Silhouette Segmentation for Near-Fall Detection Through Analysis of Human Movements in Surveillance Videos

Imane Bouraoui[1] and Jean Meunier[2]

[1]Department of Electronics, Faculty of Sciences and Technology, UMSB, Jijel, Algeria

[2]Department of Computer Science and Operations Research, University of Montreal, Canada

Keywords: Near-Fall, Silhouette, Background Substruction, Mask R-CNN, SVM, Autoencoder.

Abstract: The detection of near-fall incidents is crucial in surveillance systems to improve safety, prevent future more serious falls and ensure rapid intervention. the main objective of this paper is the detection of movement anomalies in a series of video sequences through silhouette segmentation. First, we begin by isolating the person from the background, keeping only the person's silhouette. This is achieved through two methods: the first involves median pixel, while the second utilizes an algorithm based on pre-trained Mask Regional Convolutional Neural Network (Mask R-CNN) model. the second step involves movement calculation and noise effect minimization. Finally, we conclude by classifying normal and abnormal movement signals obtained using two different classifiers: Support Vector Machine (SVM) and Autoencoder (AE). We then compare the results to determine the most efficient and rapid system for detecting near-falls. the experimental results demonstrate the effectiveness of the proposed approach in detecting near-fall incidents. Specifically, the Mask R-CNN approach outperformed the median pixel method in silhouette extraction, enhancing anomaly detection accuracy. AE surpassed SVM in accuracy and performance, making it suitable for real-time near-fall detection in surveillance applications.

1 INTRODUCTION

the human body can be prone to falls due to factors such as aging, tiredness, medical conditions such as osteoporosis, Parkinson's disease, and vertigo, as well as medications that cause dizziness or sleepiness. These conditions impact balance and coordination, significantly increasing the risk of falling. Falls are the leading cause of fatal accidents among seniors, and can cause serious physical damages including head injuries (Mobsite *et al.*, 2023). According to the World Health Organization (WHO), approximately 684,000 fatal falls occur annually, making them the second leading cause of accidental injury deaths. Additionally, around 37.3 million falls require medical attention each year (Silva *et al.*, 2024). Consequently, it is crucial to develop and implement effective fall prevention strategies.

to solve this issue, we propose a video surveillance system that can be implemented in various locations, including residences, hospitals, airports, and factories. This system is designed to evaluate fall risk by detecting near-falls and identifying mobility issues through anomaly detection, which is favoured due to the high variability of abnormal mobility. the system helps in identifying potential emergencies and activates alarms when danger is detected.

Our objective is to find an efficient and rapid system for detecting near-falls by creating a fall prevention system that identifies movement anomalies in video sequences using silhouette segmentation. We achieve this by isolating the person from the background to simplify the analysis, retaining only the silhouette through two methods: the median pixel method and a pre-trained Mask Regional Convolutional Neural Network (Mask R-CNN) algorithm. the median pixel method computes the median value of pixels across frames to get a background model, while Mask R-CNN is a deep learning-based approach that accurately segments objects, (He *et al.*, 2017). We then compute movement using background subtraction, which highlights moving regions and minimizes noise effects that may result from small environmental variations or inaccuracies in segmentation. Finally, we classify normal and abnormal movement signals using two types of classifiers: a supervised learning

Bouraoui, I. and Meunier, J.
Silhouette Segmentation for Near-Fall Detection Through Analysis of Human Movements in Surveillance Videos.
DOI: 10.5220/0013153500003905

In *Proceedings of the 14th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2025)*, pages 620-627
ISBN: 978-989-758-730-6; ISSN: 2184-4313

algorithm, Support Vector Machine (SVM), and an unsupervised neural network, Autoencoder (AE), (Neloy *et al.*, 2024).

the rest of the paper is organized as follows: Section 2 reviews the literature on human fall detection and highlights the limitations of current systems. Section 3 explains the two methods used for detecting human bodies. Section 4 describes a dataset of simulated normal and abnormal activities in a realistic apartment-laboratory and tests it with the proposed algorithms. Section 5 discusses the findings and compares the results obtained by the two proposed methods. Finally, Section 6 wraps up this work by outlining the limitations and suggesting possible directions for future research.

