

Quantification vectorielle d'images couleur par la théorie de l'évidence

Christophe CHARRIER¹

Anne-Laure JOUSSELME²

¹ LUSAC EA 2607 - équipe Vision et Analyse d'Image
Université de Caen Basse-Normandie
120, rue de l'exode 50000 - Saint-Lô
c.charrier@chbg.unicaen.fr

²Section Systèmes d'Aides à la Décision
R & D pour la Défense Canada - Valcartier
2459, Blvd Pie XI Nord, Val-Bélair, QC, G3J 1X5, Canada
Anne-Laure.Jousselme@drdc-rddc.gc.ca

Résumé

La technique de compression par quantification vectorielle peut être améliorée soit en agissant directement sur l'étape de construction du dictionnaire soit en agissant sur l'étape de quantification des vecteurs d'entrée. Dans cette étude, une amélioration de la deuxième étape a été proposée. Par défaut, la quantification s'effectue selon la règle du plus proche voisin qui ne permet pas de prendre en compte le voisinage du vecteur à quantifier ni la sensibilité du système visuel humain (SVH) aux trois primaires colorimétriques. Afin de tenir compte de ce voisinage et de la sensibilité du SVH, la règle des k plus proches voisins à été utilisée sur chacun des axes pris séparément. Les trois classifications étant considérées comme trois sources d'information indépendantes sont alors combinées dans le cadre de la théorie de l'évidence. Le meilleur vecteur code est enfin sélectionné après avoir ainsi pris en considération l'incertitude liée en particulier à la sensibilité du système visuel humain.

Des tests psychophysiques de la qualité ont permis de montrer que la qualité des images a été sensiblement améliorée.

Mots clefs

Quantification vectorielle, perception, théorie de l'évidence, qualité.

1 Introduction

Par définition, la compression par Quantification Vectorielle (QV) [1] accepte un vecteur d'entrée \vec{x} de dimension n et le remplace par un vecteur \vec{y} de même dimension appartenant à un dictionnaire qui est un ensemble fini $(\omega_j)_{j \in [1, \dots, N]}$ de vecteurs code, appelés aussi classes, ou barycentres puisque ceux-ci sont calculés par une moyenne itérative des vecteurs \vec{x} .

L'étape de quantification (selon un dictionnaire construit à partir d'un ensemble d'apprentissage) repose sur la règle

du plus proche voisin : un vecteur \vec{x} à classifier sera affecté à une des classes $(\omega_j)_{j \in [1, \dots, N]}$ sous la condition que cette affectation engendre la plus petite distorsion. Or, ce type d'affectation implique un choix binaire, *i.e.*, le vecteur \vec{x} doit nécessairement appartenir à la classe ω_j dont le barycentre est son plus proche voisin (au sens de la distance euclidienne). Une telle règle d'affectation peut se révéler trop drastique dans les cas où les distances entre le vecteur \vec{x} et deux barycentres sont très proches. Dans ce cas, l'un ou l'autre des barycentres pourrait convenir, selon un ou plusieurs critères qui peuvent être basés sur quelques propriétés du SVH ou sur des critères statistiques autres que la seule distance euclidienne.

Une des améliorations possibles pour éviter cette décision dure, serait de considérer chacune des composantes couleur de manière indépendante, pour obtenir ainsi une classification par composante. Dans notre cadre d'étude, les composantes R, V et B sont alors considérées comme trois sources d'information indépendantes. Dans un premier temps, la règle des k plus proches voisins est appliquée sur chacune des trois composantes, générant alors trois séries de classes potentielles. Cette étape, tenant compte de k voisins et non pas un, permet de considérer l'incertitude selon chacune des composantes couleur et de repousser la prise de décision. Enfin, la décision de la classe d'affectation finale de \vec{x} se fait après avoir combiné ces trois classifications.

