Reconnaissance de la sémantique émotionnelle portée par les images basée sur la théorie de l'évidence.

N. Liu[1], E. Dellandrea[1], B. Tellez[2], L. Chen[1].

[1]Université de Lyon, CNRS, École Centrale, LIRIS, UMR5205, F-69134, France

[2]Université de Lyon, CNRS, Université Lyon1, LIRIS, UMR5205, F-69622, France

{ningning.liu, emmanuel.dellandrea,liming.chen}@ec-lyon.fr,
bruno.tellez@liris.cnrs.fr

Résumé

La reconnaissance de la sémantique émotionnelle d'une image prend une place de plus en plus importante dans la communauté de recherche. Elle offre en effet des perspectives nouvelles et motivantes pour retrouver et classer des images selon la charge émotionnelle qu'elles peuvent porter. Cependant, comme tout sujet émergeant, les contributions sur ce thème demeurent relativement rares et beaucoup de pistes doivent être étudiées. Dans cet article, nous nous proposons d'évaluer l'efficacité de différents types de descripteurs et de classificateurs pour la reconnaissance d'émotions visuelles dans les images. Dans un second temps, nous proposerons l'utilisation de la théorie de l'évidence de Dempster-Shafer qui permet la manipulation et la fusion de connaissances ambigües et incertaines telles que celles rencontrées dans le traitement des émotions. Les expérimentations menées sur la base d'images IAPS mettent en évidence l'efficacité de cette approche.

Mots-clefs

Emotion, image, classification, théorie de l'évidence, descripteurs visuels.

1 Introduction

Un des buts de l'informatique, et particulièrement de l'intelligence artificielle est d'élaborer des ordinateurs intelligents qui ont la capacité d'interagir avec des êtres humains de façon naturelle. Dès lors, une des questions essentielles est de permettre aux ordinateurs de reconnaitre, de comprendre et d'exprimer des émotions [1]. Plusieurs travaux ont été faits depuis plusieurs années sur ces aspects en informatique mais également en robotique. Quand il s'agit de reconnaitre des émotions (voir [2] pour un tour d'horizon très complet), les recherches portent principalement sur la reconnaissance d'affects dans des données audio (parole ou musique) et sur la reconnaissance visuelle d'expressions faciales. Très peu de contributions traitent de la reconnaissance de la sémantique émotionnelle portée globalement par les images que ce soit par ses couleurs, sa composition ou tout autre élément qui peut provoquer une émotion. Face à ce sujet de recherche émergeant, un grand nombre de questions doivent être abordées concernant principalement les trois problèmes suivants : la représentation des émotions, l'extraction de caractéristiques visuelles nécessaire à la reconnaissance des émotions et les modèles de classification pour traiter les différentes propriétés des émotions [3, 4, 5]. En effet, comme dans tous les autres problèmes de vision par ordinateur, la principale difficulté consiste à franchir le fossé sémantique qui existe entre les descripteurs bas-niveau extraits des images et les concepts sémantiques de haut-niveau qui sont dans notre cas les émotions.

Dans cet article, nous nous proposons d'étudier l'efficacité de différents types de descripteurs visuels ainsi que les classificateurs nécessaires à la reconnaissance d'émotions dans les images. De plus, nous proposerons d'utiliser la théorie

de l'évidence de Dempster-Shafer [15,16], qui permet la manipulation de connaissances ambigües et incertaines comme celles relatives aux émotions.

Le reste de l'article est organisé de la façon suivante. Les différents modèles pour représenter les émotions sont décrits dans la partie 2. Les propriétés des images utilisées pour caractériser les émotions sont présentées dans la partie 3. Les classificateurs testés et choisis pour reconnaitre les émotions sont détaillés dans la partie 4. Les expérimentations sont présentées dans la partie 5. Enfin, nous ferons une synthèse de notre étude dans la partie 6.