2 RESEARCH LITERATURE

This research literature explores various technologies and methods to improve the accuracy and efficiency of fall detection systems for the elderly. (Anderson *et al.*, 2006) employed video-based silhouettes to detect falls by creating binary maps of body positions, which were used to train hidden Markov models for activity recognition. (Rougier *et al.*, 2007) proposed a fall detection method using video surveillance, combining movement history and human shape variation to identify falls in seniors. the approach aims to improve safety and quality of life for the elderly, showing promising results on video sequences of daily activities and simulated falls. (Zigel *et al.*, 2009) introduced a fall detection system that utilizes floor vibrations and sound sensing, relying on signal processing and pattern recognition without the need for wearable devices, making it effective even if the person is unconscious or stressed. (Chen *et al.*, 2010) developed a real-time video-based system that detects falls by combining skeleton features with human shape variations, achieving high accuracy in distinguishing actual falls from similar activities. (Chua *et al.*, 2013) used an uncalibrated camera to detect falls by analyzing human shape and head movements, incorporating new ellipse-based and head shape models to improve detection accuracy. (Yang and Lin, 2014) proposed a depth image processing approach to detect falls, particularly effective when pedestrians are partially obscured. Their method accurately distinguishes between humans and objects, adjusts for lighting variations, and measures tilt angles. (Kwolek and Kepski, 2014) combined depth maps with a wireless accelerometer to detect falls and reduce false alarms, using movement and acceleration data analyzed by a Support Vector Machine (SVM) classifier for reliable fall detection while preserving privacy. (Nizam *et al.*, 2017) utilized a Microsoft Kinect Sensor to track joints and measure velocity for fall detection, identifying falls based on abnormal joint positions and sudden changes in velocity.

Recent advancements in fall detection systems have significantly benefited from innovations in artificial intelligence and sensor technologies. (Asif *et al.*, 2020) proposed a deep learning framework for privacy-preserving human fall detection using RGB video data. Addressing the critical issue of fall detection for the elderly, their system utilizes synthetic data to train models that recognize falls from real-world footage. By focusing on human skeleton and segmentation rather than raw images, the framework ensures privacy by anonymizing personal information. (Zhu *et al.*, 2021) developed a fall detection algorithm that utilizes deep learning, computer vision, and human skeleton keypoints. This method employs OpenPose to extract skeleton data and applies deep learning for classifying falls, aiming to improve detection accuracy for the elderly. (Chen *et al.*, 2022) designed a system that combines vision-based fall detection with Building Information Modeling (BIM) for rescue routing. the system includes modules for fall detection, communication via a cloud server, and rescue route planning, aiming to improve response times and reduce injury severity. (Mobsite *et al.*, 2023) introduced a privacy-preserving camera-based fall detection system using semantic segmentation. the system extracts human silhouettes with a Multi-Scale Skip Connection Segmentation Network (MSSkip) and analyzes them using a ConvLSTM network, achieving high accuracy in detecting and classifying falls. (Gao *et al.*, 2023) proposed a fall detection method based on human pose estimation and a lightweight neural network. OpenPose extracts human keypoints, which are processed by a modified MobileNetV2 network to improve fall detection accuracy by correcting keypoint labeling errors. (Duong *et al.*, 2023) reviewed deep learning techniques for video surveillance anomaly detection, categorizing approaches based on objectives and metrics. the review highlights the significance of generative models and feature engineering, and discusses challenges and future research directions. (Alanazi *et al.*, 2024) developed a vision-based fall detection system using a 4-stream 3D convolutional neural network (4S-3DCNN) and image fusion. (Silva *et al.*, 2024) evaluated wearable fall detection systems, noting that performance significantly decreases when transitioning from simulated to real-world conditions.

the study emphasizes the need for more realistic testing to improve system effectiveness. (Gharghan and Hashim, 2024) reviewed elderly fall detection systems, focusing on the impact of wireless communication and AI technologies. Their study categorizes traditional and AI-based methods, evaluating system architectures, sensors, and performance to help researchers select effective fall detection solutions. (Das Chagas *et al.*, 2024) proposed a fall risk detection method using Channel State Information (CSI) from wireless networks and IoT devices. This method employs k-Nearest Neighbors (kNN) to detect fall risks by monitoring changes in wireless signals, achieving high accuracy in hospital settings.

Few studies have specifically focused on near-fall detection. (Dubois and Charpillet, 2014) developed a markerless fall prevention system using Microsoft Kinect to track human movement and analyze gait parameters like step length and gait speed to assess fall risk in the elderly, proving reliable in real-world scenarios. (Yang *et al.*, 2016) introduced a semi-supervised system using wearable inertial measurement units (WIMUs) to detect near-miss falls in ironworkers, using a one-class support vector machine to automatically identify near-misses without disrupting work. (Tripathy *et al.*, 2018) created an eigen posture-based system with Kinect sensors to assess fall risk by analysing postural instability.

