Region based visual object categorization using segment features and polynomial image modeling

Huanzhang Fu, Alain Pujol, Emmanuel Dellandrea, and Liming Chen

LIRIS, UMR 5205 CNRS
Ecole Centrale de Lyon - 36 Av. Guy de Collongue - 69134 Ecully Cedex - France
{huanzhang.fu,alain.pujol,
emmanuel.dellandrea,liming.chen}@ec-lyon.fr

Abstract. This paper presents a novel approach for visual object classification. Based on Gestalt theory, we propose to extract features from coarse regions carrying visually significant information such as line segments and/or color and to include neighborhood information in them. We also introduce a new classification method based on the polynomial modeling of feature distribution which avoids the drawbacks of a popular approach, namely "bag of keypoints". Moreover we show that by separating features extracted from different sources in different "channel", and then to combine them using a late fusion, we can limit the impact of feature dimensionality and actually improve classification performance. Using this classifier, experiments reveal that our features lead to better results than the popular SIFT descriptors, but also that they can be combined with SIFT features to reinforce performance, suggesting that our features managed to extract information which is complementary to the one of SIFT features.

1 INTRODUCTION

Generic Visual object classification is one of the most challenging problems in computer vision. Indeed the number of real world object types which need to be discriminated, as well as variations in view, imaging, lighting and occlusion which also pose a serious problem. To this we must add the difficulty induced by intra-class variations, typical of semantic classes of everyday objects. As such it has attracted a lot of attention in the past years [1].

1.1 RELATED WORK

Most works in the literature make use of a "bag of features" kind of approach [2, 3] which tries to adapt the "bag-of-words" representation for text categorization to "Visual Object Categorization" *VOC* problem and has shown its effectiveness, obtaining the best performance in Pascal VOC contest [1]. These methods view images as an orderless distribution of local image features, typically using the popular SIFT feature [4], extracted from salient image regions, called interest "points" [4–6] or more simply from points extracted using a grid [7]. The set of these local features is then characterized by a histogram of "visual keywords" from a visual vocabulary which is learned from the training set by a hard assignment (quantization) or a soft assignment through GMMs. These distributions can thus be compared to estimate the similarities between images and categorized through a machine learning process, for instance SVM.

Although the "bag-of-local features" approach has achieved the best performance in the last Pascal VOC contests, the overall performance, with an average precision around 60% over 20 classes achieved by the best classifier, is still far from real application-oriented requirements. In particular, the size of visual vocabulary which is the basis of "bag-of-local features" approach is hard to be fixed as there are no evident similar concepts in images as compared to a textual document. The basic problem is that the "bag-of-local features" approach, while adapting the best practice from text categorization, does not necessarily correspond to a human visual perception process which is ruled by some Gestalt principles according to several studies on visual perception [8, 9] and supposed to perform a holistic analysis combined with a local one through a fusion process. Moreover, the schemes so far proposed in the literature for automatic generic visual object classification also suffer from well-known machine learning problems, namely the curse of dimensionality when increasing feature vector size which leads to exponential learning complexity as well as a small and biased training dataset, in particular one with an imbalanced ratio of positives versus negative samples.

1.2 OUR APPROACH

Our basic hypothesis is that effective visual object classification or detection should be inspired by some basic human image interpretation principles. In this paper we propose overcoming the shortfalls of the popular "bag-of-local features" approach and make use of some basic principles from the Gestalt theory, in particular the well known Gestalt laws of Perceptual Organization which suggest both the grouping of pixels into homogeneous regions as well as the interaction between regions.

Desolneux et al. have given in [10] a comprehensive introduction to Gestalt theory in an image analysis perspective. Gestalt theory starts with the assumption of active grouping laws in visual perception which recursively cluster basic primitives into a new, larger visual object, a gestalt. These grouping laws follow criterion such as spatial proximity, color similarity. These laws also highlight the interaction between regions. This interaction is also mentioned by Navon [11] who showed the preponderance of global perception over local perception. This is also an important claim that motivated our approach. On the other hand, as we mentioned previously, the currently most successful approaches are based on the "bag of keywords" framework [2, 3]. We will mention the work of Barnard et al. [12] which is a region-based approach where regions are labeled with probable categories.

