

# Improvement of natural image search engines results by emotional filtering

Patrice Denis<sup>1</sup>, Vincent Courboulay<sup>1,\*</sup>, Arnaud Revel<sup>1</sup>,  
Syntyche Gbèhounou<sup>2</sup>, François Lecellier<sup>2</sup>, Christine Fernandez-Maloigne<sup>2</sup>

<sup>1</sup> Laboratoire Informatique, Image et Interaction (L3i)  
Pôle Sciences et Technologie - Université de La Rochelle  
Avenue Michel Crépeau - 17042 La Rochelle Cedex 1 - France

<sup>2</sup> Département SIC du Laboratoire XLIM - UMR CNRS 7252  
Université de Poitiers - Bât SP2MI Téléport 2 - 11 Bd Marie et Pierre Curie  
BP 30179 - 86962 Futuroscope Chasseneuil Cedex - France

## Abstract

With the Internet 2.0 era, managing user emotions is a problem that more and more actors are interested in. Historically, the first notions of emotion sharing were expressed and defined with emoticons. They allowed users to show their emotional status to others in an impersonal and emotionless digital world. Now, in the Internet of social media, every day users share lots of content with each other on Facebook, Twitter, Google+ and so on. Several new popular web sites like FlickrR, Picassa, Pinterest, Instagram or DeviantArt are now specifically based on sharing image content as well as personal emotional status. This kind of information is economically very valuable as it can for instance help commercial companies sell more efficiently. In fact, with this kind of emotional information, business can be made where companies will better target their customers needs, and/or even sell them more products.

Research has been and is still interested in the mining of emotional information from user data since then. In this paper, we focus on the impact of emotions from images that have been collected from search image engines. More specifically our proposition is the creation of a filtering layer applied on the results of such image search engines. Our peculiarity relies in the fact that it is the first attempt from our knowledge to filter image search engines results with an emotional filtering approach.

Received on 14 April, 2015; accepted on 24 January, 2016; published on 25 April, 2016

**Keywords:** creative/affective experiences, creative technologies, creative user experience, emotions, image search engine, natural images, image retrieval, image web browsing, bottom-up saliency, regions of interest

Copyright © 2016 P. Denis *et al.*, licensed to EAI. This is an open access article distributed under the terms of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi:10.4108/eai.25-4-2016.151164

## 1. Introduction

Needs from users of the Web are evolving with new technologies. Nowadays, people are accustomed to participate and give their opinions on websites or social networks. For example, they share everyday emoticons, *likes* on Facebook, *+1's* on Google+, short messages on Twitter. They even share their professional networks connections on LinkedIn or Viadeo. The *Internet of Things* is a technology that now allows to interact

with computers not only with keyboard and mouse but in new user-oriented devices like connected watches, pedometers, refrigerators, electronic scales or Google glasses for instance.

Web search engines users are often lost in the huge amount of answers they got from anonymous requests and want specific and personalized replies to their own needs. Semantic Web, sometimes called Web 3.0, is evolving to focus more and more on structuring its data to answer more precisely and personally to these users.

In this context, emotional status of users is now taken into account by recent works. Business industries are in

\*Corresponding author. Email: [vcourbou@univ-lr.fr](mailto:vcourbou@univ-lr.fr)

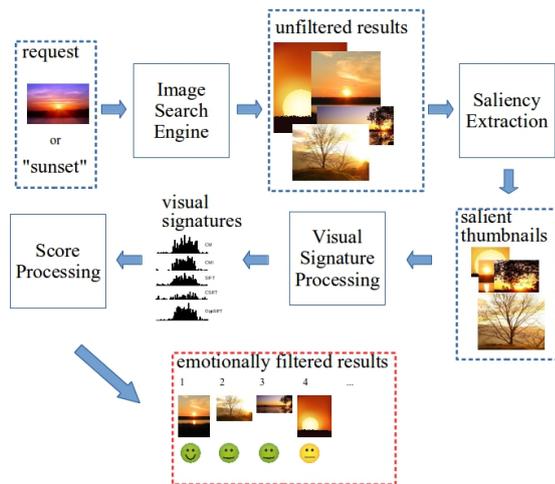


Figure 1. Global emotional filtering model.

because they think they could increase their customers market delivering better customized products arousing positive emotions. This paper proposes to contribute to these researches introducing a new layer to image search engines filters according them to add an image emotion status. We do not know any search engine that has addressed these issues the way we have.

We first introduce in Section 2 background on image search engines (2.1). Emotion definitions and modelling are given in 2.2. We discuss in Section 2.2 how emotions have been extracted from digital documents in the past. More details are given specifically on texts and images. 3 is devoted to the description of our proposed system. We first present Support Vector Machines technology (SVMs), used for our classifiers. Then, an image database, called SENSE, is presented and we justify its usage to train our system. Next, we expose a bottom-up saliency model we used to focus on the most important emotional information of each image. Then we detail the whole processing string our system. We show some of its results on our Internet crowded image database of cooking recipes and compare it to other systems. Last section, (4), is kept for conclusions and perspectives.

## 2. Background and related works

### 2.1. Image search engines filtering

Image research engines often allow filtering their results. This is usually done by extraction and comparison of image features. Those can be deduced from bottom-up or top-down approaches [15]. Bottom-up methods extract low level features on image data, while top-down methods are task-driven ones using supervised learning. Common filtering techniques get information from direct access within the image file metadata. Examples are bottom-up features like the

size; dimension proportion ratio: height, width; style: photo, drawing. Top-down classical features extraction from metadata includes: characters, faces, portraits with head and shoulders... The number of indexed images where features were extracted is also a criteria to compare search engines filters. Table 1 shows a comparison on common filters.

Here is a non exhaustive list of several famous and not so famous image search engines:

*Google images*<sup>1</sup>; *Bing images (Microsoft)*<sup>2</sup>; *TinEye*<sup>3</sup>; *Picsearch*<sup>4</sup>; *oSkope visual search*<sup>5</sup>; *Search22*<sup>6</sup>.

The following section is about emotion modelling and extraction within digital documents.

### 2.2. Emotion and digital images

Human emotions are answers of human beings to their environment. This environment is interpreted by their brains through several sensors: eyes, ears, nose, mouth, skin... People use their emotions to interact with other humans. They rely on their empathy to interpret these emotions correctly.

**Emotion modelling.** Many definitions of emotion have been published and differ from different schools of psychology. Referring to computational approach of emotion recognition, two theories are considered [25]:

- *Basic emotion theories* [39, 45] establishing the existence of basic emotions. These basic emotions are also referred as fundamental or primary.
- *Theories of evaluation* defining emotion as a set of appraisal states that occur when a human being is faced with an external stimulus.

These two theories have defined the two more used emotion classification:

1. **Discrete approach** [10, 37]: emotional process can be explained with a set of basic or fundamental emotions, innate and common to all human (sadness, anger, happiness, disgust, fear, ...). There is no consensus about the nature and number of these fundamental emotions. This modelling is usually preferred in emotion extraction based on facial expressions. An example of discret classification of emotion is given in the Figure 2(a); it represents the Plutchik's wheel of emotions [37]. The author defines eight basic emotions; combinations of primary emotions define secondary ones.

