Article

TRANSACTIONS OF THE INSTITUTE OF MEASUREMENT & CONTROL

Transactions of the Institute of Measurement and Control
2022, Vol. 44(4) 735–743

DOI: 10.1177/0142331220984350
journals.sagepub.com/home/tim
SAGE

Multimodal data fusion framework enhanced robot-assisted minimally invasive surgery

Wen Qi[1], Hang Su[1], Ke Fan[1], Ziyang Chen[1], Jiehao Li[1], Xuanyi Zhou[1], Yingbai Hu[2], Longbin Zhang[3], Giancarlo Ferrigno[1] and Elena De Momi[1]

Abstract

The generous application of robot-assisted minimally invasive surgery (RAMIS) promotes human-machine interaction (HMI). Identifying various behaviors of doctors can enhance the RAMIS procedure for the redundant robot. It bridges intelligent robot control and activity recognition strategies in the operating room, including hand gestures and human activities. In this paper, to enhance identification in a dynamic situation, we propose a multimodal data fusion framework to provide multiple information for accuracy enhancement. Firstly, a multi-sensors based hardware structure is designed to capture varied data from various devices, including depth camera and smartphone. Furthermore, in different surgical tasks, the robot control mechanism can shift automatically. The experimental results evaluate the efficiency of developing the multimodal framework for RAMIS by comparing it with a single sensor system. Implementing the KUKA LWR4 + in a surgical robot environment indicates that the surgical robot systems can work with medical staff in the future.

Keywords

Multimodal data fusion, human activity recognition, minimally invasive surgery, redundant manipulator, event-based control

Introduction

Background of robot-assisted minimally invasive surgery

Surgical robots can perform many kinds of operations with higher accuracy and flexibility, which is significant to achieve accurate, safe, and minimally invasive surgery. Therefore, robot-assisted minimally invasive surgery (RAMIS) (Su et al., 2021a) has attracted more attention over the past decades. The RAMIS system consists of three parts: a teleoperated console, a slave manipulator system, and the teleoperation control system (Caccianiga et al., 2020). Doctors use local manipulators to remotely operate surgical instruments in patients and observe the operating environment through a three-dimensional (3-D) endoscopic camera (Milstein et al., 2018). Compared with traditional minimally invasive surgery, minimally invasive surgery (RAMIS) has a more delicate operation, clearer vision, and a more comfortable operation process. Besides, patients obtain all the benefits of traditional minimally invasive surgery, such as small incision size, short wound healing time, less pain, and low risk of surgical infection (Enayati et al., 2016), which promotes the broad application of RAMIS. However, compared with traditional minimally invasive surgery, the evidence of improved prognosis in patients with RAMIS is unclear in some procedures, and the application of RAMIS in other operations is still limited. At present, some limitations of RAMIS can be alleviated by optimizing the controller of the system and the guidance approaches so that the doctor's operation will be more precise (He et al., 2020). Compared with traditional open surgery, RAMIS has made significant progress in a large number of intervention surgeries. RAMIS provides a fantastic potential for further improvement of MIS. Many RAMIS platforms on the market, such as Da Vinci, Blu ray, and versions. They are convenient for operation with high-definition 3D vision and various intelligent surgical tools (Konstantinova et al., 2014). However, the high cost of custom-made surgical robots limits the hospital's purchase of them. Industrial robots with redundant manipulators, such as the KUKA robot, have been successfully developed with high-performance controllers for decades. The lower cost of specialized surgical robots has increased their medical application prospect Chen et al. (2019, 2020), especially in RAMIS.

[1]Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Italy
[2]Department of Informatics, Technical University of Munich, Germany
[3]Department of Mechanics, KTH Royal Institute of Technology, Sweden

Corresponding author:
Hang Su, Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Piazza Leonardo da Vinci, 32, Milano, 20133, Italy.
Email: hang.su@polimi.it

