Retrieving Images for Remote Sensing Applications

Neela Sawant, Sharat Chandran, and B. Krishna Mohan

Department of Computer Science and Engineering
Indian Institute of Technology Bombay
http://www.cse.iitb.ac.in/∼{neela, sharat}
http://www.csre.iitb.ac.in/bkmohan

Abstract. A unique way in which content based image retrieval (CBIR) for remote sensing differs widely from traditional CBIR is the widespread occurrences of *weak textures*. The task of representing the weak textures becomes even more challenging especially if image properties like scale, illumination or the viewing geometry are not known.

In this work, we have proposed the use of a new feature *'texton histogram'* to capture the weak-textured nature of remote sensing images. Combined with an automatic classifier, our texton histograms are robust to variations in scale, orientation and illumination conditions as illustrated experimentally. The classification accuracy is further improved using additional image driven features obtained by the application of a feature selection procedure.

1 Introduction

For many years, information extracted from remote sensing image archives has been exploited for specialized applications like monitoring land cover and land usage, identifying cases of floods or fires, urbanization, deforestation, and so on. Building such applications have been relatively easy with existing domain knowledge and readily available information about image properties like scale, orientation, and illumination conditions. This scenario is changing rapidly as technological advances such as Google Earth demand a generic framework to satisfy unpredictable 'casual' user queries with possibly unknown image properties.

Remote sensing images are essentially textured images with lands, grass, forests, mountain ranges, water, clouds, snow, buildings, and the like. Majority of these categories exhibit weak textures and a highly irregular structure. Hence the focus of remote sensing CBIR systems should be on identifying the texture features correctly. In the past, many CBIR systems have tried capturing characteristic textures using features like local texture patterns [1], Gabor multiscale features [2,3,4], Markov random field (MRF) textures [5,6], Gibbs Markov models [7], and wavelet features [8,9]. The SIMPLIcity (Semantics-Sensitive Integrated Matching for Picture Libraries) [10] system uses a combination of texture and color features.

P. Kalra and S. Peleg (Eds.): ICVGIP 2006, LNCS 4338, pp. 849–860, 2006.
© Springer-Verlag Berlin Heidelberg 2006

Based on our experiments, we find that the success of these methods limited owing to the problems in handling unknown imaging conditions and the inability to capture weak textures effectively. Gabor features for example, respond well to strong textures but are not able to capture the weak textures effectively. Multi-scale filter based techniques, like Gabor or wavelet based approaches, extract features at multiple scales and try to find the best match across them. Considering the weak-textured nature of remote sensing images, it is often difficult to get distinguished texture readings across scales. Moreover, features from different texture categories at different scales may falsely appear similar, thus limiting the classification accuracy further. MRF features represent weak textures well, but they are not scale independent.

A classic problem faced by most of the existing systems is the *misleading image appearances*. The color and texture appearances of the same surface vary significantly with the changes in illumination and camera angle properties. Fig. 1 shows an example where water appears green in one image and blue in another. To a human observer, there is no confusion regarding the presence of water. However this similarity will not be detected if only low level features are used.

Fig. 1. The color of ocean water exhibits a spectrum from green to dark blue

The effect of imaging condition on textures is explained in [11,12,13] using the CUReT textures database. Different textures might appear very similar resulting in large inter-class similarities or the same surface might exhibit different textures leading to large intra-class variations. Misleading image appearances is a common problem for remote sensing applications as the illumination varies with the time and season. The pose of camera does not vary much for the satellite images taken from great heights but it plays a significant role for aerial images taken from surveillance helicopters.

1.1 Our Contributions

1. We propose a new texture feature, the *'texton histogram'* to represent the characteristic weak textures of remote sensing images. We have shown that this feature is largely robust to the problems of unknown scaling, orientation, and global illumination.
2. We develop a classifier system to identify image contents semantically, using the texton histogram as the base feature. The accuracy of semantic classification is further improved using additional features obtained from an extensive feature selection procedure. We show that our system can handle the problems of misleading image appearances as well as that of unknown image properties.

