

Diplôme d'ingénieur POLYTECH Nice Sophia-Antipolis

Mathématiques Appliquées et Modélisation

Rapport de Stage 5^{ème} année

Outil d'aide à la validation de données Et Algorithmes de détection d'alignements de points

Du 12 Mars au 07 Septembre 2018

À Thales Alenia Space, Cannes



Julien GUEGAN

Cédric BOULBE – Didier AUROUX

Cécile MONFORT – Francis OLIVIER

Remerciements

Je tiens à remercier toute personne qui a pu contribuer, de près ou de loin, à la réalisation de mon stage.

D'abord, je remercie Thales Alenia Space ainsi que toutes les personnes qui m'ont accueilli sur le site de Cannes durant mon stage.

Ensuite, j'aimerais remercier pleinement Cécile, ma tutrice de stage, de m'avoir fait confiance pour ce projet. Ce fut un réel plaisir de travailler avec elle, j'espère que mon travail lui aura donné satisfaction.

Finalement, je souhaiterai remercier tout particulièrement les personnes travaillant dans l'équipe Qualité Image avec qui j'ai partagé ces quelques mois : Francis, Marine, Vincent, Emeric, Kevin.

En bref, je remercie tous le monde.

Table des matières

I	Présentation de l'organisme.....	5
I.1	Groupe Thales	5
I.2	Thales Alenia Space	6
I.3	Domaine observation sciences et service Architecture Chaîne Image	6
II	Contexte	8
II.1	Programme MTG	8
II.2	Instruments FCI et LI.....	8
II.3	Instrument Quality Tool (IQT)	9
III	Sujet et Objectifs	10
IV	Outil d'aide à la validation de données.....	11
IV.1	Logiciel BEST	11
IV.2	Utilisation de la bibliothèque EAST.....	13
V	Algorithmes de détection d'alignements de points.....	19
V.1	Transformée de Hough.....	20
V.2	Méthode A Contrario pour la détection d'alignements de points 2D.....	21
VI	Conclusion.....	31
VII	Bibliographie	32
VIII	Annexes.....	34
VIII.1	Déroulement du Stage.....	34
VIII.2	Programmes Météosat.....	36
VIII.3	Exemple de décodage de données	37
VIII.4	Notions complémentaires sur la partie V.2.....	38

Table des Acronymes

Acronyme	Mot complet
TAS	Thales Alenia Space
ESA	European Space Agency
MOP	Meteosat Operational Program
MSG	Meteosat Second Generation
MTG	Meteosat Third Generation
IQT	Instrument Quality Tool
CCSDS	Consultative Committee for Space Data Systems
CNES	Centre National d'Etudes Spatiales
EAST	Enhanced Ada SubSet
DEDSL	Data Entity Dictionary Specification Language
FCI	Flexible Combined Imager
LI	Lightning Imager
OC	Optical Channel
ASIC	Application Specific Integrated Circuit
DT	Detected Transient
FT	False Transient
TT	True Transient
NFA	Nombre de Fausses Alarmes

THALES ALENIA SPACE OPEN

Ce document ne peut être reproduit, modifié, adapté, publié, traduit d'une quelconque façon, en tout ou partie, ni divulgué à un tiers sans l'accord préalable et écrit de Thales Alenia Space.

© 2018, Thales Alenia Space

Modèle : 83230326-D0C-TAS-FR/003

I Présentation de l'organisme

I.1 Groupe Thales

Le groupe Thales est une société exerçant dans l'aérospatial, les transports, la défense et les technologies de l'information. Avec une forte dimension internationale, le groupe compte 65 000 collaborateurs répartis dans 56 pays. Ses origines remontent à 1998 lorsque les branches spécialisées dans les activités militaires d'Alcatel, Dassault Electronique et Thomson CSF sont réunies pour former une nouvelle société.

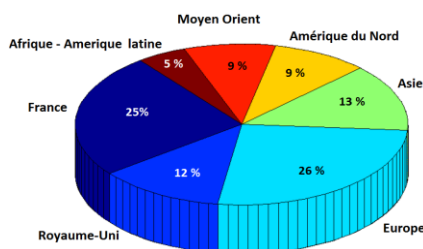


Figure 1 - Répartition géographique du groupe Thales

Le chiffre d'affaire réalisé par le groupe en 2013 s'élevait à 14, 2 milliards d'euros. Le groupe Thales est un organisme porté vers l'innovation. En 2013, 20% des revenus (soit 2.5 milliards d'euros) étaient consacré à la recherche et au développement. Le capital est partagé par l'état français indirectement à 27% et le groupe Dassault Aviation à 26% tandis que le reste (47%) est flottant.

L'entreprise est divisée en six secteurs d'activité :

- « **Systèmes d'information & de communications sécurisés** » : Produits de radiocommunication, Sécurité des technologies de l'information, Systèmes de réseaux et d'infrastructures, Systèmes de protection, Systèmes d'information critiques.
- « **Systèmes terrestres & aériens** » : Radar de surface, Gestion du trafic aérien, Opérations militaires aériennes, Systèmes d'armes avancés, Électronique de missiles, Optronique, Armement & munitions, Véhicules protégés.
- « **Systèmes de mission de défense** » : Systèmes navals de combat de surface, Systèmes sous-marins, Systèmes de combat électroniques, Renseignement, Surveillance, Reconnaissance.
- « **Avionique** » (n° 1 européen) : Avionique civile, Avionique militaire, Avionique Hélicoptères, Multimédia de bord, Systèmes électriques, Formation et simulation, Sous-systèmes hyperfréquence & d'imagerie.
- « **Systèmes de transport terrestre** » : Signalisation ferroviaire grandes lignes, Signalisation ferroviaire pour réseaux ferrés urbains, Systèmes intégrés de communication & de supervision, Systèmes de billettique.

THALES ALENIA SPACE OPEN

Ce document ne peut être reproduit, modifié, adapté, publié, traduit d'une quelconque façon, en tout ou partie, ni divulgué à un tiers sans l'accord préalable et écrit de Thales Alenia Space.

© 2018, Thales Alenia Space

Modèle : 83230326-DOC-TAS-FR/003

Le sixième secteur d'activité est lié à l'industrie spatiale à laquelle appartient Thales Alenia Space et c'est au sein de celle-ci que s'est déroulé mon stage.

I.2 Thales Alenia Space

Thales Alenia Space (TAS) est une co-entreprise, née le 10 avril 2007, et résultant de l'entrée des activités spatiales d'Alcatel-Lucent dans le groupe Thales. Elle fut créée par association de Thales et Finmeccanica, pour respectivement 67% et 33% du capital, et forme avec le groupe italien Telespazio, la coalition « Space Alliance ». TAS est une société basée à Cannes, composée de 7500 salariés et répartie sur 9 sites industriels, dans 5 pays d'Europe. En 2014, son chiffre d'affaire s'élevait à plus de 2 milliards d'euros. TAS est un acteur international incontournable dans plusieurs domaines d'activités tels que :

- les satellites de télécommunication avec des programmes militaires (Syracuse 3, Sicral) ou civils (Globalstar, O3B ou Iridium Next). Ce marché représente 50 % de son activité.
- les projets environnementaux axés sur l'observation de la Terre tels que le programme Global Monitoring for Environment and Security ou Meteosat.
- l'exploration avec sa participation à de nombreux programmes tels qu'ExoMars ou Herschel & Planck.
- les infrastructures orbitales : Thales Alenia Space a conçu et réalisé 50% du volume pressurisé de la Station Spatiale Internationale et a participé à la mise en place de vols spatiaux.
- la Navigation pour le positionnement avec des satellites comme Galileo.

Avec 40 ans d'expérience dans le domaine spatial, plus de 200 programmes spatiaux et satellites ont été menés par TAS. Le siège opérationnel de Thales Alenia Space est basé à Cannes, il compte aujourd'hui plus de 2000 salariés dont la grande majorité sont des ingénieurs et cadres. Cet établissement regroupe plusieurs activités comme la réalisation de satellites à but scientifiques (observation du globe terrestre, météorologie ...), la conception de satellites de télécommunication, l'assemblage intégration et tests des satellites à l'intérieur de salles blanches, le centre européen d'optique spatiale, le support des opérations et le contrôle des satellites en orbite.

I.3 Domaine observation sciences et service Architecture Chaîne Image

Le domaine d'observation et science est la branche de Thales Alenia Space qui prend en charge les satellites d'observation spatiale et terrestre. Parmi les projets en cours on retrouve les satellites de météorologie Meteosat (MSG et MTG), le programme Jupiter Icy Moon Explorer (JUICE) qui survolera les trois lunes glacées de Jupiter, et les programmes pour l'environnement

THALES ALENIA SPACE OPEN

Ce document ne peut être reproduit, modifié, adapté, publié, traduit d'une quelconque façon, en tout ou partie, ni divulgué à un tiers sans l'accord préalable et écrit de Thales Alenia Space.

© 2018, Thales Alenia Space

Modèle : 83230326-DOC-TAS-FR/003

comme JASON ou Sentinel-3 qui doivent surveiller les océans. L'organisation du domaine Observation et Science est détaillée ci-dessous en Figure 2.

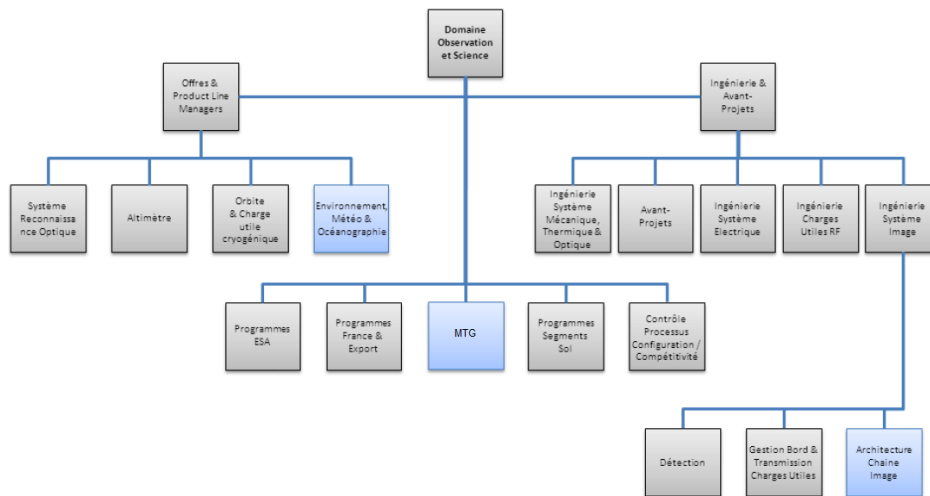


Figure 2 - Organisation du domaine Observation et Science

Le service Architecture Chaîne Image est en charge de plusieurs projets qui ont pour objectifs d'écrire des spécifications d'architecture de la chaîne Image et de développer des blocs de ces architectures, l'objectif est d'établir des spécifications des images acquises par les instruments et concevoir les traitements à appliquer comme par exemple corriger les altérations subies par les images (bruits radiométriques, déformations géométriques...). De plus, l'équipe du service Architecture Chaîne Image dans laquelle j'étais intégré est en charge de la partie imagerie du programme MTG s'organisant comme illustrer en Figure 3.

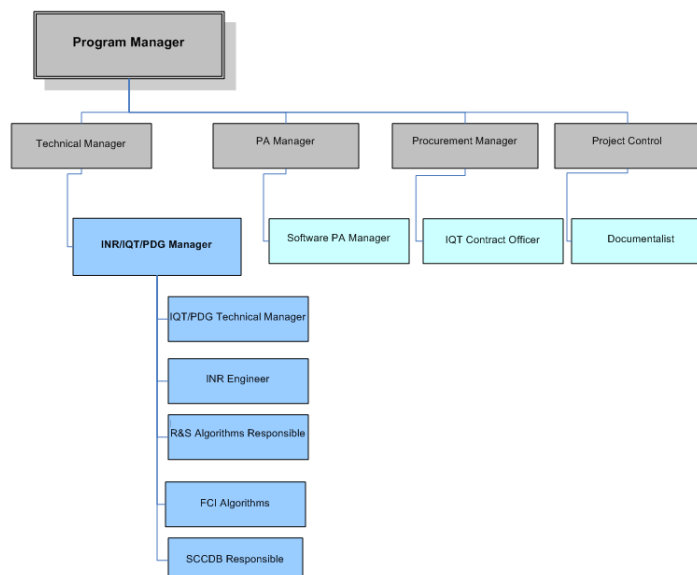


Figure 3 - Organisation hiérarchique du programme MTG

THALES ALENIA SPACE OPEN

Ce document ne peut être reproduit, modifié, adapté, publié, traduit d'une quelconque façon, en tout ou partie, ni divulgué à un tiers sans l'accord préalable et écrit de Thales Alenia Space.

© 2018, Thales Alenia Space

Modèle : 83230326-D0C-TAS-FR/003

II Contexte

II.1 Programme MTG

Le programme Météosat a été mis en place par l'Organisation météorologique mondiale au milieu des années 1970 et le premier satellite de ce programme fut lancé le 23 novembre 1977. Ce programme est aujourd'hui une famille de satellites météorologiques réalisés notamment sous maîtrise d'œuvre de l'Agence spatiale européenne (ESA). Géostationnaires ils scannent en continu une zone précise du globe et envoient régulièrement sur Terre de nombreuses images numériques de la surface terrestre qui permettent aux météorologues à partir des observations (anticyclones, dépressions atmosphériques, masses nuageuses ...) de faire des prévisions en utilisant des modèles météorologiques. Outre les bulletins météo quotidiens, ces données sont également essentielles pour anticiper certains phénomènes météorologiques ravageurs tels que les ouragans et tempêtes mais également pour le suivi global de l'évolution du climat de la planète. Actuellement, le programme Météosat comprend 3 générations de satellites : en 1977 MOP (1 mission d'observation, 3 canaux, satellite spinné), en 2002 MSG (2 missions d'observation, 12 canaux, satellite spinné), 2021 MTG (5 missions, 17 canaux, satellite 3-axes).

Le programme MTG (Meteosat Third Generation) actuellement en développement est donc la 3^{ème} génération de satellite du programme Météo européen, il est un vrai défi technologique puisqu'il améliore sur bien des aspects les performances des générations précédentes (voir Annexe VIII.1 Figure 36). Il est un système composé de 6 satellites : 4 MTG-I et 2 MTG-S (≈ 3500 kg/satellite) qui aura pour rôle de remplir 5 missions principales qui sont l'imagerie du globe, le sondage infrarouge de l'atmosphère, la détection des éclairs, la collecte de données pour la recherche/sauvetage et le sondage ultra-violet.

