Action Gist Based Automatic Segmentation for Periodic In-hand Manipulation Movement Learning

Gang Cheng, Norman Hendrich, Jianwei Zhang

***Abstract*— We consider in-hand manipulation tasks that consists of periodic movements. In order to improve the manipulation learning ability of a robot with a human-like hand, this paper introduces a segmentation method based on the techniques of action gist. Action gist is the key motion information in manipulation with the property of semantics. In the techniques of in-hand manipulation action gist, there is a Meta Motion Occurrence Histogram describing the motion information in the demonstration set. This paper proposes an algorithm related to the Meta Motion Occurrence Histogram to maximize the common motions in each segment, so as to figure out the best segmentation solution in the in-hand manipulation sequence. The experiments illustrate the performance of the proposed method, and discuss the possibility of segmentation fusing with the information from tactile sensor.**

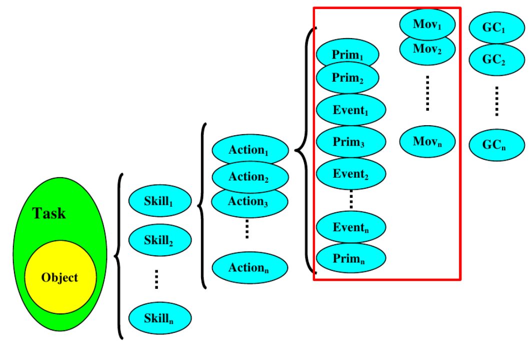

Fig. 1. A hierarchical structure of the in-hand manipulation [1]. This paper concentrates on the primitive-movement level, to analyze the periodic variation of the movements.

I. INTRODUCTION

Human in-hand manipulation involves the synergy of five fingers and the palm, the feedback produced from visual and haptic neurons, and the central processing unit – the brain. When we assign a manipulation job to a robot, especially one with a human-like hand, it is natural to regard this procedure as the knowledge transition from human to robot. Compared with designing a specific solution for the robotic hand for every single application, it is easier to supervise the humanoid hand by *learning from human demonstration* because in this case we can find teachers all over the world.

In the process of learning in-hand manipulation skills from humans, a necessary step is to divide a long in-hand manipulating sequence into several short segments. One reason for this is that for complicated hand movements rooted in multiple joints and a high degree of freedom, it is inconvenient to describe and memorize the corresponding kinematic form. Another reason is that as the aim of this kind of manipulation is to transform the state of the in-hand object, it is easier to solve several subtasks step by step than to tackle one big problem as long as we create a set of proper mileposts. This hierarchical idea is well adopted by several European projects [1] as Fig.1.

To understand and extract the hand movement patterns, we need a set of sensors to perceive the state of the hand and the object, as well as the contact information between them [2]. Besides multiple sensors, we should also be clear about the specific contextual information according to the scenario [3] in order to take the correct actions. When we have enough data, it is time to analyze the entire manipulation process. Rather than an analysis of simple manipulation tasks [4] [5], we concentrate more on the detailed finger gaiting application. In this field, a slight variation of a finger joint can be a key point in the process of manipulation. Since there are many joints in the hand, it is necessary to find an effective tool to represent variation and perform segmentation.

Specifically to deal with periodic in-hand manipulation movements (e.g. rotating a screwdriver, turning pages for several times in Section VI), this paper proposes a kind of segmentation method based on the techniques in *action gist*, which describes the information of key motions widely adapting to different-sized hands. The method segments the time-series manipulation process based on the similarity of each segment.

The structure of this paper is the following: After presenting the state-of-the-art and related work, we give a short introduction to the techniques of in-hand manipulation action gist. Then a segmentation algorithm based on a tool in the techniques is given in Section IV. After that we discuss the possibility of fusing the segmentation result with tactile information in Section V. Section VI illustrates the segmentation results by comparing the results from manual partition, and discusses how to integrate the segmentations in multiple sensors to improve in-hand manipulation learning. And the final part gives the conclusion and future work in Section VII.

II. RELATED WORK

Since the process of in-hand manipulation contains a series of hand movements, it is common to segment the entire process into several smaller parts. In this way, the manipulation process becomes easily understandable so that we can concentrate on abstracting the interesting information in the segments. One kind of segmentation method depends on the hand gesture, which is based on the fact that the whole manipulation process can be understood as the translation of several significant grasping gestures. Several works of research such as [6] have succeeded in reducing the scale of the realistic human hand gestures using principal components analysis (PCA) and discriminant functions. In this case, it is possible to use finite key hand poses to represent the entire manipulation process and guide the robotic hand in task execution.

Meanwhile it is possible to segment the value sequence according to the joint angle local minima or maxima, and then concentrate on the local extremes to study the periodicity [7]. However, as in-hand manipulation consists of the synergies of many joints, it is difficult to segment the movement only at this level.

More information can definitely give us more help, such as considering the hand and the object posture together as [8] to instruct the regrasping movement. This work does not concern the whole hand postures but only the area where the object and the hand interact. Besids taking the object into consideration, another solution is to add sensors and understanding the manipulation in multiple channels [2], [4], [5]. Force sensing is also an important criterion of the manipulation state transition [9]. Without a sense of force feedback, humans are unable improve their manipulation skills. [10] segment the manipulation process as the contact region and the measured force on a specifically designed pencil.

