
Extraction et analyse de l'impact émotionnel des images

**Syntyche Gbèhounou, François Lecellier,
Christine Fernandez-Maloigne**

*Département SIC du Laboratoire XLIM, UMR CNRS 7252, Université de Poitiers
Bât. SP2MI, Téléport 2, Boulevard Marie et Pierre Curie, BP 30179
F-86962 Futuroscope Chasseneuil cedex
{syntyche.gbèhounou,francois.lecellier,christine.fernandez}@univ-poitiers.fr*

RÉSUMÉ. Cet article propose une méthode d'extraction de l'impact émotionnel des images à partir de descripteurs bas niveau. Nous avons émis l'hypothèse que la précision de ces derniers encodait des informations haut niveau intéressantes voire discriminantes pour les émotions. Il n'existe à ce jour aucun descripteur particulièrement adapté à l'étude de l'impact émotionnel des images. Les émotions ressenties dépendent, en effet, de plusieurs informations dans l'image mais également de sa nature (très sémantique ou non) ou encore de la durée d'observation. Plus ce temps est long plus l'interprétation sémantique de l'image prend le dessus sur l'émotion « primaire ». À ces descripteurs nous avons associé deux classifieurs performants, particulièrement adaptés à des discriminations d'informations complexes. Il s'agit d'un réseau de neurones multicouche et d'un SVM dans son extension multiclasse basée sur la stratégie « un contre un ». Très souvent, on associe les émotions à l'expression du visage, mais nous avons décidé de ne pas faire de cette information la principale caractéristique émotionnelle des images naturelles, qui en général ne contiennent pas de visages. Nous avons, à cet effet, effectué nos tests sur une base diversifiée de 350 images, construite à partir d'images à faible contenu sémantique. Notre choix de descripteurs est basé sur des supposés liens entre les émotions et le contenu des images mais également sur la précision qu'offrent certains descripteurs de traitement d'images en indexation ou en catégorisation. La complexité des émotions a été prise en compte en intégrant, au processus de classification, les résultats de tests psychovisuels que nous avons mis en place. Nous avons défini trois classes d'émotions. Les taux de réussite moyens obtenus lors de la classification sont de 56,15 % pour le réseau de neurones et 55,25 % pour le SVM. Ces résultats sont encourageants au regard des résultats de la littérature. Ces tests confirment aussi l'hypothèse que les descripteurs choisis sont complémentaires dans notre processus d'extraction des émotions.

ABSTRACT. This paper proposes a method to extract the emotional impact of images based on accurate and low level features. We supposed their accuracy could also implicitly encode high-level interesting or discriminant information for emotional impact extraction. Emotions are often associated with facial expressions, but we decided not to consider this feature as first emotional characteristic of natural images, which, in general, does not contain faces. Using

this statement, our tests have been done on a new image database composed of low semantic diversified images. The complexity of emotion modeling was considered in classification process through psycho-visual tests. The twenty five observers assessed the nature and the power of emotions felt. For the nature of the emotion they had the choice between "Negative", "Neutral" and "Positive" and the power ranged from "Low" to "High". With the nature of emotions, we made a classification in three classes of emotions. The average success rate is 56.15 % for artificial neural network and 55.25 % for the SVM classifier; that is really relevant regarding the equivalent results in the literature.

MOTS-CLÉS : émotions, classification, réseaux de neurones, SVM, tests psychovisuels, images couleur.

KEYWORDS: emotions, classification, artificial neural network, SVM, psycho-visual tests, color images.

DOI:10.3166/TS.29.409-432 © 2012 Lavoisier

Extended abstract

This paper proposes a method to extract the emotional impact of images based on accurate and low level features. We supposed their accuracy could also implicitly encode high-level interesting or discriminant information for emotional impact extraction.

There are no suitable image processing features for emotional impact study and analysis. In the literature, there is no consensus about the features used. The chosen features are often linked to the application. In fact, the emotions depend on different information in the image but also the nature of the image (high semantic or not) or the observation time. A long observation period may induce a semantic interpretation, in this case it is not a "primary" emotion. For the classification process we used two classifiers particularly adapted for complex information classification. These are multilayer artificial neural network and SVM classifiers, the last one in its extension for multi-class with the "One against One" strategy.

Emotions are often associated with facial expressions, but we decided not to consider this feature as first emotional characteristic of natural images, which, in general, do not contain faces. Using this statement, we built a new image database for our tests. It is composed of 350 low semantic and diversified images with very few faces.

We chose our features based on the supposed links between emotions and images content but also on the accuracy of some image processing features on indexation and categorization. These features are colours, textures, local and global descriptors. We used some characteristics with fidelity to human vision properties. For example, the different colours are extracted with a region growing segmentation in CIELAB colour space, which is known to be perceptually uniform. One of our two texture features is the energy of Gabor features. For this last one, our filters configuration is one proposed to be close to the vision of the human cortex. Wave Atoms are our second type of

features for textures characterisation. We also used SIFT, OpponentSIFT and GIST as local and global descriptors.

The complexity of emotion modeling was considered in a classification process through psycho-visual tests. We have organized subjective evaluations on 48 images of our database. These tests are divided into two parts separated by a week (24 images scored by test). There are in the two tests some images repeated but with some changes; colour balance modifications or geometric transformations. This is to observe the results of some invariant images processing in the emotional impact extraction. The twenty five observers assessed the nature and the power of emotions felt. For the nature of the emotion they had the choice between "Negative", "Neutral" and "Positive" and the power ranged from "Low" to "High". We chose this test protocol and emotion modeling because according to us, it is easy to score emotion by this way. Our modeling is also equivalent to the dimensional one Valence/Arousal in which, the valence allows to distinguish positive and negative emotions and the arousal that varies from "Low" to "High" defines the degree of arousal. Dimensional models represent more emotions than discrete ones and seem to be more interesting on natural scenes. To limit the influence of the observation time, observers had 8 seconds to evaluate each image. Because our images are low semantic, we do not need to be careful about the interaction between the emotional impact of two consecutive images in the psycho-visual assessments.

With the nature of emotions, we made a classification in three classes of emotions. The average success rate is 56.15 % for artificial neural network and 55.25 % for the SVM classifier; that is really relevant regarding the equivalent results in the literature. We made a marginal classification; one classification process for each category of features. So we had 6 classifiers. The results of the classification process revealed that the features are complementary. Even if some descriptors do not have the best classification rate they stay interesting for some categories of emotions. For example, the success rate of OpponentSIFT is lower than SIFT's but they better recognize negative emotions. So each category of features we extracted covers different information for a best classification. Another interesting analysis is about a link between the percentage of observers who scored an image in one class and the number of classifiers which score the same image in the desired class. In our process when at least 50 % of the observers put an image in one class, at least 3 of our 6 classifiers find the good class on this image.

1. Introduction

Les images sont de réels vecteurs émotionnels. C'est pourquoi elles sont le support privilégié des médias, notamment dans un cadre publicitaire. Nous comprenons aisément que c'est l'un des moyens les plus performants pour véhiculer un message. Essayer d'identifier l'émotion d'une image est une tâche ambitieuse. En effet, différentes informations contenues dans une image (couleurs, contenu, textures...) peuvent être des vecteurs émotionnels. Mais beaucoup d'*a priori* existent sur les liens entre les

images et leur contenu émotionnel notamment sur les couleurs et leur impact. De plus, l'émotion d'une image est fortement liée à son contenu puisque certains lieux (plage, cimetière...) nous renvoient à des émotions différentes. D'autres facteurs encore plus complexes que le contenu ou la couleur globale entrent en jeu dans l'interprétation émotionnelle d'une image. Il s'agit notamment de tous les facteurs culturels et personnels. Ces derniers sont d'autant plus difficiles à prendre en compte puisqu'ils ne pourraient être décrits par des descripteurs objectifs.

