biblio.ugent.be

The UGent Institutional Repository is the electronic archiving and dissemination platform for all UGent research publications. Ghent University has implemented a mandate stipulating that all academic publications of UGent researchers should be deposited and archived in this repository. Except for items where current copyright restrictions apply, these papers are available in Open Access.

This item is the archived peer-reviewed author-version of:

Perceptual-Based Textures for Scene Labeling: a Bottom-Up and a Top-Down Approach

Gaëtan Martens, Chris Poppe, Peter Lambert, Rik Van de Walle

In: IEEE Proceedings of the 5th International Conference on Future Information Technology

ISBN: 978-1-4244-6949-9

Optional: link to the article

To refer to or to cite this work, please use the citation to the published version:

Gaëtan Martens, Chris Poppe, Peter Lambert, Rik Van de Walle. Perceptual-Based Textures for Scene Labeling: a Bottom-Up and a Top-Down Approach. IEEE Proceedings of the 5th International Conference on Future Information Technology. Busan, Korea. 20-24 May. 2010

Perceptual-Based Textures for Scene Labeling: a Bottom-Up and a Top-Down Approach

Gaëtan Martens, Chris Poppe, Peter Lambert, and Rik Van de Walle

Ghent University – IBBT – Multimedia Lab
Gaston Crommenlaan 8 bus 201
B-9050 Ledeberg-Ghent, Belgium
{gaetan.martens,chris.poppe,peter.lambert,rik.vandewalle}@ugent.be

***Abstract*— Due to the semantic gap, the automatic interpretation of digital images is a very challenging task. Both the segmentation and classification are intricate because of the high variation of the data. Therefore, the application of appropriate features is of utter importance. This paper presents biologically inspired texture features for material classification and interpreting outdoor scenery images. Experiments show that the presented texture features obtain the best classification results for material recognition compared to other well-known texture features, with an average classification rate of 93.0%. For scene analysis, both a bottom-up and top-down strategy are employed to bridge the semantic gap. At first, images are segmented into regions based on the perceptual texture and next, a semantic label is calculated for these regions. Since this emerging interpretation is still error prone, domain knowledge is ingested to achieve a more accurate description of the depicted scene. By applying both strategies, 91.9% of the pixels from outdoor scenery images obtained a correct label.**

I. Introduction

Over the last decades, the development of techniques for the digitization of visual information, together with the decreasing costs and increasing capacity of digital storage media, has led to an explosion of digital content. According to a study of the International Data Corporation (IDC), there will be a continuous growth of digital information over the next years [5]. Further, this study suggests that the biggest growth in data is visual in nature, from devices such as digital cameras, digital surveillance cameras, and digital televisions. To deal with this explosion of digital images in size and complexity, one of the main imperatives IT organizations face, is the need for new tools and standards for data search and analytics. In particular, there's an increasing need for the development of automated image content analysis and description techniques in order to retrieve images based on their visual content efficiently from large collections. In describing the visual content of images for content-based image retrieval, the use of primitive image features (e.g., color, edges, shapes) may not be sufficient due to the *semantic gap*. The semantic gap is a major discrepancy in content-based information retrieval. Smeulders et al. [21] describe the semantic gap as: "the lack of coincidence between the information that one can extract from the data and the interpretation that the same data have for a user in a given situation". This means that a user wants to retrieve data on a semantic level, but the characterizations can only provide a low-level similarity. In this context, it is particularly important to use content descriptors that are robust to the accidental variance introduced by the image creation process (e.g., the variation of the illuminant in visual data). Therefore, the application of appropriate features is of utter importance.

