Object Classification in Videos

BTech Report
by
Subodh Misra
B.Tech, 4th year
Y6481
Guide: Prof. Amitabha Mukerjee
Dept of Computer Science and Engineering
IIT Kanpur - 208 016
INDIA

November 8, 2009

CONTENTS

1 VISUAL OBJECT CLASSIFICATION IN VIDEOS

Can we identify the objects that are moving in this video sample?

Figure 1: Cars moving through a parking area

It is very easy for the human brain to locate all the objects that are in motion in a video sample and classify them under different categories. For example, in the above video clips we can see a red object moving across the frame and we can clearly identify it as a car. This is an easy task for a human being but if we want a computer system to do this job then it becomes an extremely difficult task.

This is my Btech project to build a software tool that can classify all the moving objects in a video clip.

2 CHALLENGES IN OBJECT CLASSIFICATION IN VIDEOS

2.1 The detection of a moving object in a video

We detect foreground objects in a video by detecting all the pixels whose content is being changed in consecutive frames. This is a difficult task as even the small change in the intensity of pixels can be counted as belonging to a moving object and so there is a lot of noise while detecting foreground objects.

Also there is not just a single object moving in a video and it may be the case that two moving objects overlap in which case it becomes difficult to classify the hidden object because the key features of that object might have been hidden which makes it difficult for classifier to categorize that object. This is called Occlusion.

2.2 Classification of extracted objects

After the foreground object extraction, classification is not easy because of background clutter and varying imaging conditions as well as large number of object categories in the universe.

3 DATASETS

The Caltech-101 dataset compiled by Fei-Fei et al. (2004) is used for training the classifier to classify images under one of these 101 categories.

PETS2000 video, IIT gate video will be the test videos. Objects from these videos will be extracted and classified in one of caltect 101 object categories.

4 PRESENT IMAGE CLASSIFICATION MODELS

4.1 SERRE POGGIO CLASSIFIER

This is the biologically inspired model of object recognition. The way human brain works is that it builds hierarchical models invariant of size and position but has feature specificity. There are simple cells which select specific features and there are complex cells which perform the pooling operation to get size and position invaiance. In this way we are able to recognize objects according to its features not the size and position. In serre poggio classifier visual cortex mechanism is implemented by alternating layers of simple and complex cells.

Simple cells use convolution with local filters and compute higher order features.

Complex cells use pooling mechanism to get invariance.

Following layers are calculated to get the feature vector of an image.

- **S1 layer : Gabor filter of different size is used to get S1 features from original image**
- **C1 layer : S1 features are convolved with 2D max filter to get C1 features to get spatial invariance.**
- **S2 layer : Now there are d2 prototypes from training images. we perfrom template matching between each of d2 prototypes and C1 features. We get S2 features for each of d2 prototypes.**
- **C2 layer : Now we perform global MAX operation and we get array of d2 scalars.**

Finally we perform all pairs linear SVM classification.

4.1.1 Saurabh Daptardar's thesis

Mtech student of IITkanpur Y7 batch had imlemented Serre poggio classifier as part of his thesis work "Explorations on a neurologically plausible model of image object classification".

4.2 Mutch and Lowe classifier

This is also the biologically inspired model of object recognition as was Serre Poggio model but there are some changes in this algorithm.

Following layers are calculated to get the feature vector of an image.

- **Image layer : Image pyramid of 10 scales is created**

- **S1 layer : Gabor filter of size 11 X 11 is used to get S1 features from original pyramid**
- **C1 layer : S1 pyramid is convolved with 3D max filter of size 10 X 10 and depth 2 units to get C1 pyramid.**
- **S2 layer : Now there are d2 prototypes from training images. we perfrom template matching between each of d2 prototypes and C1 pyramid. We get S2 pyramid for each of d2 prototypes.**
- **C2 layer : Now we perform global MAX operation and we get array of d2 scalars.**

Finally we perform all pairs linear SVM classification.

Sparse features are used in this mechanism and only the features that are highly wieghted by SVM are considered.

Differences between Serre Poggio and Mutch n Lowe algorithm are:

- **Different sized gabor filters are used in Mutch And Lowe for S1 feature calculation and pyramid approach is used to extract S1 features.**
- **Image height is always scaled to 140 in Serre Poggio while in Mutch n Lowe shorter edge is scaled to 140 while maintaining the aspect ratio.**
- **C1 samples do not overlap in serre poggio model.**

4.3 Bosch and zissermann technique of object classification

This technique estimates the Region of Interest(ROI) in an image in order to reduce background noise and hence better feature calculation of the object in the image. It improves the image classifiction by not just calculating the bag of visual words but also the orientation gradient to get the edge distribution in the object and better feature calculation. For classification of images, it uses Random forest and ferns classifier and not the SVM classifier because former is easy to train and test the images.

5 OBJECTIVE OF BTP

The objective of BTP is to get the foreground objects from a video sample and then use the one of the image classification mechanisms to classify under one of caltech 101 object categories.

For this purpose I will be doing the following:

- **Use the foreground object extraction code to extract the moving objects in a video.**
- **Implement the Mutch and Lowe code to classify extracted objects. I will use the code written by Saurabh Daptardar to replace basicserrepoggio() method implementation by mutch and lowe method. All pairs linear SVM classifier will be used which is used by Saurabh's code for the classification of images.**

I am imlementing Mutch and Lowe code because it has been proved to be more efficient than Serre poggio experimentally and so I would like to implement it rather than using Serre poggio code.

Bosch and Zissermann technique is not used because they are using Random forest classifier instead of SVM and they use the technique of ROI which is not useful in our case as we are already extracting the object from the video and we have to classify that object only.

6 WORK DONE SO FAR

- **In this semester I have read various image classification models : Serre Poggio, Mutch and Lowe, Bosch and zissermann. Then according to the requirements of the project, I selected the most appropriate model to be implemented.**
- **Got the code fom Saurabh Daptardar and studied it and figured out which all methods I need to replace with my implementation.**
- **Studied the codes of PHOG and PHOW generation to get the feel of feature calculation of images.**
- **Tried to run Saurabh's code for objects extracted from PETS2000 video to see the performance of serre poggio model but could only calculate S2 features of those images as there was some problem to find a 'PatchPosFile' in saurabh's code which are random patches from trained images of caltech 101 dataset.**

References

[1] **A. Bosch, A. Zisserman, and X. Munoz. Image classification using random forests and ferns. 2007.**

[2] **"Saurabh Daptartdar". *"Explorations on a neurologically plausible model of image object classification"*. PhD thesis, 2009.**

[3] **L. Wolf T. Poggio H. Jhuang, T. Serre. Biologically inspired system for action recognition. 2007.**

[4] **Christoph Lampert. Detecting objects in large image collection and videos by efficient subimage retrieval. 2009.**

[5] **Jim Mutch and David Lowe. Object recognition with sparse, localized features. 2006.**

[6] **Poggio" "Riesenhuber. "hierrachical model of object recognition in cortex". 1999.**

[7] **T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio. Robust object recognition with cortex-like mechanisms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 411–426, 2007.**