

# Depth Vision Guided Human Activity Recognition in Surgical Procedure using Wearable Multisensor

Wen Qi<sup>1</sup>, Hang Su<sup>1</sup>, Fei Chen<sup>2</sup>, Xuanyi Zhou<sup>3</sup>, Yan Shi<sup>4</sup>, Giancarlo Ferrigno<sup>1</sup> and Elena De Momi<sup>1</sup>

**Abstract**—The increasing complexity of modern surgery rooms brings many challenges. Human activity recognition (HAR) plays a significant part in healthcare, telemedicine, long-term treatment, and even surgery by using wearable inertial sensors or depth cameras. Although the development of artificial intelligence techniques provide various machine learning (ML) methods to identify activities, it is a time-consuming implementation and high work burden to collect and label the large data set. To fascinate efficient data collection and labeling, we propose a novel depth vision-guided HAR architecture to obtain the labels of the collected raw data from the inertial measurement unit (IMU) sensors using depth data automatically. Experimental results show that the novel depth vision-guided interface can be utilized for identifying activities without labeling the data in advance.

## I. INTRODUCTION

During complex surgical procedures, the captured sensor data from the operating room (OR) includes multiple information that can be applied for modeling, processing, and predicting surgical workflow. Human activity recognition (HAR) is the most popular technique to observe patients' anatomical structure [1]. In a surgery procedure, monitoring the doctor's behavior provide practical guidance. Consequently, there is a growing interest in the medical community to build a HAR recognized system that provides clinical support from merely presenting the related information to help surgeons make surgical decisions in the complex surgical procedures [2], [3]. To support the demands for the intraoperative, the wearable multisensor system, intelligent presentation of various information sources, and context-specific support of surgeons are widely mentioned for future system application [4]. It provides much useful operating guidance by monitoring the movement of a doctor in real-time during a surgery procedure [5]. Nowadays, by carrying wearable and portable smart devices can identify human activities with a high recognition rate [6]. However, several complex transfer activities are challenging to be identified, such as from sitting to standing.

\*This work was supported by the European Commission Horizon 2020 research and innovation program, under the project SMARTSurg, grant agreement No. 732515.

<sup>1</sup>Wen Qi, Hang Su, Giancarlo Ferrigno and Elena De Momi are with the Department of Electronics, Information and Bioengineering, Politecnico di Milano, 20133, Milan, Italy. wen.qi, hang.su, giancarlo.ferrigno, elena.demomi@polimi.it.

<sup>2</sup>Fei Chen is with the Active Perception and Robot Interactive Learning Laboratory, Department of Advanced Robotics, Istituto Italiano di Tecnologia, Via Morego, 30 16163 Genova, Italy. fei.chen@iit.it.

<sup>3</sup>Xuanyi Zhou is with State Key Laboratory of High Performance Complicated, Central South University, 410012, Changsha, China.

<sup>4</sup>Yan Shi is with School of Automation Science and Electrical Engineering, Beihang University, 100083, Beijing, China. shiyan@buaa.edu.cn.

Using a signal type of sensor cannot provide more effective information to enhance the recognition rate [7]. As an active and comfortable research field, wearable or portable smart devices for HAR are widespread used [8].

In the past decade, many studies have achieved to monitor human activities using wearable sensors or depth cameras. Most of these techniques tried to explore different methods of sensing technologies and extracting features to improve the identification ability of HAR using smartphones. A better hybrid filter and wrapper (HFW) method is proposed in [9] to extract features from accelerometer and gyroscope sensors by comparing with Principal Component Analysis (PCA) [10], Fast Correlation-based Filter (FCBF) [10], Fast Correlation-based Filter (FCBF) [11], and sequential forward selection (Wrapper) [12] methods. Khan et al. [13] proposed Kernel Discriminant Analysis (KDA) to recognize five human activities sampled from the smartphone on five body positions, which obtains a higher classification accuracy than Signal Magnitude Area (SMA) and Linear Discriminant Analysis (LDA). Three motion sensors (linear accelerometer, accelerometer, and gyroscope) at both wrist and pocket positions are applied to recognize less-repetitive human activities, for instance, drinking coffee, eating, and smoking in [14]. The phase trajectory matrix has been constructed, and the PCA algorithm has been applied by extracting features in the first period to split the time series online [15]. The recognition ability is compared in [16] on different extracted features sets, for instance, frequency domain (wavelet transform), time-domain (statistical and mathematical parameters), and discrete domain (symbolic representations). Although several chosen IMU sensors and feature extraction methods provide important information to identify activities, they ignore the affection of artificial movement and changes of direction and position of the smartphone.

