Journal of Engineering and Sustainable Development
https://jeasd.uomustansiriyah.edu.iq/index.php/jeasd
https://doi.org/10.31272/jeasd.26.5.3

Vol. 26, No. 05, September 2022
ISSN 2520-0917

CLUSTERING SURGEMES USING PROTOTYPES FROM ROBOTIC KINEMATIC INFORMATION

***Safaa Albasri[1]** saaxfc@umsystem.edu
Omar Ibrahim[1] oai9bc@mail.missouri.edu
Mihail Popescu[2] PopescuM@health.missouri.edu
James Keller[1] KellerJ@missouri.edu

1) Electrical Engineering and Computer Science, University of Missouri-Columbia, Columbia, MO 65211 USA
2) Health Management and Informatics, University of Missouri-Columbia, Columbia, MO 65211 USA

Submitted 22/5/2022 ***Accepted in revised form 16/8/2022*** ***Published 1/9/2022***

Abstract: Training a surgeon to be skilled and competent to perform a given surgical procedure is essential in providing a high quality of care and reducing the risk of complications. However, existing training techniques limit us from conducting in-depth analyses of surgical motions to evaluate these skills accurately. We develop a method to identify the gestures by applying unsupervised methods to cluster the surgical activities learned directly from raw kinematic data. We design an unsupervised method to determine the surgical motions in a Suturing procedure based on predefined surgical gestures. The first step is to find the prototypes by clustering the surgemes of the expert surgeon from all the same expert trials. Then, we map the other surgeons surgemes to the nearest representative of the prototypes and report the clustering accuracy by employing the rand index technique. We utilize four techniques in our proposed unsupervised approach for gesture clustering based on Hierarchical and FCM algorithms. In addition, we highlight the advantages of representing time series data before clustering in terms of computation time saving and system complexity reduction, respectively.

Keywords: *DTW, RMIS, FCM, Ward, Rand-Index, Surgemes, Calinski Harabasz, Xie-Beni, and Clustering.*

1. Introduction

The innovations in surgical robotic platforms have opened new training and education capabilities for surgeons to provide high-quality surgical care in the operation room. In addition, the information captured by robotic minimally invasive surgery (RMIS) delivers program-based insights that could potentially help enhance patient outcomes and care costs [1]. Also, the accessibility of the driven data representing the movement of the surgeons gives the opportunities to create and build models for objective methods and assessments that deliver feedback during a surgical task [2].

Most techniques are supervised classification based on predefined or pre-segmented gesture data. These surgical gestures (surgemes) are annotated manually by chief surgeons, consuming more time and being susceptible to human mistakes by missing parts (surgemes) or inconsistently applied criteria throughout a surgical task [3, 4]. Several works intended to identify surgical activities from unsupervised viewpoints without prior knowledge of gestures [2, 5, 6]. Despinoy et al. [5] proposed a framework for segmentation and recognition surgemes from kinematic data. They first applied unsupervised segmentation by finding a relevant selection of dexemes (a numerical representation of subgestures to perform a surgemes). Secondly, they used learning features from dexemes to associate them with corresponding surgemes (composed of a set of dexemes) [5]. Another

**Corresponding Author: saaxfc@umsystem.edu*

approach introduced by Fard et al. [7] is known as soft boundary unsupervised surgemes segmentation. The temporal sequence of surgemes segment and merge based on some criteria, and then the boundaries between parts are smoothed. A recent deep-learning approach was proposed by Murali et al [8], based on a deep convolution network using both kinematic and video data for surgical gesture segmentation.

This paper aims two folds: i) we propose an unsupervised method to identify the surgemes of the surgeons based on clustering algorithms. ii) we re-represent the segments by utilizing the mean of the feature instead of using all the time frames to reduce the complexity and computational time and make the proposed approach more feasible.

2. Methodology

2.1. Surgemes Clustering Framework

Fig. 1 shows the overall flow diagram of our proposed approach for surgical gesture clustering based on raw kinematic data. First, the surgemes prototypes were obtained from the expert surgeon by clustering their surgemes from all trials using unsupervised algorithms. In the second step, we utilize the medoid method to individually locate the representative surgeme for each surgeme prototype. Next, we mapped every gesture of the trainee surgeon per trial on the representative surgemes by measuring the distance with all the representative gestures and assigned it to the cluster with the smallest distance. Finally, we assess the performance of our unsupervised approach using the rand-index between the ground truth and the predicted labels of the clustering approach. More details about each step will be discussed in the following sections.

This approach starts by normalizing each surgeme using mean and variance to ensure that the data are scale and shift invariants which allow reasonable comparison between them.

Let X_i be a time series, then the corresponding normalized signal $\hat{X}_i$ is:

$$\hat{X}_i = \frac{(X_i - \mu_i)}{\sigma^2{}_i} \quad (1)$$

Where μ_i and σ_i are the arithmetic mean and standard deviation of time series i, respectively. Next, to form the prototype surgemes, we employ unsupervised methods on one expert surgeon gestures using hierarchical and Fuzzy c-means (FCM) Algorithms.

