

# TECHNIQUES FOR IMAGE CLASSIFICATION, OBJECT DETECTION AND OBJECT SEGMENTATION

Ville Viitaniemi and Jorma Laaksonen



TEKNILLINEN KORKEAKOULU  
TEKNISKA HÖGSKOLAN  
HELSINKI UNIVERSITY OF TECHNOLOGY  
TECHNISCHE UNIVERSITÄT HELSINKI  
UNIVERSITE DE TECHNOLOGIE D'HELSINKI



# TECHNIQUES FOR IMAGE CLASSIFICATION, OBJECT DETECTION AND OBJECT SEGMENTATION

Ville Viitaniemi and Jorma Laaksonen

Helsinki University of Technology  
Faculty of Information and Natural Sciences  
Department of Information and Computer Science

Teknillinen korkeakoulu  
Informaatio- ja luonnontieteiden tiedekunta  
Tietojenkäsittelytieteen laitos

Distribution:

Helsinki University of Technology  
Faculty of Information and Natural Sciences  
Department of Information and Computer Science  
P.O.Box 5400  
FI-02015 TKK  
FINLAND  
URL: <http://ics.tkk.fi>  
Tel. +358 9 451 1  
Fax +358 9 451 3369  
E-mail: [series@ics.tkk.fi](mailto:series@ics.tkk.fi)

© Ville Viitaniemi and Jorma Laaksonen

ISBN 978-951-22-9421-3 (Print)  
ISBN 978-951-22-9422-0 (Online)  
ISSN 1797-5034 (Print)  
ISSN 1797-5042 (Online)  
URL: <http://lib.tkk.fi/Reports/2008/isbn9789512294220.pdf>

TKK ICS  
Espoo 2008

**ABSTRACT:** In this paper we document the techniques which we used to participate in the PASCAL NoE VOC Challenge 2007 image analysis performance evaluation campaign. We took part in three of the image analysis competitions: image classification, object detection and object segmentation. In the classification task of the evaluation our method produced comparatively good performance, the 4th best of 19 submissions. In contrast, our detection results were quite modest. Our method's segmentation accuracy was the best of all submissions. Our approach for the classification task is based on fused classifications by numerous global image features, including histograms of local features. The object detection combines similar classification of automatically extracted image segments and the previously obtained scene type classifications. The object segmentations are obtained in a straightforward fashion from the detection results.

**KEYWORDS:** image classification, object detection, object segmentation, benchmarking



## CONTENTS

1	Introduction	1
2	Image Content Analysis Tasks of VOC Challenge 2007	1
2.1	Image Data . . . . .	2
2.2	Learning Tasks and Performance Measures . . . . .	3
3	Framework for Object Classification	3
4	Approach to the Classification Task	5
5	Approach to the Detection Task	6
5.1	Image Segmentation . . . . .	7
5.2	Features . . . . .	8
6	Approach to the Segmentation Task	8
7	VOC Challenge 2007 Results	9
8	Conclusions and Discussion	9
	References	10



## 1 INTRODUCTION

Huge masses of digital visual information are produced nowadays, both automatically and also by the ever increasing numbers of ordinary people with access to easily usable tools for creating personal digital content. Automatic image analysis techniques are called for to analyse and organise these overwhelming sea of information. Especially useful would be methods that could automatically analyse the semantic contents of images and videos as it is just the content that determines the relevance in most of the potential uses. One important aspect of image content is the object composition: the identities and positions of the objects the images contain. This paper discusses techniques for recognising and locating objects of some particular semantic class in images when example images of that semantic class are given.

Quite often data analysis researchers work with data sets of their own. The problems the individual researchers try to solve may be very similar and there could be significant synergy in the solutions. However, due to different data sets the relative effectiveness of various approaches may be difficult to compare. Standardised benchmark tasks and performance evaluation campaigns are partial solutions to this problem.

In the remaining parts of this paper we first briefly describe the tasks of PASCAL NoE VOC Challenge 2007 image analysis performance evaluation campaign that focuses on the object composition of photorealistic images. This is done in Section 2. We then document the image content analysis techniques that we used to participate the campaign. The techniques are to a large part collected from our earlier papers [12, 13]. However, many improvements and adaptations to the current tasks are also described. Section 3 outlines our generic software framework for image similarity assessment. In the subsequent Sections 4, 5 and 6 we concretise how the framework is applied to image classification, object detection and object segmentation tasks, respectively, of the VOC Challenge 2007. The results of the challenge are summarised in Section 7. Finally in Section 8 we present some conclusions and discussion.