2 Représentation des émotions

Plusieurs modèles ont été étudiés dans la littérature pour représenter les émotions [2]. Les deux principales approches sont le modèle discret et le modèle dimensionnel. Le premier modèle consiste à choisir des noms ou des adjectifs pour décrire les émotions, tels que le bonheur, la tristesse, la peur, la colère, le dégoût et la surprise. Le second modèle décrit les émotions selon une ou plusieurs dimensions où chacune représente une caractérisation de l'émotion, les plus utilisées étant l'appréciation, l'activité ou le contrôle. Ce deuxième modèle permet de représenter un plus large éventail d'émotions que le premier.

Le choix de la représentation émotionnelle est généralement guidé par l'application. Ainsi, les deux approches sont utiles et peuvent même être combinées, car elles peuvent apporter des informations complémentaires. Dans cet article, nous proposons une représentation hybride comme l'illustre la figure 1. Chaque image est ainsi représentée comme un point de l'espace constitué des deux dimensions que sont l'appréciation (variant de très déplaisante à très plaisante) et l'activité (variant de très calme à très dynamique). Cet espace est divisé en quatre quadrants permettant d'obtenir quatre types d'émotions distinctes afin de caractériser la charge émotionnelle de chaque image.

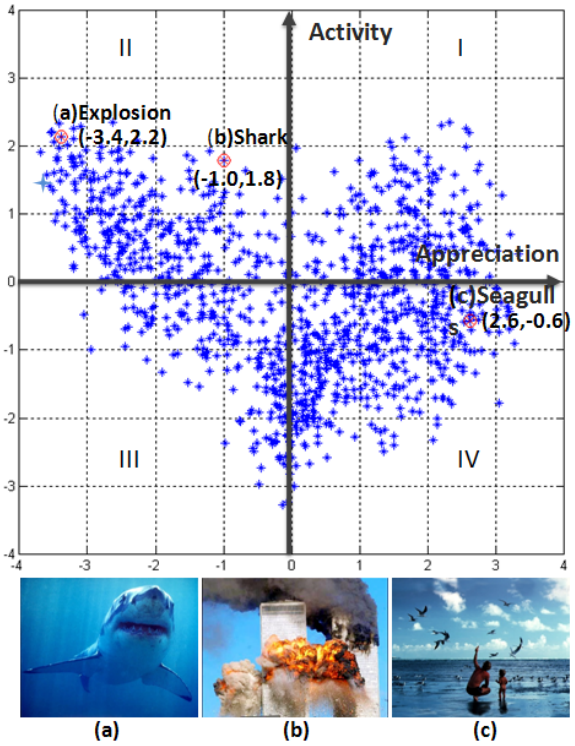

Figure 1 –*Représentation des images de la base IAPS selon des critères d'activité et d'appréciation [8].*

A terme, il serait même envisageable de réaliser un découpage plus fin de cet espace et d'associer à chaque région, un label spécifique représentant une émotion.

3 Descripteurs d'images pour la reconnaissance des émotions

L'extraction des caractéristiques propres d'une image est une question clé pour la reconnaissance de concepts dans des images, et en particulier, les émotions. Ces caractéristiques doivent porter les informations nécessaires pour permettre la reconnaissance des différents concepts. Comme la reconnaissance des émotions dans les images est un domaine de recherche émergent, très peu de travaux ont été réalisés pour identifier les caractéristiques de l'image qui sont les plus efficaces dans ce contexte.

3.1 **Descripteurs d'images traditionnels**

La plupart des travaux traitant de la reconnaissance des émotions utilisent les descripteurs qui le sont généralement également pour d'autres problèmes de vision par ordinateur. Les trois principales catégories de descripteurs d'images sont basées sur la couleur, la texture et la forme. En ce qui concerne la couleur, des études ont montré que l'espace HSV (Hue, Saturation, Value) est un espace de couleur qui est mieux adapté à la perception réelle des couleurs par l'homme que d'autres espaces tels que l'espace RGB traditionnel. Ainsi, sur la base de cet espace de couleur, plusieurs façons de décrire le contenu couleur des images peuvent être considérées tels que les moments de couleurs, les corrélogrammes et histogrammes de couleur ainsi que les histogrammes relatifs à la température de la couleur [10, 11].