The hierarchical clustering method has been shown to be effective and efficient at separating human activities, which is well-suited for time series clustering. [9, 10]. Then, we employ the minimum variance algorithm (Ward) on a pairwise distance matrix which is obtained by computing the distance between two segments to create the prototype surgemes. The distance d_{rs} between two clusters Cr and Cs is defined as the

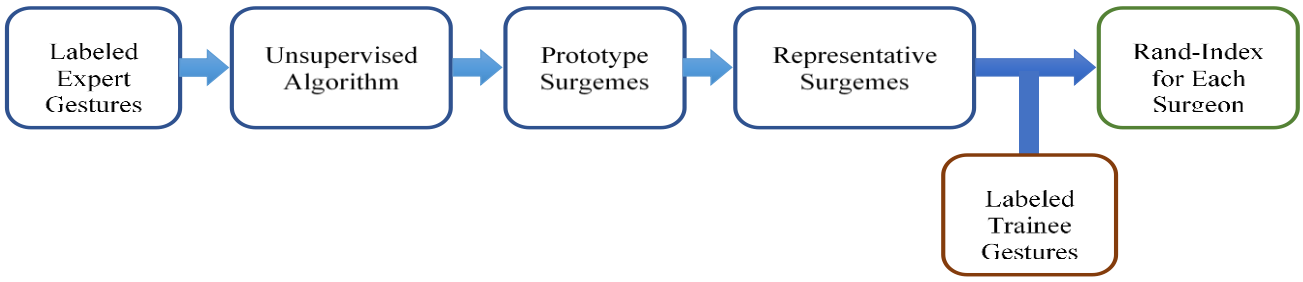


Figure 1: Overview of our Clustering Surgemes approach using Rand-Index for each surgeon.

distance of their centroid is equivalent to the following equation:

$$d_{rs} = \frac{n_r n_s}{n_r + n_s} \left\| \overline{X}_r - \overline{X}_s \right\|^2 \qquad (2)$$

Where $\overline{X}_r$ and $\overline{X}_s$ are the centroids of the two clusters, $n_r$and n_s are the number of objects in cluster Cr and Cs, respectively [11]. Additionally, we used fuzzy c-means (FCM) to partition the expert surgeon's surgemes into a predefined number of clusters equal to the number of distinct surgemes in each surgical task.

The FCM partition membership needs to meet the following constraint to prevent the trivial solution by allocating all the cluster memberships to zero:

$$\sum_{i=1}^{C} u_{ij} = 1 \qquad \forall\, j = 1,2 \ldots n \qquad (3)$$

The objective function of the FCM that meets the criteria can be formulated as follow:

$$J(U,V) = \sum_{j=1}^{n} \sum_{i=1}^{C} u_{ij}^{m}\, d^2(x_j, v_i) \qquad (4)$$

The parameter m>1 is the fuzzifier that controls the rate of the membership value. The values of partition membership u_{ij} and prototype centers that require minimizing J and the distance between data sample x_j and the set of cluster centers v_i can be determined by the following equations [12]:

$$u_{ij} = \frac{(1/d(x_j, v_i))^{2/(m-1)}}{\sum_{k=1}^{C} (1/d(x_j, v_k))^{2/(m-1)}} \qquad (5)$$

$$v_i = \frac{\sum_{j=1}^{n} u_{ij}^{m}\, x_j}{\sum_{j=1}^{n} u_{ij}^{m}} \qquad (6)$$

The distance measure plays a vital role in time series clustering, where the data are grouped based on their similarity. The Euclidean distance and dynamic time warping (DTW) is the most frequently used similarity measurement due to its effectiveness and efficiency in determining the similarity between objects. Euclidian distance is simple, fast, and parameter-free. However, it is sensitive to noise and shifts in the time axis. It also requires both signals to have equal time lengths (one-to-one matching) [13]

On the other hand, DTW employs a one-to-many matching between time axes without considering local and global shifting issues in the time series data, which overcomes these restrictions. By resolving this time scale issue (local shift), it is possible to match time series data similar in pattern but have a different time axis.[14].

DTW distance can be implemented using dynamic programming in which the accumulated distance formula recursively computes the optimal warp path between two segments $X_i = [x_1 \ldots, x_M]$ and $Y_i = [y_1 \ldots, y_N]$:

$$D(X_i, Y_j) = \delta(x_i, y_j) + \min \begin{Bmatrix} D(x_{i-1}, y_{j-1}), \\ D(x_i, y_{j-1}), \\ D(x_{i-1}, y_j) \end{Bmatrix} \qquad (7)$$

where $\delta(x_i, y_j)$ is the Euclidian distance between the two aligned segments of the warp path (that give the minimum distance) [15].

To evaluate how well our proposed approach works for clustering surgemes, we compare the resulting predicted labels with ground truth labels using the rand-index. The higher the rand-index value, the better performing.

2.2. Rand-Index Performance Evaluation

Unsupervised performance evaluation is not an easy task to assess the cluster results in the absence of data labels. However, for the dataset [16], a senior specialist in robotic surgery provided the manually segmented references.

