



# Annotation sémantique 2D/3D d'images spatialisées pour la documentation et l'analyse d'objets patrimoniaux

Adeline Manuel

## ► To cite this version:

Adeline Manuel. Annotation sémantique 2D/3D d'images spatialisées pour la documentation et l'analyse d'objets patrimoniaux. Génie des procédés. Ecole nationale supérieure d'arts et métiers - ENSAM, 2016. Français. <NNT : 2016ENAM0009>. <tel-01304776>

**HAL Id: tel-01304776**

**<https://pastel.archives-ouvertes.fr/tel-01304776>**

Submitted on 20 Apr 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

École doctorale n° 432 : Science des Métiers de l'ingénieur

**Doctorat ParisTech**

**T H È S E**

pour obtenir le grade de docteur délivré par

**l'École Nationale Supérieure d'Arts et Métiers**

**Spécialité Conception**

*présentée et soutenue publiquement par*

**Adeline MANUEL**

le 22 Mars 2016

**Annotation sémantique 2D/3D d'images spatialisées  
pour la documentation et l'analyse d'objets patrimoniaux**

Directeur de thèse : **Livio DE LUCA**  
Co-Directeur de thèse : **Philippe VÉRON**

**Jury**

**M. Frédéric MERIENNE**, Professeur des Universités, Le2i, Arts et Métiers ParisTech

**Mme Michela SPAGNUOLO**, CNR Senior Research, IMATI, CNR Italie

**M. Marc PIERROT-DESEILLIGNY**, Directeur de Recherche, IGN

**M. Jean-Marc VALLET**, Ingénieur de Recherche Ministère de la Culture et de la Communication, CICRP

**M. Livio DE LUCA**, Directeur de Recherche CNRS, MAP, CNRS

**M. Philippe VÉRON**, Professeur des Universités, LSIS, Arts et Métiers ParisTech

Président

Rapporteur

Rapporteur

Examineur

Examineur

Examineur



## ANNOTATION SEMANTIQUE 2D/3D D'IMAGES SPATIALISEES POUR LA DOCUMENTATION ET L'ANALYSE D'OBJETS PATRIMONIAUX

**RESUME :** L'introduction des technologies numériques dans les méthodes de documentation pour le patrimoine culturel a permis de concevoir de nouveaux outils pour l'acquisition et la gestion des informations collectées pour des études pluridisciplinaires. Ces outils commencent à s'affirmer comme des supports privilégiés pour la description, l'analyse et la compréhension des objets d'étude. Les solutions existantes d'annotation sémantique d'images, de modèle 3D ou encore hybrides 2D/3D se révèlent aujourd'hui encore insuffisantes pour aborder ce problème complexe qu'est l'annotation d'artéfacts patrimoniaux. La complexité provient du fait que l'étude d'objets patrimoniaux ne peut pas se réduire à une simple analyse bidimensionnelle mais doit aussi s'articuler avec une véritable analyse tridimensionnelle. La description des objets doit donc pouvoir s'appuyer sur un support de représentation plus sophistiqué que seulement des images ou seulement des modèles 3D en explicitant, d'une part, la complexité morphologique de l'objet et en reflétant, d'autre part, les différents aspects véhiculés par les acquisitions d'imagerie scientifique.

Cette thèse propose une approche permettant de conduire des annotations sur les différents supports bidimensionnels (photographies et imagerie scientifique) tout en facilitant le travail d'annotation par une propagation automatique de ces annotations entre les différentes représentations (2D ou 3D) de l'objet en se basant sur la notion de référencement spatial et sur l'établissement d'une relation projective 2D/3D. L'objectif de l'approche est de définir un continuum d'information entre toutes les phases du processus allant de l'acquisition d'images et de données spatiales jusqu'à la construction de représentations 3D sémantiquement enrichies en intégrant les aspects multi-supports, multi-échelles, multi-couches et multi-temporels. L'idée est de pouvoir insérer des éléments de sémantique au niveau de toutes les phases de traitements des données tout en assurant la corrélation de jeux d'annotations sur des ensembles d'images d'un point de vue à la fois spatial, temporel et sémantique. Ce travail a abouti à la définition et le développement d'un ensemble de modules informatiques pouvant être utilisés par des spécialistes de la conservation du patrimoine culturel comme par le grand public via une interface web.

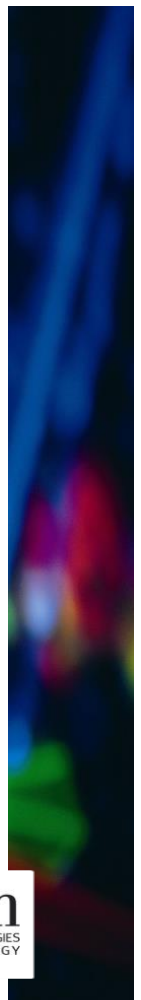
**Mots clés :** Patrimoine culturel, description sémantique, annotation d'images, spatialisation d'informations, multi-représentations, multi-temporels

### 2D/3D SEMANTIC ANNOTATION OF SPATIALIZED IMAGES FOR THE DOCUMENTATION AND ANALYSIS OF CULTURAL HERITAGE

**ABSTRACT:** The introduction of digital technologies in the documentation methods for cultural heritage has helped to design new tools for the acquisition and management of information collected for multidisciplinary studies. These tools are beginning to emerge as the preferred media for describing, analyzing and understanding the objects of study. Existing solutions for semantic annotation on images, on 3D models or with 2D/3D hybrid methods still reveal itself today insufficient to tackle the complex problem of annotating heritage artifacts. The complexity comes from the fact that the study of cultural heritage objects cannot be reduced to a simple bi-dimensional analysis but must also be linked with a true three-dimensional analysis. The description of the objects must be able to rely on a more sophisticated representation than only images or only 3D models by on one hand making explicit the morphological complexity of the object and on the other hand by reflecting all aspects conveyed by the acquisitions of scientific imaging.

This thesis proposes an approach for conducting annotations on the different bi-dimensional media (photography and scientific imaging) while facilitating the annotation work with an automatic propagation of these annotations between the different representations (2D or 3D) of the object. It is based on the concept of spatial referencing and the establishment of a projective relationship 2D / 3D. The goal of the approach is to define a continuum of information between all phases of the process ranging from the acquisition of images and spatial data to go until the building of 3D representations semantically enriched by incorporating multi-media, multi-scale, multi-layered and multi-temporal aspects. The idea is to insert semantic elements at all phases of data processing while ensuring correlation of annotations groups on sets of images from a spatial, temporal and semantically point of view. This work resulted in the definition and development of a set of software modules that can be used by specialists in the conservation of cultural heritage as well as the general public via a web interface.

**Keywords:** Cultural heritage, semantic description, image annotations, information spatialization, multi-representation, multi-temporal







## Remerciements

Je tiens à remercier toutes les personnes qui, de près ou de loin, m'ont aidé et ont contribué à l'élaboration de ce travail de recherche et la réussite de ses trois années de thèse.

Je souhaite remercier *Livio DE LUCA*, en tant que directeur de thèse, pour m'avoir accueillie au sein du laboratoire MAP, tout d'abord en tant que stagiaire puis en tant que doctorante, et pour ses encouragements tout au long de ses années ; *Philippe VERON*, en tant que co-directeur de thèse, pour ses conseils et la liberté qu'il m'a accordée au cours de ces années. Je tiens à les remercier également tous les deux pour leur disponibilité malgré leurs emplois du temps très chargés et de m'avoir donné l'envie et l'occasion de poursuivre mes études dans le domaine de la recherche.

Je remercie également les membres du jury d'avoir accepté de lire et d'évaluer mon travail. Merci à *Frédéric MERIENNE* d'avoir présidé ce jury de thèse. Merci à *Michela SPAGNUOLO* et Marc *PIERROT-DESEILLIGNY* d'avoir été les rapporteurs attentifs de mon travail. Merci à *Jean-Marc VALLET* pour son regard professionnel et minutieux ainsi que ses encouragements rassurants.

Je remercie aussi tous les partenaires des projets C3DC et MONUMENTUM pour l'intérêt qu'ils ont porté à mes travaux, leur aide et leurs remarques constructives quant à l'élaboration de mon système.

Je tiens à remercier les membres du laboratoire MAP, anciens et nouveaux, qui ont contribué par leur bonne humeur à créer une ambiance agréable tout au long de ses années de thèse et mais aussi pour l'aide et les conseils que chacun a pu m'apporter : *Aurélie, Anthony, Benjamin, Chiara, Christine, David, Eloi, Estelle, Fabien, Francesca, Frieda, Iwona, Jean-Yves, Julie, Kévin, Laurent, Léo, Marine, Menehould, Michel Berthelot, Michel Florenzano, Micheline, Nicola, Nicolas Martin-Beaumont, Nicolas Nony, Noémie, Nouha, Pascal, Philippe, Tommy, Valéria*.

Mes remerciements vont aussi à mes amis et à ma famille, notamment à mes parents *Éric* et *Anne*, ainsi qu'à *Victor*, qui ont supporté mes humeurs, m'ont soutenu et encouragé au cours de la réalisation de ce travail.

Merci à toutes et à tous !



## Résumé

L'introduction des technologies numériques dans les méthodes de documentation pour le patrimoine culturel a permis de concevoir de nouveaux outils pour l'acquisition et la gestion des informations collectées pour des études pluridisciplinaires. Ces outils commencent à s'affirmer comme des supports privilégiés pour la description, l'analyse et la compréhension des objets d'étude. Les solutions existantes d'annotation sémantique d'images, de modèle 3D ou encore hybrides 2D/3D se révèlent aujourd'hui encore insuffisantes pour aborder ce problème complexe qu'est l'annotation d'artéfacts patrimoniaux. La complexité provient du fait que l'étude d'objets patrimoniaux ne peut pas se réduire à une simple analyse bidimensionnelle mais doit aussi s'articuler avec une véritable analyse tridimensionnelle. La description des objets doit donc pouvoir s'appuyer sur un support de représentation plus sophistiqué que seulement des images ou seulement des modèles 3D en explicitant, d'une part, la complexité morphologique de l'objet et en reflétant, d'autre part, les différents aspects véhiculés par les acquisitions d'imagerie scientifique.

Cette thèse propose une approche permettant de conduire des annotations sur les différents supports bidimensionnels (photographies et imagerie scientifique) tout en facilitant le travail d'annotation par une propagation automatique de ces annotations entre les différentes représentations (2D ou 3D) de l'objet en se basant sur la notion de référencement spatial et sur l'établissement d'une relation projective 2D/3D. L'objectif de l'approche est de définir un continuum d'information entre toutes les phases du processus allant de l'acquisition d'images et de données spatiales jusqu'à la construction de représentations 3D sémantiquement enrichies en intégrant les aspects multi-supports, multi-échelles, multi-couches et multi-temporels. L'idée est de pouvoir insérer des éléments de sémantique au niveau de toutes les phases de traitements des données tout en assurant la corrélation de jeux d'annotations sur des ensembles d'images d'un point de vue à la fois spatial, temporel et sémantique. Ce travail a abouti à la définition et le développement d'un ensemble de modules informatiques pouvant être utilisés par des spécialistes de la conservation du patrimoine culturel comme par le grand public via une interface web.

**Mots-clés :** Patrimoine culturel, description sémantique, annotation d'images, spatialisation d'informations, multi-représentations, multi-temporels



## Abstract

The introduction of digital technologies in the documentation methods for cultural heritage has helped to design new tools for the acquisition and management of information collected for multidisciplinary studies. These tools are beginning to emerge as the preferred media for describing, analyzing and understanding the objects of study. Existing solutions for semantic annotation on images, on 3D models or with 2D/3D hybrid methods still reveal itself today insufficient to tackle the complex problem of annotating heritage artifacts. The complexity comes from the fact that the study of cultural heritage objects cannot be reduced to a simple bi-dimensional analysis but must also be linked with a true three-dimensional analysis. The description of the objects must be able to rely on a more sophisticated representation than only images or only 3D models by on one hand making explicit the morphological complexity of the object and on the other hand by reflecting all aspects conveyed by the acquisitions of scientific imaging.

This thesis proposes an approach for conducting annotations on the different bi-dimensional media (photography and scientific imaging) while facilitating the annotation work with an automatic propagation of these annotations between the different representations (2D or 3D) of the object. It is based on the concept of spatial referencing and the establishment of a projective relationship 2D / 3D. The goal of the approach is to define a continuum of information between all phases of the process ranging from the acquisition of images and spatial data to go until the building of 3D representations semantically enriched by incorporating multi-media, multi-scale, multi-layered and multi-temporal aspects. The idea is to insert semantic elements at all phases of data processing while ensuring correlation of annotations groups on sets of images from a spatial, temporal and semantically point of view. This work resulted in the definition and development of a set of software modules that can be used by specialists in the conservation of cultural heritage as well as the general public via a web interface.

**Keywords:** Architectural heritage, semantic description, image annotations, information spatialization, multi-representation, multi-temporal



---

## Table des matières

<b>I. Introduction .....</b>	<b>13</b>
I.A. Contexte .....	13
I.B. Objectif .....	14
I.C. Portées, limites, apports principaux.....	15
I.D. Cadre de la thèse .....	15
I.D.1. Culture 3D Clouds.....	15
I.D.2. Monumentum : MOdélisation NUMérique et gEstion de doNnées pour la conservaTion de strUctures Maconnées .....	17
I.E. Terrains d'expérimentations .....	18
I.F. Structure du document .....	19
<b>II. Outils pour l'acquisition de données.....</b>	<b>21</b>
II.A. Les différents capteurs pour l'acquisition.....	21
II.A.1. Capteurs photographiques.....	21
II.A.2. Scanners .....	24
II.A.2.a. Scanner laser .....	24
II.A.2.a.i. Les scanners longue-portée .....	24
II.A.2.a.ii. Les scanners par triangulation.....	25
II.A.2.b. Scanner à lumière structurée .....	25
II.A.2.c. Précision et niveau de détails .....	26
II.B. Les types de projection .....	27
II.B.1. Projection perspective et images photographiques .....	27
II.B.2. Projection orthogonale et redressement d'images.....	29
II.B.3. Projection panoramique et capteurs 3D .....	30
II.B.3.a. Projection cylindrique .....	31
II.B.3.b. Projection sphérique.....	31
II.B.3.c. Projection cubique.....	32
II.B.4. Bilan .....	33
II.C. Extraction d'informations .....	33
II.C.1. Informations colorimétriques.....	34
II.C.2. Extractions de descripteurs images .....	34
II.C.2.a. Caractéristiques globales.....	34
II.C.2.a.i. Histogrammes.....	34

---



II.C.2.a.ii. Moments d'images.....	35
II.C.2.b. Caractéristiques locales.....	35
II.C.2.b.i. Détection de points d'intérêt.....	35
II.C.2.b.ii. Descripteurs locaux.....	36
II.C.3. Informations géométriques.....	38
II.C.3.a. Mesures métriques.....	38
II.C.3.b. Génération de coordonnées 3D à partir de relevé photographique .....	39
II.C.4. Extractions de descripteurs 3D.....	41
II.C.4.a. Caractéristiques globales.....	41
II.C.4.b. Caractéristiques locales.....	42
II.C.5. Informations morphologiques .....	43
II.C.6. Métadonnées .....	44
II.D. Conclusion sur les méthodes d'acquisition.....	46
<b>III. Méthodes d'annotations sémantiques .....</b>	<b>49</b>
III.A. Annotations d'images.....	50
III.A.1. Annotations manuelles .....	50
III.A.2. Approches guidées par une ontologie.....	52
III.A.3. Utilisation des informations textuelles de contexte .....	54
III.A.4. Annotations automatiques .....	54
III.A.4.a. Approches probabilistes.....	55
III.A.4.b. Approches par classifications .....	56
III.A.4.c. Approches par les graphes .....	59
III.A.4.d. Approches paramétriques .....	59
III.A.4.e. Conclusion sur les méthodes d'annotation automatique.....	59
III.A.5. Méthodes semi-automatiques .....	61
III.B. Annotation de modèles 3D .....	62
III.B.1. Reconstruction 3D .....	62
III.B.2. Annotations 3D de modèles .....	63
III.C. Méthodes d'annotations hybrides 2D/3D .....	65
III.C.1. Phototourism/Photosynth.....	66
III.C.2. Nubes Imago .....	67
III.D. Constat sur les méthodes actuelles d'annotation pour la documentation du patrimoine .....	68
<b>IV. Approche pour la caractérisation sémantique d'objets patrimoniaux à partir d'images spatialisées .....</b>	<b>71</b>
IV.A. Approche proposée.....	73

---

IV.B. Problèmes à traiter .....	75
IV.B.1. Génération d'informations liées aux images (Chapitre V) .....	75
IV.B.2. Ajout de nouvelles images (Chapitre V) .....	75
IV.B.3. Transfert d'annotations entre les images (Chapitre VI).....	76
IV.B.4. Outils d'annotation (Chapitre VII) .....	76
IV.B.5. Outils d'extraction d'informations (Chapitre VII) .....	76
IV.C. Développements informatiques .....	76
IV.C.1. MATLAB.....	77
IV.C.2. C++ et QT .....	77
IV.C.3. HTML5 et WebGL .....	78
<b>V. Lier l'information 2D à l'information 3D : spatialisation des informations .....</b>	<b>79</b>
V.A. Méthodes de référencement spatial .....	79
V.A.1. Référencement image-coordonnées .....	80
V.A.2. Référencement image-modèle 3D .....	81
V.A.3. Référencement image-image .....	84
V.A.4. Analyse des méthodes actuelles.....	84
V.A.5. Solution retenue .....	85
V.B. Relation entre pixels et coordonnées 3D .....	89
V.B.1. Données d'entrées.....	89
V.B.2. Représentation géométrique d'une image photographique.....	91
V.B.2.a. Transformation entre le repère monde et le repère caméra .....	92
V.B.2.b. Transformation entre le repère caméra et le repère capteur .....	92
V.B.2.c. Transformation entre le repère capteur et le repère image .....	93
V.B.2.d. Distorsion dans le repère image .....	93
V.B.2.e. Exploitation du résultat .....	94
V.B.3. Cas des orthophotographies .....	97
V.B.4. Scans Laser .....	99
V.B.5. Projections et relation 2D/3D.....	100
V.C. Gestion des données.....	101
V.D. Ajout de nouvelles images.....	104
V.D.1. Orientation de l'image .....	104
V.D.1.a. Orientation directe .....	104
V.D.1.b. Orientation indirecte .....	106
V.D.1.c. Orientation d'images à différents niveaux de détails.....	107

---

V.D.2. Résolution d'image et densité du nuage de points.....	109
V.D.3. Ajout d'images de différents états temporels .....	110
<b>VI. Processus hybride d'annotation d'images.....</b>	<b>113</b>
VI.A. Transfert d'annotation entre les images .....	113
VI.A.1. Principe de la méthode .....	114
VI.A.2. Définition de l'annotation .....	116
VI.A.3. Recherche des coordonnées 3D.....	117
VI.A.4. Transfert sur les autres images .....	119
VI.B. Transfert multi temporel.....	121
<b>VII. Outils d'annotations et d'extractions d'informations.....</b>	<b>125</b>
VII.A. Segmentation d'images.....	126
VII.A.1. Méthodes sans connaissance de l'organisation des pixels .....	127
VII.A.2. Méthodes avec connaissance de l'organisation des pixels.....	128
VII.A.2.a. Méthodes basées sur les régions.....	128
VII.A.2.b. Méthodes basées sur l'énergie .....	129
VII.A.2.c. Méthodes par lignes de partage des eaux (Watershed) .....	130
VII.A.3. Segmentation RGBD.....	130
VII.A.4. Evaluation des méthodes.....	131
VII.B. Outils de sélection semi-automatique .....	132
VII.B.1. Sélection par seuil : « Baguette Magique ».....	133
VII.B.2. Sélection sur la segmentation .....	135
VII.B.2.a. Segmentation des images grâce aux informations liées .....	135
VII.B.2.b. Outils de sélection semi-automatique à l'aide de la segmentation.....	138
VII.C. Outils d'analyses quantitatives.....	139
VII.C.1. Analyses sur les données bidimensionnelles.....	140
VII.C.2. Analyses sur les données tridimensionnelles .....	142
VII.C.2.a. Caractérisation morphologique.....	143
VII.C.2.b. Cartographie de profondeur .....	144
VII.C.2.c. Mesures dimensionnelles.....	144
VII.C.3. Conclusion.....	145
<b>VIII. Interface d'annotations : Développement d'Aïoli.....</b>	<b>147</b>
VIII.A. Structuration des données .....	147
VIII.B. Interface MATLAB .....	149
VIII.C. Interface C++/QT .....	150
VIII.D. Interface HTML5/WebGL.....	151

IX. Conclusion.....	155
IX.A. Limites.....	155
IX.A.1. Protocole d'acquisition des photographies.....	156
IX.A.2. Intégration d'images de détails.....	156
IX.A.3. Densité du nuage de points et objets cachés.....	156
IX.A.4. Gestion d'états temporels .....	156
IX.B. Perspectives de recherche.....	157
IX.B.1. Evolutions futures.....	157
IX.B.2. Application mobile pour l'annotation sur site .....	158
Bibliographie .....	161
Liste des publications .....	179
Table des illustrations .....	181
Annexe A : Structuration des Données .....	185
Annexe B : Interfaces d'Annotations .....	189
Annexe C : Modèle de Données Nubes Data .....	193
Annexe D : Fonctionnement de Aïoli.....	199



## I. Introduction

### I.A. Contexte

Aujourd'hui, dans le domaine de la documentation du patrimoine culturel, grâce aux avancées de l'informatique, les experts bénéficient de nouveaux scénarii pour étudier le patrimoine culturel. L'introduction des technologies numériques dans les méthodes de documentation a permis de concevoir des nouveaux outils pour l'acquisition et la gestion des informations collectées pour des études pluridisciplinaires. Ces outils commencent à s'affirmer aujourd'hui comme des supports privilégiés pour la description, l'analyse et la compréhension des objets d'étude.

Les méthodes de numérisation et de reconstruction 3D ont notamment connu des évolutions importantes avec le développement des scanners laser 3D et de logiciels de reconstruction 3D à partir d'images photographiques. C'est pourquoi aujourd'hui le modèle 3D est à considérer comme le support graphique à privilégier pour la documentation et la diffusion culturelle. Mais, si les avancées récentes en matière de reconstruction 3D permettent d'obtenir des représentations graphiques métriquement précises et géométriquement et visuellement très détaillées, une véritable exploitation de ces nouvelles représentations au sein des méthodologies d'étude et des modes opératoires propres aux chercheurs et aux professionnels du patrimoine semble encore lointaine.

Par ailleurs, la 3D est à considérer comme une des nombreuses sources iconographiques manipulées par les spécialistes du patrimoine. L'iconographie (peinture, croquis, photographies, dessin, ... etc.) occupe certainement une place centrale au sein des ressources documentaires généralement utilisées pour la connaissance d'un objet patrimonial. Les ressources iconographiques constituent en effet des témoignages essentiels pour la connaissance de la géométrie et des matériaux des objets tout au long de leur histoire.

Parmi tous les types de sources iconographiques, les photographies occupent une place de tout premier plan pour la documentation de l'état de conservation et des transformations temporelles d'un objet patrimonial. Au vu de sa grande facilité d'utilisation et de sa diffusion au sein des communautés scientifiques et professionnelles du patrimoine, la photographie peut aisément être considérée comme l'instrument le plus efficace pour collecter, observer, analyser, enregistrer et diffuser de l'informations sur les objets patrimoniaux.

Les photographies contiennent un grand nombre d'informations à propos de l'objet représenté tant au niveau des couleurs qu'au niveau des formes. Les photographies peuvent être facilement annotées et peuvent servir de support pour identifier (et éventuellement extraire des informations sur) des dimensions, des matériaux, des phénomènes d'altération des surfaces, etc. Enfin, grâce aux progrès récents de la photogrammétrie et de la vision par ordinateur, de collections d'images photographiques peuvent être orientées et corrélées automatiquement afin de produire des représentations 3D des scènes photographiées.

Par conséquent, la photographie se révèle (encore une fois dans son histoire) comme un moyen essentiel pour documenter et étudier la morphologie et de l'état de conservation des objets patrimoniaux. Cependant, afin d'exploiter pleinement les nouvelles possibilités d'analyse offertes par ces avancées récentes, ainsi que d'obtenir une documentation exhaustive d'un objet patrimonial, un grand nombre d'images (des dizaines, des centaines ou des milliers en fonction de la taille et de la complexité de l'objet) est généralement à acquérir sur le terrain. Chacune de ces images peut représenter un point de vue important, un aspect particulier, un détail essentiel, ou encore la mémoire d'une observation sur le terrain qui

contribue à la connaissance et à la compréhension de l'objet et de son interaction avec le temps. Ainsi, au vu du grand nombre d'images acquises et manipulées pendant l'étude d'un objet patrimoine, il semble essentiel d'étudier aujourd'hui les conditions au travers lesquelles les annotations faites par un utilisateur sur toutes ces images puissent être propagées et corrélées au sein d'un référentiel spatial et temporel.

La prise en compte de ce problème, d'une manière globale et cohérente sur le plan méthodologique, consiste en une avancée claire par rapport aux outils et méthodes existantes. En effet, si d'un côté des solutions d'annotation d'images existent et commencent à se diffuser dans les secteurs du web sémantique, ces méthodes (manuelles, semi-automatiques ou automatiques) se révèlent inadaptées à la description de la complexité des objets patrimoniaux (niveaux de détails, variété de formes et de matériaux, variété d'aspects à analyser, etc...). Parallèlement, même si le domaine de l'annotation d'images permet d'aborder le développement de nouveaux moteurs de recherche pour le catalogage, la richesse informationnelle des images ouvre des nouveaux scénarii d'analyse et de corrélation à explorer à l'interface entre la vision par ordinateur et l'analyse géométrique.

## I.B. Objectif

L'objectif de ce travail de thèse est de définir une approche de sémantisation 2D/3D basé sur le référencement spatial et d'annotations d'images garantissant une continuité d'informations depuis l'acquisition jusqu'à la construction de représentations 3D sémantiquement enrichies.

Il s'agit donc d'identifier **des solutions pour corréler (d'un point de vue spatial, temporel et sémantique) des jeux d'annotations au sein d'un jeu d'images**. Cette problématique suppose que l'on soit capable, d'une part, de gérer les informations relatives au relevé photographique et, d'autre part, de structurer et de représenter ces informations et connaissances. L'approche a été définie suite à une recherche bibliographique sur les méthodes actuelles d'annotations.

Le travail effectué prend en compte les problèmes suivants :

Concernant **la spatialisation des informations**, l'approche prend en compte l'utilisation des moyens actuels d'acquisition de données 3D par photogrammétrie ou par acquisition laser. Ces deux méthodes permettent d'obtenir des nuages de points de l'objet relevé. La photogrammétrie calcule notamment les positions relatives des images les unes par rapport aux autres mais aussi par rapport au nuage de point.

Concernant **la gestion des informations**, l'approche prend en compte les différents niveaux d'informations 2D et 3D et de structurer les données en fonction. Les données produites par photogrammétrie permettent d'une part de spatialiser les images mais aussi de créer une relation bijective entre des informations 3D (nuage de point, normales, etc.) et des informations 2D (images). Les données obtenues par lasergrammétrie permettent aussi d'obtenir une telle relation. Cette relation sert de base pour gérer la propagation des informations entre les différentes représentations et permettre ainsi de mettre en correspondance différents types de support d'informations.

Concernant **l'enrichissement des informations**, l'approche est capable de supporter les différentes représentations de l'objet (2D et 3D) et des outils permettent de générer des informations à lier aux images ou aux objets des images et de les gérer. Ces différentes informations peuvent correspondre à différents niveaux de consultation et de compréhension en fonction du profil des utilisateurs et de leurs préoccupations. Le système permet donc d'associer des attributs sémantiques aux parties des images afin de décrire les formes et d'enregistrer les observations.

Concernant **l'exploitation des informations**, l'approche assure la restitution et la consultation des données par l'utilisateur par visualisation sur les images des résultats. Cependant au vu du lien établi entre les informations 2D et les informations 3D, une visualisation complémentaire en 3D des informations peut être aussi mise en place afin de compléter l'approche.

Plusieurs types d'**interface** intégrant les divers aspects de l'approche seront présentés.

### I.C. Portées, limites, apports principaux

Ce sujet se situe dans le domaine de la sémantisation des ressources iconographiques relatives à l'étude et à la documentation des objets patrimoniaux. Au vu de la quantité disponible de sources iconographiques (dessins, peintures, croquis, etc) pour un objet, il a été décidé de ne se concentrer pour l'instant sur l'utilisation seule de photographies numériques actuelles (acquises par des capteurs numériques), l'ajout de photographies historiques étant envisagé pour le futur mais nécessitant de plus amples recherches.

La problématique générale comprend plusieurs préoccupations puisque l'idée est de s'intéresser à la relation entre des sources bidimensionnelles et des représentations tridimensionnelles, tout en incluant un aspect de description sémantique devant être relatif aux deux dimensions de représentations. L'apport principal est donc d'avoir défini un processus permettant, par l'annotation d'images, de mettre en relation ces aspects sur le plan géométrique et informatique.

L'approche se base sur un processus par phases. Chacune de ses phases relève de problèmes spécifiques qui seront traités à travers la définition de méthodes et de développements informatiques.

En conséquence, ce document est organisé en plusieurs parties présentant premièrement l'utilisation des données, deuxièmement l'approche proposée, puis la définition d'outils innovants pour l'aide à la sélection des informations et enfin les apports informatiques.

### I.D. Cadre de la thèse

#### I.D.1. Culture 3D Clouds

Cette thèse s'inscrit dans le projet CULTURE 3D CLOUDS (C3DC) [[C3DC](#)]. Ce projet est financé sur le Fond Unique Interministériel (FUI).

Ce projet a pour objectif de développer une plateforme d'informatique en nuage (Figure 1) pour la numérisation 3D, la documentation, la conservation et la diffusion du patrimoine culturel.



Le projet est fondé sur un partenariat comprenant des instituts culturels (RMN<sup>1</sup>, CMN<sup>2</sup>), des laboratoires de recherche (MAP<sup>3</sup>, IGN<sup>4</sup>, ETIS<sup>5</sup>, INRIA<sup>6</sup>, CEA-LIST<sup>7</sup>, Institut Telecom<sup>8</sup>) ainsi que des entreprises (HPC Project, BeIngenious, Reciproque, ValEISTI, Silkan) spécialisées dans le développement informatique ainsi que dans la numérisation 3D et la valorisation du patrimoine culturel, et un institut de formation (EISTI<sup>9</sup>).

L'architecture de la plateforme se compose de quatre parties principales en relation avec les activités qu'elle proposera :

- **Numérisation** : Ce sous-projet se concentre sur le développement des fonctionnalités de numérisation 3D de la plateforme afin d'offrir des solutions aux photographes pour pratiquer la numérisation 3D.
- **Catalogage** : Ce sous-projet se concentre sur le développement des fonctionnalités de la plateforme pour offrir aux institutions culturelles un moyen de stocker et de mettre à disposition des numérisations 3D en s'intéressant notamment aux problèmes d'archivage, d'indexation, de recherche, et d'exploitation de ces représentations.
- **Expérimentation** : Ce sous-projet se concentre sur la mise en œuvre de valorisation permettant au public un accès aux contenus numérisés tout en permettant de bénéficier de retour d'expérience des utilisateurs potentiels.
- **Diffusion et communication** : Ce sous-projet se concentre sur l'organisation d'actions de communications et de diffusion autour du projet.

Le sujet de cette thèse s'intègre dans la partie de catalogage du projet.

---

<sup>1</sup> RMN : Réunion des Monuments Nationaux

<sup>2</sup> CMN : Centre des Monuments Nationaux

<sup>3</sup> MAP (CNRS/MCC) : Modèles et Simulations pour l'Architecture, l'Urbanisme et le Paysage

<sup>4</sup> IGN : Institut National de l'Information Géographique et Forestière

<sup>5</sup> ETIS (CNRS/UCP) : Equipe Traitement de l'Information et Systèmes

<sup>6</sup> INRIA (Geomatrica) : Institut National de Recherche en Informatique et Automatique

<sup>7</sup> CEA-LIST : Laboratoire d'Intégration des Systèmes et des Technologies

<sup>8</sup> Institut Telecom : Département Artemis, IT-DS

<sup>9</sup> Ecole Internationale des Sciences du Traitement de l'Information

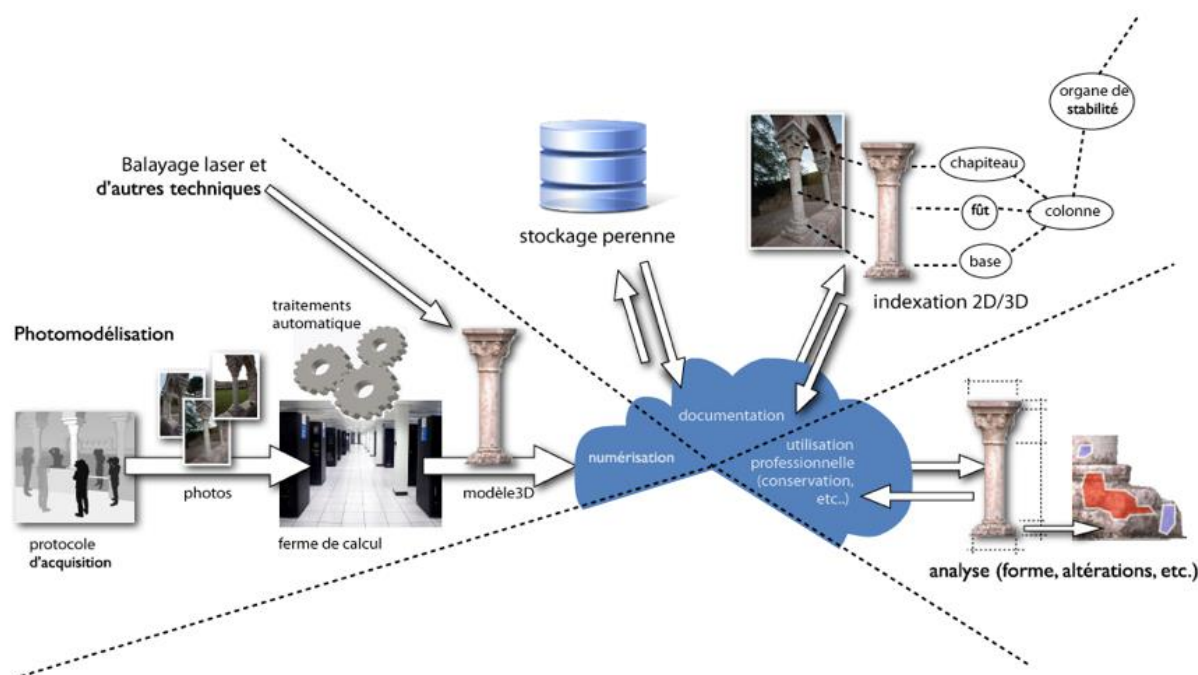


Figure 1 : Schéma de fonctionnement de la plateforme C3DC

### I.D.2. Monumentum : MODélisation NUMérique et gEstion de doNnées pour la conservaTion de strUctures Maconnées

Les travaux développés ont aussi été intégrés au projet Monumentum qui est un projet financé par l'Agence Nationale de la Recherche (ANR).

Ce projet a pour but de concevoir et de développer une plateforme logicielle ouverte et extensible pour la capitalisation et la gestion de connaissances favorisant la compréhension et l'analyse des phénomènes de dégradation qui affectent les édifices historiques.

Ce projet est constitué sur un partenariat entre différents acteurs de plusieurs domaines distincts (CICRP<sup>10</sup>, LMGC<sup>11</sup>, IGN, MAP, CRMD<sup>12</sup>, DRAC<sup>13</sup>).

Le projet se divise en quatre phases distinctes pour lesquels les différents acteurs interviendront à différents niveaux (Figure 2):

- **Acquisition et fusion des données :** Ce sous-projet se concentre sur l'aspect d'acquisition des données et sur les possibilités de fusion de ces données. Ce sous-projet s'inspire des travaux élaborés dans le cadre du projet C3DC.
- **Modélisation et structuration sémantique :** Ce sous-projet se concentre sur l'identification morphologique de l'objet à travers une modélisation et une structuration sémantique par annotation d'images. Ceci permet de passer de données brutes à un modèle 3D intelligible au travers de la définition d'aspects sémantiques caractérisant l'ouvrage, de relations topologiques entre les éléments et enfin de la géométrie.

<sup>10</sup> CICRP : Centre Interdisciplinaire de Conservation et Restauration du Patrimoine

<sup>11</sup> LMGC : Laboratoire de Mécanique et Génie Civil

<sup>12</sup> CRMD : Centre de Recherche sur la Matière Divisée

<sup>13</sup> DRAC : Direction Régionale des Affaires Culturelles

- **Modélisation et analyse du comportement** : Ce sous-projet se concentre sur l'élaboration de simulation mécanique du comportement de l'objet à partir des résultats des deux premières phases.
- **Analyse et aide à la décision** : Ce sous-projet se concentre sur la mise en relation des différentes données afin de permettre le suivi de conservation de l'objet patrimonial et d'offrir des outils pour l'aide à la décision.

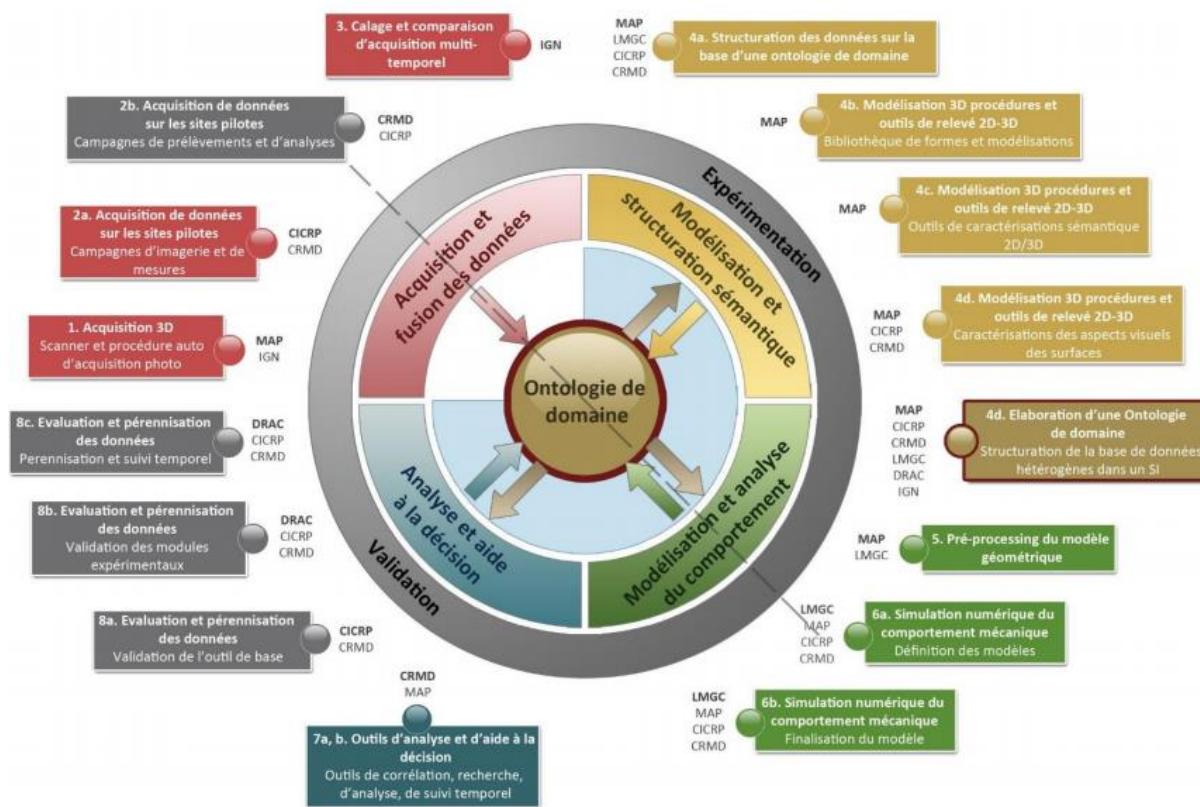


Figure 2 : Contributions scientifiques du projet Monumentum

## I.E. Terrains d'expérimentations

Pour valider la dimension méthodologique et informatique de cette approche, plusieurs sujets d'étude, rencontrés sur des expériences conduites au sein du laboratoire MAP-GAMSAU dans le cadre de plusieurs projets, ont été utilisés.

L'approche a notamment cherché à être appliquée sur des objets à différentes échelles allant de l'objet (hauteur d'une dizaine de centimètres) jusqu'au bâtiment complet. Cet aspect d'échelle peut aussi être considéré au sein d'un même objet où une zone aura été plus détaillée que l'ensemble de l'objet patrimonial.

Des cas d'objets considérés généralement comme des représentations planes, tel que des peintures, ont aussi été considéré. En effet, si l'aspect des problèmes de conservation est intégré sur ces objets, il ne faut pas seulement considérer ce que la scène représentée renvoie mais aussi le support et les variations d'épaisseurs de la couche picturale, dues à la technique employée ou à l'état de dégradation, ce qui introduit un aspect tridimensionnel au problème.

Les images de ces différents terrains d'expérimentation illustrent différentes parties du manuscrit.

## I.F. Structure du document

Ce document est structuré de la façon suivante. Dans le chapitre 2, on s'intéresse aux différents outils d'acquisition dont les experts disposent pour leur relevé et les informations qui sont nécessaires pour l'analyse d'un objet patrimonial. La dernière partie de ce chapitre s'intéresse notamment à la comparaison entre les deux méthodes principales de relevé.

Le chapitre 3 se focalise sur l'étude bibliographique relative aux méthodes actuelles de sémantisation pour les objets patrimoniaux en s'appuyant plus particulièrement au domaine de la documentation d'objets patrimoniaux.

A partir de l'analyse des travaux décrits au sein du chapitre 3, le chapitre 4 présente l'approche de sémantisation d'images développée. Cette approche cherche à s'adapter aux problèmes de la documentation patrimoniale mais pourrait être aisément applicable à d'autres domaines. Ce chapitre se termine avec la mise en évidence des problèmes relatifs aux différentes phases de l'approche et qui seront abordés de façon spécifique dans les chapitres suivants.

Les chapitres 5 et 6 s'intéressent aux problèmes identifiés dans la définition de l'approche générale. Chaque chapitre traite d'une partie du processus pour l'organisation et l'exploitation des informations. Le chapitre 5 traite du problème de la spatialisation des informations. Il aborde tout d'abord sur les techniques actuelles d'orientation d'images dans l'espace puis se concentrera ensuite sur la mise en relation des informations 2D et les informations 3D. Le chapitre 6 est centré sur l'exploitation de cette relation pour établir une continuité d'informations entre les différentes représentations. Il traite notamment le problème de mise à jour de la base d'images ainsi que les différentes préoccupations qu'une telle mise à jour peut entraîner.

Le chapitre 7 définit des outils pour l'aide à la saisie et à l'analyse. Ce chapitre s'intéresse d'abord aux méthodes de segmentation existantes pouvant permettre de définir des zones de caractéristiques proches, puis sur l'exploitation d'outils semi-automatique pour l'aide à la sélection. Il finit par la définition d'outils quantitatifs pour l'aide à l'analyse.

Le chapitre 8 présente les différentes implémentations d'interface du système.

Le chapitre 9 présente la conclusion, les limites du travail réalisé et les perspectives de recherche.



## II. Outils pour l'acquisition de données

Les nouvelles technologies de l'information offrent aujourd'hui de nouvelles possibilités aux experts pour acquérir de grandes quantités de données. Les technologies élaborées pour l'acquisition de ses données sont aujourd'hui plébiscitées pour la définition de nouvelles méthodologies d'analyse. En effet, les données relevées contiennent directement un certain nombre d'informations utiles à l'analyse ou peuvent servir de base pour en produire d'autres sur des méthodes qui sont généralement automatiques.

Ce chapitre s'intéresse donc, tout d'abord à la phase d'acquisition des données avec la description des différents capteurs à disposition des experts, puis sur l'enregistrement de ces données qui est fonction du type de projection considérée par le dispositif d'acquisition. Enfin la dernière partie s'intéresse aux informations qui peuvent être extraites de ces données que ce soit de manière directe ou par traitement de ces données.

### II.A. Les différents capteurs pour l'acquisition

Les experts disposent aujourd'hui de plusieurs types de capteurs pour relever l'état d'un objet patrimonial en fonction des besoins. Ces capteurs permettent de rendre en deux dimensions ou en trois dimensions les informations acquises.

L'acquisition des données par les capteurs peut se faire de manière passive [[Remondino F. et al, 2006](#)] ou de manière active [[Vosselmant G. et al, 2010](#)]. Les capteurs passifs captent les émissions naturelles de l'environnement physique. Les capteurs actifs, pour leur part, nécessitent une source externe. Ainsi ces capteurs sont généralement couplés à un transmetteur émettant un signal, qui sera capté par le récepteur après réflexion de ce signal sur l'objet à acquérir.

#### II.A.1. Capteurs photographiques

Les capteurs photographiques sont des composants électroniques photosensibles convertissant un rayonnement électromagnétique en un signal analogique. Ces capteurs utilisent l'effet photoélectrique pour restituer la lumière en image. Ils sont constitués de cellules photovoltaïques qui mesurent l'intensité de la lumière [[Lanh T.V. et al, 2007](#)]. Cette intensité lumineuse est ensuite transformée en courant électrique. Chaque point du capteur, qui compose une partie d'un pixel, enregistre l'intensité lumineuse pour produire une image.

Un capteur est composé de deux éléments superposés :

- Un élément composé de cellules photosensibles qui possèdent chacune trois filtres étant chacun réceptif à une couleur (rouge, vert et bleu)
- Un élément composé d'un dispositif à transfert de charge

On trouve aujourd'hui deux types de capteurs, le CCD (*Charged Coupled Service*) [[Janesick J.R., 2001](#)] et le CMOS (*Complementary Metal Oxyde Semi-Conductor*) [[Bigas M. et al, 2006](#)]. Ce dernier est le plus répandu sur les appareils photo numériques de type Reflex, tandis que le CCD équipe plutôt les compacts. La différence entre les deux tient dans le transfert de la charge. Le CCD transfère la charge à un collecteur qui va à son tour transférer

l'ensemble des charges vers le convertisseur tandis que le CMOS transfère la charge directement au convertisseur.

Les cellules photosensibles composant le capteur ne sont sensibles qu'à l'intensité et non à la couleur. C'est pourquoi un filtre est placé devant chaque photodiode, ne laissant passer qu'une seule des composantes (rouge, vert, bleu). Le filtre le plus couramment utilisé est le filtre de Bayer (Figure 3) [Bayer B., 1976] composé à 25% de filtres rouges, 25% de filtres bleus et 50% de filtres verts de sorte à imiter la physiologie de l'œil humain plus sensible aux longueurs d'ondes correspondant au vert. Le processeur d'image associé au capteur photosensible combine ces trois couleurs primaires pour créer, à la suite de plusieurs étapes de traitement [Ramanath R. et al, 2005], une image couleur. Aujourd'hui la plupart des constructeurs utilisent leur propre adaptation de grille.

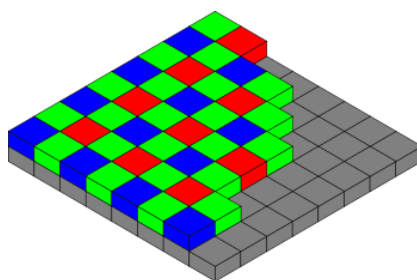


Figure 3 : Disposition de Bayer des filtres de couleur

Un rayonnement électromagnétique est une variation de champs électriques et magnétiques qui peut être décrit sous la forme d'une onde électromagnétique [Bonn F. et al, 1992]. Une analyse spectrale permet de décomposer cette onde en ondes monochromatiques de longueur d'ondes  $\lambda$ . Le spectre électromagnétique est la répartition de ces différentes longueurs d'ondes et peut théoriquement s'étendre de zéro à l'infini. Ce spectre est divisé en plusieurs classes de rayonnements (Figure 4) qui nécessitent chacune des moyens particuliers pour être étudiées. Ainsi la lumière du visible, que l'humain est capable de discerner, ne représente qu'une petite portion de ce spectre.

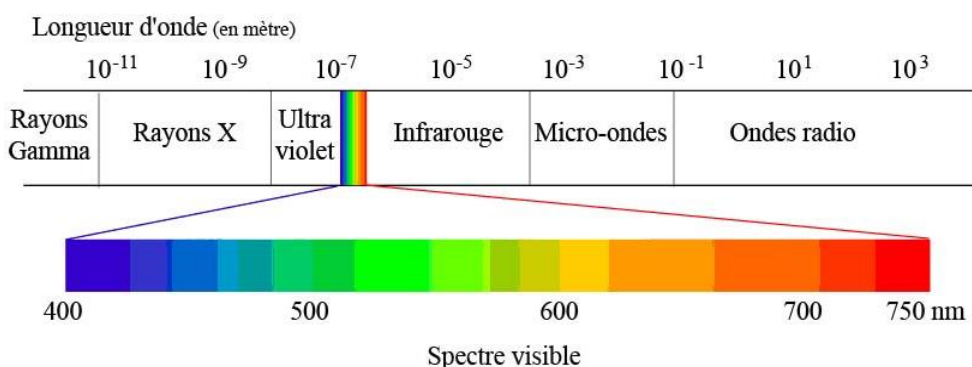


Figure 4 : Spectre électromagnétique et classes de rayonnement

Les rayonnements électromagnétiques reçus par les capteurs ne se réduisent ainsi pas seulement aux rayonnements du visible. Toutes les longueurs d'ondes du spectre électromagnétique peuvent potentiellement être captées. Cependant les appareils de photographies numériques sont équipés de filtres [Lanh T.V. et al, 2007] afin d'éviter de capter ces rayonnements pouvant atténuer le rendu de l'image numérique mais il existe aussi



des appareils plus spécifiques destinés à ce type de relevé tels que la caméra multi-spectrale. Cette caméra enregistre en une seule prise de vue plusieurs longueurs d'onde à l'aide de filtres spécifiques [Helling S. et al, 2004] et en ne se limitant pas seulement aux longueurs d'onde du domaine visible. Cette technique permet d'extraire des détails beaucoup plus fins et la visualisation de détails non visibles à l'œil nu. Ainsi l'enregistrement, par exemple, des rayonnements infrarouges permettent de faire ressortir des détails contenus plus profondément dans l'objet. Une étude a notamment été réalisée sur La Joconde [Elias M. et al, 2008] à l'aide de ces techniques et a permis de pouvoir faire ressortir des détails inconnus ou connus seulement par les historiens de l'art.

Les capteurs photographiques ont différentes tailles qui vont de 24x36mm pour les Full Frame à 4,29x5,76mm pour certains compacts (Figure 5).

Full Frame	APS-H	APS-C	4:3	1/1.7"	1/2"	1/2,3"	1/2,5"
24x36mm	18,1x28,7mm	15,8x23,6 ∞ 14,8x22,1mm	13x17,3mm	5,7x7,6mm	4,8x6,4mm	4,62x6,16mm	4,29x5,76mm

Figure 5 : Les différentes tailles de capteurs photographiques

La taille du capteur a une incidence directe sur l'angle de champ. Un grand capteur a ainsi un plus grand champ qu'un capteur plus petit mais aussi une profondeur de champ moins importante.

Si le nombre de pixels est élevé, la résolution d'une photographie est bonne. Cependant, la densité des cellules photosensibles par millimètres carré du capteur a une forte influence sur la qualité des images. Cette qualité dépend de l'intensité de la lumière que le capteur peut recevoir sur chacune de ses cellules photosensibles. Ainsi un grand capteur peut potentiellement contenir plus de cellules photosensibles qu'un petit capteur et ainsi donner une meilleure qualité d'image.

Afin d'enregistrer une image sur le capteur, la lumière doit être focalisée sur un point précis. Les objectifs sont donc composés de groupes de lentilles se déplaçant les unes par rapport aux autres selon la distance de mise au point. Deux types d'objectifs sont disponibles : à focale fixe et à focale variable. La longueur focale prend part à la détermination de l'angle de champ (Figure 6). Une longueur de focale courte permet d'obtenir un champ plus grand qu'une longueur de focale longue.

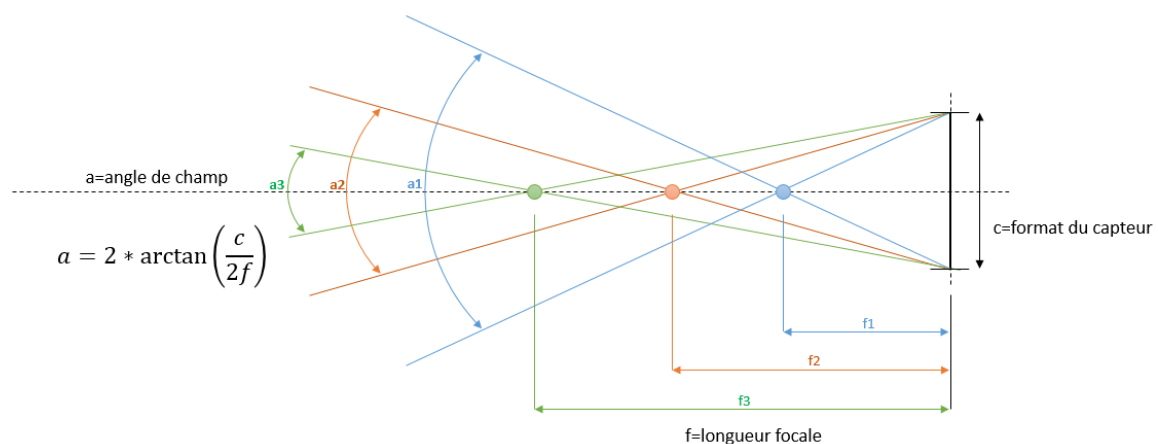


Figure 6 : Influence de la focale sur l'angle de champ pour un même capteur



## II.A.2. Scanners

Plusieurs dispositifs de scan ont été développés permettant la génération de nuage de points. Ces nuages de points peuvent être par la suite traités et intégrés à d'autres données telles que des images afin de créer un modèle 3D complet qui pourrait être utilisé dans de nombreuses applications comme le relevé, la navigation, la réalité augmentée, le suivi de déformation ... Bien que la précision et le détail des nuages de points obtenus soient de très bonne qualité, le coût de ces dispositifs reste un frein important à leur utilisation.

Le balayage laser terrestre est « *l'utilisation d'un dispositif basé au sol, qui utilise un laser pour mesurer les coordonnées tridimensionnelles d'une région donnée de la surface d'un objet de façon automatique, dans un ordre systématique et à un taux élevé de vitesse, près du temps réel* » [Boehler et al, 2002a]. Ainsi les modules de balayage laser permettent une acquisition directe en temps réel de milliers de coordonnées spatiales (x,y,z) par seconde.

