Silhouette Classification Using Pixel and Voxel Features for Improved Elder Monitoring in Dynamic Environments

Erik E. Stone
Department of Electrical and Computer Engineering
University of Missouri
Columbia, MO
ees6c6@mizzou.edu

Marjorie Skubic
Department of Electrical and Computer Engineering
University of Missouri
Columbia, MO
skubicm@missouri.edu

Abstract— **We present a method for improving human segmentation results in calibrated, multi-view environments using features derived from both pixel (image) and voxel (volume) space. The main focus of this work is to develop a low-cost, vision-based system for passive activity monitoring of older adults in the home, to capture early signs of illness and functional decline and allow seniors to live independently. Silhouettes are extracted to address privacy concerns. Specific embedded assessment goals include daily gait, fall risk, and overall activity, as well as fall detection. To achieve these goals, accurate, robust segmentation of human subjects (silhouette extraction) from captured video data is required. We present a simple technique that makes use of features acquired from background subtraction results (silhouettes) of multiple calibrated cameras, along with the 3D voxel object formed from the intersection of those multiple silhouettes in a volume space to improve human segmentation results in dynamic environments; moving objects, non-human objects, and lighting changes often complicate this task. The technique is qualitatively evaluated on three data sequences, two of which were captured in an independent living facility for older adults.**

Keywords – activity monitoring, eldercare, smart environments, silhouette extraction

I. Introduction

As older adults are living longer, research has focused on developing new technologies to allow them to continue living independently. These technologies have aimed to monitor a broad array of activities and biological signals, with the ultimate goal of detecting changes, both short and long term, in older adults' physical and/or cognitive function. Such detection mechanisms would facilitate medical interventions when needed by older adults, while allowing them to continue living in their preferred settings and reducing the strain on and need for expensive care facilities.

Recent work has focused on the use of passive infrared motion sensors in the home [1-3], for monitoring overall activity as well as walking speed. Such systems have shown promising results and addressed issues of privacy, but are somewhat limited by the coarseness of their measurements. Additionally, these systems alone are not able to quickly detect adverse events, such as falls, that may impact older adults; thus, requiring additional sensors for this task.

This work focuses on developing a low-cost, vision-based system for passive monitoring of older adults. A vision-based system could provide a more accurate, detailed assessment of an individual's daily and long term activity in the home, while also addressing the need for quick detection of adverse events, such as falls.

A major concern of vision-based monitoring systems is the need to maintain the privacy of those being monitored. Specifically, raw images cannot be stored. However, research has shown that video data anonymized through the use of silhouettes may address the privacy concerns of older adults to such video-based monitoring technologies [4]. This, along with automatic recognition algorithms, requires the robust segmentation of people from video data.

This paper describes a method for improving automatic segmentation of humans from video data in dynamic environments with multiple, calibrated, static cameras, from which 3D voxel objects can be created, without the use of a high level tracking algorithm. Section II of this paper looks at related work. Section III gives a description of our video based monitoring system. Section IV discusses our method for improved human segmentation in calibrated, multi-view environments. Section V contains qualitative results of our method on three data sequences. Finally, Section VI contains concluding remarks and a discussion of future work.

II. Related Work

The construction of 3D voxel objects from multiple silhouettes, often referred to as shape from silhouettes, has been widely investigated [5-8]. The main application of such techniques has, for the most part, been in the field of markerless human motion analysis, with the goal of fitting articulated humanoid models of various complexities to the resulting 3D voxel objects [9]. Such systems have generally used three or more cameras in a constrained environment in order to obtain accurate results.

The separate issue of acquiring silhouette images from video data has also been widely investigated, with the main approach being background subtraction [11]. Background subtraction techniques include mixture of Gaussians [12], Eigen-backgrounds [13], and Wallflower [14]. Such techniques aim to detect pixels or image blocks which vary

978-1-61284-936-2/11/$26.00 ©2011 IEEE

from a background model of the scene. Modifications have been introduced to handle shadows and other artifacts.

By themselves, background subtraction techniques cannot infer the type of object (human, lighting, non-human, etc.) responsible for the resulting foreground segmentation. A higher-level of reasoning is required. Simple connected component based size and number thresholds, along with shape features, are generally insufficient given significant changes in the range and pose of humans and objects from the camera, or multiple people in the scene. Often, simple to sophisticated tracking algorithms which attempt to model human motion are employed to deal with this problem [10,15-18,22]. In some cases, tracking ability is limited to humans only in certain poses. Such systems have used both single and multiple cameras to attempt to track people. Additionally, work has been done in using vision-based tracking systems (single and multi-camera) for fall detection [19-21, 24] and good results have been reported.