We feel that lacking these principles, "bag of features" approaches deprive themselves of meaningful information. Thus, instead of SIFT like local features, we propose some region-based meaningful features extracted from image regions with neighborhood information. These region based features result from perceptually significant "Gestalts" segmented according to some basic Gestalt grouping laws. Because there are some cases where region segmentation cannot be consistent with object boundaries, we will not try to label individual regions. Regions produce a feature vector which is supposed to have no meaning on its own but that can contribute to one or more classes. Regarding features, we propose using visually meaningful features, such as color and line segment based features [13] which we will extend to provide information from neighboring regions. We intend to compare the use of these region based features to popular SIFT features but also to check the efficiency of the combination of our features with SIFT features to evaluate their complementarity.

We also propose a polynomial representation to model image feature distribution. Its interest is three-fold. We can circumvent the difficulty of fixing the size of visual vocabulary, we avoid the inaccurate assumption of Gaussian repartition of features and we can cope with a smaller number of feature vectors per image (which is the case with our features).

Finally, we also study and compare two fusion strategies, namely early fusion strategy by grouping all the features together and fed into a single classifier, and late fusion strategy which makes use of "channels" with a separate classifier for each kind of features, the outputs of these classifiers being merged later [14] in a process similar to boosting [15]. Experiments carried out on a subset of Pascal VOC dataset show that our features not only perform adequately (providing better classification results than the popular SIFT features), but also can be combined with SIFT features to provide better overall performance. Moreover, the separation of the different channels by a late fusion strategy performs better than early fusion strategy. Late fusion induces lower-dimension input feature vectors for each classifier.

The rest of the paper is organized as follows. In section 2 we will describe our region segmentation algorithm and the features we extract, section 3 will introduce our classification method while section 4 contains experimental results and discussion. We will have some concluding remarks and evoke future research prospects in section 5.

2 IMAGE FEATURES AND FEATURE STRUCTURE

2.1 REGION BASED FEATURE EXTRACTION

As we have seen in the previous section, studies on human perception strongly hint at a region based approach. On the other hand, introducing region segmentation brings about a host of new problems regarding segmentation robustness and accuracy so while this approach suits human perception better, we have no guarantees its benefits will overcome its drawbacks. In our approach we specifically designed a robust region segmentation method that aims at automatically producing coarse regions from which we can consistently extract feature vectors [13]. We will now briefly describe the outline of the algorithm.

The principle of our region segmentation algorithm is to segment an image into partial gestalts for further visual object recognition. We thus made use of the following Gestalt basic grouping laws in our gestalt construction process: The color constancy law stating that connected regions where color does not vary strongly are unified; the similarity law leading to group similar objects into higher scale object; the vicinity law suggesting grouping close primitives with respect to the others; and finally good continuation law saying that reconstructed amodal object, i.e partially perceived physical structure which is reconstructed through understanding, should be as homogenous as possible. Because those laws are defined between regions and their context, at each step we assess the possibility to merge according to global information.

The algorithm is based on color clustering but also includes an extra post-processing step to ensure spatial consistency of the regions. In order to apply previously mentioned Gestalt laws, we defined a 3-step process: first we filter the image and reduce color depth, then we perform adaptive determination of the number of clusters and cluster color data and finally we perform spatial processing to split unconnected clusters and merge smaller regions.

Images are first filtered for robustness to noise, colors are then quantified by following a first, fast color reduction scheme using an accumulator array in CIELab color space to agglomerate colors that are perceptually similar. In the second step we use an iterative algorithm to determine a good color count which limits the quantization error. Indeed, quantization error measured by MSE between original and quantized color evolves as per figure 1 according to the number of clusters.

