PERGAMON

Pattern Recognition 35 (2002) 1759–1769

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

A hierarchical classification strategy for digital documents

R. Schettini[a,*], C. Brambilla[b], G. Ciocca[a], A. Valsasna[a], M. De Ponti[c]

[a] *ITIM, Consiglio Nazionale delle Ricerche, Via Ampere 56, 20131 Milano, Italy*
[b] *IAMI, Consiglio Nazionale delle Ricerche, Via Ampere 56, 20131 Milano, Italy*
[c] *ST Microelectronics TPA Group, Printer Division, Via Olivetti 2, 20041 Agrate Brianza, Italy*

Received 5 July 2001

Abstract

The effective classification of image contents allows us to adopt strategies that can meet the increasing demand for quality, speed and ease of use in imaging applications. We report here on our experience in the use of CART classifiers for the classification of images indexed by low-level perceptual features such as color, texture, and shape. The problem addressed is the complex matter of distinguishing among photographs, graphics, texts, and compound documents. To cope with the great variety of compound documents we have designed a hierarchical classification strategy which first classifies images as compound or non-compound by verifying the homogeneity of the sub-images in terms of low-level features. Non-compound images are then classified as photographs, graphics, or texts. The results are reported and discussed.

Keywords: CART methodology; Compound documents; Graphics; Image classification; Low-level features; Photographs; Texts

1. Introduction

Internet and the web have become the key enablers of the revolution in the management of the digital imaging workflow, in both the domestic and the working environment. This emerging workflow structure depends upon the effective realization of three fundamental steps: image acquisition (the digital way in); image reuse (digital recirculation) and cross-device image rendering (the digital way out). We believe that content-based image classification to be mandatory for the accurate description and use of digitized images. The effective classification of image (or sub-image) contents allows us to adopt the most appropriate strategies for image enhancement, color processing, compression, and rendering to meet the increasing demand for image quality, speed, and ease of use. This is particularly the case of cross-media color reproduction. Recognizing the class to which a processed image is likely to belong would allow the Color Management System to process the image according to specific strategies for text, graphics and photo's images, to automatically perform color adjustments, or to obtain a more pleasant (or preferred) color reproduction without requiring user interaction. Digital document classification also allows optimization of image data size, providing the best compression and data representation (text data, for instance, are best represented with high-resolution, low bit-depth, lossless-compressed images, while pictorial information requires lower-resolution, high bit-depth, and can tolerate compression with some visual loss). This avoids quality trade-offs, and provides better interoperability, by adding relevant side information to the images to be acquired, broadcasted, or rendered.

In this paper we address the problem of image classification using low-level features, such as color, edge distribution, and image composition. The hierarchical strategy we propose has been designed to address the high-level problem of distinguishing among photographs, graphics,

* Corresponding author. Tel.: +39-02-706-43288; fax: +39-02-706-43292.
E-mail address: schettini@itim.mi.cnr.it (R. Schettini).

0031-3203/02/$22.00
PII: S0031-3203(01)00168-6

texts, and compound documents. It is based on the use of tree classifiers built with the CART methodology [1].

Due to the great variability of the images to be classified, we first built and validated a "classifier engine" for the classification of photographs, graphics, and texts, and then used that to derive a compound vs. non-compound classifier. The "classifier engine" was obtained by generating multiple tree classifiers and by combining these through a majority vote.

The low-level features we used to describe the images were derived from a general purpose image indexing library, and, in designing such a library we have considered perceptual similarity (the feature distance between two images are large only if the images are not "similar"), efficiency (the features can be rapidly computed) and economy (their dimensions must be small in order not to affect classification efficiency).

There have been very few efforts to automate the classification of digital color documents to date. Athitsos and Swain [2], and Gever et al. [3] have proposed automated systems for distinguishing photographs from graphics on the World Wide Web. Schettini et al. [4,5] have defined a method for distinguishing photographs from graphics and texts purely on the basis of a rather high number (389) of low-level features.

Szummer and Picard [6] have designed algorithms for indoor/outdoor image classification. They have systematically studied color histograms computed in the Otha color space, multiresolution autoregressive model parameters, and coefficients of shift invariant discrete cosine transform computed on the whole image and on sub-blocks. They have reported a correct classification rate of 90.3% using color histograms and multiresolution autoregressive model parameters on a database of over 1300 consumer image provided by Kodak.

Vailaya et al. [7] have considered the hierarchical classification of vacation images using binary Baysian classifiers: at the highest level images are classified as indoor or outdoor; outdoor images are further classified as city or landscape, and finally, landscape images are classified as sunset, forest, or mountain scenes.

The present paper is organized as follows. The features used to index the images are outlined in Section 2. Section 3 gives a synthetic description of the CART methodology. The hierarchical classification strategy we propose is described in Section 4, while Section 5 reports the results of our experiments on a database of over 35,000 images collected from various sources, such as images downloaded from the web, or acquired by scanner, and bitmap versions of electronic pages. Section 6 presents our conclusions.

