CHAPTER 5

Applications

This chapter, we apply five proposed methods, i.e., sgFCMed, sgFPCMed, sgPFCMed, sgUPCMed and sgUPFCMed for real applications, i.e., Thai sign language translation and identification of cardio-pulmonary resuscitation activity in medical simulation videos.

5.1 Thai sign language translation

The one of the communication method for Thai hearing impaired people is Thai sign language (TSL) [60]. TSL is a way to communicate with deaf people through hand gestures. Since the most hearing people cannot understand their sign language, the communication between deaf and hearing people is difficult. To solve this problem, hand sign language translation is expected to assist the deaf or hearing impaired people communicate with hearing people.

Our objective of this study is to improve dynamic Thai sign language translation system with video caption without prior hand region detection and segmentation using Scale Invariant Feature Transform (SIFT) [61] and various string grammar fuzzy clustering. The SIFT use to match test frame with symbols in the signature library. String grammar fuzzy clustering, i.e., sgFCMed, sgFPCMed, sgPFCMed, sgUPCMed, sgUPFCMed and fuzzy K-nearest neighbor (FKNN) [62] are used to find a matched sign words.

5.1.1 Dataset

In our experiment dataset (training and test video data set) is collected from 25 subjects at different times of day for several days. The dataset consists of 10 hand sign words (classes), i.e., ‘‘elder’’, ‘‘grandfather’’, ‘‘grandmother’’, ‘‘gratitude’’, ‘‘female’’, ‘‘male’’, ‘‘glad’’, ‘‘thank you’’, ‘‘understand’’, ‘‘miss’’, the number of samples for each hand sign are shown in Table 5.1. In each video, the signers are requested to wear a black shirt with long sleeves and stand in front of dark background for subjects 1-20 and signers can wear any shirt and stand in front of complex background for subjects 21-25 while they are performing the Thai sign language.

5.1.2 System description

We first collect keyframes of 31 hand gestures of 10 Thai words in signature library from five subjects. We manually select the representative frame (RFrame) of each video file in each word of Thai hand sign language, the number of image of each hand gesture is also shown in Figure 5.1 and we then manually selected the hand part of each frame with the size of 190×190 pixels. After that, we compute keypoint descriptors of each frame using SIFT and then we keep these keypoint descriptors in signature library database. Our Signature library contains 730 keypoint descriptors from each subject and 3,650 keypoint descriptors in total.

Table 5.1 Number of words in the training data set from subjects 1-15 and test data set from subject 1-25

	Subjects	elder	grand-father	grand-mother	female	male	glad	under-stand	miss
Training data set	1a	36	36	36	12	36	36	36	36
	2a-15a	32	32	32	32	32	32	32	32
Test data set	1b	12	12	12	12	12	12	12	12
	2b-15b	8	8	8	8	8	8	8	8
	16-19	20	20	20	20	20	20	20	20
	20	10	10	10	10	10	10	10	10
	21-25	5	5	5	5	5	5	5	5

In order to translate Thai hand sign language, we proposed SIFT [61] and string grammar fuzzy clustering method for improving the classification result. The diagram of the proposed system is given in Figure 5.2 which composed mainly of three steps, i.e., string representation, string grammar clustering, and classification.

Figure 5.1 Examples of 31 hand gestures.

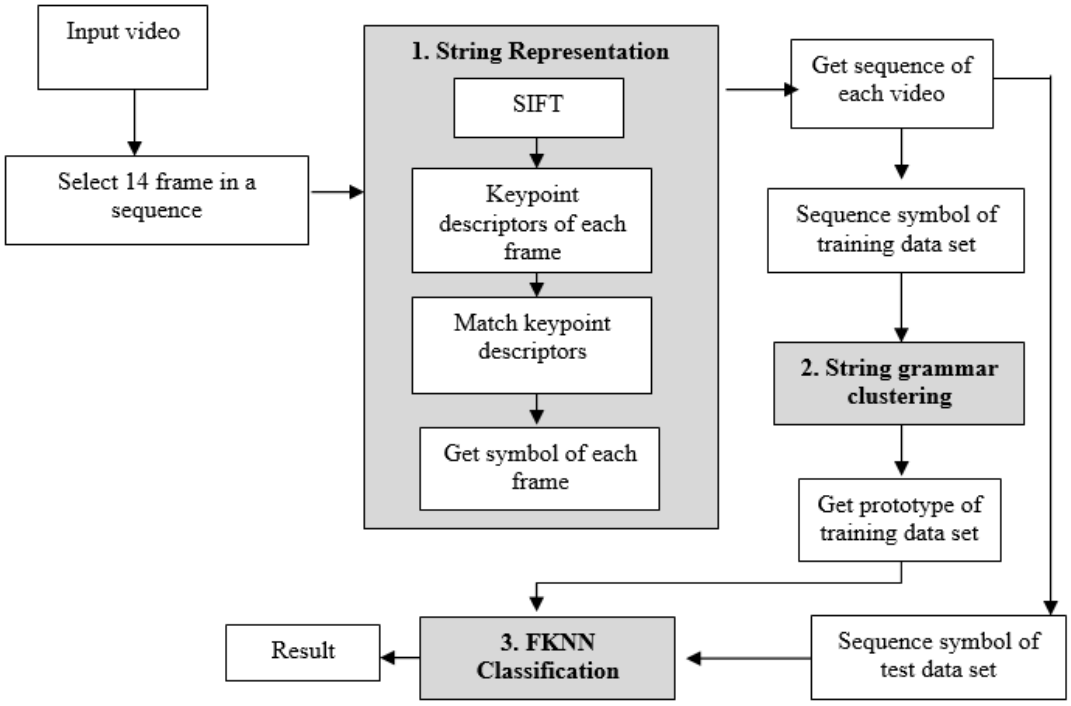


Figure 5.2 System overview of Thai sign language translation.

1) String representation.

To transform image sequence into a string of symbols, we can describe as the following step.

1.1) Extract video frame of each video to JPEG file with the size of 720×576 pixels.

1.2) We manually select only 14 image frame from each image frame sequence.

1.3) SIFT is utilized to extract some interesting point from images. The example of keypoints found on keyframe of SIFT [61] process is shown in Figure 5.3, while Figure 5.4(a)–(c) shows examples of keypoint descriptors of three hand gestures.

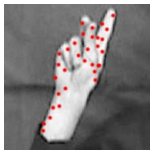
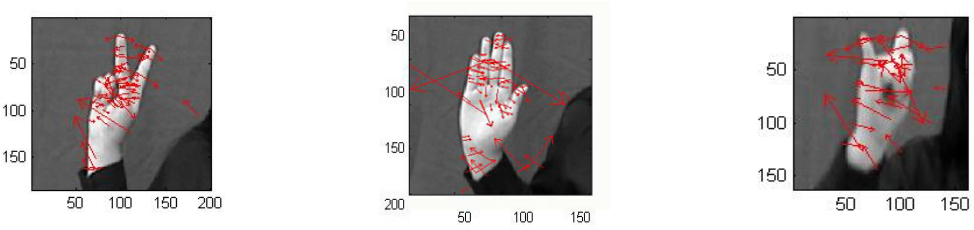


Figure 5.3 Keypoints found on a keyframe.



(a) (b) (c)

Figure 5.4 Keypoint descriptors found on hand gesture.

1.4) Match keypoint of test image frame to each other in the signature library by identifying their nearest neighbor with 0.65, 0.7 and 0.75 SIFT threshold. Matching is performed by comparing each local extrema based on the associated descriptors using the Euclidean distances that measure between each other keypoint descriptors in the signature library database and the current keypoint descriptor.

The matched symbol is selected by choosing the one that gives the maximum of the average number of matched keypoints per keyframes (Avg_Match) [63] of each symbol. Since, the keyframes for each symbol in the signature library may have different numbers of keypoints, The equation of Avg_Match is as follow:

$$Avg\_Match = \frac{\text{No.of matched keypoints of the symbol}}{\text{No. of key frames of the symbol}} \tag{5.1}$$

An example of the computed Avg_Match of the symbol ‘‘b’’ is shown in Figure 5.5. The matched symbol in Figure 5.5 is ‘‘b’’ with Avg_Match of 4.89.

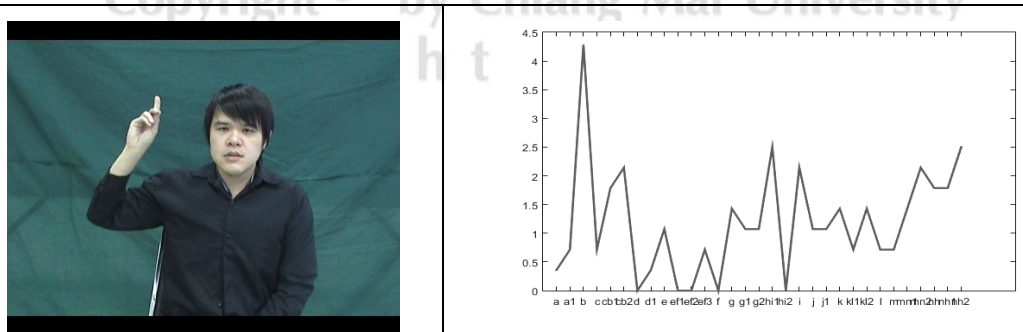
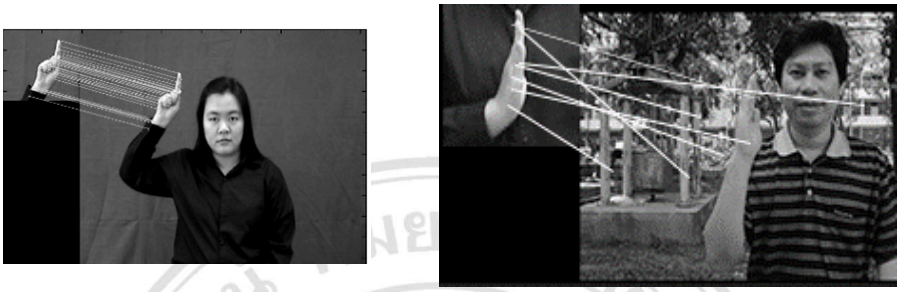


Figure 5.5 Avg_Match of symbols “b” matched symbol is “b”

The process of matching between keypoint descriptors of the test keyframe and the keyframe in signature library database are shown in Figure 5.6(a) and (b). Hence, we get the sequence of symbol of each frame in the sequence used as primitives in our string grammar fuzzy clustering algorithms.



(a) (b)

Figure 5.6 The hand gesture (a) “b” assigned to test image using SIFT and test images with constraint, (b) “k” is assigned to test image using SIFT and test frames without constraint.

2) String grammar clustering.