When we consider the in-hand manipulation segmentation as a motion segmentation, we can learn more from similar topics. Basically, motion segmentation methods can be classified as online and offline. The online method can be like [11], it can yield segmentation feedback to improve the robot real-time reaction. But here we aim at analyzing the human demonstration, so we have enough time to process the acquired data. Thus we are more interested in the offline ideas.

For a time-series motion segmentation, we can consider it as a kind of clustering. Each segment is separated as the local relations of the elements. [12] discussed Principle Component Analysis (PCA), Probabilistic PCA and the Gaussian mixture model in high level motion segmentation through the body joint angle variation. And then [13] continued their work, clustered human motions based on k-means [14], and refined the classification by a global minimization algorithm. For the segmentation, key information is extracted as the criterion to divide the time series. Along with the clustering idea, we even can apply a general clustering model for time series data as lately proposed by [15].

Anyway, we need to define the segmentational feature as the criterion to maintain the segment quality, such as rotation-invariant features [16]. We believe that in our in-hand manipulation case, we can find more specific semantic features.

For the hand movement recognition, [17] imported Empirical Copula to accurately detect the scenario. Besides, [18] imported a Fuzzy Active curve axis Gaussian Mixture Model (FAcaGMM) to detect the scenario fast. Based on the data-glove value, these methods can analyze what kind of hand movement is performed even if the training set includes only a few samples. However, it is not clear whether both methods can automatically segment the long manipulation process including multiple specific operations.

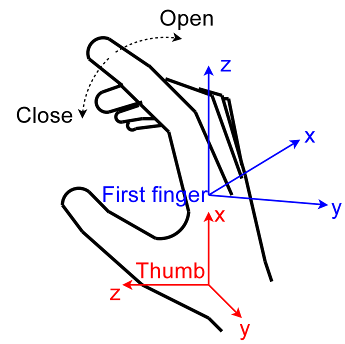

Fig. 2. The finger coordinates with the palm. The thumb in the red coordinate differs from the other four fingers due to its special location in the hand.

[19] proposed a segmentation method independent from prior knowledge on motion characteristics, and it is very effective in one-dimensional cases. But because of the recursive estimation algorithm they applied, the accuracy runs low with increasing number of joints.

Different from the classical segmentation criteria, our method concentrates mainly on the hand movement itself. It is a kind of semantic analysis based on the similar motions which periodic manipulation presents. We use the data-glove to generalize the in-hand manipulation action gist with respect to the application, and based on this kind of semantic information we complete the segmentation.

III. IN-HAND MANIPULATION ACTION GIST

In-hand manipulation action gist is a kinetic concept which represents the key finger motions in a manipulation task and widely adapts to different hands. Take a simple grasping example: In most cases, we "close" our fingers but do not "open" them to grasp the object. That means, we use a kind of action gist to manipulate the object, and another person who has a different-sized hand can also apply the same action gist and successfully manipulate it. With in-hand manipulation action gist we can learn from the demonstration and record concise information, then instruct the robotic hand executing the movement by the gist guidelines.

Meta motion is the basic unit in the in-hand manipulation action gist. It is related to the kinetic direction on each finger including closing, opening and a projected direction relevant to the palm. As Fig.2 and Fig.3 show, every finger has 8 kinetic meta motions and 1 idle meta motion. By Gaussian Markov Random Field presenting the data relation in different frame distances, the finger joint angles (e.g. as measured by a data-glove) are translated into a meta motion sequence.

Generally speaking, experienced humans seldom repeat the same meta motion in a short manipulation episode. They will try their best to avoid any collision and approach the target with as few meta motions as possible.

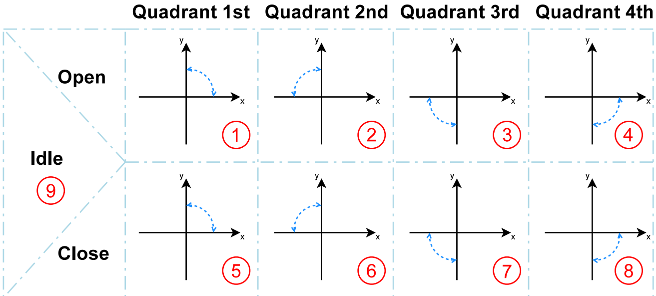

Fig. 3. Nine types of meta motion in each finger. Two flex/ext-joints are modeled as one parameter as open or close, and the abduction angle variation cooperates with the metacarpal-proximal angle variation to form a 2D projected direction for the finger's movements. The idle motion is specifically set apart and labeled as 9.

When we have multiple trials of a specific manipulation application, we can generalize the action gist by evaluating the popularity of each sample through a Meta Motion Occurrence Histogram (mostly abbreviated to "Histogram" in the remaining part of this paper). The Histogram is a statistical matrix of the demonstration set, it describes the meta motion occurrence frequency according to the position but not relevant to the duration in the real trial. For example, when the first finger performs meta motion 7, 3, 4 in 3 seconds, 10 seconds and 1 second respectively, the Histogram will sequentially record meta motion 7, 3, 4.

The form of the Meta Motion Occurrence Histogram is described as follows:

$$H(a,r,l) = \sum_{\eta(m_i^s,a,r,l)=1} G(\psi(m_i^s),a,\sigma_s) \tag{1}$$

where meta motion m_i^s is the i-th element from action gist $\mathbf{m}^s$ in the demonstration set $\mathbf{M}$. $\eta(\cdot) = 1$ if and only if m_i^s belongs to finger r, labeled as meta motion l, and position a locates near m_i^s but no other motion on finger r. $\psi(m_i^s) \in [0,1]$ indicates the normalized order position when the meta motion begins, so position a is near position $\psi(\cdot)$. $G(t_1,t_2,\sigma_s) = \frac{1}{\sigma_s\sqrt{2\pi}} e^{-\frac{(t_1-t_2)^2}{2\sigma_s^2}}$ is the typical Gaussian distribution form, and σ_s is a parameter that controls the impact factor reduction, it is set as reciprocal to the length of sequence $\mathbf{m}^s$. Considering this is a discrete numeric processing, the histogram has a resolution. The normalized a will finally be scaled as an integer form during calculation.