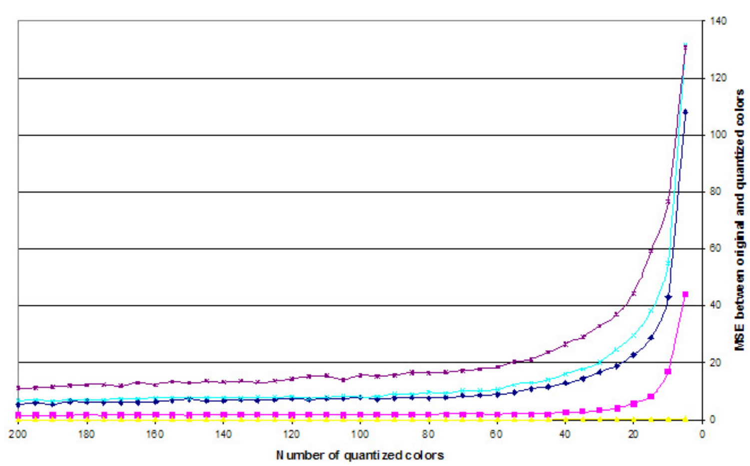

Fig. 1. Evolution of MSE between quantized colors and original colors

This clearly shows a threshold cluster number under which quantization MSE begins to rise sharply. By performing several fast coarse clustering operations using Neural Gas algorithm [16], which is fast and less sensitive to initialization than its counterparts such as K-means, we are able to compute the corresponding MSE values and generate a target cluster count. We then use hierarchical ascendant clustering which is more accurate but much slower thus executed only once in our case, to achieve segmentation. The third step consists in splitting spatially unconnected regions, merging similar regions and constraining segmentation coarseness. Merging of similar regions is achieved through the use of the squared Fisher's distance (used for a similar task in [17]).

$$D(R_1, R_2) = \frac{(n_1 + n_2)(\mu_1 - \mu_2)^2}{n_1 \sigma_1^2 n_2 \sigma_2^2} \tag{1}$$

Where n_i, μ_i, σ_i^2 are respectively the number of pixels, the average color and the variance of colors within region i. This distance still stays independent towards image dynamics as it involves intra-cluster distance vs. inter-cluster distances. Finally, regions which are too small to provide significant features are discarded.

With this algorithm we obtain consistent coarse regions that can be used for our classification system. Sample segmentation results on Pascal challenge dataset images can be seen on figure 2.

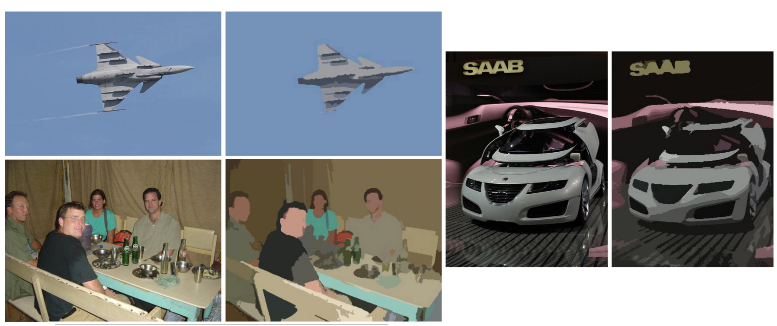

Fig. 2. Sample segmented Images

2.2 OUR FEATURES

With the purpose of testing region based features and validating our segment based features, we will be using only two kinds of features for this work: color features and shape features. We will use region based color features in the form of color moments (mean, variance and skewness) [18] for each color channel. These features are quite compact and have proven as efficient as a high dimension histogram [19]. Various color spaces were experimented for the computation of these features and best results were achieved in the CIELch color space which is derived from CIELab.

As stated earlier, local shape features have proven their efficiency for content based indexing and, more specifically, have performed well in the Pascal challenge. We thus developed segment based features relying on a fast connective Hough transform [20] that performed well in global image classification [21] and more specifically provided more significant information than gradient based features. These features are relevant regarding our approach of following human visual interpretation as, most of the time, there are few segments within a region but, on the other hand, they represent features that stand out visually and their simple presence is significant.