L'émotion est une réaction complexe qui engage à la fois le corps et l'esprit. Elle est déclenchée par une évaluation de la signification personnelle du sens de ce qui se produit dans la situation. Elle peut être définie comme étant une réaction affective transitoire d'assez grande intensité, habituellement provoquée par une stimulation venue de l'environnement¹.

On retrouve beaucoup d'autres définitions de l'émotion en fonction des différentes écoles. En effet, le concept d'émotion est utilisé de diverses manières selon qu'il est envisagé en référence à l'aspect stimulus, à l'expérience subjective, à une phase d'un processus, à une variable intermédiaire ou à une réponse. Cependant, globalement, on distingue dans la littérature deux théories psychologiques principales de l'émotion :

- les théories des émotions de base ;
- les théories de l'évaluation.

Dans les théories des émotions de base dont les pères sont, entre autres, Darwin, Ekman, Izard, Plutchik (Rigoulot, 2008 ; Tayari *et al.*, 2009), des émotions dites basiques ou fondamentales sont énumérées. Il faut noter que seulement 5 émotions de base sont communes aux différentes propositions (la tristesse, la colère, la joie, le dégoût et la peur).

Dans les théories de l'évaluation (Rigoulot, 2008) aussi appelées « théories de l'appraisal », les émotions sont définies comme un ensemble limité dans le temps de changements d'états en réponse à l'évaluation d'un stimulus externe ou interne particulièrement pertinent pour l'organisme. Dans ces théories, il est admis qu'un même événement peut provoquer des émotions différentes selon la situation. On s'intéresse donc au déclenchement de l'émotion plus qu'à l'émotion spécifique (problème dit de différenciation) suscitée.

Quelle que soit la théorie, on retrouve généralement deux classifications des émotions (Liu *et al.*, 2011) illustrées par deux exemples sur la figure 1 :

- l'approche catégorielle ou discrète ;
- l'approche dimensionnelle.

Dans l'approche catégorielle, les processus émotionnels peuvent être expliqués par un ensemble d'émotions basiques ou fondamentales, qui seraient innées et communes

1. Définition donnée par le Larousse en ligne <http://www.larousse.fr/dictionnaires/francais/émotion>

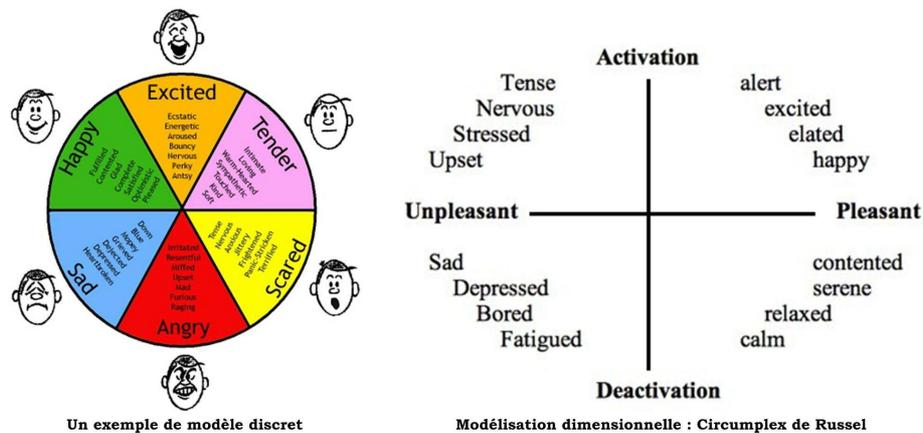


Figure 1. Deux exemples de modélisation des émotions

à toute l'espèce humaine. Il n'y a pas de consensus quant au nombre et à la nature de ces émotions dites fondamentales.

Les modèles de l'approche dimensionnelle se différencient de ceux de l'approche catégorielle par le fait que les émotions résultent d'un nombre fixé de concepts représentés dans un espace multidimensionnel (Liu *et al.*, 2011). Les dimensions peuvent être un axe de plaisir, d'éveil, de puissance. Ces dimensions varient en fonction des besoins du modèle. Le modèle le plus utilisé est celui de Russell avec les dimensions *valence* et *activation* (ou encore *valence-arousal*) représenté par le circumplex de Russel sur la figure 1 :

- **la valence** représente la manière dont se sent une personne quand elle regarde, par exemple, une image. Cette dimension varie du positif au négatif et permet de distinguer les émotions négatives et les émotions agréables ;
- **l'activation** correspond au niveau d'excitation corporel.

L'avantage de ces modèles est de permettre de représenter un très grand nombre d'émotions. Elles sont représentées dans un espace bidimensionnel dont les dimensions varient d'une information trop présente à pas assez. Certaines émotions peuvent être confondues (la peur et la colère par exemple) ou non représentées (entre autre la surprise) dans un modèle bidimensionnel de type Valence/Activation.

Dans la littérature, beaucoup de travaux se basent sur une modélisation discrète des émotions par exemple ceux de (Wei *et al.*, 2008), (Paleari, Huet, 2008), (Kaya, Epps, 2004).

Par ailleurs, plusieurs approches ont été développées pour l'extraction des émotions des images. L'une des techniques utilisée le plus souvent est la détection de visages. Beaucoup de travaux reposent sur l'expression des visages pour déduire l'émotion contenue dans les images. Une émotion est alors associée à des traits du visage (sourcils, lèvres entre autres). De nombreux travaux portent ainsi sur le décodage de

l'expression faciale émotionnelle ; ceux de (Tomkims, 1962), (Scherer, Ekman, 1984) ou encore (Ekman, 1992) (le « mentor » de la célèbre série américaine « Lie to me »). On rappellera également la théorie de Todorov (Todorov, Oosterhof, 2009) selon laquelle les visages auxquels on fait le plus confiance sont joyeux et féminins.

L'autre grande famille de travaux est celle de la détection des émotions basée sur les caractéristiques de l'image (Wei *et al.*, 2008 ; Ou *et al.*, 2004a ; Lucassen *et al.*, 2010 ; Liu *et al.*, 2011). On retrouve le plus souvent les caractéristiques suivantes :

- couleurs ;
- textures ;
- formes présentes (détection de visage, d'animaux).

(Lucassen *et al.*, 2010) par exemple, définissent quatre émotions relatives aux textures qui sont :

- Chaud-Froid ;
- Féminin-Masculin ;
- Dur-Doux ;
- Lourd-Léger.

À partir des résultats obtenus de l'analyse de la variance et des études précédentes faites par (Ou *et al.*, 2004a) (dans lesquelles on retrouve des similarités dans les échelles émotionnelles), (Lucassen *et al.*, 2010) ont mis en place des fonctions permettant d'avoir une explication des variances sur l'échelle de l'émotion couleur. Ils ont établi entre autres des liens entre leurs différentes émotions et les paramètres couleur L^* , c^* et h^* .

Toujours dans le principe d'extraction de caractéristiques, certains auteurs, par exemple (Wang, Yu, 2005), se sont intéressés à la description sémantique des couleurs. À partir de cette information, ils associent à une image une sémantique émotionnelle.

Enfin, l'extraction des formes permet entre autres de pouvoir, là encore, exploiter les expressions du visage (d'êtres humains ou d'animaux) éventuellement contenues dans l'image. Dans certains travaux, les différentes lignes contenues dans l'image et leur orientation sont prises en compte. Par exemple (Liu *et al.*, 2011), évoquent l'idée selon laquelle des lignes obliques communiqueraient du dynamisme et de l'action alors que des lignes horizontales ou verticales du calme et de la relaxation.

Au vu de la littérature, nous avons choisi de nous inscrire dans cette famille d'approches d'extraction de caractéristiques. Pour intégrer les différents aspects difficiles à modéliser (influence culturelle ou encore vécu personnel), nous avons décidé de mettre en place des tests psychovisuels afin de constituer une référence à notre classifieur.

