2023

Transfer Learning Using Infrared and Optical Full Motion Video Data for Gender Classification

Alexander M. Glandon
Old Dominion University

Joe Zalameda
Old Dominion University, jzala001@odu.edu

Khan M. Iftekharuddin
Old Dominion University, kiftekha@odu.edu

Gabor F. Fulop (Ed.)

David Z. Ting (Ed.)

See next page for additional authors

Original Publication Citation

Glandon, A. M., Zalameda, J., & Iftekharuddin, K. M. (2023). Transfer learning using infrared and optical full motion video data for gender classification. In G. L. Fulop, D. Z. Ting, & L. L. Zheng (Eds.), *Infrared Technology and Applications XLIX, Proceedings of SPIE 12534 (1253418).* SPIE. *https://doi.org/10.1117/12.2663972*

Invited Paper

Transfer Learning using Infrared and Optical Full Motion Video Data for Gender Classification

Alexander M. Glandon, Joe Zalameda, Khan M. Iftekharuddin
Electrical and Computer Engineering Dept., Old Dominion University, Norfolk, VA, USA 23529

ABSTRACT

This work is a review and extension of our ongoing research in human recognition analysis using multimodality motion sensor data. We review our work on hand crafted feature engineering for motion capture skeleton (MoCap) data, from the Air Force Research Lab for human gender followed by depth scan based skeleton extraction using LIDAR data from the Army Night Vision Lab for person identification. We then build on these works to demonstrate a transfer learning sensor fusion approach for using the larger MoCap and smaller LIDAR data for gender classification.

Keywords: security, gender recognition, person re-identification, motion capture, LIDAR, transfer learning, small data, deep learning

1. INTRODUCTION

The ODU Vision Lab has ongoing work in gender classification and person classification using optical and infrared special purpose sensors [1-5]. Infrared sensors compliment the abilities of visual range sensors for human subject analysis including face recognition [6-9], action recognition [10-14], and gender [15-20] and identity [20-26] recognition. These sensors may be deployed in the field with a small amount of training data available to establish recognition models. Our group began this project using motion capture (MoCap) data collected at the Air Force Research Lab. Our initial work used feature engineering to capture human motion gait signatures in walking data that are used to classify gender in walking and running subjects [1, 5]. Using kinematic features, this work was extended for gender classification in asymmetric limp and aperiodic throwing actions [4]. We also analyzed an infrared LIDAR dataset collected at the Army Night Vision Lab. Using novel silhouette, gait, and depth features we extract a skeleton representation that resembles the MoCap skeleton [2]. We also used this small LIDAR dataset to learn identity by augmenting the data with generative adversarial network samples [3].

In this work we train a deep network to perform gender classification using data from MoCap to pretrain a deep model for the LIDAR data. Typically, deep learning requires a large dataset of training data to prepare a deep model with a large number of initially unknown parameters. In real world situations with special purpose sensors, there may not be a practical way to gather large datasets. However, deep learning can still be applied in these cases through transfer learning. We demonstrate a use case of transfer learning by using MoCap to learn a gender classification model that is transferred to the infrared LIDAR domain for a comparable gender classification task. Infrared LIDAR is a depth and intensity sensor that can measure a point cloud using time-of-flight of a laser and can also generate an infrared intensity image based on magnitude of reflected laser light. This sensor data was provided by the Army Night Vision and Electronic Sensors Directorate. MoCap data provides 3D video in a controlled setting where subjects walk in straight lines. The MoCap data is relatively low noise and contains many samples at a high frame rate. The LIDAR data has several kinds of challenging noise artifacts and has lower frame rate. Therefore, MoCap data is used to improve the gender recognition ability for LIDAR data. The ability to classify subject gender is one of many tasks that can help a real-world security effort to recognize or analyze human identity and behavior. This work continues an effort by our lab [1-5] to analyze human subject data using special sensors including MoCap and LIDAR.

2. BACKGROUND

The advances in deep learning training algorithms are understood to have been enabled by high performance parallel computer processing and collection of large datasets. Crowd sourcing and brute force labeling have been employed to generate extremely large datasets for image and video understanding problems. In addition, special purpose sensors can

Infrared Technology and Applications XLIX, edited by Gabor F. Fulop, David Z. Ting,
Lucy L. Zheng, Proc. of SPIE Vol. 12534, 1253418 · © 2023 SPIE
0277-786X · doi: 10.1117/12.2663972

be employed in a constrained environment to collect a dataset for training a deep learning model. These models can subsequently be deployed in real-world environments. However, in many cases collection of large datasets is not practical.

Small data learning can be divided into two approaches. Machine learning based on handcrafted features, and deep learning based on automated feature (representation [27]) learning. In prior work [2], we established the potential for skeleton extraction using the information rich data modality, LIDAR, and applied the extracted skeletons for human identification. In addition [2], we apply a matrix completion algorithm to estimate the position of occluded human limbs. There is a body of existing work using LIDAR with deep learning method. Given a larger dataset such as LIDAR or other 3D data, a deep learning model can be trained [28-30]. Our given dataset is small, consisting of 10 subjects with a total of 5326 frames of walking activity. There is also existing work using handcrafted features with LIDAR and other 3D data. In [31] the authors extract silhouette features from LIDAR for pedestrian detection. In [32] the authors compliment silhouette with other features from LIDAR for object recognition including pedestrians. In [33] several handcrafted features are compared for human identification. Gait analysis is performed by other authors using motion capture data for human subjects [1, 5, 34]. We demonstrate in [2] that using motion capture type skeleton representation for feature extraction combined with occlusion completion improves human identification performance with our small in house LIDAR data, in comparison to existing human identification work. Using the data rich LIDAR modality, we extract a skeleton representation for robust small data learning. In this work, we approach a small data learning problem, but now we apply deep learning methods.