Moreover, the artificial intelligence technique provides more methods based on machine learning (ML) or deep learning (DL) to identify more complex activities and increase the classification accuracy [17]. A k-Nearest Neighbor (k-NN) method is proved to have a higher recognition rate than Naïve Bayes (NB) in predicting four types of motions (stand, sit, walk and run) [18]. A mixture-of-experts model is designed in [19] to process incomplete and uncertain data. Reyes-Ortiz et al. [20] proposed a Transition-Aware (TA) HAR system using the Support Vector Machine (SVM) algorithm to classify less than 33 activities while wearing a smartphone or wearable devices in real-time. A Hierarchical Hidden Markov Model (HHMM) method is presented in [21] to recognize human activities (go stairs, standing, walking,

running, and moving arms) as well as some daily activities (taking a bus and shopping). The HHMM method can obtain higher accuracy than the Artificial Neural Network (ANN) and the Hidden Markov Model (HMM). Afterward, a new two-stage continuous HMM algorithm is present to decrease the number of feature subsets for decomposing the complex human activities into some more straightforward activities. The ConfAdaBoost.M1 ensemble algorithm is applied in [22] to obtain a higher prediction precision on recognizing six human activities, while the algorithm needs to extract 561 features. An adaptive slide window approach is presented to classify human activities by calculating the probability of signals, including an adjustable window length. Nevertheless, the long signal segment is brought more misclassification due to the possibility of having multiple human activities. In our previous works, an FR-DCNN method is proposed to recognize 12 complex activities by carrying a smartphone on the waist [23]. Compared with other DL algorithms, such as Bi-LSTM, LSTM, and CNN, the designed FR-DCNN model outperformed by 1% on average. This contribution not only achieves fast computation but also propose a new data compressing method for solving the problem of time-consuming. Subsequently, we present a multisensor based HAR model by using both respiratory and acceleration sensors [24]. However, several types of interference (i.e., object occlusion and artifact) caused by exercise or unstable movement always decrease classification accuracy. Furthermore, using a single sensor, e.g., depth camera or gyroscope, is challenging to label complex activities.

In this paper, we proposed a novel depth vision-guided HAR system using wearable multisensor to monitor the doctor's activities in a surgery room. It includes an automatic labeling method by clustering the depth data, and a recognition approach of processing inertial measurement unit (IMU) signals measured by carrying an IMU sensor on the waist. The depth data are used to identify activities and provide labels of the IMU segments by designing a hierarchical k-medoids (Hk-medoids) clustering algorithm. A multiple layer neural network (MNN) classifier is trained to identify five possible activities in SR, namely sitting, standing, walking, bend over forward, and several transfer motions (e.g., from lying left to the right side). The innovation of this paper is as below:

- 1) A novel autonomous learning framework is presented to integrate the benefits of both depth vision and IMU signals.
- 2) An MNN classifier is utilized to improve accuracy with the practical application during the surgery procedure.
- 3) The designed signal processing approaches based on IMU signals avoid the disturbances from the changes in position and orientation

This paper is organized as follows. Section II introduces the proposed methodology and the designed hardware system. The performance of Hk-medoids and MNN classifier is analyzed in Section III. Section IV discusses the achievements and delineates avenues for further work on this topic.

## II. METHODOLOGY

### A. Depth Vision based Labeling

The captured depth data from the Kinect V2 device is used to a label by the designed three layers Hk-medoids algorithm automatically. It aims to divide the mentioned five complex activities layer by layer by using the 3D joints depth data. All of the 25 nodes data are used to label the activities, which can be expressed as  $V \in \mathbb{R}^{75}$ . After removing the noises by the wavelet denoising approach, the sliding windows strategy is adopted to split the raw data into  $N$  segments, namely  $V = \{v_i\}, i = 1, 2, \dots, N$ , with a fixed detection length  $L_d$  and overlap  $L_o$ . To enhance the accuracy of clustering the dynamic and static activities, we calculate average ( $\bar{v}_i$ ), standard deviation ( $\sigma(v_i)$ ), maximum ( $\max(v_i)$ ), and minimum ( $\min(v_i)$ ) of the depth data to label the five activities based on the Hk-medoids approach.

