Hand Posture Recognition with Co-Training

Yikai Fang[1,2] Jian Cheng[1] Jinqiao Wang[1] Kongqiao Wang[2] Jing Liu[1] Hanqing Lu[1]

[1]National Lab of Pattern Recognition
Institute of Automation,
Chinese Academy of Sciences, Beijing, 100080
{ykfang, jcheng, jqwang, jliu, luhq}@nlpr.ia.ac.cn

[2]Nokia Research Center
No.5 Dong Huan Zhonglu, BDA
Beijing, 100176
{yikai.fang, kongqiao.wang}@nokia.com

Abstract

As an emerging human-computer interaction approachvision based hand interaction is more natural and efficient. Howeverin order to achieve high accuracy, most of the existing hand posture recognition methods need a large number of labeled samples which is expensive or unavailable in practice. In this paper, a co-training based method is proposed to recognize different hand postures with a small quantity of labeled data. Hand postures examples are represented with different features and disparate classifiers are trained simultaneously with labeled data. Then the semi-supervised learning treats each new posture as unlabeled data and updates the classifiers in a co-training framework. Experiments show that the proposed method outperforms the traditional methods with much less labeled examples.

1. Introduction

With the development of computing technology, computers are well integrated in our daily life. Many computer applications require more and more Human-Computer Interaction (HCI). However, the interaction nowadays is usually done using dedicated devices such as mouse, keyboard and pen. These devices of interaction are always not natural and cumbersome. The easiest way to interact with the machine would be the ways of daily communication between human and human. Hand gesture is frequently used in people's day-to-day life. It's also an important component of body languages in linguistics. Compared with those devices mentioned above, hand gestures are more natural in interaction. Integrating the use of hand gestures in HCI will be of great benefit and make the interaction much more natural and intuitive.

In order to use hand gestures in HCI, it is indispensable to make hand gestures interpreted by computers. A common technique is to use magnetic sensors, acoustic or inertial trackers and data gloves. These extra sensors and instruments may be easy to collect hand configuration and motion and give accurate results. However, these equipments are usually expensive and bring much cumbersome experience to users. The less intrusive methods for gesture interaction are vision based gesture interaction, which have many appealing characteristics. The prominent one is that it realizes a natural interaction between human and computers independent of external dedicated devices. Moreover, vision based gesture interaction has the advantage of being unobtrusive.

In this paper, we focus on the problem of hand posture recognition. Here the hand posture means the pose or configuration of the hand in one single image. There have been a number of research efforts on hand posture classification or recognition in recent years. Ong and Bowden[8] distinguished hand postures with boosted classifier tree and obtained fairly good results. However, the classifier in their method too complicated and time-consuming. In addition, the samples are with simple and similar backgrounds and the training requires thousands of labeled samples. Kolsch[6] employed fanned boosting detection for classification and got nearly real time results. While the training process is extremely time exhausting. Just et al introduce modified census transform (MCT) into hand gesture classification[5]. Their method gives fairly good average accuracy above 80 percent with the classifier trained with more than 2,000 samples per posture, while the performance in recognition experiments under complex background was not much satisfactory. These prevalent methods for recognition are learning based with specific features. However, thousands of labeled samples are indispensable in most of these methods. And the training is usually time costly. Furthermore, the traditional methods convert or normalize a variety of features into a unified feature space, which ignores the distinct attributes of the different features.

To address the problems mentioned above, a hand posture recognition approach with Co-training strategy is proposed in this paper. The main idea is to train two disparate classifiers with each other and improve the performance of both classifiers with unlabeled samples. This proposed method improves the recognition performance with less labeled data in a semi-supervised way based on Co-training framework[1] .The rest of this paper is organized as follows. Section 2 briefly introduces some related works about Co-training. In section 3, the proposed method is described in detail. Then experiments results on the standard dataset are provided section 4. Finally, we will conclude and propose some future directions.

2. Related work

Co-training was proposed by Blum and Mitchell[1] as a method for training a pair of learning algorithms. The basic assumption is that the two learning algorithms use two different views of the data. The key property is that some examples which would have been confidently labeled using one classifier would be misclassified by the other classifier. The classifiers go through unlabeled examples, label them, and add the most confident predictions to the labeled set of the other classifier. Therefore the classifiers train each other by providing additional informative examples from unlabeled data. After Co-training, the final classifiers, which are trained on labeled and unlabeled data, are significantly improved.

