

Caractérisation d'une région d'intérêt dans les images

Yves Caron*, Pascal Makris*
Nicole Vincent**

* Laboratoire d'Informatique EA 2101
Université François – Rabelais de Tours
64, avenue Jean Portalis
37200 TOURS

yves.caron@etu.univ-tours.fr, makris@univ-tours.fr
<http://www.li.univ-tours.fr>

** Laboratoire CRIP5-SIP
Université Paris 5

45, rue des Saints Pères
75270 PARIS CEDEX 06
nicole.vincent@math-info.univ-paris5.fr
<http://www.math-info.univ-paris5.fr>

Résumé. Une image est un support d'information qui a montré son efficacité. Néanmoins une image comporte souvent plusieurs zones, l'arrière plan et une zone d'intérêt privilégiée. La vision humaine permet la segmentation de manière naturelle et intégrant toute la connaissance que le sujet peut avoir de l'objectif visé par l'image. Nous proposons ici une méthode de détermination des régions d'intérêt d'une image numérique comme zones saillantes. Les lois de Zipf et Zipf inverse sont adaptées au traitement des images et permettent d'évaluer la complexité structurelle d'une image. Une comparaison des modèles locaux évalués sur des imagerie permet de mettre en évidence une région de l'image. Deux méthodes de classification ont été utilisées pour la détermination de la région d'intérêt : la partition d'un nuage de points représentant les caractéristiques associées aux imagerie, et les réseaux de neurones. Cette méthode de détection permet d'obtenir des zones d'intérêt conformes à la perception humaine. On opère une hiérarchisation sur les zones en fonction de la structuration de l'information élémentaire, les pixels.

1 Introduction

Dans notre monde moderne, dominé par les nouvelles technologies de l'information, l'image a pris une importance considérable en tant que vecteur de la connaissance. De nos jours, l'image tend même parfois à remplacer l'écrit comme moyen privilégié de communication. Par rapport au texte, l'image a pour avantage d'être perceptible de manière plus immédiate et d'être davantage indépendante du contexte culturel, notamment de la langue maternelle de l'individu, même si cette indépendance n'est que relative. L'utilisation de l'image devient donc particulièrement intéressante dans un monde caractérisé par l'intensification des échanges internationaux. Cependant, si une image est un moyen de représenter une information, toutes les zones d'une image n'apportent pas des informations du même ordre. En effet, le plus souvent seule une partie de l'image apporte une

connaissance nouvelle pour l'observateur, le reste peut être considéré comme un arrière-plan qui présente un intérêt nettement moindre. On retrouve le même phénomène dans un texte où les informations importantes se trouvent au sein d'un discours de moindre importance. Il est alors utile de pouvoir extraire la zone d'intérêt de l'arrière-plan de l'image. Pour un être humain, l'identification d'une zone d'intérêt est une tâche naturelle, réalisée de manière instantanée par notre cerveau. En revanche, pour une machine la tâche est beaucoup plus ardue. En effet, la notion même de zone d'intérêt est fondamentalement subjective, pour une même image des individus différents pourront voir des zones d'intérêt différentes, selon leur culture, leur personnalité, leurs centres d'intérêts ou leur humeur du moment. De plus, elle fait appel à des référents culturels dont une machine ne dispose pas. C'est pourquoi, dans le cadre d'une détection automatisée, il est nécessaire de se limiter à des cas particuliers. En fait, nous cherchons à extraire de l'image une région saillante, c'est à dire qui se distingue nettement de l'arrière-plan, sans préjuger du contenu sémantique de la scène. Ce processus se rapproche de celui utilisé par la vision humaine de bas niveau. De nombreux travaux (Yarbus 1967), (Marr 1982), (Finlay 1980) ont été effectués sur le système visuel humain, pour déterminer quelles sont les caractéristiques visuelles de la scène observée qui influencent le plus la perception humaine, en se basant notamment sur les mouvements des yeux de la personne qui observe une scène. Ces caractéristiques sont notamment le contraste, les couleurs, la taille et la forme des objets. La perception visuelle a également été étudiée dans le cadre de la théorie de la Gestalt (Wertheimer, 1922), (Koffka, 1934). Selon cette théorie, la vision humaine perçoit les formes globalement par rapport à un contexte, elle construit les formes en regroupant les éléments visuels de l'image selon des propriétés de proximité, de similarité, de courbure, de continuité et de symétrie. Différentes méthodes ont été proposées pour la détection de régions saillantes dans une image, elles se basent notamment sur le calcul de caractéristiques perceptuelles des régions de l'image (Osberger, 1998), les dissimilarités locales de l'image (Stentiford 2001), l'analyse multi-échelle de caractéristiques de bas niveau (Itti et al, 1998), les transformées en ondelettes (Wang 2001), les mesures d'entropie multi-échelles (Kadir 2000) , la dimension fractale (Carlotto 1990) ou encore la fusion par chaînes de Markov de caractéristiques issues de la théorie de la Gestalt (Massmann et al. 1997).

