In-hand Manipulation Action Gist Extraction from a Data-glove

Gang Cheng, Norman Hendrich, and Jianwei Zhang

Department of Informatics, Group TAMS, University of Hamburg, D-22527 Germany
{cheng,hendrich,zhang}@informatik.uni-hamburg.de

Abstract. The process of different human manipulating a specific object in hand obeys very similar operating steps. The hand movement can be modeled and generalized into action gist to guide other human or robots to execute the specific in-hand manipulation task. This paper suggests a kind of action gist similar to the way humans learn to represent the five finger hand motions in in-hand manipulation. Our method is based on Gaussian Markov Random Field that processes data-glove values to obtain the action gist. Several experiments are carried out to discuss the performance of the proposed methods.

1 Introduction

The word *gist* means the essential part of an idea or experience. Different from *hand gesture*, the *in-hand manipulation action gist* is a concept with kinetic property. It represents the key hand motions in any given manipulation task and widely adapts to different hands. The manipulation process is generalized as several compact meta motions. On the one hand, this makes it easy to remember, on the other hand it can be translated from one entity to another, just as the knowledge passing from the teacher to the student.

As we know, in the mechanism of the human hand, the motions and forces are governed by the neuromuscluar apparatus, refer to [15]. The movement of the hand is continuous, but according to human cognition, it can be classified as infinite types of motions in the brain. For example, as the muscles tightening up and relaxing, or the fingers closing and opening. Then in the specific application, the possible solution sequence is recalled and executed. The object in question is touched and released by the hand components over time. When the touching motion is executed, an interacting force is generated between the object and the hand, and the neuromuscluar system keeps the hand in a proper *force applying state* that does not damage the hand itself but still holds the object firmly.

In our lab we have a five-finger air muscle hand from the Shadow Robot Company [1], it is very similar to the human hand and better protected against damage even when overforce is applied. With a humanoid hand, a robot can implement much more human-like object manipulation than before. Because of

[1] http://www.shadowrobot.com/

the high degree-of-freedom, a multi-finger robot hand can perform more dexterous skills rather than grasping, holding or translating the object from one place to another. It can rotate, or shift objects and perform other advanced in-hand movements. These manipulation skills depend on the cooperation of five fingers and the palm, and in the process of in-hand manipulation, the roles are hand and object. The hand plays the role of control, and changing the object state is the aim of the manipulation. Therefore here the manipulation process is considered as a *State-Action Model* [8][4][6], meaning that the whole process is divided into states which are changed through actions. The action is equal to hand movement, and the state is supposed to be the criterion of how the process proceeds. Hand movement can be considered as a continuous hand joint angle variation, with countless angle combinations between each joint pair. The movement leads the manipulation process from one state to another state until the final target of the application is achieved.

The method can be applied in both human analysis and the control of robots with humanoid hands. However, it is unrealistic to map the motion exactly as from the demonstrator because of the different hand sizes. It can be imagined different-sized hands can interact with the object from different distances, obviously it can result in different gaps with a same pose. Actually in developing their hand skills, humans have the ability to learn from others and to practice by themselves. Nobody can memorize the detailed joint angles of their hands, but they can remember the key motions which are related to the moving tendency of each finger; this is defined as *in-hand manipulation action gist.*

This paper proposes a cognitively feasible in-hand manipulation action gist definition for a robot with an extremely life-like humanoid hand, to enable it to learn in-hand manipulation with a small amount of key information. The action gist is expected to be universal for all in-hand movements regardless of whether it is simple (grasping) or complex (finger-gaiting). The structure of this paper is organized into several sections. After the following related work, the definition of meta motion is given, which is an element of the in-hand manipulation action gist. Then the modeling process is introduced, and experiments are carried out to discuss the performance of the algorithms. The final part is the conclusion and future work.

2 Related Work

There are multiple ways to generate a manipulation model.