The frequent possible meta motion takes a higher value in the element of the Histogram. As a result, for every action gist in the demonstration set, there is a simple way to evaluate its popularity as in the following equation:

$$Score(\mathbf{m}^s) = \sum_i H(\psi(m_i^s), \tau_{finger}(m_i^s), \tau_{label}(m_i^s)) \tag{2}$$

where $\tau_{finger}(\cdot)$ indicates the meta motion belongs to, and $\tau_{label}(\cdot)$ indicates the meta motion type.

One kind of action gist can adapt to different kinds of hands with one manipulation task. However, one manipulation task can be done by countless correct methods. Fortunately and specifically for one person, it is common that the movement to solve one manipulation task is more or less similar.

IV. PERIODIC IN-HAND MANIPULATION MOVEMENT SEGMENTATION

In the real world the paradigms of periodic manipulation can be like rotating a key, turning pages, or other movements operating repeatedly. In the process of *Learning from Demonstration*, we can decompose the entire continuous movement into several parts, with each part being a loop. In this way, we only need to show the robot the complete part dozens of times, and the robot will extract the necessary information for future practice. The problem of this method is that we have to cut the movement by subjective judgment, as the connective information between adjacent movements may be missing. Thus if the demonstrator has the chance to perform periodic movement without interruption, it is a more natural way to acquire knowledge about the skill.

We intend to make use of the techniques of in-hand manipulation action gist to automatically segment the entire manipulation process. Because the meta motion semantically reflects the finger movement, the reduplicative motion patterns can more or less imply the periodic information.

Firstly we give the definition of the segment in a trial of periodic in-hand manipulation movements.

Definition 1: If in a segment $\mathbf{m}^s$ in the periodic in-hand manipulation sequence, the same-labeled meta motions ω_i are located before ω_j in the entire sequence and they belong to the same finger, then $\mathbf{m}^s$ starts from the beginning position of ω_i and end before the beginning position of ω_j. In this case, the next segment $\mathbf{m}^{s+1}$ starts from the beginning position of ω_j if the same-labeled and same-fingered meta motion ω_k exists behind ω_j.

In this case, every segmentation must begin with the meta motion of equal type. And then the segment boundary selection speeds up. The assignment as Def.1 may result in mistakes, but the majority should share the same sequence head regarding the statistical point of view. Therefore, as long as we have done enough periodic manipulation demonstration, it is possible to extract the key information in the periodic movement.

Secondly we give the criterion of a good segmentation.

Definition 2: $|\mathbf{m}^{s_1} \bigcap \mathbf{m}^{s_2}|$ represents the quantity of common same-fingered meta motions at similar positions in both segments $\mathbf{m}^{s_1}$ and $\mathbf{m}^{s_2}$.

For a good segmentation, we expect $|\mathbf{m}^{s_1} \bigcap \mathbf{m}^{s_2}|$ to be as high as possible. Here the Meta Motion Occurrence Histogram is able to generalize the segmentational result and provide us with the evaluation of the periodicity of the segmentation.

Definition 3: To a 3 dimensional Meta Motion Occurrence Histogram $\mathbf{H}$, the corresponding Frobenius norm is defined as

$$\|\mathbf{H}\|_F = \sqrt{\sum_a \sum_r \sum_l |H(a,r,l)|^2} \tag{3}$$

Theorem 1: $\max \sum_{s_1} \sum_{s_2} |\mathbf{m}^{s_1} \bigcap \mathbf{m}^{s_2}| \Leftrightarrow \max \|\mathbf{H}\|_F$

Proof: According to Eq.1,

$$\begin{aligned} & \max \|\mathbf{H}\|_F \\ \Leftrightarrow \quad & \max \sum_a \sum_r \sum_l \left(\sum_\eta G\left(\psi(m_i^s), a, \sigma_s\right) \right)^2 \end{aligned}$$

Here we can see that, once m_i^s exists, it will contribute to several elements in $\mathbf{H}$. In the meantime, the constraint $\eta(m_i^s, a, r, l) = 1$ controls the meta motion number of the contribution. $m_i^s \in \mathbf{m}^s$, and $\mathbf{m} \subset \mathbf{M}$. According to Def.1, unless m_i^s impacts $H(a, r, l)$ alone, it has to increase the sum of $H(a, r, l)$ with other meta motions.

Obviously, when we neglect the specific a, r and l,

$$\begin{aligned} & \left(\sum_\eta G\left(\psi(m_i^s), a, \sigma_s\right) \right)^2 \\ \doteq \quad & \left(\sum_\iota G_\iota\right)^2 \\ \doteq \quad & \left(\sum_i G_i + \sum_j G_j \right)^2 \\ > \quad & \left(\sum_i G_i\right)^2 + \left(\sum_j G_j \right)^2 \end{aligned}$$

Supposed in this case $\mathbf{i}$ meta motions remain a contribution to the specific element of the Histogram, but $\mathbf{j}$ meta motions work alone. The part $\left(\sum_i G_i\right)^2$ is considered as the new sums of the original element. The other part $\left(\sum_j G_j \right)^2$ is the new sums of where the meta motions jump to. So it means the original form takes a higher sum.

