Hierarchical Image Classification

Stefan Kuthan
PRIP
Institute of Computer-Aided Automation
Vienna University of Technology
Favoritenstr. 9/1832
A-1040 Vienna Austria
stefan.kuthan@gmx.at

Allan Hanbury
PRIP
Institute of Computer-Aided Automation
Vienna University of Technology
Favoritenstr. 9/1832
A-1040 Vienna Austria
hanbury@prip.tuwien.ac.at

ABSTRACT

A framework for deriving high-level scene attributes from low-level image features is presented. The assignment of the attributes to images is done by a hierarchical classification of the low level features, which capture colour, texture and spatial information. A system for image classification is implemented, which aids in the evaluation of the different methods available. A detailed analysis of the best features for different classification tasks is presented. Classification and retrieval results on the ImagEVAL image dataset are provided.

Categories and Subject Descriptors

I.4.7 [**Image Processing and Computer Vision**]: Feature Measurement; I.5.4 [**Pattern recognition**]: Applications—*Computer Vision*

General Terms

Experimentation, Performance

1. INTRODUCTION

This paper presents our image classification system entered into Task 5 of the ImagEVAL 2006 campaign. It concerns the extraction of image semantic types (e.g. landscape photograph, clip art) from low-level image features.

A variety of applications for image classification and feature extraction can be found in *Content Based Image Retrieval* (CBIR). An application especially suited to the classification under consideration here is the automatic colour correction of consumer photos during film development [5, 7]. Another application could be the automatic classification of images in large electronic-form art collections, such as those maintained by museums or image archives of print media / television. Generally speaking, such a classification is useful everywhere where a manual classification or sorting process is infeasible because of the number of images under consideration.

There exists much work on this sort of image classification [1, 3, 7, 8, 9, 11], however papers often concentrate on a small subset of the classes given or even just a binary classification. Each evaluation is usually done on a different set of images, making it difficult to judge the effectiveness of the methods. This paper contributes by analysing the effectiveness of a large number of features for the tasks listed above. An effective feature combination method and hierarchical clustering approach is presented.

Sections 2 and 3 describe the system used to classify the images for the ImagEVAL campaign, where Section 2 presents an overview of the features extracted, while Section 3 describes the classification methods used. Section 4 presents detailed results of feature selection experiments performed on the ImagEVAL training data. The classification and retrieval results are presented and discussed in Section 5. More detailed information on the system and features used can be found in [4].

2. FEATURES

All input images are encoded in the RGB colour space. Therefore it would be of advantage to work with RGB since no conversion is needed. The drawback however is that this space is ill-suited for most classification based on colour. For example, different illumination will change the perceived colour. While the human eye will make adjustments to accommodate for this, it is hard to construct a metric for which an image has the same (pixel) values regardless of lighting conditions. The luminance information is more important to our perception than the chroma, a difficult fact to consider when using a colour-space where luminance is not directly available, rather being a combination of all three channels.

To capture colour information, histograms are calculated in several colour spaces. This section shows why the particular conversions were considered and details on the parameters chosen. The number of bins per channel is 20.

RGB Histogram. Although the RGB space was expected to perform worse than other colour spaces for the reasons mentioned above, there are good reasons for calculating a feature vector based on this space. An advantage is that no conversion errors are introduced. The classification of images into the nature and urban class was also expected to benefit from this space when considering the green channel which is expected to show higher values for the nature class.

Ohta Histogram. The Ohta colour space is proposed for indoor-outdoor classification in [7]. The first channel of this space captures brightness information as it is the sum of the three channels of RGB.

CIELUV / CIELAB Histogram. An advantage of both the CIELUV and the CIELAB colour spaces is that the Euclidean distance between two sets of colour coordinates approximates the human perception of colour difference. The luminance information is directly available in the first channel.

srgb Histogram. The calculation of the normalized RGB colour space[1] is performed as proposed in [1]. The "intensity free" image is computed by dividing each channel of RGB by the intensity at each pixel. The calculation of the intensities is as follows:

$$I = (299 * R + 587 * G + 114 * B)/1000 \qquad (1)$$

HSV. The HSV colour space, representing hue, saturation and colour value (brightness) has the shape of a hexagonal cone. The angle is given by the hue, the distance from the centre of the cone by the saturation and the vertical position by the value. This colour space is used for a part of the **colour statistics** shown in the following list:

- Illuminant: this value indicates the colour of the light source. It is calculated in two versions, through the "Grey-world algorithm" and the "White patch algorithm". The former is calculated by the mean of the three colour channels, which is assumed to be "grey" (multiplied by 2 to get white), the latter is calculated by assuming that a white patch is always visible in an image, therefore taking the maximum value of each channel.

- Unique colours: this value is calculated by transformation into the HSV-space and counting the unique values in the Hue channel.

