Trace Transform Based Method for Color Image Domain Identification

Igor G. Olaizola*, Member, IEEE*, Marco Quartulli*, Member, IEEE*, Julián Flórez*, Member, IEEE*, and Basilio Sierra*, Member, IEEE*

***Abstract*—Context categorization is a fundamental pre-requisite for multi-domain multimedia content analysis applications. Most feature extraction methods require prior knowledge to decide if they are suitable for a specific domain and to optimize their input parameters. In this paper, we introduce a new color image context categorization method (DITEC) based on the trace transform. The problem of dimensionality reduction of the obtained trace transform signal is addressed through statistical descriptors of its frequency representation that keep the underlying information. We also analyze the distortions produced by the parameters that determine the sampling of the discrete trace transform. Moreover, Feature Subset Selection (FSS) is applied to both, improve the classification performance and compact the final length of the descriptor that will be provided to the classifier. These extracted features offer a highly discriminant behavior for content categorization without prior knowledge requirements. The method has been experimentally validated through two different datasets.**

***Index Terms*—CBIR, image domain identification, pattern recognition, trace transform.**

I. Introduction

THE importance of semantic context is very well known in Content Based Image Retrieval (CBIR) [1], [2]. This is especially relevant for broad-domain data intensive multimedia retrieval activities such as TV production and marketing or large-scale earth observation archive navigation and exploitation. Most modeling approaches rely on local low-level features, based on shape, texture, color etc. The drawback of these methods is that the characterization of the context requires prior contextual information, introducing a chicken-and-egg problem [3]. A possible approach to reduce this dependency involves the exploitation of global image context characterization for semantic domain inference. This prior information on scene context could represent a valuable asset in computer vision for purposes ranging from regularization to the pre-selection of local primitive feature extractors [4].

Manuscript received November 26, 2012; revised June 19, 2013 and October 21, 2013; accepted December 06, 2013. Date of publication January 23, 2014; date of current version March 13, 2014. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Sen-Ching Cheung.

I. G. Olaizola, M. Quartulli, and J. Flórez are with the Department of Digital TV & Multimedia Services, San Sebastián 20009, Spain (e-mail: iolaizola@vicomtech.org; mquartulli@vicomtech.org; jflorez@vicomtech.org).

B. Sierra is with the University of the Basque Country, San Sebastián 20008, Spain (e-mail: b.sierra@ehu.es).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2014.2300843

Novel semantic approaches that try to overcome the current existing limitation derived from fixed taxonomies and manual annotations, rely on automatic or semi-automatic ingestion processes. These processes minimize the *semantic gap* by introducing *semantic middleware* [5] layers based on a combination of:

- explicit information provided by human made taxonomies.
- relevance feedback data and knowledge extracted from manual annotations.
- implicit information obtained by data mining techniques through training processes.

In this paper we introduce a new method for global feature extraction based on a statistical modeling of the trace transform in the frequency domain. This method is very suitable for semantic context classification, especially for those cases where the lack of prior knowledge does not allow the effective use of specific local features.

A. Related Works

Research contributions related to the approach proposed in this paper are outlined in this section.

Local features have been used broadly for context categorization [6], [7]. SIFT [8] and SURF [9] are among most popular choices in this respect. A two step approach for the efficient use of local features has been proposed by several authors such as Ravinovich *et al.* [10] and Choi *et al.* [11]. Olaizola *et al.* [12] have proposed an architecture for hypothesis reinforcement based on an initial analysis of low-level features for context categorization and further hypothesis creation. This architecture can exploit context specific feature extractors to validate or refuse the initial context hypothesis. This stresses the value of global descriptors for initial domain categorization purposes. Once a specific domain has been identified, different low-level features can be extracted. However, these features cannot be combined in a simple way and the obtained multi-attribute spaces must be normalized in order to be used in similarity search or retrieval tasks [13].

Among different global descriptors such as histograms of several local features [14], texture features or self similarity, there are also some specific algorithms in the literature which have shown great potential: GIST [15], [16] is probably one of the most popular ones. In typical operational implementations, all these algorithms are combined with other global or local features.

