Online Human In-Hand Manipulation Skill Recognition and Learning *

Disi Chen$^{1[0000-0001-9297-0802]}$, Zhaojie Ju[1], Dalin Zhou[1], Gongfa Li[2], and Honghai Liu[1]

[1] University of Portsmouth, Portsmouth, PO1 3HE, UK
[2] Wuhan University of Science and Technology, 430080, Wuhan, China

Abstract. This work intends to contribute to the development of recognition technologies in human in-hand manipulation skills. This work proposed a probabilistic framework for both human skill representation and high efficient recognition. Gaussian Mixture Model (GMM) as a probabilistic model, is highly applicable in clustering, data fitting and classification. The human in-hand motions were perceived by a wearable data glove, CyberGlove, the motion trajectory data proposed and represented by GMMs. Firstly, only certain amount of motion data were used for batch learning the parameters of GMMs. Then, the newly coming data of human motions will help to update the parameters of the GMMs without observation of the historical training data, through our proposed incremental parameter estimation framework. Recognition in the research takes full advantages of probabilistic model, when the GMMs were trained, the log-likelihood of a candidate trajectory can be used as a measurement to achieve human in-hand manipulation skill recognition. The recognition results of the online trained GMMs show a steady increase in the accuracy, which proved that the incremental learning process improved the performance of human in-hand manipulation skill recognition.

Keywords: In-hand manipulation skills · GMMs · Online learning.

1 INTRODUCTION

In-hand manipulation is defined as a kind of capability of human to interact with an object within one hand [1], during which the object can be moved/rotated or deformed. Recently recognition of human in-hand manipulation, becomes a hot research area in human-machine interaction. As for human beings, in-hand manipulation skills are usually learned by us when we grow up and they play an important role in our daily life. However, to study these skills is a challenging task for the intelligent system or robots. Firstly, to percept the motion of human hand in a high precision is difficult, since the structures of human hand are rather complex, with 27 bones and up to 25 degree of freedom (DoF) [2]. The huge amount of sensory information also leaves a gap in analyzing the

* Supported by grant of the EU Seventh Framework Programme (Grant No. 611391)

actions performed by such a complex physical structure. Additionally, collecting a dataset with an ideal number of samples for classifier training, usually requires massive human actions, this both take a lot of memory and tedious human involvement. To overcome the difficulties mentioned above, we proposed a novel online learning framework for human in-hand manipulation skill recognition. In this framework, fixed component Gaussian Mixture Models (GMM) [3] were applied for trajectory data representation, an Expectation Maximization (EM) [19] inspired incremental learning algorithm helps to update the parameters of GMMs, which released human candidates from tedious repeating one kind of in-hand manipulation at one experiment. In this paper, we also discussed the recognition experiment results based on the motion data of human in-hand manipulations collected by the CyberGlove, to show the performance of our proposed method.

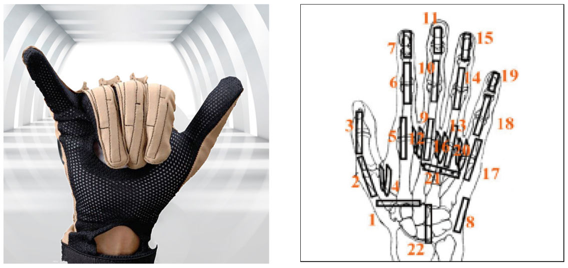
2 RELATED WORK

2.1 In-Hand Motion Capturing

Recently, many literatures studied the human in-hand manipulations [5][6], those researches revealed the relations between the object and human hands in a kinematic way and gives a guidance for designing a sensory system to collect human motion data in manipulations. There are many sensing systems designed for human hand motion measurement, such as data glove [7], camera system [8][9] and electromyography (EMG) system [10][11]. However, due to the high precision and better anti-interference ability compared with EMG system, data glove proved to be one of the most important sensing devices for in-hand motion perception, which measures the degree of bending of wearer's fingers and sent the corresponding signal to the computer. Kuroda [12] et al, proposed an innovative intelligent low-cost data glove named StrinGlove, which is capable to measure the displacements of all 24 DoF of one human hand with 24 inductcoders and 9 contact sensors. Fahn & Sun [13] presented a data glove with only five sensors properly mounted on the palmer suface, which is able to measure 10 DoF of a hand. Using data glove can not only continuously measure the angles of finger joints, but also record the haptic information on the finger tips. In this paper, we choose Cyberglove (Fig. 1(a)) as the sensor for measuring, since this data glove is capable to capture the movements of fingers and palm with 22 built in sensors as shown in Fig. 1(b).

