Real-time recognition of suicidal behavior using an RGB-D camera

Bo Li[1], Wassim Bouachir[2], Rafik Gouiaa[3] and Rita Noumeir[4]
[1 3 4] École de technologie supèrieure
e-mail:bo.Li.1@ens.etsmtl.ca
[2] LICEF research center, TÉLUQ University, Montréal (QC), Canada
e-mail: rita.noumeir@etsmtl.ca, wassim.bouachir@teluq.ca , rafik.gouiaa.1@ens.etsmtl.ca

***Abstract*— Inmates in solitary confinement may attempt to harm themselves in many ways, resulting in trivial to mortal injuries. In this context, suicide by hanging is one of the major causes of death among the incarcerated. The Rapid detection of suicide can reduce the mortality rate. Recently, several technologies have been developed to detect suicide by hanging attempts, but most of them use bulky devices, or they are greatly depending on human attention. In this paper, we propose a computer vision based system to automatically detect suicide by hanging attempts. Our method is based on modeling suicidal actions using pose and motion features, by exploiting the body joints' positions. The proposed video surveillance system analyses depth images provided by an RGB-D camera to detect the event of interest in real-time, regardeless of illumination conditions. The experimental results obtained on a realistic dataset demonstrated the high precision of our system in detecting suicide by hanging.**

***Keywords*— Suicide detection, video surveillance, Kinect, depth images.**

I. Introduction

Suicide attempts have been prevalent within correctional settings. In Canada, suicide has been documented as the leading cause of unnatural death among federal inmates between 1994 and 2014 [1]. On the whole, suicide attempts tend to occur by hanging using bedding, shoelaces or clothing, when prisoners are being held in isolation, and during the night or the weekend when the number of staff in service is minimal.

With the increasing use of technology, video surveillance systems (e.g CCTV systems) have been established as an alternative to the direct observation used by the security staff to monitor actively suicidal inmates. However, camera blind spots coupled with busy camera operators can severely limit the efficiency of such systems. In fact, nearly one-third of the attempts continue to occur in full view of camera equipment, which raises questions about the feasibility of vision-based solutions for suicide prevention [2]. In the literature, different other measures have been taken to automatically detect and prevent hanging events. For instance [3], a system based on a series of sensory strips on the floor and bed of the jail cell has been invented. This system works under the principle of "weight off" and an alarm would be triggered if the inmate is not laying on his bunk or standing on the floor. This invention did not really succeed presumably because many victims committed suicide in either the standing or sitting position on the floor. Moreover, other solutions have been used such as special protective clothes (Safety smocks and blankets) to be worn by actively suicidal inmates [3], a top door alarm [4] which alerts the correctional staff if the door is used as a ligature point and suicide-resistant jail cells [5]. These systems require either wearing a bulky equipment or they are dedicated to particular cases. In addition, if the inmate simply removes the equipment (bracelet, clothes etc), a false alarm would probably go off and an emergency response would also be called.

Recently, with the increasing progress made toward camera-based human action recognition, researchers tend to improve the existing video-surveillance systems (e.g CCTV systems) by automatically monitoring and preventing suicide attempts without the camera operator's intervention. In this context, Lee et al [6] proposed a system based on analyzing 3D images captured by an Asus Xtion Pro camera to detect a suicide by hanging event. Nevertheless, this is a preliminary and limited study, where only the case of a partial suspension hanging was considered, without addressing real-world scenario difficulties such as occlusion and scale change. Besides, only a few video samples (150 samples) with short duration of 3 seconds each one, have been used to perform the experiments. Lately, our previous work [7] presented an intelligent video surveillance system for automated detection of suicide by hanging attempts. Unlike in [6], we considered a large dataset captured by an RGB-D camera where 21 persons are invited to perform different scenarios for unsuspected behavior and suicide attempts. In addition, real-world settings difficulties such as partial occlusion and scale change have been considered. With the invisible illumination principle used by a Kinect camera, our system can operate day and night without bothering inmates. Although our algorithm achieved a high accuracy rate, it had difficulties to correctly classify some activities of daily living such as wearing or removing pieces of clothing, which triggers false alarms in a real scenario.

