Identifying unintentional falls in action videos using the 3D Cylindrical Trace Transform

Georgios Goudelis*, Georgios Tsatiris*, Kostas Karpouzis*, Stefanos Kollias*†
*School of Electrical and Computer Engineering
National Technical University of Athens
9, Heroon Politechniou str., 15780, Athens, Greece
{ggoudelis, gtsatiris}@image.ntua.gr, kkarpou@cs.ntua.gr
†School of Computer Science
University of Lincoln
Brayford Pool, Lincoln, Lincolnshire, UK
skollias@lincoln.ac.uk

***Abstract*—Identification of unintentional falls is a critical application in smart assistive environments, especially in the context of elderly care. However, visually discriminating between falls and fall-like, intentional activities is a challenging task. In this paper, we propose the utilization of a novel feature extraction scheme based on the newly formulated *3D Cylindrical Trace Transform*, on spatio-temporal interest points, for the task of fall detection. Using this pipeline, we are able to produce features invariant to occlusion, viewpoint, camera placement and other distortions. Experimentation on two publicly available datasets, on a number of different conditions, proved the efficiency of the proposed methodology for the task at hand.**

I. Introduction

The need to automatically detect falls has mainly arisen from the tendency of elder people to live alone or spend a lot of time unattended. Care for the elderly has traditionally been the responsibility of family members and was provided within a home environment. Increasingly in modern societies, state or charitable institutions are also involved in the process. Decreasing family size, the greater life expectancy of elderly people, the geographical dispersion of families and changes in work and education habits have attributed to this [3]. These changes have affected European and North American countries but are now increasingly affecting Asian countries as well [4].

Research is focused on the autonomy of elderly people which tend to live alone or are not able to indulge themselves in the luxury of an attendance person. Falls are a major public health issue among the elderly and recently, in this context, there has been an increasing focus on fall detecting systems.

A division of fall detection techniques could be into two main categories: wearable sensor based and vision based techniques. The first category is based on wearable devices such as accelerometers and gyroscopes, or on smartphones that contain this kind of sensors and are mainly carried continuously by subjects. The second category is based on 2D or 3D cameras, involving image analysis and pattern recognition techniques of high computational complexity. Methods in the latter category present the advantage that a continually carried device is not required. Of course, multiple modalities may be joined to produced composite methods. A characteristic example of a multimodal approach is given in [7]. In other studies, researchers in [8] divide fall detectors in three main categories: wearable device based, ambiance sensor based and camera (vision) based, while, from a different perspective, researchers in [9] make distinctions based on whether a specific method measures acceleration or not.

Primary attempts on providing a general overview of the fall detection status are presented in [10] and in [9]. However, as the advancement of technology on this area is rapidly growing, these reviews are mostly outdated. A newer, comparative study and more extensive literature review is provided in [6]. This article aims to serve as a reference for both clinical and biomedical engineers planning or conducting investigations on the field. The authors are mostly trying to identify real-world performance challenges and the current trends on the field. A more detailed discussion is provided in [8] but lacks references to new trends, such as smartphone-based techniques.

In the direction of vision based solutions, such as the one presented in this paper, researchers in [11], placed the camera on the ceiling and analyzed the segmented silhouette and the 2D velocity of the subject. The determination of a fall is achieved by an experienced thresholding. Authors in [12], in order to draw a distinction between falling and other fall-like activities, such as sitting, added the extra information of noise. However a sound-based system cannot be very robust as most of the environments where such solutions are applied are noisy. Another approach, presented in [13], is based on a combination of motion history and human shape variation. To cover large areas, wall cameras have been mounted and the final decision is made by thresholding the extracted features. In the study presented in [14], the classification between every day activities and fall events is achieved by extracting eigenmotion and by applying multi-class Support Vector Machines.