En plus de ces coordonnées, certains dispositifs de balayage permettent aussi d'acquérir les valeurs d'intensité (nuance de 255 niveaux de gris) et les valeurs colorimétriques (valeurs RVB extraites à partir d'une caméra). Cependant, presque aucun système de balayage laser ne peut accomplir toutes les exigences d'un relevé d'architecture. En effet, un édifice peut présenter une variété d'objets qui s'étendent des éléments architecturaux aux décors sculptés.

### II.A.2.a. Scanner laser

Les scanners laser sont des capteurs actifs pouvant générer des nuages de points denses décrivant l'objet en détails. Ces dispositifs sont principalement montés sur un élément fixe [Zogg H.M., 2008]. Ils sont utilisés généralement pour mesurer des objets de différentes tailles allant du bâtiment et d'éléments architecturaux jusqu'à des petites pièces tels que des bijoux.

Les modules de relevé scanner 3D pouvant être utilisés pour le relevé d'édifice se divisent en deux catégories : les scanner longue-portée et les scanners à triangulation. Les deux principes sont différents soit en termes d'exactitude, soit en termes de portée. Une description des principes de fonctionnement ont été décrits par [Marbs A. et al, 2001].

#### II.A.2.a.i. Les scanners longue-portée

Les scanners à longue portée mesurent des angles horizontaux et verticaux et mesurent la distance en calculant le temps de vol ou en comparant le changement de phase de l'onde transmise et reçue d'un signal modulé (Figure 7).

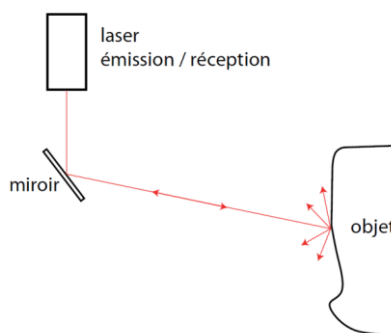


Figure 7 : Principe de fonctionnement des scanners longue portée [Marbs A. et al, 2001]

**Scanner à temps de vol :** Une impulsion est envoyée sur l'objet étudié. En comptant le temps nécessaire au trajet aller-retour de l'impulsion du faisceau laser réfléchi, la distance entre l'émetteur et la surface qui a réfléchi l'impulsion peut être calculée. Ce système détecte seulement un point à la fois dans la direction sur laquelle il est pointé. Pour scanner

l'environnement complet, le scanner doit donc répéter cette opération point par point et changer sa direction de vue entre chaque mesure.

*Scanner à décalage de phase* (Figure 8) : Sur ces scanners, le faisceau émis est modulé par une onde harmonique. Après réflexion sur l'objet, le signal reçu laisse apparaître un décalage dans la phase par rapport à l'onde émise. Connaissant la différence entre les deux phases, la distance peut être calculée.



Figure 8 : Exemple de scanner à décalage de phase : Faro Focus 3D

#### II.A.2.a.ii. Les scanners par triangulation

Les scanners par triangulation utilisent également la lumière laser pour sonder l'environnement. Cependant le retour est capté par une caméra CCD (Figure 9). La caméra, le laser et le point projeté sur l'objet forment ainsi un triangle. En fonction de la distance, le point apparaît à un endroit différent dans le champ de vision de l'appareil. A partir des paramètres connus de configuration du dispositif, les dimensions du triangle peuvent être déterminées et donnent la position du point laser. Une variante de ce dispositif passe par l'utilisation de deux caméras DTC positionnées à chaque extrémité de la base du module. Cette variante utilise le même principe géométrique mais permet d'obtenir un taux d'acquisition plus élevé.

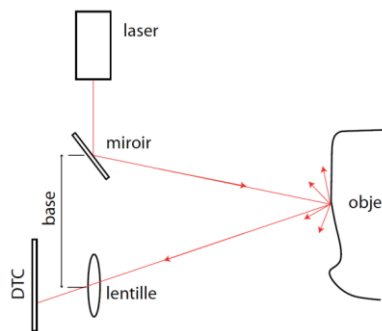


Figure 9: Principe de fonctionnement des scanners par triangulation [Marbs A. et al., 2001]

#### II.A.2.b. Scanner à lumière structurée

Ce type de scanner est composé d'un projecteur et de une ou plusieurs caméras. Le projecteur diffuse un motif sur l'objet (Figure 10) [Gupta M. et al, 2011]. La déformation du motif est captée par les caméras puis traduite par des méthodes de triangulation afin d'obtenir une représentation 3D.

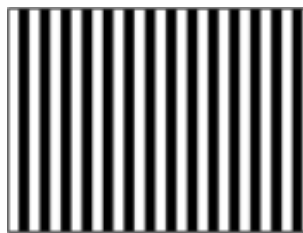


Figure 10 : Motif projeté sur l'objet [Gupta M. et al, 2011]

Le Kinect (Microsoft XBox), est un scanner à lumière structurée développé à l'origine pour des applications liées au divertissement mais intéresse aujourd'hui beaucoup la communauté scientifique, notamment de par son faible coût. Cette technologie est aujourd'hui utilisée dans la robotique ([El-Iaithy R.A. et al, 2012] [Henry P. et al, 2012] [Oliver A. et al, 2012]) mais commence aussi à intéresser les chercheurs pour quelques applications à l'étude d'objets patrimoniaux. Ici, la projection du motif se fait à l'aide d'une lumière infrarouge qui sera captée par une caméra infrarouge (Figure 11) [Zhang Z., 2012]. Les rayonnements infrarouges vont être réfléchis sur l'ensemble des surfaces touchées. Plus l'objet sera loin, plus la quantité de rayonnement infrarouge réfléchi sera faible. Ainsi la distance à l'objet sera estimée en fonction de l'intensité infrarouge reçue.

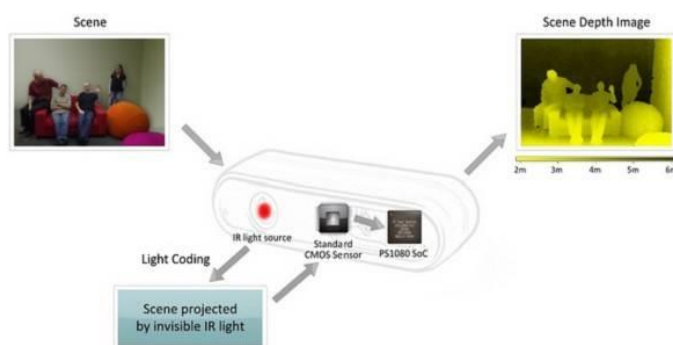


Figure 11 : Principe de la technologie Kinect

Cependant cette technique reste très largement insuffisante en termes de précision par rapport aux besoins pour l'analyse et la documentation d'objets patrimoniaux car elle a été prévue pour des environnements intérieurs très éclairés.

#### II.A.2.c. Précision et niveau de détails

Aujourd'hui de nombreuses compagnies proposent leur propre solution de laser scanner. Chacune de ces techniques possède des forces et des faiblesses rendant leurs utilisations appropriées pour des situations différentes. [Bohler W. et al, 2003] présentent une analyse comparative de la précision des scanner 3D.

L'avantage des scanner par temps de vol est leur longue portée permettant de travailler sur de longues distances de l'ordre plusieurs kilomètres avec une précision de l'ordre du centimètre. A l'opposé, les scanner par triangulation ont une faible portée de quelques mètres mais avec une précision de l'ordre du dixième de micromètre. Les méthodes par lumière structurée (en dehors du Kinect) nécessite de se mettre dans le noir et ne peuvent scanner que de petits objets mais ont une précision de l'ordre du micromètre. Ainsi dans le domaine de la documentation du patrimoine, le choix du procédé est donc dépendant de la taille de l'objet à numériser et en fonction de la distance d'acquisition.

La densité des points acquis est dépendante de l'orientation du module. Ajouté à cela, une acquisition d'un seul point de vue n'est jamais suffisante. Il existe toujours des zones d'ombres qui restent invisibles au scanner et ne sont donc pas relevées. C'est pourquoi il est nécessaire généralement de réaliser différentes acquisitions à partir de plusieurs points de vue. Ceci nécessite de pouvoir ramener les différents relevés dans un même système de coordonnées en utilisant différentes stratégies (points d'appui, cibles, ou procédures automatiques de superposition de surfaces) [Boehler W. et al. 2002b]. Le recalage des scans se fait ainsi de façon plus ou moins précise en fonction de la méthode choisie.

## II.B. Les types de projection

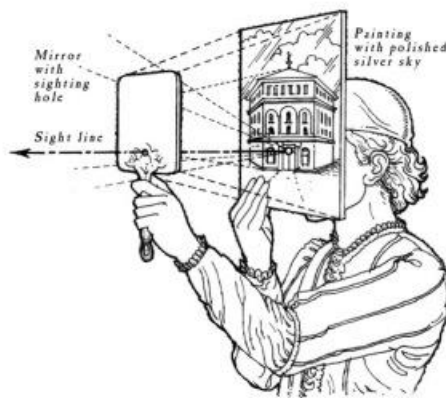
Les méthodes de représentation de la réalité ont beaucoup évolué au cours de l'histoire permettant de rendre cette réalité de diverse manière. La base de ses représentations s'inspire de méthodes géométriques de construction et de contrôle sur l'espace et les formes. Ainsi, différents types de projections ont été utilisés au cours des siècles pour la représentation des formes.

Chacun des capteurs présentés précédemment repose sur un type de projection spécifique lié au type d'acquisition. Ainsi les appareils photographiques enregistrent le réel par une transformation perspective. Les scanner-laser à longue portée utilisent, pour leur part, des projections panoramiques là où les scanner-laser à courte portée reposent sur une projection perspective. D'autre part, il est aussi possible d'obtenir des images reposant sur une projection orthographique par des opérations de redressement d'images.

### II.B.1. Projection perspective et images photographiques

Le principe de la projection perspective repose sur l'obtention d'un effet visuel semblable à celui perçu par l'œil humain. Le principe est de considérer la distance entre l'objet et le point d'observation. Les objets lointains apparaissent plus petits que des objets proches dans cette représentation.

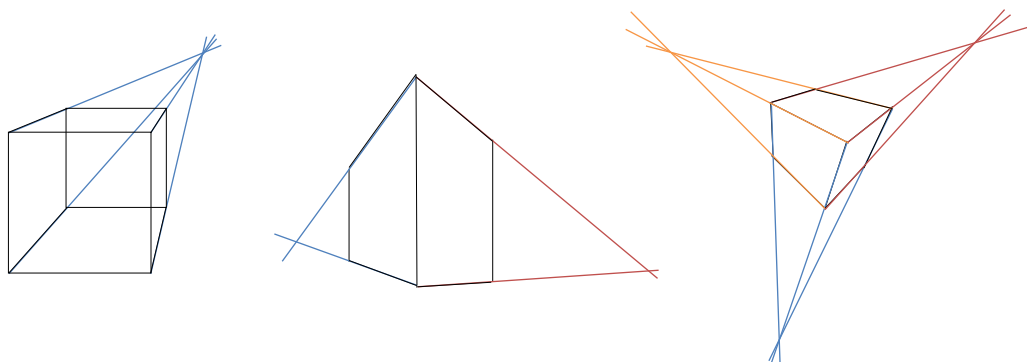
Cette méthode a été codifiée par Filippo Brunelleschi au XVe siècle à travers deux expériences (Figure 12). Il observa que, à partir d'un seul point de vue fixe, les lignes semblent converger vers un point unique. Ce constat lui a permis de calculer la taille que les objets devraient avoir sur une peinture pour la rendre réaliste. Cette méthode est devenue le modèle standard de rendu réaliste à la Renaissance. Cependant la construction géométrique de l'espace n'étant pas rigoureuse sur ses représentations, elles ne peuvent pas être utilisées comme bases pour l'extraction de mesure métrique de l'objet représenté.



*Figure 12 : Dispositif de perspective de Filippo Brunelleschi*

Un objet situé à l'infini est assimilable à un point, appelé point de fuite. Deux droites parallèles convergent sur le même point de fuite. En fonction du point de vue choisi, une projection peut avoir jusqu'à 3 points de fuite (Figure 13) [Fyfe D., 2007] :

- un seul point si la projection est parallèle à une des faces du cube de référence.
- deux points de fuite si la projection est perpendiculaire à une des faces du cubes de référence.
- trois points de fuite si la projection n'est perpendiculaire à aucune des faces du cube de référence.



*Figure 13 : Les différentes projections perspectives avec un, deux ou trois points de fuite*

Le principe des appareils photographiques permet de se placer dans ce cas de projection. La projection photographique peut s'approximer par une perspective dont le point de vue est le centre du diaphragme, assimilé au centre optique de l'objectif [Gabriel C., 2015]. L'enregistrement des données vérifie donc une loi géométrique qui permet de transformer l'espace 3D en espace 2D par la projection perspective. Cependant, les différentes lentilles composant l'appareil photographique introduisent des aberrations géométriques de champ ou d'inclinaison appelées distorsion [Taillet R. et al, 2009]. Cette distorsion apparait lorsque les conditions de Gauss ne sont plus respectées et provoque des déformations sur l'image qui peuvent être « en coussinet » ou en « barillet » (Figure 14). Ces déformations sont généralement plus visibles sur les lignes droites.

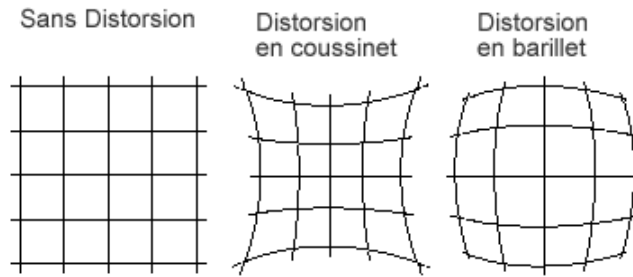


Figure 14 : Distorsions

Ainsi il faut, en plus de la loi géométrique définissant la projection perspective, introduire une loi de distorsion modélisant le déplacement des points sur l'image. Ces deux principes permettent de pouvoir définir les transformations subies par les rayons lumineux avant d'atteindre le capteur de l'appareil photographique et ainsi pouvoir enregistrer les données sous forme d'image.

### II.B.2. Projection orthogonale et redressement d'images

Une projection parallèle est une forme de représentation en deux dimensions d'un objet en trois dimensions. Cette représentation a pour objectif de conserver l'impression de volume et de relief en conservant les rapports entre toute longueur prise selon une direction de l'espace. Ainsi cette projection est très utile pour mesurer des distances.

A la différence de la projection perspective, cette projection ne cherche pas à rendre fidèlement l'objet tel que le voit l'œil humain. Quel que soit la distance de l'objet par rapport au plan de projection, il aura toujours la même taille sur la projection. La projection parallèle peut être considérée comme une projection perspective où les points de fuite sont à l'infini et est définie par une direction de projection sur un plan.

La projection orthogonale est une projection parallèle particulière où un des axes de l'objet est parallèle à la direction de projection ce qui permet d'obtenir des vues de face, de dessus, ou de profil d'un objet (Figure 15).

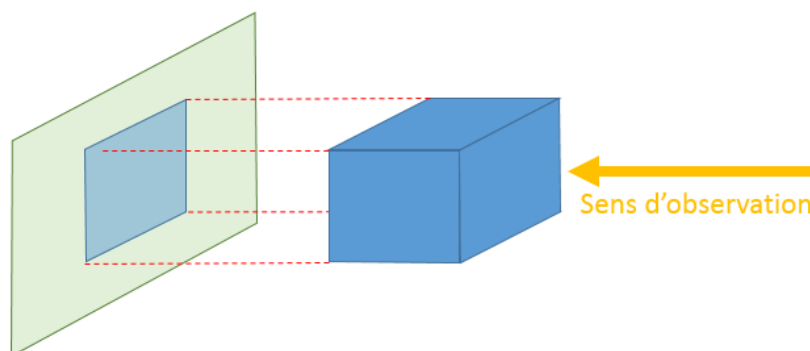
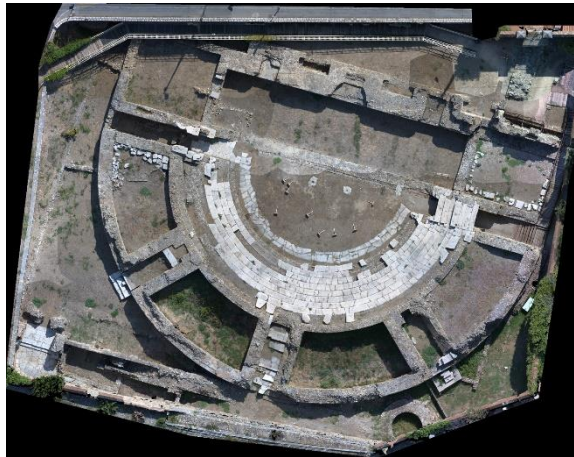


Figure 15 : Principe de la projection orthogonale

Ce type de projection est le plus utilisé dans l'élaboration de dessins techniques et dans les dessins d'édifice (comme des plans d'élévation). Ce principe se retrouve aussi dans les cas de d'images redressées ou d'orthophotographie.



Le redressement d'images repose sur une transformation géométrique convertissant une projection perspective en projection orthogonale [Kraus K. et al, 1997]. Il est ainsi nécessaire d'établir une relation spatiale entre la caméra et un plan représentant la face dominante de l'objet. Cette relation permet la projection des points de la photographie sur le plan. Cependant en fonction de l'objet, il peut être nécessaire de définir plusieurs plans pour chacune des parties de l'objet. Dans ce cas, on parle d'orthophotographie (Figure 16). Sur des objets très étendus, par exemple des façades, la génération d'une orthophotographie reposera sur l'utilisation de plusieurs images redressées. L'intérêt de l'utilisation d'orthophotographie, pour le domaine qui nous intéresse, est de disposer d'un support alliant information métrique et information visuelle très détaillée.



*Figure 16 : Exemple d'orthophotographie [Saleri R. et al, 2013]*

### II.B.3. Projection panoramique et capteurs 3D

Les images panoramiques sont un type de représentation qui consiste à créer des images avec des champs extrêmement larges [Conrat J.M. et al, 2010]. L'angle de prise de vue peut atteindre 360° en horizontal et 180° en vertical. Ces images peuvent être créées :

- soit avec un appareil photographique panoramique
- soit avec un appareil photographique numérique et un logiciel pour assembler plusieurs vues.

L'acquisition d'une image panoramique à partir de plusieurs images repose sur la notion de recouvrement. L'assemblage d'une image panoramique consiste à connecter les caractéristiques communes des images (Figure 17) [De Luca, 2009]. Des méthodes permettent d'effectuer cette mise en correspondance des images de façon de plus en plus automatique [Bodin B. et al, 2004]. Pour faciliter la recherche de zones communes, une règle de recouvrement doit être définie entre les images et le mieux, pour l'acquisition de ces images, est d'utiliser un trépied afin que les images aient toutes le même centre de projection. A partir de ces points homologues, les paramètres d'orientation et les caractéristiques optiques peuvent être estimées.

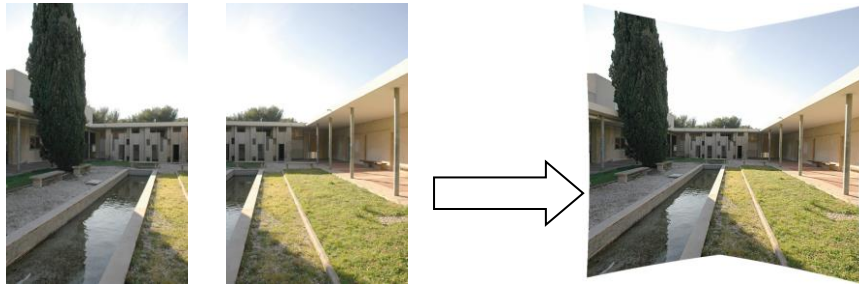


Figure 17 : Assemblage de deux images [De Luca, 2009]

Ensuite les images sont projetées et fusionnées dans l'espace panoramique. Dans le cas où le champ de visibilité est inférieur à  $180^\circ$ , il est possible d'utiliser le même type de projection que pour les photographies, c'est-à-dire une projection perspective (voir II.B.1. **Projection perspective**). Dans les autres cas (où l'angle du champ de vision est très important ou supérieur à  $180^\circ$ ), le problème consiste à rendre des informations tridimensionnelles dans un espace bidimensionnel (image). Plusieurs types de projection sont proposés pour ces cas : projection cylindrique, projection sphérique, projection cubique [De Luca, 2009].

#### II.B.3.a. Projection cylindrique

Sur cette projection, les directions de l'espace sont projetées sur un cylindre (Figure 18). L'image est projetée à partir du centre optique sur un cylindre dont la hauteur correspond à la dimension verticale de l'image et dont la circonférence correspond à la dimension horizontale de l'image.

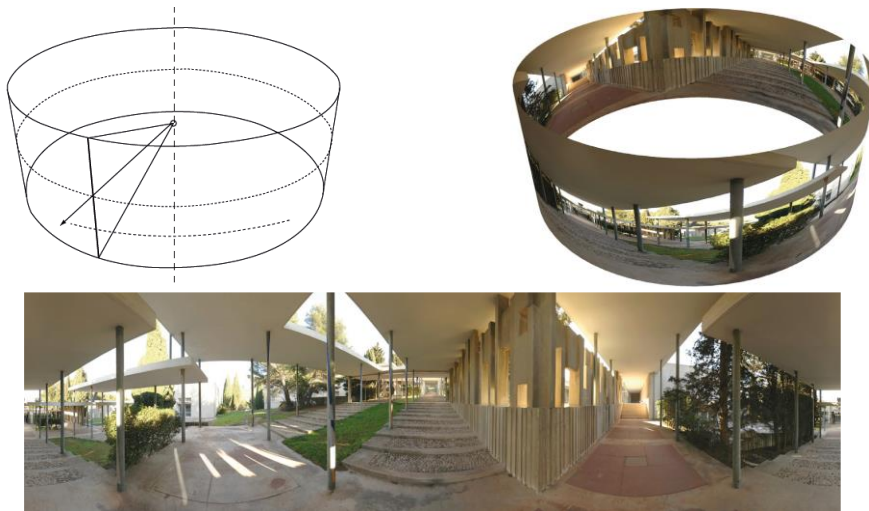


Figure 18 : Projection cylindrique d'une scène [De Luca, 2009]

Ce type de projection limite l'étendue verticale de l'image. Elle est donc plus adaptée pour la représentation d'environnements situés sur une ligne d'horizon sans élévations importantes.

#### II.B.3.b. Projection sphérique

Sur cette projection, les directions de l'espace sont projetées sur une sphère (Figure 19). L'image est projetée à partir du centre optique sur une sphère dont la coordonnée horizontale correspond à la dimension horizontale de l'image (distribuée sur  $360^\circ$ ) et dont la coordonnée verticale correspond à la dimension verticale de l'image (distribuée sur  $180^\circ$ ).



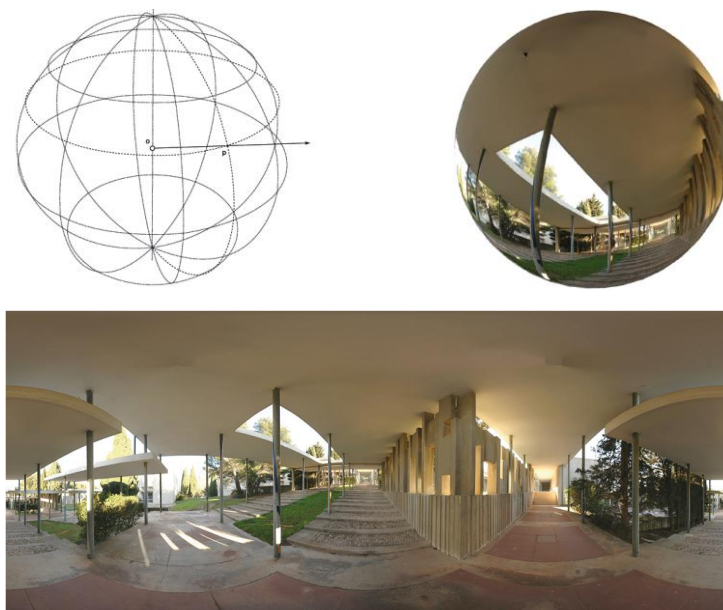


Figure 19 : Projection sphérique d'une scène [De Luca, 2009]

Cette projection a l'avantage de représenter la totalité de l'espace environnant. La position d'un pixel sur la sphère dépend d'une valeur angulaire horizontale et d'une valeur angulaire verticale. Les lasers scanner pouvant acquérir les informations colorimétrique de la scène rendent cette dimension colorimétrique sous la forme d'une image sphérique liée aux informations géométriques relevées.

### II.B.3.c. Projection cubique

Sur cette projection, les dimensions de l'espace sont projetées sur les six faces d'un cube (Figure 20). Ces projection se fait de manière « classique » à l'aide d'une projection perspective (II.B.1. **Projection perspective**). Une projection cubique correspond à une représentation en déplié des faces du cube.

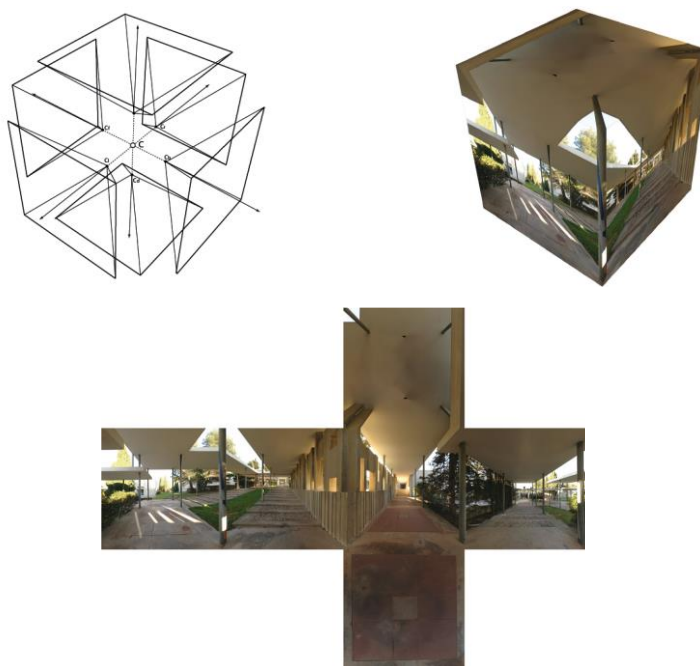


Figure 20 : Projection cubique d'une scène [De Luca, 2009]

Comme pour la projection sphérique, cette projection permet de projeter l'ensemble de l'espace environnant. Cependant cette représentation est moins homogène puisqu'il existe des discontinuités au niveau des arêtes du cube.

#### II.B.4. Bilan

Le dispositif d'acquisition est donc ce qui détermine comment les données vont être enregistrées et par conséquent de quelle manière celles-ci vont rendre de la réalité. Ainsi que l'on utilise un appareil photographique ou un laser scanner, le rendu des images obtenues ne sera pas le même. Afin de pouvoir réinterpréter les informations contenues dans les images, il faut donc pouvoir connaître le système de projection utilisé.

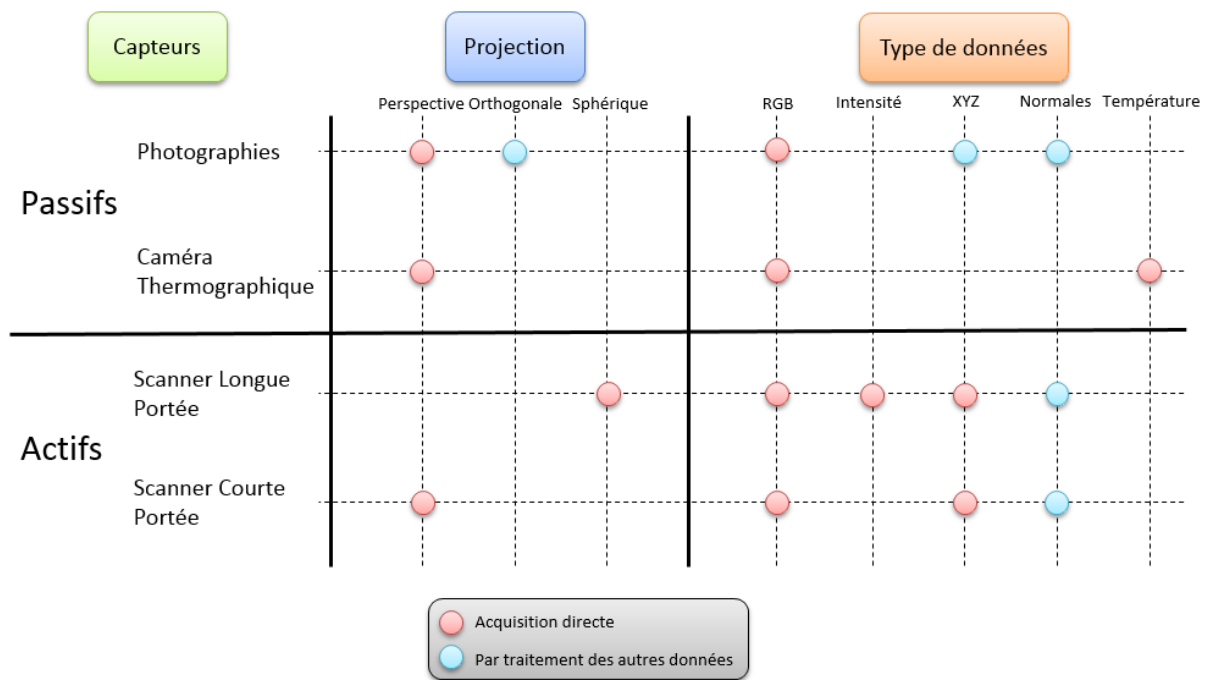


Figure 21 : Répartition des systèmes d'acquisition et des projections par rapport au type de données enregistrées et pouvant être générées

La Figure 21 résume la relation qu'il y a entre l'acquisition des données et la manière dont les données sont enregistrées en structure en fonction d'un type de projection. On constate donc que les experts se retrouvent avec de nombreuses données obtenues avec les différentes méthodes d'acquisition et issues de plusieurs types de projection.

#### II.C. Extraction d'informations

Tous les différents supports d'analyse, quel que soit leur type de projection, contiennent un certain nombre d'informations utiles aux experts. Afin d'extraire ces informations, il est nécessaire d'analyser l'image. Cette analyse peut être simplement réalisée par une personne qui va décrire le contenu de l'image en fonction de sa propre expérience.

Cependant, avec le développement de l'informatique, l'analyse d'image a pu connaître de nombreux développements notamment grâce aux avancées faites dans le domaine du traitement du signal.

### II.C.1. Informations colorimétriques

La couleur est la perception visuelle de la lumière du visible [Vandenbroucke N., 2000] [Kowalski P., 1990]. Quel que soit sa représentation, elle est là pour introduire une information que ce soit, en art pour évoquer une sensation ou rendre la réalité, ou en chimie pour définir la composition d'un pigment. Cependant il est aussi possible de rendre en couleur visible les informations véhiculées par la lumière en dehors du domaine visible.

La colorimétrie est une science relativement récente. C'est une technique qui se révèle très utile dans le domaine de la conservation et de la restauration [Boust C. et al, 2007]. Dans le cadre d'études pour la conservation, la couleur va permettre donner des informations sur l'état du bâtiment ou de l'objet. Une couleur montre un type de composant.

Un changement de couleur ou une absence de couleur permet d'attester d'une modification de l'objet par un certain nombre de mécanismes (par exemple des dégradations). Une première étude seulement sur la couleur peut donc permettre de révéler des évolutions (voulues ou non) de l'objet ou simplement des informations de composition.

Pour les images numériques, la codification des couleurs passe par une représentation sous forme d'un tableau de pixel. Chaque cellule du tableau contient une valeur rendant la sémantique dépendant du type de signal codé.

### II.C.2. Extractions de descripteurs images

Afin de retrouver les éléments composants une image, il est nécessaire d'extraire des caractéristiques de cette image qui seront par la suite analysées par le processus d'annotations automatiques. Les caractéristiques d'une image peuvent soit être associée à la globalité de l'image (caractéristiques globales), soit associée à des parties de l'image (caractéristiques locales). Un important inconvénient des caractéristiques globales est que les descripteurs extraits ne peuvent pas distinguer l'objet et l'arrière-plan de l'image. A l'inverse, les caractéristiques locales représentent des parties de l'image et leur implémentation requiert la détection de multiples points caractéristiques sur l'image d'entrée. Ainsi les caractéristiques locales sont plus efficaces avec les phénomènes d'occlusions.

#### II.C.2.a. Caractéristiques globales

Les caractéristiques globales d'une image sont liées à des représentations par histogramme ou par l'utilisation de moments d'image.

##### II.C.2.a.i. Histogrammes

Les histogrammes sont des représentations d'un décompte sur les données organisé par groupes. Ces histogrammes peuvent être utilisés pour décrire l'image en comptant la fréquence de chaque valeur d'intensité sur les pixels de l'image. Ainsi sur une image noir et blanc, l'histogramme construit les groupes sur l'intensité des valeurs (de 0 à 255) et compte le nombre de pixels pour chaque groupe (Figure 22). La comparaison de deux images passe alors par l'utilisation de fonctions d'analyse de distances. Les histogrammes peuvent donc révéler des informations quantitatives sur l'intensité des pixels mais pas sur le contenu de l'objet représenté dans l'image.

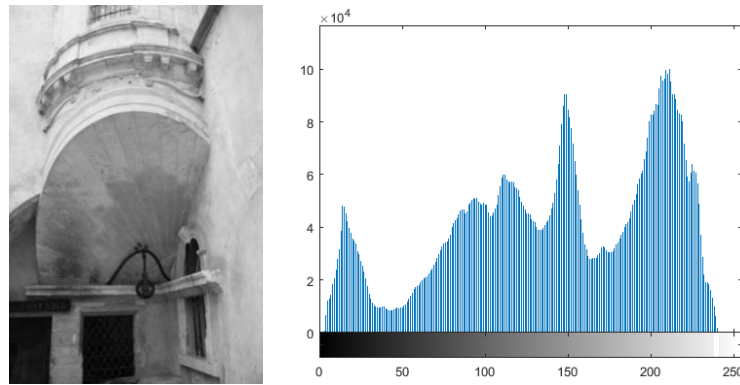


Figure 22 : Histogramme d'une image en niveaux de gris

#### II.C.2.a.ii. Moments d'images

Les moments d'image sont des quantités scalaires utilisées pour définir une fonction et faire ressortir ses caractéristiques importantes. Ces moments peuvent être utilisés sur des propriétés simples de l'image telles que l'aire ou l'orientation.

En 1962, [Hu M.K., 1962] fut le premier à introduire l'utilisation de moments géométriques pour la caractérisation images 2D en définissant sept moments invariants en translation, en rotation et en taille. Depuis, de nombreux types de moments invariants ont été proposés ([Teague M.R., 1980] [Li Y., 1992] [Wallin A. et al, 1995] [Flusser J., 2000] [Flusser J., 2002]). Des invariants photométriques ont aussi été introduit par [Van Gool L. et al, 1996] pour la description de données multi-spectrales.

#### II.C.2.b. Caractéristiques locales

L'extraction de caractéristiques locales a connu des évolutions importantes depuis plusieurs années. De nombreuses méthodes d'extraction de ces caractéristiques existent.

##### II.C.2.b.i. Détection de points d'intérêt

Les descripteurs locaux sont souvent accompagnés d'une détection de points d'intérêt sur lesquels les algorithmes d'extraction vont se baser pour définir des régions à analyser. L'utilisation de ces points d'intérêt permet d'accélérer les temps de calculs afin d'éviter d'avoir à appliquer le processus d'extraction sur tous les pixels de l'image. Cependant de nombreux descripteurs utilisent des grilles ou un ensemble de pixels au lieu d'utiliser les points d'intérêt.

D'après [Haralick R.M. et al, 1991], les points d'intérêt doivent être sélectionnés selon leur invariance, leur robustesse, leur rareté et leur interprétabilité. Plus spécifiquement les points d'intérêt doivent être uniques et distinct de l'arrière plan. De plus, ils doivent être invariants aux distorsions géométriques et radiométriques et robustes face au bruit. [Schmid C. et al, 2000] ont publié une évaluation des détecteurs les plus utilisés par rapport à ces critères.

[Moravec H.P., 1977] fut l'un des premiers à introduire un détecteur de points d'intérêt. Le principe est de calculer la somme des carrés des différences d'intensité sur des pixels adjacents par groupe carré de quatre à huit pixels. La différence des intensités est calculée par rapport à certaines directions et une carte est construite en utilisant la valeur de variation minimale pour chaque pixel. Le problème de ce procédé est qu'il n'est pas invariant en

rotation et est très sensible au bruit. De plus, les points d'intérêt ont tendance à être détectés le long des arêtes.

L'opérateur de [Förstner W. et al, 1987] [Förstner W., 1994] détecte les points d'intérêt en calculant la matrice de covariance des pixels. Chaque matrice décrit la distribution des gradients autour des pixels. Les points d'intérêt sont choisis parmi ceux ayant une grande différence de gradient.

[Harris C. et al, 1988] ont proposé une méthode similaire à celle de [Moravec H.P., 1977]. Cependant la recherche se fait selon toutes les directions. Le bruit est réduit à l'aide d'un lissage gaussien des images afin d'éviter les faux positifs. Des optimisations de ce système ont été proposées par [Shi J. et al, 1994] et [Mikolajczyk K. et al, 2004].

L'opérateur laplacien est un algorithme très utilisé dans le traitement des images. Le LoG (*Laplacian of Gaussian*) est une méthode qui applique l'opérateur laplacien sur une région traitée par un filtre gaussien. Le DoG (Difference of Gaussian) est une approximation du LoG qui utilise la différence entre deux images filtrées à différents niveaux par le filtre gaussien pour détecter les extrema locaux. Le DoG est utilisé notamment dans le processus SIFT (voir II.C.2.b.ii. Descripteurs locaux) défini par [Lowe D.G., 1999].

L'algorithme SUSAN, introduit par [Smith S.M. et al, 1997], détecte les caractéristiques en appliquant à l'image un masque circulaire dont le centre est appelé nucleus. Les zones dans le masque qui ont la même ou une luminosité similaire au nucleus sont appelés USAN (*Univalve Segment Assimilating Nucleus*). Cet algorithme permet de détecter les coins et les arêtes et ne requiert pas de réduction de bruit.

FAST (*Features from Accelerated Segment Test*) ([Rosten E. et al, 2006]) est une modification de SUSAN. Cet algorithme est un des plus efficaces en termes d'efficacité et de répétabilité. Ses performances sont basées sur AST (*Accelerated Segment Test*). Il a été amélioré par [Mair E. et al, 2010] en utilisant le détecteur AGAST.

Enfin, TILDE (*Temporally Invariant Learned Detector*) est un nouveau détecteur de points d'intérêt introduit par [Verdie Y. et al, 2014]. Cet algorithme permet notamment de résoudre les problèmes liés aux changements entre les images à cause de conditions d'éclairage ou météorologiques.

#### II.C.2.b.ii. Descripteurs locaux

En 1999, [Lowe D.G., 1999] a introduit, dans la communauté de la vision par ordinateur, l'algorithme SIFT (*Scale Invariant Feature Transform*) qui a pour but de détecter, de décrire et d'extraire des caractéristiques qui seront utilisées pour la comparaison d'images et la reconnaissance d'objets. SIFT utilise des caractéristiques invariantes par rapport à la taille de l'image, mais aussi en rotation et en translation. Ces caractéristiques sont aussi partiellement invariantes par rapport aux changements de lumières, les projections affines et le bruit des images. Chaque caractéristique est décrite par un vecteur. L'algorithme passe notamment par la construction d'une pyramide de DoG pour extraire les points d'intérêt et le calcul de l'orientation des gradients (Figure 23) pour la définition de descripteurs de point d'intérêt.

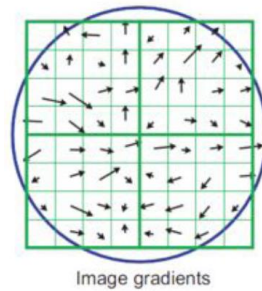


Figure 23 : Gradients d'images

De nombreuses extensions de l'algorithme ont été proposées. [Ke Y. Et al, 2004] ont proposé PCA-SIFT qui utilise des cartes de magnitude du gradient au lieu de l'orientation des gradients. SIFT a aussi été étendu pour une utilisation sur des images couleurs par de nombreux auteurs ([Bosch A. et al, 2006] [Van De Weijer J. et al, 2006] [Brown M. et al, 2011]). Une évaluation des descripteurs basés sur les couleurs a été réalisée par [Burghouts G.J., et al, 2009].

Une alternative à SIFT, Dense SIFT, n'utilise pas la détection de point d'intérêt mais utilise tous les pixels ou une grille de pixels dans l'image comme base pour la définition des caractéristiques. De la même manière, le descripteur HOG (*Histogram of Oriented Gradients*) ([Dalal N. et al, 2005]) utilise une grille spécifique de pixels pour construire un histogramme des directions de gradients. Le descripteur GLOH (*Gradient Location and Orientation Histogram*) ([Mikolajczyk K. et al, 2005]) qui s'inspire du descripteur HOG mais utilise une grille spécifique non rectangulaire (Figure 24).

[Bay H. et al, 2006] ont introduit l'algorithme SURF (*Speeded Up Robust Features*) qui offre des résultats intéressants en termes de répétabilité, de distinction, de robustesse et de rapidité par rapport aux algorithmes jusqu'alors proposés. La détection utilise les images intégrales, définies par [Viola P. et al, 2001], à travers une approximation de matrices hessiennes.

BRIEF (*Binary Robust Independent Elementary Features*) a été proposé par [Calonder M. et al, 2010]. Cet algorithme est le premier à proposer l'utilisation de valeurs binaires mesurées directement sur des portions des images. Les valeurs binaires sont calculées par comparaison de paires d'intensités de pixels le long de lignes. Les descripteurs binaires offrent une efficacité importante en termes de stockage et de coût de calculs. Plusieurs évolutions de BRIEF ont été proposées (D-BRIEF [Trzcinski T. et al, 2012], BinBoost [Trzcinski T. et al, 2013]).

DAISY est un autre descripteur local introduit par [Tola E. et al, 2010] qui a beaucoup de similarité avec SIFT et GLOH. Les descripteurs sont basés sur des cartes d'orientation convoluée de l'image originale (Figure 24).

BRISK (*Binary Robust Invariant Keypoints*) est une méthode introduite par [Leutenegger S. et al, 2011] qui combine détection de points d'intérêt, description et correspondance. La partie détection est inspirée par les détecteurs FAST et AGAST. Le modèle d'échantillonnage (Figure 24) ressemble fortement à celui utilisé dans DAISY mais est plus efficace en termes de stockage et de temps de calculs. Enfin, la comparaison passe par le calcul de la distance de Hamming, de la même façon que BRIEF.



ORB (*Oriented Fast and Rotated BRIEF*) ([Rubblee E. et al, 2011]) est une alternative pour SIFT et SURF. L'algorithme est une combinaison de FAST pour la détection de points d'intérêt modifié pour donner les informations d'orientation (oFAST) et des descripteurs BRIEF modifiés pour diminuer les faibles performances de ce descripteur au niveau de l'invariance en rotation (rBRIEF).

FREAK (*Fast RetinA Keypoints*) est un descripteur inspiré par le système de vision humain. Introduit en 2012 par [Alahi A. et al, 2012], c'est un descripteur basé sur l'utilisation de valeurs binaires qui sont traitées pour comparer les intensités d'images à travers un motif rétinien. En comparaison des autres descripteurs qui utilisent des motifs circulaires, le motif utilisé a une densité de points plus importante près du centre (Figure 24). D'autre part, la taille des noyaux évolue de manière exponentielle et certaines zones se superposent.

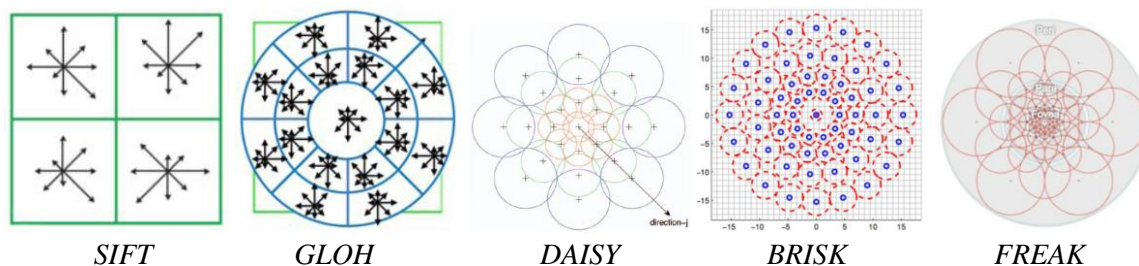


Figure 24 : Exemples de descripteurs locaux

### II.C.3. Informations géométriques

Les informations géométriques sont des ensembles de données repérées dans l'espace. Ces informations sont des mesures de l'objet et peuvent être obtenue de différentes manières.

#### II.C.3.a. Mesures métriques

Lors de l'étude d'objets patrimoniaux, la connaissance de mesures est une des premières questions qui apparaît. Tous les procédés du relevé jusqu'à la méthodologie de l'étude vont généralement être conditionnés par ces mesures. A partir d'un relevé succinct des mesures d'un objet, il est possible de directement commencer à tracer un aperçu des proportions d'un objet à l'aide de la position de points caractéristiques. Mais ces mesures peuvent par la suite être exploitées pour engager la modélisation 3D de l'objet.

Parmi les sources iconographiques disponibles pour les objets patrimoniaux, les documents métriques fournissent les mesures et les proportions de différentes parties. Ce sont des représentations (généralement orthogonales) de l'espace, exécutée à la main ou à l'aide d'outil technique à différentes échelles. Dans ces sources, on retrouve les plans, les cadastres (Figure 25), les coupes, les élévations, etc.



Figure 25 : Exemple de cadastre

Les mesures sont donc une étape nécessaire à la restitution d'un objet patrimonial et peuvent être de plusieurs sortes : prises de coordonnées, mesure d'angles, mesure de distances, mesure de dénivellements, axes de cotes [De Luca L., 2009].

La collecte de coordonnées passe par le relevé d'un ensemble de points caractéristiques directement sur site. L'intégration de ces données relevées par des méthodes traditionnelles à des photographies permet de mettre en lien, sous un aspect métrique, différentes vues d'un objet patrimonial.

Une distance se définit comme la mesure entre deux points dans l'espace. La mesure d'une distance peut se faire de manière directe en connaissant les coordonnées des deux points aux extrémités ou de manière déduite en se servant de caractéristiques géométrique de l'objet et de ses parties.

Un angle est une portion de plan délimité par deux droites sécantes exprimé généralement en degré ou en radians. La mesure d'un angle nécessite soit trois points, soit deux points et un axe, soit un point et deux axes.

La mesure des dénivellements permet de mesurer des différences de niveau. Elle nécessite la définition d'un plan de référence à partir duquel les distances des éléments seront définies. Cette mesure permet généralement de définir l'altitude de points caractéristiques par rapport au sol.

La mesure de cote permet de documenter l'ensemble des relations métriques entre les éléments qui composent un objet. Ces mesures peuvent se classer selon trois catégories : totales (liée à l'ensemble du relevé), totales relatives (liées aux éléments qui définissent la composition de l'objet) ou partielles (liées à des parties significatives de l'objet).

### *II.C.3.b. Génération de coordonnées 3D à partir de relevé photographique*

Aujourd'hui il existe des méthodes pour générer automatiquement ces mesures. Les relevés par laser scanner fournissent notamment directement des coordonnées 3D (voir II.A.2. *Scanners*). Cependant il existe aussi des méthodes indirectes qui se basent sur l'exploitation



du support photographique pour extraire les coordonnées spatiales à l'aide de différentes techniques de restitution ([Remondino F. et al, 2006] [Sturm P. et al, 2011]).

La photogrammétrie se définit comme « *la science ou la technique permettant d'obtenir des informations fiables sur l'espace naturel ou sur des objets physiques par l'enregistrement, la mesure et l'interprétation d'images photographiques ....* » [American Society of Photogrammetry, 1980].

Pendant longtemps, des chambres métriques et semi-métriques ont été utilisées pour l'acquisition photogrammétrique [Carbonnel M., 1968]. Ainsi selon [Grussenmeyer P. et al., 2001], les chambres métriques appartiennent à trois catégories : les chambres métriques, les chambres semi-métriques et les chambres à DTC (CCD). Avec l'évolution des appareils numériques et des capteurs CCD, des images numériques peuvent être aujourd'hui directement acquises pour servir de bases à l'acquisition photogrammétrique et aux traitements sous-jacents.

La mesure se base sur le calcul des paramètres d'orientation internes et des paramètres d'orientation externes de l'appareil photographique [Kraus K. et al, 1997] (Figure 26). L'orientation interne consiste à calibrer l'appareil c'est-à-dire à fixer la position de la focale et à estimer la distorsion de l'image. L'orientation externe détermine la position et l'orientation de l'appareil dans l'espace-objet. Cette orientation est donc composé de six paramètres : trois pour la position (coordonnées suivant les trois axes: X, Y, Z) et trois pour l'orientation (angles autour des trois axes :  $\Theta$ ,  $\psi$ ,  $\omega$ ).

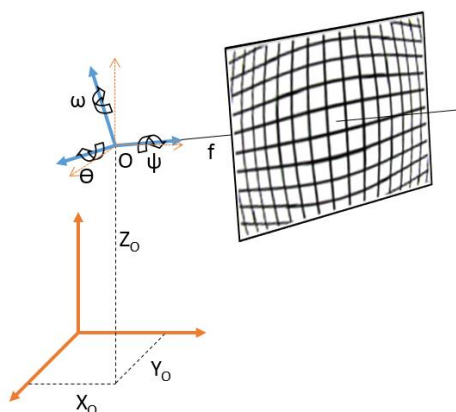


Figure 26 : Paramètres de calibration interne et externe

A partir de ces paramètres, la restitution photogrammétrique va extraire un ensemble de coordonnées dans l'espace. Le nombre d'images utilisées va déterminer la méthode : mono, stéréo ou multi-images (Figure 27) [Grussenmeyer P. et al.,2001].

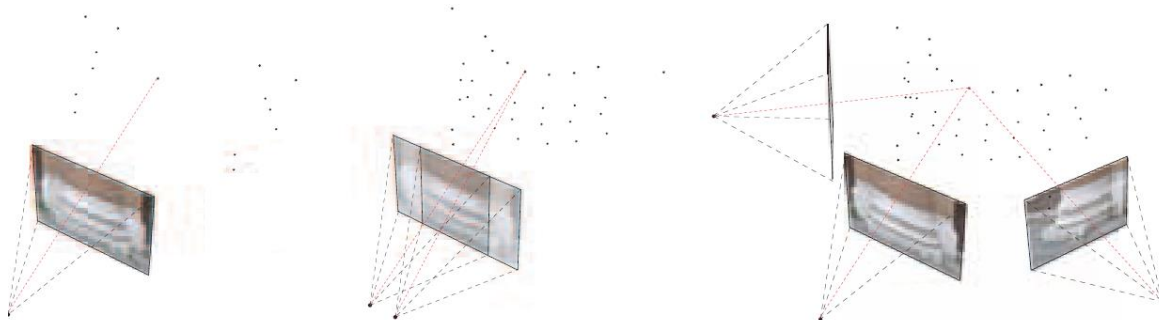


Figure 27 : Les différentes techniques de restitution photogrammétrique. De gauche à droite : mono, stéréo, multi-images

La restitution à partir d'une seule image se base sur la mesure de plusieurs points puis le redressement est estimé par approximation en considérant que les surfaces sont identifiées par un minimum de quatre points d'appui. Cependant l'extraction de mesures est contrainte à l'orientation du plan choisi pour la rectification.

La restitution à partir d'un couple d'images utilise une configuration stéréo. Les coordonnées 3D peuvent être retrouvées en cherchant l'intersection dans l'espace de deux points homologues sélectionnés entre les images.

Pour la restitution à partir de plusieurs images, la méthode se base sur l'ajustement des faisceaux à partir d'un ensemble de correspondances sélectionnées sur les photographies. Sur cette méthode, le calcul des paramètres interne et externe peut être affiné par minimisation d'erreur.

Ces trois méthodes ont été, à leur début, relativement manuelles, notamment sur la sélection des points homologues entre les images sur les deux dernières techniques. Aujourd'hui des solutions automatiques se développent pour les détectations des points de correspondances et d'extraction de caractéristiques (voir [II.C.2. Extractions de descripteurs images](#)) permettant de développer des chaînes de reconstruction automatique à partir d'images non-orientées afin de générer des nuages de points à différents degrés de densité. Parmi ces développements, on peut noter les travaux de l'IGN (Institut Géographique National) [[Pierrot-Desseiligny M. et al, 2011a](#)] [[Pierrot-Desseiligny M. et al, 2011b](#)] pour la suite de logiciels open-source MicMAC, l'application [[123D Catch](#)] développé par Autodesk, PhotoScan [[PhotoScan](#)] de Agisoft, ou encore Bundler [[Bundler](#)].

#### II.C.4. Extractions de descripteurs 3D

Les modèles 3D comportent de nombreuses informations géométriques qu'il est nécessaire d'étudier par l'extraction de caractéristiques. De même que pour les images, les modèles 3D peuvent être décrits par un ensemble de descripteurs liés à la géométrie de l'objet. Ces descripteurs doivent par conséquent être aussi robustes à certaines transformations telles que la rotation, le bruit sur les surfaces, la taille, ... et peuvent être associés de manière globale ou locale au modèle 3D.

##### *II.C.4.a. Caractéristiques globales*

Les descripteurs globaux ont pendant longtemps dominé parmi les recherches effectuées dans ce domaine. De nombreux exemples existent dont les plus connus sont ceux de Fourier ou harmoniques sphériques ([[Saupe D. et al, 2001](#)] [[Kazhdan M. et al, 2003](#)]), les moments de forme ([[Saupe D. et al, 2001](#)]) et les distributions de forme ([[Osada R. et al, 2002](#)]).

Ce dernier propose un descripteur basé sur une distribution des formes de l'objet. Cette distribution est extraite à partir d'aspects simples de géométrie de l'objet. Plus précisément, des modèles 3D arbitraires sont transformés en fonctions paramétriques et une probabilité de distribution est générée pour chacune des fonctions de forme. Cette distribution sert par la suite à la recherche de similarité (Figure 28).

