



SPIM

Thèse de Doctorat



UFC

école doctorale **sciences pour l'ingénieur et microtechniques**
UNIVERSITÉ DE FRANCHE-COMTÉ

Titre

Sous-titre

■ Gilles PERROT

SPIM

Thèse de Doctorat



école doctorale **sciences pour l'ingénieur et microtechniques**
UNIVERSITÉ DE FRANCHE-COMTÉ

N° | X | X | X |

THÈSE présentée par

Gilles PERROT

pour obtenir le

Grade de Docteur de
l'Université de Franche-Comté

Spécialité : **Informatique**

Titre

Sous-titre

Soutenue le 17 septembre 2012 devant le Jury :

Incroyable HULK	Rapporteur	Professeur à l'Université de Gotham City Commentaire secondaire
Super MAN	Examineur	Professeur à l'Université de Gotham City
Bat MAN	Directeur de thèse	Professeur à l'Université de Gotham City

SOMMAIRE

1	Introduction	7
2	Le traitement des images bruitées	11
2.1	Modèle d'image bruitée	12
2.2	Modèles de bruit	12
2.2.1	Le bruit gaussien	12
2.2.2	Le speckle	12
2.2.3	Le bruit "sel et poivre"	13
2.2.4	Le bruit de Poisson	13
2.3	Les techniques de réduction de bruit	13
2.4	Les techniques de segmentation	14
2.4.1	Analyse d'histogramme	14
2.4.2	Analyse de graphe	15
2.4.3	kernel-means, mean-shift et dérivés	17
2.4.4	Les contours actifs, ou <i>snakes</i>	20
2.4.5	Méthodes hybrides	22
2.5	L'état de l'art des implémentations GPU	22
3	La segmentation orientée régions dans les images bruitées	23
3.1	Présentation - existant	23
3.2	La parallélisation du snake polygonal	23
4	Le filtrage des images sur GPU	25
4.1	Algorithme de réduction de bruit par recherche des lignes de niveaux	25
4.2	Filtre médian	25
4.3	Filtres de convolution	25
5	Conclusion générale	27

INTRODUCTION

Nous percevons le monde au travers le filtre de nos cinq sens, parmi lesquels la vue et plus largement la vision joue un rôle essentiel et unique. Elle nous fournit de précieuses données sur les objets qui nous entourent et leur comportement, comme leur couleur, leur forme, leur volume, leur texture mais aussi leur position, vitesse et accélération relative à notre propre mouvement.

Notre vision est aussi l'appareil sensoriel grâce auquel ont pu naître la lecture et l'écriture, au travers de la reconnaissance des signes qui composent les mots et la ponctuation, ainsi que des groupes de signes qui forment les mots et des groupes de mots qui donnent le sens au texte écrit.

Très tôt, l'être social qu'est l'homme a exprimé le désir et le besoin de partager avec d'autres hommes ce qu'il voyait ou avait vu. Dès lors, il a simplifié, modélisé et inventé des représentations pour les adapter à sa perception et aussi aux contraintes externes. Ses représentations visuelles ont tout d'abord été statiques, exécutées sur des murs de roche, puis il a un jour été capable de les emporter avec lui pour transmettre des messages, partager des informations et des émotions.

Les images prennent ainsi diverses formes selon le rôle qu'on leur attribue : véhiculer des informations textuelles, géométriques ou techniques au travers de schémas conceptuels, ou reproduire au mieux une réalité observée comme le fait la photographie, mais aussi provoquer des émotions au travers de représentations abstraites comme des peintures.

Certains traitements appliqués à ces représentations sont déjà très anciens ; le fait d'y dissimuler des informations supplémentaires (stéganographie) remonte par exemple à l'antiquité et à été développé sous de multiples formes depuis, tant la nécessité de transmettre secrètement des informations à toujours été un enjeu particulièrement important.

Toutefois, la découverte et la maîtrise de l'électricité ont ouvert des voies nouvelles au traitement technique de l'information et ont permis l'avènement de ce que l'on nomme désormais le traitement du signal, dont les application à l'image forment un des domaines aux contours et méthodes particuliers.

Durant une période que l'on peut juger, a posteriori, assez courte, la chaîne d'imagerie électronique fut entièrement analogique, de la caméra vidéo à l'écran cathodique par exemple, la télévision étant un des moteurs des progrès techniques dans le domaine. Peut-être qu'aujourd'hui, à l'époque du numérique roi, les jeunes gens ne se le figurent pas, mais de nombreuses opérations étaient déjà appliquées à ces signaux analogiques qui véhiculaient les images.

Parmi les tous premiers furent les traitements visant à améliorer la qualité visuelle, c'est

à dire à réduire l'effet des perturbations imputables aux dispositifs d'acquisition ou bien ajuster les niveaux de luminosité et contraste d'une prise de vue effectuée dans des conditions difficiles. Bien d'autres, plus complexes, furent aussi maîtrisés comme l'incrustation d'éléments synthétiques ou le codage permettant de réserver la visualisation des images aux détenteurs d'un décodeur adapté. Ce dernier paraîtrait archaïque de nos jours tant son principe était simple : permuter les lignes de chaque image selon une séquence prédéterminée. La capacité de calcul d'un simple téléphone portable actuel serait suffisante pour en décrypter la séquence et décoder les images en temps réel.

Naturellement, la représentation numérique des signaux à de nouveau fait émerger de nouvelles perspectives que l'on ne cerne vraisemblablement que partiellement, pour en vivre les évolutions au jour le jour. L'informatique est le coeur de ces techniques récentes et a déjà permis le développement d'opérations très complexes, quasi irréalisables en analogique, mais aujourd'hui mises à la portée de tout possesseur d'ordinateur personnel. Citons en exemple la segmentation, capacité à distinguer les différentes zones d'une image, ou bien la reconnaissance de formes qui intervient, entre autres, dans l'interprétation de texte manuscrit, ou encore la poursuite (tracking) permettant de suivre l'évolution d'objets en mouvement dans une séquence d'images. Malgré tout, même si certains algorithmes sont capables d'opérations impossibles à la vision humaine, comme par exemple l'extraction d'information dans des images très fortement bruitées, d'autres demeurent très difficiles à automatiser alors que notre cerveau semble les effectuer sans effort, comme la segmentation.