We use five string grammar clustering algorithms, i.e., sgFCMed, sgFPCMed, sgPFCMed, sgUPCMed and sgUPFCMed for finding multi-prototypes of Thai hand sign language training dataset. For our five string grammar clustering algorithms which are utilized for finding multi-prototypes of each Thai hand sign language class, more details can be referred in Chapter 3. The parameters setting for Thai hand sign language dataset is shown in Table 5.2.

Table 5.2 The parameter setting of our algorithms for Thai hand sign language data set.

parameter	sgFCMed	sgFPCMed	sgPFCMed	sgUPCMed	sgUPFCMed
m	2	2	2	2	2
η	-	2	2	-	2
γ	-	-	2	-	-
a	-	-	1	-	1
b	-	-	6	-	6
stopping criteria	0.1	0.1	0.1	0.1	0.1
maximum number of iterations	100	100	100	100	100

3) Fuzzy K-nearest neighbor classification.

For testing process, after the multi-prototypes, i.e. $SC = \left\{ sc_1^1, ..., sc_{N_1}^1, sc_1^2, ..., sc_{N_2}^2, sc_1^C, ..., sc_{N_C}^C \right\}$ where sc_k^j is string prototype *k* of class *j* for each class from sgFCMed, sgFPCMed, sgPFCMed, sgUPCMed and sgUPFCMed are created. We apply the fuzzy K-nearest neighbor [61] to classify the test dataset. FKNN is similar to K-nearest neighbor (KNN) The data point from FKNN can belong to multiple classes with different membership functions associated to these classes. For each *s*, the membership value u_i in class *i* can be calculated as the following formulation:

$$u_i(s) = \left\{ \frac{\sum_{j=1}^{K} u_{ij} \left(\frac{1}{Lev\left(sc_j^q, s\right)} \right)^{1/(m-1)}}{\sum_{j=1}^{K} \left(\frac{1}{Lev\left(sc_j^q, s\right)} \right)^{1/(m-1)}} \right\}, \tag{5.2}$$

where u_{ij} is the membership value of prototype sc_j^q in the class *i* ,*c* number of class and *K* is number of nearest neighbors. From the equation (5.2), we can see the assigned memberships of *s* are affected by the inverse of the levenshtein distances from the nearest neighbors and their class memberships. Hence the result of classification is determined by the assigned membership of the string *s*, *s* is assigned to class *i* if $u_i(s) > u_j(s)$ for all $j \neq i$.

In our experiment, we used the crisp membership to assign each prototype since we have been known class of each prototype already. We set membership is 1 for sc_j^q in class *q* and we set memberships are zero for all other classes. The parameter *m* is used to determine how heavily the distance is weighted when calculating each neighbor's contribution to the membership value, and its value is chosen for our experiment as m=2.

5.1.3 Result and Discussion

The videos datasets are recorded from subjects 1–25 that are described in Table 5.1. Each video is decimated, left only 14 frames. We divided our experiments into three parts: training with subject 1a, training with subjects 1a–5a and training with subjects 1a–15a. There are three groups of blind test, i.e., dataset 1b-15b, dataset 16-20 (used to represent signer-independent), and 21-25 (with various complex natural backgrounds). For all training and test dataset, we assign symbols to each frame in the training data set using the SIFT and we use the threshold values for the experiment are varied from 0.65, 0.7, and 0.75. We then use several multi-prototype string grammar clustering algorithms to classify ten hand sign words which the lengths of each string representation are 14. After we create multi-prototypes in terms of a sequence of primitives already, the test string is assigned to a word the closest prototype belongs to according to the FKNN based on Levenshtein distance.

1) Comparison the performance of the five string grammar fuzzy clustering on validation training set.

For comparison the performance of the five string grammar fuzzy clustering, we implement the 4-folds cross-validation on the training set and we implement each string grammar fuzzy clustering algorithm, i.e., sgFCMed, sgFPCMed, sgPFCMed, sgUPCMed and sgUPFCMed with 4, 8, and 12 clusters on each class separately to create multi-prototypes for each class. Then the FKNN with K = 1, 3, 5, 7 and 9 are implemented as a classifier on each set of created multi-prototypes. Tables 5.3 to 5.5 show the best correct classification of the validation set of training with 1a, training with 1a–5a and training with 1a–15a for FKNN with K = 1, 3, 5, 7 and 9, respectively.

From Table 5.3, we can see that the best average classification of 100.00% on the validation set when training with 1a is from the sgUPFCMed with 12 prototypes on 0.75 sift threshold using K = 9. For sgUPFCMed with 4 and 8 prototypes yield 90.48% (with K = 7) and 98.81% (with K = 9), respectively. Whereas, the best average of

sgFCMed, sgFPCMed, sgPFCMed and sgUPCMed provide 92.86%, 95.24%, 97.62% and 98.81%, respectively.

We also can see that on 4-folds cross-validation when training with data set 1a–5a for FKNN on 0.65 sift threshold from sgUPFCMed gives the best average classification rate of 90.10% (with $K = 5$), 92.33% (with $K = 7$) and 95.79% (with $K = 9$), using 4, 8 and 12, respectively in Table 5.4. Whereas, the best average of sgFCMed, sgFPCMed, sgPFCMed and sgUPCMed with 12-prototypes provide 90.10%, 93.81%, 95.30% and 94.55%, respectively.

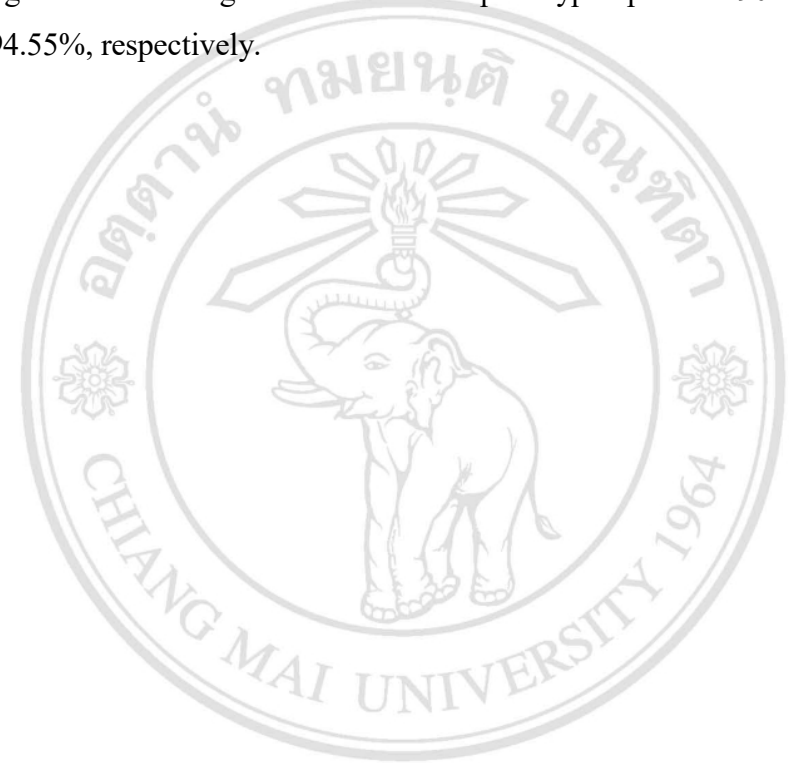


Table 5.3 The best classification rate (%) of the validation set from 4-fold cross validation when training with data set 1a for FKNN with K = 1, 3, 5, 7 and 9.

K of FKNN	Sift threshold	Number of prototype in each class													
		4					8					12				
		sgFCMed	sgFPCMed	sgPFCMed	sgUPCMed	sgUPFCMed	sgFCMed	sgFPCMed	sgPFCMed	sgUPCMed	sgUPFCMed	sgFCMed	sgFPCMed	sgPFCMed	sgUPCMed	sgUPFCMed
1	0.65	75.00	78.57	79.76	78.57	79.76	79.76	84.52	89.29	89.29	90.48	82.14	83.33	84.52	91.67	92.86
	0.7	76.19	80.95	78.57	79.76	78.57	85.71	86.90	90.48	89.29	92.86	86.90	91.67	94.05	92.86	94.05
	0.75	75.00	76.19	82.14	83.33	82.14	78.57	84.52	86.90	88.10	90.48	79.76	83.33	85.71	92.86	94.05
3	0.65	80.95	80.95	84.52	86.90	84.52	80.95	86.90	89.29	88.10	90.48	89.29	91.67	95.24	89.29	92.86
	0.7	78.57	79.76	82.14	82.14	82.14	83.33	86.90	90.48	88.10	91.67	89.29	91.67	96.43	92.86	94.05
	0.75	79.62	80.95	78.57	76.19	78.57	80.95	85.71	90.48	90.48	91.67	83.33	85.71	86.90	89.29	92.86
5	0.65	82.14	83.33	88.10	86.90	88.10	86.90	89.29	95.24	96.43	97.62	89.29	91.67	95.24	96.43	98.81
	0.7	79.76	83.33	88.10	88.10	88.10	89.29	92.86	98.81	96.43	90.48	91.67	95.24	95.24	91.67	92.86
	0.75	72.76	76.19	84.52	85.71	84.52	84.52	86.90	91.67	89.29	92.86	89.29	91.67	96.43	91.67	95.24
7	0.65	73.81	79.76	84.52	85.71	84.52	83.33	88.10	92.86	92.86	95.24	88.10	91.67	96.43	95.24	98.81
	0.7	80.95	85.71	90.48	90.48	90.48	84.52	85.71	86.90	89.29	90.48	89.29	92.86	96.43	90.48	94.05
	0.75	75.00	76.19	78.57	80.95	88.10	79.76	83.33	85.71	88.10	89.29	84.52	85.71	88.10	90.48	91.67
9	0.65	70.24	76.19	79.76	79.76	79.76	76.19	77.38	82.14	82.14	85.71	92.86	91.67	94.05	84.52	86.90
	0.7	76.19	78.57	82.14	80.95	82.14	80.95	82.14	86.90	88.10	89.29	94.05	91.67	97.62	89.29	91.67
	0.75	67.86	71.43	77.38	78.57	79.76	76.19	80.95	83.33	89.29	98.81	95.24	95.24	95.24	98.81	**100.00**

Table 5.4 The best classification rate of the validation set from 4-fold cross validation when training with data set 1a–5a for FKNN with K = 1, 3, 5, 7 and 9.