3D information extracted using depth sensors, such as the Microsoft Kinect, is shown to provide efficiency on partial occlusion and viewpoint problems. Thus, a number of works based on leveraging such information have been published. In [15], a velocity based method is presented, that takes into account the contraction or expansion of the width, height and depth of a 3D bounding box. A priori knowledge of the scene is not required as the set of captured 3D information is adequate to complete the process of fall detection. Another approach creates two feature parameters: the orientation of the body and the height information of the spine, using either image or world coordinates, based on captured Kinect data [16]. The Kinect sensor is also used in [17], where the proposed algorithm is based on the speed of the silhouette head (previously detected), the body centroid and their distance from the ground. Because it incorporates positions of both the body centroid and the head, this technique is regarded to be less affected by the centroid fluctuation. Finally, a statistical method based on Kinect as proposed in [18]. The decision is made based on information about how the human moved during the last few frames. This method combines a set of proposed features under a Bayesian framework. This study's main focus is to create a technique that, while it has been trained by data captured from a specific viewpoint, is also able to classify falls that have been captured by a different one.

A. The proposed work: A preample

The work introduced in this paper is a follow-up study to the one presented in [22], which was based on the pipeline proposed in [19]. However, this study focused on modeling actions in a per-frame fashion, not taking into account any temporal interlinking between prominent features in the action sequence. Although they show resilience to occlusion, this may reduce their applicability on highly occluded environments, where spatial information can be distorted. Moreover, without any mechanism to cope with the different lengths of action sequences, these techniques could not accurately incorporate any information regarding rapid position changes and velocity, which is vital in discriminating between similar actions. For instance, in an unintentional fall, we observe more abrupt position changes of the subject than when performing a crouching action or lying down. The methodology presented in [22] lacks the properties to take this information into account.

In this work, we take advantage of a new extension of the Trace Transform to the 3D domain, called the 3D Cylindrical Trace Transform [25] and its advanced properties, to surpass the limitations of previous methods and a solution to a practical problem.

II. Overview of the proposed system

The Trace transform can be seen as a generalization of the Radon Transform [20]. While the Radon transform of an image is a 2D representation of the image in coordinates ϕ and p with the value of the integral of the image computed along the corresponding (ϕ, p) line, Trace calculates functional T along this tracing line. This functional may not necessarily be the integral. The final transform is created by tracing an image with straight lines and calculating certain functionals of the image values along these lines. Transforms for different action snapshots are given in Figure 1. A detailed overview of the fundamental theory behind the Trace transform can be found in [21] and its use in action recognition can be seen in [19].

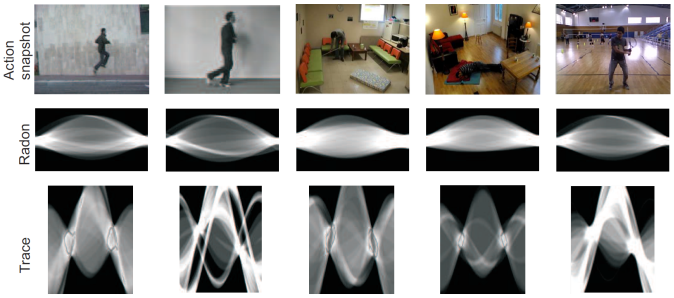

Fig. 1. Examples of Radon and Trace transforms created from the silhouettes of different action snapshots taken from various datasets.

A. The 3D Cylindrical Trace Transform

The definition of the 3D Cylindrical Trace Transform (thereby called 3D-CTT) has been inspired by the 3D Radon transform [23], as it is formulated in [24]. The 3D-CTT is an extension of the Trace Transform [21] to the 3D space.

As explained thoroughly in [25], the Cylindrical Trace Transform CTT_f is applied on spatio-temporal volumes such as spatio-temporal interest point (STIP) meshes [2][1]. If we consider a spatio-temporal mesh as a 3D model M, the 3D-CTT associates a functional T to each line tracing the model, placed at cell (p, φ, θ), with distance p and angle φ characterizing uniquely a line and the plane that forms angle θ with the origin. In essence, Trace transforms are continuously calculated on planes rotating in the direction of polar axis A, cutting the 3D mesh M.