This paper presents an extension of our previous work [7], where significant improvements are achieved.

- A more efficient scaling method is proposed to alleviate the effect of morphological difference within candidates and keep the features invariant with respect to people.
- A feature selection approach is applied in order to speed up our algorithm for real-time application and improve its generalization capacity.

- An efficient implementation is elaborated for meeting the real-time requirements of suicide detection application.

We review recent work on action recognition in section 2. In section 3, we describe the proposed method for suicide by hanging detection. Experimental results are provided in section 4, and we finalize by a conclusion in section 5.

II. Related works

Human activity recognition from videos serves as a mandatory prerequisite step for several computer vision applications including patient monitoring systems, ambient assisted living systems, and especially intelligent surveillance systems. A traditional human activity recognition system that uses input information from a camera operates typically on two main steps: 1) feature extraction which consists on extracting visual cues that are relevant with respect to human activities, 2) action learning and classification which is based on learning statistical models from the extracted features, and using them to classify new feature observation. According to the type of extracted features, previous work on human activity recognition can be generally divided into three categories [8]: The first category uses global models which focus on detecting the full visible human body, without the detection and labeling of individual body parts, using background subtraction or tracking techniques. Different types of features such as dense or sparse optical flow [9], silhouettes [10] or contours [11] are usually used for representing the localized parts of body. Most approaches in this category operate on the whole human body and do not explore how to adapt global representations to deal with challenging cases such as viewpoint changes, partial occlusion, appearance variations and camera movement, which limit their performance on a real scenario.

Methods in the second category use local representations of activities instead, where the image/video is decomposed into small regions (patches), regardless of the body parts annotation, illumination changes or body localization. These small patches catch the regions of high variations in time and spatial domains and involve appearance and/or motion information. Based on this, Space-time interest points [12] were derived to generalize the interest points and local descriptors [13] and apply them to the case of activities recognition from videos. Despite their effectiveness to overcome some global representation limitations such as sensitivity to noise, background subtraction defaults, and partial occlusion, such approaches still have difficulties in analyzing complex human activities because of the limited semantics they represent [14].

The methods in the third category adopt model-based pose estimation approaches to represent a human activity as a sequence of poses in time, by employing the spatial configurations of human body articulations. This representation follows the principle, in [15], describing how humans observe actions. This work demonstrated that humans can easily recognize activities from the motion of a few human body joints. Based on this assumption, several algorithms have been proposed to address the human body joints localization. For instance, [16] and [17] both proposed methods for joints localization and human activity recognition using images captured from an uncontrolled RGB camera. However, pose-based activity recognition can be very challenging because of the complexity of estimating high quality poses from RGB action videos, except in special cases (e.g static and calibrated cameras and simple backgrounds), which are still computationally expensive tasks. Pose-based activities recognition is still an active research area, which requires significant improvement to overcome the mentioned limitations.

In this work, we are particularly interested in techniques of the latter category, as pose-based approaches are more suitable to represent complex real life activities. However, to deal with these approaches' limitations, we propose to use images captured by cost effective RGB-D cameras. Such devices provide the 3D spatial information on human body, in addition to the colored image of the scene. Moreover, pose estimation can be achieved in real-time under various illumination conditions. The proposed method is detailed in the next section.

III. Proposed method

The additional information provided by RGB-D cameras open an alternative line of work to tackle the human pose estimation problem in real time, and therefore dealing with human activity recognition. Our method relies on exploiting the relative distance between human joints' position in 3D spaces to extract pose and motion features. To detect suicidal behavior, features corresponding to the current observation is fed into a machine learning model to perform a binary classification. A suicide by hanging attempt is finally announced if the percentage of positive observations exceeds a certain threshold during a sliding temporal window.

