An Effective Dynamic Gesture Recognition System Based on the Feature Vector Reduction for SURF and LCS

Pablo V.A. Barros, Nestor T.M. Júnior, Juvenal M.M. Bisneto,
Bruno J.T. Fernandes, Byron L.D. Bezerra, and Sérgio M.M. Fernandes

Polytechnics School - University of Pernambuco - Recife, Brazil
{pvab,ntmj,jmmb,bjtf,byronleite,smmf}@ecomp.poli.br

Abstract. Speed Up Robust Feature (SURF) and Local Contour Sequence(LCS) are methods used for feature extraction techniques for dynamic gesture recognition. A problem presented by these techniques is the large amount of data in the output vector which difficult the classification task. This paper presents a novel method for dimensionality reduction of the features extracted by SURF and LCS, called Convexity Approach. The proposed method is evaluated in a gesture recognition task and improves the recognition rate of LCS while SURF while decreases the amount of data in the output vector.

1 Introduction

Hand gesture recognition system provides a natural, user-friendly way of interaction with the computer which is more familiar to the human beings. Gesture recognition has a wide area of application including human machine interaction, sign language and video game technology. Most of the dynamic gesture recognition systems are divided in two steps: feature extraction and pattern classification. Several techniques are applied for feature extraction but many of them might fall in dimensionality curse as show in the survey published by Bilal et al[3]. The dimensionality curse says that the numerical approximation of a function will require more computation as the number of active variables grows [9].

Meena [11] uses the Local Contour Sequence (LCS) algorithm to calculate the feature vector for gesture recognition. This technique generates an output with many features by the calculation of the distance of all the points that compound a hand posture, which can difficult the learning process. Bao et al. [1] use the Speed Up Robust Features (SURF) to recognize gestures trough tracking. In their work, they use the SURF points in adjacent frames to help describing a hand trajectory. SURF generates many points, as it is applied in all the image an returns the interest points based in different image transformations and not in the observed object. Jiang et al. [8] describe full body gestures using SURF and Bag of Video Worlds Model to normalize the interest points extracted by SURF. This method uses a set of SURF descriptors for each interest point

V. Mladenov et al. (Eds.): ICANN 2013, LNCS 8131, pp. 412–419, 2013.

extracted. This descriptors are obtained summing the HAAR wavelets response around the interest points and thus generates a large feature vector, containing each description vector for all the interest points. Yao et al. [15] use SURF to extract the key points of a hand posture and uses Adaboost to decrease the computational cost for the training. Yao et al. method increase the classification efficiency by using a hybrid classification model, but the feature vector still contain the same amount of features, and it could be minimized.

This paper introduces a novel method to reduce the feature vector size. The Convexity approach extends LCS and SURF and shows a better result in classification of dynamic gestures. A experiment in gesture recognition is performed and demonstrate the efficiency and efficacy of the approach with three different pattern classification methods: Dynamic Time Wrapper [13], Hidden Markov Model [12] and Elman Recurrent Neural Network [7].

This paper is structured as follows: Section II describes the Convexity Approach algorithm. Section III presents the experimental results. Finally, in Section IV, the conclusions and some future work are given.

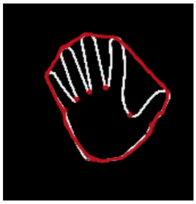
2 Convexity Approach

The Convexity Approach reduces a feature vector, choosing the smallest group of points that can represent the hand posture. This algorithm is applied in Local Contour Sequence (LCS), created by Gupta [4], and in Speed Up Robust Features (SURF) described by Bay [2] . The input for the Convexity Approach is a set of points that represents the hand posture. The first step of the Convexity Approach algorithm is to minimize the hand posture. The second step is to find the convex hull of the previously selected points. The last step uses the points that composes the convex hull for a feature calculation based on point distance. Figure 1 shows the execution illustration of the Convexity Approach.

Fig. 1. Convexity Approach execution illustration

The Douglas-Peucker algorithm [6] is used in the first step to create an approximation curve of the external points, forming a minimized polygon for hand posture. In this algorithm, the two extreme endpoints of a set of points are connected with a straight line as the initial rough approximation of the polygon. Then, it approximates the whole polygon by computing the distance from all intermediate polygons vertices to that line segment. If all these distances are less than the specified tolerance T, then the approximation is good, the endpoints are retained, and the other vertices are eliminated. However, if any of these distances exceeds the T tolerance, then the approximation is not good enough.

