Sidewalk-level People Tracking with a Low-cost 3D LIDAR System

Constantin Savtchenko and John R. Spletzer

Abstract— **Three-dimensional (3D) LIDAR systems are becoming the sensor of choice for many mobile robotics applications. This can be attributed to their accuracy, robustness, and strong invariance to ambient illumination levels. In many applications, adoption is only limited by the high system cost. In this work, we investigate the use of low-cost 3D LIDAR systems in a people tracking application for a smart wheelchair system. The limited spatial resolution of these systems proved challenging in the people tracking task. To solve this problem, we employed a k-Nearest-Neighbor (k-NN) appearance classifier in conjunction with an extended Kalman Filter (EKF)-based motion classifier. Preliminary experimental results indicated a successful tracking rate of over 95%.**

I. INTRODUCTION

The year 2010 is considered by some to be the tipping point for the mobile robotics field [1]. Educational robots such as iRobot's Create platform are ubiquitous at universities [], Willow Garage's Robot Operating System (ROS) continues to gain momentum [2], and low cost 3D sensors like Microsoft's Kinect enable a range of indoor applications at a previously unheard of price point [3]. Indeed, one of the most significant enabling technologies for outdoor mobile robotics is the recent development of three-dimensional (3D) LIDAR systems. The case for 3D LIDARs was made convincingly in the 2007 DARPA Urban Challenge. Only 6 vehicles (of 89 original entries) completed the race, but all 6 relied upon (then) new 3D LIDAR systems [4].

The focus of our own research is enhancing the autonomy of smart wheelchair systems in urban environments. Urban areas and sidewalks in particular are often congested with pedestrian traffic. Since our wheelchair systems will share this space, safe operation is a paramount concern. As a result, we have significant interest in being able to reliably detect and track persons in a range of environmental conditions. While much work has focused upon 3D LIDARS for use on automobiles, little has been done with field and service robots operating at the sidewalk level. This is not unexpected, as the high cost of 3D LIDARs makes them not economically viable for many applications. However, the recent release of low-cost/resolution 3D flash LIDARs offers the potential for robust 3D perception in outdoor environments.

The objective of this work is to investigate the suitability of these systems for real-time people tracking. The results have potentially broad applications to field and service robotics, in cases where human-robot interactions are of necessity.

C. Savtchenko is with the Computer Science and Engineering, Lehigh University, Bethlehem, PA, USA cas210@lehigh.edu

J. Spletzer is an Associate Professor of Computer Science and Engineering, Lehigh University, Bethlehem, PA, USA spletzer@cse.lehigh.edu

II. RELATED WORK

There has been extensive work using camera systems for pedestrian tracking, a survey of which by Geronimo *et al* can be found at [5]. Fortunately, much of the limitations they identified with camera-based people tracking systems (*e.g.*, ground plane estimation, region of interest extraction, *etc.*) can be resolved with three dimensional data.

There has also been recent work with LIDAR-based approaches. Spinello and Siegwart used a Velodyne 3D LIDAR to approach pedestrian tracking with in parts [6]. Their approach split a pedestrian into layers, and each layer voted to determine if the object is a pedestrian. Their approach is powerful in that they ignore the ground plane (*e.g.*, when dealing with multi-level ground planes, sidewalks to roads, *etc.*) and that it can leverage techniques developed and expanded upon in 2D LIDAR research. Douillard *et al* combined a 3D lidar with a camera for urban classification []. Objects such as trees, cars, and pedestrians were classified using a rule-based system approach.

Navarro-Serment *et al* used a combination of multiple 2D and 3D scanners [7]. Point clouds gathered by the 3D scanner were projected into 2D after ground plane removal. Classification was then done using thresholds set for an object geometry, velocity, and distance. Tracking was done using a simple rule; if the object was inside the bounding box of an object from the previous frame, they wre the same object. In later work [8], a two- dimensional approach with two linear SVMs was similar. This paper is similar in that uses a Geometric Score (GS) and Motion Score (MS) to make classifications.

Prokhorov used a Recurrent Neural Network to achieve a high vehicle detection rate [9]. His approach relied on the RNN to do the segmentation based on the temporal order of points received from a Velodyne LIDAR. Object detection was high with a low amount of false positives. This paper stands out in its unique classification and segmentation method.