The trace transform has already been proposed for several computer vision applications. Indeed, a method based on this transform has been included in the MPEG-7 [17] standard specification for image fingerprinting [18], [19]. Other applications (mostly with monochrome images) include face recognition [20], [21], character recognition [22] and sign recognition [23]. The proposed approach based on a recursive application of the trace transform to reduce the dimensionality of the obtained feature space (known as the *triple feature*), offers an excellent performance for image fingerprinting, but does not offer good semantic discriminative characteristics for domain characterization due to the high data loss incurred by the diametrical and circus functionals [24]. The approach proposed by Liu and Wang [21] reduces the number of attributes using Principal Component Analysis (PCA) to select the most relevant coefficient and reduce the dimensionality of the feature space. However, this approach does not take into account the frequency relationships among the different coefficients and increases the feature extraction complexity as it requires the covariance matrix information of all previous samples. Moreover, the feature relevance of each individual DCT coefficient is too low and also sensitive to noise and variations.

Li *et al.* [25] have proposed a generalization of the Radon transform and trace transform by introducing prior knowledge of specific identification or fingerprinting tasks and extending the geometric sets from straight lines to arbitrary choices. This approach provides a complete set of resources for non-rigid object identification and has been successfully tested for pedestrian recognition, part segmentation and video retrieval. However, the broad set of configuration parameters and pre-processing tasks is not suitable for domain identification purposes where the lack of *a priori* knowledge is one of the main issues.

II. General Description of the DITEC Method

The DITEC method is composed of four main steps where an observed image D is estimated as $\hat{C}$ of the unknown global image semantic concept C. The four DITEC steps are thus the following:

Sensor modeling: image acquisition and pre-processing (radiometric noise, color space, geometric quantization and image lattice finiteness effects).

Data transformation: trace transform (detailed in 2.2) applied to the pre-processed image I. The result will depend on the chosen functional and on the selected geometric parameters (detailed in Section 2.2.3). The outcome T of the trace transform of an image is a two-dimensional signal represented by means of sinusoids with a particular amplitude, phase, frequency and intensity. This characterization process represents one of the key steps in the overall information extraction process.

Feature extraction: summarization of the extracted features T, compressed and adapted into a manageable set E of object-based descriptors. It aims to reduce the considered descriptor space dimensionality while preserving essential information in order to allow a good performance in the subsequent classification process.

Class assignment: vectors obtained in the previous step are processed to improve the performance of classifiers in the defined feature space. All the obtained vectors are statistically analyzed to select their most representative attributes. Then the supervised classification process is carried out to obtain an estimate $\hat{C}$ of the unknown global image semantic concept C.

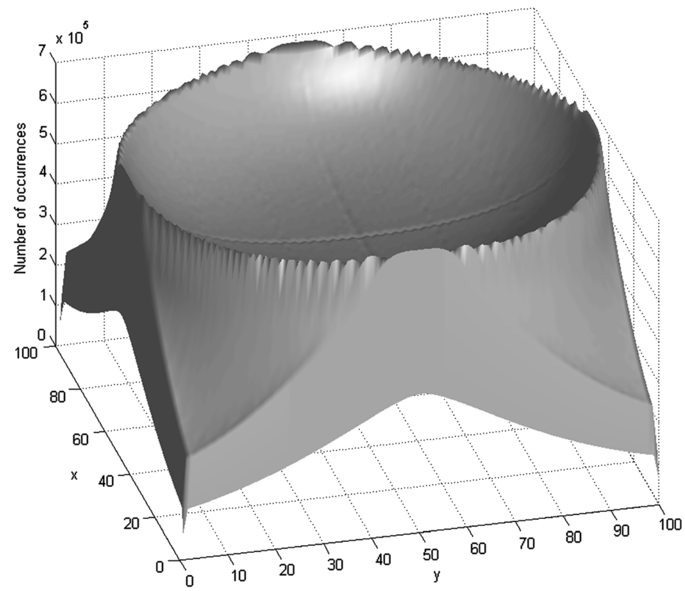


Fig. 1. Trace transform contribution mask at very high resolution parameters (Image resolution: 100×100 px. $n_\phi = 1000, n\rho = 1000, n_\xi = 5000$).

A. Sensor Modeling

The first pre-processing step transforms the RGB color space into YC_bC_r [26]. The luminance channel (Y) will be used as the most relevant channel to encode shape related features. Color distribution information is encoded by processing the chrominance channels (C_b, C_r).