The principle of our segment based feature extractor is the following. As for any other Hough transform we start from an edge map of the processed image. Because we wish to avoid problems related to edge thickness, we use a Canny Edge Detector [22] to process our image in order to ensure a one pixel thickness for our edge map. For an edge point on the edge map, we examine its neighborhood identified by its relative angular position (r, θ): each direction θ is processed while a connected edge is found at distance $r + 1$, which gives us a list of segments by orientation for this edge point. Once we have this list, we store the longest segment and remove it from the edge map. To avoid hindering intersecting segment detection, we use two separate edge maps: one for segment source point detection and one for connected points detection. Removed segments are only removed from the source point map, which avoids detecting the same segment twice while preserving intersecting segments. These segment features are extracted once for the whole image.

During this extraction step, we can build a map from image coordinates to the corresponding segments, therefore we can quickly detect segments within a region. For validation purposes, our "segment" shape features are a simple histogram combining length and orientation. In order to obtain scale invariant features, we normalize lengths by dividing them by the longest segment's length. We then obtain rotation invariance by computing an average orientation in order to have a stable average and by expressing all angles with respect to this average direction. We therefore obtain a feature that is invariant to translation, scale as well as rotation. The size of the histograms was experimentally determined and set to 6 bins for orientation and 4 for length.

Finally in order to include neighborhood information, our region based features (color moments and Hough segment features), are expressed at four different levels: original region, region + neighbors, region + neighbors + neighbor's neighbors, etc. Those levels are concatenated in the final feature vector. This is a basic way to integrate spatial relationship but also to include global information in each feature vector. On most images the fourth level will represent features extracted over the whole image. Characteristic values are also informative: segment features at first level (region) may sometimes have a 0 value which is a telling region characteristic. Image with few regions will also cover the whole image quickly, leading to constant values for higher levels. This indicates the overall complexity of the image and, we manage to capture this characteristic by our polynomial features which express the variation of each feature vector's dimension within an image as we will see in the next section.

3 CLASSIFICATION

3.1 POLYNOMIAL MODELING BASED IMAGE REPRESENTATION

In this section, we propose a novel method of image representation and classification using the features mentioned in the previous section. Each image produces a set of feature vectors; current approaches include building a "visual vocabulary" by using a clustering algorithm or the use of Gaussian Mixture Models (GMM). A drawback of these approaches is that it is really difficult to estimate the optimal vocabulary size. Regarding GMM, if the number of Gaussians is too small then it can't supply enough normal distributions for a large amount of diversified feature vectors to be modeled, while a too high number of Gaussians suffers from an insufficient number of feature vectors to optimize the parameters of the model. Moreover, our region-based approach generates generally few feature vectors and in very different numbers from one image to another. This makes modeling an image by GMM very difficult. Therefore we propose our approach of polynomial modeling based image representation to avoid this drawback.

The basic idea of this image representation is to model the histogram characterizing the distribution of each feature of the image feature set with a polynomial The coefficients of these polynomials will then be considered to form a new feature vector which will be the image representation used for object categorization in the next step.

The polynomial model for a given feature histogram is computed as follow. Given the set D of histogram values $D = \{(x_1, y_1), ..., (x_M, y_M)\}$ (M being the number of values), a polynomial $f(x)$ of degree N, described by its set of coefficients $P = \{p_1, p_2, ...p_{n+1}\}$, is computed to interpolate the data, by fitting $f(x_i)$ to y_i in a least squares sense. Thus, the vector P can characterize the distribution of D. An example is given in figure 3. Once the distribution of each feature from the feature set has been modeled thanks to a polynomial of degree N, a new image feature vector Q is produced by concatenating the coefficients of all polynomials.

Assuming that we have L features in a feature set and use polynomials of degree N, then the vector Q has a dimension of (N+1)*L, which ranges generally from hundreds to thousands. A vector of such high dimensionality used for classification generally leads to the "curse of dimensionality" [23]. Consequently, the dimension of these vectors has to be reduced. Numerous feature selection methods can be envisaged [24]. We have chosen the canonical discriminant analysis [25] as it is a quick algorithm which allows to reduce the dimension by producing a new representation space which distinguishes the best the different classes. Its

