3D LIDAR Human Skeleton Estimation for Long Range Identification*

A. Glandon[1], L. Vidyaratne[1], N. Sadeghzadehyazdi[2], Nibir K. Dhar[3], Jide O. Familoni[3], Scott T. Acton[2], K. M. Iftekharuddin[1]

1 Electrical and Computer Engineering, Old Dominion University

2 Electrical and Computer Engineering, University of Virginia

3 Night Vision and Electronic Sensors Directorate, Fort Belvoir

Introduction

This research project is the continuation of an effort by the ODU Vision Lab to understand human gait and motion using special-purpose imaging sensors and novel computer vision algorithms.

The current phase of this project extends earlier work with Motion Capture (MoCAP) imaging to long range LIDAR imaging. Using 3D LIDAR data, we designed an algorithm to identify human subjects at a range of up to 60 ft.

Imaging Modalities

MoCAP

Motion Capture (MoCap) is a 3D imaging system which measures 3D human skeleton information in real-time by attaching markers to a human subject and viewing the human motion with a set of cameras at different angles. The first phase of this work used these skeletons to determine the gender of a human subject while walking.

3D LIDAR

LIDAR scans the surface of objects and gives a 3D depth map of the objects in real-time. We developed a computer vision system that can estimate a human skeleton from Lidar data, which resembles the MoCap data format. Using these computed 3D skeletons, we perform human identification.