This paper focuses on the segmentation and interpretation of outdoor scenes and, herewith related, the recognition of materials. A major weakness of image retrieval systems is their lack of domain knowledge. Consequently, many systems are error prone when it comes to detection of high-level concepts. Most approaches to extract the semantics of scenery images using low-level features use an image partitioning as an intermediate step. Wang et al. [24] use a codebook to segment an image based on the statistics of the regions' color and texture features. At pixel level, color-texture classification is used to form the codebook. This codebook is in the next stage used to segment an image into regions. The context and content of these regions are defined at image level. Zhu et al. [26] partition the image into equally sized blocks and indexes the regions using a codebook whose entries are obtained from the features extracted from a block. The method of Li et al. [10] uses 2-dimensional hidden Markov models to associate the image and a textual description. Depalov et al. [4] use a quantized color and texture segmentation algorithm to segment images depicting natural scenes. The features of the obtained regions are used as medium level descriptors to extract semantic labels at region level and later at scene level. However, the use of quantized colors may result in weaker segmentations. Yuan et al. use spatial context constraints to label image regions [25]. Segmented image regions are first regularized into a 2-dimensional lattice layout to represent a graphical model for learning and inference. However, their learning is supervised and the parameters of the support vector machines and conditional random fields are estimated sequentially rather than simultaneously. Athanasiadis et al. [1] associate a region with a fuzzy set of candidate concepts stored in an ontological knowledge base. A merging process is performed based on new similarity measures and merging criteria that are defined at the semantic level with the use of fuzzy sets operations. Schober et al. [20] applied domain knowledge for the interpretation of landscape images. They relate the extracted low-level features with concepts and

then generate rules which define the coherences between the concepts. An ontology defines the spatial relations between the concepts to remove the incorrect assignments. A detailed overview of content-based image retrieval techniques which include semantics is given by Liu et al. [11]. They divide these methods into five categories: (i) employing ontologies to define high-level concepts, (ii) applying machine learning on low-level features, (iii) using relevance feedback to account for the users' action, (iv) generating semantic templates to assist high-level information retrieval, and (v) using both visual content and surrounding text. The approach presented in this paper uses techniques from the first and second category. We apply machine learning on perceptual textures and ingest domain knowledge.

Humans and primates outperform the best machine vision systems in many aspects. Humans are very good at getting the conceptual category and layout of a scene within a single fixation. So, building a system that emulates the recognition tasks of the cortex has always been a challenging and attractive idea. Next to color, the human visual system (HVS) is best trained to texture perception. Since texture is much more robust than color with respect to lighting conditions, it could play an important role in such kind of application. Indeed, Renninger and Malik already have concluded that a texture analysis provides useful information for rapid scene identification [19]. However, in computer vision the use of visual neuroscience has often been limited to a tuning of Gabor filter banks. No real attention has been given to biological features of higher complexity. In this paper, we propose a set of biologically inspired texture features for scene analysis. We present a bottom-up approach to link these low-level features to semantic concepts using both unsupervised and semi-unsupervised machine learning algorithms. In the first stage, the proposed features are used to segment an image into similar regions without supervision. In the second stage, a previously trained classifier is then used to label the obtained segments using a semi-supervised learning technique. In the last stage, a top-down approach is used to apply domain knowledge on the obtained labels which results in a more accurate scene description. The paper is organized as follows. In Sect. II, we briefly describe the computational model to calculate the biologically inspired texture features and Sect. III outlines the feature extraction and explains the data preprocessing. Section IV explains the consecutive steps of our methodology. Experiments on semi-supervised (material) classification and semantic labeling are then described in Sect. V. Finally, concluding remarks and some future work appear in Sect. VI.

II. Computational model

The computational model of the texture features we propose for classification is described in this section. At first, we take a closer look in Sect. II-A at the configuration of the Gabor filter which is at the basic level of our method. The model of Petkov and Kruizinga is briefly explained in Sect. II-B to compute enhanced grating cell responses. Finally, Sect. II-C considers the spatial smoothing of Gabor responses with regard to texture analysis.

A. Gabor filter

In the spatial domain, a Gabor function is a Gaussian modulated by a sinusoid. To model the receptive fields of simple cells in the visual cortex, the real part of the following family of 2-dimensional Gabor filters are used as proposed by Daugman [3]:

$$g_{\lambda,\theta,\varphi}(x,y) = \frac{\cos\left(2\pi\frac{x'}{\lambda}+\varphi\right)\exp\left(-\frac{1}{2}\left[\frac{x'^2}{\sigma_x^2}+\frac{\gamma^2 y'^2}{\sigma_y^2}\right]\right)}{2\pi\sigma_x\sigma_y} \tag{1}$$

where

$$\begin{cases} x' = x\cos\theta - y\sin\theta \\ y' = x\sin\theta + y\cos\theta. \end{cases}$$