Compared to these works, our approach differs in several ways. First and foremost, we are interested in people tracking at the sidewalk level using low-cost 3D LIDARs. The limited resolution of such LIDARs introduces significant challenges to the classification process. To meet this goal, we employ a k-Nearest-Neighbor (k-NN) appearance classifier in conjunction with an EKF-based motion classifier. The former proved to be advantageous in dealing with imperfect data,

while the latter served to improve tracking frequency and reduce false negatives. Details on the approach now follow.

III. TECHNICAL APPROACH

A. Development Platform

The development platform for this project was based upon the smart wheelchair system (SWS) developed for the Automated Transport and Retrieval System (ATRS) shown at Figure 1 [10]. Odometry measurements are provided by high-resolution quadrature encoders (8,192 CPR), with gyroscope corrections from a Microstrain 3DM-GX1 IMU []. For exteroceptive sensing, the SWS integrates an IFM Effector O3D200 flash LIDAR. The IFM provides 3D measurements of the environment at reasonable cost (<$1,500US). The tradeoff for the low price point is fairly coarse resolution (64x50) over a field of view of 40° x 30° , respectively.

Fig. 1. Prototype smart wheelchair system integrating the the flash LIDAR (encircled white).

B. Object Segmentation

Fundamental to our approach is the ability to automatically segment objects of interest (*i.e.*, persons vs. non-persons) from the background scene. The first step towards accomplishing this was to establish an estimate for the relative orientation of the local ground plane. This served as a reference elevation when segmenting objects of interest. Taking inspiration from [11], we employed an iterative re-weighted least squares (IRLS) approach.

The strength of the IRLS formulation when compared to more traditional ground plane tracking approaches (*e.g.*, RANSAC [12]) is that it integrates all *a priori* knowledge of the ground plane orientation through temporal filtering and regularization. To illustrate this, let

$$\Pi_{k-1} = a_{k-1}x + b_{k-1}y + c_{k-1}z + d_{k-1} = 0 \qquad (1)$$

denote our estimate for the ground plane at time $k-1$, and $P_k \in \mathbb{R}^{3\times m}$ denote the m points recovered from the LIDAR at time k. We assume that the ground plane orientation changes with time, but slowly when compared to the scan rate of the LIDAR. In other words, if a point $\vec{p}_k = (x_k, y_k, z_k) \in P_k$ were on the ground plane at time k, then its distance from Π_{k-1} should be small in practice. IRLS exploits this constraint by solving a problem of the form

$$\min_{a,b,c,d} \sum_{i=1}^{m} W(\vec{p}_{k_i}, \Pi_{k-1})(ax_{k_i} + by_{k_i} + cz_{k_i} + d)^2 \qquad (2)$$

where $\vec{p}_{k_i}$ denotes the i^{th} of the m points from the LIDAR scan at time k, and $W : \mathbb{R}^3 \times \Pi \rightarrow \mathbb{R}$ is a weight function based upon the normal distance of a point in the LIDAR scan to the estimated ground plane. In this work, W was a logistic function of the form

$$W(x) = \frac{1-A}{1+B*e^{(Cx-D)}} + A \qquad (3)$$

We found empirically that values of $A = 0.02, B = 0.9, C = 15$, and $D = 0.1$ worked well in practice. The resulting weight function is illustrated at Figure 2.

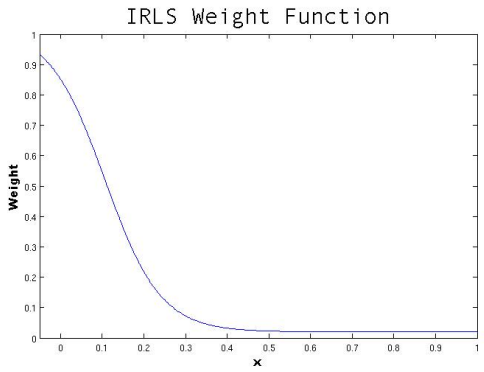

Fig. 2. Weighting function used for the IRLS algorithm.

