



Interpolation multi-échelle des attributs des images

Marine Campedel
Jean-François Aujol
Yann Gousseau
Henri Maître
Michel Roux
Luo Bin

2005D009

Mai 2005

Département Traitement du Signal et des Images
Groupe Traitement et Interprétation des Images

Ecole Nationale Supérieure des Télécommunications

Groupe des Ecoles des Télécommunications - membre de ParisTech
46, rue Barrault - 75634 Paris Cedex 13 - Tél. + 33 (0)1 45 81 77 77 - www.enst.fr
Département TSI



**Rapport à mi-parcours de l'étude DCT/SI/AP-04-522
effectuée à la demande du CNES**

Interpolation multi-échelle des attributs des images

**Marine Campedel, Jean-François Aujol, Yann Gousseau, Henri Maître,
Michel Roux, Luo Bin**

16 mai 2005

Département Traitement du Signal et des Images
GET - Télécom Paris

Table des matières

Introduction	5
1 Echelles et résolution	9
1.1 Introduction	9
1.1.1 Quelles résolutions ?	9
1.1.2 L'enjeu de la diversité en fouille d'image	9
1.1.3 La diversité des images	11
1.2 Que faire avec la diversité de résolution ?	12
1.2.1 Définitions et notations	12
1.2.2 Un panorama des stratégies	13
2 Robustesse des primitives au changement d'échelle	17
2.1 Protocole expérimental	17
2.2 Extraction des caractéristiques	18
2.3 Expériences	18
2.3.1 La base de données de texture	18
2.3.2 Les vignettes de résolutions différentes	19
2.3.3 La validation croisée	20
2.4 Résultats	21
2.4.1 Robustesse des caractéristiques extraites par les méthodes GABOR, HARALICK et QMF	21
2.4.2 Résultats sur les caractéristiques sélectionnées	21
2.5 Conclusion	23
3 Espace multi-échelle linéaire	29
3.1 Mesure d'échelle dans l'espace multi-échelle linéaire	29
3.2 Mesure d'échelle basée sur la variation totale	30
3.2.1 Rappel : espace BV des fonctions à variations bornées	30
3.2.2 Principe	31
3.2.3 Expériences et résultats	31
3.3 Mesure basée sur le flot de la variation totale (TV)	35
3.3.1 Expériences et résultats	36
3.3.2 Comparaison avec la méthode basée sur la variation totale et pers- pective	38

4	Comportement structure/texture suivant la résolution de l'image	43
4.1	Introduction	43
4.2	Protocole	43
4.3	Résultats	44
4.4	Mesure de l'échelle de certains éléments de l'image	45
	Conclusion	55
	Bibliographie	57

Avertissement

Ce document est le rapport à mi-parcours de l'étude : **DCT/SI/AP-04-522 « Interpolation Multi-échelle des attributs des images »** . Il accompagne un rapport concernant l'étude : **DCT/SI/AP-04-521 « Etude des attributs image pertinents de l'imagerie PLEIADES »**. Ces deux études sont exécutées à la demande du CNES par le département TSI de Télécom-Paris, et dans le cadre du Centre de Compétence CNES-DLR-ENST.

Ces deux rapports sont très complémentaires, les travaux qu'ils relatent étant intimement imbriqués, mais présentés ici de façon séparée pour des raisons contractuelles. Ces travaux font suite à deux rapports [8], [9], aux conclusions regroupés dans un document public : [7].

Ces travaux ont également bénéficié de l'assistance de Ivan Kyrgyzov et Mihai Costache pour la constitution des bases de données et l'expérimentation et des conseils d'Eric Moulines pour l'optimisation des systèmes d'apprentissage.

Alain Giros (CNES) et Mihai Datcu (DLR) ont contribué au choix des méthodes présentées, ainsi qu'aux discussions des résultats.

Introduction

Dans ce rapport nous présentons les premiers résultats de l'étude conduite sur l'impact des résolutions variées affectant les images satellitaires, sur la constitution d'archives et sur l'indexation de ces archives.

Dans une première partie (chapitre 1), nous présentons quelques réflexions générales associées à la notion de résolution. Elles s'appuient tout d'abord sur quelques remarques concernant la résolution, puis sur un inventaire des principaux capteurs d'imagerie haute résolution disponibles ou à venir, conduisant au constat de la diversité des résolutions disponibles aujourd'hui, pour aboutir à quelques propositions d'actions qui ont guidé nos travaux.

Dans une seconde étude (chapitre 2), nous nous penchons sur le problème de la robustesse des primitives au changement d'échelle. Cette étude est conduite de façon à déterminer quel champ d'échelles peut être couvert par un même jeu de primitives mesurées à une échelle précise. L'objectif de ce travail est de limiter le domaine où l'on sera amené à mettre en place une politique d'adaptation aux changements d'échelle, par exemple par recalcul des index ou par interpolation/extrapolation des index.

Puis, nous nous intéressons à deux problèmes fondamentaux en traitement d'image qui l'un et l'autre abordent le problème de l'échelle, mais par des voies sensiblement différentes.

Le premier (chapitre 3) est celui de la mesure de l'échelle des éléments présents dans une image. Il s'agit de proposer une approche rigoureuse qui permette, étant donnée une image, de dire quelle est la taille caractéristique des éléments qui la composent. Pour cette mesure d'échelle, une difficulté supplémentaire vient du fait qu'il s'agit à la fois d'une notion locale et globale : localement, nous voulons savoir quelle taille d'élément caractéristique associer à chaque pixel, et globalement, nous voulons être en mesure d'associer une échelle à une image. Pour proposer cette mesure intrinsèque de l'image, nous sommes amenés à introduire un critère numérique.

Dans la littérature, comme nous le verrons dans la suite dans la Section 3, il existe déjà des méthodes de mesure d'échelle [18, 6]. Nous montrons les idées de quelques unes de ces approches, et nous proposons deux nouvelles méthodes. L'une est basée sur la variation totale (TV), l'autre s'appuie sur le flot de la variation totale. La variation totale d'une image régulière correspond essentiellement à la valeur du module de son gradient. La vitesse du flot de la variation totale (TV) dépend de la taille et de la forme des régions dans l'image. Chacune des deux méthodes consiste essentiellement en deux étapes. On représente d'abord l'image dans l'espace multi-échelle linéaire, puis on calcule les critères numériques (la variation totale ou la vitesse du flot de la variation totale dans notre cas) sur chaque échelle. On peut ainsi déduire une échelle caractéristique à l'aide du graphe de cette mesure en fonction de l'échelle. Nous reviendrons plus en détails dans les Sections 3.2 et 3.3 sur ces notions. L'intérêt de la variation totale en traitement d'images est apparu avec le travail [28] en res-

tauration d'images. On sait que l'espace associé des fonctions à variations bornées modélise particulièrement bien la géométrie des images [23, 4].

Le deuxième problème (chapitre 4) est celui du comportement structure/texture d'une image. Selon la résolution à laquelle on considère une image, elle peut apparaître comme une texture ou de la géométrie. Typiquement, des images de zone urbaine à une résolution de 10 mètres ont un comportement essentiellement d'images texturées, alors que les mêmes scènes à 25 centimètres ont essentiellement un comportement d'image géométrique. Il est facile de deviner que ce deuxième problème est lié à celui du calcul de l'échelle des éléments d'une image. Il nous faut là aussi proposer un critère numérique pour préciser le comportement d'une image. C'est ce que nous faisons dans la Section 4 : nous proposons une méthode pour déterminer le comportement géométrie/texture d'une image suivant la résolution, et nous faisons le lien avec les mesures d'échelle des sections précédentes.

Le plan du rapport est le suivant.

Dans le premier chapitre, nous examinerons quelques propriétés associées à la diversité des images satellitaires (section 1.1.1), puis nous proposerons plusieurs stratégies pour prendre en compte ces diversités (section 1.2).

Dans le second chapitre, consacré à l'étude expérimentale de la robustesse des primitives d'indexation aux variations d'échelle, nous présenterons tout d'abord le protocole mis en place : 2.1, puis nous présenterons les expériences conduites (section 2.3), enfin les résultats et les conclusions qui établissent la robustesse des primitives (section 2.4).

Dans la Section 3, nous rappellerons quelques éléments sur la construction d'espaces multi-échelles linéaires. La Section 3.1 présentera quelques notions fondamentales des mesures d'échelles, et introduira une mesure largement utilisée dans la littérature. Dans la Section 3.2, nous proposerons une méthode de calcul de l'échelle d'une image basée sur la variation totale. Dans la Section 3.3, nous comparerons cette approche avec une version améliorée de celle de [6] basée sur la vitesse du flot de la variation totale. Dans la Section 4, nous nous intéresserons au problème du comportement géométrie/texture d'une image, et nous ferons le lien avec les mesures d'échelles proposées précédemment.

Nous terminerons par quelques remarques et perspectives en guise de conclusion.

Chapitre 1

Echelles et résolution

1.1 Introduction

L'une des problématiques actuelles de l'interprétation des images issues des archives satellitaires concerne la multiplication des satellites et des capteurs. Si l'on s'intéresse aux seuls capteurs métriques ou décimétriques (de la classe de SPOT), on trouve surtout aujourd'hui, outre les satellites militaires dont les images sont très peu disponibles, des images issues des satellites SPOT 1 à 4, à résolution de 10 m, de SPOT 5 à résolution de 5m ou 2,5 m, de Quickbird, Ikonos ou Orbview en dessous d'un mètre. De nombreux autres satellites sont attendus dans les années à venir, et en particulier la famille des Pléiades à résolution sub-métrique. On passera en revue rapidement ces satellites au paragraphe 1.1.1.

Quels sont les problèmes spécifiques associés à la multiplicité de ces capteurs dans la problématique de l'indexation et de la fouille de données ? Ceci sera examiné au paragraphe 1.1.2.

Nous verrons ensuite diverses stratégies pour prendre en compte ces diversités, Cela constituera la dernière section (section 1.2) de ce chapitre.

1.1.1 Quelles résolutions ?

Cette étude est essentiellement centrée autour des satellites modernes à haute résolution spatiale, et typiquement autour des satellites SPOT 5 et Pléiades, donc des résolutions évoluant de 70 cm à 5 m. Si l'on veut étendre la gamme des satellites concernés, on inclura les satellites militaires ou les satellites à venir dont la résolution sera de l'ordre de 30 cm (en "one foot resolution satellites") et vers les basses résolution on s'arrêtera Spot 1-3 multibandes et aux meilleurs Landsat 7, soit environ 20 m. C'est donc une échelle de l'ordre de 100 qu'il faudra pouvoir couvrir.

Le tableau 1.1 regroupe les principaux satellites civils (ou à usage dual) dans la gamme de résolution qui nous concerne.

1.1.2 L'enjeu de la diversité en fouille d'image

Tout d'abord la spécificité des capteurs (sensibilité, résolution, agilité, périodicité de repassage, époque de vie) fait que les images sont porteuses d'information potentiellement

appareil	pixel au sol (en m)
MOMS-01 (sur la navette challenger)	20
MOMS-2P (sur MIR PRIRODA)	6
Proba HRC (anglais)	18
Landsat 7	15
JERS-1 (Japon) avec OPT	18
AVNIR-2 (lancement 2005)	10
SPOT 1-4 Panchro et multispectral	10 et 20
SPOT 5 Panchro et multispectral	2,50 et 5
RESURS F1 (Russe) avec chambre KFA1000	10
RESURS F1 (Russe) avec chambre KFA3000	2
CBERS 1 et 2 (Sino-brésiliens)	20
CBERS 3 et 4 à venir	3
IRS 1C et 1D (Indien) (capteur PAN)	5,80
IRS P6 (Indien)(capteur LISS IV	6
IRS2A (lancement 2005)	5,10
Ikonos 2	1
futur Ikonos	0,50
QuickBird 2	0,82
Orbview 3	1 à 3
Orbview 7 (lancement prévu en 2007)	0,41
EROS A+ et B (Israel)	180 et 0,70
Pleïades (prévu en 2007)	0,70
Formosat-2 (Taiwan)	2
SPIN-2 (Russie)	2

TAB. 1.1 – Les principaux satellites à résolution métrique ou décimétrique (ne sont pas portés les satellites militaires), d’après [26].

complémentaire. Cette variété est une richesse que l’utilisateur souhaite utiliser, soit pour chaque image individuellement, soit en recoupant ces diverses images (ce que l’on regroupe souvent sous le vocable **fusion**).

Mais cette diversité est également un facteur de confusion et d’erreur car les apparences de ces diverses images peuvent rendre plus difficile leur exploitation. Typiquement lors d’une fouille de données, les critères qui permettent de rappeler des scènes d’intérêt peuvent fonctionner pour un capteur mais pas pour un autre et l’on peut ainsi ignorer des informations d’intérêt. C’est particulièrement vrai lorsque l’on fait des requêtes par l’exemple, puisque les exemples sont issus d’une famille de capteurs et peuvent rendre difficile le rappel d’images d’autres capteurs.

Nous allons donc examiner comment l’on peut espérer résoudre certains problèmes associés à cette diversité (parmi lesquels la résolution est l’une des sources de diversité les plus évidentes). Tout d’abord nous examinerons les sources de diversité, puis nous ferons un inventaire des techniques possibles pour prendre en compte spécifiquement la variété des résolutions.

1.1.3 La diversité des images

On peut classer de nombreuses façons les éléments de variabilité entre images. Nous avons choisi de distinguer celles qui sont intrinsèques à la scène, de celles qui sont liées au système d'imagerie.

La diversité due à la scène observée

La relative permanence des traits géomorphologiques et des grandes implantations anthropiques donne à la plupart des paysages une constance sur la durée des observations satellitaires (les 25 dernières années) que l'on souhaite mettre à profit pour assurer une très bonne robustesse des techniques de fouille d'image.

Néanmoins on ne peut négliger les évolutions des paysages et ces évolutions sont même souvent un élément important de caractérisation des zones comme il a été montré dans [16] en particulier. Ces évolutions peuvent être rangées selon diverses échelles de temps dans lesquelles se passent les phénomènes, comme c'est détaillé dans [16] :

- les grandes évolutions des boisements, des cours d'eau, de l'urbanisation, prenant place sur des décennies ;
- l'évolution du parcellaire, les constructions urbaines et les modifications routières et des infrastructures, dans l'échelle de l'année ;
- les phénomènes saisonniers foliation, maturation des cultures, et les accidents naturels (feux, inondations) sur l'échelle du trimestre,
- les cycles des activités anthropiques : fumées d'usines, travaux agricoles, coupes de forêts, évoluant de la semaine au trimestre ;
- les phénomènes météorologiques : pluies, neige, brouillards, à l'échelle du jour ou du mois ;
- les nuages et les effets du vent (tempêtes, cyclones), les activités volcaniques, mais aussi la hauteur du soleil, à l'échelle du jour ;
- les phénomènes "instantanés" : avions, éclairs, voitures, bateaux, marées, etc.

Ces variations peuvent être périodiques (ou quasiment) et parfois prévisibles car on dispose généralement d'éléments permettant de les faire intervenir lors de la fouille ou de l'indexation (date de prise de vue, couverture nuageuse, bulletins météo, activités volcaniques, hauteur des marées, niveau des rivières, ...), mais on dispose cependant rarement de capacité pour les modéliser (sauf pour des problèmes simples (inondations, nuages, brouillards, fumées, incendies). A notre connaissance elles n'ont jamais été prises en compte pour améliorer les capacités des techniques de recherche dans les bases de données, aussi bien sous des formes très simples que sur des modèles complexes de type GMES).

La diversité due aux capteurs

Les capteurs interviennent également très fortement pour induire une diversité dans les images collectées.

Tout d'abord la (ou les) gamme de sensibilité des capteurs diffèrent. Sans parler de capteurs utilisant des longueurs d'ondes complètement différentes comme les radar et les optiques, ou même les IR thermiques, il est rare qu'il y ait une correspondance exacte dans les sensibilités spectrales. Là encore, à notre connaissance, peu d'efforts ont été faits pour

modéliser les différences et - par exemple - calculer l'apparence d'une image SPOT5 comme si elle était issue d'un capteur avec une sensibilité spectrale d'IKONOS ou de Landsat. Très difficile pour ce qui concerne les images panchromatiques (à moins de disposer d'une segmentation thématique et de connaître pour chaque classe les réponses spectrales), cette opération serait probablement plus accessible à partir d'images multispectrales, et encore bien meilleures à partir d'images hyperspectrales.

Une deuxième différence provient de la visée adoptée par le satellite. L'accroissement de l'agilité des satellites autorise aujourd'hui de très importants dépointés par rapport au nadir qui entraînent trois types de différences :

1. des changements d'apparence dus à la parallaxe, en particulier dans les milieux urbains, mais aussi sur toutes les zones fortement montagneuses ou à déclivité importante (falaises, carrières, barrages, etc.) ;
2. des changements de luminosité dus à l'angle d'observation ; ces différences sont généralement mineures et tout à fait incluses dans les diversités soulignées en section 1.1.3, sauf pour des aspects particuliers liés aux réflexions spéculaires et aux *hot spots*
3. des changements dus à la résolution (et en règle générale à la réponse impulsionnelle) des capteurs qui "filtrent" la scène, en laissant plus ou moins apparents des contenus fréquentiels différents.

