Extraction of emotional impact in colour images

Syntyche Gbèhounou, François Lecellier and Christine Fernandez-Maloigne; University of Poitiers, Department SIC of XLIM Laboratory, UMR CNRS 7252; Poitiers, France

Abstract

This paper proposes a method to extract the emotional impact of images.
Emotions are often associated with facial expressions, but we decided consider other features as first emotional characteristic of natural images, which, in general, does not contain faces. For a seek of generally we have built a new image database composed of a large variety of low semantic images. We used colour images because often colours and emotions are supposed to be linked.
For the modelling of the emotions, we considered colours features completed with other recent and efficient descriptors. We supposed that different features used could also implicitly encode high-level information about emotions. The concept of emotion is not easy to model. The perception of emotion is not only influenced by the content and the colour of the images. It is also modified by some personal experiences like cultural aspects and personal semantic associated to some colours or objects.
The complexity of emotion modelling was considered in classification process through psycho-visual tests. The twenty-five observers assessed the nature and the power of emotions they felt. These tests allowed us to distinguish three classes of emotions, which are "Negative", "Neutral" and "Positive" emotions.
We used a Support Vector Machine for classification and the average success rate is 51,75%; that is really relevant regarding the equivalent results in the literature.

Introduction

Images are known as emotional vectors, that's why they are the preferred media through advertising. We can easily understand that they are one of the most powerful tools to transmit information. Try to extract the emotional impact in images is an ambitious task, since different information contained in an image (content, texture, colours, semantic, ...) can be emotional vector. There are many *a priori* about different relations between the images and their emotional content particularly about colours and their impact. Furthermore the emotion of an image is strongly linked with its content because some places (beach, cemetery, ...) refer to different emotions. Other factors, including cultural aspects, more complex than the content and the overall colour are considered in the emotional interpretation of an image.
The emotion can be defined as psychological state that arises spontaneously rather than through conscious effort and is sometimes accompanied by physiological changes; a feeling. There are many other definitions of emotion according to the different schools of psychology. In fact, the concept of emotion is used in different ways as it is considered in reference to the stimulus aspect, the subjective experience, a phase of process, an intermediate variable or a response. Nevertheless, in the literature there are two major psychological theories of the emotion[2]:

- Theories of basic emotions,
- Theories of evaluation.

In the theories of basic emotions supported among others by Darwin, Ekman, Izard, Plutnick [16, 17], basic or fundamental emotions are listed. Note that just five are common to the different authors, which are: sadness, anger, happiness, disgust and fear.
In the theories of evaluation also called theories of appraisal, emotions are extracted from our evaluations of events that cause specific reactions in different people [16].
Whatever the theory of emotion, usually two methodologies of emotion classification are found [2, 17]:

- Discrete approach,
- Dimensional approach.

In the discrete approach, emotional process can be explained with a set of basic or fundamental emotions, innate and common to all human. There is no consensus about the nature and the number of these fundamental emotions [2].
The models of dimensional approach are different from those of the discrete approach by the fact that the emotions are the result of fixed number of concepts represented in a dimensional space [2]. The dimensions can be an axis of pleasure, arousal and power. These dimensions vary depending the needs of the model. The most used dimensional model is Russel's with the dimensions valence and arousal:

- ***The valence*** corresponds to the way a person feels when she looks such as a picture. This dimension varies from negative to positive and allows to distinguish between negative emotions and pleasant.
- ***The arousal*** represents the activation level of the human body.

The advantage of these models is to define a large number of emotions. Despite this, some emotions can be confused (such as fear and anger in Russel model) or unrepresented (among others surprise in Russel model).
In the literature, a lot of works are based on the discrete modelling of the emotions, for example those of Wei and al.[20], Paleari and Huet[15], Kaya and Epps[7].
Several approaches are developed to extract emotion of the images. They usually consider among other techniques a face detection step. An emotion is then associated with facial features (such as eyebrows, lips).
The other major approach is the detection of emotions from the characteristics of the image [20, 14, 11, 2]. The most used features are:

- Colours,
- Textures,
- Shapes (facial detection, animals detection ...).