III. System Overview

Our system, shown in Figure 1, consists of two inexpensive web cameras, capturing 640x480 images, which are used to monitor the environment (room). The cameras are positioned to be roughly orthogonal, and the intrinsic and extrinsic calibration parameters of the cameras are estimated *a priori*. In order to preserve privacy, silhouettes are extracted from the captured video data using a background subtraction technique which fuses color and texture features, as described in [19].

Given extracted, corresponding silhouettes from each of the cameras, a 3D object, formed from the intersection of the projection of the two silhouettes into volume (voxel) space, can be constructed. For our work, the voxel space is discretized into 1x1x1 inch cubic elements and rooms as large as 30 ft. by 17 ft. by 8 ft. have been used for testing.

Prior work has looked at the accuracy of fall detection, gait assessment, and body sway measurement algorithms using 3D objects constructed using our system, in relatively controlled environments, and excellent results have been achieved [24-26]. These results include comparisons of walking speed, left/right stride length, and left/right stride time with a GAITRite electronic mat, and body sway measurements with a Vicon marker based motion capture system. The algorithms have not tried to fit articulated, high degree-of-freedom human models to the 3D objects, in an effort to keep the computational complexity low. Instead, the algorithms are used to extract information directly from the 3D voxel objects with minimal processing.

Previous efforts have not generally addressed the issue of dynamic, un-controlled environments due to the challenges involved in robustly differentiating human silhouettes/3D objects from those of lighting changes, moving non-human objects, un-identified shadows, and other noise sources. Ultimately, operation in such conditions will be required for a practical system that can be used in a dynamic home setting.

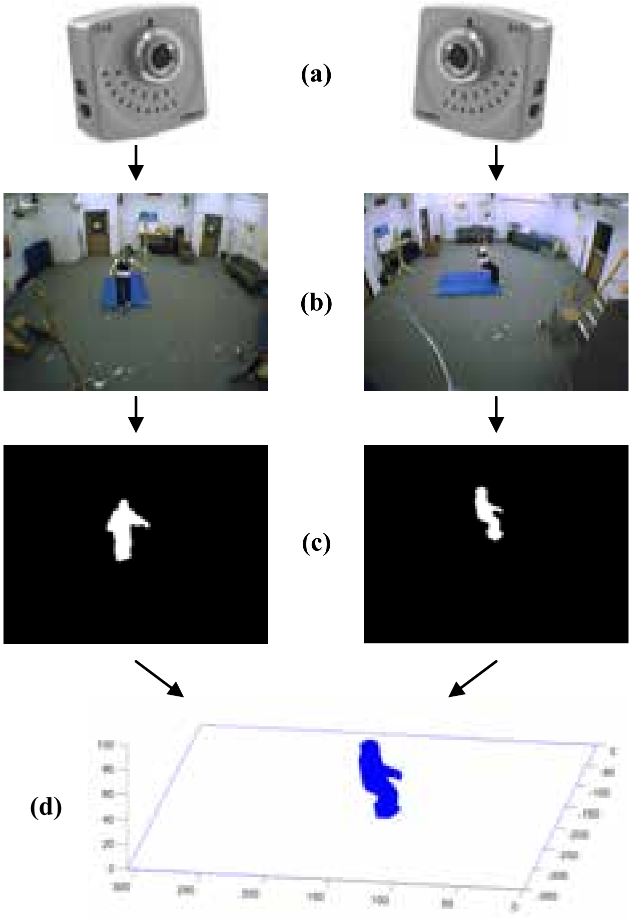


Fig. 1. System overview. (a) Two cameras positioned orthogonally in the environment. (b) Camera views of the same scene. (c) Extracted silhouette images. (d) Three dimensional voxel object formed by the intersection of silhouette projections in voxel space.

IV. Methodology

In previous work validating the accuracy of our system for extracting gait parameter and other activity parameters, it was assumed that good silhouettes were already available. Specifically, the 3D voxel object used for analysis was taken to be the intersection resulting from the projection of the largest connected component from each of the silhouette images into voxel space. Given controlled settings, this assumption works well. However, in dynamic, uncontrolled environments, lighting changes, moved objects, and multiple people can result in many connected components in the extracted silhouette images. Furthermore, the size of each of these resulting connected components in image space is dependent on the position of the camera with respect to the object. Thus, a more detailed analysis is necessary.