2. CART classifiers

Generally speaking, CART classifiers are trees constructed by recursively partitioning the predictor space, each split being based on conditions related to the predictor values. The process is binary: the predictor space and each subset of it are split exactly in two (see Fig. 1). In tree terminology the subsets are called nodes: the predictor space is the root node, terminal subsets are terminal nodes, and so on. The construction process is based on training sets of cases for which class $j \in \{1,\ldots,J\}$ is known. In our problem the predictors are the features indexing the images, and the training sets are composed of images for which the semantic class is known. Once a tree has been constructed, a class is assigned to each of the terminal nodes, and it is this that makes the tree a classifier: when a new case is processed by the tree, its predicted class is the one associated with the terminal node into which the case finally moves on the basis of its predictor values.

The class assigned to each terminal node t is the one that minimizes the estimated misclassification cost within the node, which is given by

$$r(t) = \min_i \sum_j c(i \mid j)\, p(j \mid t), \tag{1}$$

where $c(i \mid j)$ is the cost of misclassifying a class j case as a class i case, and $p(j \mid t)$ is the estimated probability of the class j in node t.

The performance of a tree is evaluated in terms of its overall misclassification probability, or misclassification cost, which, if T denotes the tree, is estimated by

$$R(T) = \sum_{t\ terminal\ node} r(t)\, p(t), \tag{2}$$

where $p(t)$ is the estimated probability of a case being assigned to node t.

The critical problems of the splitting process are essentially two: how to identify candidate splits, and how to define the goodness of the splits. Candidate splits are generated by a set of admissible questions regarding the values of the predictors. These questions differ according to the nature of the predictors themselves. In the case of a category predictor, for example, all splits that assign the values of the predictor to two different groups are considered candidates. At each step of the process, all the predictors are searched one by one, and the best split, in the sense defined below, is found for each predictor. The best splits are then compared, and the best of these selected.

The idea central to the goodness of splits is that of selecting the splits so that the data in the descendant nodes are purer than the data in the original ones. To do so, different functions of impurity of the nodes, $i(t)$, are introduced, and the decrease in value of the chosen function produced by a split is taken as a measure of the goodness of the split itself. For a node t and its descendant nodes t_l and t_r, this is

$$\Delta i(s,t) = i(t) - p_l i(t_l) - p_r i(t_r), \tag{3}$$

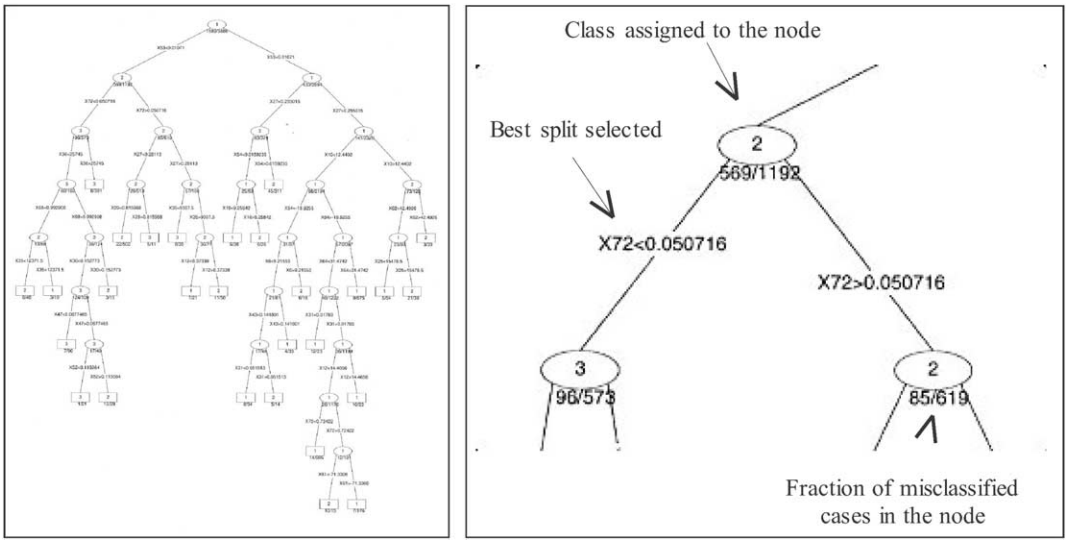

Fig. 1. An example of tree classifier and node details.

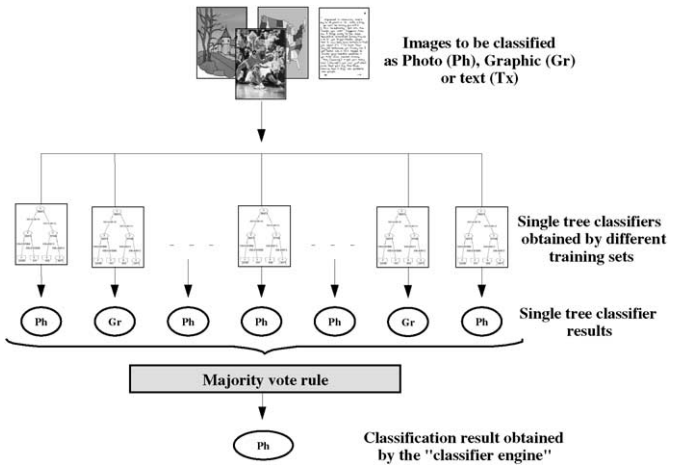

Fig. 2. Our "classifier engine".

where p_l and p_r are the proportions of the cases of t falling in t_l and t_r, respectively, according to split s.

