

Gesture-based Object Recognition using Histograms of Guiding Strokes

Amir Sadeghipour¹
sadeghipour@uni-bielefeld.de

Louis-Philippe Morency²
morency@ict.usc.edu

Stefan Kopp¹
skopp@techfak.uni-bielefeld.de

¹ Cognitive Interaction Technology -
Center of Excellence,
Bielefeld University,
Bielefeld, Germany

² Institute for Creative Technologies,
University of Southern California,
Los Angeles, CA, USA

Abstract

Humans perform iconic gestures to refer to entities through embodying their shapes. For instance, people often gesture the outline of an object (e.g. a circle for a ball) when referring to it during communication. In this paper, we present a gesture-based object recognition algorithm that enables natural human-computer interaction involving iconic gestures. Based on our analysis of multiple gesture performances, we propose a new 3D motion description of iconic gestures, called Histograms of Guiding Strokes (HoGS), which successfully summarizes hand dynamic during gestures. Our gesture-based object recognition algorithm compares favorably to human judgment performance and outperforms most conventional gesture recognition approaches.

1 Introduction

Gestures are widely used during human interaction. As a frequent communicative behavior, hand gestures can be used to express semantic content similar to speech. McNeill [7, 8] proposed a categorization of hand gestures, which consists of four types (or dimensions): iconic, metaphoric, deictic and beat gestures. Iconic gestures are pictorial gestures, which depict entities, objects or actions. In this study, we focus on iconic gestures that refer to physical objects. Such iconic gestures carry visual characteristics of the referred object in hand dynamic. Comparing to other gesture types, this makes iconic gestures less conventional and more flexible and adaptable to different frames of references. Their flexibility hinders generalization over different performances, since there are different ways to perform an iconic gesture for a specific object. For instance, in order to refer to a ball, one can draw a circle with the right hand, or two semi-circles with both hands simultaneously. Moreover, hand movement direction or velocity may differ from one another without affecting the referential intention behind the gesture. However, in spite of all variations among iconic gestures performances, one looks for the common shape-related features, which indicate the referred object (e.g. recognizing the shape of a circle). In this context, distinguishing between the relevant similarities and differences among gesture performances represents the main challenge in iconic gesture recognition.

© 2011. The copyright of this document resides with its authors.
It may be distributed unchanged freely in print or electronic forms.

In this paper, we present a new gesture descriptor, called Histograms of Guiding Strokes (HoGS), well-suited for automatic recognition of iconic gesture depicting 3D objects. We created a dataset, called 3D Iconic Gesture dataset (in short, 3DIG), which contains 29 subjects performing more than 1700 iconic gestures. Analyzing the recorded gesture performances, we show that people use very different techniques when performing gestures. Thus, the main challenge addressed by our HoGS descriptor is to tolerate the variations among the gesture performances, while enabling robust discriminative classification. In other words, our HoGS descriptor generalizes over intra-class variations while maintaining the inter-class variabilities. We performed extensive experiments to evaluate our approach with respect to: alternative classification features, human judgment performance, and other conventional approaches of gesture recognition (such as HMM using instantaneous features).

In the next Section, we review some related work investigating iconic gestures. Section 3 first describes how the 3DIG dataset has been recorded. Then, based on the data analysis, we explain how the computation of HoGS descriptor fulfills the raised requirements. In Section 4 the results of evaluation experiments are reported and discussed.

2 Related work

Aiming at designing an efficient and natural human-computer interface, gesture recognition has gained in importance. Specifically, recent technological advances in motion tracking systems have opened new perspectives in recognition of non-verbal behavior. Marker-free motion tracking devices such as MS KinectTM provide low-priced and convenient solutions compared to older tracking systems. Generally, the work done on gesture recognition focuses mostly on sign languages or on predefined hand movements as controlling means for navigation or manipulation, as it was reported by Karam and Schraefel [4]. Accordingly, they suggested a taxonomy of gestures in human computer interactions, based on a literature review of gesture recognition. Mitra [9] provided a survey on the approaches of gesture recognition, ranging from statistical modeling, computer vision, pattern recognition and connectionist systems. Given the limited space in this paper, we focus our review on iconic gesture analysis and recognition.