In practice, the IRLS algorithm input was constrained to used LIDAR points within a range of 5 meters. To mitigate the effect of a greater density of points being returned from closer ranges, the input was binned into 10cm x 10cm cells in the LIDAR's x-z (horizontal) plane, and the point with minimum y (vertical) value was used. The choice of minimum was done to further reduce the impact of large vertical obstacles (*e.g.*, walls). Finally, a set of seeds corresponding to points on the default ground plane was also added to (2) to act as a regularization component. Representative results from this process are illustrated at Figure 3, which shows the raw 3D scan from the IFM, along with the recovered ground plane. q

Once Π was estimated, all points within a distance threshold of 5 cm were eliminated from the scan. The remaining points within 10 meters of the LIDAR are then converted to a 64x50x100 voxel image where the x-z plane is relative to Π. The voxel image was then pre-processed using connected component labeling to identify the set $\mathcal{C} = \{C_1, \ldots, C_n\}$ of objects of interest [13]. The scan is now ready to be classified.

C. An Appearance-based Classifier

The appearance-based classifier assigns an object of interest to one of two classes: person $\mathcal{P}$ or non-person $\bar{\mathcal{P}}$.

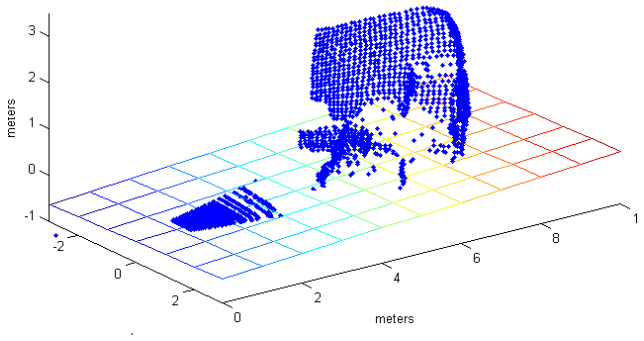

Fig. 3. Ground plane tracking with the IFM. The inclined ground plane is clearly tracked.

To accomplish this each connected component $C_k \in \mathcal{C}$ recovered from the segmentation approach described in Section III-B was first transformed into a compact six-dimensional feature vector: $\mathbf{x}_k = [h_k, w_k, d_k, v_k, \lambda_k, \rho_k]^T$. The first four elements correspond to the maximum y-value (height), width, depth, and volume of the bounding box of C_k, respectively. λ denotes the percentage of pixels above mid-height, to model the tendency of people to be top heavy. Finally, ρ represents the density of the x-y projection of C_k, which is defined as the number of pixels in the projection divided by the total number of pixels in the associated bounding box. This process is illustrated at Figure 4.

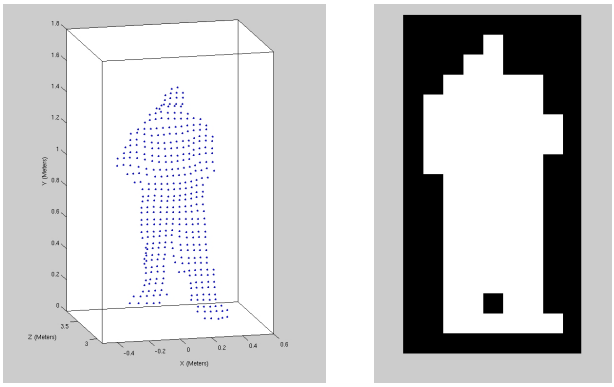

Fig. 4. Three-dimensional point cloud (left) and its associated x-y projection in the voxelized image. Object density ρ is defined as the number of pixels in the projection divided by the number of pixels in its bounding box.

Our objective in this work is the discovery of observations that are elements of the person class $\mathcal{P}$. Note that $\mathcal{P}$ is a multi-modal distribution: persons can be viewed from a variety of orientations, can exhibit different degrees of motion, can be partially occluded, certain clothes can affect sensor performance, *etc.*. As such, a simple thresholding is not sufficiently robust enough for our needs. We chose instead to investigate a k-Nearest Neighbors (k-NN) classifier [14]. k-NN compares an observed vector $\mathbf{x}'$ to a set of labeled prototype vectors $\mathbf{X} = \{\mathbf{x_1}, \ldots, \mathbf{x_n}\}$ that are known *a priori*, and assigns $\mathbf{x}'$ to the class that occurs most frequently among its k closest neighbors in $\mathcal{X}$.