En ce qui concerne la texture, la principale caractéristique demeure les matrices de cooccurrences [11,12]. Toutefois, les descripteurs de Tamura [12] peuvent également représenter une alternative intéressante. En effet, des descripteurs tels que la granularité, le contraste ou la directionnalité se sont avérés fortement corrélés avec la perception visuelle de l'homme.

Enfin, la description des formes peut être envisagée grâce à l'extraction des contours permettant l'obtention de l'histogramme d'orientation des lignes [6,12] ou encore les descripteurs de Haar [10,13].

3.2 **Descripteurs sémantiques de l'image pour la reconnaissance des émotions**

Certaines tentatives ont été faites pour identifier des descripteurs de plus haut-niveau liés aux émotions. En effet, les études sur les peintures ont mis en évidence la portée sémantique des couleurs et des lignes qui y apparaissent, comme cela est rappelé dans les travaux de [6] où sont proposés des descripteurs d'images plus corrélés aux émotions grâce à l'exploitation de ces informations. Ainsi, en utilisant la théorie des couleurs d'Itten, une signification émotionnelle des couleurs peut être dégagée. Tout d'abord, comme mentionné plus haut, les couleurs sont décrites en terme de teinte, de luminance et de saturation grâce à l'espace de couleur HSV, afin de se rapprocher de la perception humaine des couleurs. Ces couleurs sont ensuite projetées sur un cercle chromatique, appelé cercle d'Itten où les couleurs fortement contrastées ont des coordonnées opposées par rapport au centre du cercle. Itten a montré que les combinaisons de couleurs peuvent produire des effets tels qu'une harmonie, une disharmonie, du calme ou de l'excitation. Ainsi, l'harmonie sera détectée sur le cercle d'Itten si les positions des couleurs connectées entre elles constituent un polygone régulier comme montré dans la figure 2. Le descripteur correspondant à cette hypothèse est obtenu en mesurant la distance entre le centre du cercle d'Itten et le centre du polygone reliant les couleurs dominantes de l'image. Ces dernières sont préalablement obtenues par un algorithme basé sur les k-means.

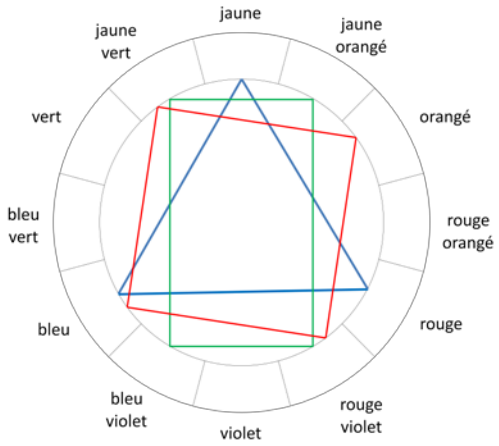

Figure 2. *Cercle d'Itten et exemples d'associations de couleurs harmonieuses (en rouge, vert ou bleu).*

Les lignes portent également une information sémantique importante sur les images. En effet, des lignes obliques suggèrent le dynamisme et l'action tandis que les lignes horizontales ou verticales communiquent plutôt le calme et la détente. Pour exprimer cela en terme de descripteurs d'images, les lignes sont d'abord extraites grâce à une transformée de Hough, puis le rapport entre le nombre de lignes obliques et

le nombre total de lignes dans une image est calculé.

4 Modèles de classification pour la reconnaissance des émotions

La plupart des travaux traitant de la classification des émotions dans les images reposent sur des approches traditionnelles de classification largement utilisée dans d'autres problèmes de vision par ordinateur. Malheureusement, elles ne sont pas toujours appropriées pour traiter de la spécificité des émotions. Parmi ces approches, on peut citer les réseaux de neurones [14], les machines à vecteurs supports (SVM) [9,10,12] ou les modèles par mélange de gaussiennes [10].