The standard deviations σ_x and σ_y of the Gaussian factor determine the effective size of the surrounding of a pixel in which the summation takes place. A circular Gaussian is preferred so that there is a constant spatial extent in all directions, therefore $\sigma_x = \sigma_y (= \sigma)$. The parameter λ is the wavelength of the sinusoid, and the ratio σ/λ determines the bandwidth of the filter. Experiments indicate that the frequency bandwidth of simple cells is about one octave [17], thus $\sigma/\lambda \approx 0.56$. The spatial aspect ratio γ determines the eccentricity and herewith the eccentricity of the receptive field ellipse. According to Jones and Palmer [7], it has been found that γ varies in a limited range of $0.23 < \gamma < 0.92$ and is set to a constant value of 0.5. Further, the orientation of the filter is denoted by $\theta \in [0, \pi[$. This is the normal to the parallel lobes of the filter in the spatial-frequency domain, denoted by x' in equation (1). Finally, the phase offset φ affects the symmetry of the function. For $\varphi = 0$ or $\varphi = \pi$ the filter is symmetric while for $\varphi = \pi/2$ or $\varphi = -\pi/2$ the filter is anti-symmetric. The response of the receptive field function of a simple cell, tuned to orientation θ and frequency $1/\lambda$, to the luminance channel of an input image $I(x,y)$ is then given by:

$$r_{\lambda,\theta,\varphi}(x,y) = \iint I(s,t)\, g_{\lambda,\theta,\varphi}(x-s, y-t)\, ds\, dt. \tag{2}$$

B. Enhanced grating cell operator

Grating cells respond to bar gratings of a given orientation and periodicity, but not to single bars. In order to better distinguish the salient texture-specific periodicities and to obtain an improved texture discrimination, an enhanced image $\overline{I}(x,y)$ is created by applying a histogram equalization to the original input image $I(x,y)$. Histogram equalization is a well-known technique that rescales the range of the pixel values to produce an image whose pixel values are more uniformly distributed which results in an image with a higher contrast. In previous work, we found that applying a histogram equalization increases the performance of texture segmentation when using grating cell outputs [12].

To model the non-linear behavior of the grating cells, we make use of the model of Kruizinga and Petkov [9]. This

model first computes the output of a simple cell of the visual cortex $s_{\lambda,\theta,\varphi}(x,y)$, tuned to a specific orientation θ and frequency $1/\lambda$, to input $\overline{I}(x,y)$

$$s_{\lambda,\theta,\varphi}(x,y) = \begin{cases} 0 & \text{if } a_\lambda(x,y) = 0 \\ \chi\left(\frac{\frac{r_{\lambda,\theta,\varphi}(x,y)}{a_\lambda(x,y)}R}{\frac{r_{\lambda,\theta,\varphi}(x,y)}{a_\lambda(x,y)}+C}\right) & \text{otherwise,} \end{cases} \quad (3)$$

where the average gray value of the receptive field is given by

$$a_\lambda(x,y) = \iint \overline{I}(s,t) \exp \frac{(x-s)^2+\gamma^2(y-t)^2}{2\sigma^2} dsdt.$$

and R denotes the maximum response level, C is the semi-saturation constant, $\chi(t) = t$ for $t \geq 0$ and $\chi(t) = 0$ for $t < 0$.

This output is then used to calculate the activity of a grating subunit. A grating subunit will be activated if for the preferred orientation θ and spatial-frequency $1/\lambda$, the function $s_{\lambda,\theta,\varphi_n}$ is alternately activated in intervals of length $\lambda/2$ for $n = -3,-2,..,2$ and this along a line segment of length 3λ centered on point (x,y). In other words, a grating subunit is thus activated if at least 3 parallel bars with spacing λ and orientation θ of the normal to them are encountered. In the final stage, the response of the grating cell operator $w_{\lambda,\theta}$ is obtained by summing up the grating subunits for a given θ and λ. The operator is made symmetric by considering the opposite direction $\theta + \pi$. Using $\overline{I}(x,y)$ as input, we obtain the enhanced grating cell operator $\overline{w}_{\lambda,\theta}$. For more details, we refer to [9].

C. Spatial smoothing

Textures which do not have sufficiently narrow bandwidths may suffer from leakage. The effects of leakage can be reduced by post-filtering the channel amplitudes with Gaussian filters having the same shape as the corresponding channel filters but greater spatial extents. Therefore, smoothed Gabor responses are known to improve the performance for texture analysis [2]. There exists a physiological reason for utilizing smoothing since it mimics characteristics of the HVS. Hall and Hall [6] describe the existence of sustained channels in the visual system, indicating that the HVS not only considers pixels in the field of view, but also pixels in the vicinity.