(a) Polygon Minimization



(b) Convex Hull and inner points selection

Fig. 2. Outputs of step one (a) and two (b) of Convexity Approach

In this case, it chooses the point that is furthest away as a new vertex subdividing the original set points into two set points. This procedure is repeated recursively on these two shorter set points. If at any time, all of the intermediate distances are less than the T threshold, then all the intermediate points are eliminated. The routine continues until all possible points have been eliminated. Figure 2(a) shows the output of this step.

The second step is to select the most significant points for the specifically hand posture. We run the Sklankys [14] algorithm in the last step output.The algorithm consists in the following sequence:

- The convex vertex of the polygon is found.
- The remaining n-1 vertexes are named in clockwise order starting at P0.
- Select P0, P1 and P2 vertices and call then “Back”, “Center” and “Front” respectively
- Execute the follow algorithm:
  - **while** “Front” is not on vertex P0 and “Back”, “Center” and “Front” form a right turn **do**
    if “Back”, “Center” and “Front” form a left turn or are collinear vertex **then**
    change “Back” to the vertex ahead of “Front”. Relabel “Back” to “Front”, “Front” to “Center” and “Center” to “Back”.
    else if “Back”, “Center” and “Front” turn left **then**
    change “Center” to the vertex behind “Back”, Remove the vertex and associated edges that “Center” was on and relabel “Center” to “Back” and “Back” to “Center”
    end if
    end while
- For each pair of selected points the algorithm traces a line. The most distant point of this line is selected as an inner point. Figure 2(b) shows the resultant points of the algorithm in a input image.

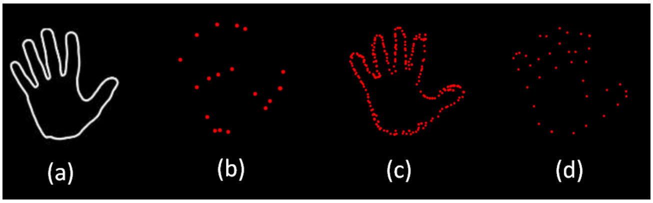


Fig. 3. Applying the Convexity Approach on LCS (a) produces a new output vector (b). Applying the Convexity Approach on SURF (c) produces a new output vector (d).

The last step of convexity approach is the feature extraction based on distance calculation. A line is formed by each pair of the external points chosen by step two. The distance between this line and the closer inner point is calculated and added to the output vector. Figure 3 shows the result obtained by LCS and SURF and the application of Convexity Approach, respectively CLCS and CSURF.

For some classification techniques, such as Neural Networks, the feature vector must be normalized. To solve this problem we propose a method. First, the number of normalized distances is defined. Then if the image has fewer points than the determined one, "0" are added in the feature vector until matches the desired length. The outputs with more points than the desired length are normalized using a selection algorithm. This algorithm consists in calculation of a window, W, through the division of the output length for the desired length. The output vector is traversed, and each W position is added to the new vector of outputs. If the new output vector is smaller than the desired length, the remaining positions are randomly visited and used to compose the new output vector until the desired length is achieved.

3 Dynamic Gesture Recognition

The efficiency and effectiveness of the Convexity Approach applied to LCS and SURF algorithms is evaluated in this task. We compare the results of the classification of dynamic gestures using the original version of the LCS and SURF with the extended one applying the Convexity Approach, called CLCS and CSURF, respectively.

3.1 Methodological Protocol

The database used for this test is the RPPDI Gesture Database[1]. It contains a set of four different gestures: Click, Grasp, No and Goodbye.

To test the algorithms a dynamic gesture recognition system is used and it is composed by two modules: extraction and classification. Four methods for feature extraction are evaluated: SURF, LCS, CSURF and CLCS. Three classification techniques are build: Dynamic Time Wrapper, Hidden Markov Model

[1] Available at `http://rppdi.ecomp.poli.br/gesture/database/`

Table 1. Best Techniques Configuration Combination with and without Convexity Approach

Technique	Parameters	CLCS	CSURF	LCS	SURF
Elman RNN	Neurons Input Layer	140	210	5600	1400
	Hidden Layers	12	14	75	37
	Neurons Output Layer	4	4	4	4
	Execution Average Time	21.843ms	27.953ms	72.662ms	76.543ms
	Taining Average Time	210824ms	180351ms	400298ms	550221ms
HMM	States	3	3	3	3
	Baum-Welch Iterations	100	100	10	10
	Features	10	10	5600	1400
	Execution Average Time	181.959ms	276.724ms	-	-
	Taining Average Time	0.446ms	0.690ms	-	-
DTW	Features	140	210	5600	1400
	Execution Average Time	45.743ms	164.964ms	640.713ms	1722.289ms
	Taining Average Time	129.015ms	133.620ms	156.048ms	208.191ms

and an Elman Recurrent Neural Network. The experiments are evaluated with the combination of each extraction technique with each classification technique.