For the cases where $\mathbf{i}$ meta motions remain but $\mathbf{j}$ motions jump to cooperate with another element, there is no rule to judge which form is better. Nevertheless, it is not vital that more than one meta motion fit the constraint $\eta(\cdot) = 1$. All of these cases will be taken into the competition. We will select the winner with the highest sum.

Then we can say the more meta motions join to work together, the higher the total is. On the other side of the theorem,

$$\begin{aligned} & \max \sum_{s_1} \sum_{s_2} |\mathbf{m}^{s_1} \bigcap \mathbf{m}^{s_2}| \\ \Leftrightarrow \quad & \text{More } \mathbf{m}^s \text{ fits the constraint } \eta(\cdot) \end{aligned}$$

Therefore,

$$\begin{aligned} & \max \|\mathbf{H}\|_F \\ \Leftrightarrow \quad & \max \sum_{s_1} \sum_{s_2} |\mathbf{m}^{s_1} \bigcap \mathbf{m}^{s_2}| \end{aligned}$$

■

So when the segmentation generates a corresponding meta motion sequence set $\{\mathbf{m}^s\}$, we can justify whether this is the best segmentation by examining the corresponding $\|\mathbf{H}\|_F$. Based on this theorem, we propose an algorithm to automatically segment the periodic movement in an in-hand manipulation meta motion sequence as Algorithm 1. The current algorithm is a linear enumerating method to segment the sequence, where each segment begins with the same-labeled meta motion. Later it will be improved as an iterative or head-independent algorithm after we have enough criteria to prove it.

In this way, we can naturally present a repeated in-hand demonstration to the robot. During the process of analysis, the motion sequence is segmented as the proposed algorithm, and evaluated by Eq.2. Meanwhile in practical processing, there is the risk of errors or mistaken movements being included in the coherent movement. But as long as the positive data is the majority, we can refer to the evaluation from Eq.2 and believe that the segmented movement with a high score is acceptable. The one having a higher score indicates its popularity in the periodic movement, so we can use it to reproduce the manipulation.

Algorithm 1 Segment the periodic movement of an in-hand manipulation demonstration with the techniques in action gist

Require:
The extracted meta motion sequence $\mathbf{M}$

Find the same meta motion l on each finger r, store their starting positions as $\mathbf{P}_{l,r} = \left\{ P_{l,r}^1, P_{l,r}^2, P_{l,r}^3, \cdots \right\}$;
$Score_{max} \leftarrow 0$, segmentation solution $\mathbf{Z} \leftarrow \{\}$;
for all $\mathbf{P}_{l,r} \neq \{\}$ **do**
  $\mathbf{m}^s \leftarrow$ The meta motion sequence ranging at the positions of $\left[P_{l,r}^s, P_{l,r}^{s+1} - 1 \right]$;
  Demonstration set $\mathbf{M}_{tmp} \leftarrow \{\mathbf{m}^s\}$;
  Calculate the Histogram $\mathbf{H}$ of $\mathbf{M}_{tmp}$;
  if $\|\mathbf{H}\|_F > Score_{max}$ **then**
    $Score_{max} \leftarrow \|\mathbf{H}\|_F$, $\mathbf{Z} \leftarrow \mathbf{P}_{l,r}$;
  end if
end for
return $\mathbf{Z}$;

V. PERIODIC MOVEMENT SEGMENTATION FUSION WITH TACTILE SENSOR

Generally speaking, with more sensors the segmentation will become more accurate. Meanwhile, the segmentation process is equal to design a goal function and then to optimize it. However, more sensors make the decision complex, we need to weigh and consider balance among many choices. So far, the common form of the fusion obeys the rule like following equation [20][21][22]:

$$P_{fusion} = \frac{\sum_i c_i P_i}{\sum_i c_i} \tag{4}$$

where c_i is the weight of the sensory segmentational component P_i. Regarding this equation we can imagine that the number of components may increase the uncertainty of segmentation. Therefore, we only address how to deal with gloved and tactile data without weighting in this paper.

Because tactile information is an important criterion in hand manipulation, we intend to integrate our segmentation method with tactile perception. Considering Fig.8 from Section VI-B we believe that tactile information also can be refined as periodic criteria. And it is possible to apply the same method to the tactile state sequence as the techniques of manipulation action gist. But according to our current experimental experience, the tactile segmentation is not as reliable as the meta motion segmentation. The reason is related to the different sensitivities of tactile cells, and the complexity of the tactile sensory structure. And we have to point out, that the meta motion is the direction of the finger, but the contact always changes as the finger touches / leaves the object. That means the segments begin and end based on different mechanism, we can not use interpolation between the boundaries of two kinds of segments.

Therefore, we consider the tactile information as a support to the current segmentation method. The workflow is as Algorithm 2.