The most commonly used function of node impurity is the Gini diversity index

$$i(t) = \sum_{i \neq j} p(i \mid t)\, p(j \mid t) = 1 - \sum_{j} p^2(j \mid t), \tag{4}$$

which can be interpreted in terms of variances of Bernoulli variables. If, for each class j, we consider the random variable Y_j , which is 1 (success) if a case of t belongs to class j and 0 (failure) otherwise, it can be modeled as a Bernoulli variable with probability of success $p(j \mid t)$, and in this case the quantity

$$1 - \sum_{j} p^2(j \mid t) \tag{5}$$

is the sum of the estimated variances of such variables.

The goodness of a split can also be evaluated by the reduction in deviance [8] produced by the split. For a node t, the deviance is defined as

$$D(t) = -2 \sum_{j} n_{tj} \log p(j \mid t), \tag{6}$$

where n_{tj} is the frequency of class j cases in node t. The underlying idea is that the n_{tj} cases of the training set belonging to a node t constitute a random sample from the multinomial distribution specified by $p(j \mid t)$. $D(t)$ is proportional to the entropy function of the variable class within the node. Generally speaking, the deviance is a function quantifying the discrepancy between a fit and the data [9].

Fig. 3. Skin regions detection.

Since the process goes on until some stopping rule is satisfied, the trees can be very big and overfit the data. One of the major innovations of the CART methodology is the possibility of pruning process based on the idea of finding a trade-off between the complexity and the accuracy of the trees. For a tree T, the pruning process generates a sequence $\{T_l\}_{l\in\{1,\ldots,L\}}$ of subtrees decreasing in size, each of which is the best, in its size range, according to a cost-complexity measure defined as

$$R_\alpha(T) = R(T) + \alpha|T|, \tag{7}$$

where $|T|$ is the number of terminal nodes, and α ($\geqslant 0$) is the unit cost of complexity per terminal node. The subtrees are evaluated in terms of their overall misclassification probability, or misclassification cost, on the basis of test sets, or by means of cross-validation, and the best subtree is then selected. Choosing the tree to use for classification in this way reduces the strong dependence of the classification itself on the training data, and provides a more parsimonious classifier.

Recent work [10] has shown that the accuracy of CART classifiers can be improved by perturbing and combining methods. This means generating multiple versions of a classifier by perturbing the training set, or the construction method, and then combining these multiple versions to produce a single classifier. The most natural way to combine a different classifiers is by majority vote. We have called a classifier obtained by perturbing and combining a "classifier engine"; Fig. 2 shows the one we have used.

3. Image description using pictorial features

The following features were used to index the images:

Color distribution, described in terms of the moments of inertia (i.e. the mean, variance, skewness and kurtosis) of the distribution of hue, saturation and value [11].

Color information: The feature entries are: (i) the percentage of "colored" pixels of the image, that is the pixels having a saturation value higher a given threshold, and (ii) the number of distinct colors present in the image.

Edge distribution, the statistical information on image edges extracted by Canny's algorithm: (i) the percentages of low, medium, and high contrast edge pixels in the image; (ii) the parametric thresholds on the gradient strength corresponding to medium and high contrast edges; (iii) the number of connected regions identified by closed high contrast contours; (iv) the percentage of medium contrast edge pixels connected to high contrast edges [12].

Wavelets: Multiresolution wavelet analysis provides representations of the image data in which both spatial and frequency information are present. It has recently been used in content-based retrieval for similarity retrieval and target search, e.g. Ref. [13]. In multiresolution wavelet analysis we have four bands for each level of resolution: a low-pass filtered version of the processed image, and three bands of details. Each band corresponds to a coefficient matrix one-fourth the size of the processed image. In our procedure the features are extracted from the luminance image using a three-step Daubechies multiresolution wavelet expansion producing 10 sub-bands [14]. Two energy features, the mean and variance, are then computed for each subband. These features provide a concise description of the image's texture and shape.

Texture: estimate of texture features are based on the neighborhood graytone difference matrix (NGTDM), i.e. coarseness, contrast, busyness, complexity, and strength [15,16].

Image composition: The HSV color space was partitioned into 11 color zones corresponding to basic color names. This partitioning was defined and validated empirically by different groups of examiners. The spatial composition of the color regions identified by the process of quantization was described in terms of [17]:

(i) fragmentation (the number of color regions);
(ii) distribution of the color regions with respect to the center of the image;
(iii) distribution of the color regions with respect to the x-axis, and with respect to the y-axis.