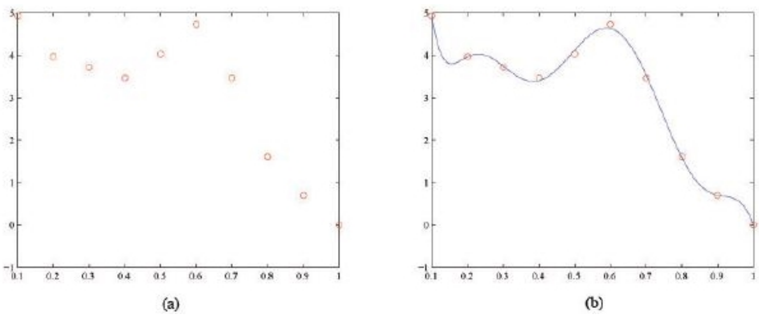

Fig. 3. (a) Histogram for one feature of the image feature set. (b) A polynomial curve models the histogram in (a)

principle is to produce a series of uncorrelated discriminating variables, in order to have individuals in the same class projected on these axes as close as possible and individuals from different classes as distant as possible. In most cases, we obtain K-1 axes where K is the number of classes. Thus with the help of this method, the overall feature vector Q becomes a much more simplified vector called polynomial modeling based image representation.

3.2 CLASSIFICATION PROCESS

Given an image to classify, we first detect interest points or regions from which the features that we need are extracted. These features are then transformed to form a new feature vector through polynomial modeling based image representation using the method introduced in the previous subsection. Finally, this new feature vector will pass through the classifier beforehand trained or pass through a set of such classifiers, which concerns the fusion strategy presented in the next paragraph, to judge whether this image contains the specified object. Any classifier can be used for categorization of this image representation, such as SVM or Neural networks. In our evaluation, we have chosen a simple multilayer perceptron.

Fusion strategy is usually used in multimedia data analysis because generally three modalities exist in video, namely the auditory, the textual, and the visual modality, so that a fusion step is necessary to combine the results of the analysis of these modalities [14]. However, the same idea can be employed in visual object categorization, since we can also extracted different types of features from the same image to form several information streams for fusion, such as SIFT, Region based Color Moments (RCM) and Region based Histogram of Segments (RHS) in our case. This fusion of different types of features can follow several strategies: an early fusion is obtained when a single feature set is composed of all the features; a late fusion is obtained when a single decision is taken from the intermediate decision taken from each type of features. A problem generally encountered when considering early fusion is that the features of the feature set have different natures and consequently the feature vector is not homogeneous. Between these two strategies, numerous intermediate strategies can be conceivable which consist in generating intermediate classes from different sources and to take a final decision based on these intermediate classes. In our experiment, we evaluate the two main strategies: early fusion and late fusion. Their schemes are illustrated respectively in figure 4 (a) and 4 (b).

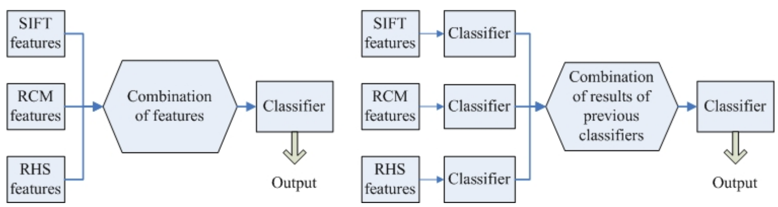

Fig. 4. (a) General scheme for early fusion (b) General scheme for late fusion

4 EXPERIMENTAL RESULTS

We have used in our experiments the database of PASCAL Challenge 2007 [1]. This database consists in 20 object categories and totally 2501 images taken in real world are provided for training and 2510 for validation. As a first experimental evaluation of our classification approach, we have chosen 5 semantic representative classes namely airplane (248 images), bicycle (243 images), bus (186 images), horse (287 images) and person (2008 images). One versus all multilayer perceptrons have been built for each class with a 4 fold-cross-validation. The structure of these perceptrons is composed of one hidden layer for all the tests. However the number of neurons in this layer varies according to the number of inputs, which ranges from 2 to 15. Finally, the degree of polynomial for modeling has been empirically set to 8.