The spatially smoothed Gabor responses we use, are obtained by convolving symmetric Gabor responses with a Gaussian with standard deviation $\sigma' = 2\sigma$:

$$\tilde{r}_{\lambda,\theta} = [r_{\lambda,\theta,0} * gauss](x,y) \quad (4)$$

where

$$gauss(x,y) = \frac{1}{2\pi\sigma'^2} \exp\left(-\frac{x^2+y^2}{2\sigma'^2}\right).$$

III. Features

The texture features consist of enhanced grating cell features $\overline{w}_{\lambda,\theta}$ and the spatially smoothed Gabor responses $\tilde{r}_{\lambda,\theta}$. The frequencies for the filters are $\sqrt{2}, 2\sqrt{2}, 4\sqrt{2}, 8\sqrt{2}$, and $16\sqrt{2}$ cycles per image and we use 8 orientations (θ=0, $\frac{\pi}{8}$,..,$\frac{7\pi}{8}$), what results in an 80-dimensional texture feature vector. Since $\overline{w}_{\lambda,\theta}$ have a different range than $\tilde{r}_{\lambda,\theta}$, scaling of the feature vectors is of special importance, otherwise bigger variables tend to dominate the others. Therefore, normalization is required.

IV. Methodology

Our approach consists of 3 steps. Instead of directly assigning a label to each pixel, we first apply an intermediate segmentation step. Based on the perceptual texture, the image is segmented into similar regions using no supervision (Sect. IV-A). Secondly, the texture features of the region are used for material identification in order to obtain a label (Sect. IV-B). Finally, we apply domain knowledge on the computed intermediate results to achieve more accurate image descriptions (Sect. IV-C). Thus, our methodology employs two strategies:

1) a bottom-up strategy to compute semantically relevant information from the low-level image data (IV-A and IV-B),
2) a top-down strategy that ingests domain knowledge to increase the accuracy of the obtained interpretation (IV-C).

A. Segmentation

To segment a scenery image into regions, we make use of a Self-Organizing Map (SOM). The SOM is a single layer artificial neural network that simulates the process of unsupervised self-organization with a simple, yet effective numerical algorithm [8]. There exists a lot of neurophysiologic evidence to support the idea that the SOM captures some of the fundamental processing principles of the human (both visual and auditory) cortex. An important property of the SOM is that it clusters similar data vectors and projects dissimilar ones far from each other on the map. A SOM includes a grid of nodes and each node is associated with a parametric real vector, called the model vector. For a given input, the model vectors are updated according to the following rule: (i) find the best matching unit (BMU) using a predefined metric, and (ii) change the model vectors in a neighborhood of the BMU (the size of the neighborhood is a decreasing function of time). The computed texture features are used to train a 10×10 SOM. As a result of this training process, pixels which belong to the same texture, are assigned to the same or adjacent nodes. For more details about unsupervised image segmentation, we refer to our previous work [12], [13].

B. Semi-supervised Classification

In the second step, the obtained regions are assigned a label using a semi-supervised learning algorithm. To classify the extracted feature vectors, we make use of a hierarchical variant of the SOM. At first, a 2-dimensional SOM of a predefined size is trained using a labeled training set. The labels of the training data correspond to the semantics the data describe. However, we have experienced that when some textures are relatively similar to each other compared to several other textures in the training data, it is possible that these textures

are not distinguished by the SOM and, consequently, they are assigned to the same node. To tackle this issue, we employ a hierarchical approach utilizing the labels of the training data as some means of supervision. The training vectors associated with such a node are used to train a new, smaller SOM. This process is iteratively repeated until a certain stopping criterion is reached or no progress in the classification is obtained. Suppose that a node N is related to a set V of training vectors v of k classes c_i, $i = 0..k-1$: $N \leftarrow V = \{v_{c_0,1} \,..\, v_{c_0,m_0}, \,..\, v_{c_j,1} \,..\, v_{c_{k-1},m_{k-1}}\}$ where m_i denotes the number of training vectors of texture c_i assigned to N. The stopping criterion is defined by a threshold $0 < \tau \leq 1$:

$$\frac{\max_{i=0..k-1}\{m_i\}}{\sum_{i=0}^{k-1} m_i} \geq \tau \tag{5}$$

If (5) isn't satisfied, V is used to train a new SOM. This process is then repeated for the nodes of the resulting SOM. The label of an unknown test sample is easily obtained by calculating its BMU. If the BMU is an empty node (no vectors were assigned to this node during the training phase), the label of its closest node is used. The parameter τ is empirically set to 0.95 and the dimensions of the SOM for K texture classes are chosen as follows: 4×4 for $K = 2$ or 3, 8×8 for $K = 4$, 10×10 for $K = 5$. The label of an image region is then easily computed by calculating the BMU of each pixel of the region and then assigning the label with the highest count.