There are four main techniques that allow deep models to be trained on small datasets. First is synthetic data generation [35-37]. If meaningful data can be artificially computer generated, this can reduce the need for timely manual collection and labeling. Second is model regularization. This is used to fit the model complexity to the amount of data available, such as the dropout method [38]. Neither synthetic data nor model regularization solve the problem of having an incomplete picture of the target distribution due of small data. The next two methods use large datasets related to the target small dataset to try and reduce the space of distributions (constrain or place a prior on the problem) using information gained from the adjacent domain. Transfer learning is one such method [39]. A deep learning model can first be trained on a large, similar dataset to learn a related problem distribution, or prior. Then the model can be further updated by training on a dataset from the target domain, thereby learning a posterior distribution. Transfer learning can be combined with data augmentation as in [40]. The other method is unsupervised learning [41]. This takes unlabeled data, which may be far less expensive to acquire than labeled data and learns underlying trends in the distribution. This information can be applied next to a model that performs a classification activity with a small subset of the data with labels.

In this work we demonstrate hand crafted feature engineering for MoCap skeleton data from the Air Force Research Lab for human gender recognition in walking, running, limping, and throwing. Next, we develop a LIDAR skeleton modality from raw LIDAR data from the Army Night Vision Lab. This extracted 3D skeleton is initially used for human identification. The LIDAR skeleton is then used with a generative network to augment the dataset and perform human subject identification. Finally, the MoCap skeleton is used in transfer learning as a larger dataset to transfer knowledge to the smaller LIDAR dataset for gender classification.

3. METHOD

This work builds upon related ongoing works at the ODU Vision Lab [1-5] in multimodal image based learning for human subject classification tasks. The first two projects in section 3.1 and section 3.2 are based on a MoCap dataset collected at the Air Force Research Lab. These MoCap videos give the 3D location of 31 locations or joints of the human subject. This data was used to develop models for gender classification in walking, running, limping (asymmetric action), and throwing (aperiodic action). The next two projects add an Army Night Vision Lab LIDAR video dataset giving 3D depth scans of human subjects walking in front of an infrared laser. Section 3.3 describes a novel algorithm for skeleton data extraction to emulate MoCap data using the depth scans. These LIDAR skeletons are used for person re-identification. Section 3.4 describes training a generative model using the extracted LIDAR skeleton data to create synthetic examples for improving classification accuracy with our small LIDAR skeleton dataset. Finally, section 3.5

describes the latest work using the larger Airforce MoCap skeleton data in transfer learning to fuse MoCap and LIDAR image modalities to improve the accuracy of gender classification from the smaller Army LIDAR skeleton dataset.

3.1 MoCap Gender Recognition in Periodic Walking and Running

The initial work [1, 5] in gender analysis used skeleton data collected using MoCap sensors at the Air Force Research Lab. Fig. 1 shows an example of a MoCap data frame from the video sequence. The skeleton data takes the form of a 93-dimensional vector for each frame in time. This contains the information from 31 motion capture location markers attached to the human subject in 3D space. In [1] principal component analysis (PCA) is used to reduce the dimension of

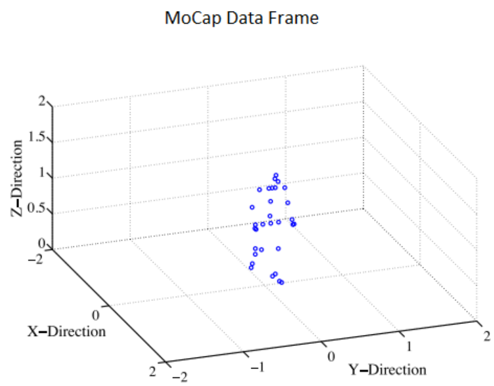

Fig. 1a MoCap Skeleton Frame [1]

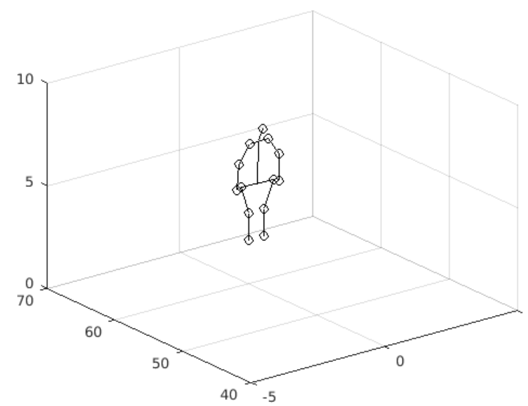

Fig 1b LIDAR Skeleton Frame

the data for the videos for each subject. The PCA weights features from non-pathological gait studies [42, 43] based on the statistical significance in relation to gait analysis as shown in table 1. The size of each skeleton is normalized so that gender classification is not biased by skeleton height. Periodicity information including FFT frequency and phase

Table 1. Features extracted for asymmetrical limping and aperiodic throwing actions [1].