In order to achieve the goal of a real-time, low-cost vision system, we would like to avoid the use of computationally intensive tracking algorithms (mentioned in Section II) which attempt to model human motion, or at least have such algorithms be independent of the silhouette extraction process. Therefore, we have adopted a simple scheme that fuses image and voxel based features (leveraging the 3D ability of our system) into the update procedure of our background subtraction model for each camera in order to achieve more robust human segmentations that are independent of high leveling tracking algorithms. Such tracking algorithms could operate on the extracted silhouettes and voxel objects at a later time, in a non-real-time fashion. The steps of our approach are outlined below.

A. Connected Component Identification

The first step of our technique is to identify and label the individual connected components, c_j^i, $1 \leq i \leq C_j$, $1 \leq j \leq J$, in the extracted silhouette images. Let C*j* denote the number of connected components in silhouette image *j*, and let *J* denote the number of silhouette images (cameras). Given the number of connected components in each silhouette image, we can then compute the number of 3D objects, N_o, formed by the intersection of the projection of a single connected component from each of the silhouette images as:

$$N_o = \prod_{j=1}^{J} C_j$$

For each of the N_o voxel objects, some (if not many) of which may have an empty voxel space intersection, a set of features will be used to classify the voxel object as either human or non-human.

B. Feature 1 – Connected Component Usage

The first feature, connected component usage (*CCU*) illustrated in Figure 2, is used to identify the percentage of each connected component associated with a 3D voxel object that forms the voxel space intersection for the object. Thus, each voxel space object will have a *CCU* value for each silhouette image. Let us denote the connected component from silhouette image *j* used in the construction of object *k* as $c_j^{i_k}$.

The *CCU* value for object *k*, silhouette image *j*, is obtained by first back-projecting the 3D object, o_k, into each camera view, evaluating the cardinality of the intersection of the back-projection, $b_{j,k}$, and the original connected component, and, finally, normalizing by the cardinality of the original component as shown below:

$$CCU_{k,j} = \frac{\left| b_{j,k} \cap c_j^{i_k} \right|}{\left| c_j^{i_k} \right|}$$

For objects that are well-segmented in each camera view, the *CCU* values associated with the corresponding voxel object should be near 1.0. For objects that are not well-segmented in each of the camera views, the *CCU* value from one view may be high, while the other lower (as per the example in Figure 2). For voxel objects that are the result of intersecting the segmentation of different real world objects from each silhouette image, the *CCU* values for all silhouette images should be low.

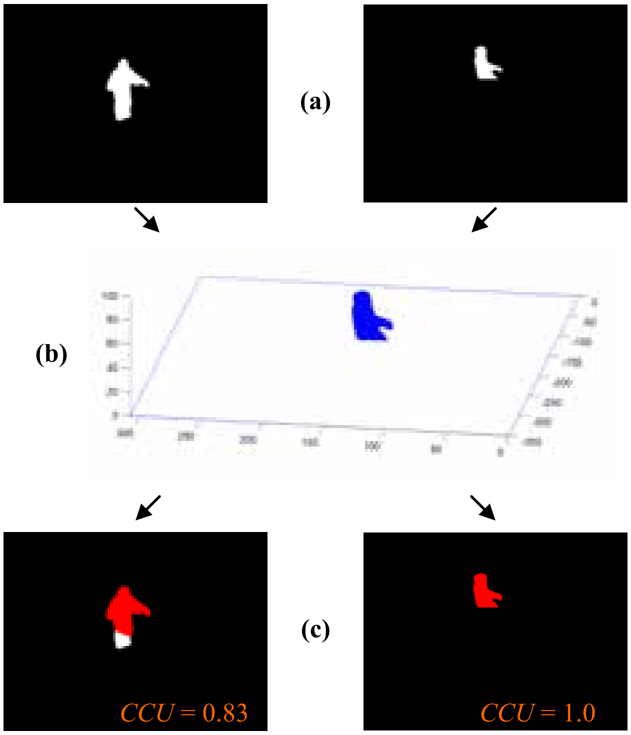


Fig. 2. Illustration of the *CCU* feature. (a) Connected components from silhouette images. Person is well segmented in left silhouette image, while part of the legs are missing in the right silhouette image. (b) Voxel space intersection for voxel object formed by projection of connected components. (c) Back-projection (red) of voxel object intersection overlaid on original connected components, with computed *CCU* values.

C. Feature II – Position Corrected Volume

The second feature, position corrected volume (*PCV*) uses the view vector of each connected component associated with a voxel object to adjust the volume of the object based on its position with respect to the cameras.