Cette technique fait référence aux méthodes de fusion de données. Parmi tous les outils qui s'offrent à nous dans ce domaine, nous décidons d'utiliser la théorie de l'évidence (aussi appelée théorie de Dempster-Shafer ou théorie des fonctions de croyance) [2], qui permet d'une part de traiter des informations incertaines (un maximum de k classes possible sur chacun des axes couleur), et d'autre part de combiner des informations provenant de plusieurs sources (chacun des axes couleur). Dans le cadre de cette théorie, plusieurs règles de décision sont définies pour nous permettre de sélectionner la classe à laquelle sera finalement

affecté \vec{x} .

2 Utilisation de la théorie de l'évidence

2.1 Principe de base

Soit $\Omega = \{\omega_1, \dots, \omega_N\}$, l'ensemble des N classes possibles pour \vec{x} , appelé cadre de discernement et correspondant au dictionnaire dans notre application. Au lieu de restreindre ses mesures à cet ensemble (comme le ferait la théorie des probabilités contrainte par son axiome d'additivité), la théorie de l'évidence s'étend sur l'ensemble puissance de Ω , noté 2^Ω , l'ensemble des 2^N sous-ensembles de Ω . On définit alors une fonction de masse initiale m de 2^Ω dans $[0, 1]$ vérifiant les conditions suivantes :

$$\sum_{A \subseteq \Omega} m(A) = 1 \quad \text{et} \quad m(\emptyset) = 0 \quad (1)$$

où \emptyset est l'ensemble vide. $m(A)$ quantifie la croyance que l'on accorde au fait que la classe cherchée appartienne au sous-ensemble A de Ω (et à aucun autre sous-ensemble de A). Les sous-ensembles A tels que $m(A) > 0$ sont appelés éléments focaux.

Deux fonctions de masse initiale m_1 et m_2 représentant les informations respectives de deux sources différentes, peuvent être combinées selon la règle de Dempster [3] :

$$m(A) = \frac{\sum_{B \cap C = A} m_1(B)m_2(C)}{1 - K}, \quad \forall A \in 2^\Omega, \quad A \neq \emptyset. \quad (2)$$

K est appelé le facteur de conflit et représente le désaccord entre les deux sources. Il correspond à la masse de l'ensemble vide $K = \sum_{B \cap C = \emptyset} m_1(B)m_2(C)$. Notons que la combinaison de Dempster, appelée aussi somme orthogonale et notée $m = m_1 \oplus m_2$ est commutative et associative. Après combinaison, il reste à prendre une décision quant à l'élément le plus "probable" de Ω . Plusieurs règles de décision sont alors possibles, mais une des plus utilisées reste celle du maximum de probabilité "pignistique". Cette règle de décision présentée par Smets [4] utilise la transformation pignistique, qui permet de répartir uniformément la masse associée à un sous-ensemble de Ω sur chacun de ses éléments :

$$\text{BetP}(\omega) = \sum_{\omega \in A \subseteq \Omega} \frac{m(A)}{|A|}, \quad \forall \omega \in \Omega \quad (3)$$

$|A|$ étant le cardinal de A . La décision va alors vers l'élément de Ω dont cette valeur est la plus grande :

$$\omega^* = \text{Arg} \left\{ \max_{\omega \in \Omega} [\text{BetP}(\omega)] \right\}. \quad (4)$$

2.2 Application à la quantification vectorielle

On propose de représenter l'information fournie par chaque classification indépendante (*i. e.* selon chacune des

trois composantes (R,V,B)) par une fonction de masse initiale. Ces trois fonctions de masse initiale $(m_i)_{i \in \{R,V,B\}}$ sont donc créées après calcul des k plus proches voisins et avant décision du vote majoritaire. Elles permettent de prendre en compte l'incertitude associée à chacun des axes en tenant compte en particulier la sensibilité de chacun des axes. Ainsi, on regroupe dans le même élément focal les classes qui se trouvent très proches les unes des autres sur un même axe, et la prise de décision ne s'effectue qu'après avoir combiné les résultats des deux autres projections.

