

Eigenregions for Image Classification

Clément Fredembach, Michael Schröder, Sabine Süsstrunk

Abstract

For certain databases and classification tasks, analyzing images based on region features instead of image features results in more accurate classifications. We introduce *eigenregions*, which are geometrical features that encompass area, location and shape properties of an image region, even if the region is spatially incoherent. Eigenregions are calculated using principal component analysis (PCA). On a database of 77'000 different regions obtained through the segmentation of 13'500 real-scene photographic images taken by non-professionals, eigenregions improved the detection of localized image classes by a noticeable amount. Additionally, eigenregions allow us to prove that the largest variance in natural image region geometry is due to its area, and not to shape or position.

Keywords

Eigenregions, Image Classification, Region Analysis, Image Features

I. INTRODUCTION

Many imaging applications, such as automatic color correction (ACC) or content-based image retrieval (CBIR), are based on successful image classification [2, 4, 14, 18]. It is evident that the performance of these algorithms is directly linked to the performance of the classification. Thus, in order to be effective, image classification algorithms need features that well express relevant image properties. These features are often calculated based on the whole image. However, dependent on the database and task, a region-based approach that uses more localized information can improve image classification results [5–7, 13].

In this paper, we introduce *eigenregions*, which are geometrical features that encompass area, location and shape properties of a region. Eigenregions are obtained by analyzing segmented image regions with Principal Component Analysis (PCA). Principal component analysis has already successfully been implemented in image classification for many tasks (see [15, 19] for examples), but usually

on the whole image. As opposed to other geometrical region features [8, 10, 17], eigenregions can be used and result in significant classification improvement even if the image regions are spatially incoherent. They are also visually significant and computationally efficient. Another key result obtained with eigenregions is that for a large dataset of natural images, the largest variance in region geometry is due to the area and not to shape or position.

We tested the performance of eigenregions in an image classification experiment, where the goal was to correctly identify semantic image classes, such as “blue sky,” “skin tone,” and “vegetation.” The data set consisted of 13’500 real-scene photographic images taken by non-professionals. After segmenting the image into regions, we calculate region features, classify the regions, calculate the corresponding image features, and finally classify the images (see Fig. 1). We found that the addition of eigenregions to the feature vector improves region classification results for localized classes, i.e. for region classes usually localized in a specific part of an image, such as sky regions. We also show the relevance of eigenregions by classifying regions using only eigenregions as features, where they achieved a very respectable success rate. Finally, the image classification rate based on region features is significantly better than the classification based on image features only.

This article is organized as follows: Section 2 discusses the image segmentation algorithms used. Section 3 discusses the drawback of existing geometrical features, explains how eigenregions are obtained, and demonstrates their significance. Section 4 illustrates the usefulness of eigenregions for region classification, Section 5 explains how to obtain image classification from region classification, and Section 6 concludes the article.

II. IMAGE SEGMENTATION

To be able to extract region features, images need first to be segmented into meaningful regions. We chose the Dominant Colors in Lab (DC Lab) method, based on the MPEG-7 guidelines [12]. It uses a k-means powered algorithm to segment images according to (CIE) Lab clusters [21]. The k