The Recurrent Neural Network chosen for the classification module is the Elman Recurrent Neural Network(Elman RNN). The Backprogation with Simulated Annealing training strategy is used as it showed good results for dynamic sequences training in the work of Zhang [16]. The Hidden Markov Model technique uses a K-Means Clustering [5] to find the best initial approximation. The Baum-Welch algorithm [10] is used to train the HMM, resulting in a fast training process.

All experiments are repeated 30 times in a database division containing 2/3 of the sequences in each gesture class for training and 1/3 for testing, randomly chosen. The results present the mean among all repetitions. The best configuration results are showed in Table 1.

3.2 Experimental Results

The 5600 features of the LCS are too excessive to calculate the HMM probabilities, and it could not converge in a result. The same happens when using SURF. The HMM used with CLCS and CSURF reached a recognition rate of 91%, for both, as showed in Table 2.

Elman Recurrent Neural Network recognized 77% of the gestures using the LCS. Using SURF the recognition rate dropped to 72%. The CLCS recognition rate was 90% and the CSURF recognition rate was 92%. Table 3 shows the confusion matrix of all the extraction techniques combined with the RNN.

Table 2. Confusion Matrix using CLCS and CSURF as feature extraction technique techniques and HMM as classification technique

	CLCS				CSURF			
	Gest. 1	Gest 2.	Gest. 3	Gest. 4	Gest. 1	Gest 2.	Gest. 3	Gest. 4
Gest. 1	0	1	4	0	1	0	4	0
Gest. 2	1	6	1	0	0	8	0	0
Gest. 3	0	0	11	0	2	0	9	0
Gest. 4	0	0	0	6	0	0	0	6

Table 3. Confusion Matrix using LCS, CLCS, SURF and CSURF as feature extraction technique techniques and RNN as classification technique

	LCS				CLCS			
	Gest. 1	Gest 2.	Gest. 3	Gest. 4	Gest. 1	Gest 2.	Gest. 3	Gest. 4
Gest. 1	4	3	1	0	7	1	0	0
Gest. 2	2	6	0	0	1	7	0	0
Gest. 3	2	0	9	0	1	0	10	0
Gest. 4	0	0	0	6	0	0	0	6
	SURF				CSURF			
	Gest. 1	Gest 2.	Gest. 3	Gest. 4	Gest. 1	Gest 2.	Gest. 3	Gest. 4
Gest. 1	4	3	1	0	7	0	1	0
Gest. 2	2	6	0	0	0	8	0	0
Gest. 3	3	0	8	0	2	0	9	0
Gest. 4	0	0	0	6	0	0	0	6

Table 4. Confusion Matrix using LCS, CLCS, SURF and CSURF as feature extraction technique techniques and DTW as classification technique

	LCS				CLCS			
	Gest. 1	Gest 2.	Gest. 3	Gest. 4	Gest. 1	Gest 2.	Gest. 3	Gest. 4
Gest. 1	12	3	0	0	13	0	2	0
Gest. 2	2	11	0	3	2	16	0	0
Gest. 3	5	0	14	2	2	0	21	0
Gest. 4	0	0	0	12	0	0	0	12
	SURF				CSURF			
	Gest. 1	Gest 2.	Gest. 3	Gest. 4	Gest. 1	Gest 2.	Gest. 3	Gest. 4
Gest. 1	12	3	0	0	13	2	0	0
Gest. 2	2	11	0	3	1	15	1	0
Gest. 3	5	0	13	3	1	20	0	0
Gest. 4	0	0	0	12	0	0	0	12

Table 5. Classication rate resume for all the Feature Extraction/Classification techniques combinations

Combination	Clas(%)	Combination	Clas(%)	Combination	Clas(%)
HMM + LCS	-	RNN + LCS	77.0	DTW + LCS	77.0
HMM + SURF	-	RNN + SURF	72.0	DTW + SURF	75.0
HMM + CLCS	91.0	RNN + CLCS	90.0	DTW + CLCS	97.0
HMM + CSURF	91.0	RNN + CSURF	92.0	DTW + CSURF	93.0

Using DTW to classify the distances between the sequences provided by the LCS technique generated a recognition rate of 77%. Using the SURF sequences 75% of the gestures were recognized correctly. Using CLCS the recognition rate was 97% and CSURF got 93% of the correct gestures recognized. Table 4 shows the confusion matrix.

