Region-Based Image Classification for Automatic Color Correction

Clément Fredembach[1], Michael Schröder[2], Sabine Süsstrunk[1]
[1]Audiovisual Communications Laboratory (LCAV)
Swiss Federal Institute of Technology (EPFL),
Lausanne, Switzerland
[2]Swiss Technology Consulting Group AG
Zurich, Switzerland

Abstract

This work presents a framework for image classification based on region information. The main application of this framework is for automatic color correction of real-scene images. Class specific color correction is already implemented in some systems, but image classification usually relies on global image features only. Our region-based approach shows a significant improvement over previously achieved classification for a variety of critical image classes. Improved image classification can directly be translated into improved automatic color correction.

1. Introduction

The presented work is based on the idea that region-level analysis can improve image scene analysis and content description. The proposed framework is comprised of several successive steps, starting with the image to be analyzed and ending with its classification:

1. Image Segmentation
2. Region Features Calculation
3. Region Classification
4. Image Features Calculation
5. Image Classification

The primary goal of our approach is to improve the performance of class-specific Automatic Color Correction (ACC) algorithms [1]. To achieve a good correction, an ACC algorithm has to be able to distinguish between color cast and scene content. For example, it needs to distinguish between a blue object and a bluish color cast in case of scene illuminant estimation. One of the current problems with "simple" correction algorithms (such as *max RGB* [2], *grayworld* [3] or *retinex* [4]) is that each one performs well for some classes of images but poorly for others [1]. If an algorithm is tuned such that it improves the quality of all classes of pictures, it will result in a weak correction and the image improvement will only be marginal. The class-specific approach first attempts to identify the image content to determine which "simple" correction has to be applied (Fig. 1).

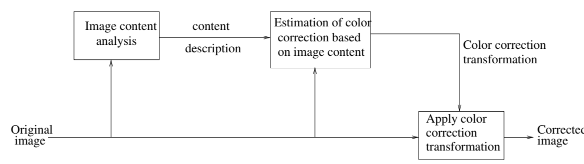

Figure 1: Design of a fully automatic color correction. The image content is directly used to determine which algorithm is appropriate to obtain the transformation applied to the images.

2. Image Segmentation

Segmentation is a process undertaken for every image. Thus, it has to be computationally effective and relevant. Since region analysis is based on regions as defined by the segmentation step, no robust classification is possible without a satisfactory segmentation method.

We chose Dominant Colors in Lab (DC Lab) to segment our images. DC Lab is a segmentation based on MPEG-7 [5]. It uses a k-means powered algorithm to segment images according to Lab clusters. The k parameter is fixed to 8 and another distance parameter d has to be specified. This parameter represents the minimal distance between two cluster centroids under which the clusters are merged. Its simplicity, given that it doesn't utilize any texture features, makes it very fast while at the same time allowing good quality segmentations. Arguments for using this algorithm over other available ones [6,7] include its computational speed (real time processing) as well as its ease of use.

3. Region Features

Segmenting images permits to identify regions, but classification also requires the presence of features to discriminate between different regions. To be effective, a feature should, via a single value or vector of variables, express a region. Region features usually belong to one of the following categories: *color*, *texture* or *geometry*.

3.1. Color Features

Color features are an essential group of region features. They are the easiest features to implement while still being relevant for natural scene classification. We implement color features in two different color spaces: First Lab, for it is the space in which the images are segmented and because it is designed as a perceptual space, thus facilitating the semantic interpretations of the regions. Then sRGB since it is the color encoding in which images are obtained. The color features are the mean and the variance of each channel (L, a, b, R, G, B), resulting in twelve color features for each region. A 13^{th} feature, the mean ΔE $(M_{\Delta E})$ of a region, can be obtained with those values:

$$M_{\Delta E} = mean(\sqrt{(L-L_i)^2 + ((a-a_i)^2 + ((b-b_i)^2})$$

L_i represents the luminance of an individual pixel *i*, and L is the mean luminance of the region. The mean is calculated over all the pixels belonging to a region.