Après cette introduction, l'article est organisé de la façon suivante. Le cadre de travail est exposé dans la section 2 ; nous y décrivons notre base d'images et les tests psychovisuels mis en œuvre. La section 3 est consacrée aux différents descripteurs

utilisés pour la reconnaissance des émotions. Le processus de classification et les résultats sont détaillés dans la section 4. Enfin, nous faisons une synthèse de notre étude et nous donnons des perspectives à nos travaux dans la section 5.

2. Cadre de travail

2.1. Base d'images

Puisqu'aucune base universelle d'images de tests n'existe encore dans le domaine de l'extraction de l'impact émotionnel, il est très important de pouvoir constituer la base la plus diversifiée possible.

Notre base d'images a ainsi été construite à partir d'images libres de droit et gratuites dans sa grande majorité. La base d'images utilisée dans (Gbèhounou *et al.*, 2012)² a été étendue à 350 images. Elle comprend des paysages, des animaux, des personnages, des aliments, des bâtiments, comme l'illustre la figure 2. C'est une base faiblement sémantique qui présente l'avantage de présenter des images du quotidien. Certes, elle se différencie des autres bases utilisées puisqu'elle ne contient pas d'émotions suggérées. De ce fait, elle nous permet de pouvoir organiser des tests sans nous soucier des interactions possibles entre des émotions suscitées par des images successives. En effet, des émotions trop fortes influencent l'évaluation des images suivantes. Les différentes bases d'images souvent utilisées dans l'étude des émotions se distinguent entre elles par leur contenu. Il existe entre autres des bases d'images abstraites et de photographies (Machajdik, Hanbury, 2010) ou encore des bases d'images utilisées en psychologie comme la base IAPS (Lang *et al.*, 2008). La base IAPS (*International Affective Picture System*) du Centre d'études de l'émotion et de l'attention de l'Université de Floride³ est très sémantique alors que celles des images abstraites de Machajdik et Hanbury nécessitent une sensibilité à l'art pour l'évaluation. Notre but est de pouvoir effectuer nos évaluations sur une base la plus accessible possible et surtout composée d'images naturelles⁴.

Actuellement, on retrouve dans la base plusieurs images identiques par rapport au contenu mais différentes puisqu'ayant subi une transformation particulière (transformation géométrique, modification de la balance des couleurs). Même si ces opérations altèrent, dans certains cas, le caractère « naturel » de l'image, elles nous permettent d'évaluer le comportement de certains descripteurs classiques de traitement d'images que nous avons choisis.

Pour associer à chaque image une « vérité terrain » quant à son contenu émotionnel, nous avons décidé de mettre en place des évaluations psychosensorielles significatives.

2. Le présent article étant une version étendue de ce dernier.

3. <http://csea.php.ufl.edu/media.html>

4. Les résultats des tests menés sur la base d'images sont accessibles sur <http://francois.lecellier.org/emotions>.



Figure 2. Quelques images de la base

2.2. Tests psychovisuels et leur exploitation

Vingt-cinq observateurs ont participé volontairement aux différents tests ; 28 % étaient des femmes et 72 % des hommes. La moitié était âgée de 18 à 24 ans, les autres âgés de 25 ans à plus de 50 ans. Au cours de ces différents tests les observateurs ont évalué les mêmes images dans les mêmes conditions. La stratégie de nos tests est intimement liée au modèle d'émotions choisi. Nous avons décidé de travailler sur un modèle bidimensionnel dont les deux dimensions sont :

- la nature de l'émotion ;
- la puissance de l'émotion.

Ce choix de modélisation se justifie par notre souhait de décorréler l'évaluation de l'impact émotionnel des mots traduisant les émotions basées essentiellement sur l'expression du visage (par exemple la surprise, la joie, la colère, la tristesse). En effet, l'expression du(des) visage(s) éventuellement contenu(s) dans les images n'est pas l'information principale sur laquelle nous nous sommes appuyés. Aussi cette modélisation nous permet d'avoir un protocole de tests facile et accessible à tous les observateurs en fonction des images de notre base. Il faut garder à l'esprit qu'un modèle discret peut avoir des résultats qui varient d'une culture à une autre. En effet, d'une langue à une autre ou d'une culture à une autre, on peut utiliser des mots ou des expressions différentes pour exprimer les mêmes émotions. Nous avons mis en place deux séries de tests effectuées à une semaine d'intervalle. Les observateurs ont évalué la nature et la puissance de l'émotion suscitée par les images.

Comme on peut le voir sur la figure 3, pour définir la nature de l'émotion, les utilisateurs avaient le choix entre « Négative », « Neutre » ou « Positive ». Quant à la puissance, elle variait linéairement de faible à forte.

Notre modélisation émotionnelle est équivalente au modèle Valence/Activation, dans lequel la valence permet de distinguer les émotions négatives et agréables tandis que l'activation, variant de faible à forte, permet de définir le niveau d'excitation corporelle. Les résultats de nos évaluations peuvent être très facilement représentés dans le modèle Valence/Activation comme on peut le voir sur la figure 4. Entre les émotions positives et négatives, se trouvent les émotions neutres. Dans l'idéal ces dernières sont



Figure 3. Application de tests d'évaluation

situées sur l'axe « Low High » puisque l'on se situerait au juste milieu de l'axe « Unpleasant Pleasant ». Mais en réalité à cause des incertitudes on peut également les retrouver de temps en temps sur l'axe « Unpleasant Pleasant ».

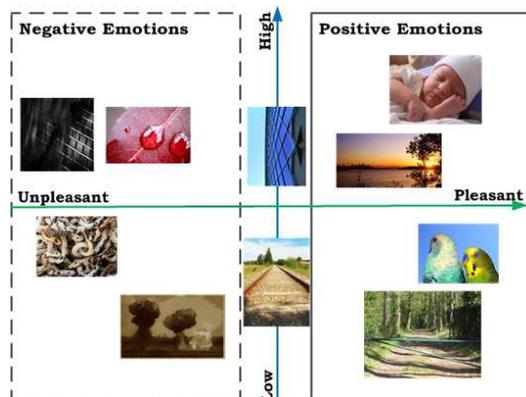


Figure 4. Représentation des résultats de nos évaluations dans le modèle Valence/Activation

Vingt-quatre images étaient présentées dans chacune des deux séries qui composaient les évaluations psychovisuelles. L'ensemble des 48 images évaluées ont été classées de la façon suivante : 5 images neutres, 17 négatives et 26 positives. Les participants disposaient de 8 secondes pour noter chacune d'elles. Si le temps paraît aussi court c'est tout simplement pour augmenter les chances de recueillir des émotions primaires et non des émotions qui découlent d'une éventuelle interprétation trop poussée de la sémantique du contenu de ces images.

Notons que le nombre d'observateurs pendant nos tests est cohérent par rapport au nombre d'images évaluées. Dans notre cas, chacune des 48 images concernées a été évaluée par les 25 observateurs. En effet, sur la base IAPS (Lang *et al.*, 2008) qui est considérée comme une référence dans les études sur les émotions, chacune des 1 196 images de leur base a été évaluée par environ 100 participants. La stratégie de test est également différente puisque dans le cas des évaluations sur IAPS, les participants ont noté 60 images par tests. L'image à évaluer était affichée pendant 6 secondes et ils disposaient de 15 secondes pour les adultes, 20 pour les enfants, pour indiquer leurs émotions selon le système de notation SAM (*Self Assessment Mannequin*).

Dans nos deux séries de tests, certaines images étaient identiques mais présentées avec un traitement différent d'une série à l'autre, voire au sein de la même série, ou encore présentées avec deux traitements différents dans une même série. Les deux traitements appliqués aux images sont :

- le changement de la dynamique des couleurs ;
- la rotation.