More recently, (Choi *et al.*, 2022) employed inertial measurement units and Directed Acyclic Graph Convolutional Neural Networks (DAG-CNN) for near-fall detection, emphasizing the need for further research in pre-emptive fall prevention. (Tran *et al.*, 2022) proposed video surveillance systems, employing machine learning algorithms like Isolation Forest and One-Class SVM to detect near-falls in seniors with high accuracy, offering early fall risk warnings based on the velocities of a simplified 3-joint skeleton. (Ebrahimi *et al.*, 2024) used an autoencoder to detect mobility anomalies, particularly near-falls, by identifying high reconstruction errors. They used a set of 20 features of seven key points on a skeleton, encompassing joint positions, velocities, accelerations, angles, and angular accelerations, to train their model. (Silva *et al.*, 2024) pointed out the limitations of fall detection systems trained on simulated falls, noting significant performance drops when tested on real-world data while (Yu *et al.*, 2024) proposed Semi-PFD, a semi-supervised model for pre-impact fall detection that outperformed supervised models, especially with limited fall data, highlighting its practical potential for injury prevention.

Despite these successes, near-fall detection remains a challenge. the best results obtained by (Ebrahimi *et al.*, 2024) and (Tran *et al.*, 2022) relied on skeleton extraction that was not always reliable and involved some filtering and fine-tuning steps for good results and several tests to obtain the best model parameters. In our work we propose to use the silhouette of the person instead of the skeleton for a simpler body representation and better near-fall detection.

3 METHODOLOGIES

3.1 Dataset

the dataset used in this work comprises realistic and challenging videos intended for near-fall detection. Several RGB and RGBD videos were captured using an Intel RealSense Depth Camera D455, mounted in a ceiling corner to provide a full view of an individual performing various daily activities in the living room of an apartment laboratory. For this study, 150 short RGB videos of one subject were selected from the dataset, each lasting about 15 to 30 seconds (the other videos of the dataset involved another subject and other activities not relevant for our study because without realistic near-fall situations). Among these, 100 videos show the subject performing various activities of daily living, such as walking, dusting, tying shoelaces, turning on the television with a remote control, opening a door, picking up a book and reading, and more, while the remaining 50 videos present realistic near-fall situations, such as tripping over a mat, colliding with a table leg, or losing balance (figure 1).

3.2 Silhouette Segmentation

3.2.1 Median Pixel Method

the median pixel algorithm is a method used to remove backgrounds from images, especially in fixed video or time-lapse photography, where the background remains the same and only the foreground changes, (Sobral and Vacavant, 2014).

As shown in figure 2, a background model is first created by calculating the median pixel values from all input images from the dataset described in the previous section, creating an 8-bit grayscale image used as the reference for background subtraction. Each new frame is then subtracted from this median image. Otsu's thresholding method is applied to separate the foreground from the background, producing a binary mask that highlights moving objects or people. This process repeats for each new image in the sequence, enabling accurate foreground extraction.

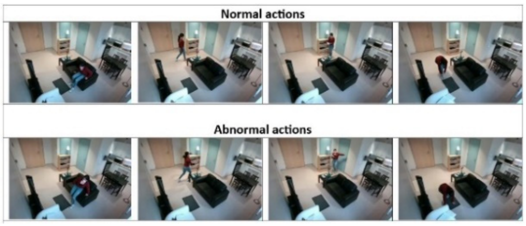

Figure 1: Examples of normal and abnormal movements in the dataset.

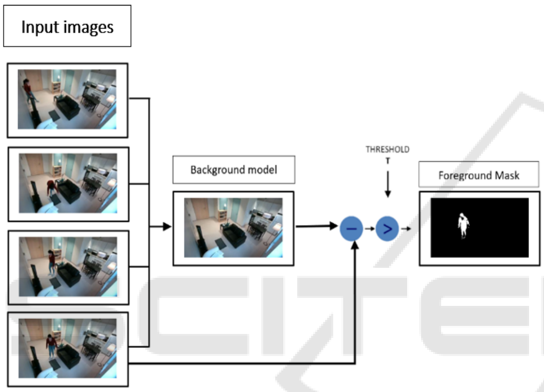

Figure 2: Diagram of the median pixel model for silhouette detection.