The Co-training strategy has been introduced to some classic problems in computer vision community. Recently, Co-training has been used in object tracking and detection. Levin[7] trains two classifiers to detect vehicles with the same feature of grey and background-subtracted images. Javed[4] improves the performance of two boosting classifiers with Co-training, in which the based classifiers are Bayes classifiers and features are derived from PCA of training samples. Tang[10] obtains two SVM classifiers with color histogram and HOG (histogram of oriented gradient) as two independent views of object. The two SVM classifiers are combined to get the location of object in video and new samples are generated to update the SVMs online.

3. Posture recognition with Co-training

Inspired by Blum and Mitchell's work in webpage classification[1], a co-training based hand posture recognition method is proposed in this paper. Two different views of samples, haar and HOG features are used to train disparate classifiers for hand posture recognition. These features describe different type of information in samples. Simple combination of these classifiers ignores the complementation characteristics of different features. To improve the performance, these disparate classifiers are trained together in a uniform cooperation process.

3.1 Multiple views

One approach for object representation is to use Haar features. The advantage of using the Haar features is that they can be calculated very efficientlyAnd the boosting learning based on Haar features has many successful applications in face detection and recognition. LBP (Local Binary Pattern) or MCT[5] are more robust to illumination variation and employed in face recognition with higher accuracy. However, the binary pattern histogram in LBP methods[5] lose its statistical significance with only a small training set available. As the intention in this paper is to train classifiers with less labeled data, LBP based features are not suitable for the proposed method.

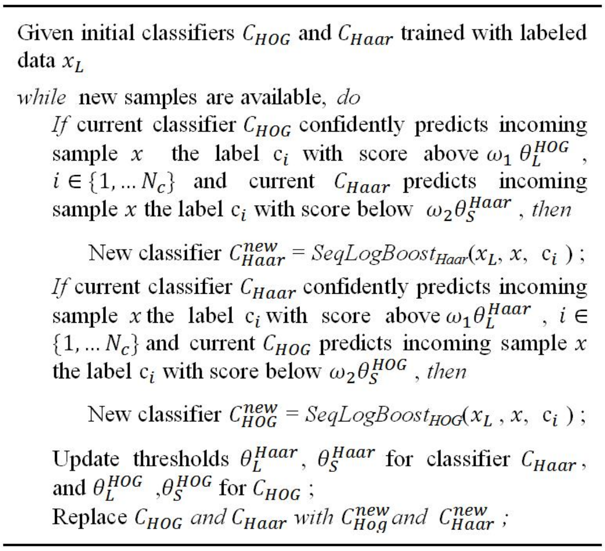
Histogram of oriented gradients (or HOG) is another feature for object representation and is frequently used in pedestrian detection [3]. Since this feature descriptor operates on a dense grid of localized cells and reflects the shape or appearance information of object, we apply it in hand posture recognition with LDA(linear discriminant analysis) as weak classifiers.

3.2 The learning of posture classifiers

In the proposed method, two boosting based classifiers are co-trained to differentiate one hand posture against others. One classifier uses haar features and the other use HOG feature.

One important aspect to keep the co-training effective is that each classifier labels only those unlabeled samples on which it can make a confident prediction and add them to the training data of the other classifier. So the additional samples should have large margins on one classifier. In our method, the sequential logistic regression algorithm[2] is used to obtain the boosting classifiers. Denote the outputs from the weak classifiers by the vector $\vec{h} \in [-1, -1]^n$ and the weights associated with these classifiers by $\vec{\alpha} \in [-1, 1]^n$, $\sum_{j=1}^{n} \vec{\alpha_j} = 1$, n is the number of weak classifiers. According to Schapire[9], large margins on the training set imply correct classification on test data both experimentally and theoretically. That is, if there is some real number $\theta_L > 0$ such that the probability that $\vec{h} \cdot \vec{\alpha} > \theta_L$ is significant and the conditional probability that $\vec{h}$ corresponds to a confident prediction given that $\vec{h} \cdot \vec{\alpha} > \theta_L$ is close to 1. Schapire[9] concludes that there exists θ_L estimated on the training or validation set, for which the risk of misclassification on test data is very low. So the samples confidently labeled with threshold criteria can be added to the training data of another classifier.