II.2 Instruments FCI et LI

Le Flexible Combined Imager (FCI) est l'un des principaux instruments embarqués du satellite MTG-I. Il permet le scan complet du disque terrestre en 10 minutes ainsi qu'un rapide scan de l'Europe (un quart du disque terrestre) en 2,5 minutes. Le FCI effectue des mesures sur 16 canaux, dont huit sont placés dans le domaine spectral solaire entre $0,4 \mu\text{m}$ et $2,2 \mu\text{m}$, délivrant des données avec une résolution spatiale de 1 km. Les huit canaux supplémentaires se situent dans le domaine spectral thermique entre $3,8 \mu\text{m}$ et $13,3 \mu\text{m}$, fournissant des données avec une résolution spatiale de 2 km. Le FCI apportera un soutien accru aux prévisions immédiates des conditions météorologiques à fort impact telles que les orages ou le brouillard. Il offrira également de meilleures capacités de surveillance des aérosols, ce qui contribuera grandement à la surveillance de la qualité de l'air. Grâce à sa capacité de résolution plus élevée dans l'infrarouge thermique, FCI permettra d'améliorer la détection des incendies et d'augmenter la qualité de divers produits liés au climat.



Figure 4 – L'instrument FCI dans un centre de test. [18]

Le Lightning Imager (LI) est un imageur qui fournit une localisation et une détection en temps réel de la foudre pendant la nuit et le jour. À bord des satellites MTG-I, il mesurera en continu à une longueur d'onde de $777,4 \pm 1$ nm (la longueur d'onde de la foudre), avec une résolution spatiale de 4,5 km au point de sous-satellite, déclenchée par un seuil variable – des impulsions optiques déclenchées par la foudre d'une énergie d'émission de plus de 4 à 7 $\mu\text{Jm}^{-2}\text{sr}^{-1}$. Le champ de vision (FOV) de l'instrument LI est couvert par quatre caméras identiques de 1000*1170 pixels, chacune couvrant un domaine sur quatre sur le disque terrestre observable. De plus, la fréquence d'image de l'instrument est de 1 KHz ce qui signifie qu'il est capable de traiter une image toute les millisecondes.



Figure 5 - Modèle structurel du LI. [18]

II.3 Instrument Quality Tool (IQT)

L'IQT est un ensemble de logiciels dédiés à l'évaluation de la qualité de l'image (spectrale, radiométrique, géométrique) des instruments de MTG. Avant le lancement, l'IQT permet de valider les performances géométriques en simulant des données réalistes à partir d'images de différents satellites existants. Après le lancement du satellite, il permet de valider toutes les performances de qualité d'image en utilisant les données de vol. L'IQT sera également utilisé pour l'optimisation des performances du système et pour l'investigation des anomalies en vol ayant un impact sur les performances.

Il est composé de 3 principales applications logicielles :

- **Simulation d'images et de données auxiliaires (IDS)** générant des données L0¹ que fourniront les instruments en vol;
- **Processeur prototype de niveau 1 (L1PP)** regroupant les traitements dédiés à la correction radiométrique, la correction spectrale et la correction géométrique qui permettent de générer des produits de niveau 1 à partir du niveau 0 et des données auxiliaires associées.
- **Evaluation des performances (PAT)** comprenant des outils dédiés à l'évaluation des performances géométriques, radiométriques et spectrales.

¹ Les données L0 sont les données scientifiques générées par un instrument donné, telles qu'elles sont reçues au sol, après restauration de la séquence chronologique des données par paquets.

III Sujet et Objectifs

En termes de responsabilités, le développement de l'IQT est partagé entre d'un côté Thales Alenia Space qui est responsable devant son client l'ESA de fournir un outil IQT répondant aux besoins spécifiés et de l'autre côté un sous-traitant spécialisé en développement de software (sélectionné conjointement par Thales Alenia Space et l'ESA) en charge de développer le code industriel de l'IQT. Dans ce cadre, l'équipe Qualité Image définit et valide, via des prototypes Matlab, les algorithmes qui doivent être implémentés dans les applications citées plus haut comme par exemple l'outil IDS qui simule l'ensemble des données L0 que fourniront les différents instruments une fois en vol, ces données sont par la suite exploitées par les autres applications de l'outil IQT pour évaluer les performances de la qualité de l'image. L'équipe Qualité Image est également garante du bon fonctionnement de l'outil IQT développé par le sous-traitant et doit donc s'assurer en particulier que les paquets de données L0 générés par l'IDS sont conformes en termes de format et de contenu. Il est donc nécessaire pour l'équipe Qualité Image, de disposer d'un outil capable de décoder ces données indépendamment de la façon dont elles ont été créées, de valider que la structure des paquets répond bien aux normes exigées et de vérifier le contenu scientifique. Le développement d'un tel outil a été la problématique de la première partie du stage.

Ensuite, toujours dans le cadre de l'IQT, une thématique récurrente est l'optimisation du temps de calcul réalisé par les différents traitements exécutés par le logiciel IQT. En effet, le logiciel traite d'importants volumes de données (comme des images) et chacune des applications du logiciel doit pouvoir s'exécuter dans des temps raisonnables (spécifiés par l'ESA). Notamment, la fréquence d'image de l'instrument LI est de 1 KHz, ce qu'il signifie qu'il doit traiter une image toute les millisecondes. Or, l'une des méthodes testée dans le cadre du filtrage des fausses détections d'éclairs du LI est actuellement trop longue à l'exécution. La deuxième partie du stage a donc été l'investigation d'une nouvelle méthode pour la détection de lignes pour le traitement des images de niveau L1b², dans le but de réduire le temps de calcul réalisé par la méthode actuellement utilisée.

² Les données L1b sont les données post-traitées qui sont radiométriquement et spectralement calibrées.

IV Outil d'aide à la validation de données

L'objectif principal de cette première partie de stage était donc de développer un outil capable de valider ces paquets de données et décoder des données "sciences" (des images, des dates, des angles ...). Pour ce faire, l'utilisation d'un logiciel développé par le CNES a été préconisée. Il a donc fallu installer, configurer, utiliser et comprendre le fonctionnement de cet outil ainsi que rédiger un guide d'utilisation pour assurer la pérennité de son utilisation. D'autre part, il a été compris que le CNES avait également développé une bibliothèque s'interfaçant avec cet outil et donc l'utilisation de cette bibliothèque en C pour manipuler, interpréter et valider les données décodées, ce qui fut une part importante du travail pour fournir un outil complet.

IV.1 Logiciel BEST

BEST est un ensemble d'outils développés en accord commun entre le CNES et l'ESA destinés à manipuler le cycle de vie entier des données (définition, génération, édition, vérification, extraction ...). Ses objectifs principaux sont de fournir une description des données qui soit complète, pérenne et facilement évolutive, accéder aux valeurs des données sans avoir à écrire de code spécifique.

BEST s'appuie notamment sur 2 recommandations internationales :

- EAST (CCSDS 644.0-B-1), langage conçu pour créer des descriptions non ambiguës des formats de données, y compris des informations syntaxiques (logiques et physiques)
- DEDSL (CCSDS 647.3-B-1), langage conçu pour décrire des informations sémantiques des données au moyen d'attributs sémantiques.

Le logiciel via une interface graphique permet à l'utilisateur de décrire ses données facilement sans connaissance de langages de programmation puis générer automatiquement un fichier décrivant ses données écrit en EAST ou en DEDSL. Dans le cadre de notre application, nous porterons un intérêt à la description EAST de nos données pour le décodage. La première étape était donc d'utiliser un des modules de BEST appelé "OASIS modeller" pour créer facilement des modèles de données correspondants à la façon dont les données L0 ont été codées. En effet, cet outil permet d'ajouter au fur et à mesure à notre modèle des éléments caractérisés par leur mode d'encodage, ici binaire et leur taille en bits, ainsi qu'éventuellement l'intervalle de valeur qu'ils peuvent prendre.

Dans le cadre du programme MTG et de l'IQT, les données sont générées suivant un protocole précis. En effet, depuis les années 1980 le Comité Consultatif pour les Systèmes de Données Spatiales (CCSDS)³ a développé une norme internationale ayant pour objectif d'envoyer efficacement la télémétrie en utilisant une unité de longueur variable appelée "paquet⁴ source" (avant chaque mission développait son propre système de données). Cette norme est utilisée par de nombreuses missions spatiales leur permettant de partager les données des équipements embarqués et des équipements de traitement des données au sol. Ces paquets de données sont donc écrits de façon à respecter la norme du CCSDS, et respectent tous la structure générique décrite Figure 6. Le packet Header est composé de 32 bits et contient des champs servant notamment à identifier notre paquet. Il est suivi du packet Data Field qui est de longueur variable qui est formé d'un Data Field Header de 96 bits caractérisant la nature des données du paquet puis le train de bits suivant encode les données en elles-mêmes générés par un instrument (des images, des mesures, des dates etc...) et enfin le paquet se termine avec 16 bits de contrôle (permettant par exemple d'identifier s'il y a une erreur dans la réception du paquet.)

Packet Header (48 Bits)						Packet Data Field (Variable)		
TM_PH						TM_PDF		
Packet ID			Packet Sequence Control		Packet Length	Data Field Header	Source Data	Packet Error Control
TM_PID			TM_PSC		TM_PLEN	TM_DFH	TM_SDF	TM_PEC
Version Number	Type	Data Field Header Flag	Application Process ID	Grouping Flags	Source Sequence Count			
TM_VER	TM_CAT	TM_DFH	TM_API	TM_GRP	TM_SEQ			
3 bits	1 bit	1 bit	11 bits	2 bits	14 bits			
16 bits			16 bits		16 bits	96 bits	Variable	16 bits

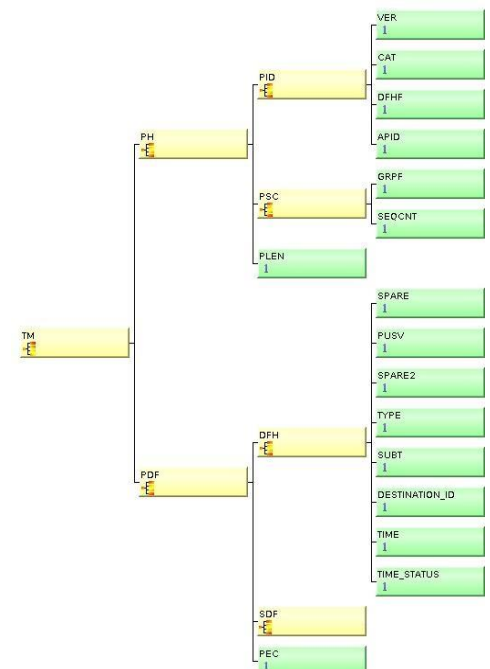


Figure 6 - Structure des paquets de télémétrie spécifiée par le CCSDS

Figure 7 - Modèle des données décrit avec BEST

Pour décrire les données sources ("Source Data" dans la Figure 6), les différentes documentations du programme MTG spécifiant le traitement des données de la mission permettent d'ajouter chaque donnée et le nombre de bits associé selon la nature du paquet (décrit par les Headers du paquet) conformément à la façon dont ces données sont codées. Une fois que notre modèle de données est complet, le logiciel BEST nous offre la possibilité de

³ Organisme composé des principales agences spatiales mondiales qui développe des normes pour éviter les problèmes récurrents des données spatiales et ainsi permettre l'interopérabilité.

⁴ Un paquet spatial est une unité de données de longueur variable (de 7 à 65542 octets), délimitée, alignée en octet qui encapsule un bloc de données d'observation (et des données auxiliaires).

vérifier la consistance entre le modèle que l'on a décrit et un fichier de données L0, et donc de détecter de premières anomalies si elles existent. Nous pouvons aussi afficher la valeur de toutes ces données via un ASCII dump⁵. Cependant pour valider nos données, nous avons besoin de pouvoir manipuler un minimum ces informations pour par exemple reconstituer des images au format bitmap, transformer des secondes en date calendrier ou encore appliquer des formules de conversion.

IV.2 Utilisation de la bibliothèque EAST

Le CNES met à disposition en même temps que son logiciel BEST une bibliothèque appelé EAST qui permet de récupérer certaines données spécifiées et donc les manipuler par la suite en C pour reconstituer des données plus facilement interprétables que des chiffres comme des images, des dates... La procédure pour accéder à une donnée via les services de la bibliothèque EAST en C est alors la suivante : initialiser (*start*), désigner le descriptif EAST⁶ utilisé (*select_DDR*) c'est-à-dire le fichier généré à partir de BEST et du modèle de données créé, charger en mémoire d'un bloc de notre fichier de donnée (*load_next_data_block_from_file*), préparer les accès futurs (*prepare_accesses_for_current_block*), accès aux données (*get_data_entity*), mettre fin aux services (*stop*).

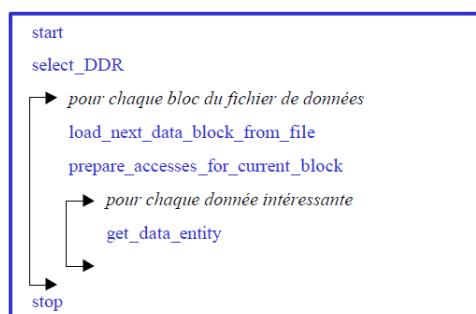


Figure 8 - Procédure d'appels aux méthodes de bibliothèque EAST pour accéder à une donnée

Au final, en ayant réalisé un modèle de données conforme au fichier de données L0, la création d'images bitmap a pu être possible en décodant les valeurs des pixels contenus dans un fichier binaire. Nous nous sommes intéressés dans un premier temps au décodage des images codées dans les paquets de données générées par l'instrument FCI (cf. II.2). Le FCI scanne entièrement le disque terrestre en 17 canaux spectraux (du visible à l'infrarouge) toutes les 10 minutes avec une résolution de 0.5 à 2 km selon le canal. Le scan d'un disque complet de la Terre est appelé un *Repeat Cycle* et est effectué en faisant des allers-retours appelé *swath* d'Est en Ouest et d'Ouest en Est (voir Figure 9).

⁵ Une simple conversion ASCII des données imprimée dans un fichier texte

⁶ Fichier généré à partir de BEST et du modèle de données créé qui décrit l'information syntaxique de nos données. EAST est un langage spécifique pour décrire des données et recommandé par le CCSDS.