Given the calibrated 3D view vector of each pixel from each camera, relative to the global coordinate system, the view vector of connected component c_j^i, denoted v_j^i, is taken to be the average of the view vectors of each pixel in the connected component. Assuming *J* cameras, the *PCV* for object *k*, is calculated as:

$$PCV_k = \eta_k * volume(o_k)$$

$$\eta_k = \max_{\substack{j,n \in J \\ n \neq j}} \left[\left(1 - \left| \frac{v_j^{i_k} \bullet v_n^{i_k}}{\left\| v_j^{i_k} \right\| \left\| v_n^{i_k} \right\|} \right| \right)^{\alpha} \right]$$

The *PCV* is a correction to the raw voxel space volume based on the orthogonality of the view vectors of the connected components used to form the object. If the view vectors are perfectly orthogonal then the volume of the object is left unchanged. However, as the view vectors become more parallel the volume is corrected (reduced) to counteract the effects of construction error, as detailed in [27]. This adjustment results in more stable volume measurements as objects move about the environment. Given more than two cameras, the *PCV* is based on the pair of cameras whose connected component view vectors are the closest to orthogonal. (We have used $\alpha = 0.3$).

D. Classification – Human vs. Non-human

A set of heuristic rules is used to classify each voxel object as human or non-human, based on the image space and voxel space features described previously. This technique is not meant to completely solve the problem of identifying humans in multi-view environments, but to act as a filtering step in a silhouette extraction process that is independent of any high-level tracking or human motion modeling. Furthermore, we would like the system to classify voxel objects which consist of two connected humans, or humans in contact with small to medium sized objects as human for the purpose of segmentation. As a result, the rules are defined to err on the side of classifying objects as human, allowing a higher level of reasoning to later determine cases which are not clear.

Rules for Voxel Object Classification

for each voxel object, $1 \leq k \leq N_o$

Rule			
1: If	$PCV_k < 4,000\ in^3$		**Then** *non-human*
2: If	$PCV_k > 80,000\ in^3$		**Then** *non-human*
3: If	$\min_{j \in J} \left(CCU_{k,j} \right) < 0.15$		**Then** *non-human*
4: If	$\max_{j \in J} \left(CCU_{k,j} \right) < 0.8$		**Then** *non-human*
5: If	$\max_{\substack{j,n \in J \\ n \neq j}} \left(CCU_{k,j} + CCU_{k,n} \right) < 1.1$		**Then** *non-human*
6: If	*!(R1 | R2 | R3 | R4 | R5)*		**Then** *human*

Rule 1 acts to remove voxel objects which are too small to be human based on the object's computed *PCV*. This takes care of small moving objects such as books, magazines, and cups. As it is based on the *PCV* of the voxel intersection, (assuming a reasonable segmentation) it will not be dependent on the position of the object with respect to the camera. In addition, rule 1 also handles empty voxel objects, which are often the result of attempting to intersect connected components that don't correspond to the same real world object. Rule 2 acts to remove voxel objects which are too large to be human based on the computed *PCV*. This takes care of segmentations resulting from quick, large scene changes, such as lighting, and large moving objects.

Rules 3, 4, and 5, in general, act to remove voxel objects that are not empty, but are the result of intersecting connected components from different real world objects. If an object is reasonably segmented in each camera view, even with significant (50%) occlusion in one view, one of the *CCU* values should be near 1.0 (i.e., > 0.8), and the others, though lower, should not be near zero (i.e., > 0.15).

If any of the first five rules fire, then the voxel object is classified as non-human. If none of the first five rules fire, then the object is classified as human, as stated in rule 6.

Lastly, after the initial classification step, an attempt is made to determine if any non-human voxel objects are actually detached parts of human voxel objects. This determination is based on position and distance from the human voxel object, and whether the non-human voxel object is formed using one of the connected components used to form the human voxel object. If it is determined that a non-human voxel object is a detached part of a human, then the connected components of the two objects for each silhouette image are combined, as well as the voxel space intersections.

E. Background Model Updating

The last step of the technique integrates the human vs. non-human classification into the update procedure of the background model for each camera. The base level of the update procedure uses frame-to-frame motion estimated with overlapped blocks. Specifically, any blocks which are determined to contain motion are prevented from updating.

Next, for each silhouette image, connected components that are associated with a human classified object are treated as follows: first, the convex hull of the connected component in image space is computed; second, the convex hull is dilated using a 5x5 kernel; third, all of the pixels in the dilated convex hull are prevented from updating; thus, connected components associated with human classified voxel objects will persist indefinitely even if not moving.