The most common quality measure in the domain of time series clustering is the Rand index [17]. We used the Rand index criteria between the predicted result labels of our proposed framework and the ground truth surgeme labels. The Rand index values are between 0 and 1, where one indicates the two surgemes are identical or precisely the same. The Rand index criteria between the ground truth labels and the predicted labels are defined as the measure of the ratio of the correct decisions taken by the approach. In other viewpoints, it can be defined as the number of agreements between two groups, G and Y, over the total number of pairs (agreements and disagreements), which can be calculated using the following equation [17, 18]:

$$\mathrm{RI(G,Y)} = \frac{\mathrm{T_P} + \mathrm{T_N}}{\mathrm{T_P} + \mathrm{F_p} + \mathrm{T_N} + \mathrm{F_N}} \qquad (8)$$

where T_P, F_p, T_N and F_N are the corresponding number of true positives, false positives, true negatives, and false negative results, respectively.

2.3. Calinski Harabasz Validity Index

The variance ratio criterion (VRC), known as the Calinski Harabasz (CH) index, is computed for K clusters and N data points as:

$$VRC = \frac{trace_B}{trace_W} \times \frac{(N-K)}{(K-1)} \qquad (9)$$

Where $trace_B$ and $trace_W$ are the overall between-cluster variance and within cluster variance, respectively.

The overall between-cluster $trace_B$ can be written as [19]:

$$trace_B = \sum_{i=1}^{K} n_i \|C_i - C\|^2 \qquad (10)$$

Where $C_i$is the centroid of cluster i, n_i is the number of observations in cluster i, and C is the centroid of the entire sample data.

The overall within-cluster variance $trace_W$ is defined as:

$$trace_W = \sum_{i=1}^{K} \sum_{j=1}^{n_i} \|x_j - C_i\|^2 \qquad (11)$$

Clusters that are well-defined clusters will have a high variance between clusters variance $trace_B$ and a small variance within-cluster $trace_W$. The higher the CH ratio, the better the data partitioning will be. The solution with the highest Calinski-Harabasz index value is the one that has the optimal number of clusters [19, 20].

2.4. Xie-Beni Validity Index

The ratio of the compactness of the fuzzy c-partition to its separation is called the compactness and separation validity function or well-known as the Xie-Beni index, which can be computed as [21]:

$$XB = \frac{\sum_{i=1}^{C} \sum_{j=1}^{n} u_{ij}^2 \left\|V_i - X_j\right\|^2}{n \min_{i,j} \left\|V_i - V_j\right\|^2} \qquad (12)$$

Where n is the number of data points. The fuzzy centroid V_i $(i = 1, 2 \ldots c)$ and the fuzzy membership u_{ij} of X_j belonging to cluster i are calculated using

$$V_i = \frac{\sum_{j=1}^{n} u_{ij}^m X_j}{\sum_{j=1}^{n} u_{ij}^m} \qquad (13)$$

$$u_{ij} = \frac{\left(\frac{1}{\|X_j - V_i\|}\right)^{\frac{1}{m-1}}}{\sum_{i=1}^{c} \left(\frac{1}{\|X_j - V_i\|}\right)^{\frac{1}{m-1}}} \qquad (14)$$

A lower value of XB implies a partition in which all the clusters are compact and separate. Thus, the smaller values of XB correspond to the optimal number of clusters [19, 21].

3. Experimental Results

3.1. Dataset

The experiments on the proposed approach are conducted with a general and public surgery dataset [16]. This dataset includes both kinematic and video information from eight different expert surgeon levels: expert, intermediate, and novice surgeons. Each surgeon repetitively performed three basic surgery tasks (suturing, needle passing, knot tying) five times (known as a trial). We used only the raw kinematic data captured at 30Hz from the da Vinci robotic surgery system with different trial frame lengths. There is 76 dimensions or variables information to describe the kinematics for all four manipulators. Each manipulator has 19 variables that consist of 3 cartesian positions, nine rotation matrices, three linear velocities, three angular velocities, and one gripper angle [16].

This dataset has manually annotated ground truth segments (surgemes) for each trial at every task. Each annotation provides the label of the surgeme, the start, and the end frames in the kinematic data allocated for each trial. In particular, the common vocabulary of potential surgemes comprises 15 elements and is listed with their description in Table 1 for the three surgical tasks. Some surgemes seem to be in more than one task; however, the background environment differs between tasks [16]. Therefore, even though there are a combined total of 15 surgemes, not necessarily all of them show up in one surgical task. For example, suturing, needle passing, and knot tying include 10, 8, and 6 of the 15 surgemes, respectively.

Table 1: Surgemes Vocabulary for all the surgical tasks [16].

Surgeme index	Gesture description	Suturing	Needle-Passing	Knot-Tying
G1	Reaching for a needle with right hand	☑	☑	☑
G2	Positioning needle	☑	☑	
G3	Pushing a needle through tissue	☑	☑	
G4	Transferring needle from left to right	☑	☑	
G5	Moving to the center with the needle in the grip	☑	☑	
G6	Pulling suture with the left hand	☑	☑	
G8	Orienting needle	☑	☑	
G9	Using the right hand to help tighten the suture	☑		
G10	Loosening more suture	☑		
G11	Dropping suture at the end and moving to endpoints	☑	☑	☑
G12	Reaching for a needle with left hand			☑
G13	Making C loop around the right hand			☑
G14	Reaching for suture with the right hand			☑
G15	Pulling suture with both hands			☑
	Number of Gestures in each task	**10**	**8**	**6**