- Histogram sparseness: a histogram is calculated and bins containing counts higher than a fixed cut-off value counted.

- Pixel saturation: this is calculated as a ratio between the number of highly saturated and unsaturated pixels in the HSV colour space [1].

- Variance in and between each channel of the RGB space.

The following texture features are calculated:

Edge direction. This feature is used to compare the frequency of occurrence of edge directions. As with colour, a histogram is used to discretise the values. For a greyscale image the gradient is calculated in two directions by convolution with the horizontal and vertical Prewitt kernels. The next step is the calculation of the magnitude and direction at each pixel x:

$$m(x) = \sqrt{f_h(x)^2 + f_v(x)^2} \qquad (2)$$

$$\theta(x) = arctan\left(\frac{f_v(x)}{f_h(x)}\right) \qquad (3)$$

where f_h and f_v are the horizontal and vertical edges.

Edge direction coherence vector. The calculation of the edge direction coherence vector is accomplished by a morphological closing of the magnitude image with a line segment followed by a morphological opening with a small disk. Thereby the dominating structures are enforced while degenerate "edges" – isolated pixels – are removed. As above, a greyscale image is used for the input. In both cases the result is a histogram of the direction image multiplied (masked) by the thresholded magnitude image. The 37 bins represent 5 degree intervals from -90° to 90°. The number of edge pixels found is stored in an extra bin of the histogram. Normalization with the image size is also performed.

Edge Statistics. This feature is used to determine whether the edges in the image result from intensity changes, as is the case with natural images, or from changes in hue, a method employed in paintings [1]. The intensity edges are found as above. The colour edges are found by first transforming the image into the srgb space, resulting in normalised RGB components. The colour edges of the resulting "intensity-free" image are then determined by applying the edge detector to the three colour channels and fusing the results by taking the maximum. The feature extracted is the fraction of pure intensity-edge pixels.

Wavelets. The Haar transform [5] is used to decompose an image into frequency bands. To extract an image feature this transform is applied to the L component of a CIELUV image. The square root of the second order moment of wavelet coefficients in the three high-frequency bands is computed. This image feature captures variations in different directions. In the implementation of the system 4 levels are computed. This yields a feature vector of length 12.

Gabor filter. The Gabor filter is a quadrature filter. It selects a certain wavelength range (bandwidth) around the centre wavelength using the Gaussian function. This is similar to using the windowed Fourier transform with a Gaussian window function. The feature vector is constructed by calculating the mean and standard deviation of the magnitude of the transform coefficients at several scales and orientations [6, 10]. This means that the fast Fourier transform (FFT) is applied to an image and then the Gabor filter, specific to this scale and orientation, is applied. Now the inverse of the FFT is taken and the mean and standard de-

[1] This is not the sRGB as defined by IEC 61966-2-1 "Default RGB Colour Space".

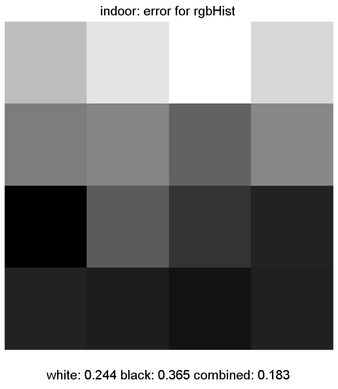


Figure 1: Using image tessellation to capture Spatial Information: Indoor-Outdoor

viation calculated. For the system this filter is applied at 6 orientations and at 4 scales. Two values are collected at each point; therefore the feature vector has the length 48.

3. CLASSIFICATION

For implementation of the system Matlab Version 6.5 was used. The library PRTools[2] Version 4.0.14 [2] is used to construct the classifier. The results reported in the next section were obtained with the k-NN classifier, where the number of neighbours is set to 5. Other tested classifiers are not used due to their complexity, sharply increasing computation time (neural net, Mixture of Gaussians), or because of their lower performance, probably because of the inability to model complex distributions (Linear and Quadratic Bayes and Parzen classifier). The Bagging classifier, based on k-NN and the Decision trees proved to be competitive but not as robust as the k-NN classifier.

3.1 Spatial Information

To capture spatial information, each image is divided into 16 sub-images. This 4×4 image tessellation is of benefit because image regions can be weighted according to their importance. For each sub-block a feature vector is calculated separately. A simple concatenation of these would increase the dimensionality by a factor of 16. To keep the classification simpler the following method is used: a classifier is built for each sub-block and a combining classifier, described in the next section, effectively weights the results of these.

A drawback of this approach is that only simple concepts can be captured through this method (e.g. blue sky at the top for outdoor images). Complex concepts, such as XOR cannot be solved. As an example for successful weighting, Figure 1 shows the error rate for indoor-outdoor classification based on the RGB histogram, averaged over the sub-blocks of 1000 test images when trained with 2000 images. In Figure 1, white represents the best error rate of 0.244% and black the worst with 0.365%. As can be observed the classification is better for the blocks in the upper part of the images, probably capturing the "sky" information. Also the combination of the results of the individual sub-blocks brings an improvement to an overall error rate of 0.183%.