Lucassen and al.[11] have defined four emotions related to textures that are:

- Warm-Cool,
- Masculine-Feminine,

- Hard-Soft,
- Heavy-Light.

From the results obtained from the analysis of the variance and the previous studies made by Ou and al.[14] (in which there are similarities in the emotional scales), Lucassen and al.[11] have implemented functions for an explanation of variances on the scale of colour emotion. They have established links between their different emotions and colour parameters L*, c* and h*.
One the other side, Wang and Yu [19], use the semantic description of colours. With this information, they associate an emotional semantic with an image.
The orientation of the different lines contained in the images is sometimes considered. According to Dellandréa and al.[2], oblique lines could be associated with dynamism and action and horizontal and vertical lines with calm and relaxation.
In view of the literature, we chose the approach used images features in association with subjective evaluations. In order to consider the different aspects (cultural influence, personal experience), which are not easy to model, we decided to organize psycho-visual tests. They allowed us to have a reference for our emotional classifier of colour images.
This paper is organized as follow. We describe the image database and the different psycho-visual tests organized in part 2. The part 3 is about characteristics of the image used for emotion recognition. The classification process is explained in part 4. Finally, in part 5 we summarize our study and we provide some perspectives.

Image database and psycho-visual tests

Image database

There is no universal image database in the domain of emotion extraction. Then, it is very important to build a diversified database. Our database is composed of 218 images free to use for research and publication. It is composed of landscapes, animals, food and drink, historic and touristic monuments as shown in figure 1. We also applied geometric transformations and changes

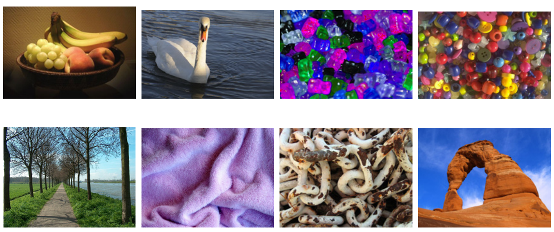


Figure 1. *Some images from our database*

in the colour balance on some images of the database.
We chose to organize psycho-visual tests in order to associate a subjective response with the emotional content of the images.

Psycho-visual tests

Twenty-five subjects participated in the experiments, 28% women and 72% men. Half were aged 18 to 24 years. They took part on voluntary basis and did not receive a financial reward.
The test strategy is closely linked to the model of emotions we have chosen. We decided to work with a dimensional model which dimensions are:

- The nature of the emotion,
- The power of the emotion.

Our emotion modelling is equivalent to Valence/Arousal model in which, the valence allows to distinguish positive and negative emotions and the arousal that varies from low to high defines the arousal body. This parameter is to describe the intensity of the emotion associated with the nature of emotion chosen.
We have organized two series of tests conducted at weekly intervals. During the different tests, observers evaluated the nature and the power of the emotion aroused by the image. As shown on the figure 2, for the nature of the emotion, the participants had choice between "Negative", "Neutral" and "Positive". The power is from low to high.

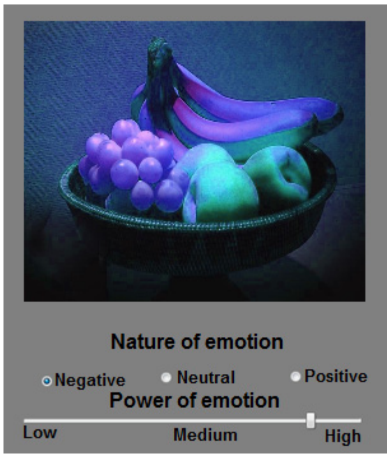
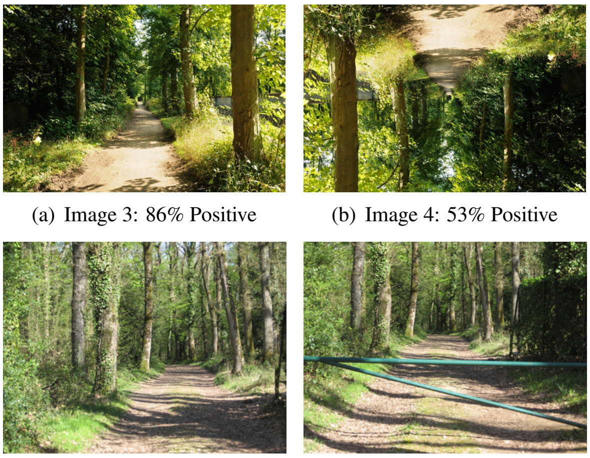