## 2 IMAGE CONTENT ANALYSIS TASKS OF VOC CHALLENGE 2007

The image content analysis techniques addressed in this paper have been used to participate the Pascal Visual Object Classes Challenge 2007 [3]. In the challenge—organised also in 2005 and 2006— machine learning systems are compared by their ability to recognise objects from a number of visual object classes in realistic scenes. The problem is formulated as a supervised learning problem in which a training set of labelled images is provided. The 2007 challenge defines four competitions: two main competitions in image classification and object detection, and two smaller-scale taster competitions in object segmentation and person layout detection. We participated only the first three competitions, not the person layout taster. The following description of data and the tasks thus excludes aspects relevant for that task only.

Table 1: The 20 object classes of VOC Challenge 2007

Category	object classes
Person	person
Animal	bird, cat, cow, dog, horse, sheep
Vehicle	aeroplane, bicycle, boat, bus, car, motorbike, train
Indoor	bottle, chair, dining table, potted plant, sofa, tv/monitor

## 2.1 Image Data

The image collection for the 2007 challenge consists of 9963 photographic images of natural scenes. Each of the images contains at least one occurrence of the 20 object classes detailed in Table 1. In many images there are objects of several classes present. Altogether the images contain 24640 objects. The most common object class (“person”) is present in 40% of the images, the rarest (sheep) in 1.9%. Often there are several objects of the same kind in an image, e.g. 2.3 persons and 2.7 sheep on average among the images where the named objects are present at all. Figure 1 shows some examples of the images and objects.



Figure 1: Examples of VOC Challenge 2007 images and their annotations

For the main competitions, the image collection is partitioned equally into training (denoted “trainval”) and test sets. The training half of the data is further divided evenly into two sets (“train” and “val”) that the organisers of the challenge suggest to be used for development and validation of techniques. The images of the training set have been manually annotated with the bounding boxes of all the occurrences of the 20 object classes. Furthermore, some of the bounding boxes include information about the pose of the object and some have been tagged as “truncated” and/or “difficult” by the human annotators. The “difficult” objects were left out when evaluating the results, and we excluded them also from the training of our methods. We did not use the pose or truncation information at all. Similar annotation of the test data has been released after the challenge by its organisers.

Manually annotated pixel-wise segmentation masks of approximately 8% of the training images were additionally provided as training material for the segmentation taster competition. Apparently a somewhat smaller proportion of the test images were manually segmented in a similar fashion. All the 20 object classes are present in the segmented images. In our approach to the segmentation task we did not use the pixel mask information for training, but relied on the bounding box annotations only.

## 2.2 Learning Tasks and Performance Measures

We consider here three different types of tasks on the image sets. In the classification task 20 rankings of the test set images are produced, each one corresponding to one of the object classes. In each ranking the goal is to order the test images according to descending likelihood of presence of the corresponding object in the images. In the detection task similar rankings are produced for the bounding boxes of the object classes. There is no limitation on the number of bounding boxes that can be reported. In the segmentation task the goal is to label each pixel of the test images to one of the 20 object classes or to the background class.

The rankings produced in the main competitions are compared using precision-recall (PR) curves. To generate the curves, the boolean valued correctness of each entry in the ranking has to be determined. In the classification task this is a matter of straightforward comparison with the ground truth annotations. In the detection task the predicted bounding boxes are considered as true or false positives based on the relative area of overlap with ground truth bounding boxes. To be regarded as a correct detection, the area of overlap  $a_o$  between the predicted bounding box  $B_p$  and ground truth bounding box  $B_{gt}$  must exceed 50% by the formula

$$a_o = \frac{\text{area}(B_p \cap B_{gt})}{\text{area}(B_p \cup B_{gt})}. \quad (1)$$

Subsequent detections after the first one overlapping the same ground truth bounding box are considered false detections. For the rankings, interpolated average precision (AP) is used as the quantitative evaluation measure.

The segmentation accuracy is calculated for each object class by determining the percentage of actual pixels of the class that have received the correct label. Overall accuracy is evaluated as the mean of the accuracies over the 20 object classes and the background class.

