Proceedings of the 2007 IEEE/RSJ International
Conference on Intelligent Robots and Systems
San Diego, CA, USA, Oct 29 - Nov 2, 2007

Feature selection for grasp recognition from optical markers

Lillian Y. Chang, Nancy S. Pollard, Tom M. Mitchell, and Eric P. Xing

Abstract— **Although the human hand is a complex biomechanical system, only a small set of features may be necessary for observation learning of functional grasp classes. We explore how to methodically select a minimal set of hand pose features from optical marker data for grasp recognition. Supervised feature selection is used to determine a reduced feature set of surface marker locations on the hand that is appropriate for grasp classification of individual hand poses. Classifiers trained on the reduced feature set of five markers retain at least 92% of the prediction accuracy of classifiers trained on a full feature set of thirty markers. The reduced model also generalizes better to new subjects. The dramatic reduction of the marker set size and the success of a linear classifier from local marker coordinates recommend optical marker techniques as a practical alternative to data glove methods for observation learning of grasping.**

I. Introduction

The human hand has amazing flexibility as a manipulator, but the complex movement can be challenging to measure, model, and imitate. Manually-programming manipulation tasks for a multi-fingered robotic system can be time-consuming and result in inflexibility with respect to specific task parameters. The cost of adding new behaviors could be significantly reduced if the robot system has the ability to learn from observing a human teacher. In observation learning, examples provided by a human demonstrator are used to automatically synthesize grasps for a robot manipulator [1, 2]. The observation method should provide the system with a minimum set of features that can represent the type of grasp performed by the demonstrator. It is also desirable that the demonstrator be able to perform the action as naturally as possible to provide a good quality example.

In previous approaches to grasp recognition for observation learning, the human demonstrator wears a data glove while performing the example grasp [2–5]. Sensors attached to the data glove may measure the finger joint angles or the position of selected points on the hand. The direct measurement of the glove configuration allows for the grasp features to be detected easily. However, the data glove obstructs the demonstrator's contact with the object and may prevent a natural grasp. Additionally, the accuracy of the measured joint angles depends on how well the glove fits the individual's hand, and this aspect creates difficulties in particular for demonstrators with smaller hands.

This work was supported by the National Science Foundation (CCF-0343161, IIS-0326322, ECS-0325383, CNS-0423546, and CCF-0702443). L. Y. Chang is supported by the National Science Foundation Graduate Research Fellowship.

L. Y. Chang and N. S. Pollard are with The Robotics Institute, and T. M. Mitchell and E. P. Xing are with the Machine Learning Department, at the School of Computer Science, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213, USA {lillianc,nsp,Tom.Mitchell,epxing}@cs.cmu.edu

Vision systems may also be used to observe the demonstrated grasp. Several researchers have developed hand pose estimation methods for vision-based gesture interfaces [6–9]. These systems observe hand pose without a data glove and allow for natural motion. Applying a vision strategy to grasping observation, though, is challenging due to occlusion of the fingers by the grasped object which complicates segmentation of the hand from the rest of the image.

Another observation technique is marker-based motion capture, where optical markers attached to the hand are used to track the observed movement in a controlled environment with calibrated cameras. The addition of surface markers simplifies the detection of key interest points, without affecting the natural grasping motion nor obstructing contact with the object as data gloves may. Marker-based capture of hand pose has been used previously for computer animation applications, where a full set of markers tracking all finger segments is used to measure example grasps [10, 11]. Reconstructing the complete hand pose can still be difficult because of incorrect marker correspondences and occlusions that result from using a full set of markers to track all the finger segments.

We propose to use a reduced marker protocol to simplify the capture procedure and describe the hand configuration in a low-dimensional space. This is based on the idea, suggested by previous studies, that the recognition of a discrete set of functional grasps may not require measuring the complete configuration of the hand. The work of Santello et al. [12] and Mason et al. [13] found that mimed reach-to-grasp motions can be represented by just a few principal components in the joint angle space. Dejmal and Zacksenhouse [4] also use principal component analysis for recognizing discrete classes of manipulation movements from data glove input. However, using principal components to find the dominant synergies in the input feature space requires that all input degrees of freedom be measured and does not allow for simplification of a marker-based protocol. In computer animation, Chai and Hodgins [14] and Liu et al. [15] build locally-linear models of multiple behaviors to reconstruct full-body motion from a small set of control marker inputs. Liu et al. [15] select a reduced marker set from a full optical marker set as we do, but the markers are chosen as the features which maximize the variance in the lower-dimensional space such that they can be used to reconstruct the full-dimensional representation.