3.2 Combining Features

The method used for incorporating spatial information is extended for several features straightforwardly. For each sub-block and for each feature a classifier is trained using a subset of the data available. Depending on the number of features used, between 16 (for one feature) and 64 (for 4 features) classifiers have to be trained.

The training of the sub-blocks with a subset of the data is done to introduce "unseen" data for the combining classifier. This avoids overfitting the combining classifier.

The output when applying a classifier is a value signifying the confidence with which each image belongs to the class under consideration. The trained classifiers are applied to all of the training data independently.

In the next step their outputs are concatenated to a feature vector and the combining classifier trained. The number of classifiers for each sub-problem is therefore the number of blocks times the number of features plus one.

Experiments were also carried out with the possibilities for combining classifiers provided by PRTools. These are: Product, Mean, Median, Maximum, Minimum and Voting combiner. However classification with these combiners generally shows an error rate higher than that achieved with the scheme above.

3.3 Hierarchical Classification

A hierarchical classification similar to that described in [8] is implemented. The classifier for the whole problem is organised in the hierarchy shown in Figure 2.

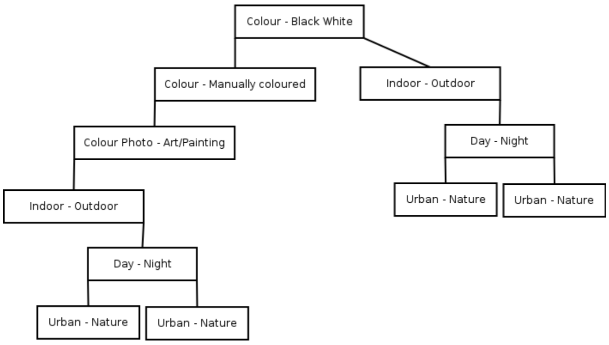


Figure 2: Hierarchy of Classifiers

At each node the training or application of a classifier takes place. Only the appropriate sub-sample of images, as determined by the node, is passed to the children nodes. At leaf nodes training or classification stops. This is a divide and-conquer strategy with several advantages. One advantage, compared to a classification of all attributes at once, is reduced complexity through reduction to two-class problems. Also there is no need for a third class of images belonging to none of the classes under consideration.

[2] Pattern Recognition Tools: available at `http://www.prtools.org/`

Each node can be configured individually. The system currently has settings for: enabling/ disabling classification, list of low-level features selected, prior probabilities, chosen combining scheme (classifier, voting scheme) and the list of children, if any. This structure could be extended for parameters specifying the type of classifier (k-NN, decision trees etc.) and parameters to use. During the training phase the obtained classifiers are also stored in this structure.

This scheme also helps to keep the feature-vector used for training and during classification as small as possible, for example for day-night classification only one feature is used.

The logic of the problem-domain is easy to implement through the setting of the "children" list. This allows for a relatively easy extension to other attributes. Through this integration of the logic, inherent in the targets, a plausibility-check is not needed for the class labels (e.g. a setting of two contradicting labels does not lead to an error). The hierarchy shown in Figure 2 was obtained through analysis of the problem domain.

When applying the classifier, classification stops at the leaf nodes. This leads to an increase of speed and could be further exploited to only extract the needed features for each image.

Each of the nodes can be analysed separately. Figures such as the one shown in Figure 1 are available for each attribute and feature pair and help to interpret performance at each node.

4. FEATURE SELECTION

We compare the capability of each of the features described in Section 2 to successfully perform one of the required binary classifications. For the comparison of features and also as a means to test their variance, box plots were created with a (smaller) sample of 700 training and 200 test images. Figure 3 shows these box plots. On the *x-axis* the following features are shown: histograms in five colour spaces, colour statistics, edge statistics, edge direction histogram and coherence vector, wavelets, Gabor filter and the combination of all features. On the *y-axis* the error-rate can be read off. Each box is limited by the lower quartile (25% of the data) and the upper quartile (75%). The median is indicated by a horizontal line. Whiskers and crosses show the extent of remaining data. These results are used to manually select the best features for each sub-problem.

To measure accuracy and retrieval effectiveness, the following statistics are collected for each classification task:

tn true negatives: the instances correctly classified as negative;

tp true positives: the instances correctly classified as positive;

fp false positives: negative instances wrongly classified as positive;

fn false negatives: positive instances misclassified as negative.

These four values are summarised in a confusion matrix. This matrix has the following form:

a	b	<– classified as
tn	fp	a
fn	tp	b

As can be observed the sums of the rows of each class show how many instances belong to either class, whereas the sums of the columns show how many instances are classified to belong to each class.