k-NN requires supervised learning, which in the context of our problem amounted to establishing the $\mathcal{P}$ and $\bar{\mathcal{P}}$ classes. To this end, a total of 200 persons were imaged by the LIDAR from a range of distances and relative orientations while driving the wheelchair through South Bethlehem. Additionally, a total of 350 non-person prototypes were also imaged to represent our non-person class. Note that $|\bar{\mathcal{P}}| > |\mathcal{P}|$ to reflect the greater diversity of non-person objects. Each image was then synthesized to a 6-D feature vector as outlined above.

k-NN classifiers require a distance metric for defining "nearness." We investigated both Minkowski and Mahalanobis distance metrics. For the latter, we examined employing the covariance matrix of the separate classes, the pooled classes, and combinations thereof. Ultimately, we chose a Mahalanobis metric

$$d(\mathbf{x} - \mathbf{x}')^2 = (\mathbf{x}' - \mathbf{x_i})^T \Sigma^{-1} (\mathbf{x}' - \mathbf{x}) \tag{4}$$

where $\Sigma = (\Sigma_{\mathcal{P}} + \Sigma_{\bar{\mathcal{P}}})/2$ was the average covariance of the two classes. This modeled equal contributions from each of the two classes, and provided the best empirical results of the alternatives that were considered.

Traditional k-NN classifiers assigns a unit vote to each prototype in the k nearest neighbor set. We instead employed a weighted voting scheme where the each of the k nearest prototypes received a number of votes equal to the squared distance to the test sample, *i.e.*,

$$W_{\mathcal{P}} = \sum_{\mathbf{x} \in \mathcal{P}} d(\mathbf{x} - \mathbf{x}')^2, \qquad W_{\bar{\mathcal{P}}} = \sum_{\mathbf{x} \in \bar{\mathcal{P}}} d(\mathbf{x} - \mathbf{x}')^2 \tag{5}$$

The test point $\mathbf{x}'$ was then assigned to the class with the greatest weight. The ratio of these weights $Q(\mathbf{x}') = W_{\mathcal{P}}/(W_{\mathcal{P}} + W_{\bar{\mathcal{P}}}) \in [0, 1]$ was also used as a quality metric, and provided a confidence measure with respect to class assignment. $Q = 1$ indicates that all k neighbors $\in \mathcal{P}$, while $Q = 0$ is indicative of all k neighbors being from $\bar{\mathcal{P}}$. The use of Q is discussed in more detail in Section III-D.

D. A Motion-based Classifier (MC)

Initial testing with our appearance-based classifier indicated it was capable of identifying people with a high probability. However, the frequencies of both false negatives and false positives were not insignificant. In an attempt to mitigate these, a second-stage motion-based classifier (MC) was added to the people detection process. As input, the MC takes the set of objects $\mathbf{X}' = \{\mathbf{x'_1}, \ldots, \mathbf{x'_j}\}$ output by the appearance classifier, along with their associated confidence scores $\mathbf{Q} = \{Q(\mathbf{x'_1}), \ldots, Q(\mathbf{x'_j})\}$.

The objective of the MC is to associate tracks with candidate persons. In explaining its operation, let us first consider the case at time k where no tracks are yet established. For each $\mathbf{x}'(k) \in \mathbf{X}'(k)$, if $Q(\mathbf{x}'(k)) \geq Q_{\mathcal{P}} \implies \mathbf{x}'(k) \in \mathcal{P}$, and the MC immediately associates a track $\mathbf{t}(k) \in \mathbf{T}$ with the person $\mathbf{x}(k)$. For $Q(\mathbf{x}'(k)) < Q_{\mathcal{P}} \implies \mathbf{x}'(k) \in \bar{\mathcal{P}}$, and the non-person object is subsequently ignored.

