

Introduction aux modèles Graphiques

Etude et implémentation de l'article « Blobworld: Image Segmentation Using Expectation-Maximization and Its Application to Image Querying » de Chad Carson, Serge Belongie, Hayit Greenspan, and Jitendra Malik, Member (IEEE)

Jérémt Rapin, Olivier Jais-Nielsen

Introduction

La plateforme Blobworld propose une solution de recherche dans une base d'images.

L'approche adoptée par la plate-forme Blobworld pour l'indexation d'une base d'images est la suivante : on cherche à segmenter les images en régions non nécessairement cohérente d'un point de vue sémantique mais au moins cohérente d'un point de vue visuel pour l'utilisateur ; ce dernier doit pouvoir sélectionner des régions proposées pour décrire sa recherche. On cherche alors dans la base d'images interrogées, les images dont la segmentation comporte des régions proches.

La première étape consiste à transformer les pixels formant une image en un ensemble de points dans un espace de descripteurs de pixels, chaque descripteur représentant des propriétés locales de chaque pixel. Cet espace est classifié en régions auxquelles sont associés des descripteurs de régions ; la projection de ces régions dans l'espace de l'image initiale fournit ainsi une segmentation de l'image et les descripteurs permettent la comparaison des régions avec celle issue de la base interrogée.

On s'intéresse ici à la transformation dans l'espace des descripteurs de pixels et à la classification dans cet espace.

On illustrera les différentes étapes et résultats à l'aide d'images typiquement présentes dans une base d'images basée sur Blobworld.



Figure 1 Exemple d'image



Figure 2 Exemple d'image

Descripteurs de pixels

L'image doit être segmentée en régions cohérentes d'un point de vue spatial, du point de vue de la couleur et enfin du point de vue de sa texture. Ces trois aspects justifient la définition de l'espace des descripteurs de pixels.

Un descripteur d'un pixel est un vecteur de \mathbb{R}^8 . Deux coordonnées caractérisent la position du pixel dans l'image, trois coordonnées sont consacrées à la description de la couleur locale du pixel et trois autres à la texture locale du pixel.

Analyse multi-échelle

L'un des paramètres locaux pertinent dans la description d'une zone de texture est l'échelle spatiale de la texture. On estime alors en chaque point différentes grandeurs caractérisant des propriétés locales de l'image, pour un intervalle d'échelles. L'analyse de l'évolution de ces propriétés par rapport aux échelles étudiées permet d'extraire une échelle locale. Cette échelle locale et les grandeurs caractéristiques associées servent à définir les descripteurs locaux.

Estimer une mesure locale en un point (x, y) à une certaine échelle s revient à intégrer la mesure sur une fenêtre carrée de « rayon » 3σ centrée au point (x, y) en pondérant par une distribution gaussienne de variance σ^2 et normalisée sur le carré.

On considère les échelles de l'intervalle $[0,3.5]$; la distance caractéristique maximale pour laquelle un motif sera considéré comme une texture est alors de l'ordre de 10 pixels.

Direction principale

On considère la clarté (première composante dans le système colorimétrique $L^*a^*b^*$) au point (x, y) que l'on notera $l(x, y)$.

On estime alors la matrice du moment d'ordre 2 au point (x, y) et à l'échelle σ par $M_\sigma(x, y) = [G_\sigma * \nabla l(\nabla l)^T](x, y)$ où $\nabla l(x, y)$ est le gradient de l au point considéré et G_σ un filtre gaussien de divergence σ .

La direction principale $\phi_\sigma(x, y)$ au point et à l'échelle considérés est donnée par la direction du vecteur propre de plus grande valeur propre de la matrice du moment d'ordre 2.

