

AR 2005
AR/05/F : TMS/SYSTR/00005/FR

Projet Techno-Vision ROBIN : Spécification de la base d'images fournie par MBDA-F

Diffusion : M/T/SM/T/V (N. CAMBOU, L. AGRANIER, J-M. FEVRE, A. BILBEAU)
(A. KOTENKOFF)
M/T/SM/T (M. LEMOINE)

Classification militaire et industrielle Document NC Page descriptive NC Date de déclassification	N° Date	
	<hr/> N° CONTRAT CLIENT	
	<hr/> PROGRAMME AR 2005 MATERIEL TECHNIQUE Traitement d'Images	
TITRE Projet Techno-Vision ROBIN : Spécification de la base d'images fournie par MBDA-F Tapez le titre du document dans la zone Titre, le nom de l'auteur dans la zone Auteur et les mots clés (en minuscule) dans la zone Mots Clés		
RESUME <p>Le projet ROBIN (Recherche d'Objets dans des Bases d'Images Numériques), retenu dans le cadre de l'appel à proposition Techno-Vision (technologies de la vision), est une initiative des ministères de la recherche et de la défense pour promouvoir le benchmarking dans le domaine de la vision par ordinateur. M/T/SM/T/V participe à ce projet en qualité d'expert de l'évaluation d'algorithmes de vision et en fournissant des bases de données d'apprentissage et de tests. Le présent document définit la base d'images qui va être livrée aux participants du projet.</p>		
AUTEUR (S) Nom KOTENKOFF A. Sigle M/T/SM/T/V Visa		APPROBATION(S) Nom CAMBOU N. Sigle M/T/SM/T/V Visa
Indice	Date	Justifications de l'évolution du document
A	06/06/2005	Création
Nature du document	MOTS CLES	Nb pages : 14
Langue FR	Benchmarking d'algorithmes Détection / Reconnaissance automatiques d'objets	Nb figures Nb annexes
ARCHIVAGE	N° interne	REFERENCES AUTRES DOCUMENTS
Localisation Année destruction OBSERVATIONS		

SOMMAIRE

SOMMAIRE.....3

.1 INTRODUCTION

La vision par ordinateur s'est attachée depuis ses débuts à développer des méthodes d'interprétation automatique des images, en particulier, des méthodes permettant de reconnaître ou d'identifier des objets présents dans ces images. Des premières méthodes exploitant le formalisme des systèmes experts jusqu'aux méthodes récentes mêlant apprentissage et vision, pas une année ne s'est écoulée sans qu'une conférence majeure du domaine ne publie des articles novateurs consacrés à ce sujet. Cependant, malgré des efforts continus et quelques résultats encourageants, la reconnaissance d'objets reste une question ouverte. Son enjeu est pourtant certain : les domaines de la robotique, de la conception d'IHM, de la vidéosurveillance, de la gestion de bases de données image, de l'imagerie aérienne ou satellitaire, intègrent une telle fonctionnalité comme finalité ou comme étape intermédiaire d'une chaîne de traitement. L'objectif de la proposition **ROBIN**, soumise dans le cadre de l'appel à proposition **Techno-Vision** du ministère de la recherche et du ministère de la défense, est d'apporter une contribution à l'analyse de ce problème difficile par la mise en place d'une structure d'évaluation d'algorithmes sur des données d'intérêt opérationnel. En effet, sept parmi les acteurs industriels majeurs ayant des activités en vision par ordinateur et une trentaine de laboratoires publics de recherche ont allié leurs forces pour proposer une campagne d'évaluation ambitieuse. La campagne porte sur l'évaluation d'une tâche unique: la recherche d'objet dans des images. Les algorithmes recevront en entrée une liste d'images chargée de modéliser les objets cible ; ils devront délivrer la liste des positions où se trouvent ces objets dans la ou les images testées (détection) et donner un indicateur de la nature de l'objet présent dans l'image (catégorisation). Cette formulation permet de rendre compte des aspects les plus notables de la reconnaissance d'objets en un format unifié. La diversité des contextes applicatifs sera représentée par des images aéroportées (avion/satellite), des images terrestres (caméra fixe ou mobile) ou des images issues de séquences vidéo.