3. We develop an efficient end-to-end system that retrieves results containing similar semantic contents in about 100ms (Matlab based, database size of 400 images).

1.2 Proposed Approach

The problem of similarity retrieval is posed as a semantic matching problem where an image is represented as a composition of high level concepts. We use six frequent remote sensing categories, viz., *bushes (forest), clouds, plains, snow, urban, and water* as the high level concepts. The application of a semantic approach helps in identifying image contents independent of the scale, orientation conditions as well as the intra-class feature variations and the inter-class feature similarities.

The mapping from low-level features to high-level concepts is done by Support Vector Machine (SVM) classifiers trained using multiple-instance learning approach. A feature selection technique, the *gain-ratio* method is used to choose concept-specific selective low-level features from the feature-space of color and texture features.

1.3 Organization of Paper

The paper is organized as follows. Sec. 2 discusses the features chosen to represent remote sensing imagery. The focus of this section is on the construction of the texton histogram, followed by a discussion of its ability to detect weak textures, irrespective of illumination, scale and orientation conditions. Sec. 3 discusses our semantic learning approach. The overall system architecture is described in Sec. 4. Experimental results are given in Sec. 5 followed by concluding remarks in the last section.

2 Features

The accuracy of a CBIR system can only be as good as the features used to represent images. If only gray-scale texture features are used, water and snow covers might be indistinguishable. Similarly, if only color is used, snow and clouds might appear indistinguishable. Hence, it is better to use a combination of carefully chosen multiple features to distinguish a category from another. In our experiments, we selected category-specific features from a feature-space of color, weak textures and strong texture features and used them to train a single SVM classifier for that category.

2.1 Texton Histogram

Textons are the putative units of preattentive human texture perception [14]. Different definitions are given in different works to compute textons. [15] gives an operational definition where textons are computed as the frequently co-occurring combinations of oriented linear filter responses. [13] defines textons as the joint

distribution of intensity values over extremely compact neighborhoods. Our definition of textons is inspired by the work in [13]. Our design is equally focused on local property that is a function of a 3 x 3 neighborhood *and* the *texton histogram* which is more global in nature. Based on an extensive set of experiments with one thousand seven hundred 128 x 128 image tiles, a *texton dictionary* is learned using an unsupervised process. Each item in the dictionary (a texton) is a pixel label computed from a large number of 3 x 3 local neighborhoods of various pixels. The process is summarized in Fig. 2 and Fig. 3.

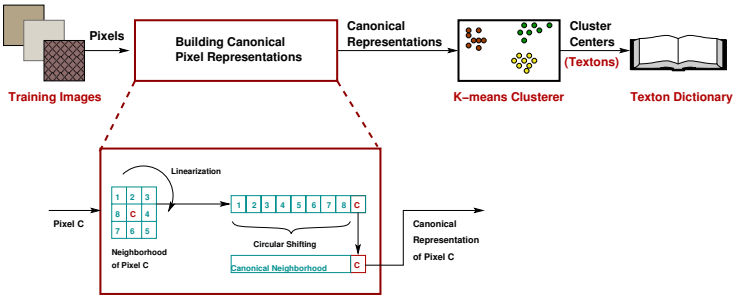


Fig. 2. Constructing a texton dictionary

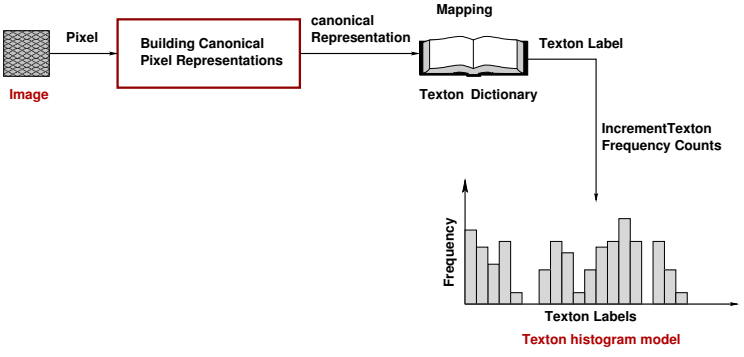


Fig. 3. Constructing a texton histogram. A texton histogram captures the local property of weak texture, but is also invariant to various effects.

