Use of Image Regions in Context-Adaptive Image Classification*

Ville Viitaniemi and Jorma Laaksonen

Laboratory of Computer and Information Science, Helsinki University of Technology,
P.O.Box 5400, FIN-02015 TKK, Finland
{ville.viitaniemi, jorma.laaksonen}@tkk.fi

Abstract. In this paper we describe and discuss our existing PicSOM software framework from the point of view of context-adaptive analysis of image contents, especially its method for using automatic image segmentation. We describe and experimentally validate a modification to the segment-using procedure that both essentially reduces the computational cost and slightly improves classification accuracy. Finally, we apply the segment-using methodology in qualitatively investigating the roles of primary objects and their context in classifying the images of the Pascal VOC Challenge 2006 database.

1 Introduction

People are nowadays faced with constant and overwhelming stream of digital image and video content. Furthermore, the amounts of data being generated seems to be constantly increasing. Therefore, automatic methods are highly desirable to analyse and index the large data masses. Especially useful would be methods that could automatically interpret the semantic contents of images and videos as it is just the content that determines the usefulness of them for most purposes.

Our PicSOM software framework [8,11] is aimed at automatically organising and ordering large unannotated databases of audio-visual objects according to the similarity of their contents. The databases may include images, videos, texts and multi-part objects, such as mobile multimedia messages, emails and web pages. The objects in the database may be hierarchically related to each other, such as different parts of a multimedia message. In this paper, we focus on the case of image databases where the hierarchy consists of images and their segments. The underlying principle of the framework is to extract a large number of visual features from the database images. Then the aim is to find statistical dependencies between the visual features and the current goals of image analysis. The dependencies are not required to be deterministic rules, such as blue always

* Supported by the Academy of Finland in the projects *Neural methods in information retrieval based on automatic content analysis and relevance feedback* and *Finnish Centre of Excellence in Adaptive Informatics Research.*

Y. Avrithis et al. (Eds.): SAMT 2006, LNCS 4306, pp. 169–183, 2006.

corresponding to sky. On the contrary, even weak probabilistic correlations are sought after.

A natural approach to describing an image would be to list what different parts the image contains, and then possibly describe the major parts in more detail. In this light, partitioning the image to disparate parts and describing the image in terms of the content and relationship of these parts appears to be a promising approach. Indeed, image segmentation often is crucial in image understanding. The PicSOM method of using image segments is in line with the statistical approach to image analysis: a set of alternative segmentations is formed and the system is adaptively allowed to find the segmentations most beneficial to the image analysis task at hand.

On different occasions, the similarity of image content arises from different visual features, such as colour or shapes. Only a subset of all the imaginable features is relevant in a given image analysis task. It is the context of image analysis that determines the set of relevant visual features. For purposes of the PicSOM framework, we divide the context into two levels of specificity. The distinction is made to allow two different mechanisms of adaptation of the framework to the context. We call the broader level of context database context. With database we understand the collection of the images the content analysis is targeted at, whether or not they are actually collected in a database. We consider the database part of the context to be fixed, so the system may be adapted to the database level context beforehand, without needing to take the suitability of this specific system to other databases. A schematic example of database level adaptation to context is given by an imaginary database consisting solely of red and green apples. In the context of this database, it is not wise to allocate resources to a colour feature telling whether an object is blue. Task level context is the part of the context that changes so frequently that completely re-building the image analysis system would be impractical. For example, we may think of a database of vehicles of different colour. Different visual features are relevant, depending on whether we want to identify motorbikes, red automobiles, or any yellow vehicle in the database.

In this paper we consider the use of the adaptive PicSOM image analysis framework in the context of image classification task detailed in Section 2. Section 3 outlines the working principles of the framework in general, whereas in Section 4 we discuss the use of image segments more in detail and experimentally investigate of the usefulness in the image classification task. In Section 5 we describe how the method for using segments may be moved from the on-line part of the system to the off-line preprocessing and made computationally more lightweight.

Having performed the image classification task with help of image segments, we may reverse the direction of the analysis. In Section 6 we compare the the contribution of different image segments to the classification in the context of this image database. In Section 7 we present conclusions and discussion.

2 Image Analysis Task

To evaluate and motivate the techniques we discuss later in this paper, we consider a concrete image analysis task. The task consists of classifying images of the portion of the Pascal Visual Object Classes Challenge 2006[1] image set whose ground truth classifications have been made public at the time of this writing. This set of 2618 images contains realistic images of ten classes. The classes are defined by the absence or presence of an object, e.g. cat, in the images. The classes are partially overlapping, i.e. several different objects may appear in the same image.