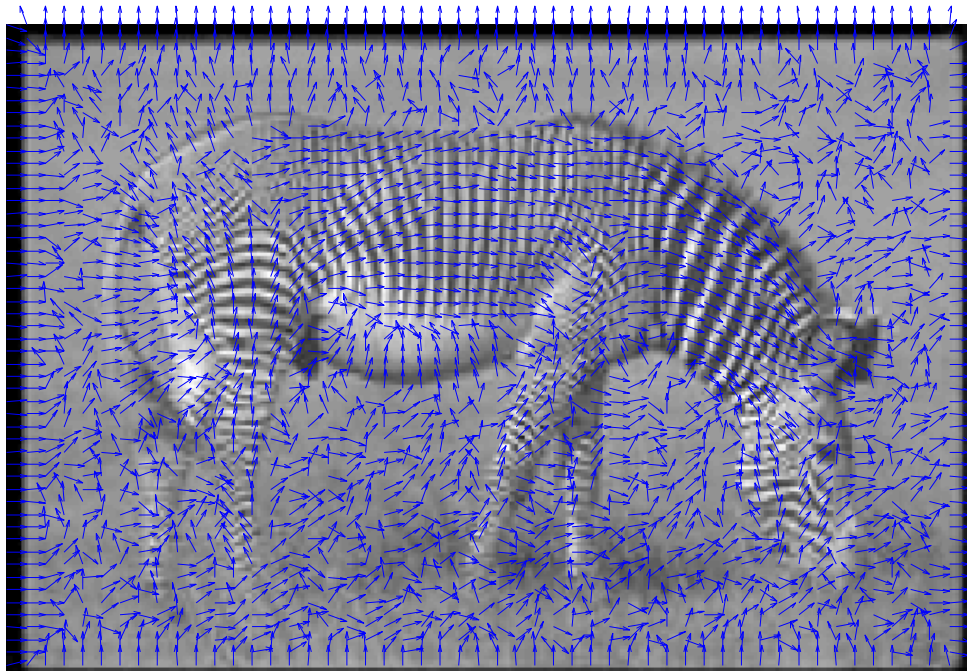


Figure 3 Direction principale locale.

Polarité

On considère ensuite en tout point (x, y) et à toute échelle σ la polarité $p_\sigma(x, y)$ définie par :

$$\begin{cases} E_+(x, y) = [G_\sigma * [\nabla l \cdot n]_+](x, y) \\ E_-(x, y) = [G_\sigma * [\nabla l \cdot n]_-](x, y) \\ p_\sigma = \frac{|E_+ - E_-|}{E_+ + E_-} \end{cases}$$

Avec $n(x, y)$ un vecteur orthogonal à $\phi_\sigma(x, y)$. La polarité est un pourcentage étant d'autant plus élevé qu'une direction est majoritaire sur le voisinage défini par l'échelle.

Une zone uniforme à une très faible polarité à toutes les échelles, une zone de contour à une polarité élevée à toutes les échelles et une zone de texture à une polarité faible pour les échelles inférieures à l'échelle caractéristique de la texture, puis maximale pour cette échelle caractéristique, stable pour les échelles supérieures immédiates.