Skin's pixels, the percentage of skin pixels. We used a statistical skin color detector (see Fig. 3) based on the r, g chromaticities of the pixel; a training set of 30,000 color

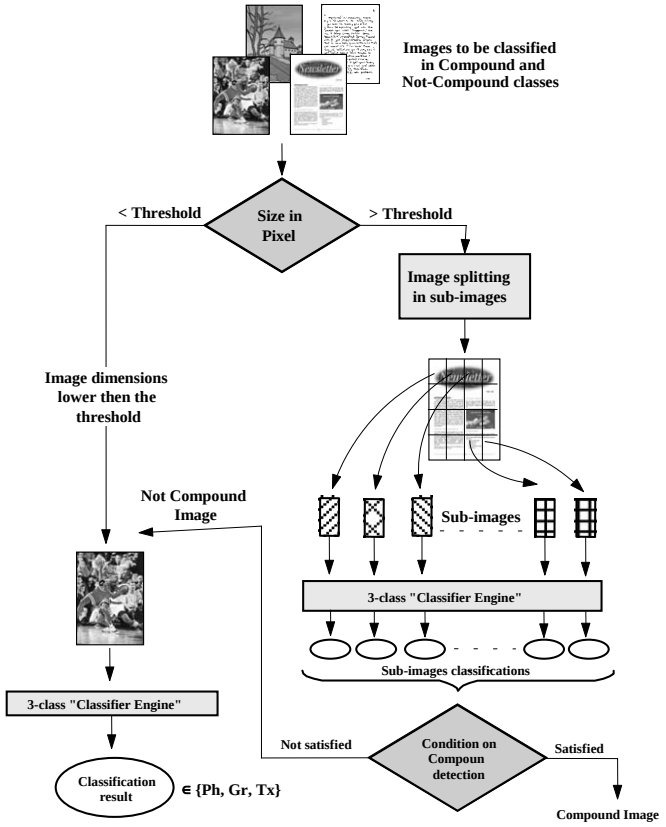

Fig. 4. Document classification strategy.

skin data was used to model the probability distribution of the skin color [18] (see the appendix).

The widely differing natures of the indices limit the risk of having different images correspond to very close points in the feature space. However, while all the features must be computed for the images in the training sets, only the features actually used by the classifier need to be computed for the images in the test sets and for new images to be processed. In our experimentation the features used in the classifiers we obtained are less than one-third of the original ones.

4. Document classification strategy

The problem addressed was that of classifying a color document as photo, graphic, text, or compound [19]. The photo class included photographs of indoor and outdoor scenes, landscapes, people, and objects. The graphic class included banners, logos, tables, maps, sketches and photo-realistic graphics. The text class included digitized handwritten texts, as well as colored and black and white texts both scanned and computer generated, in various fonts. Compound images were those containing data (homogeneous regions) of various types, namely text, photographs, and graphics. Examples of compound documents are structured articles such as journals, newspapers, newsletters, and documents with an unconstrained layout, such as advertisements and the covers of CDs, books and journals, together with non-traditional documents, such as web pages and video frames. We define the non-compound class as the union of photo, text and graphic classes.

The more straightforward way to address a classification problem with four classes would have been to use a four-class classifier. However, the great variety and complexity of compound images would have required the definition of a huge training set without guaranteeing its completeness. Consequently, we discarded this approach and defined the classification strategy described below and shown in Fig. 4. We first built and validated a "classifier engine" for the classification of photographs, graphics, and texts. We then used it to derive a compound vs. non-compound classifier: the images were subdivided into a given number of disjoint sub-images, and these were classified as photo, graphics, or text by

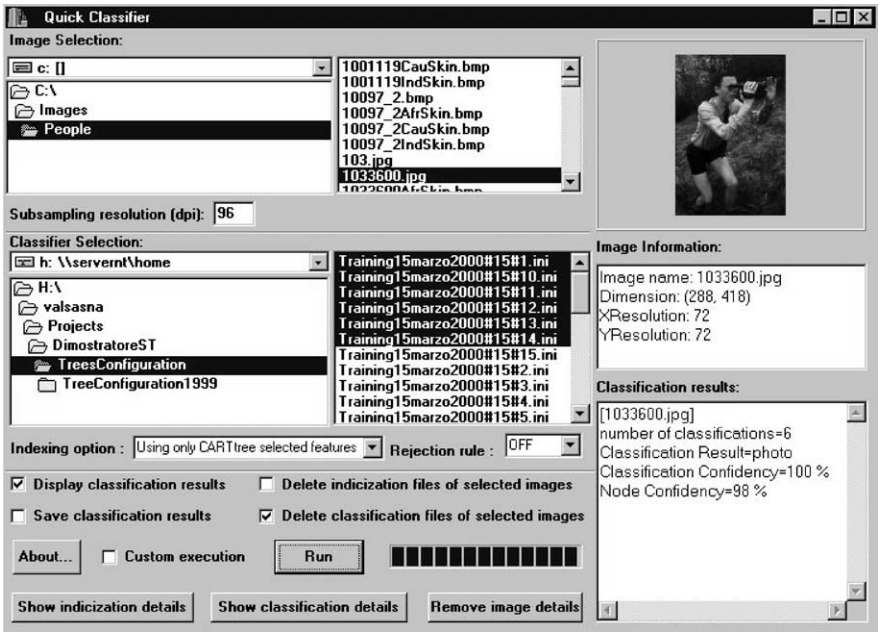

Fig. 5. The user interface of our system.

the "classifier engine". A measure of confidence for the classification of each sub-images was provided by the percentage of trees, combined in the "classifier engine", that contributed to the result. The whole image was classified as compound if, with a "good" level of confidence, its sub-images were classified in at least two of the three different classes. Non-compound images were then globally classified as photograph, graphic or text.