Each of the following sub-sections considers one of the binary classification problems to be solved. Below each heading, the best features for the task (features chosen) as well as the statistics on the correctly and incorrectly classified images. Note that the total number of images to be classified can be smaller for the classification nodes lower in the hierarchy. The baseline is calculated by division of the size of the bigger class by the total number of instances. This is the best result possible when guessing the class, without any feature available. A discussion of the features follows.

4.1 Black and White

Features Chosen:	Lab, cStat	
Correctly Classified Instances:	990	99.0 %
Incorrectly Classified Instances:	10	1.0 %
Total Number of Instances:	1000	
Baseline:	79.7%	

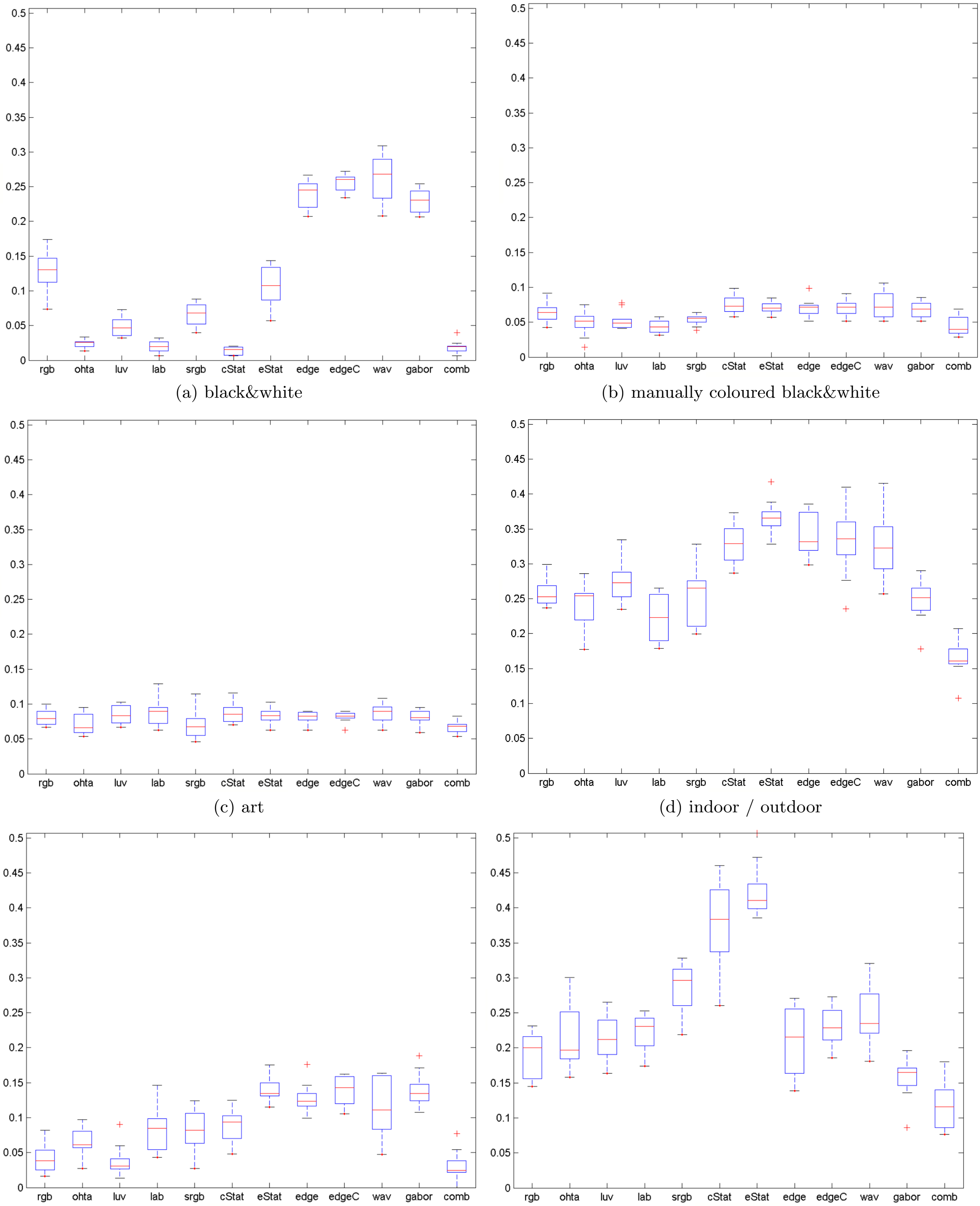
Confusion Matrix

other types	black white	<– classified as
793	4	other types
6	197	black white

The features used for the colour - black and white classifier are the colour statistics and the CIELAB histogram. The results achieved are very good, supporting the decision to choose these features. An interesting result shown in the box plot (Figure 3a) is that pure texture features perform significantly worse than colour features. While this result is not surprising it does show that the content (i.e. objects, image composition) of the images of the two classes is quite similar.