One kind of model is to plan the motion in continuous space including the position and the speed of each relative component. The major stream is the dynamic movement primitive (DMP) framework introduced by [3] and [12], in which the movement is recorded and represented with a set of differential equations. The position and the speed is controlled in terms of the immediate position and speed feedback. [9] expanded the model into a manipulation control application so that the hand can grasp and place the object in the destined area. To include obstacle avoidance in this job, an extra item is added in the system equation, which causes the form of the framework to change with the task. Different from the separate models to deal with multiple tasks, [1] applied Locally Weighted Regression to generate the movement, and the manipulating process is divided into several steps by the perceptual input. Rather than generalizing a trajectory in Cartesian or joint angle space, [2] considered the joint velocity space and enables the robot to accomplish similar tasks. As a result, this method can produce smoother trajectories than others.

The above frameworks consist of models depending on precise perception of spatial manipulator trajectories. However, for muscle control, it tracks the trajectory related to the moving tendency, not the position. Therefore, DMP does not offer any significant advantages to the target of this paper.

Another branch but a relatively older one is the generalized motor program (GMP), see [14] and [13]; here the overall process is guided by invariant features. [7] extended this model with the symbolic motion structure representation (SMSR) algorithm. The body movement is tracked and segmented according to the joint angles, and then the values are used to plan a novel similar application. However, the SMSR only extracts the body motion into simple joint angle variations such as increasing, decreasing and stationary. Therefore, it would have difficulties when dealing with the multiple links cooperation application because it does not consider this kind of application so much. Different from simply defining the motion, it is possible to have a related higher semantic model. [10] applied Fuzzy-Logic Control to execute the motion sequence, and this idea was examined in a 2D five-segment body model by simulation. The above methods suppose that the motion sequence to an application is fixed, but actually humans can have many ways of completing a specific application. What we need are the most effective or common methods of the teacher.

[5] indicated that humans learn motion by way of muscle control, not the position perception, therefore to know the posture variation (joint angle) is more important than the absolute posture. Thus the motion tendency oriented model is more feasible than DMP.

Once the model is decided on, the next problem is how to sense the movement. Many studies concentrate on sensing from the robot, for example, [9], [1], or [2]. However, for fingers, it is not convenient to directly move the robotic fingers to find the result. Another channel is vision, the components are tracked to complete the motion behavior model. For example, [11] employed color pattern on the demonstrator to track the human motion. It is promising to use vision to analyze hand motion, but the visual processing itself is a challenging topic which increases the difficulty of model generation.

A quick way to know the finger movement is using a data-glove, it can sense every finger joint relation in each data frame. Based on this kind of sensing channel, our study intends to generate an action gist model to represent human in-hand manipulation behavior.

3 Meta Motion Definition

To establish a set of hand motions which presents the hand posture transformation in in-hand manipulation, we intend to construct the model as follows.

1. It covers all possible movements of a hand
2. Each motion in the set is unambiguous from other motions
3. The motion involves the relative joint angle variation but no absolute position information

An exception to the above is the idle pose. When the motion remains static for a while, we have to decide whether it is "move, stop and move again", or consider it as moving continuously. Our strategy is to analyze the movement without static motions first, and in the second loop to find the static section following certain rules.

Supposed that the hand has the form of five fingers and one palm, the palm stays still, then the movement is equal to the cooperation of the five fingers. The basic movement of each finger can be classified as open or close, and in terms of the moving direction at the *proximal phalange end* related to the palm, every finger has the same motion definition. Specifically, the coordinate origin of the thumb is different from the other four fingers because of its diverse position on the palm.

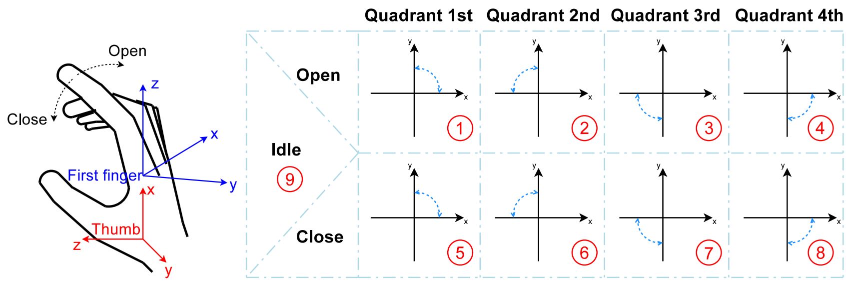


Fig. 1. Meta motion definition. Five fingers move related to the palm, so we describe their moving directions in coordinates. The thumb in the red coordinate is different from the other four fingers due to its special location in the hand. In each finger, two flex/ext-joints are modeled as one parameter as open or close, and the abduction angle cooperates with the metacarpal-proximal angle to form a 2D projected direction (blue arc indicates the directional range in the quadrant). The idle motion is specifically set apart and labeled as 9