Notre approche se base sur un modèle issu de l'analyse linguistique : la loi de Zipf (Zipf 1949). Cette loi modélise les fréquences d'apparition des mots d'un texte à l'aide des lois puissance. Les modèles en loi puissance ont été utilisés non seulement en linguistique (Miller 1958), (Cohen 1996) mais aussi en économie (Pareto 1897), pour l'étude de la démographie urbaine (Hill 1970), (Gabaix 1999) ou pour la modélisation du trafic Internet (Breslau et al., 1999). En matière d'analyse d'images, la loi de Zipf a déjà été utilisée pour l'évaluation de la qualité d'images compressées (Vincent et al., 2000) et pour la détection d'objets artificiels dans les milieux naturels (Caron et al., 2002). Notre approche se base sur la comparaison des caractéristiques locales des modèles de loi puissance calculés sur une image segmentée en imagettes. Elle permet ainsi de mettre en évidence une région saillante de l'image qui se distingue de l'arrière-plan par sa complexité structurelle, et qui sera considérée comme région d'intérêt. La connaissance intéressante est plus complexe qu'un discours sans contenu. Après avoir présenté les modèles de lois puissance utilisés, les loi de Zipf et de Zipf inverse, nous décrivons les codages des motifs de l'image. Enfin, nous détaillerons la méthode de détection et les résultats obtenus avec deux méthodes de classification, une classification non supervisée par partition d'un nuage de points, et une classification par réseaux de neurones.

2 Les modèles de loi puissance

2.1 La loi de Zipf

La loi de Zipf (Zipf 1949) est une loi empirique qui décrit la répartition des fréquences d'apparition des différents éléments d'un ensemble de symboles topologiquement ordonnés comme les mots d'un texte. Cet ensemble est organisé de telle manière qu'il soit possible de définir des n-uplets de symboles et de les dénombrer. La loi de Zipf décrit la distribution des fréquences d'apparition de chacun de ces n-uplets, c'est à dire le nombre d'occurrences d'un même n-uplet. Ces fréquences d'apparition ne sont pas aléatoires, elles s'expriment par une loi puissance en fonction du rang des n-uplets classés par ordre décroissant de leur fréquence d'apparition. La fréquence d'apparition $N_{\sigma(i)}$, du n-uplet U_i de rang i dans la séquence est donnée par la loi puissance :

$$N_{\sigma(i)} = k \cdot i^{-\alpha} \quad (1)$$

Dans cette formule, k et α sont des constantes positives. La valeur de l'exposant α caractérise la loi puissance, dans le cas des textes en langage naturel comme le français ou l'anglais, cette valeur est en général proche de 1. La valeur de α peut être estimée par la pente $-\alpha$ de la droite obtenue par la représentation graphique dans un repère bi-logarithmique de $(i, N_{\sigma(i)})$. On appelle courbe de Zipf cette représentation graphique.

2.2 La loi de Zipf inverse

La loi de Zipf inverse (Zipf 1935) est également une loi empirique qui décrit les fréquences d'apparition des n-uplets d'un ensemble. On s'intéresse ici aux nombres de n-uplets différents ayant la même fréquence. Ces n-uplets sont définis de la même manière qu'avec la loi de Zipf. D'après la loi de Zipf inverse, le nombre I de n-uplets distincts ayant une même fréquence f peut être modélisé par la formule suivante :

$$I(f) = A f^{-\gamma} \quad (2)$$

Dans cette formule, A et γ sont des constantes positives. Dans le cas des textes en langage naturel, cette loi se vérifie pour les mots les moins fréquents et la valeur de l'exposant γ est proche de 2. Comme la loi de Zipf, la loi de Zipf inverse peut être représentée graphiquement en échelle bi-logarithmique, on appelle courbe de Zipf inverse une telle représentation. Les caractéristiques de la courbe de Zipf inverse permettent de caractériser le contenu du texte : plus le vocabulaire est riche et plus la pente et l'ordonnée à l'origine de la courbe de Zipf inverse seront élevées.

