# Orientation Estimation and Grasp Type Detection of Household Objects for Upper Limb Prostheses with Dynamic Vision Sensor

Siyi Tang*, Rohan Ghosh*, Nitish V. Thakor*, Sunil L. Kukreja*
*Singapore Institute for Neurotechnology (SiNAPSE), National University of Singapore, Singapore 117456.
Email: sunilkukreja.sinapse@gmail.com

***Abstract*—Although the past decade has seen important advances in prosthetic technologies, grasping household objects with an artificial hand still requires significant skill and effort for an amputee to regulating hand behaviour. A solution to this problem is to automate the process by using vision sensors that determine the object's orientation and optimal grasp procedure. In this paper, we use a neuromorphic dynamic vision sensor (DVS) to assist amputees with object grasping. Event-driven sensors such as the DVS have gained popularity in recent years as an alternative to conventional frame-based sensors due to their low-power consumption and low-latency. Here, we use event data from a DVS to find a grasp-appropriate orientation for the object and subsequently its optimal grasp type. Our estimation technique exploits general assumptions such as object symmetry and grasp preference to be along the smallest major dimension of an object. The grasp type is determined through a combination of multiple convolutional neural network (CNN) classifiers. We evaluated our grasp estimation methodology on a set of 20 household objects. The results of this study show that 96.25% of the estimated orientations were within ±10° of the actual orientations. In addition, our grasp detection method yielded a 99.47% accuracy on unseen object classes.**

## I. Introduction

In recent years the prosthetic market has been growing rapidly due to many technological advances [1]. With modern and increasingly sophisticated upper limb prostheses amputees have the ability to lead normal lives. Approaches that implement electrode recordings of electromyographic (EMG) activity for grasp control with prosthetic hands have revolutionized the functionality and adoption of artificial limbs [2]–[4]. However, these approaches are invasive and prone to both inter-subject and intra-subject variabilities [5]. For these reasons, less user intensive approaches are preferred [6]. Therefore, an effective non-invasive approach imposing minimal cognitive load on amputees to facilitate grasping tasks remains an open problem. We propose a vision-based framework for grasping of household objects with a dynamic vision sensor (DVS) mounted on an upper limb prosthesis.

The DVS is an asynchronous, event-based vision sensor inspired by the human retina [7]. It operates at a high temporal resolution of $1\mu$s with a $15\mu$s latency and contains a pixel array of size $128 \times 128$. In contrast to conventional vision sensors that transmit the whole image at fixed frame rates, the pixels of DVS asynchronously transmit information associated with visual changes at precise times. As a result, the data transmitted by a DVS is much less redundant, leading to a significant reduction in computational burden and lower power consumption. This allows for rapid processing and classification for tasks such as object recognition. Other applications of DVS include fast visual tracking [8], visual flow computation [9] and traffic monitoring [10].

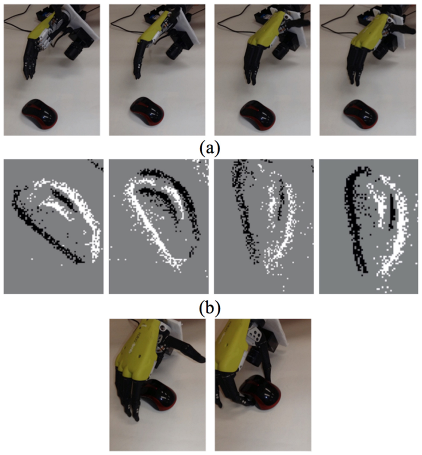

Fig. 1: An example demonstration of the orientation estimation and grasp-type detection procedures for a computer mouse object. (a) The prosthetic hand is rotated to the estimated grasp-appropriate orientation. (b) The DVS output corresponds to each orientation in (a). (c) The prosthetic hand changes to the estimated tripod grasp posture and executes the grasp.