Feature	Statistical p-Value
Cadence: Steps per minute	0.03
Step Width: Gap between the feet	0.046
Pelvic Obliquity: Pelvis range of motion along the frontal plane	0.003
Torso Sway: Torso range of motion	0.01
Shoulder Excursion	0.044

information is appended to the vector to add motion information for each subject. Different classifiers are applied including linear discriminant analysis (LDA) and support vector machine (SVM) and nonlinear SVM. Finally, features are selected from body part groupings to determine if certain skeleton joints are more representative of the gender specific gait. This gait based identification work is expanded in [5] where a more challenging action of running subject are studied using similar features derived from PCA with linear classifiers LDA and SVM and nonlinear classifiers SVM with modified kernel and decision stump with AdaBoost. In that work it is found that linear classification is sufficient for gender recognition using these gait-based skeleton features.

3.2 MoCap Gender Recognition in Asymmetrical Limping and Aperiodic Throwing

The next part of our work [4] uses the same skeleton dataset for gender classification, but changes the underlying action of interest from walking and running to asymmetrical limping and aperiodic throwing actions. Kinematic features for throwing based on angles and velocities of the upper body throwing arm are extracted. Features for limping focusing on lower body and gait cycle that are also extracted. After these kinematic features are identified and collected, a short list of statistical features including mean, median, maximum, minimum, standard deviation, variance, sum of values, and length is calculated. An addition longer list is calculated using the python library tsfresh. The following machine learning models are applied to subsets of the features: SVM, SVM with radial basis function (RBF), decision tree, decision tree with AdaBoost, and K-nearest neighbor. Regular SVM (without RBF) gives the highest accuracy for both throwing and limping. The best feature subset for throwing returns 99.75% gender classification accuracy. The best feature subset for limping returns 98.53% gender classification accuracy.

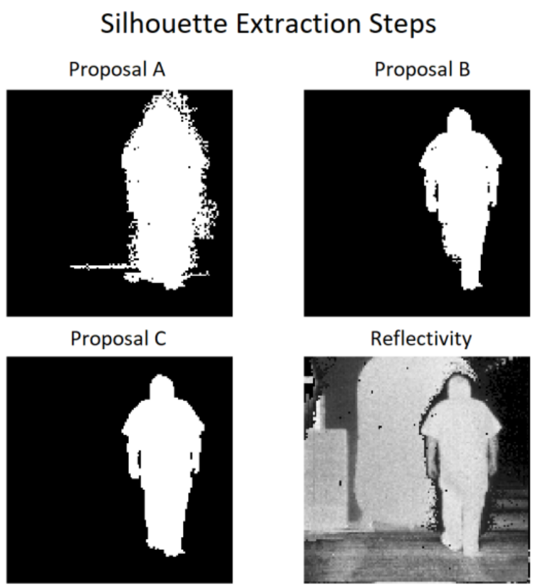

Fig. 2 Silhouette Extraction from LIDAR [2]

3.3 LIDAR data skeleton extraction and human identification

The next step our group took was to examine a LIDAR dataset of depth scanning of human subjects. Many other groups performing depth analysis of human subject motion are focusing on Kinect depth sensors. The advantage of our sensor is the range of operation is increased from 11.5 feet for Kinect to 60+ feet for the LIDAR from the Army Night Vision Lab. This is a great advantage for security applications where distance of perception is an important factor.

Taking inspiration from the MoCap skeleton data modality, we developed a computational model for skeleton extraction from LIDAR video. Fig 1a. gives an example of the MoCap skeleton and Fig 1b. the LIDAR skeleton. This model begins with a novel silhouette extraction algorithm using the depth and intensity scans of a human subject. It is based on a set of 4 proposed silhouettes that successively eliminate different components of the LIDAR noise as shown in Fig 2. Next walk direction vectors are extracted using the silhouette for side-to-side motion and the depth scan for back-and-forth motion. These extracted 2D walk vectors give an idea of the orientation of the walking subject. Next the walk is decomposed into phases based on the orientation of the subject's legs, arms, and body as a whole. A phase specific algorithm is used to extract the lower and upper arms, lower and upper legs, and spine and head. Using the phase information and a novel combination of morphological operations and constrained Hough transform in the silhouette space, and augmenting with depth information, 13 joints in total are extracted in 3D space over each frame as shown in Fig 1b. These joints are used to extract bone length, which can overcome a scenario where a subject is trying to mask their identity, as bone length cannot be easily faked. Using the skeleton and silhouette features together, an SVM is able to perform 91.69% identity recognition in a gallery of 10 subjects.

3.4 Synthetic Data Augmentation for LIDAR Skeleton Human Identification

The LIDAR data is next used to train a deep network called Hierarchical Co-occurrence Network (HCN) [44]. An augmented dataset is developed using generative adversarial network (GAN) for human re-identification [3]. The GAN generator and discriminator architectures are shown in Fig. 3. Using a 50% synthetic dataset, we are able to increase accuracy from 74.49% baseline to 78.22% accuracy on 10 subject gallery in LIDAR data. This demonstrates that deep networks can be improved for small data modeling using data augmentation.