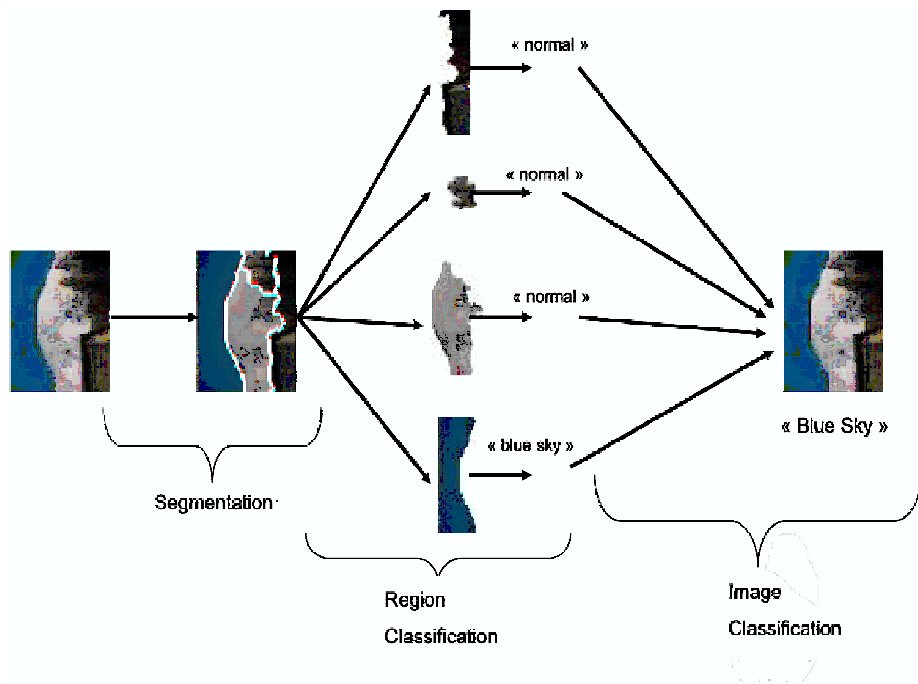


Fig. 1. This figure illustrates the major steps of the classification framework. First, the image is segmented into several regions. Using region-specific features, the regions are classified in several semantic classes. The results from region classification are then used to obtain a robust image classification.



Fig. 2. An image segmented with edgeflow (middle) and DC Lab (right). Contrarily to Edgeflow, DC Lab regions, represented by different gray levels, are not spatially coherent.

parameter, representing the maximal number of possible clusters, is fixed to 8. Arguments for using this algorithm over other available ones [1, 11] include its computational speed (real time processing) and its ease of use, while still providing good quality segmentation. However, DC Lab regions are not spatially coherent (see Fig. 2), which prevents the use of many existing geometrical features (see Section III).

For comparison purposes, we also calculated eigenregions with image regions obtained via edgeflow segmentation [11]. Since edgeflow regions are visually different from DC Lab, we compared eigenregions to find if the geometrical characteristics of the regions are due to the segmentation algorithm or not. We found that the eigenregions are similar in both cases, suggesting that these structural features are

independent of the segmentation. The results presented in the rest of the article are based on DC Lab regions.

III. EIGENREGIONS

One of the major requirements for a robust classification is to have a set of relevant and reliable features that will conveniently describe a region. Color and texture features have received far more attention and usually perform better than geometrical features. It was found that the latter can either be too general or too precise [10]. Some geometrical properties, such as region position, does not in general fulfill continuous or gaussian assumptions required by some classification algorithms [17]. Moment analysis could be used, but it is difficult to determine a priori how many are necessary for a good classification [8, 10]. Another drawback is that if the image segmentation algorithm results in regions that are not spatially coherent, the perimeter or radii of a region cannot be used as parameters [3]. Other geometrical properties could be utilized in this case, such as the compactness or the positions of a region in the image. These features are, however, too limited to improve an already high classification rate on our image database [5], whereas the proposed eigenregions do significantly improve the image classification (see Section IV).

Eigenregions are obtained by calculating the principal components of the region locations. The segmented images have a size of 64×48 (3072) pixels, resulting in a dimensionality too large to undertake a PCA. We therefore started by downsampling the regions to a 5×5 area, such as shown in Fig. 3. Since the area of interest is the coverage of a pixel by a region, the downsampling procedure determines which percentage of a new “pixel” was originally covered by the region. Let R be a region, I the image to which the region belongs and ρ a pixel. Consider the downsampling procedure:

$$\begin{aligned} \forall \rho \in I : I(\rho) &= 1 \text{ if } \rho \in R, 0 \text{ otherwise} \\ I_{\text{down}}(i) &= \frac{1}{\beta} \sum_{j=1}^{\beta} I(\beta(i-1) + j) \end{aligned}$$