Certains invariants demeurent toutefois, comme le fait que l'image numérisée est naturellement altérée par des *bruits* de nature et intensité variables trouvant leurs origines dans les dispositifs d'acquisition et l'éclairage des scènes. La technologie des capteurs (CMOS), l'accroissement de leur densité en pixels et l'augmentation des fréquences de balayage concourent d'ailleurs à en intensifier les effets, justifiant la recherche de méthodes de réduction de bruit adaptées. De nombreuses solutions ont été proposées et expérimentées au fil des années, toutefois, à ce jour, aucune ne s'est imposée comme universelle par ses propriétés ou son domaine d'application, tant les caractéristiques des images et des perturbations sont variées. Par ailleurs, aucune ne fait sauter le verrou du compromis qualité - vitesse et ainsi les méthodes qui génèrent des résultats de grande qualité ne sont dans la pratique pas utilisables du fait de leur temps de calcul trop long. Paradoxalement, ce type de limitation ne se résout pas automatiquement et systématiquement avec l'accroissement régulier des capacités de calcul des ordinateurs car la résolution des capteurs grandit en suivant une courbe quasi identique. La recherche de performance dans le traitement des images doit donc continuer de diversifier les voies qu'elle emprunte pour faire émerger des structures adaptées. Dans ce contexte, répondant à la demande du grand public pour un rendu toujours plus réaliste des jeux vidéos, les fabricants d'adaptateurs graphiques (cartes graphiques) ont progressivement étendu les capacités de leurs produits pour les doter de moyens de traitement très spécialisés mais pouvant être exécutés à grande vitesse, permettant d'envisager la gestion en temps réel de flux vidéos en haute définition, ce que le processeur central d'un ordinateur personnel ne peut, le plus souvent, pas garantir. Ces cartes graphiques sont ainsi devenues de véritables assistants processeurs graphiques, ou coprocesseur, appelés aujourd'hui GPU pour *graphical processing unit*.

La technologie de fabrication des GPU est identique à celle utilisée pour la fabrication des microprocesseurs classiques qui pilotent nos ordinateurs (CPU) : il s'agit de la photolithographie de motifs sur un substrat de silicium. Les éléments fonctionnels

de base de ces motifs sont des transistors CMOS (Complementary Metal Oxyde semiconductor) mesurant aujourd'hui 22 nm. Pour parvenir à traiter les flux d'image plus rapidement que ne peuvent le faire les CPU, les constructeurs de GPU ont organisé différemment les transistors de sorte à disposer d'un grand nombre de petites unités de calcul, chacune avec une faible capacité de stockage individuelle. Ce faisant, les GPU ne disposent plus de toutes les fonctionnalités de gestion et de contrôle que possèdent les CPU et doivent donc être employés différemment pour tirer parti de cette structure particulière. Pour tenter de comprendre l'organisation d'un GPU actuel, il faut s'imaginer des millions d'exécutants (les fils d'exécution ou threads) regroupés en équipes de travail (les blocs), comportant jusqu'à un millier d'exécutant. Au sein d'une même équipe, les exécutants peuvent communiquer entre eux et échanger des données grâce à un espace de stockage proche et rapide. En revanche, la communication inter équipe est difficile et très lente, associée à un espace de stockage, certes volumineux mais éloigné et dont l'usage s'avère donc très pénalisant. Cette organisation très particulière est naturellement adaptée aux opérations d'affichage, pour lesquelles elle a été conçue, mais considérant les débits permis par ces processeurs graphiques, les chercheurs et développeurs ont très vite voulu profiter de ces performances pour traduire des algorithmes dont les implémentations étaient jusqu'alors trop lentes. Toutefois, l'adaptation à l'architecture des GPU s'est avérée souvent délicate et n'a parfois apporté aucune amélioration des performances.

C'est dans ce cadre que s'inscrivent les travaux présentés ici, axés sur la recherche de méthodes performantes pour le traitement sur GPU d'images numériques bruitées. Nos premiers travaux ont porté sur la segmentation en parallélisant un algorithme existant pour CPU, tout en lui conférant la capacité de traiter des images de plus grande taille (jusqu'à 16 millions de pixels). Nous nous sommes ensuite intéressés à la réduction de bruit et proposé un algorithme original et adaptable à une grande variété de perturbations, du bruit additif gaussien au bruit multiplicatif gamma. La conception de cet algorithme a été guidée par la volonté de fournir un élément fonctionnel performant sur GPU, pas nécessairement optimal au sens de la qualité du traitement, mais permettant une utilisation temps réel tout en apportant une amélioration qualitative par rapport aux filtres rapides simples. Dans le même esprit, nous nous sommes penchés sur l'implémentation du très employé et très étudié filtre médian pour en proposer l'implémentation la plus performante connue à ce jour avec un débit de plus de 1,85 milliards de pixels à la seconde. À cette occasion, nous avons appliqué des principes d'utilisation de la mémoire qui semblent pouvoir être transposables avec succès à d'autres classes d'algorithmes. Nous l'avons ainsi montré pour les filtres de convolutions.

Ces techniques permettent des gains substantiels en terme de vitesse, mais imposent en contrepartie un style de programmation très inhabituel et fastidieux, ôtant parfois toute versatilité aux blocs fonctionnels ainsi conçus. C'est pourquoi enfin, nous avons développé une application en ligne, accessible à tous et permettant de générer simplement le code source des fonctions souhaitées (médiens, convolutions) en sélectionnant les options et valeurs des paramètres adaptés au besoin.

Nous nous attacherons dans la suite à effectuer un survol des techniques de traitement d'images numériques et des principales classes d'algorithmes employés. Ce chapitre sera également l'occasion de définir les notations générales applicables à tous le manuscrit. Dans le chapitre ??, nous présenterons les principaux travaux de recherche en lien avec ceux que nous avons menés et qui seront présentés dans les chapitres ?? et suivants, en commençant par notre contribution à la segmentation orientée régions, suivie

de celles concernant la réduction de bruit sous les formes du filtre contraint par lignes de niveaux, du filtre médian, puis des filtres de convolution.

LE TRAITEMENT DES IMAGES BRUITÉES

L'étendue des techniques applicables aux images numériques est aujourd'hui si vaste qu'il serait illusoire de chercher à les décrire ici. Ce chapitre présente plus spécifiquement les algorithmes utilisés en présence d'images (fortement) bruitées. Le bruit rend potentiellement délicate l'extraction des informations utiles contenues dans les images perturbées ou en complique l'interprétation, qu'elle soit automatique ou confiée à la vision humaine. L'intuition nous incite donc à chercher des méthodes efficaces de prétraitement pour réduire la puissance du bruit afin de permettre aux traitements de plus haut niveau comme la segmentation, d'opérer ensuite dans de meilleures conditions.