Previous work in the robotics community has investigated grasp recognition using non-linear classification models. Bernardin et al. [16] classify the entire reach-to-grasp movement trajectory by training a hidden Markov model (HMM)

1-4244-0912-8/07/$25.00 ©2007 IEEE.

on the 16 data glove joint angles and 13 tactile sensors on the glove. Similarly, Ekvall and Kragic [3] also used a HMM to classify grasping sequences observed from data glove measurements of four positions on the back of the hand and the fingertips. Another approach by Moussa and Kamel [17] uses artificial neural networks for predicting manufacturing grasps from the taxonomy proposed in Cutkosky [18] based on features of the object and task rather than a demonstrated hand pose. We find that a linear classifier is sufficient to predict grasps from local marker coordinate positions describing the hand pose for demonstrated grasps of several objects.

The following sections present how we determine an appropriate reduced marker protocol for grasp recognition. Supervised feature selection is first used to identify a subset of features from a full set of marker positions. The grasp classifier is then trained on the reduced feature set resulting from the feature selection experiments. We evaluate the method on grasp data which includes examples from multiple demonstrators on multiple objects.

II. Problem Definition

This study investigates feature selection of the surface marker positions for the purpose of grasp classification. The classification goal is to predict the grasp class for a single hand pose given the positions of a specified set of markers.

The input feature vector $\mathbf{x}$ for a marker set with M markers is a $3M$ column vector consisting of the three-dimensional marker positions which represents the hand pose at a single time sample. The marker positions are expressed with respect to a local coordinate system attached to the back of the hand (Fig. 1), such that the description of the hand pose is invariant to the orientation and position of the hand in the external coordinate system. The classification output is the single predicted grasp Y, which can take one of K possible grasp class values from the set $\{y_1, y_2, \ldots, y_K\}$ (Fig. 2).

The purpose of feature selection is to determine a subset of markers that will support accurate decoding of the grasp class.

III. Method

A. Grasp classification model

Our approach uses a linear logistic regression classifier for evaluating candidate marker sets in supervised feature selection and then for predicting grasp from the final trained model. Although the fingers exhibit nonlinear kinematics relative to the palm, the anatomic constraints on hand motion will limit each surface marker to a continuous region of reachable positions. The location of a single marker will be further constrained to a sub-region within the overall reachable space depending on the type of grasp. Thus, we believe that hand poses as represented by local marker coordinates will be compactly clustered according to the grasp class, such that a classifier with linear decision boundaries can be successful for predicting grasp types from surface marker data.

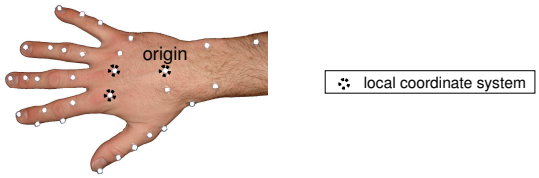


Fig. 1. A total of 31 markers are attached to the hand. Three markers on the rigid portion of the back of the hand define the local coordinate system for the marker positions. The origin marker is excluded in the feature selection tests, such that the full marker set consists of the remaining 30 markers.

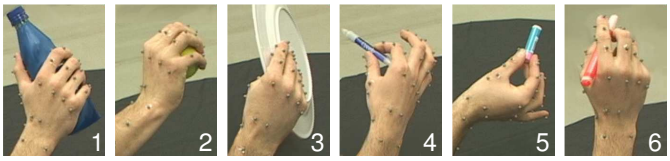


Fig. 2. Six different functional grasps were considered for grasp classification: (a) cylindrical grasp, (b) spherical (or circular) grasp, (c) lumbrical grasp, (d) two-finger pinch (or pad-to-pad) grasp, (e) tripod (or three-jaw chuck), and (f) lateral tripod grasp.

We give a brief overview of multiclass linear logistic regression, which is used as our baseline classifier. Please see, e.g., Bishop [19] for more details.

In multiclass logistic regression, the posterior probabilities of a class y_k given the input features are modeled by the softmax function:

$$p_k(\mathbf{x}, A) = p(Y = y_k|\mathbf{x}) = \frac{\exp(\mathbf{a}_k^{\mathrm{T}}\mathbf{x})}{\sum_{j=1}^{K}\exp(\mathbf{a}_j^{\mathrm{T}}\mathbf{x})}, \quad (1)$$

where the vectors in the matrix $A = (\mathbf{a}_1, \ldots, \mathbf{a}_K)$ are the model weights for the K classes.