Images of dimensions (in pixels) smaller than a minimum threshold were excluded from the compound vs. non-compound classification, and classified directly globally as photograph, graphic, or text. This constraint was set because no strategy for compound document processing or analysis, such as region segmentation, or zone classification, could be useful or feasible in the case of images smaller than the chosen threshold.

Fig. 5 shows the user interface of the system which we implemented to perform the classification strategy.

Table 1
Average classification accuracy obtained on the training sets

	Predicted class		
True class	Photo	Graphic	Text
Photo	0.97	0.03	0
Graphic	0.02	0.93	0.05
Text	0	0.02	0.98

Table 2
Average classification accuracy obtained on the test sets

	Predicted class		
True class	Photo	Graphic	Text
Photo	0.95	0.04	0.01
Graphic	0.03	0.88	0.09
Text	0	0.08	0.92

5. Experimental results

The image database used in our experiments consisted of over 36,000 images collected from various sources: images downloaded from the web, or acquired by scanner, and bitmap versions of electronic pages. It contained some 30,000 photos, 4000 graphics, 1500 texts, and 1000 compound images. All this material varied in size (ranging from 120×120 pixels to 3500×3500 pixels), resolution, and tonal depth.

To address the three-class classification problem (photo, graphic, and text), we built several trees using independent training sets of some 4600 images (about 2500 photos, 1500 graphics and 600 texts) randomlyextracted from the available database with no replacement. All the image typologies present in the three classes were identically represented in each training set, and the three classes were assumed to have equal misclassification costs. The trees were all pruned by a cross-validation process. For each single tree, the test set was formed by

Table 3
Average classification accuracy obtained on the training sets by using the "classifier engines"

	Predicted class		
True class	Photo	Graphic	Text
Photo	0.99	0.1	0
Graphic	0.03	0.94	0.03
Text	0	0.1	0.99

Table 4
Average classification accuracy obtained on the test sets by using the "classifier engines"

	Predicted class		
True class	Photo	Graphic	Text
Photo	0.97	0.03	0
Graphic	0.03	0.93	0.03
Text	0	0.04	0.96

all the images not belonging to the training set used to build it, and included some 31,000 images (about 28,000 photos, 2500 graphics and 800 texts).

We generated several "classifier engines", combining, by majority vote, these single trees in groups of 15. For each "classifier engine" the training set was composed of all the images belonging to the training set of at least one of the trees combined in the classifier engine, employing in all some 26,000 images (22,000 photos, 2700 graphics, and 1200 texts). The test set was composed of all the images not belonging to any training set of the trees combined in the "classifier engine", and totalled some 9600 images (8000 photos, 1300 graphics, and 300 texts).

Tables 1 and 2 show the average classification accuracy of the three-class tree classifiers on the training and test sets, respectively.

Tables 3 and 4 show the average classification accuracy of the "classifier engines" obtained on the training and test sets, respectively.

As can be seen, the accuracy of the classification obtained by the single trees on the training sets is already very good; the "classifier engine" brings only a slight additional improvement. But the application of the "classifier engine" produces a considerable improvement in classification accuracy on the test sets, for graphics (5%) and texts (4%) in particular.

Table 5 shows the classification accuracy obtained for particular image typologies of the three classes considered. The typologies are: *building*, *people*, *animal*, *landscape*, *object* and *mixed* for the photo class, *easy clip art*, *smooth clip art*, *table* and *map* for the graphic class and, *black and white*, *colored background* and *colored text* for the text class.

If we look in detail at the classification results for the different image typologies, we see that the greatest improvements are achieved on those images most inaccurately classified by single trees. On the whole database, the accuracy for photographs of people and objects increases from 89% to 96% and from 87% to 93%, respectively; for smooth clip art, graphic tables and maps, from 87% to 94%, from 84% to 95%, and from 81% to 89%, respectively. For text images with a colored background or colored text the accuracy increases from 83% to 94% and from 87% to 98%, respectively.

The performance, evaluated in terms of overall classification accuracy, of the different "classifier engines" are very similar, in our application, that is related to

Table 5
The average classification accuracy obtained on particular image typologies[a]

Subclasses	Single classifier												CE		
	Training				Test				Whole DB				Whole DB		
	N	ph	gr	tx	*N*	ph	gr	tx	*N*	ph	gr	tx	ph	gr	tx
ph building	482	98	2	0	4840	96	3	1	5322	96	3	1	99	1	0
ph people	474	92	6	2	1973	89	9	3	2447	89	8	2	96	4	0
ph animal	529	99	1	0	4211	98	2	0	4640	98	1	0	100	0	0
ph landscape	491	98	1	0	10568	98	2	0	10779	98	2	0	99	1	0
ph object	383	92	8	1	1778	85	13	2	2161	87	12	1	93	6	1
ph mixed	0	///	///	///	3463	96	3	0	3463	96	3	0	98	2	0
gr easy clip art	171	2	96	2	2199	1	96	3	2370	1	96	3	0	99	1
gr smooth clip art	167	10	86	4	314	11	87	2	481	11	87	3	5	94	1
gr table	294	4	87	10	166	3	79	18	460	3	84	13	0	95	5
gr map	115	3	85	12	470	2	80	18	585	2	81	17	2	89	10
tx black and white	266	1	1	98	532	0	2	98	798	1	1	98	0	0	100
tx colored background	312	7	3	90	271	8	15	76	583	8	9	83	0	6	94
tx colored text	87	5	8	87	0	///	///	///	87	5	8	87	0	2	98

[a]CE: classifier engine; *N*: the number of images considered; and ph, gr and tx: photo, graphics and text, respectively.