C'est cette dernière source de disparité, la plus importante puisqu'elle affecte directement le contenu de la scène, que nous allons maintenant prendre en compte.

1.2 Que faire avec la diversité de résolution ?

1.2.1 Définitions et notations

Introduisons tout d'abord quelques définitions et notations qui nous seront utiles par la suite. Il est en particulier important de clarifier le vocabulaire que nous utiliserons.

La grandeur directement accessible pour tout satellite est la **distance au sol entre pixels** (par exemple 10 m pour Spot 1 à 4). Nous l'appellerons plus brièvement **pixel au sol**.

Plus complexe est le terme de résolution.

Il est commun de dire : « Le satellite Spot-1 a une résolution de 5 m, tandis que Spot-5 a une résolution de 2,5 m. ». De même on dit : « Spot 5 est un satellite à plus haute résolution que Spot-1. » Ces 2 phrases ne peuvent être simultanément vraies. La résolution est, par définition [5], la mesure des plus petits détails distinguables, et cette définition valide la première phrase. Mais l'usage a solidement établi la recherche de haute résolution comme moyen d'accès aux fins détails et valide en ce sens la seconde phrase. C'est sur cet usage que nous nous alignerons par la suite.

La **résolution** d'un capteur s'exprime donc par sa capacité à distinguer des détails fins et divers critères peuvent être proposés (celui de Rayleigh, déduit de la figure d'Airy, est le plus connu [27]¹). Des critères de résolution exprimés par un chiffre seul ne sont pas significatifs pour comparer finement des capteurs dont les réponses impulsionnelles sont

¹bien adapté à des pupilles circulaires entachées de la seule limitation de la diffraction, le critère d'Airy n'est pas très adapté aux capteurs CCD.

différentes, mais ils sont très commodes. Afin de ne pas entrer dans une étude détaillée de tous les capteurs satellitaires et de leurs réponses impulsionnelles après correction au sol, nous considérerons par la suite que les images fournies par les opérateurs des satellites (après les traitements au sol) vérifient exactement le critère de Nyquist et donc la condition limite d'échantillonnage de Shannon [22]².

Sous ces hypothèses, nous exprimerons la résolution r de tous les satellites comme la fréquence maximale au sol (exprimée en m^{-1}) qu'ils peuvent analyser, donc à partir du pas d'échantillonnage au sol (le pixel au sol) d :

$$r = \frac{1}{2d}$$

Une telle définition ne préjuge donc pas des capacités de deux satellites A et B de même pixel au sol de distinguer de façon très différente des détails fins puisqu'à l'intérieur de la bande passante commune, ils peuvent respecter plus ou moins bien le contenu fréquentiel des objets de la scène. Par contre, avec notre définition, tout satellite dont le pixel au sol est de $2m$, a une résolution de $1m^{-1}$ et ne verra aucune fréquence supérieure à $1m^{-1}$.

Considérons inversement un objet de la scène, dont une dimension caractéristique est δ , (par exemple une cuve de diamètre δ). Il ne sera perçu comme un objet de dimension δ par un satellite de pixel au sol d que si $\delta > d$. Les objets de dimension $\delta < d$ contribuent à l'énergie du pixel auquel ils appartiennent à hauteur de leur contenu spectral dans la bande $]-\frac{1}{2d}, \frac{1}{2d}[$, les autres fréquences étant filtrées par le dispositif d'acquisition. Ils apparaissent donc comme des objets de taille d .

1.2.2 Un panorama des stratégies

Confronté à la diversité des sources d'images, l'utilisateur qui consruit un système d'archivage, d'indexation et de fouille d'images peut adopter plusieurs points de vue plus ou moins ambitieux.

Des archives séparées

Dans une première approche, le concepteur peut choisir le point de vue du fournisseur de données, pour n'indexer que les seules images qu'il produit, et à la limite du raisonnement, pour indexer de façon spécifique chaque source d'images séparément.

Dans une telle approche, on s'appuie sur les spécificités du capteur et du marché qu'il engendre pour indexer au mieux (c'est à dire avec les performances de rappel et de pertinence les meilleures) les seules images issues, par exemple, de Spot-5.

Dans une telle situation, la résolution est généralement fixée par les capacités du capteur. Le seul paramètre qui peut varier est alors celui lié à l'angle de visée qui induit des variations modestes, mais réelles entre les images. Ce problème fait l'objet de l'étude du chapitre 2.

²Nous noterons que, pour un capteur solide, cette hypothèse est pratiquement impossible à satisfaire puisque les matrices solides induisent un repliement de spectre qui ne peut être supprimé que par un filtrage par l'optique de réception et que celle ci ne permet pas un filtrage sans perte. Un satellite optique est donc soit aliasé soit redondant.

On évoluera, à terme, vers des systèmes d'indexation spécifiques à chaque type d'images (Spot, Ikonos, Pleiades, etc.) et bien sûr des systèmes de reconnaissance ou de fouille séparés. On pourrait concevoir qu'un principe unique soit partagé par ces systèmes (par exemple l'indexation par les descripteurs d'Haralick, ou la classification par SVM linéaire), mais il serait surprenant que l'on aboutisse à des choix de paramètres identiques et la "factorisation" des moyens serait somme toute réduite.

Une approche normalisée

Dans une autre approche, partant d'un tableau des satellites existants ou à venir, comme le tableau 1.1, on peut choisir quelques échelles de référence auxquelles seront extraites toutes les primitives quelle que soit la source des images.

Par exemple on pourrait concevoir que l'on prend pour référence des images "virtuelles" à pixel au sol de :

1 m, 2,5 m, 5 m, 10 m, 20 m

et bâtir un jeu de primitives pour chacune de ces échelles. On aurait bien sûr tout intérêt à construire ces primitives à partir d'un schéma auto-reproduisant de façon à permettre son extension à de nouvelles échelles si besoin est (par exemple comme le font les primitives détectées par filtrage QMF). L'apprentissage se ferait également aux divers niveaux disponibles sur les exemples d'apprentissage (par exemple sur des exemples issus de Spot 5 THR, on commencerait au niveau 2,5 m) et la recherche se ferait naturellement vers les seules images qui disposent de descripteurs à ce niveau : Ikonos, Quickbird, Spot 5, mais pas SPOT 4).

Lorsqu'un satellite à haute résolution apparaîtra, ses images seront indexées bien sûr à la résolution la meilleure de la hiérarchie, mais aussi aux résolutions dégradées, soit par calcul des primitives à basse résolution à partir des primitives à haute résolution (si c'est possible), soit par mesure directe sur l'image HR, soit enfin par reconstruction d'une image basse résolution et mesure sur cette image.

Si une image a une résolution intermédiaire entre deux résolutions de la hiérarchie de normalisation (comme CBERS 3 à 3 m - cf. tableau 1.1), se pose un intéressant problème : comment tirer profit au mieux de sa résolution sans dégrader l'image. Est-il possible d'extrapoler les descripteurs à la résolution supérieure inconnue (ici 2,5 m) ? Cette question est en partie abordée au chapitre 2.

Des questions délicates apparaissent sur le choix des résolutions de référence. Nous avons présenté ci-dessus un choix « dyadique » qui couvre de façon particulièrement économique la gamme d'intérêt et qui s'appuie sur la résolution à 10 m en raison de la grande disponibilité des images de Spot 1 à 4. Un autre choix pourrait s'appuyer sur la résolution des images des satellites à venir (comme Orfeo), puisque c'est à leur intention que l'on étudie le système. On tomberait alors sur une succession d'échelles du type :

0,75 m, 1,50 m, 3 m, 6 m, 12 m, 24 m.

Enfin on pourrait s'accorder sur des résolutions non régulières, mais "servant" au mieux l'existant et le futur, et s'appuyant donc sur les systèmes d'imagerie les plus diffusés, du genre :

0,75 m, 1,50 m, 2,5 m, 5 m, 10 m, 15m.

Il est clair que ces décisions seront difficilement prises de façon unilatérale si l'organisme qui les prend ne dispose d'une très solide avance technologique.

Une approche par invariants

Très répandue dans le domaine de la fouille de données du multimédia, l'approche par invariants consiste à extraire des primitives qui se retrouvent identiques à des échelles diverses. Compte tenu des spécificités de l'imagerie satellitaires, les seules invariances importantes sont l'invariance par rotation et l'invariance d'échelle et non l'invariance perspective qui rend le calcul d'invariants particulièrement délicat dans le domaine du multimédia.

Pour notre application, les outils de l'analyse linéaire (comme les ondelettes de Gabor ou les QMF par exemple, ainsi que les moments) permettent de construire assez aisément de telles primitives.

Ces approches doivent être considérées avec une grande prudence car elles supposent que l'objet cherché est rare et que sa manifestation sous des apparences différentes reflète la diversité des capteurs seulement. Or ce n'est pas toujours le cas : des vergers d'essence différentes (par exemple des pêchers et des noyers) peuvent avoir des aspects très semblables, l'un avec un pixel d'un mètre, l'autre avec un pixel de trois mètres. Les primitives invariante perdront ces nuances.

Une approche « orientée objet »

Dans cette approche, on oublie les sources d'images existantes et l'on s'intéresse exclusivement aux scènes observées. Elles appartiennent bien sûr à des paysages très variés : zones vierges (forêts, savanes, déserts, toundra, ..), zones agricoles (vergers, parcellaires, cultures extensives, ...), zones urbanisées (centre ville, villages, banlieues, zones résidentielles) ou aménagées (usines, centres commerciaux, serres, chantiers, ...). chaque type de paysage se caractérise par des échelles et des résolutions spécifiques, mais également par des primitives pertinentes. C'est à ces échelles, si elles sont présentes dans l'image, que seront faites les mesures de primitives sur toute image qui se présente. Les étapes d'apprentissage, ainsi que la fouille reprendront également ces échelles et l'on sélectionnera les seules images qui offrent une résolution suffisante pour la tâche attendue.

On sent que la détermination d'échelles significatives pour un paysage donné est une démarche très proche d'une indexation sémantique puisqu'elle associe des mesures à des classes et probablement de façon implicite des classes à des mesures.

Très satisfaisante pour l'esprit, cette approche pose quand même un certain nombre de problèmes assez difficilement solubles en l'état de nos connaissances.

1. Existe t'il aujourd'hui un "inventaire" incontesté des paysages d'intérêt ? Cet inventaire est-il susceptible de satisfaire des communautés très différentes concernées par : la défense, l'aménagement du territoire, la prévention des risques, l'agriculture, la prospection géologique ou forestière, etc. ? Imagine t'on aujourd'hui que nos classifications satisferont également les utilisateurs du futur ?
2. Les objets sont-ils eux-mêmes liés à quelques échelles significatives qui s'imposent ? Ces échelles ne dépendent-elles pas de l'espace et du temps ?

3. est-on capable d'identifier, lors de l'indexation de l'image, les divers paysages rencontrés de façon à mesurer les bonnes primitives en chaque point ?

Il semble important de pouvoir explorer un peu cette piste et nous l'avons fait de deux façons :

1. en démarrant une étude sur la répartition des paysages dans les imageries satellitaires. Ces travaux sont en cours, ils ont déjà vu la collecte de plus de 4 000 échantillons de paysages significatifs et par ailleurs ils ont permis de construire une base de données pour recevoir et manipuler ces échantillons (et les concepts sous-jacents) (stage de J. Siyar). Ces travaux seront présentés ultérieurement.
2. en essayant de déterminer quelles sont les échelles les plus significatives pour un paysage donné. Ces travaux sont présentés aux chapitres 3 et 4.

Les travaux identifiés au point 1 ci-dessus sont de longue haleine. Ils devraient s'appuyer sur plusieurs types de travaux :

- sur une bonne connaissance de l'espace multidimensionnel des primitives extraites, et en particulier de leur distribution (clusters, hiérarchies, etc.) ;
- sur l'inventaire des méthodologies de classification des paysages selon les divers domaines thématiques ;

ils pourraient aboutir, à terme à des propositions raisonnées de représentation des paysages tels qu'ils se manifestent en télédétection, et éventuellement à leur structuration en ontologie.

Chapitre 2

Robustesse des primitives au changement d'échelle

Dans ce chapitre, nous allons essayer de répondre à une question précise : considérant un ensemble homogène d'images (en l'occurrence celles de Spot Panchro à 5 m de résolution) considérant une procédure de recherche fixée à une résolution nominale (comportant l'indexation, l'apprentissage puis la fouille), peut-on plonger des images à une résolution différente, sans aucune précaution, et obtenir des performances de rappel satisfaisantes ? On voit que l'on répondra ici, en particulier, à 2 interrogations que nous nous sommes posées au chapitre précédent :

1. faut-il tenir compte de la résolution exacte des images lorsque le satellite a une visée légèrement oblique, ou lorsqu'il change d'altitude ?
2. dans le cas où l'on choisit des échelles de référence pour indexer les images, peut-on appliquer les procédures définies pour une résolution r de référence à un satellite de résolution $r + \epsilon$?

2.1 Protocole expérimental

On cherche donc à classifier les images satellitaires de différentes résolutions à partir d'une base de données d'apprentissage construite sur une seule résolution. La classification est basée sur la méthode SVM [29] à l'aide des caractéristiques extraites sur les images satellitaires. De plus on se concentre sur l'indexation par les textures telles que définies dans nos études antérieures [7]. Enfin on s'attachera à des classes simples : nuage, mer, désert, forêt, ville et champs.

Il existe de nombreuses méthodes d'extraction des caractéristiques de texture, méthodes basées sur le filtrage [11] [20] [21] [19], méthodes basées sur les informations statistiques de l'image [11] [15], etc. Dans nos travaux antérieurs, nous avons sélectionné un certain nombre de méthodes particulièrement efficaces et nous avons également défini les paramètres qui convenaient. Pour notre objectif, il est intéressant de vérifier si les caractéristiques extraites à l'aide de ces méthodes sont robustes au changement de résolution. On étudie donc ici la robustesse des caractéristiques que l'on a extraites à l'aide des méthodes de GABOR, de HARALICK et de QMF par rapport au changement de résolution.

Le schéma de travail est le suivant : disposant d'images d'apprentissage à la résolution nominale (dans notre cas, $5m/pixel$), on extrait, à cette résolution, les primitives les plus adaptées à l'indexation (comme nous l'avons fait dans nos travaux antérieurs), puis on entraîne les classifieurs (SVM) à partir de ces primitives. Par ailleurs, disposant d'images à des résolutions différentes (allant de $2.5m/pixel$ à $9.3m/pixel$), on les indexe à l'aide de primitives obtenues par les mêmes algorithmes (sans tenir compte donc de la différence de résolution). On utilise alors les classifieurs à résolution nominale pour classifier ces images à résolution différente, et on évalue les performances des classifications.

2.2 Extraction des caractéristiques

Dans cette section, trois méthodes d'extraction de caractéristiques sont introduites : GABOR, QMF et HARALICK.

La méthode de GABOR est basée sur les filtres de Gabor. Ce sont effectivement des fonctions gaussiennes centrées à diverses fréquences et de différentes largeurs. On peut partitionner le domaine fréquentiel à l'aide de ces filtres en réglant les centres et les largeurs. On utilise les moyennes et les variances des images ainsi filtrées comme caractéristiques. Les détails de cette méthode peuvent être trouvés dans [7].

La méthode de HARALICK est basée sur la matrice de cooccurrence. Celles-ci comptent le nombre de paires de pixels séparés d'un certain déplacement fixé. Chaque élément dans cette matrice représente le nombre de paires de pixels dont les niveaux de gris correspondent à la position de cet élément dans la matrice. Ensuite on calcule des propriétés statistiques de ces matrices, e.g. le contraste, le moment angulaire, etc., que l'on prend comme caractéristiques. On peut trouver les détails dans [7].

La méthode QMF procède par une décomposition en ondelettes de l'image. On calcule les moyennes et les variances des coefficients d'ondelettes qui sont utilisées comme caractéristiques, cf. [7]

2.3 Expériences

2.3.1 La base de données de texture

Afin de procéder de façon très semblable à nos travaux antérieurs, nous avons extrait 6 classes de textures des images SPOT 5 originales (à la résolution de $5m/pixel$). Les 6 classes sont : *nuage*, *mer*, *désert*, *ville*, *champs* et *forêt*. On note c_0 la classe *nuage*, c_1 la classe *mer*, etc. Pour chaque classe, il y a 100 vignettes, i.e. des petites images de taille 64×64 pixels. On note alors I_{100k+n} , $1 \leq n \leq 100$ la n ème vignette de la classe c_k de résolution $5m/pixel$. Par exemple, I_{28} correspond à la 28ème vignette dans la classe *nuage*, et I_{245} correspond à la 45ème vignette de la classe *désert*. Un exemple des vignettes de SPOT5 est montré dans la figure 2.3.2.