2.2 Motion Skill Representation and Learning

Hand motion data of in-hand manipulation are temporal and spatial coupled, with a high dimension. To achieve recognition of in-hand manipulations, representation learning is important in recognition tasks. Statistical modelling is a bunch of powerful tools for representation learning, the models most frequently applied are known as GMMs. Considering the features of human in-hand manipulation data, we introduce GMMs to model the data. Calinon & Billard [14]



(a) (b)

Fig. 1. Cyberglove and its sersor positions

proposed an approach to teach incrementally human gestures to a humanoid robot, in which the motion data were projected to another space and encoded into a GMM. Kalgaonkar & Raj [15] successfully identified 100 hand gestures based on ultrasound data with a high accuracy using a GMM.

Incremental parameter estimation of GMMs has already been achieved with various methods, most of them assume that novel data arrives in blocks as opposed to a single datum at a time. Hall et al. [16] merged Gaussian components in a pair-wise manner by considering volumes of the corresponding hyperellipsoids. Song and Wang [17] who used the W statistic and the Hotelling's T2 statistic for judging the equivalence of each Gaussian component before merging. However, they do not fully exploit the available probabilistic information. EM-inspired online parameter estimation method will try to explain it in the context of other novel data, affecting the accuracy of the fitting [18].

3 THE INCREMENTAL LEARNING FRAMEWORK FOR GAUSSIAN MIXTURE MODEL

For a compact motion data representation, we resized the time-domain signals of all channels collected by the Cyberglove to be the same length, as shown in Fig. 2, each resized signal $\{\eta_t\}_{t=1}^{T}$ includes the bending information in the joint space. Firstly, for each in-hand motion, we used the human motion data to build the initial dataset D_{ini} for batch learning the parameters of a 2-dimensional GMM with 5 Gaussian components. Assuming there are m independent trajectories form different channels in one dataset, which is denoted $D_{ini} = \{d_i\}_{i=1}^{m}$ and for each $d_i = \{(t, \eta_t)\}_{t=1}^{T}$, where the subscript m is the index of channels and t represents the time step. For a concise representation in the following section, one sample used for training GMM can be denote as $s_t = (t, \eta_t)$. Then, more data from human performance will be collected and processed in the same form, denoted as D_{new}, these newly arrived data will be used for online updating the historical GMMs.

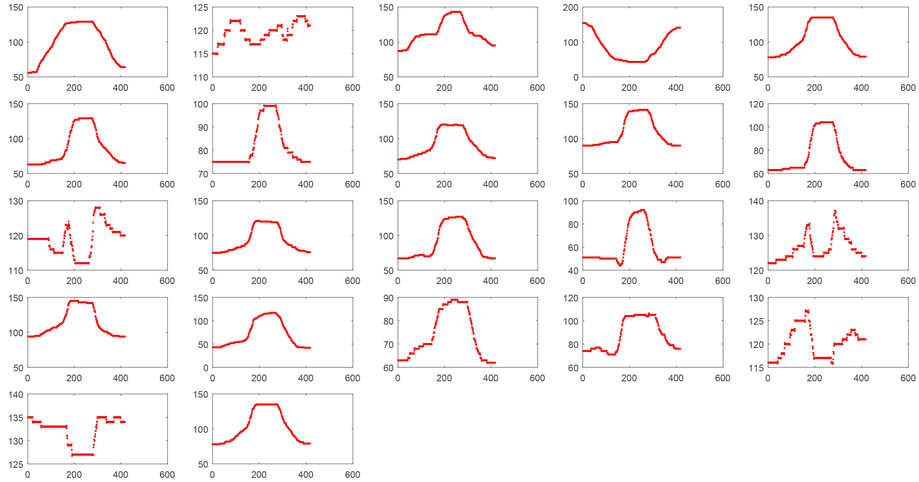


Fig. 2. Resized trajectories from human motion data

3.1 Batch Learning of GMM Using Standard EM Algorithm

The standard EM algorithm was applied to estimate the parameters of each GMM g_m with samples generated by one single channel m. The probability density function (PDF) is,