Fig. 6. Examples of photos misclassified as graphics by the "classifier engine".

cross-media color image reproduction, we have chosen the "classifier engine" having the best performance on photos. Examples of misclassified photos, graphics and texts are shown in Figs. 6, 7 and 8. The photographs misclassified as graphics are mostly of small dimensions and low resolution, or object portraits with a uniform background. Graphics misclassified as photos are graphic illustrations with a photo realistic intent, or smooth clip art, while the graphics misclassified as texts are maps or tables with overlaid text. Texts misclassified as graphics present a few colored words in large fonts, or busy backgrounds.

For the detection of compound images, we experimented only on documents with a horizontal and vertical size exceeding the experimentally set threshold of 500 pixels. All the images in our database that satisfied this condition (about 1000 compound images, 1000 photographs, 500 graphics and 500 of text), were subdivided into disjoint sub-images of variable size, by a 4×4 equally spaced grid. The threshold for the acceptance of sub-image classification was set at 90%.

Compound and non-compound documents were correctly classified with an accuracy of 90% and 83%, respectively. Among the non-compound documents, 10% of the photographs were misclassified as compound, while the misclassification figure for graphics and text was 20%. We also observed that about 40% of the graphic and text images misclassified as compound, were misclassified by the three-class classification as well. Examples of misclassified compound images are given in Fig. 9 (see Table 6).

The two biggest problems in document subdivision performed for compound document detection are: first, that a sub-image of a compound may also be a compound image itself, rendering its classification in photo, graphic, and text classes a badly posed problem; and, second, that a non-compound sub-image may be misclassified, while the whole image is not. Both these problems can be handled by prior analysis of the document to roughly detect the position of any homogeneous regions constituting it, and utilize these as sub-images. Varied subdivisions into sub-images could also be used, and the results of the corresponding classifications compared. We plan to experiment these refinements of the strategy in the near future.

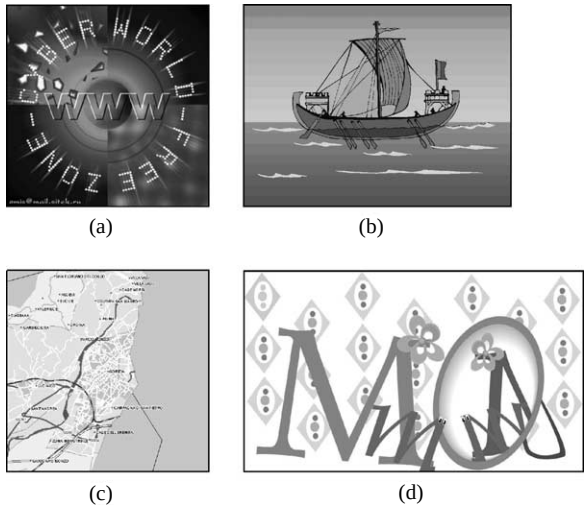

Fig. 7. Examples of graphics misclassified as photos (a, b) and texts (c, d) by the "classifier engine".

6. Conclusions

Digital imaging workflows have become increasingly complicated in the last few years. Many factors have driven the increased complexity of this arena: many

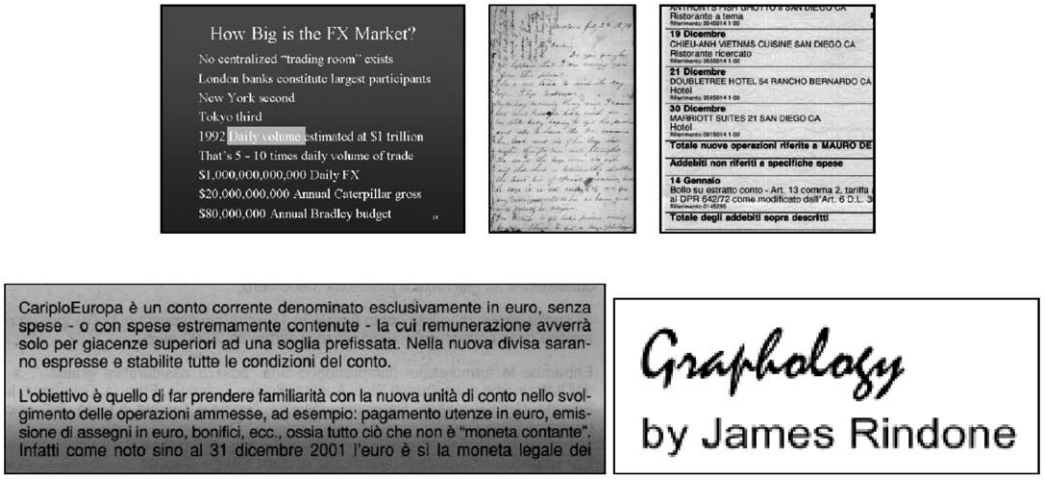

Fig. 8. Examples of texts misclassified as graphics by the "classifier engine".