Algorithm 2 Segment the periodic movement of an in-hand manipulation demonstration with multiple information

---

Require:
Ranked segmentation solution $\{\mathbf{Z_i}\}$,
so $\|\mathbf{H_i}\|_F > \|\mathbf{H_k}\|_F$, $i < k$
Tactile segmentation solution $\{\mathbf{T_j}\}$

```
1:  S_m ← +∞;
2:  for i = 1 do
3:    for all T_j do
4:      if Z_i and T_j have similar number of the segments then
5:        Sum up the position difference between the nearest segment pair, one in Z_i, and the other in T_j;
6:        if S_m > the calculated sum then
7:          p_m ← j;
8:        end if
9:      end if
10:   end for
11:   if S_m ≠ +∞ then
12:     goto 16;
13:   end if
14:   i ← i + 1
15: end for
16: Update Z_i by T_{p_m} with the closest segment positions, store it in Z_new;
17: return Z_new;
```

We consider Algorithm 2 as a kind of compromise between both kinds of segmentation plans. The reason is that the result calculated by action gist techniques based method is always close to the manual segmentation. We had better to keep the scale and the distances of the segments not far from the original solution.

VI. EXPERIMENT AND DISCUSSION

We employ the data-glove for the input of hand movements. The values from the data-glove are processed as described in Section III into meta motions to semantically present the in-hand movement. In this section, we firstly compare the glove-value segmentation results with manual segmentation results, and then discuss the possibility to fuse the segmentation from data-glove and tactile sensor.

A. Experimental Examination by Multiple Scenarios

We have an integrated system to record the Cyberglove data with synchronized visual data [23]. By this tool, we are able to compare the segmentation from our proposed algorithm with manual results. We have 4 scenarios shown in Fig.4 to examine the performance of our method. The demonstrator performs the experiments and repeatly moves the corresponding object. Each application is demonstrated many times. After that, through maximizing the Frobenius norm of each segmentation, we can have the result like Fig.5.

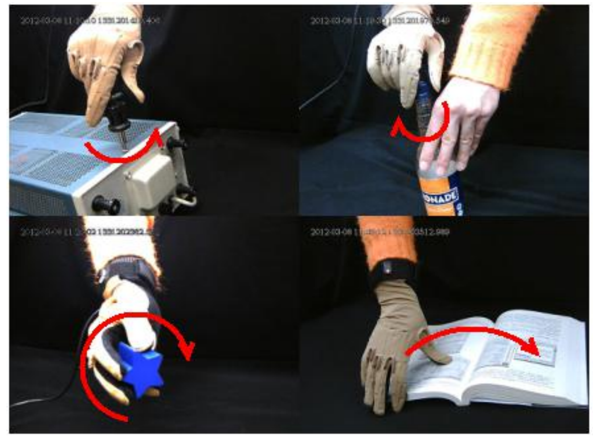

Fig. 4. Four scenarios of periodic movements to examine our proposed method. The first one is to use a screwdriver to fix the screw. The second one is to rotate the cover to open the bottle. The third one is to play a star-like toy. The fourth one is to turn the pages of a book. The red arrows indicate the operating directions of the corresponding objects.

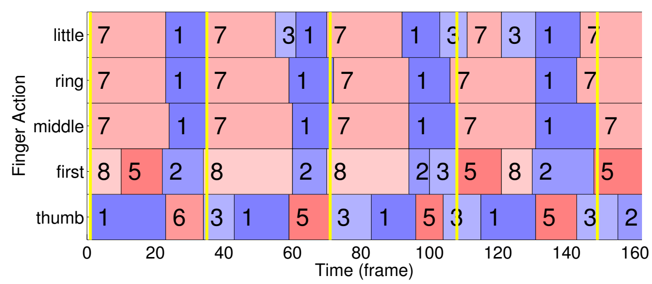

Fig. 5. A segmentation example with the corresponding meta motion sequence from the screwdriver scenario. The figure shows the meta motions of all five fingers, each type of meta motion is represented by color rectangles with the corresponding type number. Specifically, the closing motions employ warm colors but the opening motions apply cool colors at different saturation levels. The x-axis is a time axis indicating the cyberglove frame number. The yellow lines segment the entire sequence into several parts. This example is a segmentation by Meta Motion 7 in the middle finger.

We have recorded synchronized visual data at the frame rate of 30 fps, and the frame rate of our Cyberglove is set as 15 fps. We spend 10 seconds for each demonstration of performing the periodic movements, except 20 seconds for the page turnings. By comparing with the timestamps of the visual sensor and the Cyberglove, we evaluate the performance of the proposed method as Tab.I and Fig.6.

In Tab.I, “Repeat” indicates how many times the demonstrator actually performs the motions. “Miss” indicates the number of the segments that the automatical segmentation fails to find. Moreover, “Exceed” counts the extra segments that the automatical segmentation finds but which are not real in the demonstration. In the table, we can see that the automatical segmentation of the star-like block rotation and the page turning have mistakes. These two tasks are more complicated than the other two, and without training

TABLE I
THE PERFORMANCE OF DATA-GLOVE BASED SEGMENTATION

Scenario	Repeat	Miss	Exceed
screwdriver	4 times	0 times	0 times
	4 times	0 times	0 times
	5 times	0 times	0 times
cover opening	7 times	0 times	0 times
	6 times	0 times	0 times
star rotation	6 times	2 times	0 times
	6 times	0 times	0 times
page turning	5 times	0 times	2 times
	8 times	0 times	0 times
	8 times	0 times	0 times

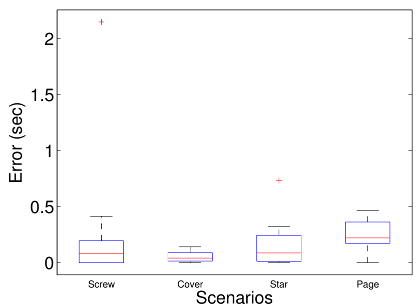

Fig. 6. The errors of the automatical segmentation. Compared with the manual segmentation, we calculate the related errors measured by second. The blue boxes indicate the main variances of the errors. The black boundaries indicate the minimal and maximal errors of the demonstration, and the red lines in the boxes represent the averages. The red crosses indicate the outliers.

the demonstrator uses different movements to achieve the manipulation. But regarding that the trial time increases, the demonstrator becomes experienced and then this case will not easily happen.