Considering each cutting plane as a 2D function $\xi(x, y)$ formed by the projection of the M on that plane, its Trace transform $\breve{g}(p, \varphi)$ can be given by evaluating a functional T along all lines (p, φ) tracing ξ:

$$\breve{g}(p, \varphi) = T(\xi(x, y)\delta(p - x\cos\varphi - y\sin\varphi)) \tag{1}$$

The final representation of the 3D model, the proposed 3D-CTT, is also a 2D function of parameters (p, φ). It is given by the sum of the individual calculations of Trace transforms on planes rotated by angle θ relative to the origin:

$$CTT_f(p, \varphi) = \sum_{n=1}^{N} \breve{g}_n(p, \varphi). \tag{2}$$

where $\breve{g}_n$ is the n^{th} Trace transform, i.e. the transform calculated on the 2D planar projection of M on the plane that forms angle θ_n with the origin. Also: $N \geq 2$, $0 < \theta_n \leq \theta_{max}$ and $\theta_{max} = 180^o$. An illustration of the 3D Cylindrical Trace transform on a STIP mesh is given in Figure 2.

B. Feature extraction scheme

1) 3D Cylindrical Trace Transform on STIP meshes: As stated in more detail in [25], the 3D-CTT is applied on STIP meshes. Specifically, the STIP acquisition implementation

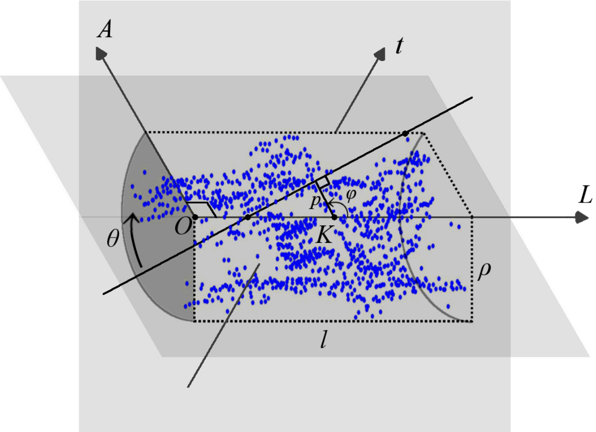

Fig. 2. 3D CTT calculation on Selective STIPs extracted from an action video sequence. t denotes the direction of time.

followed in the proposed pipeline is the one presented in [1], the so-called Selective Spatio-Temporal Interest Points (SSTIPs). This methodology focuses on global motion instead of local spatio-temporal information, thus preventing the erroneous detection of interest points due to cluttered backgrounds and camera motion. Furthermore, the authors show that their method produces stable, repeatable STIP meshes, robust to the local scene properties and suitable for action recognition tasks. An example of extracted SSTIPs from an action sequence is given in Figure 3.

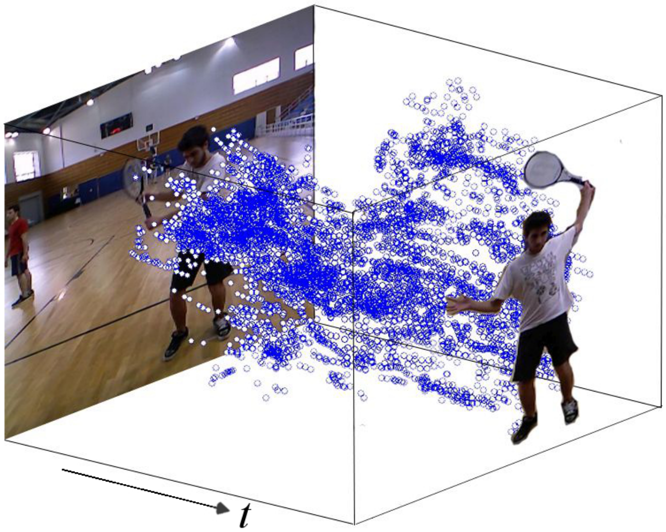

Fig. 3. Selective STIPs extracted from an action sequence. t denotes the direction of time.

By applying different functionals to the SSTIPs mesh M, a set of $CTT_{f_i}(p, \varphi)$ transforms is produced, where $i = 1...I$ and I is the number of transforms one chooses to calculate.

An improvement compared with previous techniques is the ability of this pipeline to encode variations in the length of action sequences. In other words, if the speed at which an action is performed plays a significant role in the classification of that action, this pipeline has the ability to incorporate it in the extracted feature set. We call this property *time-sensitivity*. Results on the fall detection datasets in section III-B demonstrate the potentness of this property in distinguishing between very similar spatio-temporal volumes of different, however, length. A typical example of such volumes are the ones generated by lying down and falling. As already mentioned in section I-A, previous Trace transform based techniques in [19] and [22] fail to take this information into account because they are not time-sensitive.