Pixels which are classified as background in the silhouette extraction step, and are not blocked from updating, are updated using a slow rate, $\delta \approx \frac{1}{600(fps)}$ (where frames per second ≈ 5). Pixels, along with their neighbors, which belong to a connected component not associated with a human classified object, and are not prevented from updating, are updated using a faster rate, $\delta \approx \frac{1}{(fps)}$. This faster update rate acts to quickly absorb these foreground segmentations into the background model.

Finally, because connected components associated with human classified voxel objects will persist indefinitely, and because our classification method is not 100% accurate, there must a procedure for getting rid of segmentations due to incorrect classifications or they could eventually clutter the images. Although the previously described classification rules would eventually take care of such cases (when other segmentations interact with the incorrect ones, or when lighting changes occur), we do include an override that is thrown if the percentage of foreground pixels is greater than a predetermined threshold. When this occurs, all of the pixels in the image are updated at the fast rate, except those in motion blocks. (We have used a threshold of 40%, although this should be adjusted based on the camera field of view and placement.)

V. Experimental Results

We have included results of our algorithm on three data sequences (fps = 5), two of which were captured in an independent living facility for older adults and the third in a lab setting. Evaluation of the technique is difficult from two standpoints: 1) how to quantify the results without hand segmenting humans in the video frames, which is prohibitively time intensive; 2) determining what other algorithm(s) (and, thus, with what parameter settings) to compare it against. Ultimately, we have decided to limit the scope of the results presented here to qualitative evaluation, and to simply show our results against our baseline background subtraction with no update procedure. Future efforts will provide a more comprehensive analysis.

A. Sequence 1

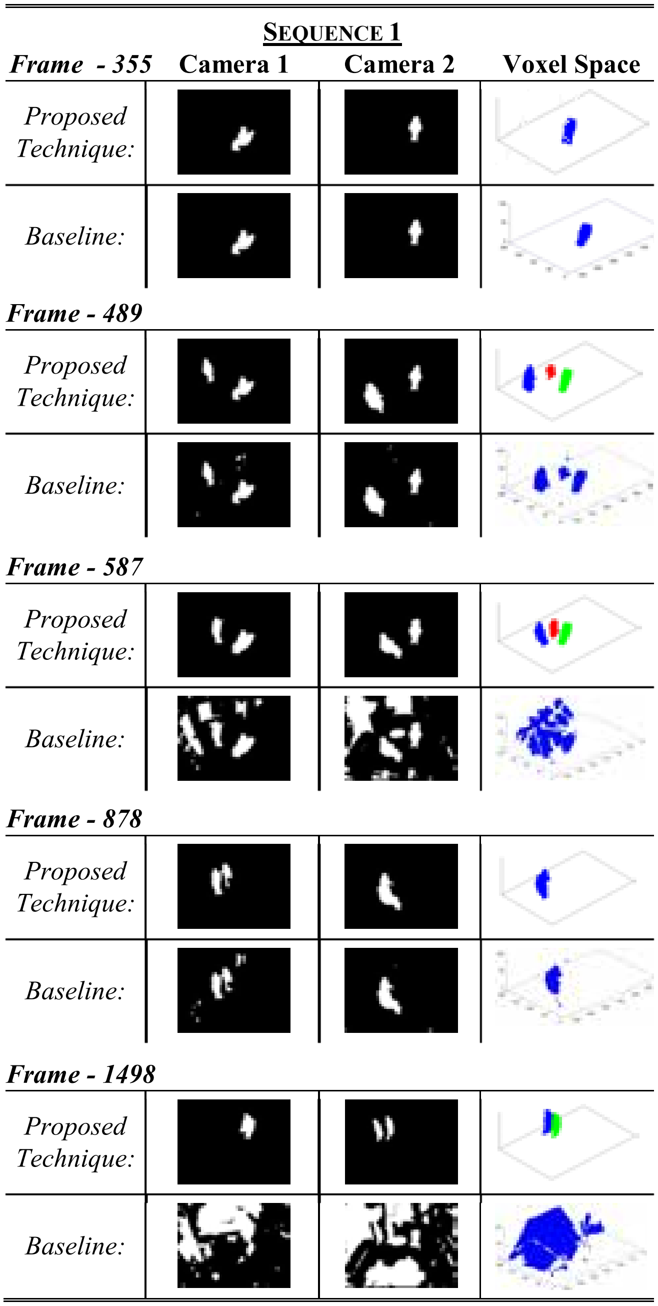
The first sequence, captured in an independent living facility for older adults called TigerPlace, consists of two adults acting out a scripted scenario. Person one walks in and sits down (frame 355). Person two enters the room (frame 489), hugs person one, and sits down. As person two is sitting down, a relatively quick (although not instantaneous) lighting change occurs due to the sun through an outside glass door (frame 587). Subsequently, person one gets up to make tea for person two, hands person two the tea (frame 878), then sits back down. During this time, the lighting continues to fluctuate. Finally, they both exit the room with person one picking up a jacket (frame 1498). As the silhouette images from sequence 1 illustrate, our technique is able to update the effects of the lighting change, which greatly impacts the baseline system, while still segmenting the people, resulting in accurate silhouette images. However, when person one and person two are sitting in close proximity, one of the voxel space objects formed by the intersection of the non-corresponding connected components from each silhouette image is classified as human. Therefore, although the individual silhouette images are correct, three separate objects are shown in the voxel space representation during this time.

