

Indexation et appariement d'images par modèle de mélange gaussien des couleurs

Christophe BIERNACKI, Roger MOHR

Laboratoire Gravir, projet Movi de l'Inria Rhône-Alpes
655 avenue de l'Europe, 38330 Montbonnot Saint Martin, France
Christophe.Biernacki@inrialpes.fr, Roger.Mohr@inrialpes.fr

Résumé – Nous proposons d'approximer la distribution de certains descripteurs d'image, comme la couleur ou la couleur associée à une information de gradient, par d'une part des histogrammes et d'autre part un mélange de gaussiennes. La dimension de cette dernière signature étant choisie automatiquement par des critères statistiques, le gain sur la taille de l'index est souvent important. Concernant la qualité de l'appariement, les histogrammes semblent plus performants.

Abstract – Distribution of some image descriptors, as color or color combined with gradient information, is modeled by histograms or Gaussian mixtures. Dimension of Gaussian mixture signature is often smaller because number of components is automatically chosen by statistical criteria. But histograms lead to higher matching rate.

1 Introduction

Que se soit pour rechercher des images similaires dans des bases d'images, ou pour retrouver des objets dans des scènes vidéos, on cherche à décrire des images par des signatures qui autorisent une classification et une recherche efficaces (cf. le groupe de travail correspondant du GDR ISIS). Les histogrammes sont souvent utilisés à travers la distance du χ^2 (cf. [8]) pour l'efficacité et la simplicité de leur mise en œuvre.

Nous proposons ici une alternative intéressante aux histogrammes, considérant une modélisation par des mélanges de distributions gaussiennes. Dans ce cadre, des outils de choix de modèles permettent de déterminer de manière automatique le nombre de composantes gaussiennes, en d'autres termes la taille de la signature.

En premier lieu, nous présentons et expérimentons les deux approches dans le cadre de la signature des couleurs. Ensuite, nous étendons cette étude en couplant l'information de couleur à l'information spatiale du gradient. Enfin, nous concluons sur les avantages et les limites de notre proposition.

2 Approximation de la distribution des couleurs

2.1 Approximation par histogramme

Une image est composée de n pixels, chacun ayant une couleur c_i ($i = 1, \dots, n$). Cette couleur est un scalaire lorsque l'image est en niveau de gris et un vecteur à trois composantes lorsque l'image est en couleur.

L'histogramme est une approximation de cette distribution des couleurs c_1, \dots, c_n [8]. Cette signature peut s'interpréter en terme de mélange probabiliste: un histogramme en K classes est une loi mélange f avec K lois uni-

formes. On a

$$f(c|\theta) = \sum_{k=1}^K p_k U(c|a_k) \quad (1)$$

avec p_k la proportion de la classe k ($p_k \geq 0$ et $\sum_k p_k = 1$), $U(\cdot|a_k)$ la loi uniforme de support a_k . Le paramètre global θ regroupe l'ensemble des p_k et a_k .

Généralement, le support est fixé a priori par maillage régulier et seules les proportions sont estimées. Si les n couleurs sont des réalisations i.i.d. de f , l'estimateur du maximum de vraisemblance conduit à la notion classique de fréquence. L'histogramme est alors dit « normalisé ».

2.2 Approximation par mélange gaussien

Principe Comme alternative aux mélanges de lois unimodales, on peut considérer des mélanges de lois gaussiennes. Formellement, le mélange gaussien f s'écrit

$$f(c|\theta) = \sum_{k=1}^K p_k N(c|a_k), \quad (2)$$

avec $a_k = (\mu_k, \Sigma_k)$, μ_k et Σ_k respectivement le centre et la matrice de variance de la k ème composante normale $N(\cdot|a_k)$. Cette fois, non seulement les proportions p_k mais aussi les positions-dispersions a_k peuvent être estimées car le mélange est identifiable. L'estimation du paramètre global θ peut se faire en maximisant la vraisemblance $L(\theta) = \sum_{i=1}^n \ln f(c_i|\theta)$ par l'algorithme EM (*expectation - maximisation*) [3].