Même si ces traitements modifient le caractère naturel de certaines images, ils nous permettent de pouvoir tester l'efficacité de certains descripteurs.

Dans les images des séries de tests, il y avait très peu de visages humains (6,25 %). C'était une façon pour nous d'éviter l'interprétation des émotions liées à ces derniers. L'émotion est une information d'autant plus complexe à modéliser qu'un simple détail sur une image peut tout changer. Ceci est illustré par les images de la figure 5.

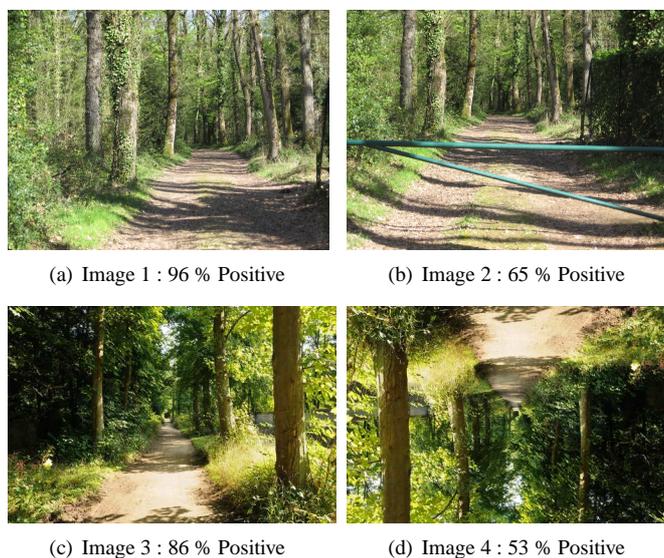


Figure 5. Puissance de l'émotion sur des images naturelles (a)-(b)-(c) et transformée (d)

La présence de la barrière (image 5(b)) sur l'image 5(a) change le taux de classification de cette image lors des tests. Pour un nombre significatif de nos observateurs, la barrière est synonyme d'interdiction de franchissement ; ce qui a donc changé l'émotion qui passe de 96 % à seulement 65 % d'impression positive. La rotation (image 5(d)) appliquée sur l'image 5(c) modifie aussi le taux de classification par les observateurs.

Les résultats des tests ont été intégrés au processus de classification. Une partie a servi pour l'apprentissage. Seules les images ayant un taux de classification dans une catégorie d'émotion (neutre, négative ou positive) supérieur d'au moins 10 % aux autres dans cette catégorie, sont considérées comme classées dans celle-ci. S'il arrivait qu'une image soit classée à 50 % dans 2 classes (cette situation ne s'est présentée dans aucun des tests effectués), elle ne serait pas conservée pour l'exploitation dans le processus de classification.

3. Descripteurs utilisés pour la reconnaissance des émotions

Passons maintenant à l'extraction des paramètres qui nous ont permis d'envisager une classification authentique des images selon les « vérités terrain » données par les tests psychovisuels.

Les couleurs sont les premières caractéristiques discriminantes des émotions. Les différentes hypothèses qui ressortent des études bibliographiques sur le lien entre les émotions et les couleurs concernent le plus souvent le rouge, le vert, le bleu et le jaune. Par exemple, selon (Beresniak, 1990), le rouge est une couleur vivante et excitante, par contre la combinaison rouge+gris provoquerait un sentiment tragique. Le jaune est la couleur la plus gaie, la plus claire, rayonnante et jeune. C'est aussi une couleur tonique et éclatante. La couleur la plus dynamique selon lui est le orange. Cette couleur combine la gaité du jaune et l'action du rouge. Le bleu est une couleur profonde et mystique qui appellerait au calme. En effet, les couleurs correspondent principalement à une certaine interprétation d'une sémantique liée à des situations, à des phénomènes mais aussi à la culture. Elles ne sauraient donc être les seules caractéristiques discriminantes des émotions.

Les textures ont aussi une place importante dans le ressenti émotionnel. Une grille par exemple, quelle que soit sa couleur, a une sémantique d'enfermement ; ou encore, le métal ne fait pas le même effet, que ce soit au toucher ou à la vue, qu'un brin d'herbe. Enfin les descripteurs liés à la forme et aux contours des objets de l'image sont également significatifs. Nous avons aussi supposé que par leur précision, ces descripteurs encoderaient des informations de haut niveau intéressantes pour l'extraction des émotions.

3.1. Couleurs

Pour identifier les différentes couleurs, nous avons utilisé une segmentation couleur par croissance de régions (Fernandez-Maloigne *et al.*, 2004). Pour initialiser les

germes de la croissance de régions, nous avons procédé à une analyse d'histogramme en niveaux de gris. La conversion en niveaux de gris est faite conformément au standard NTSC dont la relation est donnée par l'équation (1).

$$gray = 0.299R + 0.587G + 0.114B \quad (1)$$

L'analyse d'histogramme a été faite sur l'image en niveaux de gris afin de gagner en temps de calcul dans la recherche des zones homogènes. Les germes considérés sont les maxima de l'histogramme en niveaux de gris. Si cette détection des germes a été faite en niveaux de gris, la croissance quant à elle s'est faite dans l'espace couleur CIELAB pour minimiser les problèmes d'apparition de fausses couleurs au moment de la comparaison, par la distance euclidienne, entre la couleur du pixel à agglomérer et la couleur moyenne de la région croissante. Nous n'avons conservé que la couleur moyenne des différentes régions.

3.2. Textures

La caractérisation des textures a été faite sur les images en niveaux de gris (calculées avec l'équation (1)), à l'aide de deux types de descripteurs. Le premier descripteur est l'énergie des caractéristiques de Gabor basée sur la décomposition en bancs de filtres de Gabor (Grigorescu *et al.*, 2002). Dans le domaine spatial, un filtre de Gabor est le produit d'une sinusoïde complexe et d'une enveloppe gaussienne. Un filtre de Gabor 2D est défini en continu par la fonction $g_{\lambda, \Theta, \varphi}(x, y)$ de la façon suivante :

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

avec $x' = x \cos \Theta + y \sin \Theta$ et $y' = y \cos \Theta - x \sin \Theta$.

Pour calculer les caractéristiques de Gabor, nous avons considéré 12 angles différents $\Theta \in [0, \pi[$ tous les $\frac{\pi}{12}$ et deux phases différentes $\varphi \in \{0, -\frac{\pi}{2}\}$ (0 pour le cas symétrique et $-\frac{\pi}{2}$ pour le cas anti-symétrique). Ce qui revient à un total de 24 filtres. Nous avons considéré une gaussienne isotrope ($\gamma = 1$), d'écart type $\sigma = 0,56\lambda$. Ce dernier choix est motivé par les propriétés des cellules du cortex visuel qui peuvent être modélisées par le filtre de Gabor comme le soulignent (Grigorescu *et al.*, 2002). L'énergie des caractéristiques de Gabor est la combinaison des résultats des 12 filtres pour chacune des phases considérées.

Les coefficients Wave Atoms sont le second type de descripteurs que nous avons utilisés pour caractériser les textures. Ils sont basés sur la décomposition introduite par (Demagnet, Ying, 2009).

Les Wave Atoms sont, en première approximation, une variante de paquets d'ondelettes 2D avec une longueur d'onde d'échelle parabolique. C'est la transformée la plus parcimonieuse pour la description des textures localement oscillantes et la seule actuellement pour laquelle un tel résultat est disponible (Demagnet, Ying, 2009).