However, it would be inefficient to add each confidently labeled sample to training data of the other classifier. The samples labeled during co-training improve the performance of the boosted classifier only if the samples have small or negative margin. If the samples have been labeled confidently by the boosted classifier with large margin, add it to training will have little effect on the final classifier. Thus the unlabeled samples which are confidently labeled by current classifier and have a small margin simultaneously are indispensable. The threshold θ_s on the score of the boosted classifier can also be established through the training or validation set. Once a sample has been labeled and if it has a small margin, it's used for updating the boosting classifiers.



```
Given initial classifiers $C_{HOG}$ and $C_{Haar}$ trained with labeled data $x_L$
while new samples are available, do
    If current classifier $C_{HOG}$ confidently predicts incoming sample $x$ the label $c_i$ with score above $\omega_1 \theta_L^{HOG}$ , $i \in \{1, \dots N_c\}$ and current $C_{Haar}$ predicts incoming sample $x$ the label $c_i$ with score below $\omega_2 \theta_S^{Haar}$ , then
        New classifier $C_{Haar}^{new} = SeqLogBoost_{Haar}(x_L, x, c_i)$ ;
    If current classifier $C_{Haar}$ confidently predicts incoming sample $x$ the label $c_i$ with score above $\omega_1 \theta_L^{Haar}$ , $i \in \{1, \dots N_c\}$ and current $C_{HOG}$ predicts incoming sample $x$ the label $c_i$ with score below $\omega_2 \theta_S^{HOG}$ , then
        New classifier $C_{HOG}^{new} = SeqLogBoost_{HOG}(x_L , x, c_i)$ ;
    Update thresholds $\theta_L^{Haar}$, $\theta_S^{Haar}$ for classifier $C_{Haar}$, and $\theta_L^{HOG}$ ,$\theta_S^{HOG}$ for $C_{HOG}$ ;
    Replace $C_{HOG}$ and $C_{Haar}$ with $C_{Hog}^{new}$ and $C_{Haar}^{new}$ ;
```

Figure 1. The learning of posture classifiers

The co-training and classifier update algorithm in the proposed method is shown in Figure 1. C_{HOG} and C_{Haar} are classifiers with HOG and haar respectively. The number of classes is denoted with N_c. The function *SeqLogBoost* implements the sequential logistic regression algorithm.

The update of classifier C_{Haar} and C_{HOG} is shown in Figure 1. To simplify the threshold selection, the threshold θ_L for the large margin is selected as the highest score achieved by negative samples in labeled and unlabeled data, and θ_S for the small margin is the lowest score achieved. To be more conservative, ω_1 and ω_2 are set to 1.5 and 2 in our experiments.

4. Experiments

The dataset used to validate the proposed method is the benchmark dataset in the field of hand posture recognition -Triesch dataset[11]. It consists of 10 hand signs performed by 24 different people against different backgrounds. The backgrounds are of three types: uniform light, uniform dark and complex. Figure 2 gives some posture examples.

To train and test the proposed method with these images, they are cropped and resized to 128x128, followed by histogram normalization. As the samples of each posture contain only 72 images, which is insufficient for the training of boosting based classifier. So it's necessary to increase the available number of samples. Here some little perturbations are added to the initial images. The images are shifted, scaled and rotated. Then 35 images are generated for each original image. Thus there are totally 25200 images, 2520 images for each hand posture.

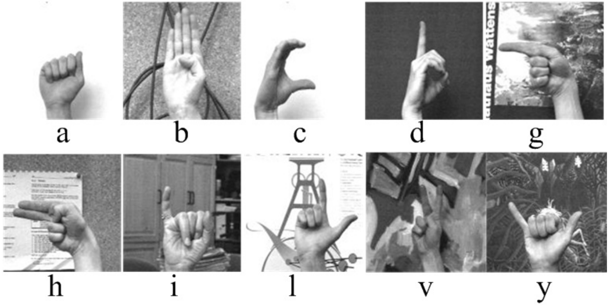


Figure 2. Hand posture examples

The dataset is divided into 2 subsets: training set, and test set. For each hand posture, there are 1050 images in training set and 1470 images in test set. One classifier is trained for each posture. For a given posture test image, all classifiers are applied and the classifier giving the highest score is selected to label the test posture. The negative samples which are necessary to each posture classifier comprise samples of nine other postures and have equivalent number of images with positive samples.

The initial two classifiers of haar and HOG feature are trained with 150 positive samples. Then in each step of classifier update, 300 samples are added to training set as unlabeled data, which contains 150 positive samples mixed with the same amount of negative samples, until the 1050 positive samples for each posture are used up. The recognition accuracies for 10 postures are shown in Figure 3. Each plot in Figure 3 gives recognition accuracy on one posture.