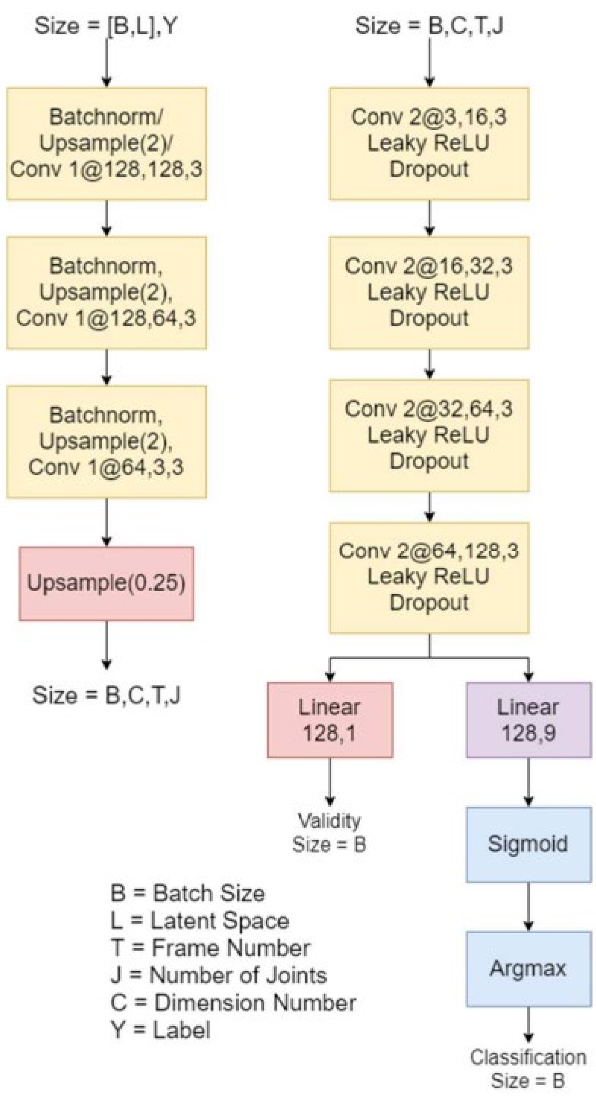

Fig. 3 GAN Generator and Discriminator [3]

3.5 Transfer Learning using MoCap Skeleton and LIDAR Skeleton Data

In addition to data augmentation, transfer learning is another paradigm that can be used to apply deep networks to smaller datasets. The LIDAR data is challenging for the following reasons: The data contains different noise artifacts and even some dead pixels. The intensity returned is grayscale for a single laser frequency being reflected. The ground truth is not labeled for the joint locations, which prohibits supervised learning. The dataset is small with only 5326 frames in total for all 10 subjects. Lastly, the resolution is small at 128x128 pixels with human subjects at a distance in the field of view. The MoCap dataset is a larger dataset with less noise than the LIDAR data. We pre-train the HCN convolutional neural network on MoCap data to perform gender classification in walking data. Then the LIDAR data is used to fine tune the HCN deep network. The MoCap data is pretrained on gender classification using 10-fold cross validation. For each fold, after the MoCap data is used for pretraining the HCN, the LIDAR dataset is then used for fine tuning. After both experiments are complete, we obtain 10 results for 10 folds MoCap, each with a train/test LIDAR 90/10 split. Both pretraining and fine-tuning are run for 100 epochs. The same is repeated for a simple CNN with just two convolutional layers because of small datasets in this study. The results and discussion of this combined experiment are given as follows.

4. RESULTS AND DISCUSSION

First, we present the results for 10-fold cross validation on the HCN model. For MoCap training, we obtain a 10-fold average of 81.80% gender classification accuracy +/- 3.13 standard deviation. For MoCap testing, we obtain 61.82% +- 20.94. Figure 4a shows the loss curve for the first fold of MoCap training. Figure 4b gives the training accuracy for the same fold. When we use a 90/10 split for training and testing using LIDAR dataset, we obtain training accuracy of 95.63% +/- 0.44% over the 10 folds from MoCap pretraining. The testing accuracy over the same folds is 93.04% +/- 1.56%. Figure 5 gives the loss curve and accuracy curve for the LIDAR training, respectively.

Next for a simple CNN, we report complimentary transfer learning results for the LIDAR dataset. For MoCap training, we obtain 10-fold average of 98.34% gender classification accuracy +/- 0.68 standard deviation. For MoCap testing, we obtain 77.57% +/- 16.01. Figure 6a gives the loss curve for the first fold of MoCap training. Figure 6b shows the training accuracy for the same fold. When we use a 90/10 split for training and testing using LIDAR dataset, we obtain training accuracy of 96.92% +/- 0.99% over the 10 folds from MoCap pretraining. The testing accuracy over the same folds is 95.63% +/- 0.94%. Figure 7 gives the loss curve and accuracy curve for the LIDAR training.

As shown in Figs. 4 and 5 the HCN training curves are very jagged. This is possibly due to the higher model complexity of HCN. In Figures 6 and 7 the simple CNN training curves are relatively smooth for the small parameter network. We also notice that the MoCap data for pretraining has a larger distance between training and testing accuracy for both HCN and simple CNN. Finally, we report ROC curves for HCN and CNN on the fine-tuned LIDAR models in Figure 8. We note that the lower complexity simple CNN model gives better true positive rate for lower false positive rates.