The k-medoids model is to find clusters (labels) by choosing a set of  $i_k$  members of  $V$  as "medoids." Then, the unselected member of  $V$  is assigned to its closest medoid. The partitioning properties of a valid k-medoids clustering in  $V$  are:

- $v_i \neq \emptyset, 1 \leq i \leq k$ ;
- $\cup_{i=1}^k v_i = X$ ;
- $v_i \cap v_j = \emptyset, i \neq j, 1 \leq i, j \leq k$ .

It needs to distinguish similar and dissimilar objects quantitatively. The whole procedure of the k-medoids clustering with a given k can be written as follows:

- 1) Set a k value as the number of medoids and compute the Euclidean distance to find the nearest medoid
- 2) A common criterion is adopted by swapping the medoid and un-medoid in each step, which is called the sum of intra-cluster variation (SICV) expressed in Eq. 1
- 3) The lowest error is selected as the best configuration
- 4) Repeat the first three steps until there are no more changes

$$SICV = \sum_{i=1}^N \sum_{j=1}^k d^2(v_i, c_j) \quad (1)$$

The cluster centres is contained as  $C = \{c_j\}, j = 1, 2, \dots, k$ .  $c_j$  denotes the  $i$ th cluster centre, and  $d^2(v_i, c_j)$  is the Euclidean distance of  $v_i$  and  $c_j$ .

Algorithm 1 describes the details of the Hk-medoids method. The first layer aims to divide dynamic and static activities by using standard deviation and extrema, namely  $F_1^c = \{\sigma(v_i), \max(v_i), \min(v_i)\}$ . The second layer can separate sitting and standing by comparing average  $F_2^c = \bar{v}_i$ . The last layer aims to divide the rest three dynamic activities based on feature  $F_3^c = \{\sigma(v_i), \max(v_i), \min(v_i), \bar{v}_i\}$ .

### B. Multisensor based classifier for HAR

The proposed multisensor based HAR method aims to train a classifier using the MNN approach. The inputs are IMU signals, while the outputs are the acquired labels  $y^*$  by the Hk-medoids method.

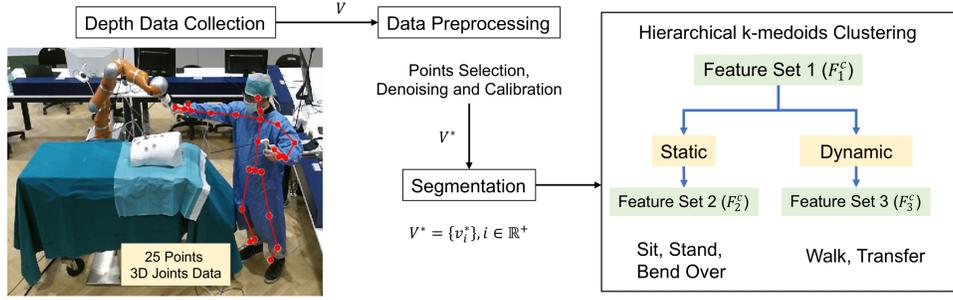


Fig. 1. The flow chart of self-labeling using the designed Hk-medoids clustering method.

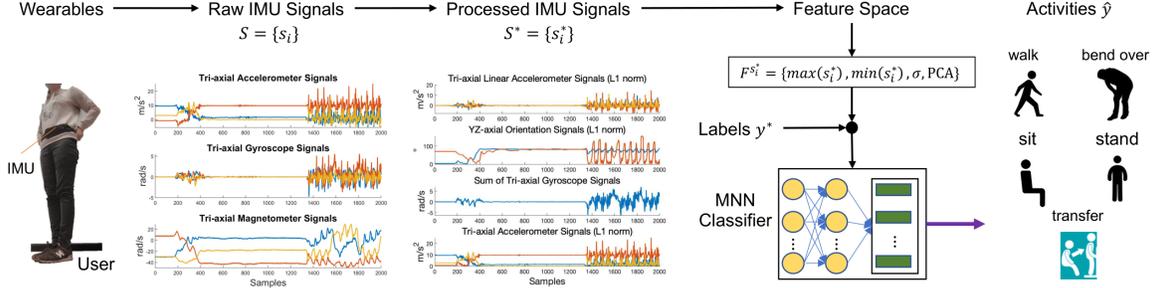


Fig. 2. The data stream procedure of signal processing and modeling for HAR.