Marsh and Watt [6] performed a study to test whether iconic gestures could be employed as an intuitive technique in human-computer interaction. They asked the subjects to express or describe different complex 2D and 3D objects through non-verbal communication. Data analysis showed that the subjects referred to the primitive objects by drawing the outline of their shapes. The complex objects were additionally described through mimicking actions performed on the referred objects. Sowa [10] undertook a study to investigate co-verbal iconic gestures and proposed a model for their shape-related semantic representation. He analyzed 383 recorded iconic gestures [11] which were performed for five relatively complex objects. He reported some gesture techniques such as: decomposing a complex object into its components and referring to them sequentially; or the so called “dimensional underspecification”, when only the dominant object extents are expressed. The wide range of techniques that people can use to perform gestures raises the main challenge of gesture recognition, namely supporting the big variation among performances. For instance, Vatavu et al. [13] referred to this problem and proposed an approach based on elastic deformation of motion trajectories. Thereby, it supports variations in gesture performances such as translation, scale and rotation.

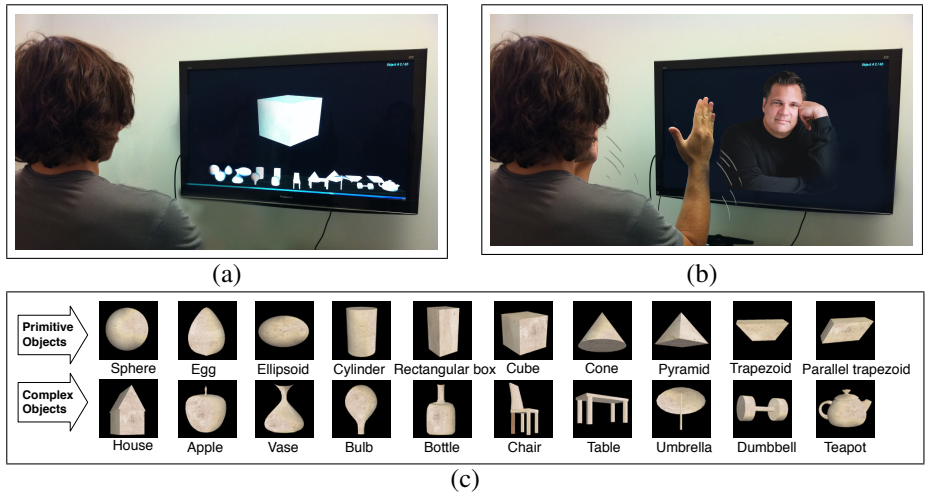


Figure 1: The experiment setup. In (a) a participant is watching a cube rotating in the middle of the screen. In (b) the participant is performing a gesture to refer to the previously seen cube for an imaginary interlocutor. (c) shows all the primitive and complex virtual objects.

3 Gesture-based object recognition

The existing studies on iconic gestures neither provide a public dataset nor sufficient kinematic specification of iconic gestures. Hence, we created an extensive dataset, called 3D Iconic Gestures dataset, in which we tracked subjects' body motion while performing gestures illustrating 3D objects. Based on the analysis of intra-class and inter-class variations in this dataset, we propose HoGS motion descriptor. HoGS are invariant to the intra-class variabilities among different performances referring to the same object, while keeping distinctive features with respect to inter-class variations. The following section describes our 3DIG dataset while Section 3.2 describes the design principles behind HoGS.

3.1 Three-dimensional Iconic Gesture (3DIG) dataset

We captured gesture performances of 29 subjects (20 males and 9 females), while performing iconic gestures to refer to 20 different 3D objects. For this study, subjects stand in front of a big display, mounted at about the height of their heads (see Figure 1). They were asked to refer to the virtual objects by gestures. The subjects were structured to perform the gesture so that people watching the video could recognize the object.