Fig.6 indicates the errors of the segmentation positions are mostly under 0.2s. Actually to the normal speed manipulations, with human eyes it is difficult to distinguish the movement difference in this duration level. So we assume the segmentation results are acceptable.

Among the demonstrations mentioned in Tab.I, we have 4 failed segmentations. We investigate the raw data and find two reasons for the failures. One is because the pause between two periodic segmentation is long, meanwhile the fingers look staying idle but actually slightly move. In this case the meta motion parser finds some unexpected motions which disturb the segmentation. The other reason is that the demonstrator applies multiple methods to carry out the manipulation, then the algorithm can not detect the segmentation correctly.

Therefore, it is helpful that more sensors participate in the manipulation analysis.

B. Periodic In-hand Manipulation Segmentation by a Tactile Sensor and the Techniques in Action Gist

In this part we aim at analyzing the manipulation skill from multiple sensors. Even though the proposed algorithm in this paper is based on the information processing of the

Fig. 7. The experimental setup for hand pose and tactile information aquistion [24]. The demonstrator wears the Cyberglove for the hand posture sensing. The front side of the hand equiped with a Tekscan Grip system receives the contact information.

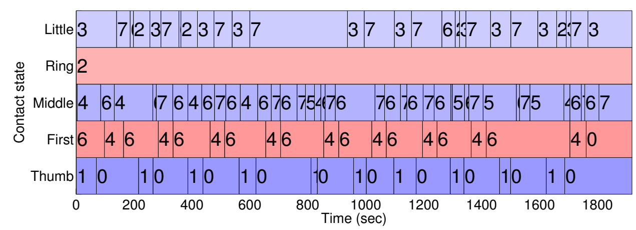

Fig. 8. A segmentation example based on the tactile information. The scenario is star-like toy rotation. It is assumed that the contact region of each joint has a threshold value indicating whether it is touched. We use the 0 to present the non-touched state, and 1 for the touched state. For each finger we sum up the corresponding joint states, and represent it in decimal digits instead of binary form. For example, the thumb in digit 1 indicates that the distal joint is touched, but the proximal joint is non-touched; The first finger in digit 6 indicates that only the corresponding distal joint is non-touched, but medial and proximal joints are touched. We can count that the thumb spends 10 periods in state 1. Furthermore, in this demonstration, the demonstrator does rotate the block 10 times. However, the demonstrator does move the ring finger with touching and non-touching 10 times, but we can not see that the ring finger has any change in the figure. Therefore we currently only use tactile information as assistance and consider the contact state transition as a future research.

data-glove, we can have the experiment carried out using several devices including a stereo camera, a magnetic tracker, Cyberglove, the Tekscan Grip system (for details of the set up please refer to [25] and [24], or the applications [2][5][4][26]). Too many devices installed on the hand restrain the natural movement of the demonstrator, but it provides a chance to compare the segmentation result.

Fig.7 shows the experimental set up enclosing the hand. The image sequence of the manipulating process, the hand joint angle and the tactile information are available and synchronized in this set up. Besides, the movement of rotating a star-like block is a typical periodic in-hand manipulation. We are going to study the block rotation movement by manual, tactile and finger-action-semantic segmentation.

The tactile sensor Tekscan consists of arrays of haptic cells attached to each the finger joint. Each cell contains a value presented by an unsigned byte indicating the intensity of the contact pressure. Basically we can segment the sequence by the different contact area combination. The state definition is similar to [26], but we do not need the palm contact information. One reason is that the palm does not participate in the rotation, another reason is that we find too much noise

TABLE II

THE STAR ROTATION PERFORMANCE OF FUSION BASED SEGMENTATION

Repeat	Glove-based Miss / Exceed	Fusion-based Miss / Exceed
10	0 / 1	0 / 0
10	0 / 1	0 / 0

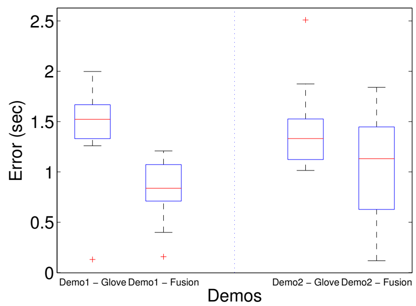

Fig. 9. The errors of glove-based method and fusion-based method comparing with manual segmentation. Because of the experimental set up, and for the demonstrator can not naturally perform the manipulation with wearing too many devices, the errors are higher than Fig.6. But anyway we can see the fusion-based segmentation is better than the single sensor based segmentation.

in this application when the parts in and around the palm rub reciprocally. Meanwhile the tactile information involves many factors, even as [27] indicates, grip force is affected by the hand posture. Therefore, here we only need to consider whether a force is applied to a specific area.

Therefore, we separate the in-hand manipulation state by the contact force variation with respect to each finger joint. Through summing up the intensity of corresponding cells, smoothing the totals, threshold filtering to separate the high and low value, and a series of post processing, an example of the segmentation is shown in Fig.8. Then we can see as each finger holds a different transition form, there are many possibilities in the entire process.