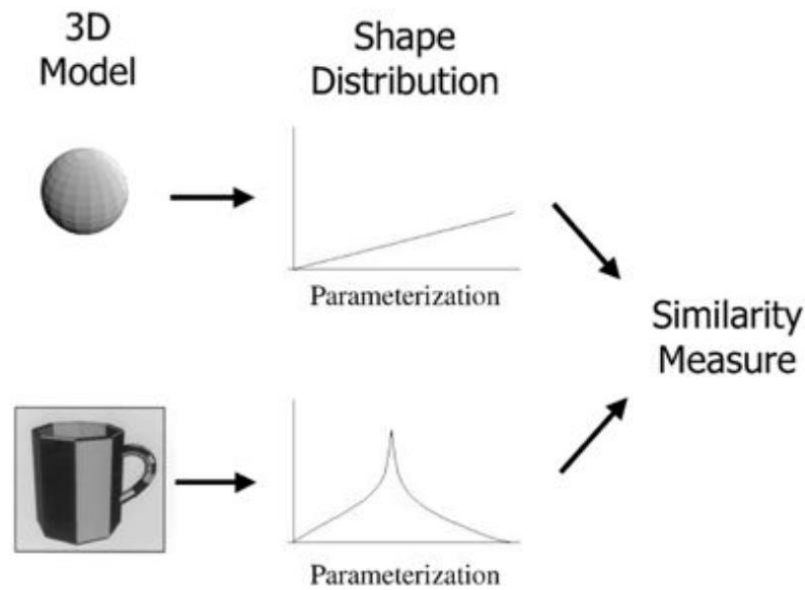


Figure 28 : Distribution des formes et comparaison [Osada R. et al, 2002]

L'aspect important de cette méthode est cette fonction de formes dont la distribution peut fournir une signature descriptive d'un modèle 3D. L'avantage de cette méthode de comparaison est qu'elle ne nécessite pas d'enregistrement de vues, de correspondance de formes ou d'ajustement de modèle.

Ces différentes approches posent plusieurs problèmes. Le premier problème est que ces caractéristiques globales ne peuvent pas servir à comparer des formes partielles ou des modèles segmentés. Dans le domaine de la conservation du patrimoine, les objets sont souvent endommagés et seules des parties de l'objet original persistent. Il est donc difficile de trouver des similitudes avec des modèles de la base de données causant ainsi de fausses classifications. Le second problème est que cette méthode ne fonctionne pas dans le cas d'occlusions. Enfin, de fortes variations au sein d'une classe affectent très fortement la performance de ces descripteurs.

#### II.C.4.b. Caractéristiques locales

Les descripteurs locaux ont montré récemment qu'ils possédaient de nombreux avantages par rapport aux descripteurs globaux, étant donné qu'ils permettent la comparaison partielle des modèles et prennent mieux en compte les problèmes d'occlusions.

Un des plus anciens descripteur local, introduit par [Johnson A.E. et al, 1999], est les images spin SI (*Spin Images*). Ces descripteurs enregistrent un histogramme représentant l'occupation spatiale du modèle 3D autour d'un point 3D. Ainsi une surface est décrite par trois composants : un nuage de points dense, les normales à la surface et une image liée à chaque point qui encode les propriétés globales de la surface dans un système d'orientation lié à l'objet plutôt qu'à la vue (Figure 29).

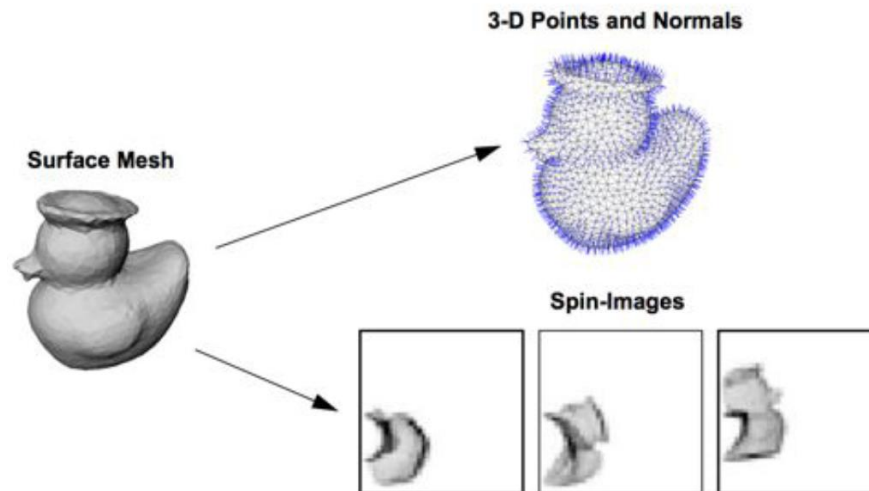


Figure 29 : Composant d'une représentation par SI

Comme ces images représentent une description locale de la forme globale, le descripteur est invariant sur les transformations rigides. La comparaison est ensuite effectuée en comparant ces images et ainsi chercher des correspondances entre les points.

Dans la littérature récente, de nombreuses méthodes d'extraction de caractéristiques locales 3D ont été proposées ([Mian A.S. et al, 2006] [Sun J. et al, 2009] [Gelfand N. et al, 2005] [Pottmann H. et al, 2009] [Novatnack J. et al, 2008]).

[Knopp J. et al, 2010] ont proposé une extension du descripteur 2D SURF pour les formes 3D. La méthode combine la détection de points d'intérêt avec une extraction de descripteurs basée sur la géométrie des voisins proches des points d'intérêt. 3D SURF est une adaptation du modèle proposé par [Willems G. et al, 2008]. La différence entre les deux méthodes est que 3D SURF applique la même échelle spatiale sur les trois dimensions. Pour la détection des points d'intérêt 3D, le modèle est voxellisé en cubes 3D. A partir de ces points d'intérêt, le descripteur 3D SURF est calculé en estimant l'orientation des points d'intérêt, puis en utilisant cette orientation pour construire un système de coordonnées locales pour chaque point.

L'utilisation de descripteurs 3D locaux, tout comme les descripteurs 2D locaux, génère un nombre arbitraire de descripteurs. Afin de résoudre ce problème, une méthode appelée « sac de mots » BoW (*Bag of Words*) permet de décrire par un histogramme la fréquence à laquelle chaque élément de vocabulaire apparaît dans le modèle.

### II.C.5. Informations morphologiques

La morphologie est l'étude des formes. Ces caractères sont établis à partir d'un ensemble de descripteurs. A partir de ses informations, il est possible de découper la morphologie d'un objet. Les descripteurs étant choisis selon l'objectif d'analyse, le découpage de la morphologie est par conséquent aussi relatif à cet objectif. Par exemple, dans le vocabulaire de l'architecte [Pérouse de Montclos J.M., 2004], le bâti est organisé selon une structure signifiante qui n'est pas seulement caractérisée par une apparence visuelle ou une nature géométrique. Plusieurs méthodes existent pour pouvoir décrire et classer ces formes. Selon [Blaise J.Y., 2003], dans le cas de l'analyse architecturale, les approches par intension (association d'un modèle

théorique à un élément réel) sont incompatibles avec les exigences de l'étude du patrimoine construit, alors que des approches par extension (extraction et groupage de descripteurs communs à plusieurs éléments) semblent plus pertinentes.

En géométrie, les descripteurs des formes peuvent être de plusieurs types. Les plus simples à mettre en œuvre sont les normales et la courbure d'une surface. Les normales sont des propriétés importantes des surfaces et sont utilisées dans de nombreux domaines comme l'infographie afin d'appliquer des sources de lumières correctes pour le rendu d'un modèle 3D. De nombreuses méthodes d'estimation des normales existent. [Klasing K. et al, 2009] proposent une comparaison de différentes méthodes d'estimation de normales. Une des méthodes les plus simples d'estimation des normales est celle proposée par [Berkmann J. et al, 1994]. Le problème de déterminer la normale d'un point d'une surface est approximé par le problème d'estimation d'une normale à un plan tangent à la surface, ce qui se rapporte à une estimation par les moindres carrés d'une surface [Shakarji C., 1998]. Cela nécessite donc d'avoir pu d'abord détecter les points voisins au point considéré pour estimer cette surface [Rusu R.B., 2009]. Avec l'estimation des normales, il est possible par la suite d'avoir une estimation de la courbure.

L'analyse des normales et des courbures d'un objet peuvent permettre de faire ressortir des informations importantes sur des objets. L'analyse des normales peut facilement permettre de détecter des surfaces planes ou parallèles puisque de telles surfaces ont toutes la même direction de normales. [Adrian J. et al, 2014] et [Pamart A., et al, 2015] ont, par exemple, conduit la comparaison d'un ensemble de colonnes extraites du cloître de l'abbaye de St Michel de Cuxa (à Codalet, situé au sud de la France) à partir de l'analyse de la courbure des colonnes.

## II.C.6. Métadonnées

Le contenu d'un fichier ne permet pas de décrire toutes les conditions de signification, de contexte ou d'organisation du fichier ou de son contenu. La nécessité de décrire des données, quel que soient leur support (papier ou électronique) est depuis longtemps établie. Tous les établissements devant gérer de l'information ont déjà une longue pratique de la codification des contenus des documents, par exemple dans la pratique de gestion des fiches et des notices documentaires de bibliothèques. Avant l'arrivée de l'informatique, ces données étaient enregistrées et codifiées sur des supports physiques distincts de la donnée qu'ils décrivaient. L'arrivée de l'informatique a permis de numériser ces données afin de faciliter la gestion et de les intégrer directement dans les documents eux-mêmes, permettant d'éviter le risque de perte.

Le terme de *métadonnée* est apparu dans les années 1990 dans le cadre de la description de ressources sur Internet. Une métadonnée est une donnée qui définit ou décrit une autre donnée (*méta* : préfixe grec indiquant l'autoréférence) [Sicilia M.A., 2006]. Elles désignent ainsi un ensemble d'informations standardisées relatives à un fichier.

Les types et les fonctions des métadonnées peuvent être classés en trois catégories générales [Niso, 2004]:

- *Métadonnées descriptives* : décrivent et identifient les ressources d'informations
- *Métadonnées structurelles* : facilitent la navigation et la présentation des ressources

- *Métadonnées administratives* : facilitent la gestion et le traitement à court et long termes des collections numériques

Ces catégories n'ont pas de frontières bien définies. Ainsi un attribut peut être relatif à plusieurs catégories.

Les métadonnées correspondent ainsi à des marqueurs introduits dans les fichiers. Cette structuration de l'information d'une ressource permet d'améliorer les recherches d'informations et s'applique à tous les types de ressources numériques. Les métadonnées ont donc un fort potentiel pour l'interopérabilité de ressources informatiques.

En fonction du support ou de l'application, des nombreux standards de description de métadonnées existent. Les métadonnées sont notamment au cœur de l'architecture du Web sémantique [Berners-Lee T. et al, 2001]. Développé par le W3C, RDF (Resource Description Framework) [Klyne G. et al, 2004], est le langage de base du Web sémantique. Dans le cadre d'images numériques, chaque format implémente une façon spécifique de stocker les métadonnées. Cependant il existe certaines normes communes comme le standard EXIF (*EXchangeable Image File*) [EXIF, 2010] ou le standard IPTC (*International Press and Telecommunication Council*) [IPTC, 2014]. Ainsi pour une image acquise à l'aide d'un appareil numérique, les métadonnées peuvent contenir des informations sur la date et l'heure de la prise de vue, les dimensions de l'image, les données GPS, les paramètres de prises de vue, la marque de l'appareil, le modèle de l'appareil, ... (Figure 30).



Figure 30 : Exemple de métadonnées associées à une même image numérique



La pratique des métadonnées est aujourd'hui sujette à des nombreuses réserves car la transmission d'un document entraîne la transmission de ces métadonnées contenant généralement des informations assez détaillées voire parfois personnelles que l'utilisateur ne souhaite pas toujours transmettre. Cependant la quantité d'informations complémentaires qu'elles contiennent est généralement utile pour l'exploitation des documents associés et pour la mise en place de systèmes de recherche documentaire efficace. A ce jour, l'utilisation des métadonnées reste encore assez peu exploitée dans de nombreux domaines. Que la transmission des informations contenues dans les métadonnées soit voulue ou non, le plus grand risque est d'ignorer l'importance de ces métadonnées.

## II.D. Conclusion sur les méthodes d'acquisition

Les principales techniques utilisées pour le relevé permettent aux experts d'obtenir les informations nécessaires pour entreprendre des analyses d'objets patrimoniaux. Elles sont aujourd'hui généralement basées sur les techniques de balayage laser ou sur les techniques de photogrammétrie. Plusieurs études ont déjà cherché à comparer les deux techniques ([Guarnieri A. et al, 2004] [Fuchs A. et al, 2004]). Ces auteurs ont évalué ces méthodes par rapport au problème de la restitution 3D complète. Dans l'approche proposée, l'évaluation de ces méthodes nécessite seulement d'évaluer la qualité du nuage de points produit à la suite de l'acquisition. Afin d'évaluer ces méthodes, plusieurs critères sont à envisager.

**Facilité de mise en œuvre :** La mise en place d'une acquisition par photogrammétrie est plus simple et plus rapide que l'acquisition utilisant un laser scanner. Dans les deux cas, il est nécessaire de pouvoir se déplacer autour de l'objet à numériser afin d'acquérir plusieurs points de vue. Au vu de la légèreté du dispositif photographique, celui-ci permet donc d'acquérir de façon beaucoup plus rapide plusieurs points de vue autour de l'objet que par acquisition lasergrammétrique. De plus, ce faible encombrement de l'appareil photographique permet de pouvoir être embarqué à bord de systèmes mobiles (par exemple sur des drones ou des bateaux) ce qui lui permet, contrairement au relevé laser, de pouvoir acquérir des zones difficiles d'accès. Cependant, l'acquisition lasergrammétrique donne directement des informations géométriques là où la photogrammétrie nécessite d'extraire ces informations par des algorithmes souvent lourds en calculs.

**Qualité du nuage de points :** Comme montré par [El-Hakim S. et al. 2003], il y a aujourd'hui une réelle nécessité à comparer les méthodes de reconstruction basée sur l'image. Plusieurs études ont cherché à pouvoir évaluer la précision des nuages de points obtenus à partir de techniques utilisant les images ([Ahmadabadian A.H. et al. 2013] [Koutsoudis A. et al. 2013] [Remondino F. et al, 2012]). Beaucoup de ses études cherchent à comparer le résultat obtenu à partir des images à un relevé par laser. [Toschi I. et al. 2014] ont notamment cherché à aborder le problème de l'évaluation d'un point de vue métrologique montrant que le protocole d'acquisitions d'images jouait un rôle important dans l'obtention de résultats. Cependant, ils ont aussi montré que les résultats du nuage de points obtenus avec le logiciel développé par l'IGN, avec le bon protocole d'acquisition, pouvaient avoir des résultats, en termes de précision, proches de ceux obtenus par lasergrammétrique.

**Quantité d'informations :** La quantité d'informations enregistrées par les deux dispositifs est énorme. Les lasers peuvent acquérir un grand nombre d'informations très rapidement (jusqu'à 976 000 points par secondes pour le FARO Focus 3D). Mais d'autre part, la quantité d'informations contenues dans une photographie (de 24 millions à 50 millions de pixels par image sur des appareils photographiques numériques reflex) peut être encore plus importante. Cependant, même si la photogrammétrie nécessite de lourds calculs pour obtenir des informations géométriques, là où le laser donne directement ces informations qui sont contraintes à la phase d'acquisition sur le terrain, elle peut extraire ces informations à différents niveaux d'approximation en fonction des besoins de restitutions à posteriori de l'acquisition.

**Variété d'informations :** Si les scanners permettent d'acquérir par leur automatisation des représentations exhaustives de la dimension géométrique des objets, leur équipement, en termes d'acquisitions photométriques, reste encore imparfait. Or dans le cas de la documentation de l'état de conservation d'objets patrimoniaux, ces informations sont très importantes et la photographie procure la véracité nécessaire à l'étude de ces objets.

Au-delà de l'acquisition de données tridimensionnelles, la documentation, l'étude, la conservation et la valorisation des objets patrimoniaux reposent sur l'acquisition d'une grande quantité de données hétérogènes. Ces données supposent donc l'intégration d'autres techniques d'acquisition et de traitement. En effet, à côté de la représentation tridimensionnelle, l'acquisition de données quantitatives telles que la thermographie infrarouge ou d'images UV est un domaine bien diffus au sein des méthodes d'analyse. Il apparaît donc essentiel d'étudier les conditions au travers desquelles les informations extraites de la numérisation 3D et des données hétérogènes d'analyse soient intégrées au sein d'un support d'observation et d'analyse.





### III. Méthodes d'annotations sémantiques

Dans le chapitre précédent, nous nous sommes intéressés aux outils et aux données qu'utilisent les experts pour documenter un objet patrimonial. Dans ce chapitre, nous nous intéresserons aux méthodes et aux solutions existantes pour l'annotation sémantique des différents supports.

Les experts disposent de plusieurs types de représentations qui peuvent être classées selon deux grandes catégories : les représentations 2D et les représentations 3D. Chacune de ses représentations peut être enrichie sémantiquement par un processus d'annotation. Les méthodes d'annotation s'adaptent donc en fonction de la représentation choisie.

Le principe de l'annotation repose sur la mise en relation entre un document et des informations complémentaires. L'annotation sémantique est le processus qui crée des étiquettes sémantiques visant à soutenir des recherches avancées. En outre, l'annotation est utilisée pour convertir des structures syntaxiques en structure de connaissances [Slimani T., 2013]. Les annotations ainsi définies permettent donc de créer une structuration complexe de l'information. Cette structuration a une influence importante sur les données pouvant être affichées par l'utilisateur et la manière dont elles sont affichées. On distingue ainsi quatre modèles [Andrews P. et al, 2012] (Figure 31) :

- **Tags** : Un tag est un mot-clé (un mot ou une séquence de caractères sans espaces) ou un terme assigné à une ressource qui décrit implicitement une propriété particulière de la ressource. L'ordinateur ne connaît pas, dans ce cas, la signification sous-entendue par le commentateur.
- **Attributs** : Un attribut est une paire d'éléments : le nom de l'attribut (définissant la propriété de la ressource) et sa valeur.
- **Relations** : Une relation est une paire de composants : le nom de la relation et la ressource associée. La ressource annotée est donc liée par rapport au nom de la relation. Ainsi les relations sont une extension des attributs pour les ressources.
- **Ontologies** : Une ontologie est un modèle qui consiste à aligner une ressource ou une partie de celle-ci avec une description de certaines de ses propriétés et caractéristiques selon un modèle conceptuel formel (ou ontologie). Comme défini par [Studer R. et al, 1998] une ontologie est donc « une spécification explicite d'une conceptualisation partagée ».

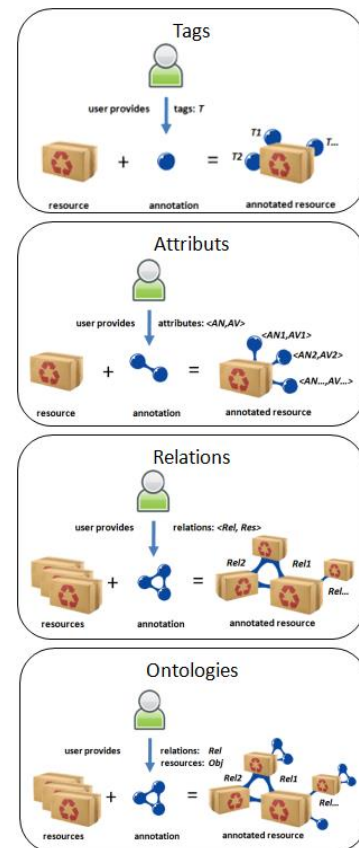


Figure 31 : Modèles d'annotations [Andrews P. et al, 2012]

Si les documents sont enrichis par des informations sémantiques, l'interopérabilité est assurée et permet par la suite de faire des recherches sémantiques plutôt que des recherches sur de simples chaînes de caractères [Uren V.S. et al, 2006] [Ma Y. et al, 2013].

### III.A. Annotations d'images

L'ajout d'un contenu sémantique à des images à travers l'utilisation de mots-clés ou d'ontologies s'intitule « annotation d'images » [Busayarat C., 2010]. Dans le domaine de la conservation du patrimoine, elle se révèle comme un moyen important pour documenter le contenu des images. Ce procédé d'annotation passe par la définition de zones d'intérêt sur l'image auxquelles sont alors associées à une description à l'aide de mots-clés ou d'ontologies. L'annotation est principalement utilisée pour la gestion de base de données d'images, notamment afin d'améliorer les possibilités de la recherche [Rodden K., 1999].

Les annotations peuvent être rattachées aux photos de deux façons :

- **Annotation globale** : l'annotation est associée à l'ensemble du document, elle décrit le contenu du document
- **Annotation locale** : l'annotation est associée à des zones du document (points, courbes, polygones, etc.), chaque zone est annotée séparément des autres

Les techniques d'annotations d'images sont classifiées selon plusieurs catégories [Kharkate S.K. et al, 2013] :

- Annotations manuelles
- Approches guidées par une ontologie
- Utilisation des informations textuelles de contexte
- Annotations automatiques
- Annotations semi-automatiques

Chacune de ces méthodes sont décrites dans les paragraphes suivants.

#### III.A.1. Annotations manuelles

L'annotation manuelle est la méthode d'annotation la plus répandue dans les systèmes de bases de données [Gong Y. et al, 1994]. Cette méthode permet aux utilisateurs d'associer des mots-clés aux images. Les mots-clés ne demandent pas de formalisation. N'importe quel mot-clé peut être associé par l'utilisateur, du moment qu'il le juge adéquat à la description du contenu de l'image.

Beaucoup de sites internet, proposant de créer des albums photos en ligne, permettent de faire ses propres annotations sur les photos. C'est le cas notamment du site Flickr [FLICKR] où les utilisateurs peuvent ajouter des « tags » (étiquettes) à leurs photos ou de Marqueeed [Marqueeed, 2014] qui vise un aspect collaboratif de la description d'images (Figure 32). Des zones (seulement rectangles pour Flickr) peuvent être tracées autour des objets sur les photos et une description ou des commentaires peuvent leur être associés. L'annotation se fait ici photo par photo, ce qui peut être très long à mettre en place.

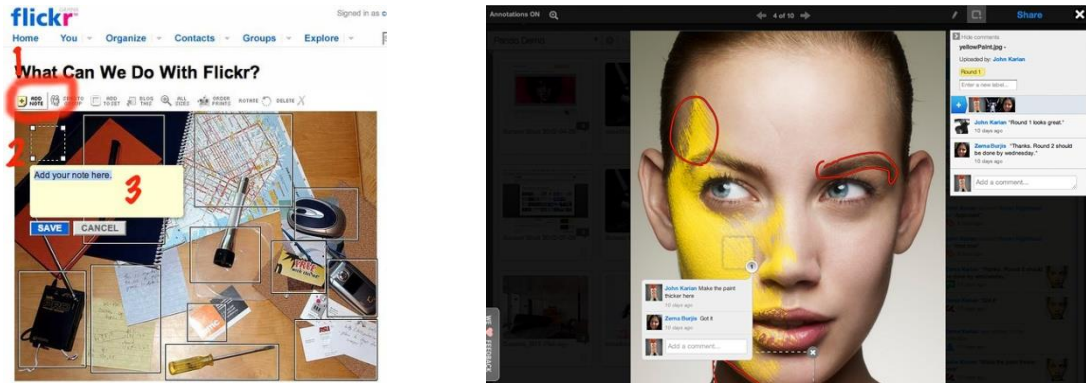


Figure 32 : Exemple d'annotation avec [FLICKR] à gauche et avec [Marqued, 2014] à droite

Afin d'améliorer le processus d'annotation manuelle, des interfaces ont été développées pour pouvoir associer les annotations à des blocs d'images. Parmi elles on peut citer *Photofinder* [Kang H. et al., 2000] qui a évolué pour devenir *Photomesa* [Bederson B.B., 2001] (Figure 33). Plusieurs photos sont sélectionnées et le même mot-clé peut être associé à chaque photographie en une seule fois. Mais cette technique ne peut être utilisée que sur des annotations globales (concernant toute l'image et non une zone). Bien qu'elle évite une grosse partie de travail d'annotation, cette méthode reste toujours manuelle et demande beaucoup de manipulations.

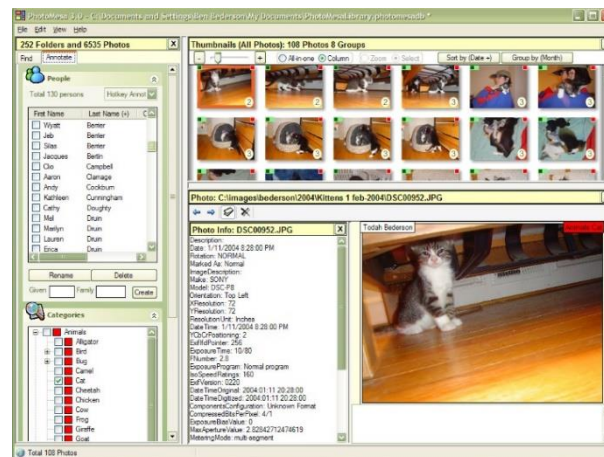


Figure 33 : Interface de Photomesa [Bederson B.B., 2001]

Les séquences vidéo étant des séquences d'images, certains travaux ont cherché à utiliser les mêmes principes d'annotations aux vidéos. Dans *VideoAnnEx* [Lin C-Y et al. 2003] (Figure 34), la séquence vidéo est divisée en différentes parties qui sont chacune annotée. Ici, les annotations sont globales sur chaque partie de la séquence.

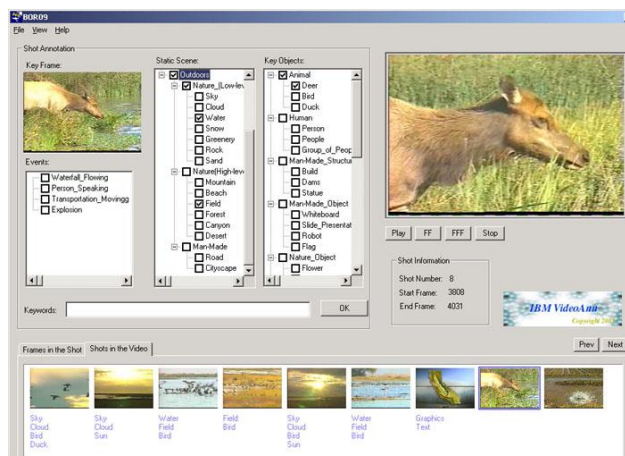


Figure 34 : Interface de VideoAnnEx [Lin C-Y et al. 2003]

L'annotation manuelle est celle qui reste la plus précise en termes d'exactitude de description du contenu des images. La définition des mots-clés à associer est basée sur le jugement humain. L'aspect d'analyse humaine est ce qui fait que les termes sont plus justes que pour d'autres méthodes mais c'est aussi ce qui peut valoir des oublis de la part de l'utilisateur. Le processus est de plus, par sa mise en place, très long.

L'utilisation de mots-clés reste d'autre part très subjective. Un concept est souvent décrit par plusieurs mots synonymes. L'utilisateur peut avoir utilisé l'un des mots sur une image et un autre sur une autre image. Cependant, le concept reste le même entre les deux images. Ainsi, si les mots associés aux images sont analysés, il n'y aura pas de correspondance entre les deux images car le lien entre les deux mots n'aura pas été établi informatiquement mais seulement conceptuellement. D'autre part, un utilisateur peut choisir de n'associer qu'un seul mot-clé par image ou régions d'images alors qu'un autre utilisateur en utilisera plusieurs. La subjectivité rentre en compte aussi dans le fait qu'un utilisateur ou une autre n'aura pas les mêmes objectifs de descriptions. De plus, l'utilisation de mots-clés ne permet pas de mise à jour facile. Le seul moyen de mettre à jour les mots-clés est de remplacer les termes image par image.

### III.A.2. Approches guidées par une ontologie

Les ontologies peuvent aussi servir pour l'annotation d'images. Pour des annotations basées sur l'ontologie, les étiquettes sont définies par un domaine de connaissances. Cette formalisation des connaissances limite l'utilisateur à l'utilisation d'un certain nombre de descriptions liés à l'ontologie.

[Bannour H. et al, 2011] ont exploré le rôle des ontologies et des approches basées sur les connaissances pour la compréhension de la sémantique des images. L'utilisation de représentations explicites de ces connaissances, c'est-à-dire les approches guidées par des ontologies, est puissante car elle fournit une structuration formelle qui contient des définitions sémantiques explicites qui peuvent être directement traitées par une machine et permet en même temps de décrire des connaissances implicites par inférences automatiques. Là où l'utilisation de mots-clés cherche à être facile d'accès pour l'utilisateur et donne des résultats satisfaisants en termes de recherche, les approches guidées par ontologies se concentrent sur



la nécessité d'avoir des descriptions complètes de l'image et offrent des solutions de recherche plus avancées.

Plusieurs travaux ont cherché à définir des interfaces permettant l'annotation d'image à partir de l'utilisation d'ontologies. *Photostuff* [Halaschek-Wiener C. et al, 2005] propose à l'utilisateur d'annoter des régions d'images et d'éditer les métadonnées. L'annotation se fait ici à l'aide d'ontologies mais ne peut être associée qu'à des formes rectangulaires sur l'image. Dans *M-OntoMat-Annotizer* [Petridis K et al, 2006] (Figure 35) devenu *Video Image Annotation Tool*, les zones peuvent être définies plus précisément. En effet, l'utilisateur peut dessiner des segments (i.e. des régions), choisir le domaine de connaissances pour les ontologies et annoter les régions voulues.

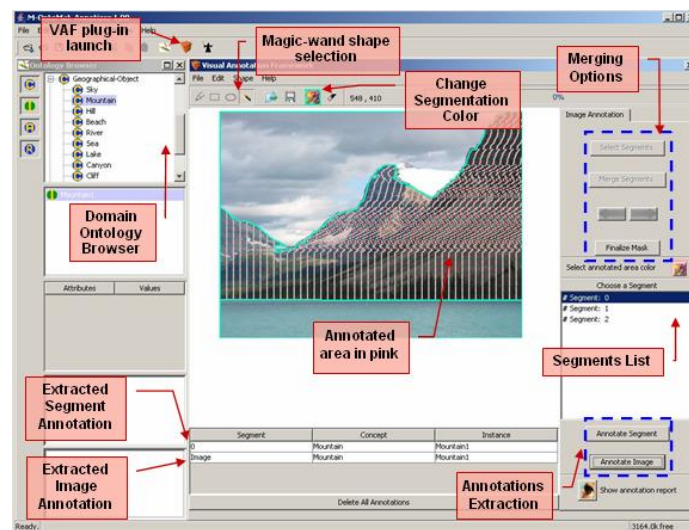


Figure 35 : Interface de M-OntoMat-Annotizer [Petridis K et al, 2006]

Un système d'annotation d'image appliqué au domaine de l'architecture patrimoniale a été développé par [Azzi Y., 2008] (Figure 36). L'utilisateur peut dessiner sur les images la délimitation des objets à décrire et associer à chaque région des termes du vocabulaire architectural.

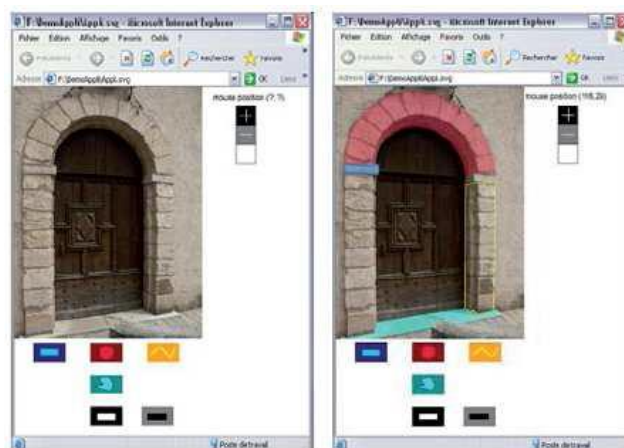


Figure 36 : Annotation sémantique d'images d'architecture patrimoniale [Azzi Y., 2008]

Dans le domaine de la documentation, cette méthode a été adoptée par de nombreux experts, notamment par sa simplicité de mise en œuvre et au vu des supports utilisés.

Cependant tout comme pour l'utilisation de mots-clés, elle ne peut pas être applicable dans le cas où la collection d'images se compose de plusieurs centaines, voire de plusieurs milliers d'images, ce qui est souvent le cas dans le cadre de relevé pour l'étude du patrimoine.

### III.A.3. Utilisation des informations textuelles de contexte

Internet regroupe un grand nombre d'images. Afin d'organiser et de retrouver ces images, les informations des éléments les accompagnant peuvent être utilisées pour l'annotation. Automatiquement il est possible d'associer du contenu sémantique pour des images Web. Des contextes similaires peuvent être assignés à des images en passant par une analyse de la page Web dans laquelle elles sont contenues. Cependant par cette méthode, l'efficacité de l'association sémantique est très mauvaise et peut associer de mauvaises descriptions aux images. Cette faible efficacité provient du fait que les images peuvent dans ce cas être utilisées par tous et qu'il n'existe aucun standard définissant les relations entre les images d'une page et son contenu. D'autre part, les images sont intégrées au Web par différents groupes de personnes avec des objectifs très différents les uns des autres et la qualité des images peuvent varier de façon très significative.

### III.A.4. Annotations automatiques

Afin de chercher à éviter à l'utilisateur de trop grandes manipulations sur les données, des techniques automatiques d'association de contenu sémantique ont été développées. Les méthodes basées sur l'analyse du contenu de l'image (*content-based image analysis techniques*) [Flickner M. et al, 1995] cherchent à extraire des images un certain nombre de caractéristiques de bas niveaux telles que les couleurs, la texture ou la forme afin de reconnaître les éléments qui composent l'image et de les associer à des descriptions. D'autre part, des méthodes plus récentes se basent sur l'utilisation de techniques d'apprentissage pour permettre d'obtenir de meilleurs résultats.

Le concept général du processus automatique d'annotation d'images tel que décrit par [Pandya D. et al, 2014], se compose de cinq phases (Figure 37) :

- **Segmentation** : A cette étape, l'image est partitionnée en groupes de pixels homogènes. La segmentation d'image extrait des caractéristiques visuelles de l'image qui peuvent être séparées ou regroupées afin de construire des zones d'intérêts sur les images. De nombreuses approches existent pour la segmentation d'images (voir Chapitre VII).
- **Extraction de caractéristiques** : Les informations visuelles de bas niveau des images segmentées sont extraites à l'aide de descripteurs tels que la couleur, la texture ou la forme, la couleur et la texture étant les plus expressives pour extraire de caractéristiques visuelles sur les images (voir Chapitre II).
- **Sélection de caractéristiques** : Cette étape consiste à réduire l'espace des caractéristiques par l'utilisation de techniques statistiques telles que l'analyse du composant principal (*Principal Component Analysis*) [Song F., et al, 2010] ou l'algorithme d'optimisation par essaim de particules (*Particle Swarm Optimization Algorithm*) [Jafar O.A.M., et al, 2012].
- **Regroupement/Classification** : Dans cette étape, le groupe de vecteurs de caractéristiques est formé en fonction de techniques de regroupements efficaces comme le K-Means (voir Chapitre VII). Le regroupement inclut une partition du groupe de vecteurs sur la base de caractéristiques communes spécifiées et diverses mesures de similarités.

- **Modèle d'annotation :** L'annotation des images est faite à partir du modèle d'annotation défini de façon à transférer les données de tests sur les images. Les modèles d'annotations peuvent être soit probabilistes, soit de classification, soit paramétriques ou encore basés sur les graphes.

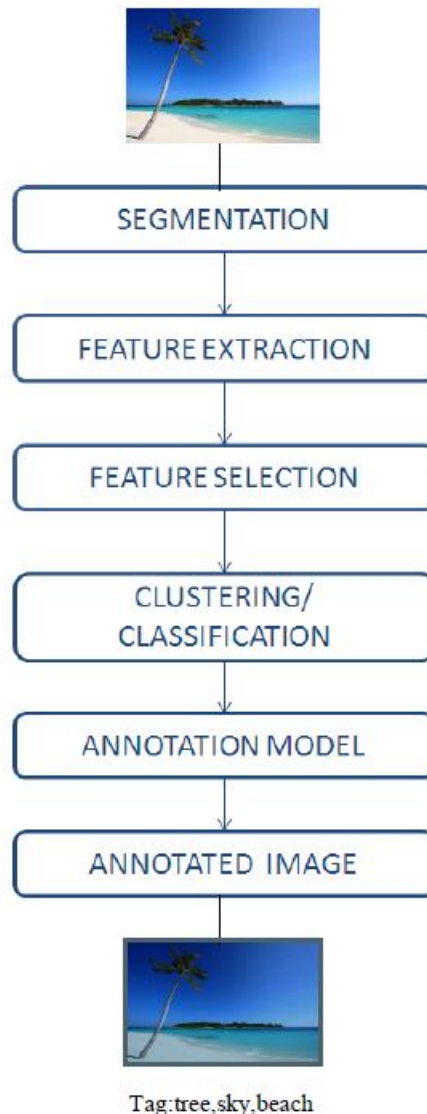


Figure 37 : Concept de l'annotation automatique d'images [Pandya D. et al, 2014]

Les modèles d'annotation automatiques peuvent être classifiés selon quatre approches : probabiliste, classification, graphes ou paramétriques.

#### III.A.4.a. Approches probabilistes

Dans ces approches, l'annotation de l'image est réalisée par l'estimation de la probabilité conjointe d'une image avec un ensemble de mots. [Dyugulu P. et al, 2002] utilisent le modèle de traduction automatique statistique SMT (*Statistical Machine Translation*) et appliquent l'algorithme EM (*Expectation Maximization*) afin d'apprendre un maximum d'associations vraisemblables de mots à des régions d'images en utilisant un corpus bilingue. Le jeu de données prétraitées fourni par [Dyugulu P. et al, 2002] est devenu une référence largement utilisée et populaire des systèmes d'annotation. Le principal aspect de cet algorithme est qu'il



annote l'image en partitionnant les segments en *blobs* et trouve l'association des mots avec les *blobs* en sélectionnant les mots présentant la plus haute probabilité.

[Feng S.L. et al, 2004] proposent un modèle probabiliste générateur qui est basé sur le processus de Bernoulli pour générer les mots et l'estimation de la densité des noyaux pour produire les caractéristiques des images. Il apprend simultanément les probabilités conjointes de l'association de mots avec les caractéristiques des images en utilisant un ensemble d'images d'entraînement contenant des mots-clés. Puis il génère des multiples annotations probables pour chaque image. Cette approche utilise une segmentation multiple et l'extraction des caractéristiques est réalisée par les couleurs et la texture de l'image. Cette approche se concentre plus sur la présence ou l'absence de mots dans une annotation plutôt que sur son importance.

[Mori Y. et al, 1999] proposent un modèle de cooccurrences de mots avec les régions des images en utilisant une grille régulière. Le processus d'annotation commence par partitionner l'image en rectangles de même taille. Chaque rectangle permet d'extraire un descripteur basé sur la couleur et sur la texture. Tous les descripteurs vont alors être regroupés dans des ensembles représentés par leur centroïde. Chaque rectangle hérite de tous les labels de l'image originale, puis l'estimation de la probabilité d'un label est établie par la cooccurrence de ce label.

[Wang B. et al, 2007] proposent un modèle progressif pour estimer la probabilité conjointe de mots pour une donnée d'une image. Le mot avec la plus grande probabilité est d'abord annoté. Puis les mots sont successivement annotés en intégrant les informations des mots préalablement annotés.

#### *III.A.4.b. Approches par classifications*

Dans ces approches, les fonctionnalités sont introduites directement par un classificateur binaire classique qui donne un vote par oui ou par non. Les outils d'apprentissage par la machine les plus communs incluent les SVM (Support Machine Vector), les réseaux de neurones artificiels et les arbres de décision.

#### ***SVM (Support Vector Machine) :***

Le mécanisme des classificateurs SVM fonctionne en cherchant un hyperplan à partir d'un ensemble d'échantillons d'apprentissage afin de les séparer. Un vecteur de caractéristiques et une étiquette est associée à chaque échantillon. Le SVM est fondamentalement un classificateur binaire. La sortie du classificateur est le concept sémantique qui est utilisé pour l'annotation de l'image [Zhang D. et al, 2012]. L'objectif est définir un hyperplan qui partitionne l'ensemble des exemples de façon à ce que tous les points d'un même label soient du même côté de l'hyperplan. [Chapelle O. et al, 1999] utilise cette technique pour entraîner 14 classificateurs SVM pour 14 concepts d'image différents. Afin d'entraîner un SVM pour un concept particulier, les images d'apprentissage correspondant à ce concept sont considérées comme des réponses positives alors que les autres sont vues comme des réponses négatives. Par conséquent, chaque classificateur peut être considéré comme « un contre tous ».

Pendant le test, chaque classificateur génère une décision probabiliste et la classe avec une probabilité maximale est sélectionnée comme étant le concept de l'image testée.

### **Réseaux de neurones artificiels ANN (Artificial Neural Network) :**

Un réseau de neurones artificiels est un modèle de calcul dont la conception est inspirée du fonctionnement des neurones biologiques. Ce sont généralement des réseaux d'apprentissage qui peut apprendre à partir d'exemples et peuvent prendre des décisions sur un nouvel échantillon. Ils se composent de plusieurs couches de nœuds interconnectés, appelés neurones. La première couche correspond à la couche d'entrée et a autant de neurones que de données d'entrée. La couche de sortie a par sa part un nombre de neurones correspondant au nombre de classes.

Plusieurs types de réseaux neuronaux existent et ont été exploité dans le processus d'annotation automatique d'images. [Socher R. et al, 2011] utilisent un réseau neuronal récurrent RNN (*recursif neural network*) pour détecter des éléments d'une scène naturelle (Figure 38).



Figure 38 : Exemples de résultats de détection par réseaux neuronaux pour [Socher R. et al, 2011]

Le Google Lab Research a pour sa part utilisé des réseaux neuronaux convolutionnels CNN (*convolutionnal neural networks*) [GoogleLabResearch, 2014] permettant de générer une phrase décrivant le contenu de l'image (Figure 39).

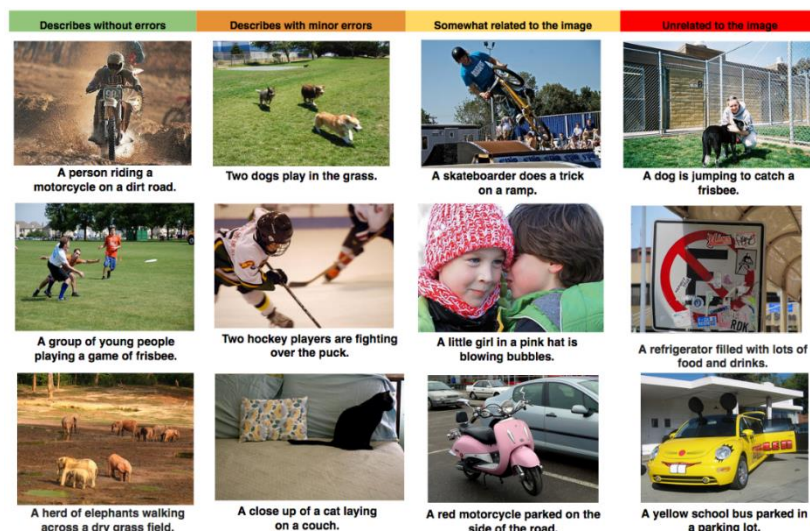


Figure 39 : Exemples de résultats obtenus par le système du Google Lab Research [GoogleLabResearch, 2014] classés en fonction de l'évaluation de la description

Les performances sur les résultats de l'annotation automatique sont fortement affectées par les résultats de la segmentation. Ainsi afin d'éviter la segmentation, [Zhao Y. et al, 2008] ont proposé une approche appelée LSA (*Latent Semantic Analysis*) basée sur le modèle de réseaux neuronaux qui consiste en trois phases : recherche de corrélation contextuelle entre les mots-clés, recherche de caractéristiques à partir des images d'apprentissage et annotation de l'image.

### Arbre de décision DT (*Decision Tree*) :

Un arbre de décision définit une stratégie de décision en plusieurs étapes. En fonction du nombre de décisions prises à chaque nœud, un arbre de décision peut être appelé arbre binaire ou n-aire. Pendant la phase d'apprentissage, l'arbre est construit en divisant de manière récursive les échantillons d'apprentissage en plusieurs ensembles qui ne se chevauchent pas et à chaque fois que les échantillons sont divisés, l'attribut utilisé pour la division est supprimé. La procédure se poursuit jusqu'à ce que tous les échantillons d'un groupe ou d'une classe atteignent sa profondeur maximale ou lorsqu'il ne reste plus aucun attribut de divisions.

[Wan S., 2011] propose un algorithme d'arbre de décision simple SDT (*Simple Decision Tree*) construisant l'arbre de manière récursive de haut en bas. Dans cette approche, l'algorithme SDT est utilisé pour améliorer la décision en calculant un modèle par une recherche heuristique dans l'espace des modèles. Pour identifier un nouvel échantillon, l'arbre est parcouru depuis le nœud à la racine jusqu'à un nœud « de feuille » en utilisant la valeur d'attribut du nouvel échantillon. La décision pour l'échantillon est le nœud « de feuille » lorsque l'échantillon l'atteint. Cet algorithme est facile à interpréter et peut apprendre avec un petit nombre d'échantillon. Il est également robuste pour des données incomplètes ou bruitées. L'avantage de ce type d'approche est que la recherche est efficace. L'inconvénient est qu'elle ne tient pas compte du fait que de nombreuses images peuvent appartenir à plusieurs catégories.

#### *III.A.4.c. Approches par les graphes*

[Liu J. et al, 2009] ont défini une approche basée sur l'apprentissage par les graphes pour les images et pour les mots. Pour une donnée annotée de l'ensemble d'apprentissage, et à partir des caractéristiques visuelles de l'image, l'apprentissage par les graphes pour l'image vise à propager les étiquettes à partir des images annotées vers les images non-annotées à l'aide de leurs similitudes visuelles. Ce type d'apprentissage ne se concentre que sur les similitudes entre images et ne cherche pas les corrélations possibles entre les mots. Deux mots à forte cooccurrence dans l'ensemble des données d'apprentissage conduiront à une forte probabilité d'annoter certaines images conjointement. Trois sortes de corrélation peuvent être exploitées. L'une est la cooccurrence des mots dans l'ensemble d'apprentissage et les deux autres proviennent du contexte Web.

#### *III.A.4.d. Approches paramétriques*

Dans ces approches, l'espace des caractéristiques est supposé suivre un certain type de distribution continue connue. La probabilité conditionnelle est modélisée en utilisant une distribution gaussienne multivariée.

[Li J. et al, 2006] décomposent les images d'apprentissage en fonction des concepts par des régions qui sont représentées par les couleurs LUV (L : clarté, U et V : chrominance) et les caractéristiques de textures d'ondelettes. Puis ils regroupent les régions dans des groupes qu'ils appellent prototypes. Pour chaque prototype, un modèle gaussien est défini. Finalement un modèle de mélange gaussien GMM (*Gaussian Mixture Model*) est construit pour chaque concept en faisant la moyenne des prototypes individuels dans les concepts. Pour annoter une image inconnue, les caractéristiques des régions sont extraites et la probabilité que l'image appartienne à un concept est calculée à nouveau sur la base du GMM. L'inconvénient de cette méthode est que l'estimation des paramètres pour les modèles gaussiens est complexe.

[Carneiro G. et al, 2007] ne partitionnent pas l'image en régions mais ils supposent que les caractéristiques de l'image suivent certaines distributions gaussiennes et apprennent directement les GMM pour chaque image d'apprentissage avec l'algorithme EM (*Expectation Maximization*). Ceci équivaut à simultanément segmenter et apprendre du modèle. Dans la phase d'annotation, un GMM est construit pour chaque image inconnue, puis est adapté aux GMM de chacun des modèles des concepts. Les concepts avec la plus grande correspondance sont sélectionnés pour annoter l'image.

#### *III.A.4.e. Conclusion sur les méthodes d'annotation automatique*

Ces méthodes d'annotation automatique sont beaucoup plus rapides que les méthodes d'annotations manuelles et ne nécessite pas d'intervention humaine. [Pandya D. et al, 2014] ont établi un tableau comparatif de ces méthodes (Figure 40).

APPROACH	ADVANTAGE	DISADVANTAGE
Probabilistic	Multiple labelling	Large no of parameters
	Fast ,Model free	Annotation results are highly affected by segmentation
Classification	Does not rely on segmentation	Cannot handle single object belonging to multiple category
	Robust and accurate	Multiple classifier needed for large sample for data
Parametric	Does not rely on segmentation	Estimation of the parameters is complex
	Multilabelling	Training time is expensive
Graph	Parameters are easy to tune	Does not consider the correlation between words

Figure 40 : Comparaison des approches d'annotation automatique d'images [Pandya D. et al, 2014]

Même si ces méthodes montrent des résultats intéressants, pour le domaine d'application qui nous intéresse, elles ne sont pas suffisamment précises à cause de la quantité de formes hétérogènes dans des applications pour le patrimoine. En effet, les objets patrimoniaux offrent souvent des variations morphologiques minimales. D'autre part, il est généralement nécessaire d'avoir plusieurs vues afin de pouvoir classifier de manière précise un objet.

L'étape de segmentation pose beaucoup de problème sur des sujets aussi détaillés que des objets patrimoniaux. Le problème de la détection des formes sur des objets patrimoniaux est particulièrement difficile à traiter (Figure 41). Les nombreux algorithmes existants (voir [Chapitre VII](#)) ne sont généralement applicables qu'à des études de cas spécifiques.



Figure 41 : Exemples d'objets patrimoniaux complexes [3DIcons]

Le problème de l'apprentissage, et donc de l'identification, est que cela nécessite l'utilisation d'un très grand nombre de références pour pouvoir comparer les éléments. Par conséquent cela nécessite d'avoir de très grandes quantités de données répertoriant les

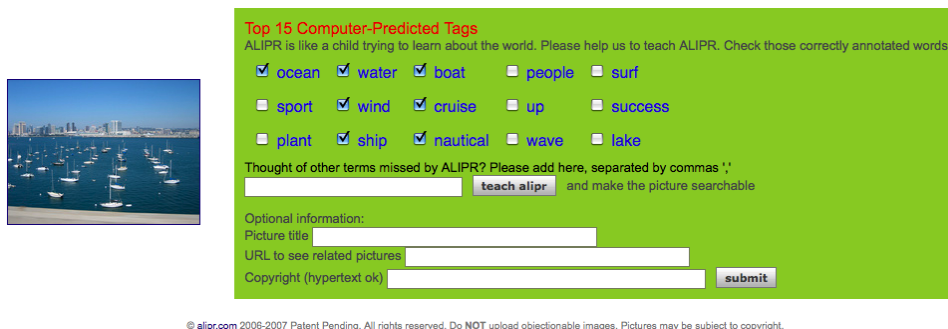


différentes formes. Pour le domaine d'application envisagé, il n'est pas possible d'imaginer d'avoir ce genre de base de données au vu du nombre d'instances important et par conséquent l'absence de modèles théoriques fiables. En effet, les objets patrimoniaux ne font pas l'objet de production en série et sont donc uniques.

### III.A.5. Méthodes semi-automatiques

Les méthodes semi-automatiques sont des combinaisons des méthodes manuelles et des méthodes automatiques. A l'aide d'algorithmes de reconnaissance de formes tels que ceux décrits précédemment (voir [III.A.4. Annotations automatiques](#)), ou d'analyse d'informations textuelles associées à l'image, un ensemble de mots-clés semblant pertinents est sélectionné. Un contrôle par l'utilisateur des mots-clés trouvés permet d'affiner et de faciliter le travail d'annotation.

De nombreux sites gérant des bases de données d'image proposent ce système afin de simplifier le travail d'annotation des utilisateurs. Parmi eux, ALIPR (*Automatic Photo Tagging and Visual Image Search*) [Li J. et al, 2006] permet d'annoter automatiquement ses photographies et de rechercher des photographies par mots-clés. L'image est chargée sur le site puis une liste de mots-clés est proposée à l'utilisateur, qui va valider les mots-clés qu'il juge pertinent (Figure 42). Cependant, l'utilisateur peut tout de même ajouter des mots-clés supplémentaires qui ne sont pas proposés par le site pour permettre au site d'enrichir ses possibilités de descriptions et de continuer son apprentissage.



The screenshot shows the ALIPR web interface. On the left is a small image of a harbor with many sailboats. To the right is a green box containing the following text and form elements:

**Top 15 Computer-Predicted Tags**  
 ALIPR is like a child trying to learn about the world. Please help us to teach ALIPR. Check those correctly annotated words.

☒ ocean   ☒ water   ☒ boat   ☐ people   ☐ surf  
☐ sport   ☒ wind   ☒ cruise   ☐ up   ☐ success  
☐ plant   ☒ ship   ☒ nautical   ☐ wave   ☐ lake

Thought of other terms missed by ALIPR? Please add here, separated by commas ''  
  and make the picture searchable

Optional information:  
 Picture title   
 URL to see related pictures   
 Copyright (hypertext ok)

© alipr.com 2006-2007 Patent Pending. All rights reserved. Do NOT upload objectionable images. Pictures may be subject to copyright.

Figure 42 : Résultats de détection d'ALIPR [Li J. et al, 2006]

Les méthodes semi-automatiques permettent d'obtenir des résultats nettement meilleurs que les méthodes tout-automatiques car elles prennent en compte l'interprétation humaine en demandant validation par un utilisateur. Elles restent aussi plus rapides que les méthodes manuelles car l'utilisateur n'a pas à chercher lui-même les mots-clés pour décrire le sujet de l'image.

Cependant ces méthodes demandent tout de même un travail de vérification et de validation important. De plus, les limitations de méthodes automatiques restent car les algorithmes de reconnaissance de formes sont les mêmes.

Tout en restant les méthodes les plus performantes, il n'est pas envisageable d'utiliser ces méthodes dans le domaine de la documentation patrimoniale car le problème de la complexité des formes persiste. En effet, au vu de l'hétérogénéité des formes des objets patrimoniaux, il est impossible de pouvoir envisager la constitution d'une base de modèles pour la reconnaissance.

### III.B. Annotation de modèles 3D

L'utilisation de représentation 3D comme support pour la documentation du patrimoine est aujourd'hui bien ancrée dans les pratiques. De telles représentations permettent de naviguer autour de l'objet et de se rendre compte de la disposition des éléments entre eux. Ces modèles sont facilement diffusables au sein d'un ensemble de partenaires. C'est pourquoi il est nécessaire de pouvoir annoter un modèle 3D.

Comme vu dans le [Chapitre II](#), il existe plusieurs méthodes permettant d'acquérir des objets patrimoniaux. Cependant ces méthodes génèrent des nuages de points et les méthodes actuelles d'annotation de représentation 3D nécessitent d'avoir un modèle surfacique ou maillé sans perte d'informations.

#### III.B.1. Reconstruction 3D

La manière la plus simple de construire un modèle 3D est par la géométrie (*geometry-based 3D*). Chaque partie de l'objet est construite à partir d'éléments géométriques basiques que l'on peut généralement modifier pour qu'ils représentent au mieux l'objet que l'on veut représenter. Cette manière de construire le modèle oblige l'utilisateur à tout définir. Des nuages de points peuvent être utilisés pour servir à guider l'utilisateur dans sa reconstruction. La géométrie résultante sera par contre réduite à une représentation plus approximative.