The classification task is considered in a supervised setting: approximately one half of the images is used as training set and the rest as the test set. Table 1 shows the class statistics of the image sets. The classification performance is evaluated in terms of Area Under Curve (AUC) property of the classifier Receiver Operating Characteristic (ROC) curves.

Table 1. Statistics of image sets. Columns correspond to different object classes.

	bicycle	bus	car	cat	cow	dog	horse	motorbike	person	sheep	total
training set	127	93	271	192	102	189	129	118	319	119	1277
test set	143	81	282	194	104	176	118	117	347	132	1341

3 PicSOM Image Analysis Framework

3.1 Outline

In the PicSOM framework the database objects are divided into two groups: example and target objects. In a supervised learning context these correspond to training and test sets, respectively. In an interactive image retrieval setting the set of example and target objects is adapted dynamically during a retrieval session as a result of the user giving relevance feedback on the target objects the system retrieves. The framework is used to rank the target objects according to their similarity to a given set of positive example objects, simultaneously combined with the dissimilarity to a set of negative example objects.

The similarity of the objects is evaluated in terms of a large set of visual features of statistical nature. For this purpose the example and target images are pre-processed in the same manner: the images are first automatically segmented and a large collection of statistical features is extracted from both the segments and whole images. Several different segmentations of the same images can be used in parallel. Procedures for feature extraction and image segmentation are discussed in detail in Sections 3.2 and 4, respectively. In the current experiments, we consider the use of 290 segmentation–feature combinations.

[1] http://www.pascal-network.org/challenges/VOC/

The features extracted from images and image segments can be interpreted as feature vectors in multi-dimensional feature spaces. Each of the resulting feature spaces is quantised with a tree-structured variant of the Self-Organising Map (SOM) [6], a TS-SOM [7]. SOM is an unsupervised neural algorithm that adaptively forms a mapping from a high-dimensional input space onto a two dimensional grid. The database level adaptation to context in the PicSOM framework relies just on the adaptiveness of the SOM: the quantisation of the feature spaces concentrates attention on feature values that are actually common in the database.

The quantisation forms representations of the feature spaces where points on the two-dimensional TS-SOM surfaces correspond to images and image segments. For the current experiments, 64x64 TS-SOMs are used. The organisation of the image database produced by one of the feature TS-SOMs is shown in Figures 1 and 2.

Due to the topology preserving property of the TS-SOM mapping, assessment of image similarity in terms of each of the individual feature spaces can be performed by evaluating the distance of the representation of an object on the TS-SOM grid to the representations of positive and negative example objects. Technically this is done by placing triangular convolution kernels to the locations of example objects and evaluating their value at the locations of the target objects. The kernels placed at positive examples are normalised to sum to unity, as are the kernels placed at negative examples.

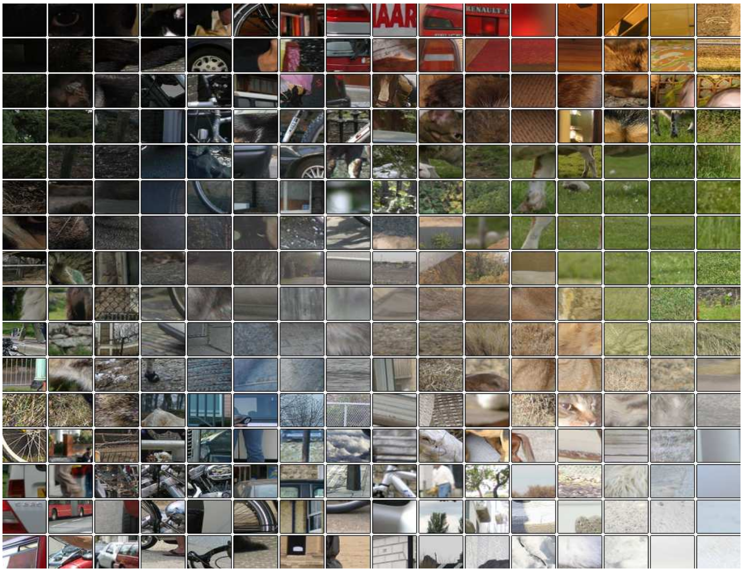

Fig. 1. A TS-SOM organised by colour moments of image patches. For sake of clarity, a 16x16 intermediate level of the TS-SOM structure is shown. For actual image analysis, only the 64x64 bottom level is used.

