Coordinates with Dynamic Time Warping (CC-DTW, 11.50% of EER), Row and Column Profiles with Dynamic Time Warping (RCP-DTW, 10.75% of EER), Contour Coordinates with Modified Hausdorff Distance (CC-MHD, 9.25% of EER), and Shape Contexts with Modified Hausdorff Distance (SC-MHD, 15.25% of EER).

Fusion between shape and texture: is carried out among the best shape-based approaches and the best texture-based approaches, in terms of verification following the frontal protocol. It is carried out at the score level, following the sum rule, having previously normalized scores to the same range. As performance of shape-based approaches is always worse compared to texture-based approaches, a weighted fusion is conducted. Weights are empirically estimated, giving more importance to texture than to shape. Concretely these weights are set to 0.8 and 0.2 for texture and shape, respectively. Table II shows the fusion results attained for verification. As can be seen, even if shape-based approaches perform poorly compared to texture-based approaches, they are able to complement the evidence given by the latter ones, slightly outperforming the results. The best fusion results are achieved when combining Multialgorithmic Fusion with CC-MHD, or the Multimodal CNN-Level Fusion with RCP-DTW reaching EER of 1.5%. When comparing those results to the reference work carried by Alefs *et al.* [2], we found that our best schemes outperformed the performance reported there. Concretely, the verification rates obtained at $FAR = 0.001$ were of 92% and 88% for the Multialgorithm Fusion with CC-MHD, and the Multimodal CNN-Level Fusion with RCP-DTW, respectively (see also Fig. 3). Notice that those results are obtained with a more realistic scenario than the one used in the reference work.

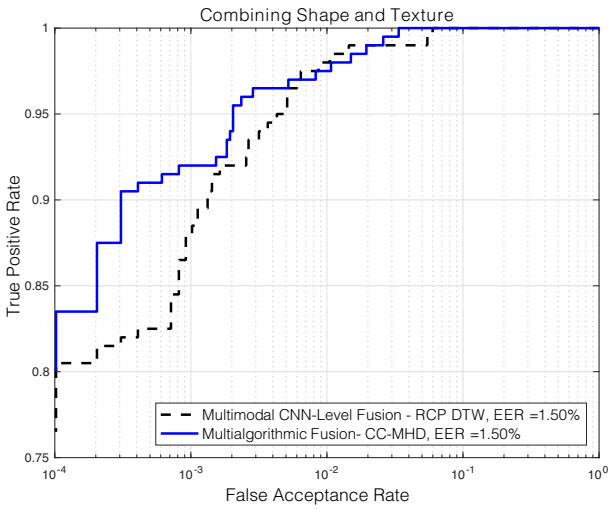

Figure 3. **ROC curves of the best Shape and Texture configurations**. Abbreviations used: Contour Coordinates (CC); Row Column Profiles (RCP); Dynamic Time Warping (DTW); Modified Hausdorff Distance (MHD), Convolutional Neural Networks (CNN).

VI. CONCLUSION

In this work we have studied the problem of person recognition through millimeter wave images. In particular, we have found that shape and texture information can provide additional performance improvements in comparison with their standalone performance rates. For large scale application, further experiments will be needed, including experiments with larger datasets. For future work, we plan to perform experiments using active images, which are the ones used in the mmW scanners deployed in airports. In any case, even with practical limitations, the present work has shown the feasibility of using mmW body texture for Person Recognition, and a quantitative analysis of the key factors affecting its performance, in this way, showing that current mmW security deployments may be improved with biometrics functionalities.

ACKNOWLEDGMENT

This work has been partially supported by project CogniMetrics TEC2015-70627-R (MINECO/FEDER), and the SPATEK network (TEC2015-68766-REDC). E. Gonzalez-Sosa was supported by a PhD scholarship from Universidad

Autonoma de Madrid. Vishal M. Patel was partially supported by US Office of Naval Research (ONR) Grant YIP N00014-16-1-3134. Authors wish to thank also TNO for providing access to the database.

REFERENCES

[1] T. Ahonen, A. Hadid, and M. Pietikainen. Face Description with Local Binary Patterns: Application to Face Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12):2037–2041, 2006.

[2] B. Alefs, R. den Hollander, F. Nennie, E. van der Houwen, M. Bruijn, W. van der Mark, and J. Noordam. Thorax biometrics from millimetre-wave images. *Pattern Recognition Letters*, 31(15):2357–2363, 2010.

[3] R. Appleby and R. N. Anderton. Millimeter-wave and submillimeter-wave imaging for security and surveillance. *Proceedings of IEEE*, 95(8):1683–1690, 2007.

[4] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4):509–522, 2002.

[5] H. Chen and A. K. Jain. Dental biometrics. *Encyclopedia of Biometrics*, pages 343–351, 2015.

[6] P. Dollár. Piotr's Computer Vision Matlab Toolbox (PMT). https://github.com/pdollar/toolbox, 2016.

[7] A. Eitel, J. T. Springenberg, L. Spinello, M. Riedmiller, and W. Burgard. Multimodal deep learning for robust RGB-d object recognition. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 681–687, 2015.

[8] E. Gonzalez-Sosa, R. Vera-Rodriguez, J. Fierrez, and V. M. Patel. Exploring body shape from mmw images for person recognition. *IEEE Transactions on Information Forensics and Security*, 2017.

[9] E. Gonzalez-Sosa, R. Vera-Rodriguez, J. Fierrez, and V. M. Patel. Millimetre wave person recognition: Hand-crafted vs learned features. In *Proceedings of IEEE International Conference on Identity, Security and Behavior Analysis*, pages 1–6, 2017.

[10] A. Hadid, N. Evans, S. Marcel, and J. Fierrez. Biometrics systems under spoofing attack: an evaluation methodology and lessons learned. *IEEE Signal Processing Magazine*, 32(5):20–30, 2015.

[11] S. Hu, N. J. Short, B. S. Riggan, C. Gordon, K. P. Gurton, M. Thielke, P. Gurram, and A. L. Chan. A polarimetric thermal database for face recognition research. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 119–126, 2016.

[12] A. K. Jain, K. Nandakumar, and A. Ross. 50 years of biometric research: Accomplishments, challenges, and opportunities. *Pattern Recognition Letters*, 79:80–105, 2016.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems (NIPS)*, pages 1097–1105, 2012.

[14] H. Maeng, S. Liao, D. Kang, S.-W. Lee, and A. K. Jain. Nighttime face recognition at long distance: Cross-distance and cross-spectral matching. In *Computer Vision*, pages 708–721. Springer, 2012.

[15] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep Face Recognition. In *British Machine Vision Conference (BMVC)*, 2015.

[16] V. M. Patel and J. N. Mait. Passive millimeter-wave imaging with extended depth of field and sparse data. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2521–2524, 2012.

[17] V. M. Patel, J. N. Mait, D. W. Prather, and A. S. Hedden. Computational millimeter wave imaging: problems, progress and prospects. *IEEE Signal Processing Magazine*, 33(5):109–118, 2016.

[18] N. Simhi and G. Yovel. The contribution of the body and motion to whole person recognition. *Vision Research*, 122:12–20, 2016.

[19] H. Zhang, V. M. Patel, B. S. Riggan, and S. Hu. Generative adversarial network-based synthesis of visible faces from polarimetric thermal faces. *arXiv preprint arXiv:1708.02681*, 2017.