La figure 1 montre les deux différentes approximations (histogramme et mélange gaussien) de la distribution des couleurs rouge-vert.

2.2.1 Justification

On peut comprendre chaque composante gaussienne comme une couleur dominante avec sa variabilité. Si une telle

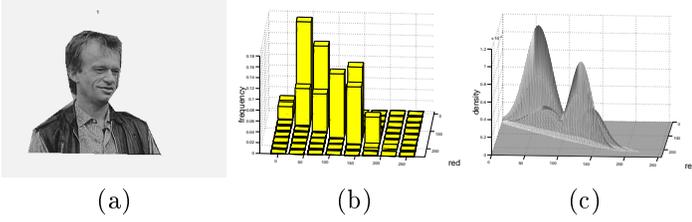


FIG. 1 – (a) une image, (b) son histogramme des couleurs rouge-vert, (c) le mélange gaussien correspondant.

composante existe, les accidents de réflexions, d’ombrages, etc., peuvent faire que la distribution s’éloigne de la loi gaussienne, mais une telle modélisation serait trop complexe, et très dépendante des matériaux considérés. Reste alors le choix du nombre de telles couleurs dominantes.

2.2.2 Choix du nombre de classes

Le nombre de composantes peut aussi être estimé en retenant la valeur qui minimise le critère bayésien BIC [7] (connu aussi sous le nom MDL) ou bien encore le critère bayésien classifiant ICL [1]. Le critère BIC s’écrit

$$\text{BIC} = -2L(\hat{\theta}) + dl \ln n \quad (3)$$

avec $\hat{\theta}$ l’estimateur du maximum de vraisemblance de θ et dl le nombre de degrés de liberté dans θ . Le critère ICL correspond au critère BIC pénalisé par l’entropie E de la partition des couleurs :

$$\text{ICL} = \text{BIC} + 2E. \quad (4)$$

Il a l’avantage d’être plus robuste et plus réaliste que BIC car il sélectionne moins de classes (voir la figure 2).

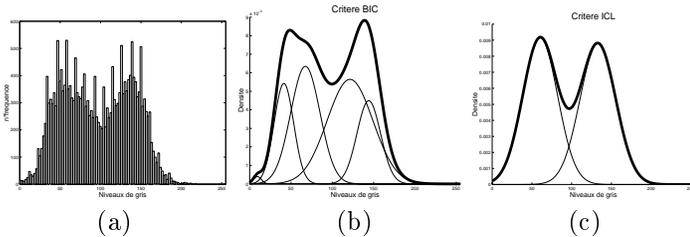


FIG. 2 – (a) une image, (b) les composantes gaussiennes (trait fin) et la densité mélange (trait épais) obtenus avec le critère BIC, (c) idem avec le critère ICL.

2.3 Extension à d’autres distributions

Jusqu’à présent, nous avons utilisé les histogrammes et les mélanges gaussiens uniquement pour approximer la distribution des couleurs. Cependant, il est bien connu [5] que les couleurs peuvent être insuffisantes seules pour l’appariement, par exemple à cause de leur variabilité à des changements de luminosité. Délaissant les méthodes de normalisation des couleurs (normalisation interne et/ou externe) [4], nous proposons d’associer les couleurs à l’information spatiale apportée par l’orientation et/ou la norme du gradient. Les histogrammes et les mélanges gaussiens pourront ensuite approximer la distribution des couleurs combinées à cette orientation et/ou à cette norme.

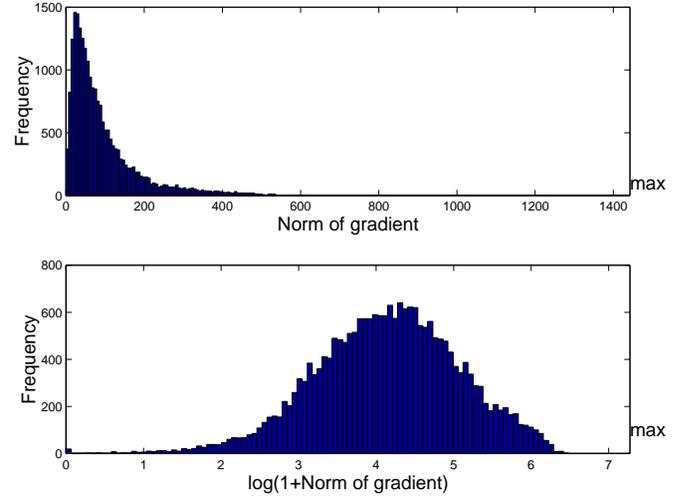


FIG. 3 – Distribution de la norme du gradient et de son logarithme.