Human activity recognition using multi-sensor fusion

Sensor-based human activity recognition (HAR) has been widely used in various fields, including intelligent medicine, smart home, and sports activities. In the past decade, with the development of sensor technology and the reduction of sensor equipment cost, various sensors are widely used in human activity recognition (Mario, 2018). HAR based on multi-sensor fusion has essential application value for guiding RAMIS. The rapid development of artificial intelligence (AI) technologies provides many effective methods to build a HAR classifier. Many HAR models have been proposed during the last 10 years by manually extracting typical features from the raw signals. A deep convolutional neural network for efficient HAR using smartphone sensors was proposed in Tufek et al. (2019), and features were automatically extracted from raw data. A deep belief network (DBN) was trained with features extraction from raw data of human activities by kernel principal component analysis (KPCA) and linear discriminant analysis (LDA) in Hassan et al. (2018). A kernel extreme learning machine (QPSO-KELM) based on kernel discriminant analysis (KDA) and quantum-behaved particle swarm optimization was designed to extract features and enhance the accuracy of the HAR system (Alharthi et al., 2019). A classifier based on multiple support vector machines (SVMs) was proposed, and specific SVM models were trained respectively for each type of feature (Mondéjar-Guerra et al., 2019). A transition-aware human activity recognition (TAHAR) system architecture based on SVMs was proposed, realizing real-time classification with a series of inertial sensors (Reyes-Ortiz et al., 2016). A new fusion model was proposed in Huynh-The et al. (2020). In the fusion model, manual features and in-depth features are fused by a multi-class SVM classifier. Essentially, SVM is of a convex quadratic optimization problem with linear inequal constraints. SVM can separate binary class data from one class to another through searching for the optimal separating hyperplane (Mathur and Foody, 2008). SVM was used to classify the spatiotemporal parameters of the preoperative gait of patients with knee osteoarthritis (Naik et al., 2018). An augmented feature space was established by using a combination of SVM and hidden Markov model (HMM) in robot-assisted surgical systems (Tatinati et al., 2014). Based on multi-sensor HAR, the data collected by multi-sensor need to be fused. Researchers have proposed different data fusion strategies for multi-sensor data fusion (Gravina et al., 2017). However, using traditional machine learning (ML) methods for data fusion, each data fusion strategy has advantages and disadvantages. A single sensor's ability to recognize that human activity is limited, and multiple sensors can provide more recognition ability. More complex activities can be identified by processing and analyzing the data collected by various sensors (Wang et al., 2018). However, when using multiple sensors for activity recognition, some sensor data play a decisive role in the activity recognition model, while others may have negative effects (such as noise or damaged signals) in the learning process (Norgaard et al., 2019). Moreover, when using multiple sensors to identify activities, some sensors are not effective for some activities. Some redundant data will increase the amount of calculation and may lead to an overfitting phenomenon, which may deteriorate the classification model (Ehatisham-Ul-Haq et al., 2019). Therefore, multi-sensor data fusion is of great significance.

Multimodal framework for HAR

Therefore, this paper discusses a multi-sensor fusion HAR based on the SVM algorithm. Human activities will be collected and preprocessed by a multimodal sensor fusion system. Then, we use SVM-based model to classify human behaviors and intentions. The ensemble learning-based SVM by using the fusion of data preprocessing and features extraction is explored. Some studies work about imbalanced data classification with SVM has been achieved. The study for a multi-class SVM ensemble learning algorithm has been explored on HAR's RAMIS application occasion. In the robot's operational space (Gao et al., 2021), the robot's control strategy is automatically switched according to different surgical tasks and human activities. The proposed HAR-RAMIS method is a step forward in exploring the higher-level surgical knowledge given by artificial intelligence technology so that more surgical robots can carry out the intelligent and efficient operation and cooperation. In our previous works, a multi-sensors-based HAR system was proposed to monitor breathing patterns during different activity (Qi and Aliverti, 2019). The framework integrated both physiological and physical sensors to capture multiple data. Also, a smartphone-based HAR framework is proposed by combining multiple signals collected from the inertial measurement unit (IMU) sensors (Qi et al., 2020).

In this article, we propose a multimodal data fusion framework to identify elaborate behaviors and hand gestures. A hardware wireless connection system is designed to collect 3D joints and IMU data simultaneously for providing more information to increase the recognition accuracy. The following items list the contributions of the proposed multimodal data fusion framework:

- A wireless connection hardware system, is designed to capture multiple data simultaneously;
- A multimodal data fusion framework is proposed to process and analyze the raw data.

The paper is organized into the following four sections. The general architecture of the proposed multimodal data fusion framework and the implemented wireless connection system are described in Section 2. Section 4 shows the performance evaluation of the proposed SVM-based modal by comparing with other ML models. Section 5 summarizes the achievements of this article and delineates further work.

Methodology

The multimodal data fusion framework in RAMIS

The scenario of the multimodal data fusion operating room is depicted in Figure 1. It uses Kinect cameras, IMU sensors, and virtual reality (VR) techniques to capture the mentioned data, combined to achieve complex instrument operation procedures. To enhance the recognition rate, we collect various

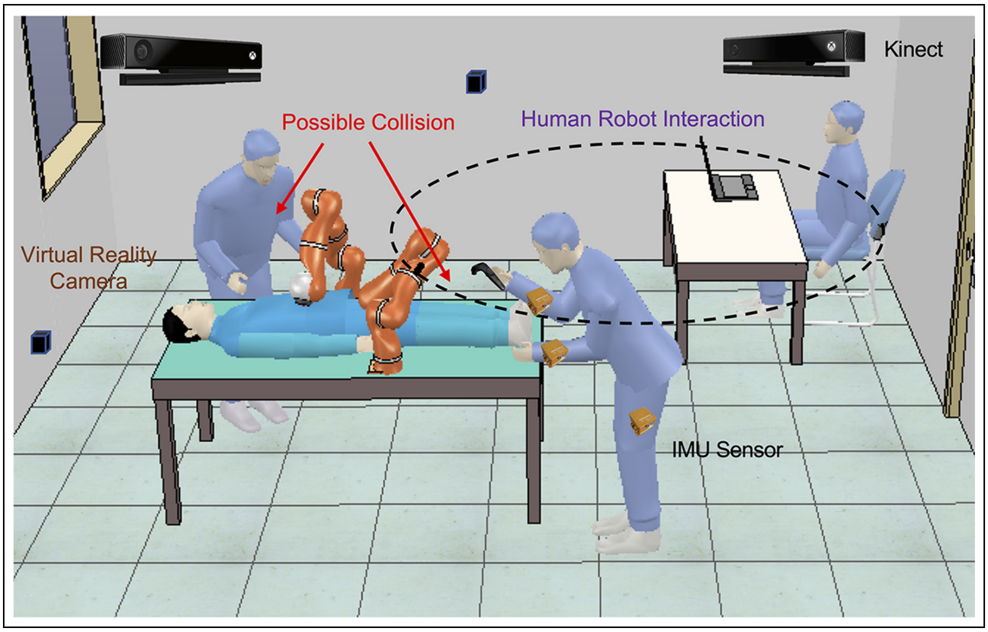

Figure 1. The general scenario in the operating room for human activity and hand gesture recognition using multimodal fusion-powered devices.

