Boosting Colored Local Features For Generic Object Recognition

D. Hegazy, J. Denzler
Chair of Computer Vision, Institute of Computer Science,
Friedrich-Schiller-University, Jena, Germany
{hegazy,denzler}@informatik.uni-jena.de

Inclusion of local color information in generic object recognition is ignored by almost all the approaches, although it is important and can improve the recognition performance. In this paper, we present an object-class recognition approach using boosting as a learning technique. Simple local color descriptors combined with the SIFT descriptors are used. Experiments using benchmark and complex object-class datasets are performed and good performance is obtained.

Introduction

The object recognition problem has challenged the computer vision community for long time due to the huge change in the scale, occlusion and lighting conditions which has a great effect on the appearance of the objects. The problem of generic object recognition (also called object-class recognition) inherits the difficulties of the object recognition problem in addition to the intra-class and inter-class variability problems. Despite the difficulties of the generic object recognition problem many approaches appeared trying to provide a solution to this problem [1, 2, 3, 4, 5, 6, 7]. Most of the approaches do not include color information in their recognition. In this paper, we propose a generic object recognition model using one layer boosting as the underlaying learning technique. Combination of the SIFT descriptors and simple local color descriptors [8] is used.

The Recognition Model

In our generic recognition model objects from a certain class in still images are to be recognized. The objects are not segmented before the learning process nor information about the location or position of the objects within the images are given in the learning. Figure 1 gives a brief description of our recognition approach. In the first step, interest regions are detected in the training images. We used the Hessian-affine interest point detector[1]. Then, local descriptors are extracted from the interest regions. Two types of local descriptors are used: The SIFT descriptors [9] and local color descriptors presented in [8]. The local descriptors together with the labels of the

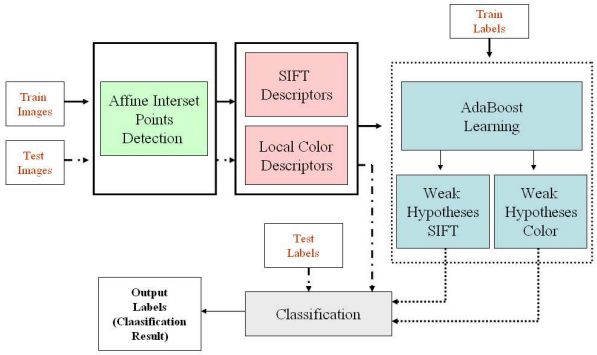

Figure 1: The proposed object-class recognition model

[1] http://www.robots.ox.ac.uk/ vgg/research/affine/

training images are given to the boosting learning technique [10] and produces final classifier (final hypotheses) as an output which predicts if relevant object is presented in the new test image.

Local Descriptors

The first type of descriptors we use is the Scale invariant Feature Transform (SIFT) descriptor which is widely used texture-based feature introduced by Lowe [9].
The second type is local color descriptor presented by van de Weijer and Shmid [8]. They introduced a set of local color descriptors with different criteria such as photometric robustness, geometric robustness, photometric stability and generality.
Among those descriptors introduced in [8], we chose to use opponent angle color descriptors as it is robust with respect to both geometrical variations caused by changes in viewpoint, zoom and object orientations and photometric variantions caused by shadows, shading and specularities. Brief description of how to construct it given (according to [8]) as follows :

$$ang_x^O = \arctan(\frac{O1_x}{O2_x}) \qquad (1)$$

where $O1_x$ and $O2_x$ are the derivatives of opponent colors and are given by:

$$O1_x = \frac{1}{\sqrt{2}}(R_x - G_x), O2_x = \frac{1}{\sqrt{6}}(R_x + G_x - 2B_x) \qquad (2)$$

and R_x, G_x and B_x are the derivative of color channels. The opponent colors and their derivatives are proven to be invariant with respect to specular variantion [8].
Before computing the previously mentioned invariant, color illumination normalization should be first done as described in [8] To construct the opponent angle descriptor, the derived invariant is transformed into robust local histogram. This is done by adjusting the weight of a color value in the histogram according to its certainty as in [8] (photometric stability). The resulting opponent angle descriptor of is length 37.