2.3.1 Orientation du gradient

Pour chaque pixel de l’image, l’orientation est calculée sur le gradient « Sobel » de l’image en *niveaux de gris* (même si l’image est en couleur) après filtrage gaussien d’écart-type 0.5 et de masque 3×3 .

2.3.2 Norme du gradient

La norme du gradient est calculée avec le même filtre. Cependant la distribution de cette norme est, comme on le voit sur la figure 3, peu uniformément distribuée sur le support et peu gaussienne également. Ces deux particularités sont gênantes respectivement pour les histogrammes et pour les mélanges gaussiens.

Pour remédier à ces deux inconvénients, nous retenons alors une fonction logarithmique de la norme : $\ln(1+\text{norme})$. Le résultat est visible sur la figure 3.

3 Expérimentation

3.1 La séquence « violetta »

Pour la phase d’expérimentation, nous considérons la séquence des 97 personnages segmentés de la figure 4. Ces images sont extraites de la séquence vidéo « violetta ». Les personnages d’un même plan avec le même label sont similaires mais il existe aussi des similarités entre plans différents. L’ensemble des similarités entre numéro de personnages est : 3 avec 7, 4 avec 8, 6 avec 9, 10, 14 et 16, 12 avec 15.

3.2 Les signatures comparées

Nous indexons uniquement les 16 personnages désignés par « LEARN » sur cette figure ; ces personnages correspondent au premier de chaque plan vidéo.

De chaque personnage, huit descripteurs sont extraits :

1. **gray** : niveaux de gris ;
2. **gray+orient** : gray associé à l’orientation du gradient (comme décrite à la section 2.3.1) ;



FIG. 4 – Images utilisées pour l’expérience. Les objets de la base sont notés « LEARN »

3. **gray+norm**: gray associé à la transformée logarithmique de la norme du gradient (comme décrite à la section 2.3.2);
4. **gray+orient+norm**: gray associé à orient et à norm;
5. **RGB**: les trois couleurs;
6. **RGB+orient**: RGB associé à orient;
7. **RGB+norm**: RGB associé orient;
8. **RGB+orient+norm**: RGB associé à orient et à norm.

L’indexation se fait en approximant la distribution de chacun des huit descripteurs précédents soit par des histogrammes à 2 et 8 classes par canal, soit par des mélanges gaussiens. Dans ce dernier cas, le nombre de classes est choisi automatiquement par BIC ou ICL dans une plage de 1 à 6 pour les niveaux de gris et une plage de 1 à 8 pour les autres descripteurs. L’estimation des paramètres du mélange se fait en initialisant l’algorithme CEM [2] 10 fois au hasard puis en lançant l’algorithme EM avec les paramètres initiaux fournis par la meilleure itération de l’algorithme CEM au sens du critère qu’il optimise. Nous limitons ainsi la dépendance de EM au choix des paramètres initiaux. Enfin, l’estimation ne se fait qu’avec

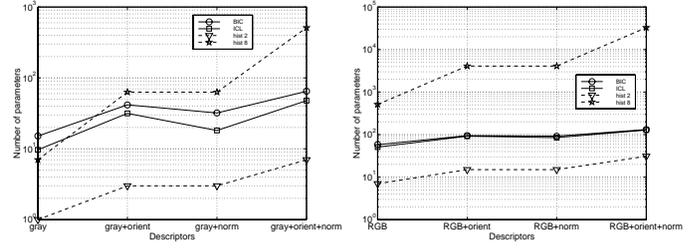


FIG. 5 – Nombre moyen de paramètres libres pour chacune des méthodes d’approximation et chacune des combinaisons de descripteurs.