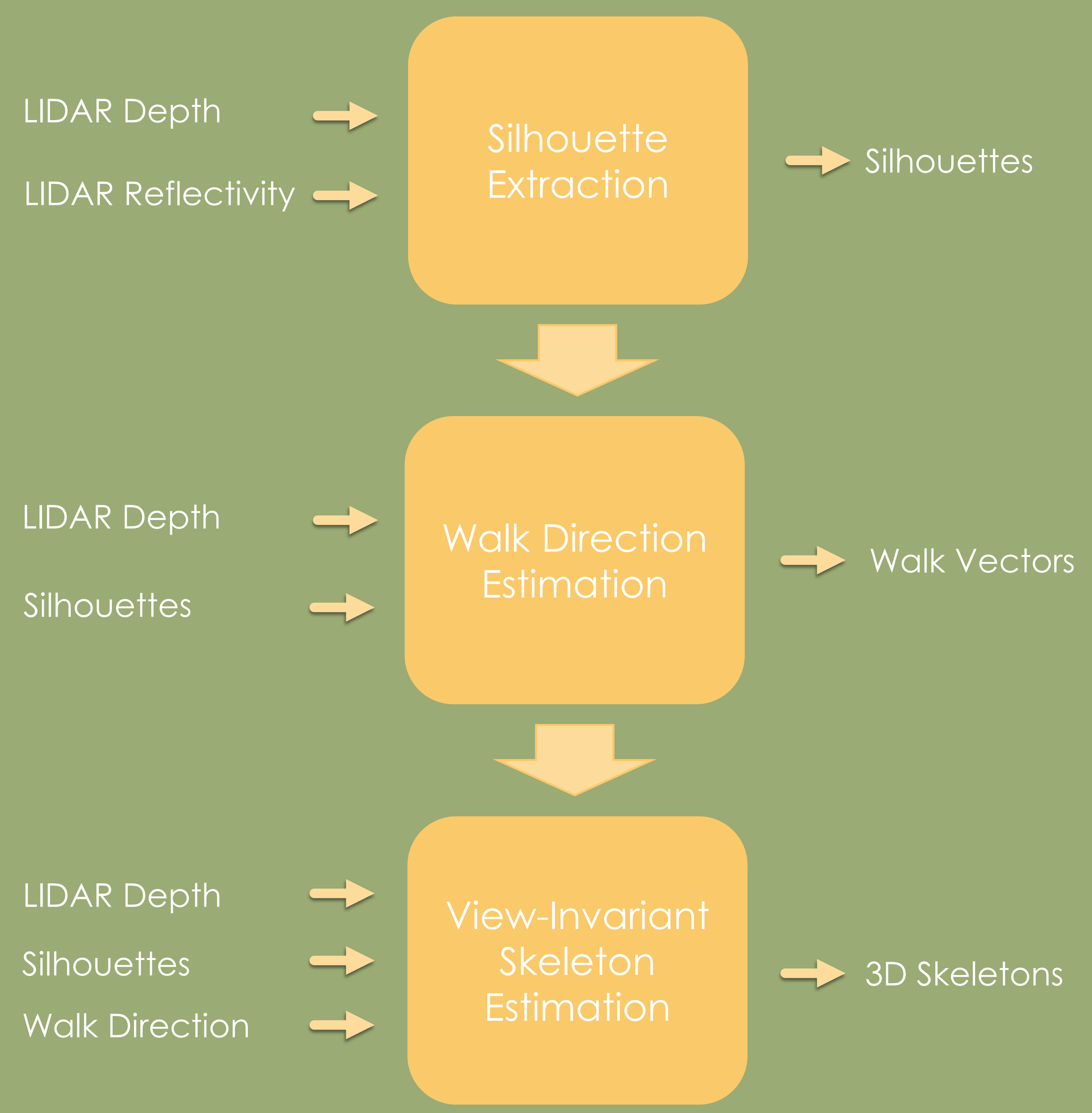
Shortwave Infrared

Works at night where objects reflect ambient glow from atmosphere not visible to human eyes. Also capable of imaging through fog or haze due to potential to transmit through water.

Longwave Infrared

This is the traditional "thermal" based infrared imaging, where temperature differences are detected and, like shortwave, is also useful in a night setting.

LIDAR Algorithm



Silhouette extraction is based on a novel algorithm using a series of refined silhouette proposals using LIDAR sensor reflectivity and depth channels. Walk vector estimation is based on digital signal processing using the extracted silhouettes and the LIDAR depth. Finally, 3D skeleton is extracted using image processing tools including iterative thinning, a specially constrained Hough line transform, and novel joint specific algorithms. Body parts that are occluded from view of the LIDAR are also estimated using a matrix completion algorithm.

Results Comparison*

Algorithm	Data	Activity	Number of Subjects in Dataset	Features / Model	Top-1 Recognition Performance
Recurrent Attention Model	IAS-Lab RGB/D	freely moving	11	Combined LSTM and CNN Attention Model	64.4%
Recurrent Attention Model	IIT PAVIS RGB/D	not specified	79	Combined LSTM and CNN Attention Model	43.0%
Point Cloud Matching	BIWI RGB/D	walking	50	Warped Point Cloud Matching / Kinect Features	90.2%
Point Cloud Matching	IAS-Lab RGB/D	freely moving	11	Warped Point Cloud Matching / Kinect Features	86.3%
OpenPose	NVESD gray/D	walking	10	2D Intensity Pose Model and Silhouette / SVM	85.00%
Proposed Computational Modeling	**NVESD gray/D**	**walking**	**10**	**3D Computational Model + Silhouette / SVM**	**91.69%**

In comparison to a deep recurrent model, our method uses small data without a deep network and achieves higher identification accuracy for a similar number of subjects. In comparison to point cloud matching, our method gives competitive performance to their 640x640 resolution with our LIDAR resolution of only128x128 pixels. We also tested an existing OpenPose model on our data and find better identification accuracy using our skeleton extraction algorithm.

* Glandon, A., Vidyaratne, L., Sadeghzadehyazdi, N., Dhar, N. K., Familoni, J. O., Acton, S. T., & Iftekharuddin, K. M. (2019, July). 3d skeleton estimation and human identity recognition using lidar full motion video. In 2019 International Joint Conference on Neural Networks (IJCNN) (pp. 1-8). IEEE.

Future Work

The next phase of this project proposes shortwave and longwave infrared sensors to recognize human action. We are working to adapt an unsupervised learning method called domain adaptation. There already exists large datasets of visual imaging human actions. Given the size of these datasets, we can develop a representation of human action that can be adapted to infrared data. The goal will be to find a representation that captures the information about human action in both domains, so that the visual learning can be reapplied to the infrared images.

Acknowledgements

Support provided by US Army NVESD, CERDEC through a grant #100659 and also providing 3D Lidar data.

Contact Email: aglan001@odu.edu (Alex Glandon)