FIG. 2.1 – Une image de SPOT5 à la résolution de $5m$ sur les champs ©CNES

2.3.2 Les vignettes de résolutions différentes

Pour tester la robustesse des caractéristiques sur l'échelle, on va classifier les vignettes de différentes résolutions par les classifieurs qui sont obtenus à partir des vignettes d'une certaine résolution fixée. Dans notre cas, les vignettes d'apprentissage sont toujours à la résolution de $5m/pixel$, et les vignettes à classifier ont des résolutions variant de $2.5m/pixel$ jusqu'à $9.3m/pixel$. Au sein de cet intervalle, on utilise 28 résolutions différentes.

Pour chaque échantillon, la résolution est calculée par :

$$s_i = \frac{5}{\text{round}[64/(32 \times 1.158^i)]} m/pixel$$

où i prend les valeurs $0, 1, 2, \dots, 9$ et où la fonction $\text{round}(r) : R \rightarrow Z$ est égale à l'entier le plus proche de r .

Malheureusement, nous ne disposons que d'une très petite séquence d'images qui corresponde à toutes les résolutions s_i sur un même site (la séquence sur Salon de Provence). Elle ne contient pas une variété suffisante de chacun des paysages que nous souhaitons distinguer. Nous ne pouvons donc pas conduire l'expérience en utilisant des images réelles à chaque étape.

Nous procédons donc à partir de la base de données que nous avons constituée à l'échelle nominale de $5m$, et nous créons les images aux échelles différentes par filtrage ou par interpolation.

Les images filtrées (à résolution plus faible) sont obtenues par un filtre passe-bas parfait à partir de la scène originale. Les images ainsi obtenues de 64×64 pixels couvrent donc un champ plus large que l'imagette nominale correspondante.

Pour les résolutions plus hautes, on utilise plusieurs méthodes d'interpolation : le plus proche voisin, l'interpolation bilinéaire et les interpolations par splines d'ordre 2, 3 et 4, de façon à étudier l'influence de l'interpolation dans l'indexation.

On notera m_0 la méthode par plus proche voisin, m_1 la méthode bilinéaire. m_2 , m_3 et m_4 correspondent respectivement aux méthodes de spline d'ordre 2, 3 et 4.

On peut donc noter $I'_{i,j,100k+n}$ la vignette à la résolution s_i obtenue par la méthode d'interpolation m_j à partir de la vignette I_{100k+n} , e.g. la vignette $I'_{0,1,245}$ est la vignette à la résolution de $2.5m/pixel$ obtenue par interpolation bilinéaire à partir de la vignette I_{245} .

Des exemples des vignettes aux résolutions de $2.5m$ et $9.3m$ sont montrés respectivement dans les figures 2.2 et 2.3.

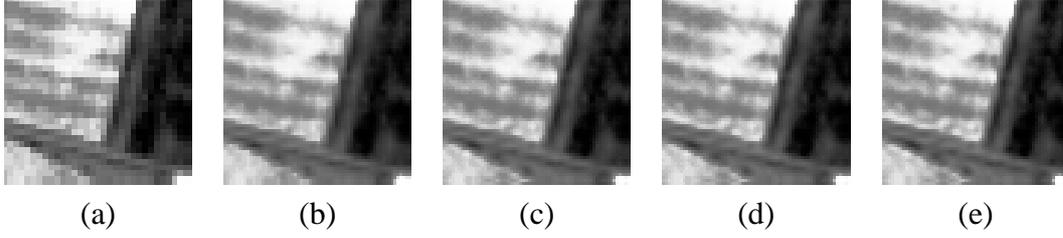


FIG. 2.2 – Les images à la résolution de $2.5m$ obtenues à partir de l'image de résolution $5m$ à l'aide de l'interpolation par le plus proche voisin (a), bilinéaire (b), la spline d'ordre 2 (c), 3 (d) et 4 (e)

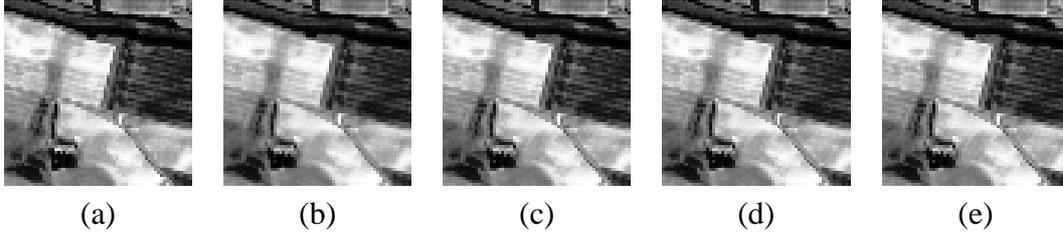


FIG. 2.3 – Les images à la résolution de $9.3m$ obtenues à partir de l'image de résolution $5m$ à l'aide de l'interpolation par le plus proche voisin (a), bilinéaire (b), la spline d'ordre 2(c), 3(d) et 4(e)

2.3.3 La validation croisée

Pour éviter l'effet du choix des vignettes d'apprentissage, lors de la validation, on procède par tirage aléatoire et on fait 5 expériences. Chaque fois, on utilise 80 vignettes d'apprentissage et 20 vignettes pour le test. Plus précisément, supposons que l'ensemble $N' = \{n'_l, 1 \leq l \leq 100 \quad l \in \mathbb{Z}\}$ contient tous les nombres entiers entre 1 et 100 rangés dans un ordre aléatoire. On construit 5 ensembles d'apprentissage :

$$TR_1 = \{I_{100k+n'_l}, \quad 1 \leq l \leq 80 \quad \text{et} \quad 0 \leq k \leq 5\}$$

$$TR_2 = \{I_{100k+n'_l}, \quad 21 \leq l \leq 100 \quad \text{et} \quad 0 \leq k \leq 5\}$$

$$TR_3 = \{I_{100k+n'_l}, \quad 41 \leq l \leq 100 \quad \text{ou} \quad 1 \leq l \leq 20 \quad \text{et} \quad 0 \leq k \leq 5\}$$

$$TR_4 = \{I_{100k+n'_l}, \quad 61 \leq l \leq 100 \quad \text{ou} \quad 1 \leq l \leq 40 \quad \text{et} \quad 0 \leq k \leq 5\}$$

$$TR_5 = \{I_{100k+n'_l}, \quad 81 \leq l \leq 100 \quad \text{ou} \quad 1 \leq l \leq 60 \quad \text{et} \quad 0 \leq k \leq 5\}$$

De plus pour la résolution s_i et la méthode d'interpolation m_j , on forme 5 ensembles de test :

$$TE_{i,j,1} = \{I'_{i,j,100k+n'_l}, \quad 1 \leq l \leq 20 \quad \text{et} \quad 0 \leq k \leq 5\}$$

$$TE_{i,j,2} = \{I'_{i,j,100k+n'_l}, \quad 21 \leq l \leq 40 \quad \text{et} \quad 0 \leq k \leq 5\}$$

$$TE_{i,j,3} = \{I'_{i,j,100k+n'_l}, \quad 41 \leq l \leq 60 \quad \text{et} \quad 0 \leq k \leq 5\}$$

$$TE_{i,j,4} = \{I'_{i,j,100k+n'_l}, \quad 61 \leq l \leq 80 \quad \text{et} \quad 0 \leq k \leq 5\}$$

$$TE_{i,j,5} = \{I'_{i,j,100k+n'_l}, \quad 81 \leq l \leq 100 \quad \text{et} \quad 0 \leq k \leq 5\}$$

On classe les images dans $TE_{i,j,p}$ par le classifieur entraîné par les images dans TR_p , où $1 \leq p \leq 5$ et $p \in Z$. Enfin on calcule la moyenne $me_{i,j}$ d'erreur de classification pour la résolution s_i et la méthode d'interpolation m_j .

2.4 Résultats

2.4.1 Robustesse des caractéristiques extraites par les méthodes GABOR, HARALICK et QMF

Dans cette section, les résultats de la classification sont présentés dans les figures 2.4, 2.5 et 2.6. Pour la méthode de GABOR, on utilise toujours 3 échelles et 4 orientations, donc la dimension de caractéristiques est $3 \times 4 \times 2 = 24$. Pour HARALICK, on utilise 13 caractéristiques sur 4 directions et on calcule la moyenne et la variance de ces 13 caractéristiques, donc il y a au total $(2 + 4) \times 13 = 78$ caractéristiques pour chaque image. La distance de HARALICK est toujours 3 pixels. Pour QMF, on utilise 3 échelles, donc il y a $3 \times 3 \times 2 = 18$ caractéristiques. De plus, les classifications sont réalisées à l'aide de la machine à vecteurs de support (SVM).

Dans ces figures, on peut constater que :

1. Quelle que soit la méthode d'extraction des caractéristiques et pour les méthodes d'interpolation bilinéaire ou au plus proche voisin, quand la résolution est proche de $5m/pixel$, on peut toujours obtenir un résultat relativement bon. Par contre quand la résolution s'éloigne de $5m/pixel$, le résultat se dégrade un peu, modérément pour la méthode HARALICK et GABOR, plus nettement pour QMF.

2. La méthode de HARALICK est la plus robuste non seulement par rapport à la résolution mais aussi par rapport à la méthode d'interpolation. Par contre, pour les deux autres méthodes, quand on utilise les interpolations splines, les résultats se dégradent considérablement. La raison est que ces deux méthodes sont basées sur le filtrage, et l'interpolation spline est également aussi un filtrage passe-bas, il peut diminuer beaucoup les irrégularités des images. Comme on utilise les moyennes et les variances comme caractéristiques pour la méthode de GABOR et la méthode QMF, l'interpolation spline a donc une grande influence sur les caractéristiques.

2.4.2 Résultats sur les caractéristiques sélectionnées

On a procédé au paragraphe précédent en conservant toutes les primitives pour indexer et classifier les images, mais on sait (cf. notre précédent rapport [7]) que l'on peut considérablement réduire le nombre de caractéristiques utilisées par une sélection lors de l'apprentissage. Est-ce que l'on peut encore utiliser un petit nombre seulement de primitives si l'on ne travaille plus à la résolution nominale ?

Comme on a vu ci-dessus que la méthode de HARALICK est la plus robuste par rapport au changement de résolution, on va ici s'intéresser plus précisément à cette méthode et à la robustesse des caractéristiques de HARALICK. La question peut être reformulée de la façon suivante :

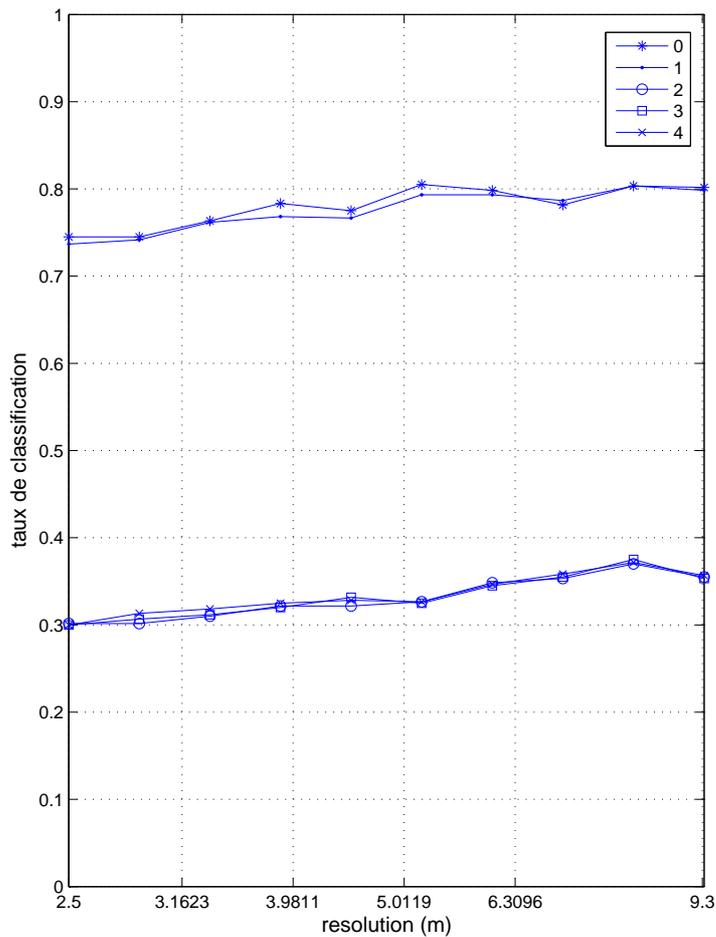


FIG. 2.4 – Résultats des classifications à partir de 24 caractéristiques de la méthode de GABOR par rapport aux différentes résolutions. L'axe horizontal est la résolution et l'axe vertical représente les taux de classification. Légende : 0 - interpolation au plus proche voisin, 1 - interpolation bilinéaire, 2, 3 et 4 - interpolation spline d'ordre 2, 3 et 4. On note que les interpolations splines associées au classifieur par GABOR sont inacceptables.

Si l'on utilise A caractéristiques de HARALICK parmi les 78, est-ce que les résultats restent toujours comme ceux montrés dans la figure (2.5) ?

On utilise trois méthodes qui sont présentées dans [7] pour sélectionner 15 parmi les 78 caractéristiques :

1. la méthode de FISHER (qui est basée sur l'analyse discriminante linéaire de Fisher),
2. la méthode de RFE (qui est basée sur l'estimation de poids relatifs à l'optimisation d'un problème de discrimination linéaire qui peut être résolu à l'aide d'SVM),
3. la méthode de SVC_FS (qui est basée sur le *Support Vector Clustering*).

Les figures 2.7, 2.8 et 2.9 nous confirment que les caractéristiques sélectionnées sont aussi

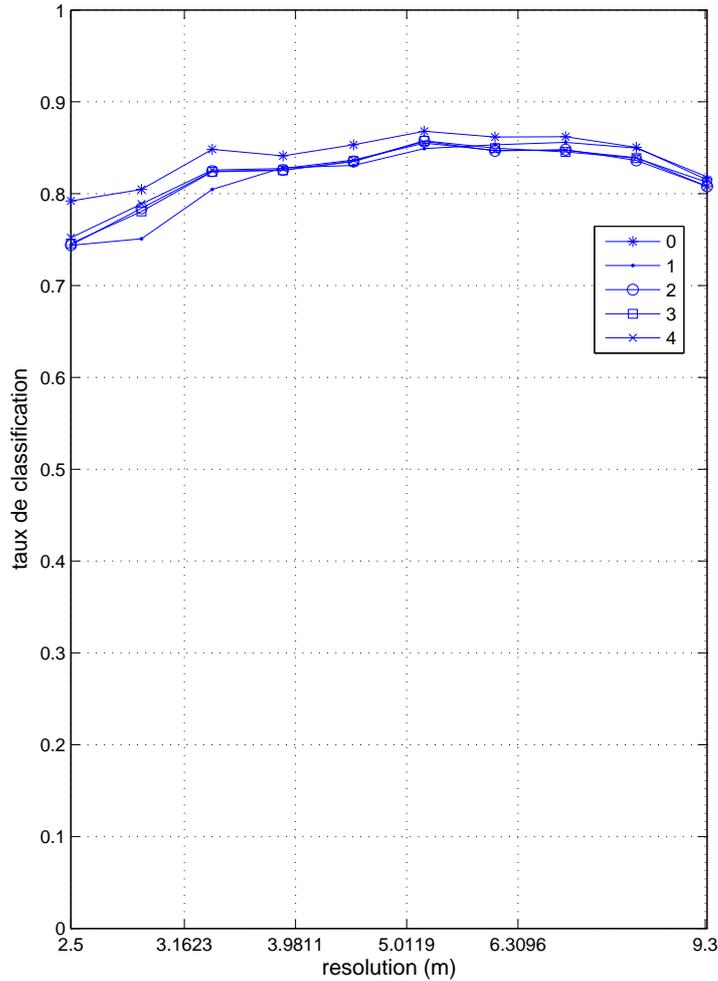


FIG. 2.5 – Résultats des classifications à partir de 78 caractéristiques de la méthode de HARALICK par rapport aux différentes résolutions. Légende : 0 - interpolation au plus proche voisin, 1 - interpolation bilinéaire, 2, 3 et 4 - interpolation spline d'ordre 2, 3 et 4

robustes par rapport au changement de résolution.