Comme dans toutes les décompositions multi-échelles (les ondelettes par exemple), on y retrouve plusieurs informations de différents niveaux. Avec la transformée en Wave Atoms, à chaque échelle, le paramètre d'orientation est implicite et varie de manière plus fine qu'avec d'autres transformées (ondelettes par exemple). Le nombre de coefficients par orientation dépend du niveau de décomposition. Avant d'appliquer la décomposition en Wave Atoms, nous avons redimensionné toutes les images en 256×256 en prenant soin de faire du « zero padding » si nécessaire. Avec cette taille d'image, nous obtenons 5 niveaux de décomposition. Nous n'avons conservé que l'échelle 4 qui nous offre un compromis entre le niveau de description et la parcimonie. Elle est composée de 91 orientations ; chaque orientation ayant $2^4 * 2^4$ soit 256 coefficients.

3.3. Descripteurs locaux et globaux

Nous avons utilisé les descripteurs locaux SIFT (dans leur version originale) introduits par (Lowe, 1999 ; 2004) qui ont montré leur efficacité dans de nombreux travaux. L'image est résumée par un ensemble de vecteurs de caractéristiques locales invariantes :

- à la translation ;
- au changement d'échelle ;
- à la rotation ;
- aux changements d'éclairage ;
- aux projections affines ou 3D.

Nous avons aussi testé l'apport des OpponentSIFT dans l'extraction de l'impact émotionnel des images. Ces derniers sont préconisés par (Sande *et al.*, 2010) lorsque l'on n'a pas d'*a priori* sur une base d'images. Ces caractéristiques décrivent tous les canaux couleurs dans un espace « antagoniste » $(0_1, 0_2, 0_3)$ défini par l'équation (3) en utilisant les SIFT. Même si ces derniers sont calculés à partir d'une approche marginale (chaque composante couleur a été traitée indépendamment des autres), ce sont des descripteurs couleurs.

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

Les descripteurs globaux « GIST » ont été introduits par (Oliva, Torralba, 2001). Ils permettent de représenter la structure dominante spatiale de la scène à partir d'un ensemble de dimensions perceptives (la naturalité de la scène, son ouverture, sa rugosité, son expansion, sa robustesse). Dans un souci de réduction de dimension, ces 5 dimensions « perceptives » ne sont pas calculées mais plutôt estimées à partir des informations spectrales et des informations grossièrement localisées.

La première étape pour avoir les GIST est d'opérer un filtrage sur les images avec un banc de filtre de Gabor. Ensuite, ces descripteurs sont calculés sur des imagettes (régions locales) de dimensions paramétrables.

Les descripteurs « GIST » et les OpponentSIFT sont des caractéristiques couleurs puisqu'ils ont été calculés sur les trois plans couleurs que ce soit dans un espace « antagoniste » pour les OpponentSIFT ou dans RGB pour les GIST. Toutes les autres caractéristiques (textures et SIFT) ont été calculées sur les images converties en niveaux de gris, comme décrit précédemment dans l'équation (1).

4. Classification des images

La classification des images est la dernière étape de notre processus d'extraction de l'impact émotionnel des images. Cette étape comporte deux phases :

- *une phase de représentation compacte de l'information* : cette phase est primordiale car après les calculs des différentes caractéristiques, nous obtenons des vecteurs de dimensions importantes qui ne peuvent être directement exploités ;

- *une phase de discrimination des attributs* : la discrimination se fait à l'aide d'un classifieur. Nous avons à cette étape testé deux classifieurs ; dans un premier temps un réseau de neurones multicouches et ensuite un SVM (Chang, Lin, 2011) avec un noyau linéaire. En ce qui concerne le réseau de neurones, le nombre de couches dépendra des dimensions finales des informations à discriminer, suivant la catégorie des caractéristiques.

4.1. Représentation compacte des vecteurs de caractéristiques

Après l'extraction des caractéristiques, comme décrit dans la section précédente, nous disposons d'un nombre important d'informations, comme on peut le voir dans le tableau 1, qui donne la taille moyenne des vecteurs d'attributs par catégorie de descripteurs et par image.

Tableau 1. Taille moyenne par image des vecteurs de caractéristiques avant la réduction de dimension

<i>Catégories de descripteurs</i>	<i>Nombre moyen de descripteurs</i>	<i>Taille des vecteurs</i>
<i>Couleurs</i>	151	3
<i>OpponentSIFT</i>	500	384
<i>GIST</i>	960	1
<i>Gabor</i>	47 648	12
<i>Wave Atoms (Échelle 4)</i>	91	256
<i>SIFT</i>	469	128

Cette étape qui peut s'apparenter à une réduction de dimensions des vecteurs d'attributs est donc très importante pour pouvoir utiliser un classifieur de type réseau de neurones ou SVM sans augmenter considérablement les temps de calcul.

Les différents vecteurs d'attributs ont été normalisés (normalisation L2) avant la phase de réduction de dimensions. Plusieurs solutions s'offrent à nous pour réduire les dimensions, mais elles ne sont pas toutes exploitables dans le cas de l'extraction des émotions. Par exemple, une analyse par composantes principales (ACP) n'est pas très facilement exploitable. En effet, les SIFT et les attributs couleurs n'offrent pas le même nombre de caractéristiques sur toutes les images. Aussi, faire une ACP nous obligerait à avoir un espace de référence et donc à choisir des images de référence (ce choix étant délicat).

La solution que nous avons finalement adoptée, pour réduire les dimensions avant l'étape de classification, est une quantification vectorielle avec un algorithme de type *k-means*. La relation entre le nombre de centroïdes et le nombre de descripteurs est donnée par :

$$k = \sqrt[4]{N * d}, \quad (4)$$

avec k le nombre de centroïdes, N le nombre de descripteurs et d la taille du vecteur de caractéristiques. L'ensemble des centroïdes constitue donc k vecteurs de dimension d . Les dimensions des centroïdes sont récapitulées dans le tableau 2 pour chaque catégorie de descripteurs.

Tableau 2. Configuration des différents centroïdes obtenus après l'algorithme des *k-means*

Catégories de descripteurs	Nombre de centroïdes	Taille des vecteurs
Couleurs	13	3
GIST	15	1
Gabor	72	12
Wave Atoms (Échelle 4)	32	256
SIFT	41	128
OpponentSIFT	55	384

L'hypothèse que nous avons émise en optant pour les centroïdes est qu'en constituant la base la plus diversifiée possible, on pourrait résumer en ces centroïdes des informations assez larges, de manière à pouvoir projeter n'importe quelle image sur ces derniers. Ceci permettrait donc de ne pas recalculer les centroïdes en fonction de la base d'images.

Une fois ces centroïdes calculés, pour chaque image et suivant chaque catégorie de descripteurs, nous avons effectué une quantification vectorielle. Pour la discrimination lors de la classification, nous avons utilisé les histogrammes de répartition, qui équivalent à des « Sacs De Mots » (*Bags of words*) (Jégou *et al.*, 2010) les mots de notre dictionnaire étant les centroïdes. Pour chaque image, à travers le vecteur d'attributs

d'une catégorie de descripteurs donnée, on compte le nombre de fois qu'un centroïde est représenté dans l'image⁵.

Comme les vecteurs de caractéristiques, les histogrammes sont normalisés en utilisant une norme L2 avant la classification.

4.2. Discrimination des images

Comme nous l'avons évoqué dans l'introduction, les tests ont été effectués pour constituer une référence à nos classifications finales. Une partie des tests (59 %) a servi pour l'apprentissage du système de classification et l'autre partie comme séquence de généralisation. Une première étape d'analyse des résultats a été réalisée pour déterminer les différents pourcentages de classification des images dans les trois classes considérées (émotions négatives, neutres et positives). Nous avons effectué une classification marginale (chaque catégorie de descripteurs a été considérée indépendamment sans fusion ni des attributs ni des classifications). Même si les SVM sont conçus à la base pour des problèmes binaires nous avons utilisé une extension en problème multiclasse basée sur une stratégie « un contre un ». Le choix du SVM se justifie surtout parce qu'il se présente dans de nombreux travaux, par exemple ceux de (Liu *et al.*, 2011) comme étant le meilleur classifieur pour les émotions.