$$g_m(s_t) = \sum_{k=1}^{K} P(k)p(s_t|k) = \sum_{k=1}^{K} \pi_k \mathcal{N}(s_t|\mu_k, \Sigma_k) \tag{1}$$

where $P(k)$ is the prior of each Gaussian component, their value π_k are also called mixing coefficients, s_t is a data point in the trajectory d_i, $p(s_t|k)$ is the conditional PDF of a sample s_t, given k, $\mathcal{N}(\cdot)$ means a 2-dimentional normal distribution in this problem. All the parameters of the model g_m are $\Theta = \{\pi_k, \mu_k, \Sigma_k\}_{k=1}^{K}$, which were optimized by E-step and M-step in standard EM algorithm according to maximum Likelihood Estimation (MLE). In E-step, we need to calculate the component likelihoods $p(k|s_t)$,

$$p(k|s_t, \Theta) = \frac{p(s_t, k|\Theta)}{\sum_j^K p(s_t, j|\Theta)} \tag{2}$$

and

$$p(s_t, k|\Theta) = p(s_t|k, \Theta) \cdot p(k|\Theta) = \mathcal{N}(s_t|\mu_k, \Sigma_k) \cdot \pi_k \tag{3}$$

where, the parameters, μ_k, Σ_k and π_k, are either initialized by human or from M-step during last iteration. The EM algorithm stopped when a local optimization reached, the optimal parameter set is Θ_{ini}.

3.2 Incremental Parameter Learning of GMMs

The newly coming data, denoted as D_{new}, usually include new information of the motion. To enroll those new in formation into the GMM, standard EM algorithm needs to explore both the historical data D_{ini} and the new data D_{new} during iteration, this is neither convenient nor practicable. Incremental learning is a considerable way to bridge this gap, it can update the historical parameter set Θ_{ini} without visiting the old dataset D_{ini} only used the data in D_{new}.

Considering that, after training a GMM, the old sample set D_0 is no longer accessible, only the parameters of a GMM are stored in the memory. Then, human continue to perform an action for several times, in each rollout, the trajectory $d_i^{new} \in D_{new}$ is recorded. To incrementally update the GMM, we introduced an assumption, the component likelihood $p(k|s_t)$ is almost do not change when only one new sample s_t^{new} arrives, so we modified the M-step in EM algorithm [19]. The parameter set Θ_{new} can be updated with component likelihoods and Θ_{ini}. For a clear expression, the old component likelihood denoted as $p(k|s_t) = \Phi_k$, the Θ_{new} is updated by,

$$\pi_k^{new} = \frac{\Phi_k + p(k|s_t^{new})}{N_s + 1} \tag{4}$$

$$\mu_k^{new} = \frac{\mu_i \Phi_k + s_t^{new} p(k|s_t^{new})}{\Phi_k + p(k|s_t^{new})} \tag{5}$$

$$\Sigma_k^{new} = \frac{(\Sigma_k + \mu_k \mu_k^\top - \mu_k \mu_k^{new\top} - \mu_k^{new} \mu_k^\top + \mu_k^{new} \mu_k^{new\top})\Phi_k + (s_t^{new} - \mu_k^{new})(s_t^{new} - \mu_k^{new})^\top p(k|s_t^{new})}{\Phi_k + p(k|s_t^{new})} \tag{6}$$

where N_s is the number of total samples s_t used for learning the historical parameter set Θ_{ini}. When the incremental learning process is finished the up-to-date model can be updated again when new data arrives.