Toutefois, il faut également considérer que les opérations préalables de réduction de bruit apportent des modifications statistiques aux images et influent donc potentiellement sur les caractéristiques que l'on cherche à mettre en évidence grâce au traitement principal. En ce sens, il peut-être préférable de chercher à employer des algorithmes de haut niveau travaillant directement sur les images bruitées pour minimiser les effets des altérations apportées par les filtres débruiteurs et conserver toute l'information contenue dans les images perturbées.

Les images auxquelles nous nous intéressons sont généralement les images numériques allant des images naturelles telles que définies par Caselles [] aux images d'amplitude issues de l'imagerie radar à ouverture synthétique (ROS ou en anglais SAR) [], de l'imagerie médicale à ultrasons (échographie) ou encore biologique dans le cas de la microscopie électronique. Ces dispositifs d'acquisition sont naturellement, et par essence, générateurs de bruits divers, inhérents aux technologies mises en œuvre au sein de ces systèmes et qui viennent dégrader l'image idéale de la scène que l'on cherche à représenter ou analyser. On sait aujourd'hui caractériser de manière assez précise ces bruits et la section ?? en détaille les origines physiques ainsi que les propriétés statistiques qui en découlent. On peut d'ores et déjà avancer que la connaissance de l'origine d'une image et donc des propriétés des bruits associés qui en corrompent les informations, est un atout permettant de concevoir des techniques de filtrage adaptées à chaque situation. Toutefois, la recherche d'un filtre universel, bien qu'encore illusoire, n'est pas abandonnée, tant les besoins sont nombreux, divers et souvent complexes.

2.1/ MODÈLE D'IMAGE BRUITÉE

On considère qu'une image observée est un ensemble de N observations sur un domaine Ω à deux dimensions ($\Omega \subset \mathbb{Z}^2$). À chaque élément de Ω , aussi appelé *pixel*, est associé un indice unique $k \in \llbracket 1; N \rrbracket$, une position $x_k = (i, j)_k \in \Omega$ et une valeur observée $v_k = v(i, j)_k$. La valeur observée peut, selon les cas, être de dimension 1 pour les images représentées en niveaux de gris ou de dimension 3 pour les images couleur représentées au format RVB. Les dimensions supérieures, pour la représentation des images hyperspectrales n'est pas abordé. L'image observée peut ainsi être considérée comme un vecteur à N éléments $\bar{v} = (v_k)_{k \in \llbracket 1; N \rrbracket}$. Les divers traitements appliqués aux images observées ont souvent pour but d'accéder aux informations contenues dans une image sous-jacente, débarrassée de toute perturbation, dont nous faisons l'hypothèse qu'elle partage le même support Ω et que nous notons \bar{u} . Le lien entre \bar{u} et \bar{v} peut être exprimé généralement par la relation $\bar{v} = \bar{u} + \sigma \epsilon$, où $\epsilon \in \mathbb{R}^N$ représente le modèle de perturbation appliquée à \bar{u} et σ représente la puissance de cette perturbation qui a mené à l'observation de \bar{v} . Dans le cas général, ϵ dépend de \bar{u} et est caractérisé par la densité de probabilité (PDF pour probability density function) $p(v|u)$.

2.2/ MODÈLES DE BRUIT

2.2.1/ LE BRUIT GAUSSIEN

Le bruit gaussien est historiquement le plus étudié et celui auquel sont dédiées le plus de techniques de débruitage. La génération des images numériques au travers les capteurs CMOS et CCD ?? est le siège de nombreuses perturbations dues à la technologie de fabrication et à la nature du rayonnement dont ils mesurent l'intensité en différents zones de leur surface, appelées *photosites*. On distingue en particulier les bruits suivants selon leur origine physique :

- la non uniformité de réponse des photosites.
- le bruit de photon
- le bruit de courant d'obscurité
- le bruit de lecture
- le bruit de non uniformité d'amplification des gains des photosites.

Des descriptions détaillées des mécanismes concourant à la génération de ces bruits sont fournies dans ?? Dans un certain intervalle usuel d'intensité lumineuse, il est toutefois admis que l'ensemble des ces perturbations peut être représenté par un seul bruit blanc gaussien, de type *additif* (AWGN), dont la densité de probabilité suit une loi normale de moyenne nulle et de variance σ^2 . On a alors l'expression suivante, où $\sigma > 0$

$$p(v|u) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(v-u)^2}{2\sigma^2}}$$

2.2.2/ LE SPECKLE

En imagerie radar, sonar ou médicale, les surfaces que l'on veut observer sont "éclairées" par des sources cohérentes. Les propriétés locales de ces surfaces sont le siège de

réflexions multiples qui interfèrent entre elles pour générer un bruit de tavelures, ou speckle, dont l'intensité dépend de l'information contenue dans le signal observé.

Le speckle est ainsi un bruit de type *multiplicatif* qui confère aux observations une très grande variance qui peut-être réduite en moyennant plusieurs observations, ou vues, de la même scène. Si L est le nombre de vues, le speckle est traditionnellement modélisé par la PDF suivante :

$$p(v | u) = \frac{L^2 v^{L-1} e^{-L \frac{v}{u}}}{\Gamma(L) u^L}$$

L'espérance vaut $E[v] = u$ et la variance $\sigma^2 = \frac{u^2}{L}$ est effectivement inversement proportionnelle à L , mais pour le cas mono vue où $L = 1$, la variance vaut u^2 , soit un écart type du signal v égal à sa moyenne.