K of FKNN	Sift threshold	Number of prototype in each class														
		4					8					12				
		sgFCMed	sgFPCMed	sgPFCMed	sgUPCMed	sgUPFCMed	sgFCMed	sgFPCMed	sgPFCMed	sgUPCMed	sgUPFCMed	sgFCMed	sgFPCMed	sgPFCMed	sgUPCMed	sgUPFCMed
1	0.65	87.13	88.37	89.11	88.86	89.85	87.62	88.12	88.86	88.61	89.60	88.37	88.37	89.60	89.11	90.10
	0.7	86.14	87.38	88.61	88.12	89.36	86.63	87.62	88.61	87.87	88.86	87.62	88.61	89.85	89.36	90.10
	0.75	85.89	86.39	87.62	87.13	88.37	86.63	87.87	88.61	88.37	89.11	86.63	87.13	87.87	87.13	88.12
3	0.65	87.62	88.37	88.37	87.87	88.86	88.37	89.11	89.85	89.60	90.59	89.85	90.59	90.84	90.10	91.09
	0.7	87.13	87.62	88.12	87.62	88.37	88.86	90.10	91.34	91.09	91.58	89.11	89.60	90.59	89.85	90.84
	0.75	86.88	87.62	87.38	86.63	87.87	86.63	87.62	88.61	87.87	89.11	89.60	89.60	90.10	89.36	90.84
5	0.65	87.38	89.36	89.60	88.86	90.10	87.62	90.10	90.59	90.10	91.09	88.86	90.35	92.57	91.83	93.32
	0.7	87.13	87.87	88.86	88.37	89.11	87.62	88.12	89.11	88.61	89.85	89.60	91.34	93.32	93.07	94.06
	0.75	86.63	89.11	89.36	88.86	89.85	87.38	89.85	90.10	89.60	90.59	87.87	88.86	91.09	90.84	91.83
7	0.65	88.37	89.36	89.36	88.86	89.60	89.60	90.35	91.58	91.34	92.33	88.86	91.34	94.31	93.56	94.06
	0.7	87.62	88.12	88.61	88.12	88.86	88.86	91.09	92.08	91.83	90.59	89.60	91.34	92.57	93.07	90.84
	0.75	88.12	88.86	89.36	88.86	89.85	88.61	89.36	89.85	89.11	90.59	89.60	93.56	93.81	93.56	94.06
9	0.65	86.88	87.62	88.61	88.37	89.11	87.62	87.87	89.60	88.86	90.35	88.61	93.81	95.30	94.55	**95.79**
	0.7	86.88	87.62	88.12	87.38	88.37	87.13	88.12	90.35	89.60	91.09	90.10	88.61	94.06	93.32	94.80
	0.75	85.89	86.63	87.38	88.86	89.85	86.14	86.88	88.86	88.61	89.60	87.87	90.84	92.82	89.11	90.10

Table 5.5 The best classification rate of the validation set from 4-fold cross validation when training with data set 1a–15a for FKNN with K = 1, 3, 5, 7 and 9.

K of FKNN	Sift threshold	Number of prototype in each class														
		4					8					12				
		sgFCMed	sgFPCMed	sgPFCMed	sgUPCMed	sgUPFCMed	sgFCMed	sgFPCMed	sgPFCMed	sgUPCMed	sgUPFCMed	sgFCMed	sgFPCMed	sgPFCMed	sgUPCMed	sgUPFCMed
1	0.65	84.88	85.38	85.80	85.63	85.96	85.38	85.88	86.21	85.96	86.30	85.96	86.38	86.71	86.63	86.96
	0.7	85.38	86.05	86.38	86.30	86.46	85.88	86.30	86.79	86.63	87.04	86.30	86.88	87.29	87.04	87.54
	0.75	83.14	83.72	83.97	83.89	84.14	83.72	83.97	84.22	84.05	84.47	83.89	84.22	84.80	84.72	84.88
3	0.65	85.30	85.30	85.22	85.13	85.30	85.80	86.38	86.71	86.46	86.96	86.05	86.54	86.96	86.79	87.21
	0.7	85.80	85.47	85.63	85.47	85.88	86.05	86.54	86.96	86.88	87.04	87.04	87.29	87.54	87.38	87.62
	0.75	83.06	83.39	83.22	82.97	83.39	83.64	84.14	84.55	84.30	84.63	84.30	84.72	85.05	84.97	85.22
5	0.65	84.55	85.13	85.63	85.38	85.88	85.13	85.63	86.05	85.96	86.21	86.54	87.04	87.54	87.29	87.62
	0.7	84.88	85.47	85.63	85.55	85.88	85.30	85.71	86.21	86.05	86.30	86.79	87.38	87.96	87.71	88.04
	0.75	82.89	83.31	83.55	83.31	83.72	83.14	83.72	84.05	83.97	84.14	84.63	85.13	87.46	87.38	87.54
7	0.65	86.21	86.21	86.38	86.13	86.63	86.79	86.79	87.46	87.38	87.62	87.13	87.38	88.04	87.79	88.12
	0.7	86.71	87.71	87.79	87.54	88.04	87.13	87.96	88.12	87.96	88.21	87.62	88.21	85.38	85.22	85.47
	0.75	84.72	84.88	84.97	84.72	85.13	84.97	85.22	85.63	85.47	85.80	85.30	85.55	85.96	85.71	86.05
9	0.65	85.96	85.88	85.96	85.71	86.21	86.54	86.63	87.04	86.79	87.29	86.96	87.38	87.96	87.71	88.12
	0.7	86.46	87.62	87.54	87.46	87.71	86.96	87.71	88.04	87.79	88.12	87.62	88.79	88.54	88.37	**88.79**
	0.75	84.55	84.55	84.72	84.47	84.88	84.72	85.13	85.63	85.47	85.71	85.30	85.55	85.55	85.30	85.63

Again, sgUPFCMed gives higher recognition rate than the others on 0.7 sift threshold when training with 1a-15a as shown in Table 5.5. The best average classification accuracy rates are 88.04% (with K = 7), 88.21% (with K = 7) and 88.79% (with K = 9), using 4,8 and 12, respectively. Whereas, the best average of sgFCMed, sgFPCMed, sgPFCMed and sgUPCMed with 12-prototypes provide 87.62%, 88.79%, 88.54% and 88.37%, respectively. From all experiments on the validation set from 4-folds cross-validation, we can see that if we increase the number of prototypes in the process of string grammar clustering, there is it chance that the classification rates of all type of signer will also increase. From the results in Tables 5.3 to 5.5, the 12-prototypes string grammar clustering with 9-FKNN gives the classification rate higher than the other number of prototype. Hence we use 12-prototypes string grammar clustering with 9-FKNN for the blind test dataset as shown in Tables 5.6 to 5.14.

2) Comparison the performance of the five string grammar fuzzy clustering on testing set.

We measured performances of our algorithms, including, sgFCMed, sgFPCMed, sgPFCMed, sgUPCMed and sgUPFCMed. We divided our experiments into three parts: training with 1a, training with 1a–5a and training with 1a–15a. There are 3 groups of blind test, i.e., dataset 1b-15b, dataset 16-20 (used to represent signer-independent), and 21-25 (with various complex natural backgrounds). We set number of prototype for each class equal to 12 and K of FKNN equal to 9.

In Table 5.6 to 5.8, our blind test datasets are training on the data sets 1a, the classification result of the blind test datasets 1a and 1b (signer-dependent) are shown in Table 5.6. From these tables, we can see that the blind test results from sgUPFCMed with 0.7 SIFT threshold gives the best classification as 100% and 95.85% for datasets 1a and 1b, respectively. These experiments reach 100% accuracy rate on blind test dataset 1a, since both training and testing dataset are the same dataset, that is 1a. Since, the dataset 1a and 1b is the same signer, the result of 1b dataset also high accuracy.

Table 5.6 Classification rate on test sets of signer-dependent when training with data set 1a.

test	sgFCMed			sgFPCMed			sgPFCMed			sgUPCMed			sgUPFCMed		
	Sift threshold			Sift threshold			Sift threshold			Sift threshold			Sift threshold		
	0.65	0.7	0.75	0.65	0.7	0.75	0.65	0.7	0.75	0.65	0.7	0.75	0.65	0.7	0.75
1a	98.21	98.21	98.51	98.21	98.51	99.40	98.81	93.33	95.31	97.92	97.92	97.92	99.40	**100.00**	99.40
1b	90.00	92.50	92.50	93.33	93.33	92.50	93.65	93.33	90.94	90.83	91.67	89.17	94.17	**95.83**	92.50

Table 5.7 Classification rate on test sets of signer-semi-dependent when training with data set 1a.

test	sgFCMed			sgFPCMed			sgPFCMed			sgUPCMed			sgUPFCMed		
	Sift threshold			Sift threshold			Sift threshold			Sift threshold			Sift threshold		
	0.65	0.7	0.75	0.65	0.7	0.75	0.65	0.7	0.75	0.65	0.7	0.75	0.65	0.7	0.75
2a	61.57	64.69	65.00	62.19	65.00	65.31	62.81	65.94	65.94	62.19	65.00	65.00	63.75	66.88	66.56
3a	63.44	71.88	59.37	63.75	72.81	60.31	65.00	73.44	60.94	64.69	73.13	60.31	65.31	73.75	61.56
4a	52.19	49.69	53.13	52.50	50.63	53.44	53.44	51.25	54.38	53.13	50.63	53.75	53.75	51.88	55.00
5a	55.32	61.25	51.26	55.63	61.88	51.88	56.56	62.50	52.19	55.63	61.88	51.25	57.19	63.44	52.50
2b	57.50	56.25	51.25	58.75	57.50	52.50	60.00	60.00	55.00	58.75	58.75	51.25	63.75	61.25	57.50
3b	60.00	58.75	45.00	62.50	61.25	46.25	66.25	63.75	48.75	65.00	60.00	48.75	70.00	67.50	52.50
4b	57.50	63.75	52.50	58.75	65.00	55.00	60.00	66.25	57.50	58.75	65.00	56.25	62.50	67.50	60.00
5b	57.50	62.50	63.75	58.75	63.75	66.25	61.25	66.25	68.75	57.50	63.75	67.50	62.50	68.75	72.50

Table 5.8 Classification rate on test sets of signer-independent when training with data set 1a.