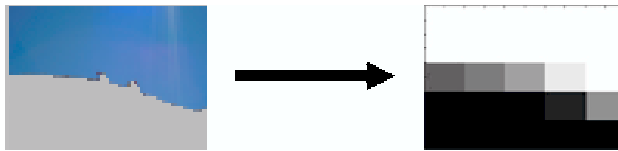


Fig. 3. Downsampling of a region (left) to a 5×5 image. The grey values of the down-sampled image display proportional coverage, from 100% (white) to 0% (black)

where $\beta = \lfloor \frac{3072}{25} \rfloor$ is the downsampling factor and I_{down} is the downsampled image that will be used in the principal component analysis. Note that we investigated a number of downsampling factors, ranging from 3×3 to the entire image. Downsampling of 5×5 and higher lead to the same visual and mathematical results. We present the results for a 5×5 downsampling, which is the most computationally efficient implementation.

The eigenregions are then calculated on a set of 77'776 (N) regions obtained via the segmentation of 13'500 real-scene photographic images taken by non-professionals, i.e. consumer images.

The general PCA analysis is as follows: let X be the data matrix (N vectors of length 25) and let \bar{X} (1×25) be the mean over all observations of X . Y is defined as X with \bar{X} subtracted from each of its columns. The covariance matrix is $C = Y \cdot Y^T$, and the single value decomposition of C yields $C = V \cdot \Lambda \cdot V^T$, where V is the eigenvector matrix and Λ is the diagonal matrix containing the eigenvalues.

Using this method, the first eigenfeature (the one corresponding to the largest eigenvalue) will be in the direction of the largest variance. Of importance to the classification is the fact that the eigenvectors are orthogonal. This implies that they can all be used as features in the selection process since they are independent.

Fig. 4 shows the 25 eigenregions obtained using this method. The reconstruction rates can be observed in Fig. 5. From a visual point of view, a semantical meaning can be given to the first 5 eigenregions. The centered vertical and horizontal split (eigenregions 2 and 3) are due for the most part to landscape photographs having a horizon line. Eigenregions 4 and 5 represent the distinction between an object and a background. Since those distinctions strongly depend on the nature of the

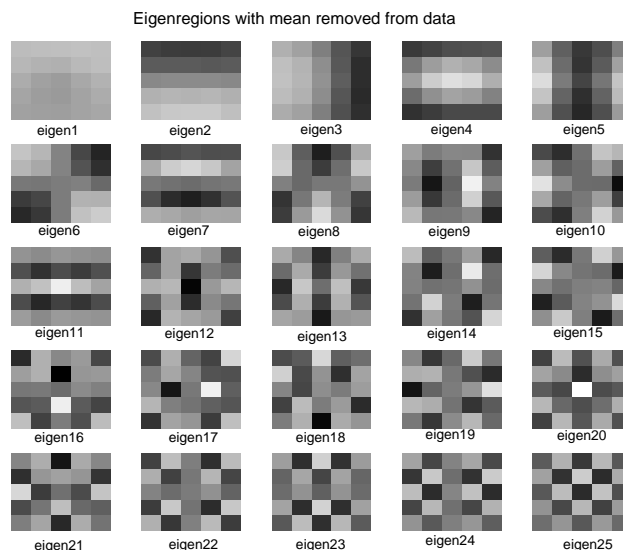


Fig. 4. DC Lab eigenregions. Coverage values range from -1 (black) to +1 (white). Those regions express the most common structure of segmented images. Relevant semantic regions include top/bottom (eigen 2), left/right (eigen 3) and various center/surround interactions (eigen 4 and 5).