Details. For a pixel p under consideration, the 3 x 3 local neighborhood without p is linearized to form an 8-element string representation s. This string is circularly shifted to yield a canonical form $\overline{S}$. The canonical representation satisfies two properties:

1. For any string representation s' of the same neighborhood, $\overline{S}$ is lexicographically smaller than or equal to s'.
2. The left and right neighbors for every element in $\overline{S}$ are the same as in s, under circular shift condition.

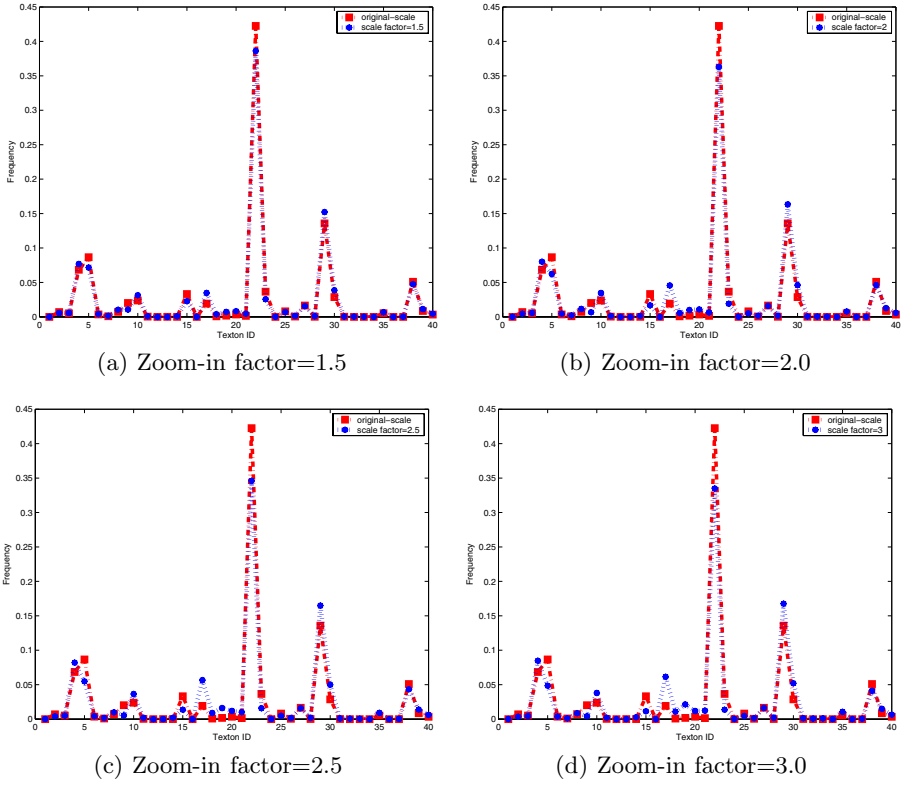
Pixel p is then appended at the end to form a 9 element vector for the next step.

The canonical representations for a large number of pixels are clustered using the K-means algorithm, where K is found automatically. The cluster centers are chosen as the representative textons to form the dictionary. Each texton is thus a 9-element array of tuples (mean, variance) corresponding to each of the 9 dimensions of the 3 x 3 local neighborhood. The textons in the dictionary are identified by a unique identifier, the texton-id. This procedure is depicted in Fig. 2. It is performed offline, and done exactly once in the system.

To summarize the weak textures for any candidate image, we build a probabilistic model, the texton histogram. After an image is intensity normalized, each image pixel is labeled with the closest item in the texton dictionary. The texton histogram feature is computed as the fraction of the total number of image pixels assigned per texton. The procedure for computing the texton histogram is shown in Fig. 3.