test	sgFCMed			sgFPCMed			sgPFCMed			sgUPCMed			sgUPFCMed		
	Sift threshold			Sift threshold			Sift threshold			Sift threshold			Sift threshold		
	0.65	0.7	0.75	0.65	0.7	0.75	0.65	0.7	0.75	0.65	0.7	0.75	0.65	0.7	0.75
6a	40.01	39.38	31.56	40.94	39.69	31.88	40.94	40.63	32.81	40.00	40.00	32.19	41.25	41.56	33.13
7a	57.82	70.32	61.88	58.13	70.94	62.19	58.75	71.25	62.81	58.44	70.94	62.50	59.06	71.88	63.44
8a	55.32	66.57	68.75	55.94	67.19	69.38	56.25	68.13	70.31	55.31	67.50	69.38	57.19	68.44	70.63
9a	28.44	29.38	26.56	28.75	29.69	26.88	30.31	30.31	27.81	29.38	29.69	27.19	31.25	31.25	28.75
10a	24.38	25.94	30.94	24.69	26.25	31.25	25.31	27.19	32.19	24.69	26.88	31.25	26.25	27.81	32.81
11a	62.19	63.44	56.56	63.13	64.38	57.19	62.81	64.69	57.50	61.88	64.06	57.19	63.75	65.63	58.44
12a	43.75	51.25	45.63	44.69	52.19	45.94	45.31	52.50	46.56	44.38	52.19	45.63	45.94	52.81	47.19
13a	31.26	39.38	35.00	31.88	40.00	35.63	32.81	40.63	36.25	31.88	39.69	35.63	33.75	41.56	36.88
14a	56.57	59.07	55.01	57.50	60.00	55.94	57.50	60.31	56.56	56.56	60.00	55.63	58.44	60.94	57.50
15a	42.81	48.44	40.63	43.75	48.75	41.25	44.06	49.06	42.19	43.75	48.75	41.56	44.69	49.38	43.13
6b	43.75	41.25	40.00	46.25	45.00	41.25	50.00	46.25	42.50	46.25	42.50	38.75	52.50	50.00	45.00
7b	66.25	80.00	75.00	70.00	81.25	78.75	73.75	85.00	82.50	71.25	83.75	80.00	76.25	87.50	83.75
8b	70.00	76.25	80.00	71.25	78.75	83.75	72.50	81.25	87.50	68.75	77.50	83.75	75.00	82.50	88.75
9b	40.00	40.00	31.25	43.75	43.75	35.00	47.50	45.00	38.75	46.25	41.25	33.75	48.75	48.75	41.25
10b	30.00	40.00	41.25	33.75	43.75	43.75	37.50	45.00	45.00	33.75	43.75	45.00	38.75	46.25	48.75

Table 5.8 Classification rate on test sets of signer-independent when training with data set 1a. (continue).

test	sgFCMed			sgFPCMed			sgPFCMed			sgUPCMed			sgUPFCMed		
	Sift threshold			Sift threshold			Sift threshold			Sift threshold			Sift threshold		
	0.65	0.7	0.75	0.65	0.7	0.75	0.65	0.7	0.75	0.65	0.7	0.75	0.65	0.7	0.75
11b	67.50	67.50	60.00	71.25	68.75	62.50	75.00	72.50	65.00	72.50	68.75	62.50	77.50	73.75	65.00
12b	51.56	57.50	53.13	55.00	60.00	57.50	56.25	61.25	58.75	52.50	57.50	58.75	60.00	65.00	62.50
13b	35.00	41.25	27.50	36.25	42.50	28.75	38.75	43.75	31.25	37.50	41.25	28.75	42.50	46.25	35.00
14b	51.25	51.25	48.75	52.50	52.50	50.00	56.25	53.75	52.50	53.75	50.00	50.00	58.75	55.00	55.00
15b	41.25	41.25	41.25	43.75	42.50	43.75	46.25	45.00	46.25	43.75	41.25	45.00	50.00	46.25	48.75
16	55.50	52.00	55.00	55.50	57.00	60.00	60.50	57.00	61.00	60.00	56.50	60.00	61.00	58.00	62.00
17	68.00	64.50	66.50	68.00	64.50	66.50	74.50	70.50	71.50	73.50	69.50	70.50	75.50	72.00	72.50
18	44.50	47.00	63.00	44.50	48.50	78.50	60.50	63.00	79.00	60.00	62.50	78.00	62.00	64.50	80.50
19	50.00	55.00	60.50	59.00	58.00	60.50	65.00	62.00	62.00	63.50	60.50	60.50	66.00	63.50	63.00
20	48.00	57.00	54.00	55.50	62.00	60.00	57.00	56.00	59.00	55.00	54.00	58.00	59.00	57.00	61.00

Table 5.9 Classification rate on test sets of signer-dependent when training with data set 1a-5a.

test	sgFCMed			sgFPCMed			sgPFCMed			sgUPCMed			sgUPFCMed		
	Sift threshold			Sift threshold			Sift threshold			Sift threshold			Sift threshold		
	0.65	0.7	0.75	0.65	0.7	0.75	0.65	0.7	0.75	0.65	0.7	0.75	0.65	0.7	0.75
1a	98.81	98.81	98.51	99.11	99.40	98.81	100.00	100.00	99.70	99.11	98.81	99.11	100.00	100.00	100.00
2a	96.88	96.88	96.25	97.50	97.50	96.56	98.13	98.44	96.88	97.50	98.13	96.25	98.75	98.75	97.81
3a	96.56	94.38	95.31	97.19	95.00	95.94	97.81	95.94	96.56	97.19	94.69	95.63	98.13	95.94	97.19
4a	94.38	90.00	84.06	95.31	90.31	85.00	96.25	90.94	85.31	95.94	90.31	84.69	97.19	91.88	86.25
5a	92.81	92.81	90.63	93.13	93.44	91.25	93.44	94.38	91.88	93.44	93.44	90.94	95.00	95.00	92.81
1b	95.83	92.50	94.17	98.33	94.16	95.00	98.33	95.83	96.67	99.17	95.83	95.00	99.17	97.50	98.33
2b	80.00	78.75	78.75	82.50	82.50	80.00	83.75	83.75	82.50	82.50	82.50	80.00	88.75	87.50	83.75
3b	96.25	92.50	92.50	98.75	96.25	96.25	98.75	98.75	97.50	97.50	93.75	96.25	98.75	98.75	98.75
4b	93.75	92.50	80.00	97.50	93.75	81.25	97.50	96.25	83.75	95.00	96.25	80.00	98.75	98.75	86.25
5b	93.75	92.50	86.25	95.00	93.75	88.75	98.75	97.50	91.25	95.00	93.75	90.00	97.50	98.75	95.00

Table 5.10 Classification rate on test sets of signer-independent when training with data set 1a-5a.

test	sgFCMed			sgFPCMed			sgPFCMed			sgUPCMed			sgUPFCMed		
	Sift threshold			Sift threshold			Sift threshold			Sift threshold			Sift threshold		
	0.65	0.7	0.75	0.65	0.7	0.75	0.65	0.7	0.75	0.65	0.7	0.75	0.65	0.7	0.75
6a	50.31	50.63	56.88	50.63	51.26	57.19	51.56	51.56	57.50	50.94	51.25	56.88	51.88	52.19	57.81
7a	75.94	79.69	75.00	76.25	80.00	75.94	76.56	80.63	76.25	75.94	79.69	75.63	77.19	80.94	76.56
8a	77.81	77.81	82.50	78.44	78.44	82.81	78.75	79.38	83.75	78.75	78.75	83.44	80.31	79.69	84.06
9a	52.50	55.63	57.81	53.13	56.57	58.75	54.06	57.19	59.69	52.81	56.88	59.06	54.06	58.13	60.63
10a	56.25	54.69	63.13	57.19	55.63	63.44	57.50	56.56	63.75	57.19	55.63	62.81	58.13	56.25	64.06
11a	70.63	72.50	71.25	71.57	73.13	72.19	72.19	73.44	72.81	72.19	72.50	72.19	73.44	73.75	73.75
12a	67.81	67.19	71.56	68.75	67.50	72.50	69.69	68.44	72.81	68.75	67.19	72.19	70.63	69.06	73.44
13a	52.81	55.31	52.81	53.76	55.63	53.44	54.69	56.56	54.06	53.13	55.31	53.44	54.38	56.25	54.38
14a	68.44	66.56	74.06	69.06	67.50	74.37	69.38	68.44	75.31	69.69	67.19	74.38	70.63	69.06	75.63
15a	78.13	75.94	75.00	78.44	76.25	75.62	79.06	77.19	75.94	78.44	75.94	75.31	80.31	77.19	76.25
6b	48.75	55.00	52.50	52.50	57.50	55.00	55.00	61.25	57.50	53.75	55.00	55.00	56.25	62.50	58.75
7b	60.00	73.75	73.75	63.75	75.00	77.50	66.25	77.50	81.25	62.50	75.00	80.00	68.75	81.25	85.00
8b	68.75	76.25	75.00	71.25	77.50	78.75	75.00	81.25	80.00	70.00	77.50	78.75	75.00	82.50	81.25
9b	47.50	43.75	47.50	48.75	47.50	50.00	52.50	48.75	51.25	50.00	47.50	50.00	52.50	51.25	53.75

Table 5.10 Classification rate on test sets of signer-independent when training with data set 1a-5a. (continue)

test	sgFCMed			sgFPCMed			sgPFCMed			sgUPCMed			sgUPFCMed		
	Sift threshold			Sift threshold			Sift threshold			Sift threshold			Sift threshold		
	0.65	0.7	0.75	0.65	0.7	0.75	0.65	0.7	0.75	0.65	0.7	0.75	0.65	0.7	0.75
10b	52.50	53.75	62.50	55.00	55.00	66.25	58.75	58.75	67.50	57.50	53.75	65.00	61.25	61.25	71.25
11b	52.50	51.25	62.50	55.00	55.00	66.25	57.50	56.25	67.50	55.00	56.25	66.25	60.00	58.75	70.00
12b	71.25	70.00	80.00	73.75	73.75	82.50	76.25	75.00	85.00	71.25	73.75	83.75	76.25	76.25	88.75
13b	53.75	51.25	53.75	55.00	53.75	57.50	58.75	55.00	58.75	56.25	52.50	57.50	58.75	60.00	62.50
14b	53.75	52.50	56.25	55.00	55.00	58.75	56.25	58.75	61.25	55.00	53.75	58.75	62.50	60.00	62.50
15b	63.75	67.50	56.88	67.50	68.75	72.50	70.00	70.00	76.25	66.25	71.25	72.50	73.75	75.00	77.50
16	54.50	62.00	65.50	55.50	63.00	66.50	57.00	64.00	67.00	55.50	63.50	66.00	57.50	65.00	67.50
17	78.50	85.00	86.50	79.50	86.00	88.00	80.50	86.50	89.00	78.50	85.50	88.00	80.50	87.50	90.00
18	54.50	57.00	73.00	55.00	57.50	73.50	56.50	58.50	75.00	54.50	58.00	74.50	56.50	59.00	76.00
19	68.00	66.00	65.50	68.50	67.00	66.00	69.00	67.50	67.00	68.00	66.00	66.50	70.50	68.50	68.00
20	54.00	56.00	62.00	57.00	58.00	64.00	58.00	61.00	67.00	57.00	59.00	64.00	62.00	63.00	69.00

Table 5.11 Classification rate on test sets of signer-dependent when training with data set 1a-15a.