## 3 FRAMEWORK FOR OBJECT CLASSIFICATION

In the following description of our software framework we use the term target object to denote training and test set images in the case of the classification task and automatically extracted image segments in the case of the object detection task. Our method for tackling the VOC Challenge tasks is based on assessing the similarity of the visual properties of the test set target objects to the properties of the training set target objects. Figure 2 schematically outlines the general system architecture used for supervised similarity assessment. As the input, the framework takes a binary partitioning of the target objects of the training set into two classes. As the output a real number is produced for each of the target objects in the test set reflecting the estimated likelihood of the target object to belong to the first one of the classes of the binary partitioning. This list of real numbers can be sorted in descending order to produce a similarity ranking of the test set target objects.

In the first stage of the architecture the visual properties of each target object are described simultaneously with several elementary feature vectors,

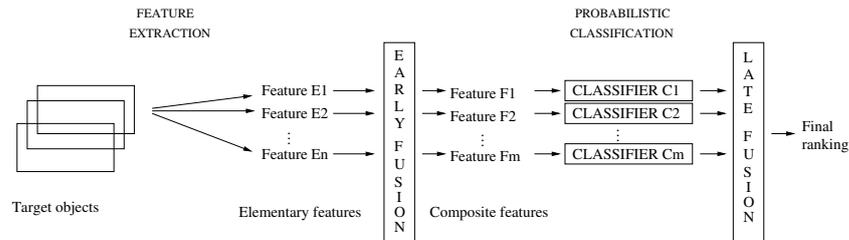


Figure 2: The general architecture of the supervised similarity ranking framework

each one reflecting a different aspect of the appearance, such as colour or texture. The same set of fixed-length feature vectors is extracted from the target objects in both training and test sets. More feature extraction details are found in the subsequent Sections 4 and 5.2.

The next stage of the system is early fusion of the elementary feature vectors to produce composite features. For the system used in VOC 2007 this is realised by forming combinations of the original features by concatenating the corresponding feature vectors.

The resulting sets of composite feature vectors are fed into supervised probabilistic classifiers. There is a separate classifier for each feature combination. Based on the partitioning of the training set, each classifier estimates the partial likelihood of the test set object in light of the corresponding composite feature. As the feature-wise classifiers we used weighted C-SVC variants of the SVM algorithm, implemented in the version 2.84 of the software package LIBSVM [2]. The outputs of the SVMs are normalised to probability estimates by using the method of [15]. As the kernel function we used the  $\chi^2$ -kernel for histogram type features and the RBF kernel for all other features. The free parameters of the C-SVC cost function and the kernel function are chosen on basis of a search procedure that aims at maximising the six-fold cross validated area under the receiver operating characteristic curve (AUC) measure in the training set. The search procedure performs importance sampling in the parameter space by a suitably scaled increasing function of AUC. In the cost function of C-SVC the weight parameter  $w$  compensates for the unbalanced class distribution in the training set. For this parameter, we use a heuristically chosen value  $w = 2$ . To limit the computational cost of the classifiers, we perform random sampling of the training set, in which we suppress the more frequent class. Some more details of the classification stage can be found in [13].

Finally, in the last stage the outputs of separate classifiers are fused together. This is achieved by concatenating the smoothed logarithms of the probability estimates for each target object into a single vector and training an additional SVM layer with RBF kernel to perform the classification using these concatenations as feature vectors. Probability estimates for the training set target objects are obtained with six-fold cross-validation. The output of the SVM is once again converted to a probability estimate [15].

Table 2: Some of the elementary features extracted from the images

Feature	Tiling	Dim.
DCT coefficients of average colour in rectangular grid	global	12
CIE L*a*b* colour of two dominant colour clusters	global	6
Histogram of local edge statistics	4 × 4	80
Haar transform of quantised HSV colour histogram	global	256
Average CIE L*a*b* colour	5	15
Three central moments of CIE L*a*b* colour distribution	5	45
Co-occurrence matrix of four Sobel edge directions	5	80
Magnitude of the 16 × 16 FFT of Sobel edge image	global	128
Histogram of four Sobel edge directions	5	20
Histogram of relative brightness of neighbouring pixels	5	40

## 4 APPROACH TO THE CLASSIFICATION TASK

In this section we describe how the framework described in the previous section was applied to the classification task described in Section 2.2. The target objects in the case of the classification task are images. The partitionings of the training images were produced in one-against-all fashion, thus resulting in one ranking per each object class.