3.2.2 Pre-Trained Mask R-CNN Model

Mask R-CNN (Mask Region-based Convolutional Neural Network) is a powerful algorithm designed for object detection, with a particular focus on instance segmentation, allowing for both detection and pixel-level segmentation of objects, (Wang *et al.*, 2019).

For person detection, we processed a series of images from the dataset described in the section 3.1 using a pre-trained Mask R-CNN model, which is based on a ResNet-50 backbone. As shown in figure 3. This model extracts features from the images, generates proposals for potentially interesting regions via the Region Proposal Network (RPN), corrects spatial misalignments with ROI Align to ensure precise feature extraction, and classifies objects while refining their bounding boxes. the masks of the detected human silhouettes are then refined using the Otsu thresholding algorithm to improve the body edges, and a final image is created by overlaying the detected silhouettes on a dark background.

3.3 Near-Fall Detection Algorithm

For the detection of a near fall, a simple movement detection algorithm is applied based on the pose of the moving person, assuming that a near fall appears as a peak compared to the signal of normal movements.

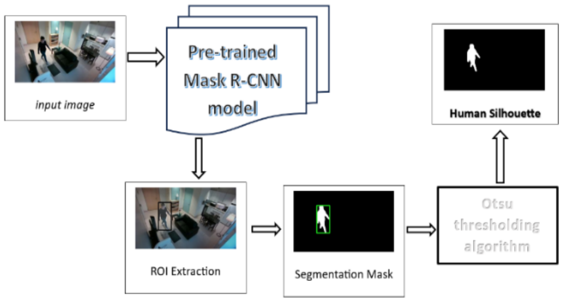

Figure 3: Diagram of the pre-trained Mask R-CNN model for silhouette detection.

However, one difficulty in movement detection lies in the fact that, in our dataset, the camera is fixed, so as the person approaches, the number of pixels increases, and as the person moves away, the number of pixels decreases. For this reason, we calculated a movement ratio which provides a normalized measure of movement by considering both the amount of movement (related to the difference between consecutive silhouettes) and the average intensity of the frames (related to the size of the silhouette). This can be useful for comparative analysis across different scenes or lighting conditions.

The algorithm for movement detection of multiple frames is presented as follows:

Data: Sequence of silhouette images I_t, extracted for each video.
Results: Normalized movement ratio $Ratio_t$ for detecting anomalies in movement patterns.

Pre-processing:

1. Silhouette images are extracted from each video and converted to 8-bit grayscale.
2. A Gaussian filter smooths each image, reducing noise and enhancing movement detection.
3. the filtered image at time t is denoted as $I_t^{filtered}$.

Algorithm Steps:

1. *Initialize previous frames*:
   5 previous frames are stored in a list to compute temporal changes:

$$prev\_frames = \{\emptyset, \emptyset, \emptyset, \emptyset, \emptyset\} \quad (1)$$

2. *Background Substruction*:
If $prev\_frames[1] \neq \emptyset$, compute the difference between the current filtered frame and the previous 5 frames to detect movement areas:

$$\Delta_t = \sum_{k=1}^{5} \left| I_t^{filtered} - prev\_frames[k] \right| \tag{2}$$

3. *Binary Thresholding*:
A binary mask B_t is created by thresholding Δ_t, where pixel intensities exceeding a predefined threshold are classified as motion pixels.

$$B_t = \begin{cases} 1 & if\ \Delta_t(x,y) > Threshold \\ 0 & otherwise \end{cases} \tag{3}$$

4. *Movement Pixel Count*:
the total count of motion pixels is calculated, representing the magnitude of detected movement.

$$Motion_t = \sum_{x,y} B_t(x,y) \tag{4}$$

5. *Average Frames Intensity*:
Compute the average intensity of the current frame and the previous 5 frames for normalization:

$$Average_t = \frac{1}{6N}\left(\sum_{k=1}^{5}\sum_{x,y} prev\_frames[k](x,y) + \sum_{x,y} I_t^{filtered}(x,y)\right) \tag{5}$$

where N is the total number of pixels in the frame.