test	sgFCMed			sgFPCMed			sgPFCMed			sgUPCMed			sgUPFCMed		
	Sift threshold			Sift threshold			Sift threshold			Sift threshold			Sift threshold		
	0.65	0.7	0.75	0.65	0.7	0.75	0.65	0.7	0.75	0.65	0.7	0.75	0.65	0.7	0.75
1a	98.51	98.21	98.81	98.81	99.11	99.11	99.70	99.70	99.40	99.11	98.81	98.81	100.00	100.00	99.70
2a	95.94	95.00	90.63	96.57	95.95	90.95	97.19	96.56	91.88	96.88	95.63	91.56	97.81	97.19	92.19
3a	92.81	92.81	92.81	93.76	93.76	93.76	94.38	94.38	94.69	94.06	93.75	94.06	95.00	95.31	95.63
4a	85.31	80.94	77.50	85.95	81.89	77.82	86.88	82.50	78.75	86.25	81.56	78.13	87.81	83.44	79.38
5a	90.31	91.25	90.00	90.95	92.20	90.64	91.88	92.50	91.56	90.94	92.19	90.63	92.50	93.13	91.88
1b	91.67	91.67	90.83	94.17	93.75	92.50	95.00	96.67	95.00	92.50	94.17	92.50	95.83	99.17	97.50
2b	91.25	90.00	83.75	93.33	91.25	85.00	96.25	95.00	87.50	95.00	93.75	86.25	96.25	97.50	90.00
3b	76.25	87.50	82.50	80.00	88.75	86.25	81.25	90.00	87.50	80.00	88.75	85.00	83.75	92.50	91.25
4b	86.25	81.25	81.25	90.00	85.00	82.50	92.50	88.75	86.25	91.25	87.50	85.00	96.25	90.00	88.75
5b	80.00	81.25	87.50	81.25	83.75	90.00	85.00	86.25	92.50	82.50	83.75	91.25	87.50	88.75	95.00

Table 5.12 Classification rate on test sets of signer-semi-independent when training with data set 1a-15a.

test	sgFCMed			sgFPCMed			sgPFCMed			sgUPCMed			sgUPFCMed		
	Sift threshold			Sift threshold			Sift threshold			Sift threshold			Sift threshold		
	0.65	0.7	0.75	0.65	0.7	0.75	0.65	0.7	0.75	0.65	0.7	0.75	0.65	0.7	0.75
6a	75.94	76.88	78.13	76.89	77.51	78.45	77.81	77.81	79.06	77.50	76.88	80.00	78.13	78.75	78.44
7a	97.81	95.63	91.25	98.45	96.57	91.89	99.06	97.19	92.19	98.75	96.88	92.81	99.69	98.13	91.25
8a	93.44	91.25	90.63	94.39	91.57	90.95	95.00	92.50	91.88	94.06	92.19	92.19	95.94	93.44	91.56
9a	79.38	81.88	77.19	79.70	82.51	78.14	80.63	83.13	78.75	80.31	82.50	79.38	81.25	83.44	77.81
10a	78.75	80.94	79.69	79.39	81.57	80.32	80.31	82.50	80.94	80.00	81.88	81.88	80.94	83.13	80.63
11a	82.50	84.06	78.44	83.45	84.39	79.07	84.06	84.69	79.38	83.44	83.75	80.00	85.00	85.00	78.44
12a	74.69	77.50	80.94	75.01	78.45	81.89	75.63	78.75	82.81	75.31	77.81	83.75	75.94	79.06	82.19
13a	81.25	82.50	81.56	81.57	83.45	82.20	82.50	84.06	83.13	81.56	83.44	83.44	82.81	85.00	82.81
14a	87.81	91.25	88.13	88.14	91.57	88.76	88.75	92.19	89.06	88.13	91.88	89.69	89.38	93.13	88.75
15a	90.63	91.88	87.19	91.26	92.51	88.14	92.19	93.44	88.75	91.88	92.81	89.69	92.50	94.38	88.44
6b	82.50	83.75	81.25	85.00	85.00	85.00	87.50	87.50	88.75	85.00	86.25	91.25	88.75	91.25	85.00
7b	81.25	91.25	91.25	83.75	92.50	95.00	85.00	95.00	96.25	83.75	92.50	98.75	87.50	98.75	92.50
8b	93.75	86.25	88.75	95.00	87.50	90.00	96.25	90.00	93.75	92.50	88.75	95.00	97.50	91.25	92.50

Table 5.12 Classification rate on test sets of signer-semi-independent when training with data set 1a-15a. (continue)

test	sgFCMed			sgFPCMed			sgPFCMed			sgUPCMed			sgUPFCMed		
	Sift threshold			Sift threshold			Sift threshold			Sift threshold			Sift threshold		
	0.65	0.7	0.75	0.65	0.7	0.75	0.65	0.7	0.75	0.65	0.7	0.75	0.65	0.7	0.75
9b	76.25	70.00	70.00	80.00	73.75	73.75	82.50	75.00	77.50	81.25	72.50	78.75	85.00	76.25	73.75
10b	68.75	60.00	65.00	70.00	62.50	66.25	73.75	65.00	68.75	70.00	61.25	72.50	76.25	66.25	66.25
11b	62.50	68.75	73.75	63.75	72.50	75.00	66.25	73.75	78.75	65.00	72.50	80.00	67.50	77.50	75.00
12b	70.00	73.75	70.00	72.50	76.25	73.75	73.75	77.50	77.50	70.00	75.00	80.00	76.25	78.75	76.25
13b	76.25	72.50	71.25	78.75	73.75	73.75	80.00	77.50	77.50	77.50	73.75	81.25	81.25	78.75	76.25
14b	76.25	75.00	77.50	78.75	76.25	80.00	81.25	78.75	81.25	77.50	76.25	82.50	82.50	80.00	78.75
15b	75.00	76.25	80.00	78.75	78.75	83.75	82.50	82.50	86.25	80.00	78.75	90.00	86.25	85.00	82.50

Table 5.13 Classification rate on test sets of signer- independent when training with data set 1a-15a.

test	sgFCMed			sgFPCMed			sgPFCMed			sgUPCMed			sgUPFCMed		
	Sift threshold			Sift threshold			Sift threshold			Sift threshold			Sift threshold		
	0.65	0.7	0.75	0.65	0.7	0.75	0.65	0.7	0.75	0.65	0.7	0.75	0.65	0.7	0.75
16	75.00	79.00	72.00	76.50	73.50	72.50	77.00	74.00	74.00	76.00	73.00	72.50	79.00	76.00	75.00
17	95.00	98.50	92.50	96.00	93.50	94.00	97.00	94.00	94.50	96.50	94.50	93.50	98.50	96.50	95.00

Table 5.13 Classification rate on test sets of signer- independent when training with data set 1a-15a. (continue)

test	sgFCMed			sgFPCMed			sgPFCMed			sgUPCMed			sgUPFCMed		
	Sift threshold			Sift threshold			Sift threshold			Sift threshold			Sift threshold		
	0.65	0.7	0.75	0.65	0.7	0.75	0.65	0.7	0.75	0.65	0.7	0.75	0.65	0.7	0.75
18	64.50	69.00	80.50	66.00	70.50	81.00	67.00	71.00	82.00	67.00	71.00	81.50	69.00	72.50	82.50
19	84.00	87.00	77.50	85.50	82.00	79.00	86.50	82.50	80.50	85.00	81.50	79.00	87.00	84.00	82.00
20	71.00	79.00	63.00	74.00	75.00	65.00	75.00	78.00	66.00	75.00	77.00	63.00	79.00	81.00	67.00

Table 5.14 Classification rate on test sets of signer- independent when training with data set 1a-15a and test with various complex natural backgrounds.

test	sgFCMed			sgFPCMed			sgPFCMed			sgUPCMed			sgUPFCMed		
	Sift threshold			Sift threshold			Sift threshold			Sift threshold			Sift threshold		
	0.65	0.7	0.75	0.65	0.7	0.75	0.65	0.7	0.75	0.65	0.7	0.75	0.65	0.7	0.75
21	61.00	60.00	55.00	71.00	69.00	55.00	71.00	74.00	58.00	72.00	76.00	66.00	74.00	76.00	72.00
22	57.00	55.00	47.00	61.00	66.00	51.00	63.00	73.00	54.00	72.00	80.00	58.00	74.00	78.00	70.00
23	63.00	57.00	47.00	63.00	58.00	63.00	66.00	69.00	57.00	65.00	66.00	64.00	66.00	68.00	68.00
24	58.00	54.00	40.00	65.00	55.00	57.00	62.00	61.00	58.00	68.00	68.00	66.00	69.00	71.00	70.00
25	57.00	57.00	42.00	57.00	57.00	56.00	59.00	58.00	56.00	54.00	56.00	62.00	64.00	66.00	64.00

Whereas, the average of the best classification results of the blind test dataset from 2a–5a, 2b–5b (signer semi-independent) is around 65% from sgUPFCMed signer-independent are not as good as those of the signer-dependent and signer-semi-dependent, the average of classification rates in each category are above 90%.The average classification rates of sgFCMed, sgFPCMed, sgPFCMed and sgUPCMed are 61.09%, 62.23%, 63.67% and 63.27% as shown in Table 5.7.

The average of the best classification results from 6a-15a, 6b-15b, and 16–20 (signer-independent) is 57.10% from sgUPFCMed as shown in Table 5.8, and the average classification rates of sgFCMed, sgFPCMed, sgPFCMed and sgUPCMed are 52.20%, 53.91%, 55.68% and 54%. The results of signer semi-independent and signer-independent provide low accuracy rate may be because the signer from blind test and training dataset are different.

From Tables 5.9, when we trained on the data sets 1a-5a, the average of the best classification rates of the blind test data sets of signer-dependent (subjects 1–5) are 97.20%, 96.28% and 93.61% with 0.65, 0.7 and 0.75 SIFT threshold, respectively from sgUPFCMed. Whereas the average classification rates of sgFCMed, sgFPCMed, sgPFCMed and sgUPCMed are 93.90%, 93.61%, 96.27% and 95.24%.

From Table 5.10, the best average classification rates of the blind test data sets of signer-independent (subjects 6 to 15) are 65.80%, 67.06% and 70.39% with 0.65, 0.7 and 0.75 SIFT threshold, respectively from sgUPFCMed, and that of signer-independent (subjects 16–20) are 65.40%, 68.60% and 74.41% with 0.65, 0.7 and 0.75 SIFT threshold, respectively from sgUPFCMed. Whereas the average classification rates of the blind test data sets of subjects 6 to 15from sgFCMed, sgFPCMed, sgPFCMed and sgUPCMed are 65.03%, 67.56%, 68.91% and 67.64% with 0.75 SIFT threshold and that of subjects 16–20 are 70.5%, 71.60%, 73% and 71.8% with 0.75 SIFT from sgFCMed, sgFPCMed, sgPFCMed and sgUPCMed, respectively

From the Tables 5.9 and 5.10 in our experiments, we can see that the classification rates for all type of signer are increased as we increase the number of signers in training process of the string grammar fuzzy clustering.