2.5 Conclusion

On a démontré ici la robustesse des caractéristiques de GABOR, HARALICK et QMF par rapport au changement de résolution et on a montré que, surtout lorsque les données sont issues d'un filtrage bilinéaire ou au plus proche voisin, les résultats de classification restent à peu près constants, et la dégradation des résultats n'est pas radicale quand la résolution s'éloigne de la valeur nominale de $5m/pixel$. On a confirmé que la méthode HARALICK est la plus robuste. On a montré que ses performances restaient pratiquement constantes

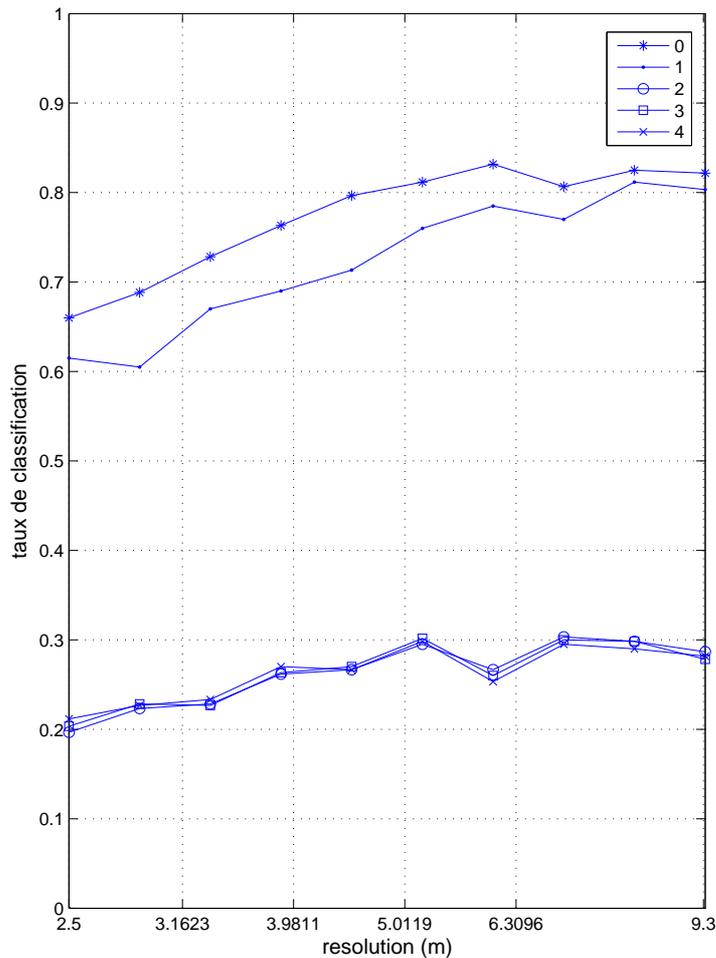


FIG. 2.6 – Résultats des classifications à partir de 18 caractéristiques de la méthode de QMF par rapport aux différentes résolutions. Légende : 0 - interpolation au plus proche voisin, 1 - interpolation bilinéaire, 2, 3 et 4 - interpolation spline d'ordre 2, 3 et 4. On note que les interpolations splines associées au classifieur par QMF sont inacceptables.

même après sélection des caractéristiques.

Nous pouvons ainsi utiliser les algorithmes d'apprentissage sur une gamme de résolution relativement large allant de la résolution moitié à la résolution double, sans modifier les traitements d'extraction des primitives. Si l'on se concentre dans cette gamme $[\frac{1}{2}, 2]$, on peut imaginer deux stratégies pour étendre la méthode :

1. tout d'abord on peut imaginer déduire des primitives mesurées à l'échelle s_i un jeu de primitives valable pour l'échelle s_0 ;
2. ensuite on peut envisager de modifier les conditions de l'apprentissage fait à l'échelle s_0 de façon qu'il permette une reconnaissance à l'échelle s_i .

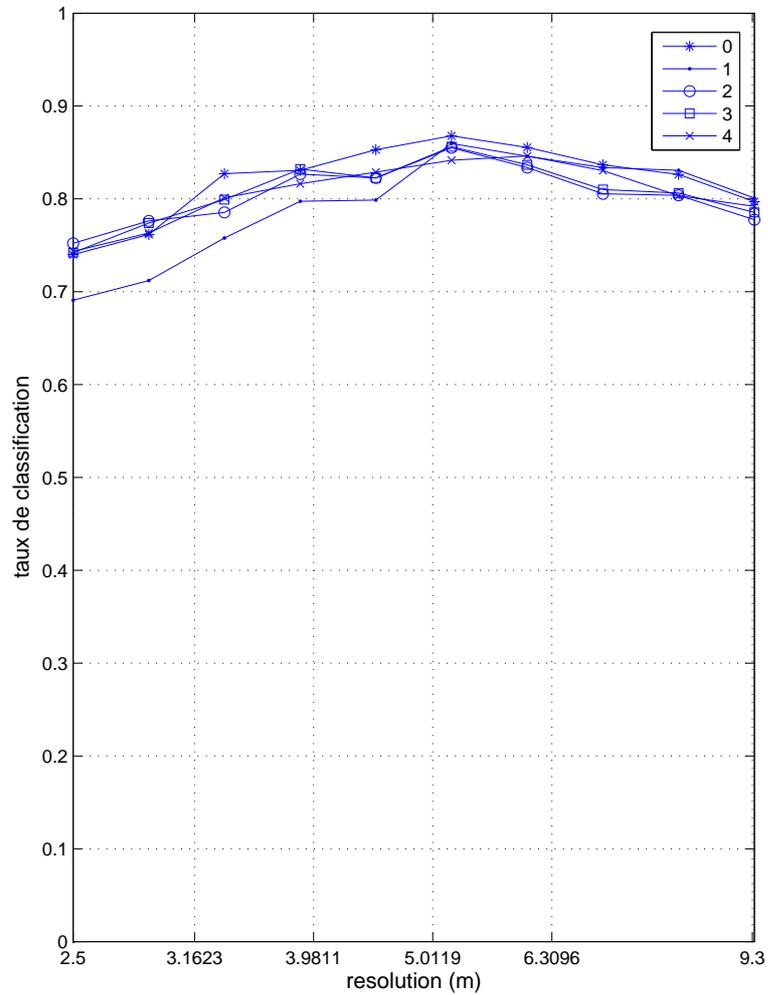


FIG. 2.7 – Résultats des classifications de 15 caractéristiques de HARALICK sélectionnées par la méthode FISHER par rapport aux différentes résolutions. Légende : 0 - interpolation au plus proche voisin, 1 - interpolation bilinéaire, 2, 3 et 4 - interpolation spline d'ordre 2, 3 et 4

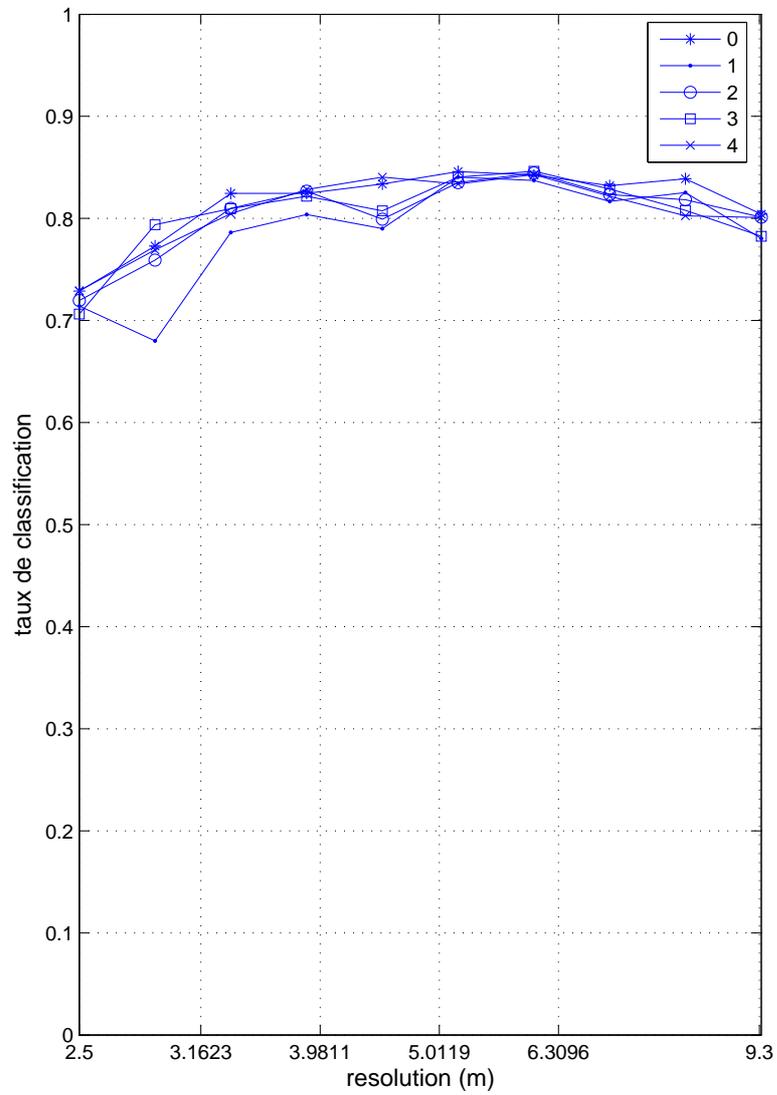


FIG. 2.8 – Résultats des classifications de 15 caractéristiques de HARALICK sélectionnées par la méthode RFE par rapport aux différentes résolutions. Légende : 0 - interpolation au plus proche voisin, 1 - interpolation bilinéaire, 2, 3 et 4 - interpolation spline d'ordre 2, 3 et 4

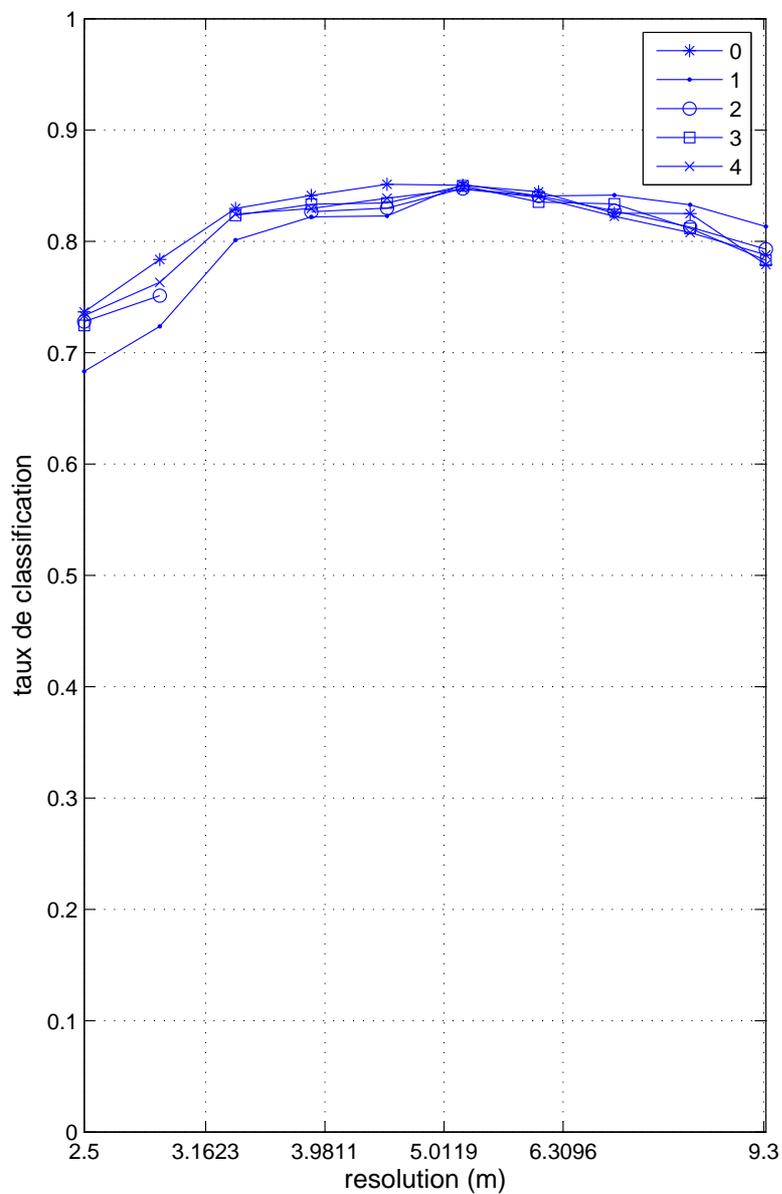


FIG. 2.9 – Les résultats des classifications de 15 caractéristiques de HARALICK sélectionnées par la méthode SVC_FS par rapport aux différentes résolutions. Légende : 0 - interpolation au plus proche voisin, 1 - interpolation bilinéaire, 2, 3 et 4 - interpolation spline d'ordre 2, 3 et 4

Chapitre 3

Espace multi-échelle linéaire

Soit : $I : \mathbb{R}^D \rightarrow \mathbb{R}$ ($D=2$ pour les images).

La représentation de l'espace échelle linéaire est définie comme la solution de l'équation de chaleur [18] :

$$\partial_t L = \frac{1}{2} \Delta L = \frac{1}{2} \sum_{i=1}^D \partial_{x_i x_i} L \quad (3.1)$$

avec la condition initiale $L(\cdot; 0) = I(\cdot)$.

La solution de (3.1) est [18] :

$$L(\cdot; t) = g(\cdot; t) \star I(\cdot) \quad (3.2)$$

où l'opérateur \star représente la convolution. et $g : \mathbb{R}^D \times \mathbb{R}_+ \rightarrow \mathbb{R}$ est un noyau gaussien d'écart-type t , i.e. :

$$g(x; t) = \frac{1}{\sqrt{2\pi t}} e^{-\sum_{i=1}^D x_i^2 / (2t^2)}$$

Remarquons que la convolution avec une gaussienne est la seule représentation d'espace multi-échelle linéaire [17]. Il est cependant possible d'utiliser d'autres représentations multi-échelles. Ce sont alors des représentations non linéaires. Parmi les constructions non linéaires, citons notamment celles basées sur des opérateurs mathématiques morphologiques ou sur le flot de la variation totale (TV) [10, 3, 6].

3.1 Mesure d'échelle dans l'espace multi-échelle linéaire

Dans cette section et les deux suivantes, nous nous intéressons à la mesure d'échelle basée sur la représentation d'espace multi-échelles linéaire.

L'idée principale de la mesure d'échelle est que l'on représente une image $I(x, y)$ dans l'espace multi-échelle linéaire. Pour chaque échelle t , on a une image $I(x, y; t)$. Pour chaque $I(x, y; t)$, on associe une mesure $F(I_t)$ (le maximum local ou la variation totale par exemple) comme la mesure de l'échelle.

De plus il nous faut distinguer la notion de résolution et celle d'échelle. La résolution est la période spatiale minimale. Autrement dit, c'est la zone couverte par un pixel de l'image. Pour la notion d'échelle, elle se définit plutôt pour un objet. On appelle échelle d'un objet

sa taille significative. Par exemple un bâtiment de largeur 30 mètres, dans une image de résolution de $5m/pixel$, a une échelle de 6 pixels [30]. Remarquons que la dernière notion est à la fois une notion locale et globale, i.e., on peut mesurer localement l'échelle sur chaque pixel, et on peut aussi associer une seule échelle caractéristique à une image de façon globale.

On commence par présenter la mesure d'échelle plutôt locale introduite dans [18] (qui est la mesure la plus utilisée). L'auteur a proposé d'utiliser le maximum local d'une dérivée normalisée par rapport aux échelles. Nous illustrons cette idée sur un exemple simple :

Supposons que l'on ait $I(x) = \sin(\omega_0 x)$. Cela implique

$$L(x; t) = e^{-\omega_0^2 t/2} \sin(\omega_0 x)$$

On peut voir que le maximum de la dérivée de $L(x; t)$ par rapport à x est

$$L_{x, \max(t)}(t) = \max \frac{\partial(L(x; t))}{\partial x} = \omega_0 e^{-\omega_0^2 t/2}$$

Evidemment, $L_{x, \max(t)}(t)$ est une fonction décroissante de l'échelle t . On introduit donc un opérateur dérivé normalisé défini sous la forme

$$\partial_{\sigma, norm} = t^{1/2} \partial_x$$

où $\sigma = x/t^{1/2}$. Le maximum par rapport à x de la dérivée normalisée de $L(x; t)$ est alors

$$L_{\sigma, \max}(t) = t^{1/2} \omega_0^2 e^{-\omega_0^2 t/2}$$

et ce maximum est atteint lorsque $t = \omega_0^{-2}$. Cela signifie que la position du maximum ne dépend que de la fréquence de $f(x)$.

A partir de cette idée, on peut mesurer l'échelle associée à chaque pixel en calculant le $L_{\sigma, \max}(t)$ correspondant. Malheureusement, pour calculer l'échelle d'une région entière, on voit que la complexité de calcul est considérable.

C'est pourquoi nous sommes amenés à introduire d'autres méthodes pour calculer l'échelle d'une région, méthodes basées sur la mesure de la variation totale d'une image [28].