As a basis for the classification, a set of elementary visual features was extracted from the images. Ten of the extracted feature types are listed in Table 2. Dimensionalities of the feature vectors are given in the rightmost column of the table. The first four rows correspond to features that more or less closely resemble the ColorLayout, DominantColor, EdgeHistogram and ScalableColor features of the MPEG-7 standard [5]. The column “Tiling” of the table shows that some of the features are calculated truly globally, such as the global colour histogram feature of the fourth row, others, such as the edge histogram feature of the third row, encode some spatial information by using a fixed image grid. These feature vectors are formed by concatenating the descriptors of individual tiles. The features calculated for five tiles employ a center-surround tiling mask where the image area is divided into four tiles by the two diagonals of the image, on top of which a circular center tile is overlaid.

In addition to the tabulated features, the set of elementary visual features was extended with histograms of interest point features. The interest points were detected using a Harris-Laplace detector [9]. The SIFT feature [8], based on local gradient orientation, was calculated for each interest point. Histograms of interest point features have proven to be efficient in image content analysis and are gaining popularity in various image analysis tasks [4, 10]. We form histograms of the SIFT features according to codebook vectors selected using the self-organising map (SOM) algorithm [6]. We include several sets of histogram features in the feature set.

As basic versions of the histogram features we quantise the interest point features with SOM codebooks of five different sizes:  $16 \times 16$ ,  $25 \times 20$ ,  $40 \times 25$ ,  $50 \times 40$  and  $80 \times 50$ . We employ also codebooks that use knowledge of the object classes. We create these codebooks by first forming separate codebooks for each object class with the SOM algorithm and then concatenating the 20 class-wise codebooks. Features resulting from the use of two such codebooks

with sizes of  $20 \times 16 \times 16$  and  $20 \times 25 \times 20$  units were included in the feature set.

Further additions to the feature set were obtained by spatially subdividing the images into fixed tiles and forming histograms for each tile. We used rectangular grids of sizes  $2 \times 2$ ,  $3 \times 3$  and  $4 \times 4$  as well as the previously described five-part center-surround tiling. Number of histograms of this kind was multiplied by the use of codebooks of different sizes and different histogram normalisation methods. Following the idea presented in [7], the spatially subdivided histograms were additionally concatenated to form spatial pyramids. Altogether five different pyramids both with and without the proposed exponentially diminishing weighting of finer resolutions.

The set of elementary features is rather large. It would probably not be expedient to extract such many features in order to just classify images in a database, although fusing large set of features seems to be a safe route towards comparatively good accuracy. When doing classifier fusion, the accuracy gain from adding new features to the feature set tends to saturate at some point whereas the computational costs of feature extraction naturally keep growing linearly. However, as we already had extracted the features anyway to experiment with different alternatives, fusing them all did not incur very significant additional costs.

The feature combinations that were formed in the early fusion stage were chosen rather ad hoc. We formed some pairs, triples and quadruples of the global and semi-global features listed in Table 2. The histogram features were passed through the early fusion stage as such. Additionally the global histogram with  $40 \times 25$ -sized codebook was paired with all the tabulated features. All in all, the classifiers were formed for 141 composite features.

Based on 6-fold cross-validated classification performance in the “train” half of the “trainval” set, we performed some feature selection. We employed greedy sequential forward/backward search, starting from empty feature set and performing feature additions and removals until the performance no longer improved. Inside the search loop we used product-based fusion instead of SVM fusion as the former is computationally essentially lighter, but performs not much worse according to our observations. The feature selection procedure is not fully satisfactory but still somewhat successful: compared to the fusion of all the formed classifiers the classification performance in the validation set remains approximately the same, but the required number of variables is greatly diminished to only 3–28 variables for a single object class. To perform the final classification of the test data, we fused together the union of the variable sets selected for each of the 20 classes, 93 variables altogether.

## 5 APPROACH TO THE DETECTION TASK

The detection task was addressed by first segmenting the images as will be detailed in Section 5.1 and then assessing the likelihood of each segment to correspond to the targeted object class. This problem is factored into product of two parts: 1) the likelihood of the image containing the segment to contain the object somewhere, and 2) the conditional probability of just the consid-

ered segment to correspond to the targeted object, given that the object is known to appear in the image. The former likelihood was readily available from the solution to the classification task. The latter one was obtained by applying the framework outlined in Section 3 with image segments as target objects.