<sup>1</sup><http://www.google.fr/imghp>

<sup>2</sup><http://www.bing.com/?scope=images>

<sup>3</sup><http://www.tineye.com>

<sup>4</sup><http://www.picsearch.com>

<sup>5</sup><http://www.oskope.com>

<sup>6</sup><http://www.search-22.com/downloads/images.php>

Filter	date	size	orientation	color	file format	licence	style	faces	image query	textual query
Google image	x	x		x	x	x	x	x	x	x
Bing image	x	x	x	x			x	x		x
TinEye	x	x				x			x	
Picsearch		x	x	x			x	x		x
Exalead		x	x	x	x	x	x	x		x

Table 1. Comparison of common filters proposed by image search engines

2. **Dimensional approach:** emotions are considered as the result of a fixed number of concepts represented in a dimensional space. The dimensions can be pleasure, arousal or power for instance. They vary depending to the needs of the model. Russel’s dimensional model, represented in the Figure 2(b), is the most used with the dimensions valence and arousal:

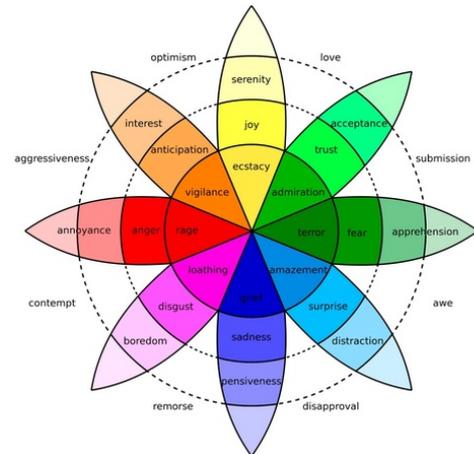
- *The valence* corresponds to the way a person feels when looking at a picture. This dimension varies from negative to positive and allows to distinguish between negative emotions and pleasant ones.

- *The arousal* represents the activation level of the human body.

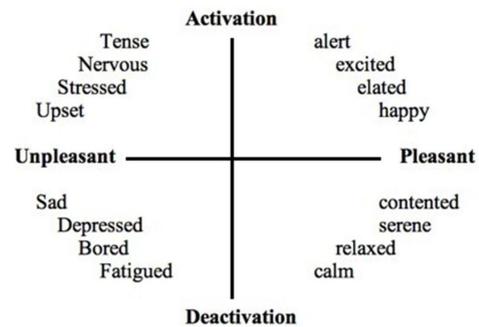
The advantage of these models is to define a large number of emotions. Despite this, some emotions can be confused (such as fear and anger in the circumplex of Russel) or unrepresented (among others surprise in Russel’s model).

**Emotion extraction.** As a reaction to our environment, emotions are linked to our five senses. But let’s face that only seight and hearing are used in our digital world of multimedia documents. Many researches have been made to achieve emotion extraction by computers and improve interaction between humans and machines. The holy grail to achieve being able to bring human-empathy to computers aptitudes. Though, the multiplicity and diversity of digital documents is still a major research challenge for computers to correctly gather and interpret emotions. Moreover, extraction of emotions can either be targeted to the content or to the user of the document. In other words, the goal may be to extract emotions intrinsically embedded into a document or those that will arouse on humans reading it.

**From audio** To decode emotions from audio files, focus is directed on our hearing aptitudes. Everyone knows that music or voice provoke emotions on people. For instance, one can often tell the mood of another person by listening her even in a foreign language.



(a) Plutchik's wheel of emotions



(b) Russell's dimensional model

Figure 2. Examples of discrete and dimensional emotion classification.

Competitive challenges like the *avec*<sup>7</sup> or the *isca-speech*<sup>8</sup> for instance are gathering new contributors every year [22, 31].

Table 2 summarizes these emotion extraction approaches on audio documents.

The majority of digital documents are captured by our eyes. Most contributions in the literature are then focused on visual documents as we do in the following.

<sup>7</sup>International Audio/Visual Emotion Challenge and Workshop ; <http://sspnet.eu/avec2014/>

<sup>8</sup>Interspeech Conferences of the International Speech Communication Association; <http://www.isca-speech.org/iscaweb/>

Method	features	extraction/transmission	context/application
Krishnan	speech	speaker's emotional state	recognition system
Shing	audio	audio emotion	recognition system

Table 2. Extraction approaches on audio documents

**From text documents** The first scrutinized documents to extract emotions were digital textual documents. Text is linked to language. If computers could interpret texts, they would be able to interpret our emotions expressed. This textual information, is still combed to decrypt human emotions. For instance, Salway and Graham used audio description text streams to extract characters emotions in movies [41]. Kamvar and Harris extracted emotional information of textual contents from bloggers all around the Internet [20].

With social media, needs to express emotions brought up emoticons. They were historically defined by small groups of characters, usually two or three, representing a facial expression (: -) :-| :- ( ^\_^) and (;\_;) are examples). Now they are single character encoded (© or ☺). Emoticons were used to dig up emotions from text documents [38, 46]. Hu et al. defined different emotional signals extracted from social media datasets (Twitter and Facebook short messages) to built an unsupervised sentiment analysis system [17]. Lin et al. proposed an emotion classification system based on five textual feature types [24]. They pointed out that authors and observers points of view on emotions are often not the same as in [21].

Table 3 summarizes these emotion extraction approaches on textual documents.

The following paragraph focuses on emotion extraction from image documents.

**From image documents** Image emotional extraction is a challenge because searchers are still learning how to link computer language and human being interpretations of images. The first approaches to compute image emotions were to find associated text. That way, previous works done on text (as in Section 2.2) helped extract emotions on images.

Emotion search engines or interface based on textual metadata and users tagging were proposed in [11, 18, 21, 52].

A large part of the literature is devoted to the links between emotions and colours [2, 6, 7, 27, 32–34, 49]. In fact, a consensus states the existence of a link between colours and particular emotions. As stated by Ou et al. [32], colours play an important role in decision-making, evoking different emotional feelings. In a serie of publications, Ou et al. [32–34] studied the relationship between emotions, preferences and colours. They established a model of emotions associated with colours from psychophysical experiments.

Emotions are also extracted from facial features (such as eyebrows, lips) using faces contained in images [35]. It seems to be the easiest way to predict emotions, since facial expressions are common to human to express basic emotional feelings (happy, fear, sadness).

More recently some authors considered the emotion recognition as a CBIR task [28, 44, 50]. The underlying idea is to use the traditional techniques of image retrieval to extract the emotional impact. Extraction of traditional image features (colours, textures, shapes) combined with a classification system allows the authors to predict the emotional impact of images after a learning step. It is often necessary to add complementary information. For example, Wang and Yu [48] used the semantic description of colours to associate an emotional semantic to an image. Liu et al. [25] stated that oblique lines could be associated with dynamism and action; horizontal and vertical ones with calm and relaxation.