### Algorithm 1 Hierarchical k-medoids Clustering.

**Input:** the segments  $v_i$  and the features sets  $F$ ;

**Output:** the labels of segments  $y_i^*$ ,  $i = 1, 2, \dots, N$ ;

- 1: **Layer 1:** divide segments  $v^*$  into dynamic  $v_D^*$  and static  $v_S^*$  classes based on  $F_1^c$ ;
- 2: **Layer 2:** separate the two static activities by  $F_2^c$ ;
- 3: **Layer 3:** divide three dynamic motions based on features set  $F_3^c$ ;

Fig. 2 describes the procedure of signal processing and modeling. The IMU signals are collected from the user who wears the IMU sensor on the waist. To remove some turbulence (e.g., movement artifact and magnetic fields) for accuracy enhancement, we combine a set of signal processing methods as follows.

Firstly, the raw IMU sensors (accelerometer  $S_\alpha$ , gyroscope  $S_\beta$ , and magnetometer  $S_\gamma$ ) are transferred into two derived sensors, namely, linear accelerometer and gravity. We use the linear L1-norm accelerometer  $|S_\alpha|$  to fix the tri-axes of acceleration as follows.

$$\|S_\alpha\| = \sum_{i=1}^{L_d} |s_{\alpha_i}| \quad (2)$$

To acquire the gravity  $|S_\alpha|_g$  and the linear acceleration ( $|S_\alpha|_l$ ) vectors, we implement third-order zero phase Low-Pass Elliptical Filter (LPEF) [25] to decompose  $|S_\alpha|$ .

Secondly, the orientation axes  $S_\Theta$  is computed based on the attitude and heading reference system algorithm (AHRS filter) [26] to improve the identification ability of static activities. It adopts the in-direct Kalman filter models the

error process,  $x$ , with a recursive update as

$$x_t = \begin{bmatrix} S_{\Theta_t} \\ S_{\alpha_t} \\ S_{\beta_t} \\ S_{\gamma_t} \end{bmatrix} = F_t \begin{bmatrix} S_{\Theta_{t-1}} \\ S_{\alpha_{t-1}} \\ S_{\beta_{t-1}} \\ S_{\gamma_{t-1}} \end{bmatrix} + \omega_t \quad (3)$$

$x_t$  is the output at time  $t$  which consists of the raw three types of IMU sensors and the new orientation composition  $S_{\Theta_t}$ .  $\omega_t$  is 12-by-1 additive noise vector and  $F_t$  is the state transition model. Thirdly, we use the sum of angular velocity ( $\sum S_\beta = S_{\beta_x} + S_{\beta_y} + S_{\beta_z}$ ) to identify dynamical motions.

Fig. 2 shows the processed IMU signals from the inputs to output. We only adopt YZ-axis orientation, tri-axial linear L1-norm acceleration, the L1-norm accelerometer and the sum of gyroscope signals to establish the MNN classifier. As it is discussed above, the selected inputs can improve clustering and classification accuracy.

Several obvious features are extracted to train the MNN classifier as follows:

- 1) Extrema means the maximum and minimum of the signal segment, namely  $\max(s_i^*)$  and  $\min(s_i^*)$ .
- 2) Standard deviation is the amount of variation can find the range and level of the activity, namely  $\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (s_i^* - \bar{s}_i)^2}$
- 3) Average  $\bar{s}_i^*$  is the mean of the processed IMU signals, which provide the reference information of the average level of human activity.
- 4) Principal component analysis (PCA) is the most popular tool for evaluating the visualized genetic distance and relatedness between different activities.

Although the capability of single layer ANN approaches has proved to establish any complex model between high

dimensions inputs and classes, they have many drawbacks of overfitting to limit the classification performance.