Le maillage polygonal est généralement la méthode la plus adaptée pour représenter correctement les données des mesures. Beaucoup de méthodes ont été développées pour créer une représentation polygonale et continue d'un maillage à partir d'un nuage de points. La conversion d'un nuage de points en une surface polygonale est généralement basée sur quatre étapes :

- **Prétraitement** : Cette phase élimine les données bruitées ou extrait un échantillon du nuage de points pour réduire la quantité de données à traiter. Ces étapes sont généralement obligatoires puisque les nuages de points relevé à l'aide de scanner ou généré à partir d'images possèdent de nombreuses redondances.
- **Topologie globale de la surface** : Cette étape établit des relations de voisinages entre les points. Cela nécessite généralement l'utilisation d'informations morphologiques telles que les normales pour la gestion des discontinuités.
- **Génération du maillage** : Des mailles sont créées répondant à certaines exigences de qualité. Cette étape produit des sommets, des arêtes et des faces qui se relient entre elles sur les bords partagés afin de représenter la surface. Plusieurs algorithmes existent pour générer ces surfaces. Un des plus connus est la reconstruction par la méthode Poisson (*Poisson Reconstruction*) [[Kazhdan M. et al., 2006](#)]. La méthode de Poisson échantillonnée (*Screened Poisson Reconstruction*) [[Kazhdan M. et al., 2013](#)] est une évolution de cette méthode qui offre des résultats plus précis (Figure 43).
- **Post-traitement** : Une fois le modèle créé, des opérations de correction sont appliquées pour affiner et perfectionner la surface polygonale obtenue.

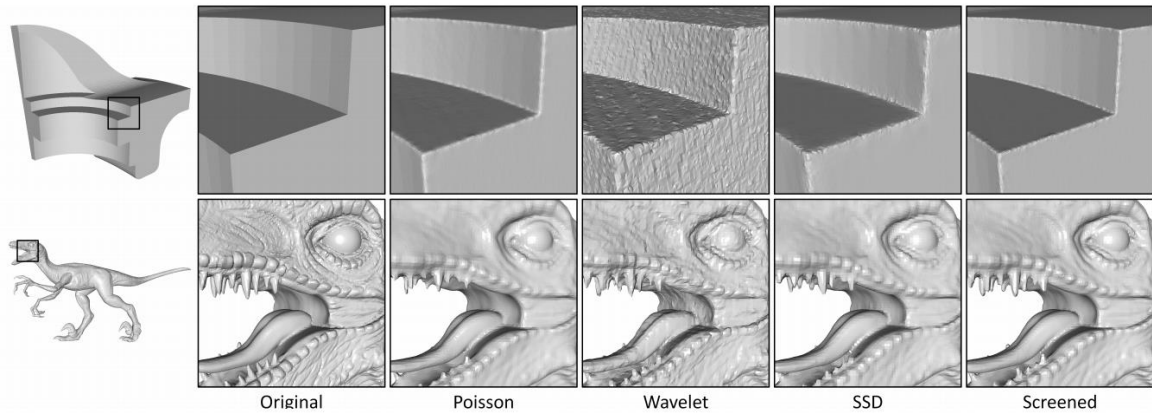


Figure 43 : Exemples de reconstruction surfacique pour différentes méthodes [Kazhdan M. et al., 2013]

### III.B.2. Annotations 3D de modèles

Les méthodes d'annotation 3D permettent d'enrichir sémantiquement les informations géométriques et morphologiques de l'objet représenté. La possibilité d'attacher du contenu sémantique sur des modèles 3D intéresse grandement les conservateurs car ils souhaitent pouvoir annoter des caractéristiques significatives relatives à la provenance de l'objet. La plupart de ses méthodes cherchent à associer aux parties du modèle des informations sémantiques.

Certaines méthodes se sont basées sur l'utilisation de standards pour conduire l'annotation sémantique. 3DSEAM (3D Semantic Annotation Model) [Bilasco I.M. et al, 2005] (Figure 44) est un modèle d'annotation sémantique d'objets 3D utilisant le standard X3D. Il vise à indexer le contenu 3D en considérant les aspects visuels géométriques et sémantiques. Il est instancié en utilisant MPEG-7, prolongé avec des repères spécifiques. Ces repères lient les caractéristiques aux fragments correspondants de X3D.

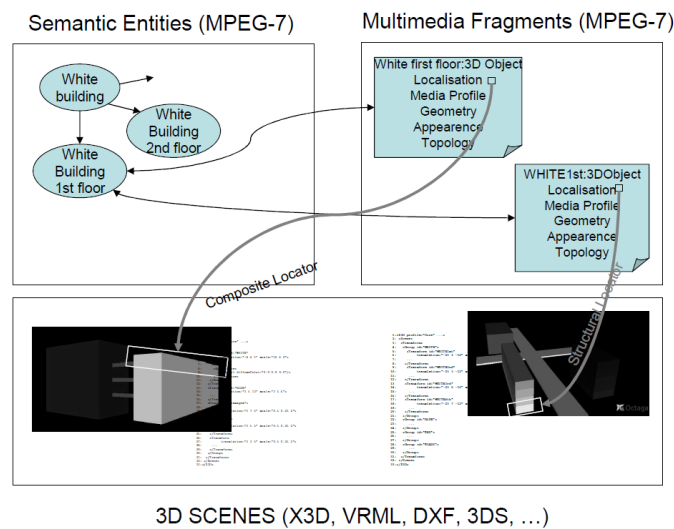


Figure 44 : Schéma global du modèle 3DSEAM [Bilasco I.M. et al, 2005]

ShapeAnnotator [Attene M. et al., 2007] [Attene M. et al., 2009] est un système qui permet d'effectuer des segmentations de mailles d'une surface 3D et d'annoter les parties détectées par des concepts exprimés par une ontologie (Figure 45). Chaque partie est connectée à une



instance qui peut être stockée dans une base afin de faciliter la recherche. Plusieurs segmentations sont proposées afin de pouvoir choisir exactement la partie considérée.

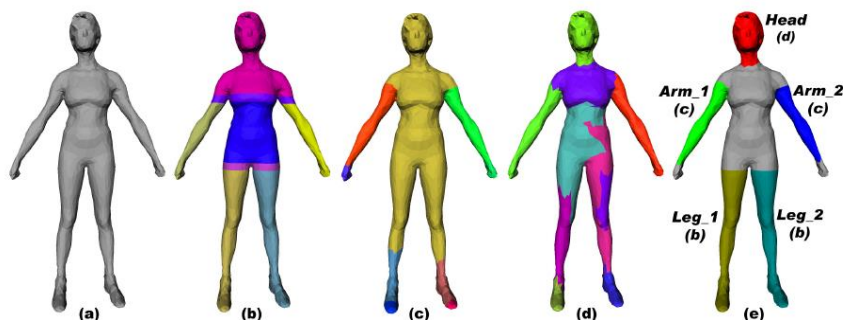


Figure 45 : Une mesh de base (a) est segmentée en utilisant plusieurs algorithmes (b), (c) et (d) puis seules les parties appropriées des chaque segmentation (b), (c) ou (d) sont sélectionnées et annotées (e) [Attene M. et al, 2009]

D'autres méthodes utilisant aussi la segmentation du modèle 3D existent. [Manferdini A.M. et al, 2010] ont développé un système permettant suite au traitement de segmentation et d'annotation de diffuser les résultats obtenus sur des outils open-source disponibles sur le Web afin de permettre l'accès aux données au plus grand nombre d'utilisateurs et afin de diffuser les connaissances (Figure 46).

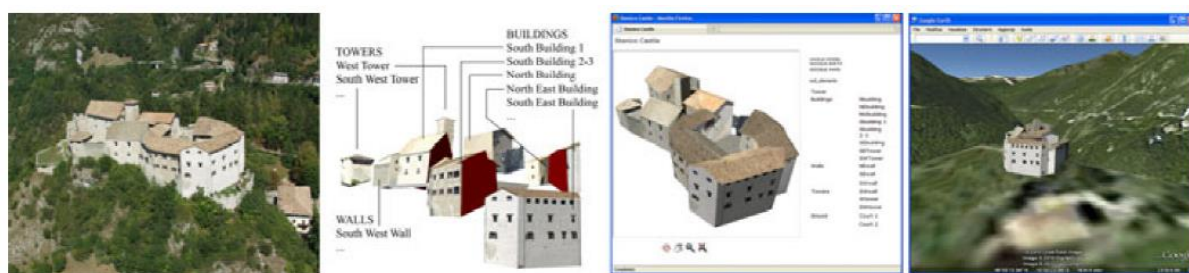


Figure 46 : De gauche à droite : (1) Modèle de base, (2) segmentation de la géométrie en sous éléments, (3) définition de l'organisation hiérarchique et (4) visualisation sur un outil open-source 3D [Manferdini A.M. et al, 2010]

Le projet Arrigo [Havemann S. et al, 2008] a élaboré un système interactif permettant aux visiteurs de l'exposition d'explorer des modèles 3D de statues et de découvrir des informations détaillées sur ces objets via des annotations attachées à des parties de l'objet 3D (Figure 47). Les annotations sont stockées en TEI/XML et basées sur le modèle CIDOC/CRM. Cependant les régions d'intérêt ne peuvent être que de formes sphériques.



Figure 47 : Aperçu de l'interface du projet Arrigo [Havemann S. et al, 2008]

Le 3DSA (*3D Semantic Annotation*) développé par [Hunter J. et al, 2011] [3DSA, 2014] est un système permettant à l'utilisateur d'attacher des annotations à des points, des surfaces, des régions ou des segments d'objets 3D (Figure 48). Ce système est basé sur l'utilisation de tags définis par une ontologie.



Figure 48 : Exemple d'objets sur [3DSA, 2014]

Les méthodes d'annotations 3D permettent donc de pouvoir sémantiser directement la géométrie de l'objet. Cependant elles sont généralement dépendantes d'algorithmes de segmentation qui, comme pour les images, donnent des résultats variées en termes de parties détectées.

Les formes pouvant supporter l'annotation sont limitées. Il faut soit sélectionner des parties de l'objet 3D, soit créer un nouvel objet dans la scène pour y attacher la description. La sélection de parties sur un modèle ne se fait pas de façon aisée. La manipulation du modèle 3D est plus difficile à mettre en œuvre que la manipulation d'images. C'est pourquoi les modèles 3D n'apparaissent pas comme des supports faciles à utiliser pour la manipulation.

### III.C. Méthodes d'annotations hybrides 2D/3D

L'utilisation séparée de supports 2D et de supports 3D pour l'annotation montrent que ces deux supports possèdent un certains nombres d'avantages et d'inconvénients. Les inconvénients de l'un deviennent généralement des avantages sur l'autre. Ainsi l'utilisation combinée de ses deux types de supports semble être une perspective intéressante. Cependant cela pose un problème de mise en relation des deux types de supports. Pour l'être humain, il est facile de situer deux photos l'une par rapport à l'autre car les connaissances et la vision dans l'espace qu'il possède lui permet de faire le lien entre les deux. Pour un ordinateur, il faut traduire ces connaissances afin qu'il puisse faire le même travail que le cerveau humain. Il faut donc lui donner des renseignements de position dans l'espace pour faire le lien entre les images.

Deux travaux principaux [Snavely N. et al, 2006] [Busayarat C., 2010] se sont déjà intéressés à l'interfaçage entre les images et le modèle 3D. Dans les deux cas, le lien entre images et modèle 3D passe par l'orientation dans l'espace des images autour du modèle.

### III.C.1. Phototourism/Photosynth

*Phototourism* [Snavely N. et al, 2006] est un système pour naviguer au sein de larges collections de photographies en 3D (Figure 49). Cette approche prend en entrée de larges collections d'images qui peuvent provenir de collections personnelles ou de site internet de partage de photographies. La position de la vue des photographies est calculée et un nuage de point léger de la scène est créé. L'interface permet à l'utilisateur de se déplacer d'une image à l'autre à travers l'espace 3D. Ce système est aujourd'hui implémenté par Microsoft dans Photosynth [PHOTOSYNTH, 2008].

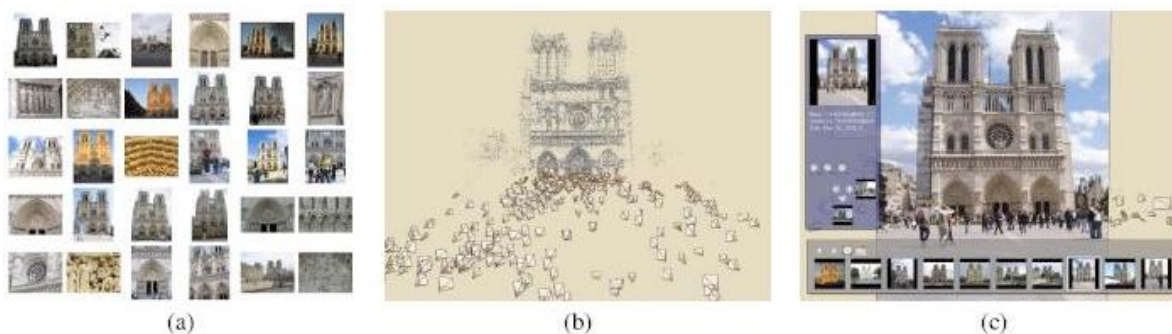


Figure 49 : Etapes de générations : (a) collections d'images, (b) génération du nuage de points, (c) interface de Photosynth [Snavely N. et al, 2006]

Grace à cette connaissance de la position des images autour du nuage de points, le système permet de faire des annotations sur les images et d'automatiquement transférer cette annotation sur les images où l'objet sélectionné apparaît. A partir d'une sélection sur une image, le système détecte un ensemble de points du modèle 3D apparaissant dans l'image. Puis ces points sont recherchés sur les autres images où ils apparaissent (Figure 50).

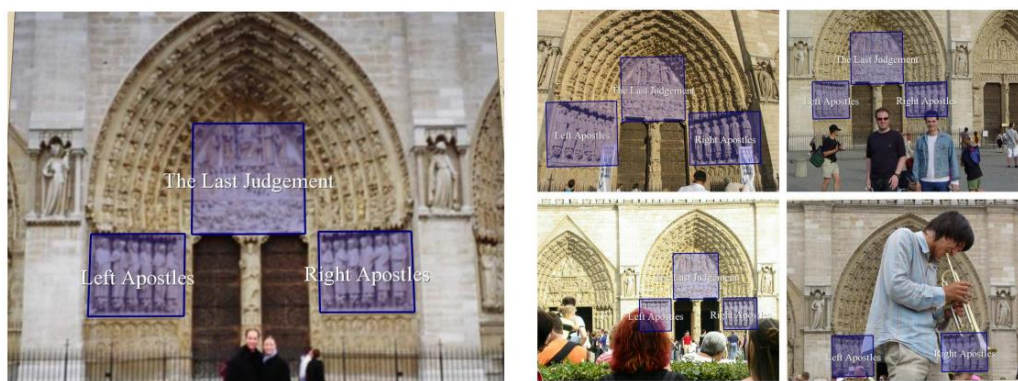


Figure 50 : Annotations sur Phototourism [Snavely N. et al, 2006]

Grâce à ce système, les annotations sont donc transférées automatiquement d'une image à l'autre. Cependant, au vu de la densité du nuage de points et donc de la quantité de points sur lesquels l'annotation peut se reposer pour le transfert, il n'est pas possible de dessiner ou de transférer d'autres formes que des rectangles. Dans le cas de la documentation d'objets patrimoniaux, cette imprécision au niveau de la zone sélectionnée peut rapidement limiter les possibilités d'analyses.



### III.C.2. Nubes Imago

[Busayarat C., 2010] a développé un système utilisant la maquette numérique comme support pour la recherche visuelle d'informations patrimoniales. L'idée est permettre la sémantisation de sources iconographiques par référencement spatial autour d'une maquette numérique.

Le système repose sur l'obtention d'une représentation complète de l'édifice à partir d'un relevé terrain. Le modèle obtenu est décomposé en éléments caractéristiques, qui sont chacun qualifié par un ensemble d'attributs sémantiques. Les images sont ensuite référencées spatialement autour de la maquette (Figure 51).

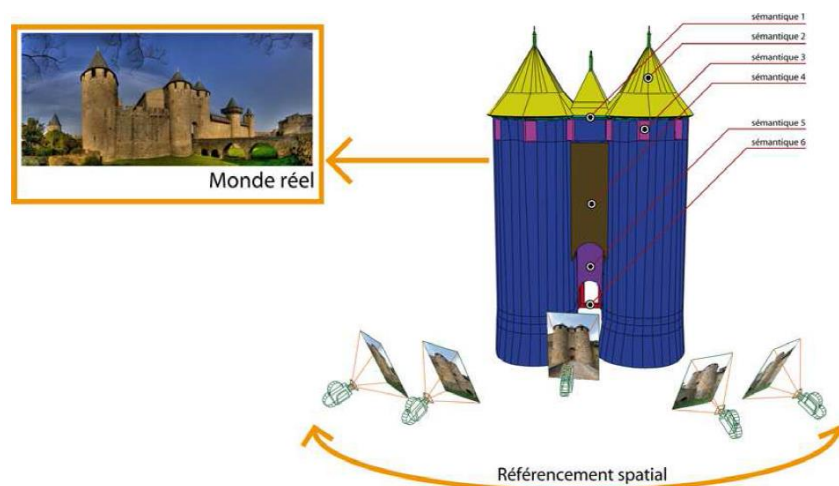


Figure 51 : Maquette numérique et référencement spatial des images [Busayarat C., 2010]

A partir de ce référencement spatial, la description sémantique du modèle 3D peut être projetée vers les images référencées spatialement. Les images reçoivent donc une couche sémantique par projection des contours des formes 3D suivant la vue de l'image (Figure 52). Cependant cette projection ne permet pas de gérer l'aspect d'occlusions dans les images.

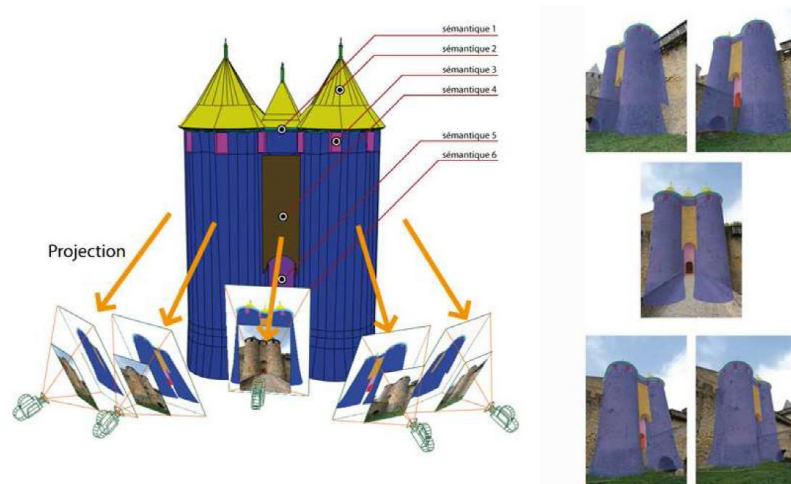


Figure 52 : Projection d'annotations du modèle 3D sur les images 2D [Busayarat C., 2010]

Ce système montre que l'utilisation de la maquette numérique comme support d'informations peut permettre à des experts d'établir des critères de recherche visuels. De plus,

les informations de la maquette numérique peuvent être projetées sur les images représentant l'objet.

D'autre part, [De Luca L. et al, 2010] ont étudié les modalités pour pouvoir associer ce système de visualisation d'annotation aux aspects temporels. Le principe repose sur l'utilisation de maquettes représentant différents états temporels (Figure 53). Ces travaux montrent que la spatialisation des informations peut aussi permettre de gérer le suivi d'évolution temporel sur un objet.

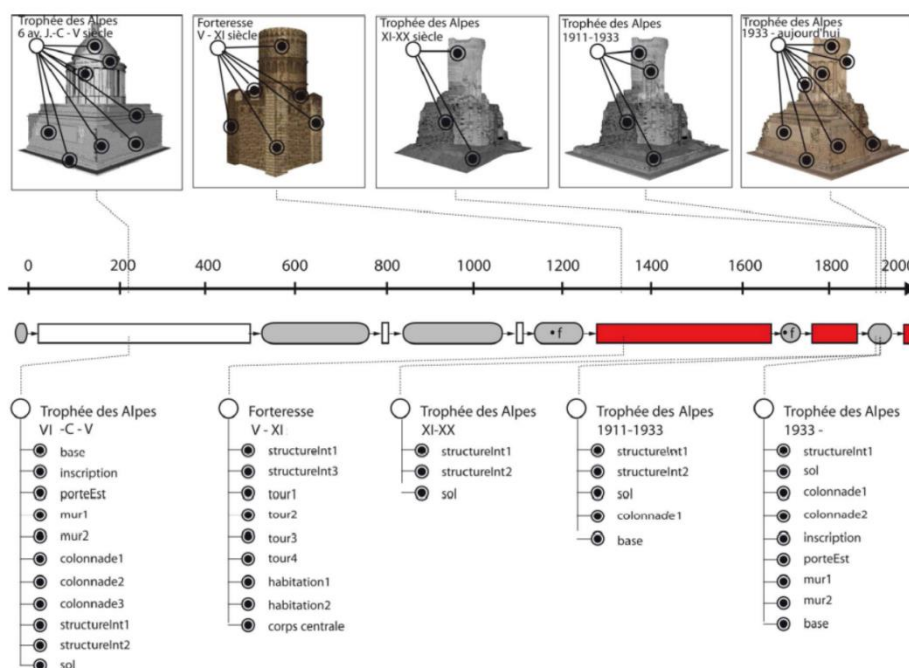


Figure 53 : Représentation spatio-temporelle des transformations d'un bâtiment [De Luca L. et al, 2010]

Le fait de devoir construire une maquette complète de l'objet est déjà en soi une étape lourde en manipulation. Les étapes de relevé et de reconstruction bien que aujourd'hui partiellement automatisées demandent beaucoup de travail. De plus, la caractérisation sémantique de la maquette suppose une manipulation et une sélection sur la maquette qui n'est pas toujours aisée. Enfin, les images ne peuvent servir que pour visualiser les informations et ne peuvent pas être utilisées pour annoter l'ensemble bien qu'elles soient un support d'annotations plus pratique que la maquette numérique.

### III.D. Constat sur les méthodes actuelles d'annotation pour la documentation du patrimoine

Les méthodes actuelles d'annotations montrent bien qu'elles sont nécessaires dans le cadre de la documentation d'un objet patrimonial. L'annotation d'image a toujours été la base pour établir des analyses que ce soit pour de la cartographie d'altérations ou pour simplement de la description d'objets. Avec le développement des modèles 3D, ceux-ci se sont avérés être des supports contenant de grands nombres d'informations sur les caractéristiques de l'objet représenté et donc indispensables à la documentation.

L'utilisation de supports 2D ou de supports 3D seuls ne suffit plus aux experts. Les deux représentations doivent pouvoir être utilisées en parallèle pour établir des analyses plus poussées des objets patrimoniaux. Les travaux de [Snavely N. et al, 2006] et [Busayarat C., 2010] montrent bien que l'utilisation des deux supports et leur mise en relation offre de nouvelles possibilités pour l'annotation sémantique.

La relation spatiale établie entre les deux supports permet de faire le lien entre toutes les différentes représentations de l'objet, qu'elles soient 2D ou 3D. Ce lien peut permettre de consolider les aspects avantageux de chacune des représentations en termes de méthodes d'annotations. Si l'annotation est plus simple à réaliser sur une image, le fait de pouvoir lier cette annotation 2D à une annotation 3D simplifie le travail d'annotation du modèle 3D qui lui est nécessaire à l'extraction d'informations géométriques. D'autre part, ce lien peut également ouvrir des perspectives pour le suivi temporel (4D) des objets.

Le problème reste de pouvoir gérer les différents aspects que cela apporte. En effet, la relation spatiale doit pouvoir faire face à des images représentant différentes échelles d'un objet et aux différentes résolutions de chacun des supports. D'autre part, les différents supports peuvent représenter l'objet dans différents états. L'hétérogénéité des données par rapport à ces aspects et la quantité qu'elles représentent impose donc la nécessité d'établir un système de propagation automatique d'annotation intégrant la possibilité de gérer ces aspects afin de simplifier le travail des experts.





## IV. Approche pour la caractérisation sémantique d'objets patrimoniaux à partir d'images spatialisées

Les deux chapitres précédents exposent les outils mis à disposition des experts de la documentation d'objets patrimoniaux. Cependant les méthodologies développées ne permettent pas encore de complètement intégrer l'aspect de continuité d'informations. En effet, la nécessité pour les experts est d'utiliser plusieurs représentations possibles (images, nuage de points, orthophotographies, images IR et UV, ...) et de pouvoir continuellement mettre à jour les données en fonction de l'évolution de l'objet tout en maintenant le lien entre les différentes représentations.

Suite à l'analyse des méthodes actuelles d'annotation (voir [Chapitre III](#)), on peut retenir que ces méthodes ne sont pas encore complètement au point pour aborder le problème plus complexe de l'annotation précise d'artéfacts patrimoniaux. Même si les méthodes semblent offrir des résultats convenables, notre domaine d'application reste difficile à traiter au vu la complexité et de l'hétérogénéité des formes et des supports de documentation. Notamment ces méthodes ne permettent pas d'obtenir de manière efficace un support d'analyse. Elles nécessitent généralement la génération d'un modèle 3D complet qui sera ensuite enrichi sémantiquement [[Busayarat C., 2010](#)] [[Stefani C. et al, 2012](#)].

La complexité provient du fait que l'analyse ne peut pas se réduire à une analyse bidimensionnelle de l'objet mais doit se constituer à partir d'une véritable analyse tridimensionnelle des formes. Cependant, les méthodes d'annotation de modèle 3D restent complexes au niveau de la manipulation surtout dans un domaine où les descriptions doivent correspondre à des zones précises de l'objet et pas seulement à une simple décomposition en formes. C'est pourquoi les travaux liant l'information 2D à l'information 3D montrent qu'il est possible d'envisager des solutions pour pallier ces difficultés liées aux différentes représentations et de plus, d'offrir à l'utilisateur une correspondance à la fois bidimensionnelle et tridimensionnelle à sa description.

Ceci montre bien que, dans le domaine de l'étude du patrimoine, l'étude et la description des éléments doit s'appuyer sur un support de représentation plus sophistiqué que seulement des images ou seulement un modèle 3D. C'est un support qui doit être capable d'explicitier la complexité morphologique de l'objet ainsi que les caractères spécifiques aux éléments qui le composent tout en permettant une manipulation aisée de l'objet et une visualisation multi-représentations afin de pouvoir pousser l'analyse à un plus haut niveau.

Les outils utilisés aujourd'hui dans le domaine du relevé du patrimoine (voir [Chapitre II](#)) permettent de collecter, de structurer des informations acquises à différents niveaux de représentations en grande quantité afin de construire des représentations tridimensionnelles d'objets patrimoniaux. Les reconstructions tridimensionnelles permettent de disposer de plusieurs niveaux de détails sur un objet patrimonial.

Une hypothèse, introduite par [[Blaise J.Y. et al, 2004](#)] considère que le modèle 3D peut être considéré comme un moyen d'accès privilégié à l'information patrimoniale. La représentation morphologique d'un édifice peut être utilisée comme source principale pour propager les attributs sémantiques sur un ensemble de sources iconographiques spatialisées

autour du modèle tridimensionnel. Les travaux de [Busayarat C., 2010] (voir Chapitre III) vont dans ce sens mais n'exploitent qu'une partie de la relation entre les images et le modèle 3D et n'offrent pas de solution pratique à l'annotation puisqu'il faut d'abord élaborer un modèle 3D sémantisé ce qui demande un travail long et peu ergonomique. D'autre part, la réelle nécessité de disposer d'un modèle 3D aussi détaillé, en termes de reconstruction, peut se poser. Les travaux de [Snaveley N. et al, 2006] montrent qu'une simple référence spatiale peut suffire même si, dans leur cas, le degré de précision et la densité d'informations ne sont pas suffisants pour offrir une solution satisfaisante à notre domaine d'application.

Il est ainsi possible d'envisager une solution où une simple référence spatiale suffisamment dense pourrait permettre de définir la description sémantique (pouvant être structurée par un ensemble de termes) comme le point commun entre les représentations 3D de la géométrie de l'objet et les acquisitions (images provenant de différents capteurs) représentant l'objet (Figure 54).

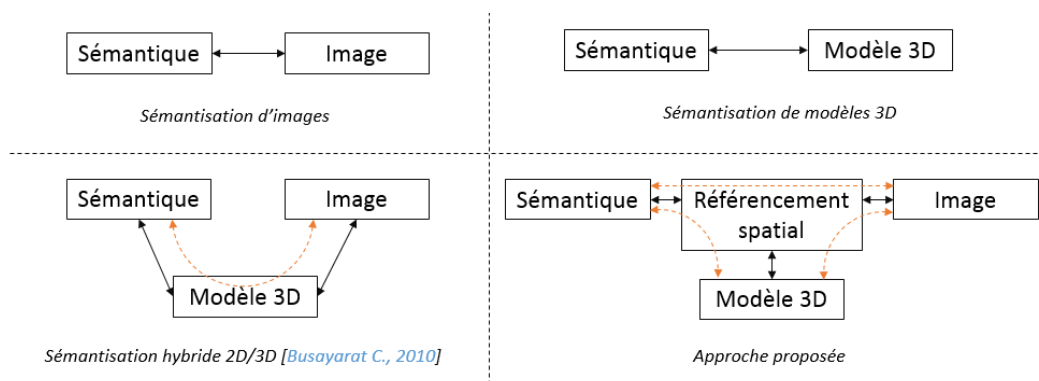


Figure 54 : Différence entre les approches actuelles de sémantisation

Un tel modèle offre une relation cohérente et continue entre les données hétérogènes utilisées pour l'analyse, la compréhension et la documentation du patrimoine bâti. L'objectif est de définir un continuum d'information entre toutes les phases du processus allant de l'acquisition d'images et de données spatiales jusqu'à la construction de représentations 3D et la structuration d'informations sémantiquement enrichies en intégrant les différents aspects utiles à l'analyse tels que l'intégration des différentes représentations (laser, images numériques, orthophotographies) à différentes échelles, la superposition de plusieurs couches d'analyses correspondant à plusieurs niveaux de lecture ou encore l'aspect temporel lié aux transformations de l'objet dans le temps et ainsi introduire le problème des acquisitions multiples dans le temps (Figure 55). L'idée est de pouvoir insérer des éléments de sémantique au niveau de toutes les phases du traitement des données et notamment dès les premières phases d'acquisition de données sur le terrain.

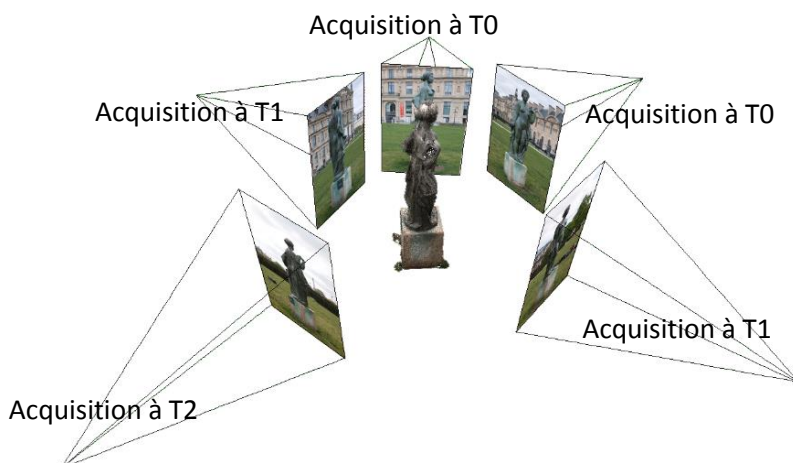


Figure 55 : Objet patrimonial et acquisitions multiples

Cela demande de mettre en place une méthodologie complète de caractérisation sémantique 2D/3D en se basant sur la notion de référencement spatial déterminant les relations projectives entre les différentes représentations (2D et 3D). Cette relation sera utilisée pour la propagation des annotations sémantiques entre les représentations et la mise en relation des données hétérogènes pour tendre vers l'objectif de fusion de données.

Même si, dans la suite du manuscrit, le sujet se concentre plus particulièrement sur l'utilisation d'images et de la photogrammétrie pour la génération du nuage de points, il sera montré que la méthode développée peut se transférer sur l'utilisation d'un relevé obtenu par laser scanner ou même par la combinaison des deux techniques. Par ailleurs, la méthode prend aussi en compte plusieurs techniques d'imagerie scientifique (infrarouge, ultraviolet, thermographie, etc).

#### IV.A. Approche proposée

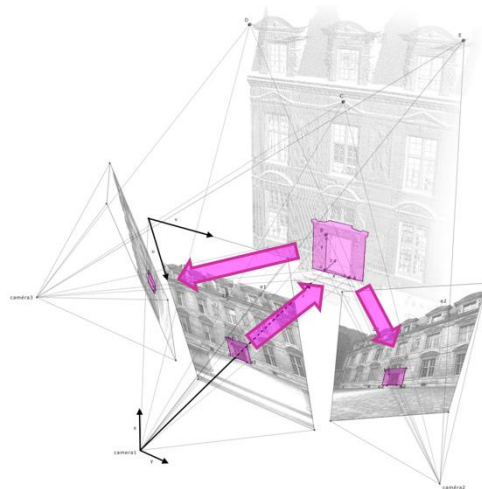
L'approche se compose de cinq étapes.

La première étape consiste en l'obtention d'une première représentation tridimensionnelle sous forme de nuage de points très dense à partir de l'acquisition d'un ensemble d'images de l'objet. Ce processus se base sur les derniers développements en termes de reconstruction par photogrammétrie (voir [Chapitre II](#)).

La deuxième étape se concentre sur la mise en relation des représentations bidimensionnelles (images de l'acquisition) et la représentation tridimensionnelle générée à la première étape. Cette étape s'appuie sur le référencement spatial des images autour du nuage de point initial (orientation) et des caractéristiques internes de l'appareil d'acquisition (calibration). Cette étape met ainsi en lien les images et la représentation 3D de l'objet et indique la position des coordonnées 3D dans les images.

La troisième étape propose un système de propagation d'annotation entre les images (et implicitement le modèle 3D). Les annotations (dessinées sur une ou plusieurs images de l'ensemble) sont automatiquement projetées sur les autres images pertinentes de l'ensemble. Le système consiste à extraire la région spatiale (sous forme de nuage de points) de l'annotation à partir de la sélection sur une image et de se servir de la relation projective entre

le nuage de point et les images pour retrouver la correspondance sur les autres images (Figure 56).



*Figure 56 : Les trois premières étapes de l'approche : un nuage point est réalisé à partir d'un ensemble de photographies ; les photographies sont spatialement référencées autour du nuage ; la relation projective entre 2D et 3D permet de transférer les annotations entre les images et le nuage de points*

La quatrième étape repose sur la caractérisation sémantique de la zone sélectionnée. Cette phase consiste à associer des tags sémantiques et des descriptions aux éléments sélectionnés sur les images. La description est alors associée à toutes ses représentations bidimensionnelles et sa représentation tridimensionnelle (extraite du nuage de point à partir de la sélection sur l'image).

Cette approche a permis la définition d'un ensemble d'outils pour l'aide à la description afin de faire ressortir des caractéristiques spécifiques utiles à la sélection et à l'analyse.

Le point fort de cette approche est de permettre un gain considérable en termes de nombre de procédures. En effet, là où les méthodes actuelles nécessitent aujourd'hui de nombreuses étapes afin d'obtenir un modèle maillé et segmenté, l'approche proposée offre une génération des données et d'extraction d'informations pertinentes beaucoup plus rapide et beaucoup plus flexible.

Le second avantage de cette approche est la facilité par laquelle l'annotation est conduite. L'annotation sémantique est conduite directement sur les images offrant un meilleur support de sélection qu'une représentation 3D, puis elle est automatiquement transférée sur le nuage de points et sur les autres images. La cohérence entre les images et la représentation tridimensionnelle de l'objet permet de segmenter les images et le modèle 3D à partir d'une seule sélection sur une image. Cependant, les possibilités de sélection ne se limitent pas à la sélection sur une seule image pour une annotation ce qui permet une flexibilité importante pour la sélection de régions et de sous-régions d'objet complexes.

Le dernier avantage de notre approche est la capacité à gérer de multiples calques d'annotation en fonction de différentes stratégies de description et à la facilité à modifier la structure de la description sémantique. Compte tenu de la relation projective, chaque changement de niveau dans la description correspondra à une mise à jour automatique de la segmentation des images et de la représentation 3D. Cette possibilité est particulièrement

importante dans le domaine de la documentation d'objets patrimoniaux où il est nécessaire de pouvoir établir différents niveaux de descriptions en fonction des exigences d'analyse (conservation, restauration, aménagement, valorisation, etc.). Dans ce cas, il est possible d'envisager des descriptions sémantiques multicouches. Cet aspect permet de servir de base pour des analyses comparatives entre les différentes couches sémantiques.

## IV.B. Problèmes à traiter

A partir de l'approche proposée, le travail effectué comprend plusieurs préoccupations.

### IV.B.1. Génération d'informations liées aux images (Chapitre V)

Notre approche repose sur l'établissement d'un lien entre les images et le nuage de points. L'espace 3D devient la référence pour établir la relation projective entre la représentation 3D et la représentation 2D. Chaque image 2D est associée à un modèle géométrique exprimant sa position, son orientation et ses paramètres de distorsion liée au capteur et à son point de vue dans le référentiel du modèle 3D.

Pour cela, nous avons besoin d'étudier les méthodes existantes d'orientation des images dans l'espace. Nous nous intéresserons ensuite à la représentation géométrique des points de vue des images par l'utilisation du modèle sténopé de l'appareil afin de pouvoir établir le lien entre les informations 2D et les informations 3D.

### IV.B.2. Ajout de nouvelles images (Chapitre V)

Dans le cas de l'utilisation de différents types de capteurs, le problème de l'ajout de nouvelles images au sein de l'ensemble d'images déjà traitées se pose. En effet, la génération du nuage de points ne peut se faire qu'à partir d'images de même type. Si l'on veut pouvoir utiliser des images provenant de différents capteurs, il faut pouvoir les ajouter à posteriori des traitements de génération du nuage de points.

Dans notre approche, il est possible de créer une relation entre les images et le nuage de points de l'objet en connaissant la position et l'orientation d'une image. On peut envisager qu'à partir du moment où cette image peut être orientée autour du nuage de points et si on connaît les paramètres de calibration du capteur, cette image peut être intégrée au sein du système. Le problème de l'estimation de l'orientation et de la calibration de cet autre capteur est donc essentiel.

La possibilité d'ajouter des images soulève aussi des problèmes liés au fait que l'acquisition de ces images n'a pas été réalisée avec les mêmes paramètres que les images déjà traitées. Ces images n'auront généralement pas les mêmes conditions d'éclairage voire même, ayant été prises à un temps différent, ne représentent pas l'objet dans le même état. Un autre problème se pose à propos du niveau de détails de l'image ajoutée. En effet, la relation projective sera plus difficile à établir si l'image présente une densité trop importante de pixels par rapport au nuage de points. L'approche cherche donc à établir des bases pour gérer ces différents problèmes.

#### IV.B.3. Transfert d'annotations entre les images (Chapitre VI)

Au vu de la quantité nécessaire de photographies pour l'analyse d'un objet, l'annotation de ces photographies ne peut pas se faire de façon manuelle. Comme vu dans le Chapitre III, les méthodes d'annotations les plus intéressantes dans notre domaine sont celles semi-automatiques. Si le lien entre les représentations 2D et 3D est établi, on peut en plus envisager de transférer les annotations entre les différentes représentations.

L'approche devra donc, à partir de la relation établie entre les images et la représentation 3D, permettre le transfert automatique d'annotations entre les images. Ainsi à partir d'une annotation sur une image, la correspondance devra être trouvée sur le nuage de points et sur les autres images lorsque la partie sélectionnée de l'objet apparaît. L'annotation devra donc pouvoir être visualisable de deux façons : des projections 2D sur les images et un ensemble de points 3D représentant l'annotation dans l'espace.

#### IV.B.4. Outils d'annotation (Chapitre VII)

Les méthodes de segmentation d'images ou de modèles 3D offrent des possibilités d'extraction d'informations rapides. Cependant les algorithmes actuels ne sont pas assez précis pour une utilisation totalement automatique dans le domaine de la conservation du patrimoine. Ils peuvent néanmoins offrir des possibilités d'analyses supplémentaires en faisant ressortir des caractéristiques complémentaires de l'objet.

L'approche cherchera donc à intégrer ces aspects de segmentation pour développer un système d'aide à la sélection et à l'analyse. Pour cela, nous nous intéressons aux méthodes de segmentation existantes et leur utilisation dans notre domaine.

#### IV.B.5. Outils d'extraction d'informations (Chapitre VII)

Pour entreprendre l'analyse d'un objet patrimonial, il est nécessaire de disposer d'outils d'analyse quantitative. Les analyses qualitatives de l'expert peuvent être supportées par ces aspects quantitatifs relatifs à l'objet. La quantification permet d'exprimer avec précision et de rendre vérifiables les aspects qualitatifs.

Le système permettant d'intégrer à la fois des aspects bidimensionnels et des aspects tridimensionnels, ces deux types d'informations peuvent être utilisés pour servir de support à l'extraction d'informations quantitatives. Notamment, l'aspect tridimensionnel peut permettre d'obtenir des informations métriques ou surfaciques sur les objets annotés. C'est pourquoi, afin de permettre aux experts de disposer d'éléments pertinents pour l'analyse, le système intègre un ensemble d'outils cherchant à faciliter le travail de l'expert.

### IV.C. Développements informatiques

Les apports méthodologiques de ce travail de thèse sont expérimentés et validés au sein de différents développements informatiques.

Pour créer ce système, nous utilisons un environnement de programmation qui permette de gérer les aspects suivants :

- Création et modification de formes géométriques dessinées sur les images



- Lecture et exploitation de données tridimensionnelles
- Lecture de fichiers XML
- Création de fonctions pour les étapes de propagation d'annotations
- Création et modification d'attributs rattachés à divers niveaux de la géométrie et des images
- Fonctions pour l'extraction d'informations et l'analyse d'images
- Interface utilisateur pour la gestion du processus d'annotation d'images
- Exportation et importation des informations vers une base de données

A partir de ces exigences, nous avons choisi MATLAB parmi les diverses solutions envisageables. En effet, MATLAB permet la gestion complète de ses aspects. Cependant, si la partie interface et la gestion de la base de données peuvent être réalisée sous MATLAB, il est aussi possible de ne garder que des briques de calculs développées sous MATLAB pour la partie de traitement et d'envisager une gestion de ces aspects par une interface développées en C++ (version locale) ou HTML5 (version web).

#### IV.C.1. MATLAB

MATLAB est un langage de programmation de haut niveau utilisé à des fins de calculs numériques. Il permet de manipuler des matrices, d'afficher des courbes, des mettre en œuvre des algorithmes, de créer des interfaces utilisateur, et peut s'interfacer avec d'autres langage comme C, C++, Java et Fortran. Il peut s'utiliser seul ou bien avec des *toolbox* intégrant divers algorithmes connus.

Ce langage permet d'avoir accès à la création d'interface utilisateur et à toutes les fonctions du logiciel (sous réserve de posséder la *toolbox* associée à la fonction). Matlab repose sur une architecture basée sur des lignes de commandes. Des séquences de commandes peuvent être sauvegardées dans un fichier texte sous la forme de script ou intégrées dans une fonction. Les commandes pour l'ensemble des fonctionnalités peuvent être accessibles à travers le langage de script liées à l'interface utilisateur (menus, outils, boîtes de dialogues, etc.). Chaque commande peut avoir plusieurs arguments et plusieurs sorties qui établissent la façon dont la commande est exécutée. Les codes développés sous forme de scripts ou de fonctions peuvent facilement être diffusés au sein de la communauté et ainsi il est possible de récupérer des algorithmes déjà implémentés par d'autres utilisateurs.

#### IV.C.2. C++ et QT

C++ est un langage standardisé de programmation de bas niveau. Il peut être considéré comme du C agrémenté de fonctionnalités supplémentaires mais il ne se limite pas à cette définition. L'utilisation de C++ comme langage de programmation offre de nombreux avantages :

- **Popularité** : C++ fait partie des langages de programmation les plus répandus, notamment car il est libre d'utilisation et dispose de nombreuses documentation faciles d'accès.
- **Rapidité** : c'est un langage qui offre des performances importantes grâce au contrôle possible sur le fonctionnement de l'ordinateur et de la gestion de la mémoire.



- **Portabilité** : un même code source peut théoriquement être compilé sous différentes plateformes sans avoir besoin de réécrire le programme.
- **Bibliothèques** : la bibliothèque standard de C++ est en grande partie un sur-ensemble des fonctions disponibles dans la bibliothèque standard de C. Il existe, de plus, de nombreuses bibliothèques pouvant être utilisées dans C++.
- **Multi-paradigmes** : Différents paradigmes de programmation sont intégrés dans C++ comme la programmation procédurale ou la programmation orientée objet.

QT est une interface de programmation orientée objet et développé en C++ qui offre des composants d'interface graphique (widgets), d'affichage 3D, d'accès aux données, de dessin, de gestion d'exécution, etc. et conçue pour être utilisée en C++. Cependant elle offre aujourd'hui des possibilités d'utilisation avec d'autres langages. Il s'agit d'un ensemble de bibliothèques donnant de nombreux outils pour développer des programmes de façon plus efficace.

#### IV.C.3. HTML5 et WebGL

HTML5 est la dernière révision majeure d'HTML. C'est le langage de base pour créer des sites internet et sert à structurer le contenu d'une page web. Il est inspiré du XML et repose sur le principe de balises imbriquées servant à guider le navigateur sur l'affichage de la page web.

HTML5 dispose de deux avantages majeurs :

- **Simplicité** : c'est un langage qui s'apprend très vite et qui ne contient qu'une centaine de balises.
- **Compatibilité** : les différents éditeurs de navigateurs intègrent HTML5 ce qui permet d'afficher un page web de la même façon quel que soit le navigateur (Firefox, Opéra, Internet Explorer, etc.), l'OS (Windows, Mac, Linux, etc.) et le support (Pc, Smartphone, Tablette, etc.).

Cependant, il ne permet de gérer que du contenu statique et non du contenu dynamique sur la page web. Il faudra pour cela l'associer à d'autres types de langages tels que CSS (pour la mise en page), JAVASCRIPT (pour agir sur le comportement de la page), PHP (pour faire le lien avec une base de données, gérer les variables, etc.) ou SQL (pour éditer la base de données).

WEBGL est une spécification d'interface de 3D dynamique pour les pages et applications HTML5. Elle permet d'afficher et de gérer dynamiquement des éléments graphiques complexes en 3D dans la fenêtre du navigateur web d'un client. Il est actuellement intégré dans la plupart des grands navigateurs modernes.

## V. Lier l'information 2D à l'information 3D : spatialisation des informations

Le chapitre précédent propose une approche pour la sémantisation d'un ensemble d'images orientées dans l'espace. Cette approche nécessite donc mettre en relation spatiale des sources iconographiques (sous formes de photographies) par rapport à une représentation 3D de l'objet en interprétant les données sous forme géométrique (Figure 57). Ce chapitre s'intéresse donc à la mise en relation entre les représentations 2D et les représentations 3D. Comme nous le verrons dans le chapitre suivant, cette relation spatiale servira à transférer des attributs sémantiques entre les images et le modèle 3D.

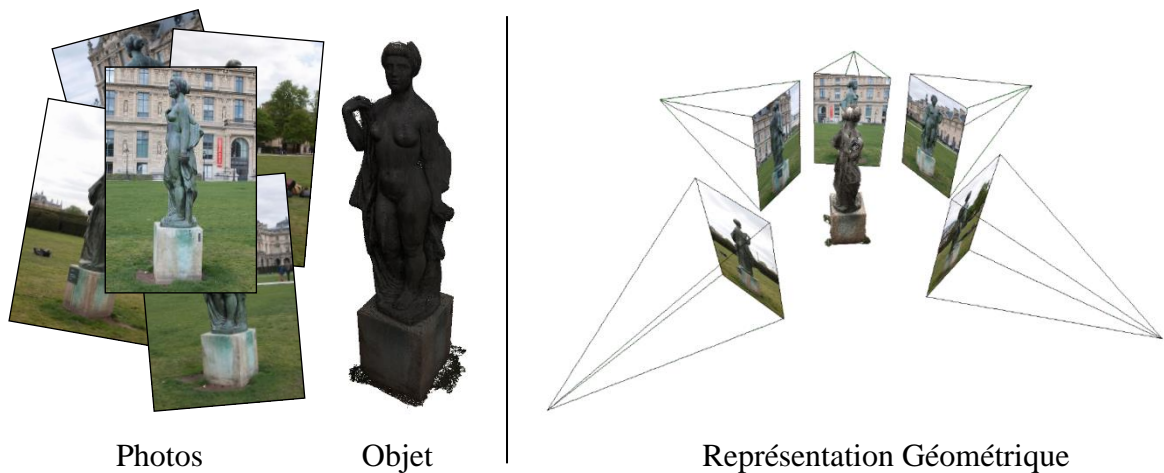


Figure 57 : Différence entre données réelles et représentation géométrique des données

Ce chapitre se divise en trois parties. Dans la première partie, nous aborderons les méthodes actuelles de référencement spatial d'images 2D sur des ensembles de coordonnées 3D. Cette opération permet de définir un modèle géométrique d'orientation et de calibration de l'appareil photographique. Dans la deuxième partie, nous nous intéresserons à la mise en relation entre les pixels des images et les coordonnées 3D à partir du référencement spatial. Enfin la dernière partie s'intéressera à la gestion des données 3D liées aux images 2D.

### V.A. Méthodes de référencement spatial

Afin de mettre en relation la morphologie d'un objet avec des représentations 2D du même objet, notre travail aborde le problème de l'organisation spatiale des photographies autour d'une représentation tridimensionnelle de l'objet. La notion de « resection spatiale », définit par [Moffitt, F.H. et al, 1980] comme un « *procédé dans lequel la position et l'orientation spatiales de la photographie sont déterminées en se basant sur des mesures photogrammétriques et sur des points de contrôles apparaissant dans la photographie* », peut constituer un point de départ pour le développement d'un système de référencement spatial de sources bidimensionnelles.

Le référencement spatial crée une relation spatiale entre l'image 2D et la représentation 3D. Il s'agit de trouver les paramètres du point de vue de la source afin d'observer le modèle 3D à

partir de ce point de vue. Ce point de vue peut être estimé de façon fine ou de façon approximative.

Les méthodes de référencement spatial permettent donc d'obtenir des informations géométriques relatives au point de vue de l'image (position, orientation, distorsion, etc.) avec différents niveaux de précision. Ces méthodes de référencement spatial peuvent par la suite être utilisées pour interroger des bases de données d'images en fonction de critères spatiaux.

Nous allons présenter un ensemble de méthodes de référencement spatial en les classant sous trois catégories : référencement image-coordonnées, référencement image-modèle 3D et référencement image-image.

Pour chacune de ces catégories, les méthodes peuvent être manuelles, semi-automatiques ou automatiques. Les méthodes manuelles de référencement spatial sont les solutions les plus simples et les plus faciles à utiliser. Les méthodes semi-automatiques de référencement spatial d'images se basent sur des solutions géométriques appelées calibration et orientation des caméras. Ces solutions utilisent des paramètres géométriques et optiques de l'appareil sous formes de valeurs numériques. Cela crée au sein de la représentation 3D une caméra virtuelle représentant le point de vue de l'image par rapport à un référentiel 3D. Ces méthodes d'orientation et de calibration prennent en compte des paramètres intrinsèques (ou internes) (distance focale, distorsion, etc.) et des paramètres extrinsèques (ou externes) (translation et orientation dans l'espace) de l'appareil au moment de la prise de vue. Enfin les méthodes automatiques de référencement spatial peuvent être de deux types : soit par l'utilisation d'un outil pour enregistrer l'information spatiale de l'image au moment de sa prise de vue, soit par l'utilisation d'algorithmes pour retrouver le point de vue de l'image sans informations spatiale.

#### V.A.1. Référencement image-coordonnées

Les méthodes de référencement image-coordonnées cherchent à placer dans l'espace les images par rapport à un repère défini. Aujourd'hui de nombreux appareils photographiques intègrent des systèmes GPS (*Global Positioning System*) ou peuvent être associés à des modules GPS additionnels. Les données de positions sont ainsi directement stockées dans les métadonnées de l'image. Cependant, pour notre méthodologie, l'utilisation seule des données GPS n'est pas assez précise, même si elle peut aider à donner une première estimation de l'orientation des images.

[Hu Z. et al. 2006] ont développé une solution d'estimation de la position et de l'orientation d'une caméra basée sur un système hybride mêlant un GPS, un gyroscope et un algorithme de vision par ordinateur pour superposer des indicateurs 3D de navigation et de trafic routier dans un environnement de réalité augmentée.

Le laboratoire MAP-GAMSAU a développé le module OMEGA [OMEGA] (Figure 58) permettant d'enregistrer de façon automatique, la position et l'orientation des photographies au moment de la prise de vue. Ce système utilise une station totale (tachéomètre) équipée d'un télémètre pour la prise de mesure. La station est aussi équipée d'une caméra interne dont la position et l'orientation relative sont connues. Le système OMEGA est équipé des prismes et de boules de repères de la station et sont utilisées comme références pour déterminer l'orientation de l'appareil photographique au travers d'un algorithme d'analyse d'image.

Comme les positions relatives entre le prisme, les boules et l'appareil photographique sont connues la position et l'orientation du module peuvent être déterminée à partir de la détection de la position des boules enregistrée par la caméra intégrée à la station totale au moment de la prise de vue.



Figure 58 : Le système OMEGA [OMEGA] pour le référencement automatique d'images

Le système APIS (*Architectural Photogrammetry Information System*) [Waldhäusl P. et al, 1994] utilise une base de données qui répertorie les bâtiments documentés par des acquisitions photogrammétriques. Le système permet d'enregistrer les informations sur les édifices et d'enregistrer les positions des photographies lors de la prise de vue.

#### V.A.2. Référencement image-modèle 3D

Dans le domaine de la documentation patrimoniale, plusieurs recherches visent à créer une relation entre les sources 2D et une représentation morphologique d'un objet en indiquant de façon approximative la position de prise de vue. Les SIG (Systèmes d'Informations Géographiques), par exemple, permettent d'indiquer sur un plan la position et parfois la direction de vue d'une photographie.

[Kadobayashi R. et al., 2003] propose un système qui permet d'effectuer des recherches d'images sur la base d'informations stockées dans les légendes des photographies tout en permettant en parallèle de manipuler des nuages de points ou des modèles 3D. Des cibles de références sont utilisées pour définir des mesures topographiques. Chaque image a son propre système de coordonnées basé sur la position du scanner au moment de l'acquisition. L'interface du système comporte une vue planimétrique, une élévation et une vue 3D (Figure 59). L'utilisateur peut rechercher des images dans la base de données par positions relatives ou par mots-clés.