3 Codage de l'image

3.1 Nécessité du codage de l'image

Nous allons maintenant devoir adapter au cas des images les modèles de lois puissance décrits précédemment, la loi de Zipf et de Zipf inverse, définies à l'origine pour l'analyse de textes. Pour cela, il nous faut définir un équivalent de la notion de mot dans le cas des images. Dans un texte, un mot est constitué d'une suite de lettres adjacentes. Dans les images, nous définissons des motifs comme des ensembles de pixels adjacents disposés dans un masque. Nous respectons la topologie de l'image en utilisant un masque constitué d'une matrice carrée de taille 3x3 centrée sur un pixel. La manière la plus simple de coder l'image serait d'utiliser directement les niveaux de gris des pixels de l'image. Cependant, dans ce cas, le trop grand nombre de motifs possibles par rapport au nombre de motifs présents dans l'image rend les fréquences d'apparition trop faibles pour être significatives. C'est pourquoi il est nécessaire d'employer des codages de l'image qui permettent de réduire le nombre de motifs différents à une valeur acceptable pour que l'on puisse en dégager une information pertinente.

3.2 Codage des N classes

Une méthode simple pour réduire le nombre de motifs différents susceptibles d'être présents dans une image est de rééchantillonner l'échelle des niveaux de gris de l'image afin de limiter le nombre de valeurs possibles de ceux-ci. Pour cela, il suffit d'affecter une même valeur aux pixels ayant des niveaux de gris contenus dans un même intervalle de valeurs. L'échelle [0, 255] des niveaux de gris est partitionnée en N intervalles, ou classes, de largeurs égales et numérotés suivant l'ordre croissant de 0 à N-1, d'où le nom de méthode des N classes donné à ce codage. Dans le cas d'intervalles de même longueur, la valeur de la classe du pixel (x,y) de niveau de gris g(x,y) est donnée par la formule suivante :

$$c(x,y)=\text{int}\left[\frac{Ng(x,y)}{255}\right] \quad (3)$$

D'autres méthodes peuvent reposer sur une classification des niveaux de gris apparaissant dans l'image. Les pixels sont désormais codés sur une échelle de N valeurs. Le nombre de motifs différents possibles n'est plus alors que de N^9 pour un motif de taille 3x3. L'utilisation de ce codage, s'il permet de réduire le nombre de motifs possibles et d'obtenir une distribution de fréquences des motifs qui ait une signification, introduit évidemment une perte d'information, puisque le nombre de valeurs possibles de l'échelle des niveaux de gris a été réduit. Il faut alors choisir une valeur de N qui soit la plus petite possible afin de réduire le nombre de motifs différents tout en étant suffisamment grande pour ne pas trop altérer la structure de l'image. Nous avons déterminé expérimentalement que le choix d'une valeur de N égale à 9 permet de préserver le contenu structurel de l'image sans que la dégradation ne soit trop visible, tout en maintenant à une valeur raisonnable le nombre de motifs différents présents dans l'image. C'est pourquoi nous retiendrons par la suite cette valeur. La figure suivante montre un exemple de motif codé avec les 9 classes.

255	210	210
25	2	34
40	2	40

8	7	7
0	0	1
1	0	1

FIG. 1 – Exemple de motif codé avec les 9 classes

3.3 Codage des rangs généraux

Le principe du codage des rangs généraux ne s'appuie pas directement sur l'information de luminance, mais les variations locales de niveaux de gris, afin de mettre en évidence les différences relatives plutôt que les valeurs absolues. Ce codage consiste à remplacer les niveaux de gris des pixels par leur rang dans un voisinage, il s'agit d'une transformation non linéaire qui, si elle fait perdre l'amplitude de la différence des niveaux de gris, conserve les relations ordinales entre les niveaux de gris des pixels voisins. L'application de ce codage à un motif consiste à affecter la valeur 0 au pixel de niveau de gris le plus bas, puis à incrémenter cette valeur dans l'ordre croissant des niveaux de gris, en affectant la même valeur si les niveaux de gris sont égaux. La figure ci-dessous montre un exemple de motif codé avec les rangs généraux.

255	210	210
25	2	34
40	2	40

5	4	4
1	0	2
3	0	3

FIG. 2 – Exemple de motif codé avec les rangs généraux

Ces codages vont maintenant être utilisés pour caractériser la complexité structurelle des images.