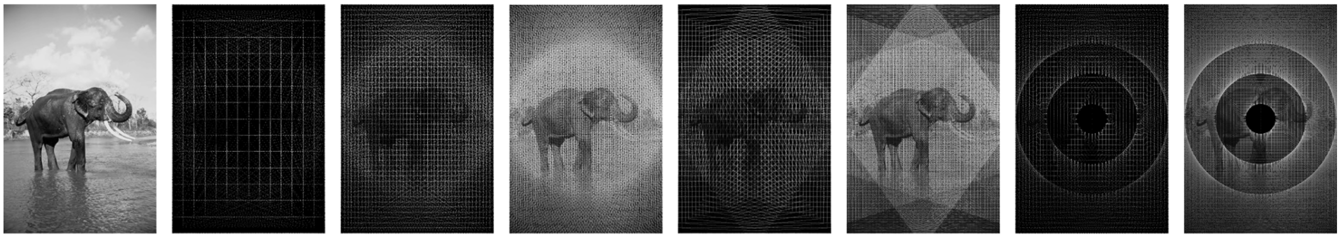
In order to reduce effects introduced by radiometric noise, image lattice and quantization, a low-pass filter is applied to each channel.

HSV [26] color space information is encoded by obtaining mean and variance values (μ, σ) of the corresponding intensity distributions in each H,S,V channel. In the Attribute Selection process, this (μ, σ) information is introduced into the obtained descriptor E.

B. Data Transformation

The data transformation process is carried out through the trace transform. The trace transform extends the Radon transform by enabling the definition of the functional and thus enhancing the control on the feature space [27]. These features can be set up to show scale, rotation/affine transformation invariance or high discriminance for specific content domains. The outcome T of the trace transform of a 2D image is another 2D signal composed of a set of sinusoidal shapes that vary in amplitude, phase, frequency, intensity and thickness. These sinusoidal signals encode the pre-processed image I with a given level of distortion depending on the functional and quantization parameters.

1) Functionals: A functional Ξ of a function $\xi(x)$ evaluated along the line L will have different properties depending on the features of function $\xi(x)$ (e.g.: invariance to rotation, translation and scaling [28]). Kadirov *et al.* [29] propose several functionals with different invariance or sensitiveness properties. The selec-



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 2. Pixels relevance in trace transform scanning process with different parameters (n_ϕ, n_ρ, n_ξ). Original image resolution $= 384 \times 256$. (a) Original. (b) (64,64,15). (c) (64,64,45). (d) (64,64,185). (e) (5,300,45). (f) (5,300,151). (g) (300,5,45). (h) (300,5,151).

tion of functionals depends on the content inherent features and characterization criteria. The functionals used in this paper have been $IF_2 \to (\int |\xi(t)|^q dt)^r$ and $\frac{\int \xi(t)dt}{\Delta_t}$.

2) Geometrical Constraints: The result of the discrete trace transform strongly depends on the selected geometrical parameters. The three resolution parameters denoted by $\Delta\phi, \Delta\rho, \xi(\Delta L)$ respectively for angle, radius and the sampling rate along the line L, establish distortions and aliasing effects that will affect the final result of the trace transform.

The final resolution of the sinogram T obtained by applying the trace transform will be defined by n_ϕ and n_ρ where:

$$n_\phi = \frac{2\pi}{\Delta\phi} \tag{1}$$

$$n_\rho = \frac{\min(X, Y)}{\Delta\rho} \tag{2}$$

with X and Y denoting the horizontal and vertical resolutions of the image I_l.

Low (n_ϕ, n_ρ, n_ξ) values will have a non-linear downsampling effect on the original image, where n_ξ is defined as:

$$n_\xi = \frac{1}{\Delta L} \tag{3}$$

The set of points used to evaluate each functional is described (assuming (0,0) as the center of the image) by:

$$y = 2\rho\sin(\phi) - \frac{x}{\tan(\phi)} \tag{4}$$

A singularity can be observed at $\phi = 0$ and $\phi = \pi$. For these cases it can be assumed that:

$$\begin{aligned} x &= \rho \quad \forall y \quad if\ \phi = 0 \\ x &= -\rho \quad \forall y \quad if\ \phi = \pi \end{aligned} \tag{5}$$

X and Y are the horizontal and vertical resolutions of the image. Equation (4) shows a symmetrical result since the same lines are obtained for $\phi \in [0, \pi]$ and $\phi \in [\pi, 2\pi]$. However this is only true for functionals that are not considering the position (such as the Radon transform). Depending on the selected functional and on the desired properties of the trace transform (e.g: rotational invariance), the ranges of ϕ and ρ can be modified to: $\phi \in [0, \pi]$ or $\rho \in [0, r]$.