We used an i-Limb prosthetic hand for our experiments [11]. To assist with grasping tasks we initially estimate the relative orientation of an object then use this information to determine a grasp-appropriate positioning. Moreover, we directly work with grasp type detection instead of inferring object category, which enables the algorithm to generalize well across unseen classes. Previously, event-based sensors have been used for human orientation detection with a convolutional neural network (CNN) [12]. We do not use a CNN based orientation estimation approach since it's performance is dependent on the dataset used for training.

In this paper we assume the following. (1) Household objects exhibit a high degree of symmetry [13]. (2) Grasping is most suitable along the smallest dimension of an object, as observed with human subjects [14]. (3) Background scene is absent of any distractors. (4) Power, pinch and tripod grasp gestures of the i-Limb provide the most optimal grasps necessary for most household objects. This is in contrast to utilizing only power and pinch grasps [15]. (5) Amputees approach objects from a consistent direction, which facilitates optimal grasp.

Fig. 1 shows the different steps involved in an example grasping routine of the i-Limb using our algorithm. Fig. 1 (a) shows the hand rotating to the grasp-appropriate orientation estimated by our algorithm. Fig. 1 (b) displays the DVS output corresponding to each orientation in Fig. 1 (a). Fig. 1 (c) shows the prosthetic hand adjusting it's grasp method based on the corresponding grasp-type detected (tripod grasp in this instance).

This paper is an extension of our previous work where object orientation was estimated through a set of category-specific supervised convolutional neural networks (CNN) and, therefore, was limited by the extent of the object corpus chosen [10]. In this work, we propose a more generalized algorithm to estimate orientation and implement multiple CNN classifiers to directly achieve grasp type recognition. Notice that this methodology avoids a separate object recognition step. Our main contributions can be summarized as follows.

- Our algorithm to estimate orientation achieves good performance without prior knowledge of the object class. It is also superior to a principal components analysis (PCA, [16]) approach to orientation estimation.
- Our classifier for grasp-type estimation generalizes well for unseen object catagories, achieving a lower error rate than a k-nearest neighbour based classification [17].

The paper is organized as follows. Section II presents our combined orientation estimation and grasp type detection algorithm. Section III describes our experimental paradigm while Section IV presents the results of our algorithm. Lastly, Section V concludes the paper and provides guidance for future work.

## II. Methods

The algorithm presented in this section consists of two parts, namely (i) orientation estimation and (ii) grasp type detection.

### A. Spatial-temporal ROI Estimation

To reconstruct frames from recorded spike-events we employed a constant event window to obtain the temporal region of interest (ROI) [10]. In addition, a rectangular spatial ROI was estimated according to a dynamic bounding box that enclosed 98% of the events on every side (up, down, left and right) of a centroid. The spatial-ROI is then resized to $60 \times 60$ before it is input as a image for the steps that follow. All the events outside of the ROIs are not utilized by our algorithm.

### B. Orientation Estimation

Fig. 2 shows examples of objects in our classification group with varying degrees of symmetry. The polarity information is not used by our approach since it is dependent on motion direction. Let $I$ denote an image obtained after the filtering methods described in the previous section. A smoothed version of $I$, denoted $I_s$, is obtained through the distance transform function $f_d$ as

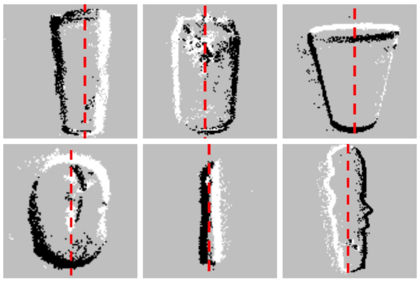

Fig. 2: Examples of objects as observed by a DVS with varying degrees of symmetry. White/black pixels represent an increase/decrease in polarity. The red vertical line represents the geometric symmetry axis.