3.2. Gestures Representation and Visualization

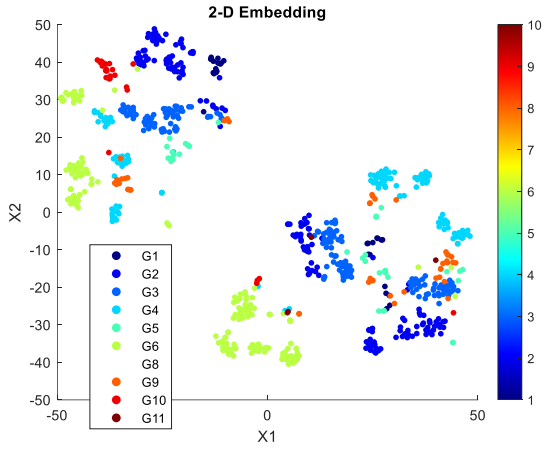
We re-represent the surgemes of each trial for every surgical task by computing the mean of each 76 variables separately (rather than using the whole segment time frames for each component. Averaging each segment simplifies the computation and eliminates the bias of the time. Furthermore, this approach allows us to use any similarity measure rather than only DTW since surgemes have identical dimensions.

We visualize the surgemes for all the trials of the suturing task in Fig. 2 using the t-Distributed Stochastic Neighbor Embedding (t-SNE) by reducing the high-dimensional space of the surgemes to a low-dimensional map of two or three dimensions [22]. Fig. 2 illustrates a reasonable and precise separation of the surgemes, even though some gestures occur in multiple locations because it is also dependent on the surgeon's skill level.

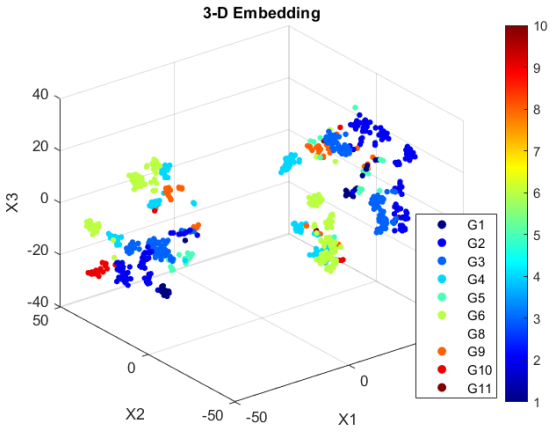
3.3. Hierarchical Clustering Results

We develop our surgeme clustering approach directly onto the raw kinematic data to prevent excessive pre-processing. We run three experiments based on hierarchical clustering (with Ward linkage) on the JIGSAWS dataset. In the first two sets, we used the DTW as a distance measure between the surgemes, while in the third set, we employed the Euclidean distance. The results are reported based on the framework discussed in **Error! Reference source not found.**1 by applying the unsupervised learning approach.

In the first experiment, we cluster all the surgemes of one expert surgeon utilizing the Ward Hierarchical clustering to find the prototype surgical gestures for each cluster. The raw surgemes are a time series with 76-dimensional and different time lengths. We have chosen the surgemes of an expert surgeon with the highest Global Rating Scores (GRS) among the expert surgeons. Because the higher scores of expert surgeons have resulted from the consistency and smoothness of their trajectories during the surgical task, leading to a better clustering outcome. Hence, we employed DTW as a practical and feasible pairwise distance measure between any two surgemes in this case.



(a)



(b)

Figure 2: t-SNE visualization of surgeme labels in Suturing task (a) 2-dimension, and (b) 3-dimension embedding.

The rand-index measures the agreement between the ground truth and clustered gestures. But first, we evaluate the optimal number of clusters using the Calinski-Harabasz clustering evaluation criterion, as illustrated in **Error! Reference source not found.**3(a). The plot shows that the highest Calinski-Harabasz value occurs at

eleven, suggesting that the optimal number of clusters is eleven in this case of Ward linkage clustering. **Error! Reference source not found.**3(b) shows the Rand-Index values change with the number of clusters. We can observe that the highest value of the Rand-Index is achieved when we use eleven clusters in the Hierarchical algorithm that fits with the Calinski-Harabasz criterion. The rand index resulting from clustering the expert data intended to find the prototypes is 92%.

At this point, each prototype has candidate surgemes to be a representative surgical gesture within the cluster. We used the medoid to locate the representative for each prototype by computing the DTW distance among the same cluster members and then finding the minimum distance relevant to the elected representative surgeme. The Medoid technique is used here because the surgemes have different lengths. Therefore, employing centers instead of the medoid to assign representatives are not allowable. Each cluster has a group of surgemes and one representative surgeme representing this group. Finally, we stream each trainee surgeon's sample point (surgeme) per trial. We measure the DTW distance between each new data sample and the prototype representative and assign it to the nearest cluster.

Error! Reference source not found.4(a) presents the average rand-index results of the proposed method for clustering the surgemes for each trainee surgeon intended for the suturing task. At the same time, the rand-index results per trial for each surgeon are shown in **Error! Reference source not found.**(b). For example, surgeon "E" has the highest average rand-index of 96% accuracy because this surgeon was the prototype clustering surgeon. Also, from this figure, we observed that our proposed method could cluster the surgemes of the surgeon who mapped each trial to the representative surgemes.