3.2 Mesure d'échelle basée sur la variation totale

3.2.1 Rappel : espace BV des fonctions à variations bornées

On rappelle ici la définition de l'espace $BV(\Omega)$ (on suppose que Ω , le domaine de l'image, est un ouvert borné Lipschitz de \mathbb{R}^2) :

Définition : $BV(\Omega)$ est le sous-espace des fonctions $I \in L^1(\Omega)$ telle que la quantité suivante soit finie :

$$TV(I) = \sup \left\{ \int_{\Omega} I(x) \operatorname{div}(\xi(x)) dx / \xi \in C_c^1(\Omega; \mathbb{R}^2), \|\xi\|_{L^\infty(\Omega)} \leq 1 \right\} \quad (3.3)$$

$BV(\Omega)$ muni de la norme

$$\|I\|_{BV(\Omega)} = \|I\|_{L^1(\Omega)} + TV(I) \quad (3.4)$$

est un espace de Banach.

Pour plus de détails sur $BV(\Omega)$, on renvoie le lecteur à [2, 12].

Remarque : Si I est C^1 , alors $TV(I) = \int_{\Omega} |\nabla I|$.

Interprétation : L'intérêt de l'espace BV en traitement d'image vient essentiellement du fait qu'il autorise les discontinuités le long de courbes (plus généralement, en dimension N , le long de surfaces de codimension $N - 1$). En particulier, il autorise la présence de bords. C'est un espace qui est particulièrement bien adapté pour les images géométriques [23, 4] (même si par-contre on sait que BV ne convient pas pour modéliser certaines images naturelles [1]). Le succès de BV en traitement d'image est étroitement lié à celui du modèle de Rudin-Osher-Fatemi [28] en restauration d'images.

3.2.2 Principe

L'idée de cette mesure d'échelle est que, si on part d'une image où les objets ont une grande échelle, la variation totale normalisée par la taille de l'image est relativement petite. Ensuite si on réduit la taille de l'image (c'est à dire si on fait un sous-échantillonnage après convolution par une gaussienne), les objets sont de plus en plus petits, et la variation totale (TV) normalisée par la taille de l'image augmente. Mais après une certaine échelle, comme la convolution par une gaussienne mélange de plus en plus des objets, la TV normalisée va se mettre à décroître. Il est donc possible d'obtenir une échelle critique correspondant au maximum de la variation totale normalisée. C'est cette échelle critique que nous définissons comme étant l'échelle des objets dans l'image.

3.2.3 Expériences et résultats

L'image originale $I(x, y)$ est tout d'abord convolée avec un noyau gaussien d'écart-type t : on obtient l'image $I(x, y; t)$. Ensuite on fait le sous-échantillonnage d'un facteur t sur $I(x, y; t)$. On calcule la TV sur l'image sous-échantillonnée et on la normalise par la taille de l'image sous-échantillonnée. Nous obtenons ainsi le graphe de la variation totale (TV) en fonction de l'échelle t . La position des maxima nous donne les différentes échelles globales présentes dans l'image.

Résultats sur les images synthétiques

Tout d'abord, on montre les résultats sur les images synthétiques pour vérifier la validité de la mesure de l'échelle.

Dans la figure 3.1(a), on a des carrés de largeur 25 pixels. De plus, la période spatiale entre les carrés est de 50 pixels. L'image originale est de taille 1024×1024 et on la réduit jusqu'à une taille de 16×16 . La figure 3.1(b) montre la mesure d'échelle de cette image. On peut voir que pour l'échelle $t = 11$ on obtient le maximum global de la variation totale : il s'agit de l'échelle de l'image.

Dans la figure 3.2(a), on a des rectangles de taille 13×45 pixels. L'image originale est de taille 1024×1024 et on la réduit jusqu'à une taille de 16×16 . La figure 3.2(b) montre la mesure d'échelle de cette image. On peut voir que pour l'échelle $t = 9$ on a le maximum global qui correspond à l'échelle de cette image.

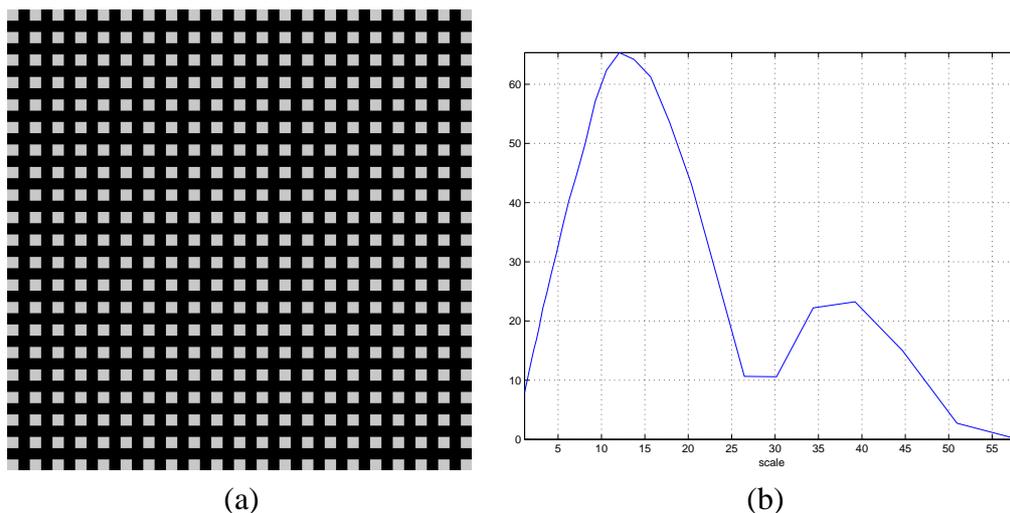


FIG. 3.1 – (a) Image synthétique des carrés de largeur 25 pixels ; (b) mesure d'échelle basée sur la TV. Notons que quand $t = 11$ la TV atteint son maximum.

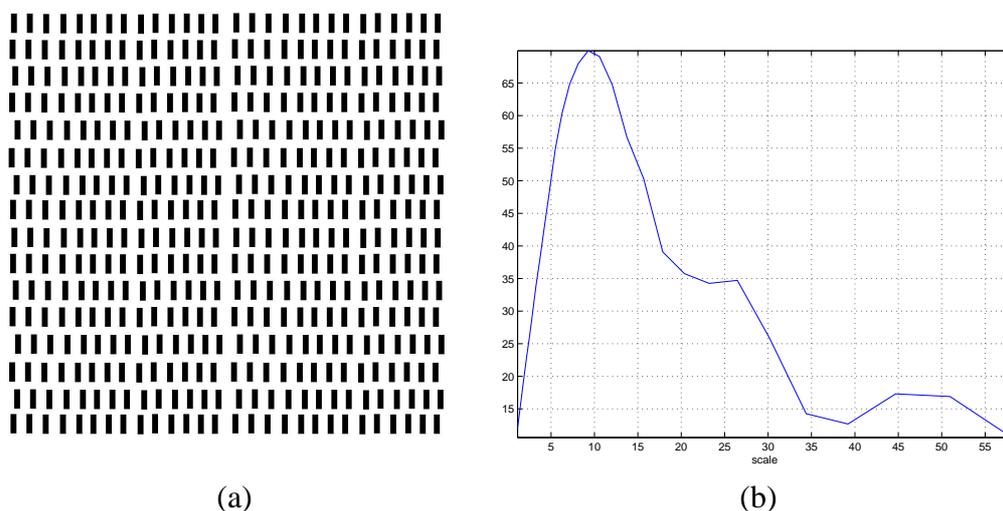


FIG. 3.2 – (a) Image synthétique des rectangles de taille 13×45 pixels ; (b) mesure d'échelle basée sur la TV. Notons que quand $t = 9$ la TV atteint son maximum.

Dans la figure 3.3(a), on a des cercles de rayon 18 pixels. L'image originale est de taille 512×512 et on la réduit jusqu'à une taille de 16×16 . La figure 3.3(b) montre la mesure d'échelle de cette image. On peut voir que pour l'échelle $t = 9$ on a le maximum global qui correspond à l'échelle de cette image.

Résultats sur les images réelles

Dans cette section, les résultats de la mesure par la TV sur les images aériennes sont présentés. On a mesuré les échelles pour trois classes de terrains, i.e. la forêt, la ville et les champs. Et pour la ville, on montre les résultats sur trois images différentes (car les tailles des bâtiments varient considérablement selon la ville considérée).

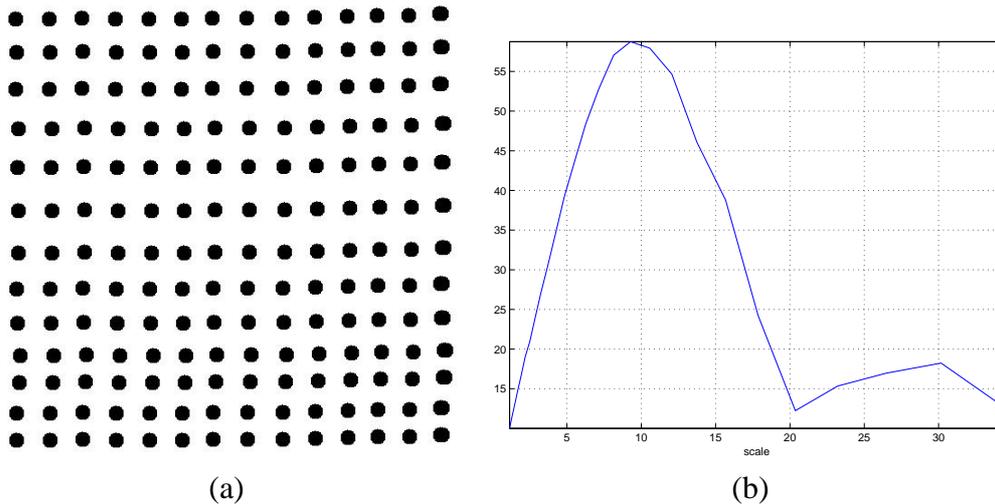


FIG. 3.3 – (a) Image synthétique des cercles de rayon 18 pixels ; (b) mesure d'échelle basée sur la TV. Notons que quand $t = 9$ la TV atteint son maximum.

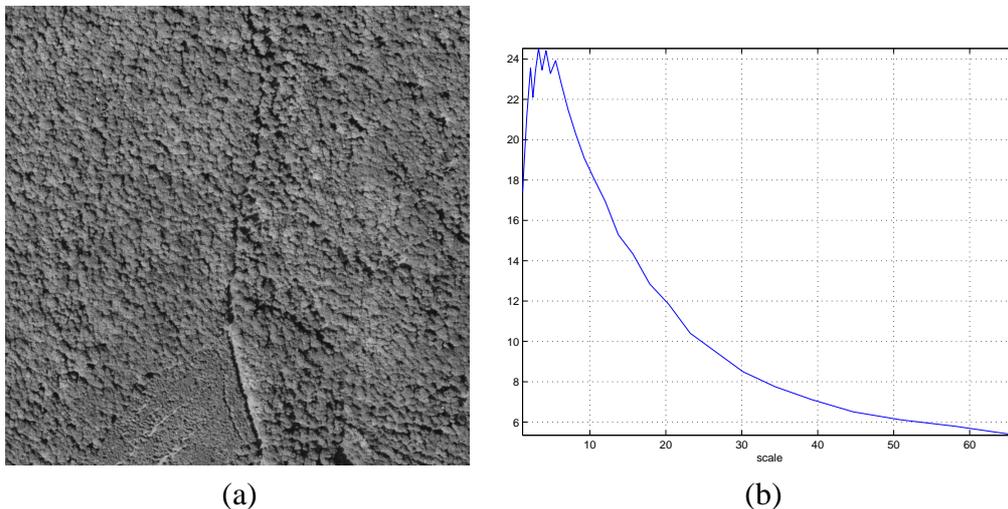


FIG. 3.4 – (a) Image aérienne sur la forêt ©CNES ; (b) mesure d'échelle basée sur la TV. Notons que quand $t \in [3, 5]$, il y a plusieurs pics voisins.

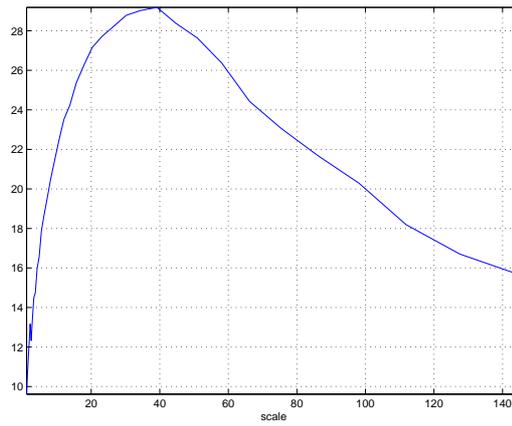
La figure 3.4(a) est une image de forêt dont la taille originale est 1024×1024 . La figure 3.4(b) montre le résultat de la mesure d'échelle. On peut constater que quand l'échelle $t \in [3, 5]$, on obtient plusieurs pics.

Les figures suivantes (3.5(a), 3.6(a), et 3.7(a)) sont des images aériennes respectivement sur Paris, Lyon et Tunis. Les tailles des images originales sont toujours 2048×2048 et on les réduit jusqu'à 16×16 . Les tailles des bâtiments de ces 3 villes sont complètement différentes. On peut voir que sur Paris, la courbe de la TV atteint son maximum pour $t = 40$, par contre pour les deux autres villes, les positions des pics sont toujours autour de $t = 20$. Et cela correspond aux différences de tailles des bâtiments entre les trois villes (les bâtiments sont plus gros sur l'image de Paris).

Enfin on a une image de champs (figure 3.8(a)). La taille de l'image originale est $1024 \times$



(a)

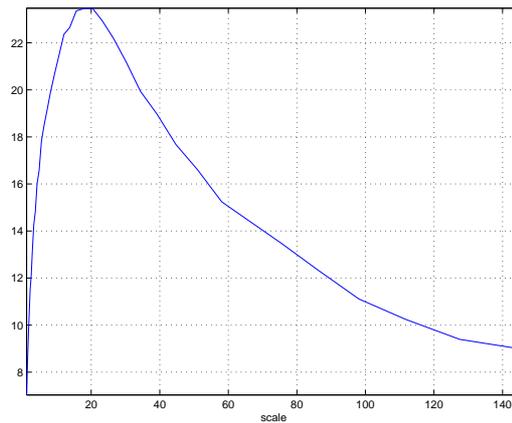


(b)

FIG. 3.5 – (a) Image aérienne sur la ville de Paris ©CNES ; (b) mesure d'échelle basée sur la TV. Notons que quand $t = 40$, la TV atteint son maximum.



(a)



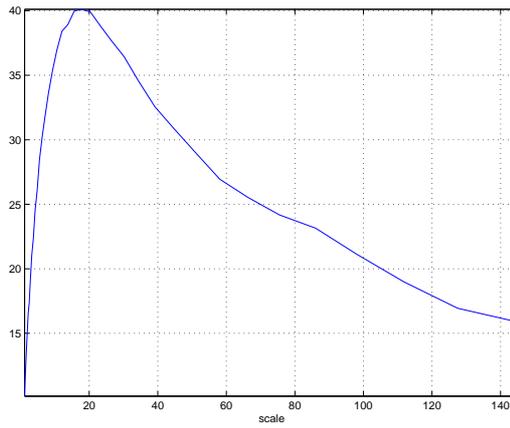
(b)

FIG. 3.6 – (a) Image aérienne sur la ville de Lyon ©CNES ; (b) mesure d'échelle basée sur la TV. Notons que quand $t = 20$, la TV atteint son maximum.

1024 et on la réduit jusqu'à 16×16 . Il n'y a pas de maximum significatif sur la courbe de la TV par rapport à l'échelle, comme illustré sur la figure 3.8(b) (bien qu'il y ait un maximum global, mais comme il n'y a pas une tendance de décroissance, la maximum global de la courbe ne peut pas décrire l'échelle de l'image). En fait, le problème ici vient du fait que l'échelle des champs est plus grande que l'échelle la plus grande à laquelle on mesure la TV (i.e. $t \geq 64$).



(a)

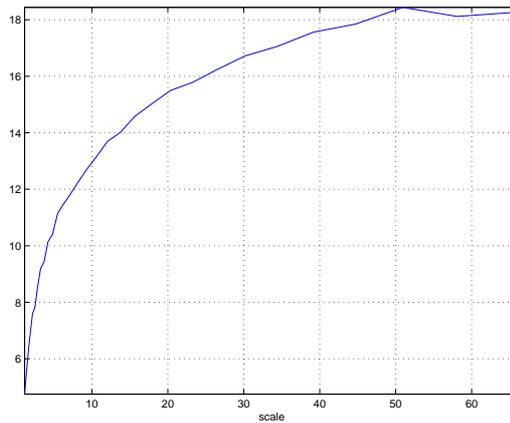


(b)

FIG. 3.7 – (a) Image aérienne sur la ville de Tunis ©Sup’Com ; (b) mesure d’échelle basée sur la TV. Notons que quand $t = 20$, la TV atteint son maximum.