2.2.3/ LE BRUIT "SEL ET POIVRE"

Le bruit *sel et poivre*, ou bruit *impulsionnel* trouve son origine dans les pixels défectueux des capteurs ou dans les erreurs de transmission. Il tire son nom de l'aspect visuel de la dégradation qu'il produit : des pixels noirs et blancs répartis dans l'image. Le bruit impulsionnel se caractérise par la probabilité P d'un pixel d'être corrompu. La PDF peut alors être exprimée par parties comme suit, pour le cas d'images en 256 niveaux de gris (8 bits) :

$$p(v | u) = \begin{cases} \frac{P}{2} + (1 - P) & \text{si } v = 0 \text{ et } u = 0 \\ \frac{P}{2} + (1 - P) & \text{si } v = 255 \text{ et } u = 255 \\ \frac{P}{2} & \text{si } v = 0 \text{ et } u \neq 0 \\ \frac{P}{2} & \text{si } v = 255 \text{ et } u \neq 255 \\ (1 - P) & \text{si } v = u \text{ et } u \notin \{0, 255\} \\ 0 & \text{sinon} \end{cases}$$

2.2.4/ LE BRUIT DE POISSON

Aussi appelé *bruit de grenaille* (shot noise), ce type de bruit est inhérent aux dispositifs de détection des photons. Il devient prépondérant dans des conditions de faible éclairément, lorsque la variabilité naturelle du nombre de photons reçus par un photosite par intervalle d'intégration influe sur les propriétés statistiques du signal. Le bruit de grenaille est de type multiplicatif et suit une loi de Poisson. La PDF peut s'écrire comme suit :

$$p(v | u) = e^{-\frac{u}{v}} \frac{u^v}{v!}$$

2.3/ LES TECHNIQUES DE RÉDUCTION DE BRUIT

La très grande majorité des algorithmes de réduction de bruit fait l'hypothèse que la perturbation est de type gaussien, même si le développement des systèmes d'imagerie radar et médicale a poussé les chercheurs vers l'étude des bruits multiplicatifs du type *speckle*

ou *Poisson*. Un très grand nombre de travaux proposant des méthodes de réduction de ces bruits ont été menés, ainsi que beaucoup d'états de l'art et d'études comparatives de ces diverses techniques, que nous n'avons pas la prétention d'égaliser.

Les techniques et implémentations que nous aborderons dans le chapitre suivant sont celles qui ont un lien direct avec les travaux que nous avons menés. Nous présenterons donc les principales classes d'algorithmes de réduction de bruit et les implémentations GPU qui leur ont été consacrées.

2.4/ LES TECHNIQUES DE SEGMENTATION

La segmentation représente également un enjeu important dans le domaine du traitement d'image et à ce titre a fait l'objet d'abondants travaux et publications touchant les nombreux cas d'analyse dans lesquels une segmentation est utilisée. On peut citer la reconnaissance de formes, la détections et/ou la poursuite de cibles, la cartographie, le diagnostique médical, l'interaction Homme-machine, la discrimination d'arrière plan, etc.

On pourrait donner de la segmentation une définition spécifique par type d'usage, mais dans un souci d'unification, on propose la formulation générique suivante : "La segmentation consiste à distinguer les zones homogènes au sein d'une image". Dans cette définition, le caractère *homogène* s'entend au sens d'un critère pré établi, adapté aux contraintes particulières de traitement comme le type de bruit corrompant les images, ou bien la dimension du signal observé \bar{v} selon que l'image est en couleur ou non. Un tel critère peut ainsi être un simple seuil de niveau de gris ou bien nécessiter de coûteux calculs statistiques dont certains seront détaillés dans les chapitres suivants.

Devant la diversité des cas à traiter et des objectifs à atteindre, on sait aujourd'hui qu'à l'instar du filtre unique, la méthode universelle de segmentation n'existe pas et qu'une bonne segmentation est celle qui conduit effectivement à l'extraction des structures pertinentes d'une image selon l'interprétation qui doit en être faite.

Les éléments constitutifs de la segmentation sont soit des régions, soit des contours. Les deux notions sont complémentaires étant donné que les contours délimitent des régions, mais les techniques de calcul basés sur l'un ou l'autre de ces éléments relèvent d'abord différents. Les algorithmes de segmentation orientés régions s'appuient pour beaucoup sur des techniques de regroupement, ou *clustering*, pour l'identification et le peuplement des régions. Ce lien trouve son origine dans la psychologie du *gestalt* ?? où l'on considère que la perception conceptuelle s'élabore au travers de regroupements visuel d'éléments. La plupart des approches proposées jusqu'à très récemment consistent à minimiser une fonction d'énergie qui n'a pas de solution formelle et que l'on résout donc à l'aide de techniques numériques, souvent itératives.

2.4.1/ ANALYSE D'HISTOGRAMME

Les techniques les plus simples à mettre en œuvre en segmentation sont les techniques de seuillage, basées sur une analyse de l'histogramme des niveaux de gris (ou de couleurs) et cherchant à en distinguer les différentes classes comme autant d'occurrences représentant des *régions* homogènes. Différents critères peuvent être appliqués pour cette analyse, visant par exemple à maximiser la variance ?? ou encore à maximiser

le contraste pour déterminer les valeurs pertinentes des seuils. Malgré la multitude de variantes proposées, ces méthodes demeurent toutefois peu robustes et présentent l'inconvénient majeur de ne pas garantir la connexité des régions déterminées. On les réserve à des applications très spécifiques où, par exemple, on dispose d'une image de référence dont l'histogramme peut être comparé à celui des images à traiter. C'est le cas de certaines applications de contrôle industriel où la simplicité algorithmique permet de surcroît des implémentations très rapides, voire câblées. Ces techniques sont aujourd'hui considérées comme rudimentaires mais les calculs d'histogrammes et les analyses associées interviennent dans beaucoup d'algorithmes récents parmi les plus évolués et performants. La figure 2.1 illustre le traitement typique de l'histogramme de l'image d'entrée 2.1(a) dans le but de distinguer les deux régions du fond et du cochon (la cible). La première étape consiste à dresser l'histogramme des niveaux de gris sur tout le domaine de l'image 2.1(b). Il faut ensuite identifier le seuil de séparation des deux régions supposées, ici, homogènes au sens des valeurs de niveau de gris. Une estimation visuelle peut-être faite, mais on voit immédiatement que même dans une situation aussi claire, le choix du seuil n'est pas évident. Pour un traitement automatique, on peut par exemple proposer la technique itérative présentée par l'Algorithme 1 qui conduit à la segmentation de la figure 2.1(c). L'image 2.1(d) est l'image initiale, corrompue par un bruit gaussien de moyenne nulle et d'écart type 25. Les résultats de la segmentation (2.1(c) et 2.1(f)) de cette image sont clairement insuffisants le segment de la cible comporte des discontinuités et dans le cas de l'image bruitée, des pixels orphelins épars demeurent en quantité. Cette technique nécessiterait une étape supplémentaire pour disposer d'une segmentation pertinente.