4 Application des modèles de lois puissance aux images

4.1 La loi de Zipf

L'analyse d'une image reposant sur la loi de Zipf consiste à balayer l'image par un masque 3x3 en codant les motifs rencontrés puis à associer à chaque motif sa fréquence d'apparition dans l'image. Les motifs sont rangés dans un tableau, à chaque itération le motif courant est comparé aux motifs déjà présents dans le tableau. Si le motif existe déjà, sa fréquence est incrémentée d'une unité, sinon le motif est ajouté dans le tableau. Les motifs sont ensuite triés dans l'ordre décroissant de leur fréquence d'apparition à l'aide d'un algorithme de tri rapide. La complexité de l'algorithme est en $O(n^2)$, où n représente le nombre total de motifs présents dans l'image. La courbe rang-fréquence, ou courbe de Zipf, est ensuite tracée dans un repère bi-logarithmique. La figure 4 montre un exemple de courbe de Zipf associée à une image avec chacun des deux codages. Les résultats sont différents selon le codage utilisé, chacun des codages mettant en évidence des informations différentes sur le contenu de l'image. Avec le codage des 9 classes, la répartition des fréquences ne semble pas suivre la loi de Zipf : la courbe n'est pas linéaire. En réalité, la courbe peut être considérée comme linéaire par morceaux, deux parties apparaissant. La première partie de la courbe regroupe les motifs les plus fréquents, il s'agit des motifs homogènes, où tous les pixels appartiennent à la même classe. La courbe rang-fréquence est linéaire pour les motifs

les plus fréquents de cette classe, et s'incurve ensuite fortement vers le bas. La répartition des fréquences des motifs homogènes obéit donc à une loi puissance sauf pour les motifs les plus rares de cette classe. Ce résultat est conforme à ceux qui sont habituellement obtenus dans les cas où le cardinal de l'ensemble des n-uplets est petit, comme par exemple dans le cas des lettres de l'alphabet. La partie de la courbe regroupe, quant à elle, les motifs non homogènes comme les détails fins ou les contours, pour ces motifs la courbe rang-fréquence est linéaire, la répartition des fréquences des motifs non homogènes obéit donc aussi à une loi de Zipf, comme c'est le cas pour les mots d'un texte. Le codage des 9 classes permet donc de mettre en évidence deux catégories de motifs. La répartition statistique des fréquences des motifs pour chacune de ces catégories suit séparément une loi puissance. Dans le cas du codage des rangs généraux, la courbe apparaît comme relativement linéaire, il est donc possible d'approximer la répartition des fréquences des motifs par une loi puissance. Les motifs les plus fréquents de la courbe se trouvent sur les contours principaux ou dans des zones saturées de l'image, et non dans les zones homogènes. En effet, les zones homogènes comportent de faibles variations de niveaux de gris qui sont mises en évidence par le codage des rangs généraux.

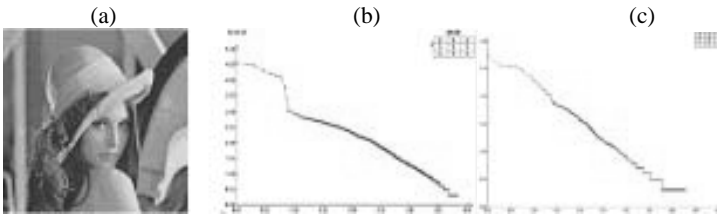


FIG. 3 – Courbe de Zipf d'une image (a) avec les 9 classes (b) et les rangs généraux (c)

4.2 La loi de Zipf inverse

L'application de la loi de Zipf inverse consiste à balayer l'image par un masque et à coder les motifs en utilisant les mêmes codages que pour la loi de Zipf, et à dénombrer les motifs distincts pour chaque valeur de la fréquence d'apparition. La figure 5 montre un exemple de courbe de Zipf inverse obtenue avec une même image pour les deux codages.

Dans les deux cas, la courbe de Zipf inverse d'une image naturelle est linéaire pour les motifs les moins fréquents de l'image. La distribution du nombre de motifs distincts ayant une fréquence d'apparition donnée suit donc une loi puissance, la loi de Zipf inverse est donc vérifiée en ce qui concerne les images naturelles. Avec le codage des 9 classes, les motifs les moins fréquents de l'image se concentrent essentiellement dans les zones riches en détails. L'utilisation de ce codage permet donc de mettre en évidence la complexité des détails de l'image. En effet, plus l'image est riche en détails et plus la pente et l'ordonnée à l'origine de la courbe seront élevées, l'image comportant un plus grand nombre de motifs qui apparaissent plus fréquemment dans l'image. Par contre, en utilisant le codage des rangs généraux, on remarque que la quasi-totalité des motifs de l'image n'apparaissent qu'une seule fois dans l'image, ce qui fait que l'utilisation de ce codage pour la loi de Zipf inverse n'apporte pas une grande connaissance quant au contenu de l'image, ce qui fait que son utilisation avec la loi de Zipf inverse n'a pas un grand intérêt pratique.