10% des pixels choisis au hasard pour l’ensemble des descripteurs, excepté les niveaux de gris, ceci pour limiter le temps de calcul.

La figure 5 donne le nombre moyen de paramètres libres sur les 16 objets pour chacune des méthodes.

3.3 Appariement avec les 16 objets indexés

Ensuite, les 97 imagettes sont tour à tour présentées comme images requêtes et les 16 imagettes de la base sont classées par ordre d’appariement pour chacune d’entre elles. Pour chacune des 97 imagettes, on déduit de ce classement le taux de bon appariement à un certain rang par le rapport du nombre de bons appariements réalisés jusqu’à ce rang et du nombre d’appariements idéalement possibles jusqu’à ce rang.

Dans le cas des histogrammes, l’ordre d’appariement est bâti sur la distance du χ^2 , Schiele [6] ayant montré de meilleures performances qu’avec l’intersection d’histogrammes ou bien encore la distance quadratique. Cette distance entre deux histogrammes de paramètres respectifs θ et θ' s’exprime par

$$\chi^2(\theta, \theta') = \begin{cases} \sum_{k=1}^K (p_k - p'_k)^2 / (p_k + p'_k) & \text{si } p_k + p'_k \neq 0 \\ 0 & \text{sinon.} \end{cases} \quad (5)$$

Dans le cas des mélanges gaussiens, l’ordre d’appariement est bâti sur la probabilité conditionnelle $\Pr(B|R)$ qu’un objet B de la base soit similaire à un objet requête R . Cette probabilité est donnée par

$$\Pr(B|R) = \frac{\Pr(R|B) \Pr(B)}{\Pr(R)}. \quad (6)$$

Supposant tous les objets de la base équiprobables a priori, on obtient alors $\Pr(B|R) \propto \Pr(R|B)$. Notant c_1, \dots, c_n les descripteurs (par exemple les couleurs) de R et $\hat{\theta}$ l’index de B , on pose

$$\Pr(R|B) = f(c_1, \dots, c_n | \hat{\theta}) = \exp L(\hat{\theta}), \quad (7)$$

ce qui revient simplement à appairer R avec l’objet B dont le paramètre $\hat{\theta}$ est le plus vraisemblable.

Les figures 6 et 7 donnent les taux de bons appariements aux quatre premiers rangs pour les quatre premières et les quatre dernières combinaisons de descripteurs respectivement.

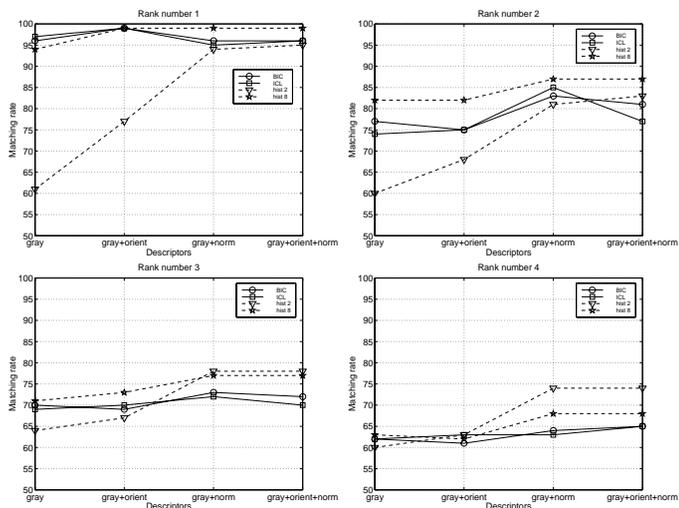


FIG. 6 – Quatre premières combinaisons de descripteurs : taux de bon match pour les quatre premiers rangs pour chacune des quatre approximations.