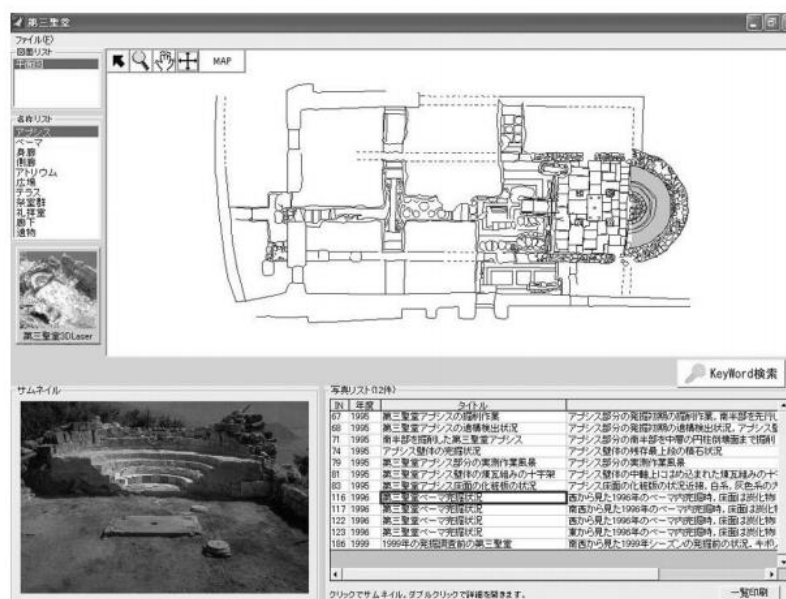


Figure 59 : L'interface du système de [Kadobayashi R. et al, 2003]

Dans son système, [Busayarat C., 2010] propose un outil de référencement spatial par superposition manuelle des sources directement sur le modèle 3D en fonction de la projection considérée de la source iconographique. Après avoir été chargée dans une base de données, l'image à orienter est placée en avant de la scène 3D en transparence. L'interactivité du système permet ensuite à l'utilisateur de manipuler un plan 2D (sur lequel l'image est plaquée) en le contrôlant dans l'espace 3D en liaison avec la caméra de navigation (Figure 60). L'utilisateur peut donc retrouver le point de vue de l'image à l'aide de contrôle de translation, de rotation et de l'angle de champ (distance focale) afin d'adapter la vision de la scène 3D à la perspective de l'image.



Figure 60 : Interface pour le référencement manuel de [Busayarat C., 2010]

La méthode DLT (*Direct Linear Transformation*) est une méthode semi-automatique apportée par [AbdelAziz et al, 1971]. Elle emploie un ensemble de points de commande dont les coordonnées d'espace-plan d'objet sont déjà connues. Le problème de cette méthode est que les paramètres de calibrage ne sont pas mutuellement indépendants.

Dans le domaine de la photogrammétrie ou de la vision par ordinateur, de nombreuses méthodes semi-automatiques sont utilisées pour orienter une image sur un ensemble de coordonnées connues ([Faugeras, 1992], [Fitzgibbon et al, 1998], [Heikkila et al, 1997], [Kumar R. et al, 1994]). La plus utilisée est celle proposée par [Tsai, 1987]. Elle utilise des correspondances de points sélectionnés respectivement dans l'espace 3D et dans l'espace 2D (image). Cette méthode se base sur deux phases : tout d'abord on calcule la position et l'orientation de l'image, puis on calcule ses paramètres internes.

[Johansson B. et al, 2002] décrit un système automatique d'estimation de la position et de l'orientation d'une image simple représentant une scène urbaine. Chaque édifice de l'image doit avoir un modèle. Une homographie est estimée de façon automatique pour chaque édifice. Le modèle ayant le meilleur score de superposition définit la position et l'orientation du point de vue de la caméra (Figure 61). Ce système suppose que les édifices sont composés principalement par des plans et que des plans ont des arêtes verticales et horizontales bien visibles.

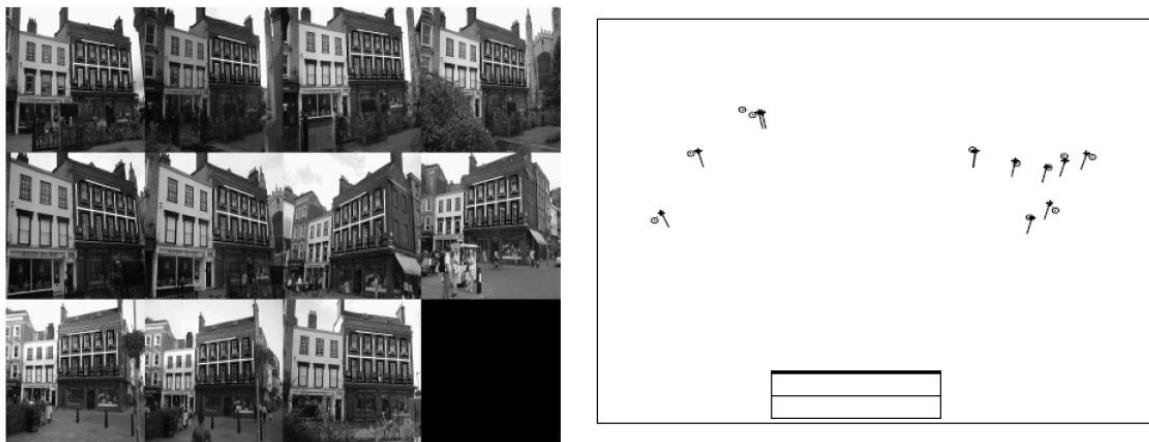


Figure 61 : A gauche, images de différents points de vue, à droite, le résultat de l'estimation des positions [Johansson B., 2002]

[Palma G., et al, 2010] ont défini un algorithme permettant le calcul de la position d'une image autour d'un modèle 3D par information mutuelle (*Mutual Information*). Le principe repose sur la génération de cartes de gradient pour l'image et pour le modèle 3D qui sont comparées afin de pouvoir déterminer les paramètres de la caméra (Figure 62). L'utilisation de ces cartes permet de ne pas avoir à prendre en compte les informations de textures ou de conditions de lumières différentes.

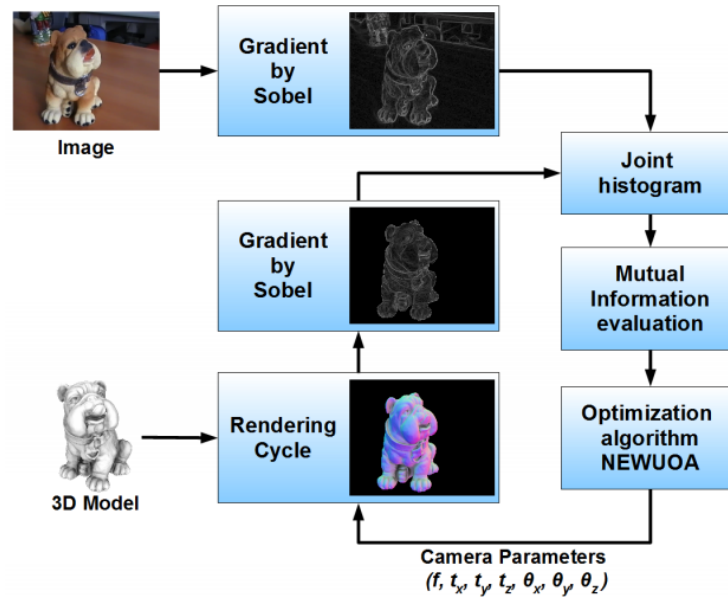


Figure 62 : Principe de l'algorithme défini par [Palma G. et al, 2010]

### V.A.3. Référencement image-image

Au niveau des méthodes algorithmiques basées sur l'analyse d'image, [Lee S.C. et al, 2002] définissent une méthode pour estimer les positions des caméras à partir d'une extraction automatique de points de fuite. Les points de fuite sont détectés à partir d'un ensemble de lignes parallèles dans l'image. Grâce à la connaissance de ces points de fuite, les paramètres des caméras peuvent être estimés.

Le logiciel Photosynth [PHOTOSYNTH, 2008], déjà présenté au Chapitre III, repose sur l'orientation de photographies numériques dans l'espace et génère un nuage de points correspondants aux coordonnées extraites à partir de points homologues. Le système permet, d'autre part, la mise à jour des collections d'images à l'aide d'une procédure demandant tout d'abord à l'utilisateur de superposer de façon manuelle la nouvelle image sur les images déjà orientées dans l'espace. A partir de cette position approximative, un algorithme d'extraction de points homologues crée des correspondances entre la nouvelle image et une ou plusieurs images orientées. Avec ces correspondances la nouvelle image peut être orientée à l'aide de calculs d'ajustement de faisceaux.

La suite logiciel MICMAC [Pierrot-Desseiligny M. et al, 2011a] [Pierrot-Desseiligny M. et al., 2011b], développé par l'IGN, permet aussi l'orientation d'un ensemble d'images et le génération d'un nuage de points très dense à partir de ces images. La mise à jour de la collection d'images orientées est aussi possible. Le principe général est la même que pour Photosynth mais il se sert des points homologues pour estimer les images les plus proches et ne demande pas d'intervention manuelle de la part de l'utilisateur.

### V.A.4. Analyse des méthodes actuelles

A partir de l'analyse des méthodes de référencement spatial présentées, l'étude de leur adaptabilité à notre méthodologie doit être engagée. Celle-ci impose de pouvoir avoir une orientation et une position précise des images autour de la représentation 3D.



Même si les méthodes de référencement manuel des photographies sont les plus simples à utiliser et à manipuler pour les utilisateurs d'un système, cela génère des erreurs dues à la détermination humaine, réduisant la précision du système. De plus, dans le cas de photographies, l'aspect de la distorsion due à la géométrie du système d'acquisition est difficile à estimer. Ce sont des méthodes qui demandent beaucoup de temps de travail pour arriver à un niveau de précision suffisant pour qu'il soit utilisable dans notre méthodologie.

Malgré la flexibilité et la facilité de l'utilisation de données GPS, ces méthodes ne peuvent pas être prises en compte pour nos exigences à cause de leur faible précision. En effet, notre méthodologie demande une précision importante dans la cohérence géométrique entre le modèle 3D et la position des images. De plus, les systèmes GPS ne donnent généralement que la position dans l'espace de l'image et non son orientation autour de cette position. Cependant, ces données GPS peuvent servir pour une estimation rapide de la position et alléger d'autres calculs plus fins d'orientation et de position.

Les dispositifs d'enregistrement tels que le système OMEGA permettent un référencement en termes de positionnement et d'orientation compatibles avec les niveaux de cohérence nécessaires à une orientation en « solution approchée » des images photographiques sur le modèle 3D. Cependant, ils limitent l'utilisation à des images prises dans des conditions de prises de vue spécifiques. Dans ce sens, ces méthodes ne seront pas une solution effective pour toute source photographique. En effet, il ne serait pas possible d'utiliser des photographies ayant simplement été prises en dehors de la campagne de relevé à l'aide du système.

Les méthodes par resection spatiale s'avèrent les plus adaptées. En effet, ces solutions déterminent la superposition image/modèle en estimant à la fois les paramètres intrinsèques et les paramètres extrinsèques avec des niveaux d'approximation différents. Ces solutions restent aussi adaptées au référencement spatial de types de représentation bidimensionnelle autre que photographique (dessins, peinture, etc.). Cependant l'orientation de ces représentations spécifiques demande tout de même une forte approximation. Les solutions logiciels offrent de plus des implémentations de ses méthodes qui ne demandent pas d'intervention de l'utilisateur. Les méthodes les plus avancées établissent de façon automatique la mise en relation entre les points 3D et les images. De plus, pour les méthodes permettant la génération du nuage de points à partir des images, une représentation est directement extraite des images à partir des paramètres internes et externes préalablement estimés. Ainsi, le modèle 3D sera toujours à un haut niveau de cohérence géométrique avec les images photographiques.

#### V.A.5. Solution retenue

Afin d'implémenter la mise en relation des données 3D et des données 2D (images) dans notre système, plusieurs aspects ont été pris en considération.

Le premier aspect concerne la grande hétérogénéité d'images disponibles pour un objet. Plusieurs supports, plusieurs formats, plusieurs résolutions et plusieurs états de conservation peuvent être représentés. Dans ce sens, la première contrainte est d'envisager une solution qui ne limitera pas l'utilisateur au niveau de ces différences entre images.

Le second aspect concerne le fait que, comme vu dans l'approche au [Chapitre IV](#), le système nécessite d'avoir une représentation 3D de l'objet conforme avec la réalité autour de laquelle les images doivent pouvoir être référencées spatialement de façon précise. Cependant, il n'est pas nécessaire d'avoir une représentation maillée de l'objet, une représentation en simples coordonnées de points suffit. C'est pourquoi une solution qui permet la génération du nuage de points en plus de l'orientation et la calibration des images semble la plus intéressante à considérer.

A partir de ces constats et de l'étude des méthodes existantes, nous avons choisi d'utiliser la suite de logiciels open source MICMAC [[Pierrot-Desseiligny M. et al, 2011a](#)] [[Pierrot-Desseiligny M. et al, 2011b](#)] développé par l'IGN. Cette boîte à outil offre plusieurs avantages.

Les modules de MICMAC permettent de générer des nuages de points à partir d'un ensemble d'images acquises selon des protocoles précis. Le nuage de points généré a donc un niveau de cohérence très important avec les images utilisées et possède un niveau de précision par rapport à la réalité presque aussi élevé que les relevés par scanner laser. Les protocoles à suivre pour l'acquisition limitent l'acquisition à des photographes avertis mais ont été détaillés de façon précise afin de pouvoir être applicables par des non-experts de la photographie [[Martin-Beaumont N. et al, 2012](#)].

Le procédé de traitements des images, détaillé dans [[Pierrot-Desseiligny M. et al, 2011a](#)], consiste en trois phases :

- **Calibration et orientation des images :** La méthode est basée sur les procédés open-source PASTIS et APERO [[Pierrot-Desseiligny M. et al, 2011b](#)]. Ils sont constitués de modules pour l'extraction de points communs (SIFT [[Lowe D.G., 1999](#)] adapté par [[Vedaldi A., 2006](#)]) (Figure 63), l'estimation des paramètres internes et externes de la géométrie de l'appareil, un calcul de la solution initiale et un ajustement de faisceaux pour les orientations relatives et absolues. Ces méthodes permettent ainsi le calcul de la position et de l'orientation dans l'espace des images (Figure 64).

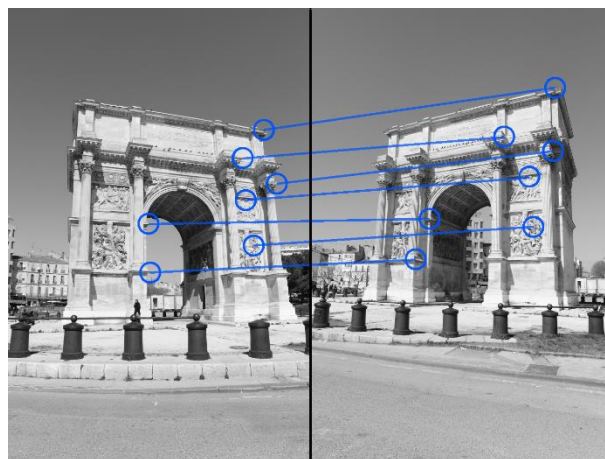
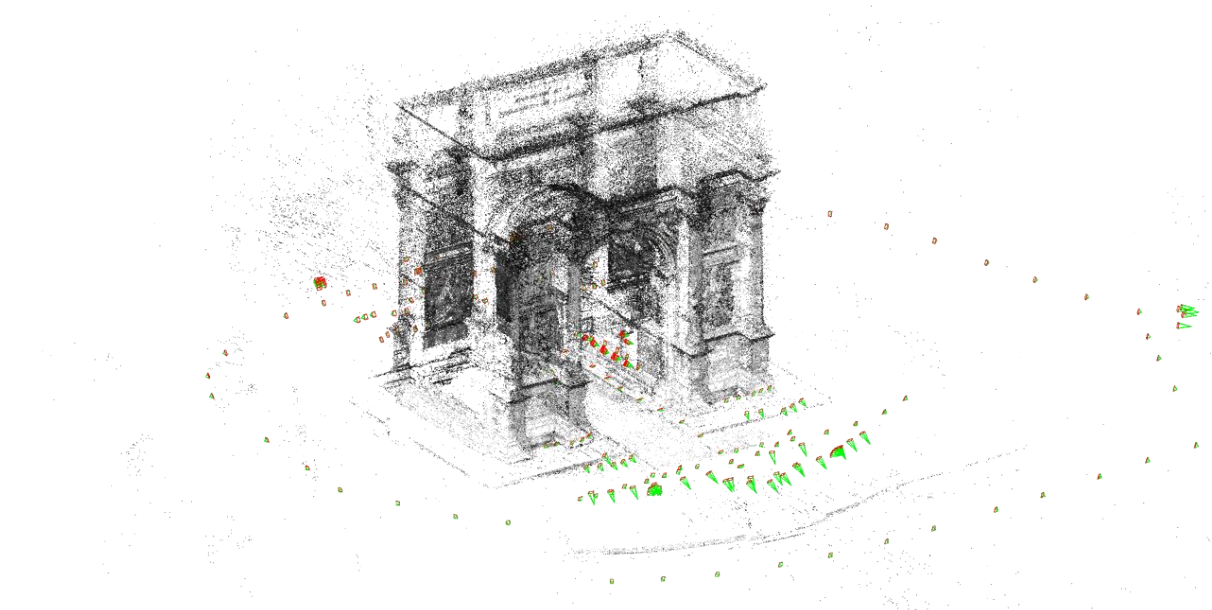
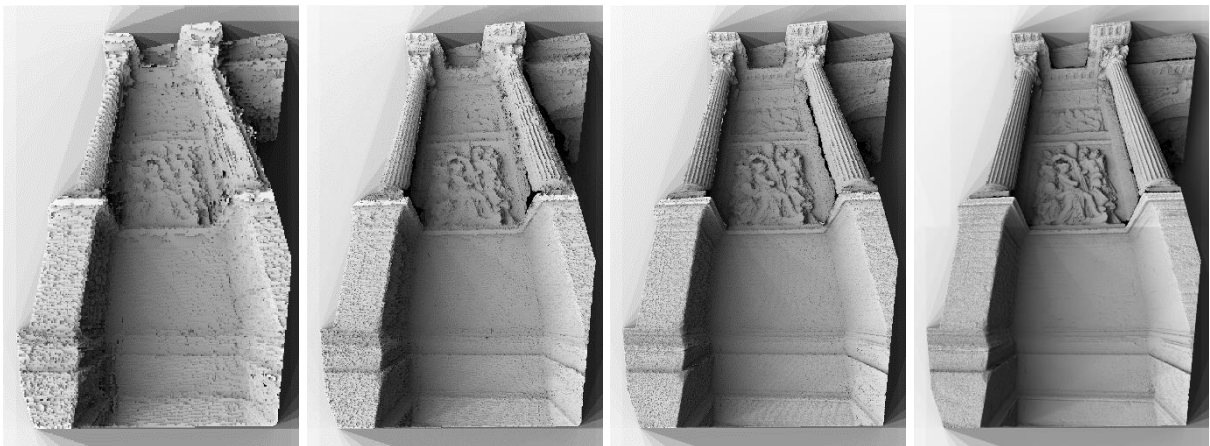


Figure 63 : Recherche des points SIFT entre deux images



*Figure 64 : Orientation des images autour des points SIFT*

- **Corrélation dense multi-vues :** A partir de la position des images, une méthode de reconstruction par appariement dense [Pierrot-Desseiligny M. et al, 2006] génère des cartes de profondeur sur les images. L'appariement à une approche pyramidale (Figure 65). Lors de cette étape, la profondeur des pixels peut être calculée pour des groupes de pixels et seules les parties les plus pertinentes pour la corrélation de chacune des images sont considérées.



*Figure 65 : Approche pyramidale : Evolution des cartes de profondeur durant l'appariement dense*

- **Génération du nuage de points :** A partir des résultats obtenus aux deux premières étapes, les cartes de profondeurs sont converties en points 3D métriques. Cette conversion est basée sur la projection dans l'espace des pixels de l'image en tenant compte de l'orientation et de la position. Ainsi chaque image sert à générer un morceau du nuage de points. C'est la superposition de tous ces morceaux de nuage de points qui permet de constituer le nuage de point complet de l'objet (Figure 66).



*Figure 66 : Nuage de points final*

D'autre part, le procédé permet par la suite de générer des orthophotographies. Au vu de la densité très importante du nuage de points produits, la génération d'orthophotographies est basée sur la projection orthogonale des résultats sur un plan. La résolution finale est calculée en fonction de la densité du nuage de points.

Pour générer le nuage de points, MICMAC nécessite l'estimation des paramètres intrinsèques et extrinsèques. Ces paramètres peuvent facilement être récupérés pour des utilisations ultérieures, ce qui n'est pas toujours possible sur d'autres solutions même automatiques. Les paramètres de calibration de l'appareil, donnant en particulier la distorsion, disposent de différents modèles d'approximation. De plus, ces paramètres étant utilisés pour la génération du nuage de points, ils offrent un haut niveau de cohérence avec le nuage de points. D'autre part, l'estimation de ces paramètres peut se faire par groupe de photographies permettant de calculer ces paramètres de façon plus précise si ceux-ci diffèrent beaucoup d'un groupe à l'autre (par exemple lors de l'utilisation de différents capteurs ou de différents objectifs).

L'aspect intéressant de cette solution est donc de pouvoir disposer des paramètres internes et externes de photographies liés à un nuage de points généré à partir de ces paramètres. Il est aussi possible d'envisager d'orienter des images provenant de différents capteurs au sein du même référentiel.

D'autre part, cette solution offre la possibilité d'orienter des images supplémentaires à posteriori des traitements et de la génération du nuage de points. Les nouvelles images sont donc simplement orientées autour du nuage de points déjà généré et les paramètres de calibration et d'orientation sont donc estimés dans le même référentiel.

La génération du nuage de points et d'orthophotographies passe par plusieurs étapes appelant différents modules de MICMAC (Figure 67) :



- TAPIOCA : Module pour le calcul des points de liaisons intégré au module PASTIS. Il s'agit d'une implémentation de l'algorithme SIFT intégrant une approche multi-échelle.
- TAPAS : Module tout automatique pour le calcul des orientations interne et externe des images intégré au module APERO utilisant les résultats générés par TAPIOCA.
- C3DC : Module de génération du nuage de points à partir des résultats de TAPIOCA et TAPAS par corrélation dense et puis conversion de cartes de profondeur.
- PIMS2MNT : Module de conversion des données en modèle numérique de terrain et définition du plan moyen.
- TAWNY : Module de génération d'orthophotographies à partir des résultats de PIMS2MNT.

```
Tapioca MulScale ".*JPG" 300 1500
Tapas FraserBasic ".*JPG" Out=Calib
Tapas AutoCal ".*JPG" In=Calib Out=All
C3DC MicMac ".*JPG" All ZoomF=4 Out=C3DC_Zoom4.ply
```

*Figure 67 : Exemple de lignes de commandes pour la génération d'un nuage de points à partir d'images au format JPG*

Dans les paragraphes suivants, nous présentons la mise en relation entre les pixels des images et la représentation 3D à partir des résultats obtenus à la fin du traitement par les différents modules de MICMAC.

## V.B. Relation entre pixels et coordonnées 3D

A partir des résultats fournis à la suite de TAPIOCA, TAPAS et C3DC (nuage de points et paramètres intrinsèques et extrinsèques de la caméra), la relation entre les images et le nuage de points peut être établie. Ainsi l'étape suivante de l'approche développée consiste à établir ce lien par une transformation géométrique dépendant du type de projection.

### V.B.1. Données d'entrées

Afin de pouvoir lier les pixels à des informations 3D, il faut d'abord comprendre les paramètres liés à la photographie.

Les paramètres extrinsèques ou externes représentent la position et l'orientation de l'appareil photographique au moment de la prise de vue. La position est définie par une translation dans l'espace et l'orientation par une rotation d'un plan dans l'espace. Ainsi cela représentent six composantes au total : trois pour la position (coordonnées suivant les trois axes: X, Y, Z) et trois pour l'orientation (angles autour des trois axes :  $\Theta$ ,  $\psi$ ,  $\omega$ ).

Les paramètres intrinsèques ou internes expriment les propriétés projectives de l'appareil. Il s'agit de la distance focale (distance entre un des plans principaux et le foyer), des coordonnées du point principal (centre projeté sur l'image) et de la distorsion (erreurs dues à l'imperfection de la lentille).

Le module TAPAS donne tous ces paramètres au sein de fichiers XML : un fichier pour l'orientation et un fichier pour la calibration.

Chaque image de l'ensemble possède un fichier d'orientation (Figure 68). Ces fichiers contiennent donc les informations de position et d'orientation de l'image dans l'espace. Le paramètre <Centre> représente la position dans l'espace du centre optique. Le paramètre <CodageMatr> représente l'orientation de l'image dans l'espace sous forme de matrice 3X3 comportant les trois vecteurs orientant l'image. Les deux premières colonnes de cette matrice définissent les vecteurs définissant le plan de l'image, et la dernière colonne représente le vecteur définissant la direction de la caméra (Figure 69).

```
<Centre>0.7149167456980563 -0.047206268472707279 -0.17731977256958339</Centre>
<IncCentre>1 1 1</IncCentre>
<ParamRotation>
- <CodageMatr>
  <L1>0.99695473105578203 -0.0092295676102741222 -0.077434354825366827</L1>
  <L2>-0.010665217494657396 -0.99977844168547469 -0.018147194735154613</L2>
  <L3>-0.077249707839480444 0.018917785882373982 -0.99683228279185643</L3>
</CodageMatr>
</ParamRotation>
```

Figure 68 : Extrait du fichier XML d'orientation d'une image donné par TAPAS

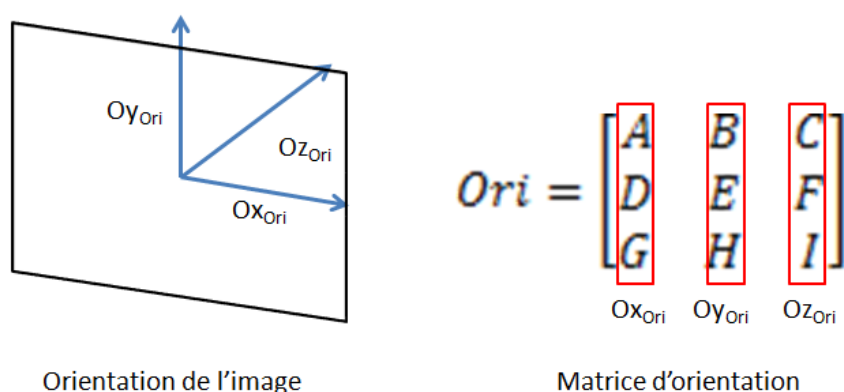


Figure 69 : Interprétation géométrique de la matrice d'orientation

Un fichier de calibration (Figure 70) peut être associé à plusieurs images à partir du moment où les images ont été prises dans les mêmes configurations, c'est-à-dire à même capteurs, à même objectif et même réglages. Le paramètre <PP> représente le point principal exprimé en position pixel dans l'image. Ce point ne correspond généralement pas exactement au centre de l'image. Le paramètre <F> donne la focale exprimé en quantité de pixels. La distorsion est définie par les paramètres sous le paramètre <CalibDistorsion>. La distorsion peut être définie par différents modèles plus ou moins précis : modèle radial (*Radial Model*), modèle standard photogrammétrique (*Photogrammetric Standard Model*), modèles unifiés polynomial (*Unified Polynomial Models*) ou modèle fisheye (*FishEye Model*).

```

<CalibrationInternConique>
  <KnownConv>eConvApero_DistM2C</KnownConv>
  <PP>2053.7923518371854 3045.0901875590912</PP>
  <F>7677.7264721636502</F>
  <SzIm>4000 6000</SzIm>
  - <CalibDistortion>
    - <ModPhgrStd>
      - <RadialePart>
        <CDist>2053.7923518371854 3045.0901875590912</CDist>
        <CoeffDist>-2.4138490009699822e-009</CoeffDist>
        <CoeffDist>1.0620382256110552e-016</CoeffDist>
        <CoeffDist>-3.1889734511483231e-024</CoeffDist>
        <CoeffDist>1.361302257971862e-031</CoeffDist>
        <PPaEqPPs>true</PPaEqPPs>
      </RadialePart>
      <P1>7.6844034811046237e-008</P1>
      <P2>-8.9726356704622132e-008</P2>
      <b1>-0.0010014988533445376</b1>
      <b2>0.00068559956909833536</b2>
    </ModPhgrStd>
  </CalibDistortion>
</CalibrationInternConique>
    
```

Figure 70 : Extrait du fichier XML de calibration d'une image donné par TAPAS

La prise en compte de la distorsion est très importante pour pouvoir faire le lien entre 2D et 3D puisqu'elle peut, suivant l'objectif utilisé, très fortement déformer l'image (Figure 71). A cause de la distorsion, une simple projection de points sur un plan ne donnera pas de résultats précis. La distorsion est généralement plus visible sur des lignes droites et sur les bords de l'image.

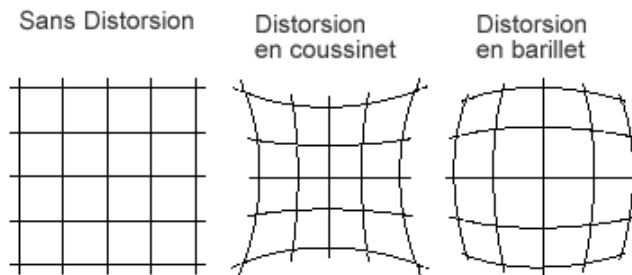


Figure 71 : Exemples de distorsion

### V.B.2. Représentation géométrique d'une image photographique

Le modèle sténopé (*pinhole*) modélise une caméra par une projection perspective. C'est un modèle qui transforme un point 3D de l'espace  $P_m$  en un point de l'image  $P_i$  par une suite de transformations élémentaires ((1), (2), (3) Figure 72). Cependant ce modèle n'intègre pas la transformation liée à la distorsion qui doit aussi être prise en compte dans notre cas ((4) Figure 72) :

- Transformation (1) entre le repère monde ( $R_m$ ) et le repère caméra ( $R_c$ )
- Transformation (2) entre le repère caméra ( $R_c$ ) et le repère capteur (plan rétinien) ( $R_r$ )
- Transformation (3) entre le repère capteur ( $R_r$ ) et le repère image ( $R_i$ )
- Transformation (4) au sein du repère image ( $R_i$ ) pour tenir compte de la distorsion

$$P_m (X_m \ Y_m \ Z_m) \xrightarrow{T_1} P_c (X_c \ Y_c \ Z_c) \xrightarrow{T_2} (x_r \ y_r) \xrightarrow{T_3} (i_i \ j_i) \xrightarrow{T_4} P_i (i_d \ j_d)$$



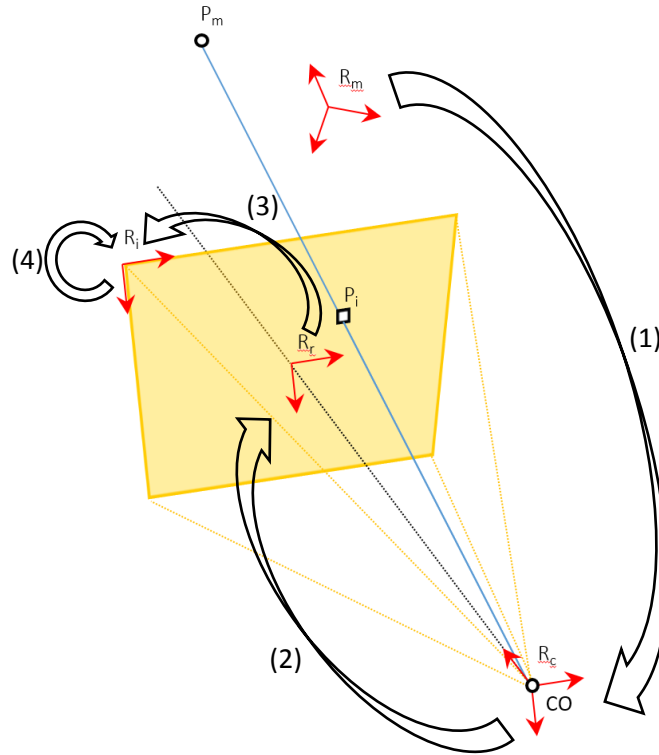


Figure 72 : Les différentes transformations du modèle sténopé

#### V.B.2.a. Transformation entre le repère monde et le repère caméra

La transformation du repère monde au repère caméra utilise les paramètres externes. Cette transformation peut se décomposer en une rotation [R] donnée par la matrice d'orientation de l'image et une translation [T] donnée par les coordonnées du centre optique CO. La matrice d'orientation exprime les vecteurs du repère caméra dans le repère monde.

D'un point de vue formel, on a :

$$P_m = Ori * P_c + CO = \begin{pmatrix} L1_1 & L1_2 & L1_3 \\ L2_1 & L2_2 & L2_3 \\ L3_1 & L3_2 & L3_3 \end{pmatrix} * P_c + \begin{pmatrix} CO_x \\ CO_y \\ CO_z \end{pmatrix} \{1\}$$

Ou, de façon équivalente (Ori étant orthogonale  $Ori^T = Ori^{-1}$ ), on a :

$$P_c = \begin{pmatrix} L1_1 & L2_1 & L3_1 \\ L1_2 & L2_2 & L3_2 \\ L1_3 & L2_3 & L3_3 \end{pmatrix} * \left( P_m - \begin{pmatrix} CO_x \\ CO_y \\ CO_z \end{pmatrix} \right) \{2\}$$

Ainsi  $P_c$  donne les coordonnées de  $P_m$  dans le repère  $R_c$ .

#### V.B.2.b. Transformation entre le repère caméra et le repère capteur

La transformation entre le repère caméra  $R_c$  et le repère capteur  $R_r$  correspond à la projection perspective transformant un point 3D ( $X_c, Y_c, Z_c$ ) en un point image ( $x_r, y_r$ ) (en unité métrique).

Ainsi on a (avec  $f$ , focale de l'objectif) :

$$x_r = f * \frac{X_c}{Z_c} \text{ et } y_r = f * \frac{Y_c}{Z_c} \{3\}$$

### V.B.2.c. Transformation entre le repère capteur et le repère image

La transformation entre le repère capteur et le repère image convertit les coordonnées images métriques ( $x_r, y_r$ ) en coordonnées images discrètes ( $i_i, j_i$ ) (pixels). Pour cette opération, il est nécessaire de prendre en compte le nombre de pixels par unité de longueur suivant les directions  $x$  et  $y$  du capteur,  $k_x$  et  $k_y$ . Dans le cas où les pixels sont carrés, ce qui est le cas pour des photographies numériques,  $k_x=k_y=k$ . Les autres paramètres à prendre en compte sont les coordonnées de l'intersection de l'axe optique avec le plan image à savoir le point principal donné en pixels ( $PP_x, PP_y$ ). Il faut, d'autre part, prendre garde à l'orientation des repères. En effet, une position d'un pixel dans l'image est définie d'abord par la ligne  $i$  puis par la colonne  $j$ . Or dans ce cas,  $i$  correspond à l'axe  $Y$  et  $j$  à l'axe  $X$ .

Ainsi, on obtient :

$$j_i = k * x_r + PP_x \text{ et } i_i = k * y_r + PP_y \quad \{4\}$$

En remplaçant {3} dans {4}, on a :

$$j_i = k * f * \frac{X_c}{Z_c} + PP_x \text{ et } i_i = k * f * \frac{Y_c}{Z_c} + PP_y \quad \{5\}$$

En notant  $F=k*f$ , la focale exprimée en densité de pixels, la formule {5} devient :

$$j_i = F * \frac{X_c}{Z_c} + PP_x \text{ et } i_i = F * \frac{Y_c}{Z_c} + PP_{xy} \quad \{6\}$$

Dans le cas où les conditions de Gauss seraient respectées,  $i_i$  et  $j_i$  donnerait directement les positions du point 3D  $P_m$  dans l'image.

### V.B.2.d. Distorsion dans le repère image

MICMAC définit une distorsion comme une composition de distorsions élémentaires. Chaque distorsion élémentaire est définie par un modèle paramétrique. Dans la majorité des configurations, une seule distorsion élémentaire aura besoin d'être considérée. Plusieurs modèles sont possibles dans MICMAC : modèle radial, modèle photogrammétrique standard, modèle grille, modèle unifié polynomial, modèle de Brown, modèle de Ebner, modèle fisheye.

Les modèles radial et photogrammétrique standard sont les plus utilisés c'est pourquoi il a été choisi pour l'instant de ne considérer que ces deux modèles au sein du système. Cependant ils offrent tout de même une modélisation assez précise de la distorsion.

**Modèle Radial** : Le modèle radial est modélisé par un centre de distorsion ( $C_x, C_y$ ) et de trois coefficients de distorsion :  $R_3, R_5$  et  $R_7$ .

$$\begin{pmatrix} j_d \\ i_d \end{pmatrix} = D_R \begin{pmatrix} j_i \\ i_i \end{pmatrix} = \begin{pmatrix} C_x \\ C_y \end{pmatrix} + (1 + R_3 * \rho^2 + R_5 * \rho^2 + R_7 * \rho^2) \quad \{7\}$$

$$\text{avec } \rho^2 = d_u^2 + d_v^2, d_u = j_i - C_x, d_v = i_i - C_y$$

**Modèle Photogrammétrique Standard** : Ce modèle intègre une partie radiale avec les mêmes paramètres que le modèle radial mais en ajoutant une partie affine de paramètres  $b_1$  et  $b_2$  et une partie dicentrique de paramètres  $P_1$  et  $P_2$ .

$$\begin{pmatrix} j_d \\ i_d \end{pmatrix} = D_P \begin{pmatrix} j_i \\ i_i \end{pmatrix} = D_R \begin{pmatrix} j_i \\ i_i \end{pmatrix} + P_1 * \begin{pmatrix} 2 * d_u^2 + \rho^2 \\ 2 * d_u * d_v \end{pmatrix} + P_2 * \begin{pmatrix} 2 * d_u * d_v \\ 2 * d_v^2 + \rho^2 \end{pmatrix} + b_1 * \begin{pmatrix} d_u \\ 0 \end{pmatrix} + b_2 * \begin{pmatrix} d_v \\ 0 \end{pmatrix}$$

#### V.B.2.e. Exploitation du résultat

Grâce à ces transformations successives, la position d'un point 3D dans une image est connue. Notre approche a donc été développée autour de cette relation géométrique entre 2D et 3D. Ainsi en répétant cette opération pour chacun des points d'un nuage de points, une liste de position dans l'image (pixels) des points de ce nuage est composée. Cependant afin de résoudre les problèmes d'unicité des correspondances sur une image, notre approche considère plusieurs étapes de rectification du résultat brut obtenu suite à la projection des points 3D sur les images.

Les différentes transformations ne considèrent jamais les dimensions de l'image. Ainsi pour un point qui est en dehors de l'image, une position pixel est tout de même obtenue. Pour pouvoir éliminer ces points de la liste des points de l'image, les composantes de la position des pixels sont comparées aux dimensions de l'image. Si une des composantes dépasse la dimension correspondante sur l'image, le point est éliminé.

D'autre part, pour une même position de pixels dans l'image, plusieurs points de l'espace peuvent correspondre. Dans ce cas, le choix du point à considérer pose problème. Ainsi, nous avons choisi une méthode simple pour déterminer le point à conserver dans ces cas grâce à une mesure de distance. En effet, la distance entre les points et le centre optique nous donne une indication importante. Un point qui est plus loin du centre optique qu'un autre tombant sur le même pixel a de grandes chances d'être caché sur la photographie. C'est pourquoi lorsque deux ou plusieurs points correspondent à une même position pixel sur l'image, le point retenu est celui qui a une distance au centre optique la plus courte (Figure 73).

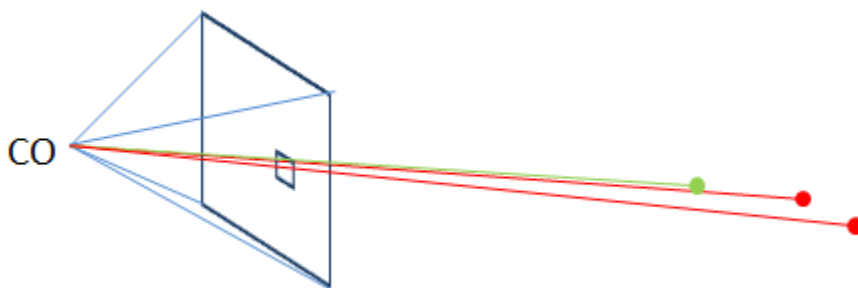


Figure 73 : Elimination de points multiples sur un pixel : seul le plus proche du centre optique (CO) est conservé, les autres points sont considérés comme non visibles sur cette image.

Un autre problème peut apparaître si le nuage de points n'est pas assez dense pour la dimension des images. En effet, dans ce cas, si les points du nuage sont projetés sur les pixels de l'image, de nombreux points de l'image n'auront pas de correspondance 3D. Pour avoir une densité de correspondance de point 3D la plus élevée sur l'image, il est possible d'envisager de projeter sur l'image redimensionnée. Ainsi si on redimensionne l'image à une dimension inférieure à la taille d'origine, la densité de correspondance au niveau des pixels de l'image aura plus de chance d'être élevée (Figure 74). Cependant comme l'image comportera moins de pixels, moins de points 3D peuvent être projetés.

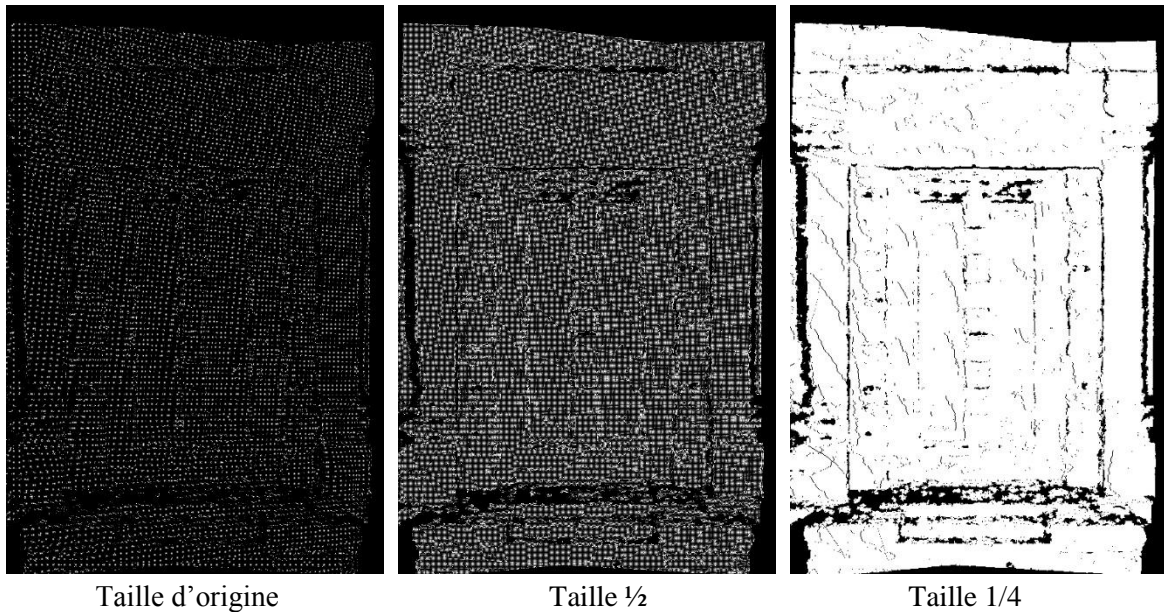


Figure 74 : Projection du nuage de points sur différentes dimensions d'une image, en blanc les pixels où une correspondance 3D a été trouvée. Plus l'image est redimensionnée pour la projection, plus la densité de correspondances 3D est élevée.

La densité du nuage de points peut poser d'autres problèmes, notamment si le nuage de points comporte des zones nettement moins denses en points que d'autres. En effet, en fonction du procédé de relevé, certaines parties peuvent être moins denses que d'autres. Au niveau de la projection des points 3D sur les images, ces zones moins denses peuvent entraîner une projection de points qui sont situés sur une partie censée être cachée derrière une autre (par exemple deux murs parallèles d'un édifice). Ainsi, si le nuage de points de la partie normalement visible n'est pas assez dense, certains points de la partie devant être cachée peuvent se retrouver projetés sur l'image alors qu'ils n'y sont pas visibles (Figure 75). En effet, pour ces points, aucun point de la zone visible n'aura été détecté sur le pixel (et par conséquent aucun point plus proche que celui de la partie cachée). Ainsi le point situé à l'arrière sera projeté sur l'image, causant des erreurs de correspondances.



Figure 75 : Sur ce nuage de points, les trous dans les parties des contreforts peuvent provoquer des erreurs de correspondances

Afin d'essayer de corriger ces erreurs, l'approche considère une analyse de la carte de profondeur des images. A partir des correspondances 3D des pixels calculées et de la position dans l'espace du centre optique de l'image, une carte en niveaux de gris représentant l'éloignement des points par rapport au centre optique peut être construite : les points les plus proches sont représentés par des points blancs et les points les plus éloignés par des points noirs. Une échelle de couleur est ainsi construite pour représenter cet éloignement. Les points présentant les erreurs affichent généralement une profondeur beaucoup plus importante que les points l'entourant. En considérant l'aspect de continuité d'un objet, ces sauts de profondeur ne devraient pas apparaître. Les points où apparaissent ces erreurs peuvent donc être détectés en étudiant les zones où des sauts de profondeurs apparaissent. Pour cela, la carte de profondeur est filtrée à l'aide d'un filtre médian qui va « lisser » la carte de profondeur. Puis par différence de la carte de profondeur « lissée » et de la carte de profondeur « brute », les points où de grands sauts de profondeurs apparaissent vont ressortir (Figure 76). Ainsi la carte de profondeur peut être corrigée pour éliminer ces points et par conséquent éliminer ces points des correspondances 3D.



*Carte de profondeur « brute »      Carte de profondeur « lissée »      Carte de profondeur corrigée*

*Figure 76 : Correction des points par l'étude de la carte de profondeur*

Cette méthode fonctionne bien pour des zones où seul un pixel ou un groupe de pixels (moins de dix pixels) présentent ce saut. Néanmoins dans des zones où le nuage de points est très peu dense, ces erreurs peuvent représenter des groupes de pixels trop conséquents pour pouvoir considérer l'aspect des voisins dans l'image. Dans ces cas, il faudrait aller jusqu'à une triangulation du nuage de points pour pouvoir éliminer les erreurs. Cependant un tel procédé demande d'établir des calculs relativement lourds.

D'autre part, les normales peuvent aussi aider à discriminer ces points. En effet, en testant la direction de la normale d'un point par rapport à ses voisins, il est possible d'évaluer si ce point est cohérent avec les autres points qui l'entourent. Si un point a une normale de direction inverse à tous les autres points qui l'entourent, la probabilité qu'il n'appartienne pas à la même surface est grande et par conséquent qu'il n'est pas censé être projeté sur l'image.

Pour la partie corrélation des images, MICMAC utilise des masques de corrélations (Figure 77). Ces masques déterminent les points de l'image à projeter dans l'espace. Seules certaines zones des images sont choisies pour la corrélation en fonction de différents critères. Ainsi une image ne servira qu'en partie à générer le nuage de points. C'est la superposition de



l'extraction de points de toutes les images qui donnera le nuage de points global. Ces cartes donnent donc une indication sur les pixels qui ont effectivement un point 3D dans le nuage. Ainsi, les pixels des images ayant servi à la génération du nuage de points n'ont pas besoin d'être testés puisqu'ils ont nécessairement une correspondance dans le nuage de points.



*Figure 77 : Masques de corrélation et image associée*

Une combinaison de ces trois critères pourrait permettre d'éliminer les fausses projections dues à la densité du nuage de points. Pour l'heure, seul le critère de profondeur a été implémenté et nous sommes partis sur l'hypothèse que le nuage de points est assez dense pour ne pas causer d'erreurs de correspondances trop importantes (des zones de plusieurs pixels) sur les images.

Ainsi, à la fin de cette étape de projection, on obtient pour chaque image, les correspondances entre pixels et points 3D.

### V.B.3. Cas des orthophotographies

Les modules PIMS2MNT et TAWNY de MICMAC servent à générer, à partir des photographies orientées, des orthophotographies (Figure 78). Ce sont des images de projection sur un plan spécifique (Figure 78). Ces images permettent d'avoir un bon point de vue pour les annotations.





Figure 78 : Orthophotographie

Ces images ne se placent pas dans un modèle de projection perspectif. Ici, la projection est de type orthogonal. Il faut donc pouvoir connaître le plan de projection et la position de l'image orthophotographique par rapport au nuage de points.

Pour ces images, les modules génèrent deux fichiers XML donnant ces informations :

- Un fichier Repère donnant la position du coin en haut à gauche de l'image et l'orientation du repère pour la projection (Figure 79)
- Un fichier MTD (*Modèle de Terrain de Données*) donnant la position du coin en bas à droite de l'image (Figure 80)

```
<?xml version="1.0"?>
- <RepereLoc>
  - <RepereCartesien>
    <Ori>3.15875866358997115 -2.58435761255958951 0.893859757024263346</Ori>
    <Ox>0.00883108961089469624 0.999889500730426817 0.0119581848682047773</Ox>
    <Oy>-0.0162170325070229715 -0.0118138690412011877 0.999798699916609968</Oy>
    <Oz>0.99982949532055243 -0.00902323818455226696 0.01611091120062208</Oz>
  </RepereCartesien>
</RepereLoc>
```

Figure 79 : Fichier de repère

```
<?xml version="1.0"?>
- <FileOriMnt>
  <NameFileMnt>NO</NameFileMnt>
  <NombrePixels>7312 5344</NombrePixels>
  <OriginePlani>-1.384 4.04</OriginePlani>
  <ResolutionPlani>0.001 -0.001</ResolutionPlani>
  <OrigineAlti>0</OrigineAlti>
  <ResolutionAlti>0.004</ResolutionAlti>
  <Geometrie>eGeomMNT Euclid</Geometrie>
</FileOriMnt>
```

Figure 80 : Fichier MTD

Dans le fichier repère, le paramètre <Ori> donne la position dans l'espace du coin en haut à gauche de l'image. Les paramètres <Ox>, <Oy>, <Oz> définissent le repère local pour la projection. <Ox> et <Oy> détermine les directions définissant le plan de projection.

A partir de ces premières données, le fichier MTD donne les autres informations pour pouvoir finir de placer l'image orthophotographique dans l'espace par rapport au nuage de points. Le paramètre <OriginePlani> donne la position du coin de l'image en bas à droite par

rapport à la position du coin en haut à gauche dans le repère local. Ainsi à partir de ces données, la position de l'orthophotographie dans l'espace est déterminée (Figure 81).



*Figure 81 : Position de l'orthophotographie dans l'espace par rapport au nuage de points*

Au final, il ne reste plus qu'à projeter les points du nuage de points de façon orthogonale sur les pixels positionnés dans l'espace pour connaître la correspondance 2D/3D. Les correspondances établies entre le nuage de points et les pixels peuvent être contrôlées de la même manière que dans le paragraphe précédent à l'aide des critères de profondeur ou de normales (voir [V.B.2.e. Exploitation du résultat](#)).

#### V.B.4. Scans Laser

Dans le cas d'un relevé 3D obtenu par scanner laser et d'images acquises sur le même appareil (Figure 82), la correspondance 3D est déjà établie entre chacun des nuages de points et leur image. En effet, la relation est directement établie entre les points 3D d'un des nuages de points du relevé et l'image associée par rapport à la projection considérée (voir [II.B.3.b. Projection sphérique](#)). Cependant il faut garder en mémoire que chaque image n'est associée qu'à un nuage de points du relevé en particulier.



*Figure 82 : Image (luminance) extraite d'un nuage de point relevé par scanner laser 3D Faro (projection sphérique)*

D'autre part, il est possible aussi d'utiliser des images qui n'ont pas été acquises par le scanner laser. Dans ce cas, les images peuvent tout de même être orientées autour du nuage de points global du relevé en définissant des correspondances entre les pixels des images et des points 3D (voir [V.A.2. Référencement image-modèle 3D](#)). Cependant, la définition d'une calibration précise nécessite d'avoir plusieurs photographies prises avec les mêmes paramètres. Les images ainsi orientées peuvent être mises en relation avec le nuage de points par projection perspective (voir [V.B.2. Représentation géométrique d'une image photographique](#)).

### V.B.5. Projections et relation 2D/3D

Les trois paragraphes précédents décrivent la manière de créer la relation pixels/points 3D en fonction de différents types de projection du réel sur une image. La Figure 83 résume le lien entre les différentes images et la projection de points 3D sur les pixels de ces images.

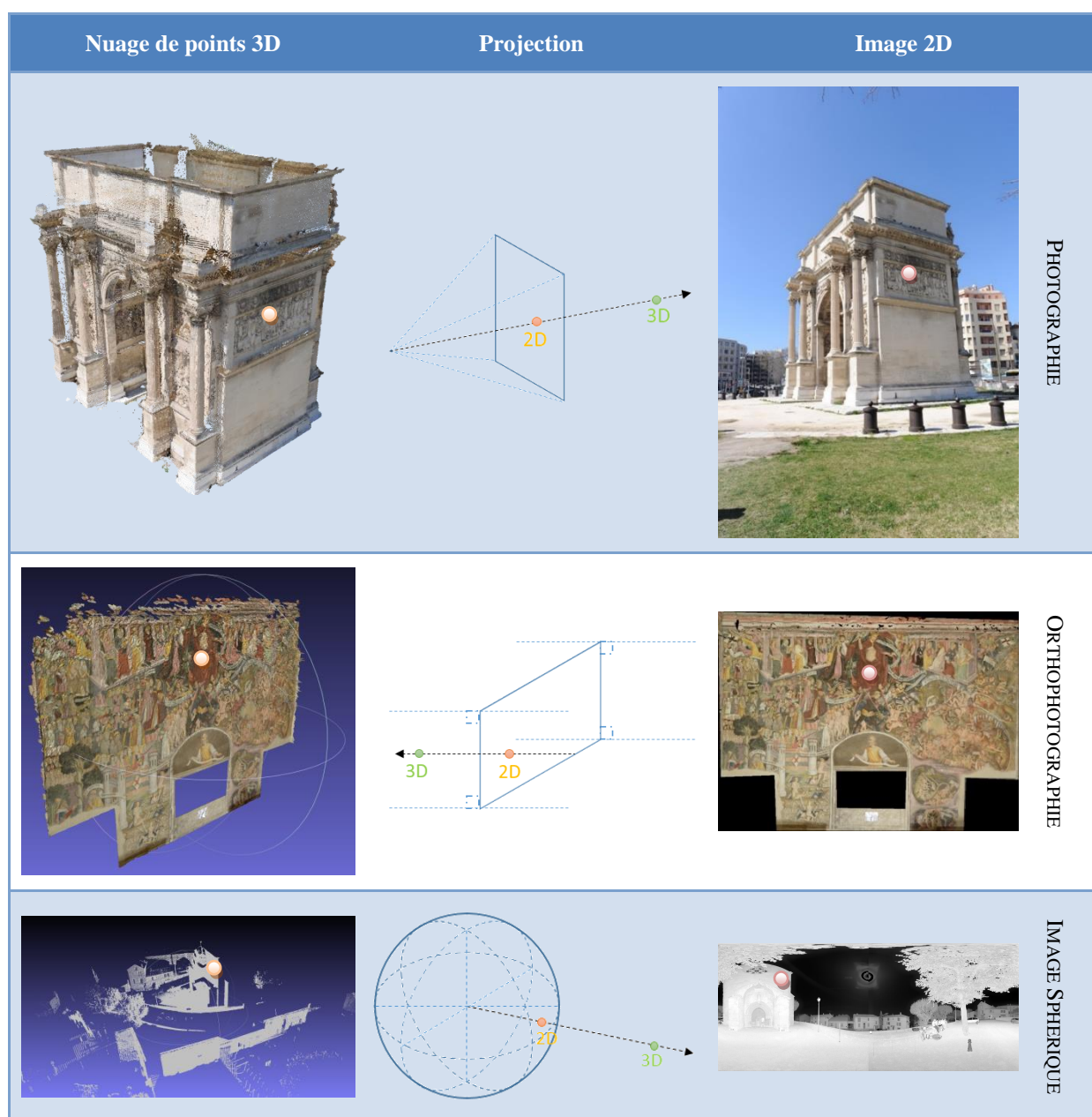


Figure 83 : Projections inverses



En définissant ainsi une conversion 3D vers 2D pour chacune des projections, l'approche proposée permet de définir une relation 2D/3D entre un nuage de points et une image. Une fois la correspondance 2D/3D établie, le type de projection des images n'est plus un critère limitant pour la mise en relation de ces images. De plus, étant donné que quel que soit le type de projection, la conversion 3D vers 2D associe à un pixel une coordonnée 3D, la structuration des données de correspondances 2D/3D peut être commune aux différents types de projection (voir paragraphe suivant [V.C. Gestion des données](#)).