3) Sampling Effects: Digital images are affected by two main effects during trace transformation:

- some pixels might never be used by the functional given the geometrical setup of the transform, and to its integration nature.
- there may be some pixels that have much higher cumulated effect than the others into the functional.

In this section we will analyze the effects that need to be taken into account in order to preserve the homogeneity of the results, avoiding pixels or areas with higher relevance than others. Even for very high (n_ϕ, n_ρ, n_ξ) values in relation to the original image resolution, the trace transform introduces a contribution intensity map that encodes the relevance of the different regions of the input picture. As shown in Fig. 1, high resolution values of the trace transform parameters tend to create a convex contribution intensity map. Therefore, high parameter values do not necessarily imply optimal image content representation on the trace transform.

High values of n_ϕ improve the rotational invariance of the trace transform (although in such a manner that it is dependent on the selected functional) while very low values $n_\phi < 5$ cannot be considered as producing a valid trace transform since there is not enough angular information.

Fig. 2 shows some cases applied to a real image and the convex contribution intensity mask effect for different sampling values.

C. Feature Extraction

Diametric and *Circus* transform have been used mostly in the literature [18], [29], [30], [31] to reduce the set of descriptors. However, even if this approach provides good results for similarity search or image hashing, the diametric and circus transform do not preserve the information. In fact, there is no inverse transform for these two operators.

In order to characterize the sinograms obtained from the trace transform (T), we propose the frequency analysis of the obtained signal and a representation based on statistical descriptor of the frequency distribution. To do this, Discrete Cosine Transform (DCT) is be applied due to its energy compaction and decorrelation properties [32].

1) Statistical Descriptors: As a consequence of the properties of the DCT and of the nature of the 2D signals resulting from the trace transform, the 2D DCT stores more energy in its lower frequencies.

In order to reduce the dimensionality of the obtained coefficients the n first orthogonal straight lines to the main diagonal of the transformed signal T are statistically characterized. These coefficients which correspond to similar frequency bands can be computed very efficiently and provide a high dimensionality reduction ratio.

To study these statistical properties, over 50000 sample vectors have been analyzed using the 1000 sample images of Corel 1000 dataset (described in Section 3.1). The analysis of obtained histograms shows strong leptokurtic distributions for all samples. Equation (6) defines the kurtosis of a distribution which is represented by (7) for a discrete set of elements. A distribution is considered leptokurtic when $k > 3$. For all analyzed distributions the minimum kurtosis value has been greater than 30.

$$k = \frac{E(x-\mu)^4}{\sigma^4} \tag{6}$$

$$k = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^4}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^2} \tag{7}$$

Assuming the leptokurtic nature of the obtained distributions, the list of values can be represented by the mean value and the kurtosis of each vector. This pair of descriptors (μ, k) of the first element (corresponding to the DC value of the DCT) is substituted by the mean and variance of the original image in HSV space. Considering that the mean and kurtosis values encode the information of coefficients corresponding to approximately similar frequencies. The obtained dimensionality of the transformed (μ, k) pairs is given by (8).

$$nDims = \sqrt{n_\phi^2 + n_\rho^2} \cdot n_c \cdot n_f \tag{8}$$

where n_c is the number of channels of the original image and n_f the number of features extracted from each vector (2 in the case of using $[\mu, k]$). Thus, the dimensionality reduction is given by (9).

$$r_f = \frac{n_\phi n_\rho}{\sqrt{n_\phi^2 + n_\rho^2} \cdot n_f} \tag{9}$$

For square resolutions and considering $n_f = 2$ the reduction factor increases linearly with the resolution (10).

$$r_f = \frac{n^2}{n \cdot n_f \sqrt{2}} = \frac{n}{2\sqrt{2}} \tag{10}$$

D. Class Assignment

After the feature extraction process explained in the previous section, a set of features E is obtained. The dimensionality of E has been reduced by applying a Feature Subset Selection (FSS) algorithm. Once the objective is fixed, FSS can be understood as a search problem with each state in the search space specifying a subset of the possible features of the task. Exhaustive evaluation of possible feature subsets has a extremely high computational complexity. Many search techniques have been proposed to solve FSS problems carrying out an intelligent search in the space of possible solutions. In this paper an Estimation Distribution Algorithm (EDA) [33] has been used to this end. Then, the obtained feature space has been processed with Support Vector Machines (SVM) and Bayesian Networks.