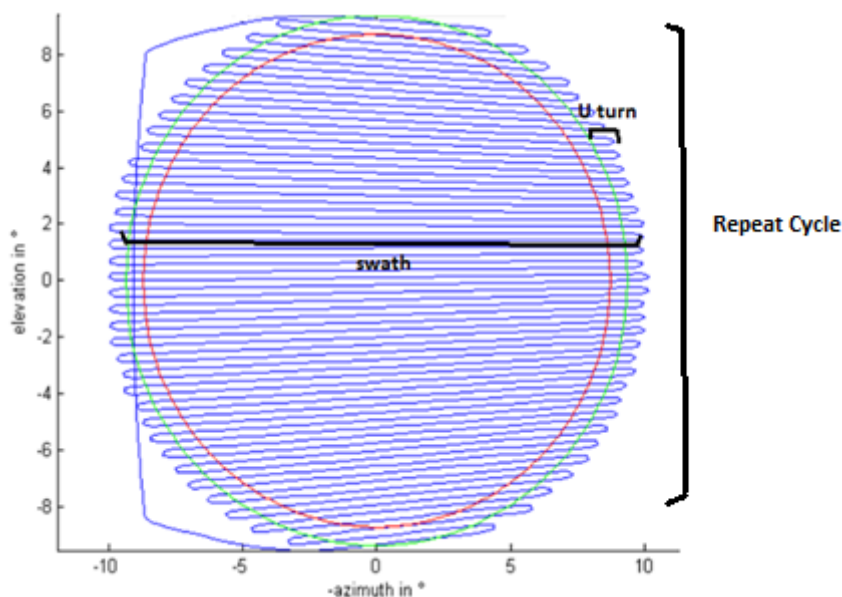


Figure 9 - Illustration de la loi de scan du FCI

Un fichier de données L0 est généré pour chacun des swaths et ce fichier contient un nombre variable de paquet (Figure 6) qui peut être parcouru entièrement grâce au service de la bibliothèque East. Un paquet contient une colonne de pixels de notre image pour une partie des canaux spectraux de l'instrument (il faut 4 paquets pour obtenir une colonne de pixels dans tous les canaux spectraux). En parcourant chaque paquets les uns après les autres, il faut récupérer colonnes par colonnes les bonnes valeurs de pixels correspondants au canal spectral auquel on s'intéresse. Par exemple, pour le fichier de données L0 correspondant au swath 12 de l'instrument FCI, on peut choisir de décoder les valeurs des pixels du canal visible de longueur d'onde 800 nm grâce aux services de la bibliothèque EAST, puis à partir de ces valeurs reconstruire une image au format bitmap. On obtient alors l'image présentée en Figure 10.



Figure 10 - Image du swath 12 du canal VIS 0.8

Puis en réalisant la même procédure pour tous les fichiers de données L0 du Repeat Cycle, soit pour les 70 swaths et en assemblant chacune des images obtenues les unes par-dessus les autres, une image complète du disque terrestre est créée (cf. Figure 11 et Figure 12). On remarque que les images obtenues sont allongées et que les swaths sont un peu décalés, c'est principalement dû au fait que 2 swaths consécutifs se chevauchent et possèdent en fait des pixels décrivant la même information. Au final, grâce à cet outil, on peut valider visuellement l'information codée dans les fichiers de données binaires et vérifier qu'ils contiennent bien ce à quoi l'on s'attend.

THALES ALENIA SPACE OPEN

Ce document ne peut être reproduit, modifié, adapté, publié, traduit d'une quelconque façon, en tout ou partie, ni divulgué à un tiers sans l'accord préalable et écrit de Thales Alenia Space.

© 2018, Thales Alenia Space

Modèle : 83230326-D0C-TAS-FR/003

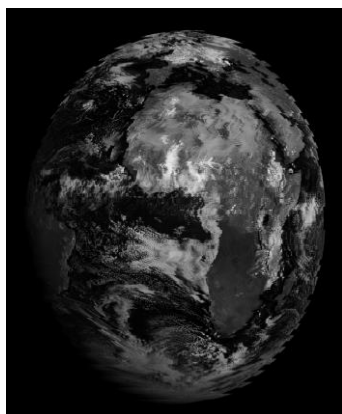


Figure 12 - Image complète du disque du canal VIS 0.8

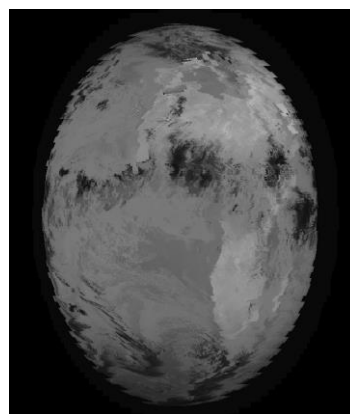


Figure 11 - Image complète du disque du canal IR 8.7

De la même manière, le décodage des images de l'instrument LI (cf. II.2) a pu être effectué. L'instrument LI a pour rôle de détecter et localiser continuellement la foudre sur presque l'ensemble du disque avec une résolution de 10 km. Il est composé de 4 canaux optiques (OC) qui sont inclinés d'un certain angle (voir Figure 13) dans des soucis d'optimisation de design.

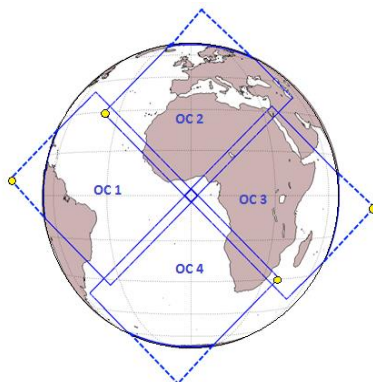


Figure 13 - Couverture terrestre des OC du LI

De plus, le détecteur de chaque OC est en réalité lu par 4 circuits intégrés spécifiques (ASIC) et chacun d'eux lit un quadrant de 500 lignes et 585 colonnes et sont numérotés selon les conventions présentées dans la Figure 14. D'autre part, pour un ASIC donné, l'image est envoyée au sol sous la forme de 2250 imagettes de 10x13 pixels (Figure 15).

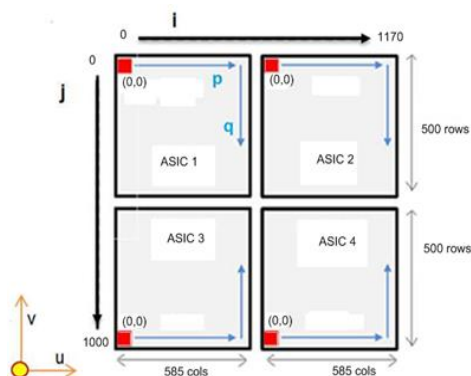


Figure 14 - Coordonnées des ASICs

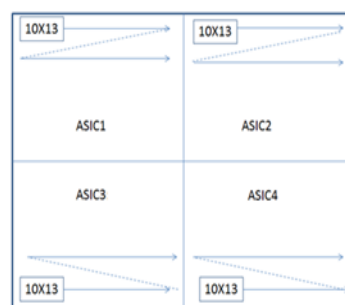


Figure 15 - Structure des imagettes dans les ASICs

THALES ALENIA SPACE OPEN

Ce document ne peut être reproduit, modifié, adapté, publié, traduit d'une quelconque façon, en tout ou partie, ni divulgué à un tiers sans l'accord préalable et écrit de Thales Alenia Space.

© 2018, Thales Alenia Space

Modèle : 83230326-DOC-TAS-FR/003

Un fichier de données L0 du LI contient une image complète de la Terre mais chaque paquet contient une des imagettes de 10x13 pixels qui ne sont pas forcément organisées les uns après les autres dans le fichier binaire. En parcourant chaque paquet les uns après les autres, il faut donc également récupérer le numéro de l'OC, de l'ASIC et de l'imagette pour pouvoir placer chaque pixel au bon endroit et reconstituer correctement l'image. Ainsi, l'image présentée Figure 16 ci-dessous a pu être décodée. Comme décrit en Figure 13, les parties du disque terrestre décodées se chevauchent d'où la présence d'un décalage sur la figure décodée. L'objectif de ce décodage est, comme pour les images précédentes, de valider visuellement l'information codée dans ces fichiers binaires.

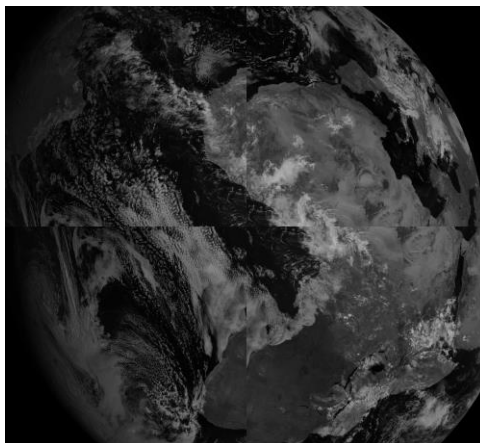


Figure 16 - Image décodée d'un fichier de données du LI

Parallèlement à ce travail sur le décodage des images des instruments LI et FCI, le décodage et la validation d'autres données ont été effectués comme par exemple les valeurs des quaternions⁷ qui sont, eux, générés à partir de la plateforme du satellite, ils représentent l'orientation et la rotation de l'engin spatial dans l'espace.

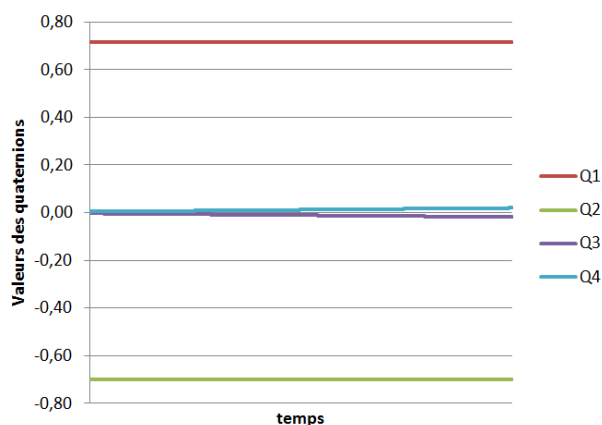


Figure 17 - Valeurs des quaternions décodées d'un swath

⁷ Nombre dans un sens généralisé car englobe les réels et complexes, la multiplication n'est plus commutative. Comparés aux angles d'Euler, ils sont plus simple à composer et évitent le problème du blocage de cardan (perte d'un degré de liberté). Comparés aux matrices de rotations, ils sont plus stables numériquement et peuvent se révéler plus efficaces.

Le décodage des parties Header (cf. Figure 6) générique à tous les types de paquets a également été mis en place pour permettre selon le besoin d'identifier la nature et la fonction du paquet.

Le développement de cet outil de validation des données L0 a permis de mettre à jour plusieurs anomalies qui ont ensuite été remontées au sous-traitant. Par exemple, en décodant la date de début d'un Repeat Cycle et en la convertissant au format UTC comme spécifiée par l'ESA, nous avons pu constater que cette donnée était codée incorrectement (mauvais format). De plus, les angles de scan (schéma Figure 9) du FCI ont également été signalés comme incorrectement codés puisqu'ils sont censés être compris entre -6° et $+6^\circ$ alors que leurs valeurs au décodage étaient hors de cet intervalle. Finalement on a également constaté que les paquets du LI n'étaient pas ordonnés correctement dans le fichier de données L0 puisque la valeur de l'Absolute Time des paquets de données montrait des inversions de pente alors qu'elle doit normalement augmenter au cours du temps (cf. Figure 18).

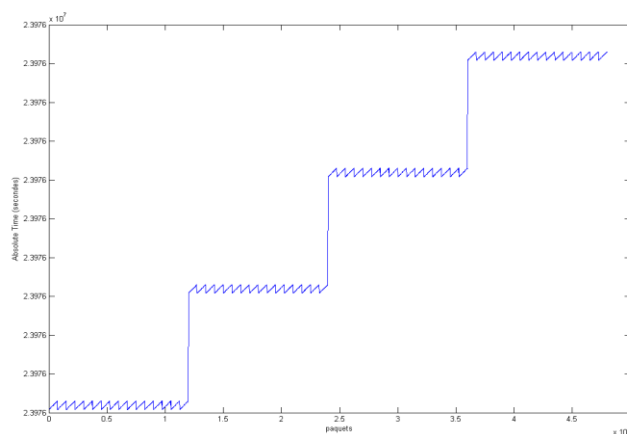


Figure 18 - Evolution de l'Absolute Time décodé

C'est dans de tels cas que l'utilisation d'un tel outil a du sens puisqu'il a permis de détecter d'importantes anomalies dans les données et ceci indépendamment de la façon dont elles ont été générées, ces anomalies peuvent être corrigées par la suite dans des versions ultérieures du projet. Par exemple, la valeur des angles de Scan a été signalée auprès de l'équipe sous-traitante en charge du développement du code industriel qui avait auparavant mal généré ce type de donnée, elle a donc été corrigée et finalement validée par l'outil présenté ici.

Au final, cet outil pourrait permettre de valider la conformité des paquets de données L0 de n'importe quelle autre mission. De plus, il pourrait être envisagé d'avoir comme objectif la description de nos données de façon complète, pérenne et évolutive, il faudrait alors enrichir le modèle de nos paquets de données en par exemple renseignant des champs supplémentaires (que j'ai ignoré jusqu'ici car inutile pour notre problématique de décodage) comme les définitions sémantiques ou encore en respectant en modifiant les noms des éléments pour qu'ils respectent les normes CCSDS. De plus, le décodage des valeurs des pixels des images d'un Repeat Cycle complet peut prendre jusqu'à 1 heure car on a alors plusieurs giga octets de données et on n'a pas accès aux différentes fonctions de la librairie qui permettent ce décodage. On a, dans

notre cas, choisi de pouvoir décoder 1 pixel sur N pour gagner en temps de décodage mais il pourrait être intéressant de chercher d'autres solutions. Par exemple, j'avais commencé à tenter d'utiliser le calcul parallèle avec Open MP et l'instruction `#pragma omp parallel` qui pourrait permettre de décoder les différents fichiers de données en parallèle en utilisant les threads disponibles (cependant, la librairie East ne semblait pas être capable de gérer les ouvertures multiples de fichiers en parallèle). De plus, il existe des méthodes dans la librairie East qui semblent permettre de traiter de gros volumes de données efficacement en enregistrant des séquences d'appels, cette stratégie pourrait potentiellement accélérer l'accès aux nombreux pixels codés dans un fichier de donnée (mais pas évident car les paquets de données de notre fichier de donnée ne sont pas identiques cf. [1]). Ces quelques stratégies pourraient être un axe de travail futur pour l'optimisation du temps de décodage de données volumineuses.

V Algorithmes de détection d'alignements de points

L'instrument LI (cf. II.2) a pour but de détecter les éclairs, il analyse le signal de chaque pixel de chaque OC (cf. Figure 13) et détecte les pixels dépassant le seuil radiométrique de déclenchement de la foudre, les données de ces événements déclenchés sont ensuite envoyées au sol. Cependant, les pixels détectés (DTs) par le système ne correspondent pas toujours réellement à des éclairs. En effet, plusieurs phénomènes font que les pixels détectés sont en fait de fausses alarmes (FTs). On peut citer par exemple le bruit RTS⁸ qui est un bruit électronique qui se produit aléatoirement dans les semi-conducteurs, l'effet "ghost" qui est la conséquence de lumières réfléchies/réfractées non voulues apparaissant dans le plan de l'image, le phénomène "jitter" (sautillement) qui résulte à cause de micro vibrations du satellite et fait varier le champ de vue de l'instrument. Par conséquent, des méthodes spécifiques à ces phénomènes sont développées afin de filtrer un maximum de ces fausses alarmes.