2) Volumetric triple features: In The fundamental work in [21], the formulation of Trace transform-based *triple features* is proposed in the following manner:

1) The Trace transform of a 2D function is produced, using a *Trace functional* T.
2) By calculating a *diametric functional* P along the columns of the 2D function's Trace transform, a circus function is obtained.
3) The final triple feature is ultimately produced by applying a *circus functional* Φ along the resulting vector of numbers from step 2.

In [19] it is shown that *ratios of pairs* of different triple features, constructed by using different functionals T, P and Φ on the frames of an action sequence video, create feature suitable for action recognition. Work in [22] also indicates their efficacy in the unintentional fall detection task.

The novel formulation of 3D-CTT in [25] proposed a modification of the aforementioned scheme, called *Volumetric Triple Features* (VTFs). Instead of a per-frame fashion, the triple feature extraction scheme is applied on the results of the 3D Trace transform of the complete spatio-temporal volume. For every $CTT_{f_i}(p, \varphi)$, calculated using functional T_i, a diametric functional P_i is applied along the columns of the transform. Then, a circus functional Φ_i is evaluated along the resulting string. This way, a set of Π triple features is computed. The procedure is illustrated in Figure 4.

All Π features are then divided by each other, to produce a new set of independent features. So, the given action sequence is finally encoded into a vector $\mathbf{v}$, the so-called VTF vector, which is essentially the set of all calculated triple feature ratios, based on the set of the different CTT_{f_i} applied on the SSTIPs of the action sequence.

$$\mathbf{v} = (\Pi_{rat_1}, \Pi_{rat_2}, ..., \Pi_{rat_{h-1}}, \Pi_{rat_h}) \qquad (3)$$

where Π_{rat} is the ratio of two triple features and h the number of calculated ratios.

The use of a dimensionality reduction technique on the resulting vectors is considered necessary, due to the fact that not all features in the vector share the same discriminatory power. In the proposed pipeline, Principal Component Analysis (PCA) is applied on the VTF vectors, in order to construct an appropriate subset of the features that is suitable for classification. In our experiments, only a small fraction of the initial VTF vector survives this task (typically between 25 and 40 features).

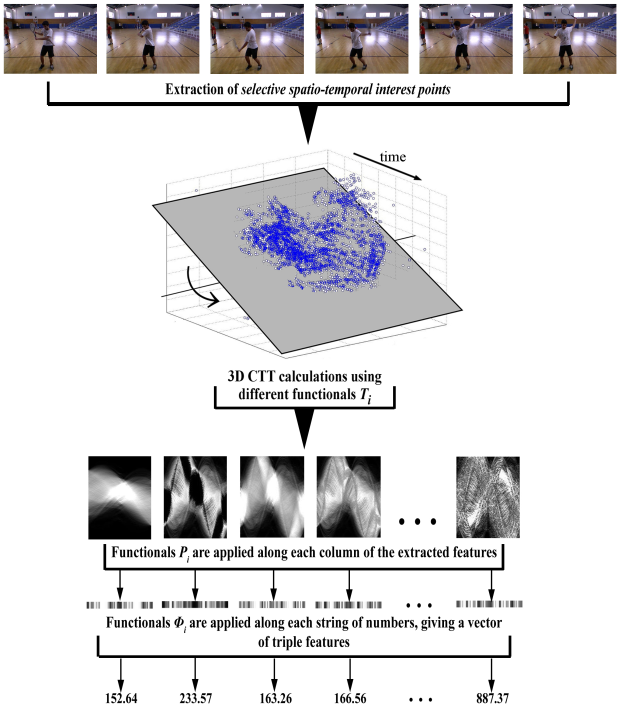
Extraction of *selective spatio-temporal interest points*
time
3D CTT calculations using different functionals T_i
Functionals P_i are applied along each column of the extracted features
Functionals Φ_i are applied along each string of numbers, giving a vector of triple features
152.64
233.57
163.26
166.56
887.37

Fig. 4. Triple feature extraction from a spatio-temporal volume.