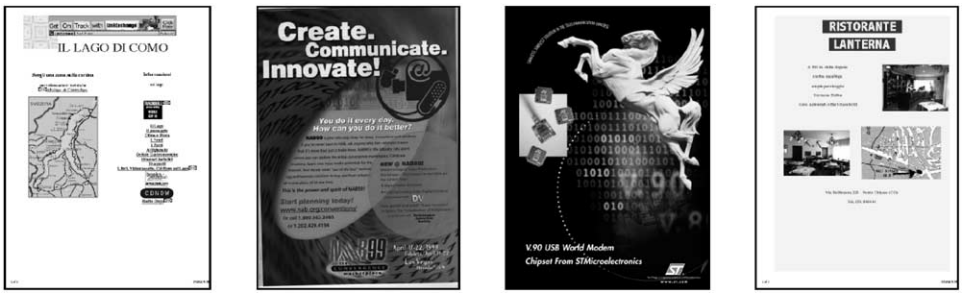

Fig. 9. Examples of misclassified compound images.

Table 6
Average classification accuracy of compound vs. non-compound classification, evaluated on image classes

	Predicted class	
True class	Non-compound	Compound
Photo	0.9	0.1
Graphic	0.8	0.2
Text	0.8	0.2
Compound	0.1	0.9

different kinds of imaging devices are now available (Inkjet and Laser Printers, Scanners, Digital Copiers, Digital Still Cameras, Internet Faxes, Monitors, and Multifunctional products), and for each type of device there are many different subcategories (taking printers, for example, we have high and low-end, networked and standalone, PC-centric and peer-to-peer products, etc.). Different driver-peripherals couples may also partition features differently, and functionalities and complex design vectors, such as speed, resolution, or the user-interface, must also be taken into account. Consequently, next generation designs in this field must address several issues, such as versatility (devices must have more and more features, and be easier to use), data size (increased resolution means more data to manage, calling for better compression and data representation schemes), quality, processing speed and ease of insertion of devices in complex home and office networks (interoperability, plug-and-play, cross-device optimization). We believe that content-based image classification will play an important role here: being able to properly classify text, graphics, photo and compound images will allow the unsupervised optimization of image data size and rendering intent using specific processing strategies. In this context tree classifiers built with the CART methodology present several advantages: (i) they can handle the co-existence of different relationships between the features in different regions of the feature space in a very natural way; (ii) they give a clear characterization of the conditions that determine when an image belongs to one class rather than to another, thereby detecting the most discriminant features for the problem addressed and unmasking redundancy; (iii) they do not require

assumptions about the probability distribution of the features; (iv) they not only provide a classification rule, but also allow the assignment of a degree of confidence in the classification; and (v) they may be very easily combined to derive an even more accurate classifier, as we have done.

Acknowledgements

Our investigation was performed as a part of a ST Microelectronics' research contract (Intelligent color printers). The authors thank ST Microelectronics for the permission to present this paper.

Appendix. Skin-tone detector

In Ref. [18] a statistical model of the skin-tone color class S has been proposed. It is based on the chromaticities (r, g) of the pixel, computed by

$$r = \frac{R}{R+G+B}, \qquad g = \frac{G}{R+G+B}.$$

Let $\underline{x}_{ij}$ be the vector of the chromaticities r, g of the pixel at side (i, j).

We have modeled the conditional probability function of $\underline{x}_{ij}$, belonging to the skin class S, by a bivariate normal distribution

$$p(\underline{x}_{ij} \mid S) = \frac{\exp(-\frac{1}{2}(\underline{x}_{ij} - \underline{\mu}_S)^{\mathrm{T}} \Sigma_S^{-1} (\underline{x}_{ij} - \underline{\mu}_S))}{2\pi |\Sigma_S|^{1/2}}.$$

We obtained an estimate of the above probability by estimating the mean $\underline{\mu}_S$ and the covariance matrix Σ_S from a training set of 30,000 skin-tone examples.

Given the above hypothesis the quadratic form

$$U(\underline{x}_{ij}; \underline{\mu}_S, \Sigma_S) = (\underline{x}_{ij} - \underline{\mu}_S)^{\mathrm{T}} \Sigma_S^{-1} (\underline{x}_{ij} - \underline{\mu}_S)$$

has a χ_2^2 probability distribution.

Therefore, given a confidence value of $\alpha \in [0, 1]$, it is possible to select the elliptic region

$$U_\alpha(\underline{x}; \underline{\mu}_S, \Sigma_S) < \lambda_\alpha,$$

which includes, with probability α, the skin-tone colors.