A. Pose representation and analysis

We represent the human body as an articulated structure consisting of various segments connected by joints. Thus, a suicide attempt is considered as the temporal evolution of joints' spatial configuration. Using a depth camera, RGB-D data facilitates the identification of the 3D locations of joints which can be easily obtained in real time as proposed in [18]. In this algorithm, a per-pixel classification is firstly performed to infer the body parts from a single depth image, where the parts defined to be spatially localized near skeletal joints of interest. The inferred parts are then reprojected into the world space to generate spatial modes specifying proposals for the 3D locations of each skeletal joint. Finally, a mean shift is applied on these modes to generate the 3D position of joints. In our method, this algorithm is applied as the first stage of the pipeline to obtain the 3D joints' locations of the human body. In view of the fact that the lower body parts seems to be irrelevant for detecting a suicide by hanging actions, the tracking of the upper joints movement is only considered. Based on this, we consider a subset of $N = 16$ upper body

joints (See Fig. 1). For each frame t, the 3D joint coordinates are noted as:

$$X_t = \{J_t^i = (x_i, y_i, z_i) | i = 1 \cdots N\} \quad (1)$$

Where (x_i, y_i, z_i) are the 3D coordinates of the ith joint J at time t which is noted as J_t^i.

B. Pose and motion features

Generally, a suicide by hanging attempt involves the action of placing a strangling object around the neck, which typically requires to move the hands to the neck from top to bottom. Based on this, we exploit the extracted joints position to derive two different features vectors as follows:

- $P_t = \{dist(J_t^i, J_t^j) | i, j = 1 \cdots N; i \neq j\}$, which is pairwise disjoint distances between the list of joints used to describe the pose at frame t.
- $M_t = \{dist(J_{t-1}^i, J_t^j) | i, j = 1 \cdots N\}$, which is pairwise distances between the list of joints in frame t and frame (t-1) used to describe the motion performed between 2 subsequent frames.

For each frame t, we combine both subsets P_t and M_t having respectively $C_N^2 = 120$ and $N^2 = 256$ components in a single 376-dimensional vector $F_t = (P_t, M_t)$. We mention that these features vectors are invariant to the rotation, since we perform pairwise comparisons instead of directly using joint positions (in the 3D camera coordinate system). Furthermore, we normalize these distances with respect to the distance between the neck and the spine middle joints in order to avoid the variation in morphology of persons and achieve the scale invariance. In the next section, we detail how to determine the scaling parameter.

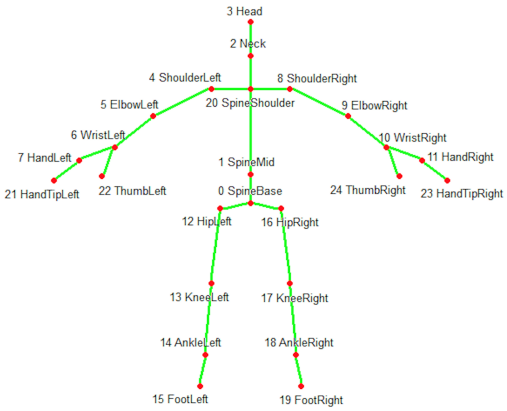

Fig. 1: Joints given by the method of Shotton et al. [18].

C. Scaling parameter estimation

It is straightforward that distances between the 3D joints position within one frame or across multiple frames depend on people morphology. In fact, for people with large body size the relative distance between shoulder and elbow is greater than that of people with small body size. Such variation is undesirable in designing a based-vision system for action recognition, as it leads to scaling variant features perturbing the system. To overcome this limit, we propose to normalize the extracted features and remove the influence of body size. The normalization process is achieved by considering the distance from neck to spine middle. Three main reasons can explain our choice:

- The distance between these two joints is proportionate to the person's height.
- For one specific person, the distance between the neck and the spine middle is stable and does not dramatically changed with respect to his current pose.
- The neck and middle spine joints are always observable for kinect cameras.