C. Domain Knowledge

Describing high-level concepts with low-level features is a challenging task and, consequently, any bottom-up approach is error prone. In order to cope with errors and to obtain a more plausible interpretation of the depicted scene, we ingest domain knowledge. Therefore, an ontology is created that describes the conditions and restrictions of the depicted concepts (i.e., the concepts that are present in the training set of the previous step). In this way, image regions can be merged or misclassifications and illogical compositions can be removed or altered. However, one should pay special attention to the knowledge modeling phase to avoid false rules or rules that are not universally applicable within the given context.

The created ontology consists of a few simple, but effective rules. Given the concepts of our training set, the following rules are iteratively applied:

1) neighboring regions with the same label, are merged
2) no blob can exist in *sky*
3) no blob of *sky* can exist in *water*
4) no *water* can be above *sky*
5) a region should be at least 8 pixels wide and high
6) regions that not obey to the above rules, are relabeled (the new label is obtained by the ith BMU of the region, where i denotes the iteration number)

V. Experiments

To test the material classification and the image labeling, we create a training set consisting of 100 real-life textures which are manually collected from the World Wide Web (WWW)[1]. Each collected texture belongs to one of these five classes: (i) *Branches*, (ii) *Bricks*, (iii) *Grass*, (iv) *Sky*, and (v) *Water*. Every class contains 20 samples. Figure 1 depicts some example textures from the training set. We carry out

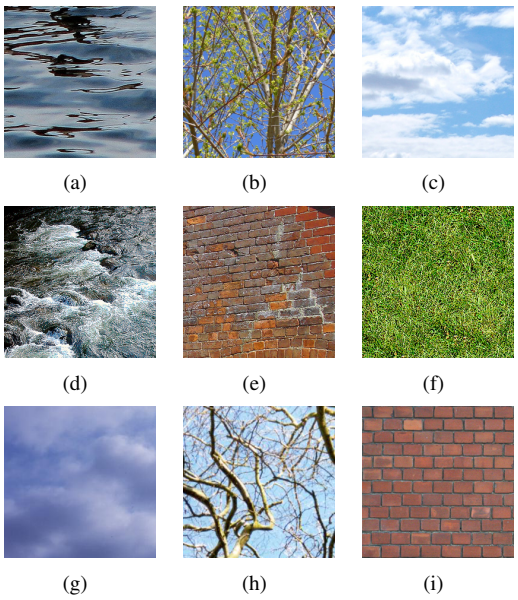

Fig. 1. Examples of real-life textures: water (a, d); branches (b, h); sky (c, g); bricks (e, i); grass (f).

two type of experiments. At first, we test the performance of our SOM-based classification for material detection (Sect. V-A) using the texture features presented in Sect. III. In the second experiment, we apply semantic labeling of scenery images (Sect. V-B). Further, we use the SOM Toolbox [22] to create the SOMs, and the Euclidean distance is employed as distance metric. Remark that the ground-truth images are created manually and therefore they should be interpreted as an approximation rather than a certainty.

A. Material classification

In this experiment, we test the proposed texture features for material classification. The performance and robustness of our SOM-based classification is analyzed using the proposed texture features which are obtained from the presented texture set. From each texture class, we leave out one image that we use as test image, the other images are then used to train the classifier. This process is repeated for every image (cross-validation).

We also tackle this classification problem using GMRF, multi-scale LBP, and Gaussian smoothed Gabor features (see equation 4). The implementation of the GMRF features is obtained from the MeasTex site [14] and the features are computed using the standard symmetric masks. The 73-dimensional GMRF feature vectors are obtained by concatenating feature vectors of GMRF models of order 1 to 7.