Figure 2. *Application of psycho-visual tests*

During each test run, subjects evaluated 24 images. So 48 images were evaluated. Observers had 8 seconds to score each image. Evaluation time is short in order to increase the chances of having primary emotions not emotions provided by the semantic interpretation of the images.
In both two series of tests, some images have the same content but modified with some transformations. We applied rotations and changes in colour balance. Even if they change the natural aspects of some pictures, they allowed us to see the results of some invariants features used to recognise the emotion.
There are very few human faces (6,25%) in our tests images. It was a way to avoid the interpretation associated with them.
Emotions are very complex to model. A little detail (figures 3(c) and 3(d)) or a rotation (figures 3(a) and 3(b)) can change its nature. For a significant number of observers, the barrier means nocrossing. The results of psycho-visual tests were used to build a reference for our classifier. 59% of the images evaluated during the tests form the learning basis. Only images with a classification rate in a category of emotion 10% higher than others in this category, are considered classified in this one. If the classification of an image by subject in two different categories of emotions is 50% (In fact we did not meet this case), it will be remove from the database of classification.

Different descriptors used for emotions recognition

Colours are the first discriminated characteristic of images for the extraction of the emotions. Many works in psychology make hypothesis about relationship between colours and emotions. For example, according to Daniel Beresniak [1], red is vibrant and exciting, by against the combination of "red + grey" causes a tragic sense. Yellow is the most cheerful colour, the most clear, radiant and youthful. He says that the most dynamic colour of is orange. This colour combines the cheerfulness of yellow and the action of red. The blue colour is deep and mysti-



(a) Image 3: 86% Positive (b) Image 4: 53% Positive

(c) Image 1: 96% Positive (d) Image 2: 65% Positive

Figure 3. *Power of the emotion on natural pictures(a)-(c) and transformed (b)-(d)*

cal call for calm.
Despite these assumptions, no consensus exists about the link between colours and emotions. Often, colours reflect the interpretation of the semantic linked to some situations, phenomena and also culture. Colours are not the only emotional information carried by an image.
Textures are also important for emotional analysis of an image. For example, a grid regardless of its colour has a semantic of confinement. We have finally supposed that local and global descriptors could also implicitly encode high-level data because of their accuracy.

Colours

To identify the different colours in the images, we have used colour segmentation by region growing [4]. For the initialisation of the seeds, we performed an analysis of greyscale histogram.
The analysis of a histogram was made in greyscale to save time in homogeneous areas. The seeds are the maxima of the greyscale histogram. The region growing was performed in CIE Lab colour space in order to have a Euclidean distance correlated with the perceptual distance. We have retained only the average colour of different regions.

Textures

For texture extraction we converted images to greyscale. We have used two kinds of texture descriptors.
The first texture descriptors we have performed are the energy of Gabor features based on Gabor filters [5] which are directly related to Gabor wavelets.
The two-dimensional Gabor filter is defined by the function $g_{\lambda,\Theta,\varphi}(x,y)$ as the multiplication of a cosine/sine (even/odd) wave with a Gaussian windows, as follows:

$$g_{\lambda,\Theta,\varphi}(x,y) = \cos\left(2\pi\frac{x'}{\lambda}+\varphi\right)\exp\left(\frac{-(x'^2+\gamma^2y'^2)}{2\sigma^2}\right), \quad (1)$$