M/T/SM/T/V participe au projet ROBIN d'une part en tant qu'expert de l'évaluation dans le processus de définition des méthodologies, des protocoles, des métriques et des épreuves, d'autre part en fournissant des données d'apprentissage et de tests.

L'objet de ce document est de préciser la proposition faite par M/T/SM/T/V dans le cadre du projet ROBIN [4], en définissant plus en détail les bases de données que MBDA s'est engagé à livrer aux laboratoires participants.

.2 REFERENCES

- [1] Technologie de la Vision - Recherche d'Objets dans des Bases d'Images Numériques - Frédéric Jurie - novembre 2004 - <http://lear.inrialpes.fr/lear/people/jurie/robin/>
- [3] Demande de Subvention - commercial proposal réf. n° M/GB/F/F/JC04265 - 20/10/2004
- [2] Fiche d'Autofinancement - conforme à la proposition commerciale réf. n° M/GB/F/F/JC04265 - 22/12/2004
- [4] Proposition et chiffrage pour la participation de MBDA-F au consortium Techno-Vision ROBIN - M/T/SM/T/V-2004-0139-A - 19/10/2004

.3 NATURE DES DONNÉES FOURNIES

Il s'agira d'une base de données d'images de synthèse, complétée par une sous-base d'images réelles issue d'une campagne d'enregistrement réalisée pour le projet ROBIN à partir de points hauts (ponts, collines...).

Les images réelles tout comme les images de synthèse seront représentatives par leur taille et leurs conditions de prise de vue variées d'un imageur embarqué à bord d'un système terrestre ou aérien.

Pour les images réelles comme pour les images de synthèse, les trois bandes spectrales retenues sont :

la bande IR 3-5 μ

la bande IR 8-12 μ

la bande visible panchromatique

..1 Description des images réelles

Les images réelles proviendront de trois capteurs matriciels distincts, associées aux trois bandes spectrales traitées :

une caméra pour la bande Visible panchromatique

une caméra ORION de taille 256² pour la bande IR 3-5 μ

une caméra micro-bolomètres de taille 320x240 pour la bande IR 8-12 μ

Il s'agira d'images unitaires et non pas de séquences d'images, en visée oblique représentatives de prises de vue aéroportées et en visée frontale pour le contexte terrestre. Elles seront utilisées en appoint, lors de la phase d'évaluation uniquement, pour étudier le passage des données de synthèse aux données réelles et analyser les problèmes liés à la modélisation.

..2 Description des images de synthèse

Les images de synthèse seront relativement proches, par leur taille, leurs conditions de prise de vue, des images réelles. Pour obtenir une modélisation la plus réaliste possible, la physique d'un capteur matriciel réel est prise en compte. Les images seront générées sous formes de séquences.

Nature du capteur

L'imageur sera matriciel de taille typique 512² pixels avec une résolution imposée. Les bandes spectrales modélisées sont les mêmes que pour les images réelles.

Les images seront codées sur une dynamique fixe, donc sans contrôle automatique de gain. Il y aura un seul niveau de bruit capteur, d'amplitude raisonnable, mais pouvant accuser de petites variations. La FTM sera nominale et réaliste.

Facteurs environnementaux

Divers paramètres influents seront modélisés : les paramètres de visibilité et de contraste comme la transmission atmosphérique (liée à la météo) et l'éclairage (lié à la hauteur solaire) d'une part, d'autre part la présence d'éléments perturbateurs comme les ombres, les fumées, les masquages et le fouillis au sol. Par contre, une modélisation réaliste du ciel n'est pas prévue, s'il apparaît dans les images.

Douze réglages différents des paramètres de visibilité et de contraste seront définis, au moyen de l'heure, de la date et des conditions météorologiques, pour traduire l'influence de l'éclairage et de la météo, à travers les facteurs suivants :

température ambiante :

- temps hivernal froid,
- temps printanier doux,
- temps estival chaud,

visibilité atmosphérique :

- réduite (brouillard),
- acceptable (brume légère),
- parfaite (ciel net),

éclairage :

- apogée solaire estivale,
- hauteur solaire intermédiaire,
- crépuscule,
- nuit.