2.2 Texton Histogram Properties

1. **Invariance to global illumination changes:** Preprocessing images using *mean-center intensity normalization* makes the process more robust to illumination effects.
2. **Invariance to local neighborhood orientations:** Using a canonical form to represent a pixel neighborhood ensures that any orientation of the 3 x 3 neighborhood still maps to the same texton. Strictly speaking, we must scan convert a circle and use a circular neighborhood. The 3 x 3 neighborhood we use is simply a practical measure that works well.
3. **Invariance to noise in local neighborhoods:** Clustering ensures that the textons are well separated from each other in space. By binning pixel neighborhoods to closest textons, the problem arising from small noise and intensity fluctuations is overcome. Even if some pixels are mapped to the wrong textons, it does not have a significant effect on the final texton histogram representation.
4. **Invariance to scale:** Combining a local representation with an unsupervised voting process enables scale tolerance. Unlike ours, the texton histogram feature described in [13] is capable of resolving the misleading texture problem; however, it is not scale-independent. The scale associated with a remote sensing image is quasi-global [16]. This global nature is captured in a histogram model whereas the basic unit texton captures the local textures.
   Experimental proof for this appears in Fig. 4 which shows the behavior of texton histogram feature for a sample image under scaling. The plots show the behavior of texton histograms at zoom-in factors 1.5, 2.0 and 2.5 and 3.0 of the original image size. Observe that the overall shape of the texton histogram remains similar under scaling, and especially note that the peak positions match nicely. Fig. 5 demonstrates the behavior of texton histogram for the same image at a zoom factor of 0.2 marking its robust nature under zoom-out situations. After observing a similar behavior for a large number of images we conclude that the texton histogram feature is robust to image scaling to a large extent.



(a) Zoom-in factor=1.5 (b) Zoom-in factor=2.0

(c) Zoom-in factor=2.5 (d) Zoom-in factor=3.0

Fig. 4. Effect of scaling on texton histograms. The red and blue colors correspond to the texton histograms at the original scale and at the zoom factor respectively.

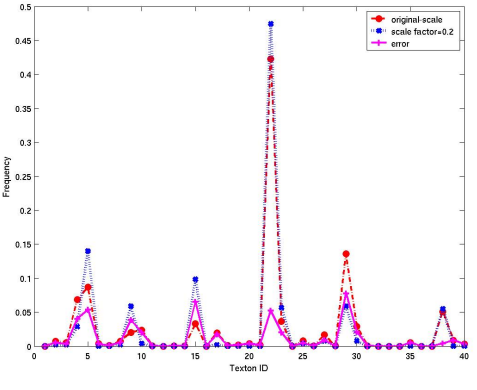


Fig. 5. Effect on texton histogram under "zoom-out" (factor=0.2)

2.3 Color and Strong Texture Features

To boost the identification of categories, we selected additional features from a feature-space encompassing color and strong texture features. In addition to the texton histogram (termed TH), we chose six color features, the dominant Y, Cb, Cr (DY, DCb, DCr) and the average Y, Cb, Cr (AY, ACb, ACr) values. The choice of YCbCr color space over RGB and HSV was made experimentally. We computed the strong texture features using a thresholded response of pixels to Sobel masks corresponding to the edges in directions 0, 45, 90, 135 degrees (termed EH0, EH45, EH90, EH135).

3 Learning Semantics

The association of distinguishing low level features to semantic categories is learned using a multiple instance learning (MIL) approach [17].

3.1 Multiple Instance Learning

In the multiple instance learning approach, an image is labeled positive for all the categories present in it. For a category, the task of learning distinguishing features reduces to identifying features which are common to the positively labeled images and absent from the negative images along with their relative weights. We use the 'gain-ratio' attribute selection [18] procedure to select an initial subset of useful features for each category. The gain-ratio method returns a ranking of all features for their discriminative capacity for the dataset under consideration. The SVM classifiers are tuned using a greedy selection [19] for these feature subsets . A binary SVM classifier is learned using the dominant features for each category. The final classifier package consists of 6 SVMs, one for each category.