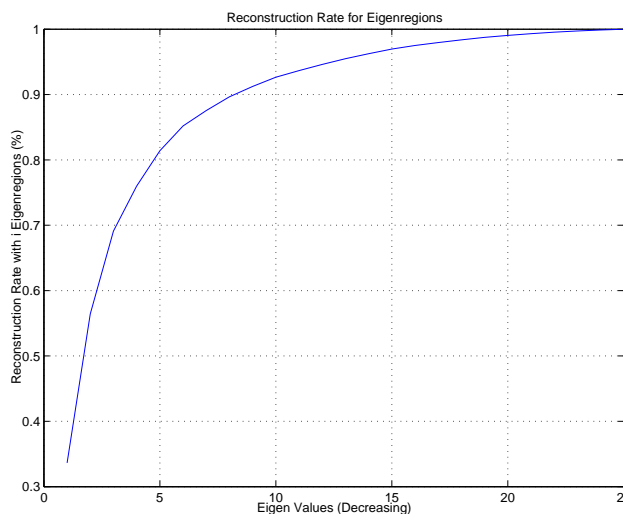


Fig. 5. Eigenregion Reconstruction Rates.

photographed object (shape, color, sharpness, etc.), it seems normal that those regions come after the landscape-type ones, as the latter are much more common in real-scene photography. The remaining eigenregions cannot easily be given semantic meaning. One explanation is that DC Lab also outputs a lot of small sized regions that can appear noisy when displayed as such.

Interestingly, the first eigenregion appears to be a “mean.” Recall that the mean of all regions was removed prior to the principal component calculation. Thus, the largest variation of the regions is in the direction of the mean, even though it is removed. By letting $C = X \cdot X^T$ instead of $Y \cdot Y^T$, we can

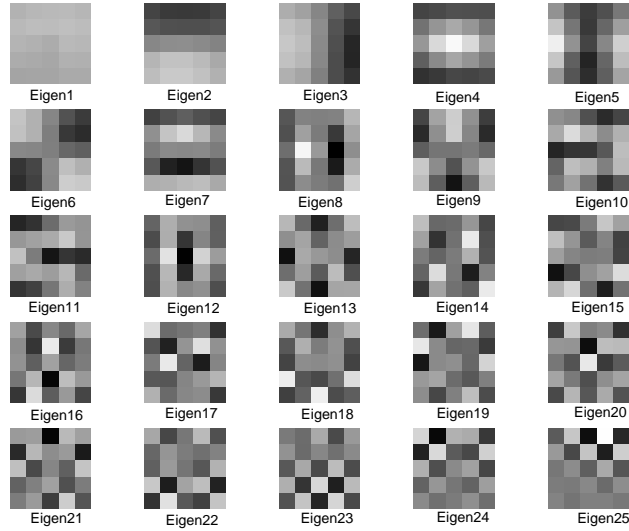


Fig. 6. DC Lab “eigenregions” *without* the mean removed from each column of X . The overall structure of the regions does not significantly change, indicating that the greatest variance of the data is in the direction of the mean.

apply a singular value decomposition instead of a standard principal component analysis to the data set. The results of this method are illustrated in Fig. 6.

Comparing Fig. 4 and Fig. 6, we can observe that removing the mean from the data does not alter the obtained eigenregions. Given that the union of an image’s regions is the image, we can compute the average region size by dividing the total number of images by the total number of regions. In our case, we have 77’776 regions from 13’512 images, yielding a value of 0.17. This value is not small enough to justify the fact that removing it bears no influence on the final result.

We can prove that for our experimental set-up, the greatest variance of a region’s structure is in the same direction as the mean. Let X, \bar{X}, Y and N be as previously defined. From the singular value decomposition, we have:

$$X \cdot X^T = U \cdot D \cdot U^T = \sum d_{ii} \cdot u_i \cdot u_i^T \quad (1)$$

$$Y \cdot Y^T = V \cdot \Lambda \cdot V^T = \sum \lambda_{ii} \cdot v_i \cdot v_i^T \quad (2)$$

