Recognition of Critical Visual Content

Andreas Opitz and Csaba Beleznai

AIT Austrian Institute of Technology GmbH
Donau-City-Straße 1, 1220 Vienna, Austria
{andreas.opitz, csaba.beleznai}@ait.ac.at

1 Introduction

Image content contained within the internet encompasses vast amounts of visual data. Some of this data, depending on the depicted content, target audience (adults, children) and the cultural background, may be inappropriate or even illegal. Consequently, authorities and content providers show an increased interest in automated ways of detecting critical visual content. This work addresses the problem and introduces methods suited for the detection of adult content and specific logo types (swastika). The proposed framework is based on a Bag-of-Words (BoW) model and employs representations encoding color and structural information. We present results and demonstrate that, despite the great visual ambiguity with respect to pose and appearance in the data, critical images can be well recognized in large datasets.

2 Methods

For the recognition of logos and adult content we employ the Scale-Invariant Feature Transform (SIFT) descriptors sampled at Difference of Gaussians (DoG) interest points. A visual vocabulary is built using hierarchical k-means clustering which serves for mapping the feature descriptors to instances of this vocabulary using soft assignment. The Term Frequency – Inverse Document Frequency (TF-IDF) weighting scheme is applied in order to emphasize the contribution of discriminative words and downweight the importance of those words which appear frequently in all images.

The novelty of our method is the combination of this approach with the Compactness Descriptor (CD) [4] and the Color Structure Descriptor (CSD) [2] incorporating (skin) color and structural information to improve the detection rate for adult content images. The classification is performed based on the results of support vector machines trained on positive and negative instances of the given object class for each feature type using a late fusion strategy.

3 Results

The FlickrLogos-32 dataset [3] was used for the evaluation of the logo filter. Both, the training and the testing set consist of 3,000 distractor images and 32 different logo classes with 30 images each. In order to test the classification

of critial content we extended the dataset by 30 cropped training images of swastikas and 30 images of swastikas with clutter which were obtained using Google image search. Figure 1 shows the performance of the method. When using single instance training based on synthesized samples of an image the performance decreases for true positive rates greater than 73%.

For the adult content classification we used the dataset from Shih et al. [4] and replaced the distractor images with those from the MIRFLICKR-25000 [1] dataset in order to increase the difficulty. For the training of the classifier we used 1,000 distractor and 100 adult content images whereas for testing we selected 10,000 distractor and 1,000 adult content images. In Figure 2 it can be seen that for high true positive rates the combinations of SIFT and CD as well as SIFT and CSD outperform the other methods.

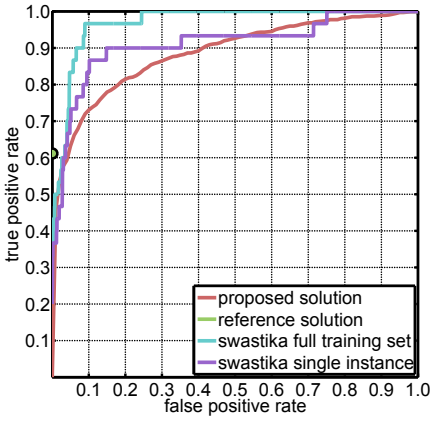

Fig. 1: FlickrLogos-32 dataset. The ROC curves represent the performance of all classes of the proposed (red) and the reference solution [3] (green marker), as well as the full (turquoise) and the single instance swastika training (violet)

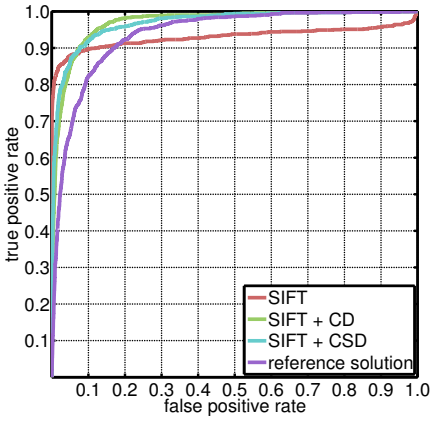

Fig. 2: Adult content dataset. The ROC curves represent the performance of the proposed solution using SIFT (red), SIFT + CD (green), SIFT + CSD (turquoise) and the reference solution [4] (violet)

References

1. Huiskes, M.J., Lew, M.S.: The MIR flickr retrieval evaluation. In: Proceedings of the ACM International Conference on Multimedia Information Retrieval (2008)
2. Messing, D.S., Beek, P., Errico, J.H.: The mpeg-7 colour structure descriptor: Image description using color and local spatial information. In: Proceedings of the International Conference on Image Processing. pp. 670–673 (2001)
3. Romberg, S., Pueyo, L.G., Lienhart, R., van Zwol, R.: Scalable logo recognition in real-world images. In: Proceedings of the 1st ACM International Conference on Multimedia Retrieval. pp. 25:1–25:8 (2011)
4. Shih, J.L., Lee, C.H., Yang, C.S.: An adult image identification system employing image retrieval technique. Pattern Recognition Letters 28(16), 2367–2374 (2007)