The joint positions provided by the method of Shutton belongs to one of the two following states: "Tracked" or "Inferred". "Tracked" means that joint is observable by the camera and its position is determined directly from the current frame , while "inferred" means that the joint is not observable but its position can be inferred from historical measurements using tracking algorithms. Fig. 2 shows the percentage of the "tracked" state of the upper body joints. We notice that three joints, which are head, neck and spine middle are always observable. This gives us an heuristic basis for choosing the distance between the neck and the spine middle as a scaling parameter noted s. This scaling parameter can be easily estimated from

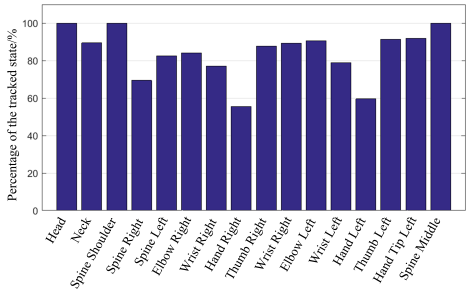

Fig. 2: Percentage of tracked instances for each joint.The total number of frames is 51003.

one frame as the distance between the neck and the spine middle. However, with multiple frames, many distance values are sequentially generated and the parameter scaling can be given by the median value, which is robust to outliers. For this, we implement an algorithm based on Min-Max heaps for estimating the median value in a constant time. Once the scaling parameter is available, current sample is scaled as:

$$F_t = \frac{F_t}{s}. \quad (2)$$

D. Feature selection and learning

Human activity recognition methods are typically based on applying a machine learning model on multi-dimensional features. However, data with high dimensionality has presented serious problems for the existing machine learning

models, i.e the curse of dimensionality [19]. The presence of a large number of features leads to the over-fitting problem and consequently affect the generalization capacity of the learning model. To deal with the problem of the curse of dimensionality, dimensionality reduction approaches have been explored, and become an active research topic within the machine learning and data mining community. In the literature, feature selection is one of the most techniques employed for reducing dimensionality. It aims to find the best subset of features from the original ones according to certain relevance criterion, which usually leads to higher generalization performance, lower running time, and better model interpretability. According to the learning style, (e.g supervised learning, unsupervised learning and semi-supervised learning), different methods have been proposed for features selection [20]. Supervised feature selection techniques can further be broadly divided into filters, wrappers and embedded techniques. In this work, we are interested in methods of the first category due to their simplicity and computational efficiency. In this case, a best subset of features are selected according to the score attributed to individual features using a scoring function such as correlation coefficients or mutual information criteria. However, selecting feature using only a scoring function can lead to rich redundancy, i.e these features are highly dependent. A popular filter method that takes into account the redundancy among the selected features is the Minimum Redundancy Maximum Relevance (mRMR) [21]. Using this algorithm, the subset features $\{x_j|j = 1 \cdots m\}$ are sequentially selected as follows:

$$\max_{x_j \in \Lambda \setminus \Gamma} = \left[I(x_j; c) - \frac{1}{|\Gamma|} \sum_{x_i \in \Gamma} I(x_i; x_j) \right| i = 1 \cdots |\Lambda|] \quad (3)$$

where $I(x_j; c)$ is the mutual information value between individual feature x_i and class c, $I(x_i; x_j)$ is the mutual information between two features x_i and x_j, Γ is the subset of the best features, Λ is the original set of features with $|\Lambda|$ cardinality. The idea behind the mRMR is to sequentially select a feature that is relevant with respect to the target class c (max-relevance) and simultaneously has a low dependence to the already selected features in Γ (min-redundancy). Both criterion are evaluated based on the mutual information.

Dimensionality reduction is very important allowing us to speed up our algorithm for a real time detection of a suicide by hanging attempt as well as to deal with the redundancy caused by the high frame rate (30 frames/second). Thus, Once the feature selection process is applied, we obtain a new observation $\hat{F}_t$ at time t, with a smaller dimension (< 376).

In our system, recognizing an activity of interest requires applying a binary classification on a single observation $\hat{F}_t$ at time t to decide whether it is a *'suicide'* attempt or *'unsuspected'* behavior. For this end, we construct a Linear Discriminant Analysis (LDA) classifier using the calculated features $\hat{F}_t$ as the classification model variables. The feature set of each class is thus modeled as a multivariate normal distribution with a common covariance matrix and 2 different mean vectors. These parameters are estimated from labeled data during the training phase. The flowchart of the training procedure is depicted in Fig. 3. In real-time detection, new observations obtained from singles frames are assigned to the class having the nearest mean vector according to the Mahalanobis distance.