HCN Training Loss and Accuracy on MoCap Pretraining

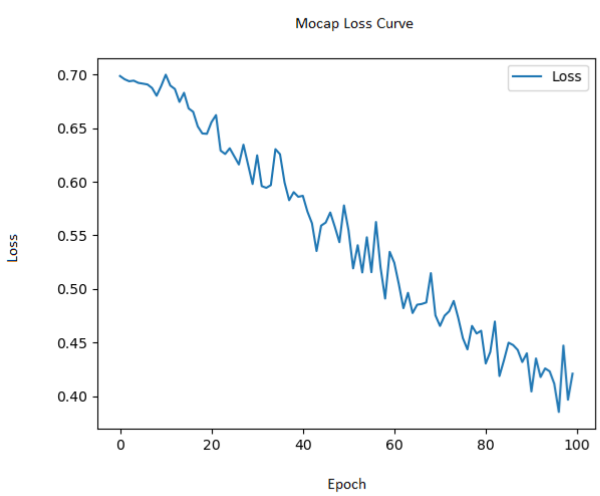

Fig 4a. MoCap Training Loss

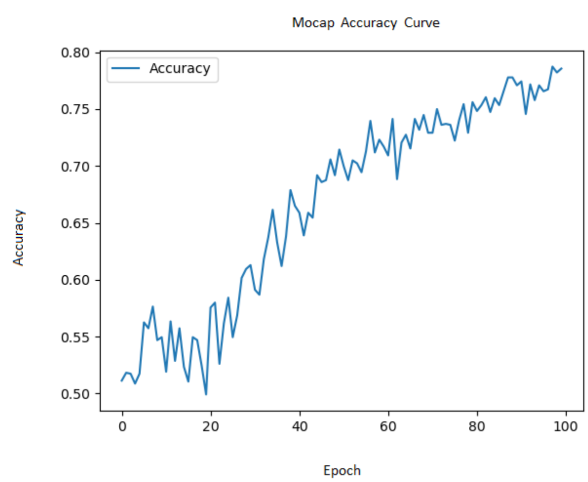

Fig 4b. MoCap Training Accuracy

HCN Training Loss and Accuracy of LIDAR Fine Tuning

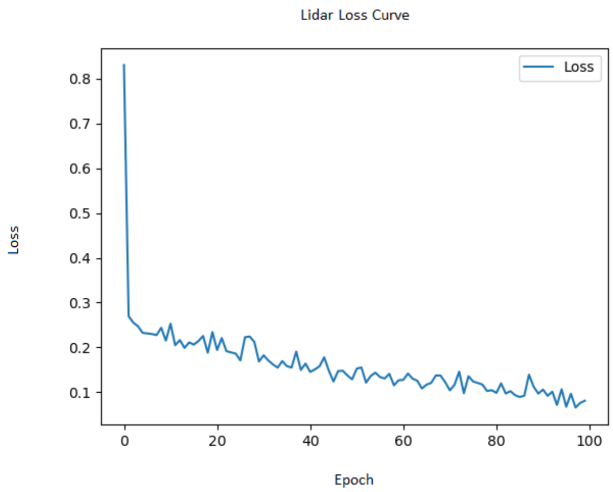

Fig 5a. LIDAR Training Loss

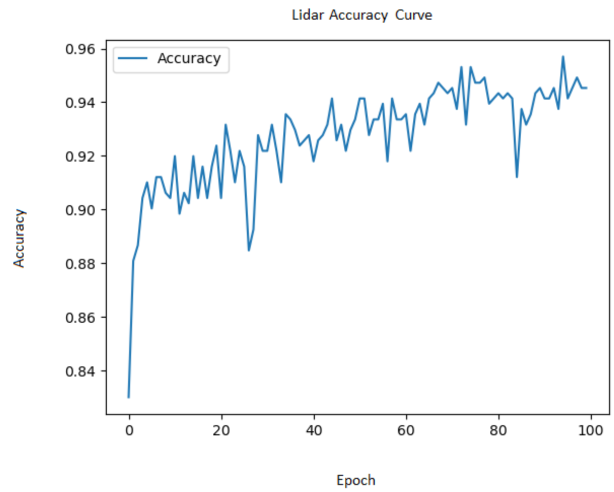

Fig 5b. LIDAR Training Accuracy

Simple CNN Training Loss and Accuracy on MoCap Pretraining

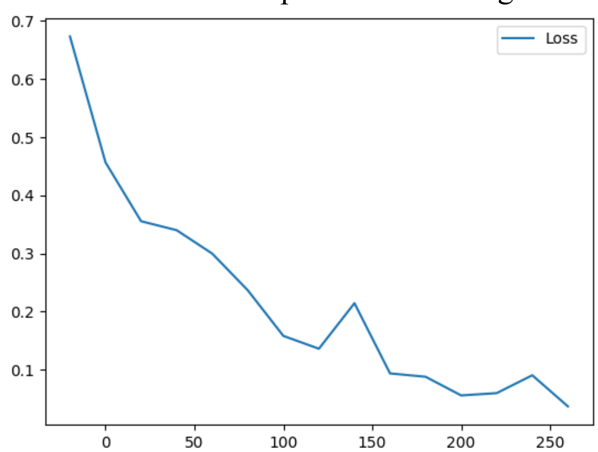

Fig 6a. MoCap Training Loss

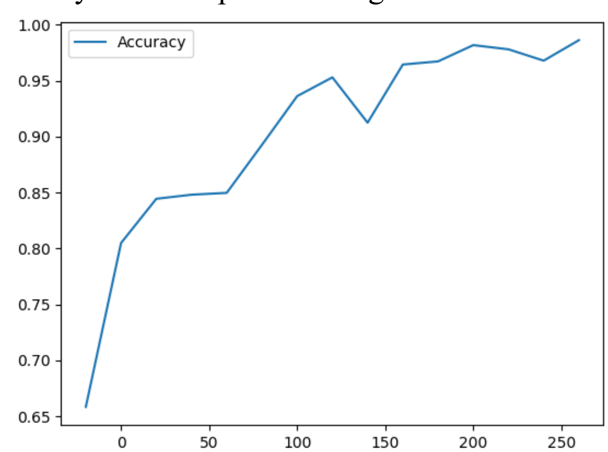

Fig 6b. MoCap Training Accuracy

Simple CNN Training Loss and Accuracy on LIDAR Fine Tuning

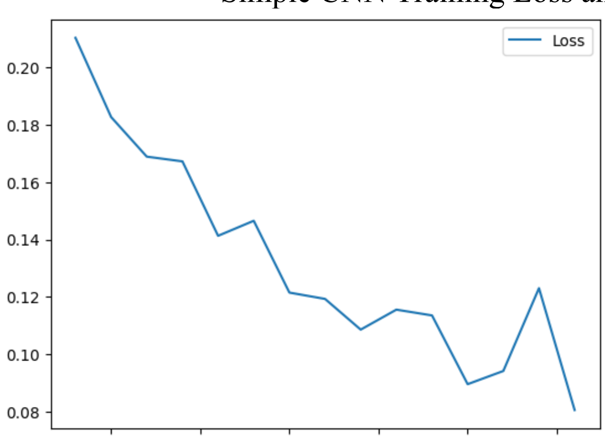

Fig 7a. LIDAR Training Loss

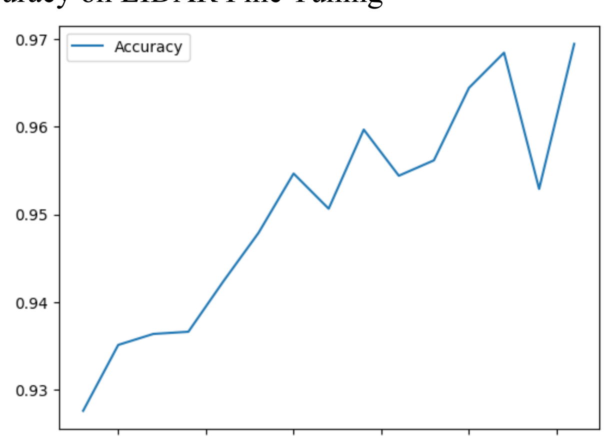

Fig 7. LIDAR Training Accuracy

ROC Curves for LIDAR Fine Tuning

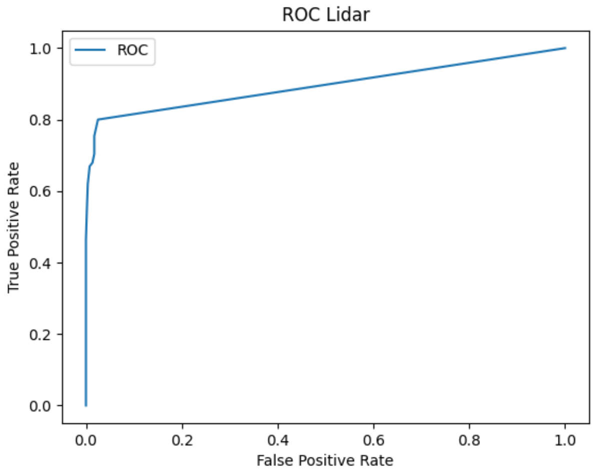

Fig 8a. HCN ROC

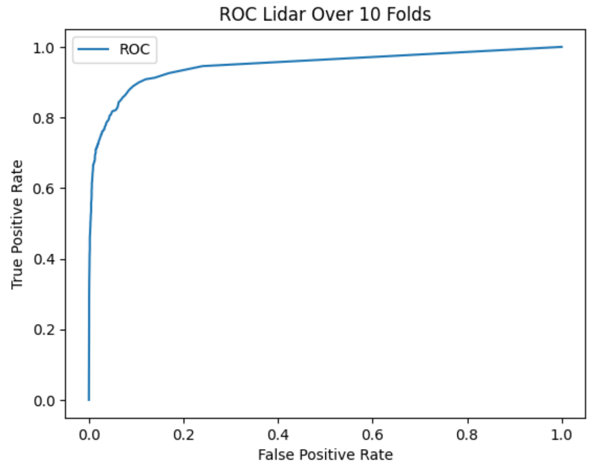

Fig 8b. Vanilla CNN ROC

5. CONCLUSION

In summary, this work is a continuation of an effort by the ODU Vision Lab to analyze human skeleton for gender and re-identification. We began with MoCap skeleton data collected by the Air Force Research Lab. Using feature engineering, we are able to extract information used to classify gender of walking and running subjects. Continuing this work, we selected asymmetric limping and aperiodic throwing to broaden the type of actions in performing gender classification. Next, we add LIDAR data collected from the Army Night Vision Lab. We are able to extract human skeletons using novel silhouette and depth feature extraction. The LIDAR skeletons are used to perform person re-identification. Next, the human skeleton extracted from the LIDAR is used to train a GAN. This augmented data improves classification performance in a re-identification deep network. Finally, using the larger MoCap data, we pretrain a HCN network and fine tune with LIDAR to perform human gender classification. We have used these two datasets with several techniques aimed at small data including feature engineering, data augmentation, and transfer learning a deep model. In future work we plan to expand the LIDAR dataset, as it currently contains only one female subject, albeit across multiple videos. The MoCap data is better balanced with nearly a 50/50 split in gender. In the future we also believe we can obtain improved results by training for more epochs, as the training curve does not reach any plateau currently in the MoCap pretraining. We believe this body of work can serve as inspiration for the many approaches that can be taken with a novel dataset to generate interesting models for a variety of recognition tasks, especially for security domain.