(a)



(b)

FIG. 3.8 – (a) Image aérienne sur les champs ©CNES ; (b) mesure d’échelle basée sur la TV. Notons qu’il n’y a pas de pic significatif de TV.

3.3 Mesure basée sur le flot de la variation totale (TV)

Les mesures d’échelle montrées précédemment sont basées sur le gradient des images. Par conséquent, l’échelle ne peut pas être mesurée sur les régions qui ne disposent pas d’un gradient significatif. De plus la mesure basée sur la TV est globale. [6] a proposé d’utiliser le flot de TV pour mesurer l’échelle qui est basée sur la taille et la forme de région dans l’image. Elle est donc à la fois globale et locale. L’idée principale est de faire évoluer l’image originale $I(x, y)$ selon l’EDP suivante

$$\partial_t u = \operatorname{div} \left(\frac{\nabla u}{|\nabla u|} \right) \quad (3.5)$$

avec la condition initiale $u_{t=0} = I$. La vitesse du flot de TV de chaque pixel dépend de l'aire de la région de même niveau de gris à laquelle ce pixel appartient (il s'agit d'un résultat théorique en dimension 1, et expérimental en dimension 2 [6]). Ainsi on peut mesurer l'échelle d'un pixel en mesurant sa vitesse d'évolution selon le flot de la variation totale. Plus concrètement, on calcule l'échelle m en utilisant la formule :

$$\frac{1}{m} = C \frac{\int_0^T |\partial_t u| dt}{T - \int_0^T \delta_{\partial_t u, 0} dt} \quad (3.6)$$

où T est le temps final de l'évolution, $\delta_{a,b} = 1$ si $a = b$, et 0 sinon. On renvoie le lecteur à [6] pour plus de détails sur cette approche.

Mais cette méthode ne considère que l'aire de la région : deux régions qui ont la même aire mais différentes formes possèdent la même mesure d'échelle. [13] a proposé une autre schéma du flot de TV qui considère non seulement l'aire de la région, mais aussi son périmètre. L'évolution est aussi basée sur (3.5). En utilisant la formule de la co-aire [2], on peut réécrire (3.5) sous la forme

$$\int_{E_\lambda} \frac{\partial u}{\partial t} dx = P(E_\lambda) \quad (3.7)$$

où $E_\lambda = \{x | u(x) < \lambda\}$, et $P(E_\lambda)$ est le périmètre de E_λ . Ensuite on calcule la vitesse d'évolution selon (3.6). Ainsi on peut voir que la vitesse d'évolution est non seulement basée sur l'aire de la région mais aussi sur son périmètre. Plus une région est grande, plus la vitesse du flot de TV est faible.

A partir de cette remarque, on peut associer à chaque pixel une échelle définie comme la vitesse du flot de TV (comme proposé dans [6]). Plutôt que d'utiliser la variation totale comme dans la section 3.2, on peut proposer une mesure globale de l'échelle définie cette fois comme la vitesse moyenne du flot de la variation totale. Dans cette section, on n'utilise que cette dernière mesure globale, et on compare avec les résultats obtenus dans la section 3.2.

En effet, si on part d'une image qui contient des objets de grande échelle, la vitesse du flot de TV normalisée par la taille de l'image doit être relativement faible. Maintenant, si on réduit la taille de l'image en faisant un sous-échantillonnage après convolution par une gaussienne, les tailles des objets diminuent, et par conséquent la vitesse du flot de TV normalisée augmente. Après une certaine échelle, les objets se mélangent, donc la vitesse du flot de TV diminue. Ainsi, le graphe de la vitesse du flot de la variation totale (TV) en fonction de l'échelle va présenter un maximum. La valeur de l'échelle correspondant à ce maximum nous donne la valeur de l'échelle de l'image.

3.3.1 Expériences et résultats

Pour calculer le flot de TV, on décompose tout d'abord l'image dans les régions de même niveau de gris en utilisant la transformée rapide d'ensemble de niveaux (*Fast Level Set Transformation (FLST)*) [25] [24] [14]). A l'aide de cette transformée, on peut ensuite calculer les aires et les périmètres des régions.

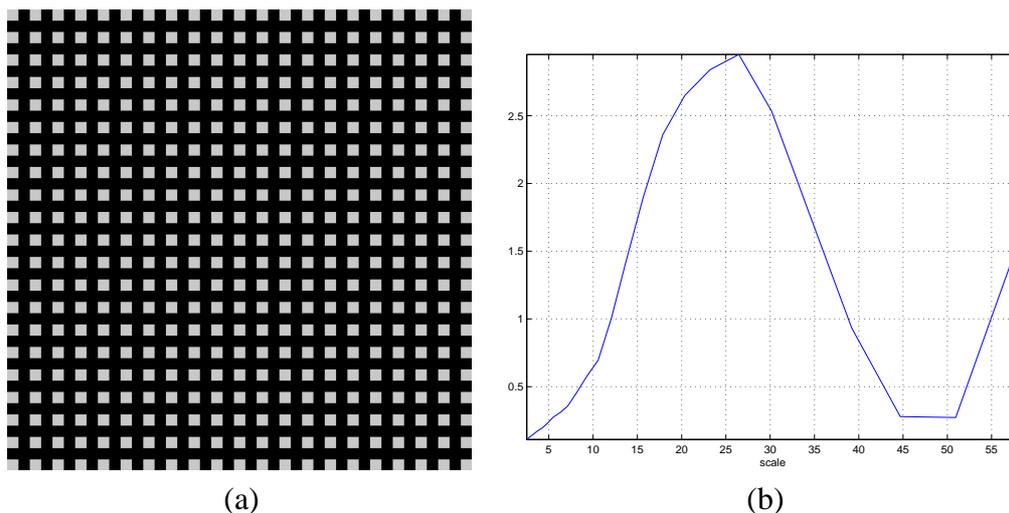


FIG. 3.9 – (a) Image synthétique des carrés de largeur 25 pixels ; (b) mesure d'échelle basée sur le flot de TV. Notons que pour $t = 26$ le flot de TV atteint son maximum.

On fait les expériences sur les mêmes images que l'on a déjà utilisées dans la section 3.2.3, i.e. 3 images synthétiques et 5 images réelles sur trois classes de terrains. Les résultats sont les courbes de la vitesse du flot de TV en fonction de l'échelle.

Résultats sur les images synthétiques

Les images 3.9(a), 3.10(a), et 3.11(a) sont les mêmes que celles présentées dans la section 3.2.3. Les figures 3.9(b), 3.10(b), et 3.11(b) sont les résultats de la mesure d'échelle en utilise le flot de TV. On peut voir que quand $t = 26$, $t = 20$ et $t = 20$ les courbes atteignent respectivement les maxima globaux. En comparant ces résultats avec ceux qui sont obtenus dans la section 3.2.3, on peut voir que la position de maximum obtenu avec TV est presque la moitié de celle obtenue avec la vitesse du flot de TV. On peut également noter que quand la taille de l'objet diminue à 2×2 pixels, la TV atteint son maximum ; par contre, pour la vitesse du flot de TV, la taille doit diminuer jusqu'à 1×1 pixel.

Résultats sur les images réelles

Dans cette section, on montre les résultats de la mesure d'échelle basée sur le flot de TV sur les images réelles. Les images sont les mêmes que celles utilisées dans la section 3.2.3. Sur la forêt (cf. figure 3.12), la vitesse du flot de TV atteint son maximum quand t est autour de 8. Sur la ville de Paris (cf. figure 3.13), la vitesse du flot de TV atteint son maximum autour de $t = 90$. Pour les villes de Lyon (figure 3.14) et Tunis (figure 3.15), les positions des maxima sont autour de $t = 30$. Et sur les champs, il n'y a pas de maximum significatif. Remarquons qu'en comparant avec les résultats de la section 3.2.3, les positions des maxima de la vitesse du flot de TV sont toujours presque le double de ceux obtenus avec TV. Cela confirme ce que l'on a vu dans la section précédente.

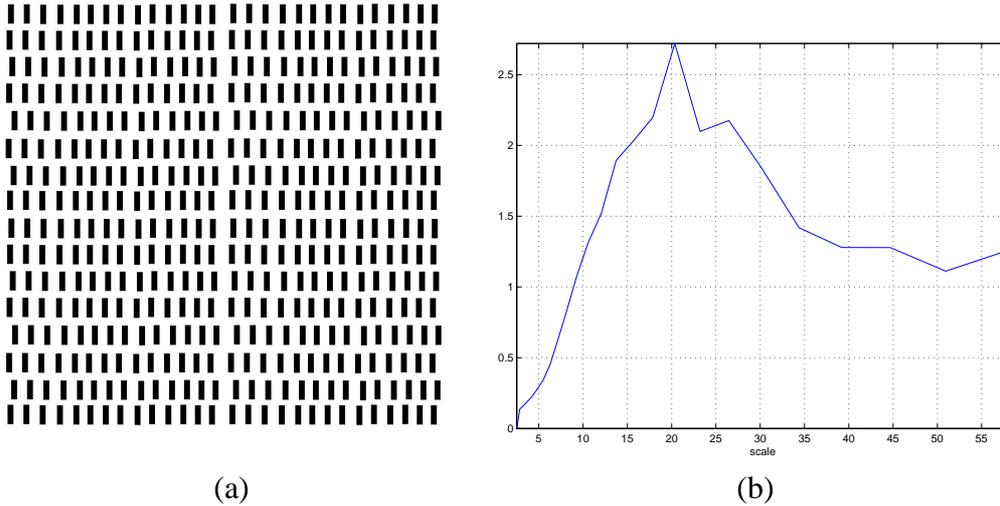


FIG. 3.10 – (a) Image synthétique des rectangles de taille 13×45 pixels ; (b) mesure d'échelle basée sur le flot de TV. Notons que pour $t = 20$ le flot de TV atteint son maximum.

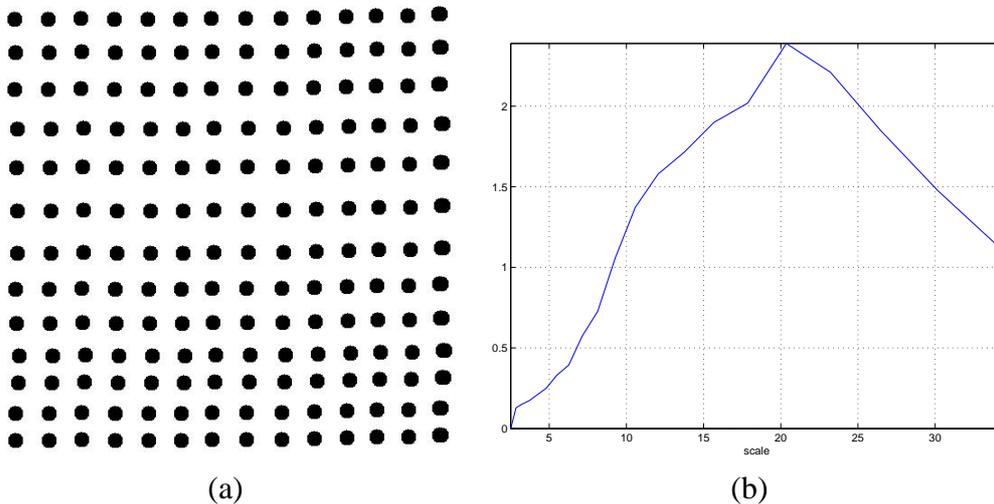
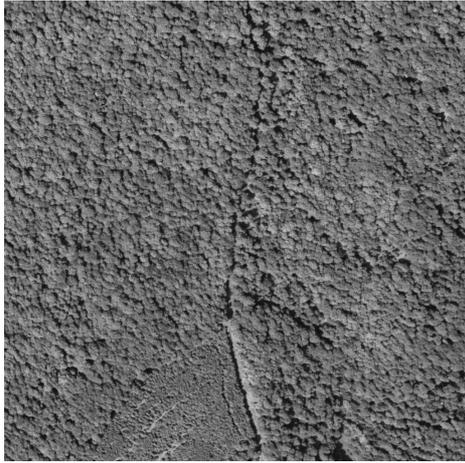


FIG. 3.11 – (a) Image synthétique des cercles de rayon 18 pixels ; (b) mesure d'échelle basée sur le flot de TV. Notons que pour $t = 20$ le flot de TV atteint son maximum.

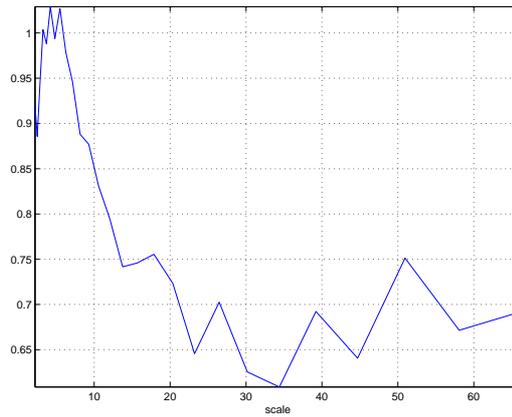
3.3.2 Comparaison avec la méthode basée sur la variation totale et perspective

On a étudié deux mesures d'échelle de l'image, i.e. la variation totale (TV) et la vitesse du flot de la variation totale (TV). Ces deux méthodes donnent des résultats satisfaisants.

La mesure utilisant le calcul de TV est essentiellement basée sur la norme du gradient de l'image ; donc si l'image ne dispose pas d'un gradient significatif, cette mesure ne donnera pas un résultat très précis. La méthode utilisant la vitesse du flot de TV est basée sur les régions dans l'image (elle prend en compte le périmètre et l'aire des régions) : cette méthode peut donc être utilisée même en l'absence de gradient, et théoriquement on peut espérer obtenir des résultats plus précis.



(a)

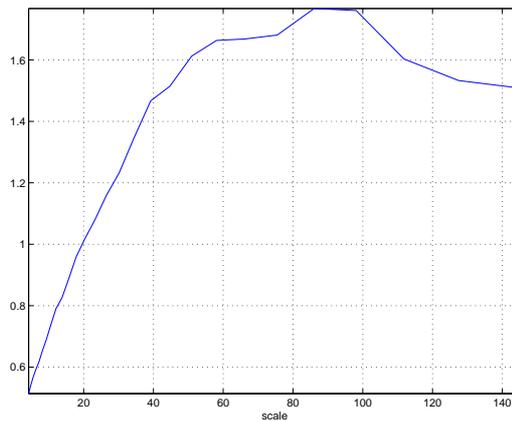


(b)

FIG. 3.12 – (a) Image aérienne sur la forêt ©CNES ; (b) mesure d'échelle basée sur le flot de TV. Notons que quand t est autour 8, le flot de TV atteint son maximum.



(a)



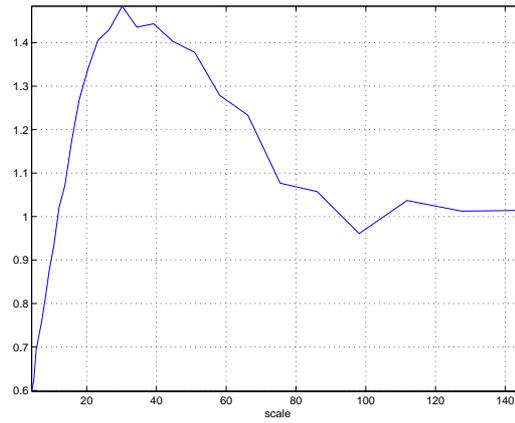
(b)

FIG. 3.13 – (a) Image aérienne sur la ville de Paris ©CNES ; (b) mesure d'échelle basée sur le flot de TV. Notons que quand $t = 90$, le flot de TV atteint son maximum.

Mais en pratique, la mesure basée sur la vitesse du flot de TV est beaucoup plus lourde à mettre en œuvre que celle basée sur TV. En effet, à chaque fois pour calculer le flot de TV, il faut décomposer l'image en utilisant la FLST (transformée rapide en ensemble de niveau). Or les résultats obtenus nous montrent que les deux mesures donnent des résultats équivalents à un facteur 2 près (la position du maximum de TV est approximativement la moitié de celui du maximum de la vitesse du flot de TV). De plus, le maximum obtenu avec TV est plus marqué que celui obtenu avec la vitesse du flot de la variation totale : la mesure avec TV est finalement non seulement beaucoup plus rapide mais aussi plus précise. Remarquons que la mesure basée sur la vitesse du flot de TV peut fournir non seulement un résultat global, mais aussi un résultat local sur chaque pixel. Mais en pratique, nous nous intéressons ici à une échelle plutôt globale, celle d'une zone homogène de l'image (i.e. ne contenant qu'une



(a)

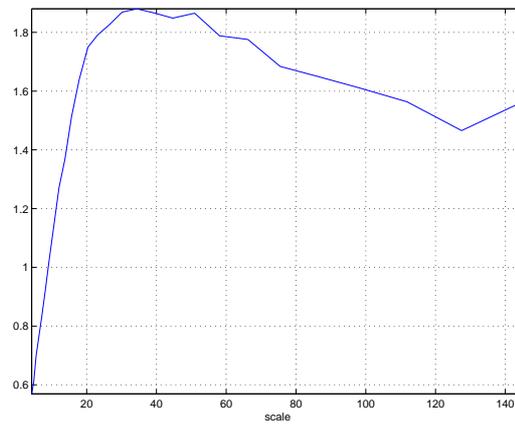


(b)

FIG. 3.14 – (a) Image aérienne sur la ville de Lyon ©CNES ; (b) mesure d'échelle basée sur le flot de TV. Notons que quand t est autour de 30, le flot de TV atteint son maximum.