with $x' = x\cos\Theta + y\sin\Theta$ and $y' = y\cos\Theta - x\sin\Theta$.
For Gabor features, we considered 12 different angles $\Theta \in [0,\pi[$ every $\frac{\pi}{12}$ and 2 phases $\varphi \in \{0, -\frac{\pi}{2}\}$ (0 symmetric case and $-\frac{\pi}{2}$ asymmetric). So there are 24 filters.
We chose an isotropic Gaussian $((\gamma = 1))$ with standard deviation $\sigma = 0,56\lambda$. This choice is justified by the properties of the visual cortex that can be model with Gabor filter as said Grigorescu and al. [5].
The energy of Gabor features is the combination of the results of the 12 filtering for each phase.

The second type of texture descriptors we used is the Wave Atoms introduced by Demanet and Ying [3]. The Wave Atoms are in first approximation a variant of 2D wavelet packets with a parabolic wavelength scale.
Like all multi-scale transforms (the wavelet transform, for example), there are several information from different levels.
The number of coefficients for each orientation depends on the decomposition level. Before applying Wave Atoms transform we resized all image to $256*256$ with zero padding. With this new size, we had 5 levels of decomposition. We just work with the scale 4 composed of 91 orientations. Each orientation contains $2^4 * 2^4 (256)$ coefficients.

Local and global descriptors

We decided to use local descriptors SIFT introduced by Lowe [9, 10] which efficiency has been demonstrated in numerous papers.
The image is featured by a set of vectors of local invariant features to:

- translation,
- image scaling,
- rotation,
- change in illumination,
- affine or 3D projection.

We also performed OpponentSIFT proposed by Van de Sande and Snoek [18]. These local descriptors are one of the different colour extensions of the greyscale SIFT. They are recommended when no prior knowledge about the dataset is available.
OpponentSIFT describes all the channels in the opponent colour space (eq. 2) using SIFT descriptors.

$$\begin{pmatrix} 0_1 \\ 0_2 \\ 0_3 \end{pmatrix} = \begin{pmatrix} \frac{R-G}{\sqrt{2}} \\ \frac{R+G-2B}{\sqrt{6}} \\ \frac{R+G+B}{\sqrt{3}} \end{pmatrix} \quad (2)$$

The information in the O_3 channel is equal to the intensity information, while the other channels describe the colour information in the image. These other channels contain some intensity information, but due to the normalization of the SIFT descriptors they are invariant to changes in light intensity. Before applying OpponentSIFT algorithm, we applied the Harris-Laplace point detector to the image because it has shown good performance for category recognition [21].

For global description of the scene, we used "GIST" which allowed us to have a low dimensional representation. Oliva and Torralba proposed them in 2001 [13].
These descriptors are obtained with a set of perceptual dimensions (naturalness, openness, roughness, expansion, ruggedness) that represent the dominant spatial structure of a scene. These dimensions are estimated using spectral and coarsely localized information.

Classification process

The classification is the last step of our method of emotional impact extraction. It is composed of two phases:

- ***The diminution of the length of features vectors***: it is very important to reduce the dimensions of the different vectors.

After the extraction, the features can not be used because they are too large.

- ***Images classification***: we used a SVM [6] for our classification. For all categories of descriptors we chose a linear kernel.

Reduction of the dimensions of features vectors

As we said, after characteristics extraction, the vectors are too large. This is confirmed in the table 1 which gives the average size of attributes for one image.

The average size of attributes for one image before the reduction of the dimensions

Characteristics	Average number of descriptors	Size of vectors
Colours	151	3
Gabor	47 648	12
GIST	960	1
OpponentSIFT	500	384
SIFT	469	128
Wave Atoms (Scale 4)	91	256

So characteristic vector dimension reduction is very important to improve classification time.
Before performing the diminution of the size of the different vectors, we applied them a L2-Norm. We have a lot of solutions to reduce the length of our set of features but they are not all usable for emotions extraction. For example, it would be very difficult to use a Principal Component Analysis (PCA). Some features (Colours and OpponentSIFT) have not the same length for all images. Then, we chose a vectorial quantization with a *k-means* algorithm.
The relationship between the number of centroids and the number of descriptors is given by:

$$k = \sqrt[4]{N*d}\ , \tag{3}$$

with k the number of centroids, N the number of descriptors and d the size of vector of characteristics. Then, the set of centroids becomes k vectors of size d.
The dimensions of centroids are summarized in the table 2.