On s'intéresse particulièrement au filtre qui traite les particules cosmiques. Des particules à hautes énergies s'impactant sur le capteur peuvent être détectées comme des DTs car elles produisent un signal avec une forte radiométrie. Ces impacts se produisant avec des angles obliques dans la majorité des cas génèrent de longues trainées avec une plus grande régularité et à une plus grande échelle que les pixels déclenchés par de vrais éclairs. D'après les expériences faites en laboratoire, ces trainées sont plus épaisses qu'un pixel et que cette épaisseur n'est pas constante, c'est pourquoi il a été supposé de considérer de longues trainées de forme trapézoïdale (cf. Figure 19). De plus, comme l'impact d'une particule est un événement rare, il a été raisonnablement supposé que seulement quelques particules peuvent s'impacter sur une seule image. L'algorithme choisi pour éliminer ces trainées utilise la transformée de Hough, c'est une technique qui sert à reconnaître des formes et notamment des lignes. Cette méthode est universellement utilisée de nos jours et est particulièrement populaire dans la communauté du traitement d'images et plusieurs implémentations existent dans la plupart des langages de programmation, elle a donc été implémentée et testée dans un travail ultérieur par les équipes de Thales Alenia Space pour filtrer ces alignements.

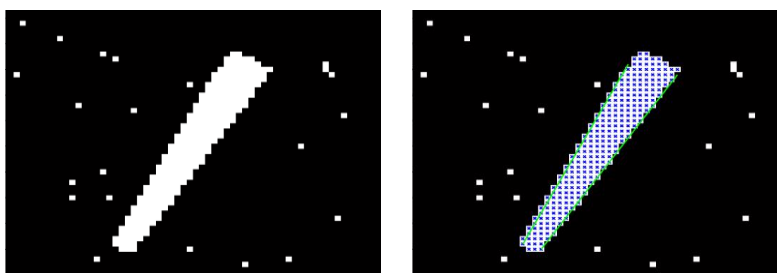


Figure 19 - Exemple d'une détection de trainée par la transformée de Hough que pourrait générer une particule cosmique. [16]

⁸ Random Telegraph Signal

V.1 Transformée de Hough

L'idée de la transformation de Hough est de représenter un point de contour de l'image dans un nouvel espace de paramètres. En effet, il est possible de caractériser une droite par deux paramètres et donc en considérant l'ensemble des droites passant par un point, l'image de cet ensemble dans cet espace de paramètres est alors une courbe. Si des points sont colinéaires, alors toutes les courbes de l'espace de paramètres se coupent au point représentant la droite en question. En pratique, on utilise les paramètres polaires (ρ, θ) pour caractériser⁹ une droite où ρ est la distance de la droite à l'origine du repère et θ est l'angle que fait la perpendiculaire à la droite avec l'axe x . Les coordonnées (x, y) d'un point appartenant à cette droite vérifient l'équation « normale » : $\rho = x \cdot \cos(\theta) + y \cdot \sin(\theta)$. Dans un premier temps, la méthode consiste à détecter les contours de l'image puisqu'on suppose que si des lignes sont présentes dans l'image elles font partie des contours de cette image. On peut par exemple utiliser des techniques de mesures de gradients locaux entre les valeurs des pixels comme les filtres de Canny ou de Sobel (voir Annexe VIII.4.a). Ensuite, chacun des points identifiés va être projeté dans le plan de la transformée de Hough : pour un point (x, y) du contour, on fait correspondre une courbe $\rho = f(\theta)$ avec $f(\theta) = x \cdot \cos(\theta) + y \cdot \sin(\theta)$ et $\theta \in [0; \pi]$ et $\rho \in [-R; R]$ où R est la grande diagonale de l'image. Une courbe ρ de l'espace de Hough associée à un point $(x; y)$ représente alors chacune des droites passant par ce point $(x; y)$ dans l'image originale.

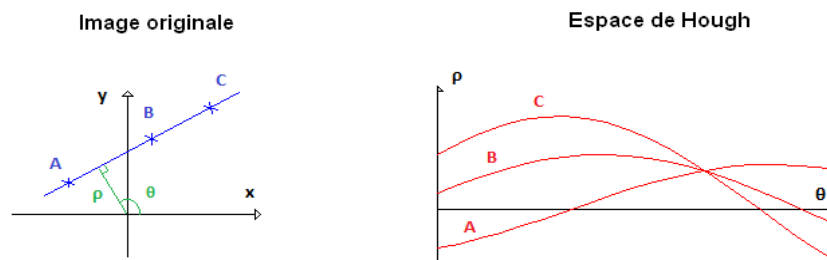


Figure 20 - Représentation de la transformée de Hough

Lorsque 2 courbes $\rho(\theta)$ correspondant à 2 points se rencontrent, le point d'intersection caractérise la droite passant par ces 2 points dans notre image. En associant à chaque point du plan (ρ, θ) une densité exprimant le nombre de courbe qui y passent, on peut détecter les alignements de point de notre image d'origine en sélectionnant les points dont cette densité est la plus importante. Cependant du fait des imperfections de l'image, les points recherchés ne sont jamais parfaitement alignés et donc les courbes ne sont pas exactement concourantes. En réalité, la méthode consiste à discrétiser l'espace de nos paramètres en petits rectangles, on compte alors pour chaque rectangle le nombre de courbe y passant. Puis, on construit une matrice dite d'accumulation qui stocke ce nombre de courbe présent dans chaque rectangle et au final, les maxima locaux de cette matrice décrivent les droites de notre image originale.

⁹ On peut également caractériser une droite par son équation cartésienne : $ax + by + 1 = 0$, les paramètres seraient alors (a, b) . L'inconvénient principale de cette approche est que les valeurs de a et b ne sont pas bornées (pour une droite verticale, $a \rightarrow \infty$).

Finalement, pour notre exemple de détection de longues trainées trapézoïdales, cette méthode a donné de bons résultats en terme de détection mais le temps de calcul réalisé est trop long (environ 9s pour une image). En fait, la complexité de cette méthode est en $O(A)$ où A est la taille de notre image. Il a donc été proposé d'étudier une autre méthode de détection d'alignement dite "*A Contrario*". Le choix de cette nouvelle méthode a été justifié par le fait, d'une part, qu'elle soit bien documentée et qu'il existe un code public open source en C pouvant servir de base de travail et d'autre part, par le fait qu'elle utilise en entrée un ensemble de points et pas une image complète (contrairement à la méthode de Hough) ce qui est plus judicieux pour notre cas d'application et pourrait donc possiblement donner de meilleurs résultats.

V.2 Méthode A Contrario pour la détection d'alignements de points 2D

Cette méthode étudiée et décrite par Lezama, Morel, Grompone et Randall dans *A Contrario 2D Point Alignment Detection* [2] s'intéresse au problème général de la détection d'alignements de points dans un bruit. Bien qu'il puisse sembler que les alignements de points sont des structures simples, la Figure 21 montre la complexité que peut avoir la détection d'un alignement de point. D'un point de vue purement factuel, le même alignement est présent dans les trois figures. Cependant, l'alignement n'est perçu comme telle par la plupart des téléspectateurs que dans le premier cas.

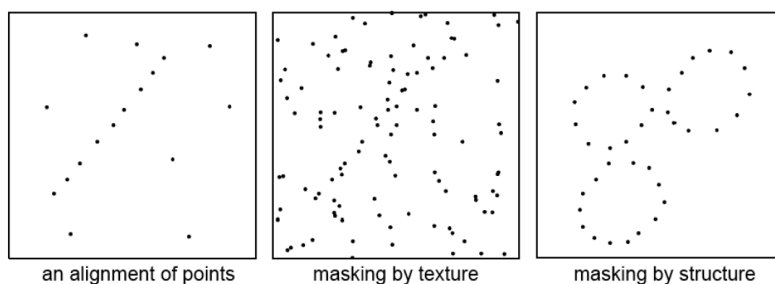


Figure 21 - Illustration de la complexité du problème de l'alignement des points [11]

La plupart des théories d'analyse d'image qui ont pour but de trouver des structures géométriques dans l'image suppose que l'image contient de telles structures et ensuite essaie de trouver leur meilleure description. Pour un tel point de vue, une approche classique est le cadre variationnel (voir Annexe VIII.4.b) où une image est une matrice et ainsi des concepts de modélisation (géométrie, formes, régularité) de fonctions peuvent être utilisés et permettent de décrire et traiter une image donnée. L'un des principaux inconvénients de ces méthodes variationnelles est qu'elles introduisent des constantes normalisées (λ, μ, \dots) et que la segmentation¹⁰ résultante dépend donc de la valeur de ces constantes. Par exemple, pour la transformée de Hough, cette constante est la valeur du seuil pour extraire les maxima de la

¹⁰ En traitement d'image, l'opération qui a pour but de rassembler des pixels entre eux suivant des critères prédéfinis.

matrice d'accumulation, même si la transformée de Hough fournit des solutions efficaces dans de nombreuses applications, il manque un critère solide pour fixer ce seuil qui doit permettre d'extraire des alignements significatifs et pertinents. Un autre inconvénient de ce type de méthode est qu'elles supposent toujours que l'image peut être segmentée même si on a une image "bruit blanc"¹¹. Ici, les auteurs de cette méthode se sont inspirés particulièrement des méthodes qui fournissent une évaluation de la signification statistique des structures alignées détectées. Un exemple en astronomie peut illustrer l'importance d'une telle évaluation : en 1980, la découverte de plusieurs alignements très précis de quasars dans le ciel a soulevé la question d'une théorie expliquant cette présence. Ces alignements ont finalement été rejetés par une analyse statistique, d'abord par des simulations et ensuite analytiquement en montrant que des alignements d'une telle précision pouvaient facilement se produire par hasard. La méthode développée ici est dérivée de la méthodologie A Contrario proposée par Desolneux, Moisan et Morel dans *Meaningful Alignments* [3] qui est une formalisation mathématique du principe de non accidentalité (voir Annexe VIII.4.d). En résumé, une structure observée est pertinente si elle n'est que rarement le fruit du hasard. C'est-à-dire, en déterminant la probabilité qu'une structure présente dans l'image aurait pu naître par accident, on peut en conclure sur sa nature : plus cette probabilité est faible et plus la structure est susceptible d'avoir une interprétation.

Rectangles candidats

Les alignements candidats seront formés en prenant chaque paire de points et en construisant un rectangle r dont l'axe principal se termine à chaque point et avec une largeur variable. Pour définir un alignement, on s'appuie sur le critère général suivant : la structure doit être allongée, de sorte qu'un rapport minimal entre la longueur et la largeur du rectangle doit être fixé et ensuite, un nombre fixe de largeurs doit être testé en diminuant géométriquement à partir de la largeur maximale. L'implémentation peut par exemple utiliser un ratio minimum longueur/largeur de 10 et une série géométrique de 8 valeurs de largeur avec un facteur de $1/\sqrt{2}$ (le choix d'une série géométrique est justifié par l'invariance de l'échelle du problème de détection).

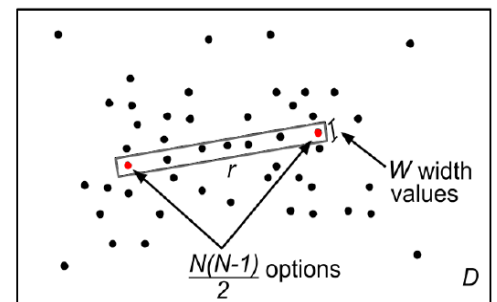


Figure 22 - Représentation schématique du rectangle évalué. Dans un domaine avec N points et en considérant W largeur possible, il y a $N(N-1)/2$ W rectangles possibles. [11]

Estimation de la densité

Dans le but d'évaluer si la configuration de points forme un alignement pertinent statistiquement, la méthode requiert d'estimer la densité de points dans la figure. Cette densité est estimée localement dans une fenêtre locale (voir Figure 23) autour du rectangle candidat car si on l'évalue globalement dans toutes l'image et qu'il y a des régions

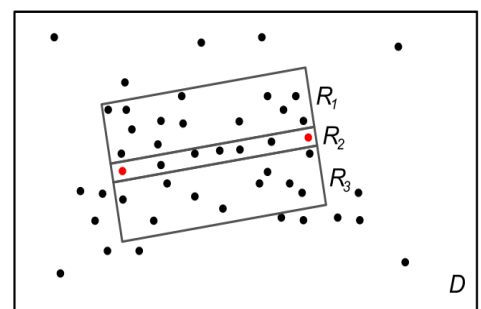


Figure 23 - Représentation schématique de la surface locale dans laquelle on estime la densité de point. [11]

¹¹ Image dont les pixels sont arrangés de façon indépendante et aléatoire selon une loi de probabilité uniforme.

avec des densités très différentes alors de fausses détections seraient créées (voir Figure 24(b)). La densité des points est donnée par le nombre de point dans le rectangle du milieu et la surface de la fenêtre locale. Le but de l'algorithme est de détecter si un nombre statistiquement surprenant de ces points sont concentrés dans le rectangle intérieur.

D'autre part, dans le but d'éviter des effets de bords qui causerait des détections dans les frontières des régions à haute densité comme montré en Figure 24(c), l'algorithme mesure séparément la densité de point de chaque côté du rectangle candidat et considère la densité de la surface locale comme étant la densité latéral maximum.

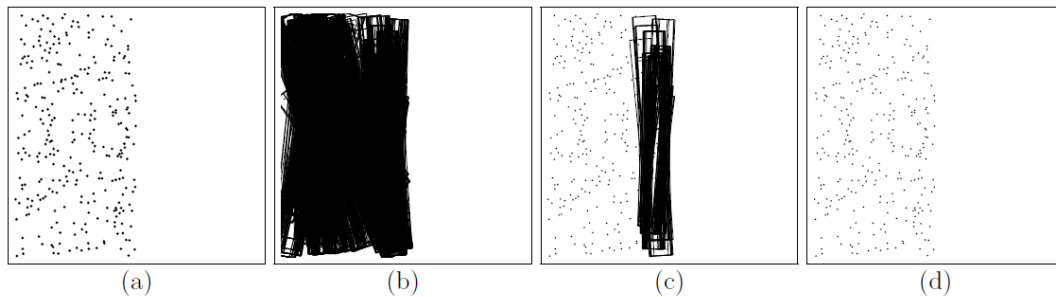


Figure 24 - (a) L'ensemble de points. (b) Alignements trouvés si on utilise une estimation globale de la densité. (c) Alignements trouvés si on utilise une estimation locale sans prendre le maximum des fenêtres latérales. (d) Pas d'alignements trouvés en utilisant la méthode proposée. [11]

Pour chaque rectangle, l'algorithme teste plusieurs tailles différentes pour cette fenêtre locale. La longueur de la fenêtre est celle du rectangle candidat et la largeur varie. La plus grande largeur testée est telle que la fenêtre locale soit un carré. Ensuite, un nombre fixé de 8 largeurs différentes sont testées dans les séries géométriques avec un facteur de $1/\sqrt{2}$.