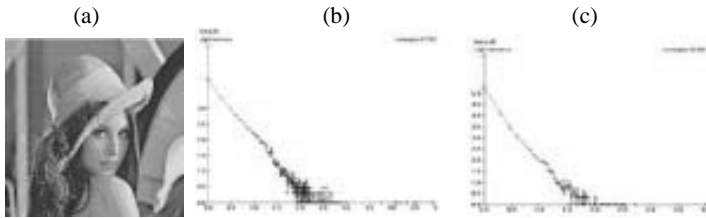


FIG. 4 – Courbe de Zipf inverse d'une image (a) obtenue avec les 9 classes (b) et les rangs généraux (c)

5 Détection de zones d'intérêt

5.1 Principe de la méthode

La détection automatique de zones d'intérêt étant un problème impossible à résoudre dans le cas général, il est nécessaire de ne considérer que certains cas particuliers. Nous cherchons à détecter une région saillante qui se distingue nettement de l'arrière-plan d'une image. On fait l'hypothèse que cette région ne comporte qu'une composante connexe et occupe au moins 20% de la surface totale de l'image, il s'agit du sujet principal de la scène. Cette région d'intérêt est en général située au premier plan de l'image, elle présente le plus souvent davantage de détails que l'arrière-plan. Les modèles de loi puissance permettant de caractériser la complexité du contenu structurel de l'image, nous pouvons les utiliser pour déterminer cette zone d'intérêt. Le codage des 9 classes est le mieux adapté à cette application, car il conserve la structure de l'image et est plus conforme à la perception humaine. Les lois de Zipf et de Zipf inverse peuvent toutes deux être utilisées. Cependant, si les modèles de loi puissance permettent de caractériser la complexité des textures d'une image, ils ne donnent par eux-mêmes aucune information spatiale quant à la position d'une éventuelle zone d'intérêt dans l'image. C'est pourquoi il est nécessaire de segmenter l'image en imagerie afin de pouvoir détecter une zone d'intérêt. Comme nous recherchons une zone d'intérêt qui occupe une part importante, la segmentation n'a pas besoin d'être très précise, on se contente d'un découpage en imagerie rectangulaires de tailles égales. Pour que la répartition des fréquences d'apparition des motifs apporte une information exploitable, les imagerie ne devront pas être trop petites, la taille optimale déterminée expérimentalement étant d'environ 5000 pixels. La méthode de détection de zone d'intérêt consiste alors à appliquer la loi de Zipf ou de Zipf inverse sur chacune des imagerie et à classer les courbes de Zipf associées à chacune des imagerie. Pour cette phase de classification nous avons choisi d'utiliser une méthode qui présente l'avantage d'être simple à mettre en œuvre et de ne pas nécessiter de base d'apprentissage explicite, ce qui la rend facilement adaptable aux images les plus variées possible. Cette méthode consiste à représenter les caractéristiques des courbes de Zipf ou de Zipf inverse associées aux imagerie par un nuage de points dans un espace à deux dimensions, puis à partitionner ce nuage en fonction de critères déterminés expérimentalement afin de caractériser la zone d'intérêt.

5.2 Utilisation de la loi de Zipf

Nous avons vu précédemment que la courbe de Zipf associée à une image codée par le codage des 9 classes peut être approximée par deux segments de droite correspondant l'un aux motifs uniformes, l'autre aux contours et aux détails. Il est donc naturel d'utiliser les pentes de ces deux parties de la courbe pour classifier une imagerie d'après les propriétés de sa courbe de Zipf. On indique alors en abscisse la pente P_1 de la première partie de la courbe et en ordonnée la pente P_2 de la seconde partie. Nous avons déterminé expérimentalement que sur la plupart des images photographiques que les zones perceptuellement saillantes de l'image se trouvent représentées par les points situés dans une partie du nuage délimitée par les droites d'équation $y=x$ et $y=1,2Gx$, où Gx est l'abscisse du centre de gravité du nuage de points. Cependant, toutes les imageries qui satisfont à cette condition ne font pas nécessairement partie de la zone d'intérêt principale de l'image, c'est pourquoi nous définissons comme zone d'intérêt la plus grande composante connexe des imageries représentées par des points tels que $P_1 > P_2$ et $P_1 > 1,2 Gx$. La figure 5 montre un exemple de zone d'intérêt détectée avec cette méthode.