Les conditions nominales de visibilité et de contraste correspondront à la combinaison : temps chaud, visibilité atmosphérique parfaite, soleil à l'apogée estivale.

Conditions de prise de vue et tailles des cibles

L'angle d'élévation des images variera entre 0 et 70°, la distance de prise d'image entre 2000 et 50 m, au minimum, ce qui se traduira par des tailles de cibles allant de la dizaine à la centaine de pixels de large.

Format des images

Les images seront codées au format 2 octets non signés compatibles SUNOS, sans entête.

.4 CONTENU DES SCÈNES

..1 Nature des objets d'intérêt

Il s'agira principalement d'objets déplaçables civils, véhicules et aéronefs au sol, plus un couple d'éléments d'infrastructure. Les objets sont rassemblés en catégories, chacune pouvant inclure plusieurs modèles assez proches entre eux par leur aspect. Les images d'évaluation contiendront les mêmes objets qu'à l'apprentissage, mais aussi d'autres objets qui n'étaient pas présents lors de cette phase, l'aspect de ces nouveaux objets restant tout de même assez similaire à celui des objets appris. Le nombre total d'objets disponibles à l'apprentissage sera de 10, le nombre total à traiter en phase d'évaluation de 15.

Parmi les objets déplaçables, il y aura les catégories suivantes :

voiture (4 modèles : 3 à l'apprentissage / 4 à l'évaluation)

camion (3 modèles : 2 à l'apprentissage / 3 à l'évaluation)

avion (3 modèles : 2 à l'apprentissage / 3 à l'évaluation)

hélicoptère (3 modèles : 2 à l'apprentissage / 3 à l'évaluation)

La seule catégorie infrastructurelle proposée est :

tour Télécom (2 modèles : 1 à l'apprentissage / 2 à l'évaluation).

Quatre grandes gammes d'échelles pixeliques correspondant à quatre niveaux de difficultés dans les épreuves seront proposés :

10-30 pixels

30-60 pixels

60-100 pixels

100-150 pixels.

Ces tailles en pixels fixent la plus grande dimension apparente de l'objet dans l'image.

Dans les images réelles, les seuls objets à reconnaître présents seront les camions et les voitures. On prévoit deux modèles de chaque, les plus proches possibles des modèles utilisés en images de synthèse.

..2 Environnement

Plusieurs types de paysages seront générés :

paysage forestier et rural,

paysage semi-urbain (présence de quelques bâtiments) et urbain dense,

paysages routiers (réseau de routes ou pistes d'aéroport),

paysages mixtes.

.5 EPREUVES

..1 Longueurs d'onde et paramètres de visibilité

Les algorithmes seront testés pour plusieurs bandes spectrales (visible, IR 3-5 μ et IR 8-12 μ). La base de donnée visible sera dupliquée partiellement en bande IR 3-5 et 8-12 μ . Une sous-partie des bases visible et IR servira à étudier l'influence des paramètres de visibilité et de contraste. Les sous-bases seront dupliquées pour tester diverses combinaisons de ces paramètres.

..2 Apprentissage

L'apprentissage se fera dans un mode quasi-nominal incluant toutefois de légères variations des paramètres d'influence et un bruit de faible amplitude. Il y aura en particulier plusieurs niveaux de contrastes, notamment une gamme de contrastes représentative des images IR nocturnes. Les objets à apprendre seront en configuration lisse, sans masquages, ni éléments perturbateurs, sous éclairage uniforme, sur fond neutre uni et moteur éteint.

..3 Validation et évaluation

Les bases de validation et de test pourront contenir de une à plusieurs classes à reconnaître dans chaque image. Il y aura aussi des scènes ne contenant aucun objet d'intérêt.