3.2. Texture Features

Texture is an essential descriptor. Whether for vegetation or sky, it is the only measure (past color) that is easily accessible. It further has been shown that texture features have a very high discriminating power, enabling distinctions that would otherwise prove impossible [8].

By analyzing a region's spectrum in the Fourier transform domain, one can obtain information about the intensity of frequency changes (i.e. a measure of homogeneity), thus allowing to distinguish between a textured region and an homogeneous one. To do so, one can use the mean amplitude spectrum of the Fourier Transform of a region's luminance (*L*) channel, which represents a texture intensity measure.

Haralick features are based on the gray level co-occurrence matrix (GLCM) [9]. It allows for the calculation of textural, spectral and contextual types of features. Based on previous work of grass texture detection [10], we implemented 7 of those features by calculating the GLCM for regions. The retained features are: *Energy, Contrast, Homogeneity, Inverse Different Moment, Entropy, Cluster Shade, Cluster Prominence*. The first 5 features have been defined by Haralick [9], the last 2 have later been added by Smits [11].

3.3. Geometrical Features

Geometry features encompass position, size and shape information. All those characteristics can be used to further refine a classification obtained through color and texture features. The classes which will the most benefit of such features are "localized" classes, such as sky regions.

The relative area of a region can have many uses. It can approximate the relative importance of a region and can provide hints about the contents of a region. The relative area is obtained by dividing the area of the region by the area of the image. In addition, its behavior can be approximated by a probabilistic function, allowing for area weighted methods to be implemented in image classification.

Eigenregions are geometrical features encompassing both position and shape information. Obtained through a principal component analysis of the regions' location, the resulting eigenvectors can then be used as geometrical features [12]. Altogether, the 10 first eigenregions are used as region features.

4. Region Classification

The experimental data to classify consists of 77'000 regions obtained through DC Lab segmentation of 13'500 real-scene images. The retained classifier is a supervised classification algorithm based on the *maximum a posteriori* rule [13]. The main advantages of supervised methods are that they usually provide better results when precise definitions of the classes are given. Directing the classification towards known labels will make it more robust and less sensitive to noise. A major drawback, however, is that only previously known elements can be found, making the classification dependent on the accuracy of defined classes [14].

We chose the following classifier because of its robustness and scalability. Let A_ν be the possible semantic classes for a region (e.g. "vegetation", "skin",...). The probability that a region having a feature vector $\vec{F}$ belongs to a class A_ν is given by

$$p(A_\nu \mid \vec{F}), \quad \nu = 1, \cdots, N \tag{1}$$

These conditional probabilities are *a priori* unknown but are related to the probability of a certain feature vector $\vec{F}$ given that the region belongs to a class A_ν $(p(\vec{F} \mid A_\nu))$, which can be obtained from the training data by Bayes' theorem.

$$p(A_\nu \mid \vec{F}) \quad = \quad p(\vec{F} \mid A_\nu)\frac{p(A_\nu)}{p(\vec{F})} \tag{2}$$

where $p(A_\nu)$ is the *prior* probability that a given region belongs to the class A_ν , i.e. the chances that a randomly selected region in our database belongs to that given class, which is obtained by ground truth.

In our implementation, the main assumption of this method is that the probability distribution of the features $p(\vec{F} \mid A_\nu)$ are of the form of a multivariate gaussian. This model is generally considered accurate enough to allow for a robust classification.

In case of supervised classification, one of the most important task is the definition of the classes. Since the data consists of real-scene images and the application is automatic color correction, our class labels are memory colors [1], for which an optimized correction is known and implemented. We classified regions belonging to the following classes: *blue sky*, *vegetation* and *skin tones*. A fourth class, "normal" was also defined for regions belonging to none of the memory colors.

Supervised classification requires a ground truth for both training and testing the classifier. Thus, one has to accurately define what does belong to the ground truth and what does not. For images, the inclusion in the ground truth is based on whether the image *contains* a certain class. To take full advantage of the scale change, however, our assumption for regions is that in order to be included in the ground truth, they have to *be* a certain class, as shown in figures 2 and 3.