Although research has been made to characterize still images emotions with low level features extraction as said above, we do not know any emotional image search engine focused only on low level extraction from image as we developed in this paper. However some studies have been made on image search engines.

Table 4 summarizes these emotion extraction approaches on images documents.

Most of these former presented papers use text metadata mining to extract emotional information about images in their search engines. A novelty of our paper is that we do not use emotional associated text, or as Solli emotional Bags of Emotions [43], but we pull out emotion in images from bags of visual words [4] based on low level descriptors extracted from pixel level processing.

### 3. Proposed system

Our proposed method starts by a Query By Image Content (QBIC) and defines an emotional filter system based on low level images features. This filter can then be applied on image search engines results in Web technologies for instance. For that, we processed an emotional score for each image in a set acquired by QBIC. This score is gained by a fusion of five classifier scores. The classifiers are based on Support Vector Machines (SVMs) [9] described in Subsection 3.1 and powered by libSVM [8]. Classifiers are trained on a low semantic database called SENSE which is described in

Method	features	extraction/transmission	context/application
Salway	audio description text	characters emotion	movies
Kamvar	textual content	emotion info	bloggers
Read, Urabe	emoticons	user emotion	machine learning
Hu	social media datasets	emotional signals	sentiment analysis
Lin	text	reader's emotion	classification

Table 3. Emotion extraction approaches on textual documents

Method	features	extraction/transmission	context/application
Knautz	textual metadata, tags	felt and/or shown emotion	animals DB/emotion search engine
Fonseca	textual metadata, tags	observer's perspective	image search interface
Yanulevskaya	texture descriptors, text	emotion	emotion categorization system
Zhang	metadata	image attractiveness	image search engine interface
Huisman	visual appeal	visual content from web pages	emotion evaluating interface

Table 4. Emotion extraction approaches on images documents

Subsection 3.2. The description and justification of the usage of a saliency model can be read in Subsection 3.3. Eventually, the whole stream of our emotional filtering system is described in details in the last Subsection 3.4.

### 3.1. Support Vector Machines

Support Vector Machines (SVMs), introduced as Support Vector Networks in [9] are supervised learning models used for classification and regression analysis.

We briefly introduce the basics of SVMs to explain how we used them in our approach. When one want to perform classification, he has to get data first. A model is learnt splitting data into training and testing sets. Each sample of the training set is described by a feature vector and labeled to the class it belongs to. The goal of SVMs is to produce a model, based on the training set which predicts the target class or label of every sample of the test set given only the its attributes (feature vector). To achieve that, samples projected in a dimensional space are separated by a linear hyperplane.

SVMs classification process is illustrated by Figure 3. Let  $(\mathbf{x}_i, y_i), i = 1, \dots, m$  where  $\mathbf{x}_i \in \mathbb{R}^n$ ,  $n$  the dimension of the feature vector and  $\mathbf{y} \in \{1, -1\}^m$  be a training set of  $m$  sample-label pairs. In both sckeches, the separating hyperplane, defined by equation  $\mathbf{w} \cdot \mathbf{x} + b = 0$ , is represented by the black line. For every sample,  $r = \frac{|\mathbf{w} \cdot \mathbf{x} + b|}{\|\mathbf{w}\|}$  is the distance to the separating hyperplane. The support vectors (rounded points) are the closest points to the separating hyperplane.  $d$  is the distance between the two classes (red and blue). It is also called the margin. The goal of SVMs is to maximize this distance. In the case of hard margin SVMs (Figure 3.a), SVM finds a linear separating hyperplane with the maximal margin. When using soft margin (Figure 3.b), an error term  $\xi$  is introduced to take classification errors and noise into account.  $C > 0$  is the penalty parameter. It

is a constant used to control the balance between the number of classification errors and the margin width.

The support vector machines (SVMs) [5, 9] require the solution of the following optimization problem:

$$\min_{\mathbf{w}, b, \xi} \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^m \xi_i$$

subject to  $y_i(\mathbf{w}^T \Phi(\mathbf{x}_i) + b) \geq 1 - \xi_i,$   
 $\xi_i \geq 0.$

Here training vectors  $\mathbf{x}_i$  are mapped into a higher (maybe infinite) dimensional space by the function  $\Phi$ .

Furthermore,  $k(\mathbf{x}_i, \mathbf{x}_j) \equiv \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_j)$  is called the kernel function. Though new kernels are often being proposed by researchers, the most known and most commonly used in the literature are the following:

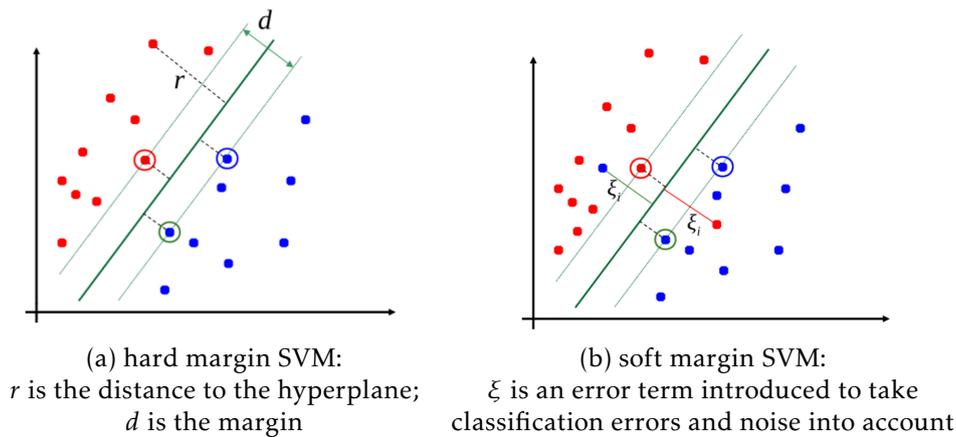
- linear:  $k(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$ ;
  - polynomial:  $k(\mathbf{x}_i, \mathbf{x}_j) = (\gamma \mathbf{x}_i^T \mathbf{x}_j + r)^d, \gamma > 0$ ;
  - RBF<sup>9</sup>:  $k(\mathbf{x}_i, \mathbf{x}_j) = \exp^{-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2}, \gamma > 0$ ;
  - sigmoid:  $k(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\gamma \mathbf{x}_i^T \mathbf{x}_j + r).$
- $\gamma \in \mathbb{R}$  being a kernel parameter.

In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces. This property of the SVMs is not described furthermore here as we used linear classification.