Pour chaque axe, on repère donc les k éléments les plus significatifs selon une distance d_i sur cet axe. La fonction de masse initiale construite selon l'axe i comporte alors trois éléments focaux $A_i, \overline{A_i}$ et Ω où $\overline{A_i}$ est le complément de $A_i \subseteq \Omega$. On construit ainsi

$$A_i = \{\omega \in \Omega, \omega = \text{Classe}(\vec{x}) | d_i(\vec{x}, \vec{x}^*) \leq \varepsilon_i d_i(\vec{x}_1, \vec{x}^*), \forall \vec{x}\}, \quad (5)$$

pour $i \in \{R, V, B\}$. \vec{x}^* est le vecteur à classifier et \vec{x}_1 son plus proche voisin (selon d_i). ε_i est une constante supérieure à 1 permettant de prendre en compte la sensibilité du système visuel humain selon l'axe i : plus la sensibilité sur l'axe i est importante, plus ε_i est élevée et plus le critère d'association de deux éléments doit être strict. À la limite, si $\varepsilon_i = 1$, alors $A_i = \{\text{Classe}(\vec{x}_1)\}$ est un singleton correspondant au plus proche voisin. Les masses sont ensuite assignées aux ensembles A_i en tenant compte de la répartition des éléments dans l'ensemble A_i que l'on représente par la distance moyenne entre deux de ces éléments. La fonction de masse initiale pour l'axe i est alors :

$$m_i(A_i) = \alpha_i e^{-\beta_i \bar{d}} \quad (6)$$

$$m_i(\overline{A_i}) = 1 - m_i(A_i) - m_i(\Omega) \quad (7)$$

$$m_i(\Omega) = 0, 01 \quad (8)$$

où α_i est une constante et $\beta_i = 1/d_{\max}$. d_{\max} est la distance maximale entre \vec{x}^* et les éléments de A_i dans l'espace (R,V,B) et \bar{d} est la distance moyenne entre chacun des éléments de A_i dans l'espace (R,V,B). Ainsi, plus \bar{d} est grand (plus les éléments sont éloignés les uns des autres), plus la masse de A_i est petite. Afin d'éviter un conflit total entre deux sources (deux axes), la masse accordée au cadre de discernement est fixée à une valeur constante.

Soient m_R, m_V et m_B , les trois fonctions de masse initiale issues des projections R, V et B respectivement. La fonction de masse résultante de la combinaison des trois fonctions m_R, m_V et m_B est obtenue à partir de l'équation (2) :

$$m = m_R \oplus m_V \oplus m_B \quad (9)$$

Finalement, la classe d'affectation de \vec{x} est sélectionnée à partir de m (équation 9) sur la base du maximum de probabilité pignistique (équation 3).

3 Mesure des performances

De façon à évaluer l'impact de cette technique sur la qualité des images, nous utilisons des tests psychophysiques de la qualité, basés sur la technique du choix-forcé.

3.1 Conditions expérimentales

Pour mener ces tests, les images utilisées sont des images naturelles dont la complexité varie de faible à texturée. La figure 1 représente un échantillon des images utilisées.

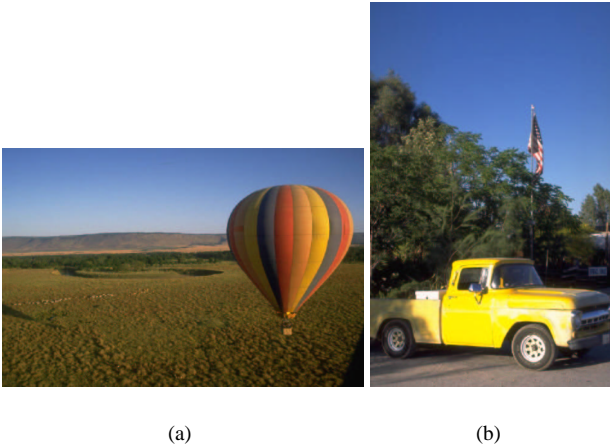


Figure 1 – échantillon d'images utilisées.