Shown in Fig.1, we project the finger motion into 2-dimensional space because the finger ends are fixed on the palm. In the X-Y plane, the finger direction is classified as 4 directions as the 4 quadrants in the Cartesian coordinate; plus with the open, close and the idle period, each finger has 9 types of meta motions. To ensure the motion model a uniform form, the X axis and the Y axis in the moving direction related to the coordinate origin is either parallel or vertical to the palm plane.

4 Action Gist from Data-glove

Here the action gist is defined as the key meta motions between two adjacent states. Guided by the action gist, the object is manipulated from the begin state to the end state.

The data-glove is a direct way to perceive the hand movement, as the data is measured by the joint angle value. Therefore, the values from the data-glove become the source for analyzing the hand movement in in-hand manipulation applications.

Corresponding to the degree of freedom, each finger has several joint values from the data-glove. However, according to the general law, the distal-intermediate and proximal-intermediate angles increase in close movement, decrease in open movement, and the varieties of metacarpal-proximal and abductional angles indicate the moving direction in the X-Y plane of the finger.

Different from the ideal environment, the acquired data-glove value can not be directly applied in the analysis. One reason for this is for the sensor noise, another one is the issue from the human operator, e.g. a hand tremor in slight operation, a short but unnecessary movement during manipulation, or at the moment the finger starts to touch the object, the value may be abnormal. Therefore a Gaussian Markov Random Field based algorithm is proposed to extract the action gist of each finger, it can effectively decrease the negative impact from the mentioned issues and provide a concise meta motion sequence. This algorithm considers each value frame from the data-glove as a node, every node can influence the other nodes on which meta motion they belong to, the nearer nodes have the stronger impacts, the criteria are based on the single meta motion similarity and the node distance, the node relationship according to the assumption is illustrated in Fig.2.

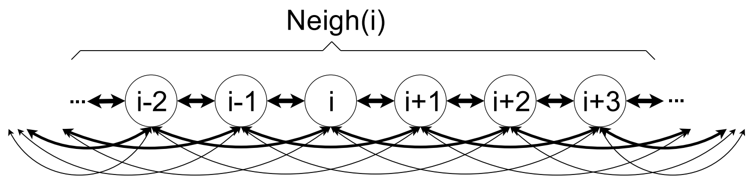


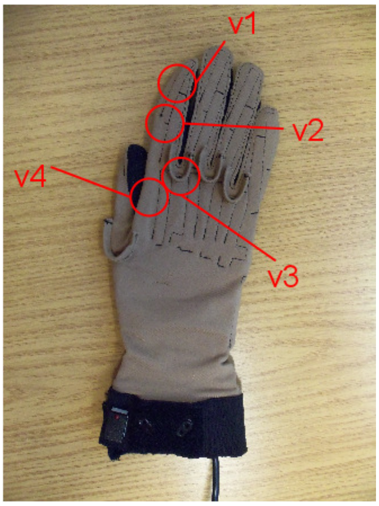
Fig. 2. Node relationship according to Gaussian MRF. Supposing each data-glove value is a node, then each node is related to other nodes in the neighboring set Neigh$(\cdot)$. With the impact factor obeying Gaussian distribution, the linewidth indicating the strength of the impact factor, we can see that the nodes sitting closer have stronger influence

The single meta motion similarity of each node can be presented as:

$$I_i^j = \begin{cases} \sum_{k \in \mathbf{F_g}} \left| v_i^k \right| + \varepsilon & , C^j_{k \in \mathbf{F_g}}(v_i^k) = 1 \\ 0 & , else \end{cases} \tag{1}$$

Here I_i^j is the intensity of node i that is similar to meta motion j, v_i^k is the k-th glove value difference (current value minus previous value) in node i. The k-th value from the data-glove sensor should belong to one finger $\mathbf{F_g}$, $\varepsilon > 0$ promises the value of intensity is always above 0. This value is not critical but should be a low value, we suggest to fix $\varepsilon = 0.05$ as experience.