Régularité des points

Pour détecter un alignement, outre le fait que la densité de point doit être localement plus élevée, la méthode doit également récupérer seulement les structures où les points sont régulièrement espacés. L'idée est de contrer le fait que la présence de groupe de points peut produire des zones de haute densité mais qui ne sont pas des alignements comme montré en Figure 25 . Pour éviter ça, cette étape consiste à diviser le rectangle r en c boîtes de mêmes

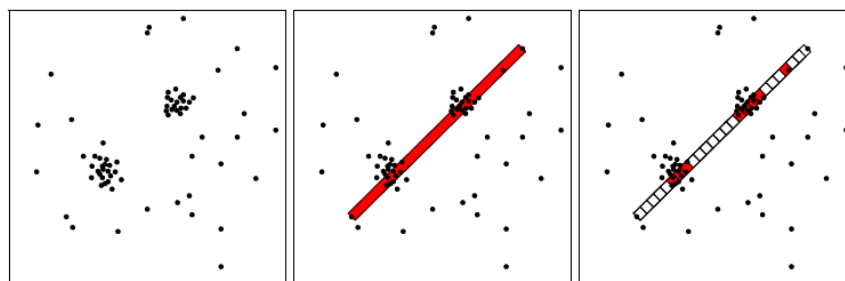


Figure 25 - A gauche : L'ensemble de points, il y a 2 groupes de points mais pas d'alignement. Au centre : Sans étudier la régularité des points, un rectangle à haute densité est détecté alors que ce n'est pas un alignement. A droite : L'algorithme divise le rectangle en boîtes et compte celles qui sont occupées évitant la fausse détection d'un alignement. [11]

tailles. Ensuite au lieu de compter le nombre de points dans le rectangle r , l'algorithme compte le nombre de boîtes occupées par au moins un point. Donc si un petit groupe de points occupe une boîte, il sera compté comme une boîte occupée, comme s'il ne s'agissait que d'un seul point. L'algorithme teste différentes valeurs pour le nombre de boîtes c dans lesquelles le rectangle est divisé. En pratique, on teste $\lfloor \sqrt{N} \rfloor$ valeurs : chaque rectangle candidat est divisé en c boîtes avec c allant de 1 à $\lfloor \sqrt{N} \rfloor$ où N est le nombre de point total de notre figure.¹²

Validation du candidats

Dans les sections précédentes, on a décrit comment est construit l'ensemble des alignements candidats. On décrit maintenant comment chaque candidat est validé comme étant une détection ou non. Comme introduit précédemment, la validation est faite en utilisant la méthodologie *A Contrario* qui consiste à supposer vraie une hypothèse nulle H_0 pour nos données où aucune détection ne devrait arriver et ensuite trouver des cas qui transgressent cette hypothèse, c'est-à-dire qui pourraient très difficilement se produire sous les hypothèses du modèle. Ici, l'hypothèse nulle H_0 est que les N points sont le résultat d'un processus aléatoire où les points sont indépendants et uniformément distribués dans le domaine. Dans les méthodes *A Contrario*, la quantité fondamentale pour valider la détection est le nombre de fausses alarmes NFA associé à un événement e et un ensemble de points. Le NFA d'un événement est notamment défini dans [4] comme l'espérance du nombre d'occurrences de cet événement. Étant donné une hypothèse *A Contrario* H_0 pour l'ensemble de points dans le domaine, ce nombre NFA d'un événement e est interprété comme une borne sur le nombre d'occurrences attendu de l'événement sous l'hypothèse H_0 , on a $NFA = N_{tests} \cdot \mathbb{P}_{H_0}(e)$ où N_{tests} est le nombre d'événements testés. Dans notre cas d'alignement de point, le NFA associé à un rectangle r et à un ensemble de point x représente une borne sur le nombre attendu de rectangles qui contiennent suffisamment de points par rapport à la densité de la fenêtre locale et régulièrement espacés pour être aussi rare que r sous l'hypothèse d'une répartition de point aléatoirement distribués. Quand le NFA associé à un rectangle est grand, ça signifie que ce rectangle est aussi attendu sous le modèle *A Contrario* (sous un ensemble de points aléatoirement répartis) et donc qu'il n'est pas assez significatif pour être considéré comme un alignement. D'un autre côté, si le NFA est très faible alors ce rectangle est rare, en d'autres mots il qui est peu susceptible de se produire par hasard dans le cadre du modèle *A Contrario* (un ensemble de points aléatoires). Dans notre cas d'application, le NFA vaut :

$$NFA(r, R, c, \mathbf{X}) = \frac{N(N-1)}{2} \cdot W \cdot L \cdot C \cdot \mathcal{B}(c, b(r, c, \mathbf{X}), p_0(r, R, c))$$

Le détail de ces résultats ainsi que les différentes grandeurs sont détaillés en annexe VIII.4.f. On veut donc garder ces alignements qui sont statistiquement significatifs : ceux qui ont un NFA faible. L'algorithme calcule donc le NFA pour chaque rectangle testé et valide seulement ceux dont le NFA est faible, il convient néanmoins de fixer un seuil de rareté ϵ . Comme montré dans [4], le résultat de la détection n'est pas très sensible à la valeur de ϵ et pour la méthode il a été choisi (suivant [4]) de fixer $\epsilon = 1$ ce qui correspond à accepter une moyenne d'au plus une fausse détection pour un ensemble de donnée dans le modèle *A Contrario*.

¹² En réalité, l'implémentation part en réalité de $c = n/2$, où n est le nombre de points à l'intérieur de l'alignement et va jusqu'à $c = \min(2n, N)$. Ce n'est qu'une accélération heuristique et n'affecte pas les performances du détecteur.

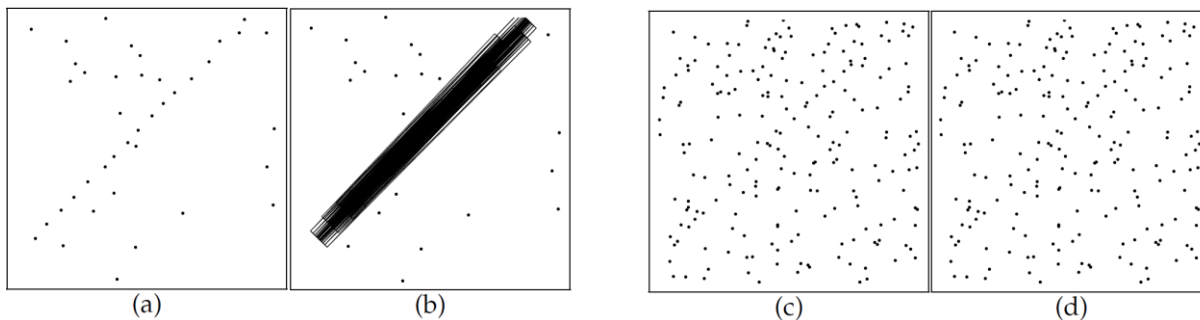


Figure 26 - Résultats produit par la méthode jusqu'ici. (a) et (c) sont les données d'entrées et contiennent le même ensemble de points alignés mais des points de bruits différents. (b) et (d) sont les résultats correspondants, chaque détection est représentée par un rectangle. [11]

Dans la Figure 26, en (d) les points alignés sont toujours présents mais très difficilement perceptibles, l'algorithme gère correctement ce phénomène de masquage et ne produit aucune détection. En (b), l'algorithme détecte correctement l'alignement évident mais on note que de multiples et redondants sont détectés. Dans la prochaine section, on explique comment prendre en compte ce phénomène.

Redondance

Comme illustré en Figure 26, la méthode jusqu'ici détecte des alignements redondants. Étant donné un alignement très significatif, plusieurs petits et larges rectangles chevauchants l'alignement principal sont détectés comme tout aussi judicieux. Ce phénomène de redondance peut impliquer des points qui font partie de l'alignement principal aussi bien que des points de fond proche de l'alignement, comme illustré en Figure 27. La question est de comment détecter le meilleur rectangle tout en expliquant et masquant les détections redondantes. Une idée similaire a été proposée par [4] sous le nom de "principe d'exclusion" : la structure observée la plus significative (celle qui a le plus petit NFA) est gardée comme détection valide. Ensuite, tous les points qui font partie de cet ensemble lui sont assignés et les structures candidates restantes ne peuvent plus utiliser ces points, le NFA des structures restantes est recalculé sans compter ces éléments exclus et ainsi les structures redondantes perdent la plupart de leur éléments et ne sont plus détectées comme des alignements.

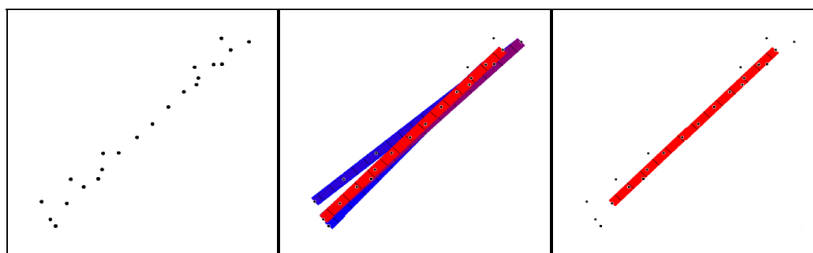


Figure 27 - Illustration de la détection redondante. A gauche : les points d'entrée. Au centre : les alignements trouvés par l'algorithme sans traiter cette redondance, en rouge la détection la plus significative et en bleu celles les moins significatives. A droite : résultat après traitement de la redondance. [2]

Complexité calculatoire

Le but principal lors de la création de cette méthode telle qu'elle a été abordée par ces auteurs était d'abord d'apporter une bonne solution au problème, en négligeant les questions d'efficacité. La méthode proposée consiste en une recherche exhaustive des candidats et leur validation incluant la partie d'élimination de redondance. Le nombre de tests effectué est le nombre théorique N_{tests} décrit plus haut, pour obtenir la complexité totale ce nombre doit être multiplié par la complexité d'un seul test. Pour calculer le NFA, on a besoin d'évaluer si chacun des N points appartient à une boîte ou non, donc la complexité d'un seul test est proportionnel au nombre de points N . Finalement, d'après l'équation page 24, la complexité totale de la recherche exhaustive est $O(\sqrt{N} N^3)$ où \sqrt{N} correspond à l'ensemble de nombre de boîtes testé C . Et dans l'étape d'élimination des redondances, les candidats supposés valides (n_{val}) doivent être comparés à ceux déjà sélectionné comme détections finales (n_{out}) et calculer leur NFA (N), cette étape finale a donc une complexité de $O(n_{val}n_{out}N)$, généralement bien plus petit que la précédente.

Les auteurs de la méthode proposent quelques solutions pour accélérer l'étape de recherche des candidats qui est celle la plus gourmande en temps de calcul. Dans un premier temps, cette recherche est hautement parallélisable, conduisant à une amélioration presque linéaire en utilisant une plateforme multiprocesseur (par exemple Open MP). Ensuite, il est largement possible d'améliorer les routines de comptage des points à l'intérieur d'un rectangle, deux techniques possibles sont l'utilisation de buckets ou d'images intégrales (structures de données ayant des propriétés spécifiques). Enfin, la recherche exhaustive peut être remplacée par une heuristique comme RANSAC [5]. Une approche utilisant des versions accélérées de la transformation de Hough peut être particulièrement importante lorsque le nombre de points augmente. L'algorithme de Figueiredo et Jain [6] peut également être utilisé comme heuristique efficace pour proposer des candidats.

Tests et Résultats

Après avoir compris le fonctionnement de la méthode, le but du travail était dans un premier temps d'implémenter cet algorithme en Matlab pour ensuite pouvoir comparer ses performances à la méthode utilisant la transformée de Hough qui a été implémenté également en Matlab dans un travail ultérieur. Il est en effet nécessaires d'utiliser le même langage de programmation et d'exécuter les tests sur la même machine pour que la comparaison entre les 2 méthodes soit cohérente. D'abord, la Figure 28 montre que l'algorithme implémenté de la méthode A Contrario a bien le comportement attendu, il détecte des alignements de points perceptibles parmi des points aléatoirement répartis. On réalise quelques cas tests présents dans l'article [2] qui permettent de vérifier que le code réécrit en Matlab retourne les mêmes résultats que le code en C initialement écrit par les auteurs de la méthode.

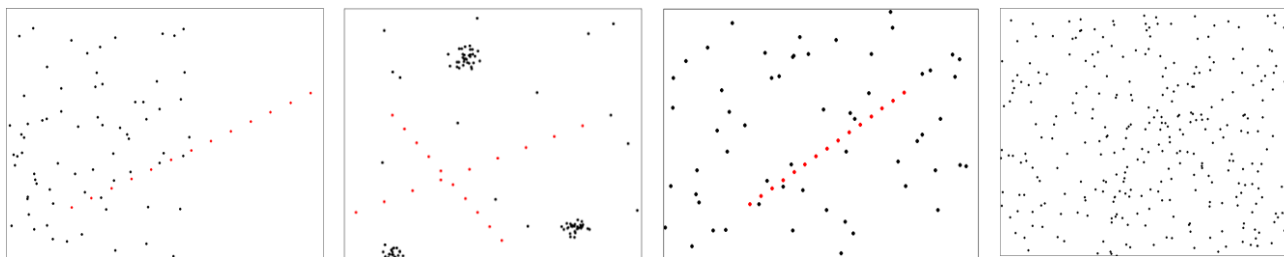


Figure 28 - Exemples de détection d'alignement avec la méthode *A Contrario* implémentée en Matlab. En rouge, la sortie de l'algorithme : les points détectés comme faisant partis d'un alignement. En noir, les autres.

On souhaite désormais, pour le même jeu de données d'entrées, comparer les performances des 2 méthodes dans le cadre¹³ qui nous intéresse c'est-à-dire des images envoyées par l'instrument LI. Ces images sont de taille 1100*1170 pixels et elles sont composées de points qui correspondent chacun à un DT de l'instrument. Sachant qu'on s'attend à avoir entre 70 et 90 DTs par frame et que les particules cosmiques sont un évènement plutôt rare, on testera donc pas plus de 5 alignements par frame. Et comme expliqué précédemment, ces particules cosmiques génèrent de longues trainées (l'alignement à détecter) de forme trapézoïdale d'une longueur et d'une épaisseur minimales et maximales (cf. Figure 19). Si on applique donc la méthode *A Contrario* à un tel cas, on se rend compte que son temps d'exécution est bien supérieur à celui de la méthode de la transformée de Hough pour un exemple où il y a 30 points bruits et une trainée composée de 50 points (cf. Figure 29).