6. ACKNOWLEDGEMENT

The authors would like to acknowledge partial support of this work by US Army NVESD, CERDEC through Grant No. 100659, DoD Center of Excellence in AI and Machine Learning (CoE-AIML) under Contract Number W911NF-20- 2-0277 with the U.S. Army Research Laboratory, and by the National Science Foundation under Grant No. 1828593, Grant No. 1950704, and Grant No. 1753793.

REFERENCES

[1] J. B. Flora, D. F. Lochtefeld, D. A. Bruening, and K. M. Iftekharuddin, "Improved gender classification using nonpathological gait kinematics in full-motion video," *IEEE Transactions on Human-Machine Systems,* vol. 45, no. 3, pp. 304-314, 2015.

[2] A. Glandon *et al.*, "3d skeleton estimation and human identity recognition using lidar full motion video," in *2019 International Joint Conference on Neural Networks (IJCNN)*, 2019: IEEE, pp. 1-8.

[3] J. G. Zalameda, B. Kruse, A. M. Glandon, M. A. Witherow, S. Shetty, and K. M. Iftekharuddin, "Generalized Adversarial and Hierarchical Co-occurrence Network based Synthetic Skeleton Generation and Human Identity Recognition," in *2022 International Joint Conference on Neural Networks (IJCNN)*, 2022: IEEE, pp. 1-8.

[4] M. Navarro, A. Glandon, N. K. Dhar, and K. M. Iftekharuddin, "Gender classification of full-body biological motion of aperiodic actions using machine learning," in *Optics and Photonics for Information Processing XIV*, 2020, vol. 11509: SPIE, pp. 21-30.

[5] C. M. Williams, J. B. Flora, and K. M. Iftekharuddin, "Gender classification of running subjects using full-body kinematics," in *Geospatial Informatics, Fusion, and Motion Video Analytics VI*, 2016, vol. 9841: SPIE, pp. 13-21.

[6] S. Guo, T. Zhang, and T. Guo, "Thermal infrared face detection method based on deep learning," in *Eighth Symposium on Novel Photoelectronic Detection Technology and Applications*, 2022, vol. 12169: SPIE, pp. 433-439.

[7] K. Ko, H. Gwak, N. Thoummala, H. Kwon, and S. Kim, "SqueezeFace: Integrative face recognition methods with LiDAR sensors," *Journal of Sensors,* vol. 2021, pp. 1-8, 2021.

[8] Y. Kim *et al.*, "CloudNet: A LiDAR-Based Face Anti-Spoofing Model That Is Robust Against Light Variation," *IEEE Access,* vol. 11, pp. 16984-16993, 2023.

[9] A. M. Srivastava, S. D. Chintaginjala, S. C. Bhogavalli, and S. Prakash, "Robust face recognition using multimodal data and transfer learning," *Journal of Electronic Imaging,* vol. 32, no. 4, p. 042105, 2022.

[10] I. Morawski and W.-N. Lie, "Two-stream deep learning architecture for action recognition by using extremely low-resolution infrared thermopile arrays," in *International Workshop on Advanced Imaging Technology (IWAIT) 2020*, 2020, vol. 11515: SPIE, pp. 164-169.

[11] D. Zhao and M. Zhi, "A review of action recognition methods based on skeleton data," in *Thirteenth International Conference on Graphics and Image Processing (ICGIP 2021)*, 2022, vol. 12083: SPIE, pp. 783-791.

[12] R. D. Brehar, M. P. Muresan, T. Mariţa, C.-C. Vancea, M. Negru, and S. Nedevschi, "Pedestrian street-cross action recognition in monocular far infrared sequences," *IEEE Access,* vol. 9, pp. 74302-74324, 2021.

[13] A. M. De Boissiere and R. Noumeir, "Infrared and 3d skeleton feature fusion for rgb-d action recognition," *IEEE Access,* vol. 8, pp. 168297-168308, 2020.

[14] J. Imran and B. Raman, "Deep residual infrared action recognition by integrating local and global spatio-temporal cues," *Infrared Physics & Technology,* vol. 102, p. 103014, 2019.

[15] M. A. Farooq and P. Corcoran, "Proof-of-Concept Techniques for Generating Synthetic Thermal Facial Data for Training of Deep Learning Models," in *2021 IEEE International Conference on Consumer Electronics (ICCE)*, 2021: IEEE, pp. 1-6.

[16] C. Chen and A. Ross, "Evaluation of gender classification methods on thermal and near-infrared face images," in *2011 International Joint Conference on Biometrics (IJCB)*, 2011: IEEE, pp. 1-8.

[17] M. Berksan, "Gender recognition and age estimation based on human gait," Başkent Üniversitesi Fen Bilimleri Enstitüsü, 2019.

[18] M. A. Farooq, H. Javidnia, and P. Corcoran, "Performance estimation of the state-of-the-art convolution neural networks for thermal images-based gender classification system," *Journal of Electronic Imaging,* vol. 29, no. 6, pp. 063004-063004, 2020.

[19] N. R. Baek, S. W. Cho, J. H. Koo, and K. R. Park, "Pedestrian Gender Recognition by Style Transfer of Visible-Light Image to Infrared-Light Image Based on an Attention-Guided Generative Adversarial Network," *Mathematics,* vol. 9, no. 20, p. 2535, 2021.