The values of the weights are determined by maximum conditional likelihood estimation from a training set (X, Y). For a data set with N independent and identically-distributed examples, the log of the conditional data likelihood is

$$l(A) = \ln p(Y|X, A) = \sum_{n=1}^{N}\sum_{k=1}^{K}\delta(Y_n = y_k)\ln p_k(\mathbf{x}_n, A). \quad (2)$$

Gradient ascent can be used to find weights which maximize the conditional data log likelihood, where the gradient of the log likelihood with respect to the weight vectors is

$$\nabla_{a_j} l(A) = \sum_{n=1}^{N}(\delta(Y_n = y_j) - p_j(\mathbf{x}_n, A))\mathbf{x}_n. \quad (3)$$

Once the set of weights A is determined from the training set, the grasp is predicted for a given hand pose by selecting the class with the maximum posterior probability:

$$y_{pred} \leftarrow \arg\max_{y_k} p(Y = y_k|\mathbf{x}). \quad (4)$$

The ratio of the maximum class probability to the second highest class probability provides a confidence measure for the prediction:

$$c(\mathbf{x}) = \frac{\max_{y_k} p(Y = y_k|\mathbf{x})}{\max_{y_k \neq y_{pred}} p(Y = y_k|\mathbf{x})}. \quad (5)$$

B. Supervised feature selection

Given a baseline classification model, we then wish to select a subset of input features which represents the hand pose in a lower-dimensional space. To avoid considering the exponential number of possible feature sets, sequential wrapper algorithms evaluate the addition or removal of a single feature at a time for locally-optimal feature selection [see, e.g., 20]. In contrast to filter approaches, where a subset is selected based on individual feature scores, wrapper algorithms score possible feature sets and thus model the interaction between features with respect to predicting the target class. In this work, we will consider two versions of greedy sequential wrappers. The forward method adds features incrementally to a reduced feature set, and the backward method discards features incrementally from a larger set of available features (Figs. 3 and 4).

We make one modification to these standard algorithms for marker-based techniques. Instead of evaluating individual position coordinates, our methods will score the features in subsets of three which correspond to the three position coordinates of one marker. This reflects the usage in the target application, where the goal is to reduce the number of markers in the protocol, rather than simply predicting grasp from a set of features that may include only one or two of the three available coordinates from a single marker.

Thus at each stage of sequential feature selection, every candidate marker consisting of a subset of three features is scored conditioned on the current selected marker set. The scoring criterion for our implementation is the classifier prediction accuracy estimated from cross-validation on the training set. In n-fold cross-validation, the training data is divided into n validation data sets. For each validation set, a classifier is trained from the examples not included in the validation set, and then the classifier is tested on the examples in the validation set. The cross-validation accuracy for a given model size is the average accuracy of the n classifiers, weighted by the number of examples in each validation set. Although there are a number of possible scoring criterion [19, 20], the classifier accuracy most directly relates to the goal of selecting an optimal marker subset for grasp classification.

The results of the sequential feature selection determine a subset of markers whose coordinates comprise the reduced feature set. Then a final classifier model is trained from a specified set of training examples and evaluated on held out test data.

Input: M, desired size of final marker set
Input: X, full set of available features (triplets of marker coordinates)
Output: S, reduced marker set of size M
// let $X(S)$ denote the set of position coordinates of markers in set S

S ←empty set
while $size(S) < M$ **do**
  for all $m \notin S$ **do**
    // for each remaining marker from the available set
    score$(m|S)$ ←accuracy of classifier $Y = f(X(S \cup m))$
    // use cross-validation accuracy to score the combined set of the current set with the candidate marker
  end for
  // select the best marker to add to the current set
  $s \leftarrow \arg\max_{m \notin S}$ score$(m|S)$
  $S \leftarrow S \cup s$
end while
return S

Fig. 3. Algorithm for forward selection of marker features.

Input: M, desired size of final marker set
Input: X, full set of available features (triplets of marker coordinates)
Output: S, reduced marker set of size M
// let $X(S)$ denote the set of position coordinates of markers in set S

S ←full set of available markers
while $size(S) > M$ **do**
  for all $m \in S$ **do**
    // for each marker in the current selected set S
    score$(m|S)$ ←accuracy of classifier $Y = f(X(S - m))$
    // use cross-validation accuracy to score the current set excluding the candidate marker
  end for
  // select the best marker to remove from the current set
  $s \leftarrow \arg\max_{m \in S}$ score$(m|S)$
  $S \leftarrow S - s$
end while
return S

Fig. 4. Algorithm for backward selection of marker features.

IV. Experimental validation and results

A. Grasp data set

Example grasps were measured using a full marker protocol with markers attached to all finger segments of the demonstrator's right hand. The three-dimensional positions of 31 markers were recorded during the grasping action (Fig. 1). Three of these markers define a local coordinate system on the rigid portion of the back of the hand [21]. The marker used as the origin is excluded from the feature vector because its local position is invariant. A single example representing the hand pose at one time sample is thus a 90-dimensional vector consisting of the local coordinates of 30 markers.