Algorithm 1 On-line suicide by hanging detection

Input: depth frame t
Output: decision result
Assumption: processing frame t with $t >= 1$
Initialization: detected = false; $\theta_t = 0$;

While detected==false **do**
- Estimate pose
- Compute joint locations
- Calculate pose feature vector P_t
- Calculate motion feature vector M_t
- $F_t = [P_t, M_t]$
- Normalizing F_t with the scaling parameter s
-Feature selection: $\hat{F}_t$
- Classify $\hat{F}_t$ using LDA
- Update θ_t
if $\theta_t \geq \theta_{min}$ **then**
    - $detected = true$
else
    - Shift temporal window by S
    - Retrieve frame t + 1
end if

E. Activity recognition

Activity recognition is carried out using the procedure summarized by the Algorithm 1. Once a trained LDA classifier is obtained, the recognition algorithm process a stream of depth images to detect whether a suicide by hanging attempt is underway. First, the body joint positions is estimated using the underlying algorithm [18]. The 3D upper body joints' position are employed to calculate the pose and motion feature vector F_t. F_t is then scaled and a feature selection algorithm is applied to generate the current local observation $\hat{F}_t$. These local observations are sequentially classified as positive or negative observations using the trained LDA. Suicide detection is based on analyzing the person's behavior during a sliding temporal window of width Δ_0., which is regularly shifted by the temporal step S at each iteration. Finally, a suicide by hanging attempt is detected and an emergency call is triggered if the percentage of positive observations θ_t exceeds a threshold θ_{min}.

IV. Experiments

A. Dataset

Since there is no publicly available datasets for suicide by hanging recognition, we created our own dataset for evaluating the designed system. 21 persons participated to perform

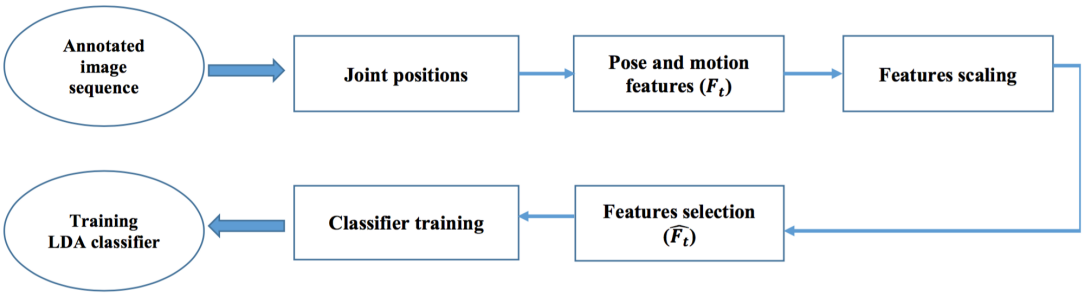

Fig. 3: A flowchart of the offline training procedure.

various actions in a room where dimensions are close to those of a prison cell. Our system consists of a Kinect v2 camera placed in an upper corner, at a distance of approximately 0.30 m from the ceiling, with a tilt angle of 35 °. For efficient body pose estimation, we suppose that the distance between the candidate and the camera is in [0.5 m, 4.5 m]. A video dataset was created, where each participant was asked to perform two scenarios:

- During the first scenario, the participant simulated a suicide attempt by hanging respectively in three steps: 1) he/she created a hangman's knot using a bed sheet (each participant can create the knot in a different way), 2) he/she attached the knot to a fixed point in the room, and finally he/she placed the knot around his/her neck
- For the second scenario, the participant asked to show some unsuspected activities from different angles with respect to the camera such as sitting, removing a piece of clothing, moving in the room, wearing clothes etc. Our dataset including 42 video sequences was used to train and evaluate the proposed system. The dataset is available from the author upon request.

B. Results

To evaluate the effectiveness and robustness of our system, we performed experiments on our dataset as follows: we randomly split the available videos into training and testing set. 32 sequences are used to train our system, while the remaining (10 sequences) are employed for the testing phase. All experiments presented in this section were carried out on a 3.6 GHz Core i74790 CPU using MATLAB R16. A real-time version has been developed using C# and the Kinect for windows SDK 2.0.