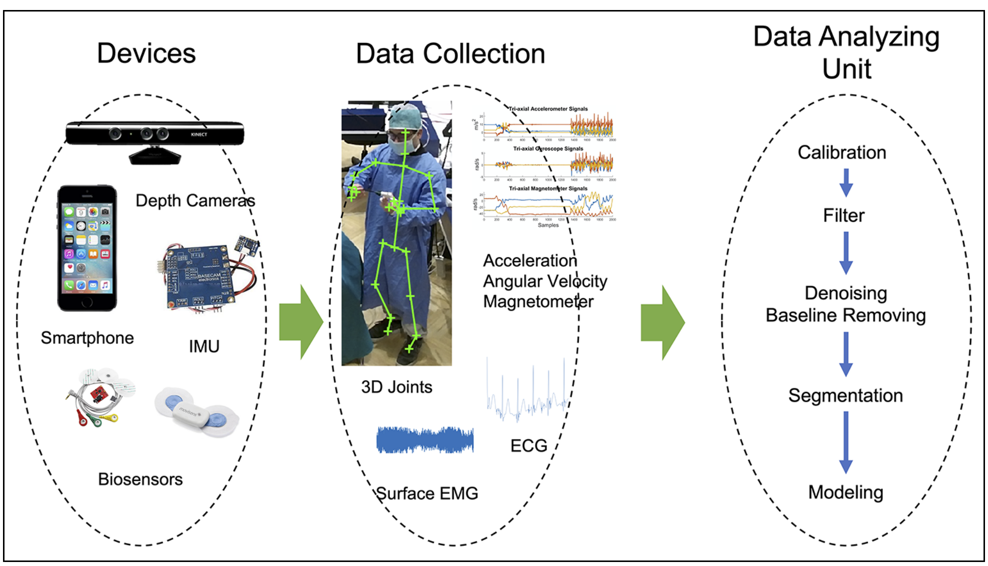

Figure 2. The procedures of data collection, processing, and modeling.

data, including 3D joints, orientation, acceleration, and angular velocity (Li et al., 2017). The surgeon's behaviors are significant to be monitored, which contributes to the intelligent behavior of the robot. In order to avoid interference and possible collision, different data based on IMU, electromyography (EMG) sensors, and depth cameras are collected to establish the classifiers. Then, the position, action intention, and direction can be predicted.

According to the given scenario in the surgery room, multiple data are collected using several devices. Figure 2 shows the procedure from data collection to analyzing. Both depth data and signals are captured from depth cameras and sensors (e.g., IMU and biosensors), including 3D joints and physical and physiological signals. In the data analyzing unit, this information is processed in a fixed procedure. First, the depth vision data should be transferred into the body frame (calibration). Then, the existing noises and baseline drifting are removed from the raw signal because they will affect computing physiological parameters' accuracy. Several signal processing algorithms are combined to solve these problems, including wavelet denoising, baseline drifting removal, Kalman filter, and particle filter (Su et al., 2020b). Notably,

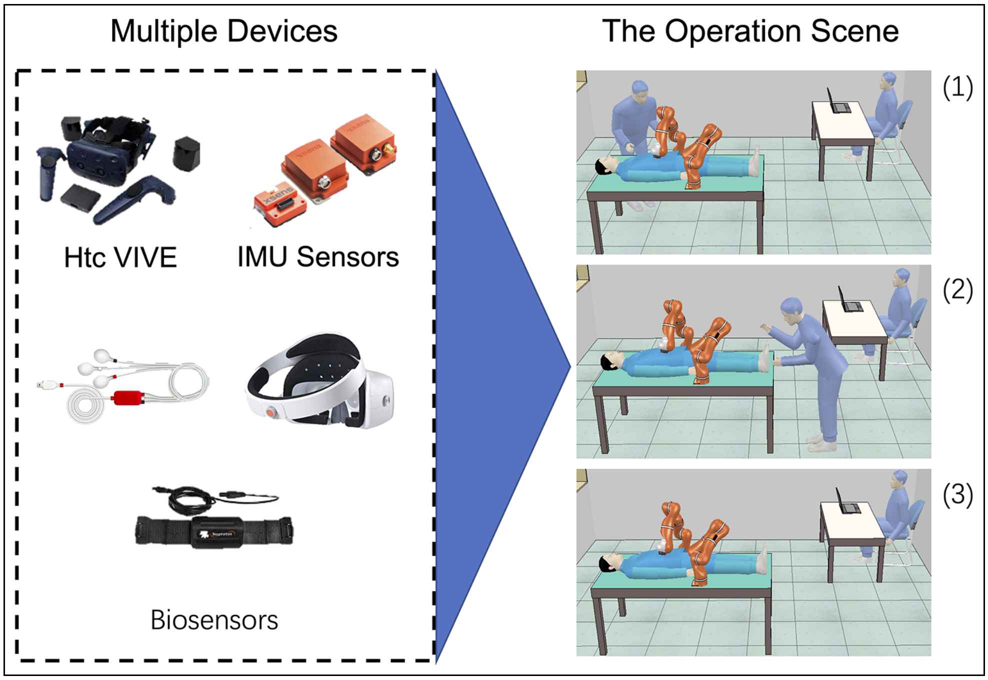

Figure 3. The multiple devices based recognition system.