La base d'évaluation contiendra de l'ordre d'une dizaine de scénarios. Un scénario se définit par une configuration déterminée de la scène et des objets d'intérêt présents à l'intérieur de celle-ci. Les scénarios seront caractérisés par :

- le nombre et la répartition des objets d'intérêt dans la scène,
- les masquages opaques ou translucides susceptibles de perturber les traitements,
- l'échauffement local des objets en IR,
- l'adjonction d'éléments perturbateurs (jet d'avion au décollage, portière ouverte sur un véhicule, fumée...),
- la complexité du fond (fond non uniforme, fouillis de sol).

L'autre source de difficulté sera liée aux conditions de prise de vue, indépendantes du scénario. La base d'évaluation inclura des conditions de prises de vue aux limites (vol rasant, faible taille pixélique) rendant les tâches de détection et de catégorisation plus difficiles.

La répartition des cas testés sera de l'ordre de 50-50 entre des configurations nominales proches de l'apprentissage et des variantes correspondant aux difficultés décrites ci-dessus.

Il pourrait être intéressant de quantifier spécifiquement l'effet du masquage. Un scénario et une trajectoire adaptés serviront à mesurer l'effet d'un masquage progressif des objets d'intérêt sur les performances des algorithmes.

.6 ORGANISATION DES BASES DE DONNÉES

Trois bases de données distinctes seront constituées :

- une base d'apprentissage,
- une base de validation, réservée au réglage des algorithmes,
- la base exhaustive d'évaluation.

La base d'apprentissage contiendra des vignettes unitaires, les bases de validation et d'évaluation des images unitaires extraites de séquences. Les séquences, livrables dans leur intégralité, seront générées à partir de trois familles de trajectoires : des faisceaux de trajectoires convergentes vers un point donné, des trajectoires de survol rectilignes et des circulaires.

..1 Apprentissage

L'apprentissage se fera à partir de vignettes et concerne un total de 10 classes d'objets. La base d'apprentissage ne contiendra que des données de synthèse. Elle sera dupliquée partiellement du visible vers les bandes IR 3-5 et 8-12. Les données d'apprentissage seront générées sur fond uni neutre, sous un éclairage uniforme.

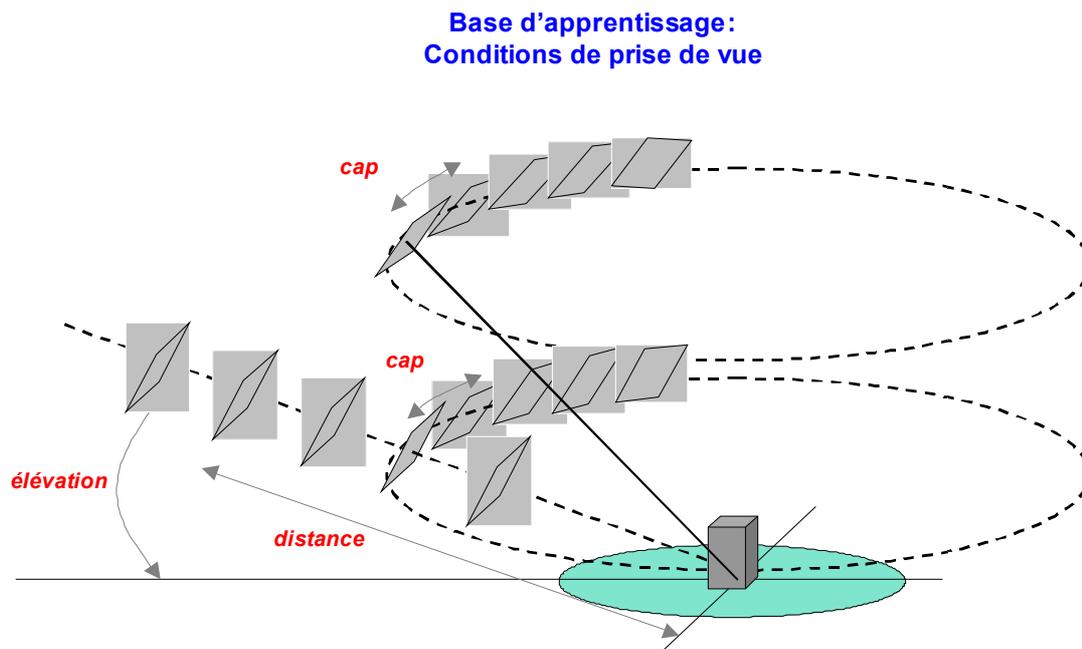
Les configurations envisagées sont les suivantes :

- 3 types de contrastes,
- 4 échelles pixéliques,
- 16 présentations en cap,
- 5 élévations.