An analysis of the CIELAB histograms shows that this colour space is well suited for this problem because the chrominance is available separately. Colour images show a Gauss-like distribution in these two channels while black and white or greyscale images show a single spike around the value representing zero or achromaticity and a very small percentage of other chrominance values. The separation of the classes in the CIELAB colour space is not perfect due to the inclusion of sepia images into the black and white class. The box plot shows that the RGB space, where chrominance as well as luminance is a product of the three channels, is not suited for this classification.

The colour statistics show good results because in black and white images there is nearly no variance between the three channels in the RGB colour space whereas colour images have a high variance. Again, sepia images are the reason for



(a) black&white

(b) manually coloured black&white

(c) art

(d) indoor / outdoor

(e) night / day

(f) urban / nature

Figure 3: Box plots showing the errors for tested features.

a small error. Both features show slightly better results for the four sub-blocks in the centre, probably because some images contain a border of a different colour/luminance. The combination of all features available did not yield a significantly better result than each of the two features chosen.

Images of the colour class misclassified as black and white have a colour distribution similar to the sepia images or are very bright with little contrast. Nearly all misclassified black and white images are sepia images reported as manually coloured black and white.

4.2 Manually Coloured Black and White

Features Chosen:		Lab
Correctly Classified Instances:	961	96.1 %
Incorrectly Classified Instances:	39	3.9 %
Total Number of Instances:	1000	
Baseline:	93.4%	

Confusion Matrix

other types	Coloured BW	<– classified as
918	16	other types
23	43	coloured BW

For the classification of images into the classes colour and manually coloured black and white images the CIELAB histogram is used. The results achieved are good, but compared to the classification in the previous section, the recall of the smaller class (manually coloured black and white) is not quite as satisfactory. This indicates that the unequal distribution of instances is used to bias the classifier towards the larger class.

The box plot (Figure 3b) shows that all features available perform nearly equally well on this classification task and that a combination of all features does not yield a significantly better result. The CIELUV as well as the CIELAB colour space seems to be best suited. An analysis of the CIELAB histograms shows that the L^* channel has a slightly different distribution for the two classes under consideration. While the colour images follow a near Gaussian distribution in this luminance channel, the distribution for the manually coloured images has two peaks. The second maximum represents noticeably higher luminance values than found in colour images. The third channel of the CIELAB histogram, representing yellow-blue chrominance also shows a significant variation between the two classes. The values for the manually coloured images show less variance around the zero value, representing achromaticity. The performance of the sub-block classifiers shows very little variance with respect to error-rate. As in the classification of black and white images, the texture features perform marginally worse than the colour features.

Misclassified images of the manually coloured class are assigned to the art or colour classes. The attribute “manually coloured image” is wrongly assigned to images of the colour class when an image contains uncommon colours, as with outdoor images with a lot of fog, but also to an aerial image. As mentioned above sepia images are difficult to classify into this class or the black and white class.

4.3 Art

Features Chosen:		srgb, wav
Correctly Classified Instances:	949	94.9 %
Incorrectly Classified Instances:	51	5.1 %
Total Number of Instances:	1000	
Baseline:	91.4%	

Confusion Matrix

other types	art	<– classified as
900	14	other types
37	49	art

The features selected for the classification of images into the classes photographic image - artistic reproduction/ paintings are the sRGB histogram and the wavelet filter. Similar to the observations made with the classification of images into the colour - manually coloured classes, the box plot (Figure 3c) does not show a single feature outperforming the others and the combination of features does not seem promising. Also the results obtained show a lower recall rate on the smaller class (paintings).

The CIELAB histogram has slightly higher luminance values for the “art” class but a higher deviation is found in the green channel of the sRGB histogram. The distribution of the histograms in this colour space is generally less spread out for images belonging to the class of paintings. The wavelet filter shows higher values for colour images, this represents texture detail, indicating that paintings are less structured than photos of (natural) scenes. Both features show little variance with regard to the results of the sub-block classifiers; however slightly better values are observed for the off-centre blocks. This could be a result of the general layout of paintings, with the subject in the centre and less detail near the borders.

As mentioned above some manually coloured black and white images are wrongly assigned to this class. Furthermore colour images of richly decorated indoor scenes (palaces, gold plating) are considered as art, as are some outdoor scenes. Difficult to interpret is the reason why many art images are misclassified as colour photos. These images mostly have a realistic colour layout and a higher level of detail.