Anyway, we can apply the segmentation method via the data-glove to understand the entire manipulation sequence. And then integrate the tactile segmentation as Algorithm 2. The results comparing with the manual segmentation is shown as Tab.II and Fig.9. Because the hand wears too many sensing devices, the demonstrator can not perform the manipulation naturally. Therefore the average errors are higher than glove-only method.

C. Popularity of the First Meta Motion in the Segment

To manipulate an object, there are countless finger-gaitings. We can get many meta motion sequences from action gist extraction, and every one will work in practice. However, for each particular manipulation task, we would like to find the common action gist. Because we think if one kind of movement is always performed by humans, it will be more stable than other movements applied in the specific scenario. For many trials from the star-like block rotation in Section VI-B, we intend to find the popular head of the segment. Thus we sum the Frobenius norm of the Histograms up and evaluate the popularities. The result is shown by Fig.10. The result indicates that when rotating a block with four fingers, the demonstrator always moves the thumb first. This criterion can be considered as a hint to the segmentation by tactile information.

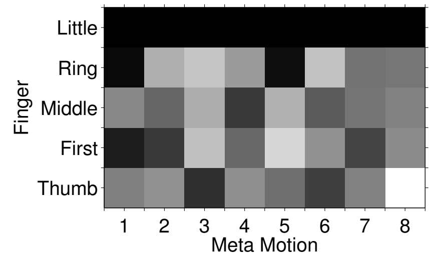

Fig. 10. The possibilities of the start meta motions in VI-B. The block intensity indicates how likely it is for the meta motion according to the finger to become the head of the segment. After several demonstrations of star-like block rotation, the meta motion 8 of the thumb wins the highest score. It implies that the demonstrator may have this behavior in the rotation scenario of a similar object.

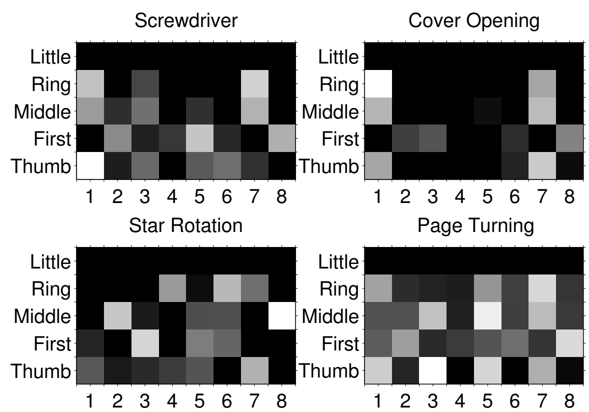

Fig. 11. The possibilities of the start meta motions in Section VI-A. In all scenarios the demonstrator uses his right hand. We can find that the first motion in the "Screwdriver" scenario is using the thumb. Considering with the fact that the screwing demonstrations are anticlockwise, this result is reasonable. And we notice that for the clockwise movement scenario "Cover Opening", the demonstrator likes to move the ring finger first as expectation. The "Star Rotation" result is different from Fig.10 because the demonstrator often uses his middle finger to keep the block at the beginning. However, we think the "Page Turning" is the most interesting one in the cases because we find more bright blocks than other scenarios. We preliminarily think that is because the demonstrator wants to use the thumb to fix the page, or other long fingers to touch the margins.

In addition, we give the analysis to other experimental scenarios in Fig.11. We hope the proposed techniques can more or less help us with the behavior understanding.

VII. CONCLUSION

We propose a segmentation method based on maximizing the Frobenius norm of the Meta Motion Occurrence Histogram, which is a technique of in-hand manipulation action gist, to find the optimized segmentation of periodic hand movements. Different from gesture segmentation, the segment is sharp at the boundary of movement variation. We believe that the proposed method does support the process of *Learning from Demonstration*. Then the robot with a human-like hand allows the demonstrator to teach naturally instead of decomposing the entire operating sequence.

The current method is based on counting the meta motion, it belongs to a kind of semantic analysis of the in-hand manipulation. The meta motion derives from the joint angles of the fingers; in the process of generation there may be some error. So it is possible to have a more precise result based on the raw data. But anyway, to display with meta motion is more understandable than to display the joint angle values. In this case, humans can more easily interfere in the learning process to improve the cognition of the robot.

In our future work, the segmentation method for periodic movement will follow the in-hand manipulation action gist in being examined on a humanoid robotic hand. Besides, another study direction is to integrate the segmentation from multiple sensors.

ACKNOWLEDGMENT

This work is partially supported by the European FP7 project HANDLE ICT-236410. We thank the group led by Jorge Dias from the Institute of Systems and Robotics, University of Coimbra, Polo II, Portugal, for providing us with the experimental devices and assisting with the data collection. Gang Cheng acknowledges financial support by the Chinese Scholarship Council.

REFERENCES

[1] "Annotated catalogue of grasp and force motion signatures," HANDLE project, D10, 2010, http://www.handle-project.eu/.

[2] N. Hendrich, D. Klimentjew, and J. Zhang, "Multi-sensor based segmentation of human manipulation tasks," in *IEEE Conference on Multisensor Fusion and Integration for Intelligent Systems*, 2010, pp. 223–229.

[3] H. Kjellström, J. Romero, and D. Kragić, "Visual object-action recognition: Inferring object affordances from human demonstration," *Computer Vision and Image Understanding*, vol. 115, no. 1, pp. 81–90, 2011.