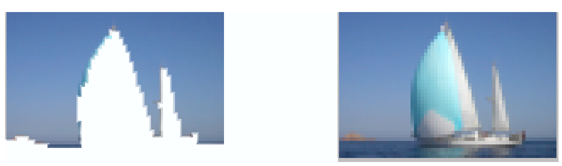

Figure 2: Example of a region belonging to the blue sky ground truth. The region is only composed of blue sky, therefore fulfilling the ground truth inclusion conditions.

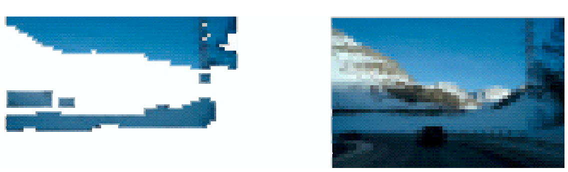

Figure 3: Example of a borderline region. While the image definitely contains blue sky, the considered region is composed of both sky and snow with a blue cast. In that case, this region will not be considered in the blue sky ground truth.

Our region ground truth consist of 10'000 regions manually annotated. Each region belongs to exactly one of the four possible classes: blue sky, vegetation, skin tones or normal (when it does not belong to any of the first three).

4.1. Region Classification Results

By performing an exhaustive search over the whole feature space, the classifier determines the best feature combination to achieve an optimal classification. At the beginning of the classification process, all available features are grouped into a feature vector $\vec{F}$ of dimension *N*. Afterwards, all the features belonging to the optimal combinations form the vector $\vec{F_S}$ of dimension $M \leq N$. Features included in $\vec{F_S}$ are the "surviving" features.

In addition to those surviving features, the other two important outcomes of the algorithm are: the *false positive* and the *false negative* rates. Those two rates and prior probabilities enable us to determine the best threshold (θ) for classification as well as the global success rate. Once the best feature combination has been found, the best thresholds will depend on the priors and on the relative weights associated with false positives and false negatives.

$$\text{False Positive Rate} : \mathrm{p(f(R) > \theta \mid gt_{R,\nu} = 0)}$$

$$\text{False Negative Rate} : \mathrm{p(f(R) \leq \theta \mid gt_{R,\nu} = 1)}$$

where

$$f(R) = \frac{p(A_\nu \mid R)}{p(A_{\neg\nu} \mid R)}$$

$A_{\neg\nu}$ represents all other classes than A_ν, and $gt_{R,\nu}$ is the ground truth value (0 or 1) of the region *R* for the class A_ν.

Actual results take the appearance of an error curve such as the ones shown in Figure 4. A value of θ can

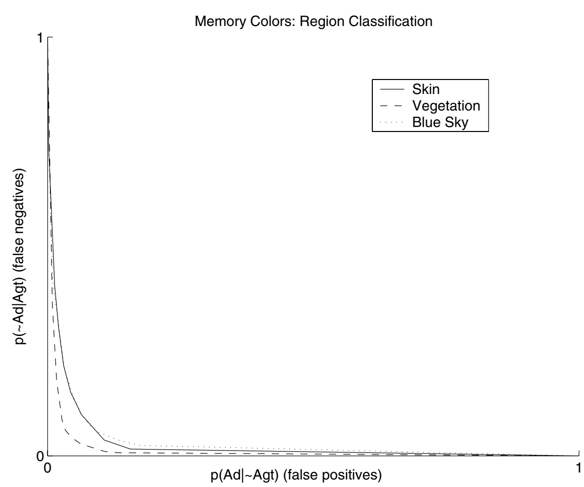

Figure 4: Error curves for memory color region classification. A threshold corresponds to each point of a curve, for which false negative and false positive rates can be obtained. Success rate for the memory colors range from 87% for skin tones to 91% for vegetation.

be associated with every point on this curve. This chosen threshold yields ratios of false positives and false negatives for the classification. Assuming that all classification mistakes carry an equal weight, the optimal value of θ is 0.5, for which the maximum a posteriori detection rule is performed (Eq. 3):

$$\frac{p(R \mid A_\nu)}{p(R \mid A_{\neg\nu})} > \frac{\theta}{1-\theta}\frac{p(A_{\neg\nu})}{p(A_\nu)} \tag{3}$$