III. Experimental evaluation

Although fall detection can be classified as a human action recognition task, the results are mostly calculated on a *yes* or *no* basis and a specific experimental approach is required. A potential fall detection system continually monitors a subject whose physical dynamic behavior is captured. This behavior is then analyzed at regular intervals and compared against a preprocessed dataset of sample actions, used to train the system. The final decision is made upon determining the relevance of the recorded incident, when compared to any one of the different samples and a *fall* or a *no fall* situation is reported. The protocol used for this particular task uses one sample for testing and the rest of the set is used for training. The decision made is binary (0 or 1) and is repeated I times where I is the number of samples in the dataset. Performance is reported as the ratio of successful classifications over I tests.

A. Experimental setup

In general, there are few available datasets dedicated to fall detection as most of the published techniques have been tested on the respective author's own datasets. However, in order to have a benchmark, we have evaluated our technique in two publicly available datasets: The UR Fall Detection [26], [27] and the Le2i Fall detection datasets [5].

The UR Fall dataset contains 60 sequences recorded with 2 Microsoft Kinect cameras and corresponding accelerometric data. Sensor data was collected using PS Move (60Hz) and x-IMU (256Hz) devices. The dataset contains sequences of depth and RGB images for two differently mounted cameras (parallel to the floor and ceiling mounted, respectively), synchronization data, and raw accelerometer data. Each video stream, both RGB and depth, is stored in separate folders in the form of png image sequences. From the specific dataset we have used the depth data provided by the ceiling mounted camera as well as the frontal RGB videos, following the experimental protocol given by authors in [26] and [27]. Frame samples taken from UR fall dataset are provided in Figure 5.

Experimentation on the UR fall detection dataset was divided into two phases. The first one aimed to evaluate the ceiling mounted depth camera scenario, corresponding to the methodology presented in [26]. More specifically, a set of 60 cropped motion sequences from the ceiling depth video subset were used. These motion sequences contained both unintentional falls, such as tripping and falling from a chair, and other everyday activities, such as walking. Artificial, fall-like activities were added to the dataset, to make the problem more challenging. Motion sequences of persons almost falling from chairs were hand-crafted and added to the dataset. Background segmentation, noise reduction and thresholding techniques were used to extract binary and depth silhouettes. Spatio-temporal points were calculated on these silhouettes.

On the second phase, we experimented with the (newly added at the time) frontal RGB videos, similarly to the experiments conducted in [27]. This was an attempt to fully evaluate the capabilities of the proposed method in environments where background segmentation is a non-trivial, error-prone procedure. For this reason, no human silhouette segmentation was performed and the experiments relied only on the spatio-temporal information from the subjects motion inside the video. This part of the dataset was used uncropped, i.e. with each video containing a full set of human actions such as entering a room, walking inside and then performing an intentional or unintentional fall. Activities closely resembling falls (not hand-crafted) were added to this set, such as crouching under a sofa, lying on a bed, bending to tie shoelaces, etc.

The Le2i Fall dataset has been captured in realistic video surveillance settings using a single RGB camera. The frame rate is 25 frames/s and the resolution is 320x240 pixels. The video data illustrates the main difficulties of realistic video sequences that can be found at an elderly home environment, as well as in a simple office room. The video sequences contain variable illumination, and typical difficulties like occlusions or cluttered and textured background. The actors performed various normal daily activities and falls. The dataset contains 130 annotated videos, with extra information representing the ground-truth of the fall position in the image sequence. The database provides different locations for testing and training, while authors in [5] have defined several protocols for the evaluation of their method. Working with the specific dataset, we have followed the protocol P1 given in the above paper, where training and test sets are built with videos from ”Home” and ”Coffee room” subsets. Samples from Le2i dataset are

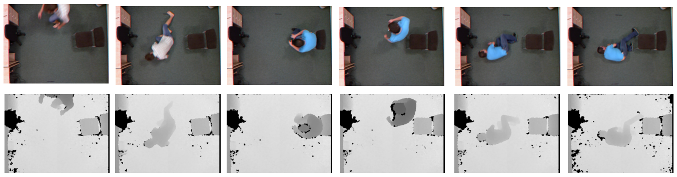

Fig. 5. Frame samples taken from the UR fall dataset for two falls. Upper row illustrates the RGB samples while the lower row provides the corresponding depth images.

