Concept Detection Using Local Binary Patterns and SVM

Duy-Dinh Le[1] and Shin'ichi Satoh[1]

National Institute of Informatics,
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, Japan 101-8430
ledduy@nii.ac.jp, satoh@nii.ac.jp

Abstract. We introduce a method for detecting concepts by using only visual information. In this method, firstly the feature vector of each frame is formed by concatenating histograms of local binary patterns extracted from overlapped and fixed-size rectangles within the frame. Then SVM classifiers are trained to classify frames as containing or not containing the concept. Experimental results show that our approach is simple but can achieve the median.

1 Introduction

Concept detection is one of the important tasks in video indexing. However, it is a challenging problem due to the following reasons:

- **Small and imbalanced training set:** the number of positive samples is very small and much smaller than the number of negative samples. Figure 1 shows the distribution of the training samples of TRECVID's concepts. These samples can be seen at http://satoh-lab.ex.nii.ac.jp/users/ledduy/Demo-TRECVID06/Concepts/AllConcepts-index.htm.
- **Semantic gaps:** Concepts containing objects such as Airplane, Building, US Flags, etc are difficult to detect due to many variations of objects in translation, occlusion, rotation, appearance, illumination, etc. Figure 2 shows examples with large variations of the concept 'Airplane'.
- **Noisy annotated data:** Concepts are highly abstraction and difficult for annotators to judge. For example, the concept 'Charts' is defined as "Shots depicting any graphics that is artificially generated such as bar graphs, line charts, etc. (maps should not be included)". However, as shown in Figure 3, some annotated samples of the last row are not followed the definition. Furthermore, several positive samples are not annotated and therefore they can be considered as negative samples that make more difficult in finding the decision boundary (Figure 4). In another case, several positive samples are not correctly annotated (Figure 5).

There are two fundamental issues when building concept detectors: concept representations and learning methods. So far, existing approaches are mainly focusing on fusion methods that try to combine different information sources (e.g

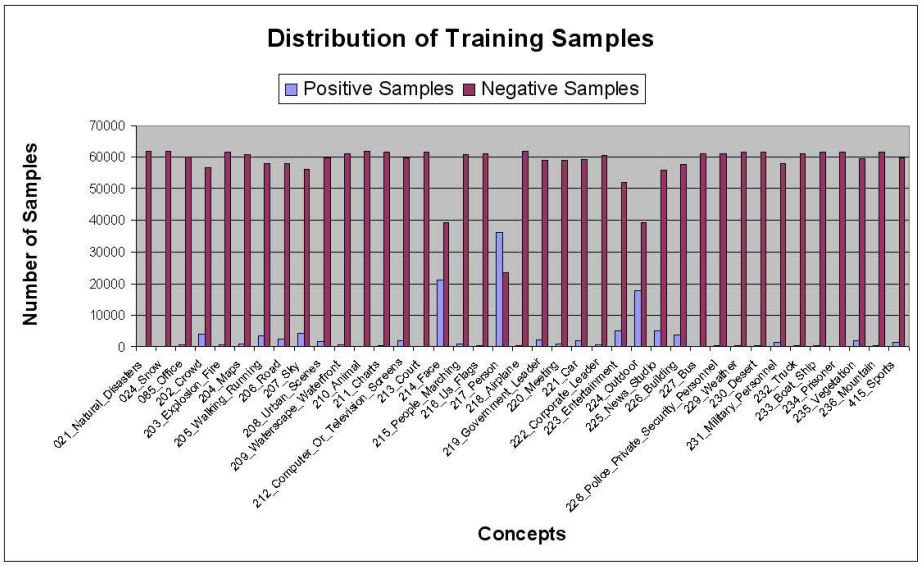

Fig. 1. Distribution of the training samples of TRECVID's concepts.

text, visual information, etc) and different abstraction level of representations. As for visual information, color based features are mainly used. Therefore, in the submission of this year 2006, we have investigated another kind of representation for visual information. Specifically, we use histograms of local binary patterns [1] extracted from overlapped rectangles within the input image to form the feature vector and then use SVM classifiers for classification.

2 Our Approach

For feature extraction, we use 196 overlapped rectangles of size 40x30 within a 320x240 frame. The overlapped size is 20x15. For each rectangle, we extract a 6-bin histogram of LBP feature as shown in Figure 6. Then these histograms are concatenated to form the feature vector. It results 1176 (196x6) features for each frame image (see Figure 7).

For each concept, we select top 20,000 negative samples from the negative image pool and maximum 1,500 positive images for training. For training SVM classifiers, we used LibSVM [2] with RBF kernel. Appropriate parameters C and γ were found by searching over the grid along with cross validation test. We found that $\gamma = 0.00390625$ and $C = 4$ are the most appropriate and used them for training all classifiers. Furthermore, we also used the probability estimation output provided by LibSVM in order to rank the samples that are classified as positives. The final results are top 2,000 samples corresponding to highest probability output.

Fig. 2. Examples of the concept 'Airplane'.

We trained 3 systems (runid: A_NII_ISM_R1_3, A_NII_ISM_R2_2, A_NII_ISM_R3_1) corresponding with 3 rounds of bootstrapping in which negative samples of the next round are accumulated by false positive samples of the previous round. For 3 rounds of training, maximum 2,000, 3,000 and 4,000 negative samples are used respectively.

The results of our group and other groups and are available on the Internet at http://satoh-lab.ex.nii.ac.jp/users/ledduy/Demo-TRECVID06/. As shown in Figure 8, our best result is approximately the median. Furthermore, by using bootstrapping, retrieval results increased from 1,643 shots to 1,913 shots over total 9,074 true shots.

3 Discussion

We have described a method that only uses the visual information represented by histograms of local binary patterns for detecting concepts. We use SVM with probability estimation output provided by LibSVM to classify and rank input images. Although the proposed approach is quite simple, its result on TRECVID's concept detection task can achieve the median.

References

1. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence **24**(7) (2002) 971–987
2. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines. (2001) Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

Fig. 3. Examples of the concept 'Charts'. It is not easy for annotators to judge these samples as 'Charts' by following the definition.

Fig. 4. Examples of incorrect annotation of the concept 'Maps'. Some samples of 'Maps' are considered as negative samples.

Fig. 5. Examples of incorrect annotation of the concept 'Airplane'. Some samples of weather forecast are annotated as 'Airplane'.

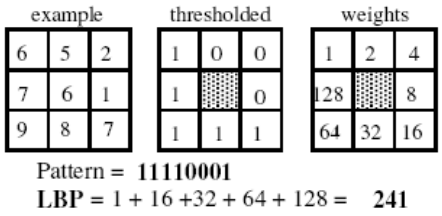

Fig. 6. Extraction of LBP features.

Fig. 7. Feature extraction for one frame image.

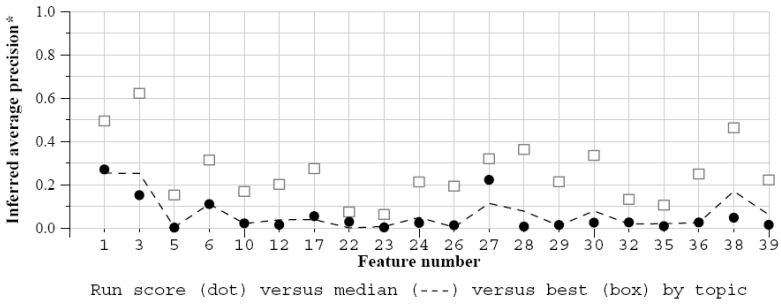

Fig. 8. The result of our best system compared to that of the other groups.