Expanding eq. 2 yields:

$$Y \cdot Y^T = (X - \bar{X}e^T) \cdot (X - \bar{X}e^T)^T \quad (3)$$

$$= XX^T - Xe\bar{X}^T - X^Te^T\bar{X} - N\bar{X}\bar{X}^T \quad (4)$$

$$= X \cdot X^T - N \cdot \bar{X} \cdot \bar{X}^T \quad (5)$$

where e is a vector of size 1×25 containing only 1's. If the first eigenvector is in the same direction as the mean, then:

$$\bar{X} \cdot \bar{X}^T = c \cdot u_1 \cdot u_1^T \quad (6)$$

where c is the norm of \bar{X} . If removing the mean does not change the direction of the first eigenvector, the other eigenvectors will not be altered either. The largest eigenvalue will decrease by $N \cdot c$ and the other eigenvalues will stay the same. Using eq. 5, we have:

$$Y \cdot Y^T = X \cdot X^T - N \cdot \bar{X} \cdot \bar{X}^T \quad (7)$$

$$\sum \lambda_{ii} \cdot v_i \cdot v_i^T = \sum d_{ii} \cdot u_i \cdot u_i^T - N \cdot c \cdot u_1 \cdot u_1^T \quad (8)$$

Which yields to, if the hypothesis is correct:

$$\lambda_{11} = d_{11} - c \cdot N \quad (9)$$

$$\forall i \geq 2, \lambda_{ii} = d_{ii} \quad (10)$$

Calculating the principal components in both cases enables us to compare the eigenvalues and experimentally verify that it corresponds to the theory. Table I regroups the eigenvalues obtained with and without the mean removed. The norm of the mean vector being $c = 0.393$, it is easy to verify that $\lambda_{11} = d_{11} - cN$. The table entries also show that all the other eigenvalues are the same, within

TABLE I

EIGENVALUES FOR THE TWO DIFFERENT VARIANTS, WITHOUT $(X \cdot X^T)$ AND WITH $(Y \cdot Y^T)$ THE MEAN REMOVED.
FOR $i > 9$, THE DIFFERENCE FOR λ_i IS ZERO.

$\Lambda (\times)$	$X \cdot X^T$	$Y \cdot Y^T$	Difference
λ_1	79'835	49'225	30'610
λ_2	24'075	24'032	43
λ_3	11'108	11'085	23
λ_4	8'519	8'519	0
λ_5	6'324	6'310	14
λ_6	4'680	4'680	0
λ_7	3'588	3'587	1
λ_8	2'794	2'794	0
λ_9	2'714	2'709	5

numerical approximation.

A. Area as the First Eigenvector

We have shown that the greatest variance in the data is in the direction of the mean. Our hypothesis is that this variance is due to the relative area of a region. Eigenregions can be seen as features encompassing shape, location, and size properties. It is, however, unlikely that either shape or location are involved in this variance since the first eigenregion has no defined shape nor location.

We verified our assumption that the first eigenvector is an indicator of area, and not shape or location by separating the regions according to their area. We define the relative area of a region a_{R_i} as $a_{R_i} = \|x_i\|$, where R_i is a region of the image I , and x_i the i^{th} column of X that corresponds to R_i . We then create i histogram bins A_i such that:

$$A_i = [a_{R_j} | b_i \leq a_{R_j} \leq c_i], \forall R_j \quad (11)$$

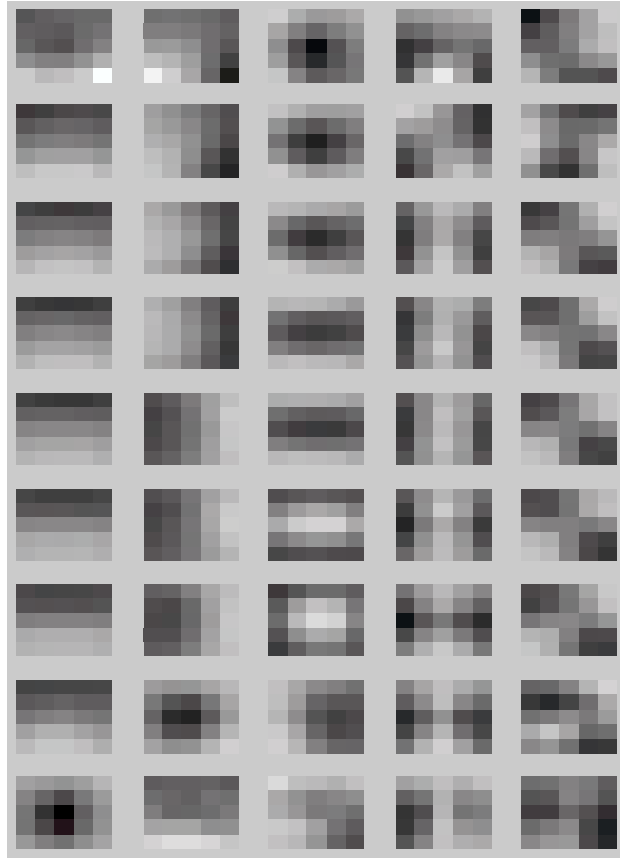


Fig. 7. First five eigenregions using PCA on 9 area bins. Bins from smallest to largest area are ordered from top to bottom. The “mean-like” eigenregion does not appear anymore and the others are similar to those obtained on all regions (see Fig. 4 and 6)

Using the area distribution of the regions, we obtained 9 bins. Expressed in terms of the regions relative area, they are: 0-0.05, 0.05-0.1, 0.1-0.15, 0.15-0.2, 0.2-0.3, 0.3-0.4, 0.4-0.6, 0.6-0.8 and 0.8-1.

Y_{A_i} denotes the data matrix of regions that are in the area bin A_i with their mean removed. For each A_i , we have $C_{A_i} = Y_{A_i} \cdot Y_{A_i}^T$ and obtain V_{A_i} and Λ_{A_i} with the principal component analysis of C_{A_i} .

Fig. 7 shows the first 5 eigenregions for all 9 area bins. The overall structure of those eigenregions is very similar to those obtained over all regions, with the exception of the first eigenregion of X . Consequently, removing a possible variation of the area also removes the mean-like eigenregion, thereby indicating that the largest variance of a region structure is in fact the region area. We further verified this by computing the inner product of V_{A_i} with the first eigenvector of X , and found them to be almost orthogonal.

As a result, we can conclude that the largest variance of natural image region structure is the region’s

area. This result is, of course, dependent on our image data set. However, this does not mean that using eigenregion features are not usable for image classification of natural images, as is shown in the next section, but that the inherent variation of natural image structures is principally due to the regions area.

IV. EIGENREGIONS IN REGION CLASSIFICATION

We tested the usefulness of eigenregions in an image classification task. The primary goal was to improve the performance of class specific automatic color correction (ACC) [16]. The considered classes *blue sky*, *skin tones*, and *vegetation* are critical for the performance of ACC. The classification algorithm used in this method is a multivariate gaussian analysis based on the maximum a posteriori rule [8], which was chosen for its robustness and scalability. This is a supervised classification algorithm, which implies that the class labels have to be defined beforehand.

A set of features is given to the classifier that then retains the relevant ones. If a feature does not actively help to differentiate two classes, it will not be selected. The “surviving” features therefore provide the best classification performance (for more details about the classification framework, see [6]).

Besides eigenregions, color and texture features were also calculated for each region. Color features include mean and standard deviation of R, G, B, L, a, and b. As texture features, we chose 7 of the Haralick features [9] that are based on a Gray-Level Co-occurrence Matrix (GLCM). Energy, Contrast, Homogeneity, Inverse Different Moment and Entropy have been defined by Haralick and two, Cluster Shade and Cluster Prominence, have been added later by Smits [20]. In addition, we also use the mean amplitude spectrum of the Fourier Transform of a region’s luminance, which represents a texture intensity measure.

We assessed the discriminative power of eigenregions by classifying regions using only eigenregions and using eigenregions in conjunction with the other acquired features. In the first case, the input to the algorithm is the approximation of the regions using the 10 first eigenregions. The classifier then

TABLE II

CLASSIFICATION SUCCESS RATES (AVERAGE OF CORRECT POSITIVE AND CORRECT NEGATIVE RATES) USING ONLY EIGENREGIONS AS FEATURES. DETECTION RATES ARE VERY GOOD FOR LOCALIZED CLASSES (SKIN, BLUE SKY), BUT NOT FOR NON-LOCALIZED ONES (VEGETATION). THE 2ND COLUMN (SUCCESS RATE) IS OBTAINED USING ALL EIGENREGIONS; THE 4TH ONE (BEST FEATURE RATE) IS THE MAXIMUM RATE USING A SINGLE EIGENREGION ONLY.