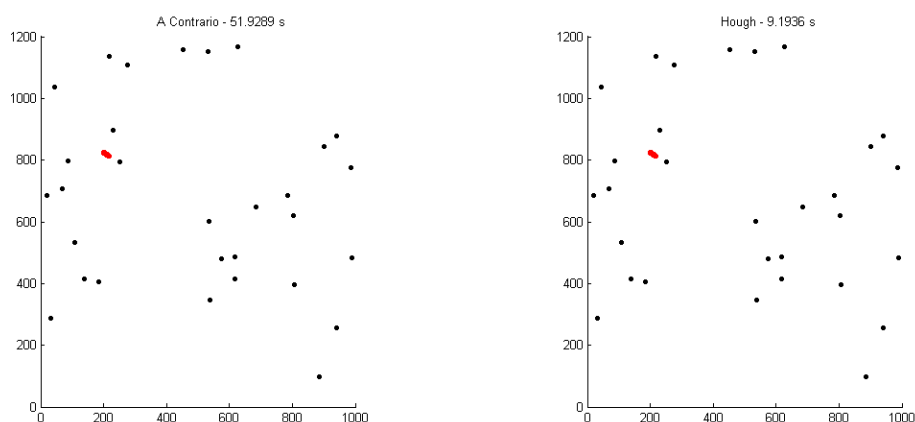


Figure 29 - Exemple de comparaison des 2 méthodes. La méthode *A Contrario* détecte l'alignement en 50 s alors que celle de Hough en 9 s.

Dans la Figure 31 et la Figure 30, on retrouve bien les comportements prévus par la complexité calculatoire explicitée dans les sections précédentes. La Figure 30 montre le temps de calcul lorsqu'il n'y a aucun alignement dans l'image puisqu'il faut que l'algorithme doit filtrer toutes les images mêmes celles qui n'ont pas d'alignements, sachant que dans un cas réel on

¹³ Dans la réalité, les images lues par le LI ne sont pas de simples points aléatoirement répartis, il y a des structures en clusters. Cependant ces clusters sont de toutes petites tailles devant nos trainées trapézoïdales c'est pourquoi on peut simplifier les hypothèses en considérant de simples points.

s'attend à plus de 60 points (les DTs), la méthode A Contrario résout le problème bien plus lentement que la méthode à transformée de Hough. Pour le cas où il y a effectivement un alignement à détecter dans l'image, on s'attend à des trainées trapézoïdales composées d'environ 50 points, d'après la Figure 31 on se trouve également bien loin des performances de la transformée de Hough.

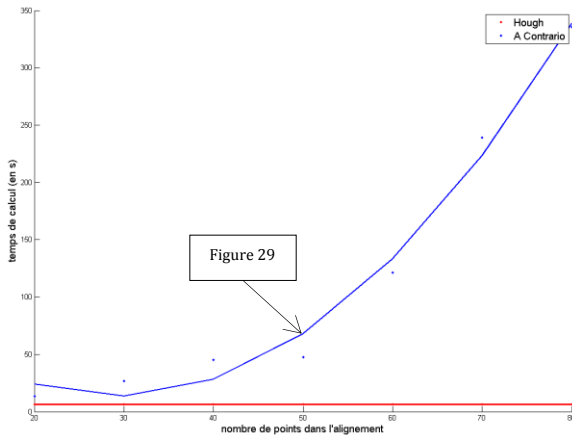


Figure 31 - Evolution du temps de calcul de la méthode A Contrario en fonction du nombre de points dans l'alignement (et 30 points bruits).

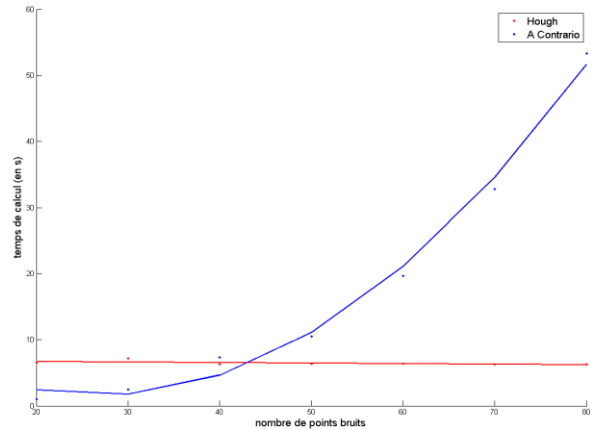


Figure 30 - Evolution du temps de calcul de la méthode A Contrario en fonction du nombre de points bruits (avec aucun alignement).

Les résultats obtenus ne sont donc pas satisfaisants pour notre exemple d'application. Il a donc été décidé d'essayer de réduire ce temps de calcul en réduisant un maximum le nombre d'alignement candidats testé dans la méthode A Contrario. En effet, la méthode a été conçue pour des cas généraux et donc, comme expliqué avant, pour chacune des paires de points d'entrée, elle vérifie la densité et la régularité de l'alignement. On a donc, en pratique, plusieurs boucles imbriquées¹⁴ : d'abord sur toutes les paires de points, ensuite sur la largeur de l'alignement, puis sur la taille de la fenêtre locale (densité), enfin sur le nombre de boîtes à tester (régularité) et on boucle encore une fois sur le nombre de points pour compter s'ils appartiennent à une boîte ou non. Mais pour notre cas présent, on a des informations sur l'alignement recherché, en effet on sait que les particules cosmiques génèrent un alignement d'une longueur bornée, d'une épaisseur bornée et régulier (chaque pixel voisin de la trainée générée est un DT). On peut donc traiter seulement les paires de points qui sont séparés d'une distance comprise entre la longueur maximum et minimum que peut générer une particule cosmique. Et de la même façon, on peut tester une unique largeur d'alignement et un seul nombre de boîtes pour réduire le nombre d'alignement candidat et ainsi gagner en temps de calcul. De plus comme on change le nombre de tests N_{tests} effectués, on adapte également la valeur du NFA pour valider les candidats testés. L'algorithme modifié de la méthode A Contrario réussit alors à détecter les alignements bien plus rapidement que la méthode de la transformée de Hough et mêmes pour des cas où il y a plus de points bruits et plus d'alignements.

¹⁴ Voir algorithmes en annexe VIII.4.g

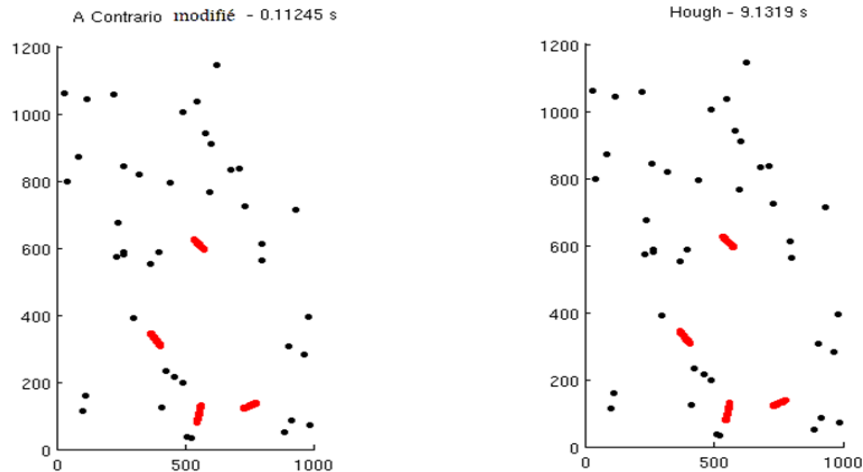


Figure 32 - Exemple de comparaison de la méthode A Contrario modifié et celle de Hough. La méthode A Contrario modifiée détecte les alignements en ~0.1 s alors que celle de Hough en 9 s.

La Figure 33 valide que la méthode A Contrario modifiée est toujours plus rapide que la transformée de Hough, le tableau contient les temps de calcul en secondes pour détecter les trainées trapézoïdales (de 1 à 5) avec des points bruits (de 40 à 80) avec en bleu ceux de la méthode A Contrario et en rouge ceux de la méthode de la transformée de Hough. Pour tous les cas effectués dans ce tableau, chacun des alignements a correctement été détectés (pour Hough et A Contrario).

Points bruits Particules cosmiques	40		50		60		70		80	
1	0.1329	9.2264	0.0239	9.1329	0.0760	9.1352	0.0304	9.0867	0.0137	9.0607
2	0.0413	9.1725	0.1070	9.1604	0.1034	9.1646	0.0435	9.1245	0.1186	9.3161
3	0.1589	9.3854	0.0611	9.5440	0.3178	9.5379	0.2143	9.2134	0.4367	9.2164
4	0.7339	9.4250	0.2539	9.2385	0.1926	9.2572	0.3213	9.2695	0.9342	9.1909
5	0.9721	9.3629	0.7512	9.2204	1.3915	9.2539	0.8681	9.3134	0.5883	9.2147

Figure 33 – Comparaison des temps de calcul (en secondes) des 2 méthodes. En rouge, la méthode A Contrario modifié. En bleu, la méthode de la transformée de Hough. Les 2 méthodes détectent dans tous les cas 100% des alignements.

En modifiant, la méthode A Contrario développée par J. Lezama, G. Randall, J. M. Morel et R. Grompone dans [2] on a pu obtenir de meilleures performances que l'algorithme auparavant implémentée utilisant la méthode de la transformée de Hough pour le cas précis où les alignements à détecter sont générés par des particules cosmiques et donc ont certaines propriétés. L'algorithme modifié ne résout pas les cas généraux de problèmes d'alignements de points, il utilise les mêmes arguments statistiques pour appliquer la résolution à notre cas d'application de détection de longues trainées trapézoïdales contrairement aux auteurs qui avaient initialement conçus la méthode pour détecter tout type d'alignement sans aucune informations *a priori*. L'utilisation de cette méthode pourrait faire gagner un temps non négligeable à la chaîne de traitement qui filtrent les fausses détections du LI. En effet, les prototypes Matlab filtrant les fausses détections du LI sont testés sur des données de référence correspondant à 100 images ce qui signifie qu'en utilisant la méthode de la transformée de Hough la détection des particules cosmiques prend environ 15 minutes alors qu'en utilisant la méthode A Contrario modifiée cette détection prend quelques secondes.

Méthodes Caractéristiques	Transformée de Hough <i>Thales</i>	A Contrario générale	A Contrario modifié
Types d'alignement	Particules cosmiques	Tous	Particules cosmiques
Temps de calcul	Moyen	Lent	Rapide
Complexité	$\mathcal{O}(A)$	$\mathcal{O}(\sqrt{N} N^3) + \mathcal{O}(n_{val}n_{out}N)$	$\mathcal{O}(.)$

Figure 34 – Résumé de la comparaison des différentes méthodes

VI Conclusion

Thales Alenia Space est un acteur technologique mondial majeur dans le domaine des satellites et des infrastructures orbitales. Les domaines d'étude qui y sont abordés sont nombreux, j'ai pour ma part découvert une partie du programme MTG au sein du service Qualité Image et de l'équipe en charge de l'IQT qui permet l'évaluation de la performance de la qualité de l'image ce qui demande de nombreuses compétences techniques et de bonnes connaissances scientifiques. Dans une optique de validation et optimisation de 2 problématiques intervenant en support de l'IQT, mon stage m'a permis d'appréhender et découvrir les difficultés que peuvent rencontrer l'ingénieur à Thales Alenia Space.

La première problématique de mon stage a été le développement d'un outil capable de décoder des données L0, de valider la conformité de la structure des paquets et de vérifier leur contenu scientifique. Le développement d'un tel outil a bien permis de valider et vérifier certaines données scientifiques et également de relever des anomalies sur la façon dont elles étaient encodées. Les principales difficultés rencontrées furent d'abord la prise en main d'un logiciel inconnu de l'équipe qu'il a fallu installer puis configurer et ensuite, lors du développement de l'outil de validation en C avec la librairie East, il a pu être parfois compliqué de savoir si la donnée à décoder était lue de la mauvaise façon par mon outil ou bien si elle était déjà mal codée en amont dans le fichier binaire généré par l'entreprise sous-traitante. Une fois ce logiciel du CNES bien pris en main, on comprend qu'il pourrait servir à décoder et valider efficacement n'importe quelles autres données d'autres programmes spatiaux ainsi que décrire complètement la donnée. Cependant un code en C s'appuyant sur la librairie fournie est nécessaire pour pouvoir interpréter et on est alors dépendant des routines utilisées. La seconde problématique de mon stage a été l'étude et les tests d'une méthode alternative dite *A Contrario* pour la détection d'alignements de points qui se base notamment sur le principe de non accidentalité et des arguments statistiques. En s'appuyant sur les hypothèses propres à notre problème et en adaptant l'algorithme de cette méthode, on a pu obtenir des résultats de tests satisfaisants puisque l'algorithme résout notre problème en un temps de calcul inférieur à la méthode originalement choisie par l'équipe pour résoudre ce problème. Cette méthode pourrait donc être implémentée dans l'IQT pour détecter les particules cosmiques au lieu de la méthode de la transformée de Hough. De plus, ce type de méthodologie *A Contrario* est intéressant puisqu'elle ne se limite pas seulement à la détection d'alignement de point mais peut être appliquée dans divers cas comme pour la détection de contours ou de clusters [4].

Finalement ce stage m'aura permis d'acquérir de l'expérience dans de nombreux domaines. J'ai en effet pu découvrir le monde de l'ingénierie spatiale et d'un programme Météo européen en étant intégré dans l'équipe Qualité Image et en participant aux réunions hebdomadaires. De plus, la lecture de la documentation technique du projet m'a permis d'appréhender et comprendre les problématiques propres à l'entreprise. Enfin, j'ai pu acquérir et appliquer des compétences technique puisque j'ai eu à développer du code en C et en Matlab, utiliser l'environnement Linux, comprendre une méthode reposant sur des concepts de mathématiques appliquées.