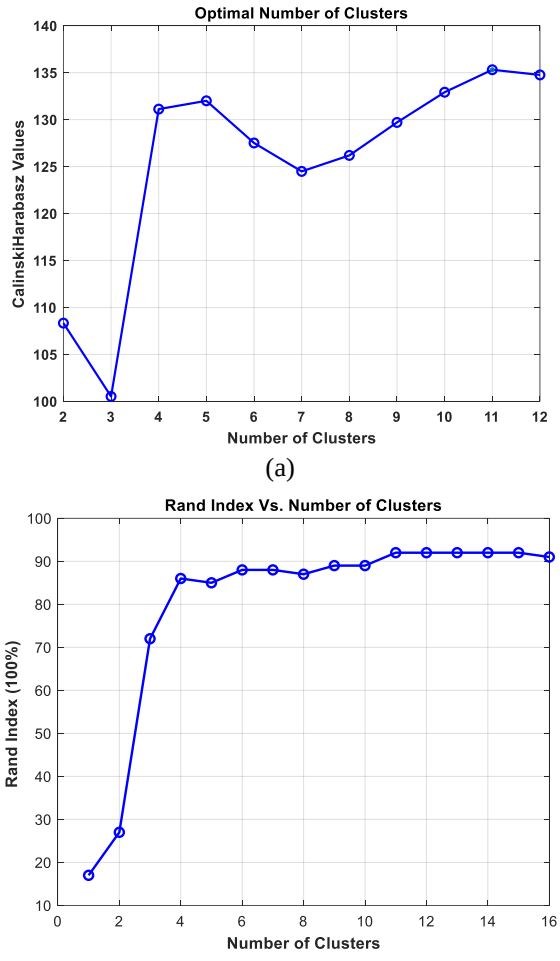
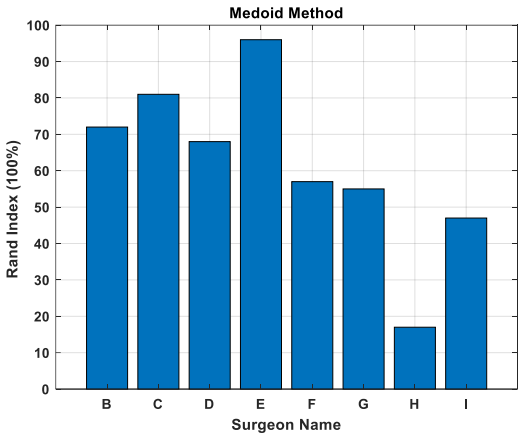
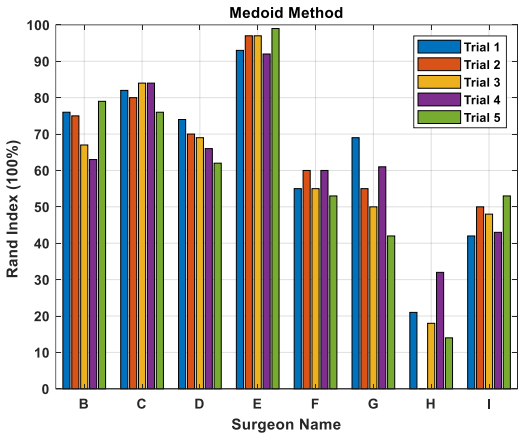


Figure 3: (a) Calinski-Harabasz clustering evaluation criterion, (b) Rand-Index plot as a function of the number of clusters.

We implemented another experiment by considering each expert's surgeme as a representative member. This can be done by using the expert's ground truth surgemes as a prototype which results in grouping them according to their labels rather than clustering them.



(a)



(b)

Figure 4: Rand-Index Results for Ward clustering using medoid (a) Average Rand Index (b) Rand-Index per trial.

First, we calculate the pairwise distance between all the surgemes of a query surgeon and each group of surgemes that belong to the expert separately. For example, let us have ten surgemes of different labels belonging to the query surgeon, and the expert surgeon has three clusters of different surgemes, each with a cluster size of 2, 4, and 5. Then, we measure the distance between the ten surgemes and the members of each cluster, which results in three distance matrices of the size of 2x10, 4x10, and 5x10 dimensions, respectively. Secondly, we average the distance matrix to each cluster, resulting in an array of 1x10. Then, we concatenate the resulted mean distance arrays in one matrix and map each surgemes to its closest group. In all the experiment steps, we applied DTW as our distance measure. Also, note that the same number of clusters were used in both experiments.

Error! Reference source not found. illustrates the comparison utilizing the average Rand index between the first experimental results that use the medoid gesture to represent each cluster. The second experiment uses all cluster members as representatives for that cluster to assign the labels of the test samples (surgemes). We can observe that the results are very close to each other for most surgeons. Still, the second approach performs better than the medoid because it considers all the members in the cluster, and it reduces the possibility of having a lousy cluster representative. Besides, the second approach uses the ground truth labels to build the prototypes of the expert surgeon instead of clustering, which results in an average Rand index of around 100%, as seen in **Error! Reference source not found.**, surgeon ("E").

We also investigate the performance of our proposed approach by employing the mean features of the surgeme mentioned before as an alternative to using all the time frames. In this case, a surgeme of length (Nx75) will be mapped to (1x75), making all the surgemes have the same size in the 75-dimensional space. Thus, we can employ Euclidean distance as our measure instead of DTW. The distance between time series is calculated using Euclidean distance because the sequences are identical in length. Using the DTW is impractical here due to its high complexity.