Time $k+1$ begins with a data association phase. If a track $\mathbf{t}(k) \in \mathbf{T}$ established at time k cannot be associated with a person $\mathbf{x}'(k+1) \in \mathcal{P}$ based upon a maximum distance threshold, the track is deleted. If a successful data association is made, an Extended Kalman Filter is established to facilitate future tracking and $\mathbf{t}(k)$ is moved to $\mathbf{T}_{EKF}$. Each track $\mathbf{t} \in \mathbf{T}_{EKF}$ is parameterized by a state vector $[x, y, \theta, v]^T$ corresponding to the position, bearing, and velocity of the person relative to the wheelchair. Note this implies that a track can only persist in $\mathbf{T}$ for a single step before being either promoted to $\mathbf{T}_{EKF}$ or deleted. Lastly, if no association was made for a person $\mathbf{x'}_j(k+1)$, a new track $\mathbf{t}_j(k+1)$ is added to $\mathbf{T}$.

The more interesting case occurs at time step $k+2$. First, all tracks in $\mathbf{T}_{EKF}$ are propagated in accordance with the EKF time update equations, *i.e.*,

$$\begin{array}{rcl} x_{k+1} & = & x_k + v_k \cos\theta_k \Delta t \\ y_{k+1} & = & y_k + v_k \sin\theta_k \Delta t \\ \theta_{k+1} & = & \theta_k \\ v_{k+1} & = & v_k \end{array} \tag{6}$$

A data association phase is then made using these updated track positions with persons in $\mathbf{X}'(k+2)$ again based upon a distance threshold. If this data association fails, the track is *not* immediately deleted. Instead, we employ a lower hysteresis threshold Q_{min} for established tracks, and attempt to associate with every $\mathbf{x}'(k+2) \in \bar{\mathcal{P}}$, but where $Q(\mathbf{x}'(k+2)) > Q_{min}$. The motivation for this is when persons become partially occluded, or are on the edge of the LIDAR's field-of-view, the appearance classifier will often identify these as non-persons. However, established tracks provide additional confidence to the motion classifier, and as a consequence it will accept a lower Q value as a result.

If still no association is made for a track $\mathbf{t} \in \mathbf{T}_{EKF}$, the track is deleted. Otherwise, once an association is made the measurement update phase of the respective EKF is run where

$$\mathbf{z_k} = [x_k, y_k, \theta_k, v_k]^T \tag{7}$$

We should note that (7) is a simplification of the actual measurement process. While the relative x-y positions of a person were estimated directly using the centroid of the bounding box from the appearance classifier, θ_k and v were estimated using a finite difference approach from successive position estimates across two time steps. However, this simplification worked well in practice. Once all tracks in $\mathbf{T}_{EKF}$ are processed, any tracks in $\mathbf{T}$ and persons in $\mathbf{X}'(k+2)$ are processed as in the previous time step.

To summarize the motion classifier update phase:

1) Any tracks in $\mathbf{T}_{EKF}$ are processed first. They are initially associated with persons in $\mathcal{P}$ as identified by the appearance classifier, and if necessary with non-persons $\mathbf{x}' \in \bar{\mathcal{P}}$ but where $Q(\mathbf{x}') > Q_{min}$. If either association succeeds, the respective EKF is updated. Otherwise, the track is deleted.
2) Any tracks in $\mathbf{T}$ are processed second. They are associated with any persons remaining in $\mathcal{P}$. If this is successful, the respective track is moved to $\mathbf{T}_{EKF}$, and the EKF is updated. Otherwise, the track is deleted.
3) Any persons remaining in $\mathcal{P}$ that were not associated with tracks in $\mathbf{T}_{EKF}$ or $\mathbf{T}$ are added to $\mathbf{T}$.

IV. EXPERIMENTAL RESULTS

All Experiments were conducted on the Lehigh University campus. For ground-truth data, both LIDAR frames and video images were simultaneously logged and timestamped for manual post-processing. The LIDAR exposures were set to 150 and 2000, which empirically provided the best performance outdoors. This resulted in a frame rate of approximately 8 Hz. Details now follow.

A. Appearance Classifier

Performance of the appearance classifier was characterized using single LIDAR frames taken while manually driving the wheelchair across Lehigh's campus. Samples in the test set were disjoint from the training set used by the k-NN classifier. The goal of the appearance classifier was to minimize the number of false positives. False negatives have the potential to be overridden by the motion classifier, as discussed in Section IV-B.

