

Détection de points d'intérêt dans les pyramides de contraste

Stéphane BRES, Jean-Michel JOLION

Laboratoire Reconnaissance de Formes et Vision
Bât 403 INSA, 20, Avenue Albert Einstein, 69621 Villeurbanne Cedex, France

sbres@rfv.insa-lyon.fr, jolion@rfv.insa-lyon.fr

Résumé – Nous présentons dans cet article une technique de détection et de localisation de points d'intérêt qui s'appuie sur la notion de contraste en multirésolution. Ce détecteur est utilisé dans le cadre de l'indexation d'images. Dans ce contexte, la notion de contraste apporte un gain de robustesse à certaines perturbations subies par l'image en regard de l'approche différentielle classique des principaux détecteurs de points d'intérêt. Nous comparons nos résultats de ce détecteur à ceux obtenus par d'autres plus classiques (Susan et Harris) par rapport au bruit particulier introduit par le codage Jpeg.

Abstract – This paper addresses the problem of detection and delineation of interest points in images as part of an automatic image and video indexing for search by content purposes project. We propose a novel key point detector based on multiresolution contrast information. The notion of contrast is more robust regarding perturbations compared to classic differential based approaches. We present some results of a robustness approach involving the Harris and Susan detectors as well as ours against the noise related to the Jpeg coding process.

1 Introduction

Ce travail porte sur la détection et la localisation de points d'intérêt dans des images en utilisant le modèle des pyramides de contraste. Il s'insère dans nos recherches actuelles dans le domaine de l'aide à l'indexation d'images¹.

La notion de points d'intérêt n'est pas récente dans le domaine de l'image. Indépendamment des travaux relevant de l'étude de la vision humaine, cette notion a surtout été développée dans le contexte de la robotique depuis les travaux de Moravec. C'est donc tout naturellement que les principaux détecteurs de points d'intérêt (Harris, Förstner, Smith) s'attachent à localiser des points où les variations différentielles sont importantes ou caractéristiques dans plus d'une direction (afin de ne pas inclure les contours). Les coins et autres éléments géométriques exhibant des courbures particulières relèvent donc des points d'intérêt dans cette approche [3].

Les études de l'attention visuelle, plus proche du mécanisme de la vision humaine, proposent des modèles radicalement différents où, si la variation dans le stimulus reste la source d'information, elle n'est pas modélisée sous une forme différentielle mais énergétique [4].

Notre travail s'insère dans cette deuxième approche.

2 Point d'intérêt et pyramide de contraste

Plus particulièrement, nous avons étudié l'apport de la mesure de contraste (définition mathématique plus faible de la variation que son approche différentielle et donc

moins contrainte). La notion de multirésolution étant maintenant communément admise comme solution au problème du choix des échelles d'analyse, nous avons tout naturellement utilisé la pyramide des contrastes pour cette étude. La pyramide C des contrastes d'une image I est définie par [5]:

$$\begin{aligned} C_k(P) &\equiv \frac{G_k(P)}{B_k(P)} \quad \text{pour } 0 \leq k \leq N-1 \\ C_N(P) &\equiv 1 \end{aligned} \quad (1)$$

où P est un pixel, k le niveau dans la pyramide, G_k le niveau k de la pyramide gaussienne construite sur I , N la hauteur de la pyramide, et B_k l'expansion du niveau $k+1$ de la pyramide G sur le niveau k (*i.e.* l'estimation du niveau k à partir du niveau $k+1$). Il a été montré dans [5] que la pyramide des contrastes est un codage exact d'une image qui permet de mettre en valeur les structures principales d'une image.

Dans ce cadre, nous qualifions de point d'intérêt tout pixel ayant un contraste significatif par rapport à son voisinage. En fonction du niveau où il sera extrait, un point d'intérêt pourra être associé à un pixel ou une région de l'image initiale. L'extraction de ces points d'intérêt nécessite une adaptation de la mesure initiale qui tienne compte de sa non symétrie dans les gammes des blancs et des noirs. On utilisera donc

$$C_k^*(P) = \text{Min} \left(\frac{|G_k(P) - B_k(P)|}{B_k(P)}, \frac{|G_k(P) - B_k(P)|}{255 - B_k(P)} \right) \quad (2)$$

qui renvoie une valeur normalisée entre 0 (absence de contraste) et 1 (contraste maximal). Les points d'intérêt seront les maxima locaux significatifs de C_k^* (cf. Figure 1).