4.1 La théorie de l'évidence

Les émotions sont des concepts de haut-niveau sémantique qui sont, par nature, hautement subjectifs et ambigus. Ainsi, afin de s'acquitter efficacement de cette tâche de reconnaissance, il est nécessaire de traiter des informations qui peuvent être incertaines, incomplètes, équivoques et pouvant conduire à des conflits. C'est la raison pour laquelle nous proposons de faire usage de la théorie de l'évidence qui gère naturellement ces difficultés.

4.1.1 *Contexte de la théorie de l'évidence*

La théorie de l'évidence de Dempster-Shafer [15,16] propose un cadre permettant un raisonnement sur des connaissances qui peuvent être incertaines, incomplètes et conduisant à des conflits. Cette théorie s'appuie sur des fonctions de masse qui sont une généralisation des probabilités et des mesures de possibilité.
Soit $\Theta = \{\theta_1, \theta_2, ..., \theta_K)$ un ensemble fini d'hypothèses possibles. Cet ensemble est nommé cadre de discernement, et l'ensemble puissance est désigné 2^{Θ}. Les concepts de base de la théorie sont les suivants :
Fonction de masse de croyance élémentaire : la fonction de masse m, associée à une source d'information donnée (un type de descripteur dans notre cas), attribue une valeur comprise dans l'intervalle [0, 1] pour toute partie A de Θ et remplit les conditions suivantes:

$$m(\emptyset) = 0 \; and \; \sum_{\mathcal{A} \subseteq \Theta} m(\mathcal{A}) = 1 \qquad (1)$$

m (A) représente la confiance, ou croyance, que nous pouvons avoir dans la réalisation d'une hypothèse A.
Les éléments focaux sont des sous-ensembles A tels que m(A)> 0. Si m (Θ) = 1 alors la source est totalement incertaine alors que si m (θ_1) = 1 alors la source est parfaite pour l'hypothèse θ_1.
Règle de combinaison : l'un des propriétés les plus intéressantes de la théorie de la preuve réside dans sa capacité à combiner les fonctions de masse différentes issues de plusieurs sources d'information. Considérons $m_1(.)$ et $m_2(.)$ deux fonctions de masse provenant de deux sources d'information indépendantes S_1 et S_2 respectivement. Dès lors, $m_1(.)$ et $m_2(.)$ peuvent être combinées pour obtenir la masse de la croyance engagée sur C ⊆ Θ, C≠ ∅; selon la formule de combinaison suivante (Shafer, 1976):

$$m(\mathcal{C}) = \frac{\sum_{\mathcal{B} \cap \mathcal{A} = \mathcal{C}} m_1(\mathcal{B}) \cdot m_2(\mathcal{A})}{1 - \sum_{\mathcal{B} \cap \mathcal{A} = \emptyset} m_1(\mathcal{B}) \cdot m_2(\mathcal{A})} \quad (2)$$

Une fois que les fonctions de masse des différentes sources d'informations à notre disposition sont combinées en une seule fonction de masse, une décision finale peut être prise en considérant l'hypothèse qui est associée à la valeur la plus élevée.

4.1.2 *Construire l'évidence*

Une des principales difficultés rencontrées lors de l'élaboration d'une méthode de classification basée sur la théorie de l'évidence concerne la manière dont les fonctions de masse sont construites à partir des descripteurs d'images. Dans ce travail, nous avons utilisé l'approche proposée dans [7] qui estime les fonctions de masse à partir de classificateurs en minimisant

l'Erreur Quadratique Moyenne entre les résultats de la classification et les sorties attendues.