### 3.2. SENSE database

Several image databases have been used to study emotions [28, 44, 50]. The most known and used

<sup>9</sup>RBF = radial basis function



**Figure 3. SVMs' principle:** The two classes (red and blue points) are separated by the SVM hyperplane (black line). Support vectors (rounded points) are the closest samples to the hyperplane.

is the International Affective Picture System (IAPS) <sup>10</sup> [23]. As there were no universal emotional image database focused on extraction of emotional impact, Gbèhounou et al. [13] proposed one. This image set is named SENSE for Studies of Emotion on Natural image dataBaS. It is composed of images as diversified as possible and with low semantic power. That is to say that the images do not shock or force a strong emotional answer during subjective evaluation in opposition to some IAPS images for example. Each image has been evaluated by users during very short period of time: eight seconds. This allows to look for primary emotion impact of the presented images. In fact, Schyns et al. demonstrated that emotions on facial expressions [42] are transmitted during the very first 200ms after the reception of the observer signal. During this very short period time, semantic interpretation can not be done by the user. On the contrary, image from the IAPS, embed high semantic power. High semantic can introduce a bias on the evaluation of multiple images as strong emotions remains longer that law emotions in the mind of evaluators [40]. The authors focused to include only low semantic images in their database in order to minimize or to remove that kind of bias during evaluation. The goal was to evaluate each image for its content and not for the remnant of the previous image emotion evaluation.

The database is free to use for research and contains 350 images of landscapes, characters, animals, food, drink, historic and tourist buildings images as shown in Figure 4. Furthermore, the database also contains very few images with human faces (4.86%). This last point reflects the desire to limit the interpretation of facial expressions which embed bigger emotional reactions.

The SENSE database was then evaluated in different subjective experiments. 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 the authors have chosen. They decided to work with a dimensional model which dimensions are:

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

This 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 intensity of the emotion associated with the nature of the chosen emotion. During the different tests, observers evaluated the nature and the power of the emotion aroused by the image. As shown on Figure 5, for the nature of the emotion, the participants had choice between "Negative", "Neutral" and "Positive". The power was chosen from low to high. During the tests, some presented images have the same content but are modified with some transformations. Rotations and changes in colour balance are applied. Even if they change the natural aspects of some pictures, they allowed to see the results of some invariants features used to recognise the emotion associated with the image. Only 2.29% images represent these transformed "non natural" content.

The next subsection will describe why we chose to process our features extraction algorithms on salient thumbnails instead of full images.

### 3.3. Bottom-up saliency model

During the emotional evaluation of images of SENSE, the authors realized that a number of pictures were

<sup>10</sup>from the Center for Emotion and Attention (CSEA) at the University of Florida



Figure 4. Images examples from the SENSE database

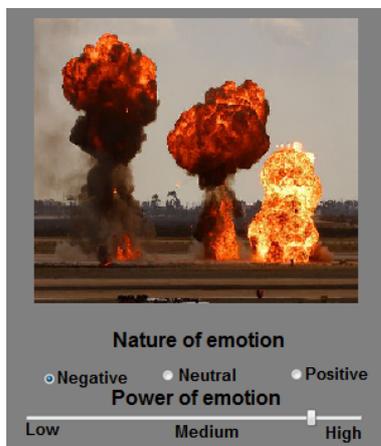


Figure 5. Emotional evaluation of SENSE images

not clearly classified by all users. This result is not surprising because of the construction of the image database. In fact, images on the database are selected so they do not cause strong emotional answers to observers. Therefore it seems logical that a number of them are not easily emotionally classifiable. A new experiment was then mounted: evaluate the most salient areas for each SENSE image. Maybe focusing on the most salient zones would allow better emotional classification? This experiment was made in [13] resulting in a novel SENSE2 database. For each SENSE image, its most salient region counterpart was added to SENSE2. SENSE2 analysis showed a correlation of emotion evaluations from each SENSE image and their salient counterpart in SENSE2.

In our proposed system, as we will see in Section 3.4, we extract emotions on images from CBIR techniques. We designed our system to work on low semantic images as we think that images answers from image search engines will not embed strong emotional impacts on users. This is why we choose to process features from image thumbnails acquired by a saliency model. The saliency model we use was developed in our

laboratory from previous works by Perreira et al.[36] and is based on a computational model of attention.

This was implemented to improve the performances of our system:

- working only on smaller images enhances time processing;
- focusing on most significant regions performs better emotional classifications.

Many researchers have worked on visual attention, one can report to [3] for a good review. A classically used model of visual attention is Laurent Itti's [19]. The first part of its architecture relies on the extraction of three conspicuity maps based on low level characteristics computation, that's correspond to the production of information on retina. These three conspicuity maps are color, intensity and orientation.

The second part of Itti's architecture proposes a medium level system which allows merging conspicuity maps and then simulates a visual attention path on the observed scene. The focus is determined by a "winner-takes-all" and an "inhibition of return" algorithms (Figure 6).

Perreira et al. proposed to substitute this second part by their optimal competitive architecture: a preys / predators system. This optimal criteria, preys / predators equations are particularly well adapted for such a task:

- preys / predators systems are dynamic, they include intrinsically time evolution of their activities. Thus, the visual focus of attention, seen as a predator, can evolve dynamically;
- without any objective (top-down information or pregnancy), choosing a method for conspicuity maps fusion is hard. A solution consists in developing a competition between conspicuity maps and waiting for a natural balance in the preys / predators system, reflecting the competition between emergence and inhibition of elements that engage or not our attention;
- discrete dynamic systems can have a chaotic behavior. Despite the fact that this property is not often interesting, it is an important one. Actually, it allows the

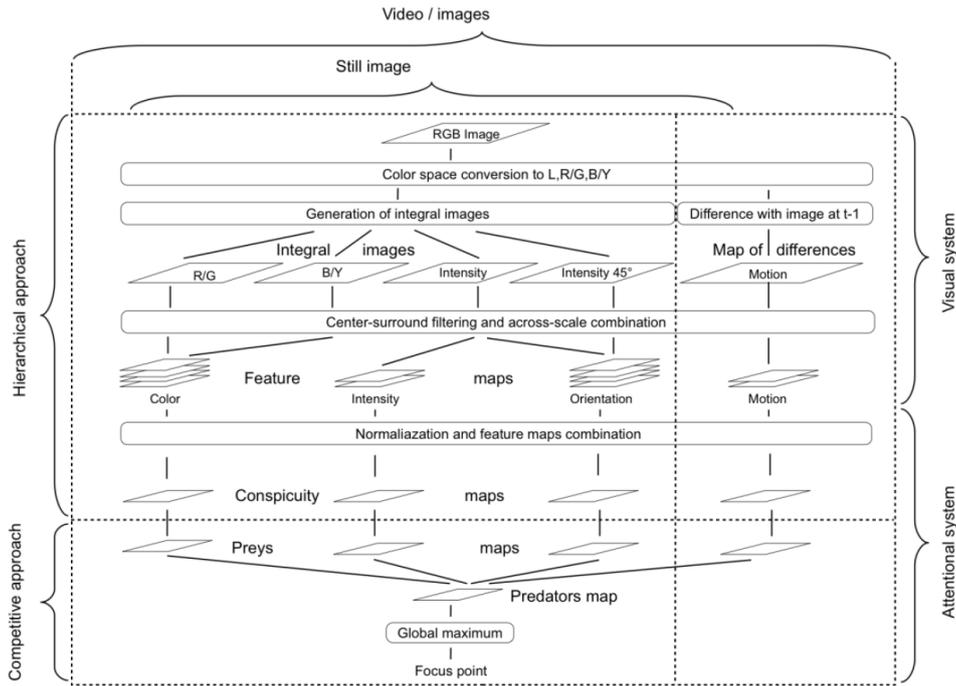


Figure 6. Architecture of the Perreira computational model of attention

emergence of original paths and exploration of visual scene, even in non salient areas, reflecting something like curiosity.