Class	Success Rate	Best Eigenregion	Best feature rate
Skin	0.864	eigen4 (center)	0.817
Vegetation	0.667	eigen2 (top/bottom)	0.58
Blue sky	0.832	eigen2 (top/bottom)	0.805

TABLE III

REGIONS CLASSIFICATION RATES (AVERAGE OF CORRECT POSITIVES AND CORRECT NEGATIVES RATES) WITH AND WITHOUT THE ADDITION OF EIGENREGIONS AS REGION FEATURES.

Class	Rate w/ out eigenregions	Rate w. eigenregions
Vegetation	0.924	0.924
Blue Sky	0.89	0.918
Skin	0.905	0.924

determines which of those features are useful for classification, as well as the resulting false positive and false negative rates. Numerical results are listed in Table II. We found that classification rates for skin and blue sky regions are better than for vegetation regions. This can be explained by the fact that vegetation is not a “localized” class. Vegetation can occur anywhere in the image, therefore rendering geometrical features less meaningful, whereas skin and sky regions will have a tendency to be located in a specific part of an image (top and center, respectively).

For the second experiment, we used a feature vector composed of the color, texture and geometrical features [5,6] to which the 10 eigenregion features were added. Table III shows that despite the already solid results obtained, the use of eigenregions permitted a noticeable improvement of the classification of sky and skin regions. The results show that even though the reconstruction rate of the eigenregions is low (see Fig. 5), their potential for discriminating between different classes is still relevant.

TABLE IV

IMAGES CLASSIFICATION RATES (AVERAGE OF CORRECT POSITIVES AND CORRECT NEGATIVES RATES) WITH *image semantic classes*. REGION-BASED CLASSIFICATION LEADS TO A BETTER OVERALL IMAGE CLASSIFICATION.

Class	Image Feature Classification	Region Feature Classification
Vegetation	0.81	0.89
Blue Sky	0.8	0.91
Skin	0.70	0.85

V. FROM REGION TO IMAGE CLASSIFICATION

Let A_ν , a semantic region class ("bluesky", "vegetation" or "skin tones"), R a region, S_ν a semantic *image* class and I the image. If we consider the output of the region classification to be $p(A_\nu|R)$, then we have [5, 6]:

$$p(S_\nu | I) = \sum_{i=1}^N \text{area}(R_i) \cdot p(A_\nu | R_i) \cdot \text{Heaviside}(p(A_\nu | R_i) - \theta) \quad (12)$$

Where θ is the chosen threshold of the multivariate gaussian.

The performance of this region-based image classification can be assessed by comparing $p(S_\nu|I)$ with the results of a previous classification on the same image set obtained by using image features only [5, 6]. The results are listed in Table IV.

VI. CONCLUSION

We have developed a geometrical feature, using Principal Component Analysis, that combines region area, shape, and position information. Using the eigenregion features in image classification for automatic color correction resulted in an improved classification for images containing localized regions.

Based on our data-base of 13'500 real-scene photographs taken by non-professionals, a key result obtained from eigenregions is that the largest variance in region geometry is due to the area and not

to the shape or position. This implies that a PCA-based method might not be the most adapted for region reconstruction purposes, since the variance is more due to randomness than to a defined structure.

Notwithstanding, using eigenregions in conjunction with color and texture features allows for a noticeable improvement of image classification rates. Despite a low reconstruction rate and random structure, eigenregions can still be used in image classification because they are discriminative enough for a variety of critical image classes.

Eigenregions being database dependent, other results may be obtained. It is possible that for image databases containing more structured scenes, eigenregions may perform even better than in our example.