4.4 Colour Photo

Correctly Classified Instances:	933	93.3 %
Incorrectly Classified Instances:	67	6.7 %
Total Number of Instances:	1000	
Baseline:	64.6%	

Confusion Matrix

other types	colour photo	<– classified as
305	49	other types
18	628	colour photo

This table is a summary of results obtained so far in that it shows the error in classification for colour photos versus the other types in question. Through the hierarchic classification, the classifications performed until this point discriminate black and white images, manually coloured images and paintings from a general “colour” class, therefore what we are left with here are photographic colour images. Not considered, but also not part of the training sample, are black and white artistic reproductions. For the further classification only colour photos and black and white photos are considered.

4.5 Outdoor - Indoor

Features Chosen:	rgb, Lab, edgeC, wav	
Correctly Classified Instances:	693	83.5 %
Incorrectly Classified Instances:	137	16.5 %
Total Number of Instances:	830	
Baseline:	63.5%	

Confusion Matrix

outdoor	indoor	<– classified as
456	69	outdoor
68	237	indoor

For the classification of images into the indoor or outdoor class the following features are selected: RGB and CIELAB histograms, coherent edge direction histogram and the wavelet filters. The result obtained in the classification process is not as good as those covered so far. However the results obtained by other authors (82% to 93%) are comparable because their training and test sets are often smaller and ambiguous images are eliminated beforehand. An interpretation of the box plot (Figure 3d) is that generally colour features seem to perform better than texture features, what is striking is that the combination of all features yields a much better result than any single feature. Also the Gabor filter performs as well as the colour features.

An analysis of the RGB and CIELAB histograms shows that indoor images have slightly less luminance and (therefore) less highly saturated pixel values. Also the sub-block classifiers for these features perform better for the upper half of the images, this can be attributed to the presence or absence of a sky or alternatively that this area best reflects lighting conditions. The values obtained through the Gabor filters show higher values for the indoor class, indicating more structure or highly textured images. For the final implementation of the system the Gabor filter was deselected because of its high computational costs, however the wavelet filters, selected instead, show a similar response for this classification. As with the Gabor filter the result of the wavelet operation shows higher values for the indoor class. The coherent edge direction histograms show higher values for the outdoor class, seemingly contradicting this observation. Both classes show peaks at the values indicating horizontal, vertical and diagonal structures -90, -45, 0, 45 and 90 degrees. This effect is somewhat more pronounced for the indoor class.

The results obtained in combining the said features are similar to the combination of all features, as indicated by the feature “comb” in the box plot. It has been indicated in several papers that a combination of features has most effect when combining features of the “colour” group with those of the “texture” group.

The reason for indoor images to be classified as outdoor often seems to be lighting conditions caused by the presence of windows or doors. Also a strong presence of green or, in the case of black and white images, a bright background seems to bias the images into this class. The outdoor images classified as indoor either show very high detail or cluttering of the image or depict outdoor scenes with lighting common to indoor images, e.g. during dawn and dusk.

4.6 Night - Day

Features Chosen:		Luv
Correctly Classified Instances:	435	96.5 %
Incorrectly Classified Instances:	16	3.5 %
Total Number of Instances:	451	
Baseline:	86.8%	

Confusion Matrix

night	day	<– classified as
36	6	night
10	399	day

The classification of images into the day - night classes is achieved using the CIELUV histograms. As can be observed in the box plot (Figure 3e), colour features perform better than texture features and the combination does not bring an improvement over using the CIELUV colour space. The accuracy achieved is acceptable and the recall rates are good for both classes. The means of the CIELUV histograms for this classification problem show a distinct deviation in the luminance channel. As can be expected photos during daylight are much brighter than night shots. Interestingly enough, the chrominance values of the night class are higher than those of the day class. This might be a conversion error due to the little luminance and therefore little hue information available or alternatively the presence of light emitting objects. The results for the sub-blocks are slightly better for the upper half of the images.

The reason for misclassification of day scenes is often a very dark sky and in one instance an underwater image with black background. Night scenes misclassified as day were taken during dusk, ambiguous even to a human observer. Three images show city scenes with man-made lighting.

4.7 Urban - Nature

Features Chosen:	rgb, edgeC, wav	
Correctly Classified Instances:	393	87.1 %
Incorrectly Classified Instances:	58	12.9 %
Total Number of Instances:	451	
Baseline:	63.2%	

Confusion Matrix

urban	nature	<– classified as
243	22	urban
36	150	nature

For the classification problem nature - urban the following low-level features where selected: RGB histogram, coherent edge direction histogram and wavelet filters. As with the indoor - outdoor classification this seems to be a harder problem. The 87% hit-rate achieved lies close to the results reported on the problem by other papers The box plot (Figure 3f) shows that colour features as well as texture features are suited for the classification, also the combination of features promises an improvement.