TABLE I
COREL 1000 DATASET CONFUSION MATRIX. GROUND TRUTH REPRESENTED IN ROWS, PREDICTED LABELS IN COLUMNS. F–MEASURE IS THE HARMONIC MEAN: $F = 2 \cdot \frac{precision \cdot recall}{precision + recall}$

Class	Precision	Recall	F–Measure
a	0.75	0.75	0.75
b	0.752	0.79	0.771
c	0.772	0.78	0.776
d	0.9	0.81	0.853
e	0.98	1	0.99
f	0.806	0.83	0.818
g	0.941	0.95	0.945
b	0.942	0.97	0.956
i	0.813	0.78	0.796
j	0.828	0.82	0.824
Average	**0.848**	**0.848**	**0.848**

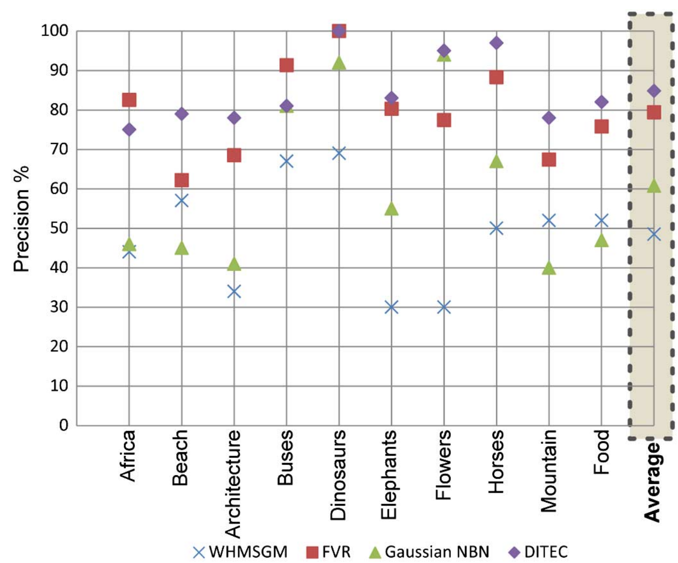


Fig. 3. Corel 1000 precision results with different feature extraction algorithms. *WHMSGM*: Mean-Shift and Gaussian Mixtures based on Weighted Color Histograms, *FVR*: Reduced Feature Vector with Relevance Feedback, *Gaussian NBN*: SIFT based Gaussian Naïve Bayesian Network.

III. EXPERIMENTAL RESULTS

The presented method has been tested with 2 different datasets. The first of them (Corel 1000 [34]) is a standard dataset which will allow the comparison of the obtained validation data with other methods existing in the literature. The second case (earth observation data), will be used to show the potential of the proposed method under diverse conditions. A 10-fold cross validation has been used in both cases to split the data into training and testing sets.

(a) (b) (c) (d) (e) (f) (g)

Fig. 4. Samples of satellite footage dataset. 256×256 px patches at different scales. (a) Athens. (b) Davis. (c) Manama. (d) Midway. (e) Nyragongo. (f) Risalpur. (g) Rome.

A. Case Study 1: Corel 1000 Dataset

The Corel 1000 dataset is composed of 1000 images distributed in 10 classes (100 instances per class). The tags of the classes are: *Africa, Beach, Architecture, Buses, Dinosaurs, Elephants, Flowers, Horses, Mountains and Food*. Even though they are semantically separated, visual similarities may be found among some of them. For example, people and trees can be found under *Africa, Beach,* and *Mountain* categories.

The following parameters have been selected: $n_\phi = 71, n_\rho = 71, n_c = 3, n_f = 2$. This choice results in 15,123 trace transform coefficients per image. By obtaining the mean values and kurtosis as described in the previous section, the number of attributes is reduced to 606 (by a factor of 25).

Based on the fact that the DCT gathers signal energy in the lower frequencies, highest coefficients are removed. Moreover, it can be assumed that chrominance channels (C_b and C_r) contain less visual information and therefore more coefficients can be removed from these channels than from the luminance signal (Y). Experimental results carried out with different combination of YC_bC_r coefficients, confirm that luminance related attributes have more relevance than chrominance related ones. The selected parameters for this example result in 202 attributes per channel. We will select the first 104 ones for Y and 60 for each C_bC_r signal, thus reducing the total amount of attributes to 224.