$$I_s = e^{\frac{-f_d(I)^2}{\sigma^2}} \tag{1}$$

where $\sigma$ is a parameter that was set to a fixed value of 10. A very high value of $\sigma$ would over-smooth the image, whereas a very low a value would result in almost no smoothing. The centroid of the events contained in $I_s$ is estimated and a fixed number of axes $\{\theta_i\}_{i=1}^{N}$ passing through the centroid at equispaced angles are generated. With a spacing of 5 degrees between consecutive axes, a total of $N = 36$ axis were generated for each image. For each axis with an orientation $\theta_i$ the planar reflective symmetry transform (PRST, [18]), $S(\theta_i)$ is computed as

$$S(\theta_i) = \frac{\sum_x \sum_y I_s(x,y)\Gamma_{I_s,\theta_i}(x,y)}{\sum_x \sum_y I_s(x,y)^2} \tag{2}$$

where $\Gamma_{I_s,\theta_i}$ is obtained by reflecting $I_s$ about the axis passing through the centroid and at an orientation of $\theta_i$. $I_s(x,y)$ and $\Gamma_{I_s,\theta_i}(x,y)$ represent the values of $I_s$ and $\Gamma_{I_s,\theta_i}$ at pixel location $(x,y)$. The axes corresponding to the local maxima of the symmetry values are considered to be the central axes of an object. For each symmetry axis, $\theta_{sym}(k)$, a mirror matrix $W_k$ is computed as

$$W_k = I_s \circ \Gamma_{I_s,\theta_{sym}(k)} \tag{3}$$

where $\circ$ represents the Hadamard product. A Sobel vertical edge detector is applied to the mirror matrix $W_k$ with a fixed threshold to eliminate cases where edges parallel to the symmetry axis have insufficient events [19]. For each symmetry axis a weighted axis-normal distance $W_D(k)$ is computed as

$$W_D(k) = \frac{\sum_x \sum_y W_k(x,y) Dist(x,y,\theta_{sym}(k))}{\sum_x \sum_y W_k(x,y)} \tag{4}$$

where $Dist(x,y,\theta_{sym}(k))$ is the perpendicular distance between each pixel $(x,y)$ in $W_k$ and the symmetry axis at an orientation of $\theta_{sym}(k)$. Since grasping an object is most suitable along the least dimension, a grasp-appropriate orientation is assigned as the orientation of symmetry axis $\theta_{sym}(k)$ having a minimal value of $W_D(k)$. Fig. 3 illustrates the steps involved in computing the PRST $S$ and the weighted distance $W_D$ for a can. Fig. 3(a) shows that this object has two symmetry axes. Fig. 3(b) displays the mirror matrix $W_k$ corresponding to the symmetry axis which is $-25°$ anti-clockwise from the vertical axis in Fig. 3(a).

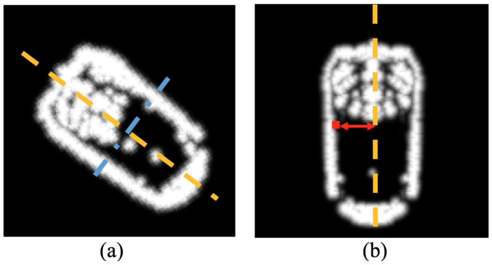

Fig. 3: (a) Smoothed image, $I_s$, of a can object. Two local maxima of symmetry values in $S(\theta)$ are depicted corresponding to the two symmetry axes shown. (b) Mirror matrix $W_k$ corresponding to $-25°$ symmetry axis. A weighted distance $W_D$ is computed as the perpendicular distance between each non-zero pixel and symmetry axis.

### C. Grasp Type Detection

Experimentally we observed that the power, pinch and tripod grasp are most suitable to common household objects. Furthermore, some object classes may have more than one potential grasp type. To account for these cases we assigned multiple grasp types to such objects. Table I shows the multiple grasp-type assignments. An estimate of grasp is computed by training three CNNs classifiers, one for each type. Inputs to the CNNs are back-rotated data based on orientations estimated from our algorithm. Each CNN has two output classes indicating whether the corresponding grasp type is applicable to the input data or not. This is equivalent to training in a one versus all manner.