the usable range should be adjusted based on the raw data/signal. The doctor's directions and positions will then be calibrated on the depth data (3D joints) for accuracy enhancement. For acquiring suitable segments, all data are divided into several data parts with the same detection length. In this paper, the SVM approach is proposed to build the classifier. The 3D joints data are used to label the segments; then, the SVM classifier can be trained based on these labels. Meanwhile, the SVM-based classifier is established using multiple data, including depth vision, IMU signals, and biosensor data. Where acceleration, angular velocity, and direction are the primary information for identifying the activities. The operator carries more IMU sensors, which can provide multiple data for recognizing complex motions or gestures (see Figure 3). Meanwhile, it reduces unnecessary interference during the surgical operation by perceiving action intention (Yang et al., 2018).

In the real operation scene, it should consider achieving smoothness, stability, and safety (Li et al., 2020). The multimodal system also needs to sense and rearrange the priority tasks. Hence, it should utilize a continuous adaptive control approach to ensure the level of switching of the tasks, shown in Figure 4. It consists of two primary modules of the event-driven and hierarchical control architecture, where the SVM algorithm is implemented to identify human activity and hand gestures by adopting multiple data.

In the RAMIS, the serial robot's redundancy can be utilized to accomplish manipulability optimization, compliant safety enhanced strategy, human-like behavior, and remote center of motion (Atawnih et al., 2014; Sandoval et al., 2018). Figure 4 displays the three levels control objectives. Firstly, it needs to ensure a successful surgery during RAMIS (Liu et al., 2020). Secondly, the small incision on the abdominal wall is respected by producing a kinematic constraint, known as a remote center of motion (RCM) (Qi et al., 2019; Wang et al., 2019). Thirdly, extra tasks are utilized by the other redundancies of the robot arm (Li et al., 2019; Qi et al., 2020; Su et al., 2020a, 2021b).

SVM-based modeling

In Figure 4, the operators' behavior needs to be monitored by the SVM-based classifier. This article considers using the SVM method to build the multi-class classifier for recognizing four hand gestures and several activities. Shortly, the SVM approach can be described by the following notations.

The collected data are divided into training and testing parts. The data couple $(x_i, y_i)_{i=1}^{N}$ is set as the training data set, where $x_i \in \mathbb{R}^d$ is the i th input matrix and $y_k \in \mathbb{R}^1$ is the one dimension output. The constructed classifier form is

$$y(x) = \text{sign}[\sum_{i=1}^{N} \alpha_i y_i \psi(x, x_i) + b] \tag{1}$$

α and b are the parameters set. The typical function of the SVM model ψ can be chosen as linear SVM, polynomial SVM of degree p, radial basis function (RBF) SVM, or two layer neural SVM.

$$\psi = \begin{cases} x_i^{\mathrm{T}} x \\ x_i^{\mathrm{T}} x + 1^p \\ exp^{-\|x-x_i\|_2^2/\sigma^2} \\ \tanh[\kappa x_i^{\mathrm{T}} + \theta] \end{cases} \tag{2}$$

where σ, κ and θ are constants, and the classifier is constructed by

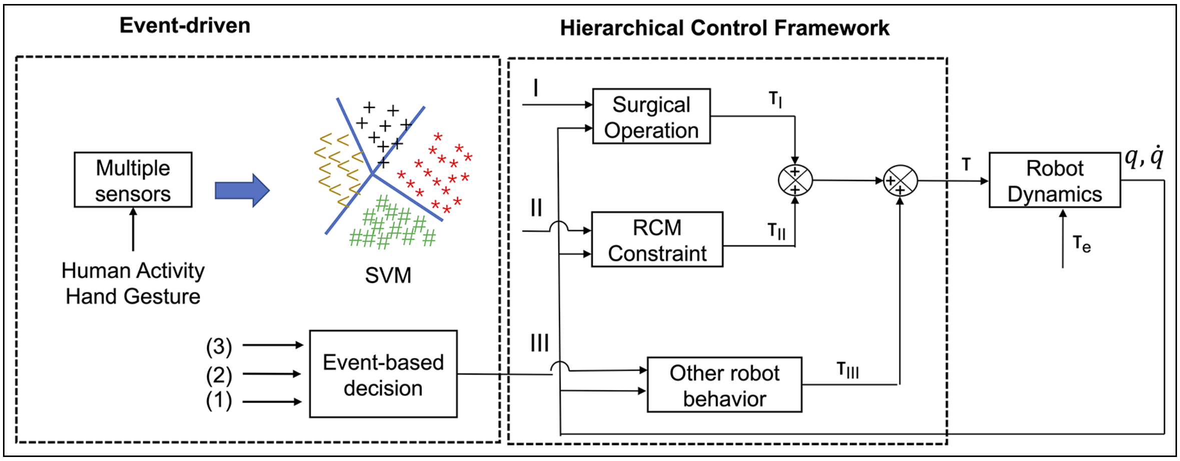

Figure 4. The event-driven scheme based on control model and SVM classifier.