3.3 In-Hand Manipulation Recognition

Manipulation recognition in our experiment settings can be regarded as a classification problem We assume there are L different manipulation motion, for each motion $\mathcal{G}_l$, we trained a set of GMMs to represent trajectories from all the channels, which denoted as $\mathcal{G}_l = \{g_1^l, g_2^l, \ldots, g_M^l\}$. After collecting the trajectories of human in-hand motions, the newly collected data D_{new} will firstly serve as testing set, which also consists of M independent trajectories from different channels on CyberGlove. GMM is a kind of typical probability model, according to the PDF of a GMM, see Eq. (1), the probability of a candidate trajectory $D_{new} = \{d_m\}_{m=1}^{M}$ generated by the GMM $\mathcal{G}_l$ can be calculated. Therefore, we defined a novel metric, based on log-likelihood, to indicate the possibility $\mathcal{H}(D_{new}|\mathcal{G}_l)$ of a trajectory belonging to any set of GMM $\mathcal{G}_l$,

$$\mathcal{H}(D_{new}|\mathcal{G}_l) = \sum_{m=1}^{M} \left\{ \sum_{t=1}^{T} log[g_m^l(s_t^{new})] \right\} \tag{7}$$

where the formula in the outside summation indicates the log-likelihood of the candidate trajectory [20], while $\mathcal{H}(D_{new}|\mathcal{G}_l)$ is no longer a probability but it can be used to evaluate the possibilities. The following equation shows how to classify D_{new} into a class $\mathcal{G}_l$ belongs to.

$$l = arg \max_{l \in L} \mathcal{H}(D_{new}|\mathcal{G}_l) \tag{8}$$

The results from the above equation will be stored and to guide the online update process.

4 EXPERIMENTS AND DISCUSSIONS

4.1 Experiment Settings

To evaluate the performance of our proposed framework, we defined 10 different in-hand manipulation skills for human to perform, as shown in Fig. 3. Human candidates were asked to finish those actions wearing a CyberGlove, For example, at the batch learning stage, each action was repeated 10 times by human to get enough data for training the initial GMMs.

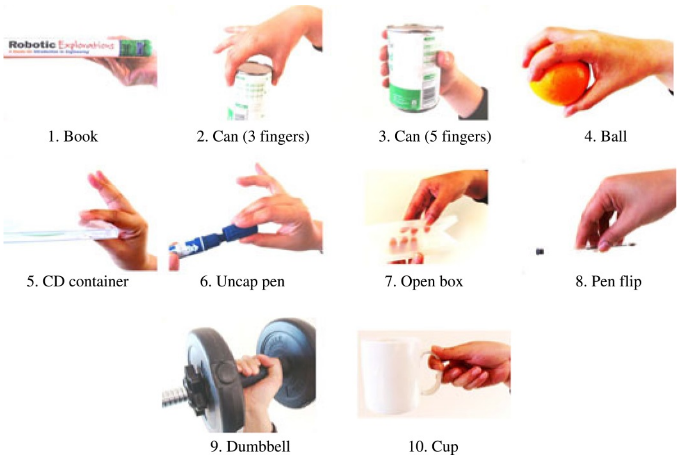

1. Book
2. Can (3 fingers)
3. Can (5 fingers)
4. Ball
5. CD container
6. Uncap pen
7. Open box
8. Pen flip
9. Dumbbell
10. Cup

Fig. 3. Pre-defined 10 different in-hand manipulation skills

In our proposed online learning and recognition framework, as shown in Fig. 4, the learning process can be divided into two stages, first one is batch learning and the second one is incremental learning. In GMM batch learning phrase, 10 set of trajectory samples of human performance D_{ini} are utilised to train an initial GMM. For a fast training, the standard EM algorithm was set to converge when the increment of the log-likelihood is less than 10^{-3}. The training result of a initial GMM is visualized in Fig. 5(a),in which the trajectories consisting