ACKNOWLEDGMENTS

The authors wish to thank Dr. Markus Naef and the IIT group of Gretag Imaging AG, Prof. Bernt Schiele, ETHZ, and Prof. Emre Telatar, EPFL.

REFERENCES

- [1] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Image Segmentation Using Expectation-Maximization and Its Application to Image Querying," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 24, pp. 1026–1038, 2002.
- [2] S.-F. Chang, "Content-Based Indexing and Retrieval of Visual Information," *IEEE Signal Processing Magazine*, Vol. 14, pp. 45–48, 1997.
- [3] Y.Q. Chen, "Novel Techniques for Image Texture Classification," *PhD Thesis, University of Southampton*, 1995.
- [4] C. Faloutsos et al., "Efficient and Effective Querying by Image Content," *Journal of Intelligent Information Systems*, Vol. 3, pp. 231–262, 1994.
- [5] C. Fredembach, "Salient Regions for Automatic Color Correction and Image Classification," MSc Thesis, EPFL, 2003.
- [6] C. Fredembach, M. Schröder, and S. Süsstrunk, "Region-Based Image Classification for Automatic Color Correction," *Proc. IS&T 11th Color Imaging Conference*, pp. 59–65, 2003.

- [7] C.-S. Fuh, S.-W. Cho, K. Essig “Hierarchical Color Image Region Segmentation for Content-Based Image Retrieval System,” *IEEE Trans. on Image Processing*, Vol. 9, pp. 156–162, 2000.
- [8] K. Fukunaga, Statistical Pattern Recognition. In C. Chen, L. Pau and P. Wang, editors, *Handbook of Pattern Recognition and Computer Vision*, pp. 33–60. World Scientific, 1993.
- [9] R. M. Haralick, K. Shanmugan and I. Dinstein, “Textural features for image classification,” *IEEE Trans. on Systems, Man, and Cybernetics*, Vol. 3, pp. 610–621, 1973.
- [10] A. K. Jain, *Fundamentals of Digital Image Processing*, Prentice-Hall International, 1989.
- [11] W. Ma and B. Manjunath, “EdgeFlow: A Technique for Boundary Detection and Image Segmentation,” *IEEE Trans. on Image Processing*, Vol. 9, pp. 1375–1388, 2000.
- [12] MPEG Requirements Group, “Description of MPEG-7 Content Set,” *ISO/IEC/JTC1/SC29/WG11/N2467*.
- [13] H. Palus, “Region-Based Colour Image Segmentation: Control Parameters and Evaluation Functions,” *Proc. IS&T First European Conference on Color in Graphics, Imaging, and Vision*, pp. 259–262, 2002.
- [14] T. Pun and D. Squire, “Statistical structuring of pictorial databases for content-based image retrieval systems,” *Pattern Recognition Letters*, Vol. 17, pp. 1299–1310, 1996.
- [15] R. Ramamoorthi, “Analytic PCA Construction for Theoretical Analysis of Lighting Variability in Images of a Lambertian Object,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 24, pp. 1322–1333, 2002.
- [16] M. Schröder and S. Moser, “Automatic Color Correction based on Generic Content Based Image Analysis,” *Proc. IS&T 9th Color Imaging Conference*, pp. 41–45, 2001.
- [17] M. Schröder, “Stochastic Modeling of Image Content in Remote Sensing Image Archives,” PhD Thesis, ETHZ, 2000.
- [18] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, “Content-Based Image Retrieval at the End of the Early Years,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 22, pp. 1349–1380, 2000.
- [19] M. Turk and A. Pentland, “Eigenfaces for Recognition,” *Journal of Cognitive Neuroscience*, Vol. 3, pp. 71–96, 1991.
- [20] P. Smits and A. Annoni, “Toward Specification-Driven Change Detection,” *IEEE Trans. on Geoscience and Remote Sensing*, Vol. 38, pp. 1484–1488, 2000.
- [21] G. Wyszecki and W. Stiles, *Color Science: Concept and Methods, Quantitative Data and Formulae*. Wiley, 1982.