Figure 5: Comparison of average Rand-Index between using medoid and mean GT in representing the surgemes in suturing task.

The histogram of the expert surgemes at the SU task is presented in **Error! Reference source not found.**6. We can observe that the expert surgeon never performed G10, and two gestures (G8 and G9) were performed just one time during the entire five trials. Consequently, the optimal number of clusters is seven instead of nine for the expert gestures.

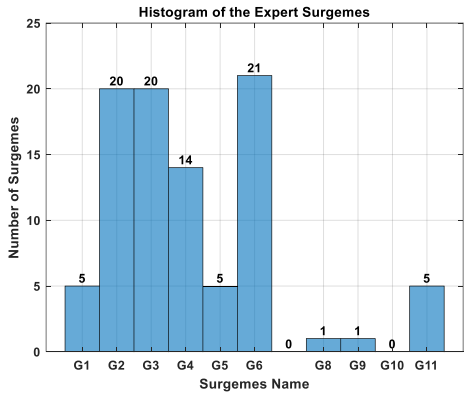
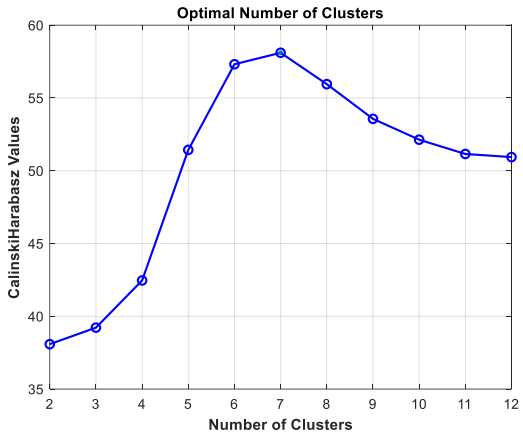
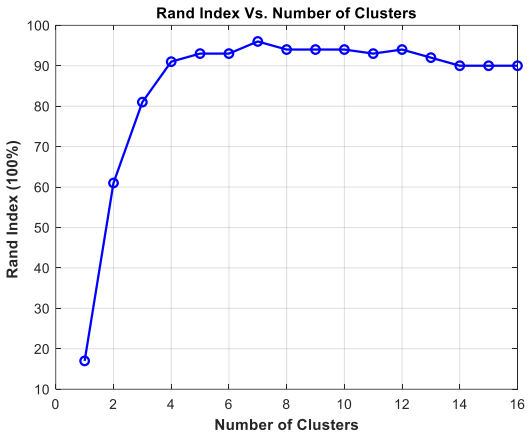


Figure 6: Histogram of the expert surgeon surgemes in SU task.

We utilize the Calinski-Harabasz clustering evaluation criterion to find the optimal number of clusters using the mean feature, as illustrated in **Error! Reference source not found.**. For example, the number of clusters chosen to be seven will give a higher average Rand-Index of 96%.



(a)



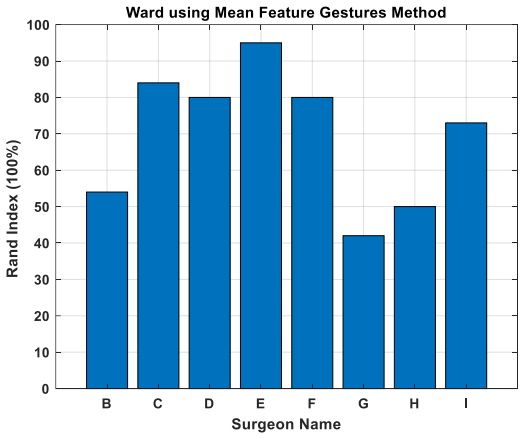
(b)

Figure 7: Mean feature of the gestures (a) Calinski-Harabasz clustering evaluation criterion, (b) Rand-Index plot as a function of the number of clusters.

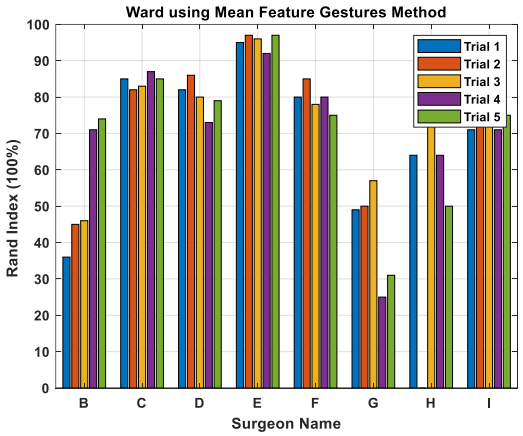
Error! Reference source not found.8 demonstrates using the mean feature technique to implement the proposed framework through average rand-index in (a) and per trial in (b) of **Error! Reference source not found.**. We can observe from the results in this figure that using the mean of the feature is more accurate than the results of the previous experiments. Another point worth mentioning is that the Rand-index results of the expert and intermediate surgeons are distinct from those of the novice surgeons.

This indicates the ability to distinguish their surgemes pattern, which is close to the model of clustering the expert surgeon. Also, the result reveals that enhanced clustering quality is

reachable without reducing the time-series features by using the mean feature technique.