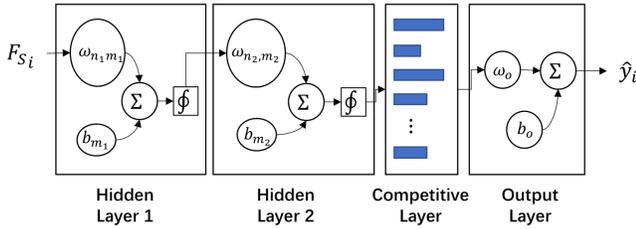


Fig. 3. The schematic diagram of designed two layers feed-forward neural network (FFNN) for classification.

In this article, a two layers MNN structure is designed with 30 and 50 nodes in each layer, respectively. It consists of two feed-forward layers and a competitive layer (see Fig. 3). The mapping networks in the feed-forward layer can be defined as:

$$\hat{y} = b_o + \omega_o \sum_{k=1}^K \Phi_k \left( \sum_{n=1}^N \sum_{m=1}^M \omega_{n, m_k} \gamma_t^{m_k} + b_{m_k} \right) \quad (4)$$

Where  $\Phi_k$  is the Broyden–Fletcher–Goldfarb–Shanno (BFGS) quasi-newton activation function. All of the  $\omega$  and  $b$  are the weights and bias in the network, and  $\gamma_t^{m_k}$  is the outputs of  $j$ th neuron. The competitive layer  $t$  classifies the input vectors into a given number of classes by the similarity between vectors.

### C. Autonomous Learning Framework

Fig. 4 illustrates the proposed autonomous learning framework. In the surgery room, 9D raw IMU signals and 45D depth vision data are synchronously captured from the IMU sensor and Kinect V2 camera. Then, they will be saved into the computing unit. After splitting the depth data  $V$  into  $N$  segments, i.e.,  $v_i, i = 1, 2, \dots, N$ , the Hk-medoids clustering approach can label these segments hierarchically. Even if the obtained classes  $y_i^*, i = 1, 2, \dots, N$  may not match the true activities  $y$ , they can be considered as the ground truth labels. The following experiments can prove this conclusion. Similarly, the acquired IMU signals  $S$  will be divided into  $N$  segment with the same detection length  $L_d$ , namely  $s_i, i = 1, 2, \dots, N$ . The approaches described above will calculate the features space  $F_{s_i}^*$ . The designed two layers neural network model can be trained by combining the inputs  $F_{s_i}^*$  and labels  $y^*$ , namely  $\hat{y} = f(F_{s_i}^*, \omega, b)$ . In the real-time demonstration, this system can predict four activities only by adopting IMU signals.

### D. Hardware Description

Fig. 5 shows the schematic diagram of a hardware system, including the data capture module and multiple computer units. The former two computers aim to collect depth vision and IMU signals from the doctor in the surgery room, while the latter is utilized to make acquisitions and process the data.

The sensors of the proposed HAR system are listed as follows:

- Kinect V2 sensor (Microsoft, USA) is composed of two cameras, an infrared (512x424 pixels) and an RGB camera (1920x1080 pixels). The framerate of the acquisition can be up to 30 Hz. And it has the feature for depth sensing of 70 degrees horizontally and 60 degrees vertically.
- The wearable 9D IMU sensor (WIT, China) consists of a tri-axis accelerometer, a tri-axis gyroscope, and a tri-axis magnetometer.

Two computers are used to obtain the data. One is equipped with i9-8950HK (2.9GHz) CPU and 16 GB RAM to perform depth image processing, while the other is with i7-4720HQ CPU (2.60GHz) processor and 8 GB RAM to obtain and process the IMU signals. Ultimately, a computer server (i9-9900K processor (3.6 GHz), 64 GB RAM and Quadro M5000 GPU) is used for data processing and HAR.

## III. EXPERIMENTS AND RESULTS

The two main procedures of the proposed framework could be evaluated separately. We adopted the supervised learning strategy to test the accuracy of the Hk-medoids clustering for automatic labeling and the classification accuracy of multisensor based MNN classifier. Ten subjects (five females and five males, between the ages of 20 and 35) were asked to make the ten human activities with a fixed order from 1 to 10. They carried an IMU sensor on the waist, which captured the IMU signals introduced in Section II-B. Each activity took at least 3 minutes. Finally, it had about 45000 samples of both depth data and the IMU signals with the same sampling frequency (50Hz).