B. Sequence 2

The second sequence, also captured in TigerPlace, consists of two adults acting out a different scripted scenario. Person one enters the room and sits, and in the process places a jacket on a cabinet, and picks up a magazine from a table (frame 555). Person one then gets up to open the door for person two before sitting back down, moving the magazine in the process (frame 985). Person two proceeds to move about the room (cleaning), until person one gets up from the chair (moving it slightly), places the magazine in a trash can (frame 1282), puts the jacket back on (frame 1469), and exits the room (frame 1581). Person two soon follows. As the illustrations of sequence 2 show, our technique is able to update the moved objects into the background quickly, and handle a small change in lighting while segmenting the people.

C. Sequence 3

The third sequence, captured in a lab setting, consists of one adult, and contains a number of environmental changes to stress the system. Initially, person one enters and sits down on a couch (frame 150). Person one then gets up; proceeds to move two small objects, move one medium

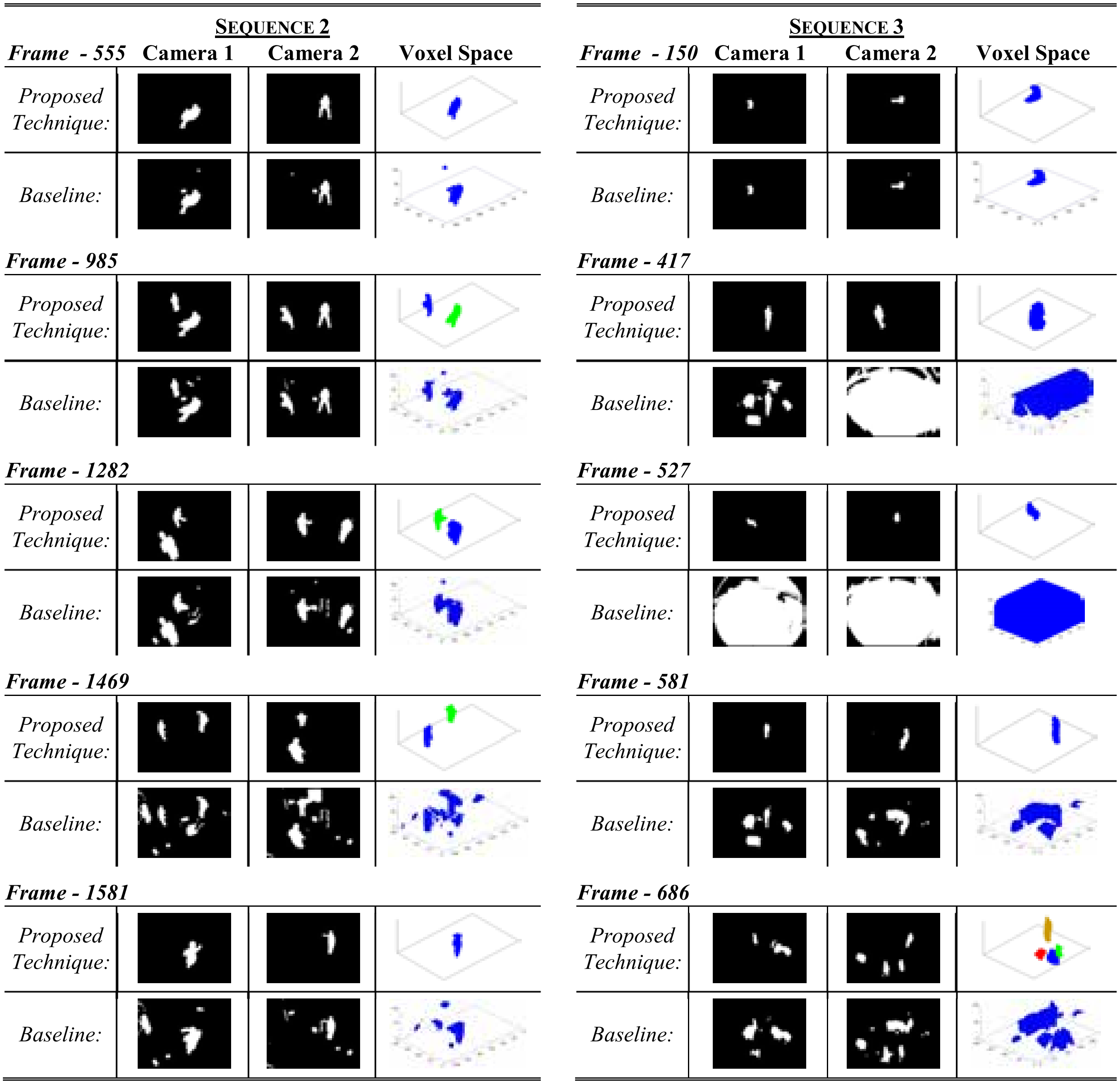


sized object, turn on a lamp, and move the couch slightly (frame 417). An overhead light is then turned off, person one walks towards the couch, the overhead light is turned back on, and person one sits down. While sitting, the overhead light is turned off for a short period of time (frame 527), and then back on. Person one then gets up, turns off the lamp (frame 581), moves two medium sized objects and one small object, and begins to exit the room (frame 686). As the illustrations of sequence 3 show, our technique is able to update the small and large moved objects into the background, and adapt well to the lighting changes due to the lamp and the overhead light, while still segmenting the person. However, movement of two medium sized objects does result in improper human segmentations.

Such improper human segmentation issues can occur when medium sized objects are moved. If they are classified as human (either where they are moved to, or where they are moved from), they may persist in the silhouette images until removed by subsequent events.

Finally, when a person is motion-less during a large lighting change, the person is updated into the background model, and their subsequent movement may result in their previous location being classified as human. In such cases, their previous location will persist in the foreground segmentation until a subsequent event eventually resolves the issue.

Although these conditions do result in temporary, improper human segmentations, they should be easily manageable by a tracking algorithm operating on the extracted silhouettes and voxel objects in a higher-level