The Learning Algorithm

In our recognition model, objects form a certain class (category) in still images are recognized. Therefor, the used learning algorithm predicts if a given image contains an instance (object) from this category or not. Adaboost is used as the leaning algorithm in our recognition model. AdaBoost is a supervised learning algorithm, which takes a training set $I = \{I_1, ..., I_N\}$ and their associated labels $l = \{l_1, ..., l_N\}$, where N is the number of the train images and $l_i = +1$ if the object in the training image I_i belongs to the class category and l_i = -1 otherwise.
Each training image is represented by a set of features $\{F_{i,j}(t_{i,j}, v_{i,j}), j = 1...n_i\}$ where n_i is the number of features in image I_i, $t_{i,j}$ indicates the type of the feature (s for SIFT and c for Color) and $v_{i,j}$ is the feature value. The AdaBoost puts weights on the training images and requires construction of a weak hypothesis h_k which, relative to the weights, has discriminative power. AdaBoost is run for a certain number of iterations T and in each iteration k one feature is selected as a weak classifier and weights of the training images are updated. In our model, the AdaBoost in each iteration selects two weak hypothesis: one for the SIFT descriptor h_k^s and one for the color descriptor h_k^c . Each weak hypothesis consists of two components: a feature vector v_k^x and a certain threshold θ_k^x (a distance threshold) where $x = s$ for the SIFT and $x = c$ for color. The threshold θ_k^x measures if an image contains a descriptor $v_{i,j}$ that is similar to v_k^x. The similarity between $v_{i,j}$, which belongs to the image I_i , and v_k^x is measured using Euclidean distance for both descriptor types .

The learning algorithm using AdaBoost is as follows:

1. The weights of the training images are initialized to 1.
2. Calculate the weak hypotheses h_k^s and h_k^c with respect to the weights using the methode descriped in [2].
3. Calculate the classification error as:

$$\varepsilon_k = \frac{\sum_{i=1}^{N}(h_k(I_i) \neq l_i)w_i}{\sum_{i=1}^{N} w_i} \qquad (3)$$

4. Update the weights: $w_{i+1} = w_i \beta^{-l_i h_k(I_i)}$ for $i = 1$ to N
   and $\beta = \sqrt{\frac{1-\varepsilon_k}{\varepsilon_k}}$.
5. Repeat the steps 2, 3 and 4 for number of iterations T.

After the T iterations, the final hypotheses (classifier) is given by

$$H_I = \begin{cases} +1 & \text{if } \sum_{k=1}^{T} (\ln \beta_k) h_k(I) > \Omega \\ -1 & \text{Otherwise} \end{cases} \quad (4)$$

where Ω is varied to get various points for the ROC curve.

Table 1: ROC-eqq-err rates of our results using the Caltech dataset

Dataset	SIFT	Opp. ang.	Combination
Motor	86%	82%	93%
Cars	90%	86%	94%
Airplanes	78%	80%	84%
Faces	96%	94%	100%

Table 2: ROC-eqq-err rates of our results using the Caltech dataset compared to other famous approaches

Dataset	Ours	[2]	[5]	[7]	[6]
Motor	93%	92.2%	92.5%	93.2%	88%
Cars	94%	91.1%	90.3%	-	86.5
Air-planes	84%	88.9%	90.2%	83.8%	-
Faces	100%	93.5%	96.4%	83.1%	93.5%

Table 3: ROC-eqq-err rates of our results using the Graz02 dataset compared to the results of opelt [2]

Dataset	SIFT	Our Comb.	[2] SIFT	[2] Comb.
Bikes	80%	80%	76.4%	77.8%
Cars	77.33%	78.62%	68.9%	70.5%
Persons	81.33%	84%	70.0%	81.2%