We applied the off-line training procedure depicted in Fig. 3 on the training set in order to extract feature vectors and generate the trained LDA classifier. Note that different simple classifiers such as LDA, Linear-SVM, bayes-naive are tried, and LDA was selected among these based on its performance on the training set. To evaluate the trained LDA, we used the algorithm 1 where its parameters were empirically fixed. The sliding temporal window Δ_0 was fixed empirically to 5 seconds and shifted at each iteration by $S = 0.07$ seconds. A suicide attempt was detected if the current threshold θ_t exceeds $\theta_{min} = .7$ for the current position of the sliding window. In addition, the optimal results have been reached empirically using only 100 features selected by the mRMR algorithm (see section III-D). Table 1 summarizes detection results for the 5 videos sequences of suicidal behavior, using the best 100 features and all 376 features. The suspected activity consists on attaching the knot and placing it around the neck. Videos were simulated by different persons that are asked to differently place the hangman's knot and change their body orientation. This allows us to evaluate the ability of our algorithm to capture the high variability in the data.

As depicted in Table 1, suicide attempts were correctly recognized in all sequences using only 100 features. Detection was achieved almost in the first few frames. In sequence 4, the detection of the suicide attempt was done in the last frame. This is caused mainly by the body's orientation with respect to the camera angle view. In fact, the participant was filmed from the side view which causes the occlusion of the upper limb and thereby affecting the joints' localization. We consider that even a delay in the detection of the suicide attempt, thus a delay in the alarm, shouldn't hinder the process of reactive intervention. Indeed, the most common cause of a death follwing a suicide by hanging is the occlusion of blood vessels and/ or airways, which takes a few minutes once the knot is sufficiently tightened around the neck.

Table 2 shows the detection results for the 5 sequences of normal (unsuspected) behavior using the best 100 features and all 376 features. The aim of this experiment is to evaluate our algorithm for recognizing some daily activities, similar to suicide attempt, requiring to move the hands around the neck. We therefore asked participants to wear or remove pieces of clothing during unsuspected scenarios. Using only 100 features, the recognition results indicate that our algorithm correctly recognized all scenarios.

As a comaprison test and to explore the effectiveness of feature selection technique in our case, we repeated the above two experiments using all features (without appliying feature selection technique). Results is illustrated in Table 1 and Table 2. According to Table 1, suicide attempts were correctly recognized as well using 376 features. Nevertheless, in this case, suicide was relatively detected late twice, in scenario 2 and 4. In addition, in Table 2, The detection results include a

Table 1: Recognition results for sequences where suicide is simulated. We present a comparison between the results obtained with feature selection (100 best features) and those using the entire feature set (376 features). For each video, the table presents: the start time of the suspected activity (Start), end time (End), detection result (Yes/No), and the detection time (Time). Times are expressed in seconds.

Video	Duration (S)		Detection (S)		
	Start	End	Yes/No	Time (S)	
				100 features	376 features
1	50	68	Yes	52	52
2	21	34	Yes	26	34
3	22	32	Yes	26	26
4	82	92	Yes	92	92
5	13	24	Yes	18	18

Table 2: A comparison between recognition, results where unsuspected actions are considered, between the best 100 features and all 376 features. For each video sequence, the table shows the detection result. In the case of false detection, the detection time is indicated in seconds.

Video	False alarm (Yes/No)		Time (S)	
	100 features	376 features	100 features	376 features
1	No	No	-	-
2	No	No	-	-
3	No	No	-	-
4	No	Yes	-	44
5	No	No	-	-

single false alarm by comparison with results using only 100 features. Based on these experiments, we can conclude that in addition to time complexity reduction, meeting the real-time requirements, the applied feature selection technique increased the classification accuracy by eliminating noisy features.

V. Conclusion

We presented an intelligent surveillance system for detecting suicide by hanging attempts in prisons. Our algorithm relies on using 3D joint's locations acquired in real-time by an RGB-D camera. Both pose and motion features are considered for representing the event of interest. A feature selection technique is applied to speed up our algorithm and improve its generalization. Once the system is trained, we apply an effective online algorithm for detecting suicide attempts from singles images. We achieved a high accuracy and a 100% sensitivity on a challenging data set.