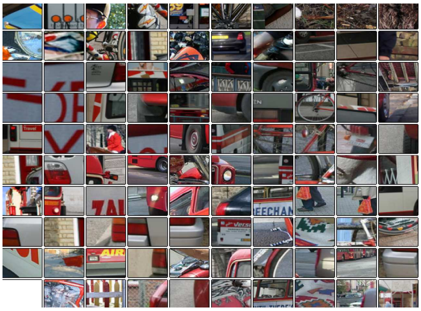

Fig. 2. A close-up of the bottom left corner of the colour moment TS-SOM that represents image patches with red colour together with light-coloured structures

A task-level adaptation mechanism of the PicSOM framework is provided by the way the similarities in different feature spaces are combined. Due to the performed normalisations, such feature spaces are effectively emphasised that perform best in discriminating the positive and negative example objects when the similarities are summed together. This is because on poorly performing feature TS-SOMs, the positive and negative kernels intermingle on the same map areas and effectively neutralise each other. On the other hand, on well discriminating TS-SOMs, the positive and negative kernels concentrate on separate areas where they amplify each other. This results in large amplitudes of similarity and dissimilarity peaks. Because the convolution kernels have a limited and fixed support, the similarity assessment procedure only takes into account local distances. This is also desired as the TS-SOM mapping only preserves local topology.

Often the goal of the system is to rank images, not just image segments according to their similarity. Segment-level similarity must thus be combined into image-level similarity. The solution used in PicSOM is straightforward: image-level similarity is obtained by summing the contributions of the segments of the image. This technique is further discussed, evaluated and extended in Sections 4 and 5.

3.2 Visual Features

A number of statistical visual features is extracted and made available for the similarity assessment algorithm. The features include MPEG-7 standard descriptors [4] as well as some non-standard descriptors. The features are extracted from image segments as well as from whole images when appropriate. Table 2 lists the used visual features.

The MPEG-7 features are extracted using the MPEG-7 Experimentation Model. For some image segments, we have replaced the MPEG-7 Color Layout and Scalable Color descriptors with our own approximate implementation of the standard for computational reasons.

In addition to the tabulated features, we have found that composite features formed from pairs of visual features are highly beneficial in our system. Currently

Table 2. Visual features extracted from image segments

MPEG-7 descriptors	non-standard descriptors
Color Layout	average colour in CIE L*a*b* colour space
Dominant Color	central moments of colour distribution
Region Shape	Fourier descriptors of segment contours
Scalable Color	histogram of Sobel edge directions
	co-occurence matrix of Sobel edge directions
	Fourier transform of local edges
	8-neighbourhood binarised intensity texture
	Zernike moments of binarised segment shape

the composite features are formed by simply concatenating the corresponding feature vectors and equalising the variance of the vector elements. We have confined us to pairs of features as the number of larger feature combinations grows very rapidly. For the same reason, we have not formed all the possible pairs but picked some most promising combinations.

4 Image Segmentation

Currently there appears to be no definitive knowledge concerning what sorts of image segmentation is beneficial or adequate for content-based image analysis. Generic, complete and to-the-pixel accurate unsupervised segmentation is generally regarded to be virtually impossible. Even if it was possible, one still has to address the issue of diversity of segmentations: one image can typically be segmented in numerous different perfectly valid ways. Choosing one segmentation over another would require interpreting the image already in an early stage of the image processing chain. A subset of the diverse segmentations of an image is formed by hierarchically related segmentations: the objects in the images often naturally decompose into sub-objects, thus defining object-part hierarchy trees.

4.1 Segmentation in PicSOM Framework

The approach PicSOM uses for image analysis is statistical in nature. In the same spirit, PicSOM also uses image segmentation in statistical manner. A number of uncertain segmentations is generated by simple means. Every single segmentation may not be correct or visually viable, but on the statistical level we hope to observe correlations between the visual properties of the segments and the sets of example images.

Besides perfect segmentations being impossible to obtain in practice, it still remains to be demonstrated how much the performance of image content analysis is compromised if less complete segmentations are used. Of course, several factors affect this question. One of them is the sophistication level of the image analysis approach. It is reasonable to suppose that if segmentations are used in a statistical manner, fairly inaccurate image segmentation may be enough.