Each example was labeled as one of six grasp types, selected from functional grasps for daily living [22] (Fig. 2). Power grasps, characterized by large contact areas with the object, included cylindrical grasp and spherical (or circular) grasp. In addition, lumbrical grasp is used to hold flat or rectangular objects [22]. Precision grasps, for fine manipulation by the fingertips, included two-finger pinch (or pad-to-pad) grasp, tripod (or three-jaw chuck) grasp, and lateral tripod grasp. The lateral tripod grasp is often used by humans for holding writing or eating utensils [22].

The data set consists of grasps demonstrated on 46 objects, which may each be grasped in multiple ways. Examples are divided into two sets according to object (Table I). Object set A consists of 38 objects corresponding to a total of 88 object-grasp pairs, and object set B consists of 8 objects corresponding to 19 object-grasp pairs. For each object-grasp pair, multiple examples with varying hand configuration and contact points were collected from the demonstrator.

Example grasps were collected from multiple demonstrators. Subjects 1 and 2 demonstrated grasps on all objects in object sets A and B. Additional test examples were recorded from Subject 3, who demonstrated grasps for object set B.

B. Feature selection results

The training set used for feature selection consisted of grasp examples from object set A performed by both subjects 1 and 2. Forward selection starts with an empty set of markers, and each iteration of the algorithm augments the current feature set by the marker whose inclusion results in the best classifier accuracy. Backward selection starts with the full set of 30 markers, and each iteration removes the marker whose omission maintains the highest accuracy.

Both wrapper algorithms were evaluated using two-fold cross validation on the training data set. Fig. 5 shows the cross-validation accuracy of the two wrapper methods for different marker set sizes. The prediction accuracy increases rapidly for a few number of markers, but there is only marginal increase for each added marker beyond five to ten markers. The plateau in the performance suggests that the number of markers could be reduced dramatically while retaining correct predictions for a large portion of examples. Using the full 30-marker set resulted in a maximum accuracy of 91.5%, but with only five markers the model could still correctly predict 86% of the grasp examples.

The cross-validation accuracy for forward selection and backward selection differ at most by 0.5% for each marker set size, but the specific rankings of the markers were not identical. For most marker set sizes, the accuracy from backward selection was higher than that from forward selection. We thus chose a final reduced marker set with five markers based on the backward selection results (Fig. 6). Note that three of the selected markers are not positioned on the fingertips. This should reduce the frequency of marker occlusions, which often occur for fingertip markers when the fingers wrap around the grasped object.

TABLE I

OBJECT-GRASP PAIRS FOR THE COLLECTED EXAMPLES IN THE TRAINING AND TEST DATA SETS.

		Grasp class					
		cyl	sph	lum	pin	tri	lat
Set	Object	1	2	3	4	5	6
A	Mug	×				×	
	Honey container	×				×	
	Mallet	×					
	Spray bottle	×					
	Oats can	×				×	
	Sunscreen	×			×	×	
	Phone	×					
	Milk jug	×				×	
	Film container	×			×	×	
	Battery	×			×	×	
	Water bottle	×				×	
	Juggling pin	×					
	Tennis ball		×		×	×	
	Foam ball		×		×	×	
	Softball		×		×	×	
	Puzzle cube		×		×	×	
	Jingle bell		×		×		
	Large egg		×		×	×	
	Medium egg		×		×	×	
	Small bowl		×			×	
	Gyro ball		×		×	×	
	Halogen bulb		×			×	
	Compact disc			×	×		
	Large plate			×	×		
	Binder			×	×		
	Calculator	×		×	×		
	Wallet	×		×	×		
	Cassette tape	×		×	×		
	Poker card	×		×	×		
	Paper			×	×		
	Key				×		
	Coin				×	×	
	Mouthwash cap				×	×	
	Thin pencil				×	×	×
	Highlighter				×	×	×
	Eraser stick				×	×	×
	Fork				×		×
	Knife				×		×
B	Chapstick	×			×	×	
	Jelly jar	×				×	
	Small egg		×		×	×	
	Book			×			
	Plastic card	×		×	×		
	Button pin				×	×	
	Thick pen				×	×	×
	Spoon				×		×

C. Evaluation of reduced marker set

The reduced marker set determined from sequential feature selection is evaluated for both single demonstrator and multiple demonstrator settings. For the single demonstrator setting, two final classifiers are trained for subject 1 and subject 2 separately, using the reduced feature set and grasp examples from object set A. In the multiple demonstrator setting, a final classifier is trained on the combined examples of subject 1 and subject 2, for grasps of object set A.

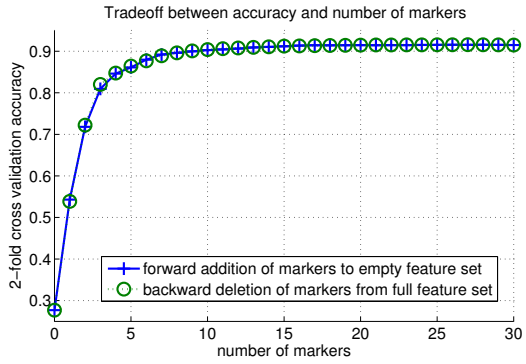


Fig. 5. Feature selection results from forward and backward wrapper methods. The plateau in the two-fold cross validation accuracy as the number of markers increases suggests that the marker set can be reduced to a small set of key marker positions.