Ces images ont ensuite été compressées en utilisant des dictionnaires construits par l'algorithme LBG [5] dont les tailles sont : 32, 64, 128 et 256. Ceci permet d'obtenir des qualités d'images reconstruites variant de mauvaise à bonne selon l'échelle de cotation définie dans la recommandation UIT-R BT.500-7 [6]. À partir de ces dictionnaires, deux catégories d'images ont été générées :

1. R1 représente les images obtenues par la quantification basée sur la théorie de l'évidence,
2. R2 représente les images obtenues par la quantification basée la règle du plus proche voisin.

Les observateurs ont été sélectionnés à l'aide de la mire de SNELLEN afin de contrôler leur acuité visuelle normale ou rendue normale par correction, et de la mire d'ISHIHARA leur pour contrôler leur vision normale des couleurs.

Lors de l'utilisation de ces tests, les observateurs doivent comparer la qualité des images du groupe R1 et R2. Chaque séance comporte une série de présentations qui sont présentées dans un ordre aléatoire ainsi que dans un ordre différent pour chacun observateur.

3.2 Évaluation de la qualité – principe

Les mesures psychophysiques de qualité permettent de définir le seuil de détection des différences entre deux images. Ce seuil caractérise la sensibilité d'un observateur λ , calculée à l'aide de la théorie de la détection.

Dans le cas d'une évaluation de la qualité basée sur le choix forcé, la *mesure de sensibilité* est définie à l'aide des estimés des deux probabilités $p(D)$ (Décision Correcte) et

$p(FA)$ (Fausse Alarme) issues du tableau 1. Cette importante mesure de sensibilité d' , représente simplement la distance entre la moyenne de la distribution du stimulus 1 et de la distribution du stimulus 2. La mesure de la sensibilité d' peut s'exprimer sous la forme suivante [7] :

$$d' = \frac{1}{\sqrt{2}}[z(p(D)) - z(p(FA))] \quad (10)$$

où $z[.]$ est la fonction inverse de la fonction de distribution normale.

Les différences entre les stimuli sont caractérisées par la valeur de d' . Cette différence est imperceptible si la valeur de d' est inférieure à 0,5. Pour une valeur de d' comprise entre 0,5 et 1,0, les différences entre les deux stimuli sont subtiles mais tout de même perceptibles. Enfin, plus la valeur de d' est grande et plus les différences sont remarquables.

	R1	R2
$\langle R1, R2 \rangle$	$p(D)$	$p(FA)$
$\langle R2, R1 \rangle$	$p(O)$	$p(RC)$

Tableau 1 – Matrice Stimulus-Réponses utilisée lors de la procédure de comparaison.

3.3 Évaluation de la qualité – résultats

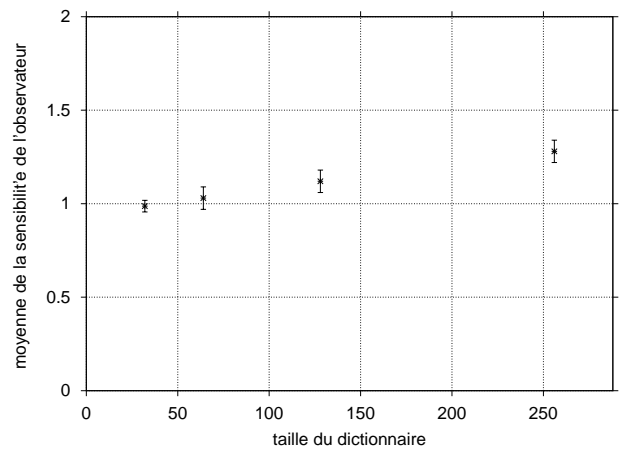


Figure 2 – Mesure de la sensibilité des observateurs en fonction du nombre de vecteurs code présents dans le dictionnaire

La figure 2 représente la valeur moyenne de d' pour chacun des dictionnaires testés. Pour un faible nombre de vecteurs code, la valeur de d' est légèrement inférieure à 1. Ceci se traduit par le fait qu'il n'existe pas de différences de qualité significatives entre les deux images. En observant la tendance générale de la courbe, on remarque que la valeur de d' croît au fur et à mesure que le nombre de vecteurs code augmente.