5 Expérimentations

Dans nos expérimentations, nous avons utilisé la base de données d'images IAPS qui est une base de référence en psychologie pour l'étude des émotions communiquées par les images [8]. Elle fournit une caractérisation des images selon trois critères en fonction de l'émotion produite : l'appréciation, l'activité et le contrôle. Cette base comporte 1192 images qui peuvent donc être représentées dans un espace dimensionnel des émotions, selon les axes d'appréciation et d'activité. Par commodité, cette représentation des émotions n'est pas utilisée directement, mais est utilisée pour définir 4 classes d'émotions correspondant aux 4 quadrants de la figure 1.
Le corpus IAPS est partitionné aléatoirement en un ensemble d'apprentissage (80% des données, 953 images) et un ensemble de test (20% des données, 239 images). Toutes les expériences sont répétées 10 fois pour obtenir un pourcentage moyen de classification correcte.
Pour évaluer la performance des différents classificateurs pour la reconnaissance des émotions dans les images, nous avons examiné quatre classificateurs représentatifs : machines à vecteurs supports (SVM), réseaux de neurones (Feed-Forward Neural Networks), Adaboost et K-plus proches voisins. Le schéma de classification que nous avons retenu consiste à utiliser deux classificateurs binaires. Le premier est entrainé pour identifier l'activité, et le second sert à identifier l'appréciation. Les résultats sont ensuite combinés pour identifier l'une des 4 classes d'émotion.

Les caractéristiques d'entrée sont générées en utilisant les techniques décrites dans la partie 3 et alignées en un seul vecteur, ce qui correspond à une fusion précoce. Les résultats de classification sont donnés dans le tableau 1. Nous pouvons observer que les classificateurs SVM avec un pourcentage moyen de classification correcte de 62,6% réalisent la meilleure performance parmi les quatre types de classificateurs, même si Adaboost atteint des performances très proches.

	NN (%)	*SVM (%)*	*Adaboost(%)*	*Knn(%)*
I	57.21	61.55	**65.02**	51.33
II	63.42	60.34	62.53	**64.42**
III	58.21	**62.61**	61.31	51.52
IV	61.72	**65.75**	64.30	61.71

Table 2 – *Pourcentages moyens de classification correcte pour 4 classes d'émotion obtenus par les 4 classificateurs.*

Un autre aspect intéressant consiste à comparer la capacité des différents types de descripteurs d'images à porter l'information relative aux émotions. Ainsi, le système de classification basé sur SVM décrit précédemment a été appliqué indépendamment pour chaque type de descripteurs. Les résultats sont donnés dans la figure 3. Cette figure présente également le pourcentage de bonne classification obtenu avec la fusion de tous les descripteurs en s'appuyant sur l'approche fondée sur la théorie de l'évidence présentée à la section 4.2. La première remarque est que la performance entre les différents descripteurs est très similaire, variant de 53,3% pour les corrélogrammes jusqu'à 58,2% pour LBP. Toutefois, parmi les différents descripteurs, il semble que la texture (LBP et la matrice de cooccurrences) soit le type de descripteurs le plus efficace. En outre, les descripteurs de plus haut-niveau (dynamisme et harmonie) même s'ils peuvent paraître moins performants au premier abord ne reposent que sur une seule valeur et donc, leur efficacité est tout à fait remarquable. Enfin, il faut mentionner que l'approche proposée pour la fusion de l'ensemble des descripteurs basée sur la théorie de l'évidence, et dont la matrice de confusion est donnée dans le tableau 2, donne les meilleurs résultats avec un pourcentage moyen de classification correcte de 64,6%. Cette valeur montre la capacité de la théorie de l'évidence à combiner différentes sources d'information et à exploiter leurs complémentarités.

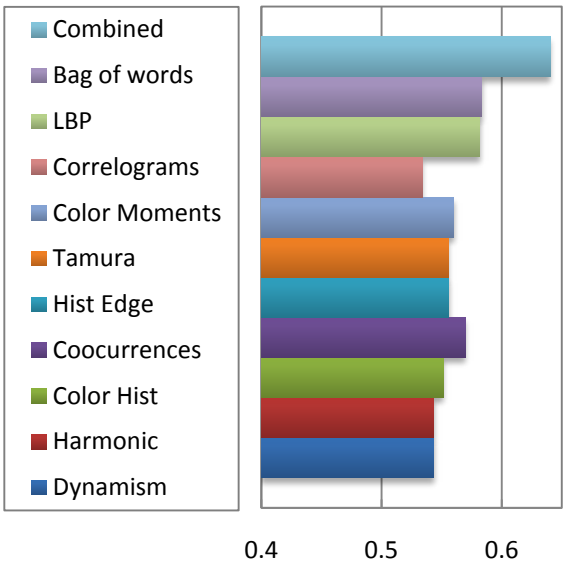

Figure 3 – *Taux de reconnaissance moyen obtenus pour chaque type de descripteurs et par fusion (combined).*