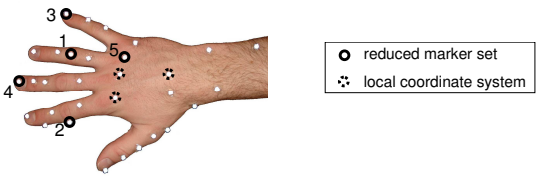


Fig. 6. The reduced marker set with five markers is determined from the feature selection results. The three markers defining the local coordinate system on the back of the hand are still required.

In addition, a fourth classifier is trained on the combined examples of subject 1 and subject 2 for all the grasps in both object sets A and B. For each training set, one classifier is trained using the reduced feature set of the selected $M = 5$ markers and another classifier is trained from the full marker set of $M = 30$ markers.

The four classifiers were evaluated on five test sets. The first and second test sets consist of grasps observed from subject 1 and subject 2, respectively, of object set A. These examples were not included in either the training set for feature selection nor the training set for the final trained classifier. The remaining three test sets consist of grasps of object set B for each of the three subjects. For each pair of a training set with a test set, the prediction accuracy is computed for both the reduced $M = 5$ and full $M = 30$ classifiers. We compare the relative performance of the two classifiers by the percent retainment which is the ratio of the accuracy of the reduced marker set classifier to the accuracy of the full marker set classifier.

D. Final classifier results

Overall, we found that the prediction accuracy was more sensitive to whether the training examples included grasps from the same subject than whether the examples included grasps for the same objects (Tables II and III). For subjects whose examples were included in the training set, the prediction accuracy was between 80–93% for the reduced marker set, corresponding to 92–97% retainment of the prediction accuracy from using the full marker set.

When the classifier is trained on examples from only a single demonstrator, the prediction accuracy for test grasps from a new subject was decreased to 21–65% for the reduced marker set classifier (Table II). However, in five of these six cases, predicting the grasp from the reduced marker set resulted in higher prediction accuracy than that from prediction based on the full feature set, corresponding to a retaininment ratio over 100%. This suggests that, although the weights trained by the classifier may not be appropriate for the new subject, the selected markers are still key features for grasp prediction that may be generalized across subjects.

The prediction accuracy for grasps of a new demonstrator improves when the training set for the classifier includes examples from multiple subjects (Table III). The prediction accuracy for the grasps observed from subject 3 increases from 22% to 70% using only the reduced marker set classifier. Importantly, comparing the bold values in Table II to those in Table III show that training the classifier on examples from multiple users results in only a marginal decrease in the prediction accuracy for grasps of subjects whose examples were included in the training set. Furthermore, the retainment of over 100% for the subject 3 test sets again shows that prediction is improved by using the reduced marker set instead of the full marker set.

Analysis of the grasp prediction errors (Fig. 7) shows the distribution of misclassified test examples for the classifier trained on the combined examples from subjects 1 and 2 for object set A. Overall, the classifier most successfully predicted cylindrical and pinch grasps for all three subjects. For subject 1, grasps labeled as spherical and lateral tripod

TABLE III

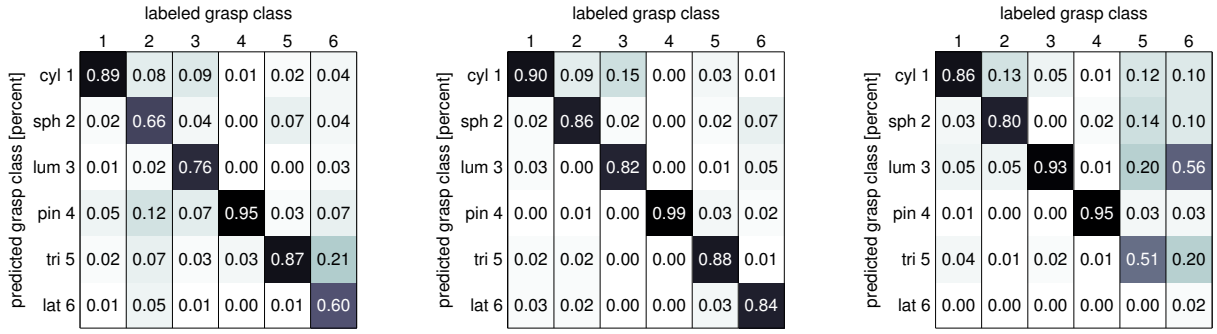
FINAL CLASSIFIER RESULTS USING TRAINING EXAMPLES FROM MULTIPLE DEMONSTRATORS. BOLD ENTRIES HIGHLIGHT CASES WHERE THE TRAINING SET INCLUDED GRASPS FROM THE SAME SUBJECT WHOSE GRASPS ARE IN THE TEST SET. PERCENT RETAINMENT MEASURES THE RATIO OF THE ACCURACY FROM THE REDUCED MARKER SET WITH $M = 5$ MARKERS TO THE ACCURACY FROM THE FULL MARKER SET WITH $M = 30$ MARKERS.