Les émotions sont d'autant plus complexes à discriminer que des images différentes peuvent appartenir à la même catégorie d'émotions. On peut observer cette grande diversité sur les histogrammes de répartition des coefficients Wave Atoms (figures 7(b) et 8(a)) de quatre images différentes de l'ensemble d'apprentissage (deux images positives (figures 6(a)-(b)) et deux négatives (figures 7(a)-(b)).

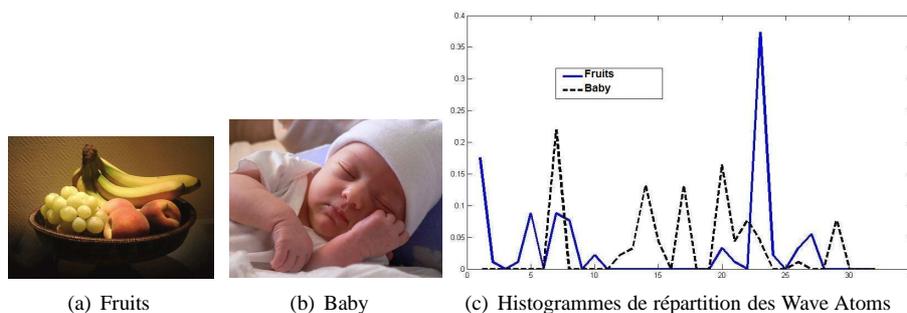


Figure 6. Deux images positives de l'ensemble d'apprentissage et leurs histogrammes de répartition des coefficients Wave Atoms

5. Un centroïde est représenté lorsqu'une ligne d'un vecteur est la plus proche possible de la définition de ce centroïde. Plusieurs centroïdes peuvent donc être représentés dans la même image.

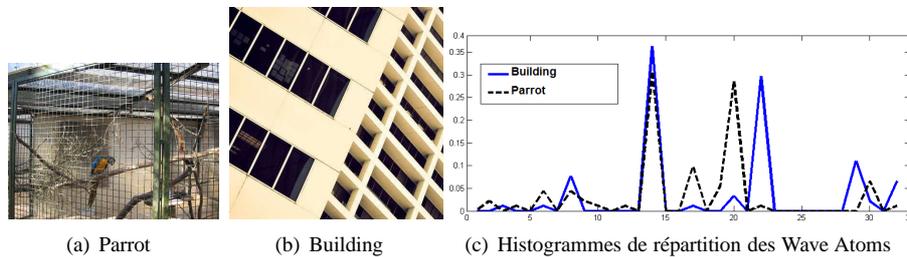


Figure 7. Deux images négatives de l'ensemble d'apprentissage et leurs histogrammes de répartition des coefficients Wave Atoms

4.3. Quelques résultats

Les taux de classification que nous présentons dans le tableau 3 sont ceux obtenus suivant chaque type de descripteurs, sur 19 images qui correspondent à celles des tests psychovisuels qui n'ont pas servi pour l'apprentissage.

Une première approche de classification a été d'utiliser un réseau de neurones. Le taux moyen de réussite sur ces images est de 56,15 % avec un réseau de neurones. C'est un taux de classification intéressant au regard de la diversité des images et de la complexité de discrimination de l'émotion qu'elles contiennent.

Tableau 3. Taux de réussite de classification obtenus avec le réseau de neurones

Catégories de descripteurs	Nombre d'erreurs	Taux de classification
Couleurs	8	57,89 %
GIST	11	42,1 %
Gabor	7	63,15 %
Wave Atoms (Échelle 4)	6	68,42 %
SIFT	8	57,89 %

Le tableau 4 donne les résultats obtenus avec le classifieur SVM. Dans ce processus de classification une légère modification a été apportée avant le calcul des histogrammes de répartition : la normalisation des centroïdes. On ne note que quelques petites différences sur les résultats obtenus (Wave Atoms 47 %, GIST 36,84 % sans normalisation des centroïdes). Le taux de classification moyen avec le SVM est de 55,25 %. Même si le classifieur SVM offre un taux de classification légèrement moins intéressant que le réseau de neurones, nous pouvons dire que les deux types de classifieurs sont équivalents, dans notre cas.

Les meilleurs descripteurs, avec le réseau de neurones, sont les coefficients Wave Atoms pour notre ensemble de test. Même si ces derniers ne sont pas les caractéristiques qui donnent le meilleur taux de classification avec le SVM, ils restent très

Tableau 4. Taux de réussite de classification obtenus avec le SVM

<i>Catégories de descripteurs</i>	<i>Nombre d'erreurs</i>	<i>Taux de classification</i>
<i>Couleurs</i>	8	57,89 %
<i>GIST</i>	9	52,63 %
<i>Gabor</i>	8	57,89 %
<i>Wave Atoms (Échelle 4)</i>	9	52,63 %
<i>SIFT</i>	8	57,89 %
<i>OpponentSIFT</i>	9	52,63 %

satisfaisants. L'échelle 4 contient donc bien suffisamment d'information dans notre cas pour caractériser convenablement les textures.

Les taux de réussite des coefficients de l'échelle 4 de la transformée en Wave Atoms et de l'énergie des caractéristiques de Gabor sont très proches. Ces deux types de descripteurs décrivent la même information dans l'image : les textures. Cette précision obtenue au moment de la classification est certainement liée au nombre d'orientations utilisé et à la configuration de la gaussienne de convolution qui se rapprochent des modèles proposés pour une description proche de la vision par le cortex visuel.

Après les caractéristiques décrivant la texture, les descripteurs locaux SIFT offrent des résultats satisfaisants. Ce sont des descripteurs assez robustes dans l'indexation des images. Ceci justifie sans doute ces résultats intéressants.

On aurait pu penser que la couleur influence énormément les émotions et que la classification suivant les couleurs donnerait les meilleurs résultats. Bien évidemment la couleur est un paramètre important dans l'interprétation des émotions mais nos taux de classification bas avec les descripteurs de couleur sont probablement liés à la nature même des descripteurs, sans doute mal choisis, et à leur exploitation. C'est d'ailleurs l'une des familles de descripteurs qu'il serait intéressant d'étudier différemment lors de la poursuite de nos travaux puisque nous sommes convaincus de pouvoir accroître les taux de classification, par une meilleure exploitation des caractéristiques colorimétriques.

Dans la suite de l'analyse des résultats de classification, nous ne considérerons que les résultats obtenus avec le classifieur SVM. Il nous paraît plus approprié, au vu du nombre d'exemples, à cause du choix d'un noyau linéaire. En effet, il est plus adéquat de choisir un noyau linéaire au vu de la taille des vecteurs de caractéristiques et du nombre d'images dont on essaie de trouver la classe d'émotions. D'une façon globale, tous les descripteurs classent bien les images positives. En observant les matrices de confusion des classifications obtenues pour chaque catégorie de descripteurs avec un SVM (tableaux 5(a) à 5(f)) on comprend mieux cette complémentarité qui existe entre les descripteurs en fonction de la nature des émotions.