[1]The texture set is available at the following location: http://www.mmlab.be/users/gmartens/textures

The multi-scale LBP, i.e. $LBP^{riu2}_{(8,1)+(16,2.4)}$, are uniform and rotation invariant and the 2-dimensional co-occurrence LBP histograms are classified using a non-parametric *L*-statistic as proposed by [15], [16].

As can be seen in Table I, the proposed features achieve the best classification results. Also the smoothed Gabor responses obtain good results. However, the multi-scale LBP have some difficulties with the *Water* and *Grass* textures because of the high variation of the data. Our experiments further indicate that GMRF features are clearly not designed for this kind of task. The GMRF have major difficulties to discriminate textures from the class *Water* and *Sky*, and textures from the class *Grass* and *Branches*.

TABLE I
MATERIAL CLASSIFICATION RATE (%)

	proposed	GMRF	LBP	smoothed Gabor
Branches	96.8	81.6	72.2	87.6
Bricks	98.5	26.5	86.11	81.9
Grass	87.7	15.5	11.1	85.8
Sky	95.9	99.9	100	95.3
Water	85.9	2.3	16.6	85.7
average	93.0	45.2	57.2	87.3

B. Labeling of image regions

In this experiment, we test the labeling of image regions of outdoor scenes using the proposed method and texture features as exemplified in Fig. 2. The labeling experiments are conducted on multiple scenery images collected from the WWW containing no other texture classes than those in our training set. After the segmentation process, our trained classifier is then used to label the computed image regions. As can be seen in Fig. 2(c) and Fig. 2(g) the result of this labeling process contains different errors. Some isolated pixels are misclassified. These errors can be removed by applying constraints on the size of the regions. However, some errors are due to reflection, e.g., the *sky*-blob in the lake of Fig. 2(c). Such type of errors can only be removed by incorporating domain knowledge. Other errors emerge from the fact that the scaling of certain textures alters due to changes in the perspective, e.g., at the edges of the mountains in Fig. 2(c). Generally, this problem is harder to tackle. In the last step, the ontology of Sect. IV-C is applied on the intermediate results. As can be seen in Fig. 2(d) and Fig. 2(h), small misclassified regions are filtered away, and the blob of *sky* in the lake is removed. After the semi-supervised classification step, 82.7% of the pixels have a corrected label. However, employing the domain knowledge enhanced the region labeling with 9.2% up to 91.9%.

VI. CONCLUSIONS

Due to the semantic gap, the automatic interpretation of images is an intricate task. In this paper, we have used a bottom-up and top-down approach for the labeling of regions from outdoor scene images. The bottom-up approach consists of two consecutive steps: (i) the extraction of features that are related to human visual perception and, (ii) the application of machine learning algorithms for unsupervised image segmentation and semi-supervised region labeling. For these tasks, we presented the use of biologically inspired texture features that correspond to cell outputs from the human visual cortex. We have shown in our experiments that these texture features obtain the best classification results for material recognition compared to other well-known texture features, with an average classification rate of 93.0%. This indicates that the application of the proposed biologically inspired features is very useful for material classification and scene interpretation. This bottom-up approach already labeled 87.9% of the pixels in scenery images correctly.

In the final stage, a top-down approach is applied. By ingesting an ontology which describes the conditions and restrictions of the considered concepts, the labeling results obtained from the bottom-up approach can be corrected. This resulted in a final classification rate of 91.9%, which means a gain of 9.2% compared to the output of the semi-unsupervised classifier.

Despite the good results, some improvements can still be made. At first, the segmentation step should be enhanced. In our experiments, we have noticed that some region boundaries are not correctly detected. The latter could be tackled by also taking edge information into account or, by using color information as well. Indeed, since color is the primary visual stimulus, we expect that introducing color information, next to texture information, could also increase the classification rate. On the other hand, in the HVS bar and grating cells seem to play an important role in boundary detection [9], [23]. In contrast to grating cells, bar cells respond only to an isolated edge or line but does not respond to any texture edge. Hence, it is possible to distinguish between an edge belonging to a textured region and a non-textured region. Another major difficulty is related to the altering of the scale of the textures due to the perspective. Since the HVS already performs feature extraction on different scales, a scale robust classifier might be inevitable.

ACKNOWLEDGMENT

The research activities as described in this paper were funded by Ghent University, the Interdisciplinary Institute for Broadband Technology (IBBT), the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT), the Fund for Scientific Research-Flanders (FWO-Flanders), and the European Union.