The analysis of the RGB histograms does not yield explicit evidence other than that all channels have higher values in the nature class. The distribution in the CIELAB colour space confirms a higher luminance for nature images. This can be attributed to a different illumination, rather counter-intuitive is that there is no abundance of green in nature images. The histograms of the coherent edge direction suggest more ordered structures in the urban class. The peaks and valleys are more pronounced for this class. While both classes have a maximum at 0°, representing horizontal detail, the urban class has maxima at -90 and 90 degrees, representing vertical detail, while the nature class has a near equally spaced distribution. This feature therefore draws on the assumption that nature images have smaller, chaotic structures than urban images. Both the Gabor and the wavelet filters show higher values for the urban class at all scales and orientations, confirming this observation.

The sub-blocks in the centre have higher accuracy for the RGB and the wavelet feature. For the coherent edge direction histogram the best results are achieved in the lower part of the images.

Natural images classified as urban show a very highly structured composition. This is caused by trees, rocks or landscape formations e.g. canyons. Some cases also show man made structures in the foreground, e.g. castles or walls. Abundance of sky or green plants as well as the presence of lakes and rivers in urban scenes seems to be the main reason for the wrong assignment of the attribute nature.

5. RESULTS

This section presents the overall classification and retrieval results using the proposed system with the features selected in the previous section.

5.1 Classification Results

For the evaluation of the system a sample size of 2000 images is chosen for training and 1000 images are used for testing. The sample sizes were chosen for the purpose of faster testing.

Table 1 shows the obtained percentage of correctly classified images compared to the baseline value for each task in the classification hierarchy. The features were chosen by analysis of the box plots (Section 4). Table 2 shows the percentage of correctly classified images achieved with the features under consideration for each of the classes.

task	features	% correct	baseline
BW	Lab colStat	99.0%	79.7%
Man.col.	Lab	96.1%	93.4%
Art	srgb wav	94.9%	91.4%
Photo	all of above	93.3%	64.6%
Indoor	rgb Lab edgeC wav	83.5%	63.5%
Night	Luv	96.5%	86.8%
Nature	rgb edgeC wav	87.1%	63.2%
Total		**71.0%**	20.0%

Table 1: Summary of Results

group	ground truth
autumn	colour Photo, outdoor, day, nature
night	colour Photo, outdoor, night, urban
kitchen	colour Photo, indoor
paintings	art
ny_city	colour Photo, outdoor, day, urban

Table 3: Selected groups from the Corel Image Database

Black and white images have little variance between the channels, a small error is made though the misclassification of sepia images. Manually coloured images have a colour distribution that is uncommon in natural images. The classifier for the attribute art draws on the observation that natural images have more structure and a more even colour distribution than images of this class. Images that are not specified as belonging to any of the classes mentioned thus far, are classified as colour photo. The error for assignment of the attribute night can be traced to ambiguous images, taken at dusk, with dark sky or underwater. Classification of images into nature and urban classes is based on strong vertical and horizontal structures in urban scenes and on colour differences, generally nature images have higher values in all three channels of the RGB histograms.

An accuracy of 71% is achieved on the whole problem, i.e. assigning up to 4 attributes to an image. The images considered to be incorrectly classified have one or more wrongly assigned attributes. This value is not simply the sum of errors reported in the previous sections. The baseline error, the assignment of the class with the highest probability, would yield 20% hit-rate with this image database; however this would also imply a recall rate of 0% for all other classes in question because all images would be assigned to the same class. The percentage of correctly classified images on the 3474 images not used for training is 72.4%. This suggests that the classifier generalises well and can be expected to perform comparably on similar data.

To compare the results on a different dataset a test run was made on a part of the Corel image database used in [5] and [11]. A sample of 500 images was selected; Table 3 shows the chosen image groups and the attributes manually assigned to them. Each group contains 100 images.

The classifier was trained on 400 images (using a random draw) and tested on the remaining 100. The percentage of correctly classified images is 80%. The discrepancy between the results can be explained by the different domain of the ImagEVAL images compared to the 500 images selected from the Corel image database.

5.2 Retrieval Results

The final version of ImagEVAL Task 5 was posed as an image retrieval problem, requiring that each image be assigned a confidence that it satisfies a query. We calculate one global confidence measure per image simply by multiplying together the confidences of the classifications in all of the nodes of the classification hierarchy which participate in the classification of the image.

Class	rgb	ohta	luv	lab	srgb	colour-stats
art	91.9	92.9	91.5	91.2	***93.1**	91.3
blackWhite	87.4	97.6	95.1	98.0	93.5	***98.6**
bw_Col	93.6	95.1	94.7	***95.7**	94.7	92.5
day	95.9	93.6	***96.2**	91.6	91.8	91.1
indoor	74.2	75.7	72.4	***77.7**	74.8	67.1
nature	*80.9	78.4	78.6	77.8	71.3	61.9

Class	edge-stats	edges	edges C.	wavelet	gabor	combined
art	91.6	*91.8	*91.8	91.2	*91.8	93.3
blackWhite	*89.3	76.1	74.4	73.7	76.9	98.1
bw_Col	93.0	93.0	92.9	92.5	*93.2	95.6
day	85.9	87.3	86.1	*88.6	85.9	96.8
indoor	63.4	65.8	66.9	67.6	*75.4	83.5
nature	57.6	78.8	77.0	75.4	***84.2**	88.1

Table 2: Comparison of Features - the percentage of correctly classified images is given; top: colour histograms, bottom: texture features. The best single feature for a class is in bold, an asterisk marks best result of each row.