Algorithme 1 : Calcul du seuil de séparation des segments de l'histogramme.

```

1 textbf :
2  $\bar{h} \leftarrow$  histogramme sur l'image ;
3  $S_{init} \leftarrow 128$  ;
4  $S_k \leftarrow S_{init}$  ;
5  $\epsilon \leftarrow 1$  ;
6 repeat
7      $\mu_{inf} = \frac{\sum_{i < S_k} h_i i}{\sum_{i < S_k} h_i}$  ;
8      $\mu_{sup} = \frac{\sum_{i \geq S_k} h_i i}{\sum_{i \geq S_k} h_i}$  ;
9      $S_k = \frac{1}{2}(\mu_{inf} + \mu_{sup})$  ;
10 until  $\|S_k - \frac{1}{2}(\mu_{inf} + \mu_{sup})\| < \epsilon$  ;
```

2.4.2/ ANALYSE DE GRAPHE

Un autre formalisme qui a généré une vaste classe d'algorithmes de segmentation est celui des graphes et repose sur l'idée que les régions de l'image sont représentées par les nœuds du graphe, alors que les liens traduisent les relations de voisinage existant entre

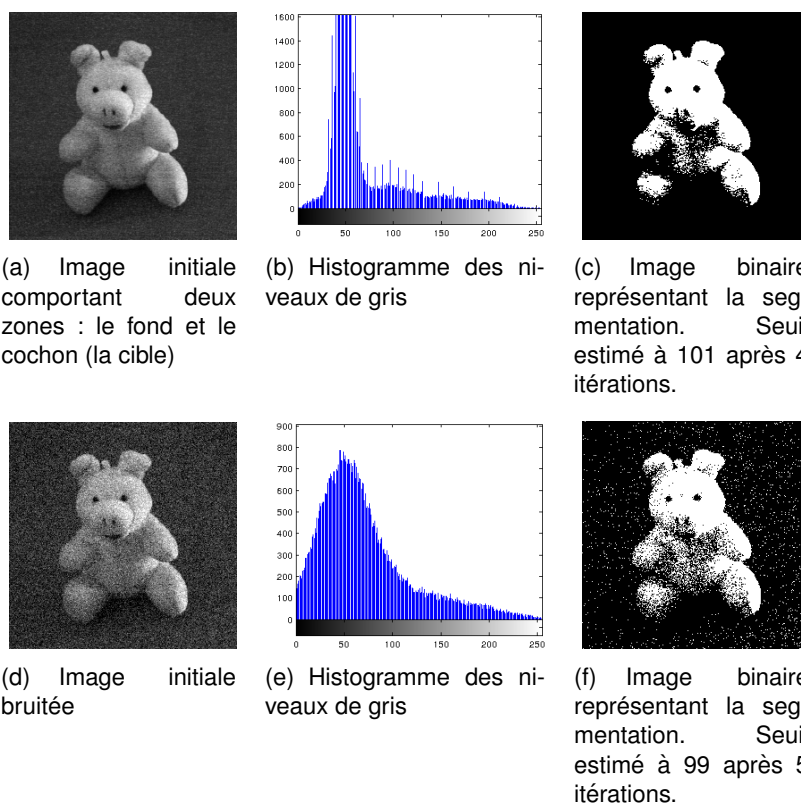


FIGURE 2.1 – Segmentation d’une image en niveaux de gris de 128×128 pixels par analyse simple d’histogramme. Colonne de gauche : image d’entrée. Colonne centrale : histogramme des niveaux de gris. Colonne de droite : résultat de la segmentation.

les régions. L'idée de base est d'initialiser le graphe avec un nœud pour chaque pixel. La segmentation est obtenue par simplification itérative du graphe, en évaluant les liens et en déterminant ceux à supprimer et ce, jusqu'à convergence. L'essentiel de la problématique réside donc dans la métrique retenue pour évaluer les liens ainsi que dans le critère de sélection et là encore, la littérature regorge d'une grande variété de propositions. Nous pouvons retenir que les premières d'entre elles, qui n'étaient pas spécifiquement dédiées à la segmentation d'images numériques mais au regroupement d'éléments répartis sur un domaine (1D ou 2D), ont été élaborées autour d'une mesure locale des liens basée sur la distance entre les éléments. La réduction du graphe est ensuite effectuée en utilisant un algorithme spécifique, comme le *minimum spanning tree*, dont l'application a été décrite dès 1970 dans ?? et où il s'agit simplement de supprimer les liens *inconsistants*, c'est à dire ceux dont le poids est significativement plus élevé que la moyenne des voisins se trouvant de chaque côté du lien en question. L'extension a rapidement été faite aux images numériques en ajoutant l'intensité des pixels au vecteur des paramètres pris en compte dans l'évaluation du poids des liens. D'autres critères de simplification ont aussi été élaborés, avec pour ambition de toujours mieux prendre en compte les caractéristiques structurelles globales des images pour prétendre à une segmentation qui conduise à une meilleure perception conceptuelle. Le principe général des solutions actuelles est proche de l'analyse en composantes principales appliquée à une matrice de similarité qui traduit les liens entre les segments. On peut citer, par ordre chronologique, les méthodes reposant sur le *graphe optimal* de Wu et Leahy ?? et plus récemment ?. Le principal point faible de ces techniques réside essentiellement dans la difficulté à trouver un compromis acceptable entre identification de structures globales et préservation des éléments de détails. Cela se traduit dans la pratique par un ensemble de paramètres à régler pour chaque type de segmentation à effectuer. Cependant, elles sont employées dans les algorithmes de haut niveau les plus récents, comme nous le verrons plus loin. La figure 2.2 montre un exemple de l'application de l'algorithme *normalized cuts* décrit dans ?? et implémenté par Cour, Yu et Shi en 2004. Cette implémentation utilise des valeurs pré-établies des paramètres de calcul de la matrice de similarité produisant de bonnes segmentations d'objets et/ou personnes dans les images naturelles, mais requiert de prédéterminer le nombre de segments à obtenir. Les images de la figure représentent les résultats obtenus avec un nombre de segments variant de 2 à 5 et montrent qu'il est difficile de trouver un compromis acceptable. Enfin, les temps d'exécutions peuvent devenir très rapidement prohibitifs, même avec des implémentations plus optimisées. Pour information, les résultats de la figure 2.2 ont été obtenus en 1.5 s environ (Matlab R2010 sur CPU intel core i5-2520M @ 2.50GHz - linux 3.2.0)