REFERENCES

[1] T. Athanasiadis and V. Tzouvaras and K. Petridis and F. Precioso and Y. Avrithis and Y. Kompatsiaris, Using a Multimedia Ontology Infrastructure for Semantic Annotation of Multimedia Content, Proc. of 5th International Workshop on Knowledge Markup and Semantic Annotation (SemAnnot '05), Galway, Ireland, 2005.

[2] A.C. Bovik, M. Clark, and W.S. Geisler, Multichannel Texture Analysis Using Localized Spatial Filters, IEEE Transactions on Pattern Analysis and Machine Intelligence, 12 (1) pp. 55–73, 1990.

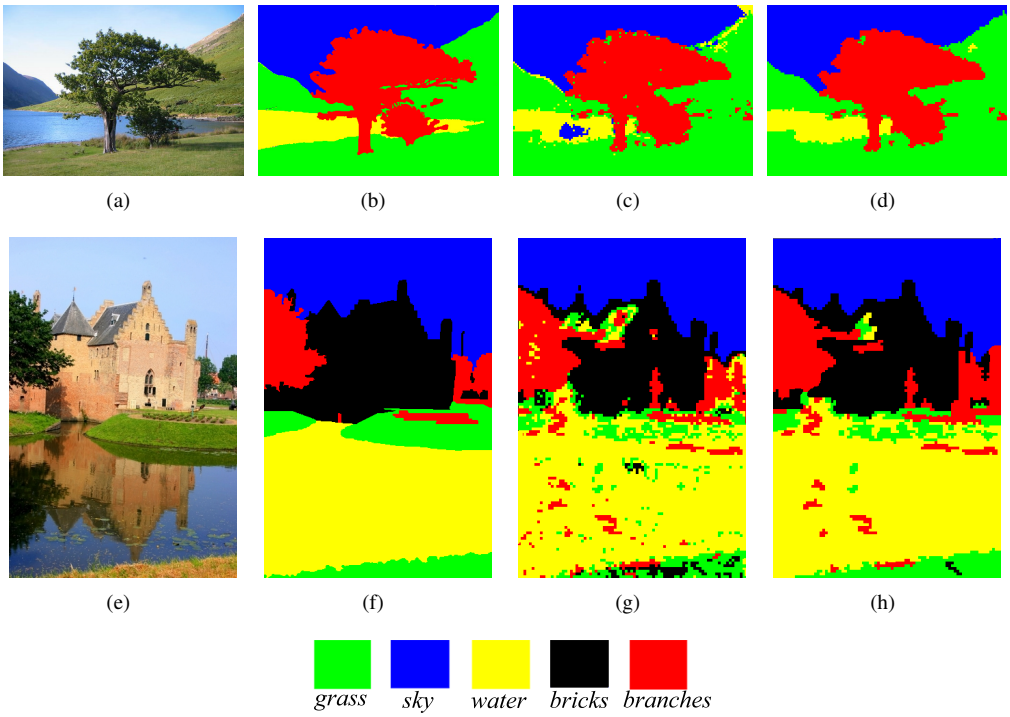

Fig. 2. Scenery image (a, e); ground truth (b, f); labeled regions (c, g); obtained regions after ingestion of domain knowledge (d, h); legend (i).

[3] J. Daugman, Uncertainty Relation for Resolution in Space, Spatial Frequency and Orientation Optimized by Two-Dimensional Visual Cortical Filters, *J. of the Optical Society of America A: Optics, Image Science, and Vision*, 2, pp. 1160–1169, 1985.

[4] D. Depalov, T. N. Pappas , D. Li, and B. Gandhi, Perceptually Based Techniques for Semantic Image Classification and Retrieval, *Proc. Human Vision and Electronic Imaging*, SPIE Conference, 2006.

[5] J. F. Gantz, C. Chute, A. Manfrediz, S. Minton, D. Reinsel, W. Schlichting, A. Toncheva, *The Diverse and Exploding Digital Universe*, IDC White Paper, March 2008.

[6] C.F. Hall, E.L. Hall, A nonlinear model for the spatial characteristics of the human visual system, IEEE Trans. Systems Man Cybern., 7 (3) pp. 161–170, 1977.