More formally, let  $r$  be an image segment,  $I_s$  and  $I_i$  binary indicator variables for the segment  $r$  and the corresponding image, respectively, belonging to certain object class. The trained classifier outputs two confidence values:  $c_i$  for the segment  $r$  being a true detection, and  $c_s$  for  $r$  being contained in an image belonging to the class. Now we write for the probability of  $r$  being a true detection

$$p(I_s = 1|c_i, c_s) = p(I_s = 1|I_i = 1, c_i, c_s)p(I_i = 1|c_i, c_s). \quad (2)$$

With rather plausible independence assumptions this can be approximated as

$$p(I_s = 1|c_i, c_s) \approx p(I_s = 1|I_i = 1, c_s)p(I_i = 1|c_i). \quad (3)$$

To estimate the first probability of the product, the framework was used with just the training image segments belonging to images of the targeted object class, thus balancing the rather high computational requirements of the detection task with numerous segments of a single image.

Partitioning the training segments to two classes according to their overlap with the ground truth classes is not all that straightforward. We investigated the matter earlier [12], with the finding that the best strategy varies with the targeted object class. The same applies to the selection of the way to segment the images. Unfortunately, within the time limits of the 2007 challenge we were not able to implement all the alternative techniques for segmentation and segment selection. Therefore we tried to select one set of techniques that would work rather well for most of the classes. The selected segmentation technique is detailed in the next section. The binary partitioning of the training sets was done so that segments meeting the 50% overlap criterion were used to form the first of the two classes.

Due to lack of time also the heuristic propagation of segment probabilities along segmentation hierarchies—found to be very beneficial in [12]—was left unimplemented. The heuristic to discourage multiple detections of the same object was implemented, however. We exponentially discounted subsequent detections when an object had already been detected in the same image. This is relevant to our segmentation strategy that produces spatially overlapping segments. To produce final rankings of bounding boxes—in contrast to irregularly shaped image segments—we simply listed the bounding boxes of the segments in the rankings.

## 5.1 Image Segmentation

The images were segmented with a generic image segmentation method which is simple and somewhat rudimentary. The method employs an area-based region merging algorithm to construct segments that are homogeneous in terms of colour and texture. The images in the database were segmented in two steps. In the first step the  $K$ -means algorithm was used to compute

an oversegmentation of the images. In the second step the segments were merged. The merging algorithm had been selected to give visually feasible results for photographs and other images in earlier applications. The merging was continued until 25 segments were left. In addition to the basic segments, we also recorded the hierarchical segmentation that resulted from continuing the region-merging algorithm until only one region remained. All the intermediate segments were considered as potential bounding box locations.

## 5.2 Features

Due to limitations in real and computer time we were able to dedicate to the detection task, the feature set that was used to describe the visual properties had to be kept quite small as the combined number of the segments of all the database images is rather large—in the order of half a million. Table 3 lists the elementary features extracted from the segments. The features are a subset of the global image features mentioned in Section 4 with the exception of two features describing the shape of the segments: Zernike moments of the segment area and Fourier descriptors of the segment contour. One of the SIFT feature histograms is formed from the interest points residing within the segment borders, the other one includes all the interest points inside the bounding box of the segment. For both of the histogram features we used the  $25 \times 20$  SOM codebook determined from all the training images. In the early fusion stage the set of features was enriched by forming twelve additional combinations of the elementary features.

Table 3: Elementary features extracted from image segments

DCT coefficients of average colour in rectangular grid
CIE L*a*b* colour of two dominant colour clusters
Average CIE L*a*b* colour
Moments of CIE L*a*b* distribution
Co-occurrence matrix of four Sobel edge directions
Histogram of four Sobel edge directions
Histogram of relative brightness of neighbouring pixels
Fourier descriptors of the segment contour
Zernike moments of the segment shape
Histogram of SIFT features within the segment
Histogram of SIFT features within bounding box of the segment

## 6 APPROACH TO THE SEGMENTATION TASK

The submitted results for the segmentation taster task were synthesised in a straightforward manner from the bounding boxes detected in the detection task. For each pixel we checked all the detected bounding boxes containing the pixel. We set the label of the pixel to be that of the bounding box with the largest estimated probability if the probability exceeded a threshold. Otherwise the pixel was labelled as background. We also noticed that our estimated probabilities of detected “person” bounding boxes were scaled incorrectly. As a quick fix we excluded the class “person” completely from the

labelling process. In unsubmitted experiments we also tried using the actual bit masks of the segments instead of their bounding boxes for labelling, but in the validation set this made the performance slightly worse.