¹. Ce travail est financé dans le cadre des projets INCO 950363 TELESUN de la communauté européenne et ACTIV de la Région Rhône-Alpes.



FIG. 1: Points d'intérêt de contraste pour trois niveaux successifs d'une image de portrait pour un seuil de significativité de 0.2.

3 Cartes de points d'intérêt

L'étape suivante consiste (si nécessaire) à combiner les points d'intérêt obtenus à différents niveaux de résolution. Deux solutions sont proposées. La première consiste à neutraliser dans le codage en contrastes multirésolution C d'une image I , tous les pixels qui ne sont pas des points d'intérêt (on force leur valeur à 1). On reconstruit alors l'image par le schéma

$$\tilde{G}_k(P) = \beta(P) \times \text{Expand}[\tilde{G}_{k+1}](P) \quad (3)$$

pour $k = N - 1, \dots, 0$, $\tilde{G}_N(P) = 1$. et où $\beta(P)$ est défini par

$$\beta(P) = \begin{cases} C_k(P) & \text{si } P \text{ max. local de } C_k^* \\ 1 & \text{sinon} \end{cases} \quad (4)$$

Le résultat obtenu (cf. Figure 2b) permet de simplifier l'image initiale à un ensemble de zones de focalisation. On peut aussi produire une carte d'énergie de contrastes normalisés en calculant la somme pyramidale des niveaux de C^* dans laquelle ont été préalablement neutralisés les non maxima locaux (en forçant leur valeur à 0) (cf. Figure 2c) par le processus suivant:

$$E_k(P) = \alpha_k \cdot \beta^*(P) + \text{Expand}[E_{k+1}](P) \quad (5)$$

pour $k = N - 1, \dots, 0$ et $E_N(P) = 0$ et où $\beta^*(P)$ est défini

par

$$\beta^*(P) = \begin{cases} C_k^*(P) & \text{si } P \text{ max. local de } C_k^* \\ 0 & \text{sinon} \end{cases} \quad (6)$$

Nous avons utilisé une pondération linéaire selon le niveau, *i.e.* $\alpha_k = \frac{1}{k+1}$ qui privilégie les contrastes locaux.

Les maxima locaux de cette dernière carte sont donc le résultat d'un processus *coarse-to-fine* dans la structure pyramidale. Plus de détails sur cette technique peuvent être trouvés dans [2]. De plus, une démonstration interactive est disponible à l'adresse

<http://rfv.insa-lyon.fr/~jolion/Cours/ptint.html>
sur des images types ou fournies par l'utilisateur.

4 Robustesse

Le cadre de ce projet est l'indexation d'images. Les points d'intérêt sont utilisés pour réduire l'information contenue dans une image stockée dans la base. Il est donc nécessaire de tester la robustesse de cet opérateur vis à vis des principales perturbations que l'on est amené à rencontrer. Parmi celles-ci, nous avons choisi de présenter ici la robustesse vis à vis du bruit de codage Jpeg. En effet, les bases classiques constituées à l'heure actuelle sont souvent formées d'images de très bonnes qualités. Cependant, la possibilité offerte à l'utilisateur d'interroger cette base avec ses propres images induit que les descripteurs issus de cette image requête doivent être robuste à un bruit de codage tel que celui introduit par Jpeg.

Notre mesure de robustesse prend en compte à la fois la variation du nombre de points ainsi que leur délocalisation. Le détail de cette comparaison peut être trouvée dans [6]. La figure 3a montre un exemple de résultat pour l'image de la figure 3b. Notre détecteur est comparé aux détecteurs de Plessey [3] et celui du projet SUSAN [8] en utilisant l'indicateur de la figure de mérite introduite par Pratt pour la comparaison des cartes de contours [7].

$$F(I, J) = \frac{1}{\max(N_I, N_J)} \sum_{i=1}^{N_J} \frac{1}{1 + d_i^2}$$

où N_I (resp. N_J) est le nombre de points contours dans la carte de contours de I (resp. J), et d_i la distance minimale entre le i ème point contour de la carte J à son homologue (le point le plus proche) dans la carte I (d_i est majorée par 5 pour réduire le temps d'évaluation de cette mesure). Par définition, cet indicateur prend ses valeurs dans l'intervalle $[0, 1]$ (la valeur 1, resp. 0, correspondant à la ressemblance maximale, resp. minimale, entre les deux cartes contours).