6. *Movement Ratio Calculation*:
the ratio of motion pixel count to average intensity provides a normalized measure of movement.

$$Ratio_t = \frac{motion_t}{Average_t} \tag{6}$$

7. *Update Previous Frames*:
the list of previous frames is updated with the latest frame, maintaining a sliding window of 5 frames.

$$\begin{cases} prev\_frames[1] = prev\_frames[2] \\ prev\_frames[2] = prev\_frames[3] \\ prev\_frames[3] = prev\_frames[4] \\ prev\_frames[4] = prev\_frames[5] \\ prev\_frames[5] = I_t^{filtered} \end{cases} \tag{7}$$

3.4 Classification

3.4.1 Support Vector Machines (SVM)

Support vector machine (SVM) is a powerful, flexible supervised learning algorithm most commonly used for classification; it can also be used for regression. the algorithm finds an optimal hyperplane to divide the datasets into different classes, (Sarang, 2023). SVMs are particularly effective for anomaly detection for several reasons: they perform well with high-dimensional data, are robust against overfitting, and aim to optimally separate normal data from anomalies. Additionally, SVMs handle imbalanced datasets effectively and can apply the kernel trick to achieve non-linear separation. Their clear decision boundaries also enhance the interpretability of anomaly classifications.

3.4.2 Autoencoder

An autoencoder is a neural network with an encoder and decoder trained to learn reconstructions close to the original input. the difference between the original input and the reconstruction output in the autoencoder is called the reconstruction error, (Torabi *et al.*, 2023). For anomaly detection, an autoencoder trained on normal data only aims to closely reconstruct inputs representing normal patterns. If the reconstruction error exceeds a predefined threshold, the input is classified as an anomaly (e.g., a near-fall); otherwise, it is considered normal. the model has two dense layers: the encoder compresses the data into a lower-dimensional latent space, and the decoder reconstructs it back to the original size. the model is trained using input and target data from the training set, aiming to minimize the reconstruction error, with the Adam optimizer and MSE loss function, (Kopčan *et al.*, 2021).

4 RESULTS AND DISCUSSION

the results were obtained using 100 videos of normal movements and 50 videos of abnormal movements (near-falls), after extracting the silhouettes of moving individuals using the Median Pixel method and the pre-trained Mask-RCNN model. Figure 4 illustrates the detection of normal movements (2 videos) and abnormal movements (1 video) with the Median Pixel method, while figure 5 presents the motion evolution over time for the same videos using Mask-RCNN. the near-fall is identified as a peak when the motion ratio exceeds 900 with the Median Pixel method and 800 with Mask-RCNN; values above these thresholds are considered anomalous movements, while those below are classified as normal.

In the SVM classifier, the dataset is split into training and test sets with 20% allocated for testing, using a random state of 42 for reproducibility. the data is reshaped into a two-dimensional form for compatibility with the SVM, which uses a linear kernel and a regularization parameter C=1, ensuring stable results with the same random state.

In the autoencoder, 50 videos of normal movements were used for training, while 100 videos were used for testing, consisting of 50 normal and 50 abnormal movements. the hyperparameters are adjusted as follows: the validation split for training is set to 10%. the latent space size in the autoencoder is set to 32, representing the dimension of the encoded layer. the learning rate, which controls the speed of weight adjustment in the Adam optimizer, is set to 0.001. the number of epochs is defined as 100, indicating how many times the entire training dataset is passed through during training. the batch size is set to 32, which represents the number of samples used to compute each weight update. Finally, the anomaly detection threshold is set to 80% of the mean squared error (MSE) of the reconstructed test data, serving to classify movements as normal or abnormal.

the classification results using SVM and AE are illustrated on the ROC curve in figure 6, where the true positive rate (TPR) indicates the proportion of the normal class correctly identified, and the false positive rate (FPR) measures the proportion of anomalous patterns misclassified as normal. Performance metrics, including area under the curve (AUC), equal error rate (EER), sensitivity, specificity, and accuracy, are summarized in table 1. Mask-RCNN with Autoencoder (AE) shows good specificity and a high AUC but has limited sensitivity. Mask-RCNN with SVM yields acceptable results but is outperformed by the AE version. In the other hand, models based on the Median Pixel method are the least effective, displaying very low sensitivities. Although this method is fast and can detect individuals in all images of a sequence as full silhouettes or contours (figure 7), its effectiveness decreases in the presence of objects, which may lead to the detection of shadows, moving furniture or partial silhouettes. In contrast, the RCNN method accurately detects the full silhouette without including the person's shadow, though it is slower and may confuse objects with the person, generating noise.