[20] N. R. Baek, S. W. Cho, J. H. Koo, N. Q. Truong, and K. R. Park, "Multimodal camera-based gender recognition using human-body image with two-step reconstruction network," *IEEE Access,* vol. 7, pp. 104025-104044, 2019.

[21] T. Yan, F. Wang, Y. Xu, and Y. Qian, "Data augmented multi-loss hybrid learning for cross-modality person re-identification," in *International Workshop on Automation, Control, and Communication Engineering (IWACCE 2022)*, 2022, vol. 12492: SPIE, pp. 268-274.

[22] J. Sun, Y. Li, H. Chen, and Y. Peng, "A sample diversity and identity consistency based cross-modality model for visible-infrared person re-identification," in *Fourteenth International Conference on Digital Image Processing (ICDIP 2022)*, 2022, vol. 12342: SPIE, pp. 250-259.
[23] Y. Li, T. Zhang, X. Liu, Q. Tian, Y. Zhang, and F. Wu, "Visible-Infrared Person Re-Identification With Modality-Specific Memory Network," *IEEE Transactions on Image Processing,* vol. 31, pp. 7165-7178, 2022.
[24] N. Huang, J. Liu, Y. Miao, Q. Zhang, and J. Han, "Deep learning for visible-infrared cross-modality person re-identification: A comprehensive review," *Information Fusion,* 2022.
[25] Y. Chen, L. Wan, Z. Li, Q. Jing, and Z. Sun, "Neural feature search for rgb-infrared person re-identification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 587-597.
[26] M. Ye, J. Shen, D. J. Crandall, L. Shao, and J. Luo, "Dynamic dual-attentive aggregation learning for visible-infrared person re-identification," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVII 16*, 2020: Springer, pp. 229-247.
[27] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE transactions on pattern analysis and machine intelligence,* vol. 35, no. 8, pp. 1798-1828, 2013.
[28] D. Maturana and S. Scherer, "Voxnet: A 3d convolutional neural network for real-time object recognition," in *2015 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, 2015: IEEE, pp. 922-928.
[29] L. Chen, H. Wei, and J. Ferryman, "A survey of human motion analysis using depth imagery," *Pattern Recognition Letters,* vol. 34, no. 15, pp. 1995-2006, 2013.
[30] E. H. Lim and D. Suter, "3D terrestrial LIDAR classifications with super-voxels and multi-scale Conditional Random Fields," *Computer-Aided Design,* vol. 41, no. 10, pp. 701-710, 2009.
[31] K. Kidono, T. Miyasaka, A. Watanabe, T. Naito, and J. Miura, "Pedestrian recognition using high-definition LIDAR," in *2011 IEEE Intelligent Vehicles Symposium (IV)*, 2011: IEEE, pp. 405-410.
[32] A. Teichman, J. Levinson, and S. Thrun, "Towards 3D object recognition via classification of arbitrary object tracks," in *2011 IEEE International Conference on Robotics and Automation*, 2011: IEEE, pp. 4034-4041.
[33] B. Gálai and C. Benedek, "Feature selection for Lidar-based gait recognition," in *2015 International workshop on computational intelligence for multimedia understanding (IWCIM)*, 2015: IEEE, pp. 1-5.
[34] N. F. Troje, "Decomposing biological motion: A framework for analysis and synthesis of human gait patterns," *Journal of vision,* vol. 2, no. 5, pp. 2-2, 2002.
[35] A. Figueira and B. Vaz, "Survey on synthetic data generation, evaluation methods and GANs," *Mathematics,* vol. 10, no. 15, p. 2733, 2022.
[36] M. Abufadda and K. Mansour, "A survey of synthetic data generation for machine learning," in *2021 22nd International Arab Conference on Information Technology (ACIT)*, 2021: IEEE, pp. 1-7.
[37] S. Lamczyk, K. Ampofo, B. Salashour, M. Cetin, and K. M. Iftekharuddin, "SURFGenerator: Generative Adversarial Network Modeling for Synthetic Flooding Video Generation," in *2022 International Joint Conference on Neural Networks (IJCNN)*, 2022: IEEE, pp. 1-8.
[38] L. Brigato and L. Iocchi, "A close look at deep learning with small data," in *2020 25th International Conference on Pattern Recognition (ICPR)*, 2021: IEEE, pp. 2490-2497.
[39] F. Zhuang *et al.*, "A comprehensive survey on transfer learning," *Proceedings of the IEEE,* vol. 109, no. 1, pp. 43-76, 2020.
[40] N. N. Tu, S. Sako, and B. Kwolek, "Fingerspelling recognition using synthetic images and deep transfer learning," in *Thirteenth International Conference on Machine Vision*, 2021, vol. 11605: SPIE, pp. 528-535.
[41] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7167-7176.
[42] R. Frimenko, C. Whitehead, and D. Bruening, "Do men and women walk differently? A review and meta-analysis of sex difference in non-pathological gait kinematics," 2014.
[43] D. A. Bruening, R. E. Frimenko, C. D. Goodyear, D. R. Bowden, and A. M. Fullenkamp, "Sex differences in whole body gait kinematics at preferred speeds," *Gait & posture,* vol. 41, no. 2, pp. 540-545, 2015.
[44] C. Li, Q. Zhong, D. Xie, and S. Pu, "Co-occurrence feature learning from skeleton data for action recognition and detection with hierarchical aggregation," *arXiv preprint arXiv:1804.06055,* 2018.