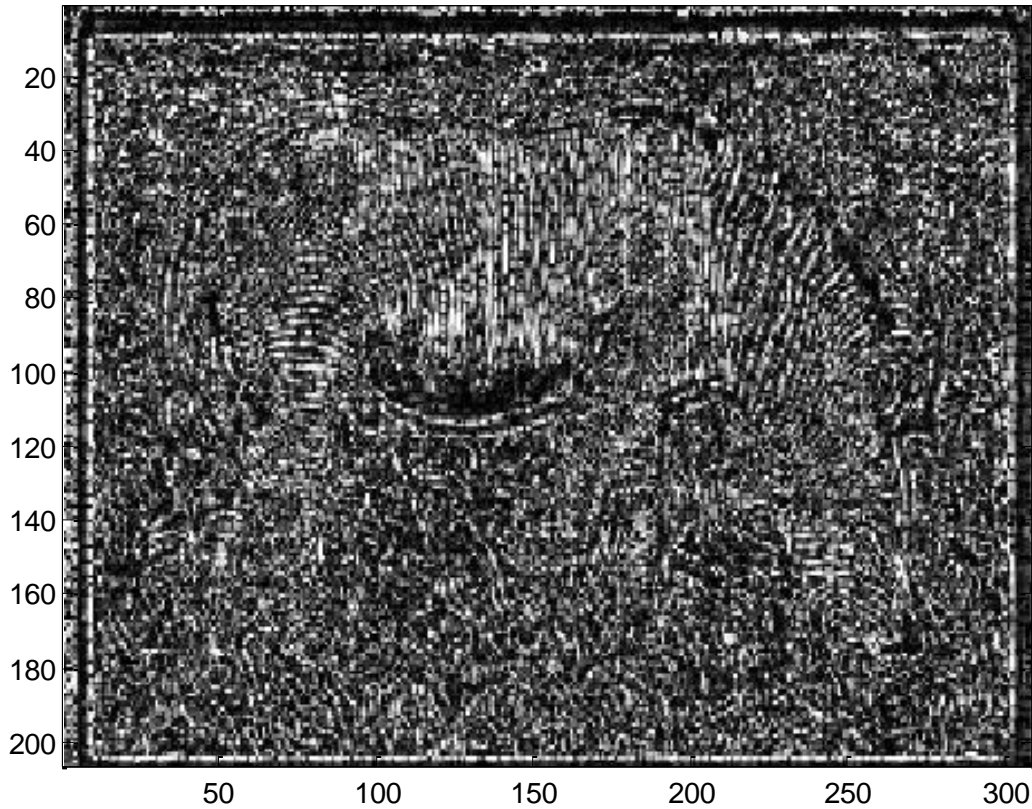


Figure 4 Carte de la polarité à une échelle faible

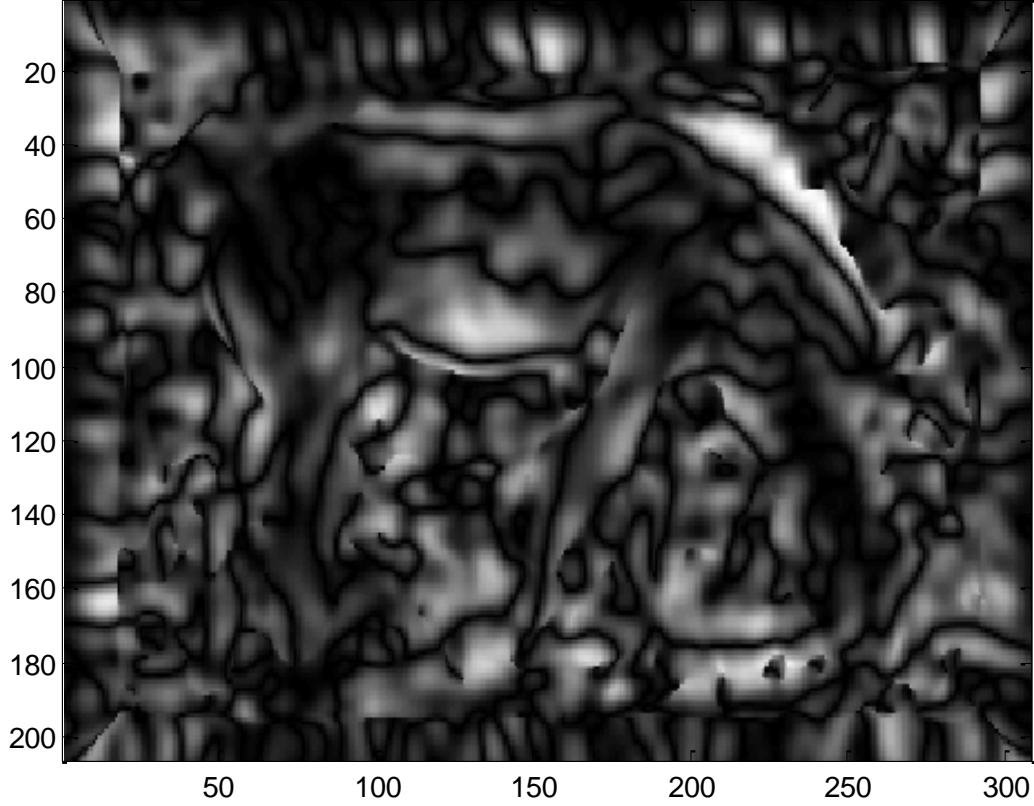


Figure 5 Carte de la polarité à une échelle élevée

Couleur locale

Sur l'image initiale, la couleur est définie en tout point (x, y) par ses trois coordonnées $(l(x, y), a(x, y), b(x, y))$ dans le système $L^*a^*b^*$. On définit la couleur en un point (x, y) et à une échelle σ par les valeurs :

$$\begin{cases} l_{\sigma}(x, y) = [G_{\sigma} * l](x, y) \\ a_{\sigma}(x, y) = [G_{\sigma} * a](x, y) \\ b_{\sigma}(x, y) = [G_{\sigma} * b](x, y) \end{cases}$$

Echelle locale

Pour déterminer l'échelle locale en un point (x, y) d'une région de texture, on considère la décroissance de la polarité lorsque l'échelle augmente. Lorsque cette décroissance cesse (deux valeurs successives ont une différence inférieure à 0.02), on considère que l'on a atteint l'échelle locale $\sigma^*(x, y)$.

Pour améliorer la robustesse de cette définition et l'homogénéité spatiale de l'échelle locale, on considère non pas les cartes de polarité à chaque échelle σ mais ces cartes après convolution avec un filtre gaussien de variance $(2\sigma)^2$.

Les zones uniformes n'ont pas d'échelle particulière ; on leur attribue arbitrairement l'échelle 0. Pour déterminer si un point (x, y) fait partie d'une zone uniforme, on calcul, à chaque échelle σ le contraste de l_σ sur le voisinage de (x, y) d'échelle σ . Si la moyenne de ces contrastes est inférieure à 0.1, le point est considéré comme faisant partie d'une zone uniforme.