(a)



(b)

Figure 8: Rand-Index results for the Ward clustering using mean features for each surgeon (a) Average Rand Index (b) Rand-Index per trial in suturing task.

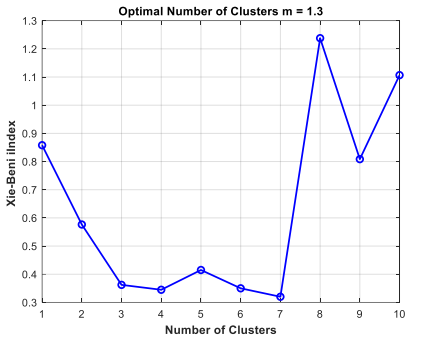
:

3.4. Fuzzy C-Mean (FCM)

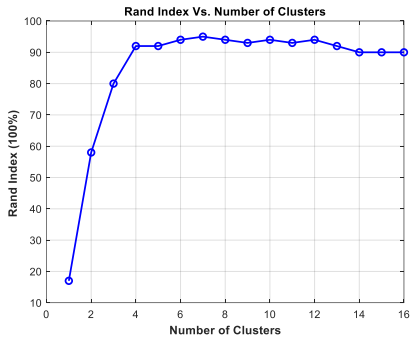
We conducted another experiment to investigate the use of the FCM algorithm to cluster the prototype surgical gestures. As mentioned previously, FCM is a clustering method wherein each surgeme belongs to multiple groups by a membership grade. In this experiment, we applied the FCM to obtain the prototype surgemes by clustering the surgemes of the expert surgeon (clustering model). We develop our proposed framework on raw kinematic data from the JIGSAWS dataset, which was used to perform suturing surgical tasks. We employed the mean features representation technique of the surgemes before clustering due to the time computation and low complexity. Therefore, Euclidean distance is used as a distance measure in this case rather than the DTW. Finally, we compare the clustering results of the surgeon surgemes with those manually annotated by a senior expert surgeon.

We employ a popular validity index in FCM, the Xie-Beni index criteria [19], to measure the optimal number of clustering as shown on the left of **Error! Reference source not found.** and the Rand-Index accuracy to the right of the exact figure. Therefore, the number of clusters chosen is seven that reached both the high Xie-Beni index and Rand-Index. The fuzzifier controlling the partitioning overlap, the small value of m approaches one means more crisp boundaries and less overlap. For the FCM algorithm, we run an experiment for different fuzzy membership m with the Xie-Beni validity index, and we set m to 1.3. For prototype surgemes, we observed that clustering of the expert surgemes achieved %95 of the rand-index accuracy compared with the ground truth.

Using the clustering method to build prototype surgemes and employing similarity measures to select representative surgeme significantly impacts the Rand index outcomes. **Error! Reference source not found.** shows the best results of the surgemes clustering using our proposed method based on the FCM algorithm. Consequently, we can see that the overall accuracy improved of the FCM compared to the experiments-based hierarchical ward clustering algorithm using DTW distance.

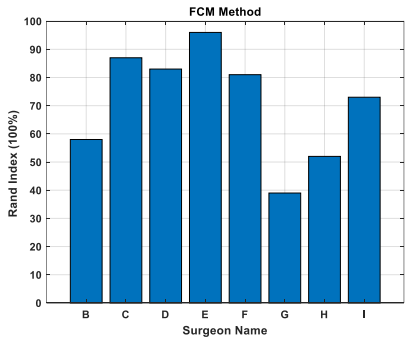


(a)

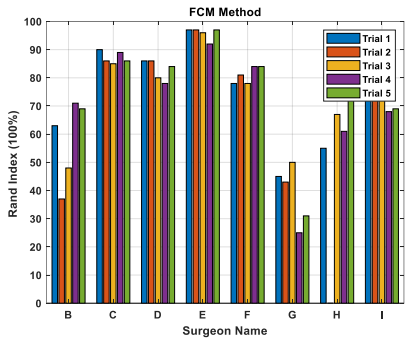


(b)

Figure 9: FCM using Mean Feature of the gestures (a) Xie-Beni validity index, (b) Rand-Index plot as a function of the number of clusters.



(a)



(b)

Figure 10: Rand-Index results for The FCM clustering using mean features for each surgeon (a) Average Rand Index (b) Rand-Index per trial in suturing task.

In **Error! Reference source not found.**, we compared the results obtained by the proposed method using all the techniques mentioned in the methodology section through the rand-index. We can observe that our approach using the mean feature method performs better than the other clustering approaches in most cases. Additionally, the clustering methods based on the mean feature surgemes representation with Euclidean distance outperform the clustering techniques that use DTW as a distance measure. It is also important to mention that the enhanced clustering quality is achievable even with the reduction in the time instance while preserving the dimension of the variable unchanged.

Furthermore, it is crucial to consider the effect of the gesture time to accomplish the surgical task, where short surgeme are challenging to discriminate, resulting in decreased clustering performance. Additionally, the insufficient data for some surgeons in specific trials makes it difficult for any model to differentiate the surgeon surgemes from the prototypes of the expert surgeon.

Figure 11: Comparison of the Rand-Index between different surgemes clustering methods for each surgeon in suturing task.