Table 5 shows the resume of this session. The combinations of feature extraction techniques with classification techniques was evaluated. Except for LCS and SURF in HMM because the amount of data was too excessive, a problem corrected by Convexity Approach. It shows that the Convexity Approach obtained a higher classification rate in all the combinations. The results shows the use of Convexity Approach decreased the size of extracted feature vector and also increased the recognition rate for all the classification techniques. It shows also that the Convexity Approach lowered the classification and execution time for all techniques.

4 Conclusion

The Convexity Approach creates a minimized feature vector for dynamic gesture recognition. It uses a dynamic selection of points, based in the point significance in the hand posture model, to reduce the total point given by a feature extraction technique.

To evaluate the Convexity Approach in a gesture recognition task, it is used in LCS and SURF, and the results used in two in two dynamic gesture recognition system, one based on HMM and one based on DTW. The results are compared and showed that Convexity Approach got better recognition rates.

The Future Works can be listed as: Apply the Convexity Approach to others gesture feature extraction techniques and test the CLCS and CSURF in benchmark gesture recognition databases.

Acknowledgments. This work was partially supported by Brazilian agencies: CNPq, CAPES and FACEPE. This project is possible due the FACEPE process number APQ-0949-1.03/10 - "Reconhecimento de gestos uma aplicação para reconhecimento de sinais de surdos com dispositivos móveis".

References

1. Bao, J., Song, A., Guo, Y., Tang, H.: Dynamic hand gesture recognition based on surf tracking. In: 2011 International Conference on Electric Information and Control Engineering (ICEICE), pp. 338–341 (April 2011)
2. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-up robust features (surf). Comput. Vis. Image Underst. 110(3), 346–359 (2008), http://dx.doi.org/10.1016/j.cviu.2007.09.014
3. Bilal, S., Akmeliawati, R., El Salami, M., Shafie, A.: Vision-based hand posture detection and recognition for sign language; a study. In: 2011 4th International Conference on Mechatronics (ICOM), pp. 1–6 (2011)
4. Gupta, L., Ma, S.: Gesture-based interaction and communication: automated classification of hand gesture contours. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 31(1), 114–120 (2001)
5. Hartigan, J.A., Wong, M.A.: A k-means clustering algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics) 28, 100–108 (1979)
6. Heckbert, P.S., Garland, M.: Survey of polygonal surface simplification algorithms (1997)
7. Jain, L.: Recurrent Neural Networks, 1st edn. CRC Press (2001)
8. Jiang, X., Sun, T., Feng, B., Jiang, C.: A space-time surf descriptor and its application to action recognition with video words. In: 2011 Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), vol. 3, pp. 1911–1915 (July 2011)
9. Kouiroukidis, N., Evangelidis, G.: The effects of dimensionality curse in high dimensional knn search. In: 2011 15th Panhellenic Conference on Informatics (PCI), pp. 41–45 (2011)
10. Baum, L., Peterie, T., Souled, G., Weiss, N.: A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. In: Proceedings of Annals of Mathematical Statistics, vol. 41, pp. 164–171 (1995)
11. Meena, S.: A Study on Hand Gesture Recognition Technique. Master's thesis, National Institute of Technology, Rourkela,India (2011)
12. Rabiner, L.R.: A tutorial on hidden Markov models and selected applications in speech recognition. In: Readings in Speech Recognition, pp. 267–296. Morgan Kaufmann Publishers Inc., San Francisco (1990), http://dl.acm.org/citation.cfm?id=108235.108253
13. Sakoe, H., Chiba, S.: Dynamic programming algorithm optimization for spoken word recognition. In: Readings in Speech Recognition, pp. 159–165. Morgan Kaufmann Publishers Inc., San Francisco (1990), http://dl.acm.org/citation.cfm?id=108235.108244
14. Sklansky, J.: Finding the convex hull of a simple polygon. Pattern Recogn. Lett. 1(2), 79–83 (1982), http://dx.doi.org/10.1016/0167-8655(82)90016-2
15. Yao, Y., Li, C.-T.: Hand posture recognition using surf with adaptive boosting. In: British Machine Vision Conference (2012)
16. Zhang, H., Wang, Y., Deng, C.: Application of gesture recognition based on simulated annealing bp neural network. In: 2011 International Conference on Electronic and Mechanical Engineering and Information Technology (EMEIT), vol. 1, pp. 178–181 (August 2011)