Also the different visual features set different needs for image segmentation. For example, colour and texture features are quite tolerant to inaccuracies in segmentation, whereas some shape features may critically depend on the quality of segmentations.

It seems natural to think that more deterministic image content analysis could lead to better results. Then there would also be need for more accurate image segmentations. However, the more accurate associations would likely be specific to specific classes of example images, such as cars. On the other hand, we want to keep our framework generic, independent of specific types of images. The framework thus has to autonomously learn the associations between image classes and their visual properties. At the current state of machine learning, learning of much more sophisticated rules seems difficult. Certainly we aim at pushing the border into that direction, but currently much more sophisticated associations would probably have to be manually specified to the system.

Ideally, the division between image segmentation and feature extraction system stages would be clear-cut. In practice, however, this is not always the case. Many visual features codify also spatial information: example of this is the MPEG-7 standard Edge Histogram feature that implicitly divides the image into 16 fixed tiles.

4.2 Implemented Segmentation Methods

The actual segmentation methods that are used to generate the alternative segmentations of the images are rather rudimentary. One of the methods geometrically divides the images to overlapping squares. Side length of the square is determined by dividing the larger of the image dimensions by ten. The overlap is achieved by first non-overlappingly tiling the image and then placing another set of tiles at the crossings of the tiling. This results in no more than 181 (for square images) square segments, on the average about 140.

As another segmentation method we employ an area-based region merging algorithm based on homogeneity in terms of colour and texture. The merging is continued until a fixed number of image segments are left. Two sets of segmentations is obtained with two different sets of region merging parameters. The first parameter values give rise to 8 segments whereas the lattar result in 25 segments. The parameters for the merging algorithm have been selected to give visually feasible results for photographs and other images in earlier applications. The diversity of segmentations is increased by recording, in addition to basic segments, the hierarchical segmentation that results from continuing the region-merging algorithm until only three regions remain. Altogether we thus have five segmentations of each database image.

4.3 Combining Segment Similarity as Image Similarity

As briefly mentioned in Section 3, in PicSOM framework the segment-level similarities are just summed together to form the image level similarities. Re-ordering summations, the similarity comparison procedure in a single feature space i can be recast as

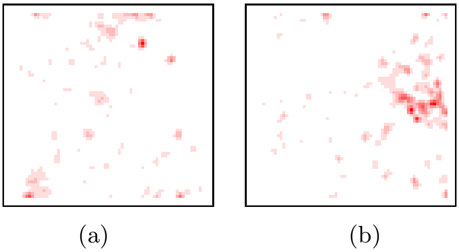

Fig. 3. Smoothed histograms of positive example segments in the quantised colour moment feature space: (a) image class "bus", (b) image class "cow". Dark shades correspond to large values.

1. Accumulating the normalised two-dimensional histograms E_i^+, E_i^- and T of positive example segments, negative example segments and target segments, respectively, using the i:th TS-SOM to define the histogram bins.
2. Kernel smoothing the subtracted histogram of example segments: $H_i = K * (E_i^+ - E_i^-)$, where K is a two-dimensional triangular kernel.
3. Interpreting the histograms as vectors and calculating the similarity as inner product: $S_i =< T_i, H_i >$.

Figure 3 shows two smoothed histograms of positive example segments in the colour moment feature space quantised with the TS-SOM of Figure 1. Among the background noise and spurious responses, we can identify several plausible concentrations from the histograms. In the left subfigure, the distribution of image class "bus" concentrates especially in the lower left corner and upper border of the SOM surface. These areas correspond to image patches with red structures. This is explained by the fact that majority of the buses are British red buses. The segments of image class "cow" concentrate on the rightmost part of the SOM that represents green image patches and image patches with structures on a green background.

Histograms summarise the probability distributions strongly: they introduce large number of invariances. Additional sources of invariances, i.e. information loss, in the present use of image segments are the complete ignorance on the spatial distribution of feature values and failure to connect the same segment in different feature spaces. Example of former would be to consider blue patches of sky in the upper part of an image to be equivalent to the blue patch of water in the lower part. Example of the latter would be to consider images with blue ball and red triangle equivalent with an image of a blue triangle with a red ball.

In some approaches [14,2,5] example and target image segments are explicitly matched with each other. With the numerous alternative segmentations and feature spaces PicSOM uses, the combinatorics of such approach soon become prohibitive. The introduced invariances, however, make the learning problem much less complex and thus more manageable. Learning with less invariances requires more learning material, i.e. example images. The resulting computational costs could easily rise to currently unrealisable level.