Experiments and Results

We evaluate our recognition model using two datasets, namely the Caltech[2] and Graz02[3] datasets.
To compare our results to the existing approaches, we first used the Caltech dataset to evaluate our recognition model. We used, for training, 100 images of the object class as positive examples and 100 images class counter-class as negative examples. For testing, 50 positive examples and 50 negative examples are used. The features of each image are clustered to 100 cluster centers using the k-means clustering algorithm.
First, we evaluated our model using only one descriptor at a time to be able to notice the benefits of combining the two descriptors together then we used a comination of them. The performance is measured in ROC-equal-error rates and is shown in table1. We did not used the background dataset as a counter-class[4] because it is not colored. We used the leaves dataset instead. It is more difficult than the background as it contains more interest points [3]. Table one shows the improvements we gained in performance from combining the two descriptors together. Table 2 shows the comparison of the performance of our recognition approach and the state-of-the-art approaches. The comparison shows that our approach outperforms the state-of-art approaches in almost all the datasets.

Further experiments on our recognition approach are performed using Graz02 dataset. Graz02 dataset is more difficult than the Caltech. The objects are shown on complex cluttered background, at different scale and with different object position. The images include high occlusion up to 50% [2]. So, we used 150 positive and 150 negative images for training and 75 positive and 75 negative images for testing as in [2]. The features of each image are clustered to 100 cluster centers using the k-means clustering algorithm. Table 3 shows

[2]http://www.vision.caltech.edu/html files/archive.html

[3]http://www.emt.tugraz.at/ pinz/data

[4]Used by almost all the recognition approaches as the counter-class.

the results and compare them to the results of combination in [2].
As shown in table 3, our results (combination) exceeds the results of [2](combination) in all the datasets. As we mentioned we use only one interest point detector with two local descriptors but in [2], three point detectors are combined with four local descriptors are used. Also, the results shows that our choice of hessian-affine as the underlaying point detector is a good choice as the results obtained using it with the SIFT exceeds the results of [2] using the DoG with the SIFT.

Conclusions

We have presented a generic object recognition model which is based on appearance information of the objects. The contribution of our model is adding simple color features to the recognition together with the SIFT descriptors which is more realistic, as color is an important part in describing the appearance of any object. Adaboost is used for learning and the performance of the model exceeds in almost all the cases the performance of the state-of-the-art generic object recognition approaches.

References

[1] P.Viola and M.Joens, Rapid object detection using a boosted cascade of simple features. // In IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR01.- 2001.-PP. 1:511-518 .

[2] A. Opelt, A. Pinz, M. Fussenegger and P. Auer, Generic object recognition with boosting. // IEEE Transactions on Pattern Analysis and Machine Intelligence.-2006.- Vol. 28, No. 3 .

[3] W. Zhang, B. Yu,G. J. Zelinsky and D. Samaras, Object class recognition using multiple layer boosting with heterogeneous features," // In IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR05.- 2005.-PP.66-73.

[4] S. Agarwal and D. Roth, Learning a sparse representation for object detection. // In 5th European Conference on Computer Vision ECCV02.- 2002.-PP.113-130.

[5] R. Fergus, and P. Perona and A. Zisserman, Object class recognition by unsupervised scale-invariant learning," // In IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR03.- 2003.

[6] M. Weber,M. Welling and P. Perona, Unsupervised learning of models for recognition," // In 4th European Conference on Computer Vision ECCV00.- 2000.-PP.264-721.

[7] J. Thureson and S. Carlsson. Appearance based qualitative image description for object class recognition. // In 7th European Conference on Computer Vision ECCV04.- 2004.- Vol.II.-PP. 518-529 .

[8] J. van de Weijer and C. Shmid. Coloring local feature extraction. // In 9th European Conference on Computer Vision ECCV06.- 2006.-PP.334-348.

[9] D. G. Lowe, Distinctive Image Features from Scale-Invariant Keypoints. // International Journal of Computer Vision.- 2004.- Vol. 60.- PP. 91-110.

[10] Y. Freund and Schapire. A decision theoretic generalization of online learning. // Computer and System science.-1997.- 55(1): 119-139.