For our experiments, we annotated 1700 image tiles of size 128 x 128 with positive/negative labels for each of the 6 categories. Full feature vectors and the corresponding labels were input to the feature selection process. The final feature dimensions selected for each concept are given in Table 1.

We observed that the inaccuracies in classification were mainly caused due to the variations in appearances of categories. The accuracy for clouds category is

Table 1. Table of concept-wise dominant features-set and classification accuracy

Concept	Dominant feature dimensions	Accuracy
bushes(forest)	TH,AY,ACb,ACr,DY	96.18
clouds	TH,DY,AY,ACr,DCb	87.19
plains	TH,ACr,ACb,DCr,DCb	90.11
snow	TH,DY	98.47
urban	TH,ACb,EH0,EH90,ACr	92.81
water	TH,ACb,ACr,AY,DY,DCr	93.23

relatively low, owing to the occasional sparse cloud nature where cloud detection is difficult and the occasional dense nature where it is confused with snow.

4 System Architecture

Fig. 6 shows the overall block diagram of the proposed retrieval system. The system can be explained in terms of three main modules: a) Learning module, b) Semantic profiles generation module and c) Query retrieval module.

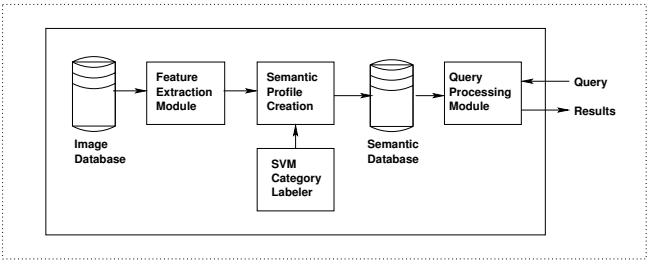


Fig. 6. Block diagram of the proposed system

Learning Module: The job of the learning module (not explicitly shown in Fig. 6) is to build the texton dictionary and the SVM category classifiers through learning. Both these tasks are done offline. The quality of texton dictionary and SVM classifier play a decisive role in assuring high quality results for the task of image classification and retrieval.

Semantic Profiles Generation: An image is divided in tiles of size 128 x 128 pixels. Features are computed for each tile and input to the SVM classifier package containing classifiers for the selected semantic categories. The output (i.e., label '1' if semantic concept is detected, '0' otherwise) of all the six classifiers is put together to construct a 6-element semantic profile for the tile. The semantic profile of the entire image is a 6-element vector where each element corresponds to the fraction of tiles voting positive for a concept. For example, if 3 out of 10 tiles vote for water and 8 out of 10 tiles contribute to land, then the semantic profile of image is {0.3 water, 0.8 land}. A tile may vote for any number of categories under consideration. A similar approach is described in [20] where an image is divided in 10 x 10 regions, each voting for a single category. Our framework differs in this aspect from [20], as for large sized images, a region is bound to contain more than one category. Hence it makes more sense to detect all of them and not restrict a region to a single label.

Query Retrieval Module: The semantic profiles for all the database images are stored in a semantic profile database. Given a user query, its semantic profile is constructed. The results are ordered based on the Euclidean distance between

the semantic profiles of the query and the candidate database image. Our approach also enables us to develop a framework for fuzzy queries, e.g., 'retrieve images containing largely water' or 'do not retrieve images containing any cloud cover'.

5 Experimental Setup and Results

To test the proposed technique, we developed a heterogeneous image database consisting of images from different on-line resources. Images showing none, one or more of the selected concepts were downloaded from the freely available image galleries of commercial satellite companies like Orbimage and Spaceimaging, and government organizations NRSA (India) and US-based NASA's 'Earth Observatory. The image database consists of 400 natural color satellite images, which are stored in JPEG format with sizes varying from 500 x 500 up to 25000 x 25000. The image resolutions vary from a few inches per pixel to a few meters per pixel. The images have been taken from across the globe, at different times of the day and across seasons making the illumination properties different. We have kept no metadata information about resolution, scale or orientation.