5 Off-Line Procedure for Using Image Segments

In the previous section, we described a method of using image segments that has been found to be useful in many image tasks. The use of image segments introduces an increased computational cost, however. For some on-line purposes such as interactive image retrieval, large time complexity may not be acceptable. Therefore, we have devised a method for precomputing a representation of the segment contents off-line as part of adaptation to the database level context and just comparing the representations on-line. For an image and a given feature space, the representations are obtained as follows:

1. A two-dimensional histogram of the feature values of the image segments are accumulated.
2. The histogram is kernel smoothed with a triangular kernel K.
3. The histogram with is downsampled with a factor of four.
4. The histograms are regarded as vectors and their dimensionality is reduced to a fixed number with principal component analysis (PCA).
5. normalise the vector components to unit variance.

For combining multiple feature spaces, we concatenate the histograms as the vectors after downsampling in step 2 and perform PCA and normalising just as for single feature spaces. Compared to only PCA, additional normalising of the components was observed to enhance image classification performance. Further processing by estimating the linear ICA model [3] from the vectors did not bring improvements in connection with the classification procedure used in the experiments.

There are two main parameters of the method: the width w of the smoothing kernel K and the number of components l_{PCA} left after the PCA. A series of experiments were performed to determine optimal values for the parameters. Unfortunately, the best values seem to be feature and image class specific. In determining a suitable compromise we decided to put more emphasis on features and image classes where the descriptor performed well in the classification task. The rationale for this is that probably the poorly performing features will not be used, after all, for that specific image class as there is a wide variety of other features available. In our experiments, we have ended up in parameter values $w = 24$ for the convolution kernel width and $l_{\mathrm{PCA}} = 64$ for the dimensionality of the PCA.

For the initial verification of the usefulness of the obtained region representations, the image classification performance resulting from their use was evaluated using a support vector machine implementation as a black-box classifier. As SVM, we use the C-SVM implementation with RBF kernels implemented in the libsvm software package [1]. With this implementation, the probability estimates needed for ROC curve formation are obtained by the pairwise coupling method described in [15].

5.1 Qualitative Validation

We have compared the image classification performance of the introduced method for obtaining the obtained off-line image representations against the traditional PicSOM on-line algorithm. For this initial experiments, we use a segmentation–feature combinations that were conveniently available. As image segmentation method we employ the square-tile-segmentation described in Section 4.2. We consider the image classification performance of the image representation derived from both 1) a single feature: color layout, and 2) a combination of three features: color layout, Fourier transform of edge distribution and color histogram/texture descriptor, the last feature being a concatenation of two features. It turns out that the choice of the specific features is somewhat unfortunate, as their performance in classifying these image classes is not very good and combining the features does not help the on-line PicSOM algorithm. However, the relative differences between the on-line and off-line variants can still be observed. Figures 4 and 5 visualise the feature space corresponding to the combined off-line representation by means of a SOM. We see that in general, some of the images cluster nicely on the SOM, for example the cluster of cars near the bottom right corner, whereas there are also some areas where images of various objects mix.

5.2 Quantitative Validation

Figure 6 compares the image classification performances of the described off-line and on-line methods for combining image segments. With the single feature, the relative order of the methods varies, the off-line variant perhaps being slightly

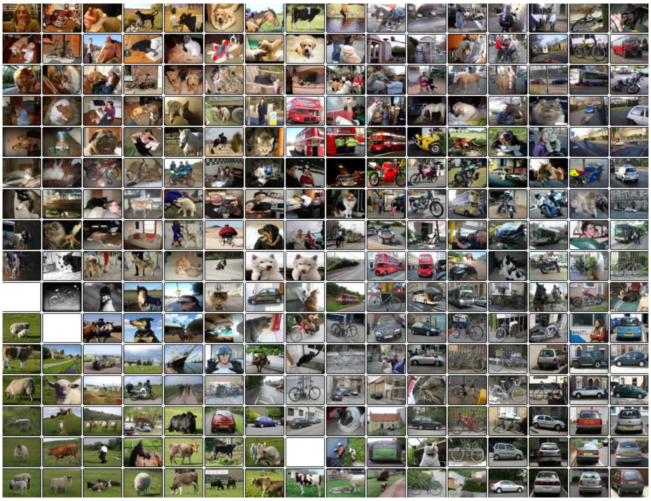

Fig. 4. An intermediate 16x16 level of a TS-SOM visualisation of the feature space defined by the offline image representation derived from the combination of three image segment features