The feature entry is defined as the percentage of pixels with chromaticies that belong to the ellipse $U_{0.75}$, i.e.

$$f_{Skin\ tone} = (N \times M)^{-1} \sum_{i=1}^{M} \sum_{j=1}^{N} I_{(\underline{x}_{i,j} \in U_{0.75,S})}.$$

References

[1] L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, Classification and Regression Trees, Wadsworth and Brooks/Cole, Belmont, CA, 1984.
[2] V. Athitsos, M. Swain, Distinguishing photographs and graphics on the World Wide Web, Proceedings of Workshop in Content-based Access to Image and Video Libraries, 1997, pp. 10–17.
[3] T. Gevers, AWM Smeulders, PicTo seek: combining color and shape invariant features for image retrieval, IEEE Trans. Image Process. 19 (1) (2000) 102–120.
[4] R. Schettini, C. Brambilla, G. Ciocca, M. De Ponti, Color image classification using tree classifiers, Proceedings of the Seventh Color Imaging Conference: Scottsdale, Arizona, 1999, pp. 269–272.
[5] R. Schettini, C. Brambilla, A. Valsasna, M. De Ponti, Content-based image classification in: G.B. Beretta, R. Schettini (Eds.), Proceedings of the Internet Imaging Conference, Proceedings of SPIE 3964 2000, pp. 28–33.
[6] M. Szummer, R. Picard, Indoor-outdoor image classification, Proceedings of the International Workshop on Content-Based Access of Image and Video databases, 1998, pp. 42–51.
[7] A. Vailaya, M. Figueiredo, A.K. Jain, H.-J. Zhang, Image classification for content-based indexing, IEEE Trans. Image Process. 10 (1) (2001) 117–130.
[8] J.M. Chambers, T.J. Hastie (Eds.), Statistical Models in S, Chapman & Hall, London, 1992.
[9] P. McCullagh, J.A. Nelder, Generalized Linear Models, Chapman & Hall, London, 1989.
[10] L. Breiman, Bagging predictors, Mach. Learning 26 (1996) 123–140.
[11] M.A. Stricker, M. Orengo, Similarity of Color Images, Proceedings of SPIE Storage and Retrieval for Image and Video Databases III Conference, 1995.
[12] J. Canny, A computational approach to edge detection, IEEE Trans. Pattern Anal. Mach. Intell. IEEE-8 (1986) 679–698.
[13] F. Idris, S. Panchanathan, Storage and retrieval of compressed images using wavelet vector quantization, J. Visual Languages Comput. 8 (1997) 289–301.
[14] P. Scheunders, S. Livens, G. Van de Wouwer, P. Vautrot, D. Van Dyck, Wavelet-based texture analysis, Int. J. Comput. Sci. Inform. Manag. 1 (2) (1998) 22–34.
[15] M. Amadasun, R. King, Textural features corresponding to textural properties, IEEE Trans. System Man Cybernet. 19 (5) (1989) 1264–1274.
[16] H. Tamura, S. Mori, T. Yamawaki, Textural features corresponding to visual perception, IEEE Trans. System Man Cybernet. 8 (1978) 460–473.
[17] P. Ciocca, R. Schettini, A relevance feedback mechanism for content-based retrieval, Inform. Process. Manag. 35 (1999) 605–632.
[18] Y. Miyake, H. Saitoh, H. Yaguchi, N. Tsukada, Facial pattern detection and color correction from television picture for newspaper printing, J. Imaging Technol. 16 (1990) 165–169.
[19] K.-C. Fan, C.-H. Liu, Y.-K. Wang, Segmentation and classification of mixed text/ graphics/ image documents, Pattern Recognition Lett. 15 (1994) 1201–1209.

About the Author—RAIMONDO SCHETTINI has been associated with the Italian National Research Council (CNR) since 1987. With 1994 he moved to the Institute of Multimedia Information Technologies, where he is currently in charge of the Image and Color Analysis Lab. He has published more than 110 refereed papers on image processing, analysis and reproduction, and on image content-based indexing and retrieval. Since 1997 he also teaches a course on multimedia design at the Faculty of Industrial Design of the Polytechnic of Milan. He is a member of the CIE TC 8/3, general co-chair of the Internet Imaging Conference since 2000, and general-co-chair of the First European Conference on Color in Graphics, Imaging and Vision (CGIV'2002).

About the Author—CARLA BRAMBILLA is a senior researcher at the Institute for Applications of Mathematics and Informatics of the Italian National Research Council (CNR). Her research focuses on classification and regression trees, generalized linear models, and survival analysis.

About the Author—GIANLUIGI CIOCCA took his Laurea degree in Computer Science at the University of Studies of Milan in 1998. He is a fellow at the Institute of Multimedia Information Technologies of the National Research Council of Italy since 1998 where his research has focused on multimedia system design, and image and on video content-based retrieval.

About the Author—ANNA VALSASNA took her Laurea degree in Mathematics at the University of Studies of Milan in 1998. She is a fellow at the Institute of Multimedia Information Technologies of the National Research Council of Italy since 1998 where her research has focused on image analysis and classification.

About the Author—MAURO DE PONTI took his degree in EE in 1987. Since then he has been working in the image processing arena, first on DTV and HDTV codec design, then on image scanners and printers. He is currently with STMicroelectronics, where he is manager of the Printer Division's Imaging and Advanced Architectures team.