Ceci peut s'expliquer par le fait que l'utilisation de la théorie de l'évidence, dès lors qu'un nombre élevé de



(a) Image compressée avec la QV classique



(b) Image compressée avec la QV associée à la théorie de l'évidence

Figure 3 – Illustration des résultats obtenus avec un dictionnaire de taille 256 selon a) la méthode de QV classique et b) la méthode proposée.

classes est mis en jeu, permet d'affiner la quantification des vecteurs en fonction de la sensibilité du système visuel humain. La figure 3 illustre ce propos. La figure 3(a) présente le résultat d'une image compressée selon le schéma de la quantification vectorielle classique ainsi qu'un détail de cette image. Les images 3(b) représentent la même image compressée avec la technique de quantification proposée. Les deux "imassettes" de détail correspondent à une partie de l'enveloppe de la montgolfière. En observant les images de détail, on aperçoit des différences de quantification des couleurs :

- dans l'image de détail de la figure 3(b), une zone grise de forme allongée est présente alors qu'elle ne l'est pas dans l'image de détail de la figure 3(a) ;
- si la quantification dans la bande jaune est analysée, des différences entre les deux images sont notables. La transition entre les différentes couleurs jaune est plus "douce" dans l'image de la figure 3(b) que dans l'image de détail de la figure 3(a) ;
- une remarque similaire peut être formulée pour la bande "rose" de l'enveloppe.

D'une manière générale, la quantification obtenue en associant la théorie de l'évidence au schéma de quantification classique permet d'obtenir une qualité globale de compression identique à la quantification vectorielle classique.

4 Conclusion

Afin d'améliorer l'étape de quantification des vecteurs d'entrée en fonction de vecteurs code présents dans un dictionnaire, l'utilisation de la théorie de l'évidence a permis d'obtenir des résultats des plus prometteurs. Dans notre cadre d'étude, une quantification vectorielle a été réalisée sur chacun des trois axes couleurs R, V et B, en fonction de la dispersion colorimétrique des k plus proches voisins. Ceci permet de prendre en compte la sensibilité du SVH aux composantes primaires.

Les résultats obtenus montrent que l'utilisation de la théorie de l'évidence lors de l'étape de quantification apporte une amélioration de la qualité des images reconstruites.

Références

- [1] A. Gersho and R. M. Gray. *Vector Quantization and Signal Compression*. Kluwer Academic Publishers, 1991.
- [2] G. Shafer. *A mathematical theory of evidence*. Princeton University Press, 1976.
- [3] A. Dempster. Upper and Lower Probabilities Induced by Multivalued Mapping. *Ann. Math. Statist.*, 38 :325–339, 1967.
- [4] P. Smets. Constructing the pignistic probability function in a context of uncertainty. *Uncertainty in Artificial Intelligence*, 5 :29–39, 1990. Elsevier Science Publishers.
- [5] Y. Linde, A. Buzo, and R. M. Gray. An algorithm for vector quantizer design. *IEEE Transactions on Communications*, COM-28(1) :84–95, January 1980.
- [6] Recommandation UIT-R BT.500-7. Méthodologie d'Évaluation subjective de la qualité des images de télévision. Technical report, UIT, 1995.
- [7] N. A. Macmillan and C. D. Creelman. *Detection theory : a user's guide*. Cambridge University Press, 1990.