Both overall accuracy (OA) and F1-measure (F1-score) are adopted to evaluate the clustering approaches and classification models. The F1-score can be calculated by considering both precision (P) and recall (R). The former ratio of correctly predicted positive observations to the total predicted positive observations, while the latter is all observations in the actual class.

$$F_1 = \frac{2 \times P \times R}{P + R} = \frac{2 \times \frac{TP}{TP+FP} \times \frac{TP}{TP+FN}}{\frac{TP}{TP+FP} + \frac{TP}{TP+FN}} \quad (5)$$

Where TP, FP, and FN are real positive, false positive, and false negative, respectively.

### A. Clustering Analyzing

The leave-one-out strategy is adopted to calculate the average of OA. Table I displays the comparison of OA between hierarchical k-medoids (Hk-medoids) with the other three popular clustering approaches, namely k-medoids, hierarchical k-means (Hk-means), and k-means methods. The Hk-means use the same three layers with the Hk-medoids structure. Both k-means and k-medoids approach cluster the depth data directly with a single layer. The Hk-medoids approach obtains the highest average of overall accuracy for

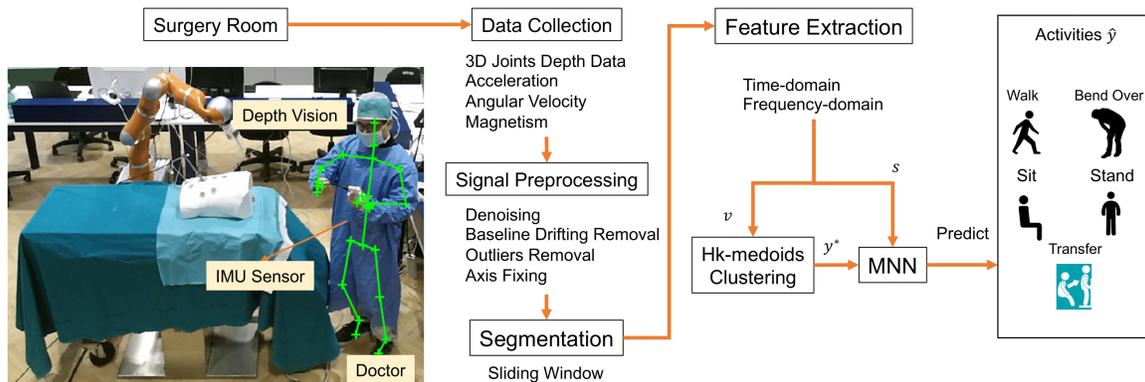


Fig. 4. The pipeline of depth vision-guided HAR using wearable multisensor in a surgery room. The Kinect device provides 3D joints depth data to label the five activities based on the designed Hk-medoids clustering method. The doctor carries the IMU sensor on the waist to afford multiple signals to train the MNN classifier for HAR.

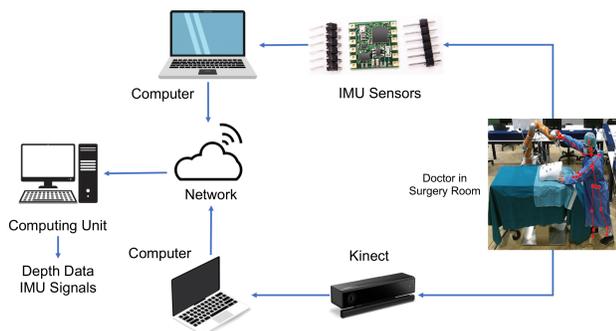


Fig. 5. The overview of hardware system structure.

labeling the five activities. However, it needs to select the specific features in each layer.

TABLE I

THE COMPARISON OF OVERALL ACCURACY AMONG FOUR CLUSTERING ALGORITHMS.

Algorithm	OA (%)	Feature Extraction
Hk-medoids	$96.55 \pm 0.04$	yes
Hk-means	$92.25 \pm 0.10$	yes
k-medoids	$86.49 \pm 0.06$	no
k-means	$58.38 \pm 0.37$	no

### B. Performance of MNN classifier Evaluation

Table 2 shows the computed F1-score of each gesture for further verification of the HSOM model's ability. Although the third gesture cannot be labeled well by HSOM, the average 90.22 is higher than that acquired by other methods.