We evaluated the performance of the proposed system in two ways. First we computed the system performance statistically giving precision values. We also compared the retrieved results with the results of *'SIMPLIicity'* system using the same underlying image database. Like in most region-based retrieval systems, in SIMPLIicity, an image is represented by a set of regions, roughly corresponding to objects, which are characterized by color, texture, shape, and location. This system classifies images into semantic categories such as textured-nontextured, city-landscape, and so on. It uses a wavelet-based approach for feature extraction and an integrated region matching technique to match the image regions.

5.1 Performance of Query Retrieval

To provide numerical results, we asked 5 human annotators to manually check the relevance of results for 18 randomly chosen sample query images. For the same images, relevance of results given by SIMPLIcity is also evaluated by the same annotators. The top ten results are considered for evaluation and the precision is computed as the fraction of images retrieved correctly as per human judgment. For each of the eighteen query images, the average precision of results given by both systems are plotted in Fig. 7. We find that on an average, the precision of the proposed system is greater than that of SIMPLIcity by 0.342.

5.2 Query Comparison

Fig. 8 shows the comparison of results between the proposed system and the SIMPLIcity for a query image in which water appears green. The top row shows our experimental results and the bottom row shows retrieval results of SIMPLIcity. The leftmost image in each row is the query image. Due to the limitation of

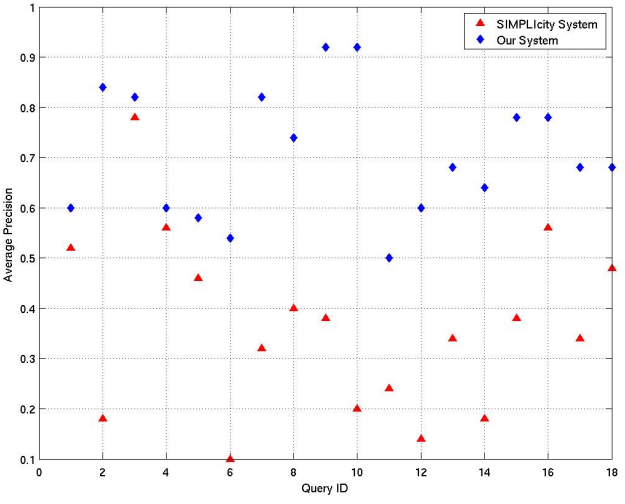


Fig. 7. Comparison of average precision values (best seen in color)

space, we have shown only top 10 matches for each query comparison. Our system has successfully identified that the query contains water and corresponding images involving water are returned irrespective of the intra-class variations in appearances.

More (favorable) results of the comparison do not appear in this version due to space limitations but are available at our website.

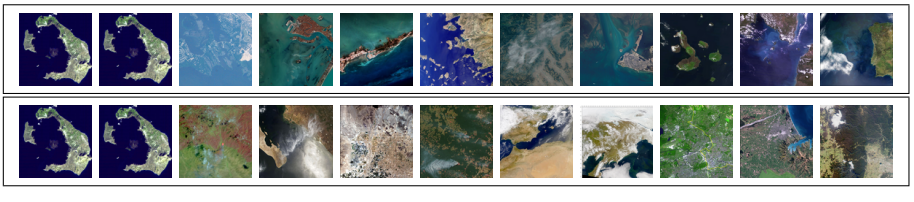


Fig. 8. Using semantic retrieval to overcome intra-class differences in appearances of a concept. The top row shows the first ten results of our system and the bottom row shows the results obtained using SIMPLIcity technique.

6 Concluding Remarks

Texton histogram is a robust feature capable of capturing the weak textured nature of remote sensing images in a scale, orientation and illumination independent manner. This feature along with other features can effectively learn the high level concepts present in remote sensing domain. Using such a semantic approach effectively counters the intra-class image variations and inter-class image

similarities. Hence, the proposed framework is able to characterize remote sensing images in a generic manner. However it should be noted that our framework does not handle spatial adjacency constraints.