The best performance has been obtained by applying a SVM classifier (accuracy $= 84.8\%$ in a k-fold 10 test). 117 attributes have been selected for the final feature space by applying FSS. The resulting confusion matrix can be observed in Table I.

Comparing the obtained results with other feature extraction approaches (Mean-Shift and Gaussian Mixtures based on Weighted Color Histograms [14], Reduced Feature Vector with Relevance Feedback [35] and SIFT based Gaussian Naïve Bayesian Network [36]), DITEC shows the best performance for most categories (Fig. 3) and the highest mean precision value. Other performance parameters (such as recall, F–Measure) have not been compared since they have not be indicated in the papers related with the rest of the methods.

B. Case Study 2: Geoeye Satellite Imagery

The Geoeye [37] dataset is composed of 1003 multi-resolution patches of Digital Globe Earth observation satellite imagery with up to ~ 1 m spatial resolution. The dataset is categorized in 7 classes corresponding to different geographical locations (Fig. 4). All the resolutions have been processed with the same trace transform parameters.

TABLE II
GEOEYE DATASET CONFUSION MATRIX. GROUND TRUTH REPRESENTED IN ROWS, PREDICTED LABELS IN COLUMNS. F–MEASURE IS THE HARMONIC MEAN: $F = 2 \cdot \frac{precision \cdot recall}{precision + recall}$

Class	Precision	Recall	F–Measure
(a) Athens	0.961	0.961	0.961
(b) Davis	1	0.943	0.971
(c) Manama	0.97	0.995	0.982
(d) Midway	1	0.954	0.976
(e) Nyragongo	0.939	0.906	0.922
(f) Risalpur	0.898	0.912	0.905
(g) Rome	0.897	0.938	0.917
Average	**0.946**	**0.845**	**0.945**

During the data mining process Bayesian networks have shown the best performance, reaching an accuracy of 94.51% in a k-fold 10 test. The final dimensionality of the feature space has been reduced to 61 attributes. Table II shows the confusion matrix of the classification results.

IV. CONCLUSION

We have shown that the DITEC method provides highly discriminant features for context categorization purposes that can be encoded as considerably short feature vectors. We have presented the geometrical constraints of the trace transform that can be optimized to efficiently represent the information contained in the original images. We also have demonstrated that the dimensionality reduction in terms of mean and kurtosis value pair of frequency coefficients results in a very robust set of features in terms of precision. For different resolution $(n_\phi, n_\rho, L(n))$ settings the accuracy has maintained around 82% for the Corel 1000 dataset and 92% for Geoeye.

Moreover, the method has successfully identified visual similarities within the datasets, and as seen in the validation section, some incorrectly classified instances are in fact visually similar to those pointed out by the classifier.

REFERENCES

[1] T. M. Strat and M. A. Fischler, "Context-based vision: recognizing objects using information from both 2 d and 3 d imagery," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 10, pp. 1050–1065, 1991.

[2] A. Torralba and P. Sinha, "Statistical context priming for object detection," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2001 [Online]. Available: http://web.mit.edu/torralba/www/iccv2001.pdf

[3] A. Torralba, K. P. Murphy, and W. T. Freeman, "Using the forest to see the trees: Exploiting context for visual object detection and localization," *Commun. ACM*, vol. 53, no. 3, pp. 107–114, Mar. 2010 [Online]. Available: http://doi.acm.org/10.1145/1666420.1666446

[4] 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. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1349–1380, Dec. 2000 [Online]. Available: http://dx.doi.org/10.1109/34.895972

[5] G. Marcos, K. Alonso, I. G. Olaizola, J. Florez, and A. Illarramendi, "Dms-1 driven data model to enable a semantic middleware for multimedia information retrieval in a broadcaster," in *Proc. 4th Int. Workshop Semantic Media Adaptation and Personalization SMAP '09*, 2009, pp. 33–37.

[6] C. G. M. Snoek and A. W. M. Smeulders, "Visual-concept search solved?," *Computer*, vol. 43, no. 6, pp. 76–78, 2010.

[7] J. C. van Gemert, C. J. Veenman, A. W. M. Smeulders, and J.-M. Geusebroek, "Visual word ambiguity," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 7, pp. 1271–1283, 2010.

[8] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 1999, vol. 2, pp. 1150–1157 [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=790410

[9] H. Bay, T. Tuytelaars, and L. V. Gool, "Surf: Speeded up robust features," in *Proc. ECCV*, 2006, pp. 404–417.