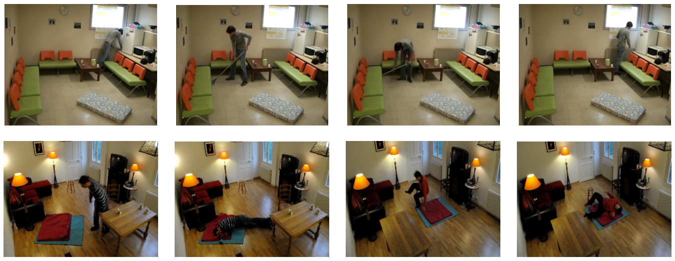

Fig. 6. Frame samples taken from the Le2i fall dataset. Upper row illustrates samples from daily actions in ”Coffee room” while lower raw provides samples from a fall occurred in ”Home”.

provided in Figure 6.

Both in the first experimental scenario on the UR dataset and the Le2i dataset, extraction of feature vectors using the proposed scheme has been preceded by human silhouette extraction. In the UR dataset case, this was handled by computing differences between depth pixels in a particular frame and their corresponding pixels in a precalculated reference frame. The reference frame was calculated by computing the median value of every depth pixel in a sliding window of 9 frames, in a total of 80 frames portraying a scene lacking human presence. Then, the mean value of every median pixel value was calculated, forming the final reference frame and eliminating a considerable amount of noise generated by the depth camera.

One can correlate the human presence in a particular frame with the occasions when the difference between depth pixel values of that frame and the reference exceeds a predefined threshold. In the case of the UR dataset, a total of four thesholds were used, to add robustness. The first two were used to filter out noisy and invalid pixels. For a pixel value to be valid (i.e. possible to be part of a human silhouette), it was required to be between 1100 and 3620 millimeters. It should be reminded that this represents distance from a ceiling mounted depth camera, such as Kinect, whose depth map values are measured in millimeters, in the $[800, 4000]$ range. Afterwards, to indicate human movement, the difference between a pixel of the current frame and a reference pixel was required to be between 50 and 2200 millimeters. These values were found to offer maximum tolerance against random noise.

Given the fact that it consisted of RGB video files of low resolution, the Le2i dataset was handled in a different way. Furthermore, light conditions in most of the cases (especially in the ”Home” subset) rendered the use of the difference between frames unreliable. In order to segment the human silhouette, the background-foreground segmentation approach proposed by Zivkovic in [28] and [29] was utilized. In this technique, a subtraction between the current frame and a background model is performed. This model is constantly updated in a per-pixel fashion, using a gaussian mixture-based approach, to contain what is considered the static part of the scene, adapting in scene changes in the video sequences.

In our experiments, sequences of both datasets have been scaled down to the spatial resolution of 320*240 pixels and have a temporal length of 26 and 12 frames on average for the UR and the Le2i dataset respectively. For the ceiling mounted camera scenario on the UR dataset and the Le2i dataset, training and testing samples were constructed by manually cropping the motion sequences to contain only the fall or fall-like activity part. The feature vector extraction pipeline used in all fall detection scenarios is described in Figure 4.

At this point, we should mention that there is no unified standard to follow for the evaluation of fall detection algorithms. In the experiments conducted in this study, a simple leave-one-out protocol was used to evaluate performance. Lack of different activity samples performed by distinguishable persons led to adapting the original leave-one-person-out protocol to a simplistic leave-one-sample-out protocol, as mentioned earlier. In every iteration, a different activity sample, regardless of whether it depicts a fall or not, is used for testing a system that is previously trained using the rest of the dataset. The results of this experimental procedure can be found in the next subsection (III-B).

B. Results

As can be seen in Table I, the proposed pipeline achieves results comparable to the state of the art techniques tested on the UR and Le2i datasets. These methods, especially the ones presented in [26] and [5], seem to be quite domain specific and rely on certain attributes of the human silhouette shape and bounding box, with a strong relation to the camera placement. This is in contrast with the 3D CTT based pipeline, which is a generalized feature extraction method that operates regardless of sensor position.