TABLE II

THE COMPARISON OF OVERALL ACCURACY AMONG MNN, SVM, LDA, AND LINEAR METHOD CLASSIFIERS.

Classifier	OA (%)
MNN	$89.77 \pm 2.89$
SNN	$84.36 \pm 4.43$
SVM	$73.49 \pm 3.44$
LDA	$68.01 \pm 3.80$

For further comparison of the different performance in each activity classification, we compare the confusion matrix between MNN and SNN classifiers. By observing the confusion matrix of the MNN classifier shown in Fig. 6, although the MNN classifier is better than the SNN method, the mentioned five activities could be identified in the same level.

		MNN (89.64%)					SNN (84.84%)				
True Labels	sit	964	26	25	22	34	966	56	45	47	46
	stand	35	909	28	26	23	35	898	33	32	42
	bend over	28	32	917	27	23	36	36	903	38	45
	walk	31	35	20	902	28	31	37	47	902	44
	transfer	32	21	29	22	909	43	38	37	32	909
		Predicted Labels					Predicted Labels				
		sit	stand	bend over	walk	transfer	sit	stand	bend over	walk	transfer

Fig. 6. The confusion matrix obtained based on MNN and SNN classifier.

## IV. CONCLUSION AND FUTURE WORK

A novel autonomous learning framework is proposed in this paper for enhancing the multisensor based HAR in the surgery room. It adopts the depth information to automatically label the five human activities automatically captured from the Kinect V2 sensor. For robustness and reducing the interference of artifact, a series of signal processing algorithms are designed to remove different types of noises. The MNN method is used to build the classifier for accuracy enhancement, which can acquire better accuracy than the other method. The proposed framework can not only be utilized for HAR using only IMU signals but also label the data based on depth data.

Although the proposed framework automatically achieves HAR, further research is needed to implement it in RAMIS fully. In our future works, a 3D segmented preoperative model will be used for AR navigation of surgeons, and hand gestures will manipulate it. Finally, HAR-based teleoperation of a surgical robot will be performed in the future.