Ainsi l'approche permet de pouvoir considérer plusieurs types d'acquisition et, par conséquent, plusieurs types de projection au sein du même système.

### V.C. Gestion des données

La recherche de correspondance 3D donne une liste de points 3D correspondant à une position de pixel sur les images, quel que soit le type de projection considéré. Ceci génère d'importantes quantités d'informations qu'il faut pouvoir stocker afin de les réutiliser par la suite dans le système. Le problème est donc de trouver une manière de gérer ces données liées aux images.

MICMAC génère directement des fichiers TIFF contenant ces informations. En effet, le principe est d'utiliser la structure multicouche des fichiers TIFF pour stocker d'autres types d'informations que les couleurs.

Les fichiers TIFF sont des fichiers images stockant l'information de couleurs sur trois couches de niveaux de gris représentant chacun une couleur élémentaire : Rouge, Vert, Bleu. Ainsi une image est représentée par trois tableaux contenant chacun des valeurs entre 0 et 255 représentant une composante couleur des pixels (Figure 84). Les tableaux sont organisés de la même façon que les pixels de l'image. Pour pouvoir accéder à chacune des couleurs d'un pixel à une position (i, j) dans l'image, il suffit de lire chacun de tableaux à cette même position pour obtenir un triplet de valeur représentant la couleur du pixel.

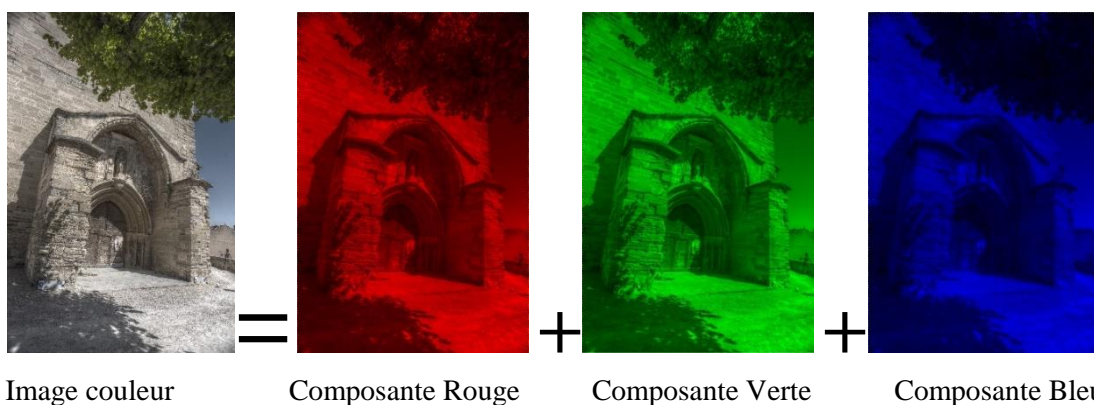


Figure 84 : Décomposition d'une image TIFF en trois couches : Rouge, Vert, Bleu (niveaux de gris colorés)

MICMAC utilise cette structure pour stocker les informations de coordonnées 3D au lieu de stocker les informations couleurs (Figure 85). Ainsi un fichier TIFF, appelé fichier XYZ, contenant les informations des coordonnées des points 3D liés aux pixels, est associé à chaque image de l'ensemble par MICMAC. Ainsi, les composantes X des pixels sont stockés dans la

couche Rouge, les composantes Y sont stockées dans la couche Vert et les composantes Z sont stockées dans la couche Bleu.

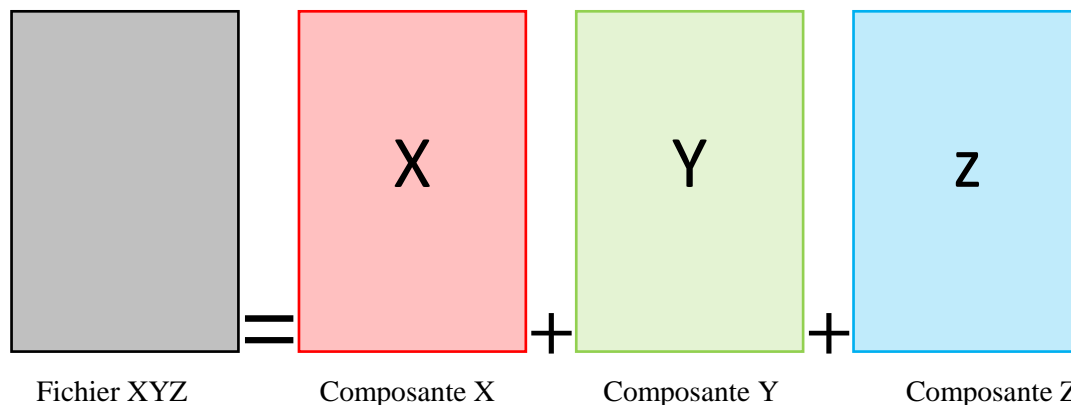


Figure 85 : Décomposition d'un fichier XYZ TIFF en trois couches : X, Y, Z (valeurs des composantes des coordonnées 3D)

Cependant il n'est pas possible d'utiliser directement ces fichiers fournis par MICMAC. Le problème vient du fait que MICMAC ne stocke dans ces fichiers qu'une partie des correspondances pixels du nuage de points sur l'image. En effet, lors de la corrélation dense pour la génération du nuage de points, il est nécessaire d'établir des masques de corrélation (voir [V.B.2.e. Exploitation du résultat](#)). Les fichiers XYZ, donnés par MICMAC, ne contiennent donc que les correspondances 3D des points utilisés dans la corrélation pour l'image associée. D'autre part, si on définit des masques de corrélation sur toute l'image, afin que les fichiers XYZ contiennent toutes les correspondances des pixels, le nuage de points obtenus comportera de nombreuses erreurs.

L'utilisation de la structure de stockage par fichier TIFF peut néanmoins être utilisée. A partir des résultats de la recherche de correspondance de points 3D par projection, ces fichiers peuvent être construits indépendamment de MICMAC. Ceci offre une structure de stockage simple qui pourrait s'étendre à d'autre type d'informations tel que le stockage des composantes des normales des pixels.

Néanmoins dans le cas d'étude d'objet patrimonial, le nombre de photographies nécessaires est relativement élevé (plusieurs centaines voire milliers de photographies pour un seul objet). L'utilisation de fichier XYZ pour ces études produirait des quantités de fichiers TIFF très importants, ce qui demanderait une gestion importante de données. Ainsi, pour simplifier, il est possible de passer par l'utilisation d'une base de données pour stocker ces informations de correspondances 3D.

Le nuage de points utilisé pour la projection étant unique, deux pixels de deux images différents correspondants au même point 3D dans l'espace auront donc les mêmes coordonnées 3D. Ainsi pour pouvoir reconnaître que ces deux pixels sont les mêmes dans l'espace, il faut comparer l'ensemble des triplets de coordonnées de ces pixels. Cependant, la comparaison de triplets de coordonnées est plus longue que la comparaison d'un nombre simple. C'est pourquoi nous avons envisagé de seulement stocker un index de points lié à une position dans l'image au lieu d'un triplet de coordonnées plus lourd à exploiter par la suite. Chaque point 3D du nuage de points est donc associé à un index. Pour cela, un tableau (Figure 86) est créé stockant les informations d'index et de données 3D (coordonnées, normales,

couleur du point sur le nuage, etc.). A ces données peut être ajouté une colonne contenant la liste des images sur lesquels le point apparaît.

Index	X	Y	Z	Nx	Ny	Nz	R	G	B	Images	...
<b>1</b>	X1	Y1	Z1	Nx1	Ny1	Nz1	R1	G1	B1	Im1, Im3, ...	...
<b>2</b>	X2	Y2	Z2	Nx2	Ny2	Nz2	R2	G2	B2	Im1, Im2, ...	...
...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...
<b>n</b>	Xn	Yn	Zn	Nxn	Nyn	Nzn	Rn	Gn	Bn	Im3, Imp, ...	...

Figure 86 : Tableaux stockant les index des points 3D associés aux données 3D

Pour les images, il n'est nécessaire de ne conserver qu'une liste d'index. Cette liste suit l'organisation des pixels de l'image et est donc composée d'autant d'index qu'il y a de pixels dans l'image. Pour les pixels où aucune correspondance 3D n'a été trouvée l'index 0 est associé. Cette liste peut donc aussi être simplement stockée dans un tableau contenant les informations relatives à chacune des images.

L'utilisation d'un index pour le stockage des informations de correspondances 3D de chacune des images suppose de partir d'un seul nuage de points que l'on projette sur l'ensemble des images. Dans les cas où l'on a une correspondance 3D obtenue d'une autre manière que par projection sur une photographie, l'utilisation des fichiers XYZ, bien que plus lourde en termes de stockage, offre plus de possibilités d'adaptation que l'indexation.

Dans le cas d'images issues directement de scanner laser, la correspondance est établie par rapport au nuage de points associé à l'image. L'indexation des points 3D pour le stockage n'a pas de sens dans ce cas car il ne faut prendre en compte que deux pixels de deux images différentes, et donc de deux nuages de points différents, mais qui semblent être au même endroit dans l'espace n'auront pas exactement les mêmes coordonnées 3D, ceci dû au fait que les deux nuages sont acquis à deux endroits différents et que l'opération de recalage des nuages entre eux provoque des variations entre les valeurs de coordonnées.

Le choix de l'utilisation des fichiers XYZ ou de l'indexation dépend donc des données d'entrées. Le seul impératif reste de pouvoir mettre en relation les données 3D avec les données 2D de façon cohérente et de pouvoir accéder à ses données de manière efficace.

Pour la suite, les résultats sont basés sur l'utilisation des modules de MICMAC et donc sur la possibilité d'indexer le nuage de points. Cependant l'utilisation de fichier XYZ ne change pas le système dans sa globalité mais seulement sur la manière de stocker ces informations de correspondances 3D. Quelques spécificités sont nécessaires à établir pour l'utilisation de ses fichiers mais ne change pas la méthodologie développée.

La mise en relation ainsi établie entre les informations 2D liées aux images (pixels) et les informations 3D liées au nuage de points (coordonnées 3D) va permettre dans la suite du procédé de mettre en œuvre de nombreuses possibilités d'aide à l'analyse d'un objet numérisé.



## V.D. Ajout de nouvelles images

Comme vu dans le [Chapitre II](#), les capteurs 2D peuvent acquérir des images qui ne soient pas uniquement des images donnant un rendu de la réalité tel que visible. En effet, des images de fluorescence sous rayonnement ultraviolet ou privilégiant le rayonnement infrarouge peuvent aussi être acquises pour un bâtiment (Figure 87). Ces images offrent d'autres informations relatives à l'objet qu'il est intéressant de pouvoir ajouter au système.

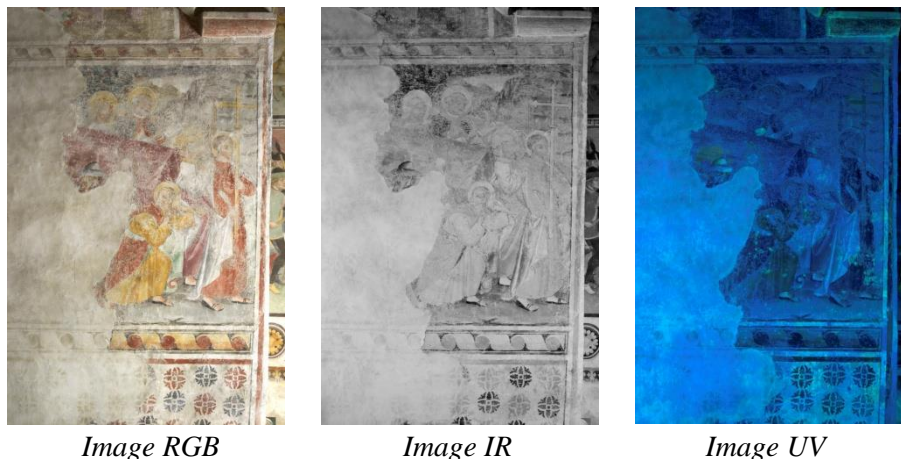


Figure 87 : Différents types d'images pouvant être acquis pour l'analyse (campagne de relevé du CICRP)

Compte tenu du système établi, l'idée est de pouvoir assurer l'orientation de ces images au sein des autres images ayant servi à générer le nuage de points. En effet, si l'image est orientée parmi les autres, il est possible de projeter les points 3D du nuage de points pour connaître la correspondance 3D avec les pixels.

### V.D.1. Orientation de l'image

L'approche expérimentée se base sur la possibilité d'orienter une image parmi les autres images déjà orientées à l'aide de TAPAS (voir [V.A.5. Solution retenue](#)). Cependant il est nécessaire que cette nouvelle image offre suffisamment de points communs avec les autres images. Cette étape de recherche étant automatique, cela peut ne pas fonctionner dans certains cas où l'image ne rend pas un aperçu proche des autres images.

D'autre part, il faut aussi considérer le fait que l'acquisition n'a pas été réalisée dans les mêmes conditions, notamment au niveau des caractéristiques de l'appareil pour la prise de vue. L'acquisition d'une nouvelle image suppose donc de prévoir, si nécessaire, d'entreprendre une étape de calibration de l'appareil. Pour cela, il est nécessaire d'avoir plusieurs images.

#### V.D.1.a. Orientation directe

Si l'image offre suffisamment de points communs détectables automatiquement par TAPIOCA (voir [V.A.5. Solution retenue](#)), la nouvelle image est orientée, à l'aide de TAPAS, par rapport aux autres images dont l'orientation est fixée et est donc orientée par rapport au nuage de points (Figure 88).

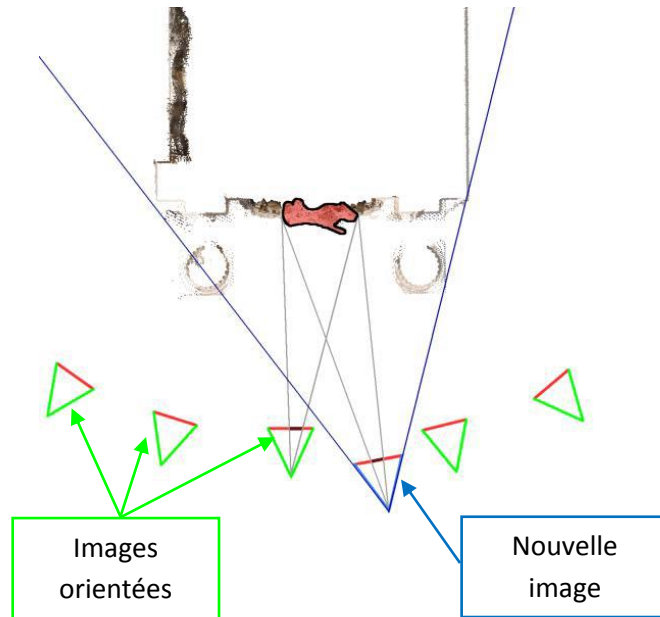


Figure 88 : Orientation d'une nouvelle image parmi les images déjà orientées

Le problème de l'orientation de l'image est dans ce cas relatif à la possibilité d'orienter une image prise dans des conditions différentes. Dans le cas d'images RGB, obtenues à partir d'un appareil photographique numérique, les conditions entre deux prises de vues peuvent avoir évoluées, notamment au niveau de l'éclairement, du temps, de la pollution, etc. Ainsi même si cette image représente le même objet, ces conditions ne rendent pas l'objet de la même manière sur la photographie (Figure 89).



Figure 89 : Différences au niveau de la diachronie pour plusieurs photographies d'un même objet

Il est donc naturel de se poser la question de la possibilité de trouver automatiquement des points communs entre les images à causes de ces différences. Cependant la pratique a montré

que TAPAS (basé sur SIFT [Lowe D.G., 1999]) était suffisamment robuste pour arriver à orienter des images présentant des différences à ce niveau. La robustesse est telle qu'il est possible d'orienter par cette méthode les images infrarouge ou ultraviolet.

#### *V.D.1.b. Orientation indirecte*

Dans le cas des images RGB, ultraviolet ou infrarouge montrée à la Figure 87, l'orientation directe de l'image fonctionne. Cependant dans le cas d'autres types d'images tel que les images thermographiques, TAPIOCA ne trouve pas de points communs entre l'image thermographique et les images RGB déjà orientées. Cependant, il est possible de passer par une autre image qui elle peut être orientée par rapport aux autres images et dont on connaît l'orientation relative par rapport à l'image dont on souhaite connaître l'orientation (Figure 90).

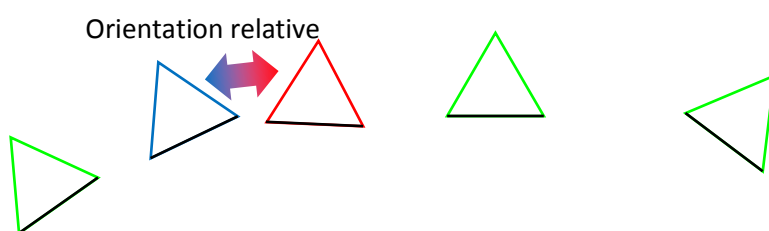


Figure 90 : Orientation d'une nouvelle image par l'intermédiaire d'une autre image : la caméra bleue est orientée par rapport aux caméras vertes, et la caméra rouge est orientée par rapport à la caméra bleue

Ainsi par transition, une image qui ne peut pas être orientée à cause de la recherche de points communs peut tout de même être orientée parmi les autres. Cependant cela suppose que l'on dispose d'une image qui puisse être orientée et dont on connaît l'orientation relative par rapport à l'image ne pouvant être orientée. Dans le cas, d'une image thermographique acquise à l'aide d'une caméra thermique, par exemple une caméra FLIR, cette image est disponible car la caméra intègre deux capteurs : un capteur d'imagerie thermique et une caméra haute sensibilité (Figure 91). Ainsi le même dispositif peut à la fois acquérir une image RGB et une image thermographique. L'ensemble étant fixe, il est possible de déterminer l'orientation relative des deux caméras. Cela suppose donc de définir une méthode pour déterminer l'orientation entre les deux images générées par l'appareil.

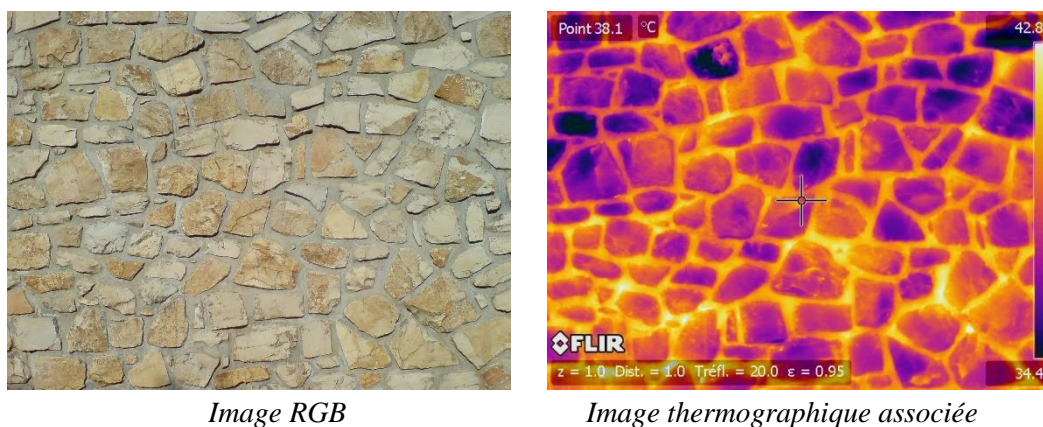


Image RGB

Image thermographique associée

Figure 91 : Images acquises par une caméra thermique

Dans le cas où le système n'intègre pas de seconde caméra avec la possibilité d'obtenir une image RGB simple, il est possible d'envisager de rajouter au système une ou plusieurs



caméras externes qui pourront servir de références. Un tel système suppose de définir précisément la position de ces caméras autour du système d'acquisition.

#### *V.D.1.c. Orientation d'images à différents niveaux de détails*

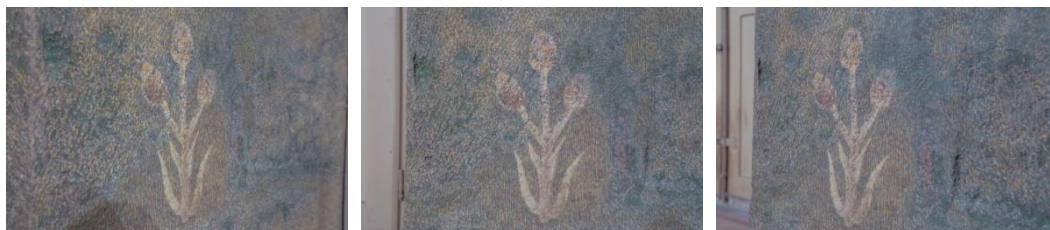
L'orientation directe ou indirecte fonctionne bien dans le cas de l'ajout d'une image qui a été prise avec un même niveau de détails (même distance de l'objet) que les images déjà orientées. Cependant si la nouvelle image a été prise à une distance beaucoup plus proche ou beaucoup plus éloignée que l'ensemble des autres images (Figure 92), TAPAS peut avoir des difficultés à orienter cette nouvelle image parmi les autres.



*Figure 92 : Images d'un même objet prises à deux niveaux de détails*

Cette impossibilité d'orienter correctement l'image vient du fait qu'une image est d'autant mieux orientée que le nombre de points communs détectable par TAPIOCA est élevé et que ces points communs sont répartis sur l'ensemble de l'image. Ici, une des deux images ne représente qu'une infime partie de l'autre. L'image à un niveau de détails plus proche peut ne représenter que quelques pixels sur l'image à niveau de détails plus éloignée. Il est donc plus difficile d'obtenir suffisamment de points communs. Sur l'exemple présenté en Figure 92, l'orientation est d'autant plus difficile puisqu'il s'agit d'un détail d'un motif qui se répète sur l'image plus éloignée.

Cependant, il n'est pas complètement impossible d'orienter cette image au sein de l'ensemble. Afin de pouvoir orienter cette image, il faut prendre en compte ce problème dès le moment de l'acquisition et prévoir l'acquisition d'images intermédiaires permettant d'effectuer une transition moins forte entre deux niveaux de détails. Ainsi en plus de la nouvelle image, ces images intermédiaires sont prises à des distances se situant entre l'image la plus éloignée et l'image la plus proche (Figure 93). Cette acquisition d'images intermédiaire va permettre de pouvoir échelonner l'orientation. Ces images seront orientées successivement parmi les images déjà orientées en partant de la plus proche, en termes de niveau de détails, des images initiales jusqu'à l'image la plus détaillée et pouvoir ainsi orienter l'image avec un niveau de détails différent.



*Figure 93 : Images intermédiaires pour l'orientation d'images de différents niveaux de détails*

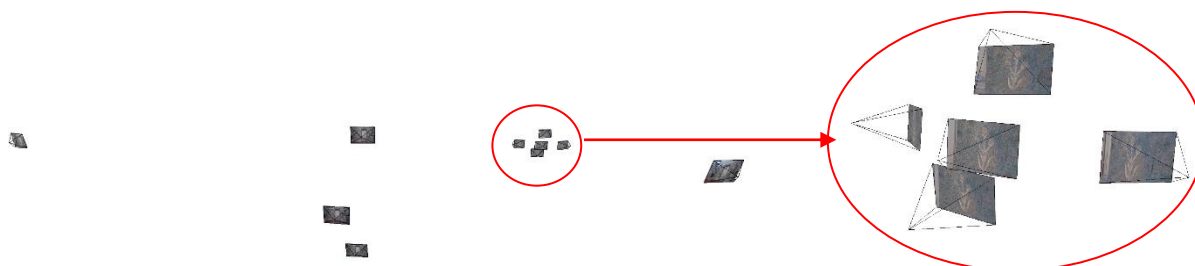
De manière plus procédurale, le principe repose sur plusieurs étapes :

- **Orientation des images constituant la base :** Les images constituant la base (images les plus éloignées de l'objet) sont orientées en premier à l'aide des modules TAPIOCA et TAPAS (Figure 94). Les orientations de ces images sont alors fixées c'est-à-dire qu'elles sont considérées comme des paramètres rigides dans les étapes suivantes du processus.



*Figure 94 : Orientation des images constituant la base*

- **Orientation des images intermédiaires :** Les images intermédiaires sont ensuite orientées par rapport aux images de la base toujours à l'aide des modules TAPIOCA et TAPAS (Figure 95). L'analyse du nombre de points communs trouvés par TAPIOCA permet de donner une indication sur la possibilité de l'orientation et de pouvoir ajuster les paramètres. Si les images intermédiaires n'ont pas été prises à l'aide du même appareil ou avec les mêmes paramètres d'acquisition, il est nécessaire de prévoir un calcul de calibration pour ces images avant l'orientation (au minimum 3 images sont nécessaires pour la calibration, idéalement 5). Si l'acquisition a nécessité d'effectuer plusieurs niveaux d'images intermédiaires, cette étape doit être répétée pour chacun des niveaux (du plus éloignée au plus proche) par rapport aux images du niveau précédent.



*Figure 95 : Orientation des images intermédiaires*

- **Orientation des images de détails :** Les images de détails sont finalement orientées par rapport aux images intermédiaires du dernier niveau de la même manière à l'aide des modules TAPIOCA et TAPAS (Figure 96). Il peut être aussi nécessaire de calibrer l'appareil à cette étape. Cependant il faut absolument avoir prévu l'acquisition d'images pour la calibration. Les images étant orientées par rapport aux images intermédiaires, elles sont alors, par transition, orientées par rapport aux images de la base (Figure 97).

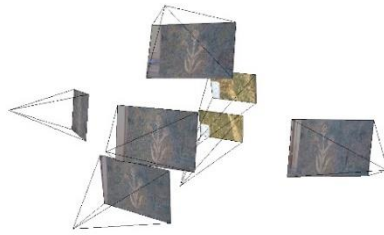


Figure 96 : Orientation des images de détails par rapport aux images intermédiaires

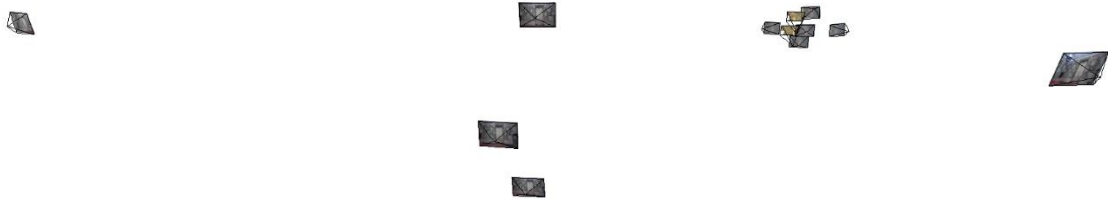


Figure 97 : Orientation des images de détails parmi toutes les images

Ainsi en procédant par étapes, les images à différents niveaux de détails peuvent aussi être intégrées au système. Cependant, cela suppose d'être rigoureux lors de la phase d'acquisition et de prévoir d'acquérir des images intermédiaires pour pouvoir permettre l'orientation et la calibration (si nécessaire) par la suite.

#### V.D.2. Résolution d'image et densité du nuage de points

La nouvelle image pouvant donc être orientée parmi les autres images, la seconde étape permettant de l'intégrer au système est de chercher les correspondances 3D avec les pixels de l'image. La méthode de projection reste la même que celle décrite dans le [Chapitre V](#). Seulement il faut prendre garde à certains paramètres qui entrent en jeu.

Dans le [Chapitre V](#), nous avons montré qu'il était nécessaire d'utiliser une redimension de l'image lors de la projection des points 3D pour la recherche des correspondances 2D/3D. Sur les images ayant servi à la génération du nuage, cette valeur de redimensionnement est aisée à évaluer puisque le nuage de point est relatif à ces images et sa densité est donc choisie en fonction des images. Pour une image ajoutée, la résolution pouvant être différente, le redimensionnement nécessaire à une bonne densité de projection peut donc être différent. Pour l'heure, l'approche demande à l'utilisateur de définir cette valeur. Ainsi c'est à l'utilisateur d'évaluer quelle dimension choisir pour la projection. Cependant, pour les développements futurs, il pourrait être intéressant de définir un critère d'évaluation de ce redimensionnement afin que la valeur puisse être choisie automatiquement.

Pour des images prises à une distance similaire des images ayant servi à générer le nuage de point, une simple utilisation du redimensionnement peut suffire à obtenir une densité de projection suffisante pour le système de propagation. Cependant, dans le cas notamment d'images de détails, ce redimensionnement peut s'avérer problématique. En effet, dans ces cas, l'image peut avoir une résolution proche des images initiales mais couvrir une zone beaucoup plus réduite. Ainsi la zone couverte sur le nuage de points est elle aussi réduite. Afin de



trouver une densité de projection correcte, il serait nécessaire de redimensionner l'image à des dimensions beaucoup trop réduites par rapport à la dimension d'origine.

Pour éviter ce redimensionnement trop important, il faut que le nuage de points soit plus dense au moins au niveau de la zone couverte par l'image de détails. L'idée est de prendre des images supplémentaires pour pouvoir établir une corrélation locale autour de ce détail et ainsi avoir une densité locale du nuage de points plus importante (Figure 98).



*Figure 98 : Génération d'une zone plus dense dans le nuage de points pour des images de détails*

L'utilisation des images intermédiaires prises pour l'orientation peut permettre cette corrélation si celles-ci ont été prises en prévoyance de l'orientation et de la corrélation. Cela suppose donc de bien prévoir la phase d'acquisition en prenant en compte ces deux aspects afin de limiter au maximum le nombre d'images nécessaires aux deux étapes.

### V.D.3. Ajout d'images de différents états temporels

Le fait de pouvoir ajouter de nouvelles images provenant d'une autre acquisition introduit le problème de la gestion de différents états temporels. En effet, la nouvelle image ajoutée peut avoir été prise à un temps où l'objet étudié a connu de nombreux changements physiques depuis l'acquisition initiale. Ainsi si l'on projette le nuage de points obtenu avant les changements, les correspondances 3D trouvées pourraient ne pas correspondre à l'état attesté par la nouvelle image.

Si l'objectif est de n'ajouter qu'une seule image présentant une évolution par rapport à l'acquisition initiale, le principe est de pouvoir indiquer à l'utilisateur que cette image n'a pas été prise au même moment et qu'elle peut donc présenter des différences par rapport au nuage de points généré et ainsi avoir une projection qui ne correspond pas à la réalité réfléchiée par l'image. Ces informations peuvent simplement être extraites à partir des métadonnées (si elles sont disponibles) en comparant les dates de prises de vues et ainsi donner cette indication à l'utilisateur. La propagation des annotations peut permettre ainsi de faire ressortir les différences entre cette image et les images de l'ensemble initial.

Si l'objectif est d'ajouter plusieurs images d'un temps différent, il peut être envisagé de construire un nouveau nuage de points pour cet état temporel. Ainsi dans le cas de l'établissement d'un suivi temporel, l'utilisation de plusieurs nuages de points (un pour chaque temps) peut être envisagée (Figure 99).

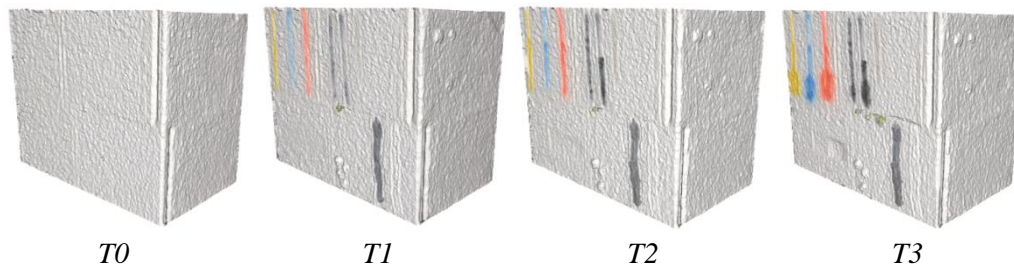


Figure 99 : Nuages de points pour différents temps d'un même objet

[Peteler F. et al, 2015] ont commencé à explorer le développement d'un processus pour comparer différents états temporels à partir d'orthophotographies. Le principe de base est de générer un nuage de points par temps à partir des images correspondantes en faisant en sorte que ces nuages soient tous orientés les uns par rapport aux autres puis de générer des orthophotographies et des cartes de profondeur liées à ces orthophotographies qui seront utilisées ensuite pour la comparaison des différents temps. Ceci est possible par l'utilisation des modules de MICMAC à la condition toujours que les images offrent suffisamment de points communs et donc que les différents temps ne montrent pas de changements trop importants entre eux. Les images de tous les temps doivent donc être orientées entre elles puis chacun des temps sert à générer un nuage de points (Figure 100). Les images sont donc orientées autour de tous les nuages de points quel que soit le temps des images et quel que soit le temps du nuage de points (Figure 101).

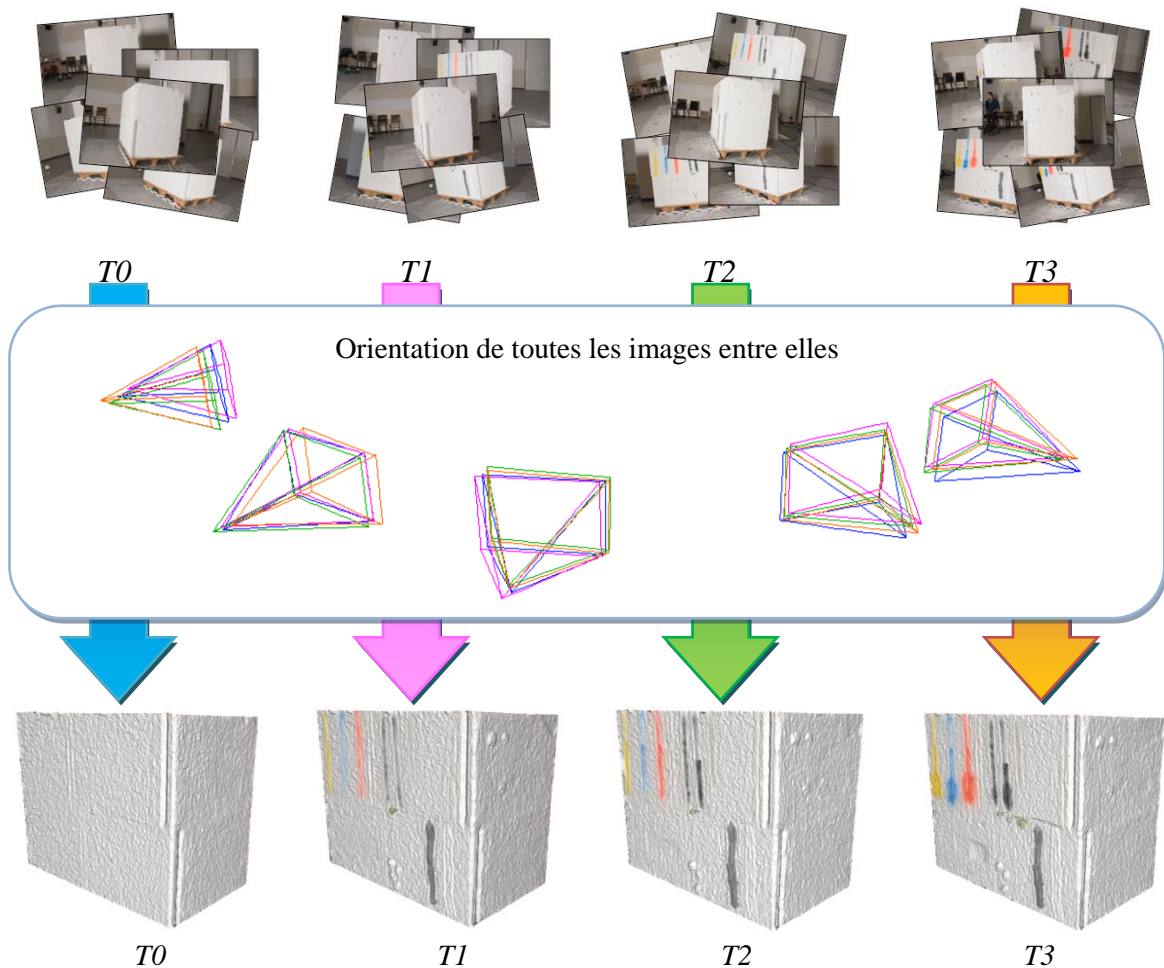


Figure 100 : Etapes du processus d'orientation multitemporelle

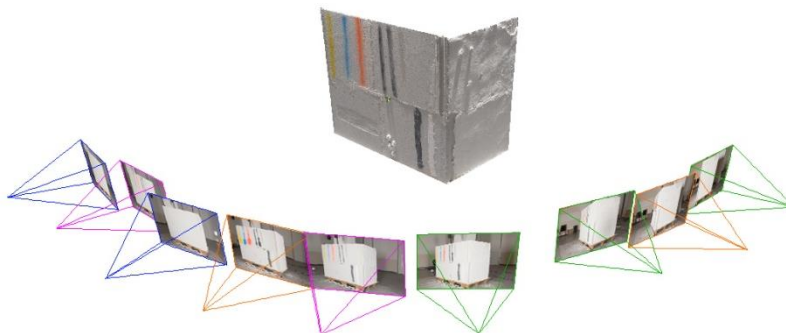


Figure 101 : Orientation d'images de différentes acquisitions autour du nuage de points d'un des temps

En partant de cette base, il est possible d'intégrer ces données à notre système. Ainsi, sur chaque image quel que soit son temps, il est possible de retrouver la correspondance 3D de chacun des nuages de points des différents temps. Chaque pixel des images possède donc autant de correspondances 3D qu'il y a de temps différents (Figure 102).


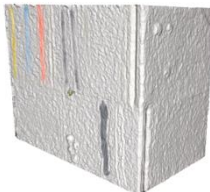
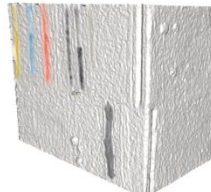
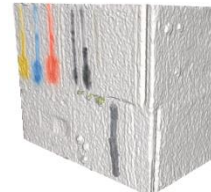











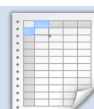



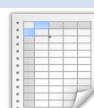
	   			
	Nuage de points T0	Nuage de points T1	Nuage de points T2	Nuage de points T3
Image T0	 Index NdpT0 sur Image T0	 Index NdpT1 sur Image T0	 Index NdpT2 sur Image T0	 Index NdpT3 sur Image T0
Image T1	 Index NdpT0 sur Image T1	 Index NdpT1 sur Image T1	 Index NdpT2 sur Image T1	 Index NdpT3 sur Image T1
Image T2	 Index NdpT0 sur Image T2	 Index NdpT1 sur Image T2	 Index NdpT2 sur Image T2	 Index NdpT3 sur Image T2
Image T3	 Index NdpT0 sur Image T3	 Index NdpT1 sur Image T3	 Index NdpT2 sur Image T3	 Index NdpT3 sur Image T3

Figure 102 : Correspondances 3D multi-temporelles

## VI. Processus hybride d'annotation d'images

Le procédé de mise en relation des entités tridimensionnelles avec des entités bidimensionnelles grâce à la spatialisation des informations, présenté au chapitre précédent, montre qu'il est possible de structurer des données 3D en fonction de données 2D quel que soit le type de relation projective.

Ce chapitre se concentre donc sur l'exploitation de cette relation pour permettre de définir un processus hybride d'annotation, fonctionnant grâce à l'interfaçage des données 3D et des données 2D.

### VI.A. Transfert d'annotation entre les images

Comme vu dans le [Chapitre III](#), l'annotation d'images ou de modèles 3D est aujourd'hui au cœur des systèmes de documentation, notamment en ce qui concerne les objets patrimoniaux. Les photographies (et plus généralement les sources iconographiques) offrent un support efficace pour entreprendre de nombreuses analyses sur les objets patrimoniaux.

Ces sources peuvent notamment être utilisées pour définir des annotations sémantiques. Ainsi des zones peuvent être dessinées sur les photographies puis liées à des descriptions sémantiques.

Cependant le relevé d'objets patrimoniaux suppose d'acquérir l'ensemble de l'objet sous différentes vues et à différents niveaux de détails. Ainsi le relevé complet d'un objet passe par l'acquisition de nombreuses images (des centaines voire des milliers sur des objets de grande taille). Les méthodes d'annotations les plus simples (annotations manuelles) supposeraient de devoir reporter chacune des annotations sur l'ensemble des images ce qui représenterait un travail très laborieux au vu de la quantité d'images. Les autres méthodes (automatiques ou semi-automatiques) reposent sur des méthodes d'analyses d'images qui sont difficiles à mettre en place sur des objets à forte complexité géométrique et surtout lorsqu'il s'agit d'annoter en fonction de différents objectifs d'analyse. Le travail d'annotation de l'expert est donc absolument nécessaire mais doit pouvoir être facilité à l'aide d'un système de transfert d'annotation automatique entre les images.

D'autre part, l'analyse d'objets patrimoniaux ne doit pas seulement se contenter d'informations bidimensionnelles. Une représentation tridimensionnelle de l'objet permet, en parallèle, d'obtenir des données géométriques de l'objet pouvant servir à extraire d'autres données d'analyse. Les méthodes d'annotations actuelles ne permettent pas de se limiter à un seul système pour établir à la fois les analyses sur les données bidimensionnelles et sur les données tridimensionnelles. Il est actuellement nécessaire d'établir les analyses sur les supports bidimensionnels et sur les supports tridimensionnels de manière indépendante et de recouper manuellement les informations extraites des deux représentations.

Notre approche cherche donc à établir un système d'annotation et d'analyse commun entre les représentations 2D et 3D par l'utilisation d'un procédé semi-automatique de transfert d'annotations entre les images et avec le nuage de points afin d'assurer le continuum



d'informations. La relation mise en place au chapitre précédent peut servir de base pour établir ce transfert.

### VI.A.1. Principe de la méthode

A partir d'un ensemble de photographies 2D orientées dont la correspondance avec un nuage de points 3D a été établie, le transfert semi-automatique d'annotation entre les images peut être établi. Cette relation permet de lier les informations 2D et les informations 3D de façon bijective. Ainsi la relation permet dans un sens de passer de la 2D vers la 3D, et dans l'autre sens de passer de la 3D vers la 2D. L'idée derrière le système de transfert est d'exploiter cette relation dans les deux sens afin de reconnaître une région dessinée sur une image sur les autres images de l'ensemble. Ainsi, grâce à cette relation, le transfert entre les images peut être envisagé.

L'approche développée établit donc le transfert en trois étapes :

- Définition de l'annotation sur une des images : (a) Figure 103
- Recherche des coordonnées de la zone par la relation 2D->3D : (b) Figure 103
- Projection sur les autres images par la relation 3D->2D : (c) Figure 103

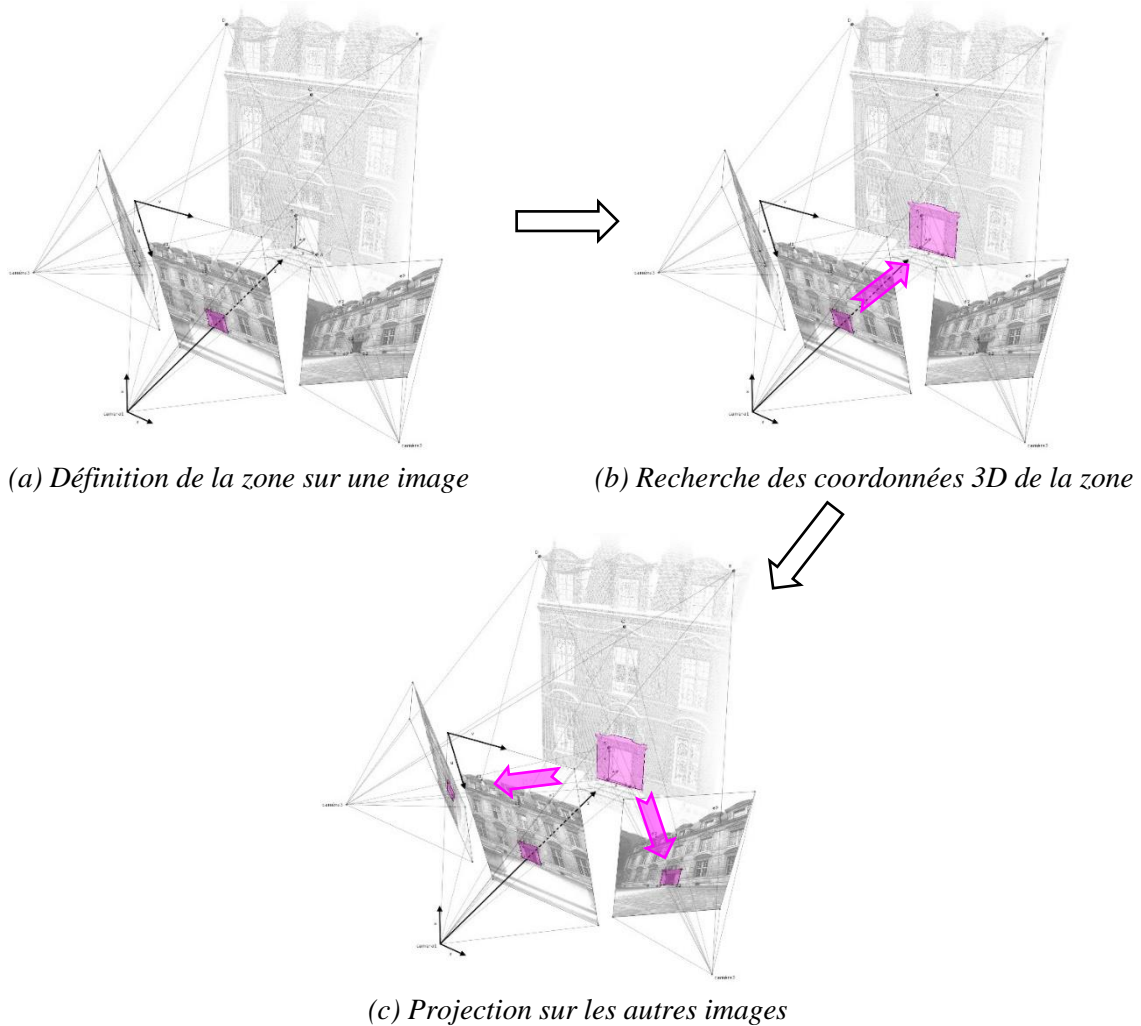
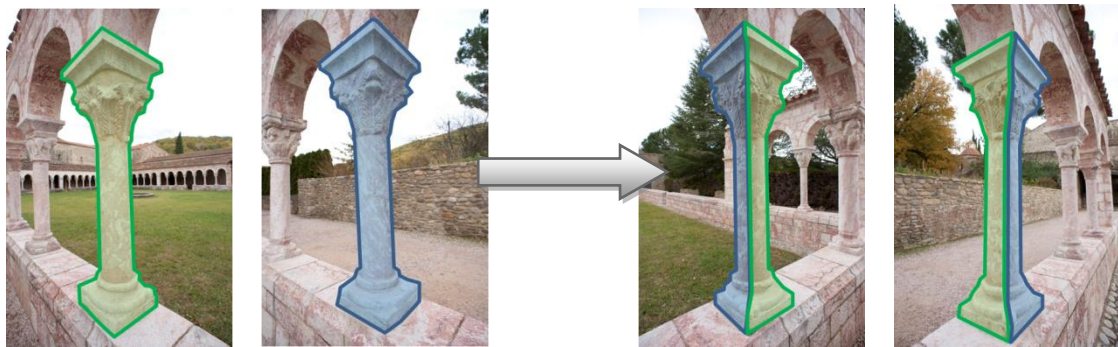


Figure 103 : Etapes du transfert d'une annotation

Chaque annotation est ainsi définie par un nuage de point et un ensemble de projections 2D sur chacune des images. Le transfert établit donc à la fois une projection sur les autres images

mais aussi une extraction du nuage de points représentant la zone définie par l'annotation. L'utilisation d'un simple nuage de points suffit à extraire par la suite d'autres données d'analyses (voir [Chapitre VIII](#)) et permet de gérer le continuum d'informations là où une représentation géométrique maillée serait plus difficile à maintenir à jour.

Outre le fait de pouvoir transférer les annotations d'une image à l'autre, la définition d'une annotation doit pouvoir utiliser plusieurs vues différentes. En effet, une annotation peut avoir besoin de plusieurs vues pour être définie. Cette situation peut se retrouver dans le cas d'une région à annoter qui peut se trouver derrière un obstacle dans des zones difficiles d'accès ou dans le cas d'une région à annoter qui nécessite de pouvoir tourner autour (par exemple, si l'on veut sélectionner une colonne) (Figure 104). Le fait de pouvoir compléter une annotation à partir d'une autre image peut aussi simplement être le fait de la stratégie de description sans forcément avoir de problèmes de visibilité.



*Sélection sous deux vues*

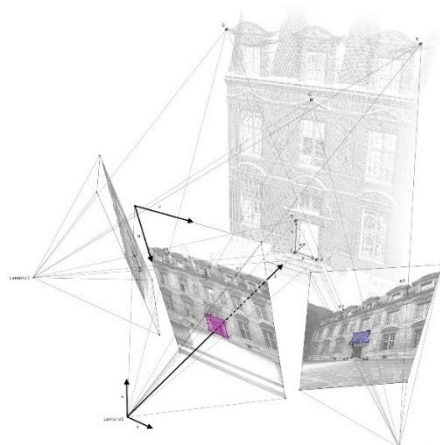
*Sélection complète de la colonne sur les autres vues*

*Figure 104 : Sélection d'une colonne complète à l'aide d'images*

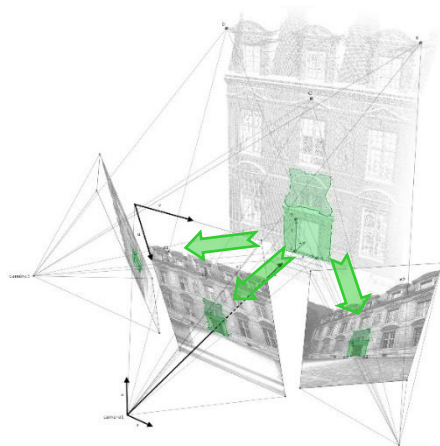
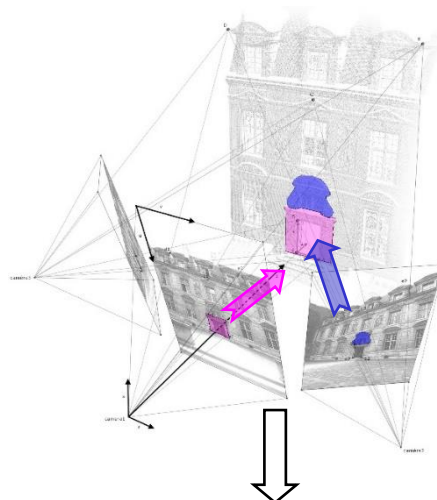
Le problème est donc de pouvoir permettre l'utilisation de plusieurs vues pour définir une annotation. Les étapes définies précédemment peuvent être adaptée à ces cas (Figure 105). Ainsi le principe de l'annotation multi-vue repose sur la définition de plusieurs zones sur plusieurs images. Chacune de ces annotations permet d'extraire un ensemble de coordonnées 3D. C'est l'ensemble de coordonnées 3D de chaque zone dessinée qui sera par la suite projetée sur toutes les images.



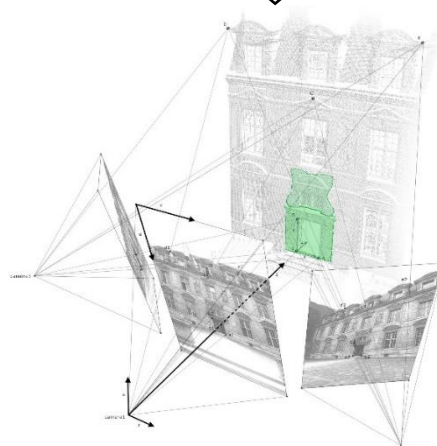
(a) Définition de l'annotation sur deux images



(b) Recherche des coordonnées 3D de chacune des zones



(d) Projection sur toutes les images



(c) Regroupement de l'ensemble des coordonnées 3D

Figure 105 : Définition d'une annotation à partir de plusieurs vues

Ainsi grâce à la relation 2D/3D, le transfert entre les images est possible. Le transfert peut donc s'effectuer dans les deux sens pour une image : **à partir de** cette image (définition d'une annotation) ou **vers** cette image (transfert à partir d'une autre image). Toutes les images peuvent être choisies pour définir l'annotation ou pour recevoir les résultats du transfert.

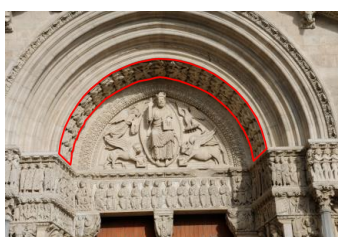
Les détails de chacune des étapes sont détaillés dans les paragraphes suivants.

#### VI.A.2. Définition de l'annotation

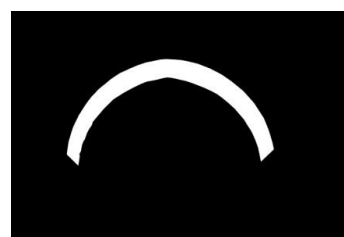
Notre méthode démarre de l'identification d'une région. A partir de la zone tracée sur une image, un masque de même dimension que l'image est extrait (Figure 106).



Image



Sélection sur l'image



Masque extrait

Figure 106 : Définition de l'annotation et masque extrait

Ce masque permet de localiser dans l'image la zone à annoter et sert pour les étapes suivantes du processus de propagation.