Le contraste sera modulé à l'aide des paramètres régissant la visibilité. En bande infrarouge, une gamme de contrastes sera représentative des images de nuit. Les tailles pixéliques sont celles qui ont été définies plus haut.

Les autres paramètres influents et le bruit seront modulés aléatoirement en parallèle.

Le volume de données sera de 960 vignettes par classe d'objet, le volume total pour l'ensemble des classes de 9600 vignettes par bande spectrale. Au total, en comptant les trois bandes, 28800 vignettes seront livrées.



..2 Evaluation

Les images de synthèse de la base de test seront extraites de séquences. La base contiendra un total de 15 classes d'objets. Elle sera déclinée dans plusieurs bandes spectrales. Une sous-partie de cette base, générée à partir du réglage nominal de la visibilité et du contraste, sera dupliquée onze fois pour les autres réglages du couple météo/éclairage.

La répartition en volumes de données sera la suivante :

Jeu de référence :

- 1500 images extraites de 150 séquences,
- bande visible,
- réglage nominal de la visibilité.

Jeu IR 3-5 :

- 1000 images extraites de 100 séquences,
- bande IR 3-5,
- réglage nominal de la visibilité,
- obtenu par duplication partielle du jeu de référence.

Jeu IR 8-12 :

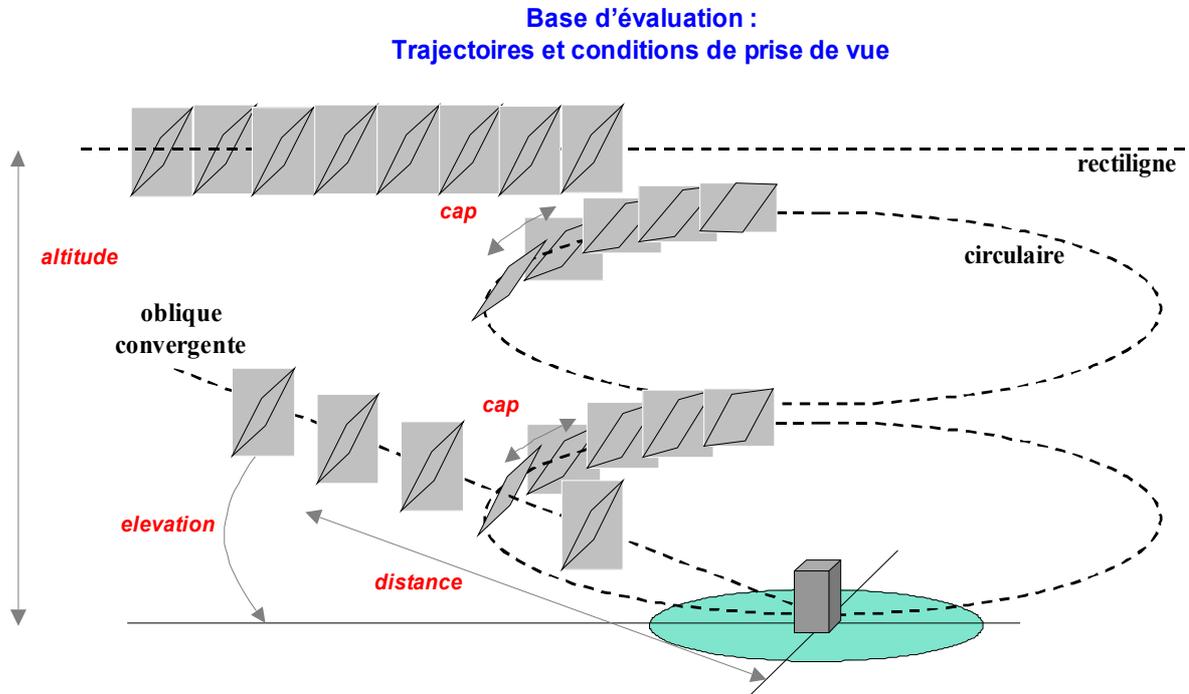
- 1000 images extraites de 100 séquences,
- bande 8-12,
- réglage nominal de la visibilité,
- obtenu par duplication complète du jeu IR 3-5.