## REFERENCES

- [1] W. Liu, Z.-J. Zha, Y. Wang, K. Lu, and D. Tao, “ $p$ -laplacian regularized sparse coding for human activity recognition,” *IEEE Transactions on Industrial Electronics*, vol. 63, no. 8, pp. 5120–5129, 2016.
- [2] R. Bharathan, R. Aggarwal, and A. Darzi, “Operating room of the future,” *Best Practice & Research Clinical Obstetrics & Gynaecology*, vol. 27, no. 3, pp. 311–322, 2013.
- [3] A. P. Twinanda, “Vision-based approaches for surgical activity recognition using laparoscopic and rgb-d videos,” Ph.D. dissertation, Strasbourg, 2017.
- [4] S. Ahmadi, N. Padoy, S. Heining, H. Feussner, M. Daumer, and N. Navab, “Introducing wearable accelerometers in the surgery room for activity detection,” *Computer-und Roboter-Assistierte Chirurgie (CURAC)*, 2008.
- [5] S. Khalifa, G. Lan, M. Hassan, A. Seneviratne, and S. K. Das, “Harke: Human activity recognition from kinetic energy harvesting data in wearable devices,” *IEEE Transactions on Mobile Computing*, vol. 17, no. 6, pp. 1353–1368, 2017.
- [6] Y. Liu, L. Nie, L. Liu, and D. S. Rosenblum, “From action to activity: sensor-based activity recognition,” *Neurocomputing*, vol. 181, pp. 108–115, 2016.
- [7] Z. Chen, L. Zhang, Z. Cao, and J. Guo, “Distilling the knowledge from handcrafted features for human activity recognition,” *IEEE Transactions on Industrial Informatics*, vol. 14, no. 10, pp. 4334–4342, 2018.
- [8] Y. Wang, L. Wei, A. V. Vasilakos, and Q. Jin, “Device-to-device based mobile social networking in proximity (msnp) on smartphones: Framework, challenges and prototype,” *Future Generation Computer Systems*, vol. 74, pp. 241–253, 2017.
- [9] A. Wang, G. Chen, J. Yang, S. Zhao, and C.-Y. Chang, “A comparative study on human activity recognition using inertial sensors in a smartphone,” *IEEE Sensors Journal*, vol. 16, no. 11, pp. 4566–4578, 2016.
- [10] J. Shlens, “A tutorial on principal component analysis,” *arXiv preprint arXiv:1404.1100*, 2014.
- [11] S. Beura, B. Majhi, and R. Dash, “Automatic characterization of mammograms using fractal texture analysis and fast correlation based filter method,” in *Proceedings of the 2nd International Conference on Perception and Machine Intelligence*. ACM, 2015, pp. 85–91.
- [12] G. Chandrashekar and F. Sahin, “A survey on feature selection methods,” *Computers & Electrical Engineering*, vol. 40, no. 1, pp. 16–28, 2014.
- [13] A. M. Khan, Y.-K. Lee, S. Lee, and T.-S. Kim, “Human activity recognition via an accelerometer-enabled-smartphone using kernel discriminant analysis,” in *Future Information Technology (FutureTech), 2010 5th International Conference on*. IEEE, 2010, pp. 1–6.
- [14] M. Shoaib, S. Bosch, O. D. Incel, H. Scholten, and P. J. Havinga, “Complex human activity recognition using smartphone and wrist-worn motion sensors,” *Sensors*, vol. 16, no. 4, p. 426, 2016.
- [15] A. D. Ignatov and V. V. Strijov, “Human activity recognition using quasiperiodic time series collected from a single tri-axial accelerometer,” *Multimedia tools and applications*, vol. 75, no. 12, pp. 7257–7270, 2016.
- [16] W. Sousa, E. Souto, J. Rodrigues, P. Sadarc, R. Jalali, and K. El-Khatib, “A comparative analysis of the impact of features on human activity recognition with smartphone sensors,” in *Proceedings of the 23rd Brazillian Symposium on Multimedia and the Web*. ACM, 2017, pp. 397–404.
- [17] X. Yang and Y. Tian, “Super normal vector for human activity recognition with depth cameras,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 5, pp. 1028–1039, 2016.
- [18] M. Kose, O. D. Incel, and C. Ersoy, “Online human activity recognition on smart phones,” in *Workshop on Mobile Sensing: From Smartphones and Wearables to Big Data*, vol. 16, no. 2012, 2012, pp. 11–15.
- [19] Y.-S. Lee and S.-B. Cho, “Activity recognition with android phone using mixture-of-experts co-trained with labeled and unlabeled data,” *Neurocomputing*, vol. 126, pp. 106–115, 2014.
- [20] J.-L. Reyes-Ortiz, L. Oneto, A. Samà, X. Parra, and D. Anguita, “Transition-aware human activity recognition using smartphones,” *Neurocomputing*, vol. 171, pp. 754–767, 2016.
- [21] Y.-S. Lee and S.-B. Cho, “Activity recognition using hierarchical hidden markov models on a smartphone with 3d accelerometer,” in *International Conference on Hybrid Artificial Intelligence Systems*. Springer, 2011, pp. 460–467.
- [22] A. Reiss, G. Hendeby, and D. Stricker, “A competitive approach for human activity recognition on smartphones,” in *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2013)*, 24–26 April, Bruges, Belgium. ESANN, 2013, pp. 455–460.
- [23] W. Qi, H. Su, C. Yang, G. Ferrigno, E. De Momi, and A. Aliverti, “A fast and robust deep convolutional neural networks for complex human activity recognition using smartphone,” *Sensors*, vol. 19, no. 17, p. 3731, 2019.
- [24] W. Qi and A. Aliverti, “A multimodal wearable system for continuous and real-time breathing pattern monitoring during daily activity,” *IEEE Journal of Biomedical and Health Informatics*, 2019.
- [25] P. Kanjiya, V. M. Khadkikar, and M. S. ElMoursi, “Adaptive low-pass filter based dc offset removal technique for three-phase plls,” *IEEE Transactions on Industrial Electronics*, 2018.
- [26] D. Roetenberg, H. J. Luinge, C. T. Baten, and P. H. Veltink, “Compensation of magnetic disturbances improves inertial and magnetic sensing of human body segment orientation,” *IEEE Transactions on neural systems and rehabilitation engineering*, vol. 13, no. 3, pp. 395–405, 2005.