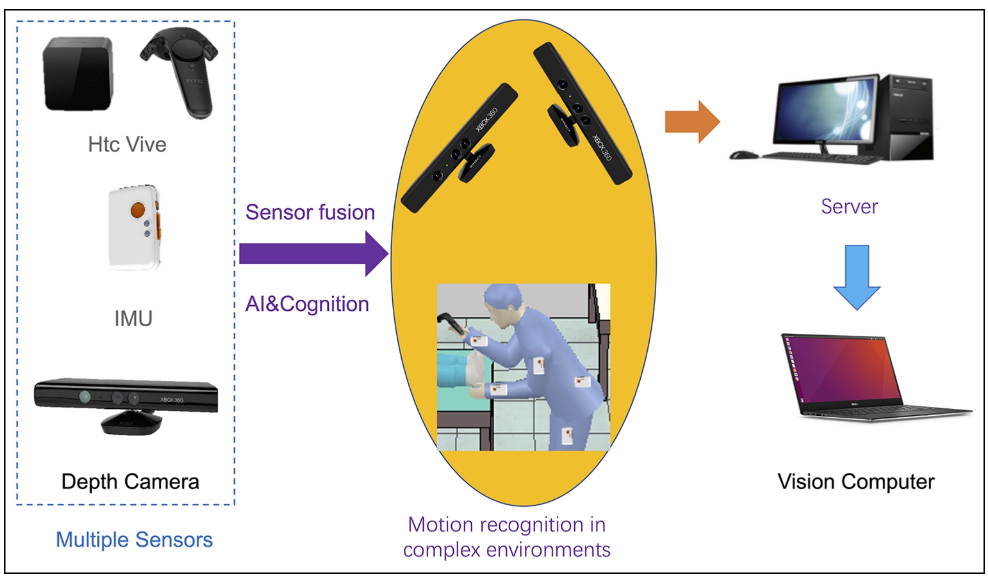

Figure 5. The hardware system structure for data collection and processing.

$$\begin{aligned} &\omega^{\mathrm{T}}\varphi(x_i) + b \geqslant 1, \text{if } y_i = +1, \\ &\omega^{\mathrm{T}}\varphi(x_i) + b \geqslant -1, \text{if } y_i = -1 \end{aligned} \tag{3}$$

which is equivalent to

$$y_i[\omega^{\mathrm{T}}\varphi(x_i) + b] \geqslant 1, i = 1, \ldots, N \tag{4}$$

where $\varphi(\cdot)$ is a nonlinear function which projects the input space into a higher dimensional space.

The multisensors wireless connection system

Figure 5 describes the data processing procedure from the capture module to the vision computer. The considered multiple signals would be captured by the three devices, namely Kinect V2 camera, IMU sensors, and Myo armband. The former two devices are used to identify human activities, while the depth vision and the Myo armband can be adopted to recognize hand gestures. Finally, the collected data are saved and processed in the server unit.

The considered sensors of this hardware system are listed as follows:

- Two cameras are embedded in the Kinect V2 sensor (Microsoft, USA). One is 1920x1080 pixels RGB camera, and the other is 512x424 pixels infrared. The horizontal and vertical depth-sensing are 70 and 60 degrees. The frame rate is set up to 30Hz.
- Three sensors (i.e., accelerometer, magnetometer, and gyroscope) are included in the 9D IMU (WIT, China).
- The server computer aims to save the captured data with 64GB RAM, i9-9900K (3.6 GHz) CPU, and Quadro M5000 GPU.

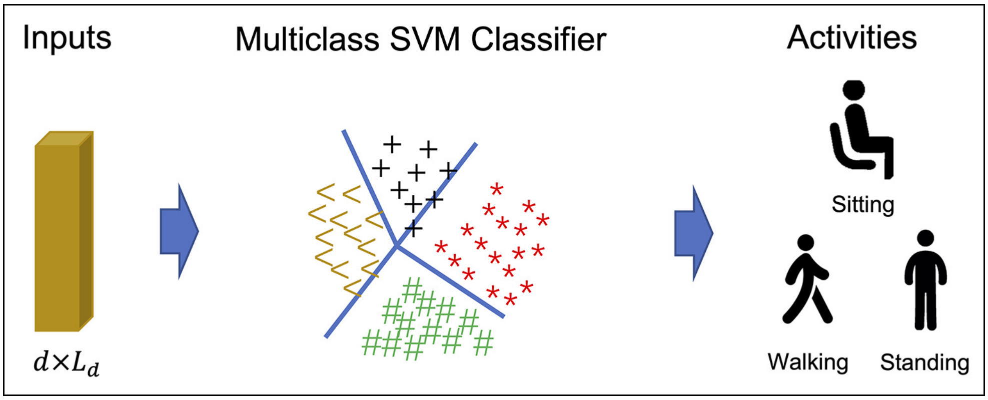

Figure 6. The schematic diagram of multi-class SVM model.

Table 1. The human activity identification results under six classifiers by combining IMU and Kinect devices.