Configuration of the different centroids after *k-means* algorithm

Categories of descriptors	Number of centroids	Size of vectors
Colours	13	3
Gabor	72	12
GIST	15	1
OpponentSIFT	55	384
SIFT	41	128
Wave Atoms (Scale 4)	32	256

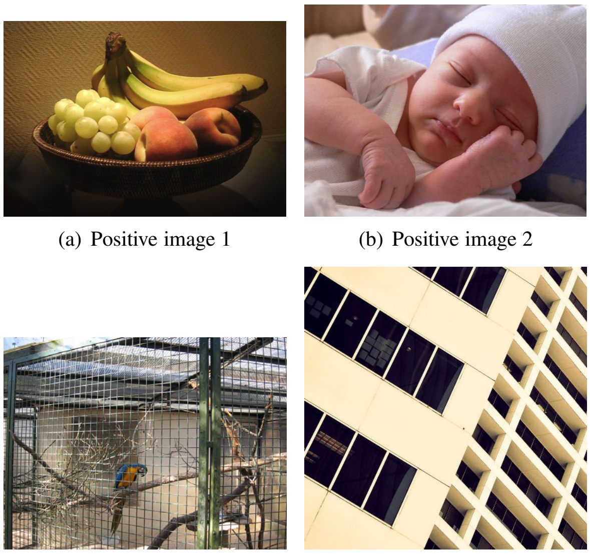
For classification we used image distribution histograms. Like the different vectors of features, the centroids and the histograms were L2-normalized.

Images classification

As we said in the introduction, the psycho-visuals tests were done to form a reference for our classifier. 59% of the 48 images scored during the tests (28 images) constitutes the learning basis. Depending the category of the features, learning is not without errors as shown in table 3, since the different images of the same class of emotion can be different both by the content and the colour or the texture. This diversity is illustrated on figure 4 with histogram of Wave Atoms distribution (figure 5).

Classification rates on learning database

Categories of descriptors	Numbers of errors	Classification rate
Colours	3	78,57%
Gabor	7	89,28%
GIST	6	78,57%
OpponentSIFT	0	100%
SIFT	0	100%
Wave Atoms (Scale 4)	9	67,86%



(a) Positive image 1 (b) Positive image 2

(c) Negative image 1 (d) Negative image 2

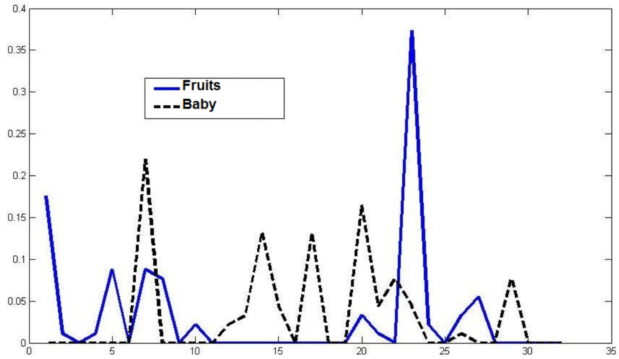
Figure 4. *Examples of learning images*

Results

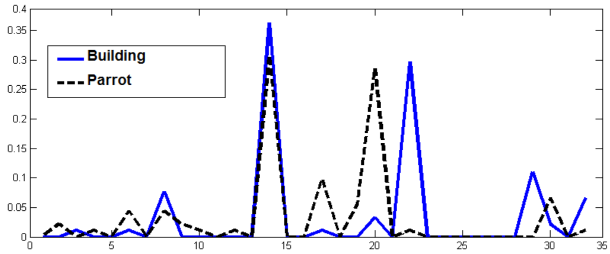
The results are presented in the table 4. The best classification rate may not be obtained with descriptors with learning without errors like SIFT or OpponentSIFT. The average classification rate is 51,75%. The best classification rate is obtained with the energy of Gabor Features. This accuracy is due to the configuration of our Gabor filters. This configuration is closed to the human visual cortex.
The two colours based descriptors also offer a good accuracy. The OpponentSIFT perform an interesting classification with 52,63%. This result confirms the recommendation of Van de