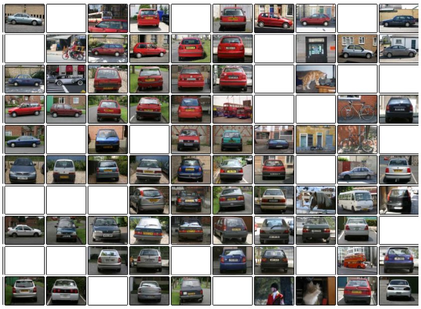

Fig. 5. A close-up of area near the bottom right corner corner of the TS-SOM visualisation of Figure 4

better overall. With the combined feature, the off-line method is consistently better than the on-line method with the exception of the accuracy being approximately equal for one image class. It is notable that the off-line method always manages to benefit from the incorporation of additional features, which is not always the case with the on-line method.

Besides the classification accuracy, an important factor in choosing between the algorithmic variant is speed. After precalculating the off-line representations, the actual comparison of the introduced region representations is observed to be more than 1300 times faster than executing the on-line algorithm. On the other hand, the on-line algorithm is more flexible. It allows the set of positive example segments to be specified freely on run-time, whereas the offline representation summarises all the segments in an image. The flexibility might be needed, for example if the framework is used for image analysis on the level of image segments. This is the case in [13] where the framework is discriminatively used to locate an object in the images an basis of examples, in [12] where image-level keywords are focused on specific image segments, and in the next section of this paper.

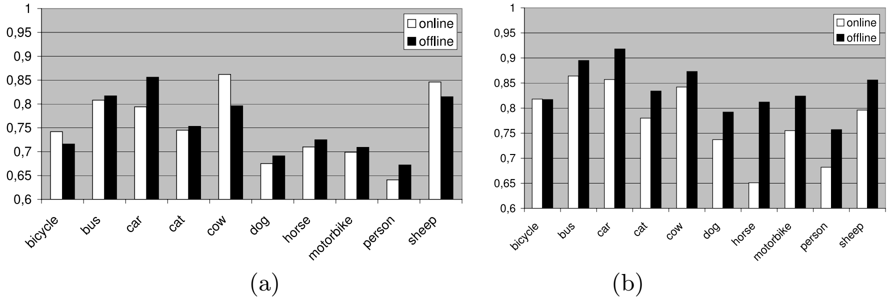

Fig. 6. AUC performance of image classification of both on-line (white bars) and off-line (black bars) methods for combining image segments for the ten VOC object classes. The two subfigures shows the accuracy resulting from use of (a) a single colour layout feature, and (b) the combination of three features.

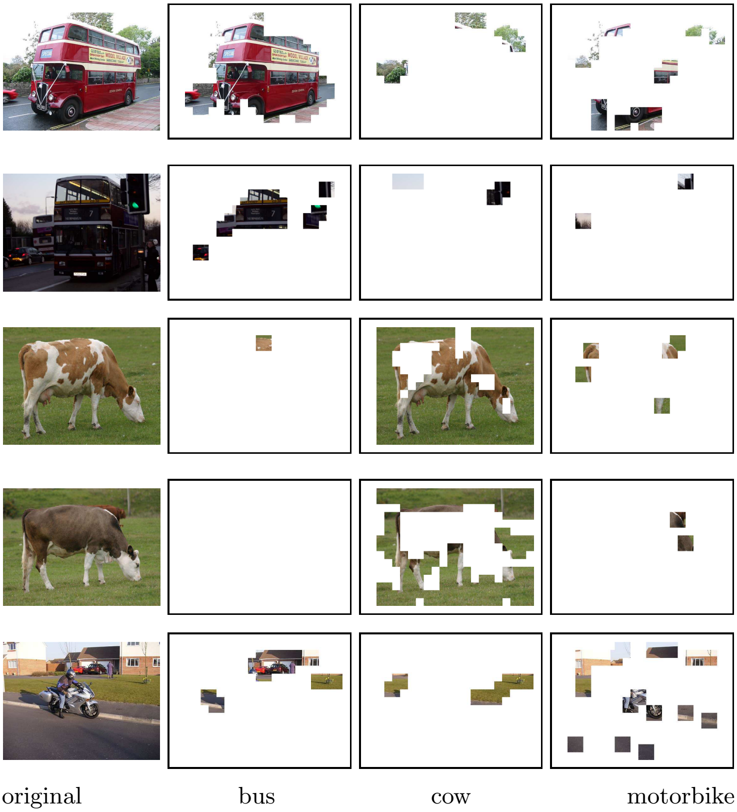

Fig. 7. Images of the image collection shown together with the collection-wide top segments that contribute most to the classification of the images as “bus”, “cow” and “motorbike”. The original images are shown in the leftmost column.