TABLE I: Grasp types and object classes associated with each.

| Grasp Types | Object Classes |
|---|---|
| Power & Tripod | Bottle, Can, Container Tube, Drinking Glass, Jar, Wallet, Pan, Remote Controller, Umbrella |
| Pinch & Tripod | Mobile Phone, Computer Mouse, Scissors, Comb, Fork, Spoon |
| Pinch | Marker, Thumb Drive, Pen, Toy Car |
| Tripod | Table Tennis Bat |

## III. Experimental Setup

Our experimental setup is illustrated in Fig. 4 (a).

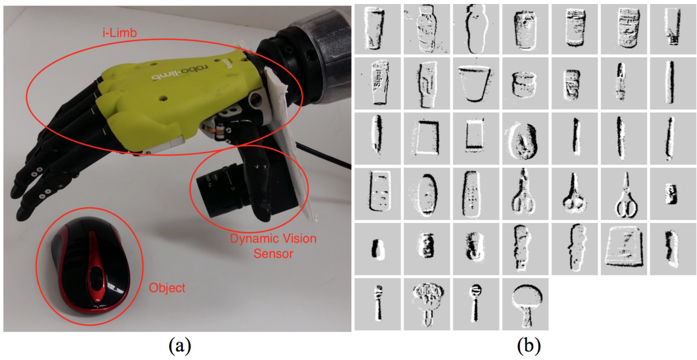

Fig. 4: (a) Experimental setup for data acquisition. The DVS was mounted on the wrist of the i-Limb prosthetic hand. The prosthetic hand was rotated from -90° to 90° with respect to the grasp-appropriate orientation. (b) Object instances (20 object classes) in our dataset, grouped together by class.

Data was recorded with a DVS attached on the wrist of the prosthetic hand, allowing it to rotate along with the wrist. Objects were placed against a white background to create a non-cluttered background. After a grasp-appropriate orientation was defined for each object, the ground truth for the orientation estimation algorithm was obtained with the i-Limb wrist at a grasp-appropriate orientation. For each object the prosthetic hand was rotated from -90° to 90° with respect to a grasp-appropriate orientation. Orientation estimation is consider correct if the difference between the estimated orientation and actual orientation is no larger than 10°.

Fig. 4 (b) shows all the object instances in our dataset. Data for 20 common household object classes were acquired. Each object class had 2-4 cases with data recorded for two poses for each (front and back). The dataset obtained was used to test our grasp detection algorithm where each image frame was rotated by an angle approximated by the orientation estimation technique. The images were then down-sampled to a dimension of $64 \times 64$, unrolled and padded to form inputs to the CNNs.

To assess the accuracy of our grasp type detection algorithm's ability to generalize across unseen objects classes, the CNNs were only trained with 15 object categories and tested on individual cases of the five remaining classes (fork, pan, comb, table tennis bat and spoon). For orientation estimation, we compare our method with a PCA based technique. The PCA approach gives the orientation along the first principal component of the events. Grasp type classification is considered correct if the estimated grasp is one of the assigned types.

## IV. Results and Discussions

### A. Orientation Estimation Performance

The results of our study to assess the robustness of our orientation estimation algorithm is provided in Table II. The table shows the estimation accuracies of the 15 object classes with 2,000 sampled events per image frame. The performance of our approach is shown next to that of the PCA counterpart. In the table our orientation estimation method is denoted as PRST+NDIST (NDIST indicates axis-normal distance $W_D$). The grasp type classification method is represented as $\mathrm{CNN}_3$. We found that 2,000 events for temporal ROI estimation were adequate to capture object edges. Our assumption that the weighted distance $W_D$ is minimized along a grasp-appropriate orientation holds for most household objects we encountered. We found that our algorithm outperformed the PCA technique for most object categories and overall shows an $\approx$ 8% improvement (96.25% versus 88.54%). As expected PCA fails to perform well for more square objects since the orientation axis passes through the diagonal in such cases. Fig. 5 shows an example of where PCA typically fails and our algorithm does not.