L'ajout des OpponentSIFT apporte une information différente puisque les erreurs commises ne sont pas les mêmes qu'avec les SIFT. Les OpponentSIFT (tableau 5(f))

Tableau 5. Matrices de confusion du classifieur SVM pour chaque catégorie de descripteurs

(a) Couleurs				(b) GIST			
	<i>Nég</i>	<i>Neut</i>	<i>Pos</i>		<i>Nég</i>	<i>Neut</i>	<i>Pos</i>
<i>Nég</i>	25 %	0	75 %	<i>Nég</i>	37,5 %	0	62,5 %
<i>Neut</i>	0	0	100 %	<i>Neut</i>	100 %	0	0
<i>Pos</i>	10 %	0	90 %	<i>Pos</i>	30 %	0	70 %

(c) Gabor				(d) Wave Atoms (Échelle 4)			
	<i>Nég</i>	<i>Neut</i>	<i>Pos</i>		<i>Nég</i>	<i>Neut</i>	<i>Pos</i>
<i>Nég</i>	25 %	0	75 %	<i>Nég</i>	37,5 %	0	62,5 %
<i>Neut</i>	100 %	0	0	<i>Neut</i>	0	100 %	0
<i>Pos</i>	10 %	0	90 %	<i>Pos</i>	40 %	0	60 %

(e) SIFT				(f) OpponentSIFT			
	<i>Nég</i>	<i>Neut</i>	<i>Pos</i>		<i>Nég</i>	<i>Neut</i>	<i>Pos</i>
<i>Nég</i>	37,5 %	0	62,5 %	<i>Nég</i>	50 %	0	50 %
<i>Neut</i>	0	0	100 %	<i>Neut</i>	100 %	0	0
<i>Pos</i>	20 %	0	80 %	<i>Pos</i>	40 %	0	60 %

Nég → émotions négatives, *Neut* → émotions neutres, *Pos* → émotions positives

semblent apporter une information plus intéressante pour classer les images négatives même si les images positives sont moins bien classées par rapport aux résultats des SIFT (tableau 5(e)). On peut noter également la difficulté pour nos descripteurs à ce stade de classer les images neutres. Elles sont indiquées soit positives soit négatives sauf avec les Wave Atoms. Ceci peut venir du fait que nous ne disposons pas de suffisamment d'images neutres à l'apprentissage et aussi dans l'ensemble de test. Il faut aussi prendre en compte le fait que la définition d'une émotion neutre n'est pas évidente et les images neutres sont très éclectiques.

Les matrices de confusion (tableaux 5(a) à 5(f)) confirment bien la complémentarité qui existe entre ces caractéristiques dans la discrimination des émotions. On peut alors imaginer que dans une optique de combinaison de la réponse de nos classifieurs, comme l'ont fait (Liu *et al.*, 2011), ces informations différentes permettraient d'arriver à une classification plus intéressante.

On note également d'après le tableau 6 que les images le plus souvent bien classées par un grand nombre d'attributs (plus de la moitié ; ce qui correspondrait dans une stratégie de combinaison de vote de la majorité à la classe finale de l'image) sont celles ayant été classées dans cette classe par au moins 50 % des observateurs. Il y a néanmoins quelques exceptions, il s'agit des images ayant subi des modifications de la balance des couleurs (panier de fruits décolorés par exemple) : malgré l'apprentissage, ces images ne sont pas bien classées. Elles sont désignées membres de la classe

Tableau 6. Pourcentage d'images bien classées par les classifieurs en fonction du pourcentage d'observateurs

Pourcentage d'observateurs	Pourcentage d'images bien classées			
	Par aucun classifieur	Par moins de 3 classifieurs	Par 3 classifieurs	Par plus de 3 classifieurs
[35 %, 50 %[28,57 %	14,29 %	28,57 %	14,29 %
[50 %, 60 %[0	28,57 %	28,57 %	42,86 %
[60 %, 70 %[0	0	0	100 %
≥ 70 %	0	50 %	0	50 %

de l'image avant modification. Ceci c'est produit sur des images qui après modification de la balance des couleurs ont été désignées négatives au lieu de positives avant cette opération. On s'en doutait bien mais les qualités d'apprentissage du système de classification dépendent fortement de la nature de l'ensemble d'apprentissage. Ceci suppose une cohérence plus importante et plus intéressante dans les réponses des observateurs malgré la grande diversité des images d'une même classe d'émotion.

4.4. Autres travaux dans le domaine de l'extraction de l'impact émotionnel

L'objectif de cette section n'est pas de déterminer quelle méthode est la meilleure sur toutes les bases mais de vérifier que nos taux de classification sont cohérents avec ceux de la littérature. En effet, il est difficile de comparer les travaux sur l'extraction de l'impact émotionnel entre eux, à cause des différences entre les stratégies de travail et les bases de tests. On peut, néanmoins, évoquer quelques taux de classification obtenus dans les travaux précédents. Par exemple les taux de classification obtenus par (Wei *et al.*, 2008) sont compris entre 42,3 % et 50,25 %. Ils proposent une classification émotionnelle des images basée sur des aspects sémantiques de la couleur.

Le taux de réussite de (Liu *et al.*, 2011) est de 52 % sur leur base de tests initiale⁶. Pour obtenir ces résultats, les caractéristiques utilisées sont les suivantes :

- les moments couleur ;
- les histogrammes de couleur ;
- les corrélogrammes de couleur ;
- les caractéristiques texture de Tamura ;
- l'histogramme des orientations des lignes et des contours ;
- les caractéristiques d'harmonie ;
- les caractéristiques dynamiques.

Ces résultats sont obtenus avec 80 % de la base de référence utilisée pour l'apprentissage et 20 % pour les tests. (Liu *et al.*, 2011) ont utilisé les mêmes descripteurs

6. Base construite à partir du site <http://www.gettyimages.com/>

sur la base IAPS (Lang *et al.*, 2008). Pour cette base de données qui est considérée comme une référence pour les études psychologiques, le taux de classification moyen après fusion avec la théorie de l'évidence est de 54,70 % contre 52,05 % pour une fusion avec une stratégie de Majority Voting. Dans leur approche, ils se basent sur une classification dimensionnelle des émotions selon des critères d'activités et d'appréciation.

(Machajdik, Hanbury, 2010) obtiennent de meilleurs résultats sur la base IAPS. Leur meilleur taux de réussite dans les différentes classes d'émotions (modélisation discrète choisie : 7 émotions) est de 68 %. Les différentes caractéristiques utilisées sont :

- caractéristiques colorimétriques ;
- caractéristiques liées à la texture ;
- caractéristiques liées à la composition de la scène (profondeur des détails) ;
- caractéristiques liées au contenu (détection de visages, de régions de peau).

Au regard de ces différentes stratégies et différentes bases de tests, nos résultats s'avèrent donc intéressants. Le gros problème pour comparer les résultats, même en ayant travaillé sur la même base, reste la répartition des images. Les résultats peuvent, en effet, être différents, si deux auteurs n'ont pas le même ensemble d'apprentissage (et donc de tests).

5. Conclusions et perspectives

L'extraction de l'impact émotionnel des images nécessite la prise en compte de plusieurs paramètres. Nous avons modélisé au mieux ces paramètres par les attributs suivants :

- les différentes couleurs présentes dans les images ;
- les différentes textures des images ;
- le contenu des images grâce à des descripteurs locaux et globaux.

Nos travaux ont la particularité d'avoir été faits sur une base d'images diversifiée que nous allons mettre à la disposition de la communauté image. Nous avons également utilisé des descripteurs récents, précis et très efficaces chacun dans leur domaine d'utilisation. Ils ont permis de déboucher sur des taux de classification très encourageants. Nous obtenons ainsi des taux de réussite moyens de 56,15 % pour le réseau de neurones et 55,25 % pour le SVM. Ces résultats sont intéressants dans la mesure où ils confirment que nos descripteurs sont cohérents par rapport à l'information émotionnelle contenue dans les images. Les résultats de classification seront également mis en ligne très prochainement pour une illustration représentative de nos travaux.

L'une des perspectives essentielles à nos travaux serait d'évaluer notre système d'extraction de l'impact émotionnel sur la base IAPS pour se comparer efficacement aux autres travaux du domaine. Aussi envisageons-nous de faire évaluer toute notre base d'images afin de pouvoir y effectuer des études entre les profils et les réponses.