The model proposed by (Tran *et al.*, 2022) delivers a better performance but they used only 55 videos from the database with activities performed at a constant distance between the subject and the camera to avoid perspective distortion to improve their results. (Ebrahimi *et al.*, 2024) generates the best overall performance due to its superior sensitivity, specificity, and accuracy, offering exceptional results. But in their study, there was a lot of manual intervention to adjust the model architecture and hyperparameters, to improve the skeleton extraction and to choose the rights handcrafted features.

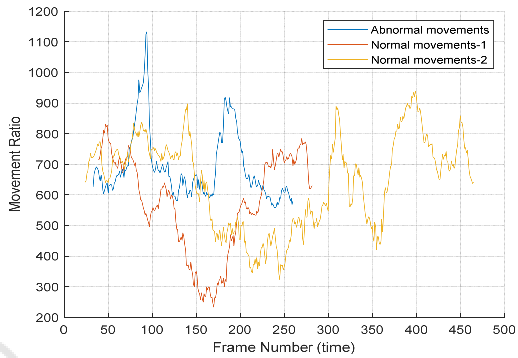

Figure 4: Results of temporal movement evolution using the Median Pixel method for silhouette detection.

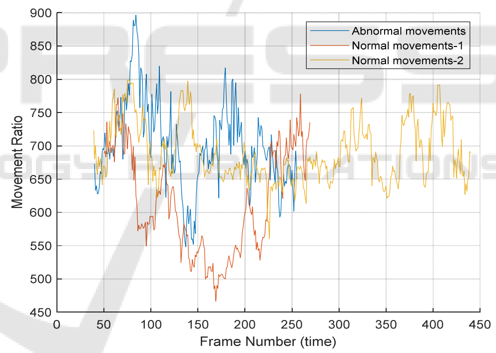

Figure 5: Results of temporal movement evolution using the pre-trained Mask R-CNN model for silhouette detection.

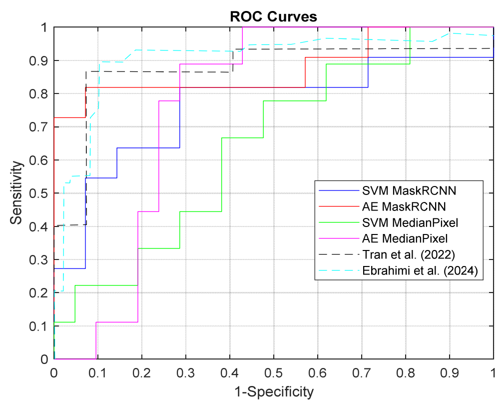

Figure 6: ROC curves of the near-fall detection systems.

Table 1: Near-fall performance.

	Sensitivity	Specificity	Accuracy	ERR	F1-score	AUC
Mask RCNN (SVM)	0.64	0.71	0.68	0.32	0.64	0.76
Mask RCNN (AE)	0.54	0.93	0.76	0.24	0.67	0.88
Median Pixel (SVM)	0.44	0.71	0.64	0.37	0.42	0.65
Median Pixel (AE)	0.22	0.81	0.64	0.37	0.27	0.77
Tran *et al.* (2022)	0.93	0.87	0.90	0.10	0.897	0.84
Ebrahimi *et al.* (2024)	0.90	0.90	0.92	0.10	0.90	0.92

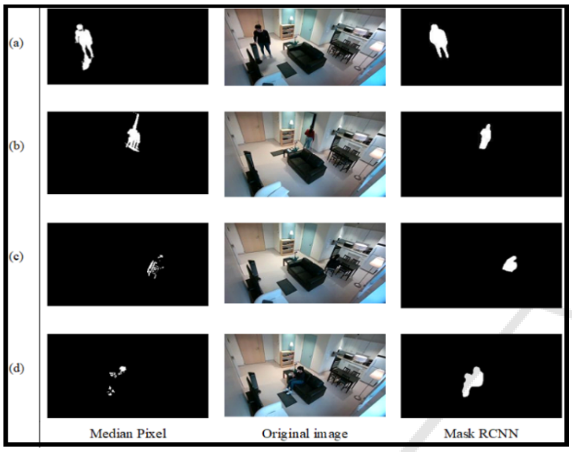

Figure 7: Limitations of median pixel and Mask_RCNN methods for silhouette extraction, (a) shadow, (b) objects, (c) and (d) incomplete silhouette.