A total of 105 images of persons and 464 non-persons were acquired. Classification results with $k = 5$ and $Q_{\mathcal{P}} = 0.5$ are presented in the confusion matrix below. Of the 105

TABLE I
CONFUSION MATRIX FOR k-NN APPEARANCE CLASSIFIER

		Decided	
		nonperson	person
Actual	nonperson	445	19
	person	8	97

persons imaged, 97 were classified as persons, for a detection rate of 92.4%. Of the 464 non-person objects imaged, 19 were classified as persons for a false-positive rate of 4.1%.

In truth, the true-positive rate was lower than hoped. Upon reviewing misclassifications, several pathological cases were identified. One we labeled the "floating torso" problem. Due to the relatively low power of the IFM LIDAR, detection of darker objects becomes problematic as range increases. In particular, we found when pedestrians were wearing dark pants, the range which they could be detected dropped dramatically. The net result was the appearance of a floating torso in space. This is illustrated in Figure 5. The frequency of these occurrences lead us to include several of these as prototypes in the training set. Still, misclassification of these instances were not uncommon. We should note that if the person continues to approach the wheelchair and a more complete LIDAR return is received, a correct classification results. We also note the same effect results as a person walks away from the LIDAR. In this event, the motion classifier is useful in maintaining a correct classification.

B. Appearance + Motion-based Classifier

The second phase of testing involved evaluating the full-classifier in continuous, dynamic operations. Our preliminary

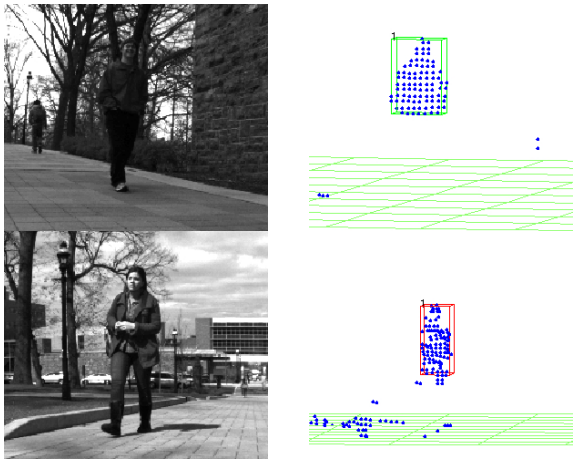

Fig. 5. Floating torso distributions. The top example was correctly classified as a person, while the more fragmented bottom example was not.

testing involved an approximately 5 minute run of driving the wheelchair across Lehigh's Campus during normal hours to ensure sufficient pedestrian traffic. Pedestrians were only counted once, meaning each entry was a unique person. Classification was determined successful if an EKF track was established on the pedestrian. For these tests, $k = 5$ and the Q parameters for the MC were set to $Q_{\mathcal{P}} = 0.5$ and $Q_{min} = 0$. The latter meant that if at least 1 of the 5 neighbors believed the object was a person and an EKF track was established, the MC could override the decision of the appearance classifier. Results are provide in the confusion matrix below. From these results, we see that 45

TABLE II

CONFUSION MATRIX FOR COMPLETE CLASSIFIER

		Decided	
		nonperson	person
Actual	nonperson	50	7
	person	2	45

of 47 persons were classified correctly, for a detection rate of 95.7%. We examined the two failure cases. Both were instances where the pedestrian never completely entered the field of view of the LIDAR. An example of this is shown at Figure 6, where the bottom half of a pedestrian appears on the edge of the LIDAR's field-of-view. If these cases are excluded, the true positive detection rate would have been 100%.

We also noted during this experiment a false positive rate of over 12%. The increase from the static test can be attributed to the fact that during driving operations, a single non-person object might be imaged 10s of times from a range of distances and orientations. If it was incorrectly classified as a person for a single frame, it was considered a false positive. In practice, associating conditional probabilities with such cases would likely prove useful.

A strength of the MC for maintaining tracks is illustrated by Figure 7. In this example, there are two persons in $\mathbf{T}_{EKF}$, meaning that EKF tracks have been established. The person on the right side begins to migrate out of the LIDAR's field of view to a point where the appearance classifier no longer associates it with the person class. Nevertheless, the MC is able to maintain a track until the person is almost entirely out of the LIDAR's view.