Devices	Parameters	SVM	k-NN	SNN1	MNN1	SNN2	MNN2
IMU (waist)	Accuracy (%)	72.36±4.39	54.35±3.22	51.27±3.03	54.26±2.92	59.77±3.26	57.36±2.52
	Time (s)	0.29	0.15	1.77	2.98	2.03	3.66
IMU (pocket)	Accuracy (%)	77.26±3.47	60.28±3.90	59.73±2.98	62.91±3.19	67.48±2.77	65.15±3.01
	Time (s)	0.31	0.19	1.57	2.55	1.99	3.83
2 IMUs	Accuracy (%)	82.91±4.04	66.38±4.89	64.29±3.11	67.49±3.72	72.35±3.26	70.00±3.53
	Time (s)	0.28	0.20	1.68	2.77	2.11	4.02
2 IMUs + Kinect	Accuracy (%)	88.17±3.77	72.49±3.13	70.93±2.99	67.49±2.01	78.46±2.74	75.50±2.11
	Time (s)	0.30	0.22	1.49	2.55	1.99	3.83

- The display equipment is the processor with i7-4720HQ CPU (2.60GHz) and 8GB RAM.

Experiments

To evaluate the performance of the proposed multi-sensors system, we design two experiments with the following protocol. The first one aims to verify HAR's identification ability based on the proposed SVM classifier using IMU or Kinect data. The second one is to prove hand gesture recognition capability by using Myo armband and Kinect. The comparison results can not only evaluate the classification ability but also prove the capability to recognize more activities or hand gestures.

HAR

We invited 10 volunteers aged 18 to 35 to participate in this experiment, including five men and five women. They were asked to do five typical activities in the operating room, that is, walking, sitting, standing, bending over forwards, and a series of transfer motions. Each activity had been done for one minute. These participants carried two IMU sensors in the left pant's pocket and on the waist. 3D joints data were collected by the Kinect camera to calibrate the results of the SVM classifier. Finally, there are 9000 samples collected based on the same sampling frequency (30Hz) of the two devices. We adopt the leave-one-out strategy to evaluate the accuracy and SVM classifier's speed of the IMU and Kinect sensors. The data collected from nine subjects is adopted to train the SVM modal, then the last one for testing. The IMU sensors were carried on the waist and pocket. The classifiers were trained based on the selected datasets of nine subjects, while the last one was used to test the classification accuracy.

We compared the recognition rate among six classifiers, namely multi-class SVM, k-nearest neighbor (k-NN), different types of single neural networks (SNN), and multiple neural networks (MNN). This paper sets 60 and 80 nodes in the single hidden layer of SNN to build the two ANN-based classifiers. We adopted two hidden layers to establish two MNN models with [40, 80] nodes and [20, 100] nodes, respectively. Figure 6 shows the designed multiple SVM classifiers to identify human activities. The multi-class SVM model used one versus one strategy.

The results will be tested 20 times for avoiding overfitting or underfitting. Also, different types of data were collected to explore the best combination of sensors. Table 1 presents the classification accuracy of five activities based on the mentioned six ML algorithms. To verify the effect of multimodal data fusion on the accuracy, we set up four kinds of data collection methods, including IMU on the waist, IMU in the pocket, two IMUs, and depth vision.

The performance of SVM is better than other classifiers by comparing the accuracy. It can be noted that data from the waist possess higher accuracy than that from the pocket. Moreover, the combination of the two data can improve the prediction accuracy of the models. Considering that depth

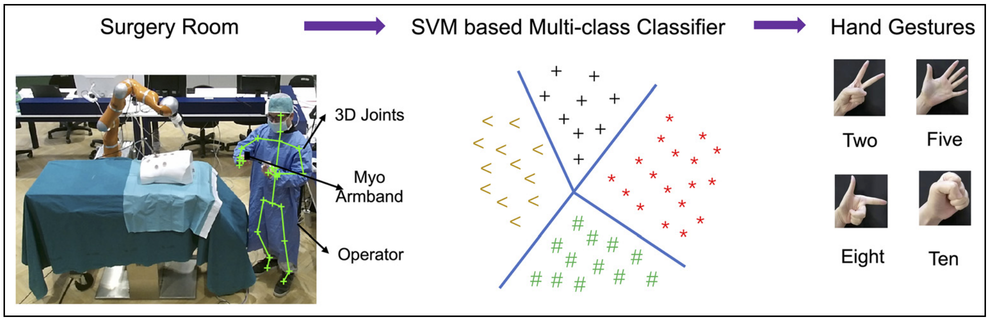

Figure 7. The diagram for recognizing four hand gestures using Myo armband and Kinect camera in the surgery environment.

Table 2. The multiple class hand gestures identification results under six classifiers by combining IMU and Kinect devices.

Devices	Parameters	SVM	k-NN	SNN1	MNN1	SNN2	MNN2
Armband	Accuracy (%)	86.35±3.99	80.03±3.26	77.58±4.78	82.54±3.03	80.03±3.31	82.17±3.05
	Time (s)	1.08	0.85	2.07	3.02	2.58	4.14
Armband & camera	Accuracy (%)	90.26±4.04	83.28±3.98	80.33±5.38	85.28±3.64	84.17±2.84	86.36±2.47
	Time (s)	0.94	0.79	2.55	3.11	2.82	4.21
IMU sensors	Not work						
2 IMUs sensors	Not work						

data can provide position information, we combined data from two IMUs and Kinect to expand the data's richness. The fourth-row results show that it can enhance these five activities' classification accuracy by using more sensors.