5. Image Features and Classification

Region classification outputs $p(A_\nu \mid R)$. Image classification, however, requires $p(S_v \mid I)$, where S_v are the image semantic classes and *I* is the considered image. Assuming that the importance of a region in the image is proportional to the image, one can use an area weighted method to obtain a robust image classification. Equation 4 explains the relationship between region and image classification.

$$\begin{aligned} p(S_v \mid I) = &\sum_{i=1}^{N} area(R_i) \cdot p(A_\nu \mid R_i) \\ &\cdot\ \mathrm{Heaviside}(\mathrm{p}(\mathrm{A}_\nu \mid \mathrm{R_i}) - \theta) \end{aligned} \tag{4}$$

The probability that an image *I* belongs to a certain semantic class S_v equals the area-weighted sum of the probabilities that its regions R_i belong to that class, provided that this probability $p(A_\nu \mid R_i)$ is greater than a certain threshold θ. The $p(A_\nu \mid R_i)$ are the direct outcome of the region classification step. The Heaviside function [13] is used to lessen the impact of a very large region having a small probability in the classification. Using such a method to calculate the $p(S_v \mid I)$ has the advantage of being direct, but it also supposes that the region classification is correct since the $p(A_\nu \mid R_i)$ are directly used for image classification.

The performance of this classifying scheme can be assessed by comparing the $p(S_v \mid I)$ with previous classification achieved using a multivariate gaussian analysis with image features only. Previous classification feature vectors are composed of the image mean *L, a* and *b* values as well as images color and density entropies. Results for both methods are obtained by testing the classification on an image database. Our database consists of 13'500 photographic "real world" images taken by consumers on a variety of digital cameras (from VGA to 3 Megapixels resolution). Those images have later been annotated by photographic experts to establish a ground truth for classification and analysis. The image feature based classifier is trained on 90% of this data set, which is not the case of the region-based one (trained on 90% of the region ground truth). Figures 5 to 7 show a performance comparison of the two methods for images containing vegetation, portraits, and blue sky. In the area of importance, the part of the curve with both low false positive and false negative rates, the region-based classification outperforms the image-based one.

Figure 5: Error Curves for Image Portrait Classification

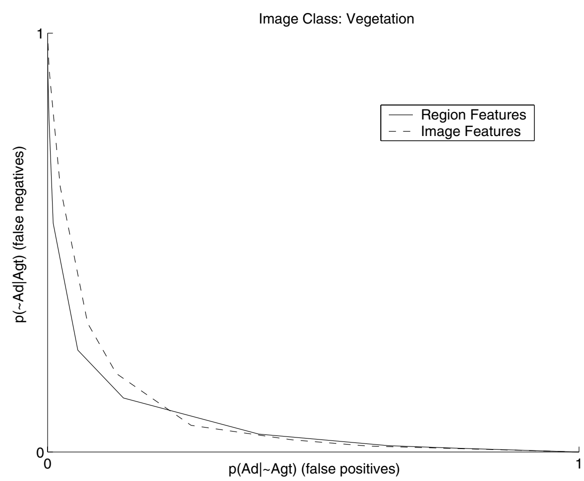

Figure 6: Error Curves for Image Vegetation Classification

6. Improvement of Image Classification

Image classification obtained through region-based methods exhibits robust results that can be used in ACC methods. Classification success rates are higher than the ones derived from image-based methods alone, but since these rates were already high, overall improvements are small. A possible approach that allows for a greater improvement is to use region-features in conjunction with image-features. Such an image analysis on several levels maximizes the global amount of information that can be extracted from an image.