During the study, the subjects could see all 20 different 3D virtual objects which were located at the bottom of the display. At each turn, one of the virtual objects was zoomed in at the center of the screen, and kept rotating so the subject could see the different perspectives. Subjects performed the gesture while a photo of an attending person was shown. As soon as the subjects retracted their hands back to the rest position, or gave a verbal signal, the next object was shown. The virtual objects were shown in a random order and each of them three times, for a total 60 gestures per subject. Body motions were recorded only while showing the photo of the attending person. Using MS KinectTM, we captured 1739 gesture performances (~ 87 gestures per object), each in three formats: color video, depth video and motion of the tracked skeleton (as 3D positions of 20 joints in 30 fps). The recorded dataset

<i>Intra-class variations</i>	<i>Description</i>	<i>Example(s)</i>
Direction	The wrists' movement direction.	Drawing the outline by moving the hand either upward or downward, clockwise or counter-clockwise.
Velocity	The wrists' movement velocity.	
Size	The size of the used spatial space for a gesture performance.	Drawing a small circle for a sphere, or using the entire feasible space.
Degree of simplification	Ignoring the objects' details or spatial dimensions (abstraction and dimensional underspecification [11]).	Whether referring to the twig of the apple or not. Depicting a square for the cube, or the whole 3D shape.
Repetition	Repeating some parts of a gestural wrist movement places emphasis on them.	Repeated curved hand movements to refer to the significant upper part of the umbrella.
Position	The location of the used spatial space for a gesture (relative to the performer).	Drawing a small triangle in front of the head or slightly on the side; or a big triangle starting from much lower spatial position.
Ordering	Objects may be decomposed into simpler components, which are referred to sequentially.	First drawing the triangular roof of the house, then the cubical part or vice versa. First drawing the bottom circular surface of a cone, then a triangle on the top or vice versa.
Handedness	The contributed hand(s).	Performing with the left or the right hand alone, or both together.

Table 1: The mostly observed intra-class variations while analyzing the color videos of the 3DIG dataset. The intra-class variations illustrate the feature dimensions whose values vary among gestures that refer to the same object.

is called 3D Iconic Gesture dataset (3DIG) and is available online ¹.

As frames of reference for recorded gestures, we chose 10 primitive and 10 complex objects (see Figure 1). As primitive objects, we took common geometrical shapes such as sphere, cone or cube. Gestures referring to such primitive shapes can generally refer to many real, more complex objects iconically (e.g. a computer is shaped like a rectangular box). The reason is that simplifying objects and referring to their abstract shapes is a commonly used strategy while performing iconic gestures [11]. Additionally, we asked subjects to perform gestures for more complex objects such as house, umbrella or chair. In this way, through investigating differently complex objects, we look for a general approach to recognize a wide range of iconic gestures.

3.2 Histograms of Guiding Strokes

For the aim of gesture based object recognition, we are looking for optimal descriptor of iconic gesture performances. To this end, the main challenge is the wide variability between subjects when illustrating the same object. This means that the optimal descriptor needs to be invariant to the intra-class variabilities, while it should still represent the features which best discriminate between different classes. In the following subsections, first we analyze the recorded dataset and discuss the inter-class and intra-class variations among gesture performances. Based on this analysis, we propose the Histograms of Guiding Strokes (HoGS) as motion descriptor, which takes both these variations into account.

¹ Available at <http://projects.ict.usc.edu/3dig>

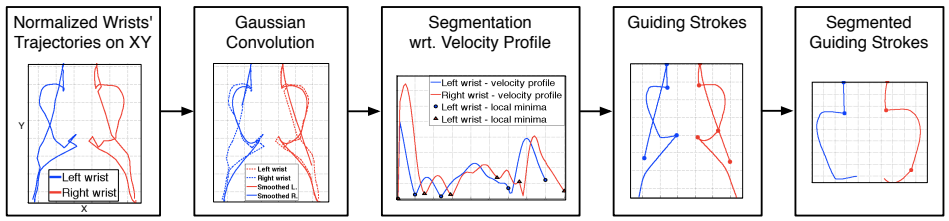


Figure 2: The data processing pipeline to extract segmented guiding strokes. As an example, a projection of a relatively clear gesture performance is illustrated (with $ID = 100$ in 3DIG, referring to “bottle”).