2.4.3/ KERNEL-MEANS, MEAN-SHIFT ET DÉRIVÉS

Parallèlement à la réduction de graphes, d'autres approches ont donné naissance à une multitude de variantes tournées vers la recherche des moindres carrés. Il s'agit simplement de minimiser l'erreur quadratique totale, ce qui peut se résumer, pour une image de N pixels, en la détermination du nombre C de segments Ω_i et leur contenu, de sorte à minimiser l'expression

$$\sum_{i \in [1..C]} \sum_{x_k \in \Omega_i} (v_k - \mu_i)^2$$

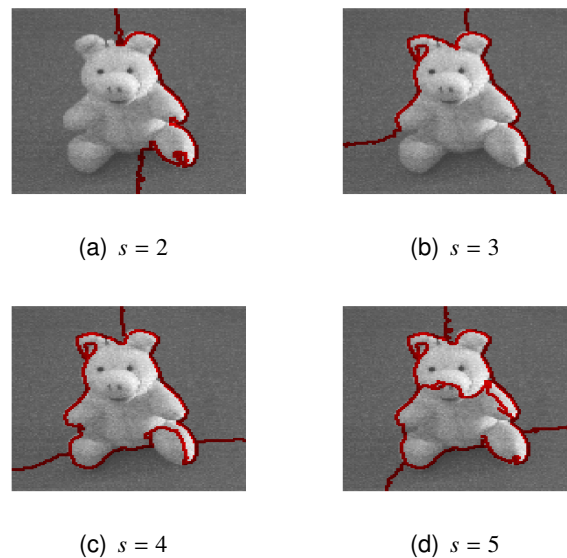


FIGURE 2.2 – Segmentation d’une image en niveaux de gris de 128×128 pixels par simplification de graphe de type *Normalized cut* pour un nombre s de segments variant de 2 à 5.

où μ_i représente la valeur affectée au segment Ω_i , i.e la valeur moyenne des observations v_k sur Ω_i , et $\bigcup_{i \in [1..C]} \Omega_i = \Omega$

Cette idée est très intuitive et simple, mais n’a pas souvent de solution explicite, d’autant que le nombre des segments est *a priori* inconnu. Dès 1965, Mac Queen a proposé l’appellation *k-means* pour cette procédure itérative de regroupement ?? qui débute avec k groupes d’un seul pixel¹ pris au hasard, puis d’ajouter chaque point au groupe dont la moyenne est la plus proche de la valeur du point à ajouter. La moyenne du groupe nouvellement agrandi doit alors être recalculée avant le prochain ajout. Cette implémentation est extrêmement simple à mettre en œuvre² mais elle possède de nombreux défaut dont le principal est qu’elle ne converge pas nécessairement vers le regroupement optimal, même si on connaît la “bonne” valeur de k . Un autre est d’être très dépendant du choix des k éléments initiaux, en nombre et en position.

Toutefois, vraisemblablement du fait de sa simplicité d’implémentation et de temps d’exécution rapides, la communauté scientifique s’est beaucoup penchée sur cette méthode pour en compenser les défauts, jusqu’à en faire une des plus employées, en particulier par les statisticiens. On compte aussi beaucoup de variantes telles les *k-centers* ?? et les *k-médians* ?? qui n’emploient pas la moyenne arithmétique comme expression du “centre” d’un segment. Des solutions ont aussi été apportées pour l’estimation de k en employant, par exemple, un critère de vraisemblance pour choisir la meilleure valeur de k dans un intervalle donné ?? . À titre d’illustration et de comparaison, l’image du cochon a été traitée par une implémentation naïve de l’algorithme original des *k-means* en donnant successivement au nombre de segments les valeurs $s = 2$ à $s = 5$. Les résultats sont reproduits à la figure 2.3 et montrent encore une fois l’influence de s sur la segmentation.

1. Dans son article, MacQueen ne parle pas de pixel mais de point. En effet, la méthode décrite ne visait pas à segmenter des images, mais des données de natures diverses.

2. Même si en 1965, rien n’était simple à programmer

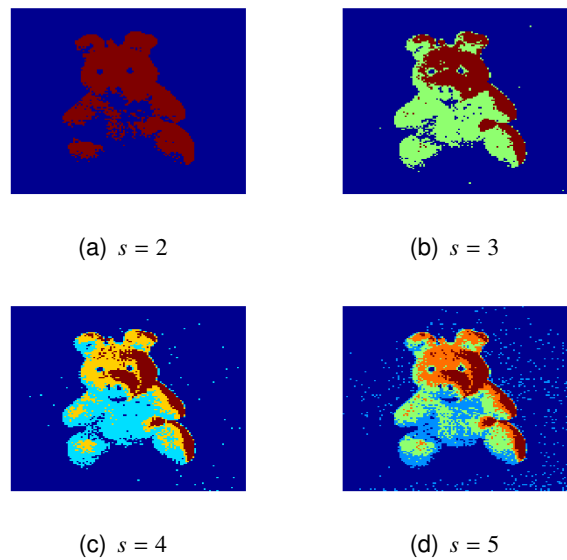


FIGURE 2.3 – Segmentation d’une image en niveaux de gris de 128×128 pixels par algorithme *k-means* pour un nombre s de segments variant de 2 à 5. Chaque couleur est associée à un segment. Les couleurs sont choisies pour une meilleure visualisation des différents segments.

Un algorithme initialement proposé en 1975 par Fukunaga et Hostetler ?? permet de manière plus générique de déterminer le nombre de segments, ou modes, ainsi que les points, ou pixels, qui les composent. Il cherche pour ce faire à localiser les k positions où le gradient de densité s’annule. Il utilise un voisinage pondéré (ou *kernel*) et détermine le centre de masse des segments en suivant itérativement le gradient de densité dans le voisinage autour de chaque élément du domaine. Lorsque l’algorithme a convergé, les k segments sont identifiés et contiennent chacun l’ensemble des points qui ont conduit à leur centre de masse respectif. Étonnemment, malgré ses qualités intrinsèques, cet algorithme du *mean-shift* est resté longtemps sans susciter de grand intérêt, jusqu’à l’étude de Cheng ?? qui en a démontré les propriétés et établi les liens avec d’autres techniques d’optimisation comme la descente/montée de gradient ou de filtrage comme le floutage. Comaniciu et Peer ont alors étendu l’étude et proposé une application à la segmentation en utilisant l’espace colorimétrique CIELUV ?? et montré qu’elle permettait une meilleure identification des modes de l’image ?. Une implémentation de la variante proposée par Keselman et Micheli-Tzanakou dans ?? appliquée à notre image de test fournit les résultats reproduits à la figure 2.4. Pour se rapprocher des traitements précédents, nous avons identifié, par essais successifs, les tailles de voisinage conduisant à des nombres de segments identiques à ceux des figures précédentes (de 2 à 5). Le volume minimal admis pour un segment a été arbitrairement fixé à 100 pixels.