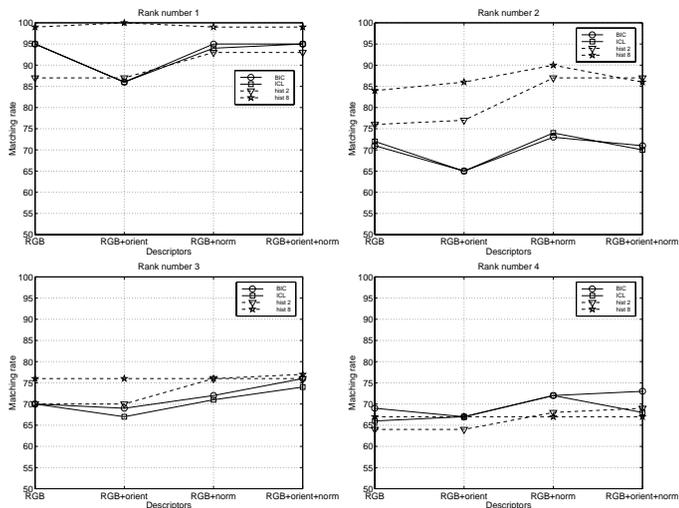


FIG. 7 – Quatre dernières combinaisons de descripteurs : taux de bon match pour les quatre premiers rangs pour chacune des quatre approximations.

3.4 Analyse des résultats

Des figures 5, 6 et 7 il ressort les constatations suivantes :

- au rang 1, la combinaison de descripteurs n’améliore que peu les performances, voire les dégrade. Pour les rangs 2 à 4, la combinaison des descripteurs améliore plus souvent les performances. La raison est certainement qu’au rang 1, le bon match se fait avec des images du même plan et que l’information de couleur ou de niveaux de gris suffisent. Au dessus du rang 1, il s’agit de matcher avec des images similaires mais dans d’autres plans et la combinaison facilite cela.
- la norme du gradient semble être une information plus pertinente que son orientation. Comme corollaire, la combinaison avec la norme et l’orientation est superflue à la seule combinaison avec la norme.
- les histogrammes à huit classes par axe sont souvent

meilleurs que les autres méthodes mais, en grande dimension, ils nécessitent plus de paramètres que les mélanges gaussiens et que les histogrammes à deux classes. Par exemple, le rapport entre les mélanges gaussiens de ICL et les histogrammes à 8 classes est d’un facteur 10 pour `gray+orient+norm` et d’un facteur 256 pour `RGB+orient+norm`.

- le critère BIC est légèrement meilleur que ICL mais ICL nécessite un peu moins de paramètres.

4 Conclusion

Indéniablement et principalement en grande dimension, les mélanges gaussiens permettent, par le choix automatique du nombre de composantes, de réduire sensiblement la taille de l’index par rapport aux histogrammes à 8 classes par axe. De plus le temps d’appariement est du même ordre entre les deux méthodes. Néanmoins, les histogrammes donnent souvent de meilleurs résultats que dans le cas gaussien. Il est vrai que les gaussiennes ne sont estimées que sur 10% des pixels d’une image, mais sinon le temps de calcul de la signature devient prohibitif.

Références

- [1] C. Biernacki, G. Celeux, and G. Govaert. Assessing a mixture model for clustering with the integrated classification likelihood. Rapport de recherche, INRIA, 1998.
- [2] G. Celeux and G. Govaert. A Classification EM Algorithm for Clustering and Two Stochastic Versions. *Computational Statistics & Data Analysis*, 14:315–332, 1992.
- [3] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data (with discussion). *Journal of the Royal Statistical Society, Series B*, 39:1–38, 1977.
- [4] B.V. Funt and G.D. Finlayson. Color constant color indexing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(5):522–529, 1995.
- [5] P. Gros, G. Mclean, R. Delon, R. Mohr, C. Schmid, and G. Mistler. Utilisation de la couleur pour l’appariement et l’indexation d’images. Technical Report 3269, INRIA, September 1997.
- [6] B. Schiele. *Reconnaissance d’objets utilisant des histogrammes multidimensionnels de champs réceptifs*. Thèse de doctorat, GRAVIR – IMAG – INRIA Rhône-Alpes, July 1997.
- [7] G. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6:461–464, 1978.
- [8] M.J. Swain and D.H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.