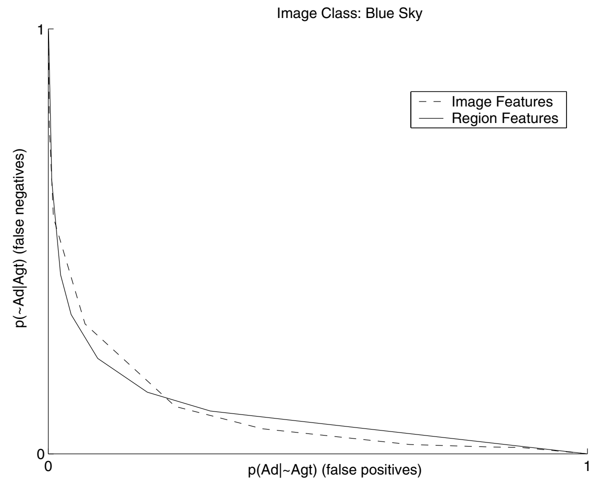

Figure 7: Error Curves for Image Blue Sky Classification

We will illustrate such an approach on two image classes: *pure blue sky* and *bluish color cast*. Those classes were given priority over the memory colors because their performance was much lower and because classifying those images correctly is critical for automatic color correction purposes.

6.1. Pure Blue Sky

Pure blue sky images are usually composed of two regions: a large region of uniform sky and another containing an object (see Fig. 8 for examples). While this class is easy to categorize visually, its high natural blue content can be the source of automatic color correction failures, since it doesn't comply with any of the scene assumptions (max RGB or grayworld).

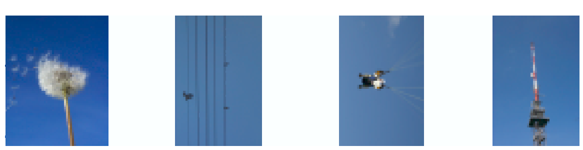

Figure 8: *Typical pure blue sky images.*

The applied ACC will often "correct" the blue sky by turning it neutral. We expect this class to profit from region-based features, considering that the sky region is easily segmented and that its uniform blueness corresponds precisely to blue sky as defined for regions. Pure blue sky images have other well defined properties at the image-level, such as low color and luminance entropy as well as a general high blue content. Using a feature vector composed of 3 original images features (mean b, entropy in ab, entropy in L) and one obtained through region classification (bluesky_p), we were able to detect those images with 98% accuracy. Using image features only, the accuracy is only 90%.

6.2. Bluish Color Cast

Whether due to wrong camera settings or to peculiar lighting conditions, many cameras generate images with color casts. Bluish casts are generally caused by the diffuse light from blue sky and can be present when reflected on a white surface (snow) or in shadowy areas. The challenge of an image classification system for automatic color correction is to detect those images whose spectral components are shifted, while leaving those having a high natural blue content untouched [15,16].

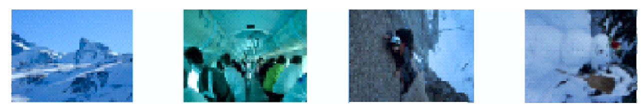

Figure 9: *Examples of images bearing a bluish color cast*

We have found it difficult to obtain a satisfactory performance using only image-level features. For instance, using the mean and variance of color values does not enable sufficient discrimination. From a region-level point of view, a color cast could be approached using correlation measures within the different regions of an image. This being, if we are to use region features in conjunction with image features, we only have the four image memory color probabilities at our disposal, given that all the other features are measured on regions only.

To take full advantage of region analysis, we decided to extend region features to the image level. Possible candidates for this extension are color and texture features. Geometrical features, while important at a region-level, are less relevant when the whole image is taken into account. The approach we chose is to measure the mean and variance of each one of the region features. Let F_I be a feature at the image level, and F_R be the same feature at the region level.

$$\mathrm{mean}(\mathrm{F_I}) = \frac{1}{N}\sum_{i=1}^{N} F_{R_i} \tag{5}$$

$$\mathrm{Var}(\mathrm{F_I}) = \frac{1}{N}\sum_{i=1}^{N} (F_{R_i} - \mathrm{mean}(\mathrm{F_I}))^2 \tag{6}$$

We additionally developed another feature based on the "blue content" of an image, i.e. the percentage of the image covered with blue regions. The parameters used for this feature are based on experimental results.

$$\lambda_{bluish} = \sum_{r \in I} \mathrm{area_r} \cdot \delta_{\mathrm{blue}} \tag{7}$$

$$\delta_{blue} = 1 \text{ if } \mathrm{B_r} > (1.2 \cdot \max(\mathrm{G_r}, \mathrm{R_r})) \tag{8}$$

These newly acquired features were used in conjunction with already existing image features, using the multivariate gaussian analysis approach. However, the results showed that no single threshold permits a satisfactory detection for both false positive and false negative rates. For our image database, using $\theta = 0.5$ minimizes the false positive rates, but at the same time only 42% of the bluish images were detected. A very low threshold ($\theta = 0.01$) would ensure the detection of almost all the bluish images while throwing away 95% of the non bluish ones. This is unfortunately unsatisfactory given the prior probabilities of the bluish/non-bluish classes (see Fig. 10).