Il est à noter que les segmentations basées sur des algorithmes de *clustering* comme ceux que l’on vient de présenter nécessitent le plus souvent une phase supplémentaire de génération des frontières inter-segments et d’affectation de la valeur de chaque segment aux éléments qui le composent. Par ailleurs, dans les deux cas du *k-means* et du *mean-shift*, chaque itération génère une réduction de la variance (due au moyennage) et on peut donc rapprocher ces techniques de celles de réduction de bruit par minimisation de

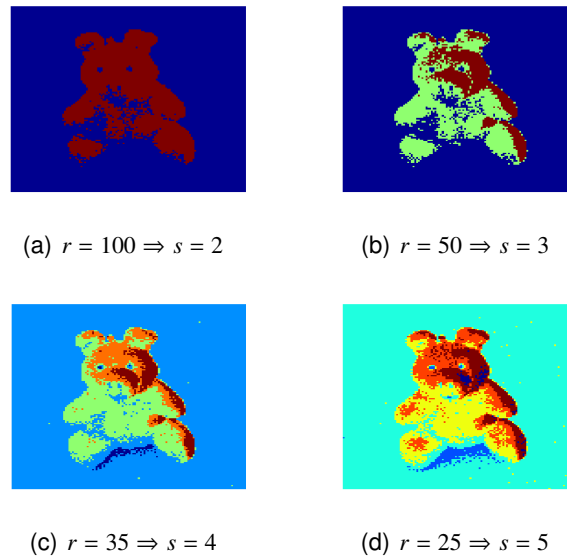


FIGURE 2.4 – Segmentation d’une image en niveaux de gris de 128×128 pixels par algorithme *mean-shift* pour un rayon de voisinage r de 100, 50, 35 et 25 pixels permettant d’obtenir un nombre s de segments variant respectivement de 2 à 5. Le volume minimal admis pour un segment est fixé à 100 pixels. Chaque couleur est associée à un segment. Les couleurs sont choisies pour une meilleure visualisation des différents segments.

variance.

2.4.4/ LES CONTOURS ACTIFS, OU *snakes*

Contrairement aux précédentes techniques et comme leur nom le laisse deviner, les éléments constitutifs de ces méthodes sont cette fois des *contours* et non plus des *régions*. De fait, ils définissent nativement une segmentation de l’image. Le principe général est de superposer une courbe paramétrique à l’image, le *snake*, puis de lui appliquer des déformations successives destinées à rapprocher le *snake* des contours de l’objet. Les déformations à appliquer sont guidées par l’évaluation d’une fonction d’énergie prenant en compte :

- l’énergie interne de la courbe, fonction de son allongement de sa courbure.
- l’énergie externe liée à l’image, fonction de la proximité de la courbe avec les zones de fort gradient et éventuellement une contrainte fixée par l’utilisateur comme des points imposés par exemple.

Ici encore, la résolution du problème revient à minimiser une fonction d’énergie sous contrainte et les diverses techniques de résolution numérique peuvent s’appliquer comme pour les autres classes d’algorithmes itératifs présentés précédemment, avec ici encore, un nombre de paramètres à régler assez important.

Dans sa version originale proposée par Kass *et al.* en 1988 ??, l’algorithme dit du *snake* présente l’intérêt de converger en un nombre d’itérations assez réduit et permet de suivre naturellement un *cible* en mouvement après une convergence initiale à une position donnée, chaque position de convergence fournissant une position initiale pertinente pour la position suivante. Toutefois, il se montre particulièrement sensible à l’état initial de la

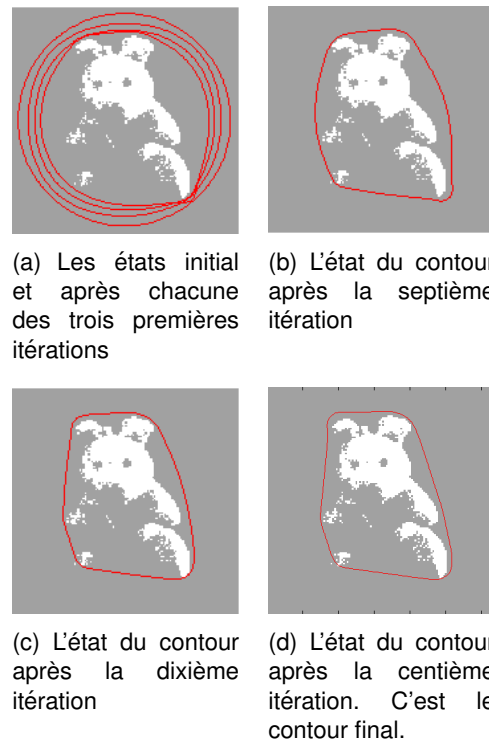


FIGURE 2.5 – Segmentation d’une image en niveaux de gris de 128×128 pixels par algorithme dit du *snake*, dans sa version originale. Les paramètres d’élasticité, de viscosité, de raideur et d’attraction ont été fixés respectivement aux valeurs 5, 0, 0.1 et 5.

courbe et requiert souvent de celle-ci qu’elle soit assez proche de l’objet à “entourer”, sous peine de se verrouiller dans un minimum local. La sensibilité au bruit n’est pas non plus très bonne du fait de la formulation locale de l’énergie. Les “concavités” étroites ou présentant un goulot d’étranglement marqué sont par ailleurs mal délimitées. Enfin, la fonction d’énergie étant calculée sur la longueur totale de la courbe, cela pénalise la bonne identification des structures de petite taille vis à vis de la longueur totale de la courbe. La figure 2.5 illustre ces défauts en montrant quelques états intermédiaires ainsi que le résultat final d’une segmentation réalisée à partir d’un contour initial circulaire et des paramètres réglés empiriquement, en employant la méthode du *snake* original. On voit que la convergence est assez rapide mais que le contour ainsi déterminé ne “colle” pas bien à l’objet que l’on s’attend à isoler.