3.2.1 Inter-class vs. intra-class variations

We analyzed intra-class and inter-class variabilities in the 3DIG dataset based on the recorded color videos. The analysis shows that people use different techniques to perform iconic gestures: for instance, either performing an action on an imaginary object (e.g. biting an apple), or to refer to the shape (e.g. by drawing the contours). We call the technique, in which the hand or the index finger illustrates the 3D or 2D contours of an object, *drawing technique*. As the most common technique, this was used for 83% of the primitive objects and 77% of the complex ones. Hence, the wrists’ movements have the most contribution to the gesture content.

The aforementioned diversity of gesture techniques leads to a wide variation among gesture performances. However, in spite of all these performance variations, many aspects of the performed gestures correlate with the characteristics of the referred object. That means, on the one hand we have the *intra-class variations* which do not change the referential intention behind a gesture. Table 1 lists the mostly observed intra-class variations in our data analysis. Many of these variations are to different degrees subject-dependent (such as direction, velocity or handedness), and some of them are to some extent object-dependent (e.g. degree of simplification or repetition). On the other hand, the shape-related *inter-class variations* play a distinctive role in iconic gesture recognition. These are mostly features that represent the geometrical characteristics such as curvature, skewness and movement orientation (table 2 lists the corresponding features as inter-class variations).

3.2.2 Extracting Guiding Strokes

To cope with the intra-class and inter-class variations, we employ the idea of decomposing the movement trajectories into strokes. Such compositional representation is known from processing motion data in different fields of studies. For instance, stroke-wise representation is applied in the field of sketch recognition of pen gestures or handwritten characters and letters (e.g. [15]). Interestingly, studies at behavioral and neural level of motor control showed that motor systems use a similar principle of decomposing complex movements into simpler elements, called motor primitives, and combine them when performing a movement [3]. For a similar segmentation, we apply the concept of *guiding strokes*, proposed by Kopp and Wachsmuth [5], which splits gestural wrists’ trajectories with respect to their spatial and kinematic features.

We define a gesture as a set of guiding strokes. Figure 2 illustrates how the guiding strokes are extracted from 3D movement trajectories of both wrists. First, these trajectory-

Sample GS.	Attribute	Domain of values
	Length (normalized)	$(0, 1]$
	Width (normalized)	$(0, 1]$
	Height (normalized)	$(0, 1]$
	Curvature = $ d / l $	$[0, 1]$
	Curvature side $\equiv \text{sign}(c_2)$	$\{L, R\}$
	Skewness $\propto d' / l $	$[-0.5, 0.5]$
	Orientation = α	$[0, \pi)$

Table 2: The attributes of a guiding stroke. These attributes represent the shape-related inter-class variations which distinct between gestures referring to different objects.

ries are normalized with respect to the player position at each frame. Second, we smooth each trajectory through Gaussian convolution to remove the spatial capturing noise. In the third step, the trajectories are split based on their velocity profiles and the local minima are considered as split points. Finally, we segment the guiding strokes, by extracting the ones which contain the semantic content of a gesture and not the preparing and retraction movement parts (cf. stroke phase in [7]). This is done automatically, through thresholding with respect to the spatiotemporal features of each guiding stroke (i.e. velocity, position and ordering). The thresholds were set empirically. In the end, we have parametrized guiding strokes, each represented as a set of attributes. Table 2 lists the attributes of a guiding stroke, which are shape-related and represent the distinctive inter-class variations. The parameterized representation of gestures as guiding strokes makes it possible to extract gesture features that reduces the intra-class variation. For instance, since we do not consider the duration or the temporal ordering of each guiding stroke, the performance velocity is repressed. Additionally, we normalize the width, height and the length of each guiding stroke with respect to the diagonal of the bounding box which contains all the guiding strokes of a gesture. The normalization step of these spatial attributes makes the representation scale invariant. Furthermore, we defined orientation of a movement with the domain of values $[0, \pi)$ as an inter-class variable feature. This is a transformed version of the direction feature (from the domain of values $[-\pi, \pi)$), which was considered as an intra-class variable feature. In this way, it is still possible to distinguish between horizontal and vertical movement orientations, while the left-right or up-down direction variabilities are ignored.