[4] D. Faria, R. Martins, J. Lobo, and J. Dias, "Manipulative tasks identification by learning and generalizing hand motions," in *Technological Innovation for Sustainability*, ser. IFIP Advances in Information and Communication Technology, L. Camarinha-Matos, Ed. Springer Boston, 2011, vol. 349, pp. 173–180.

[5] D. R. Faria, R. Martins, J. Lobo, and J. Dias, "Extracting data from human manipulation of objects towards improving autonomous robotic grasping," *Robotics and Autonomous Systems*, vol. 60, no. 3, pp. 396–410, 2012.

[6] S. Cobos, M. Ferre, M. Ángel Sánchez-Urán, J. Ortego, and R. Aracil, "Human hand descriptions and gesture recognition for object manipulation," *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 13, no. 3, pp. 305–317, 2010.

[7] A. Valtazanos, D. K. Arvind, and S. Ramamoorthy, "Comparative study of segmentation of periodic motion data for mobile gait analysis," in *Wireless Health*, 2010, pp. 145–154.

[8] P. Vinayavekhin, S. Kudohf, and K. Ikeuchi, "Towards an automatic robot regrasping movement based on human demonstration using tangle topology," in *2011 IEEE International Conference on Robotics and Automation*, 2011, pp. 3332–3339.

[9] M. Kondo, J. Ueda, and T. Ogasawara, "Recognition of in-hand manipulation by observing contact state transition for robot hand control," in *Robotics Research*, ser. Springer Tracts in Advanced Robotics, M. Kaneko and Y. Nakamura, Eds. Springer Berlin / Heidelberg, 2011, vol. 66, pp. 349–360.

[10] K. Matsuo, K. Murakami, T. Hasegawa, K. Tahara, and R. Kurazume, "Segmentation method of human manipulation task based on measurement of force imposed by a human hand on a grasped object," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2009, pp. 1767–1772.

[11] D. Kulic, W. Takano, and Y. Nakamura, "Online segmentation and clustering from continuous observation of whole body motions," *IEEE Transactions on Robotics*, vol. 25, no. 5, pp. 1158–1166, 2009.

[12] J. Barbič, A. Safonova, J.-Y. Pan, C. Faloutsos, J. K. Hodgins, and N. S. Pollard, "Segmenting motion capture data into distinct behaviors," in *Proceedings of Graphics Interface*, 2004, pp. 185–194.

[13] F. Zhou, F. Torre, and J. Hodgins, "Aligned cluster analysis for temporal segmentation of human motion," in *IEEE International Conference on Automatic Face Gesture Recognition*, 2008, pp. 1–7.

[14] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: a review," *ACM Computing Surveys*, vol. 31, pp. 264–323, 1999.

[15] T. Rakthanmanon, E. Keogh, S. Lonardi, and S. Evans, "Time series epenthesis: Clustering time series streams requires ignoring some data," in *IEEE International Conference on Data Mining*, 2011, pp. 547–556.

[16] E. Keogh, L. Wei, X. Xi, M. Vlachos, S.-H. Lee, and P. Protopapas, "Supporting exact indexing of arbitrarily rotated shapes and periodic time series under euclidean and warping distance measures," *The VLDB Journal*, vol. 18, pp. 611–630, 2009.

[17] Z. Ju and H. Liu, "Recognizing hand grasp and manipulation through empirical copula," *International Journal of Social Robotics*, vol. 2, pp. 321–328, 2010.

[18] ——, "Hand motion recognition via fuzzy active curve axis gaussian mixture models: A comparative study," in *IEEE International Conference on Fuzzy Systems*, 2011, pp. 699–705.

[19] N. Lau, B. Wong, and D. Chow, "Motion segmentation method for hybrid characteristic on human motion," *Journal of Biomechanics*, vol. 42, no. 4, pp. 436–442, 2009.

[20] J. Hackett and M. Shah, "Multi-sensor fusion: a perspective," in *IEEE International Conference on Robotics and Automation*, 1990, pp. 1324–1330 vol.2.

[21] M. Ishikawa and N. Sasaki, "Gesture recognition based on som using multiple sensors," in *International Conference on Neural Information Processing*, 2002, pp. 1300–1304 vol.3.

[22] C. Mitsantisuk, K. Ohishi, and S. Katsura, "Control of interaction force of twin direct-drive motor system using variable wire rope tension with multisensor integration," *IEEE Transactions on Industrial Electronics*, pp. 498–510, 2012.

[23] "Parameterizing and creating new actions," HANDLE project, D24, 2012, http://www.handle-project.eu/.

[24] "Tactile sensing for data-gloves," HANDLE project, D6, 2009, http://www.handle-project.eu/.

[25] "Protocol for the corpus of sensed grasp and handling data," HANDLE project, D4, 2009, http://www.handle-project.eu/.

[26] R. Martins, D. R. Faria, and J. Dias, "Symbolic level generalization of in-hand manipulation tasks from human demonstrations using tactile data information," in *IEEE/RSJ International Conference on Intelligent Robots and Systems Workshop on Grasp Planning and Task Learning by Imitation*, 2010, pp. 64–70.

[27] C. Williams, D. Shang, and H. Carnahan, "Pressure is a viable controlled output of motor programming for object manipulation tasks," in *Haptics: Generating and Perceiving Tangible Sensations*, ser. Lecture Notes in Computer Science, A. Kappers, J. van Erp, W. Bergmann Tiest, and F. van der Helm, Eds. Springer Berlin / Heidelberg, 2010, vol. 6192, pp. 339–344.