Our future work will focus on improving the capacity of the proposed algorithm for detecting suicide attempts in shorter time. This can be achieved by modeling the interaction of the human's body with the strangling object and the early detection of the action of creating a knot. This can be done by combining the depth and infrared images for detecting the knot.

Acknowledgment

This work was supported by research grants from the Natural Sciences and Engineering Research Council of Canada, MITACS, and an industrial funding from Aerosys- tems International Inc. The authors would also like to thank their collaborators from Aerosystems International Inc.

References

[1] Correctional investigator Canada, "A three year review of federal inmate suicides (2011-2014)," 2014.

[2] O. of the Correctional Investigator of Canada, "A Three Year Review of Federal Inmate Suicides (2011-2014)," Tech. Rep., September 10 2014.

[3] L. M. Hayes, "Suicide prevention in correctional facilities: Reflections and next steps," *International journal of law and psychiatry*, vol. 36, no. 3, pp. 188–194, 2013.

[4] F. E. Cook, "Door suicide alarm," Dec. 6 2011, uS Patent RE42,991.

[5] R. Reeves and A. Tamburello, "Single cells, segregated housing, and suicide in the new jersey department of corrections," *Journal of the American Academy of Psychiatry and the Law Online*, vol. 42, no. 4, pp. 484–488, 2014.

[6] S. Lee, H. Kim, S. Lee, Y. Kim, D. Lee, J. Ju, and H. Myung, "Detection of a suicide by hanging based on a 3-d image analysis," *IEEE Sensors Journal*, vol. 14, no. 9, pp. 2934–2935, Sept 2014.

[7] W. Bouachir and R. Noumeir, "Automated video surveillance for preventing suicide attempts," *7th International Conference on Imaging for Crime Detection and Prevention (ICDP 2016)*, 2016.

[8] D. Weinland, R. Ronfard, and E. Boyer, "A survey of vision-based methods for action representation, segmentation and recognition," *Computer vision and image understanding*, vol. 115, no. 2, pp. 224–241, 2011.

[9] A. Fathi and G. Mori, "Action recognition by learning mid-level motion features," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.

[10] R. Gouiaa and J. Meunier, "Human posture recognition by combining silhouette and infrared cast shadows," in *2015 International Conference on Image Processing Theory, Tools and Applications (IPTA)*, Nov 2015, pp. 49–54.

[11] S. Cheema, A. Eweiwi, C. Thurau, and C. Bauckhage, "Action recognition by learning discriminative key poses," in *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, Nov 2011, pp. 1302–1309.

[12] X. Yang and Y. Tian, "Action recognition using super sparse coding vector with spatio-temporal awareness," in *European Conference on Computer Vision*. Springer, Cham, 2014, pp. 727–741.

[13] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," *Computer vision–ECCV 2006*, pp. 404–417, 2006.

[14] S. Sadanand and J. J. Corso, "Action bank: A high-level representation of activity in video," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 1234–1241.

[15] G. Johansson, "Visual perception of biological motion and a model for its analysis," *Attention, Perception, & Psychophysics*, vol. 14, no. 2, pp. 201–211, 1973.

[16] V. Ferrari, M. Marin-Jimenez, and A. Zisserman, "Progressive search space reduction for human pose estimation," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, June 2008, pp. 1–8.

[17] M. Andriluka, S. Roth, and B. Schiele, "Monocular 3d pose estimation and tracking by detection," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 2010, pp. 623–630.

[18] J. Shotton, T. Sharp, A. Kipman, A. Fitzgibbon, M. Finocchio, A. Blake, M. Cook, and R. Moore, "Real-time human pose recognition in parts from single depth images," *Communications of the ACM*, vol. 56, no. 1, pp. 116–124, 2013.

[19] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*. Springer series in statistics New York, 2001, vol. 1.

[20] J. Tang, S. Alelyani, and H. Liu, "Feature selection for classification: A review," *Data Classification: Algorithms and Applications*, p. 37, 2014.

[21] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.