[7] J. Jones and A. Palmer, An Evaluation of the Two Dimensional Gabor Filter Model of Simple Receptive Fields in Cat Striate Cortex, J. of Neurophysiology, 58, pp. 1233–1258, 1987.

[8] T. Kohonen, Self-organizing Maps, Springer-Verlag, Berlin, Germany, 1997.

[9] P. Kruizinga and N. Petkov, Nonlinear operator for blob texture segmentation, in: Proc. NSIP'99, IEEE Workshop on Nonlinear Signal Processing, 2, A.S. Cetin, et al. (Eds.), pp. 881–885, 1999.

[10] J. Li, J. Wang, Automatic Linguistic Indexing of Pictures by a Statistical Modeling Approach, *IEEE Trans. Pattern Anal. Machine Intell.*, 25, 2003.

[11] Y. Liu, D. Zhang, G. Lu, and W. Ma, A survey of content-based image retrieval with high-level semantics. Pattern Recogn. 40 (1) pp. 262–282, 2007.

[12] G. Martens, C. Poppe, R. Van de Walle, Enhanced Grating Cell Features for Unsupervised Texture Segmentation, Performance Evaluation for Computer Vision: 31ste AAPR/OAGM Workshop 2007, Osterreichische Computer Gesellschaft, Performance Evaluation for Computer Vision, pp. 9–16, 2007.

[13] G. Martens, C. Poppe, P. Lambert, R. Van de Walle, Unsupervised Texture Segmentation and Labeling Using Biologically Inspired Features, in: Proceedings of the 2008 IEEE 10th workshop on Multimedia Signal Processing, IEEE Signal Processing Society, pp. 159–164, 2008.

[14] MeasTex Image Texture Database and Test Suite, Version 1.1 [Online]. Available: http://www.texturesynthesis.com/meastex/meastex.html.

[15] T. Ojala, M. Pietikäinen and T. Mäenpää, Multiresolution gray-scale and rotation invariant texture analysis with local binary patterns, IEEE Transactions on Pattern Analysis and Machine Intelligence, 24 (7) pp. 971–987, 2002.

[16] T. Ojala, K. Valkealahti, E. Oja, and M.M. Pietikäinen, Texture Discrimination with Multidimensional Distributions of Signed Gray Level Differences, Pattern Recognition 34 (3) pp. 727–739, 2001.

[17] D.A. Pollen and S.F. Ronner, Visual cortical neurons as localized spatial frequency filters, IEEE Trans. Systs, Man, and Cybern, 13 (5) pp. 907–916, 1983.

[18] A. Popescu, C. Millet, and P. Moëllic, Ontology driven content based image retrieval, In Proceedings of the 6th ACM international Conference on Image and Video Retrieval, Amsterdam, ACM, pp. 387–394, 2007

[19] L.W. Renninger, J. Malik, When Is Scene Recognition Just Texture Recognition?, *Vision Research*, 44, pp. 2301-2311, 2004.

[20] J.P. Schober, T. Hermes, and O. Herzog, Content-based Image Retrieval by Ontology-Object Recognition, Proceedings of the KI- Workshop on Applications of Description Logics (ADL-2004), Ulm, Germany, September 2004.

[21] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, R Jain, *Content-Based Image Retrieval at the End of the Early Years, IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22 (12) pp. 1349-1380, 2000.

[22] J. Vesanto, J. Himberg, E. Alhoniemi and J. Parhankangas, Self-Organizing Map in Matlab: the SOM Toolbox, in: Proceedings of the Matlab DSP Conference, pp. 35–40, 2000.

[23] R. von der Heydt , E. Peterhans, and M.R. Dürsteler, Periodic-pattern-selective cells in monkey visual cortex, Journal of Neuroscience, 12 (4) pp 1416-1434, 1992.

[24] W. Wang, Y. Song, and A. Zhang, Semantics Retrieval by Content and Context of Image Regions, *Proc. of the 15th International Conference on Vision Interface*, pp. 17–24, 2002.

[25] J. Yuan, J. Li, and B. Zhang, Exploiting Spatial Context Constraints for Automatic Image Region Annotation, *6th ACM International Conference on Multimedia*, pp. 595–604, 2007.

[26] L. Zhu, A. Zhang, A. Rao, and R. Srihari, Keyblock: An Approach for Content-Based Image Retrieval, *ACM Multimedia*, pp. 157–166, 2000.