(a)



(b)

FIG. 3.15 – (a) Image aérienne sur la ville de Tunis ©Sup'Com ; (b) mesure d'échelle basée sur le flot de TV. Notons que quand t est voisin de 30, le flot de TV atteint son maximum.

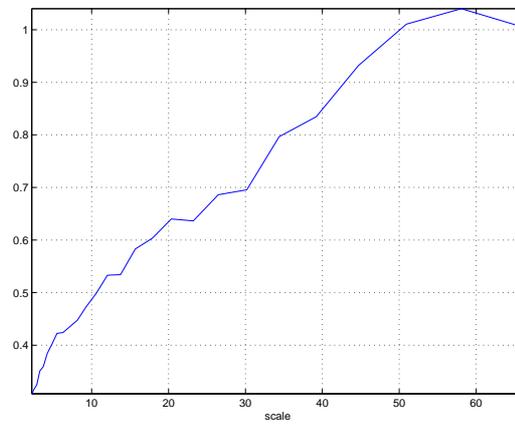
seule classe de terrain).

C'est pourquoi nous préférons plutôt la mesure basée sur la TV comme mesure d'échelle globale (dont les résultats sont confortés par ceux obtenus avec la vitesse du flot de TV).

Rappelons enfin que l'on a toujours utilisé l'espace multi-échelle linéaire. On peut donc penser à utiliser d'autres espaces non linéaires, comme par exemple l'espace obtenu à l'aide des opérateurs morphologiques. Une autre idée serait d'utiliser d'autres opérateurs tels que la courbure pour mesurer l'échelle (mais nos premiers tests numériques ne sont pas très concluants).



(a)



(b)

FIG. 3.16 – (a) Image aérienne sur les champs ©CNES ; (b) mesure d'échelle basée sur le flot de TV. Notons qu'il n'y a pas de maximum significatif du flot de TV.

Chapitre 4

Comportement structure/texture suivant la résolution de l'image

4.1 Introduction

Pour rechercher les résolutions où une image change de comportement, l'idée est de calculer la norme d'une même scène à différentes résolutions. A cette fin, le CNES nous a fourni des jeux d'images. Pour une même scène, on dispose de 33 résolutions (et pour chaque résolution, on a 3 images de différents canaux). Chaque série consiste donc en 33×3 images. La résolution va de 25 cm à 10,08 m selon une progression géométrique de raison $r = 2^{1/6}$. Cela donne les résolutions présentées dans le Tableau 4.1. En particulier, on dispose des résolutions suivantes : Orfeo (70,7 cm), Spot 5 THR (252 cm), SPOT 5 (504 cm), et SPOT 2 et 3 (10007,9 cm).

4.2 Protocole

On note \mathcal{N} une semi-norme. On calcule la semi-norme \mathcal{N} de l'image à chaque résolution (en normalisant la valeur de la norme par le nombre de pixels de l'image).

Le premier choix de semi-norme auquel on peut penser est celui de la norme L^2 normalisée : si on note I une image, on note $\mathcal{N}_{L^2} = \|I - \bar{I}\|_2$ où \bar{I} est la moyenne de I . Cependant, comme on le verra sur les graphes des Figures 4.3 et 4.7, cette semi-norme ne donne pas beaucoup de renseignements.

Un choix plus judicieux consiste à prendre la variation totale TV . On rappelle que si une

TAB. 4.1 – Valeur des résolutions simulées (en mètres). On dispose ainsi notamment de la résolution d'Orfeo (70,7 cm), de Spot 5 THR (252 cm), de SPOT 5 (504 cm), et de SPOT 2 et 3 (10007,9 cm)

0.250	0.281	0.315	0.354	0.397	0.445	0.500	0.561	0.630	0.707	0.794
0.891	1.00	1.12	1.26	1.41	1.59	1.78	2.00	2.25	2.52	2.83
3.17	3.56	4.00	4.49	5.04	5.66	6.35	7.13	8.00	8.98	10.08

fonction I est C^1 , alors $TV(I) = \int_{\Omega} |\nabla I|$. On sait en effet [23, 4] que la variation totale TV est petite pour les images géométriques, mais grande pour les images texturées. A partir du graphe de la variation totale d'une image en fonction de sa résolution, il est donc possible de dire selon la résolution si l'image a un comportement plutôt géométrique ou plutôt texturé.

4.3 Résultats

La série présentée ci-après correspond à Salon de Provence (terrain d'aviation). L'image de départ à résolution de 0,25 m, cf. Figure 4.1, étant de taille 4096 x 4096, la dernière de résolution 10,08 m est de taille 102 x 102. Les graphes des Figures 4.2 et 4.3 correspondent à l'image de Salon de Provence présentée sur la Figure 4.1.

On voit que les 3 canaux ont le même comportement (et nous avons vérifié cette propriété dans toutes nos expériences). Ils sont respectivement notés i , r et v .

On voit que la semi-norme \mathcal{N}_{L^2} ne donne pas d'information significative (Figures 4.3). Par-contre, le graphe de la variation totale TV (Figures 4.2) montre que de la résolution 0,25 m à 0,315 m, l'image est essentiellement géométrique. Si la résolution est supérieure à 5,66m, l'image a un comportement essentiellement d'image texturée. Enfin, si la résolution est entre 0,315 m et 5,66m, l'image présente à la fois des caractéristiques géométriques et texturées.

Les valeurs critiques de résolution sont donc ici de 0,315 m et de 5,66m. Notons tout de suite qu'il ne s'agit que de l'ordre de grandeur de ces valeurs, mais en aucun cas d'une mesure précise. Si la résolution est inférieure à 0,315 m, alors pour étudier l'image il faut utiliser des éléments basés sur la géométrie. Si la résolution est supérieure à 5,66m, il faut utiliser des éléments basés sur la texture. Et si la résolution est entre ces deux valeurs critiques, il faut considérer à la fois la géométrie et les textures de l'image. La Figure 4.4 présente l'image à la résolution 5,66m, et on constate effectivement que toutes les informations géométriques de l'image de départ (celle à la résolution 25cm) ont disparu.

Fort de ce premier résultat, nous ne nous intéressons plus qu'à une partie de l'image de la Figure 4.1, en haut à gauche. Nous présentons la nouvelle image sur la Figure 4.5. Contrairement à la première image (Figure 4.1), celle-ci présente à peu près partout de nombreuses structures.

Les graphes des Figures 4.6 et 4.7 correspondent à la partie de l'image de Salon de Provence présentée sur la Figure 4.5. On voit que la semi-norme \mathcal{N}_{L^2} ne donne pas d'information significative. Par-contre, la variation totale TV montre que de la résolution 0,25 m à 0,315 m, l'image est plutôt géométrique, puis qu'ensuite elle a un comportement de plus en plus d'image texturée jusqu'à la résolution de 2,25 m. Pour les résolutions de 2,25 m à 10,08 m, l'image est essentiellement texturée. La Figure 4.8 montre l'image obtenue pour la résolution de 2,25 m : on voit qu'effectivement les structures géométriques de l'image originale (Figure 4.5) ont disparu, et que l'on qualifie bien le comportement de l'image comme étant celui d'une image texturée.

Pour résumer, il y a ici deux valeurs critiques de résolution : à 0,315 m et à 2,25 m. Si la résolution est inférieure à 0,315 m, l'image est géométrique. Si la résolution est supérieure à 2,25 m, l'image est une image de texture. Enfin, si la résolution est entre 0,315 m et 2,25 m, l'image présente à la fois des caractéristiques de géométrie et de texture. A nouveau,



FIG. 4.1 – Salon de Provence à 25 cm ©CNES

rappelons qu'il ne s'agit ici que de l'ordre de grandeur des valeurs critiques. Nous cherchons d'ailleurs actuellement à développer une méthode qui permette d'obtenir ces valeurs avec une plus grande précision.

4.4 Mesure de l'échelle de certains éléments de l'image

On a mesuré les échelles sur 4 régions extraites sur l'image montrée dans la figure 4.1 dont les localisations sont montrées dans la figure 4.9. La taille de chaque région extraite est toujours 512×512 . On utilise la méthode basée sur la variation totale présentée dans la section 3.2 pour mesurer l'échelle de ces régions. Dans la figure 4.10, on obtient un pic à $t = 16$, i.e. la résolution est $4m$. Or si on regarde les graphes de la Figure 4.6, on voit que cette résolution correspond au passage à un comportement d'image totalement texturée pour l'image de la Figure 4.5. Il se trouve que les structures de la figure 4.10 sont les plus grosses de l'image de la figure 4.5. Cela signifie donc que l'on obtient un comportement d'image texturée lorsque la résolution de l'image est supérieure à la taille caractéristique des plus grosses structures de l'image. Ce résultat est cohérent avec notre intuition.

Et dans la figure 4.11, il y a plusieurs pics dans une intervalle $t \in [2, 6]$ (la résolution est

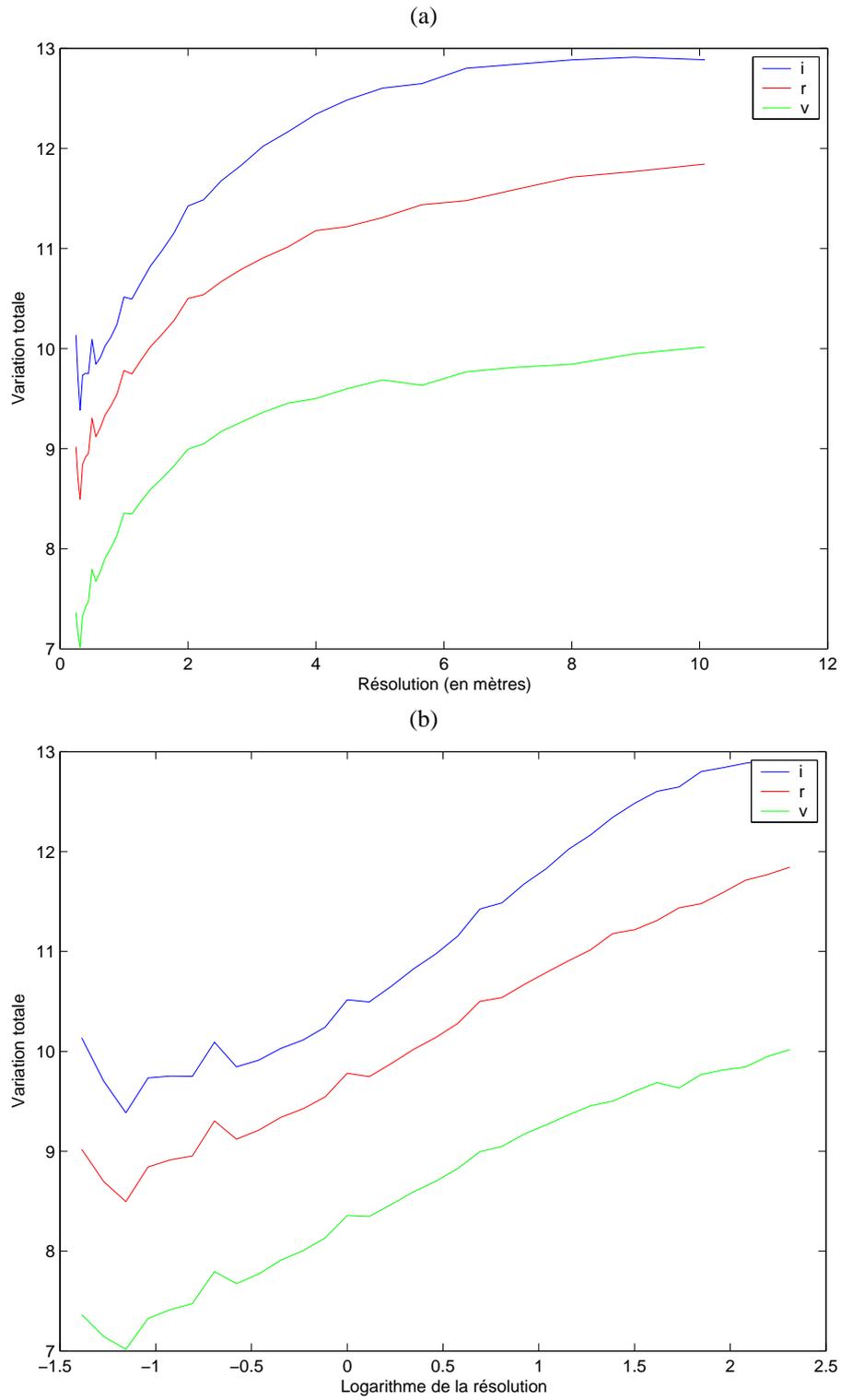


FIG. 4.2 – Variation totale de la scène de Salon de Provence (Figure 4.1) en fonction de la résolution : (a) échelle linéaire de résolution de 25 cm à 10,08 m, (b) échelle logarithmique.

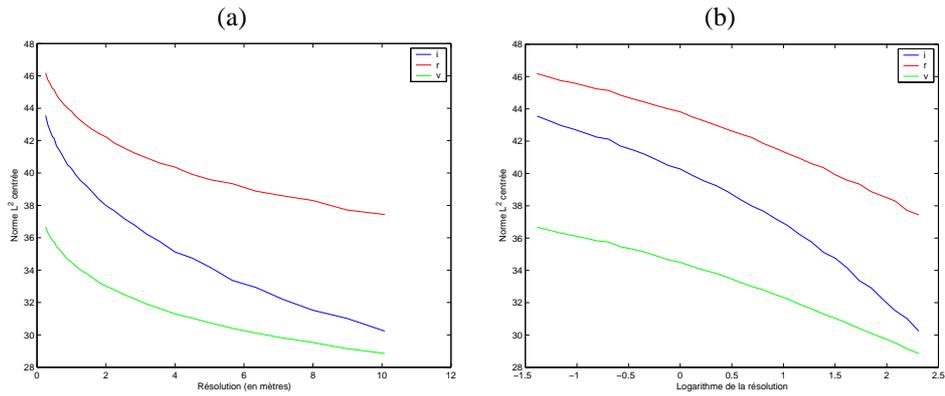


FIG. 4.3 – Norme L^2 centrée de l'image (Figure 4.1) en fonction de la résolution : (a) échelle linéaire de résolution de 25 cm à 10,08 m, (b) échelle logarithmique.



FIG. 4.4 – Image à 5,66 m ©CNES.

donc dans $[0.5, 3]m$). Pour la figure 4.12, la pic est sur $t = 4$, i.e. la résolution est presque 1m. Enfin il y a une pic sur $t = 2$ dans la figure 4.13 qui correspond à la résolution de 0.5m. Si on regarde les graphes des Figures 4.2 et 4.6, on voit que toutes ces résolutions sont dans la phase de transition d'un comportement uniquement géométrique de l'image à un comportement uniquement texturé.

On en déduit donc que la plus petite échelle d'un élément géométrique dans l'image correspond au passage d'un comportement géométrique à la phase de transition observée sur les graphes des Figures 4.2 et 4.6, et que la plus grande échelle d'un élément géométrique correspond au passage de la phase de transition à un comportement uniquement texturé.