General architecture of the Perreira model is represented in Figure 7

Starting from the “basic” version of preys / predators equations, Perreira et al. enriched processing in several ways:

- the number of parameters can be reduced by replacing  $s'$  by  $s$ . Indeed, mortality rates differences between preys and predators can be modeled by an adjustment of factors  $b$  and  $m_I$
- the original model represents the evolution of a single quantity of preys and predators over time. It can be spatially extended in order to be applied to 2D maps where each point represents the amount of preys or predators at a given place and time. Preys and predators can then “move” on this map using a classical diffusion rule, proportional to their Laplacian  $\Delta_C$  and a diffusion factor  $f$ .
- natural mortality of preys in the absence of predation is not taken into account. If the model only changes temporally, mortality is negligible when compared to predation. However, when the model is applied to a 2D map (which is the case in our system), some areas of the map may not contain any predator. Natural mortality of prey can no longer be considered negligible. A new mortality term  $-m_c$  need to be added to the model.

This yields to the following set of equations, modeling the evolution of preys and predators populations on a two dimensional map:

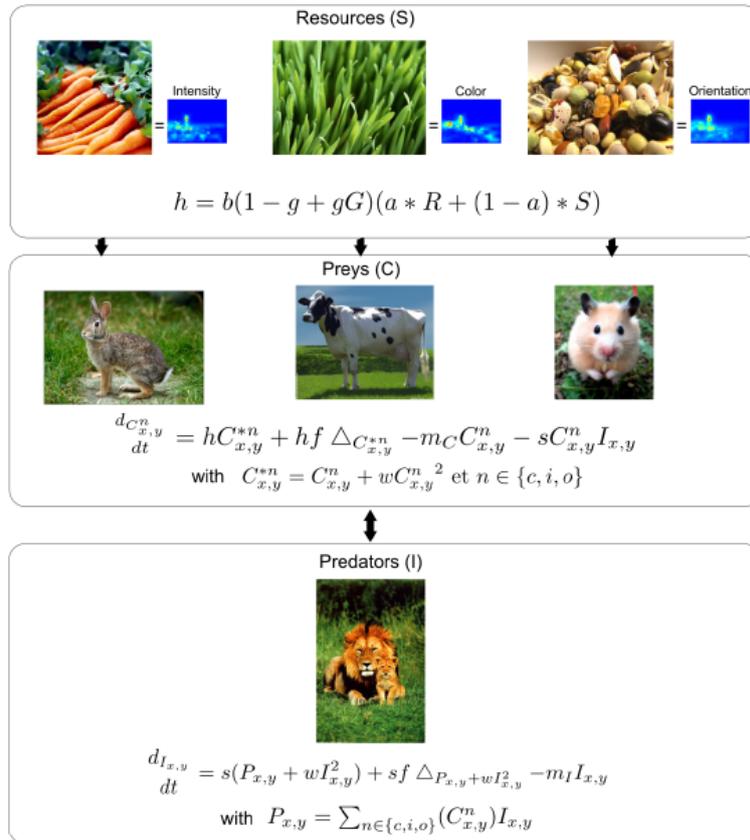
$$\begin{cases} \frac{dC_{x,y}}{dt} = bC_{x,y} + f \Delta_{C_{x,y}} - m_C C_{x,y} - sC_{x,y}I_{x,y} \\ \frac{dI_{x,y}}{dt} = sC_{x,y}I_{x,y} + sf \Delta_{P_{x,y}} - m_I I_{x,y} \end{cases}$$

A last phenomenon can be added to this model: a positive feedback, proportional to  $C^2$  or  $I^2$  and controlled by a factor  $w$ . This feedback models the fact that (provided unlimited resources) the more numerous a population is, the better it is able to grow (more efficient hunting, higher encounter rater favoring reproduction, etc.). The final preys / predators system is then:

$$\begin{cases} \frac{dC_{x,y}}{dt} = b(C_{x,y} + w(C_{x,y})^2) + f \Delta_{C_{x,y}} - m_C C_{x,y} - sC_{x,y}I_{x,y} \\ \frac{dI_{x,y}}{dt} = s(C_{x,y}I_{x,y} + w(I_{x,y})^2) + sf \Delta_{P_{x,y}} - m_I I_{x,y} \end{cases}$$

In order to simulate the evolution of the focus of attention, Perreira et al. proposed a preys / predators system (as described above) with the following features:

- the system is comprised of four types of preys and one type of predators;
- these four types of preys represent the spatial distribution of the curiosity generated by four types of conspicuity maps;



**Figure 7.** Competitive preys / predators attention model. Singularity maps are the resources that feed a set of preys which are themselves eat by predators. The maximum of the predators map represents the location of the current focus of attention.

- the predators represent the interest generated by the consumption of curiosity (preys) associated to the different conspicuity maps;
- the global maximum of the predators maps (interest) represents the focus of attention at time  $t$ .

The authors showed that despite the non deterministic behavior of preys / predators equations, the system exhibits interesting properties of stability, reproducibility and reactivity while allowing a fast and efficient exploration of the scene.

With this visual saliency model, we can focus our emotion extraction on the minimal local region of an image where the its emotion impact should be the most meaningful. In other words, this enables us to summarize every image emotion by its most salient thumbnail counterpart and enhances as well feature extraction processing time as thumbnails dimensions are smaller.

The next subsection represents the core of our system, it details how we constructed our emotional filtering structure layer by layer.

### 3.4. Emotional filtering system structure

Our emotional filter is based on an emotional score processed from a machine learning approach. The idea is to process a signature for a request image and to inject it on trained classifiers in order to output an emotional score. The global system is illustrated by Figure 1 and Figure 8.

First, we collect an input set of images by a textual or image request from any image search engine. Then we extract the main salient region for each image of the set using the computational model of attention [36] formerly described in Section 3.3. This additional processing is justified by the fact that the emotional impact of an image is correlated to its main salient region [13]. Moreover, as thumbnails, salient regions take less weight and then less processing time.

Next, each salient image (thumbnail) is processed by a Harris Laplace key-point detector[16, 29]. A dynamic algorithm is performed to ensure a minimum amount of detected key-points to describe. The size of the thumbnail to process is sometimes enlarged to ensure a minimal amount of detected key-points.