6 Context-Based Identification of Regions of Interest

In this section we qualitatively perform context-based identification of regions of interest, made possible by the described framework for using image segmentation. By methods of Section 4.3, the considered collection of images can be classified into ten classes by the statistics of the visual properties of the image segments. We now go backwards in the inference chain and ask which of the image segments contribute to the classification. For example: which parts of an image containing a cow are essential for classification of the image as such?

To perform this qualitative experiment, we restrict ourselves to a subset of the considered image database that contains approximately one tenth of the images in the full database. The class composition remains approximately equal. For the experiments we use the geometrical square tile segmentation described in Section 4.2.

Figure 7 has been generated by marking to the images the segments that contribute most to the classification of images. The threshold has been set globally to include one tenth of segments of the image database. In the figure each row corresponds to a separate image. The original image is shown in the leftmost column and the remaining columns display such segments of the image that contribute to the image's classification as "bus", "cow" and "motorbike", respectively. In Figure 8 we show a more comprehensive set of similar illustrations of image segments that contribute to classification as "bus".

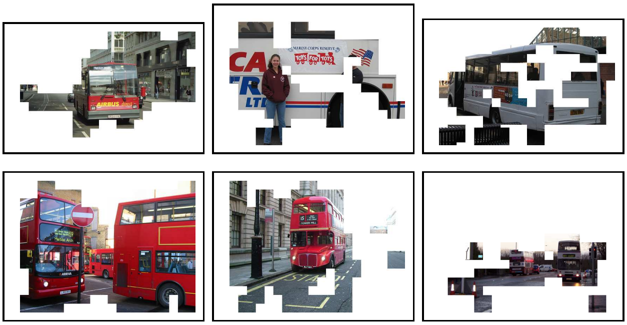

Fig. 8. The locations of image-collection-wide top bus segments in some images actually depicting buses

In an ideal case, in the figures all the image segments containing the class defining object would be highlighted, whereas all the segments outside these objects would remain unhighlighted. Then we would have directly obtained a context-based method for segmenting images contained in an image database. The method would be adaptive to the image database in two senses. Firstly, the context of the database would determine the visual elements that are discriminative in that certain database. Secondly, the segmentation method would be class-specific, producing segmentations that are discriminative in the sense of the specific object class the segmentation method was derived from.

From the images we see, however, that the situation is far from the above described ideal case. Still, from Figure 7 we see that there is strong tendency of the algorithm to identify apparently reasonable parts of the images as responsible for the classification. The algorithm is also class specific: for instance only few segments of the bus images would be likely to appear in the cow images and vice versa. This property is not deterministic, however, but probabilistic in nature.

In many cases, the context of the objects is found to be equally important as the visual properties of the object themselves. This can be seen for example in Figure 8 where the surrounding traffic or road scene often is included in the most contributing parts of the bus images. Another observation is that inside an object only some of its parts may be specific to that object class. This is exemplified, for example, by the motorbike image in Figure 7. The cow images also exemplify well these two phenomena. Both can also be explained by the limited scope of the underlying image database. The context of this type of green grass seems to be specific for the cow images in this database, although the database contains other animals such as sheep and horses. Apparently the grass in those images is somehow different. On the other hand, the brown fur part of the lower cow image does not contribute much to classification as cow. This can be explained by the database containing also other animals such as horses and dogs. In contrast, the contours of lower part of the cow, especially legs seem to be characteristic for cows.

In this section we have seen that using the above described image representation in terms of regions, we can back-track in the image classification algorithm and identify plausible regions of interest. However, the connection between the identified regions and genuine object segmentation is not straightforward. Still, integrated with other image segmentation, the regions of interest might provide a helpful cue in segmentation.

7 Discussion and Conclusions

In this paper, we have discussed the method of using image segments in the PicSOM statistical image analysis framework. A procedure was devised for moving the computational demands of this segment-using method from on-line computation to the off-line phase, thus reducing the on-line complexity to a small fraction. Still, the image classification accuracy remained uncompromised in the experiments performed. The obtained time savings are important for keeping the computational demands of the task-level context adaptation manageable. The performed experiments were initial and must be complemented with further testing with a more comprehensive set of visual features and larger image databases.