Furthermore, the prediction time of human activity shows that SVM is a faster classifier than other ANN models. Although the time of the k-NN algorithm is the fastest, its classification accuracy is low. Therefore, the multi-class SVM has the advantages of high accuracy and fast recognition, and it does not require high hardware computing ability due to its simple structure, which determines that SVM can be used as an ideal classifier of human activity.

Hand gesture recognition based on Myo armband and Kinect

Figure 7 demonstrates the recognition process of hand gestures. The same 10 participants wore the Myo armband on their forearms and performed four-hand gestures, that is, two, five, eight, and 10. In this experiment, the surface EMG signals and depth data were collected to train different classifiers. Similarly, to verify the superiority of data fusion, different types of sensors were combined to compare the six models, including Myo armband, Kinect, and IMU. The data collected from nine subjects are used for training the SVM modal, while the last one was for testing. Besides, we used the cross-validation method and tested the results 20 times to improve the data's credibility.

Table 2 presents the accuracy of four gesture recognition based on the combination of different models and sensors. It can be seen that the SVM classifier is the most accurate model to recognize hand gestures. Furthermore, the data collected by the combination of EMG signals and depth vision can train a better classifier than a single sensor. However, the data based on the IMU sensor can not be used to recognize hand gestures. On the other hand, the recognition time of hand gestures also reflects the superiority of SVM. It can balance the recognition accuracy and speedwell.

These two experimental results present that the HAR system based on multi-class SVM can distinguish human activities and recognize various hand postures accurately and quickly. Moreover, the method based on multimodal data fusion can further enhance classification accuracy.

Conclusion

A multimodal data fusion framework to identify human activities and hand gestures are proposed in this paper. It adopts multiple sensors to capture different human body data, such as IMU, Myo armband, and Kinect camera. By comparing the trained SVM classifier's performance among these sensors, multi-class SVM based on data fusion can construct classifiers of higher accuracy and faster recognition compared with other algorithms. Notably, combining multiple sensors can get better performance of the system than using a single sensor. Given the obtained results, ML-based methods are used to identify human behaviors and intentions, which can promote intelligent interaction between humans and robots and further improve operational safety and efficiency. These advantages show their potential value in future RAMIS.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the European Union's Horizon 2020 Research and Innovation Program through SMARTsurg Project under Grant 732515.

ORCID iDs

Hang Su https://orcid.org/0000-0002-6877-6783
Yingbai Hu https://orcid.org/0000-0003-2452-3570

References

Alharthi AS, Yunas SU and Ozanyan KB (2019) Deep learning for monitoring of human gait: A review. *IEEE Sensors Journal* 19(21): 9575–9591.

Atawnih A, Papageorgiou D and Doulgeri Z (2014) Reaching for redundant arms with human-like motion and compliance properties. *Robotics and Autonomous Systems* 62(12): 1731–1741.

Caccianiga G, Mariani A, De Momi E, et al. (2020) An evaluation of inanimate and virtual reality training for psychomotor skill development in robot-assisted minimally invasive surgery. *IEEE Transactions on Medical Robotics and Bionics* 2(2): 118–129.

Chen Y, Godage I, Su H, et al. (2019) Stereotactic systems for mri-guided neurosurgeries: A state-of-the-art review. *Annals Of Biomedical Engineering* 47(2): 335–353.

Chen Y, Wang L, Galloway K, et al. (2020) Modal-based kinematics and contact detection of soft robots. *Soft Robotics*. DOI: 10.1089/soro.2019.0095

Ehatisham-Ul-Haq M, Javed A, Azam MA, et al. (2019) Robust human activity recognition using multimodal feature-level fusion. *IEEE Access* 7: 60736–60751. DOI:10.1109/ACCESS.2019.2913393

Enayati N, De Momi E and Ferrigno G (2016) Haptics in robot-assisted surgery: Challenges and benefits. *IEEE Reviews in Biomedical Engineering* 9: 49–65. DOI:10.1109/RBME.2016.2538080

Gao H, Zhu J, Li X, et al. (2021) Automatic parking control of unmanned vehicle based on switching control algorithm and backstepping. *IEEE/ASME Transactions on Mechatronics*. DOI:10.1109/TMECH.2020.3037215.

Gravina R, Alinia P, Ghasemzadeh H and Fortino G (2017) Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges. *Information Fusion* 35: 68–80. DOI:10.1016/j.inffus.2016.09.005

Hassan MM, Uddin MZ, Mohamed A and Almogren A (2018) A robust human activity recognition system using smartphone sensors and deep learning. *Future Generation Computer Systems* 81: 307–313. DOI:10.1016/j.future.2017.11.029

He Y, Zhang P, Qi X, Zhao B, Li S and Hu Y (2020) Endoscopic path planning in robot-assisted endoscopic nasal surgery. *IEEE Access* 8: 17039–17048.

Huynh-The T, Hua CH, Tu NA and Kim DS (2020) Physical activity recognition with statistical-deep fusion model using multiple sensory data for smart health. *IEEE Internet of Things Journal*. DOI: 10.1109/JIOT.2020.3013272

Konstantinova J, Jiang A, Althoefer K, et al. (2014) Implementation of tactile sensing for palpation in robot-assisted minimally invasive surgery: A review. *IEEE Sensors Journal* 14(8): 2490–2501.