Further intra-class variations (i.e., ordering, handedness, repetition and degree of simplification) are addressed through computing the histogram of the relevant guiding strokes' attributes, as described in the next subsection.

3.2.3 Computing Histograms of Guiding Strokes

The idea of using histograms as a descriptor has been quite successful in past research (e.g. see Histogram of Oriented Gradients in [2]). In this paper, we propose Histograms of Guiding Strokes (HoGS) for iconic gesture recognition. HoGS represent a gesture as the normalized histograms of the guiding strokes' attributes. The following steps are performed in order to compute HoGS as a descriptor of a gesture performance. First, we extract all the segmented guiding strokes. Then, for each projection plane (xy , xz or yz), we compute the histograms of the attributes of the guiding strokes that lay on the corresponding planes. The reason for extracting features from each plane is that many 3D objects differ only with respect to specific projections, such as cone and pyramid or cylinder and rectangular box. The number of bins in the histogram of each attribute is a parameter which was determined by maximizing the Fisher score of the training data, resulting in 2 to 5 bins per attribute.

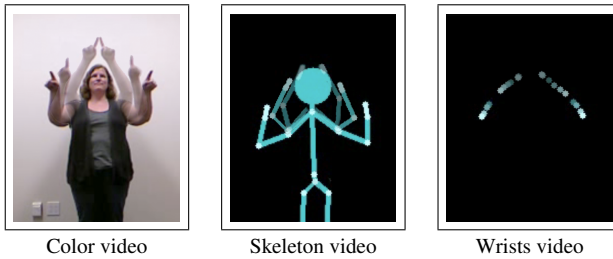


Figure 3: The captured color, skeleton and wrists videos of a gesture performance for “pyramid” ($ID = 1047$ in 3DIG). The images show three overlapped frames of the videos.

HoGS intrinsically ignores the ordering in which the guiding strokes’ attributes are given. This way, the ordering of the performed gesture parts (as an intra-class variation) is repressed. By combining the gesture trajectories performed by both left and right hands, we can address the handedness variability. The degree of simplification and repetition are other intra-class variations that affect the ratio of the meaningful parts to the whole. This ratio affects the corresponding intervals (bins) of the histograms. In this way, these intra-class variable features are not repressed completely, but they affect the representation of the gestures by affecting its significance with respect to specific features. This is the same way, how repetition and simplification of a gesture affect the recognition of a human interlocutor; namely, making the performance more or less significant.

4 Experiments

For the classification of iconic gestures in 3DIG dataset, we applied Support Vector Machine (SVM) using HoGS as descriptor. In order to validate our approach, we designed different experiments to test and compare it with respect to three aspects: first, the performance of the proposed HoGS descriptor; second, the performance of the SVM-based classifier together with HoGS features; and third, with respect to the human judgment performance.

4.1 Our approach: SVM with HoGS

For all experiments, we performed 3-Fold testing and holdout a validation subset of the data. In other words, one third of the subjects was used for testing, one third for validation and one third for training. We repeated this process three times, once for each fold. In order to classify 20 classes, we learned one-to-one SVMs and combined them through majority voting. We achieved better classification results when selecting features for each combination of objects. By doing this, we overcome the difficulty caused by shape similarity between classes (e.g. between apple, bottle, egg and sphere, or pyramid and cone). Based on the training data, we selected the best features for each combination. We used LIBSVM [1], where hyper-parameters are selected using the validation set. As classification performance metrics we use F_1 score which considers both the precision and the recall.