Classification accuracy (percent)		Training set with object set A subjects 1 and 2		
Test set		$M = 5$	$M = 30$	retainment
object set A (same)	subject 1	**83.9**	**89.4**	**93.8**
	subject 2	**90.3**	**93.5**	**96.6**
object set B (new)	subject 1	**80.6**	**86.6**	**93.0**
	subject 2	**92.0**	**95.9**	**96.0**
	subject 3	70.2	59.8	117.3
		Training set with object sets A and B		
object set B (same)	subject 3	70.2	63.9	109.8

TABLE II

FINAL CLASSIFIER RESULTS USING TRAINING EXAMPLES FROM A SINGLE DEMONSTRATOR. BOLD ENTRIES HIGHLIGHT CASES WHERE THE TRAINING SET INCLUDED EXAMPLES FROM THE SAME SUBJECT WHOSE GRASPS ARE IN THE TEST SET. PERCENT RETAINMENT MEASURES THE RATIO OF THE ACCURACY FROM THE REDUCED MARKER SET WITH $M = 5$ MARKERS TO THE ACCURACY FROM THE FULL MARKER SET WITH $M = 30$ MARKERS.

Classification accuracy (percent)		Training set with object set A					
		subject 1			subject 2		
Test set		$M = 5$	$M = 30$	retainment	$M = 5$	$M = 30$	retainment
object set A (same)	subject 1	**84.8**	**90.4**	**93.8**	57.0	44.9	126.9
	subject 2	44.6	36.3	122.8	**90.6**	**94.7**	**95.7**
object set B (new)	subject 1	**81.8**	**88.0**	**92.9**	51.2	40.5	126.2
	subject 2	45.1	41.8	107.9	**92.9**	**97.0**	**95.8**
	subject 3	21.6	23.2	93.0	64.9	52.4	123.9



a – Subject 1, object set A and B

b – Subject 2, object set A and B

c – Subject 3, object set B

Fig. 7. Prediction rates for the test set grasps for the classifier trained on examples of subjects 1 and 2 grasps for object set A (Table III). Each column shows the percentages of test examples where one grasp class was correctly predicted (values on the diagonal) or misclassified (all off-diagonal values). The results are separated by subject. (a) Classification rates for subject 1 test examples (combination of first and third test sets). (b) Classification rates for subject 2 test examples (combination of second and fourth test sets). (c) Classification rates for subject 3 test examples (fifth test set).

grasps were misclassified most frequently (Figs. 7 and 8). In particular, tripod grasp was often predicted incorrectly for lateral tripod examples, which underscores the similarity between these three-finger precision grasps (Fig. 8b). Accuracy across different grasp classes was more consistent for subject 2, with lumbrical grasp misclassified the most frequently (Fig. 8c). Classification rates for subject 3 show the first four grasps were correctly predicted in at least 80% of the examples. However, the last two grasps were misclassified for over 40% of examples, suggesting that there may be a systematic difference in the way subject 3 performs the tripod grasps compared to subjects 1 and 2 (Figs. 8d and 8e).

The confidence measure $c(\mathbf{x})$ defined in (5) can be used to improve the recognition procedure. We expect that predictions with larger values of $c(\mathbf{x})$ are more likely to be correct. If $c(\mathbf{x})$ is a small value, the example should be discarded. The subject can repeat the demonstration, or in a system which models temporal coherence, the grasp may be predicted from the remaining hand poses in the grasp trajectory. Alternatively, multiple grasp classes with the highest posterior probabilities could be returned as candidate predictions. The threshold for accepting and discarding predictions, c^*, is determined from the training data used in the feature selection experiments by the equal error operating point from receiver operating characteristic analysis [see, e.g., 23], where the rate of discarded correct predictions (false negatives) and the rate of accepted incorrect predictions (false positives) are equal. This corresponded to $c^* = 4.88$ and a prediction accuracy of 96.8% for the 73.5% of the training examples with $c(\mathbf{x}) \geq c^*$. As can be seen in Table IV, discarding examples according to the selected threshold c^* can increase the classification accuracy for the multiple demonstrator setting to over 95% from 80–92% for subjects who provided the training examples and to 83% from 70% for a new subject.