The three types of features previously introduced, namely SIFT (computed using the C# "libsift" implemented by Sebastian Nowozin [26] for their extraction), RCM and RHS, have been considered in our experiments. Moreover, two fusion strategies, early and late, have been evaluated using these feature sets, in order to evaluate their efficiency in our case of visual object categorization. Two region based features are first merged by the strategies of Early Fusion and Late Fusion, noted as EF(RCM+RHS) and LF(RCM+RHS), and SIFT is combined afterwards, noted as EF(RCM+RHS+SIFT) and LF(RCM+RHS+SIFT). We have chosen three rates, i.e. classification rate, recall rate and precision rate to evaluate the performance of our classifier. The detailed results are presented in table 1.

In table 1, there are principally 3 parts: Single Channel (SC), Early Fusion (EF) and Late Fusion (LF). From SC, we can clearly see that RCM and RHS

Table 1. Results of 5 classes for object recognition on PASCAL database 2007

Classification rate	Plane	Bicycle	Bus	Horse	Person
SIFT	65,00%	55,21%	60,75%	65,49%	58,94%
RCM	72,69%	61,57%	67,90%	65,84%	62,77%
RHS	76,60%	61,98%	66,13%	62,59%	63,54%
EF(RCM+RHS)	80,34%	63,97%	70,75%	65,63%	65,17%
EF(RCM+RHS+SIFT)	81,47%	64,63%	69,30%	66,43%	65,50%
LF(RCM+RHS)	82,02%	70,95%	91,99%	79,65%	66,74%
LF(RCM+RHS+SIFT)	85,21%	72,73%	92,74%	81,54%	69,41%
Recall rate	Plane	Bicycle	Bus	Horse	Person
SIFT	68,66%	57,90%	62,58%	71,57%	60,93%
RCM	73,45%	64,10%	68,06%	66,53%	67,27%
RHS	76,55%	68,32%	71,61%	67,09%	69,11%
EF(RCM+RHS)	80,17%	65,67%	70,86%	67,09%	68,42%
EF(RCM+RHS+SIFT)	81,43%	66,67%	70,54%	70,10%	68,57%
LF(RCM+RHS)	84,20%	73,86%	89,35%	79,48%	70,01%
LF(RCM+RHS+SIFT)	85,38%	74,86%	89,78%	83,89%	72,89%
Precision rate	Plane	Bicycle	Bus	Horse	Person
SIFT	63,98%	54,90%	60,37%	63,76%	58,60%
RCM	72,35%	60,98%	67,85%	65,56%	61,72%
RHS	76,62%	60,60%	64,53%	61,49%	62,08%
EF(RCM+RHS)	80,44%	63,47%	70,71%	65,13%	64,24%
EF(RCM+RHS+SIFT)	81,50%	64,02%	68,84%	65,25%	64,61%
LF(RCM+RHS)	80,68%	69,77%	94,32%	79,71%	65,72%
LF(RCM+RHS+SIFT)	85,09%	71,77%	95,43%	80,08%	68,14%

perform generally better than SIFT, 5%-11% augmentation recorded depending on the class, except for the horse class for which they are almost in the same level. This proves the effectiveness of our region based features RCM and RHS using the polynomial modeling based image representation. Between RCM and RHS, we find that RHS is slightly better after comparing all 3 rates and RCM tends to favor negative side. Focusing on EF and LF, we can note that the best classification rates are obtained when the 3 channels are merged using LF strategy, much better than SC and EF. The classes of bus and horse, for instance, see a classification rate increasing by about 22% and 15% respectively compared to the second higher rate obtained in EF. One of the reasons is that the fusion of several channels each of whom brings different information helps to improve the classifier's performance. Another reason might be that EF may have more chances to result in a conflict between different features which blur the boundary between classes. It can also explain that why EF performs only slightly better than SC and much worse than LF.

5 CONCLUSION

We have proposed in this paper a novel approach for visual object categorization, using polynomial modeling based image representation with new region based features. Two different fusion strategies, early and late have been considered in order to merge information from different "channels" represented by the different types of features. Results of this evaluation performed on PASCAL 2007 database have shown that good performance can be achieved with this image representation and that our segment features carry information which is complementary to SIFT features, especially when merging feature channels according to a late fusion strategy. In our future work, we envisage to improve our method, in particular by optimizing the quantification technique used for the construction of feature histograms as well as the choice of the degree of polynomial for modeling. Concerning features, we will evaluate richer representations of our region based features (e.g.: coocurrence matrix, ...) and also test more channels such as region shape and texture.