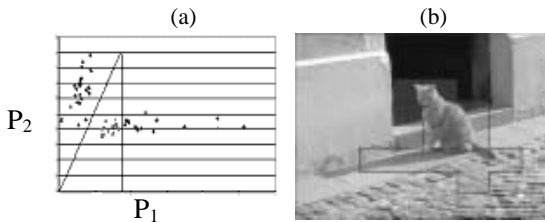


FIG. 5 – Détection de zone d'intérêt avec la loi de Zipf. Représentation du nuage de points (a) et zone d'intérêt détectée avec cette méthode (b)

Cette méthode a été évaluée sur une base de test de 100 images photographiques indépendantes de celles qui ont servi à fixer les paramètres. Les images représentent des sujets divers, elles sont choisies de manière à représenter un éventail de cas typiques que l'on peut retrouver dans la pratique. Chacune de ces images comporte une zone d'intérêt nettement visible qui a été identifiée comme telle par plusieurs observateurs humains. La zone d'intérêt est considérée comme détectée si elle correspond à celle qui a été déterminée par l'observation humaine. Les résultats sont assez moyens, la zone d'intérêt est détectée dans seulement 56% des images testées. C'est pourquoi nous allons utiliser un autre modèle de loi puissance : la loi de Zipf inverse.

5.3 Utilisation de la loi de Zipf inverse

De la même manière que la loi de Zipf, la loi de Zipf inverse peut être utilisée pour détecter les zones d'intérêt dans une image. Les caractéristiques significatives de la courbe de Zipf inverse sont la pente et l'ordonnée à l'origine. Les valeurs de ces deux caractéristiques permettent de caractériser le contenu de l'image ; plus l'image est complexe et riche en détails, plus la pente et l'ordonnée à l'origine seront élevées. Comme avec la méthode précédente, les deux caractéristiques seront représentées dans un nuage de points.

Le plus souvent, la zone d'intérêt contient plus de détails que l'arrière-plan de l'image, c'est pourquoi nous choisissons comme zone d'intérêt les imagettes pour lesquelles la pente et/ou l'ordonnée à l'origine sont les plus élevées. La zone d'intérêt initiale est constituée des imagettes qui sont représentées par les points dont l'ordonnée à l'origine est supérieure à celle du centre de gravité du nuage de points. Afin d'éliminer les détails isolés, on ne conserve que la plus grande composante connexe de ces imagettes. La zone d'intérêt déterminée avec cette méthode comportant éventuellement des trous que l'on bouche. Il est possible d'ajuster dynamiquement le seuil de séparation des deux classes du nuage en fonction de la taille désirée de la zone d'intérêt. En effet, pour qu'elle soit pertinente, la zone d'intérêt doit avoir une taille relativement importante, par exemple entre 20% et 50% de l'image. Si la zone d'intérêt initiale est trop petite ou trop grande, on ajuste les seuils et on recalcule la zone d'intérêt jusqu'à ce qu'elle ait une aire comprise entre 20% et 50% de la surface totale de l'image. La figure 6 représente un exemple de zone d'intérêt détectée à l'aide de cette méthode, en utilisant ici l'ordonnée à l'origine de la courbe de Zipf.

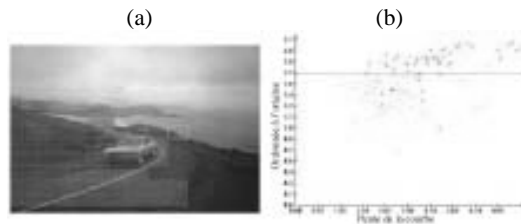


FIG. 6 – Détection de zone d'intérêt avec la loi de Zipf inverse. Représentation du nuage de points (a) et zone d'intérêt détectée avec cette méthode (b)

Jusqu'à présent nous avons considéré que la zone d'intérêt contenait davantage de détails que l'arrière-plan. Cependant il existe des images où la zone d'intérêt est au contraire plus uniforme que l'arrière-plan. Dans ce cas, la zone d'intérêt sera constituée par les imagettes représentées par les points dont l'ordonnée à l'origine et la pente sont inférieures à celles du centre de gravité du nuage. La distinction entre les deux catégories d'images peut se faire automatiquement en fonction du nombre de motifs n'apparaissant qu'une seule fois dans l'image, si plus de 50% des motifs n'apparaissent qu'une seule fois sur l'ensemble de l'image, la zone d'intérêt recherchée sera moins uniforme que l'arrière-plan de l'image.

La méthode de détection a été testée sur la même base de 100 images que la méthode précédente. Les résultats sont bien meilleurs que par la méthode utilisant la loi de Zipf. En effet, la zone d'intérêt a été détectée sur 80% des images testées si l'on utilise uniquement l'ordonnée à l'origine comme critère de séparation entre les deux classes du nuage et 84% si l'on utilise les deux caractéristiques. La zone d'intérêt détectée avec la loi de Zipf inverse est souvent plus grande que la zone d'intérêt réelle de l'image, cela est surtout vrai lorsqu'on n'utilise qu'une des deux caractéristiques des courbes de Zipf inverse.