Each key-point is then described using five low level descriptors:

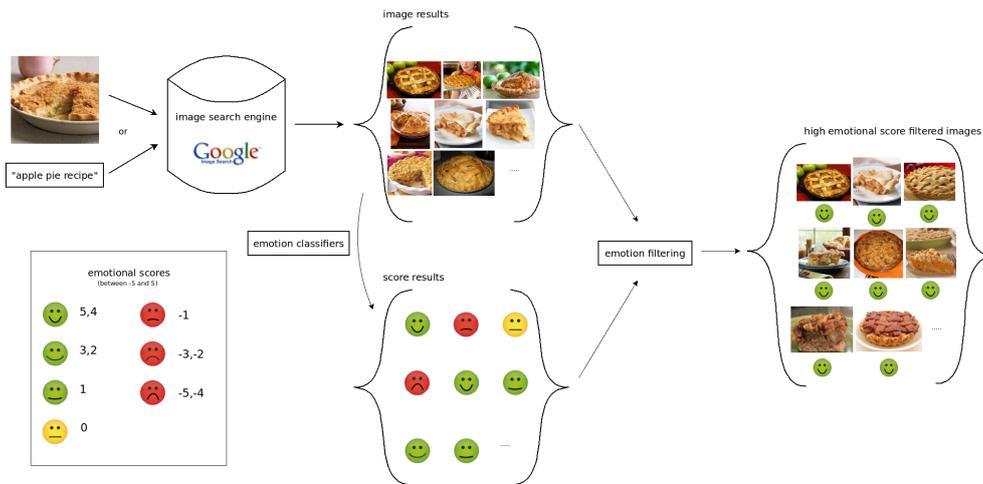


Figure 8. Emotional filtering model. In this example, only high emotional score images are displayed as a result of the filter

- color moments (CM) [51],
- color moments invariant (CMI) [30],
- scale invariant feature transform (SIFT) [26],
- color SIFT (CSIFT) [1]
- opponent SIFT (OppSIFT) [47].

These descriptors, given in [12], are justified by their properties of invariance over geometric transformations such as translations, rotations and image scaling.

Then a k-means reduction algorithm is applied on each vector to reduce the dimensionality of feature vectors. An image CBIR’s Bag of Words (BoW) method [4] is used to product visual words from these feature vectors as it was for training of the emotional classifiers in [12]. Training process is illustrated in Figure 9. Note that emotions are trained offline on the SENSE database.

So we classify each image with these five classifiers by the same training sets used with the SENSE database in [12] using Support Vector Machine classifiers (SVMs) implemented from libSVM library with linear kernels [8]. Each classifier outputs a real score between -1 and +1. -1 stands for a negative emotion impact whereas +1 stands for a positive emotion impact. All the scores are merged by a simple addition into a final emotional score. This is illustrated by Figure 10.

### 3.5. Results and discussions

A 100,000 Internet cooking recipes images database was created by a partner lab in our project. Figure 11 shows image examples of its content. That was a good opportunity to test our model on a large scale. This database is full of a large diversity in the shapes and looks of the displayed dishes. We so presumed that its content do not imply high emotional impact on viewers. This is then why we choose to use the SENSE database as a trainer for our application.

Our system can output different views. First, images can be displayed by processed emotional score order;

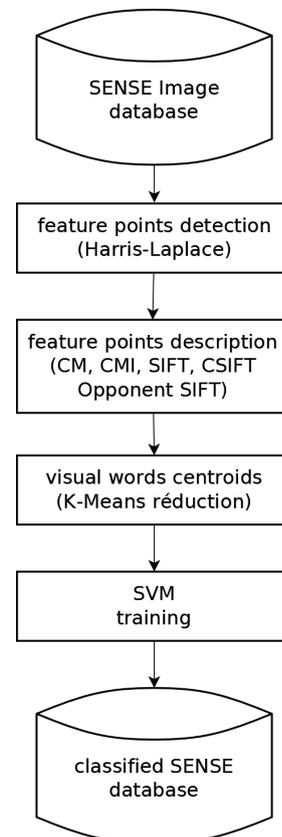


Figure 9. Emotional classifiers training

second, system can output the set of most positive emotional images to fulfill a better emotional impact on viewer when applied on research engine’s results; we can as well show negative emotional score images only.

Figure 8 shows some results of our system on this cooking database. We can see how our system is filtering images and that the result lacks of any

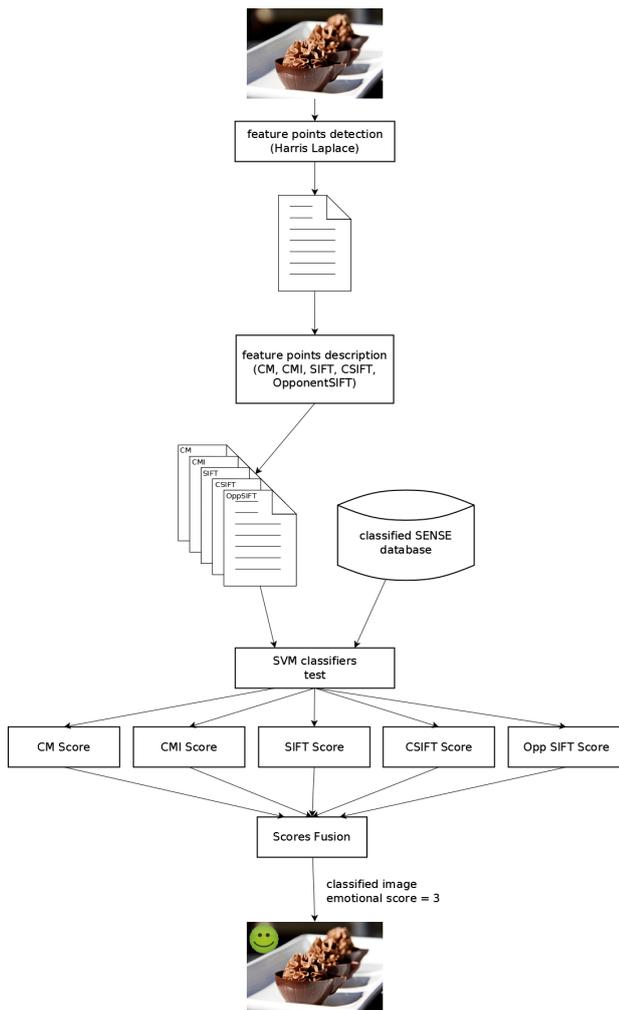


Figure 10. Emotional score processing

emotional validation. But again, the main novelty of our proposition was to apply for the first time an emotional filtering layer on the results of image search engines. This validation process will be investigated on future works. Another contribution was to use low-level features from the most salient regions only to describe image emotions.

In [14], authors used a hybrid approach to define a notion of attractiveness in web image search. This is one of the nearest works we could find from our proposed system. They used EXIF metadata associated with trained high level visual features (quality assessment, aesthetic prediction, affective classification) to define their attractiveness ranks. They also confess that their results are still widely improvable. To improve them, they add images EXIF textual metadata to their model. They also look into the image source web pages for other textual information to help in the classification process.

Our approach here is different, we want to extract the emotion information only on the image content. We are not at all interested with additional information to

enhance the performance of classification as it is usually done in other approaches. Also our system scores image sets where their scores Web pages where images can be found.

#### 4. Conclusions and Perspectives

In this paper, we focused on the impact of emotions from images collected from research image engines. We proposed for the first time to add an emotional filtering system to image search engines results.

With this new filtering technique, web users will enhance their experience on image searches.