When we are training on the data sets 1a-15a, we utilize 5 string grammar fuzzy clustering algorithms on all data sets and the classification rates are also shown in Tables 5.11 to 5.13. We can see that the best average classification rates of the blind data sets of signer-dependent (subjects 1–5) with 0.65, 0.7 and 0.75 SIFT threshold are 93.27%, 93.70% and 92.13%, respectively from sgUPFCMed, whereas that of signer-semi-dependent (subjects 6 to 15) are 84.52%, 84.86% and 81.95%, respectively from sgUPFCMed. The best average classification rates of the blind data sets of signer-independent (subjects 16–20) are 82.5%, 82% and 80.3% with 0.65, 0.7 and 0.75 SIFT threshold, respectively from sgUPFCMed.

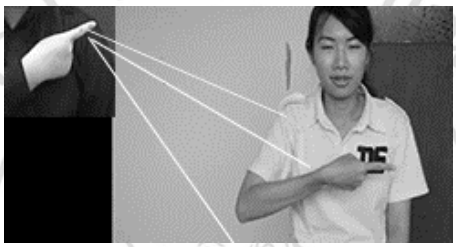


Figure 5.7 Mismatched keypoints from sift process, the hand gesture “e” is assigned to test image using SIFT and test images without constraint

Moreover, we also implement our system trained with 1a–15a on subjects 21-25 without constraint in which the signer wears any shirt and stand in the front of natural backgrounds while they take action each sign for five times of each subject. The classification results are also shown in Table 5.14. We can see that the best correct classification rates on five subjects are 76%, 78% 68%, 71% and 66%, respectively, at 0.7 SIFT threshold from sgUPFCMed. Since the signers of this testing set (subjects 21-25) are different signers from training dataset and signature library, the results of this experiment provide low classification. Furthermore, when we use SIFT for the unconstrained system with complex natural background, the keypoints might be incorrectly matched as shown Figure 5.7. We can use the equation 5.1 to find the correct symbol for each test frame even though it has some mismatched keypoints from SIFT process. However, our algorithm cannot find the right symbol if it is found that there are too many mismatched keypoints.

5.1.4 Comparison the results of string grammar fuzzy clustering with the other methods

We compare the performance of our algorithms with the reported classification rate results of Thai sign language translation system (TSL) [60] that use Hidden Markov Model (HMM) on the same dataset . The comparison can be done between this method and the best average of our translation system. Our system yields a pretty good result that is comparable with TSL for all experiment.

Table 5.15 The comparison of classification rate on test sets of our proposed method with TSL (with HMM)

	Mode	The best average of classification rate (%)		
		Sift threshold		
		0.65	0.7	0.75
TSL (with HMM)	signer-dependent	88.60	88.29	87.82
	signer-semi-dependent	80.35	80.45	80.55
	signer-independent	76.75	76.32	75.23
Proposed Method	signer-dependent	93.27	93.70	92.13
	signer-semi-dependent	84.52	84.86	81.95
	signer-independent	82.50	82.00	80.93

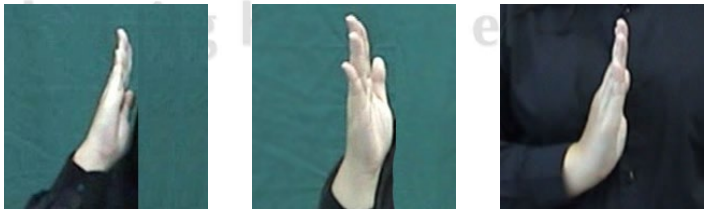
It is difficult to directly compare our method with the other methods because the sign languages from other countries are different from Thai sign language. However, we indirectly compared the performance of our algorithm with American Sign Language (ASL) [64], Arabic Sign Language (ArSL) [65] and Malaysian Sign Language (MSL) [66] as shown in Table 5.16.

We can see that our algorithm can be comparable to ArSL and give the better result than the remaining other methods. However, some misclassifications have occurred in our system. It might be a result of there are some hand gestures in the signature library that are very similar to each other as shown in Figure 5.8. Moreover, it might be some keyframes in signature library that are blurred which effect of movement of signer when

performing the action of sign language as shown in Figure 5.9. In addition, it might be because each signer perform any action is different to each other in signature library as shown in Figure 5.10 that it cause test frame has too many mismatched keypoints which can affect the accuracy in classification process.

Table 5.16 The indirect comparison of classification rate on test sets of our proposed method with the other methods

	Instrument used	Mode	Pre-process with segmentation	#of signers	Classification rate (%)
ASL [64]	None: free hand	signer-independent	Yes	3	89.09
ArSL [65]	None: free hand	signer-dependent	Yes	18	97.4
	None: free hand	signer-independent	Yes	18	90.6
MSL [66]	None: free hand	signer-dependent	N/A	36	75.33
	None: free hand	signer-semi-dependent	N/A	36	69.67
	None: free hand	signer-independent	N/A	36	74
TSL (with HMM)	None: free hand	signer-dependent	No	5	86-95 (on average)
	None: free hand	signer-semi-dependent	No	10	80 (on average)
	None: free hand	signer-independent	No	5	75-76 (on average)
Proposed Method	None: free hand	signer-dependent	No	5	92-93 (on average)
	None: free hand	signer-semi-dependent	No	10	81-85 (on average)
	None: free hand	signer-independent	No	5	80-83 (on average)



(a) (b) (c)

Figure 5.8 The hand part of similar keyframe (a) “g” (b) “g2” (c) “k”.

(a) (b) (c)

Figure 5.9 The hand part of blurred keyframe (a) “kl1” (b) “kl2” (c) “nh1”.

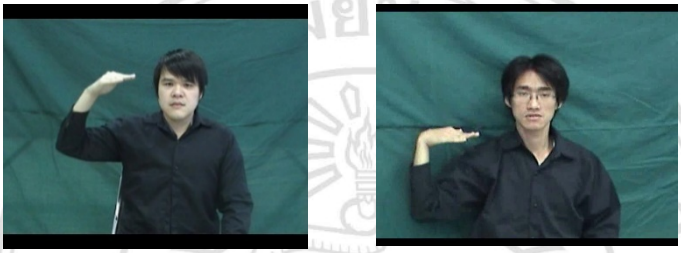


Figure 5.10 The hand part of each signer is different to each other that perform the same action

In this work, we improve dynamic Thai sign language translation system with video caption without prior hand region detection and segmentation using SIFT and various string grammar fuzzy clustering. The SIFT use to match test frame with symbols in the signature library. String Grammar Fuzzy clustering, i.e. string grammar fuzzy C-medians, string grammar fuzzy possibilistic C-Medians, string grammar possibilistic fuzzy C-medians, string grammar unsupervised possibilistic C-medians, and string grammar unsupervised possibilistic fuzzy C-medians are utilized for finding the multi-prototypes of each sign. The fuzzy K-nearest neighbor is used to find a matched sign words. We found that the best result of the blind data sets of signer-dependent is in between 92% and 93% from string grammar unsupervised possibilistic fuzzy C-medians on average and the average of that of signer-semi-independent (same subjects used in the string grammar clustering training) is around 81-85% from string grammar unsupervised possibilistic fuzzy C-medians. Whereas the best average classification rate of the blind data sets of signer-independent is 80-83% from string grammar unsupervised possibilistic fuzzy C-medians. Moreover, our system can perform translating each video without the need for any pre-processing techniques, i.e., segmentation and hand detection. The SIFT

more provides informative about position, shape and orientation of the hand and finger. This allows the system to be able to recognize the hand sign words that have similar gestures. However, when we test our algorithm without constraint on five signers (subjects 21-25) who are asked to stand in front of various complex backgrounds and can wear any shirt. The best correct classification rate in this case is around 66-78% on the average. We also compare with the method of HMM, we can see the best classification of our methods is better than HMM for all experiments. Since the HMM may create the higher misclassification than our method because there might be a chance that the model that gives the maximum value is not the right one. Whereas our method not only choose the maximum one, it utilizes string grammar fuzzy clustering for finding the multi-prototypes and after that FKNN will choose the closest string prototypes using K-nearest neighbor. Although our system provides better classification rate than previous methods for Thai hand sign language, there are still some issues that can improve its performance. For instance, string representation process could be improved by using other features because SIFT cannot extract some interesting point from images with complex natural background. Hence, the classification rate in this case is low. Moreover, our system performs without any pre-processing. The accuracy results might be reduced.

5.2 Identification of Cardio-Pulmonary Resuscitation activity in Medical Simulation Videos

Cardio-pulmonary resuscitation (CPR) [71] is an emergency lifesaving procedure are taken to restore breathing in a person who is in cardiac arrest. The CPR technique is a process that consists of chest compressions and follows by breaths. The procedures can help brain and organ damage from oxygen. The typical recommendation is 30 chest compressions and two breaths. Repeat the cycle until the patient begins to recover. However, CPR has the risk because pressing on the chest can cause a sore chest, broken ribs or a collapsed lung. Hence, the person who performs CPR must be well trained before.

In our experiment, the CPR video dataset was collected by Panicker, A (member of MRIL Lab at University of Louisville). She collected dataset from Pediatric Assessment

Resuscitation and Communication (SPARC) group of the Department of Pediatrics at University of Louisville, USA. Each simulation sessions video involves 4 to 9 trainees, and the length of each video is about 15 minutes to one hour. All video session need to be manually revised by expert or instructor to help trainees to reflect on their experiences and teach them to be more efficient during such real life scenarios. Since an amount of video scene in simulation sessions that need to be manually reviewed by an instructor is enormous. In each video does not only have CPR activity but include other activities in video. This process is hard work, and it takes a long time.

From problem mentioned above, we developed automated tools for CPR activity frames identification from an extensive CPR training sessions database. It will help the instructor to review and debrief team easily. In this work, we identify video frame that involve CPR by using spatio temporal descriptor based on three-dimensional gradients (HOG3D) [67] and self-organizing feature map (SOM) [68] for string representation and by using string grammar fuzzy clustering model, i.e., String Grammar Fuzzy C-Medians, String Grammar Fuzzy Possibilistic C-Medians, String Grammar Possibilistic Fuzzy C-Medians, String Grammar Unsupervised Possibilistic C-Medians and String Grammar Unsupervised Possibilistic Fuzzy C-Medians for finding prototypes of CPR and non-CPR. We then use fuzzy K-nearest neighbor to classify the test dataset. Our process does not require expensive optical flow computations or motion tracking for the CPR activity recognition.