Jeux paramétriques :

- 11 jeux de 300 images extraites de 30 séquences,
- bandes visible, IR 3-5 et IR 8-12,
- 11 réglages des paramètres de visibilité autres que le réglage nominal,

- générés par modulation paramétrique du jeu nominal, constitué de sous-parties prélevées parmi les trois jeux ci-dessus et regroupant chacune 100 images sur 10 séquences.

Le volume complet de la base de test sera de 6800 images, correspondant à 680 séquences.



..3 Validation

La base de validation sera organisée comme la base d'évaluation, mais elle ne contiendra que 10 classes d'objets. La variété de contenus et de réglages paramétriques des données de validation sera moins importante que pour la base de test, mais représentative des difficultés des épreuves.

La répartition en volumes sera la suivante :

Jeu visible :

- 300 images extraites de 30 séquences,
- visibilité et contraste variables.

Jeu IR 3-5 :

- 150 images extraites de 15 séquences,
- visibilité et contraste variables.

Jeu IR 8-12 :

- 150 images extraites de 15 séquences,
- visibilité et contraste variables.

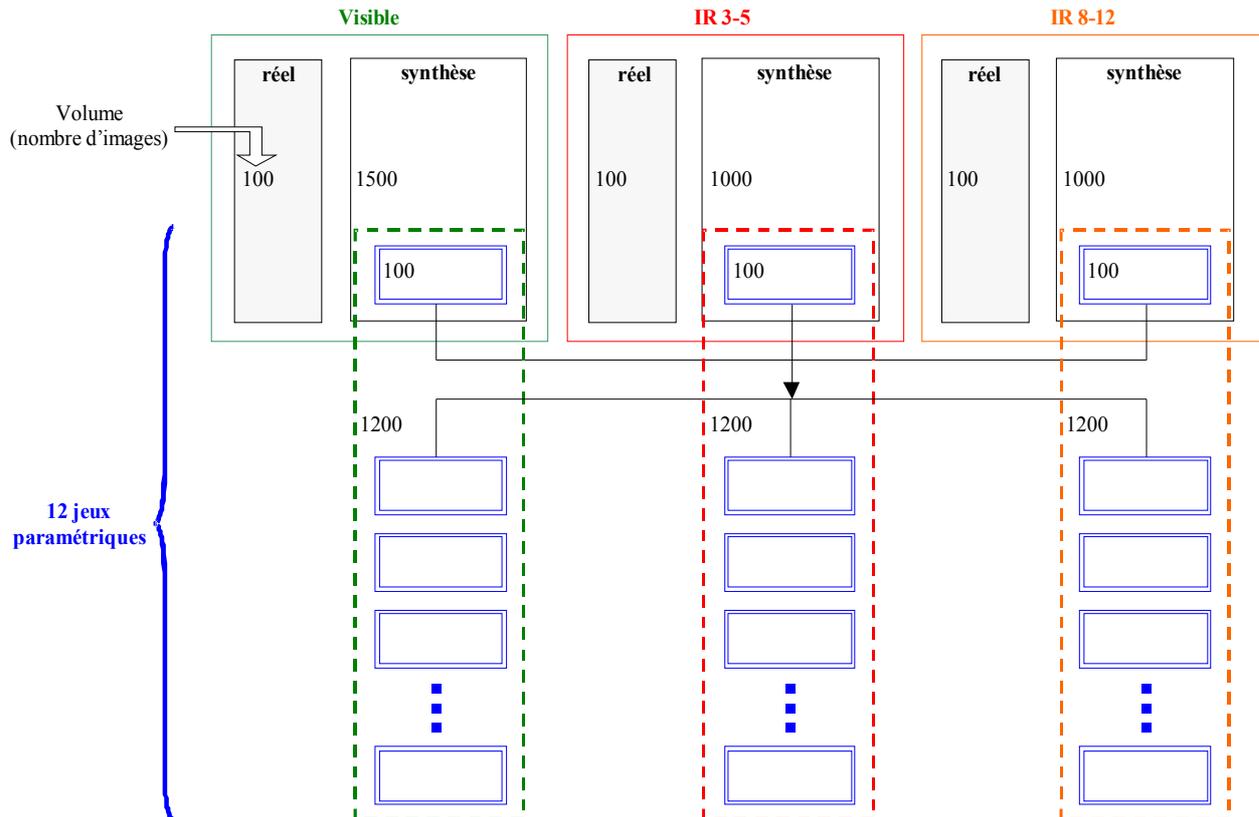
Le volume complet de la base de validation sera de 600 images correspondant à un total de 60 séquences.