VII Bibliographie

- [1] CNES, «Guide d'utilisation EAST pour utilisateurs C,» [En ligne]. Available: <https://logiciels.cnes.fr/fr/content/best>.
- [2] J. Lezama, R. Grompone, J. M. Morel et G. Randall, «A Contrario 2D Point Alignment Detection,» 6 03 2014. [En ligne]. Available: <https://hal.inria.fr/hal-00956596>.
- [3] A. Desolneux, L. Moisan et J. M. Morel, «Meaningful Alignments,» *International Journal of Computer Vision*, pp. 7-23, 2000.
- [4] A. Desolneux, L. Moisan et J. Morel, «From Gestalt Theory to Image Analysis, a Probabilistic Approach,» *Springer*, 2008.
- [5] M. A. Fischler et R. C. Bolles, «Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography,» *Communications of the ACM*, pp. 381-395, 1981.
- [6] M. Figueiredo et A. K. Jain, «Unsupervised learning of finite mixture models,» *Transactions on Pattern Analysis and Machine Intelligence*, pp. 381-396, 2002.
- [7] R. Fisher, S. Perkins, A. Walker et E. Wolfart, «Sobel Edge Detector,» 2003. [En ligne]. Available: <http://homepages.inf.ed.ac.uk/rbf/HIPR2/sobel.htm>.
- [8] J. M. Morel et S. Solimini, *Variational Methods in Image Segmentation with seven image processing experiments*, Birkhäuser, 1995.
- [9] R. Duda et P. Hart, *Use of the Hough Transformation to Detect Lines and Curves in Pictures*, Comm. ACM, 1972.
- [10] P. Hough, *Machine Analysis of Bubble Chamber Picture*, Proc. Int. Conf. High Energy, 1959.
- [11] J. Lezama, G. Randall, J. M. Morel et R. Grompone, «An Unsupervised Point Alignment Detection Algorithm,» 15 12 2015. [En ligne]. Available: <http://dx.doi.org/10.5201/ipol.2015.126>.
- [12] F. Crow, «Summed-area tables for texture mapping,» *ACM SIGGRAPH Computer Graphics*, pp. 207-212, 1984.
- [13] N. Kiryati, Y. Eldar et A. M. Bruckstein, «A probabilistic hough transform,» *Pattern recognition*, pp. 303-316, 1991.
- [14] G. Kanizsa, *Organization in Vision*, Praeger, 1979.

- [15] P. Viola et M. Jones, «Rapid object detection using a boosted cascade of simple features,» *CVPR*, 2001.
- [16] LEONARDO (L. Guinti), «MTG LI - Overall L1B Processing Add [Non-Distributable],» 2018.
- [17] I. Erlich et CNES, «BEST User Manual,» 5 Février 2010. [En ligne]. Available: <https://logiciels.cnes.fr/fr/content/best>.
- [18] «Eumetsat,» [En ligne]. Available: <https://www.eumetsat.int/website/home/Satellites/FutureSatellites/MeteosatThirdGeneration/index.html>.

VIII Annexes

VIII.1 Déroulement du Stage

Mon travail effectué durant ces 6 mois de stage m'a permis d'aborder plusieurs thématiques différents et d'acquérir de l'expérience dans de nombreux domaines. J'ai appréhendé petit à petit le fonctionnement du programme MTG et ma problématique de stage. A propos du déroulement du travail du stage, j'ai essayé d'être un maximum autonome mais je devais régulièrement poser des questions à ma tutrice pour trouver les informations nécessaires. De plus, on avait de manière hebdomadaire des réunions avec ma tutrice et/ou avec l'équipe IQT pour faire un point sur l'avancement du travail et la prévision de ce qui restait à faire. Les réunions avec ma tutrice m'ont été très instructives puisqu'elle ont permis de mieux organiser mon travail et de me rendre compte que ce genre de réunion est très utile pour avoir à expliquer son avancement et avoir un regard nouveau sur la problématique. D'un autre côté, les réunions avec l'équipe m'ont permises de comprendre le fonctionnement et le rôle d'une équipe de 8 personnes dans un projet à grande échelle et les difficultés qu'elle rencontre.

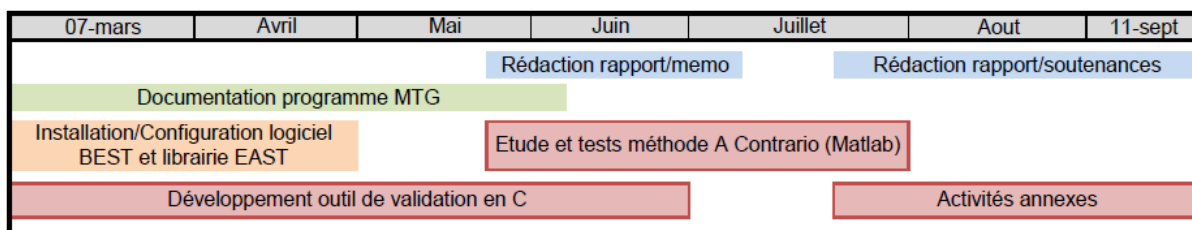


Figure 35 – Déroulement du stage

THALES ALENIA SPACE OPEN

Ce document ne peut être reproduit, modifié, adapté, publié, traduit d'une quelconque façon, en tout ou partie, ni divulgué à un tiers sans l'accord préalable et écrit de Thales Alenia Space.

© 2018, Thales Alenia Space

Modèle : 83230326-D0C-TAS-FR/003

VIII.2 Programmes Météosat

Caractéristique	METEOSAT	MSG	MTG-I	MTG-S
Statut	Retirés	Opérationnels	En développement	
Date lancement	1977-1997	2002-2015	2020-	2021-
Nombre satellites	7	4	4	2
Masse au lancement (à sec)	696 kg (320 kg)	2 040 kg	3 600 kg	3 800 kg
Énergie	200 W	600 W (fin de vie)	2 kW	
Contrôle d'attitude	Spinné		Stabilisé 3 axes	
Principaux instruments	Radiomètre MVIRI 3 canaux	Radiomètre SEVIRI 12 canaux	Radiomètre FCI 16 canaux	Sondeurs infrarouge IRS et ultraviolet UVN
Performances	Résolution 2,5 km à 5 km	Résolution de 1 à 3 km de tout l'hémisphère avec un cycle de 15 min et capacité d'observations locales aussi rapide que 2.5 min.	Résolution spatiale 0.5, 1 et 2 km.	Résolution de 8 km
	Image complète de l'hémisphère toutes les 30 minutes			
			Image complète de tout l'hémisphère en 10 min et image locale rapide en 2.5 min.	
Durée de vie contractuelle	5 ans	7 ans (MSG-1 lancé en août 2002 est encore en fonction)	8,5 ans consommables pour 10,5 ans	

Figure 36 - Comparaison des caractéristiques des 3 générations de satellites METEOSAT

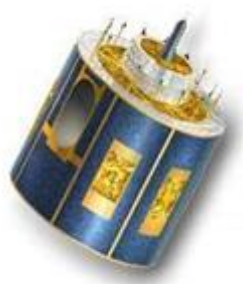


Figure 37 - Exemple d'engins spatial spinné

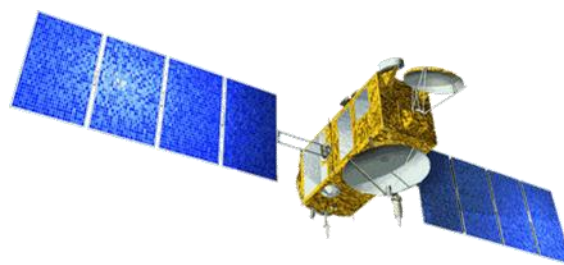


Figure 38 - Exemple de satellite à stabilisation 3-axes

THALES ALENIA SPACE OPEN

Ce document ne peut être reproduit, modifié, adapté, publié, traduit d'une quelconque façon, en tout ou partie, ni divulgué à un tiers sans l'accord préalable et écrit de Thales Alenia Space.

© 2018, Thales Alenia Space

Modèle : 83230326-D0C-TAS-FR/003

VIII.3 Exemple de décodage de données

Ci-dessous un exemple de l'utilisation de la librairie East pour décoder la valeur des angles de Scan NS d'un fichier de donnée binaire. En pratique, dans mon outil de validation, certaines lignes sont en fait des routines car répétées régulièrement dans le reste du programme. De plus quand on lit un paquet de scan, on lit également les angles EW et leur Fine Time associé et il existe également des lignes de code pour la gestion des erreurs. Ces éléments ont été ignorés dans l'exemple ci-dessous pour plus de clarté sur le déroulement des appels à la librairie East.

```
// declaration of variables
int cptSCA = 0, subtype, status, longueur, i;
OCTET *buffer;
int **NSaxis = malloc(100000 * sizeof(*NSaxis));

status = start_i(FRENCH); // East library : start services
status = select_DDR_i("dataModel.eas"); // East library : read data model

while (1) { // Loop on the number of Block = occurrence of our data model

    status = load_next_data_block_from_file_i("dataFile.bin"); // East library : read data file
    if (status == -7) //stop if it is the end of the file
        break;

    status = prepare_accesses_for_current_block_i(); // East library : prepare data block

    //get the Subtype
    status = get_data_entity_i("PDF.DFH.SUBT",&buffer,&longueur,INTEGER_32); // East library : get data
    subtype = *buffer;
    free(buffer);

    if (subtype == 31) { // if subt = 31, it is a Scan Operation packet

        for (i=0; i< 50 ; i++) { //loop on our 50 angles

            // get the NS scan angle
            char chaineNS[500];
            char tempNS[50];
            strcpy(chaineNS,"PDF.SDF.SCA.SCAN_OPERATION.BLOCK"); // build string needed
            sprintf(tempNS,"(%d).",i+1);
            strcat(chaineNS,tempNS);
            strcat(chaineNS,"NS_AXIS.ANGLE_NS");
            status = get_data_entity_i(chaineNS,&buffer,&longueur,INTEGER_32); // East library : get data
            NSaxis[cptSCA].angle[i] = *buffer;
            free(buffer);
            NSaxis[cptSCA].angle[i] = a_NSaxis[cptSCA].angle[i]*360/pow(2,25); // apply a conversion formula

        } //end for

        cptSCA++;

    } //end if

} //end while
```

Figure 39 - Exemple de code en C de décodage de données.

THALES ALENIA SPACE OPEN

Ce document ne peut être reproduit, modifié, adapté, publié, traduit d'une quelconque façon, en tout ou partie, ni divulgué à un tiers sans l'accord préalable et écrit de Thales Alenia Space.

© 2018, Thales Alenia Space

Modèle : 83230326-D0C-TAS-FR/003

VIII.4 Notions complémentaires sur la partie V.2

VIII.4.a Filtre de Sobel [7]

Ce filtre est un opérateur utilisé pour la détection de contours. Il calcule le gradient de l'intensité de chaque pixel. Ceci indique la direction de la plus forte variation du clair au sombre, ainsi que le taux de changement dans cette direction. On connaît alors les points de changement soudain de luminosité, correspondant probablement à des bords, ainsi que l'orientation de ces bords. En chaque point, le gradient pointe dans la direction du plus fort changement d'intensité, et sa longueur représente le taux de variation dans cette direction. Le gradient dans une zone d'intensité constante est donc nul. Au niveau d'un contour, le gradient traverse le contour, des intensités les plus sombres aux intensités les plus claires.

L'opérateur utilise des matrices de convolution¹⁵. La matrice (généralement de taille 3×3) subit une convolution avec l'image pour calculer des approximations des dérivées horizontale et verticale. Soit A l'image source ainsi que G_x et G_y deux images qui en chaque point contiennent des approximations respectivement de la dérivée horizontale et verticale de chaque point. On trouve G_x et G_y de telle façon :

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} * A \quad \text{et} \quad G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} * A$$

En ayant les gradients horizontaux G_x et verticaux G_y , on obtient la norme du gradient

$$G = \sqrt{G_x^2 + G_y^2}$$

On peut calculer la direction du gradient comme suit : $\theta = \text{atan2}(G_x, G_y)$ où par exemple $\theta = 0$ pour un contour vertical plus foncé à gauche.

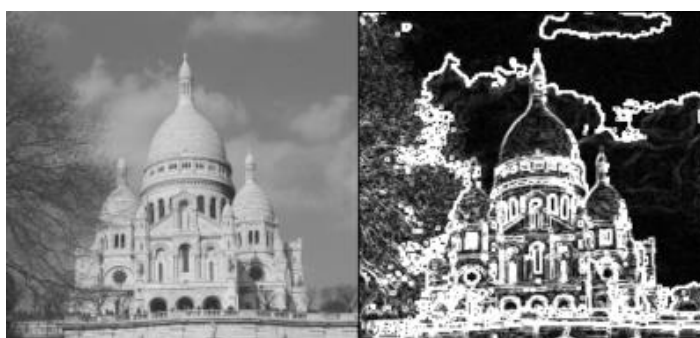


Figure 40 - Exemple d'utilisation du filtre de Sobel. *A gauche*, l'image originale. *A droite*, l'image après application du filtre.

¹⁵ Aussi appelées noyaux ou kernel, les matrices de convolution sont utilisées par la plupart des filtres de traitement des images. La convolution consiste à recalculer la valeur d'un pixel donné d'une image de départ en se basant sur la valeur du pixel lui-même et sur la valeur des pixels voisins.

VIII.4.b Cadre variationnel [8]

C'est le cadre mathématique qui permet de formaliser le problème de reconnaissance de forme, la segmentation. L'idée générale est de minimiser une fonctionnel du genre :

$$F(u, g) + R(u)$$

Où g est l'image donnée définie sur un domaine $\Omega \subset \mathbb{R}^2$ et u est l'image estimée. $F(u, g)$ est un terme de fidélité et $R(u)$ est un terme de régularité. F et R définissent un modèle *a priori*. Ci-dessous, 2 exemples de modèles utilisant le cadre variationnel.

- **Le modèle de Mumford-Shah** définit le problème de la segmentation comme un problème de lissage/détection des bords : on cherche simultanément une "image lissée par morceaux" $u(x)$ avec un ensemble K de discontinuités abruptes, les "bords" de g . Le but est alors de minimiser la fonction :

$$E(u, K) = \int_{\Omega \setminus K} (|\nabla u(x)|^2 + (u - g)^2) dx + \text{length}(K)$$

Le 1^{er} terme impose que u soit lisse à l'extérieur des bords, le 2nd que l'image lissée par morceaux $u(x)$ se rapproche en effet de $g(x)$, et le 3^{ème} que l'ensemble de discontinuité K ait une longueur minimale.



Figure 41 - Approximation d'image avec la fonction de Mumford-Shah. *Gauche* : L'image d'un œil. *Centre-gauche* : Les zones à gradient élevé dans l'image originale. *Centre-droite* : Les frontières dans le modèle de Mumford-Shah. *Droite* : La fonction lisse par morceau approximant l'image.