The property of time-sensitivity is showcased in these results, especially in the UR dataset [27], where there are many cases of fall-like activities that generated similar spatio-temporal volumes with actual falls. The key discriminating factors in such actions are the temporal interlinking between various stages of the action, the abruptness of position change and the speed. The proposed scheme appears to make use of these attributes.

What is also noteworthy is the ability of the 3D CTT based pipeline to assert the existence of an unintentional fall in RGB videos that contain other activities. The accuracy reached on this task is 95.71% and is achieved based solely on image data. In comparison, the technique proposed in [27] achieves 90% accuracy. When accelerometric data is utilized, this score rises to 98.33%. Another notable fact is the potential of the proposed pipeline to classify an event as a fall on automatically segmented human silhouettes, in arbitrarily positioned camera systems. In this task, the 3D CTT based pipeline scores 96.34% accuracy. In the same setup, the method presented in [5] achieves 95.06% accuracy, possibly minimizing the global error to 2.5%, by manually annotating the human bounding box in the scene.

TABLE I
CLASSIFICATION ACCURACY (%) ACHIEVED BY THE PROPOSED SCHEME AND OTHER PUBLISHED METHODS ON THE FALL DETECTION TASK.

	UR		Le2i
	Ceiling mounted	RGB frontal	
Our pipeline on binary shil.	**100**	**-**	**96,34**
Our pipeline on depth shil.	**95.92**	**-**	**-**
Our pipeline on RGB data	**-**	**95.71**	**-**
Kepski and Kwolek [26]	100	-	-
Kepski and Kwolek [27]	-	90	-
Charfi et al. [5]	-	-	95.06

IV. CONCLUSION

In this paper, we propose the use of a newly formulated extension of the Trace transform to the 3D space, the so-called 3D Cylindrical Trace Transform, and a novel feature extraction scheme from spatio-temporal interest points, for the task of unintentional human fall detection. Using this pipeline, action videos can be transformed into vectors of small length, which represent distortion and occlusion invariant, as well as time-sensitive features. Experimental results on two different and challenging datasets, on a variety of conditions (input data, camera placement, etc.) indicated that the method has great potential. The features created appear to be very robust in noise, illumination variation, occlusion, translation and scaling issues while at the same time the method provides the ability of adaptation to various assistive environment settings.

ACKNOWLEDGMENT

This work was supported by the 4-year EC funded project iRead (Jan 2017-Dec 2020). This project has received funding from the European Union's Horizon 2020 innovation programme under grant agreement No 731724.

REFERENCES

[1] B. Chakraborty, M. Holte, T. Moeslund, J. Gonzalez, Selective spatio-temporal interest points, Computer Vision and Image Understanding 116 (3) (2012) 396–410.

[2] I. Laptev, On space-time interest points, Int. J. Comput. Vision 64 (2005) 107–123.

[3] G. Ting, J. Woo, Elder care: is legislation of family responsibility the solution?, Asian J. Gerontol. Geriatr. 4 (2009) 72–75.

[4] H. Shirlena, T. L. Leng, T. Mika, Transnational mobilities for care: Rethinking the dynamics of care in Asia, Global Networks 12 (2) (2012) 129–134.

[5] I. Charfi, J. Miteran, J. Dubois, M. Atri, R. Tourki, Definition and performance evaluation of a robust svm based fall detection solution, in: SITIS'12, 2012, pp. 218–224.

[6] R. Igual, C. Medrano, I. Plaza, Challenges, issues and trends in fall detection systems, BioMedical Engineering OnLine 12 (1) (2013) 66.

[7] C. N. Doukas, I. Maglogiannis, Emergency fall incidents detection in assisted living environments utilizing motion, sound, and visual perceptual components, Information Technology in Biomedicine, IEEE Transactions on 15 (2) (2011) 277–289.

[8] M. Mubashir, L. Shao, L. Seed, A survey on fall detection: Principles and approaches, Neurocomputing 100 (2012) 144–152.

[9] J. T. Perry, S. Kellog, S. M. Vaidya, J.-H. Youn, H. Ali, H. Sharif, Survey and evaluation of real-time fall detection approaches, in: High-Capacity Optical Networks and Enabling Technologies (HONET), 2009 6th International Symposium on, IEEE, 2009, pp. 158–164.