### VI.A.3. Recherche des coordonnées 3D

Comme vu dans le [Chapitre V](#), les pixels de l'image sont liés à des coordonnées 3D du nuage de points. Le masque obtenu à l'étape précédente donne des positions de pixels de l'image par un codage couleur (noir ou blanc). Ainsi connaissant la position des pixels blancs du masque représentant l'annotation, il est possible d'extraire les coordonnées 3D de la zone annotée par lecture des données liées à l'image que ce soit à l'aide d'un fichier d'indexation ou d'un fichier XYZ.

Néanmoins il faut prévoir que l'extraction du masque noir et blanc ait pu être faite à partir d'une image dont les dimensions ne correspondent pas à la dimension choisie pour le stockage des informations 3D liées à l'image c'est-à-dire par rapport au redimensionnement nécessaire pour obtenir une bonne densité de correspondance 3D (voir [Chapitre V](#)). Le masque doit donc tout d'abord être redimensionné afin qu'il corresponde aux dimensions utilisées pour le stockage. Ce redimensionnement ne pose pas de problème lorsque la région dessinée est assez grande. Dans le cas où la sélection est linéaire ou ponctuelle, un redimensionnement du masque peut supprimer le peu de pixels blancs contenus dans le masque. Dans ces cas, il faut que l'extraction du masque se fasse à partir de l'image aux dimensions de stockage des informations 3D afin d'éviter le redimensionnement du masque. Ainsi, la meilleure pratique pour la sélection est d'extraire le masque directement à partir d'une image aux dimensions de stockage des informations 3D. Cependant cela oblige à redimensionner l'image originale et par conséquent, à perdre en qualité d'image.

A partir du masque aux dimensions de stockage des informations 3D de l'image, les positions  $i$  et  $j$  (lignes et colonnes) des pixels blancs sont extraites (Figure 107).

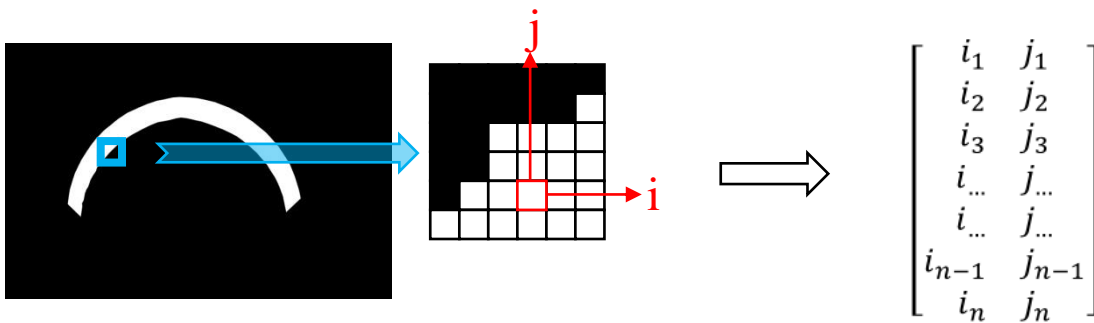


Figure 107 : Extraction des positions des pixels blancs

Avec cette liste de positions de pixels, les fichiers stockant les informations 3D liées aux images (indexation ou fichier XYZ) (voir [V.C. Gestion de données](#)) et organisés dans le même ordre que les images, sont lus à ces mêmes positions. Ainsi dans le cas de l'indexation, une liste d'index est extraite et dans le cas de fichier XYZ, une liste de triplets de coordonnées 3D est extraites (Figure 108).

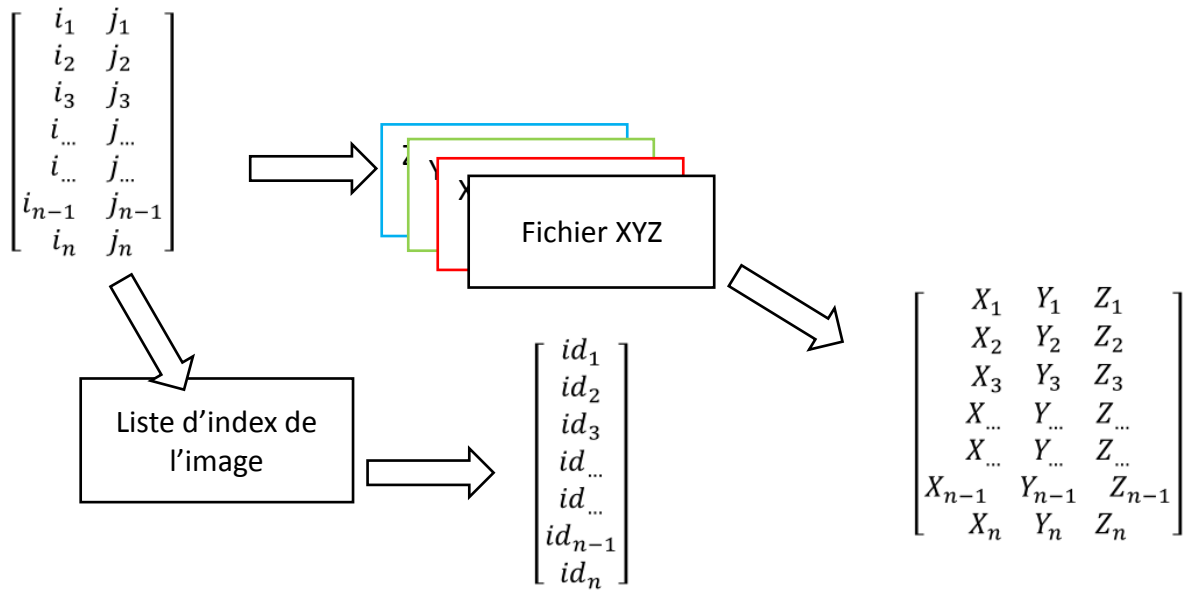


Figure 108 : Extraction des données 3D en fonction du type de stockage

Ces informations représentant des données 3D, l'annotation peut donc être visualisée par un nuage de points. A partir du fichier XYZ, les coordonnées 3D sont directement obtenues. Dans le cas d'utilisation d'indexation du nuage de points, il suffit de lire les coordonnées 3D dans le fichier d'indexation du nuage de points aux index extraits de l'annotation. Ainsi l'annotation peut être visualisée en 3D (Figure 109).

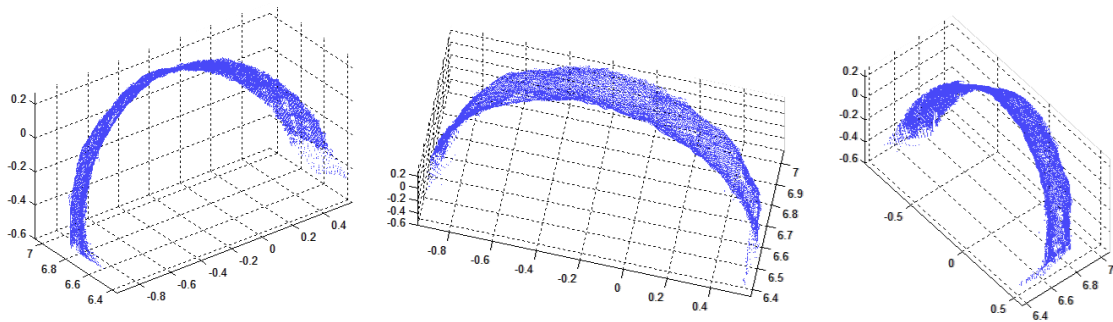


Figure 109 : Représentation 3D de l'annotation extraite à partir de la sélection sur une image

Afin d'optimiser et de simplifier les calculs, il est possible d'envisager de n'extraire qu'un échantillonnage des points de l'annotation. Cet échantillonnage ne contient que les pixels du contour et un quadrillage des points internes (Figure 110). L'utilisation d'un tel échantillonnage permet d'alléger les informations stockées pour une annotation mais pose quelques problèmes de précision au niveau de la projection.

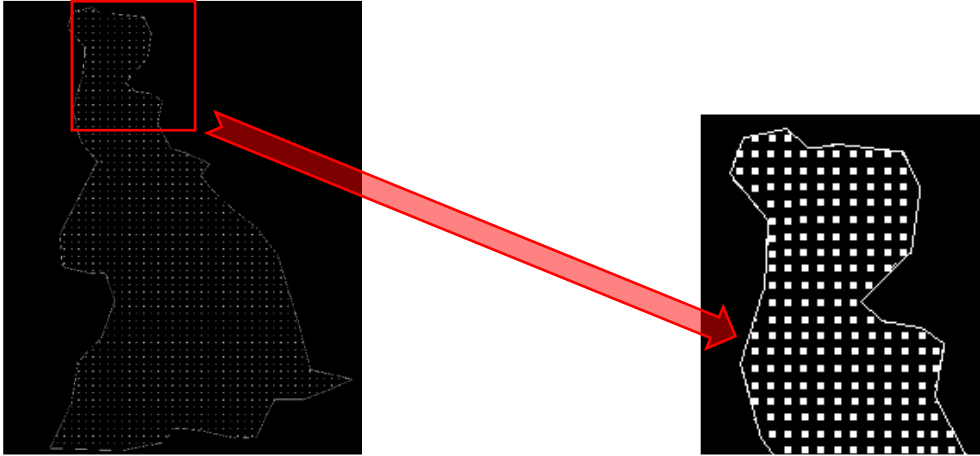


Figure 110 : Echantillonnage des pixels dans le masque

#### VI.A.4. Transfert sur les autres images

A partir des listes de données 3D extraites de l'annotation, l'étape suivante consiste à rechercher ces mêmes informations dans les fichiers de données 3D des autres images. Lors de cette recherche, les positions  $i$  et  $j$  (ligne et colonne) des triplets de coordonnées correspondants ou des index correspondants sont extraites dans chacun des fichiers de données 3D des autres images. Un masque dont les dimensions sont celles utilisées pour les images lors de la projection des données 3D est construit où les pixels blancs sont situés aux positions  $i$  et  $j$  trouvées (Figure 111). Ce masque est donc à la dimension de stockage des informations 3D pour l'image.

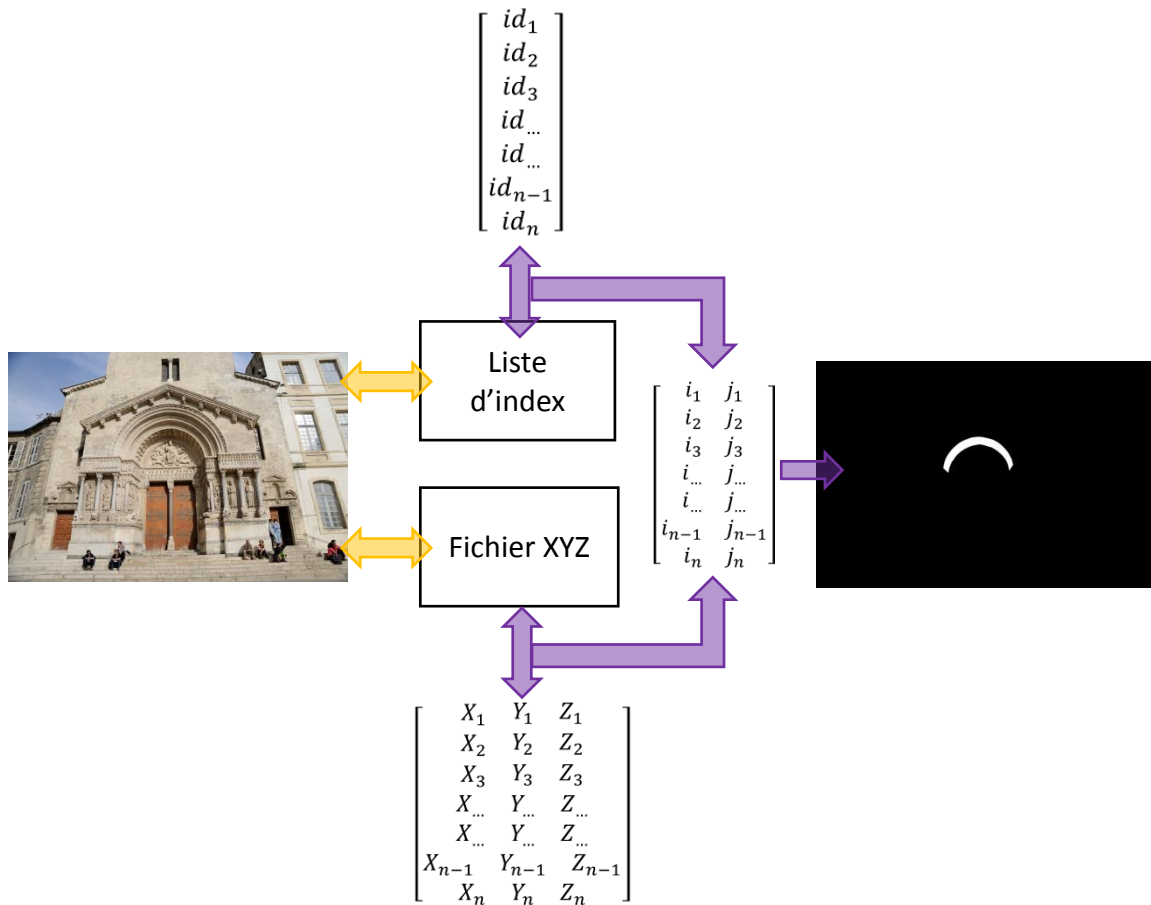


Figure 111 : Etape de projection vers les autres images

Dans le cas de l'utilisation des fichiers XYZ, notamment s'ils sont extraits à partir de données laser, il faut envisager de rechercher les autres triplets de coordonnées correspondants en intégrant une approximation sur les coordonnées. En effet, dans ce cas, il n'y pas l'assurance qu'un même point ait exactement la même coordonnée d'une image à l'autre. Ainsi, la comparaison ne se fait pas sur les coordonnées exactes mais sur un arrondi de ces coordonnées. L'approximation sur les coordonnées permet ainsi de trouver plus facilement des correspondances entre les deux listes de coordonnées et donc permet de trouver les points approximativement communs. Sans cette approximation, le système aurait peu de chances de trouver suffisamment de correspondances exactes entre les deux listes de coordonnées. Cependant cette approximation peut rajouter des erreurs au niveau de la recherche des correspondances.

L'intérêt de l'utilisation d'une indexation pour le nuage de points face à l'utilisation des fichiers XYZ apparaît. Là où les fichiers XYZ nécessitent une approximation et une comparaison sur des triplets de valeurs pour retrouver les correspondances dans les autres images, l'indexation ne nécessite qu'une comparaison de valeurs uniques, accélérant ainsi le processus de comparaison.

La superposition du masque obtenu et de l'image permet de visualiser la zone détectée (Figure 112). Cette étape doit être réévaluée pour chacune des images de l'ensemble. L'intérêt de cette approche est qu'elle permet de gérer les problèmes d'occlusions. En effet, lors de la comparaison des index de l'annotation (ou des coordonnées 3D) avec ceux de l'image, l'existence d'une correspondance entre un des index de l'annotation et un des index de l'image permet d'affirmer que le point existe dans l'image. A l'inverse, si cette correspondance n'existe pas, le point n'apparaît pas dans l'image.

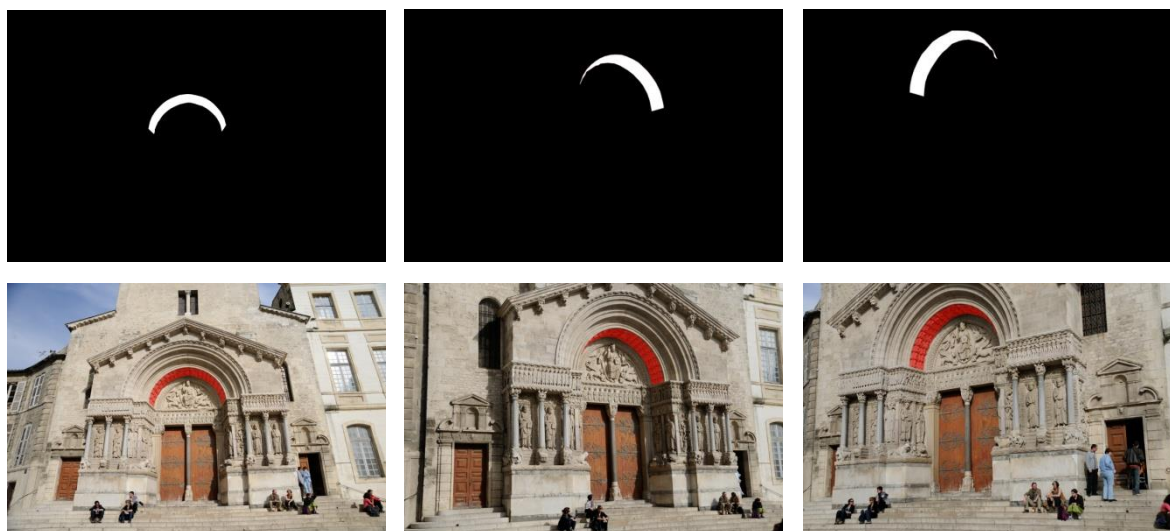


Figure 112 : Résultat de projection : masques obtenus et superposition avec l'image associée

Dans le cas de l'utilisation d'une indexation, afin d'éviter d'avoir à rechercher sur toutes les images et donc de devoir construire un masque entièrement noir par comparaison pour les images où la zone n'apparaît pas, l'utilisation de la liste des images liée à chacun des points 3D va alléger la recherche. En effet, dans le [Chapitre IV](#), nous avons montré que, lors de la projection des points 3D sur les images pour obtenir la correspondance entre pixels et



coordonnées 3D, il était possible de stocker pour chacun des points 3D la liste des images sur lesquelles le point apparaît. Ainsi, lors de la propagation, à partir de la liste des index de l'annotation, la liste des images sur lesquelles la recherche doit être faite peut être extraite par concaténation des listes d'appartenance aux images de chacun des points de l'annotation.

Dans le cas de l'utilisation d'un échantillonnage des pixels, le résultat de propagation est moins précis. En effet, étant donné que le nombre de points définissant l'annotation est moins important, la projection sera moins précise et comportera de nombreux espaces vides. Pour pouvoir combler ces espaces, une étape de correction du masque est nécessaire par une opération de dilatation suivie par une opération d'érosion (Figure 113).

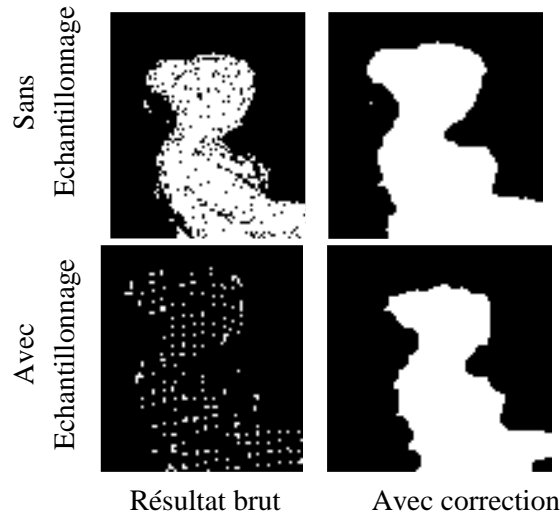


Figure 113 : Résultats du transfert avec ou sans échantillonnage, et avec ou sans correction

Ainsi l'échantillonnage offre certes des données plus légères en termes de stockage mais les résultats de projection sont beaucoup moins précis sur les images. De plus, l'étape de correction nécessite la définition de paramètres précis pour la dilatation et l'érosion. Cependant, l'échantillonnage peut être progressif afin de donner des aperçus de résultats intermédiaires.

Le processus de propagation d'annotations, ainsi défini, au sein d'un ensemble d'images spatialement orientées autour d'un nuage de points permet donc d'annoter toutes les images en même temps à partir d'une sélection d'une zone sur une ou plusieurs images.

## VI.B. Transfert multi temporel

Comme vu dans en [V.D.4. Ajout de différents états temporels](#), il est possible de créer plusieurs nuages de points pour plusieurs relevés complets à différent temps. Les images de tous les temps étant toutes orientées entre elles et donc autour de chacun de nuages de points (Figure 114), une image peut obtenir des correspondances 3D pour chacun des temps. Ainsi, à partir de l'établissement de ces correspondances, les annotations peuvent donc, en plus de pouvoir être propagées entre des images de même temps, être propagées au sein d'un ensemble d'images de différents temps.



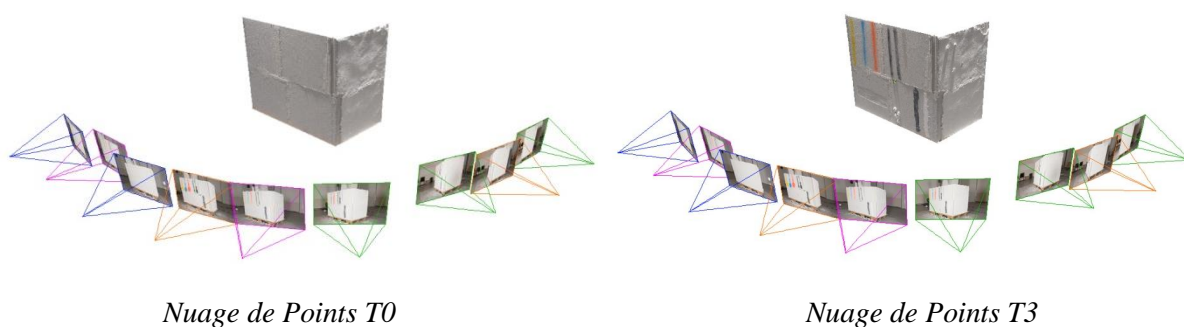


Figure 114 : Orientation d'images de différentes acquisitions autour des nuages de points à différents temps

Le fait de pouvoir lier les différents temps de cette manière offre la possibilité de pouvoir comparer des annotations faites sur des images à différents temps. Le principe reste le même. Cependant, il faut que la propagation se fasse selon un seul nuage de points de référence (Figure 115).

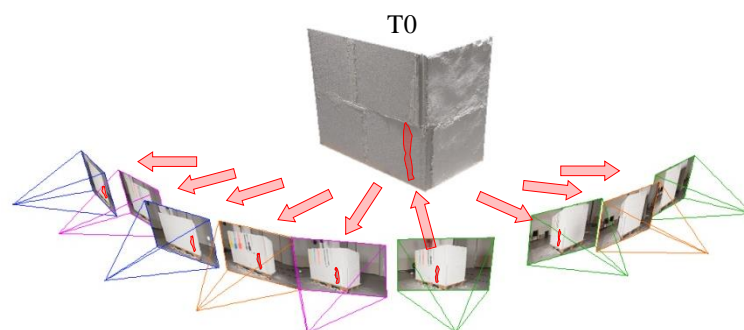


Figure 115 : Propagation multitemporelle

Le choix du temps de visualisation est donc possible. Cela peut par exemple permettre de visualiser l'évolution d'altérations au cours du temps. En effet, une altération peut avoir été annotée sur un temps et être propagée sur une image du temps initial où elle n'apparaît pas encore ou alors être comparée à son homologue d'un autre temps (Figure 116). Ainsi notre approche permet de visualiser la projection d'un état temporel sur un autre état temporel.

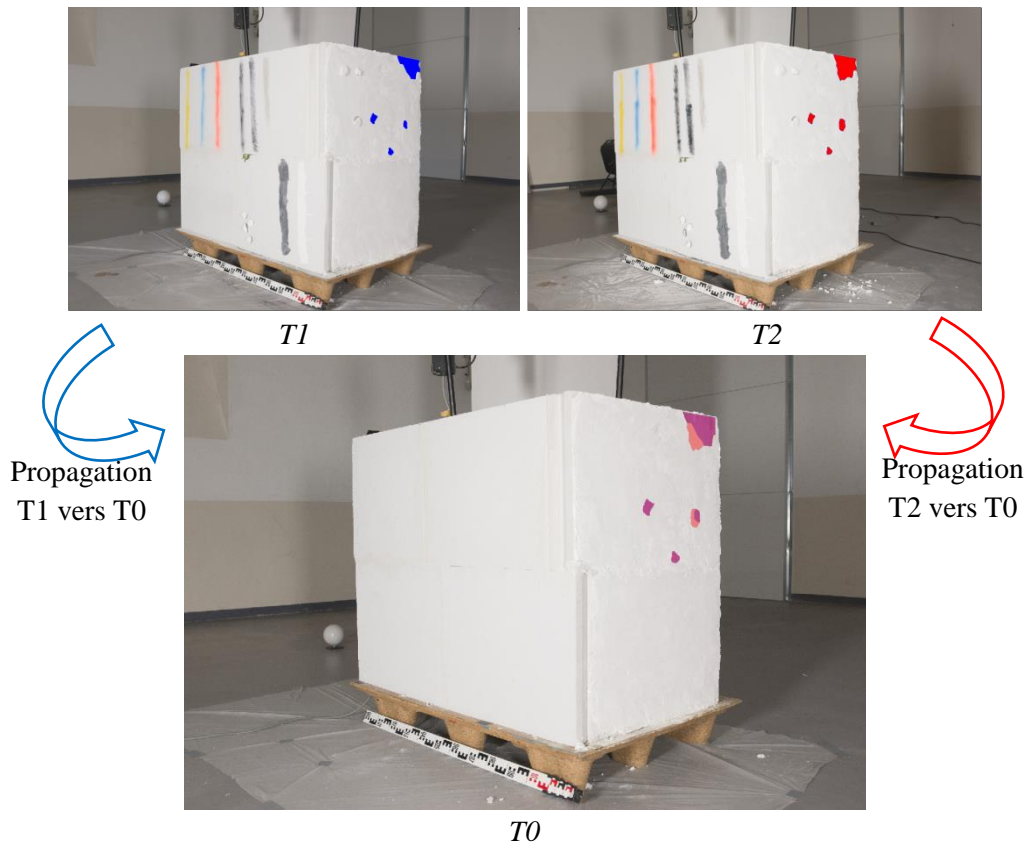


Figure 116 : Comparaison d'une altération sélectionnée au temps  $T1$  (bleu) et au temps  $T2$  (rouge) sur le temps  $T0$  (image) à l'aide de la propagation multi-temporelle (les images n'ont pas exactement la même orientation)

Pour pouvoir propager entre les différents temps, un temps de référence doit être choisi pour servir de support pour la propagation. Cependant, le choix de ce temps reste pour l'instant arbitraire. Suivant le temps de référence choisi les résultats de propagation peuvent varier. Ceci dépend des changements au niveau de la forme entre les temps. Pour éviter ces variations, une solution pourrait être d'utiliser deux propagations en parallèle en utilisant une fois le temps de l'image d'annotation et une seconde fois le temps de l'image de destination (Figure 117). Ainsi deux résultats sont obtenus et peuvent être combinés ou visualisés en parallèle pour donner le résultat final.

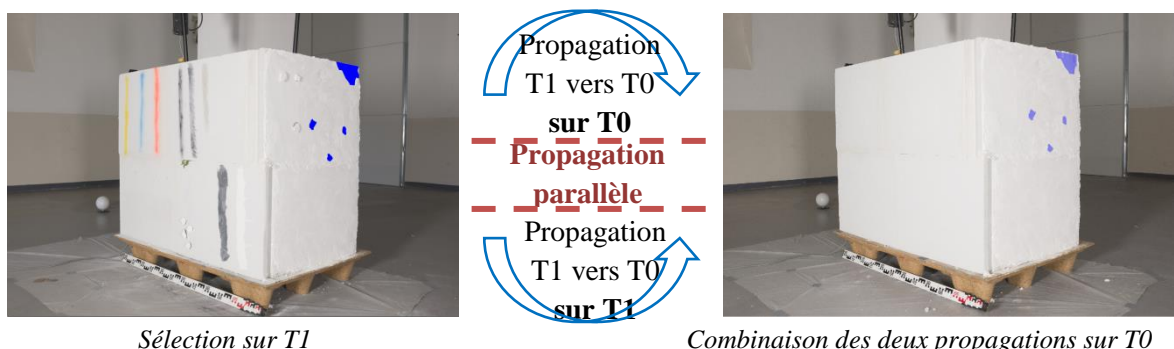


Figure 117 : Ajustement du résultat en utilisant deux temps comme intermédiaire

Le développement d'un tel système multi-temporel pose le problème de la gestion des données liées à ces différents temps. Ainsi au niveau de l'indexation des nuages de points, l'intégration d'un nouveau temps oblige à envisager une réindexation des correspondances 3D,

le nouveau nuage étant donc indexé à la suite des autres. Ainsi un état temporel est représenté dans l'espace par une liste d'index.

Cette partie de gestion de différents états temporels nécessite encore d'être étoffée. Nos premiers tests ont pu néanmoins montré qu'il était possible de lier les différents temps des images à l'aide du lien 2D/3D. Il reste néanmoins à définir de façon efficace la manière de gérer les données et la visualisation possible pour l'utilisateur. Une autre perspective intéressante serait d'utiliser des images de deux temps différents lors de la corrélation afin de détecter les différences. En effet, dans ce cas, les zones ayant subi des modifications entre les deux temps ne devrait pas pouvoir être corrélées.

## VII. Outils d'annotations et d'extractions d'informations

Le chapitre précédent a présenté la méthode permettant le transfert automatique d'annotations au sein de l'ensemble des images spatialement orientées dont on connaît la correspondance 2D/3D. Cette méthode combine donc, pour chacune des annotations, des informations bidimensionnelles liées aux pixels des images et des informations tridimensionnelles liées aux coordonnées 3D correspondantes.

A partir de ce point de départ, l'enjeu est de pouvoir définir des descripteurs pertinents liés à l'image ou à la représentation 3D afin d'aider l'annotation. Comme vu dans le [Chapitre V](#), les images utilisées peuvent contenir différentes informations sur les différentes réponses aux rayonnements (visible, infrarouge, ultraviolet, température, ...) (Figure 118). Le nuage de points, élaboré par photogrammétrie, contient des informations sur les coordonnées 3D et les normales qui peuvent servir, d'autre part, à générer des informations sur la courbure, l'occlusion ambiante ou encore la rugosité (Figure 118). Grâce au lien 2D/3D, toutes ces informations peuvent être exprimées par rapport aux images ou au nuage de points 3D, même si elles ont été générées à partir d'un seul support. Le système permet ainsi de disposer (par traitement d'images ou géométrique) de plusieurs couches additionnelles que nous avons identifiées : surfaces (coordonnées, normales, courbure), apparence visuelle (rayonnements), caractérisation géométrique (occlusion ambiante, rugosité) et caractérisation spatiale (extension et orientation globale).

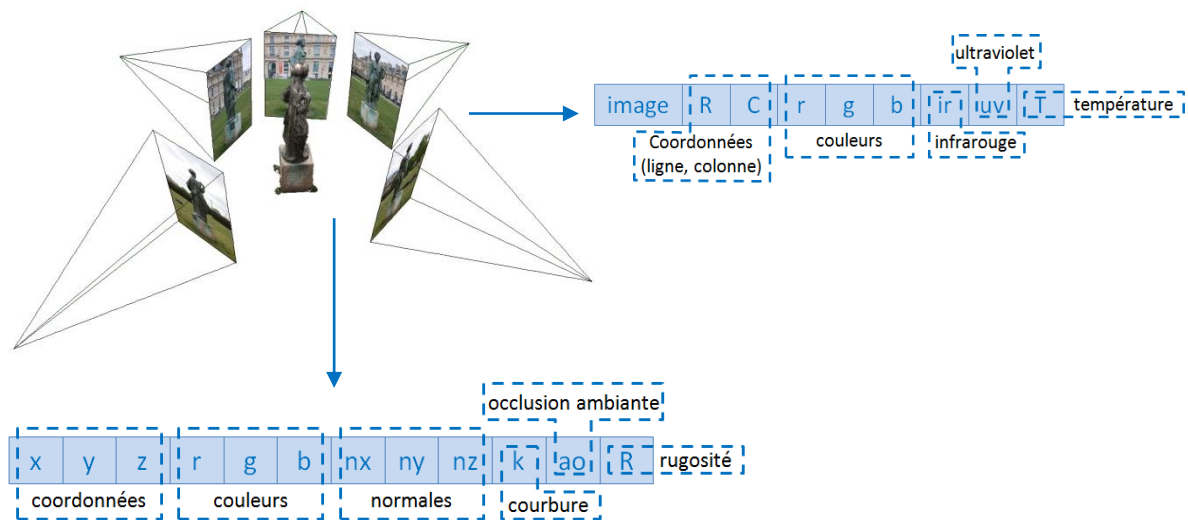


Figure 118 : Informations pouvant être liées aux points du nuage et aux pixels des images

A partir de ces données et des informations que l'on peut extraire, un certain nombre d'outils peuvent être développés pour aider à la propagation des annotations ou pour guider l'utilisateur dans son analyse. Les images d'une part peuvent être traitées par diverses méthodes de segmentation afin de faire ressortir des informations spécifiques à un type d'analyse. D'autre part, les annotations étant aussi représentées par un nuage de points, des analyses tridimensionnelles peuvent être lancées sur ce nuage et permettre d'obtenir d'autres informations sur l'annotation.

Ce chapitre se concentre donc sur une présentation des méthodes de segmentation d'images et sur la présentation de divers outils bidimensionnels et tridimensionnels cherchant à aider l'utilisateur et à pousser l'analyse à un plus haut point.

## VII.A. Segmentation d'images

Si l'homme sait naturellement séparer des objets dans une image grâce à des connaissances de haut niveau (compréhension des objets et de la scène), mettre en place des algorithmes permettant cette même interprétation reste encore aujourd'hui un des thèmes les plus courants en traitements d'images.

Le but de cette segmentation d'images est de rassembler les pixels entre eux suivants des critères prédéfinis. Les pixels ainsi regroupés forment des régions devant définir des éléments de la scène. A ce jour, de nombreuses méthodes de segmentation existent et sont généralement regroupées en quatre classes :

- Segmentation sur les régions (*region-based*)
- Segmentation sur les contours (*edge-based*)
- Segmentation par seuillage (*threshold*)
- Segmentation combinant ces trois segmentations

[Vantaram S.R. et al, 2012] expose une liste relativement exhaustive des méthodes actuelles de segmentation d'images. Il propose une classification de ces méthodes de deux façons : suivant une taxonomie de haut niveau, ou suivant une taxonomie de bas niveau.

La taxonomie de haut niveau propose une classification selon cinq critères (Figure 119) : le type d'image (couleur ou noir et blanc), l'interaction humaine (supervisée ou non), la représentation de l'image (nombre de résolutions), les attributs de l'image (un ou plusieurs) et le principe des opérations (organisation spatiale des pixels connue ou non).

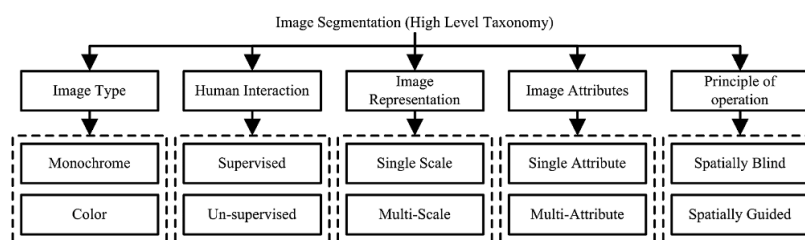


Figure 119 : Taxonomie de haut niveau des méthodes de segmentation d'images selon [Vantaram S.R. et al, 2012]

Le dernier critère reposant sur le principe des opérations sert à établir la taxonomie de bas niveau. Ainsi les méthodes de segmentation sont tout d'abord classifiées en fonction de la connaissance de l'organisation des pixels (Figure 120).

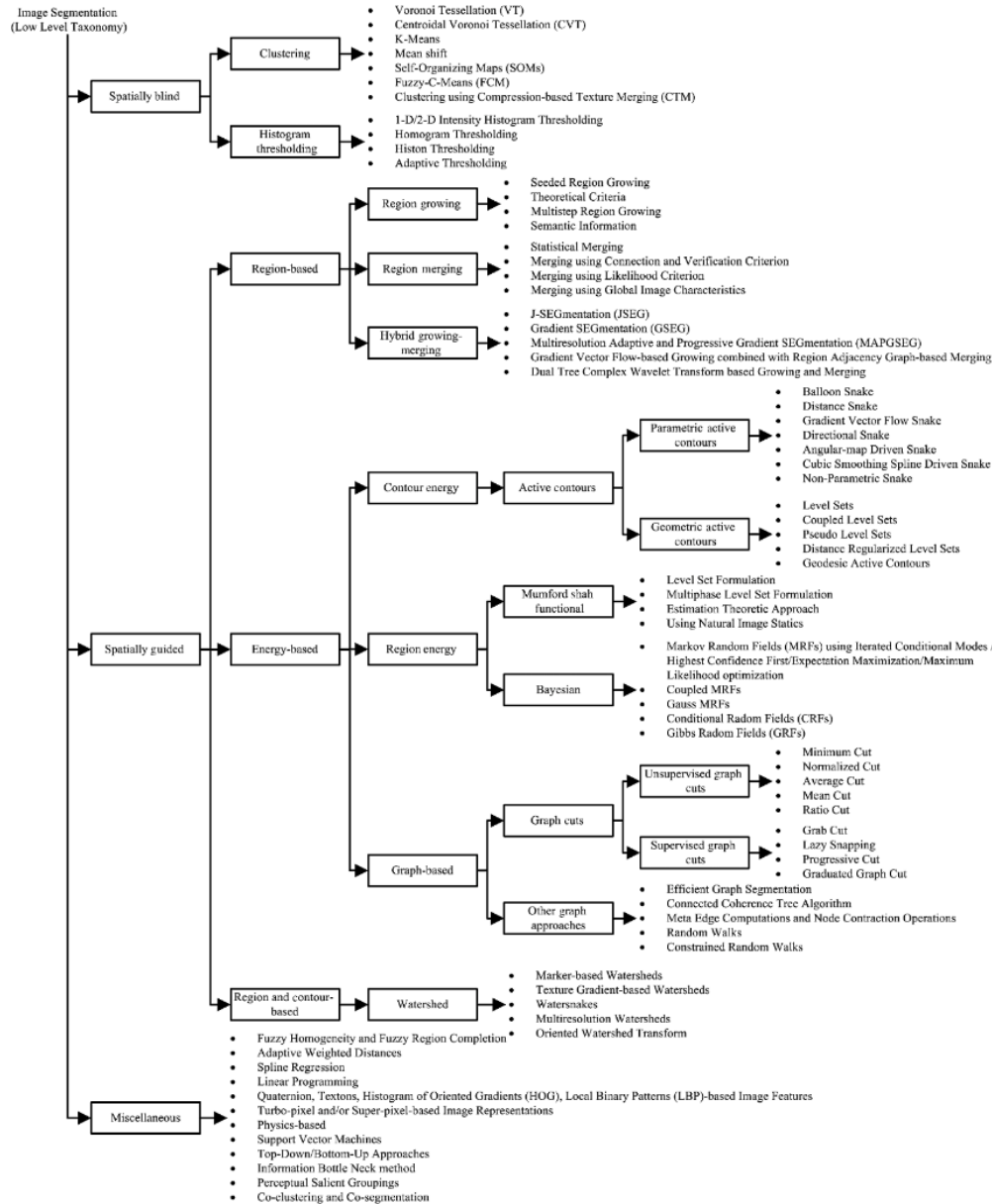


Figure 120 : Taxonomie de haut niveau des méthodes de segmentation d'images selon [Vantaram S.R. et al, 2012]

La majorité des méthodes de segmentation peut donc être catégorisée en fonction de la connaissance de l'organisation spatiale des pixels. Cependant il existe certaines techniques qui ne peuvent pas être classifiées dans une de ces parties et sont donc répertoriées à part mais donnent tout de même d'importants résultats en termes de méthodes de segmentation.

Les méthodes de segmentation d'images étant nombreuses, nous nous contenterons de présenter un échantillon des méthodes pour chacune des catégories.

### VII.A.1. Méthodes sans connaissance de l'organisation des pixels

Parmi les méthodes sans connaissances de l'organisation des pixels apparaissent les algorithmes de regroupement (*clustering*) et les algorithmes par seuillage (*threshold*).

Les méthodes par regroupement voient l'image comme un nuage de points à une dimension pour des images en niveaux de gris ou à trois dimensions pour des images couleurs



(une dimension par couleur) (Figure 121). Le principe de ces méthodes est d'analyser le nuage de point 1D ou 3D en utilisant des fonctions métriques pour identifier des regroupements significatifs de pixels.

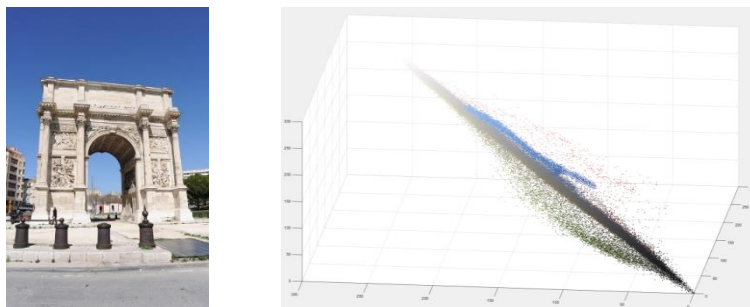


Figure 121 : Nuage de points représentant une image couleur dans l'espace des couleurs RGB

Parmi ces méthodes, la décomposition par diagramme de Voronoï ([Arbeláez P.A. et al, 2006] [Wang J. et al, 2009]) est une procédure qui décompose l'espace en groupes ou clusters en utilisant un ensemble fixe de points appelés germes.

Les méthodes K-Means ([McQueen J., 1967] [Kanungo T et al, 2002] [Chen J. et al, 2005]) reposent sur le problème de division des points en k partitions ou clusters de façon à minimiser la somme des carrées des distances.

Les méthodes Mean-Shift [Comaniciu D. et al, 2002] sont des méthodes non-paramétriques qui facilitent l'analyse d'espaces multidimensionnels en utilisant des clusters de formes arbitraires, basé sur le concept de mean-shift proposé à l'origine par [Fununaga K. et al, 1975].

Les méthodes par seuillage reposent sur des informations calculées sur toute l'image comme par exemple la moyenne des niveaux de gris permettant de construire des classes d'intensités. La construction d'histogrammes permet d'identifier les pics, les vallées et les formes donnant des informations sur l'ensemble des pixels. [Sezgin M. et al, 2004] propose une vue d'ensemble de méthodes de seuillage par histogramme.

Les méthodes, n'utilisant aucune connaissance sur les images, ont l'avantage d'être simples à mettre en place. Cependant les résultats offerts restent relativement imprécis et ne peuvent pas s'appliquer à des segmentations nécessitant un haut niveau de description.

### VII.A.2. Méthodes avec connaissance de l'organisation des pixels

Contrairement aux méthodes sans connaissance de l'organisation des pixels, les méthodes avec connaissance de l'organisation des pixels intègre l'aspect spatial de la disposition des pixels dans l'image. L'objectif est de former des groupes de pixels homogènes d'un point de vue spatial quel que soit leur relation sur des caractéristiques spécifiques. L'utilisation des informations de régions et de contours sont des méthodes courantes pour la formation de régions.

#### VII.A.2.a. Méthodes basées sur les régions

Les méthodes de segmentation qui utilisent les régions emploient des protocoles qui engagent des techniques de croissance, de séparation et de regroupement.

Les méthodes par croissance de régions ([Fan J. et al, 2001] [Fondon I. et al, 2006]) sont des procédés qui partent d'un ensemble de pixels prédéfinis (nommé *seed*) et accumulent d'autres pixels autour de chacun d'entre eux en cherchant à respecter un critère d'homogénéité (Figure 122). La croissance d'une région s'arrête lorsque les pixels autour ne satisfont plus le critère. Ces méthodes produisent généralement des sur-segmentations des images.

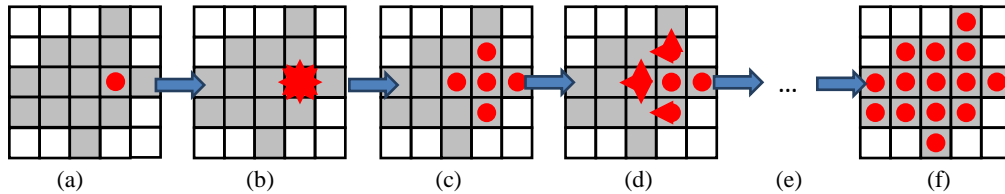


Figure 122 : Principe de la croissance de région : (a) en gris les pixels respectant le critère, en rouge le pixel *seed*, (b) les pixels autour sont testés sur le critère, (c) les pixels autour respectant le critère sont ajoutés à la région, (d) et (e) le procédé est répété jusqu'à ce que tous les pixels autour ne respectent plus le critère, (f) région finale

Les méthodes par regroupement ([Devaux J.C. et al, 2001], [Chen H.C. et al, 2004], [Kim C., 2005], [Liu Z. et al, 2011]) cherchent à grouper des sous-régions entre elles afin d'obtenir un nombre plus restreint de régions qui respecte l'organisation spatiale de l'image d'entrée.

D'autres travaux emploient des méthodes hybrides combinant les méthodes par croissance de régions et les méthodes par regroupement ([Deng Y. et al, 2001] [Wang Y. et al, 2006] [Ugarizza L.G. et al, 2009] [Vantaram S.R., et al, 2010]).

#### VII.A.2.b. Méthodes basées sur l'énergie

Les méthodes basées sur l'énergie reposent sur des techniques cherchant à minimiser une fonction de vraisemblance ou énergie. Elles peuvent être généralement classifiées entre celles qui utilisent les contours (contours actifs) et celle qui emploient les régions.

Les méthodes par contours actifs peuvent être classifiées en fonction de leur implémentation sur deux catégories. Les contours actifs paramétriques (*PAC : parametric active contours*) sont généralement représentées par des formulations lagrangiennes où les courbes évolutives sont appelées serpent (*snake*) ([Kass M. et al, 1988] [Dumitras A. et al, 2001] [Precioso F et al, 2005]). Les contours actifs géométriques (*GAC : geometric active contours*) sont représentés par des formulations euclidiennes où les courbes évolutives sont évaluées à partir d'une fonction de distance en deux dimensions ([Malladi R. et al, 1995] [Michailovich O. et al, 2007]).

Le domaine de la théorie des probabilités a permis d'introduire une classe de segmentation où les caractéristiques des régions peuvent être modélisées grâce aux champs aléatoires de Markov (*MRF : Markov Random Field*) ([Mukherjee J, 2002] [Gao J. et al, 2002]), à l'aide de champs aléatoires conditionnels (*CRF : Conditionnal Random Field*) ([Zhang L. et al, 2010] [Lee S.H., et al, 2010] ou encore avec des modèles de mélanges gaussiens (*GMM : Gaussian Mixture Models*) ([Carson C. et al, 2002] [Khan J.F. et al, 2009]).

Dans la formulation théorique des graphiques, une image peut être représentée comme un graphique représentant les pixels, leur connectivité de voisinage et leur similarité. Un critère populaire qui induit l'identification d'un ensemble de connections parcourant un chemin

spécifique dans un graphique est généralement appelé un *cut*. [Wu Z. et al, 1993] furent les premiers à développer un approche basée sur le critère du minimum de *cut*. L'introduction du standard Ncut (*Normalized Cut*) a permis le développement de nombreuses approches de segmentation ([Malik J. et al, 2001] [Tao W. et al, 2007] [Sarkar S. et al, 2000] [Boykov Y.Y. et al, 2006]).

#### VII.A.2.c. Méthodes par lignes de partage des eaux (Watershed)

Les méthodes de segmentation par ligne de partage des eaux sont des protocoles qui considèrent les images comme des reliefs topographiques intégrant les deux dimensions spatiales et un troisième paramètre relatif à un attribut spécifique (par exemple, l'intensité ou le gradient). Ces algorithmes peuvent être classés selon trois catégories : les algorithmes par inondation, les algorithmes par ruissellement et les algorithmes topologiques. De nombreux algorithmes de segmentation utilisant ces méthodes existent ([Gao H. et al, 2001] [Nguyen H.T. et al, 2003] [Kim J.B. et al, 2003] [Arbeláez P.A. et al, 2011] (Figure 123)).

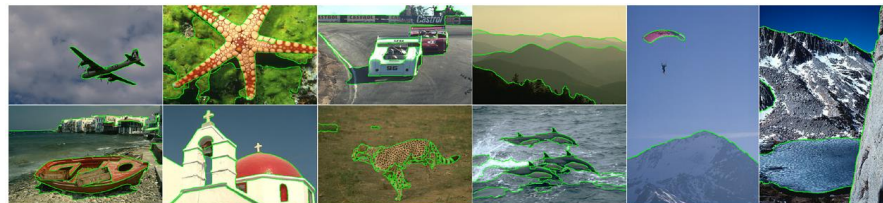


Figure 123 : Résultats de gPb-OWT-UCM [Arbeláez P.A. et al, 2011]

#### VII.A.3. Segmentation RGBD

Les méthodes présentées jusqu'à présent ne considéraient que les informations relatives aux images. Cependant avec le développement récent de capteurs RGBD tel que le Kinect (voir Chapitre II), l'acquisition de l'information de profondeur des pixels de l'image ajoute un aspect supplémentaire qui peut être intégré lors la segmentation des images. Ces méthodes ([Holtz D. et al, 2012] [Silberman N. et al, 2012] [Erdogan C. et al, 2012] [Fu H. et al, 2015]) ont notamment été développées dans le domaine de la robotique afin de pouvoir analyser la composition des scènes de manière plus efficace et détection les obstacles (Figure 124).

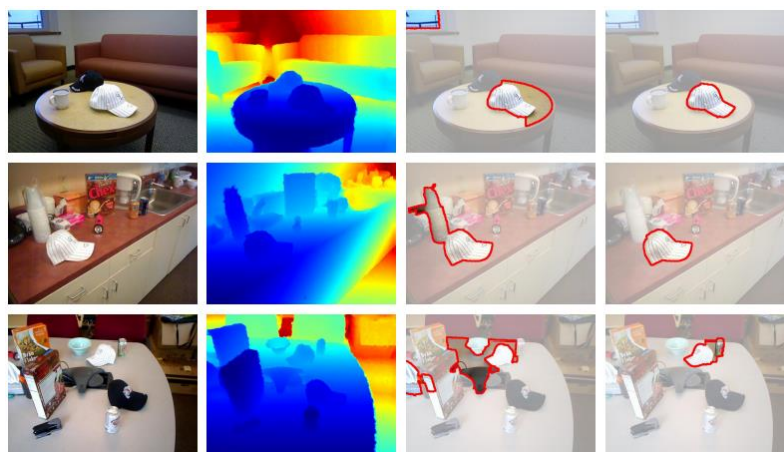


Image RGB Image de profondeur Résultats sur RGB Résultats sur RGBD

Figure 124 : Comparaison entre la co-segmentation d'image RGB et la co-segmentation d'image RGBD [Fu H. et al, 2015]

Ces travaux montrent bien que l'utilisation d'une information de profondeur permet d'obtenir des résultats intéressants en termes de segmentation des images et même d'obtenir des résultats plus précis que pour les segmentations n'utilisant que les informations relatives à l'image. Les combinaisons d'informations entre une disposition spatiale dans l'espace et une disposition spatiale dans une image offre donc une perspective intéressante pour la détection de régions sur des images.

#### VII.A.4. Evaluation des méthodes

La quantité impressionnante d'algorithmes de segmentation d'images et le large panel de techniques employées montre que la segmentation d'image est aujourd'hui toujours au cœur des problématiques. Les résultats offerts par ces méthodes montrent qu'il est possible de tirer de grandes quantités d'informations des images.

Afin d'évaluer de la qualité de ces méthodes, l'utilisation de l'index NPR (*Normalized Probabilistic Rand*) [Unnikrishnan R. et al, 2007] apparait comme une mesure précise de l'exactitude d'une segmentation. En pratique, la mesure de similarité passe par les comparaisons de segmentations issues de différents algorithmes et d'un ensemble de segmentations manuelles servant de références (Figure 125). Au vu des différentes segmentations manuelles obtenues, cet exemple illustre bien le problème de définir ce qu'est une bonne segmentation. En effet, les segmentations manuelles, pourtant toutes interprétées par un être humain, sont très différentes les unes des autres mais toutes sont justes.

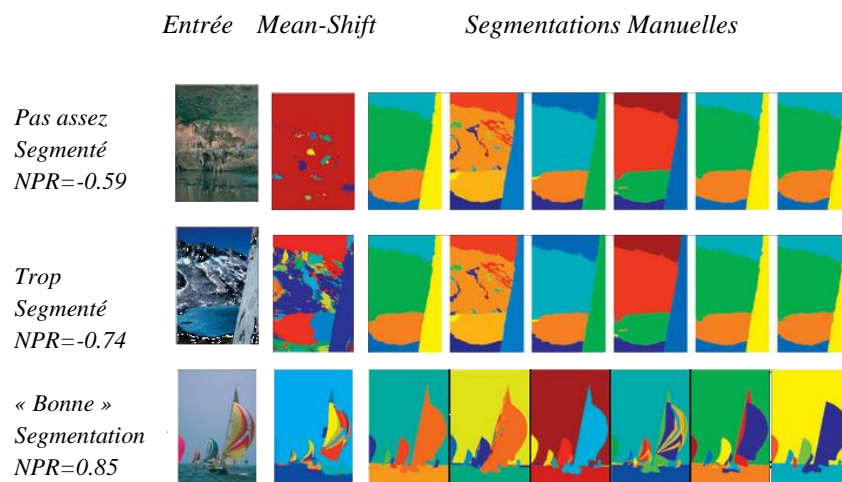


Figure 125 : Résultats de l'évaluation par l'index NPR de la segmentation Mean-Shift [Unnikrishnan R. et al, 2007]

L'utilisation de segmentation dans notre méthodologie pourrait proposer à l'utilisateur une simplicité de manipulation accrue. En effet, avec une segmentation efficace des images, il ne serait plus nécessaire de sélectionner manuellement la zone à annoter. La segmentation des images pourrait servir afin de simplement sélectionner des régions de la segmentation pour définir une annotation (voir VII.B.2.b. Outils de sélection semi-automatique à l'aide de la segmentation).

Le problème, qui se pose, est de pouvoir choisir une segmentation adaptée aux types d'images utilisées et notamment à la complexité des objets. Notre méthodologie s'intéressant à l'étude d'objets patrimoniaux, ceux-ci comportent des formes assez complexes et

nécessitent une précision très élevée pour une segmentation. Bien que les différentes méthodes de segmentation présentées montrent des résultats plutôt bons sur les jeux de données de test, dans notre cas, les formes des objets sont trop compliquées pour pouvoir appliquer une segmentation efficace qui donnerait directement le résultat attendu. De plus, l'analyse nécessite de pouvoir disposer de plusieurs niveaux de segmentation tout aussi précis les uns que les autres.