4. Conclusions and Future Work

Surgical gestures are the key elements in a surgeon's training system, and they can offer a quantitative measurement and feedback to the trainee during the robotic surgical session. We proposed a new unsupervised approach for surgemes clustering by utilizing four techniques based on Hierarchical and FCM algorithms. We evaluated our method on a real dataset by analyzing raw kinematics data from suturing tasks performed by individuals with varying levels of expertise. Also, we demonstrated the benefits of re-representing the time series data before clustering in terms of computation time reduction and system complexity.

One of the most challenging tasks in unsupervised learning is to deal with outliers. For example, some surgeons perform surgemes that the expert surgeons generally do not operate. This might adversely affect the quality of surgeme assignment to the appropriate cluster, thereby influencing the clustering algorithm's outcome.

In addition, we used a predetermined number of clusters. Therefore, future research should use a clustering technique to deal with unknown groups derived from expert prototypes. Nevertheless, this effort constitutes a step forward toward surgemes segmentation. One of the future work challenges will be to build a clustering method based on using one of the experts' surgemes for model initialization and then map the surgemes from other subjects.

Conflict of Interest

The authors confirm that the publication of this article cause no conflict of interest.

5. References

1. I. Surgical. https://www.intuitive.com/en-us (accessed.
2. Y. Gao, S. S. Vedula, G. I. Lee, M. R. Lee, S. Khudanpur, and G. D. Hager, "Unsupervised surgical data alignment with application to automatic activity annotation," 2016: IEEE, doi: 10.1109/icra.2016.7487608. [Online]. Available: https://dx.doi.org/10.1109/icra.2016.7487608
3. M. J. Fard, "Computational modeling approaches for task analysis in robotic-assisted surgery," Wayne State University, 2016 .
4. J. Reason, *Human error*. Cambridge university press, 1990.
5. F. Despinoy *et al.*, "Unsupervised trajectory segmentation for surgical gesture recognition in robotic training," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 6, pp. 1280-1291, 2015.
6. R. Dipietro and G. D. Hager , "Unsupervised Learning for Surgical Motion by Learning to Predict the Future," Springer International Publishing, 2018, pp. 281-288.
7. M. J. Fard, S. Ameri, R. B. Chinnam, and R. D. Ellis, "Soft Boundary Approach for Unsupervised Gesture Segmentation in Robotic-Assisted Surgery," *IEEE Robotics and Automation Letters*, vol. 2, no. 1, pp. 171-178, 2017, doi: 10.1109/lra.2016.2585303.
8. A. Murali *et al.*, "TSC-DL: Unsupervised trajectory segmentation of multi-modal surgical demonstrations with Deep Learning," 2016: IEEE, doi: 10.1109/icra.2016.7487607. [Online]. Available: https://dx.doi.org/10.1109/icra.2016.7487607
9. S. Hirano and S. Tsumoto, "Empirical comparison of clustering methods for long

time-series databases," in *Active Mining*: Springer, 2005 ,pp. 268-286.

10. T. Oates, M. D. Schmill, and P. R. Cohen, "A method for clustering the experiences of a mobile robot that accords with human judgments," in *AAAI/IAAI*, 2000, pp. 846-851 .
11. S. Theodoridis and K. Koutroumbas, "Pattern recognition, edition," ed: Academic Press, fourth edition Edition, 2009.
12. J. M. Keller, D. Liu, and D. B. Fogel, *Fundamentals of computational intelligence: neural networks, fuzzy systems, and evolutionary computation*. John Wiley & Sons, 2016.
13. S. Albasri, M. Popescu, and J. Keller, "A Novel Distance for Automated Surgical Skill Evaluation," in *2019 E-Health and Bioengineering Conference (EHB)*, 2019: IEEE, pp. 1-6 .
14. S. Salvador and P. Chan, "Toward accurate dynamic time warping in linear time and space," *Intelligent Data Analysis*, vol. 11, no. 5, pp. 561-580, 2007.
15. S. Albasri, M. Popescu, and J. Keller, "Surgery Task Classification Using Procrustes Analysis," in *2019 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, 2019: IEEE, pp. 1-6 .
16. Y. Gao *et al.*, "Jhu-isi gesture and skill assessment working set (jigsaws): A surgical activity dataset for human motion modeling," in *Miccai workshop: M2cai*, 2014, vol. 3, p. 3 .
17. M. Chiş, S. Banerjee, and A. E. Hassanien, "Clustering time series data: an evolutionary approach," in *Foundations of Computational, IntelligenceVolume 6*: Springer, 2009, pp. 193-207.
18. W. M. Rand, "Objective criteria for the evaluation of clustering methods," *Journal of the American Statistical association*, vol. 66, no ,336 .pp. 846-850, 1971.
19. U. Maulik and S. Bandyopadhyay, "Performance evaluation of some clustering algorithms and validity indices," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 24, no. 12, pp. 1650-1654, 2002.
20. T. Caliński and J. Harabasz, "A dendrite method for cluster analysis," *Communications in Statistics-theory and Methods*, vol. 3, no. 1, pp. 1-27, 1974.
21. X. L. Xie and G. Beni, "A validity measure for fuzzy clustering," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 13, no. 8, pp. 841-847, 1991.
22. L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of machine learning research*, vol. 9, no. 11, 2008.