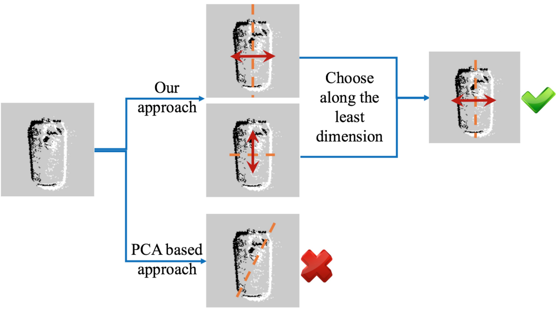

Fig. 5: A wallet object illustrates a cases where PCA estimated a diagonal to be the orientation axis. Our algorithm correctly determined a grasp-appropriate orientation.

### B. Grasp Type Detection Performance

Table III shows the performance of our grasp type detection algorithm next to that of the k-nearest neighbor classifier. We compared our technique with $k = 2$ since it performed the best for the k-NN classifier. Our method outperformed k-NN on both the seen and unseen data. Furthermore, our algorithm shows consistent performance for the unseen classes. However, for k-NN a significant reduction in accuracy was observed. The

TABLE II: Performance and comparison of the orientation estimation algorithm.

| Object Class | PRST+NDIST | PCA | Object Class | PRST+NDIST | PCA |
|---|---|---|---|---|---|
| Bottle | 97.80% | 94.57% | Comb | 100.00% | 100.00% |
| Can | 95.30% | 99.79% | Fork | 100.00% | 100.00% |
| Container Tube | 100.00% | 100.00% | Table Tennis Raquet | 96.00% | 94.67% |
| Drinking Glass | 100.00% | 81.48% | Pan | 98.50% | 100.00% |
| Jar | 77.20% | 67.93% | Spoon | 100.00% | 100.00% |
| Marker | 100.00% | 100.00% | Toy Car | 89.50% | 86.32% |
| Mobile Phone | 100.00% | 74.01% | Remote Controller | 100.00% | 95.85% |
| Computer Mouse | 80.60% | 35.48% | Scissors | 100.00% | 100.00% |
| Pen | 100.00% | 100.00% | Thumb Drive | 93.50% | 93.50% |
| Wallet | 96.60% | 47.13% | Umbrella | 100.00% | 100.00% |

TABLE III: Performance of grasp type detection algorithm.

|  | $CNN_3$ | k-NN (k=2) |
|---|---|---|
| Seen Classes | 98.36 | 95.11 |
| Unseen Classes | 99.47 | 90.93 |

a performance reduction using k-NN validates that the unseen data was sufficiently different from the seen data.

## V. Conclusion and Future Work

This work presented a combined orientation estimation and grasp type detection algorithm for common household objects using visual data from an event-based sensor placed on the wrist of an i-Limb prosthetic hand. The orientation estimation algorithm is capable of estimating the orientation without prior knowledge of the object class. Moreover, the CNN-based classifiers for grasp type detection were able to generalize well across unseen object categories. However, in this work we assumed a non-cluttered background, which is unlikely in real applications. Hence, our future work will focus on extending this algorithm for cluttered settings using approaches that facilitate real-time segmentation. Furthermore, since our orientation estimation technique is limited to handling global symmetries, extensions to this work will allow for accurately estimation of objects containing only local symmetries.

## Acknowledgment

This work is supported by a NPRP grant from the Qatar National Research Fund under the grant No. NPRP 7-673-2-251. The statements made herein are solely the responsibility of the authors.

## References

[1] D. A. Zlotolow and S. H. Kozin, “Advances in upper extremity prosthetics,” *Hand Clinics*, vol. 28, no. 4, pp. 587–593, 2016/08/29.

[2] L. R. Hochberg, M. D. Serruya, G. M. Friehs, J. A. Mukand, M. Saleh, A. H. Caplan, A. Branner, D. Chen, R. D. Penn, and J. P. Donoghue, “Neuronal ensemble control of prosthetic devices by a human with tetraplegia,” *Nature*, vol. 442, no. 7099, pp. 164–171, Jul. 2006.

[3] L. R. Hochberg, D. Bacher, B. Jarosiewicz, N. Y. Masse, J. D. Simeral, J. Vogel, S. Haddadin, J. Liu, S. S. Cash, P. van der Smagt, and J. P. Donoghue, “Reach and grasp by people with tetraplegia using a neurally controlled robotic arm,” *Nature*, vol. 485, no. 7398, pp. 372–375, May 2012.