[10] A. Rabinovich, A. Vedaldi, C. Galleguillos, E. Wiewiora, and S. Belongie, "Objects in context," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2007, pp. 1–8.

[11] M. J. Choi, J. J. Lim, A. Torralba, and A. S. Willsky, "Exploiting hierarchical context on a large database of object categories," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 129–136.

[12] I. G. Olaizola, G. Marcos, P. Kramer, J. Florez, and B. Sierra, "Architecture for semi-automatic multimedia analysis by hypothesis reinforcement," in *Proc. IEEE Int. Symp. Broadband Multimedia Systems and Broadcasting BMSB '09*, 2009, pp. 1–6.

[13] W. Scheirer, N. Kumar, P. Belhumeur, and T. Boult, "Multi-attribute spaces: Calibration for attribute fusion and similarity search," in *Proc. 2012 IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 2933–2940.

[14] M. A. Bouker and E. Hervet, "Retrieval of images using mean-shift and Gaussian mixtures based on weighted color histograms," in *Proc. 7th Int. Signal-Image Technology and Internet-Based Systems (SITIS) Conf.*, 2011, pp. 218–222.

[15] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int. J. Comput. Vision*, vol. 42, pp. 145–175, 2001.

[16] A. Torralba, R. Fergus, and Y. Weiss, "Small codes and large image databases for recognition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition CVPR*, 2008, pp. 1–8.

[17] J. M. Martinez, "MPEG-7 overview," ISO/IEC JTC1/SC29/WG11 Std., Rev., Oct. 10, 2004 [Online]. Available: http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm

[18] M. Bober and R. Oami, Description of mpeg-7 visual core experiments, ISO/IEC JTC1/SC29/WG11, 2007, Tech. Rep.

[19] R. O'Callaghan, M. Bober, P. Oami, and R. Brasnett, "Information technology - multimedia content description interface - part 3: Visual, AMENDMENT 3: Image signature tools," ISO/IEC Std. ISO/IEC TC JTC1/SC 29/WG 11 N 9581, vol. 01, 2008.

[20] S. Srisuk, M. Petrou, W. Kurutach, and A. Kadyrov, "Face authentication using the trace transform," in *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, 2003, vol. 1.

[21] N. Liu and H. Wang, "Recognition of human faces using siscrete cosine transform filtered trace features," in *Proc. 6th Int Information, Communications & Signal Processing Conf.*, 2007, pp. 1–5.

[22] M. F. Nasrudin, M. Petrou, and L. Kotoulas, "Jawi character recognition using the trace transform," in *Proc. 7th Int. Computer Graphics, Imaging and Visualization (CGIV) Conf.*, 2010, pp. 151–156.

[23] J. Turan, Z. Bojkovic, P. Filo, and L. Ovsenik, "Invariant image recognition experiment with trace transform," in *Proc. 7th Int. Telecommunications in Modern Satellite, Cable and Broadcasting Services Conf.*, 2005, vol. 1, pp. 189–192.

[24] A. Kadyrov and M. Petrou, "The trace transform and its applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 8, pp. 811–828, 2001.

[25] J. Li, S. K. Zhou, and R. Chellappa, "Appearance modeling using a geometric transform," *IEEE Trans. Image Process.*, vol. 18, no. 4, pp. 889–902, Apr. 2009 [Online]. Available: http://dx.doi.org/10.1109/TIP.2008.2011381

[26] C. Poynton, *A Technical Introduction to Digital Video*. New York, NY, USA: Wiley, 1996 [Online]. Available: http://books.google.es/books?id=jI5jQgAACAAJ

[27] A. Kadyrov and M. Petrou, "The trace transform as a tool to invariant feature construction," in *Proc. 14th Int. Pattern Recognition Conf.*, 1998, vol. 2, pp. 1037–1039.

[28] R. Fooprateepsiri, W. Kurutach, and S. Tamsumpaolerd, "An image identifier based on hausdorff shape trace transform," in *Proc. 16th Int. Conf. Neural Information Processing: Part I*, ser. ICONIP '09. Berlin, Germany: Springer-Verlag, 2009, pp. 788–797 [Online]. Available: http://dx.doi.org/10.1007/978-3-642-10677-4_90

[29] A. Kadyrov and M. Petrou, "Affine parameter estimation from the trace transform," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1631–1645, 2006.

[30] M. Petrou and A. Kadyrov, "Affine invariant features from the trace transform," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 1, pp. 30–44, 2004.