We saw that our solution still needs a validation phase. We thought of several centered on human evaluation:

- a standalone application where evaluators will validate the engine classification. Image search engine result sets will be labeled by people on an emotional perspective. Comparison between labeled ground truth samples and the system's results can then be reached;
- evaluators could compare their emotional feelings from watching direct search engines result sets with respect to emotionally filtered ones;
- users needs for individual contents may be achieved using backpropagation labelling to train classifiers on personalized emotional result sets. iv. a user interface to re-rank (drag and drop) images ordered by the emotional filtering system will take their preferred ranking positions as new labeled training sets.
- enhancement of user personalized results by deletion of images from his displayed results. This will again help the classifier training with backpropagation techniques.

New technologies such as eye-tracking can help analysing user behaviour when observing results from search engines. We already started experiments in our lab. Given two images displayed, users are asked to focus on their preferred one. The analysis of eye-tracking data allows the computer to determine the preferred one. Future works may reveal eye-tracking path patterns linked to emotion impact factors when looking to images.

Big data technologies will also be necessary to perform customizable results for users. Behavior patterns are already extracted from your navigation over the Internet. Merging Internet extracted emotional information to emotional image descriptors may also allow huge enhancements in the efficiency of image search engine filters results.

If all were implemented, Internet could eventually be focused on users and not only on data.

**Acknowledgement.** This work was supported by the ANR Visiir project, grant ANR-13-CORD-0009 of the French Agence Nationale de la Recherche.

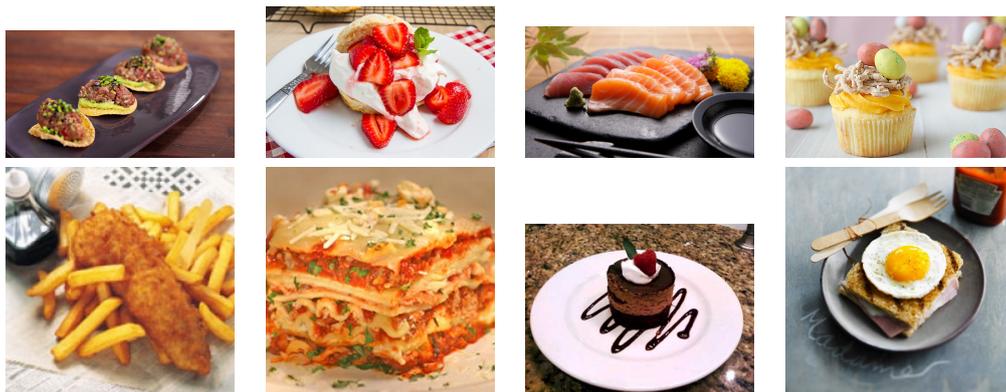


Figure 11. Images examples from the our cooking recipes database

## References

- [1] A. E. Abdel-Hakim and A. A. Farag. Csfift: A sift descriptor with color invariant characteristics. In *Comp. Society Conf. on Comp. Vis. and Pat. Rec.*, volume 2, pages 1978–1983. IEEE, 2006.
- [2] L. Beke, G. Kutas, Y. Kwak, G. Y. Sung, D. Park, and P. Bodrogi. Color preference of aged observers compared to young observers. *Color Research & Application*, 33(5):381–394, 2008.
- [3] A. Borji and L. Itti. State-of-the-art in visual attention modeling. *Trans. on Pat. Analysis and Mach. Intelligence*, 35(1):185–207, 2013.
- [4] A. Bosch, X. Muñoz, and R. Martí. Which is the best way to organize/classify images by content? *Image and vision computing*, 25(6):778–791, 2007.
- [5] B. E. Boser, I. M. Guyon, and V. N. Vapnik. A training algorithm for optimal margin classifiers. In *Proc. of the 5th annual workshop on Computational learning theory*, pages 144–152. ACM, 1992.
- [6] C. Boyatziz and R. Varghese. Children’s emotional associations with colors. *The Journ. of Genetic Psychology*, 155:77–85, 1993.
- [7] M. M. Bradley, M. Codispoti, D. Sabatinelli, and P. J. Lang. Emotion and motivation ii: Sex differences in picture processing. *Emotion*, 1(3):300–319, 2001.
- [8] C.-C. Chang and C.-J. Lin. Libsvm: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, 2(3):27:1–27:27, May 2011.
- [9] C. Cortes and V. Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.
- [10] P. Ekman. Facial expressions of emotions. *Psychological science*, 3(1):34–38, 1992.
- [11] D. Fonseca, O. García, J. Duran, M. Pifarré, and E. Villegas. An image-centred "search and indexation system" based in user’s data and perceived emotion. In *Proc. of the 3rd ACM Int. Workshop on Human-centered Computing*, HCC ’08, pages 27–34, New York, NY, USA, 2008. ACM.
- [12] S. Gbèhounou, F. Lecellier, and C. Fernandez-Maloigne. Extraction of emotional impact in colour images. In *6th European CGIV Conference*, pages 314–319, Pays-Bas, May 2012.
- [13] S. Gbèhounou, F. Lecellier, C. Fernandez-Maloigne, and V. Courboulay. Can salient interest regions resume emotional impact of an image? In R. Wilson, E. Hancock, A. Bors, and W. Smith, editors, *Computer Analysis of Images and Patterns*, volume 8047 of *Lecture Notes in Comp. Science*, pages 515–522. Springer Berlin Heidelberg, 2013.
- [14] B. Geng, L. Yang, C. Xu, X.-S. Hua, and S. Li. The role of attractiveness in web image search. In *Proc. of the 19th ACM Int. Conf. on Multimedia*, MM ’11, pages 63–72, New York, NY, USA, 2011. ACM.
- [15] J. S. Hare, P. A. Sinclair, P. H. Lewis, K. Martinez, P. G. Enser, and C. J. Sandom. Bridging the semantic gap in multimedia information retrieval: Top-down and bottom-up approaches. In *In 3rd European Semantic Web Conf. (ESWC-06)*, LNCS 4011, 2006.
- [16] C. Harris and M. Stephens. A combined corner and edge detector. In *In Proc. of 4th Alvey Vision Conf.*, pages 147–151, 1988.
- [17] X. Hu, J. Tang, H. Gao, and H. Liu. Unsupervised sentiment analysis with emotional signals. In *Proc. of the 22nd Int. Conf. on WWW*, WWW ’13, pages 607–618, Republic and Canton of Geneva, Switzerland, 2013. Int. WWW Conf. Steering Committee.
- [18] G. Huisman, M. van Hout, E. van Dijk, T. van der Geest, and D. Heylen. Lemtool: Measuring emotions in visual interfaces. In *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, CHI ’13, pages 351–360, New York, NY, USA, 2013. ACM.
- [19] L. Itti, C. Koch, E. Niebur, et al. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. on Pat. Analysis and Mach. Intelligence*, 20(11):1254–1259, 1998.
- [20] S. D. Kamvar and J. Harris. We feel fine and searching the emotional web. In *Proc. of the 4th Int. Conf. on Web Search and Web Data Mining (WSDM)*, pages 117–126. ACM, 2011.
- [21] K. Knautz, T. Siebenlist, and W. G. Stock. Memose: Search engine for emotions in multimedia documents. In *Proc. of the 33rd Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*, SIGIR ’10, pages 791–792, New York, NY, USA, 2010. ACM.