6 Classification par réseaux de neurones

Jusqu'à présent, nous avons utilisé une méthode de classification que l'on peut qualifier de simpliste car il ne s'agit que d'une partition linéaire d'un nuage de points. Mais il est également possible d'utiliser des classifieurs plus élaborés comme des réseaux de neurones. La loi de Zipf inverse donnant de meilleurs résultats en matière de détection que la loi de Zipf, nous allons donc utiliser les caractéristiques des courbes de Zipf inverse comme variables d'entrée d'un réseau de neurones. Nous utiliserons ici une partition fixe en 64 imagerie pour toutes les images de la base. Le réseau de neurones utilisé est un réseau à couches constitué d'une couche d'entrée de 128 neurones, d'une couche cachée de 128 neurones et d'une couche de sortie de 64 neurones. Le vecteur d'entrée est constitué des valeurs de la pente et de l'ordonnée à l'origine des courbes de Zipf inverse de chacune des 64 imagerie. La sortie retourne une valeur 0 ou 1 selon que l'imagerie appartient ou non à la zone d'intérêt. Les neurones ont une fonction de transfert de type log-sigmoïde et l'apprentissage se fait par rétropropagation du gradient. La base de 100 images utilisée précédemment a été divisée en une base d'apprentissage de 40 images et une base de test de 60 images. La figure 7 montre un exemple de zone d'intérêt obtenue en utilisant cette méthode.



FIG. 7 – Exemples de zones d'intérêt détectées à l'aide d'un réseau de neurones

Les résultats de ces tests sont moyennement satisfaisants, la zone d'intérêt est détectée seulement sur 66% des images testées. La région d'intérêt détectée se situe le plus souvent au voisinage du centre de l'image, proche de la position moyenne des zones d'intérêt des images d'apprentissage et elle ne correspond pas toujours à l'objet recherché. En fait, le réseau de neurones a tendance à apprendre davantage la position des régions d'intérêt sur les images d'apprentissage que les caractéristiques des courbes de Zipf inverse. Cela est dû à la grande disparité des images utilisées par notre application, qui fait que le réseau de neurones a des difficultés à généraliser à partir des exemples d'apprentissage. Afin de réduire l'influence de la position de la zone d'intérêt dans l'image, nous avons utilisé une autre architecture de réseaux de neurones, avec un réseau de neurones indépendant pour chacune des imagerie, composé d'une couche d'entrée à 2 neurones, d'une couche cachée à 3 neurones et d'un unique neurone de sortie. Cependant, les résultats ne sont pas significativement différents de ceux obtenus avec la première architecture. Les réseaux de neurones ne sont donc pas une méthode de classification très adaptée à notre problème, où l'on cherche à extraire des connaissances d'images très différentes entre elles. Dans notre cas, il est préférable d'utiliser des méthodes de classification qui ne dépendent pas explicitement d'une phase d'apprentissage.

7 Conclusion

Les modèles de lois puissance tels que les lois de Zipf et de Zipf inverse, destinés à l'origine à l'analyse linguistique, sont également utilisables pour modéliser l'information contenue dans les images. Ils permettent de caractériser la complexité structurelle d'une image au moyen de la répartition statistique des fréquences d'apparition des motifs. Ces modèles peuvent être utilisés pour extraire d'une image une zone d'intérêt définie comme une zone perceptuellement saillante de l'image qui présente une structuration différente de celle de l'arrière-plan. Après segmentation de l'image en imagettes, une classification selon les caractéristiques des modèles de lois puissance associés à chacune des imagettes permet de déterminer la zone d'intérêt de l'image. La loi de Zipf comme la loi de Zipf inverse peuvent être utilisées pour caractériser cette zone d'intérêt, la loi de Zipf inverse donnant des résultats nettement meilleurs que la loi de Zipf. Une méthode simple de classification par partition linéaire d'un nuage de points permet de détecter des régions d'intérêt qui correspondent bien à la perception d'un observateur humain. Cette méthode de classification est mieux adaptée à des images présentant une grande diversité que les méthodes dépendant d'un apprentissage comme les réseaux de neurones. Une amélioration possible de la méthode de détection de zones d'intérêt est de fusionner les caractéristiques issues des modèles de lois puissance avec des caractéristiques perceptuelles de l'image telles que des mesures de contraste ou des indices de contours. Cette méthode pourra être implémentée dans un codeur JPEG2000 afin de détecter automatiquement la zone d'intérêt d'une image à compresser. Enfin, d'autres applications des modèles de lois puissance sont envisageables dans le domaine de l'extraction de la connaissance sur les images, notamment la recherche par le contenu et l'indexation de bases d'images.