[10] N. Noury, P. Rumeau, A. Bourke, G. Olaighin, J. Lundy, A proposal for the classification and evaluation of fall detectors, IRBM journal of Alliance for engineering in Biology an Medicine 26 (6) (2008) 340–349.

[11] T. Lee, M. A., An intelligent emergency response system: preliminary development and testing of automated fall detection, Journal of telemedicine and telecare 11 (4) (2005) 194?198.

[12] B. U. Töreyin, Y. Dedeoğlu, A. E. Çetin, HMM based falling person detection using both audio and video, in: Proceedings of the 2005 International Conference on Computer Vision in Human-Computer Interaction, ICCV'05, Springer-Verlag, Berlin, Heidelberg, 2005, pp. 211–220.

[13] C. Rougier, J. Meunier, A. St-Arnaud, J. Rousseau, Fall detection from human shape and motion history using video surveillance, in: AINA Workshops (2), IEEE Computer Society, 2007, pp. 875–880.

[14] H. Foroughi, H. Yazdi, H. Pourreza, M. Javidi, An eigenspace-based approach for human fall detection using integrated time motion image and multi-class support vector machine, in: Intelligent Computer Communication and Processing, 2008. ICCP 2008. 4th International Conference on, 2008, pp. 83–90.

[15] G. Mastorakis, D. Makris, Fall detection system using Kinect's infrared sensor, Journal of Real-Time Image Processing 9 (4) (2014) 635–646.

[16] R. Planinc, M. Kampel, Introducing the use of depth data for fall detection, Personal Ubiquitous Comput. 17 (6) (2013) 1063–1072.

[17] A. Nghiem, E. Auvinet, J. Meunier, Head detection using Kinect camera and its application to fall detection, in: 11th International Conference on Information Science, Signal Processing and their Applications, ISSPA, Montreal, QC, Canada, July 2-5, 2012, pp. 164–169.

[18] Z. Zhang, W. Liu, V. Metsis, V. Athitsos, A viewpoint-independent statistical method for fall detection (2012).

[19] G. Goudelis, K. Karpouzis, S. Kollias, Exploring trace transform for robust human action recognition, Pattern Recognition 46 (12) (2013) 3238–3248.

[20] S. R. Deans, The Radon Transform and Some of Its Applications, Krieger Publishing Company, 1983.

[21] A. Kadyrov, M. Petrou, The Trace transform and its applications, IEEE Trans. Pattern Anal. Mach. Intell. 23 (2001) 811–828.

[22] G. Goudelis, G. Tsatiris, K. Karpouzis, S. Kollias. (2015, July). Fall detection using history triple features. In Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments (p. 81). ACM.

[23] A. Averbuch, Y. Shkolnisky, 3d fourier based discrete radon transform, Applied and Computational Harmonic Analysis 15 (1) (2003) 33 – 69.

[24] C. Yuan, X. Li, W. Hu, H. Ling, S. Maybank, 3d R transform on spatio-temporal interest points for action recognition, in: IEEE Conference on Computer Vision & Pattern Recognition (CVPR), 2013, pp. 724–730.

[25] G. Goudelis, G. Tsatiris, K. Karpouzis, S. Kollias, 3D Cylindrical Trace Transform based feature extraction for effective human action classification. In 2017 International Conference on Computational Intelligence and Games (CIG), IEEE.

[26] B. Kepski. M., Kwolek, Fall detection using ceiling-mounted 3d depth camera, in: Proc. 9th Int. Conf. on Computer Vision Theory and Applications (VISAPP), 2014, pp. vol. 2, 640–647.

[27] B. Kwolek, M. Kepski, Human fall detection on embedded platform using depth maps and wireless accelerometer, Computer Methods and Programs in Biomedicine 117 (3) (2014) 489–501.

[28] Z. Zivkovic, Improved adaptive gaussian mixture model for background subtraction, in: Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04) Volume 2 - Volume 02, ICPR '04, 2004, pp. 28–31.

[29] Z. Zivkovic, F. van der Heijden, Efficient adaptive density estimation per image pixel for the task of background subtraction, Pattern Recogn. Lett. 27 (7) (2006) 773–780.