- [22] A. Krishnan and M. Fernandez. System and method for recognizing emotional state from a speech signal, Nov. 2013. US Patent 8,595,005.
- [23] P. J. Lang, M. M. Bradley, B. N. Cuthbert, et al. International affective picture system (iaps): Affective ratings of pictures and instruction manual. Technical Report A-8, Univ. of Florida, Gainesville, Florida, 2008.
- [24] K. H.-Y. Lin, C. Yang, and H.-H. Chen. Emotion classification of online news articles from the reader's perspective. In *Proc. of the Int. Conf. on Web Intelligence and Intelligent Agent Technology - Volume 01, WI-IAT '08*, pages 220–226, Washington, DC, USA, 2008. IEEE Computer Society.
- [25] N. Liu, E. Dellandréa, B. Tellez, and L. Chen. Evaluation of features and combination approaches for the classification of emotional semantics in images. In *Int. Conf. on Computer Vision Theory and Applications*, 2011.
- [26] D. G. Lowe. Object recognition from local scale-invariant features. In *Proc. of the 7th Int. Conf. on Comp. Vision*, volume 2, pages 1150–1157. Ieee, 1999.
- [27] M. P. Lucassen, T. Gevers, and A. Gijsenij. Adding texture to color: quantitative analysis of color emotions. In *Proc. of CGIV*, 2010.
- [28] J. Machajdik and A. Hanbury. Affective image classification using features inspired by psychology and art theory. In *Proc. of the Int. Conf. on Multimedia*, pages 83–92. ACM, 2010.
- [29] K. Mikolajczyk and C. Schmid. Scale & affine invariant interest point detectors. *Int. Journal of Comp. Vision*, 60:63–86, 2004.
- [30] F. Mindru, T. Tuytelaars, L. V. Gool, and T. Moons. Moment invariants for recognition under changing viewpoint and illumination. *Comp. Vision and Image Understanding*, 94(1):3–27, 2004.
- [31] C. S. Ooi, K. P. Seng, A. Li-Minn, and L. W. Chew. A new approach of audio emotion recognition. *Expert Systems with Applications*, 41(13):5858 – 5869, 2014.
- [32] L. C. Ou, M. R. Luo, A. Woodcock, and A. Wright. A study of colour emotion and colour preference. part i: Colour emotions for single colours. *Color Research & Application*, 29(3):232–240, 2004.
- [33] L. C. Ou, M. R. Luo, A. Woodcock, and A. Wright. A study of colour emotion and colour preference. part ii: Colour emotions for two-colour combinations. *Color Research & Application*, 29(4):292–298, 2004.
- [34] L. C. Ou, M. R. Luo, A. Woodcock, and A. Wright. A study of colour emotion and colour preference. part iii: Colour preference modeling. *Color Research & Application*, 29(5):381–389, 2004.
- [35] M. Paleari and B. Huet. Toward emotion indexing of multimedia excerpts. *Proc. on Content-Based Multimedia Indexing, Int. Workshop*, pages 425–432, 2008.
- [36] M. Perreira Da Silva and V. Courboulay. Implementation and evaluation of a computational model of attention for computer vision. In *Developing and Applying Biologically-Inspired Vision Systems: Interdisciplinary Concepts*, pages 273–306. Hershey, Pennsylvania: IGI Global, Aug. 2012.
- [37] R. Plutchik and H. R. Conte. *Circumplex Models of Personality and Emotions*. American Psychological Association, 1997.
- [38] J. Read. Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In *Proc. of the ACL Student Research Workshop, ACLstudent '05*, pages 43–48, Stroudsburg, PA, USA, 2005. Association for Computational Linguistics.
- [39] S. Rigoulot. *Impact comportemental et électrophysiologique de l'information émotionnelle en vision périphérique*. PhD thesis, Univ. du Droit et de la Santé - Lille II, Dec. 2008.
- [40] B. Rimé. Émotion et mémoire: La rémanence des expériences émotionnelles. *Cognition et Émotions*, pages 79–96, 2004.
- [41] A. Salway and M. Graham. Extracting information about emotions in films. In *Proc. of the 11th ACM Int. Conf. on Multimedia, MULTIMEDIA '03*, pages 299–302, New York, NY, USA, 2003. ACM.
- [42] P. G. Schyns, L. S. Petro, and M. L. Smith. Transmission of facial expressions of emotion co-evolved with their efficient decoding in the brain: behavioral and brain evidence. *Plos one*, 4(5):e5625, 2009.
- [43] M. Solli and R. Lenz. Color based bags-of-emotions. In *Computer Analysis of Images and Patterns*, volume 5702 of *Lecture Notes in Comp. Science*, pages 573–580. Springer Berlin Heidelberg, 2009.
- [44] M. Solli and R. Lenz. Emotion related structures in large image databases. In *Proc. of the ACM Int. Conf. on Image and Video Retrieval, CIVR '10*, pages 398–405, New York, NY, USA, 2010. ACM.
- [45] I. Tayari, N. Le Thanh, and C. B. Amar. Modélisation des états émotionnels par un espace vectoriel multidimensionnel. Technical report, Technical report, Laboratoire Informatique, Signaux et Systèmes de Sophia Antipolis, 2009.
- [46] Y. Urabe, R. Rzepka, and K. Araki. Emoticon recommendation system for effective communication. In *Proc of the Int. Conf. on ASONAM*, pages 1460–1461, New York, USA, 2013. ACM.
- [47] K. E. Van De Sande, T. Gevers, and C. G. Snoek. Evaluating color descriptors for object and scene recognition. *Trans. on Pat. Analysis and Mach. Intelligence*, 32(9):1582–1596, 2010.
- [48] W. Wang and Y. Yu. Image emotional semantic query based on color semantic description. In *Proc of the 4th Int. Conf. on Machine Learning and Cybernetics*, volume 7, pages 4571–4576, 2005.
- [49] K. Wei, B. He, T. Zhang, and W. He. *Image Emotional Classification Based on Color Semantic Description*, volume 5139 of *Lecture Notes in Comp. Science*, pages 485–491. Springer Berlin / Heidelberg, 2008.
- [50] V. Yanulevskaya, J. Van Gemert, K. Roth, A.-K. Herbold, N. Sebe, and J.-M. Geusebroek. Emotional valence categorization using holistic image features. In *15th ICIP*, pages 101–104. IEEE, 2008.
- [51] H. Yu, M. Li, H.-J. Zhang, and J. Feng. Color texture moments for content-based image retrieval. In *ICIP*, volume 3, pages 929–932. IEEE, 2002.
- [52] L. Zhang, L. Chen, F. Jing, K. Deng, and W.-Y. Ma. Enjoyphoto: a vertical image search engine for enjoying high-quality photos. In *Proc. of the 14th annual ACM Int. Conf. on Multimedia*, pages 367–376, 2006.