Classification rates

Categories of descriptors	Numbers of errors	Classification rate
Colours	8	57,89%
Gabor	7	63,15%
GIST	12	36,84%
OpponentSIFT	9	52,63%
SIFT	8	57,89%
Wave Atoms (Scale 4)	11	42,10%



(a) Histograms of two positive images



(b) Histograms of two negative images

Figure 5. *Histograms of repartition of the images on figure 4*

Sande and Snoek [18]. Despite the good classification rate of the different colours in the image, in our future works we will try another usage of colour information for emotions.
The results of global descriptors can be explained by the diversified images in the database for the same emotion category.

Comparison of results with the previous works

It is not easy to compare the different works about the extraction of emotions because of the differences in the databases and the feature extractions. In spite of theses differences, we can compare our classification rates with some previous studies. For example, the rates obtained par Wei and al. [20] are from 43,4% to 50,25%. They perform an emotional classification of the images based on the semantic description of the images.
The success rate of Dellandréa and al. [2] is 52% on their initial database [1]. They used the following features:

- Colour moments,
- Colour histograms,
- Colour correlograms,
- Tamura features,
- Characteristics of harmony,
- Dynamic characteristics.

These results were obtained with 80% of the database used like learning database. They also test their method on IAPS database [8]. For this database considered as a reference in psychological studies, their success rate is 64,6%.
The IAPS database also increases the results of Machajdik and Handurry [12]. Their rate is 68% and they used colours, textures, composition and content features.
Compared to these different strategies and database our results are relevant.

[1]Database built with the images from `http://www.gettyimages.com/`

Conclusions and perspectives

Extraction of emotional impact of images requires the consideration of many parameters. We modelled the most of these parameters by the following attributes:

- Different colours in the images,
- Different textures,
- Images content with local and global descriptors.

Our work has the characteristic to having been made on a diversified image database that we will make available to the image community. We also considered efficient and recent descriptors. They allowed us to lead to very encouraging classification rate. We get an average success rate of 51,75%. These results are relevant and confirm that our descriptors are relatively consistent with the emotional information contained in the images.
One of the most important perspectives in our work would be to assess our system to extract emotional impact on IAPS database in order to make efficient comparison with the other works in this domain.
Furthermore in many studies that we have referenced but not limited, colours are considered as an important emotional factor. Although no link unanimous exists between different colours and emotions, they stay interesting in the extraction of emotional impact. We will continue our studies to find a adequate colour descriptors which can try to model some relationship between colours and emotions.
Saliency can be also interesting for increase our results. We have for this purpose, plans to hold new psycho-visual tests with an eye tracker. We can assess visual attention of observers based on emotions.

References

[1] Daniel Beresniak, ABC des couleurs Leurs incidences dans votre vie quotidienne, 1990, pg.11-38.

[2] Emmanuel Dellandréa, Ningning Liu, Bruno Tellez, and Liming Chen, Reconnaissance de la sémantique émotionnelle portée par les images, Technical report, Ecole Centrale de Lyon, LIRIS, UMR5205, 2010.

[3] Laurent Demanet and Lexing Ying,Wave atoms and time upscaling of wave equations, Numerische Mathematik, 113, 1-71 (2009).

[4] Christine Fernandez-Maloigne, Alain Trémeau, and Pierre Bonton, Image Numérique Couleur: De l'acquisition au traitement, 2004.