FIG. 4.5 – Image à 25 cm ©CNES

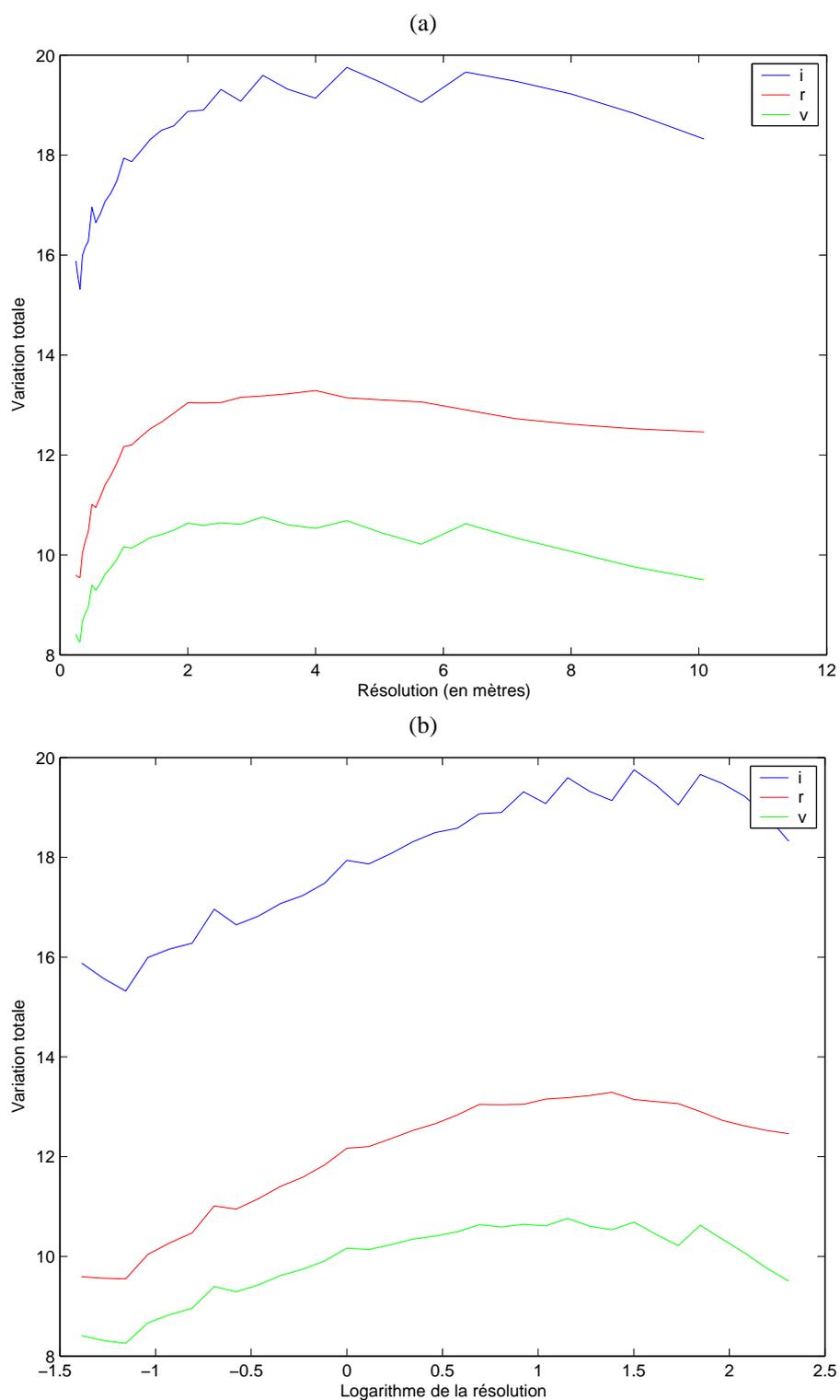


FIG. 4.6 – Variation totale de l’image (Figure 4.5) en fonction de la résolution : (a) échelle de résolution de 25 cm à 10,08 m, (b) échelle logarithmique

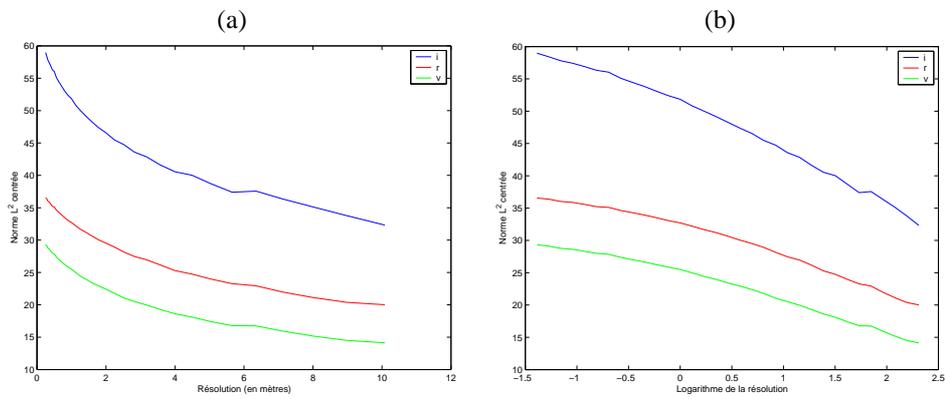


FIG. 4.7 – Norme L^2 centrée de l'image (Figure 4.5) en fonction de la résolution : (a) échelle de résolution de 25 cm à 10,08 m, (b) échelle logarithmique



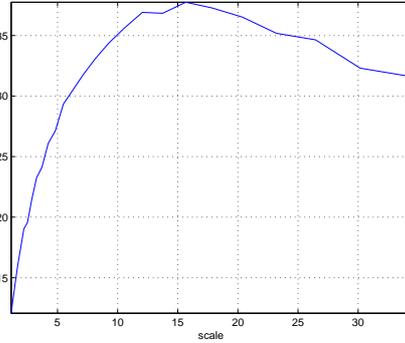
FIG. 4.8 – Image à 2,25 m ©CNES



FIG. 4.9 – Les localisations des 4 régions que l'on a extraites sur l'image montrée dans la figure 4.1



(a)

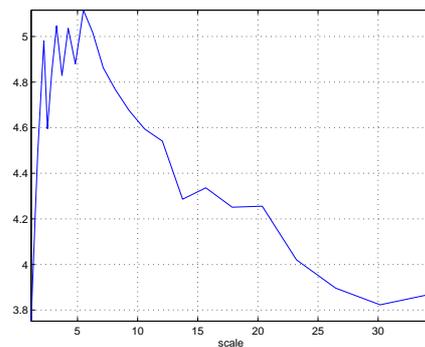


(b)

FIG. 4.10 – (a) Région extraite sur l’image 4.1 dont la localisation est indiquée dans la figure 4.9 comme *zone 1* ©CNES ; mesure d’échelle basée sur la TV. Notons que quand $t = 16$ la TV atteint le maximum.

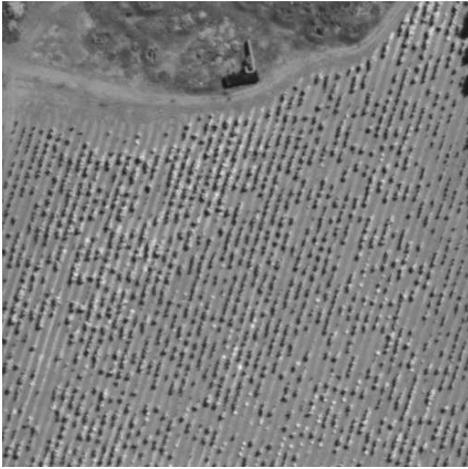


(a)

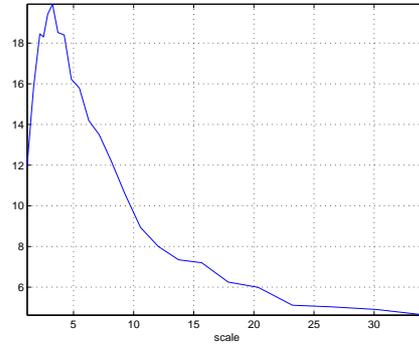


(b)

FIG. 4.11 – (a) Région extraite sur l’image 4.1 dont la localisation est indiquée dans la figure 4.9 comme *zone 2* ©CNES ; mesure d’échelle basée sur la TV. Notons que quand $t \in [2, 6]$, il y a plusieurs pics.



(a)

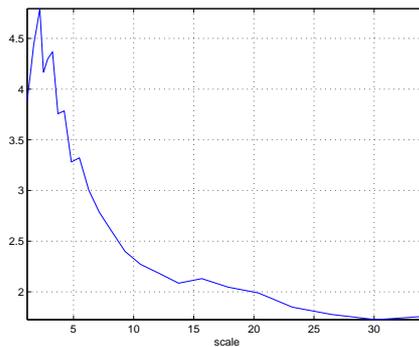


(b)

FIG. 4.12 – (a) Région extraite sur l’image 4.1 dont la localisation est indiquée dans la figure 4.9 comme *zone 3* ©CNES ; mesure d’échelle basée sur la TV. Notons que quand $t = 4$ la TV atteint le maximum.



(a)



(b)

FIG. 4.13 – (a) Région extraite sur l’image 4.1 dont la localisation est indiquée dans la figure 4.9 comme *zone 4* ©CNES ; mesure d’échelle basée sur la TV. Notons que quand $t = 2$ la TV atteint le maximum.

Conclusion et perspectives

Nous ne concluons bien sûr que de façon provisoire ce rapport puisque l'étude court encore. Nous n'avons malheureusement pas encore pu tirer profit de bases de données multiéchelle telles que celles mises en place pour la préparation d'Orfeo. Nous comptons beaucoup sur ces images pour accroître notre expérimentation.

Néanmoins, nous avons présenté ici quelques stratégies possibles pour tirer profit - et ne pas trop souffrir ! - de la diversité des résolutions qui affectent les images satellitaires. Nous avons en particulier montré que si l'on souhaite mener, très rationnellement, le choix des échelles de primitives à partir de la présence de ces échelles dans les paysages, il est indispensable de mener au préalable une étude approfondie des paysages qui composent la diversité des images satellitaires. Ceci passe par ce que nous appelons - abusivement à ce point - des ontologies de télédétection. Nous avons entrepris d'aborder - prudemment - ce domaine, en vue de présenter d'ici peu une analyse de ces perspectives. Plusieurs autres pistes de développement futur ont également été proposées qui devront faire l'objet de nouvelles discussions.

Nous avons également étudié la robustesse des primitives aux variations d'échelle et montré que l'on affectait peu les performances en autorisant l'emploi de primitives dans une gamme raisonnable d'échelles entourant l'échelle nominale. Nous avons ainsi montré qu'une excursion de plus ou moins une octave était possible avec des primitives de type Haralick. Ces travaux seront à confirmer sur de véritables bases de données multi-échelles, puisque nous n'avons pu les conduire que sur des images simulées.

Nous avons proposé deux méthodes pour calculer l'échelle dans une image en représentant l'image dans l'espace multi-échelle linéaire. La première est basée sur le calcul de la norme de la variation totale, la seconde sur la vitesse du flot de la variation totale. Nous avons montré expérimentalement l'intérêt de ces deux méthodes. Nous retenons que la première méthode, basée sur la norme de la variation totale, est finalement la plus rapide et la plus précise.

Rappelons que nos mesures se décomposent toujours en deux étapes : représentation de l'image dans un espace multi-échelle, puis calcul d'une certaine mesure sur chaque échelle. De plus, les deux mesures que nous avons considérées dans ce rapport (la TV et le flot de TV) sont calculées sur la représentation de l'image dans l'espace multi-échelle linéaire. On peut donc imaginer représenter l'image dans d'autres espaces, comme les espaces basés sur les opérateurs mathématiques morphologique par exemple. En même temps, dans la deuxième étape, d'autres mesures peuvent aussi être proposées, comme par exemple la courbure ou la norme L^2 .

Dans un deuxième temps, nous avons utilisé notre connaissance de la mesure d'échelle dans une image pour étudier le problème du comportement géométrique ou texture d'une

image en fonction de la résolution. Nos premiers résultats montrent que nous pouvons mettre en évidence deux résolutions critiques. En dessous d'une première résolution, l'image a un comportement uniquement géométrique, et au dessus de la deuxième, l'image a un comportement uniquement texturé. Entre ces deux valeurs, l'image présente à la fois des caractéristiques géométriques et texturées. Nous établissons le lien entre ces valeurs critiques et les échelles caractéristiques des éléments présents dans l'image : la première valeur critique correspond en effet à la plus petite échelle présente dans l'image, et la deuxième à la plus grande échelle.

A l'avenir, il nous faut travailler sur la précision numérique que nous avons pour calculer la valeur de ces seuils. Pour le moment, nous n'en donnons malheureusement qu'un ordre de grandeur (à un facteur 2 près).

De plus, on peut penser intégrer les mesures d'échelle et la détermination du comportement structure/texture dans le problème de classification de terrain. Plus précisément, on peut classifier les terrains de façon très grossière en fonction de l'échelle du terrain (i.e. on peut prendre la mesure d'échelle comme l'une des caractéristiques des classes que l'on considère). On peut aussi utiliser la détermination du comportement de l'image, géométrique ou texturée, pour décider quels types de caractéristiques il est plus judicieux d'utiliser pour faire la classification : on utilisera alors selon le cas des caractéristiques de géométrie ou de texture, voire les deux si on se trouve dans l'intervalle de résolution correspondant à la phase de transition entre les deux comportements.

Bibliographie

- [1] L. Alvarez, Y. Gousseau, and J.M. Morel. Scales in natural images and a consequence on their bounded variation norm. In *Scale-Space '99*, volume 1682 of *Lecture Notes in Computer Science*, 1999.
- [2] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of bounded variations and free discontinuity problems*. Oxford mathematical monographs. Oxford University Press, 2000.
- [3] G. Aubert and P. Kornprobst. *Mathematical Problems in Image Processing*, volume 147 of *Applied Mathematical Sciences*. Springer-Verlag, 2002.
- [4] J.F. Aujol and A. Chambolle. Dual norms and image decomposition models. *IJCV*, 63(1) :85–104, June 2005.
- [5] M. Born and E. Wolf. *Principles of Optics*. Pergamon, New York, fourth edition, 1970.
- [6] T. Brox and J. Weickert. A TV flow based local scale measure for texture discrimination. In *ECCV 04*, volume 2, pages 578–590, May 2004.
- [7] M. Campedel, B. Luo, H. Maître, E. Moulines, and M. Roux et I. Kyrgyzov. Indexation des images satellitaires : Détection et évaluation des caractéristiques de classification. Technical Report 2004-D-008, Télécom Paris, Paris, december 2004.
- [8] M. Campedel, B. Luo, H. Maître, E. Moulines, M. Roux, and I. Kyrgyzov. CNES - DU 402-01 - indexation des images satellitaires : Détection et évaluation des caractéristiques. Technical report, Télécom-Paris, Dept TSI, Paris, 2004.
- [9] M. Campedel, H. Maître, and E. Moulines. CNES - DU 402-02 - indexation des images satellitaires : Comparaison et évaluation des classifications. Technical report, Télécom-Paris, Dept TSI, Paris, 2004.
- [10] F. Cao. Equation aux dériveées partielles et filtrage itératif, 2003. Notes de cours du DEA MVA.
- [11] G. Van de Wouwer. *Wavelets For Multiscale Texture Analysis*. PhD thesis, Universitaire Instelling Antwerpen, 1998.
- [12] L.C. Evans and R.F. Gariepy. *Measure theory and fine properties of functions*, volume 19 of *Studies in Advanced Mathematics*. CRC Press, 1992.
- [13] G. Koepfler F. Dibos and P. Monasse. *Total variation minimization : applications to gray scale, color images and optical flow regularization*. Geometric level set methods in imaging, vision and graphics. Springer, 2003.
- [14] J. Froment and L. Moisan. Megawave2. Free software intended for image processing, <http://www.cmla.ens-cachan.fr/Cmla/Megawave/index.html>.

- [15] R. M. Haralick, K. Shanmugan, and I. Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, 3(6) :610–621, November 1973.
- [16] P. Héas. *Apprentissage bayésien de structures spatio-temporelles : application à la fouille visuelle de séries temporelles d’images de satellites*. PhD thesis, thèse de l’ENSEA, Toulouse, 2005.
- [17] J.J. Koenderink. The structure of images. *Biological Cybernetics*, 50 :363–370, 1984.
- [18] T. Lindeberg. Feature detection with automatic scale selection. *IJCV*, 30(2), August 1998.
- [19] Stephane G. Mallat. A theory for multiresolution signal decomposition : The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7) :674–693, July 1989.
- [20] B. S. Manjunath and W. Y. Ma. Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8) :837–842, 1996.
- [21] Ramchandra Manthalkar, P. K. Biswas, and B. N. Chatterji. Rotation invariant texture classification using even symmetric gabor filters. *Pattern Recognition Letters*, 24(12) :2061–2068, 2003.
- [22] H. Maître. Echantillonnage. In *Le Traitement des Images*, pages 15–45. Hermes, IC2, Paris, 2003.
- [23] Yves Meyer. Oscillating patterns in image processing and in some nonlinear evolution equations, March 2001. The Fifteenth Dean Jacqueline B. Lewis Memorial Lectures.
- [24] P. Monasse. *Contrast invariant representation of digital images and application to registration*. PhD thesis, Ceremade, University Paris 9-Dauphine, 2000.
- [25] P. Monasse and F. Guichard. Fast computation of a contrast-invariant image representation. *IEEE Trans. on Image Processing*, 9(5) :860–872, 2000.
- [26] J.-M. Nicolas. Les bases de l’imagerie satellitaire. Technical report, Télécom Paris : photocopié de cours, 2005.
- [27] J.-P. Pérez. *Optique géométrique, ondulatoire et polarisation*. Masson, Paris, 1991.
- [28] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60 :259–268, 1992.
- [29] B. Schoelkopf and A. J. Smola. *Learning with kernels support vector machines, regularization, optimization and beyond*. MIT Press, 2001.
- [30] A. Winter. *Entropie et représentations multiéchelles pour l’interprétation automatique d’images satellitaires et aériennes*. PhD thesis, ENST Paris, 1997.