As is shown in Figure 3, co-training based classifiers(hog_co and $haar_co$) obtain better results than single classifiers (hog and haar) in general. For the postures difficult to recognize, such as d, g, i and y,

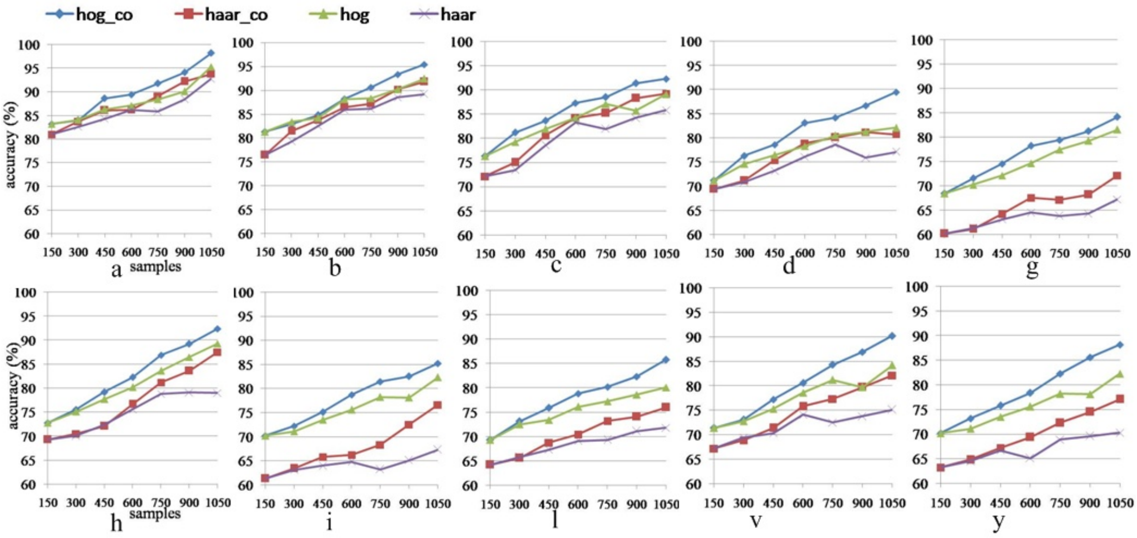


Figure 3. Accuracy of hand posture recognition

recognition accuracy is improved by 5 9 percent with co-training for both classifiers with haar and HOG. The most significant improvement appears on posture d with HOG and posture i with haar. While for the easier postures a, b and c, the improvement by co-training is 2 4 percent, not as significant as difficult postures, as the accuracy of a, b and c are rather high (above 85%) by single classifiers without co-training with little space for improvement by simple boosting training. For all postures, HOG is a more effective feature descriptor than haar. The average accuracy on the dataset by independent classifiers with HOG and haar is 85.3% and 77.4% After co-training, the accuracy is 90.1% with HOG and 82.7% with haar. Both the co-trained classifiers outperform the MCT classifiers in [5] with the amount of labeled data decreased.

5. Conclusion

In this paper, a co-training based hand posture recognition method is proposed. Haar and HOG features are used to build single classifiers respectively and the two classifiers then train each other with unlabeled data. Experiments on standard dataset show that the proposed method improves recognition accuracy of boosted classifiers with much less labeled data.

6. Acknowledgement

The research was supported by National High Technology Research and Development Program of China (Grant No. 2006AA01Z315) and Beijing Natural Science Foundation (Grant No. 4072025).

References

[1] A. Blum and T. Mitchell. Combining labeled and unlabeled data with co-training. In *COLT: Proceedings of the Workshop on Computational Learning Theory*, pages 92–100, 1998.

[2] M. Collins, R. E. Schapire, and Y. Singer. Logistic regression, adaboost and bregman distances. In *Machine Learning*, pages 158–169, 2000.

[3] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proceedings of CVPR*, pages I: 886–893, 2005.

[4] O. Javed, S. Ali, and M. Shah. Online detection and classification of moving objects using progressively improving detectors. In *Proceedings of CVPR*, pages I: 696–701, 2005.

[5] A. Just, Y. Rodriguez, and S. Marcel. Hand posture classification and recognition using the modified census transform. In *Proceedings of AFGR*, pages 351–356, 2006.

[6] M. Kolsch and M. Turk. Robust hand detection. In *Proceedings of AFGR*, pages 614–619, 2004.

[7] A. Levin, P. Viola, and Y. Freund. Unsupervised improvement of visual detectors using co-training. In *Proceedings of ICCV*, pages 626–633, 2003.

[8] E. Ong and R. Bowden. A boosted classifier tree for hand shape detection. In *Proceedings of AFGR*, pages 889–894, 2004.

[9] R. E. Schapire, Y. Freund, P. Bartlett, and W. S. Lee. Boosting the margin: a new explanation for the effectiveness of voting methods. volume 26, pages 1651–1686.

[10] F. Tang, S. Brennan, Q. Zhao, and H. Tao. Co-tracking using semi-supervised support vector machines. In *Proceedings of ICCV*, pages 1–8, 2007.

[11] J. Triesch and C. V. D. Malsburg. Robust classification of hand postures against complex backgrounds. In *Proceedings of AFGR*, pages 170–175, 1996.