a b c d e

Fig. 8. Examples from grasp classes with high rates of misclassification. (a) Subject 1 spherical grasp (Fig. 7a, column 2) was sometimes misclassified as pinch grasp, possibly due to the index finger being bent less for grasps of small objects. (b) Subject 1 lateral tripod grasp (Fig. 7a, column 6) is similar to tripod grasp when there is a straightened instead of bent middle finger. (c) Subject 2 lumbrical grasp (Fig. 7b, column 3) included a separation of the index finger from the other fingers and was misclassified as cylindrical grasp for some examples. (d) Subject 3 tripod grasp (Fig. 7c, column 5) sometimes positioned the object closer to the palm of the hand and involved a wider angle between the index and middle finger compared to the other subjects. (e) Subject 3 lateral tripod grasp (Fig. 7c, column 6) exhibited a bent index finger and straightened middle finger, compared to the straightened index and bent middle finger in the lateral tripod grasps of subjects 1 and 2.

TABLE IV

CLASSIFICATION ACCURACY INCREASES WHEN PREDICTIONS WITH CONFIDENCE MEASURES LOWER THAN THRESHOLD $c*$ ARE DISCARDED. SIZE OF THE TEST SUBSET IS REPORTED AS THE PERCENT OF EXAMPLES FROM THE FULL TEST SET (TABLE III) WITH CONFIDENCE $c(\mathbf{x}) \geq c^*$.

Classification accuracy (percent)		Training set with object set A subjects 1 and 2	
Test set, $c(\mathbf{x}) \geq c^*$		$M = 5$	test subset size
object set A (same)	subject 1	**95.9**	**66.7**
	subject 2	**95.8**	**81.6**
object set B (new)	subject 1	**96.8**	**63.2**
	subject 2	**97.1**	**79.7**
	subject 3	83.7	65.4
		Training set with object set A and B	
object set B (same)	subject 3	83.3	64.8

V. DISCUSSION

In summary, supervised feature selection has been used to methodically design a reduced marker protocol for observing demonstrated grasps. Evaluation of the selected reduced marker set on grasp examples from multiple subjects showed that using as few as five markers as input features retains over

92% as much prediction accuracy from the full set of 30 markers. In particular, not only does grasp recognition from five markers reduce the model dimensionality, but using the reduced feature set can actually generalize better to a new subject by improving the prediction accuracy compared to the full marker set. We also found that inclusion of observed examples from two subjects in the training set improved the generalization of the grasp classifier to examples from a new demonstrator and only marginally decreased prediction accuracy for the included demonstrators compared to the classifiers trained on a single subject's examples. Further investigation is required to determine how the generalization to new demonstrators can be further improved as examples from more subjects are included in training set.

Our approach identifies the number and placement of markers for one choice of reduced marker set based on sequential feature selection. We selected the markers according to the order from the backward selection results, but the forward selection cross-validation accuracy was within 0.5% of that for the backward selection for the same number of markers. In limited testing of the alternative reduced marker set from forward selection, as well as other sets of five markers selected based on prior knowledge of grasps, the cross-validation accuracies were similar to the presented results. Thus, there may be several reduced marker sets that are nearly equivalent with respect to grasp recognition.

Feature selection was investigated specifically in the context of the selected linear logistic regression classifier, which we found to be sufficient for achieving reasonable grasp prediction accuracy. Preliminary experiments also evaluated linear support vector machines as a possible classification model using an available software implementation [24]. This resulted in only a marginal difference in the prediction accuracy for the full marker set classifier but required significantly more training time, which prohibits the sequential feature selection experiments that evaluate several possible feature subsets. However, future work could investigate alternative classifiers, both linear and nonlinear, with respect to the final reduced marker set proposed here. Furthermore, we have only considered the classification of a hand pose at a single time point, and modeling temporal coherence or evolution of the grasp may also improve recognition of the demonstrator's overall action or intent.

Other directions for future work might address the robustness of the method to the number of grasp classes selected. The six classes of functional grasps considered in this work describe broad categories of functional grasps. A possible limitation is that the reduced feature set of local marker coordinates, which will vary across subjects due to different hand sizes, may be less successful for recognizing a finer discretization of grasp classes that are distinguished by only slight changes in the hand configuration. Despite this, tracking a small number of key interest points on the hand surface can provide a useful feature set for grasp recognition, is possible without data glove measurement, and could supplement machine vision systems for observation learning.

VI. Acknowledgments

The authors thank Justin Macey for his assistance with the data acquisition.

References

[1] K. Ikeuchi and T. Suehiro, "Toward an assembly plan from observation. I. Task recognition with polyhedral objects," *IEEE Trans. Robot. Automat.*, vol. 10, no. 3, pp. 368–385, Jun. 1994.