Acknowledgments

We thank Appu Shaji and Vinay Namboodiri for the cafeteria-based brainstorming sessions, and Aniruddha Joshi for his inputs on the intricacies of training SVMs. We thank J. Z. Wang who enabled comparisons by providing the SIMPLIcity executable, and data.

References

1. Kitamoto, A.: Digital typhoon: Near real-time aggregation, recombination and delivery of typhoon-related information. In: Fourth International Symposium on Digital Earth. (2005) (CD–ROM)
2. Puzicha, J., Hofmann, T., Buhmann, J.: Non-parametric similarity measures for unsupervised texture segmentation and image retrieval. In: Computer Vision and Pattern Recognition (CVPR '97), Washington, DC, USA, IEEE Computer Society (1997) 267
3. Newsam, S., Wang, L., Bhagavathy, S., Manjunath, B.S.: Using texture to annotate remote sensed datasets. In: 3rd International Symposium on Image and Signal Processing and Analysis (ISPA). (2003)
4. Newsam, S., Wang, L., Bhagavathy, S., Manjunath, B.S.: Using texture to analyze and manage large collections of remote sensed image and video data. Journal of Applied Optics: Information Processing **43** (2004) 210–217
5. Wang, L., Liu, J.: Texture classification using multiresolution markov random field models. Pattern Recogn. Lett. **20** (1999) 171–182
6. Valeriano, M.I., Escada, S.: Mining patterns of change in remote sensing image databases. In: Fifth IEEE International Conference on Data Mining (ICDM'05). (2005) 362–369
7. Schroder, M., Rehrauer, H., Seidel, K., Datcu, M.: Spatial information retrieval from remote- sensing images - part 2: Gibbs-markov random fields. IEEE Trans. Geosci. Remote Sensing (1998) 1446–1455
8. Unser, M.: Texture classification and segmentation using wavelet frames. Image Processing, IEEE Transactions on **4** (1995) 1549–1560
9. Wang, J.Z., Wiederhold, G., Firschein, O., Wei, S.X.: Content-based image indexing and searching using daubechies' wavelets. International Journal on Digital Libraries **1** (1997) 311–328
10. Wang, J., Li, J., Wiederhold, G.: SIMPLIcity: Semantics-sensitive integrated matching for picture LIbraries. IEEE Transactions on Pattern Analysis and Machine Intelligence **23** (2001) 947–963
11. Varma, M., Zisserman, A.: Classifying images of materials: Achieving viewpoint and illumination independence. In: ECCV (3). (2002) 255–271
12. Varma, M., Zisserman, A.: Statistical approaches to material classification. In: Second Indian Conference on Computer Vision, Graphics and Image Processing. (2002) 167–172

13. Varma, M., Zisserman, A.: Texture classification: Are filter banks necessary? In: International Conference on Computer Vision and Pattern recognition. (2003) 691–698
14. Julesz, B.: Textons, the elements of texture perception, and their interactions. Nature **290** (1981) 91–97
15. Malik, J., Belongie, S., Shi, J., Leung, T.K.: Textons, contours and regions: Cue integration in image segmentation. In: ICCV (2). (1999) 918–925
16. Gilles, S.: Robust description and matching of images. Technical report, University of Oxford (1998) Ph.D. Thesis.
17. Yang, C., Lozano-Perez, T.: Image database retrieval with multiple-instance learning techniques. In: Proc. International Conference on Data Engineering. (2000) 233–243
18. Quinlan, J.R.: Induction of decision trees. In Shavlik, J., Dietterich, T., eds.: Readings in Machine Learning. Morgan Kaufmann (1990) Originally published in *Machine Learning*1:81–106, 1986.
19. John, G.H., Kohavi, R., Pfleger, K.: Irrelevant features and the subset selection problem. In: International Conference on Machine Learning. (1994) 121–129
20. Vogel, J., Schiele, B.: Natural scene retrieval based on a semantic modeling step. International Conference on Image and Video Retrieval CIVR 2004, Dublin, Ireland, LNCS **3115** (2004)