5 CONCLUSIONS

In this work, we developed a fall prevention system that can detect abnormal movements, with potential uses for protecting vulnerable individuals. This system is based on silhouette segmentation and motion analysis in video sequences. We used a realistic dataset for effective model training and evaluation. Two methods were tested, successfully isolating individuals from the background to simplify the detection process: median pixel, which is fast but sensitive to background objects, and Mask R-CNN, which is more accurate but slower. Pre-trained Mask R-CNN combined with an autoencoder showed good specificity but limited sensitivity, while the version with SVM performed well but was still outperformed by the autoencoder version. the median pixel method was the least effective, with very low sensitivity, making it harder to detect near-falls.

In the near future, we would like to improve the models' sensitivity and explore other techniques to better detect anomalies in real-time and across various environments, especially by considering the velocity and acceleration of individuals.

ACKNOWLEDGEMENTS

This work was supported by the NSERC (Natural Sciences and Engineering Research Council of Canada). the authors would like to thank the research team for their assistance throughout the project, particularly in the construction of the dataset.

REFERENCES

Alanazi, T., Babutain, K., Muhammad, G. (2024). Mitigating human fall injuries: A novel system utilizing 3D 4-stream convolutional neural networks and image fusion. *Image and Vision Computing*. V. 148, August 2024, 105153.

Anderson, D., Keller, J., M., Skubic, M., Chen, X., and He, Z. (2006). Recognizing Falls from Silhouettes. *International Conference of the IEEE Engineering in Medicine and Biology Society*. 30 August 2006 - 03 September 2006. New York, NY, USA.

Asif, U., Mashford, B., von Cavallar, S., Yohanandan, S., Roy, S., Tang, J., Harrer, S. (2020). Privacy Preserving Human Fall Detection using Video Data. *In Proceedings of Machine Learning Research* 116:39–51. Published in ML4H@NeurIPS 30 April 2020. Computer Science, Engineering.

Chen, Y-T., Lin, Y-C., Fang, W-H. (2010). A hybrid human fall detection scheme. *IEEE International Conference on Image Processing*. 26-29 September 2010. Hong Kong, China.

Chen, Y., Zhang, Y., Xiao, B., Li, H. (2022). A framework for the elderly first aid system by integrating vision-based fall detection and BIM-based indoor rescue routing. *Advanced Engineering Informatics*. Vol. 54.

Choi, A., Kim, T., H., Yuhai, O., Jeong, S., Kim, K., Kim, H., and Mun, J., H. (2022). Deep Learning-Based near-Fall Detection Algorithm for Fall Risk Monitoring System Using a Single Inertial Measurement Unit. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. V.30, Page(s): 2385 - 2394. DOI: 10.1109/TNSRE.2022.3199068.

Chua, J-L., Chang, Y-C., Lim, W-K. (2013). Visual based fall detection through human shape variation and head detection. *International Conference on Multimedia, Signal Processing and Communication Technologies.* 23-25 November 2013. Aligarh, India.

das Chagas, A., O., S., Lovisolo, L., Tcheou, M., P., de Oliveira, J., B., and Souza F. (2024). A proposal for fall-risk detection of hospitalized patients from wireless channel state information using Internet of Things devices. *Engineering Applications of Artificial Intelligence*. Volume 133, Part F, 108628.

Dubois, A., Charpillet. F. (2014). A gait analysis method based on a depth camera for fall prevention. *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 26-30 August 2014. Chicago, IL, USA.

Duong, H-T., Le, V-T., Hoang, V-T. (2023). Deep Learning-Based Anomaly Detection in Video Surveillance: A Survey. *Sensors* (*Basel*). 23(11):5024.

Ebrahimi, F., Rousseau J., Meunier J. (2024), Mobility Anomaly Detection with Video Surveillance, *in Proc. of the Intl. Conf. on Image Processing, Computer Vision, & Pattern Recognition*, part of the 2024 CSCE congress, Las Vegas, NV, USA, July 2024.

Gao, M., Li, J., Zhou, D., Zhi, Y., Zhang, M., and Li, B. (2023). Fall detection based on OpenPose and MobileNetV2 network. IET Image Processing: Volume 17, Issue 3. Pages: 637-968. 12667.

Gharghan, S., K., Huda Ali Hashim, H., A. (2024). A comprehensive review of elderly fall detection using wireless communication and artificial intelligence techniques. *Measurement*. Volume 226. 114186.

He, K., Gkioxari,G., Dollár, P., Girshick, R. (2017). Mask R-CNN. *IEEE Intl. Conf. on Computer Vision* (ICCV). 22-29 October 2017. Venice, Italy.