..4 Base d'images réelles unitaires

Les jeux d'images de synthèse associés à chaque bande spectrale de la base d'évaluation seront complétés par trois lots de 100 images réelles unitaires : un lot en visible, un autre en IR 3-5 μ et un troisième en IR 8-12 μ . Le volume total des images réelles fournies sera donc de 300.

Dans ces images, les objets d'intérêt apparaîtront sous divers angles de présentation, aussi bien en élévation qu'en cap relatif. Les quatre gammes de tailles pixéliques seront représentées.

Organisation de la base d'évaluation



.7 DONNÉES D'APPOINT

..1 Images de synthèse

Le repère associé aux images de synthèse et aux vignettes d'apprentissage sera le trièdre local (X, Y, Z) = (Est, Nord, Zénith).

Pour chaque objet, le modèle géométrique dans un repère 3D comportant un axe et un point de référence sera fourni au format .DXF. Un même objet aura diverses affectations en températures possibles, chacune décrite par un fichier. Les fichiers contenant le profil thermique intrinsèque de l'objet sur plusieurs vues seront également livrés.

..1.1 Vignettes d'apprentissage

Dans les vignettes d'apprentissage, l'objet occupera toujours la même position 3D et son axe de référence sera positionné toujours de manière identique.

L'apprentissage sera paramétré à l'aide de fichiers définissant le capteur, le réglage utilisé. Ces fichiers seront livrés.

Les données fournies pour chaque vignette sont :

- la classe de l'objet,
- les conditions de prises de vues associées,
- une référence à un fichier capteur,
- une référence à un fichier de réglage,
- une référence à un profil thermique.

..2.1 Images de synthèse

Les données additionnelles fournies sont :

images :

- séquence d'origine,
- paramètres capteur,
- conditions de prises de vues associées.

environnement :

- un label météo désignant l'une des météos modélisées,
- un label désignant la nature de l'éclairage,
- une référence au fichier météo/éclairage associé.

objet d'intérêt :

- la classe de l'objet,
- le rectangle englobant de l'objet hors masquage,
- liste des pixels sur l'objet indiquant pour chacun d'eux ses coordonnées dans l'image, sa distance par rapport au capteur et une indication sur la nature du mas-

- quage éventuelle de ce pixel (absence de masquage, masquage translucide, masquage opaque),
- la localisation et l'orientation 3D de l'objet au moyen de son point et de son axe de référence modèle,
- une référence à un profil thermique.

..2 Images réelles

Les données additionnelles fournies sont :

images :

- paramètres capteur,
- les conditions de prises de vues vraies associées à la prise d'image (coordonnées GPS et angles de visée approximatifs).

environnement :

- un label météo,
- un label éclairage.

objet d'intérêt :

- la classe de l'objet,
- le rectangle englobant de l'objet, à l'exclusion des parties masquées,
- la distance télémétrée de l'objet,
- l'indication d'un masquage éventuel,
- la localisation spatiale de l'objet (coordonnées GPS) et l'orientation de son axe de référence,
- une indication de l'état thermique de l'objet (chaud, froid).