Il est cependant possible de contrôler la finesse de la segmentation mais au prix de temps de calculs qui peuvent devenir très longs. Parmi les variantes élaborées qui tentent de pallier ces défauts, les plus intéressantes sont :

- le *balloon snake*, conçu pour remédier au mauvais suivi des concavités en introduisant une force supplémentaire de pression tendant à *gonfler* le snake jusqu’à ce qu’il rencontre un contour suffisamment marqué. Cela suppose toutefois que l’état initial de la courbe la situe entièrement à l’intérieur de la zone à segmenter et est surtout employé dans des applications semi-automatiques où l’utilisateur définit au moins une position et une taille initiales pour la courbe.
- le *snake GVF* (pour Gradient Vector Flow), dont le but est de permettre qu’une initialisation lointaine de la courbe ne pénalise pas la segmentation. Une carte des lignes de gradient est établie sur tout le domaine de l’image et sert à intégrer une force

- supplémentaire dans l'énergie totale, qui attire la courbe vers la zone de fort gradient.
- les *level-sets*, dont la particularité est de ne pas employer directement une courbe paramétrique plane mais de définir l'évolution des frontières comme l'évolution temporelle de l'ensemble des points d'une surface 3D soumise à un champ de force, tels que leur élévation soit constamment nulle. Les propriétés des contours actifs par *level-sets* se sont révélées intéressantes, en particulier la faculté de se disjoindre ou de fusionner, mais les temps de calcul très pénalisants. Après la formulation initiale de Osher et Sethian en 1988 ??, plusieurs façon de réduire le coût du calcul ont été formulées, dont les plus importantes restent les techniques dites *narrow band* ?? (bande étroite) qui ne calcule à chaque itération que les points dans une bande étroite autour du plan $z = 0$ de l'itération courante et *fast marching* ?? qui s'applique dans le cas particulier d'une évolution monotone des fronts.
 - les *snake* orientés régions, qui visent essentiellement à mieux caractériser les zones à segmenter et améliorer la robustesse vis à vis du bruit en employant une formulation de l'énergie calculée sur le domaine complet de l'image ?. Les premiers résultats confirment la qualité de cette méthode, mais la nécessité d'effectuer les calculs sur l'image entière génèrait des temps de traitement prohibitifs jusqu'à ce que Bertaux et al. proposent une amélioration algorithmique exacte permettant à nouveau un calcul en 1D, le long de la courbe, moyennant une simple étape initiale générant un certain nombre d'images intégrales ?. La section ?? qui introduit notre contribution à cette technique en donnera une description détaillée.

2.4.5/ MÉTHODES HYBRIDES

Aujourd'hui, les algorithmes de segmentation les plus performants en terme de qualité emploient des techniques qui tentent de tirer le meilleur parti de plusieurs des méthodes "historiques" décrites précédemment. Le meilleur exemple, et le seul que nous citerons, est le détecteur de contour et l'algorithme de segmentation associé proposé par Arbelaez *et al.* en 2010 ?. Il compose avec la constructions d'histogrammes locaux pour générer une matrice de similitude (affinity matrix) et appliquer les techniques liées à la théorie des graphes pour réduire la dimension de l'espace de représentation (calcul des valeurs et vecteurs propres). Il utilise ensuite une technique adaptée de *ligne de partage des eaux* ?? (que l'on aurait rangée avec les mean-shift) pour regrouper les segments. Les résultats sont très bons et des implémentations efficaces ont d'ores et déjà été écrites (voir section ??).

2.5/ L'ÉTAT DE L'ART DES IMPLÉMENTATIONS GPU

LA SEGMENTATION ORIENTÉE RÉGIONS DANS LES IMAGES BRUITÉES

3.1/ PRÉSENTATION - EXISTANT

3.2/ LA PARALLÉLISATION DU SNAKE POLYGONAL

LE FILTRAGE DES IMAGES SUR GPU

- 4.1/ ALGORITHME DE RÉDUCTION DE BRUIT PAR RECHERCHE DES LIGNES DE NIVEAUX
- 4.2/ FILTRE MÉDIAN
- 4.3/ FILTRES DE CONVOLUTION

CONCLUSION GÉNÉRALE

TABLE DES FIGURES

2.1	Segmentation d'une image en niveaux de gris de 128×128 pixels par analyse simple d'histogramme. Colonne de gauche : image d'entrée. Colonne centrale : histogramme des niveaux de gris. Colonne de droite : résultat de la segmentation.	16
2.2	Segmentation d'une image en niveaux de gris de 128×128 pixels par simplification de graphe de type <i>Normalized cut</i> pour un nombre s de segments variant de 2 à 5.	18
2.3	Segmentation d'une image en niveaux de gris de 128×128 pixels par algorithme <i>k-means</i> pour un nombre s de segments variant de 2 à 5. Chaque couleur est associée à un segment. Les couleurs sont choisies pour une meilleure visualisation des différents segments.	19
2.4	Segmentation d'une image en niveaux de gris de 128×128 pixels par algorithme <i>mean-shift</i> pour un rayon de voisinage r de 100, 50, 35 et 25 pixels permettant d'obtenir un nombre s de segments variant respectivement de 2 à 5. Le volume minimal admis pour un segment est fixé à 100 pixels. Chaque couleur est associée à un segment. Les couleurs sont choisies pour une meilleure visualisation des différents segments.	20
2.5	Segmentation d'une image en niveaux de gris de 128×128 pixels par algorithme dit du <i>snake</i> , dans sa version originale. Les paramètres d'élasticité, de viscosité, de raideur et d'attraction ont été fixés respectivement aux valeurs 5, 0, 0.1 et 5.	21

LISTE DES TABLES

Résumé :

Ceci est le résumé en français

Mots-clés : Mot-clé 1, Mot-clé 2

Abstract:

This is the abstract in English

Keywords: Keyword 1, Keyword 2

The logo for SPIM (École doctorale SPIM) features a solid yellow horizontal bar on the left, followed by the letters 'S', 'P', 'I', and 'M' in a white, sans-serif font. The 'S' is slightly larger and more stylized than the other letters.

■ École doctorale SPIM 16 route de Gray F - 25030 Besançon cedex
■ tél. +33 (0)3 81 66 66 02 ■ ed-spim@univ-fcomte.fr ■ www.ed-spim.univ-fcomte.fr

The logo for the University of Franche-Comté (UFC) consists of a vertical yellow bar on the left with the text 'UNIVERSITÉ DE FRANCHE-COMTÉ' written vertically inside it. To the right of the bar, the letters 'U' and 'FC' are prominently displayed in a large, bold, black font, with 'UNIVERSITÉ DE FRANCHE-COMTÉ' written in a smaller, black font below them.