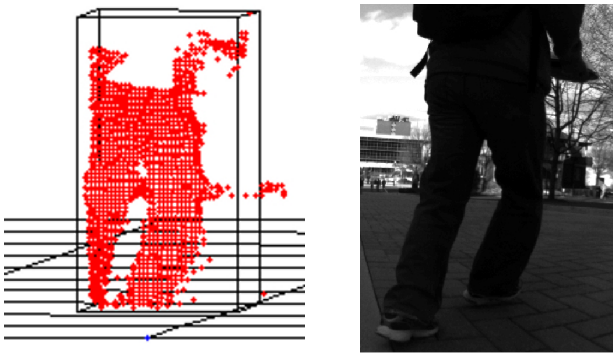

Fig. 6. A situation where the appearance classifier did not get a full view of the pedestrian, deterring tracking and classification

V. DISCUSSION

In this work, we demonstrated the potential for low-cost 3D LIDARs to be applied to a people tracking task. While providing significantly lower angular resolution than their more expensive cousins, these systems still provide the accurate distance estimates and illumination invariance associated with LIDAR systems. Our preliminary results to date indicated a successful tracking rate of over 95% during dynamic operations. Nevertheless, there is significant room for improvement.

In terms of scalability, the proposed approach should still be suitable for LIDARs of higher resolution. As the prototypes in the training set are reduced to 6D feature vectors, we expect the most computationally expensive component would be generating the 3D connected components. On this note, we should also admit that the appearance classifier is a work in progress, and there are potentially alternative/additional features that could improve performance. In addition to the 6 used in these experiments, we also examined height/width ratio, distribution eigenvector orientation, and skew. None of these had a positive effect on performance.

To this point, we have also only examined only single values for the parameters $Q_{\mathcal{P}}$ and Q_{min}. A sensitivity analysis to tune these parameters may in fact improve performance.

Perhaps the greatest limitation with the IFM LIDAR system that was used is its relatively coarse angular resolution. This made it difficult to segment pedestrians that were walking in close proximity to one another. An example is shown at Figure 8. In this instance, all three persons were lumped into a single cluster where from the ground-truth camera image they are obviously disconnected. One potential means for handling these cases would be to refine larger

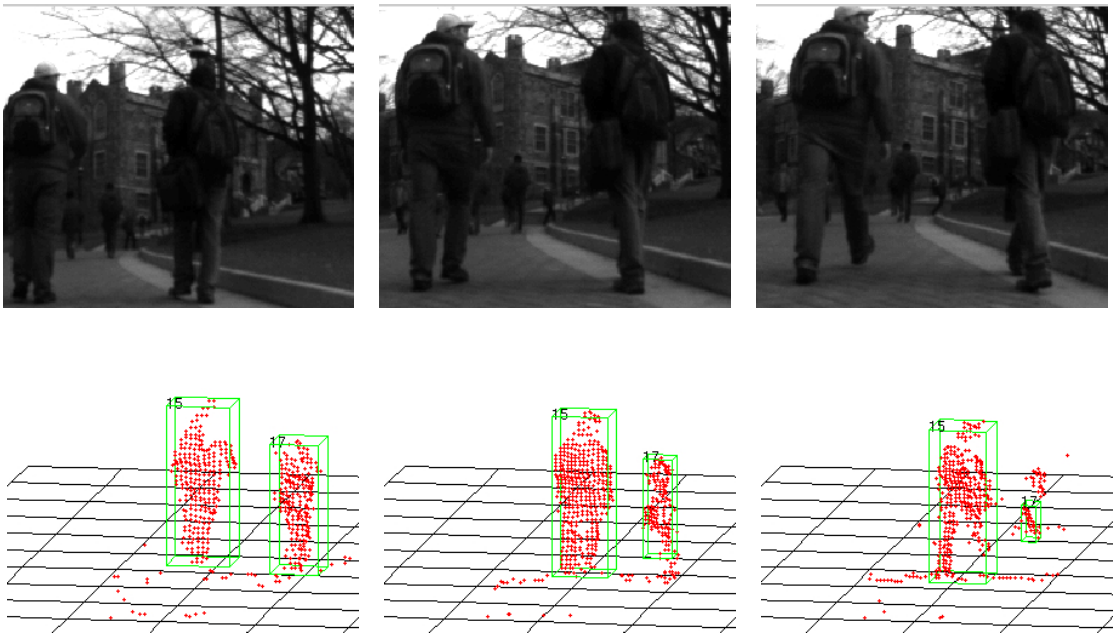

Fig. 7. A sequence of frames, showing the tracking of a person even after they fail the appearance classifier.

connected components using a k-means approach, where k could be correlated to the geometry of the bounding box. More work is required in this area.