[5] S.E. Grigorescu, N. Petkov, and P. Kruizinga, Comparison of texture features based on gabor filters, Image Processing, IEEE Transactions on, 11(10),1160-1167 (2002).

[6] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin, A practical guide to support vector classification, 2003.

[7] Naz Kaya and Helen H. Epps, Color-emotion associations: Past experience and personal preference, AIC Colors and Paints, Interim Meeting of the International Color Association, 2004.

[8] P.J. Lang, M.M. Bradley, and B.N. Cuthbert, International affective picture system (IAPS): Affective ratings of pictures and instruction manual, technical report A-8, Technical report, University of Florida, 2008.

[9] David G. Lowe, Object recognition from local scale-invariant features, Proc. International Conference on Computer Vision, 2, pg.1150-1157.(1999).

[10] David G. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision, 60, 91-110 (2004).

[11] Marcel P. Lucassen, Theo Gevers, and Arjan Gijsenij, Adding texture to color: quantitative analysis of color emotions, Proc. CGIV, (2010).

[12] Jana Machajdik and Allan Hanbury, Affective image classification using features inspired by psychology and art theory, Proc. International conference on Multimedia, pg. 83-92. (2010).

[13] Aude Oliva and Antonio Torralba, Modeling the shape of the scene: A holistic representation of the spatial envelope, International Journal of Computer Vision, 42, 145-175 (2001).

[14] Li-Chen Ou, M. Ronnier Luo, Andrée Woodcock, and Angela Wright, A study of colour emotion and colour preference. part i: Colour emotions for single colours, Color Research & Application, 29(3), 232-240 (2004).

[15] M. Paleari and B. Huet, Toward emotion indexing of multimedia excerpts, Proc. Content-Based Multimedia Indexing, International Workshop, pg. 425-432. (2008).

[16] Simon Rigoulot, Impact comportemental et électrophysiologique de l'information émotionnelle en vision périphérique, PhD thesis, Université de Lille II - école Doctorale de Biologie-Santé, 2008.

[17] Imen Tayari, Nhan Le Thanh, and Chokri Ben Amar, Modélisation des états émotionnels par un espace vectoriel multidimensionnel, Technical report, Laboratoire Informatique,Signaux et Systèmes de Sophia Antipolis, 2009.

[18] K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek, Evaluating color descriptors for object and scene recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(9), 1582-1596 (2010).

[19] Wei-Ning Wang and Ying-Lin Yu, Image emotional semantic query based on color semantic description, Proc. The Fourth International Conference on Machine Leraning and Cybernectics, 7, pg. 4571-4576. (2005).

[20] Kaiping Wei, Bin He, Tao Zhang, and Wenya He, Image Emotional Classification Based on Color Semantic Description, Lecture Notes in Computer Science, Springer Berlin/Heidelberg, 2008, 5139, pg. 485-491.

[21] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid, Local features and kernels for classification of texture and object categories: A comprehensive study, International Journal of Computer Vision, 73(2), 213-238 (2007).

Author Biography

Syntyche Gbèhounou received her MS in Computer Science and Telecommunications from the university of Poitiers (September 2011). She is doing her PhD in indexation and categorization of images and video in the department SIC of XLIM laboratory in Poitiers (France). Her work is focused on the new descriptors which can improve the indexation and categorization rate by including for example emotional and saliency information.

François Lecellier is an associate professor at the Institute of Technology, University of Poitiers since 2010. He is a member of the department SIC of the XLIM institute (UMR CNRS 7252). He obtained his PhD on Computer Science in 2009 at the University of Caen (France). His research projects are in the domain of multimedia retrieval, multimedia databases, video copy detection and image processing. He also studies region based active contours for image and video segmentation.

Christine Fernandez-Maloigne received her Master in Computer Sciences and her PhD in imaging from the University of Technology of Compiegne where she was Assistant Professor until 1997. Then she moved to University of Poitiers as a full professor where she's at the head of a laboratory dedicated to digital images with a specific team for color imaging. She was the first general chair of CGIV in 2002. She is member of IS&T, IEEE and SPIE.