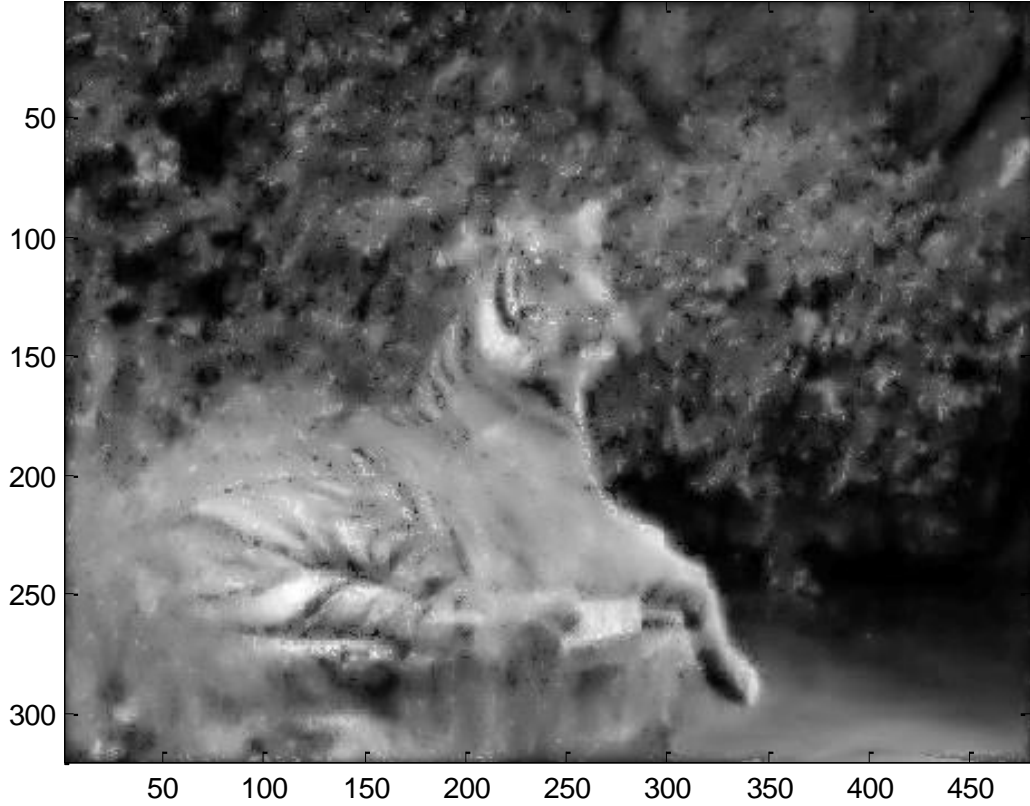


Figure 6 Image localement lissée

Une fois l'échelle locale déterminée en tout point, on peut lisser localement l'image : on associe à chaque point sa couleur dans l'image lissée à son échelle locale, c'est-à-dire, au point (x, y) les valeurs $(l_{\sigma^*(x,y)}(x, y), a_{\sigma^*(x,y)}(x, y), b_{\sigma^*(x,y)}(x, y))$. L'image lissée localement ne doit faire apparaître que les formes des régions, sans les textures. Néanmoins, on constate que certaines textures ont une échelle supérieure à 10 pixels et apparaissent comme plusieurs zones uniformes.

Descripteurs

Le descripteur associé à un pixel (x, y) est finalement défini comme :

- Les trois coordonnées de couleur locale au point (x, y) à l'échelle $\sigma^*(x, y)$.
- Trois coordonnées caractérisant la texture locale dérivée à partir de la matrice $M = M_{\sigma^*(x,y)}(x, y)$ et de la polarité $p = p_{\sigma^*(x,y)}(x, y) : (ac, pc, c)$ avec $a = 1 - \frac{\lambda_1}{\lambda_2}$ et $c = 2\sqrt{\lambda_1 + \lambda_2}$

où λ_1 et λ_2 sont les valeurs propres dans l'ordre décroissant de la matrice M . Le coefficient c est, comme p et a dans $[0,1]$ lorsque la méthode d'estimation du gradient utilise les différences finies centrées. Ce coefficient permet d'indiquer si la matrice du second ordre est significative et pondère ainsi les informations qui en sont dérivées : la polarité et le coefficient a indiquant si la direction principale est significativement dominante.

- Les deux coordonnées spatiales du pixel considéré.

Conclusion

L'approche proposée par cet article pour décrire localement la texture en un point d'une photo est particulièrement riche et peu coûteuse en calcul. Néanmoins, son champs de validité (taille caractéristique d'une texture inférieure à 10 pixels) est assez restrictif et repousser cette limite augmente le temps de calcul de façon quadratique : l'augmentation du nombre d'échelles à examiner est linéaire mais à chaque échelle est associée une taille d'un carré sur lequel intégrer les valeurs ; augmenter le rayon de ce voisinage entraîne alors une augmentation quadratique du nombre de calculs. De plus : la détection de l'échelle locale s'avère assez peu robuste car les cartes de polarité, notamment aux plus petites échelles présentent de fortes irrégularités.