Li J, Wang J, Wang S, et al. (2020) Parallel structure of six wheel-legged robot trajectory tracking control with heavy payload under uncertain physical interaction. *Assembly Automation*. DOI: 10.1108/aa-08-2019-0148

Li Z, Huang B, Ajoudani A, et al. (2017) Asymmetric bimanual control of dual-arm exoskeletons for human-cooperative manipulations. *IEEE Transactions on Robotics* 34(1): 264–271.

Li Z, Ren Z, Zhao K, et al. (2019) Human-cooperative control design of a walking exoskeleton for body weight support. *IEEE Transactions on Industrial Informatics* 16(5): 2985–2996.

Liu Z, Karimi HR and Yu J (2020) Passivity-based robust sliding mode synthesis for uncertain delayed stochastic systems via state observer. *Automatica*. DOI: 10.1016/j.automatica.2019.108596

Mario MO (2018) Human activity recognition based on single sensor square HV acceleration images and convolutional neural networks. *IEEE Sensors Journal* 19(4): 1487–1498.

Mathur A and Foody GM (2008) Multiclass and binary svm classification: Implications for training and classification users. *IEEE Geoscience and Remote Sensing Letters* 5(2): 241–245.

Milstein A, Ganel T, Berman S and Nisky I (2018) Human-centered transparency of grasping via a robot-assisted minimally invasive surgery system. *IEEE Transactions on Human-Machine Systems* 48(4): 349–358.

Mondéjar-Guerra V, Novo J, Rouco J, et al. (2019) Heartbeat classification fusing temporal and morphological information of ECGS via ensemble of classifiers. *Biomedical Signal Processing and Control* 47: 41–48. DOI:10.1016/j.bspc.2018.08.007

Naik GR, Selvan SE, Arjunan SP, et al. (2018) An ICA-EBM-based semg classifier for recognizing lower limb movements in individuals with and without knee pathology. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 26(3): 675–686.

Norgaard S, Saeedi R and Gebremedhin AH (2019) Multi-sensor time-series classification for activity tracking under variable length. *IEEE Sensors Journal* 20(5): 2701–2709.

Qi W and Aliverti A (2019) A multimodal wearable system for continuous and real-time breathing pattern monitoring during daily activity. *IEEE Journal of Biomedical and Health Informatics*. DOI: 10.1109/JBHI.2019.2963048

Qi W, Su H and Aliverti A (2020) A smartphone-based adaptive recognition and real-time monitoring system for human activities. *IEEE Transactions on Human-Machine Systems* 50(5): 414–423.

Qi W, Zong G and Karimi HR (2019) Finite-time observer-based sliding mode control for quantized semi-markov switching systems with application. *IEEE Transactions on Industrial Informatics* 16(2): 1259–1271.

Reyes-Ortiz JL, Oneto L, Samà A, et al. (2016) Transition-aware human activity recognition using smartphones. *Neurocomputing* 171: 754–767. DOI:10.1016/j.neucom.2015.07.085

Sandoval J, Su H, Vieyres P, et al. (2018) Collaborative framework for robot-assisted minimally invasive surgery using a 7-dof anthropomorphic robot. *Robotics and Autonomous Systems* 106: 95–106. DOI:10.1016/j.robot.2018.04.001

Su H, Hu Y, Karimi HR, et al. (2020a) Improved recurrent neural network-based manipulator control with remote center of motion constraints: Experimental results. *Neural Networks* 131: 291–299. DOI:10.1016/j.neunet.2020.07.033

Su H, Mariani A, Ovur Salih E, et al. (2021a) Towards teaching by demonstration for robot-assisted minimally invasive surgery. *IEEE Transactions on Automation Science and Engineering*. DOI: 10.1109/TASE.2020.3045655

Su H, Qi W, Hu Y, et al. (2021b) An incremental learning framework for human-like redundancy optimization of anthropomorphic manipulators. *IEEE Transactions on Industrial Informatics*. DOI:10.1109/TII.2020.3036693.

Su H, Qi W, Yang C, et al. (2020b) Deep neural network approach in robot tool dynamics identification for bilateral teleoperation. *IEEE Robotics and Automation Letters* 5(2): 2943–2949.

Tatinati S, Veluvolu KC and Ang WT (2014) Multistep prediction of physiological tremor based on machine learning for robotics assisted microsurgery. *IEEE Transactions on Cybernetics* 45(2): 328–339.

Tufek N, Yalcin M, Altintas M, et al. (2019) Human action recognition using deep learning methods on limited sensory data. *IEEE Sensors Journal* 20(6): 3101–3112.

Wang Y, Cang S and Yu H (2018) A data fusion-based hybrid sensory system for older people's daily activity and daily routine recognition. *IEEE Sensors Journal* 18(16): 6874–6888.

Wang Y, Karimi HR, Lam HK and Yan H (2019) Fuzzy output tracking control and filtering for nonlinear discrete-time descriptor systems under unreliable communication links. *IEEE Transactions on Cybernetics* 50(6): 2369–2379.

Yang C, Chen C, Wang N, et al. (2018) Biologically inspired motion modeling and neural control for robot learning from demonstrations. *IEEE Transactions on Cognitive and Developmental Systems* 11(2): 281–291.