A general boosting-based framework for active object recognition

Zhaoyin Jia
zj32@cornell.edu
Yao-Jen Chang
yc682@cornell.edu
Tsuhan Chen
tsuhan@ece.cornell.edu

Electrical and Computer Engineering Department
Cornell University
Ithaca, NY, USA

In recent years the problem of object recognition have been extensively studied. Under many circumstances users will have the access to control the vision system, e.g. a guided robot. In that case not only can we acquire multiple views, but also are able to *actively* control the system to pick a certain angle. This is where the context of active view recognition appears.[1][2] [3][4][6].

We propose a novel general framework with a boosting algorithm to achieve active object classification by view selection. The proposed framework actively decides the next best view for the recognition task. It evaluates different information sources for top hypotheses, generates a voting matrix for candidate views and the view selection is achieved by picking up the one with the maximum votes. Three different sources - similarity based on Implicit Shape Model, prior for model, and prior for views - are briefly presented in the followings. Moreover, we convert view selection itself into a classification problem, and propose a boosting algorithm that is able to combine these sources. Experiments show that our algorithm produces a better strategy compared to the other baseline methods.

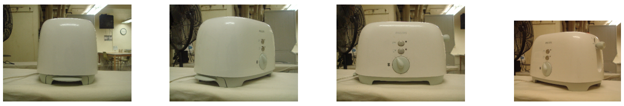
(a) 0^o (b) 45^o (c) 90^o (d) 135^o

Figure 1: Some view angles are easier to identify the object while the others are hard, e.g. 0^o is not a good view for identify this toaster. However, at 135^o, it becomes much easier. In this paper, we propose a strategy to selecting the views to maximize the recognition performance

To separate two models, the next view should correspond to models that are the least similar with each other. Therefore we measure the similarity based on the recognition model we use: the Implicit Shape Model[5]. For two ISMs, O_1 and O_2, we firstly measure the one-way similarity $Simi'(O_1,O_2)$ by treating the word $w_{1,i}$ in O_1 as a feature e in the test image, and calculate the vote of this to all the words $w_{2,j}$ in the model O_2.

$$p(O_2,\lambda|w_{1,i},l) \quad = \quad \sum_j p(O_2,\lambda|w_{2,j},l)\cdot p(O_2|w_{2,j})p(w_{2,j}|w_{1,i}) \quad (1)$$

the one-way similarity $Simi'(O_1,O_2)$ becomes:

$$Simi'(O_1,O_2) = \sum_{i,|\lambda|<rd} p(O_2,\lambda|w_{1,i},l) \qquad (2)$$

Similarity should be irrelevant of the order, and this is achieved by adding two one-way similarities together:

$$Simi(O_1,O_2) = Simi'(O_1,O_2) + Simi'(O_2,O_1) \qquad (3)$$

Whether a model itself is easy to recognize or not becomes another factor taken into consideration. A separate validation set is used for evaluating the prior of a model. We evaluate them on the implicit shape model, and sample the prior for each model $O_{c,v}$ as $Pr(O_{c,v}) = TP/(TP+FP)$, where TP and FP are the number of true positives and false positives of the model when testing the validation set.

We also sample $p(t|O_{c,v})$: the optimal views to pick-up given the ground truth $O_{c,v}$, and use it as another information source for view selection. $p(t|O_{c,v})$ is calculated in the following way: given the ground truth hypothesis $O_{c,v}$ of one testing image from the validation set, the optimal view index t is selected by picking up the view t that can maximize the corresponding true hypothesis $O_{c,v+(t-1)d}$, while minimizing all the other wrong ones. Having acquired all the optimal views, the prior for view is calculated in a Bayesian probability format: $p(t|O_{c,v}) = p(t,O_{c,v})/p(O_{c,v})$.

Also a boosting algorithm specially designed for the active recognition task has been proposed. It aims to combine different sources together and transform it into a better strategy. This algorithm takes each individual source with one specific number of accepted hypotheses K as the weak classifier, and tries to weigh each one, minimizing the risk in taking a wrong action. The classification error has been redefined in the way that, better views, although not optimal one, are still less penalized. The whole algorithm is described as followings:

Algorithm 1 Algorithm to train the view selection

```
Given $M$ training images $\{I_m, m = 1,\dots,M\}$, their corresponding optimal view sequence $\{t_{mn}\}$, and a set of weak classifiers $\{C_{u,K}\}$ with various $u$ and $K$, initialize $D_1(m) = 1/M$
for $r = 1$ to $R$ do
    evaluate every $C_{u,K}$ on all $\{I_m\}$ and calculate weighted error $\varepsilon_{u,k} = \sum_{m=1}^{M} err_{u,K}(I_m)D_r(m)$
    Select $C_r = \underset{C_{u,K}}{\mathrm{argmin}}\,\varepsilon_{u,K}$. and record the training error $\varepsilon_r$ of $C_r$.
    Choose $\alpha_r$ for $C_r$: $\alpha_r = 0.5\ln\frac{1-\varepsilon_r}{\varepsilon_r}$
    Update $D_{r+1}(m) = \frac{D_r(m)\exp(-\alpha_r \cdot Id(err(I_m)=1))}{Z_t}$
end for
```

$C_{u,K}$ represents a weak classifier from source u with parameter K. $err_{u,K}$ is the training error described above. R is the maximum iteration step.

The details of the framework and the algorithm could be found in the paper. Experiments are performed on a public dataset [7]. Results show that our proposed combined algorithm outperforms each single source and the baselines in achieving higher recognition accuracy at early steps.

[1] S. Abbasi and F. Mokhtarian. Automatic view selection in multi-view object recognition. In *ICPR*, pages Vol I: 13–16, 2000.

[2] F. Farshidi, S. Sirouspour, and T. Kirubarajan. Robust sequential view planning for object recognition using multiple cameras. *Image and Vision Computing*, 27(8):1072–1082, July 2009.

[3] Zhaoyin Jia, Yao-Jen.Chang, and Tsuhan Chen. Active view selection for object and pose recognition. In *ICCV 3DRR Workshop*, pages 1–8, 2009.

[4] C. Laporte and T. Arbel. Efficient discriminant viewpoint selection for active bayesian recognition. *International Journal of Computer Vision*, 68(3):267–287, July 2006.

[5] B. Leibe, A. Leonardis, and B. Schiele. Robust object detection with interleaved categorization and segmentation. volume 77, pages 259–289, May 2008.

[6] Lucas Paletta and Axel Pinz. Active object recognition by view integration and reinforcement learning. *Robotics and Autonomous Systems*, 31(1-2):71–86, 2000.

[7] S. Savarese and F. F. Li. 3D generic object categorization, localization and pose estimation. In *ICCV*, pages 1–8, 2007.