## 7 VOC CHALLENGE 2007 RESULTS

In this section we briefly report the results of those VOC Challenge 2007 competitions we participated. For more detailed results see [3]. In the result graphs and tables the participants are identified by abbreviations. The abbreviation for the methods described here is “TKK”.

Figure 3 shows the median APs of the classification competition participants over all the 20 object classes. In Figure 4a the submissions to the object detection competition are ranked according to number of top rankings they achieved in detecting objects of the 20 classes. Figure 4b displays the mean accuracies achieved in the segmentation taster competition. This performance measure used in the segmentation task is quite rudimentary in our opinion. However, we consider this acceptable considering the initial and interest-probing nature of this taster competition.

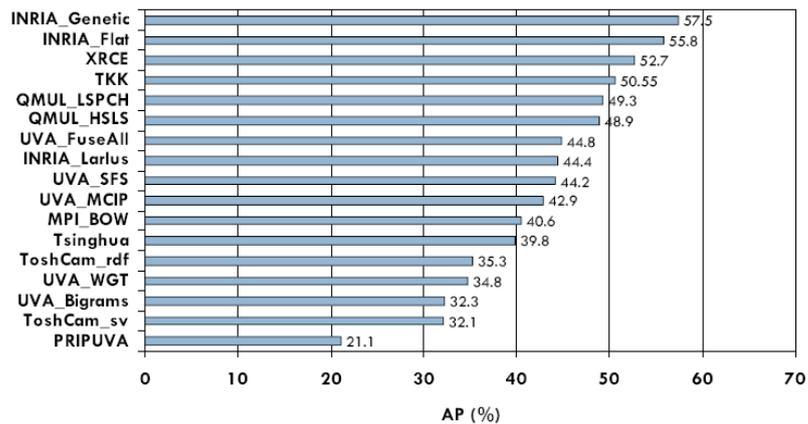


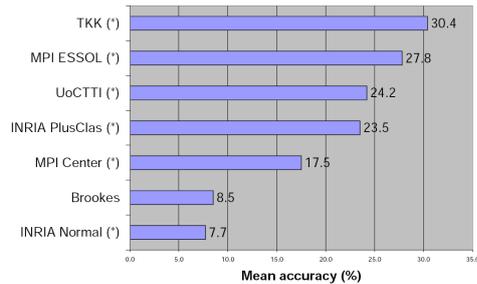
Figure 3: Results of the VOC Challenge 2007 classification competition: the median AP of all the 20 object classes. The figure is from [3].

## 8 CONCLUSIONS AND DISCUSSION

In this paper we have documented the image content analysis system that we have used to participate the VOC Challenge 2007 image analysis performance evaluation campaign. In the classification task the performance of our method was relatively competitive, the 4th best of 19 participants. It appears to be that the other top participants have used more effective variants of the histogram features and their classification, such as dense sampling of the images instead of interest points and colour information in addition to the monochromatic local shape feature. In the future we are planning to investigate the different possibilities for improving our histogram features. In contrast, the feature fusion aspects of our architecture are not as extensive in

	1st	2nd	3rd
<b>Oxford</b>	6	-	-
<b>UoCTTI</b>	6	8	1
<b>MPI_ESSOL</b>	5	0	0
<b>IRISA</b>	2	1	4
<b>INRIA_PlusClass</b>	1	6	5
<b>INRIA_Normal</b>	0	1	5
<b>MPI_Center</b>	0	2	3
<b>TKK</b>	0	2	3
<b>Darmstadt</b>	0	0	0

(a)



(b)

Figure 4: a) The overall ranking of the participants of the VOC Challenge 2007 object detection competition according to number top three rankings in detecting of individual classes. The table is from [3], b) Mean segmentation accuracies in the segmentation taster competition. Asterisk (\*) denotes the entry being automatically generated from detection task results.

the competing systems. It deserves to be mentioned, however, that also the best performing classification system derives its power from combining several histogram features cleverly with a genetic algorithm. The use of feature fusion seems to be better developed in other kinds of information retrieval systems, for example the video retrieval systems that have recently participated the TRECVID campaign [10].