4.2 Baselines

Sketch recognition illustrates a very similar classification problem to the gesture research question in this study. Willems and Niels [14] collected and proposed a set of features for pen gesture recognition. Similar to our approach, the trajectories of pen gestures consist of strokes (pen down/up) to draw sketches (including iconic shapes). The proposed space-based features consider the following characteristics of a trajectory: length, the sketched area (convex hull), direction, curvature, perpendicularity, and prototypical elements such as cups or straight lines. We extracted these features out of the trajectories of the guiding strokes. The SVM classification results using these features form a baseline for our HoGS feature set.

Furthermore, to evaluate our classification method, we compare our SVM-based classifier using HoGS features against three alternative approaches. The applied pair-wise feature selection suggests applying Decision Trees that classify on the same principle. Therefore as a baseline classifier, we applied Bag of Trees using HoGS features. We used TreeBagger from Statistics toolbox in Mathworks Matlab[®] 2011a, initialized with 500 trees.

On the other hand, methods of temporal pattern recognition (such as HMM), which work with spatiotemporal features, are commonly used classifiers for gesture recognition (see e.g. [9, 12]). For a comparison, we applied such classifiers with instantaneous features, extracted from the wrists' trajectories of the segmented guiding strokes. The resulted feature space was 15 dimensions consisting of spatial positions (3D) and directions (as 2D angles), as well as their delta and delta of delta. Using these features, we applied two baseline classifiers on the same subsets: HMM² with hyper-parameters (number of hidden states and number of Gaussian mixtures) selected using the validation set; and Nearest Neighbor (NN) with Dynamic Time Warping (DTW) as distance function.

4.2.1 Human judgment study

Human judgment represents a scoring ground-truth for this task. form the ground-truth of this dataset. For this purpose, we designed an online study, in which videos of gesture performances were shown and the subjects were asked to guess the illustrated objects. The subjects could see the images of all 20 available objects. At each turn, the video of a gesture performance was played in a loop. The subjects were instructed to either click on the image of the object best representing the gesture, or they could click on one of the alternative options "I don't know" or "It's ambiguous". These were chosen in 18% and 3% of all turns, respectively. The study was composed of three sessions, a different video modality shown during each session (see Figure 3). First, subjects guessed while watching all the color videos, then all the skeleton videos and finally all the videos of only the wrists' movements (as two moving dots). The skeleton and wrists videos were rendered in 3D, so that moving a body part towards the camera makes it larger. The bottom half of Table 3 shows the average of the judgment performance of 35 participants.

4.3 Results and discussion

The classification results of the experiments are summarized in Table 3 (using F_1 as evaluation metrics). In combination with SVM, HoGS as features outperform pen gesture features

²Using the Bayes Net Toolbox for Matlab, available at <http://code.google.com/p/bnt/>

Classifier	Descriptor	all gestures in 3DIG dataset			drawing gestures
		Recall	Precision	F_1	F_1
SVM	HoGS	0.61	0.64	0.61	0.77
SVM	pen gesture features [14]	0.54	0.56	0.53	0.57
Bag of Trees	HoGS	0.60	0.47	0.48	0.53
NN+DTW	Instantaneous	0.51	0.45	0.44	0.40
HMM	Instantaneous	0.35	0.36	0.33	0.44
Humans	Color video	0.80	0.70	0.74	0.76
Humans	Skeleton video	0.51	0.33	0.38	0.44
Humans	Wrists video	0.46	0.29	0.34	0.43

Table 3: The classification results of 20 objects (i.e. chance level is 0.05). Our approach (SVM with HoGS features) outperforms the other alternatives and compares favorably to human judgment performance.

from [14] (based on the same subsets of 3DIG dataset). Applying SVM without individual feature selection results in poor classification ($F_1 = 0.36$ for “all gestures” condition), which shows the requirement of individual features for pairs. However, this could not be achieved by a method of Decision Trees (Bag of Trees), which similarly handles pairs of classes individually with respect to specific features.

Instantaneous features, which are used in common gesture recognition approaches, do not achieve any accurate result in this dataset. This can be caused by many intra-class variations at the instantaneous level of the trajectories. Different movement directions and variate ordering together with arbitrary numbers of repetitions make the generalization at this level difficult.