Notre détecteur a un comportement similaire à celui de Plessey mais il montre cependant une plus grande stabilité pour les qualités usuelles ($\geq 75\%$). En particulier, même en cas de qualité égale à 100%, le détecteur de Plessey subit une légère dégradation car pour cette qualité, le passage dans la procédure de codage induit des écarts de ± 1 sur les pixels. De plus, notre détecteur ne nécessite l'emploi que d'un seul paramètre (le seuil final) alors que le détecteur de Plessey utilise 4 paramètres (largeur du filtre dérivatif, largeur du filtre de consolidation, paramètre k dans la formule de l'indicateur et seuil final).

Le comportement du détecteur SUSAN est bien moins bon car celui-ci a été optimisé pour un type de bruit tout à fait différent.

5 Application à l'indexation

Une application de ce détecteur dans un système de recherche d'images par similarités attentives et pré-attentive est en cours de développement. Des premiers résultats peuvent être consultés sur internet à l'adresse

<http://rfv.insa-lyon.fr/~jolion/INDEX>

Dans cette implémentation, 250 points d'intérêt sont associés à chaque image de la base. Pour chaque point d'intérêt, nous extrayons les cinq premiers invariants différentiels. La distribution spatiale de chaque invariant est ensuite résumée par un histogramme bidimensionnel. La similarité entre tout couple d'images (*i.e.* la requête et une image de la base) est obtenue par le calcul de la distance entre ces histogrammes (distance de Bhattacharyya). Un processus d'amélioration *a posteriori* permet d'ajuster la réponse du système pour un utilisateur donné.

Dans cette application, les points d'intérêt ont d'une part permis de réduire le volume de données à traiter mais aussi de focaliser l'information conservée pour chaque image sur les éléments principaux de celle-ci.

Des détails peuvent être trouvés dans [1].

Références

- [1] A. Heindricht, D. Koubaroulis, B. Levenaise-Obadia, P. Rovida et J.M. Jolion (1999) Recherche Robuste d'Images par Similarité Pré-attentive, Laboratoire Reconnaissance de Formes et Vision, Rapport de recherche RR 04.99, soumis à RFIA 2000.
- [2] C. Bres et J.M. Jolion (1998) Multiresolution contrast based detection of interest points, Laboratoire Reconnaissance de Formes et Vision, Rapport de recherche RR 02.98.
- [3] R. Deriche et G. Giraudon (1993) A Computational Approach for Corner and Vertex Detection, *Int. Journal of Computer Vision*, 10 (2), 101-124.
- [4] L. Itti, C. Koch et E. Niebur (1998) A Model of Saliency-Based Visual Attention for Rapid Scene Analysis, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(11), 1254-1259.
- [5] J.M. Jolion (1994) Analyse multirésolution du contraste dans les images numériques, *Traitement du Signal*, 11(3), 245-255.
- [6] J.M. Jolion et C. Bres (1999) Influence du codage Jpeg sur des descripteurs d'images, *Traitement du Signal*, 15(4), 309-320.
- [7] W.K.Pratt (1978) *Digital Image Processing*, New York, Wiley, Interscience.
- [8] S.M.Smith et J.M.Brady (1997) SUSAN - A New Approach to Low Level Image Processing, *Int. Journal of Computer Vision*, 23(1), 45-78.