We saw that the current method of using segments introduces many invariances. Although lifting the invariances could easily lead to need for huge amounts of example data, we are still planning to consider lifting some of the invariances, for example by considering spatial relationships of the segments [9]. The beneficiality of geometric constraints to the accuracy of image analysis has been demonstrated e.g. in [10].

We can interpret the method of generating the off-line representations of the segment distributions as a novel method of generating image features adapted to the current image analysis context, given the segment level visual features and the image segmentations. The method adapts to the database level context through the adaptive feature representations formed using SOMs. Also the segmentations may be adaptive, either on the database level or on the task level, as outlined—although not implemented—in Section 6. Comparing the merits of these adaptive features to customary, non-adaptive image features such as the MPEG-7 features remains to be performed in the future.

We qualitatively investigated the contribution of different types of segments to image classification in Section 6. It was observed that in the context of the current VOC image database, contextual information was often as important for the classification as the target object itself. This result is naturally dependent on the database at hand. If the database is sufficiently comprehensive, the significance of context in the images may become lesser, and the actual target object in the image becomes the most important contributor in image classification, as in [12].

References

1. 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`.
2. Yixin Chen and James Z. Wang. Looking beyond region boundaries: Region-based image retrieval using fuzzy feature matching. In *Multimedia Content-Based Indexing and Retrieval Workshop, September 24-25*, INRIA Rocquencourt, France, September 2001.
3. Aapo Hyvärinen, Juha Karhunen, and Erkki Oja. *Independent Component Analysis.* John Wiley & Sons, 2001.
4. ISO/IEC. Information technology - Multimedia content description interface - Part 3: Visual, 2002. 15938-3:2002(E).
5. F. Jing, M. Li, L. Zhang, H. Zhang, and B. Zhang. Learning in region-based image retrieval. In *Proceedings of International Conference on Image and Video Retrieval*, volume 2728 of *Lecture Notes in Computer Science*, pages 198–207. Springer, 2003.
6. Teuvo Kohonen. *Self-Organizing Maps*, volume 30 of *Springer Series in Information Sciences.* Springer-Verlag, Berlin, third edition, 2001.
7. Pasi Koikkalainen. Progress with the tree-structured self-organizing map. In *11th European Conference on Artificial Intelligence.* European Committee for Artificial Intelligence (ECCAI), August 1994.
8. Jorma Laaksonen, Markus Koskela, and Erkki Oja. PicSOM—Self-organizing image retrieval with MPEG-7 content descriptions. *IEEE Transactions on Neural Networks*, 13(4):841–853, July 2002.
9. Jorma Laaksonen and Ville Viitaniemi. Emergence of ontological relations from visual data with self-organizing maps. In *Proceedings of the 9th Scandinavian Conference on Artificial Intelligence Scandinavian*, Espoo, Finland, October 2006. To appear.
10. C. Millet, I. Bloch, P. Hède, and P.-A. Moëllic. Using relative spatial relationships to improve individual region recognition. In *Proceedings of 2nd European Workshop on the Integration of Knowledge, Semantic and Digital Media Technologies*, pages 119–126, London, UK, November 2005.
11. Mats Sjöberg, Jorma Laaksonen, and Ville Viitaniemi. Using image segments in PicSOM CBIR system. In *Proceedings of 13th Scandinavian Conference on Image Analysis (SCIA 2003)*, pages 1106–1113, Halmstad, Sweden, June/July 2003.
12. Ville Viitaniemi and Jorma Laaksonen. *Focusing Keywords to Automatically Extracted Image Segments Using Self-Organising Maps*, volume 210 of *Studies in Fuzziness and Soft Computing.* Springer Verlag, 2006. To appear.
13. Ville Viitaniemi and Jorma Laaksonen. Techniques for still image scene classification and object detection. In *Proceedings of 16th International Conference on Artificial Neural Networks (ICANN 2006)*, September 2006. To appear.
14. James Z. Wang, Jia Liu, and Gio Wiederhold. SIMPLIcity: Semantics-sensitive integrated matching for picture libraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(9):947–963, September 2001.
15. T.-F. Wu, C.-J. Lin, and R.C.Weng. Probability estimates for multi-class classification by pairwise coupling. *Journal of Machine Learning Research*, 5:975–1005, 2005.