The gestures performed through drawing technique are better classified in all conditions. This difference can be observed even in humans’ judgment, yet very slightly when watching color videos. Our approach (SVM classifier with HoGS features) achieved the best classification scores for both sets of drawing technique and all techniques. Comparing to human performance, these results show that wrists’ positions carry semantic content in large part. Although humans may rely on other information sources that they can only gain in the color videos, our approach (based on wrists’ movements) can best compete with human judgment performance, specially when considering only the drawing gestures. However, the mistaken gesture performances are different. We compared the F_1 scores of each object resulted by our approach and the human judgments. The results of our approach correlate mostly with the judgment based on the skeleton videos, with the correlation coefficient of 0.55, while the color videos and wrists videos achieved correlation coefficients of 0.23 and 0.15 respectively.

5 Conclusion

To address the challenging problem of gesture-based object recognition, we captured the 3D Iconic Gesture (3DIG) dataset. Based on the wrists’ trajectories, we proposed the Histograms of Guiding Strokes as optimal descriptor of iconic gestures. The evaluation experiments showed that HoGS features together with SVM classifiers outperform other approaches and best approximate the human judgment performance. Our approach shows reliable results for automatic gesture-based object recognition, especially when the gesture performance is constrained to drawing technique. As our data analysis shows, considering additional non-verbal information in future work, for instance extracting hand postures from color and depth videos, can lead to more accurate recognition.

References

- [1] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [2] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893, 2005.
- [3] Tamar Flash and Binyamin Hochner. Motor primitives in vertebrates and invertebrates. *Current Opinion in Neurobiology*, 15(6):660–666, 2005.
- [4] Maria Karam and M. C. Schraefel. A taxonomy of gestures in human computer interactions. Technical report, University of Southampton, 2005.
- [5] Stefan Kopp and Ipke Wachsmuth. Synthesizing multimodal utterances for conversational agents: Research articles. *Comput. Animat. Virtual Worlds*, 15(1):39–52, 2004.
- [6] T. Marsh and A. Watt. Shape your imagination: iconic gestural-based interaction. In *Virtual Reality Annual International Symposium, 1998. Proceedings., IEEE 1998*, pages 122–125, 1998.
- [7] David McNeill. *Hand and mind: What gestures reveal about thought*. University of Chicago Press, Chicago, 1992.
- [8] David McNeill. *Gesture and Thought*. University Of Chicago Press, 2007.
- [9] S. Mitra and T. Acharya. Gesture recognition: A survey. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 37(3):311–324, 2007.
- [10] Timo Sowa. Towards the integration of shape-related information in 3-d gestures and speech. In *Proceedings of the 8th international conference on Multimodal interfaces, ICMI '06*, pages 92–99, New York, NY, USA, 2006. ACM.
- [11] Timo Sowa and Ipke Wachsmuth. A model for the representation and processing of shape in coverbal iconic gestures. In K. Opwis and I. Penner, editors, *Proceedings of KogWis05. The German Cognitive Science Conference 2005*, pages 183–188, Basel, Schwabe., 2005.
- [12] Gineke A. ten Holt, Marcel J.T. Reinders, and Emile A. Hendriks. Multi-dimensional dynamic time warping for gesture recognition. In *Thirteenth annual conference of the Advanced School for Computing and Imaging*, 2007.
- [13] Radu-Daniel Vatavu, Laurent Grisoni, and Stefan-Gheorghe Pentiu. Gesture recognition based on elastic deformation energies. In Miguel Sales Dias, Sylvie Gibet, Marcelo Wanderley, and Rafael Bastos, editors, *Gesture-Based Human-Computer Interaction and Simulation*, volume 5085 of *Lecture Notes in Computer Science*, pages 1–12. Springer Berlin / Heidelberg, 2009.
- [14] D.J.M. Willems and R. Niels. Definitions for features used in online pen gesture recognition. Technical report, NICI, Radboud University Nijmegen, 2008.

-
- [15] Don Willems, Ralph Niels, Marcel van Gerven, and Louis Vuurpijl. Iconic and multi-stroke gesture recognition. *Pattern Recognition*, 42(12):3303–3312, 2009.