Two runs were submitted, with the difference being in the number of images used to train the classifiers for each subblock of the image tesselation and the number of images used to train the combined classifier. For run PRIP01, 5000 images were used for training the sub-block classifiers and 474 for training the combined classifiers. For run PRIP02, 4000 images were used for training the sub-block classifiers and 1474 for training the combined classifiers.

The Mean Average Precision (MAP) for run PRIP01 was 0.3676 and for run PRIP02 was 0.3141. This shows that having more training data for the sub-block classifiers is important. This is particularly visible for query 1: “Art”, which shows the most significant difference between runs (MAP of 0.4949 for run PRIP01, 0.0748 for run PRIP02).

6. CONCLUSION

A detailed analysis of a number of features commonly used for image classification is presented. The result of image classification is used to evaluate and compare the discrimination power of several features on the given problems. Secondly, conclusions about the reasons why particular features are suited to a problem are drawn. This is done through an analysis of results and variables available at sub-stages during the training and testing phases.

The best features for each binary classification are then used in a classification hierarchy to assign a set of attributes to an image. Good classification results are obtained, with 72.4% of the 3474 images used as a test set being assigned a full set of correct attributes. The poorer image retrieval results are possibly due to a poorly defined confidence measure in the overall classification. It could also be linked to the small amount of training data available for our classification-based approach. The fact that the baseline results are high for some classes (over 90% for the art and manually coloured classes, see Table 1) demonstrates the small number of samples of these classes in the training data. This is also demonstrated by the large difference in retrieval results for the art class (query 1) when the training set is increased from 4000 to 5000 images.

The hierarchical classification makes use of knowledge about the problem-domain. The attributes to be assigned to the images are mutually exclusive and cover a wide spectrum of input images.

The ambiguity of natural language, where, for example, “nature” and “urban” is not explicitly defined and leads to problems in classification of images that cannot be accurately described with either word, is an unsolved problem.

An improvement of feature extraction speed would be of advantage, not only in use of the system with large image databases, but also to rapidly test other parameter settings and low-level features.

Acknowledgements

This work was supported by the European Union Network of Excellence MUSCLE (FP6-507752), and the Austrian Science Foundation (FWF) under grant SESAME (P17189-N04).

7. REFERENCES

[1] F. Cutzu, R. Hammoud, and A. Leykin. Estimating the photorealism of images: Distinguishing paintings from photographs. In *Computer Vision and Pattern Recognition. Proceedings. 2003 IEEE Conference on*, volume 2, pages II – 305–12, June 2003.

[2] R. Duin, P. Juszczak, P. Paclik, E. Pekalska, D. de Ridder, and D. Tax. Prtools4 a matlab toolbox for pattern recognition, 2004.

[3] Q. Iqbal and J. Aggarwal. Applying perceptual grouping to content-based image retrieval: building images. *Computer Vision and Pattern Recognition, 1999. IEEE Conference on*, June 1999.

[4] S. Kuthan. Extraction of attributes, nature and context of images. Technical Report PRIP-TR-101, Vienna University of Technology, 2005.

[5] J. Li and J. Wang. Automatic linguistic indexing of pictures by a statistical modeling approach. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(9):1075 – 1088, Sept. 2003.

[6] B. S. Manjunath and W. Ma. Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI - Special issue on Digital Libraries)*, 18(8):837–42, Aug 1996.

[7] M. Szummer and R. W. Picard. Indoor-outdoor image classification. In *Content-Based Access of Image and Video Database, 1998. Proceedings., 1998 IEEE International Workshop on*, pages 42 – 51, Jan. 1998.

[8] A. Vailaya, M. A. T. Figueiredo, A. K. Jain, and H.-J. Zhang. Image classification for content-based indexing. *Image Processing, IEEE Transactions on*, 10(1):117 – 130, Jan. 2001.

[9] A. Vailaya, A. K. Jain, and H.-J. Zhang. On image classification: City vs. landscape. In *Content-Based Access of Image and Video Libraries, 1998. Proceedings. IEEE Workshop on*, pages 3 – 8, June 1998.

[10] T. Wagner. Texture analysis. *Handbook of Computer Vision and Applications, Signal Processing and Pattern Recognition*, 2:275–308, 1999.

[11] J. Wang, J. Li, and G. Wiederhold. Simplicity: semantics-sensitive integrated matching for picture libraries. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(9):947 – 963, Sept. 2001.