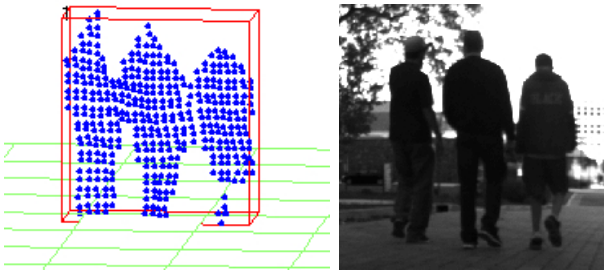

Fig. 8. Where the CCL was unable to seperate the objects of interest.

VI. Acknowledgments

The authors would like to thank Henry Baird for his insights regarding the k-NN classifier. This work was supported by the National Science Foundation (NSF) Robust Intelligence Program under Grant No. 0844585. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of NSF.

References

[1] J. Dietsch, "2010: When mobile robots reached the tipping point," *IEEE Robotics and Automation Magazine*, Dec 2010.

[2] Morgan Quigley, Ken Conley, Brian P. Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, and Andrew Y. Ng, "Ros: an open-source robot operating system," in *ICRA Workshop on Open Source Software*, 2009.

[3] Microsoft, "Xbox Kinect," `http://www.xbox.com/en-US/kinect`.

[4] M. Buehler, K. Iagnemma, and S. Singh, Eds., *Journal of Field Robotics: Special Issue on the 2007 DARPA Urban Challenge*, vol. 25, Wiley Periodicals, 2008.

[5] David Geronimo, Antonio M. Lopez, Angel D. Sappa, and Thorsten Graf, "Survey of pedestrian detection for advanced driver assistance systems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, pp. 1239–1258, 2010.

[6] L. Spinello, K. O. Arras, R. Triebel, and R. Siegwart, "A layered approach to people detection in 3d range data.," in *Proc. of The AAAI Conference on Artificial Intelligence: Physically Grounded AI Track (AAAI)*, 2010.

[7] Luis Ernesto Navarro-Serment, Christoph Mertz, Nicolas Vandapel, and Martial Hebert, "Ladar-based pedestrian detection and tracking," in *Proc. 1st. Workshop on Human Detection from Mobile Robot Platforms, IEEE ICRA 2008*. May 2008, IEEE.

[8] Luis Ernesto Navarro-Serment, Christoph Mertz, and Martial Hebert, "Pedestrian detection and tracking using three-dimensional ladar data," in *Proc. of The 7th Int. Conf. on Field and Service Robotics*, July 2009.

[9] D.V. Prokhorov, "Object recognition in 3d lidar data with recurrent neural network," in *Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on*, 2009, pp. 9 –15.

[10] C. Gao, I. Hoffman, T. Miller, T. Panzarella, and J. Spletzer, "Autonomous docking of a smart wheelchair for the automated transport and retrieval system (atrs)," *Journal of Field Robotics*, vol. 25, no. 4-5, pp. 203–222, 2008.

[11] Jon Bohren, Tully Foote, Jim Keller, Alex Kushleyev, Daniel Lee, Alex Stewart, Paul Vernaza, Jason Derenick, John Spletzer, and Brian Satterfield, "Little Ben: The Ben Franklin Racing Team's entry to the DARPA Urban Challenge," *Journal of Field Robotics*, vol. 25, no. 9, pp. 598–614, Sep 2008.

[12] M. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," in *Communications of the ACM*, 1981.

[13] Qingmao Hu, Guoyo Qian, and Wieslaw L. Nowinski, "Fast connected-component labelling in three-dimensional binary images based on iterative recursion," *Computer Vision and Image Understanding*, vol. 99, pp. 414–434, April 2005.

[14] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, John Wiley and Sons, 2001.