[4] G. C. Matrone, C. Cipriani, M. C. C. Carrozza, and G. Magenes, “Real-time myoelectric control of a multi-fingered hand prosthesis using principal components analysis.” *Journal of neuroengineering and rehabilitation*, vol. 9, pp. 40+, Jun. 2012.

[5] R. Araujo, M. Duarte, and A. Amadio, “On the inter- and intra-subject variability of the electromyographic signal in isometric contractions,” *Electromyography and clinical neurophysiology*, vol. 40, no. 4, pp. 225–229, June 2000.

[6] C. Cipriani, F. Zaccone, S. Micera, and M. C. Carrozza, “On the shared control of an emg-controlled prosthetic hand: analysis of user–prosthesis interaction,” *IEEE Transactions on Robotics*, vol. 24, no. 1, pp. 170–184, 2008.

[7] P. Lichtsteiner, C. Posch, and T. Delbruck, “A 128 × 128 120db 15$\mu$s latency asynchronous temporal contrast vision sensor,” *IEEE Journal of Solid-State Circuits*, vol. 43, no. 2, pp. 566–576, Feb 2008.

[8] X. Lagorce, C. Meyer, S.-H. Ieng, D. Filliat, and R. Benosman, “Asynchronous event-based multikernel algorithm for high-speed visual features tracking,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 8, pp. 1710–1720, Aug 2015.

[9] R. Benosman, C. Clercq, X. Lagorce, S.-H. Ieng, and C. Bartolozzi, “Event-based visual flow,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 2, pp. 407–417, Feb 2014.

[10] R. Ghosh, A. Mishra, G. Orchard, and N. Thakor, “Real-time object recognition and orientation estimation using an event-based camera and cnn,” in *Biomedical Circuits and Systems Conference (BioCAS), 2014 IEEE*, Oct 2014, pp. 544–547.

[11] Rslsteeper bebionic. [Online]. Available: http://bebionic.com

[12] J. Perez-Carrasco, B. Zhao, C. Serrano, B. Acha, T. Serrano-Gotarredona, S. Chen, and B. Linares-Barranco, “Mapping from frame-driven to frame-free event-driven vision systems by low-rate rate coding and coincidence processing–application to feedforward convnets,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 11, pp. 2706–2719, Nov 2013.

[13] A. Saxena, J. Driemeyer, and A. Y. Ng, “Learning 3-d object orientation from images,” in *Proceedings of the 2009 IEEE International Conference on Robotics and Automation*, ser. ICRA’09. Piscataway, NJ, USA: IEEE Press, 2009, pp. 4266–4272.

[14] T. Feix, I. M. Bullock, and A. M. Dollar, “Analysis of human grasping behavior: Object characteristics and grasp type,” *IEEE Transactions on Haptics*, vol. 7, no. 3, pp. 311–323, July 2014.

[15] M. Gardner, R. Woodward, R. Vaidyanathan, E. Bürdet, and B. C. Khoo, “An unobtrusive vision system to reduce the cognitive burden of hand prosthesis control,” in *13th International Conference on Control, Automation, Robotics & Vision (ICARCV), 2014n*, Dec 2014, pp. 1279–1284.

[16] J. Shlens, “A tutorial on principal component analysis,” *CoRR*, vol. abs/1404.1100, 2014.

[17] T. Cover and P. Hart, “Nearest neighbor pattern classification,” *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, January 1967.

[18] J. Podolak, P. Shilane, A. Golovinskiy, S. Rusinkiewicz, and T. Funkhouser, “A planar-reflective symmetry transform for 3d shapes,” *ACM Trans. Graph.*, vol. 25, no. 3, pp. 549–559, Jul. 2006.

[19] R. Maini and H. Aggarwal, “Study and comparison of various image edge detection techniques,” *International journal of image processing (IJIP)*, vol. 3, no. 1, pp. 1–11, 2009.