5.2.1 System Description

We first manually select few frames that contain CPR activity from the video and extract bounding box containing only the hand part of the person who is performing the CPR activity with the size of 40×40 pixels. For each CPR activity sequence includes 12 frames that will have at least 2-3 CPR cycles which are essential for capturing the rhythmic activity in the temporal dimension. For non-CPR sequences, we also manually select blocks with the size of 40 x 40 at random spatial and temporal locations that do not overlap with blocks of CPR regions.

In order to identify CPR activity frame in medical simulation videos, we proposed the method without pre-processing i.e. segmentation, tracking specifics region and motion detection to avoid error occurs in these process. The HOG3D [67] is utilized as feature descriptor, and it has been proved that HOG3D give an efficient feature for identification of Cardio-Pulmonary Resuscitation activity in medical Simulation videos better than other methods [72]. Hence we follow [72] by using the HOG3D as a feature vector descriptor to extract multi-scale spatio-temporal interest points (STIPs) from video volume which serve as the training patterns of self-organizing map (SOM) [68]. We then transform the result of SOM to string. Hence, we get string used as primitives in string grammar clustering. The diagram of the proposed system is given in Figure 5.11 which composed mainly of four steps, i.e., feature extraction using HOG3D, string representation using SOM, string grammar clustering and classification. Our approach consists of the following main steps:

1) Feature extraction using HOG3D features

For each of the CPR and non-CPR training volume with the size of 40 x 40 x12. The overview of the process for feature extraction using HOG3D [67] is depicted in Figure 5.12, and the detail is described as follows:

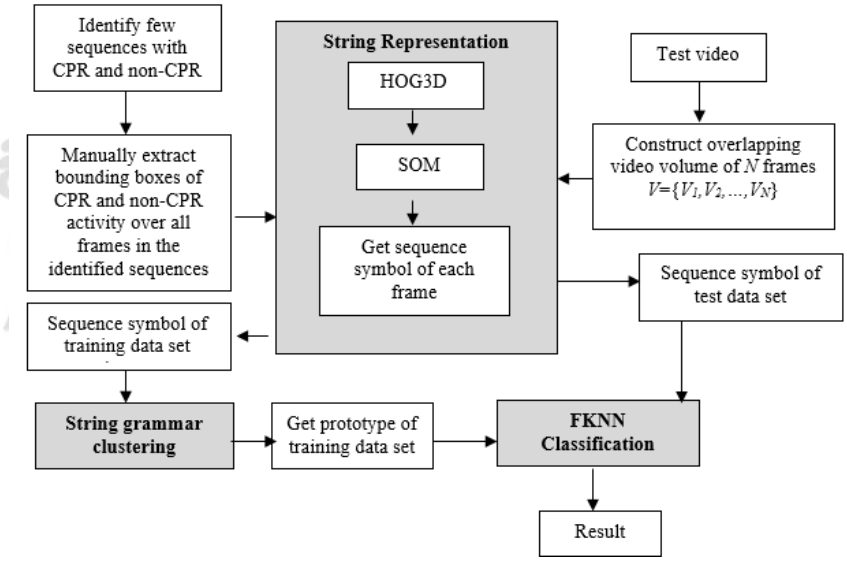
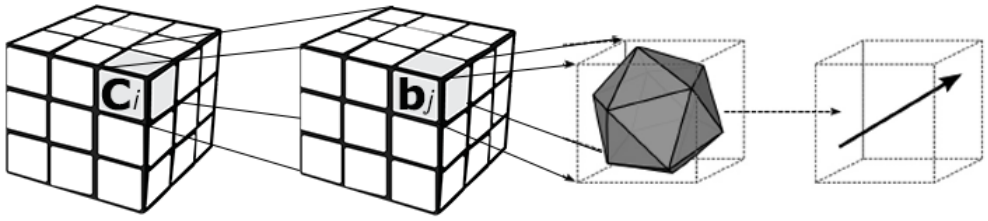


Figure 5.11 The diagram of the proposed system for identification CPR activity frame in medical simulation videos



$$\mathbf{HOG3D} = (\mathbf{h}_1, ..., \mathbf{h}_{M^2N})^{\mathrm{T}} \qquad \mathbf{h}_{\mathbf{c}i} = \sum_{j=1}^{s^3} \mathbf{q}_{\mathbf{b}j} \qquad \mathbf{q}_{\mathbf{b}j} = \qquad \bar{\mathbf{g}}_j = \begin{pmatrix} \bar{g}_{\partial x,i} \\ \bar{g}_{\partial y,i} \\ \bar{g}_{\partial t,i} \end{pmatrix}$$

(a) (b) (c) (d)

Figure 5.12 The overview of the process for feature extraction using HOG3D; (a)full descriptor of HOG3D with 3x3x3 histogram cell; (b) histogram computation over 3x3x3 sub-blocks; (c) each gradient orientation is quantized using regular polyhedrons; (d) each mean gradient is computed using integral videos.

1.1) Each volume is divided into a grid of cells with the size of M x M x N (here M=N=3). Thus, any given cell $\mathbf{c} = (x,y,t,w,h,l)^{\mathrm{T}}$, where $(x,y,t)^{\mathrm{T}}$ denote the position and w, h and l its width, height, and length, respectively.

1.2) To construct histogram, cell is divided into S^3 (here S = 3) blocks, (each cell is divided into 3 x 3 x 3 sub-blocks ($\mathbf{b}_i$))

1.3) Compute the mean gradient vectors of each sub-block ($\vec{g}_{\mathbf{b}i}$) using integral videos [69] whose the mean gradients will be quantized by projecting them on a 20-dimensional regular polygon with the gradient magnitude as its weight ($\mathbf{q}_{\mathbf{b}i}$) [70].

1.4) The histogram $\mathbf{h}_{\mathbf{c}}$ for the region $\mathbf{c}$can be computed using the sum of the quantized mean gradients $\mathbf{q}_{\mathbf{b}i}$ of all sub-blocks $\mathbf{b}_i$:

$$\mathbf{h}_{\mathbf{c}} = \sum_{i=1}^{s^3} \mathbf{q}_{\mathbf{b}i} \text{ for i=1, 2, …, } M^2N \qquad (5.3)$$

1.5) These histograms are finally concatenated to one feature of the HOG3D descriptor of 3 x 3x 3 x 20 = 540 dimensional descriptors. We will get the HOG3D features are extracted from both the training and testing videos which serve as the training patterns of SOM.

2) String Representation using Self organizing Feature Map

The self-organizing feature map (SOM) is the one of the unsupervised neural network that introduced by T. Kohonen [68]. SOM can project high dimensional data to lower dimensional lattice and can measure similarities in data as well. SOM is formed as neurons with two dimensional grid (here we use neuron spacing in hexagonal grid). Each neuron of SOM is represented by *n*-dimensional weight where *n* is the size of input data. The weight vectors of neurons are updated iteratively in training process. The best matching unit (BMU) is the winning neuron that weight vector is most similar to input. In this work, map size of SOM is 8×5. Transformation of HOG3D descriptor into a trajectory on the map is showed in Figure 5.13. The chains of the best matching units (BMUs) is used as string for string grammar fuzzy clustering step.

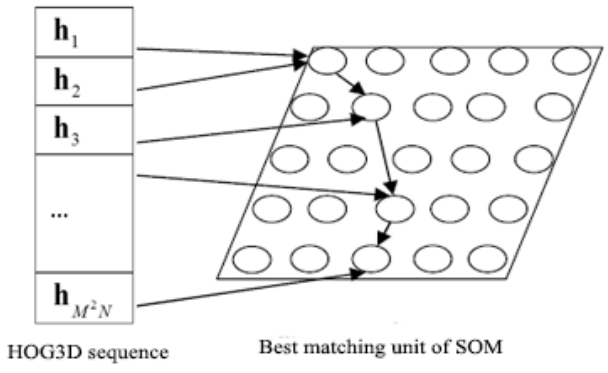


Figure 5.13 Transformation of HOG3D descriptor into a trajectory on the map.

3) String grammar fuzzy clustering Algorithm

We use five string grammar clustering algorithms, i.e., sgFCMed, sgFPCMed, sgPFCMed, sgUPCMed and sgUPFCMed for finding multi-prototypes of

CPR dataset. For our five string grammar clustering algorithms which are utilized for finding multi-prototype of CPR, more details can be referred in Chapter 3. The parameters setting for identification CPR activity frame in medical simulation videos data set is shown in Table 5.17.

Table 5.17 The parameter setting of our algorithms for CPR activity dataset.

parameter	sgFCMed	sgFPCMed	sgPFCMed	sgUPCMed	sgUPFCMed
m	2	2	2	2	2
η	-	2	2	-	2
γ	-	-	2	-	-
a	-	-	1	-	1
b	-	-	4	-	4
stopping criteria	0.1	0.1	0.1	0.1	0.1
maximum number of iterations	100	100	100	100	100

4) Classification using Fuzzy K-nearest neighbor

After we create the multi-prototypes of each class of CPR and non-CPR, i.e. $SC = \left\{ sc_1^1, ..., sc_{N_1}^1, sc_1^2, ..., sc_{N_2}^2, sc_1^C, ..., sc_{N_C}^C \right\}$ where sc_k^j is string prototype k of class j for each class from sgFCMed, sgFPCMed ,sgPFCMed,sgUPCMed, and sgUPFCMed with 5, 10, 15 and 20 clusters on each class separately to create multi-prototypes for each class are created. We apply FKNN [62] to classify the test dataset, more details can be referred in subheading 5.2.2.

5.2.2 Results and Discussion

We validate our algorithms on three CPR simulation video datasets which are provided by the (SPARC) working group at Kosair Children's Hospital, Louisville, USA. The sample of frames are shown in Figure 5.14. The duration of these simulation sessions is roughly 30 minutes, with 29 frames per sec. Each of the frames originally had a resolution of 720 x 480. Since high-resolution frames are unnecessary for edge computation, we reduce the spatial resolution to 360 x 240.