Kopčan, J., Škvarek, O., Klimo, M. (2021), Anomaly detection using Autoencoders and Deep Convolution Generative Adversarial Networks. *Transportation Research Procedia*, Volume 55, Pages 1296-1303.

Kwolek, B., Kepski, M. (2014). Human fall detection on embedded platform using depth maps and wireless accelerometer. *Computer Methods and Programs in Biomedicine*. Volume 117, Issue 3, Pages 489-501.

Mobsite, S., Alaoui, N., Boulmalf, M., Ghogho, M. (2023). Semantic segmentation-based system for fall detection and post-fall posture classification. *Engineering Applications of Artificial Intelligence*. V. 117, Part B, 105616.

Neloy, A-A., Turgeon, M. (2024), A comprehensive study of auto-encoders for anomaly detection: Efficiency and trade-offs. Machine Learning with Applications, Volume 17, 100572.

Nizam, Y., Haji Mohd, M., N., Abdul Jamil, M.M. (2017). Human Fall Detection from Depth Images using Position and Velocity of Subject. *Procedia Computer Science*. Volume 105, Pages 131-137.

Rougier, C., Meunier, J., St-Arnaud, A., Rousseau, J. (2007). Fall Detection from Human Shape and

Movement History using Video Surveillance. *21st Intl Conf. on Advanced Information Networking and Applications Workshops* (AINAW'07), 21-23 May 2007. Niagara Falls, ON, Canada, pp. 875-880.

Sarang, P. (2023). Support Vector Machines. In: Thinking Data Science. the Springer Series in Applied Machine Learning. Springer, Cham. https://doi.org/10.1007/978-3-031-02363-7_8

Silva, C., A., Casilari, E., García-Bermúdez, R. (2024). Cross-dataset evaluation of wearable fall detection systems using data from real falls and long-term monitoring of daily life. *Measurement*. Volume 235. 114992.

Sobral, A., Vacavant, A. (2014). A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos. *Computer Vision and Image Understanding*, Volume 122, Pages 4-21.

Tran, K., C., Gassi, M., Nehme, P., Rousseau, J., Meunier, J. (2022). Video surveillance for near-fall detection at home. *In proc. of the 22nd IEEE Intl Conf. on Bioinformatics and Bioengineering*. 07-09 November 2022. Taichung, Taiwan.

Tripathy, S., R., Kingshuk Chakravarty, K., Sinha, A. (2018). Eigen Posture Based Fall Risk Assessment System Using Kinect. *40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 18-21 July 2018. Honolulu, HI, USA.

Torabi, H., Mirtaheri, S. L., Greco, S. (2023). Practical autoencoder based anomaly detection by using vector reconstruction error. Cybersecurity, vol.6, no.1.

Wang, S., Miranda, F., Wang, Y., Rasheed, R. and Bhatt, T. (2022). near-Fall Detection in Unexpected Slips during over-Ground Locomovement with Body-Worn Sensors among Older Adults. *Sensors*, 22(9), 3334.

Wang, T., Hsieh, Y-Y., Wong, F-W., Chen, T-F. (2019). Mask-RCNN Based People Detection Using a Top-View Fisheye Camera. *Intl Conf. on Technologies and Applications of Artificial Intelligence*. 21-23 Nov. 2019. Kaohsiung, Taiwan.

Yang, S-W., Lin, S-K. (2014). Fall detection for multiple pedestrians using depth image processing technique. *Computer Methods and Programs in Biomedicine*. Volume 114, Issue 2, Pages 172-182.

Yang, K., Ahn, C., R., Vuran, M., C., Aria, S., S. (2016). Semi-supervised near-miss fall detection for ironworkers with a wearable inertial measurement unit. *Automation in Construction*. V. 68, Pages 194-202.

Yu, X., Wan, J., An, G., Yin, X., Xiong, S. (2024). A novel semi-supervised model for pre-impact fall detection with limited fall data. *Engineering Applications of Artificial Intelligence.* Volume 132. 108469.

Zhu, N., Zhao, G., Zhang, X., Jin, X. (2021). Falling movement detection algorithm based on deep learning. *IET Image Processing*. 16(11).

Zigel, Y., Litvak, D., Gannot, I. (2009). A Method for Automatic Fall Detection of Elderly People Using Floor Vibrations and Sound—Proof of Concept on Human Mimicking Doll Falls. *IEEE Transactions on Biomedical Engineering,* V. 56(12). Page(s): 2858 – 2867.