Additionally, $C(\mathbf{v})$ is the conditions that the finger joint angle difference stay in the range of the corresponding meta motion j. Assuming that there are always four values $v_1, v_2, v_3, v_4 \in \mathbf{v}$ standing for the joint angle variation in five fingers, they are mapped correctly with v_i^k. Commonly, v_1 is for distal-intermediate, v_2 is for proximal-intermediate, v_3 indicates abduction and v_4 is for the metacarpal-proximal angle difference. Specifically, for the thumb values in the data-glove, in order to have a uniform expression, the rotation angle is considered as v_3. Besides, the abduction value v_4 should be adjusted as an identical increasing direction according to the meta motion definition, then the conditions $C(\mathbf{v})$ are listed in the right table of Fig.3. v_1 has a less important effect here because when the object is manipulated, it is easy for the finger tips touching the object are easily to create a contra direction with v_2, but v_2 is related stably. Whether the finger is open or close mainly depends on the movement between the proximal and intermediate joints. In the table, "$\times$" means v_1 can be any value in this condition.



meta motion	v_1	v_2	v_3	v_4
1	$\times$ < 0	< 0 $= 0$	$\geqslant 0$	$\geqslant 0$
2	$\times$ < 0	< 0 $= 0$	$\leqslant 0$	$\geqslant 0$
3	$\times$ < 0	< 0 $= 0$	$\leqslant 0$	$\leqslant 0$
4	$\times$ < 0	< 0 $= 0$	$\geqslant 0$	$\leqslant 0$
5	$\times$ > 0	> 0 $= 0$	$\geqslant 0$	$\geqslant 0$
6	$\times$ > 0	> 0 $= 0$	$\leqslant 0$	$\geqslant 0$
7	$\times$ > 0	> 0 $= 0$	$\leqslant 0$	$\leqslant 0$
8	$\times$ > 0	> 0 $= 0$	$\geqslant 0$	$\leqslant 0$

Fig. 3. An example of the finger joint angle difference in the first finger. v_1 is for distal-intermediate, v_2 is for proximal-intermediate, v_3 indicates abduction between first finger and middle finger, v_4 is for metacarpal-proximal

For the data-glove, we have to mention that the abduction angle is not the absolute angle related to the palm. That means v_3 is not working perfectly, but in this study we do not consider it as a critical problem.

When the single similarities of all nodes are calculated, the influence from other nodes can be obtained by:

$$P_i^j = \sum_{t \in \mathbf{Neigh}(i)} I_t^j G(t, i, \sigma) \tag{2}$$

where $G(t, i, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-i)^2}{2\sigma^2}}$ is the typical Gaussian distribution form, $\mathbf{Neigh}(i)$ is the node set near node i (refer to Fig.2). Because the concerned action gist locates between each adjacent state pair, it is actually set as the entire glove value sequence. Besides σ is a parameter representing the area one node can primarily impact with, it also means the shortest single motion execution time corresponding to the data-glove sensing speed. Then the likelihood of meta motion j at each node can be compared to find the best meta motion segmentation.

In addition to the action gist analysis, the idle motion is processed independently from the eight kinetic motions mentioned above. The Gaussian MRF based method can also be employed here, but according to the experimental experience, to find a frequent value as high as desired in the sliding window is a better solution. To realize this method, the first step is also to have the single similarities of each node to be similar to Eq.1, but the intensity of meta motion 9 at node i becomes as $I_i^9 = 1$ and the condition becomes $C(\mathbf{v}) = 1 \iff \mathbf{v} = \mathbf{0}$. Thus the idle sections can be determined by the following condition:

$$\sum_{t \in \mathbf{Neigh}(i)} I_t^9 > threshold \tag{3}$$

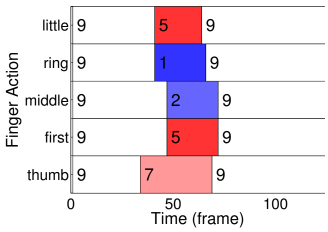
Thus node i stays idle when the sum of single intensities is larger than *threshold*. Here $\mathbf{Neigh}(i)$ is set to be at the range of d_{sw}, which is the size of the sliding window, then d_{sw} nodes are taken into consideration to find the idle section. In addition, all adjacent idle nodes are merged as an idle section, but if the length of idle sections is shorter than a single motion execution time σ, this section should be considered as not idle. Commonly, we find that $threshold = 0.90$ and $d_{sw} = 20$ fit most cases.

Using the proposed process, the action gist can be extracted from the raw data-glove values.

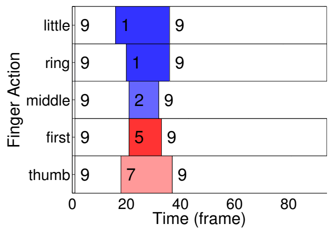
5 Experiment

Different kinds of objects are used to examine the proposed method, the hand movement involves from *clearly moving without object* to *complex finger gaiting*. The corresponding action gists are extracted from the glove values automatically, and they all agree with our expectation. In order to make a intuitive view, here we just take *bottle screw cap unscrewing* as an example. There are four different-sized bottle screw caps as Fig.4, a participant rotates the caps by four fingers for many times. After several trials we list several typical results through the proposed method as in Fig.5.

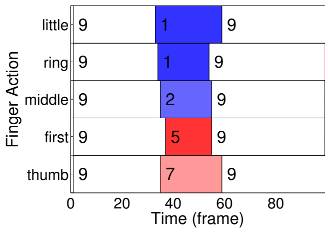
Fig. 4. Unscrewing different-sized bottle screw caps. The thumb, first, middle and ring finger participate in this scenario. Each screw cap is rotated as around 90 degree anticlockwisely, and this process is defined as a trial



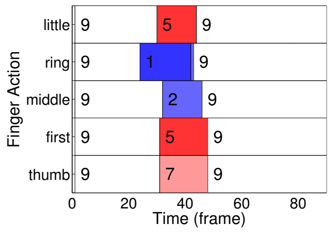
(a) Big



(b) Medium



(c) Small



(d) Tiny

Fig. 5. Action gists of bottle screw cap unscrewing. Each action gist is composed of the meta motions of five fingers, each meta motion is represented by different color rectangles with the corresponding type number. The x-axis is a time axis indicating the cyber-glove frame number. Taking idle motion into consideration, we find that the action gists from 4 trials look similar. The common meta motions are motion 7 in the thumb, 5 in the first finger, 2 in the middle finger, and 1 in the ring finger.

There are countless ways to move the fingers to reconfigure the object achieving to the goal state, but for the transfer to a robotic hand one solution is enough to guide the manipulation. Furthermore, by the result of bottle screw cap unscrewing we can see for different-sized caps unscrewing it does exist similar action gists. Therefore, we can demonstrate the scenario-specific finger-gaiting movement many times and find the popular action gist. In this case, the common one has stronger adaptability for the robotic hand, which is in different size from human hand, to complete the task.

Among the parameters in the algorithm, σ is the only one depending on the application. As we mentioned in the previous section, this parameter is relevant to movement speed and data-glove framerate. Higher σ merges more short terms, but meanwhile we risk losing critical short meta motions. On this point the value selection should be considered carefully. We enumerate all possible values and compare the extraction results for several typical applications, finally we find that the configuration of σ is not so strict. For most cases the extraction results are same, otherwise the length of meta motion changes slowly with σ variation. Thus it is not necessary to check every possible value (e.g. from 1 to 20 for 20 trials), instead, we can set $\sigma = 5$ to acquire the details, and then set $\sigma = 10$, 20 or even more to get the general context.