- **Le modèle Bayésien** : Dans les scènes réelles, les pixels voisins ont généralement des intensités similaires, les limites sont généralement lisses et les objets (par exemple de l'herbe, des arbres ...) ont des relations et des orientations préférées. L'approche consiste à articuler mathématiquement de telles régularités puis à les exploiter dans un cadre statistique pour en tirer des conclusions. Pour ce faire, la formulation Bayésienne utilise une distribution de probabilités pour représenter ces attentes *a priori*. Ce modèle s'appuie généralement sur la distribution de Gibbs $\Pi(\vec{x}) = Z^{-1} \exp(-U(\vec{x}))$ et sur le concept d'estimateur du maximum *a posteriori* (MAP).

VIII.4.c Inégalité de Hoeffding¹⁶

Soit X_1, \dots, X_n des variables aléatoires indépendantes telles que $0 \leq X_i \leq 1$. On pose $S_n = \sum_{i=1}^n X_i$ et $p = \mathbb{E} \left[\frac{S_n}{n} \right]$.

Alors pour $pn < k < n$,

$$\mathbb{P}[S_n \geq k] \leq \left(\frac{p}{k/n} \right)^{n \binom{k}{n}} \left(\frac{1-p}{1-k/n} \right)^{n(1-\frac{k}{n})}$$

En posant $r = \frac{k}{n}$,

$$\mathbb{P}[S_n \geq k] \leq e^{-n \left(r \log \frac{r}{p} + (1-r) \log \frac{1-r}{1-p} \right)}$$

$$\log \mathbb{P}[S_n \geq rn] \leq n r \log \frac{p}{r} + n(1-r) \log \frac{1-p}{1-r}$$

Pour simplifier le calcul de la valeur du NFA dans V.2, l'implémentation utilise cette approximation de Hoeffding au lieu de calculer la valeur réelle de $\mathcal{B}(n, k, p)$. On a en effet n variables aléatoires indépendantes de Bernoulli de paramètre p , leur somme $S = X_1 + \dots + X_n$ suit une distribution binomiale $\mathcal{B}(n, k, p)$. Et on la calcule tout simplement de la façon suivante :

```
static double log_bin(int n, int k, double p)
{
    double r = ( double ) k / ( double ) n;
    if( r <= p )
        return 0.0;
    else if( n == k )
        return k * log10(p);
    else
        return k * log10(p/r) + (n-k) * log10( (1-p)/(1-r) );
}
```

VIII.4.d Principe de non accidentalité / de Helmholtz

Ce principe tente de décrire quand la perception décide de grouper des objets selon un attribut (couleur, alignement, forme ...). Supposons qu'on a les objets O_1, O_2, \dots, O_n présents dans l'image et que k d'entre eux ont une caractéristique commune. On se demande alors : est-ce que cette caractéristique commune est le fruit du hasard ou est-elle significative ? Pour répondre à cette question, on fait l'expérience suivante : on suppose que la caractéristique commune étudiée a été aléatoirement et uniformément distribuée parmi tous les objets O_1, O_2, \dots, O_n de l'image, ensuite on suppose que la position de ces objets dans l'image est produite par la

¹⁶ Egalement appelé inégalité de Chernoff si on a des variables de Bernoulli

réalisation d'un processus aléatoire uniforme. Puis on se demande si la répartition observée est probable ou pas ? Le principe d'Helmholtz dit que si l'espérance dans l'image de la structure observée O_1, \dots, O_k est très petite, alors cette structure a un sens.

VIII.4.e Meaningful event [3]

Définition " ϵ -meaningful event" :

Nous disons qu'un événement est ϵ -significatif si l'espérance du nombre d'occurrences de cet événement est inférieure à ϵ sous l'hypothèse aléatoire A Contrario H_0 . Lorsque $\epsilon < 1$, nous disons simplement que l'événement est significatif. Cette définition s'appuie notamment sur le principe de Helmholtz et le raisonnement qui suit.

D'après [4], On a un ensemble de données x et on souhaite détecter une structure donnée, on teste alors un ensemble de candidats N_C . On définit alors une fonction $k(c_i, x)$ pour mesurer le degré de présence de la structure recherchée dans les données x au poste candidat c_i (dans notre cas, $k(c_i, x)$ correspond au nombre de boîtes occupées). On doit maintenant spécifier un seuil κ_i pour valider c_i quand $k(c_i, x) \geq \kappa_i$.

Le nombre total de détection est alors

$$d(x) = \sum_{i=1}^{N_C} 1_{k(c_i, x) \geq \kappa_i}$$

où 1_A est la fonction indicatrice.

L'approche A Contrario requière que le seuil κ_i soit établi tel que le nombre attendu de détection $d(X)$, où X est les données aléatoire du modèle H_0 , soit inférieur ou égal à ϵ . Comme il y a plusieurs κ_i et un unique ϵ , on choisit de répartir ϵ en parts égales entre les tests (*Bonferroni correction*), ce qui revient à autoriser le même nombre de fausses détections par candidat. On a alors

$$\kappa_i = \min \left\{ n \in \mathbb{N} \text{ tel que } \mathbb{P}[k(c_i, X) \geq n] \leq \frac{\epsilon}{N_C} \right\}.$$

D'où, le nombre attendu d'évènements observés sous H_0 est

$$\mathbb{E}[d(X)] = \mathbb{E} \left[\sum_{i=1}^{N_C} 1_{k(c_i, X) \geq \kappa_i} \right] = \sum_{i=1}^{N_C} \mathbb{P}[k(c_i, X) \geq \kappa_i]$$

Avec \mathbb{E} l'opérateur d'espérance et X les données aléatoires de H_0 . Mais par définition de κ_i , on sait que $\mathbb{P}[k(c_i, X) \geq \kappa_i] \leq \frac{\epsilon}{N_C}$.

Alors,

$$\mathbb{E}[d(X)] \leq \sum_{i=1}^{N_C} \frac{\epsilon}{N_C} = \epsilon$$

THALES ALENIA SPACE OPEN

Ce document ne peut être reproduit, modifié, adapté, publié, traduit d'une quelconque façon, en tout ou partie, ni divulgué à un tiers sans l'accord préalable et écrit de Thales Alenia Space.

© 2018, Thales Alenia Space

Modèle : 83230326-D0C-TAS-FR/003

L'approche A Contrario définit le $NFA(c_i, x) = N_C \cdot \mathbb{P}[k(c_i, X) \geq k(c_i, x)]$. Une observation (c_i, x) est ϵ -significatif si et seulement si $NFA(c_i, x) \leq \epsilon$. Ce test est bien équivalent à $\mathbb{P}[k(c_i, X) \geq k(c_i, x)] \leq \frac{\epsilon}{N_C}$, ce qui implique en retour que $k(c_i, x) \geq \kappa_i$ comme requis plus haut, et on contrôle bien l'espérance du nombre d'occurrences de l'évènement sous H_0 (ou le nombre de fausses détections).

Définition "alignement ϵ -meaningful" :

Un segment de longueur l est ϵ -meaningful s'il contient au moins $k(l)$ points ayant leur direction alignée avec celle du segment où $k(l)$ est donné par

$$k(l) = \min \left\{ k \in N \text{ tel que } \mathbb{P}[S_l \geq k] \leq \frac{\epsilon}{N_C} \right\}.$$

VIII.4.f Calcul du NFA pour la détection d'alignement

Étant donné N points $x = \{x_i\}_{i=1\dots N}$ et un rectangle r , on note $k(r, x)$ le nombre de ces points observé à l'intérieur de r . Le degré de non accidentalité d'un rectangle r détecté comme étant un alignement est mesuré en regardant à quel point la probabilité $\mathbb{P}[k(r, X) \geq k(r, x)]$ est faible où X est un ensemble de N points qui respecte H_0 (= des points aléatoirement distribués). Dans la même idée, un rectangle r' sera considéré comme au moins aussi bon que r étant donné l'observation x , si $\mathbb{P}[k(r', X) \geq k(r', x)] \leq \mathbb{P}[k(r, X) \geq k(r, x)]$.

Pour calculer cette probabilité, on doit prendre en compte les hypothèses sur la densité et la régularité des points. Soit un rectangle candidat r à l'intérieur d'une fenêtre locale R , et qui est divisé en c boîtes. Ensuite, soit $n^* = 2 \times \max(M_1, M_3) + M_2$ le nombre de points dans la surface locale en considérant la densité latérale maximum (M_1 et M_3 : nombre de points dans les parties gauche et droite de r et M_2 : nombre de points dans r). Si b est le nombre de boîtes occupées par des points, on cherche finalement à déterminer la probabilité p d'observer l'évènement **"au moins b sur c boîtes sont occupées dans r , étant donné n^* points aléatoirement distribués dans R ".**

La probabilité qu'un point tombe dans l'une des boîtes est $p_1 = S_B/S_R$ où S_B est l'aire d'une boîte et S_R l'aire de la fenêtre locale. D'où la probabilité d'avoir une boîte occupée par au moins un des n^* points est :

$$p_0 = 1 - (1 - p_1)^{n^*}.$$

Alors la probabilité qu'au moins b boîtes sur c soient occupées est :

$$p = \mathcal{B}(c, b, p_0) \text{ où } \mathcal{B}(n, k, p) = \sum_{j=k}^n \binom{n}{j} p^j (1 - p)^{n-j}.$$

Pour obtenir le NFA, on a besoin de multiplier la probabilité calculé p par le nombre de tests N_{tests} . Pour obtenir le nombre de tests, on rappelle dans un premier temps qu'on considère chaque paire de points de notre ensemble de N points, ce qui nous donne un total de

$N(N-1)/2$ paires de points¹⁷. Ensuite, on considère un nombre de W largeurs différentes à tester pour notre rectangle r et L différentes tailles de notre fenêtre locale pour l'estimation de la densité. En plus, on considère également C différents nombre de boîtes dans lesquelles r est divisé. D'où le nombre d'évènements testés est $N_{tests} = \frac{N(N-1)}{2} \cdot W \cdot L \cdot C$. Au final, en ayant calculé les quantités précédentes, le NFA d'un alignement est :

$$NFA(r, R, c, \mathbf{X}) = \frac{N(N-1)}{2} \cdot W \cdot L \cdot C \cdot \mathcal{B}(c, b(r, c, \mathbf{X}), p_0(r, R, c))$$

VIII.4.g Algorithmes de détection d'alignement de point

Algorithme : Détecteur d'alignement de points

entrée : Un ensemble \mathbf{X} de N points [$W = 8, L = 8, C = \{1 \dots \lfloor \sqrt{N} \rfloor\}$, $\epsilon = 1$]
sortie : Une liste \mathbf{I} d'alignements de point

```

for i = 1 à N-1 do
  for j = i+1 à N do
    l ← distance(xi, xj)
    w ← l/10
    for 1 à W do
      r ← rect(xi, xj, w)
      wL ← l
      for 1 à L do
        R1 ← local_win_left(xi, xj, wL)
        R3 ← local_win_right(xi, xj, wL)
        Compter M1, M2, M3 le nombre de points dans R1, r et R3 respectivement
        Calculer n'(R, X)
        for c ∈ C do
          Diviser r en c boîtes égales
          Calculer p0(r, R, c)
          Compter b(r, c, X) le nombre de boîtes occupées
          Calculer NFA(r, R, c, x)
          if NFA(r, R, c, x) ≤ ε then
            | 1 ← r
          end
        end
        wL ← wL/√2
      end
      w ← w/√2
    end
  end
end
end

```

Algorithme : Réduction de redondance

entrée : Une liste \mathbf{I} de tous les alignements significatifs
sortie : Une liste \mathbf{m} des alignements maximalelement significatifs

```

I ← sort(I) // trie par NFA, du plus faible au plus élevé
m(1) ← I(1)
for i = 2 à length(I) - 1 do
  B = I(i)
  masked = false
  for j = 1 à length(m) do
    A = m(j)
    X' = {x | x ∉ rA}
    if NFA(rB, RB, cB, X') > ε then
      masked = true
      break
    end
  end
  if masked == false then
    m ← I(i)
  end
end
end

```

Algorithme : Détecteur d'alignement de points pour les particules cosmiques

entrée : Un ensemble \mathbf{X} de N points [$W = 8, L = 8, C = \{1 \dots \lfloor \sqrt{N} \rfloor\}$, $\epsilon = 1$]
sortie : Une liste \mathbf{I} d'alignements de point

```

for i = 1 à N-1 do
  for j = i+1 à N do
    l ← distance(xi, xj)
    if l < lmin || l > lmax then
      continue
    end
    w ← l/10
    for 1 à W do
      r ← rect(xi, xj, w)
      wL ← l
      for 1 à L do
        R1 ← local_win_left(xi, xj, wL)
        R3 ← local_win_right(xi, xj, wL)
        Compter M1, M2, M3 le nombre de points dans R1, r et R3 respectivement
        Calculer n'(R, X)
        for c ∈ C do
          Diviser r en c boîtes égales
          Calculer p0(r, R, c)
          Compter b(r, c, X) le nombre de boîtes occupées
          Calculer NFA(r, R, c, x)
          if NFA(r, R, c, x) ≤ ε then
            | 1 ← r
          end
        end
        wL ← wL/√2
      end
      w ← w/√2
    end
  end
end
end

```

¹⁷ 2 parmi N possibilités = $\binom{N}{2} = \frac{N!}{N!(N-2)!} = \frac{N(N-1)}{2}$

TUTEUR DE STAGE

Nom : Monfort

Prénom : Cécile

Service : Architecture Chaîne Image
BLOEN/DOSF/OS/IO

C.C. :

MENTION DE PROTECTION

PROTECTION INDUSTRIELLE

☒ OPEN

☐ THALES ALENIA SPACE INTERNAL

☐ THALES ALENIA SPACE CONFIDENTIAL

☐ THALES ALENIA SPACE SECRET

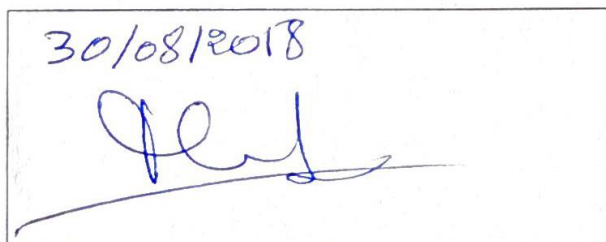
PROTECTION DE DEFENSE

☐ DIFFUSION RESTREINTE

☐ AUTRES (à préciser) :

VALIDATIONS

TUTEUR (Date et Signature)

30/08/2018


Remarques :

SECURITE (Date et Signature)

F. VERARDI
Officier de Sécurité Adjoint
ETABLISSEMENT DE CANNES
Thales Alenia Space
03/09/18

Remarques :

RAS.

THALES ALENIA SPACE OPEN

Ce document ne peut être reproduit, modifié, adapté, publié, traduit d'une quelconque façon, en tout ou partie, ni divulgué à un tiers sans l'accord préalable et écrit de Thales Alenia Space.

© 2018, Thales Alenia Space

Modèle : 83230326-D0C-TAS-FR/003