Par ailleurs, les couleurs ont été désignées comme un important vecteur émotionnel dans les images, dans de nombreux travaux que nous avons référencés de manière non exhaustive. Même si aucun résultat unanime n'existe sur le lien exact entre les différentes couleurs et les émotions, cela reste une piste intéressante dans l'extraction de l'impact émotionnel. Il serait logique de poursuivre nos travaux en explorant une autre modélisation de l'information couleur afin d'en tirer un maximum de caractéristiques discriminantes pour les émotions. Un dernier aspect qui n'est pas directement étudié, mais auquel serait liée l'émotion d'une image serait l'harmonie des couleurs. À l'harmonie s'ajouterait aussi la préférence des couleurs. Ces deux aspects ont été étudiés par (Ou *et al.*, 2004b) sur des échantillons unicolores mais également bicolores. Cette information pourrait être intégrée à nos travaux afin d'améliorer les résultats mais également de vérifier si cette hypothèse se vérifie sur des images faiblement sémantiques.

Par ailleurs, nous évaluerons des stratégies plus intéressantes et mieux adaptées d'exploitation de chaque catégorie de descripteurs afin d'utiliser les informations de la façon la plus optimale possible.

L'intégration des différents descripteurs de l'attention visuelle (Le Meur, 2005) pourrait être également intéressante pour améliorer nos résultats. Ceci pourrait être une piste à explorer dans le but d'obtenir des résultats conformes à une évaluation humaine. Nous avons, à cet effet, prévu d'organiser de nouveaux tests psychovisuels avec un oculomètre afin d'évaluer l'attention visuelle des observateurs en fonction des différentes émotions.

Bibliographie

- Beresniak D. (1990). *Abc des couleurs leurs incidences dans votre vie quotidienne*. Jacques Granger.
- Chang C.-C., Lin C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, vol. 2, p. 27:1–27:27. (Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>)
- Demagnet L., Ying L. (2009). Wave atoms and time upscaling of wave equations. *Numerische Mathematik*, vol. 113, p. 1-71.
- Ekman P. (1992). Facial expressions of emotions. *Psychological science*, vol. 3, n° 1, p. 34-38.
- Fernandez-Maloigne C., Trémeau A., Bonton P. (2004). *Image numérique couleur: De l'acquisition au traitement*. Dunod.
- Gbèhounou S., Lecellier F., Fernandez-Maloigne C. (2012). Extraction et analyse de l'impact émotionnel des images. In *Actes de la conférence RFIA 2012*, p. 470-477.
- Grigorescu S., Petkov N., Kruizinga P. (2002, oct). Comparison of texture features based on gabor filters. *Image Processing, IEEE Transactions*, vol. 11, n° 10, p. 1160-1167.
- Jégou H., Douze M., Schmid C. (2010). Représentation compacte des sacs de mots pour l'indexation d'images. In *17^e congrès Reconnaissance des Formes et Intelligence Artificielle (RFIA '10)*.

- Kaya N., Epps H. H. (2004). Color-emotion associations: Past experience and personal preference. *AIC Colors and Paints, Interim Meeting of the International Color Association*.
- Lang P., Bradley M., Cuthbert B. (2008). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual. technical report A-8*. Rapport technique. University of Florida.
- Le Meur O. (2005). *Attention sélective en visualisation d'images fixes et animées affichées sur écran : Modèles et évaluation de performances-applications*. Thèse, École polytechnique de l'Université de Nantes: École doctorale STIM.
- Liu N., Dellandréa E., Chen L. (2011). Evaluation of features and combination approaches for the classification of emotional semantics in images. In *International Conference on Computer Vision Theory and Applications*.
- Lowe D. G. (1999). Object recognition from local scale-invariant features. *International Conference on Computer Vision*, vol. 2, p. 1150-1157.
- Lowe D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, vol. 60, p. 91-110.
- Lucassen M. P., Gevers T., Gijzen A. (2010). Adding texture to color: quantitative analysis of color emotions. In *CGIV 2010 final program and proceedings*.
- Machajdik J., Hanbury A. (2010). Affective image classification using features inspired by psychology and art theory. In *Proceedings of the international conference on multimedia*, p. 83-92. New York, NY, USA.
- Oliva A., Torralba A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, vol. 42, p. 145-175.
- Ou L.-C., Luo M. R., Woodcock A., Wright A. (2004a). A study of colour emotion and colour preference. part i: Colour emotions for single colours. *Color Research & Application*, vol. 29, n° 3, p. 232-240.
- Ou L.-C., Luo M. R., Woodcock A., Wright A. (2004b). A study of colour emotion and colour preference. part iii: Colour preference modeling. *Color Research & Application*, vol. 29, n° 5, p. 381-389.
- Paleari M., Huet B. (2008). Toward emotion indexing of multimedia excerpts. In *Proceedings on Content-Based Multimedia Indexing, International Workshop*, p. 425-432.
- Rigoulot S. (2008). *Impact comportemental et électrophysiologique de l'information émotionnelle en vision périphérique*. Thèse, Université de Lille II - École Doctorale de Biologie-Santé.
- Sande K. E. A. van de, Gevers T., Snoek C. G. M. (2010). Evaluating color descriptors for object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, n° 9, p. 1582-1596.
- Scherer K. R., Ekman P. (1984). *Approaches to emotions* (L. E. Associates, Ed.). Lavoisier.
- Tayari I., Le Thanh N., Ben Amar C. (2009, Décembre). *Modélisation des états émotionnels par un espace vectoriel multidimensionnel*. Rapport technique. Laboratoire Informatique, Signaux et Systèmes de Sophia-Antipolis.
- Todorov A., Oosterhof N. N. (2009). Shared perceptual basis of emotional expressions and trustworthiness impressions from faces. *Emotion*, vol. 9, n° 1, p. 128-133.

- Tomkims S. S. (1962). *Affect imagery consciousness : Volume 1: The positive affects*. Springer Publishing Company.
- Wang W.-N., Yu Y.-L. (2005). Image emotional semantic query based on color semantic description. In *The fourth International Conference on Machine Learning and Cybernetics*, vol. 7, p. 4571-4576.
- Wei K., He B., Zhang T., He W. (2008). Image emotional classification based on color semantic description. In C. Tang, C. Ling, X. Zhou, N. Cercone, X. Li (Eds.), *Advanced Data Mining and Applications*, vol. 5139, p. 485-491. Springer Berlin/Heidelberg.

Syntyche Gbèhounou a obtenu son Master en Informatique et en Télécommunications à l'Université de Poitiers en Septembre 2011. Elle fait sa thèse en indexation et en catégorisation des images et des vidéos au Laboratoire Signal, Image et Communications (SIC), rattaché à l'UMR XLIM depuis Octobre 2011. Ses travaux sont axés sur de nouveaux descripteurs pouvant améliorer les taux de reconnaissance en indexation et en catégorisation en incluant par exemple des informations liées à l'impact émotionnel mais également à la saillance.

François Lecellier est Maître de Conférences à l'Institut Universitaire de Technologie, Université de Poitiers depuis 2010. Il est membre du Laboratoire Signal, Image et Communications (SIC), rattaché à l'UMR XLIM. Ses projets de recherche concernent la recherche d'images, les bases de données multimédias, la détection de copie vidéo et le traitement d'images. Il a aussi étudié la segmentation d'images et de vidéos basée contours actifs.

Christine Fernandez-Maloigne est Professeur à l'Université de Poitiers où elle dirige le Laboratoire Signal, Image et Communications (SIC), rattaché à l'UMR XLIM depuis 2008. Son expertise porte sur la couleur numérique, en intégrant des modèles du Système Visuel Humain, sur toute la chaîne de traitement des images et des vidéos.