reasoning component.

VI. Conclusion

In this paper, we presented a simple technique which fuses image and voxel features, from a calibrated, multi-view environment into the background model update procedure used to acquire silhouette images for the purpose of improved human segmentation. The technique is independent of any high level tracking, and results have shown it to be quite good at reducing artifacts due to lighting and moving non-human objects; which are often present in background subtraction results from dynamic, uncontrolled environments.

To improve the performance of the technique, future efforts will look to collect a set of training data, consisting of silhouette image components and voxel space intersections for both human and non-human objects in a variety of poses, settings, and occlusions, for the purpose of training a more robust classifier. Furthermore, the inclusion of additional features, such as distribution of mass, major orientation with respect to centroid height, etc., will be explored.

The benefit of incorporating a single frame, image feature based people detection and localization algorithm into the segmentation process will also be explored.

Acknowledgment

The authors would like to acknowledge the members of the Eldertech team at the University of Missouri for their help in collecting the data used in this work. This work has been supported by the U.S. National Science Foundation under Grants IIS-0703692 and CNS-0931607.

References

[1] TL. Hayes, M. Pavel, and JA. Kaye, “An unobtrusive in-home monitoring system for detection of key motor changes preceding cognitive decline,” In: *26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*; San Francisco, CA, 2004.

[2] S. Wang, M. Skubic, and Zhu Y, "Activity Density Map Dis-similarity Comparison for Eldercare Monitoring ," Proceedings, *31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Minneapolis, Minnesota, September 2-6, 2009, pp 7232-7235.

[3] S. Hagler, D. Austin, T. Hayes, J. Kaye, and M. Pavel, "Unobtrusive and Ubiquitous In-Home Monitoring: A Methodology for Continuous Assessment of Gait Velocity in Elders," *IEEE Trans Biomed Eng*, 2009.

[4] G. Demiris, O. D. Parker, J. Giger, M. Skubic, and M. Rantz, “Older adults' privacy considerations for vision based recognition methods of eldercare applications,” *Technology and Health Care*, vol. 17, pp. 41-48, 2009.

[5] W.N. Martin, J.K. Aggarwal, Volumetric descriptions of objects from multiple views. *IEEE Transactions on Pattern Analysis and Machine Intelligence*,5(2):150–158, March 1983.

[6] C. Dyer, Volumetric scene reconstruction from multiple views, Foundations of Image Understanding, pp. 469–489, Kluwer, 2001.