[2] S. B. Kang and K. Ikeuchi, "Toward automatic robot instruction from perception-mapping human grasps to manipulator grasps," *IEEE Trans. Robot. Automat.*, vol. 13, no. 1, pp. 81–95, Feb. 1997.

[3] S. Ekvall and D. Kragic, "Grasp recognition for programming by demonstration," in *Proc. 2005 IEEE Int. Conf. Robotics and Automation*, 2005, pp. 748–753.

[4] I. Dejmal and M. Zacksenhouse, "Coordinative structure of manipulative hand-movements facilitates their recognition," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 12, pp. 2455–2463, Dec. 2006.

[5] S. B. Kang and K. Ikeuchi, "A robot system that observes and replicates grasping tasks," in *Proc. 5th Int. Conf. Computer Vision*, Jun. 1995, pp. 1093–1099.

[6] J. Rehg and T. Kanade, "Visual tracking of high DOF articulated structures: An application to human hand tracking," in *Proc. 3rd Eur. Conf. Computer Vision (ECCV '94)*, vol. II, May 1994, pp. 35–46.

[7] E. Ueda, Y. Matsumoto, M. Imai, and T. Ogasawara, "A hand-pose estimation for vision-based human interfaces," *IEEE Trans. Ind. Electron.*, vol. 50, no. 4, pp. 676–684, Aug. 2003.

[8] V. Athitsos and S. Sclaroff, "Estimating 3D hand pose from a cluttered image," in *IEEE Comp. Soc. Conf. Computer Vision and Pattern Recognition*, vol. 2, Jun. 2003, pp. 432–439.

[9] C. Schwarz and N. Lobo, "Segment-based hand pose estimation," in *2nd Canadian Conf. Computer and Robot Vision*, May 2005, pp. 42–49.

[10] N. Pollard and V. B. Zordan, "Physically based grasping control from example," in *Proc. ACM SIGGRAPH/Eurographics Symp. on Computer Animation*, Aug. 2005, pp. 311–318.

[11] P. G. Kry and D. K. Pai, "Interaction capture and synthesis," *ACM Trans. Graph. (SIGGRAPH 2006)*, vol. 25, no. 3, pp. 872–880, 2006.

[12] M. Santello, M. Flanders, and J. Soechting, "Postural hand synergies for tool use," *J. Neurosci.*, no. 18, pp. 10 105–15, 1998.

[13] C. Mason, J. Gomez, and T. Ebner, "Hand synergies during reach-to-grasp," *J. Neurophys.*, no. 86, pp. 2896–2910, 2001.

[14] J. Chai and J. K. Hodgins, "Performance animation from low-dimensional control signals," *ACM Trans. Graph. (SIGGRAPH 2005)*, vol. 24, no. 3, pp. 686–696, Aug. 2005.

[15] G. Liu, J. Zhang, W. Wang, and L. McMillan, "Human motion estimation from a reduced marker set," in *Proc. 2006 Symp. Interactive 3D graphics and games*. New York, NY, USA: ACM Press, 2006, pp. 35–42.

[16] K. Bernardin, K. Ogawara, K. Ikeuchi, and R. Dillmann, "A sensor fusion approach for recognizing continuous human grasping sequences using hidden markov models," *IEEE Trans. Robot.*, vol. 21, no. 1, pp. 47–57, Feb. 2005.

[17] M. Moussa and M. Kamel, "A connectionist model of human grasps and its application to robot grasping," in *IEEE Int. Conf. Neural Networks*, vol. 5, 1995, pp. 2555–2559.

[18] M. R. Cutkosky, "On grasp choice, grasp models, and the design of hands for manufacturing tasks," *IEEE Trans. Robot. Automat.*, vol. 5, no. 3, pp. 269–279, 1989.

[19] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY: Springer, 2006, pp. 209–210.

[20] E. P. Xing, "Feature selection in microarray analysis," in *A Practical Approach to Microarray Data Analysis*, D. Berrar, W. Dubitzky, and M. Granzow, Eds. Kluwer Academic Publishers, 2003, pp. 110–131.

[21] A. E. Flatt, *The care of the rheumatoid hand*. Saint Louis: The C. V. Mosby Company, 1974, pp. 12–15.

[22] S. J. Edwards, D. J. Buckland, and J. D. McCoy-Powlen, *Developmental & Functional Hand Grasps*. Thorofare, New Jersey: Slack Incorporated, 2002.

[23] H. L. V. Trees, *Detection, Estimation, and Modulation Theory: Radar-Sonar Signal Processing and Gaussian Signals in Noise*. Melbourne, FL, USA: Krieger Publishing Co., Inc., 1992, pp. 36–46.

[24] C.-C. Chang and C.-J. Lin, *LIBSVM: a library for support vector machines*, 2001, software available at http://www.csie.ntu.edu.tw/˜cjlin/libsvm.