Références

- Yarbus A. (1967), *Eye Movements and Vision*. Plenum Press, New York NY, 1967
- Marr D. (1982), *Vision*, W.H. Freeman and Co., New York, 1982
- Findlay J. (1980), The visual stimulus for saccadic eye movement in human observers. *Perception*, 9:7–21, 1980
- Wertheimer M. (1922), *Untersuchungen zur Lehre von der Gestalt*, *Psychologische Forschung*, 1, 47-58, 1922
- Koffka K. (1935), *Principles of Gestalt Psychology*, publ. Lund Humphries, London, 1935
- Osberger W., Maeder A.J. (1998), Automatic identification of perceptually important regions in an image, 11th international conference on Pattern Recognition, no. 1, pp. 701-704
- Stentiford F. (2001), An Estimator for Visual Attention through Competitive Novelty with Application to Image Compression, *Picture Coding Symposium*, Seoul, Korea, pp. 25-27, April 2001
- Itti L., Koch C., Niebur E. (1998) A model of saliency-based visual attention for rapid scene analysis, *IEEE PAMI*. 20(11), pp. 1254-1259
- Wang J.Z. , Li J., Gray R. M., Wiederhold G. (2001) Unsupervised multiresolution segmentation for images with low depth of field, *IEEE PAMI* vol. 23, no. 1, pp. 85-90
- Kadir T., Brady M.(2001) Saliency, Scale and image Description, *International Journal of Computer Vision*, Volume 45 Issue 2 November 2001
- Carlotto M. J., Stein M. C.(1990), A Method for Searching Artificial Objects on Planetary Surfaces, *Journal of the British Interplanetary Society*, vol.43, 1990

- Massmann A., Posch S., Sagerer G., Schlüter D.(1997) Using Markov random fields for contour-based grouping. In Proceedings International Conference on Image Processing, volume II, pages 207-210, 1997
- Zipf, G.K (1949) Human Behavior and the Principle of Least Effort, Addison-Wesley, New York, 1949
- Miller G.A., Newman E.B. (1958) Test of a Statistical Explanation of the Rank-Frequency Relation for Words in Written English, American Journal of Psychology, No. 71, pp. 209-218
- Cohen A., Mantegna R.N., Havlin S. (1997) Numerical analysis of word frequencies in artificial and natural language texts, Fractals, Vol. 5 No.1, pp. 95-104
- Pareto V. (1897) Cours d'économie politique, Rouge, Lausanne et Paris, 1897
- Hill B.M. (1970) Zipf's law and prior distributions for the composition of a population, Journal of the American Statistical Association, 65, pp.1220-1232
- Gabaix X. (1999) Zipf's law for cities: an explanation, Quarterly Journal of Economics, 114:739-767, 1999
- Breslau L., Cao P., Fan L., Phillips G., Shenker S. (1999). Web caching and Zipf-like distributions: Evidence and implications. In Proceedings of IEEE Infocom 99 , New York, pp. 126-134, 1999
- Vincent N., Makris P., Brodier J. (2000) Compressed Image Quality and Zipf's Law, Proceedings of International Conference on Signal Processing (ICSP – IFIC-IAPRWCC2000) , Pékin (Chine), pp. 1077-1084.
- Caron Y., Makris P., Vincent N. (2002), A method for detecting artificial objects in natural environments, International Conference on Pattern recognition (ICPR - IAPR), Québec (Canada) pp. 600-603
- Zipf G.K. (1935), Psycho-Biology of Languages, Houghton-Mifflin, 1935

Summary

This paper presents a region of interest detection method for digital photographic images based on models adapted from linguistic analysis : Zipf law and inverse Zipf law. These models allow characterization of the frequency distribution of image patterns using power laws. The use of these models with an appropriate coding pattern allows the evaluation of the structural complexity of an image and to put into evidence image zones containing more details. The region of interest detection method consists in partitioning the image into sub-images, in applying on each sub-image Zipf and inverse Zipf laws and in classifying the sub-image in function of the characteristics of Zipf and inverse Zipf models, in order to put into evidence a region of the image which appears distinct from the background and constitutes the region of interest. Two classification methods were used for region of interest determination : the partition of a point cluster representing the characteristics associated with the sub-images, and artificial neural networks. This method allows the detection of region of interest consistent with human perception.