Prédit / Réel	I	II	III	IV
I	**63.32**	12.25	11.23	11.15
II	11.05	**61.42**	12.27	11.82
III	16.21	12.53	**66.19**	10.52
IV	10.42	13.80	10.31	**67.51**
Total	100	100	100	**100**

Table 2 – *Matrice de confusion pour les 4 classes d'émotion en utilisant la théorie de l'évidence.*

6 Conclusion

Dans cet article, nous avons étudié l'efficacité des différents types de caractéristiques et de classificateurs pour la reconnaissance des émotions dans des images. En outre, nous avons proposé une méthode de classification basée sur la théorie de l'évidence, qui présente la capacité de traiter les connaissances ambiguës et incertaines, comme celles qui peuvent caractériser les émotions. Les expériences sur la base de données IAPS ont mis en évidence que, parmi les classificateurs traditionnels, SVM obtient les meilleurs résultats, et que la texture ainsi que les descripteurs de dynamisme et d'harmonie portent des informations importantes liées aux émotions. Enfin, grâce à notre approche basée sur la théorie de l'évidence, nous avons pu atteindre un taux de reconnaissance globale de 64,6% que nous considérons comme encourageant dans un contexte aussi peu exploré que celui de l'identification de la charge émotionnelle portées par les images.

Références

[1] R.W.Picard. Affective Computing. MIT Press, Cambridge, 1997.

[2] Z. Zeng et al. A survey of affect recognition methods: audio, visual and spontaneous expressions. IEEE Transactions on PAMI, 31(1):39-58, 2009.

[3] S. Wang, X. Wang. Emotion semantics image retrieval: a brief overview. ACII, pp. 490-497, 2005.

[4] W. Wei-ning, Y. Ying-lin, and J. Sheng-ming. Image retrieval by emotional semantics: A study of emotional space and feature extraction. IEEE ICSMC, 4, 2006.

[5] W. Wang and Q. He. A survey on emotional semantic image retrieval. ICIP, pp. 117-120, 2008.

[6] C. Columbo, A. Del Bimbo, P. Pala. Semantics in visual information retrieval. IEEE Multimedia, 6(3):38-53, 1999.

[7] Al-Ani, M.Deriche. A new technique for combing multiple classifier using the Dempster Shafer theory of evidence, J. Artif. Intell. Res. 17 (2002), pp. 333-361

[8] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, the IAPS: Technical manual and affective ratings, Tech. Rep., GCR in Psychophysiology, 1999

[9] V.Yanulevskaya, J.C.Van Gemert. et al. Emotional valence categorization using holistic image features. IEEE, ICIP, pp. 101-104, 2008

[10] P. Dunker, S. Nowak, A. Begau, C. Lanz. Content-based mood classification for photos and music. ACM MIR, pp. 97-104, 2008.
[11] C.-T. Li, M.-K. Shan. Emotion-based impressionism slideshow with automatic music accompaniment. ACM Multimedia, pp. 839-842, 2007.
[12] Q. Wu, C. Zhou, C. Wang. Content-based Affective Image classification and retrieval using support vector machines. ACII, pp. 239-257, 2005.
[13] S.-B. Cho, J.-Y. Lee. A human-oriented retrieval system using interactive genetic algorithm. IEEE Transactions on systems, man and cybernetics, 32(3):452-458, 2002.
[14] K. Kuroda, M. Hagiwara. An image retrieval system by impression words and specific object names IRIS. eurocomputing, 43:259-276, 2002.
[15] A.P. Dempster. A Generalization of Bayesian Inference. J. Royal Statistical Soc. Series B, vol. 30, 1968.
[16] G. Shafer. A Mathematical Theory of Evidence. Princeton University Press, 1976.