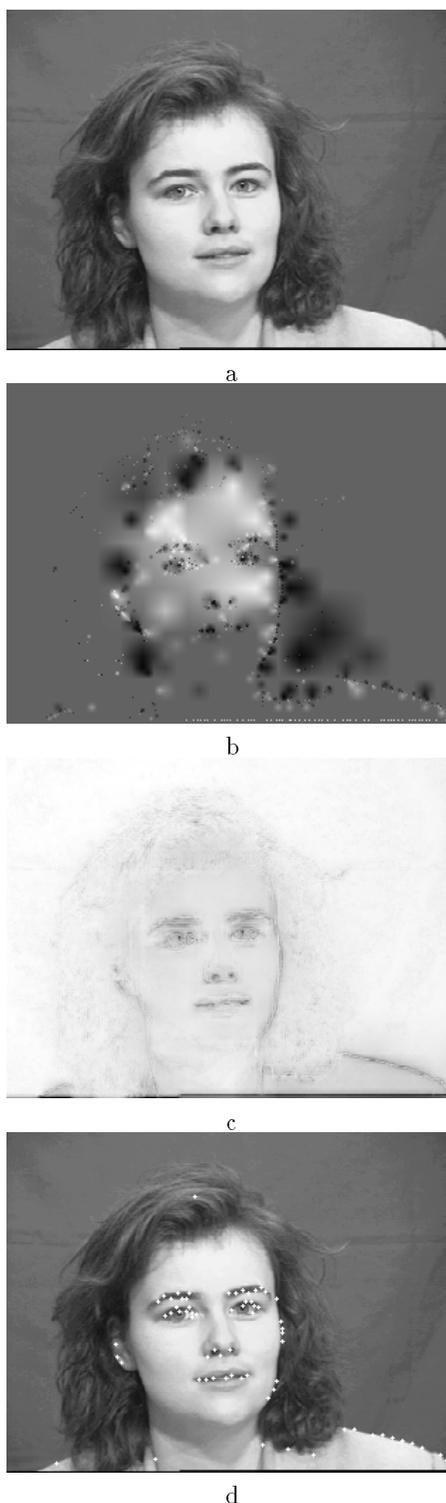
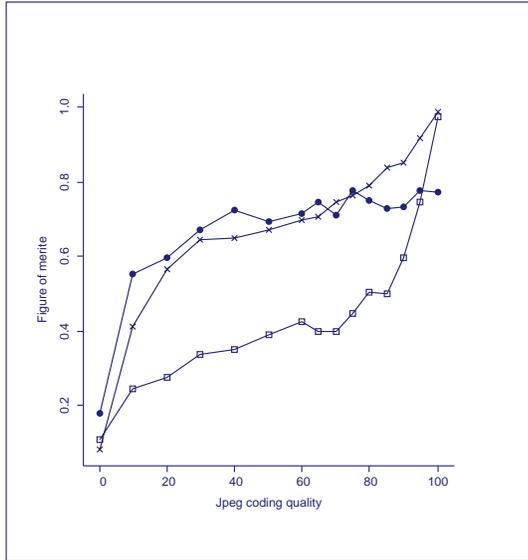
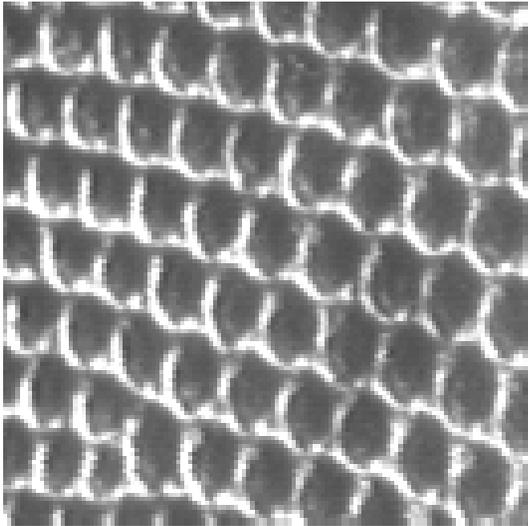


FIG. 2: Analyse d'une image grâce aux points d'intérêt. (a) Image initiale. (b) Image reconstruite en ne considérant que les points d'intérêt et un seuil de 0.1. (c) Carte d'énergie de contraste (aucun seuil). (d) Maxima locaux dans la carte d'énergie de contraste (élargis et superposés à l'image initiale).



a



b

FIG. 3: (a) Robustesse des détecteurs Susan (□), Plessey (●) et le notre (×) en regard du bruit introduit par le codage Jpeg en fonction de la qualité de ce codage. (b) L'image utilisée pour le test.