Consequently, a single detector is not able to satisfactorily fulfill the classification task. An possible improvement can be achieved by *cascading* simple classifiers. This allows for implementing several classifiers, depending on the end application.

Figure 10 shows the array of possibilities offered by cascading classifiers. In this figure, *B* represents the proportion of bluish images (proportion that is always with respect to the complete database) and *nB* represents the proportion of non-bluish images. The indicative ratio expresses the amount of bluish images with respect to non-bluish ones. Depending on the indented application, one can choose to either maximize the detection of bluish images (at the cost of a lower bluish/non-bluish ratio), or minimize the amount of false negatives, in order to ensure a more reliable correction (at the cost of detecting less bluish images).

The two more interesting classifiers are the first and the last one, as each of them fulfills a specific task. With a 10:1 ratio, the first classifier can be used in a safety detector where a correct detection is critical. The last one, composed of three detectors, was selected to supersede the current bluish classifier. Correctly detecting 75% of the color cast images while still maintaining a 3.8:1 ratio is a significant improvement over the previous classifier, as attested in figure 11.

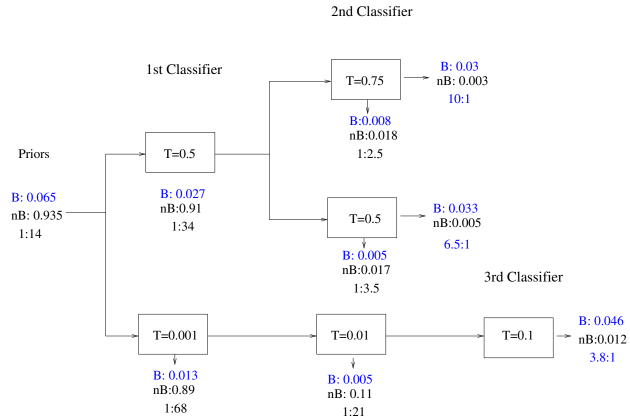

Figure 10: *Complete schema for bluish detection. At each new step, the ground truth is updated to take into account only the pictures that were detected. Based on this new ground truth, a feature selection step is done, which leads to another classifier.* B *and* nB *stand for the proportions of bluish and non-bluish images respectively, the indicative ratios are the amount of bluish images with respect to non bluish ones. On the right of the classifiers are the pictures which are detected as* bluish, *while under the classifier are the rejected ones. This is only an example of cascading classifiers, a method that can well be adapted to diverse applications.*

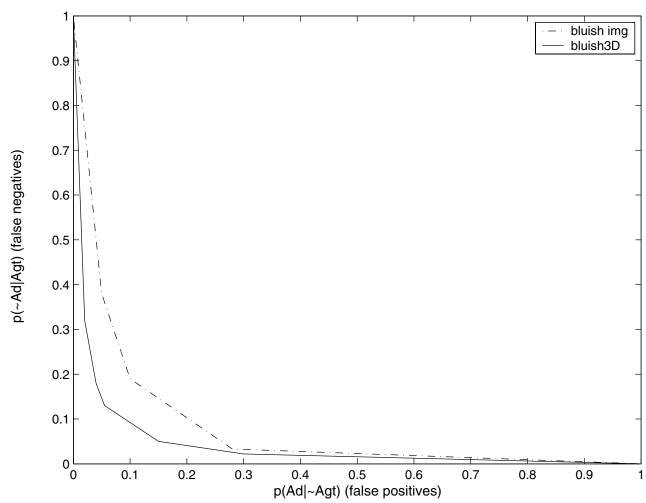

Figure 11: *Error curves for bluish detection based on images features only (dashed curve) and on the 3 level classifier using image and region features (solid curve). One can observe the continuous improvement of the latter over the former.*

7. Conclusion

We have established a framework for image classification based on region information. Results for *memory color* classes show that this approach outperforms the more conventional ones of using image-level information only. The actual cost of the region-based procedure is almost negligible. Given the resolution on which the segmentation is done and the features calculated, the whole process (from segmentation to classification) is computed in real-time on a standard computer.