6 Conclusion and Future Work

This study concentrates on action gist extraction from the demonstration of in-hand manipulation. Different from the manipulator trajectory planning, this model works in a fuzzy way to guide the finger movement. It gives the manipulator a related loosely explored space to implement the task, and from the view of human in-hand manipulation, it is more similar to the mechanism of the human hand.

This action gist model is being examined by simulation and real robot tests. When the action gist is mapped back to robotic hand control, it is supposed to work as guidelines because it provides the meta motions of each finger in order. Different-sized hands apply different joint angles to execute the manipulation, but the meta motion is always correct to indicate the finger movement direction. In every trial we give the robot quantized parameters according to a fixed meta motion sequence, and through iterations the parameters are refined to ensure the correct state transition.

The model currently is built from the value of a data-glove. One disadvantage is that the human demonstrator wearing the data-glove has a different feeling and executes the movement unnaturally, and difficult manipulation applications are hardly handled. Another drawback is related to the four abduction angles in the data-glove, which are angles between two fingers, not the absolute angle related to the palm. From this point, we do not guarantee that the finger movement perception is always correct. Hence to fuse the result from other sensors, especially tactile sensing, is another direction for developing the model further.

Acknowledgments.

This work is partially supported by the European FP7 project HANDLE ICT-236410. Gang Cheng acknowledges financial support by Chinese Scholarship Council.

References

1. Gams, A., Ude, A.: Generalization of example movements with dynamic systems. In: IEEE-RAS International Conference on Humanoid Robots. pp. 28–33 (2009)
2. Gielniak, M., Liu, C., Thomaz, A.: Stylized motion generalization through adaptation of velocity profiles. In: IEEE International Symposium on Robots and Human Interactive Communications. pp. 304–309 (sept 2010)
3. Ijspeert, A.J., Nakanishi, J., Schaal, S.: Movement imitation with nonlinear dynamical systems in humanoid robots. In: IEEE International Conference on Robotics and Automation. pp. 1398–1403 (2002)
4. Kjellström, H., Romero, J., Kragić, D.: Visual object-action recognition: Inferring object affordances from human demonstration. Computer Vision and Image Understanding 115(1), 81–90 (2011)
5. Mah, C.D., Mussa-Ivaldi, F.A.: Generalization of object manipulation skills learned without limb motion. The Journal of Neuroscience 23(12), 4821–4825 (2003)
6. Ogawara, K., Takamatsu, J., Kimura, H., Ikeuchi, K.: Modeling manipulation interactions by hidden markov models. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems. vol. 2, pp. 1096–1101 vol.2 (2002)
7. Park, W., Chaffin, D.B., Martin, B.J., Faraway, J.J.: A computer algorithm for representing spatial-temporal structure of human motion and a motion generalization method. Journal of Biomechanics 38(11), 2321–2329 (2005)
8. Pastor, P., Kalakrishnan, M., Chitta, S., Theodorou, E., Schaal, S.: Skill learning and task outcome prediction for manipulation. In: Proceedings of IEEE International Conference on Robotics and Automation. pp. 3828–3834 (may 2011)
9. Pastor, P., Hoffmann, H., Asfour, T., Schaal, S.: Learning and generalization of motor skills by learning from demonstration. In: Proceedings of IEEE International Conference on Robotics and Automation. pp. 1293–1298 (2009)
10. Qu, X., Nussbaum, M.: Simulating human lifting motions using fuzzy-logic control. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans 39(1), 109–118 (2009)
11. Riley, M., Cheng, G.: Extracting and generalizing primitive actions from sparse demonstration. In: IEEE-RAS International Conference on Humanoid Robots. pp. 630–635 (2011)
12. Schaal, S., Ijspeert, A., Billard, A.: Computational approaches to motor learning by imitation. Philosophical Transactions of the Royal Society B: Biological Sciences 358(1431), 537–547 (2003)
13. Schmidt, R., Lee, T.: Motor control and learning: a behavioral emphasis. Human Kinetics (1999)
14. Schmidt, R.A.: A schema theory of discrete motor skill learning. Psychological Review 82(4), 225–260 (1975)
15. Taylor, C.L., Schwarz, R.J.: The anatomy and mechanics of the human hand. Artificial limbs 2(2), 22–35 (1955)