[7] G. Cheung, S. Baker, T. Kanade, “Shape-from-silhouette for articulated objects and its use for human body kinematics estimation and motion capture,” In: *Computer Vision and Pattern Recognition*, Madison, Wisconsin, USA, June 16–22, 2003.

[8] TB. Moeslund, A. Hilton, V. Kruger, A survey of advances in vision-based human motion capture and analysis, Computer Vision and Image Understanding (CVIU) 104 (2–3) (2006) 90–126.

[9] F. Caillette and T. Howard. Real-Time Markerless Human Body Tracking with Multi-View 3-D Voxel Reconstruction. In *Proc BMVC*. vol. 2, pp. 597.606, 2004.

[10] S. Iwase, H. Saito, “Parallel tracking of all soccer players by integrating detected positions in multiple view images,” In: *International Conference on Pattern Recognition*, Cambridge, UK, Aug 2004.

[11] M. Piccardi, “Background subtraction techniques: a review,” in *Proc. IEEE Int. Conf. Systems, Man, Cybernetics*, 2004, pp. 3099–3104.

[12] C. Stauffer, W.E.L. Grimson, Learning patterns of activity using real-time tracking, in: Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, 2000, pp. 747–757.

[13] N.M. Oliver, B. Rosario, A.P. Pentland, A Bayesian computer vision system for modeling human interactions, in: Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, 2000, pp. 831–843.

[14] K. Toyama, J. Krumm, B. Brumitt, B. Meyers, “Wallflower: principles and practice of background maintenance,” In: *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 1, 1999, pp. 255–261.

[15] I. Haritaoglu, D. Harwood, L.S. Davis, W4: real-time surveillance of people and their activities, in: Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, 2000, pp. 809–830.

[16] C. Wren, A. Azarbayejani, T. Darrel, and A. Pentland, “Pfinder: Real-time tracking of the human body,” In: *Proc. SPIE,* Bellingham,WA, 1995.

[17] A. Mittal and L.S. Davis, “M2tracker: A multi-view approach to segmenting and tracking people in a cluttered scene using region-based stereo,” In: *Proc. 7th European Conf. Computer Vision, Kopenhagen, Danmark*, Vol. X, pages 18–33, 2002.

[18] J. Krumm et al., “Multi-Camera Multi-Person Tracking for Easy Living,” *Proc. 3rd IEEE Int’l Workshop Visual Surveillance*, IEEE Press, Piscataway, N.J., 2000, pp. 3-10.

[19] H. Nait-Charif, S. McKenna, “Activity Summarization and Fall Detection in Home Environment,” In: *Proc. of ICPR’04*, (2004) 323-326.

[20] T. Lee and A. Mihailidis. An intelligent emergency response system: preliminary development and testing of automated fall detection,” Journal of Telemedicine and Telecare, vol. 11, no. 4, 2005, pp. 194-198.

[21] A. Williams, D. Ganesan, A. Hanson, “Aging in place: Fall detection and localization in a distributed smart camera network,” In: *Proceedings of the 15th International Conference on Multimedia*, pp. 892–901, Augsburg, Germany, September 2007.

[22] Z. Zhou, X. Chen, YC. Chung, Z. He, TX. Han, and JM. Keller, "Activity Analysis, Summarization, and Visualization for Indoor Human Activity Monitoring," IEEE Transactions on Circuits and Systems for Video Technology, 18:11, 2008.

[23] R. H. Luke, “Moving Object Segmentation from Video Using Fused Color and Texture Features in Indoor Environments”, Technical Report, CIRL, University of Missouri, 2008.

[24] D. Anderson, R. H. Luke, J. M. Keller, M. Skubic, M. Rantz, and M. Aud, “Linguistic Summarization of Video for Fall Detection Using Voxel Person and Fuzzy Logic,” *Computer Vision and Image Understanding*, vol. 113, pp. 80-89, 2009.

[25] E. Stone, D. Anderson, M. Skubic, and J. Keller, "Extracting Footfalls from Voxel Data," 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Buenos Aires, Argentina, Aug 31-Sep 4, 2010.

[26] F. Wang, M. Skubic, C. Abbott, and J. Keller, "Body Sway Measurement for Fall Risk Assessment Using Inexpensive Webcams," 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Buenos Aires, Argentina, Aug 31-Sep 4, 2010.

[27] D. Anderson, R. H. Luke, E. Stone, and J. M. Keller, “Fuzzy Voxel Object,” *International Fuzzy Systems Association (IFSA)*, pp. 282-287, Lisbon, Portugal, July 2009.