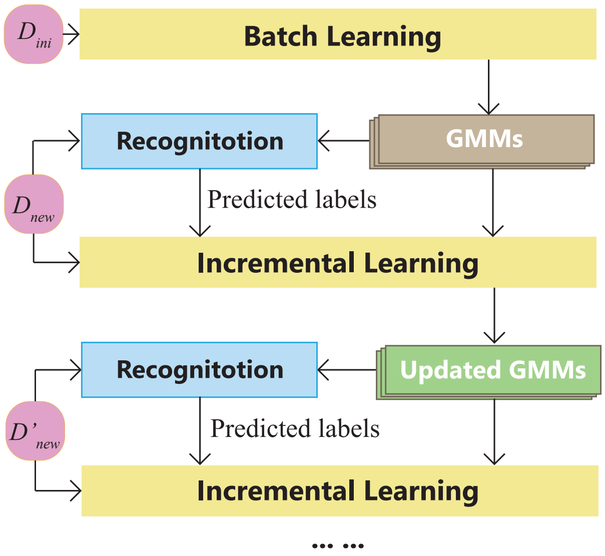


Fig. 4. The overview of online learning framework

of many blue dots are the training samples and the 5 red ellipses visualised the Gaussian functions. After training the parameters of GMMs Θ, the component likelihoods $p(k|s_t)$ and the number of the samples used for training N_s will be stored for recognition the motions in newly arrived human motion data D_{new} and online parameter updating. In incremental learning phrase, the stored parameters will be updated using the algorithm proposed in 3.2. Meanwhile, the predicted labels from recognition will guide the data to feed into corresponding GMM set $\mathcal{G}_l$. By this mean, a GMM was updated as shown in Fig. 5(b), where the blue trajectory is the newly arrived human in-hand motion data collected by CyberGlove during human demonstration, the red ellipses are the Gaussian components in old GMM and the blue ellipses indicate the updated ones using proposed incremental learning algorithm.

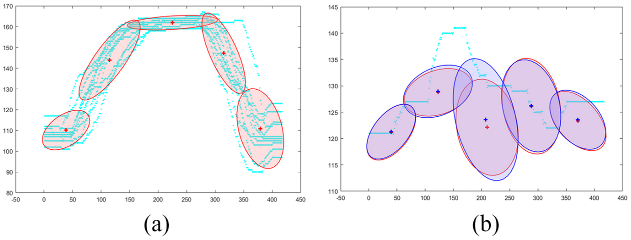


Fig. 5. Learning results of a GMM

The recognition tests conducted after each time the parameters of GMMs were updated (see Fig.4). However, except the batch learning, no ground truth labels of he data were provided during online learning for the algorithm to update the correct GMMs with the corresponding sample. This experimental setting helps to simulate the long-term learning process in real world application. The recognition accuracy witnesses a steady increasement as shown in Fig. 5, which proved the proposed incremental learning framework can achieve a good result in a online recognition task.

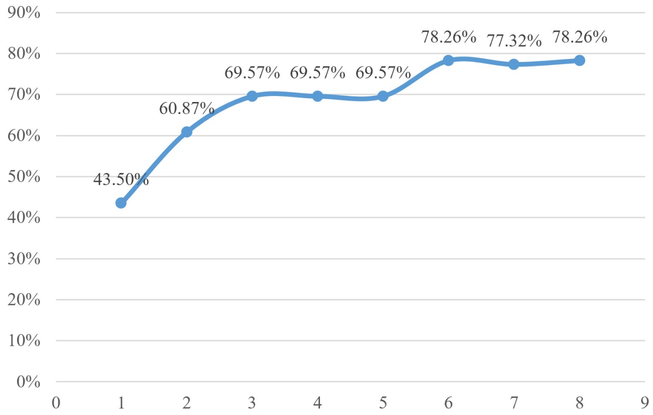


Fig. 6. Learning results of a GMM

5 CONCLUSION AND FUTURE WORK

This paper introduced a novel incremental learning framework for in-hand manipulation skill recognition. We applied GMM for representing the complex human in-hand motion data, the experimental results show the incremental learning method can improve the quality of human in-hand motion representation and optimize the recognition accuracy with the growth of the model updating times. In future work, a) the GMM can be modified to be more adaptive to various data structures, this require the learning algorithm being able to adjust the number of Gaussian components during learning process. b) In current framework, online learning needs to count the samples used for training GMM and save the component likelihoods as the extra parameters, this is not suitable for long-term online learning, in the future we are going to optimize the framework to achieve the same functions with less parameters to be stored. c) Motion recognition is a small step in human in-hand manipulation motion studies, more work can be done based on current framework, for example, applied regression method to generalize trajectories and transfer human skill to robots.