Coupling region-obtained information with image features can lead to a substantial improvement of image classification, thus allowing for a better automatic color correction. This adaptation of the general framework has been tested for two specific classes: *pure blue sky* and *bluish color cast*. The results show a large increase in correct classification rates. The presented framework can also further be refined and adapted to correctly classify a variety of real-scene classes, offering new alternatives for class-specific image enhancement algorithms.

Acknowledgement

The authors wish to thank Dr. Markus Naef of Gretag Imaging AG, now with Swiss Imaging Technologies AG, who headed the group where most of this work has taken place. This work would not have been possible without the infrastructure of Gretag Imaging AG and the support of Dr. Naef.

8. References

1. M. Schröder and S. Moser, "Automatic Color Correction based on Generic Content Based Image Analysis," *Proc. 9th Color Imaging Conference, Scottsdale*, pp. 41–45, 2001.

2. R. W. G. Hunt, *The Reproduction of Colour*. Fountain Press, 1995.

3. G. Buchsbaum, "A spatial processor model for object colour perception" *Journal of the Franklin Institute*, Vol. 310, 1–26, 1980.

4. E. Land, "The Retinex," *American Scientist*, 52:247–264, 1964.

5. MPEG Requirements Group, "Description of MPEG-7 Content Set," *ISO/IEC/JTC1 /SC29 /WG11 /N2467*.

6. C. Carson and S. Belongie and H. Greenspan and J. Malik, "Blobworld: Image Segmentation Using Expectation-Maximization and Its Application to Image Querying," *IEEE Tr. on Pattern Analysis and Machine Intelligence*, Vol. 24, Number 8, pp. 1026–1038, 2002.

7. W. Ma and B. Manjunath, "EdgeFlow: A Technique for Boundary Detection and Image Segmentation," *IEEE Tr. on Image Processing*, Vol. 9, Number 8, pp. 1375–1388, 2000.

8. A. K. Jain, *Fundamentals of Digital Image Processing,* Prentice-Hall International, 1989.

9. R. M. Haralick, K. Shanmugan, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 3, pp. 610–621, 1973.

10. M. Chamik, "Texture description and classification applied to grass detection on digital photographies," *Technical Report, Gretag Imaging AG*, 2000.

11. P. Smits and A. Annoni, "Toward Specification-Driven Change Detection," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 38, pp. 1484–1488, 2000.

12. C. Fredembach, M. Schröder and S. Süsstrunk "Eigenregions for Image Classification," *submitted to IEEE Tr. on Pattern Analysis and Machine Intelligence*, 2003.

13. Athanasios Papoulis, *Probability, Random Variables, and Stochastic Processes,* McGraw-Hill, 1991.

14. K. Fukunaga, Statistical Pattern Recognition. In C. Chen and L. Pau and P. Wang, editors, *Handbook of Pattern Recognition and Computer Vision,* pages 33–60. World Scientific, 1993.

15. G. Finlayson, S. Hordley, and P. Hubel, "Colour by correlation: a simple, unifying approach to colour constancy," *ICCV'99*, Vol. 2, pp. 835–842, 1999.

16. R. W. G. Hunt, *Measuring Colour*. John Wiley, 1987.

Biography

Clément Fredembach received his MS degree in Communication Systems from the Swiss Federal Institute of Technology in Lausanne (EPFL) in 2003. He did his master's thesis at Gretag Imaging AG and at the Swiss Federal Institute of Technology in Zürich (ETHZ), working on image classification. During his studies, he also worked 6 months at Fuji Photo Film Research in Japan on the management of graininess and sharpness in digital images. Since October 2003, he is a Ph.D. student at the University of East Anglia (UEA), Norwich, UK.