References

1. Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A.: The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results. http://www.pascal-network.org/challenges/VOC/voc2007/workshop/ (2007)
2. Dance, C., Willamowski, J., Fan, L., Bray, C., Csurka, G.: Visual categorization with bags of keypoints. In: ECCV International Workshop on Statistical Learning in Computer Vision. (2004)
3. Rothganger, F., Lazebnik, S., Schmid, C., Ponce, J.: Object modeling and recognition using local affine-invariant image descriptors and multi-view spatial contraints. International Journal of Computer Vision **66** (2006)
4. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision **60** (2004) 91–110

5. Lindeberg, T.: Feature detection with automatic scale selection. International Journal of Computer Vision **30** (1998) 79–116
6. Mikolajczyk, K., Schmid, C.: Scale & affine invariant interest point detectors. International Journal of Computer Vision **60** (2004) 63–86
7. Li, F.F., Perona, P.: A bayesian hierarchical model for learning natural scene categories. In: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). (2005) 524–531
8. Kaniza, G.: Grammatica del vedere. Il Mulino (1997)
9. Wertheirmer, M.: Untersuchungen zur lehre der gestalt ii. Psychologische Forschung **4** (1923) 301–350
10. Desolneux, A., Moisan, L., Morel, J.: From Gestalt Theory to Image Analysis: A Probabilistic Approach. Springer (2008)
11. Navon, D.: Forest before trees: The precedence of global features in visual perception. Cognitive Psychology **9** (1977) 353–383
12. Barnard, K., Duygulu, P., Guru, R., Gabbur, P., Forsyth, D.A.: The effects of segmentation and feature choice in a translation model of object recognition. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). (2003)
13. Pujol, A., Chen, L.: Coarse adaptive color image segmentation for visual object classification. In: Proceedings of the 15th International Conference on Systems, Signals and Image Processing. (2008)
14. Snoek, C.G.M., Worring, M., Smeulders, A.W.M.: Early versus late fusion in semantic video analysis. In: MULTIMEDIA '05: Proceedings of the 13th annual ACM international conference on Multimedia, ACM (2005) 399–402
15. Freund, Y., Schapire, R.: A short introduction to boosting. Journal of Japanese Society for Artificial Intelligence **14** (1999) 771–780
16. Martinetz, T., Schulten, K.: A "neural-gas" network learns topologies. Artificial Neural Networks **I** (1991) 397–402
17. Zhu, S.C., Yuille, A.: Region competition: Unifying snakes, region growing, and bayes/mdl for multiband image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **18** (1996) 884–900
18. Stricker, M.A., Orengo, M.: Similarity of color images. In: Storage and Retrieval for Image and Video Databases (SPIE). (1995) 381–392
19. Deng, Y., Manjunath, B., Kenney, C., Moore, M., Shin, H.: An efficient color representation for image retrieval. IEEE Transactions on Image Processing **10** (Jan 2001) 140–147
20. Ardabilian, M., Chen, L.: A new line extraction algorithm: Fast connective hough transform. In: proceedings of PRIP'2001. (2001) 127
21. Pujol, A., Chen, L.: Line segment based edge feature using hough transform. In: The 7th IASTED International Conference on Visualization, Imaging, and Image Processing (VIIP). (2007)
22. Canny, J.: A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence **8** (1986) 679–698
23. Bellman, R.: Adaptive Control Processes: A Guided Tour. Princeton University Press (1961)
24. Saeys, Y., Inza, I., Larranaga, P.: A review of feature selection techniques in bioinformatics. Oxford University Press (2007)
25. Fisher, R.A.: The use of multiple measurements in taxonomic problems. Annals of Eugenics **7** (1936) 179–188
26. Nowozin, S.: libsift - scale-invariant feature transform implementation. http://user.cs.tu-berlin.de/ nowozin/libsift/ (2005)