[31] P. Brasnett and M. Bober, "Fast and robust image identification," in *Proc. 19th Int. Conf. Pattern Recognition ICPR 2008*, 2008, pp. 1–5.

[32] I. E. Richardson, *Video Codec Design: Developing Image and Video Compression Systems*. New York, NY, USA: Wiley, 2002.

[33] R. Blanco, P. Larrañaga, I. Inza, and B. Sierra, "Gene selection for cancer classification using wrapper approaches," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 18, no. 8, pp. 1373–1390, Dec. 2004.

[34] [Online]. Available: http://wang.ist.psu.edu/docs/related.shtml

[35] G. Zajic, N. Kojic, N. Reljin, and B. Reljin, "Experiment with reduced feature vector in cbir system with relevance feedback," in *IET Conf. Publications*, 2008, vol. 2008, no. CP543, pp. 176–181 [Online]. Available: http://link.aip.org/link/abstract/IEECPS/v2008/iCP543/p176/s1

[36] W. Bouachir, M. Kardouchi, and N. Belacel, "Fuzzy indexing for bag of features scene categorization," in *Proc. 5th Int. I/V Communications and Mobile Network (ISVC) Symp.*, 2010, pp. 1–4.

[37] [Online]. Available: http://www.geoeye.com

Igor G. Olaizola is the head of Digital TV and Multimedia Services Department in Vicomtech (http://www.vicomtech.org). He received his MEng degree in Electronic and Control Engineering from the University of Navarra, Spain (2001). He developed his Master thesis at Fraunhofer Institut für Integrierte Schaltungen (IIS), Erlangen -Germany- 2001 and currently he is preparing his PhD in Computing Science and Artificial Intelligence at University of Basque Country. He has participated in many industrial projects related with Digital TV as well as several European research projects in the area of audiovisual content management. His current research interests include multimedia content analysis frameworks and techniques to decrease the semantic gap.

Marco Quartulli received the laurea degree in physics from the University of Bari, Italy, in 1997 and a PhD in EE and CS from the University of Siegen, Germany, in 2005. He worked from 1997 to 2010 on remote sensing ground segment engineering, image analysis, archives and mining for Advanced Computer Systems, Italy. From 2000 to 2003, he was with the Image Analysis Group at the Remote Sensing Technology Institute of the German Aerospace Center (DLR) in Germany. Since 2010, he has joined the Digital Television department of Vicomtech in Spain. His research interests include multimedia mining and asset management, multiple acquisition high-resolution optical and radar remote sensing, scene understanding, data fusion and Bayesian modeling.

Julián Flórez studied Industrial Engineering in the University of Navarra (1980), and obtained his Ph.D. in the University of Manchester, Institute of Science and Technology UMIST (1985), in the field of Adaptive Control. From 1985 to 1990, he worked as Researcher in the Centre of Study and Technical Research of Gipuzkoa (CEIT), where he collaborated in several research projects related to Electrical and Industrial Engineering with a marked industrial focus. From 1985 to 1994, he was Associate Professor in the School of Industrial Engineering of the University of Navarra, and since 1994, he is Professor at the same university. From 1990 to 1997, Dr. Flórez worked as Senior Researcher in CEIT, where he was in charge of a Department of Industrial Applications. From 1997 to 2001, he worked as Director of Corporate Development of Avanzit-SGT (Servicios Generales de Teledifusión) in the fields of Information Systems, Communications and Broadcasting infrastructure. He has a strong background in Digital Television infrastructures and was tightly involved in the deployment of one of the biggest Digital TV organizations in Spain and Europe: Quiero TV. Since 2001, he is Principal Researcher in Vicomtech. He holds some patents and has written more than 40 research papers in different areas of Industrial and Electrical Engineering.

Basilio Sierra is Full Professor in the Computer Sciences and Artificial Intelligence Department at the University of the Basque Country. He received his BSc in Computer Sciences in 1990, MSc in Computer Science and Architecture in 1992 and PhD in Computer Sciences in 2000 at the University of the Basque Country. He is the director of the Robotics and Autonomous Systems Group in Donostia-San Sebastian. Professor Sierra is presently a researcher in the fields of Robotics, Computer Vision and Machine Learning, and he is working on the development of different paradigms to improve classification behaviors; he has written more than 25 journal papers in those fields, as well as more than 100 conference contributions and more than 30 book chapters.