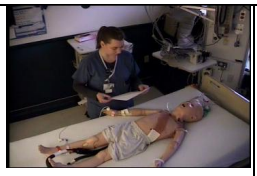
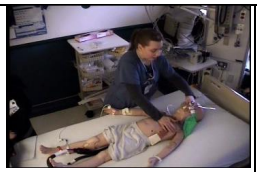
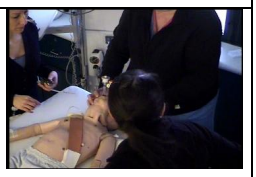
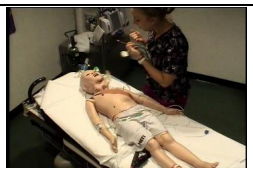
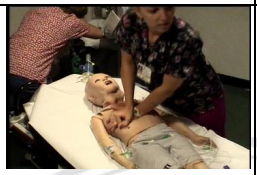
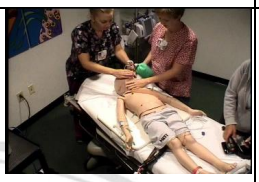
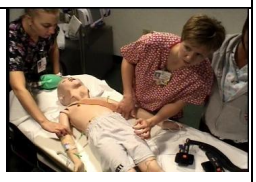
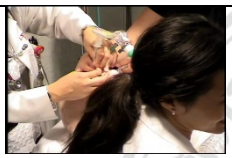
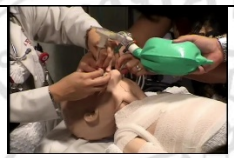
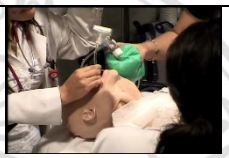
video	The sample of video frame			
CPR 1				
CPR 2				
CPR 3				

Figure 5.14 The sample of CPR video frame

For training process, 200 frames of the hand part of the person who is performing the CPR activity with the size of 40×40 pixels from video CPR1 are manually selected. We then selected 12 overlapping frames for volume constructions that will have at least 2-3 CPR cycles which are essential for capturing the rhythmic activity in the temporal dimension for each CPR activity sequence. For non-CPR sequences, we also manually select 160 blocks with the size of 40 x 40 at random spatial and temporal locations that do not overlap with blocks of CPR regions.

We use a 10-fold cross validation with the training dataset for performance evaluation of our algorithms. The predictive performances were evaluated for accuracy rate (ACC). To finding the multi-prototypes of class CPR and class non-CPR, the sgFCMed, sgFPCMed, sgPFCMed, sgUPCMed, and sgUPFCMed with 4, 8, and 12 clusters are utilized. The FKNN was selected to classify the CPR activity frame with K = 1 for 1 prototype and FKNN with K = 5 for 5, 10,15 and 20 prototypes.

Table 5.18 The average classification rate of the validation set from 10-fold cross validation.

#prototype	sgFCMed	sgFPCMed	sgPFCMed	sgUPCMed	sgUPFCMed
1	86.44	83.87	89.39	80.56	91.93
5	87.08	89.97	92.43	83.33	94.66
10	87.79	91.61	92.93	86.11	94.91
15	88.30	92.60	93.59	86.11	95.08
20	88.45	95.27	96.08	88.89	96.12

In Table 5.18, we compare the classification accuracy performance our string grammar fuzzy clustering using different number of prototypes on each class with p=1, 5, 10, 15 and 20. The FKNN classifier using sgUPFCMed with 20-prototypes gives the best average accuracy of 96.12% with only 3.88% error rate. As indicated, the predictive performance reached 91.93% when using only single-prototypes. When we use the algorithms that based on only membership such as the sgFCMed, the best accuracy rate in this case is 88.45% with 20-prototypes. The accuracy rate cannot compete with the algorithms that based on both membership and possibilistic, may be because outliers are included in our datasets.

For the indirect comparison, we compare our results with those from the HMM [71] and SVM [72], we can see that the best of our results is better than others methods as shown in Table 5.19.

Table 5.19 The comparison on the validation set with other methods.

Method	classification correct rate (%)
sgFCMed	88.45
sgFPCMed	95.27
sgPFCMed	96.08
sgUPCMed	88.89
sgUPFCMed	96.12
HMM [71]	~80
SVM [72]	~90

For testing process, we sample the frames by a factor of 10 of each video. The details of test frames of each video are shown in Table 5.20. The video frames are divided into K overlapping volumes of N frames each. Let $V = \{V_1, V_2, \ldots V_K\}$ denote the set of video volumes. From each volume V_i, where i = 1, 2, …K, sub-blocks $\{sv_1, sv_2 \ldots sv_l\}$ are extracted with 50% overlap. Hence, each volume has 160 sub-blocks in total. We also

construct N overlapping frames for volume constructions. For example, if N = 12, then V_1 will be frame1 to frame 12, V_2 will be frame 2 to frame 13 and so on as shown in Figure 5.15. The HOG3D and SOM are computed for each sub-block using the same parameters used for training. Finally, the test string is assigned to a class with the closest prototype belongs to according to the FKNN based on Levenshtein distance with K=5 and m=2 for all experiments. For identification any given video volume as a CPR activity or non-CPR activity, a filter criterion is f introduced which is that there must be at least four windows with a membership greater than 0.5 for the video volume to be classified as a CPR activity. This filtering removes the influence of most of the incorrectly classified sub-window volumes. In Table 5.21-5.23, we compare the classification accuracy performance our string grammar fuzzy clustering using different number of prototypes on each class with p =5, 10, 15 and 20 for videos CPR1, CPR2, and CPR3.

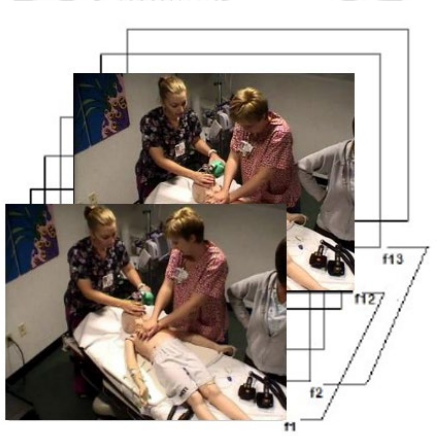


Figure 5.15 Construction of video volume

Table 5.20 The detail of three CPR simulation sessions video in our experiment.

Video	**Length of video**	**Number of test frame**
CPR1	19 minutes 28 second	4992
CPR2	16 minutes 1 second	2880
CPR3	14minutes 57 second	2772

Table 5.21 Classification rate on test sets of video CPR1

#prototype	sgFCMed	sgFPCMed	sgPFCMed	sgUPCMed	sgUPFCMed
5	83.21	89.43	89.75	89.68	90.66
10	83.68	90.02	90.45	90.42	90.98
15	87.97	92.51	93.98	93.91	94.17
20	88.49	93	94.07	94.05	95.77

From Table 5.21, The best sgUPFCMed yields 95.77%. correct classification on the video CPR1. The sgFCMed, sgFPCMed, sgPFCMed and sgUPCMed yield 88.49%, 93%, 94.07% and 94.05, respectively with 20-prototypes.

Table 5.22 Classification rate on test sets of video CPR2

#prototype	sgFCMed	sgFPCMed	sgPFCMed	sgUPCMed	sgUPFCMed
5	71.46	82.23	83.69	83.58	86.75
10	72.43	82.6	84.33	84.27	87.13
15	73.75	85.15	89.05	89.03	90.66
20	79.36	90.74	91.64	91.56	92.58

From Table 5.22, The best sgUPFCMed yields 92.58% correct classification on the blind test data set. Whereas the sgFCMed, sgFPCMed, sgPFCMed and sgUPCMed give 79.36%, 90.74%, 91.64% and 91.56% correct classification, respectively, with 20-prototypes.

Table 5.23 Classification rate on test sets of video CPR3

#prototype	sgFCMed	sgFPCMed	sgPFCMed	sgUPCMed	sgUPFCMed
5	74.67	77.79	81.69	85.68	84.67
10	77.56	79.22	83.33	85.79	87.44
15	82.48	84.72	85.62	86.47	88.65
20	84.68	86.88	89.09	86.87	90.26

From Table 5.23, again, we can see that sgUPFCMed gives 90.26% correct classification rate on video CPR3. Whereas the sgFCMed, sgFPCMed, sgPFCMed and sgUPCMed provide 84.68%, 86.88%, 89.09% and 86.87% correct classification, respectively, with 20-prototypes.

From all of the results from Tables 5.21 to 5.23, we can see the sgUPFCMed with 20-prototypes provide the best accuracy for all three videos dataset. The sgUPFCMed give 95.77%, 92.58%, and 90.26%, for videos CPR1, video CPR2 and CPR 3, respectively. The video CPR1 give the highest classification rate because few frames of video CPR1 are included in training process, but not for video CPR2 and CPR3.

In our work, we improved identification CPR activity in simulation in Medical Simulation Videos. The method of CPR action classification gets rid of the video shot detection and segmentation phase which is usually the first step of most common video classification algorithms. For CPR activity detection, we find multi-prototypes using the sgFCMed, sgFPCMed, sgPFCMed, sgUPCMed, and sgUPFCMed and classify by FKNN classifier with output string sequence from SOM with three dimensional spatio-temporal oriented gradients to discriminate between CPR and non-CPR activity. The proposed method is very straight forward and easier compared to the other methods of activity detection in medical simulations videos. It achieves a better classification accuracy as compared to the HMM and SVM classifier (Table 5.19). The proposed approach was evaluated three video simulation sessions. We have shown that our purposed methods (sgUPFCMed) can correct identity of the CPR activity with 95.77%, 92.58%, and 90.26, for videos CPR1, CPR 2 and CPR 3, respectively.

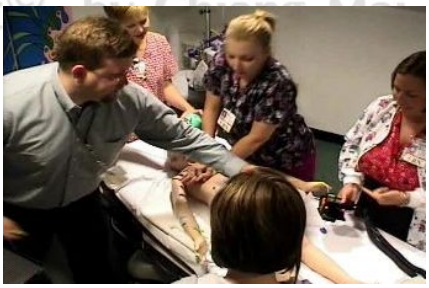


Figure 5.16 Some CPR activity sequences that have anything such as people cover the hand part of CPR activity.

However, some misclassifications have occurred in our system. It might be caused by movement of people in the scene or there are some CPR activity sequences that have anything such as people cover the hand part of CPR activity (Figure 5.16) because the CPR action is performed without noticeable movements of body or hands of the actors. Hence, the system can fail to identity CPR frame for these frames.

Although our system has been proved to be an efficient approach to classify video scenes with CPR activity, there are still some issues that can improve its performance. For instance, feature extraction could be improved by using other features because the time and space complexity of HOG3D feature is very high. In addition, our system has been trained using only some hand part of the person who is performing the CPR activity from video CPR1. If we add more data from other videos in training set, it might increase the accuracy result of our system. Moreover, our system performs without any pre-processing i.e. segmentation, tracking specifics region and motion detection. The accuracy results might be reduced. Moreover, our algorithms use multi- prototypes of each class in FKNN testing process, not all string. Choosing the number of prototypes of CPR activity and non-CPR activity is important. We should preserve discrimination between classes for improving the classification accuracy rate. We suggest choosing the number of prototypes of each class large enough that noise in the data is minimized and small enough so the samples of the other classes are not included. For this dataset, 20-prototypes provide the best accuracy for all three videos dataset.