The performance of our method in the detection task was quite modest. This was partially expected as we were not able to dedicate enough of our time to develop methods for this task. Our impression is that the segmentation method we employed is not very good at finding objects or relevant object parts. This is supported by the finding that the use of actual bit masks of the segments instead of their rectangular bounding boxes degrades the performance in the segmentation task. After the challenge, it has been demonstrated [11] that the segmentation accuracy could be markedly improved by re-segmenting the bounding boxes of our detections. Hierarchical segmentation methods have been studied quite much lately [1, 14] and it might be beneficial to update our segmentation method according to recent findings.

In the light of modest the detection performance of our system the best overall segmentation accuracy among the challenge participants we obtained is rather surprising. Partly this may be due to the segmentation competition having been presented as a “taster” and the researchers thus devoting less attention on it. It still remains somewhat mysterious why the segmentations derived automatically from our less accurate detections were better than similarly derived segmentations from more accurate detections. One explaining factor might be that our detection results consist of sets of overlapping bounding boxes that fully cover every image.

## REFERENCES

- [1] Tomasz Adamek and Noel O’Connor. Stopping region-based image segmentation at meaningful partitions. In *Proc. of SAMT 2007*, volume 4669 of *LNCS*, pages 15–27, Genova, Italy, December 2007. Springer.

- [2] Chih-Chung Chang and Chih-Jen Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [3] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results. <http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html>.
- [4] M. Everingham, A. Zisserman, C. K. I. Williams, and L. Van Gool. The PASCAL Visual Object Classes Challenge 2006 (VOC2006) Results. <http://www.pascal-network.org/challenges/VOC/voc2006/results.pdf>.
- [5] ISO/IEC. Information technology - Multimedia content description interface - Part 3: Visual, 2002. 15938-3:2002(E).
- [6] Teuvo Kohonen. *Self-Organizing Maps*, volume 30 of *Springer Series in Information Sciences*. Springer-Verlag, Berlin, third edition, 2001.
- [7] Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Proc. of IEEE CVPR*, volume 2, pages 2169–2178, 2006.
- [8] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, November 2004.
- [9] K. Mikolajczyk and C. Schmid. Scale and affine point invariant interest point detectors. *International Journal of Computer Vision*, 60(1):68–86, 2004.
- [10] Paul Over, Tzveta Ianeva, Wessel Kraaij, and Alan F. Smeaton. TRECVID 2006 - an introduction. In *TRECVID Online Proceedings*. TRECVID, November 2006. <http://www-nlpir.nist.gov/projects/tpubs/tv.pubs.org.html>.
- [11] J. Shotton, M. Johnson, and R. Cipolla. Semantic texton forests for image categorization and segmentation. In *Proc. of IEEE CVPR*, 2008. Accepted.
- [12] Ville Viitaniemi and Jorma Laaksonen. Techniques for still image scene classification and object detection. In *Proc. of ICANN 2006*, volume 2, pages 35–44, Athens, Greece, September 2006. Springer.
- [13] Ville Viitaniemi and Jorma Laaksonen. Improving the accuracy of global feature fusion based image categorisation. In *Proc. of SAMT 2007*, volume 4669 of *LNCS*, pages 1–14, Genova, Italy, December 2007. Springer.
- [14] Veronica Vilaplana and Ferran Marques. Region-based hierarchical representation for object detection. In *Proc. of International Workshop on Content-Based Multimedia Indexing*, pages 157–164, Bordeaux, France, June 2007.

- [15] T.-F. Wu, C.-J. Lin, and R.C. Weng. Probability estimates for multi-class classification by pairwise coupling. *J. of Machine Learning Research*, 5:975–1005, 2005.







TKK REPORTS IN INFORMATION AND COMPUTER SCIENCE

TKK-ICS-R1     Nikolaj Tatti, Hannes Heikinheimo  
Decomposable Families of Itemsets. May 2008.

ISBN 978-951-22-9421-3 (Print)  
ISBN 978-951-22-9422-0 (Online)  
ISSN 1797-5034 (Print)  
ISSN 1797-5042 (Online)