References

1. Yousef, H., Boukallel, M., Althoefer, K.: Tactile sensing for dexterous in-hand manipulation in robotics—A review. Sensors and Actuators A: physical, **167**(2), 171-187 (2011).
2. Ju, Z., Liu, H.: Human hand motion analysis with multisensory information. IEEE/ASME Transactions on Mechatronics, **19**(2), 456-466 (2014).
3. Chen, D., Li, G., Sun, Y., Kong, J., Jiang, G., Tang, H., ... Liu, H.: An interactive image segmentation method in hand gesture recognition. Sensors, **17**(2), 253(2017).
4. Zhang, Y., Chen, L., Ran, X.: Online incremental em training of gmm and its application to speech processing applications. In Signal Processing (ICSP), 2010 IEEE 10th International Conference on Proceeding, pp. 1309-1312. IEEE (2010).
5. Sudsang, A., Srinivasa, N.: Grasping and in-hand manipulation: Geometry and algorithms. Algorithmica, **26**(3-4), 466-493 (2000).
6. Kondo, M., Ueda, J. Ogasawara, T.: Recognition of in-hand manipulation using contact state transition for multifingered robot hand control. Robotics and Autonomous Systems, 56(1), 66-81 (2008).
7. Kumar, P., Verma, J., Prasad, S.: Hand data glove: a wearable real-time device for human-computer interaction. International Journal of Advanced Science and Technology, 43 (2012).
8. Bendels, G. H., Kahlesz, F., Klein, R.: Towards the next generation of 3D content creation. In Proceedings of the working conference on Advanced visual interfaces ,pp. 283-289. ACM (2004).
9. de La Gorce, M., Paragios, N., Fleet, D. J.: Model-based hand tracking with texture, shading and self-occlusions. In Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference On Proceeding, pp. 1-8. IEEE (2008).
10. Fukuda, O., Tsuji, T., Kaneko, M., Otsuka, A.: A human-assisting manipulator teleoperated by EMG signals and arm motions. IEEE Transactions on Robotics and Automation, **19**(2), 210-222 (2003).
11. Reddy, N. P., Gupta, V.: Toward direct biocontrol using surface EMG signals: Control of finger and wrist joint models. Medical engineering & physics, **29**(3), 398-403 (2007).
12. Kuroda, T., Tabata, Y., Goto, A., Ikuta, H., Murakami, M.: Consumer price data-glove for sign language recognition. In Proc. of 5th Intl Conf. Disability, Virtual Reality Assoc. Tech., Oxford, UK, pp. 253-258 (2004).
13. Fahn, C. S., Sun, H.: Development of a data glove with reducing sensors based on magnetic induction. IEEE Transactions on Industrial Electronics, **52**(2), 585-594 (2005).
14. Calinon, S., Guenter, F., Billard, A.: On learning, representing, and generalizing a task in a humanoid robot. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), **37**(2), 286-298 (2007).
15. Kalgaonkar, K., Raj, B.: One-handed gesture recognition using ultrasonic Doppler sonar. In Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on Proceeding, pp. 1889-1892. IEEE (2009).
16. Hall, P., Marshall, D., Martin, R.: Merging and splitting eigenspace models. IEEE Transactions on pattern analysis and machine intelligence, **22**(9), 1042-1049 (2000).
17. Song, M., Wang, H.: Highly efficient incremental estimation of Gaussian mixture models for online data stream clustering. In Intelligent Computing: Theory and Applications III, Vol. 5803, pp. 174-184. International Society for Optics and Photonics (2005).

18. Hicks, Y., Hall, P. M., Marshall, D.: A method to add Hidden Markov Models with application to learning articulated motion. In BMVC, pp. 1-10 (2003).
19. Zhang, Y., Chen, L., Ran, X.: Online incremental em training of gmm and its application to speech processing applications. In Signal Processing (ICSP), 2010 IEEE 10th International Conference on Proceeding, pp. 1309-1312. IEEE (2010).
20. Ju, Z., Liu, H.: A unified fuzzy framework for human-hand motion recognition. IEEE Transactions on Fuzzy Systems, **19**(5), 901-913 (2011).