Cependant, l'utilisation de segmentation d'images n'est pas à exclure complètement. En effet, il est possible d'envisager des outils utilisant la segmentation de l'image pour tout de même aider l'utilisateur à faire ressortir des informations liées aux images. D'autre part, les méthodes utilisant des images RGBD montre que l'association d'informations 3D permet d'établir des segmentations plus précises qu'uniquement par l'utilisation d'images RGB. Etant donné que la méthodologie permet de disposer d'informations similaires voire même plus importantes, toutes les informations 3D (coordonnées, normales, etc.) liées aux images peuvent aussi servir à la segmentation.

### VII.B. Outils de sélection semi-automatique

Les développements d'outils pour l'aide à la sélection peuvent fortement guider l'utilisateur dans son analyse et simplifier la partie sélection du système. De plus, étant donné le niveau d'informations contenu par un pixel grâce à la mise en place de la relation 2D/3D (Figure 126), l'utilisation de ces différents types d'informations dans ces outils peut aider à faire ressortir des informations supplémentaires utiles à l'analyse.

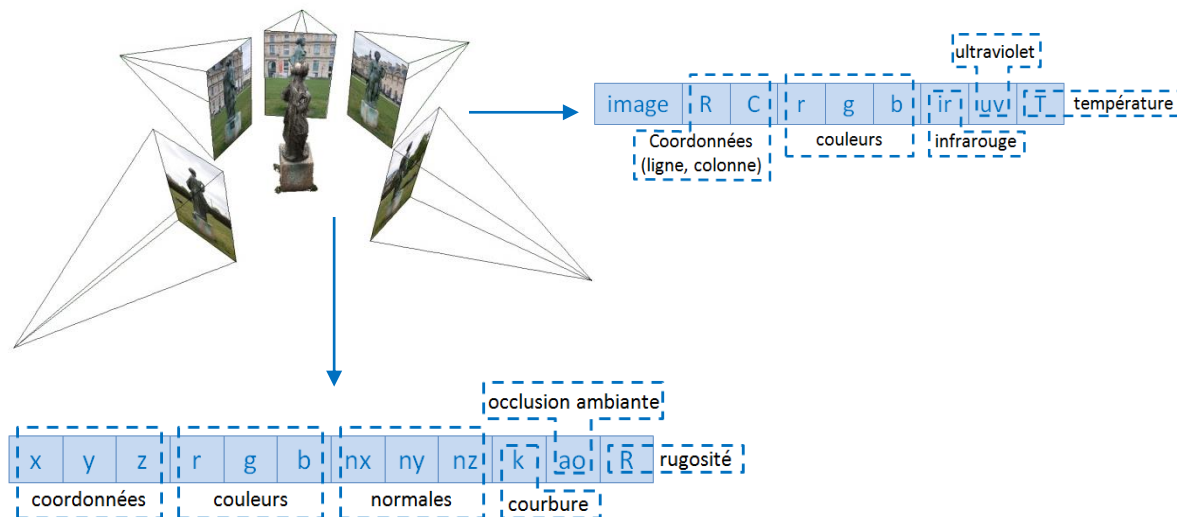


Figure 126 : Couches additionnelles d'informations

Avec l'indexation du nuage de points et les correspondances en index sur les images, chacune des informations des couches additionnelles peut être organisée en fonction des pixels des images, notamment en utilisant le principe de stockage par fichier TIFF (voir [V.C. Gestion des données](#)). Ainsi pour chaque image, une correspondance avec chacune des couches additionnelles est mise en place et peut servir pour rechercher des caractéristiques communes entre les pixels.



Deux outils ont été développés pour permettre d'extraire des informations des images de manière automatique. Un premier outil se base sur la même idée que l'outil « Baguette Magique » intégré à de nombreux éditeurs d'images, qui permet une sélection par seuil sur les couleurs, en étendant son utilisation à d'autres types d'informations. Le second outil utilise le principe de la segmentation d'images pour déterminer des régions de caractéristiques communes (clustering) et aider à la sélection.

### VII.B.1. Sélection par seuil : « Baguette Magique »

L'outil « Baguette Magique » intégré à de nombreux éditeurs d'image permet la sélection de pixels de couleurs similaires en un seul clic. Un pixel est sélectionné et sert de référence et d'autre part, un paramètre de tolérance est défini. Le paramètre de tolérance décrit le seuil à respecter par rapport à la valeur de référence. Nous avons donc développé des outils de sélection s'inspirant de cette méthode agissant sur les couleurs et sur les normales.

A partir de la sélection d'un pixel de référence par sélection de couleur  $[R_{ref}, G_{ref}, B_{ref}]$  et une valeur de tolérance  $t$ , tous les pixels de l'image sont testés. Un pixel est considéré dans le seuil de tolérance autour du point de référence si la valeur de chacune des composantes de couleurs se trouve dans l'intervalle  $[Valeur-t, Valeur+t]$ . Ainsi la recherche des valeurs des pixels se fait dans un cube de côté  $t$  dans l'espace des valeurs et de centre  $[R_{ref}, G_{ref}, B_{ref}]$  (Figure 127).

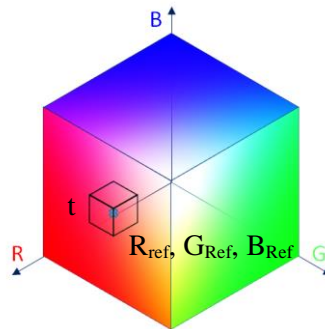


Figure 127 : Sélection des pixels répondant au critère de tolérance dans l'espace des couleurs autour du point de référence

A partir de ces critères, des pixels de l'image sont détectés et sont donc considérés comme pertinents au vu de la tolérance demandée et du point de référence. Cependant cela signifie que tous les pixels de l'image répondant à ce critère sont sélectionnés. Nous avons ajouté un critère de contiguïté afin de ne sélectionner que la région autour du point de référence (Figure 128).



Figure 128 : Résultats de la "baguette magique" sur les couleurs d'une image pour une tolérance de 22 avec ou sans contiguïté



Ce critère de sélection peut donc être utilisé sur toutes les images où la couleur représente une information telle que les images UV, les images IR ou les images de thermographies ou sur les données de valeurs telles que l'occlusion ambiante, la courbure ou la rugosité. Dans le cas d'images thermographiques, il est possible d'avoir en plus de l'image couleur, un tableau représentant la température des pixels. Dans ces cas de données de valeurs, le système peut facilement être adapté en définissant non plus une tolérance sur la couleur mais une tolérance sur la valeur. L'objectif sera donc de chercher les pixels dont la valeur de la donnée est comprise entre  $V_{Ref}-t$  et  $V_{Ref}+t$ .

Pour les images contenant les informations des normales, l'utilisation d'une tolérance telle que décrite précédemment ne rendra pas des informations pertinentes. Dans ce cas, le critère à considérer porte sur la direction des normales. Ainsi la tolérance sera définie par un angle  $\lambda$ . Les normales à sélectionner seront celle dont l'angle formé avec la direction de référence sera inférieur à  $\lambda$  (Figure 129).

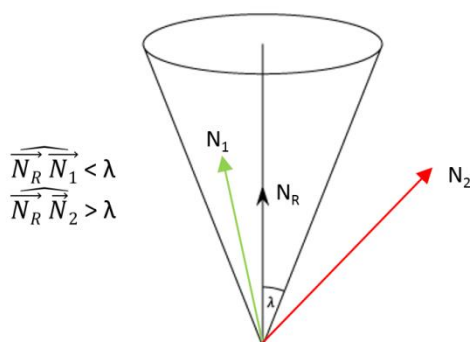


Figure 129 : Sélection des normales répondant au critère de tolérance autour de la normale de référence

Ce système de « Baguette Magique » sur les normales permet de faire ressortir notamment des zones ayant une direction de normales similaires c'est-à-dire des zones planes. De la même façon que pour les couleurs, il est possible de ne choisir que les pixels entourant le point de référence ou tous les pixels de l'image répondant au critère dans l'image (Figure 130).



Figure 130 : Résultats de la "baguette magique" sur les normales d'une image pour une tolérance de 22 degrés avec ou sans contiguïté

Ainsi grâce à ce système de sélection par seuil, l'utilisateur peut extraire, à partir d'un seul pixel sélectionné, tous les pixels de l'image ou seulement les pixels entourant le point de référence ayant une similarité soit sur les couleurs, soit sur les normales.

### VII.B.2. Sélection sur la segmentation

Une segmentation des images crée des régions de caractéristiques similaires. Comme vu en VII.A. Segmentation d'images, il existe de nombreuses méthodes de segmentation d'images. Cependant au vu des résultats obtenus par ces segmentations, elles ne peuvent pas être utilisées pour définir précisément les parties des images à annoter sur les objets qui nous intéressent.

Cependant les régions données bien que peu précises, offrent des informations sur des régions ayant des caractéristiques similaires. Si ces algorithmes de segmentation peuvent en plus être utilisés sur les images contenant des informations autres que les informations de couleurs, cela ouvre des perspectives d'analyses intéressantes.

#### VII.B.2.a. Segmentation des images grâce aux informations liées

Le choix de la segmentation à utiliser ne repose pas sur le critère de production d'une segmentation d'images précise mais de pouvoir être utilisée pour différents types d'informations. Ces informations étant de type tridimensionnel, les algorithmes de segmentation d'images qui semblent les plus pertinents à utiliser sont ceux par regroupement (*clustering*). En effet, ceux-ci considèrent l'image comme une représentation 3D dans l'espace des couleurs. Dans notre cas, cette interprétation tridimensionnelle des images est intéressante puisque l'on dispose de plusieurs images associées à des informations tridimensionnelles. Ainsi l'algorithme pourrait être utilisé sur les différents types d'informations associées aux images.

Pour les images couleurs, la dimension tridimensionnelle est l'espace des couleurs RGB. Cela représente un cube dont la valeur maximale sur chacune des directions (R, G ou B) est 255 et la valeur minimale est 0 ((a) Figure 131). Pour les images de coordonnées, la dimension tridimensionnelle est simplement l'espace 3D dans lequel sont représentées les coordonnées ((b) Figure 131). Dans le cas des images de normales, la dimension tridimensionnelle est représentée par une sphère de rayon 1 (pour des normales normées) sur laquelle les directions des normales se trouvent ((c) Figure 131).

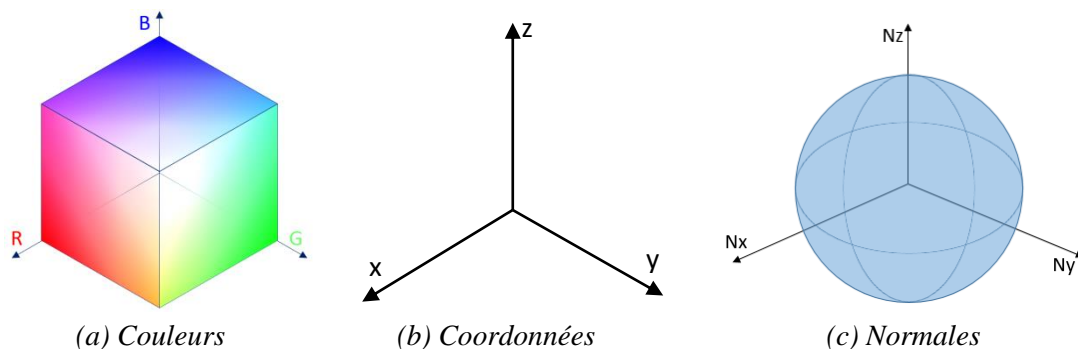


Figure 131 : Représentation des espaces 3D de chacun des types d'images

L'algorithme de K-Means a donc été testé sur les différents types d'images (Figure 132). Le fonctionnement du K-Means consiste à regrouper les points qui, dans l'espace 3D sont proches et dont la distance entre chaque groupe de points est prévue pour être minimisée. Ainsi cet algorithme se prête naturellement à notre objectif d'utilisation quel que soit le type de données.

La Figure 132 montre que la segmentation de l'image, en utilisant différents types d'informations 2D et 3D, offre des regroupements de ces informations par similarités. Notamment, la segmentation de l'image par les normales montre bien les surfaces de direction de normales similaires. Cette segmentation permet donc d'identifier les surfaces parallèles les unes aux autres ou la délimitation des surfaces. Dans le cas de l'utilisation des coordonnées 3D, l'algorithme de K-Means nécessite d'utiliser un nombre important de clusters pour espérer trouver des zones significatives. La Figure 132 montre le résultat d'une segmentation sur les couleurs de l'image représentant la réalité. Cependant cette segmentation peut très bien s'appliquer à d'autres images contenant des informations codées sur les couleurs (IR, UV, thermographique, etc.) ou seulement sur les couleurs liées au nuage de points.

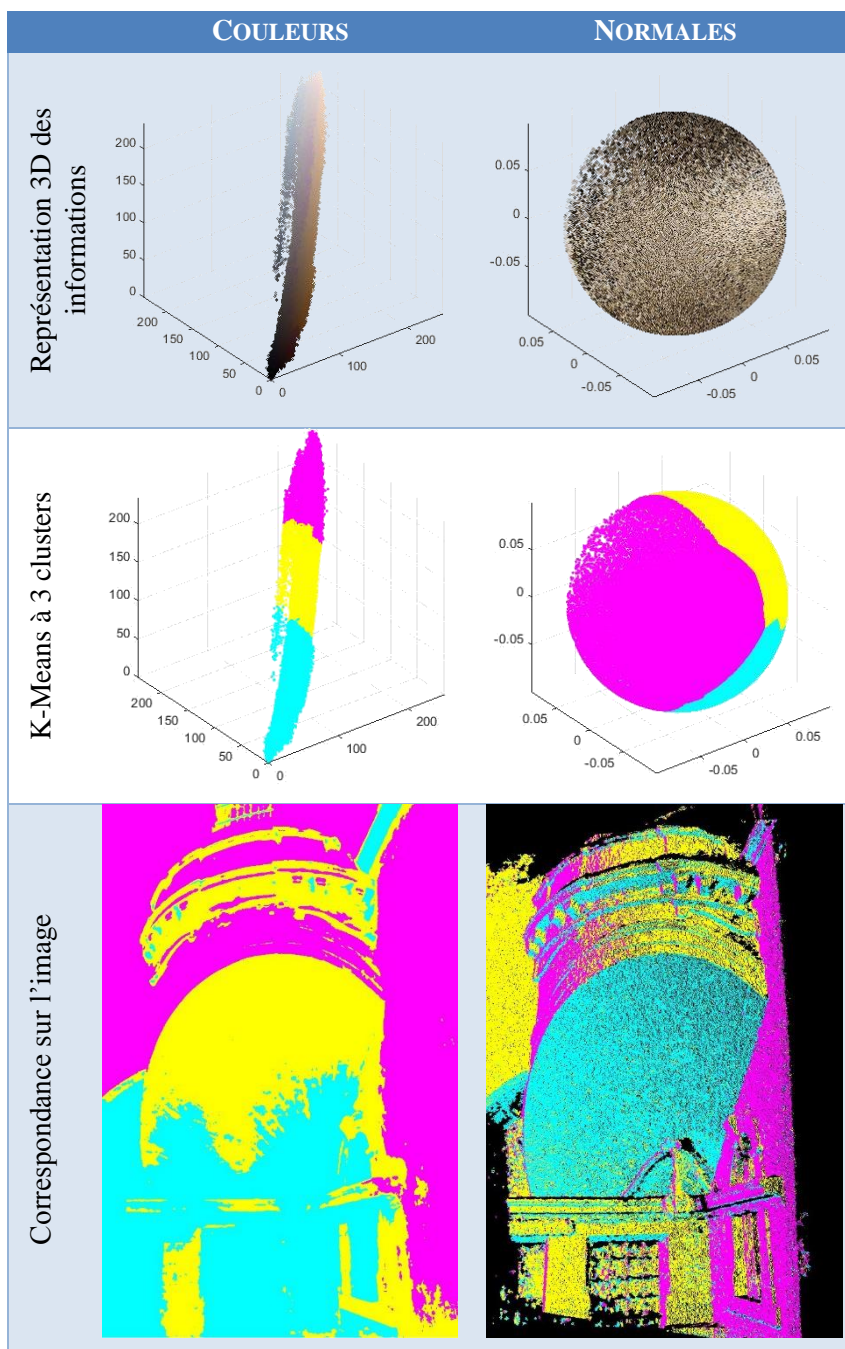


Figure 132 : Résultats de segmentation d'une même image sur les couleurs et sur les normales

La segmentation des coordonnées 3D et des normales est ici relative aux coordonnées contenues dans l'image. Il est, d'un autre côté, possible d'envisager de segmenter le nuage de points complet par des algorithmes de segmentation de nuage de points permettant de détecter plus efficacement des formes dans l'espace pour ensuite segmenter les images à partir de cette segmentation commune du nuage de points. Quelques tests ont notamment été effectués sur ShapeAnnotator [Attene M. et al., 2007] (Figure 133) et sur Plumber [Mortara M. et al, 2006] (Figure 134). Cependant ces algorithmes fonctionnent sur des maillages et non sur des nuages de points.

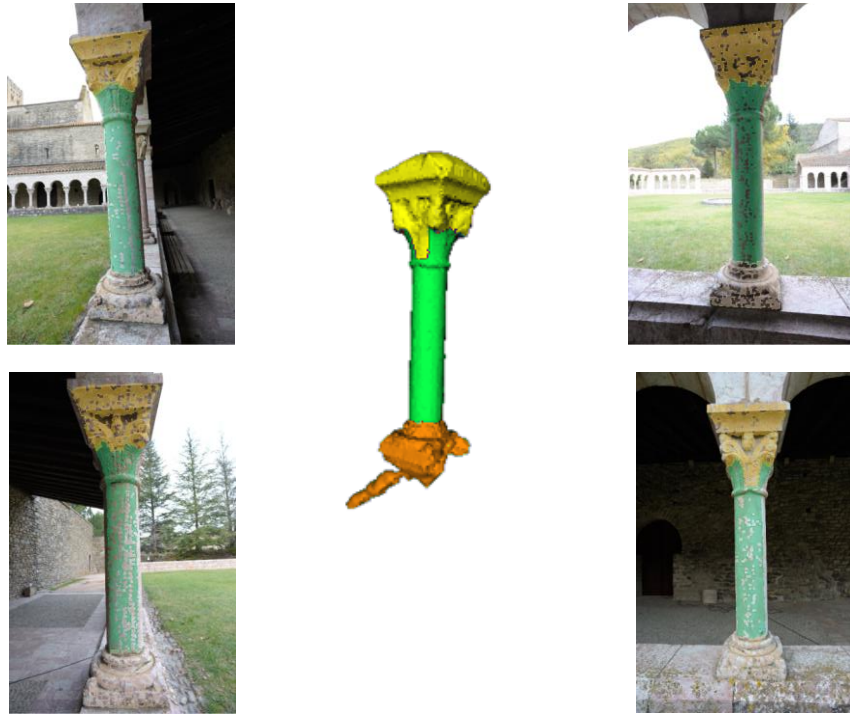


Figure 133 : Projection d'une segmentation 3D (obtenue avec ShapeAnnotator [Attene M. et al. 2007] sur les images



Figure 134 : Projection d'une segmentation 3D (obtenue avec Plumber [Mortara M. et al, 2006] sur les images

Ainsi, le fait de disposer une segmentation commune du nuage de points peut permettre d'obtenir, en plus, une relation entre les régions de segmentation d'une image à l'autre. Cependant, l'utilisation d'une segmentation du nuage de points nécessite encore une étude approfondie des algorithmes de segmentation de nuage de points afin de déterminer le ou les plus pertinents à utiliser.



### VII.B.2.b. Outils de sélection semi-automatique à l'aide de la segmentation

Les segmentations décrites au paragraphe précédent offrent des régions de caractéristiques similaires que l'utilisateur peut vouloir utiliser pour sélectionner ses annotations. Ainsi celles-ci peuvent servir de support pour la sélection afin de pouvoir affiner les régions des annotations.

Une sélection faite sur une image peut être comparée avec les régions d'une des segmentations de l'image. Cette annotation recouvre un certain nombre de régions de la segmentation. Ainsi il est possible de détecter ces régions recouvertes et d'étendre la sélection à la surface totale de toutes les régions recouvertes (Figure 135). Ainsi à partir d'une sélection très succincte, la sélection peut être affinée à l'aide des régions d'une des segmentations.

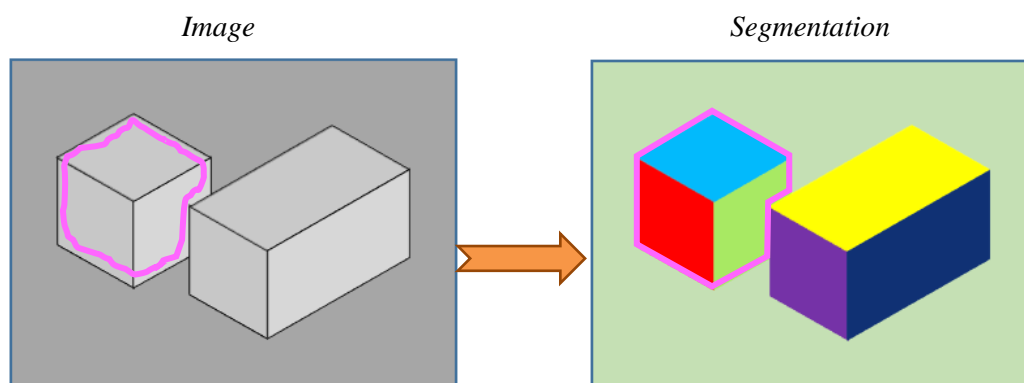


Figure 135 : A droite, sélection sur l'image, à gauche extension à l'aide de la segmentation recouverte

Le problème est que même si un seul pixel d'une région est compris dans la sélection initiale, la région de la segmentation associée sera dans son intégralité intégrée à la sélection. Ceci peut apparaître dans le cas où la segmentation est très fine et comporte de nombreuses petites régions. Dans ces cas, il est possible de prévoir d'ajouter un critère de recouvrement. Ainsi si le recouvrement d'une des régions de la segmentation par la sélection initiale est insuffisant par rapport à la surface totale de la région dans la segmentation, cette région sera exclue de la sélection (Figure 136).

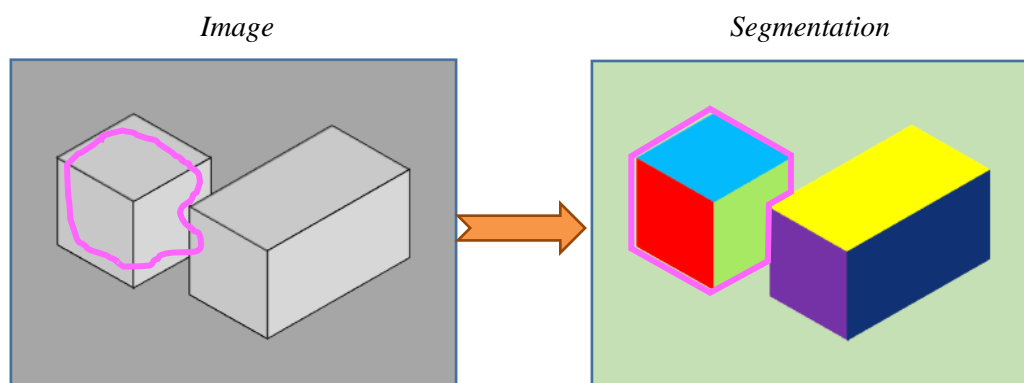


Figure 136 : A droite, sélection sur l'image, à gauche, extension à l'aide de la segmentation et avec critère de recouvrement

Sur des exemples plus concrets, cela permet de sélectionner des régions de caractéristiques communes à partir d'une sélection relativement peu précise (Figure 137).

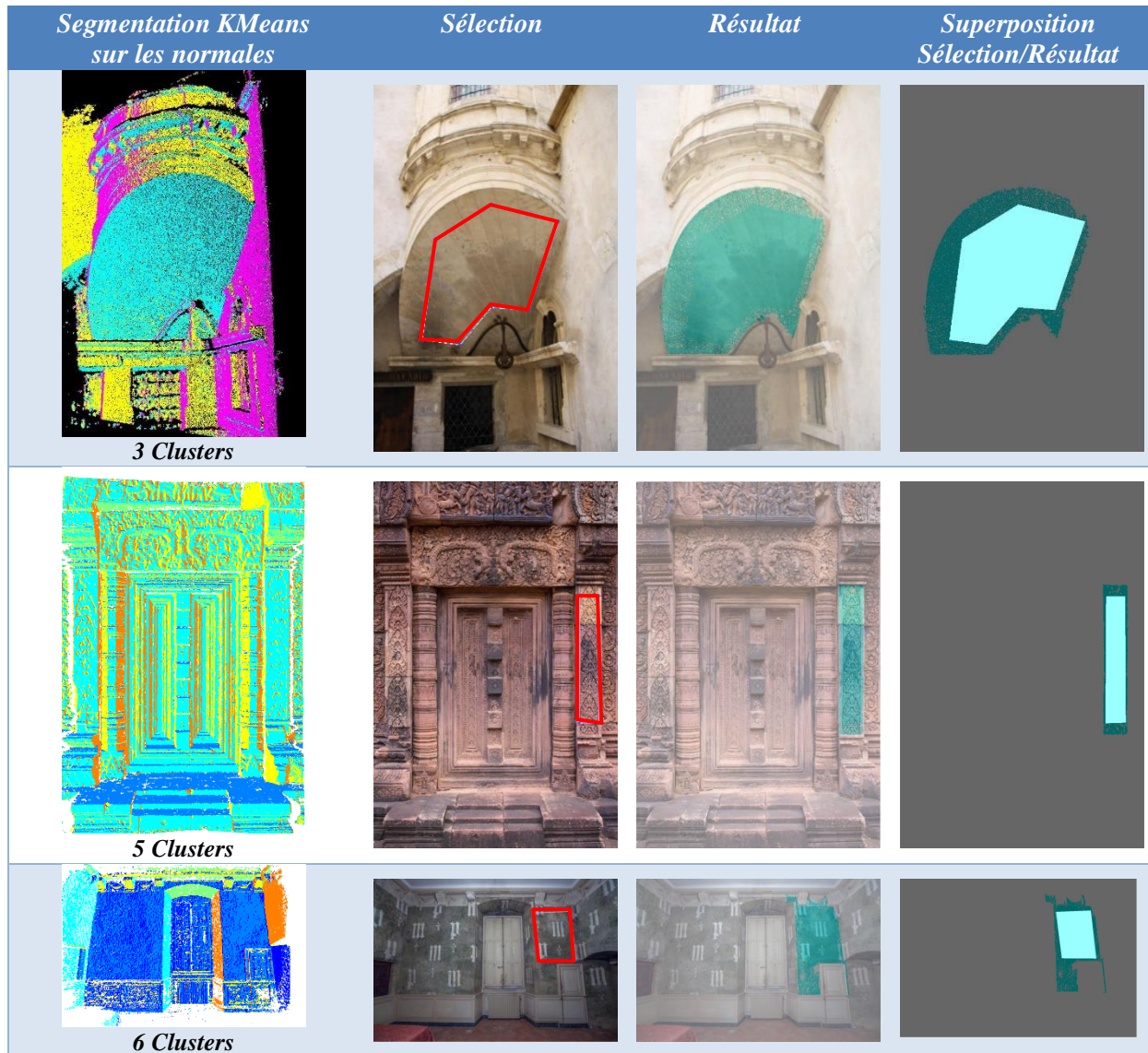


Figure 137 : Extension de sélection à l'aide d'une segmentation

Ainsi la segmentation des images permet d'aider l'utilisateur à sélectionner les annotations en fonction des segmentations faites sur les différents types de données. Il est même possible d'envisager de combiner les résultats de plusieurs segmentations en fonction des objectifs d'analyses.

Cette sélection étendue peut par la suite être utilisée comme base pour le transfert des annotations entre les images. D'autre part, il est aussi possible d'utiliser ce principe pour affiner le transfert entre les images en comparant la segmentation au masque obtenu à la fin du processus de transfert pour chacune des images.

### VII.C. Outils d'analyses quantitatives

Les annotations définies par le système intègrent plusieurs niveaux d'informations : une représentation 3D, et plusieurs projections 2D. Ces différentes informations peuvent servir de support pour entreprendre des analyses sur l'objet annoté. Les informations étant de type



bidimensionnel ou tridimensionnel, différents outils d'analyses sur ces données peuvent être envisagés.

### VII.C.1. Analyses sur les données bidimensionnelles

Chaque annotation est représentée sur les images par une projection 2D. Une analyse complète d'un objet patrimonial intègre plusieurs annotations qui peuvent être classifiées selon plusieurs calques. Un calque peut contenir plusieurs régions distinctes (un point au sein d'un calque n'appartient qu'à une seule région du calque). Ces représentations bidimensionnelles sont par ailleurs associées à une description.

Une projection 2D d'une annotation sur une image est représentée par une région sur cette image. Cette région représente une surface sur l'image et peut donner des informations de recouvrement de l'image. Ainsi il est possible de calculer un pourcentage de recouvrement de l'annotation sur chacune des images. Cette information peut être utile dans le cas d'une recherche d'image par description. En effet, si l'utilisateur interroge la base de données sur une description précise, le système peut proposer les images où apparait cette annotation et dans l'ordre décroissant du pourcentage de recouvrement. L'utilisateur doit pouvoir disposer des images où l'annotation est la plus visible c'est-à-dire avec le recouvrement le plus élevé.

Au sein d'un calque d'annotation, plusieurs régions distinctes peuvent cohabiter. Ces différentes régions possèdent chacune un pourcentage de recouvrement sur les images. Comme chacune de ces régions ne se superpose pas à une autre, puisqu'elles sont dans le même calque, un diagramme de recouvrement des régions au sein du calque peut être dessiné par rapport à chaque image. Ce type de diagramme a déjà été intégré dans la plateforme NUBES [Stefani C. et al, 2012] (Figure 138) mais les annotations étaient dans ce cas relatives aux parties des textures. Dans notre cas, une implémentation de ce système permettrait d'obtenir des informations en fonction du point de vue de l'image en cours mais aussi de façon plus globale en combinant les informations de chacune des images.

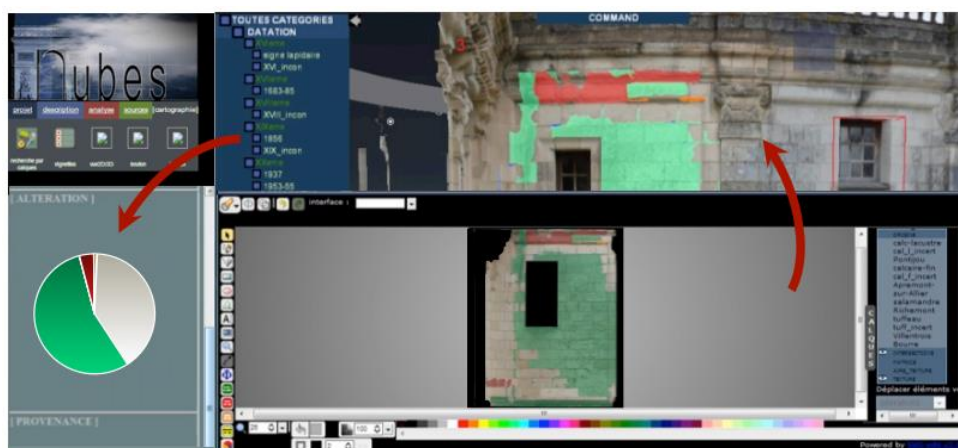


Figure 138 : Intégration d'un diagramme de recouvrement dans l'interface de Nubes [Stefani C. et al, 2012]

Deux régions de deux calques différents peuvent au contraire se superposer. Si deux calques ont été créés, c'est qu'il s'agit de deux niveaux de description sémantique différents (par exemple, un calque pour les matériaux et une autre pour les altérations). A des fins d'analyses, la comparaison de ces deux informations peut faire ressortir des corrélations ou non entre deux annotations [Messaoudi T. et al, 2014] (Figure 139). Ainsi les tests effectués

ont montré que le recouvrement entre les deux régions des deux calques permet de croiser les informations de deux calques.



Figure 139 : Superposition de trois annotations pour la recherche de corrélation [Messaoudi T. et al, 2015] : en rouge, élément architectural : contrefort, en jaune, matériau : pierre jaune de Caromb, en blanc, altération : alvéolisation

La Figure 139 montre trois types d'annotations appartenant à trois niveaux de descriptions différents (calques) : éléments architecturaux, matériaux et altérations. Ainsi le croisement de ces trois régions montre qu'il y a bien une corrélation entre le type de matériau (pierre jaune) et l'altération d'alvéolisation sur le contrefort. La gestion multicouche du système permet donc d'établir des liens entre les différents niveaux sémantiques à des fins d'analyses.

D'autre part, [Messaoudi T. et al, 2015] cherchent mettre en relation les calques de descriptions à une ontologie structurant les différentes données nécessaires pour la conservation des structures maçonnées et les mises en relations avec les concepts associées. L'association des comparaisons entre les calques de description et leurs annotations, définis par notre approche, et l'ontologie permet de mettre à disposition et de générer des informations relatives à une altération spécifique. Ainsi la superposition de deux régions sur deux calques permet, par l'ontologie, de mettre en relation les données intégrées afin de créer une instance d'altération sur un matériau particulier si toutefois ce dernier est renseigné. Par exemple, par rapport à la Figure 139, l'ontologie permet de dire que, au point de superposition de ces régions, nous constatons un ouvrage architectural de type contrefort qui a été construit avec de la pierre jaune de Caromb, un matériau touché par une altération de type alvéolisation. Ceci est une observation scientifique réalisée par un expert du domaine.

### VII.C.2. Analyses sur les données tridimensionnelles

A partir du lien entre les images et le nuage de points, chaque annotation possède en plus des représentations 2D, une représentation 3D sous forme de nuage de points (Figure 140). Cette représentation contient donc des informations dimensionnelles sur l'annotation qui peuvent être extraites à des fins d'analyses. De nombreux outils d'analyses tridimensionnelles peuvent donc être intégrés au système tout en permettant une sélection et une visualisation sur les images. Cette extraction possible des formes par rapport à une sélection pourrait par ailleurs servir de base à la construction de bibliothèques de formes.

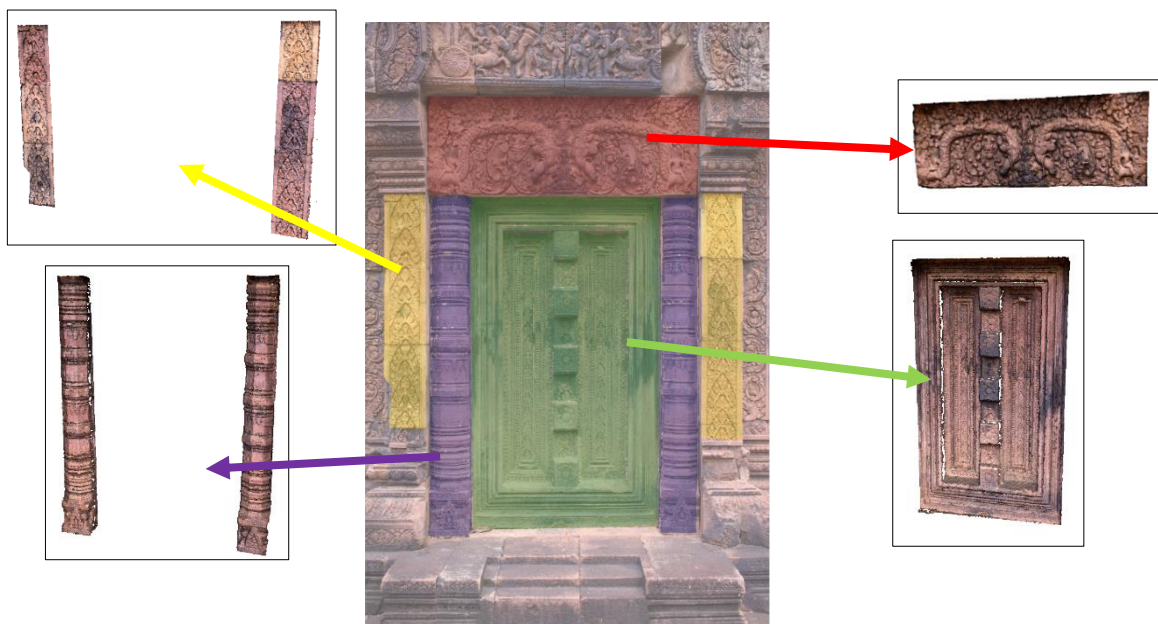


Figure 140 : Nuage de points des annotations

Une annotation peut simplement être une extraction d'un profil (Figure 141). Ainsi si la région dessinée sur une image est un simple trait, le nuage de points de l'annotation représentera un profil de l'objet suivant la direction de l'annotation. L'analyse de profil peut permettre de faire ressortir des informations caractéristiques d'un objet telles que des symétries ou encore de pouvoir extraire des informations dimensionnelles sur les formes.

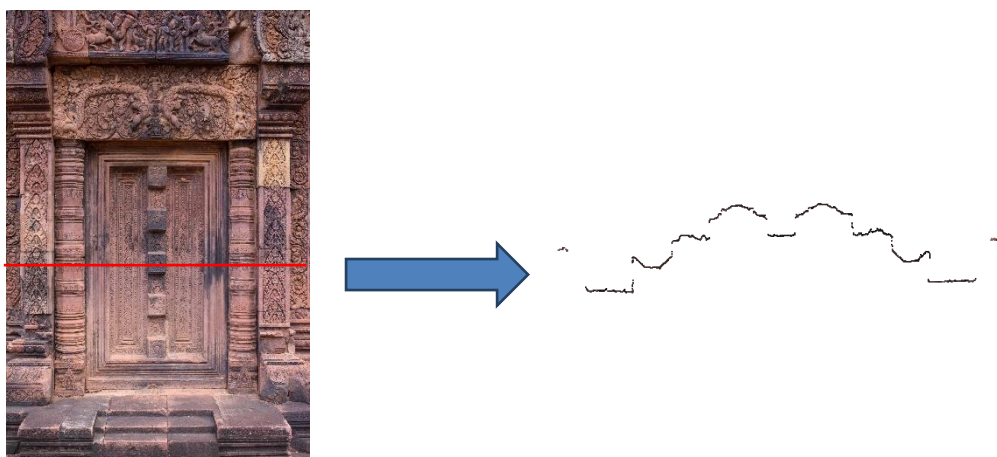


Figure 141 : Annotation et profil extrait



### VII.C.2.a. Caractérisation morphologique

A partir des données du nuage de points de l'annotation, des informations caractérisant la morphologie telles que la courbure, l'occlusion ambiante et la rugosité peuvent être générées.

La courbure est une mesure permettant de qualifier, sous forme d'une valeur, la courbe d'un objet. Par exemple, une ligne droite aura une courbure nulle alors qu'un cercle aura une courbure positive constante. Le nuage de points peut servir à calculer des informations de courbure. En effet, la courbure peut être estimée à l'aide des méthodes gaussiennes, de méthodes moyennes ou de méthodes par les normales (Figure 142).

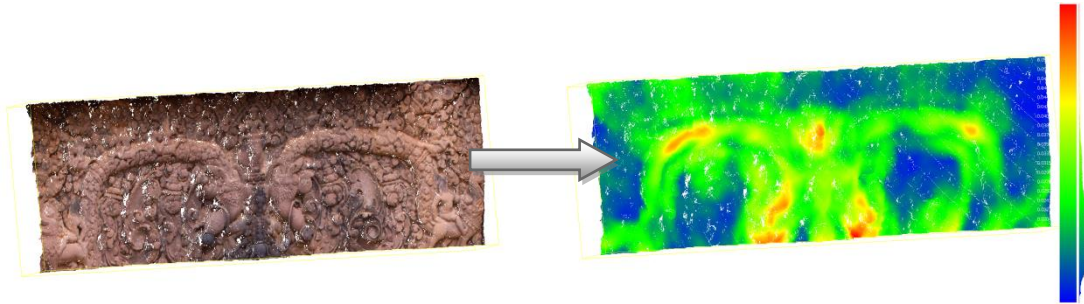


Figure 142 : Courbures extraites à partir du nuage de points d'une annotation

L'occlusion ambiante est une méthode de rendu, basée sur le *raytracing*, qui cherche à simuler une illumination globale. Cette technique assombrit les zones naturellement difficiles d'accès à la lumière. Le principe repose sur l'analyse, en chaque point, d'une sphère en testant les portions cachées de cette sphère par les autres objets de la scène. Cette méthode a donc pour effet d'assombrir les fissures, les coins et les points de contact (Figure 143).

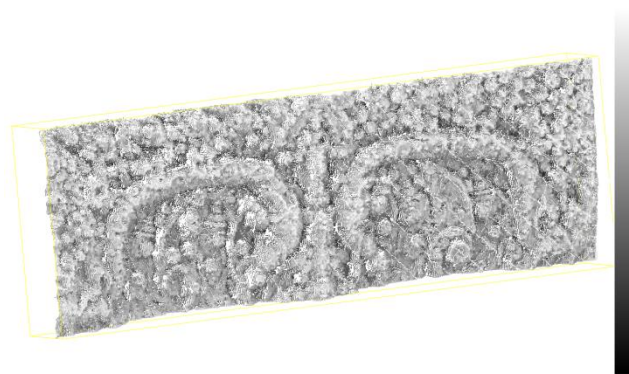


Figure 143 : Occlusion ambiante

Une autre caractéristique morphologique pouvant être extraite est la rugosité. Le calcul de la rugosité passe par la mesure de la distance entre un point et le meilleur plan passant par ses voisins (Figure 144).

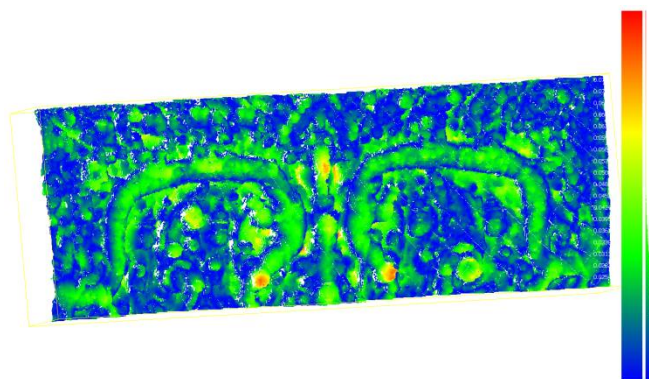


Figure 144 : Rugosité

Ces trois caractérisations morphologiques donnent une échelle de répartition des valeurs. Il est donc possible de prévoir d'utiliser les outils de sélection et de segmentation présentés en **VII.B. Outils de sélection semi-automatique**.

#### VII.C.2.b. Cartographie de profondeur

A partir d'algorithme d'estimation de plan moyen, tels que l'algorithme RANSAC ou par méthode des moindres carrés, le nuage de points extrait d'une sélection peut aussi servir de base pour la cartographie de profondeur de l'objet. En effet, l'algorithme d'estimation de plan moyen va définir un plan à partir des points extraits puis une carte de profondeur peut être construite par rapport à ce plan (Figure 145). Cette carte étant construite à partir du nuage de points, elle peut être projetée sur toutes les vues où l'objet considéré apparaît et ainsi aider l'utilisateur dans sa sélection ou dans son analyse. Cette carte peut notamment être exploitée pour l'analyse de cartographies d'altérations.

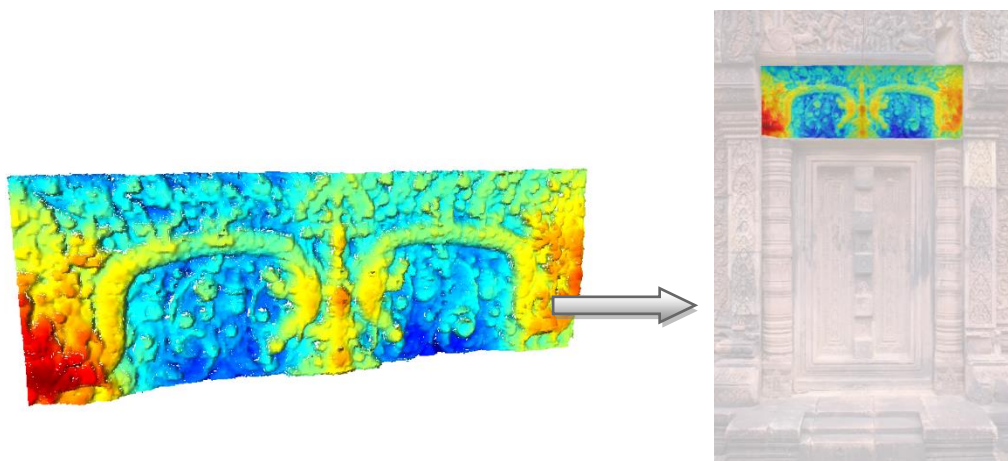


Figure 145 : Carte de profondeur à partir d'un plan par moindres carrés et reprojection sur une image

#### VII.C.2.c. Mesures dimensionnelles

Le nuage de points de l'annotation peut aussi permettre d'extraire des dimensions sur l'objet par simple calcul de distances entre deux points dans l'espace. La sélection des points extrêmes pour la mesure de la distance peut être établie directement sur les images en sélectionnant deux pixels de l'image pour lesquels nous disposons de coordonnées 3D. La

mesure donnée est relative à l'échelle choisie. Cette donnée de mesure peut aussi être transférée à d'autres images sur lesquelles les deux points extrêmes apparaissent.

[Nespeca R., 2015] a développé une méthode d'estimation rapide de surface et de volume à partir d'un nuage de points sans avoir à utiliser de reconstruction totale du modèle 3D. L'idée développée est que chaque point peut être représenté de manière locale par une surface carrée plane (ou par un cube pour les volumes) (Figure 146). Connaissant l'espacement entre les points, les dimensions de cette surface sont déterminées. Puis, en sommant la valeur de l'aire toutes les surfaces élémentaires de chacun de points, la surface représentée par le nuage de points est estimée. Pour le calcul du volume, il est nécessaire de définir un plan de référence définissant la troisième dimension du cube (par exemple le plan RANSAC). Cette méthode suppose d'avoir rééchantillonné le nuage de points afin que tous les points soient espacés d'une même distance. Cependant cela offre une manière rapide d'estimation de la surface ou du volume contenu par rapport à un plan.

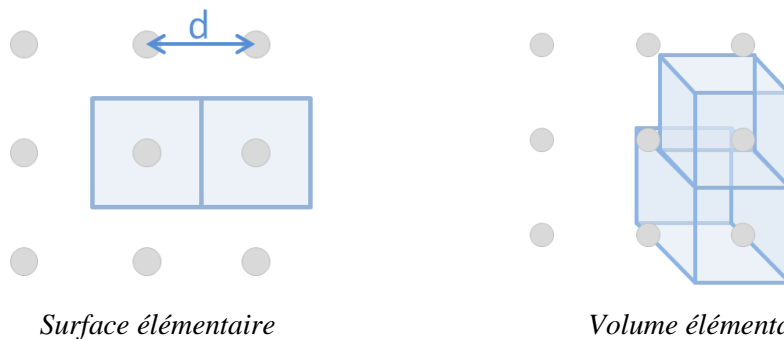


Figure 146 : Définition de surface et volume élémentaires pour l'estimation de surface et de volume d'un nuage de points [Nespeca R., 2015]

Ainsi en utilisant l'annotation sous forme de nuage de points, de nombreuses informations dimensionnelles sur l'annotation peuvent être extraites et des informations supplémentaires sur les données géométriques peuvent être extraites afin d'être exploitées pour l'analyse et mises en relation avec les données 2D contenues dans les différentes images. Il est, bien sûr, possible d'imaginer d'autres applications possibles à partir du nuage de points extraits de l'annotation.

### VII.C.3. Conclusion

Les outils, décrit en VII.C.1. [Analyses sur les données bidimensionnelles](#) et VII.C.2. [Analyses sur les données tridimensionnelles](#), montrent que les différents niveaux d'informations de chacune des représentations des annotations (projections 2D et nuage de points) peuvent servir de base à l'extraction d'informations pertinentes à l'analyse. La Figure 147 résume les informations liées à une région annotée ou pouvant être extraites de cette région. Le lien 2D/3D permet en plus de transférer les informations extraites d'une représentation de l'annotation vers l'autre représentation et ainsi de pouvoir visualiser une information sur une représentation qui ne pourrait pas servir à extraire cette information.



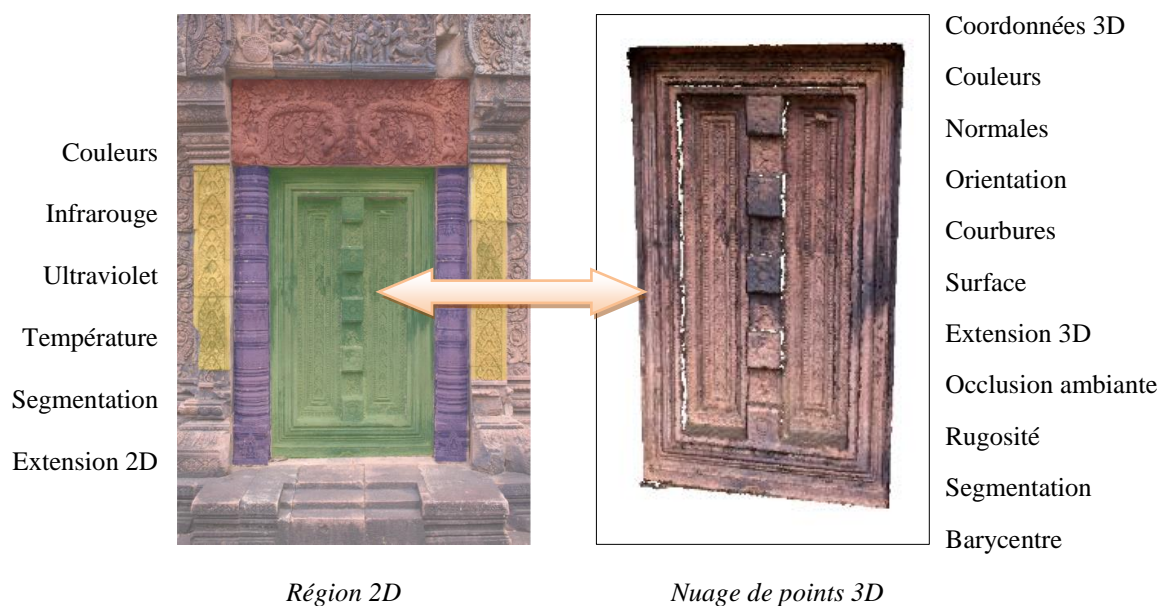


Figure 147 : Informations extraites d'une annotation

Ces informations extraites sont à la base des méthodes d'analyses et peuvent donc être exploitée pour extraire des données statistiques, spatiales (distribution spatiale) ou encore morphologiques (similarité géométrique). Ainsi notre approche permet de mettre en lien différents niveaux d'analyses qui sont habituellement difficiles à comparer.

## VIII. Interface d'annotations : Développement d'Aïoli

A partir de la méthodologie proposée, plusieurs interfaces ont commencé à être développées pour permettre une utilisation aisée par un utilisateur. Ces interfaces n'intègrent pour l'instant que l'aspect d'orientation, d'indexation, de propagation des annotations entre les images et les outils d'aide à l'annotation. Les aspects d'ajout d'images, de gestion d'états temporels et d'analyses multidimensionnelles ne sont pas encore intégrés à ces interfaces mais devraient l'être dans des implémentations futures.

Chacune de ces interfaces possède plusieurs fonctions :

- Création d'un nouveau projet
- Importation des photographies du projet
- Traitement des photographies par MICMAC
- Recherche des correspondances 3D des images avec le nuage de points
- Gestion de calques d'annotation et de régions
- Edition de régions (modifications, suppression)

Le fonctionnement de ces interfaces est détaillé en [Annexe D](#).

### VIII.A. Structuration des données

La structuration des données passe par la mise en relation au sein d'une base de données des différentes informations à gérer.

Le système se base sur un ensemble d'images traitées par photogrammétrie générant un nuage de points. Ainsi pour chaque image est associée une calibration, une orientation autour du nuage de points et une correspondance entre les coordonnées 3D et les pixels de l'image. Chaque annotation est définie par un calque et une région. Un calque peut contenir plusieurs régions qui ne doivent pas se chevaucher au sein de ce calque. Un calque peut donc être défini comme une segmentation possible de l'objet. Chaque région est associée à une représentation 3D et un ensemble de projection 2D sur les images.

Ainsi la structuration de la base de données a été définie selon plusieurs tables :

- **Projet** : informations générales du projet (nom, liste des identifiants des images du projet, auteurs, dossier de sauvegarde, date de création, date de dernière modification)
- **Nuage de Points 3D** : données 3D du nuage de points (identifiant 3D, coordonnées XYZ, normales  $N_xN_yN_z$ , couleur RGB)
- **Images** : données des images (identifiant image, nom du fichier, URL dossier des images, URL fichier d'orientation, URL fichier de calibration, liste des identifiants 3D projetés sur l'image dans l'ordre des pixels de l'image)
- **Régions 3D** : données des régions 3D (identifiant région 3D, identifiant label région, couleur de la région, liste des identifiants 3D des points de la région, identifiant du calque, liste d'identifiants région 2D associés)
- **Régions 2D** : données des régions 2D (identifiant région 2D, identifiant région 3D associée, identifiant de l'image associée, identifiant du calque, liste de positions)

des pixels de l'image définissant la région dans l'image, pourcentage de recouvrement de l'image par la région)

- **Labels (ou Tags)** : données des labels (identifiant du label, nom du label, commentaires)
- **Calques** : données des calques (identifiant du calque, nom du calque)

La Figure 148 montre le schéma global de cette base de données. Une région est décrite dans ce modèle par un label (ou tag) et un commentaire. Cependant, il est facilement envisageable de créer un modèle de description plus complexe en fonction des besoins d'analyses, en passant par exemple par une ontologie de domaine ou un thésaurus. Le tag devient ainsi un lien vers le concept.

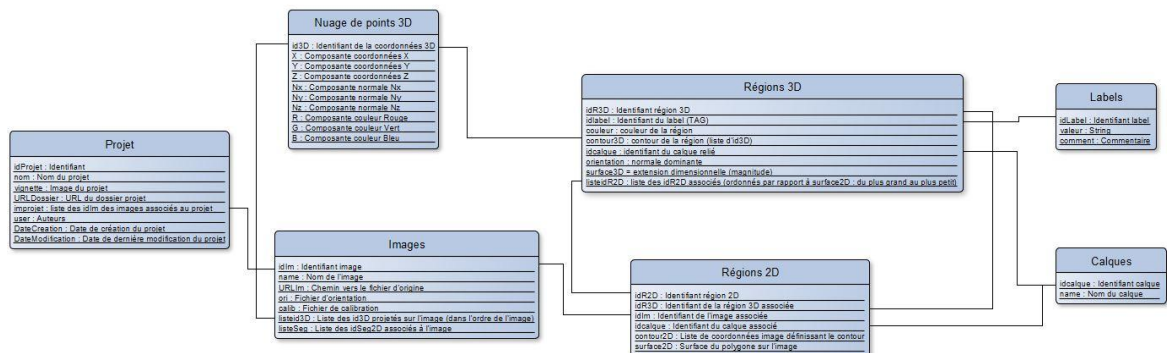


Figure 148 : Structuration des données au sein de la base de données (voir Annexe A)

En relation avec la structure générale du projet, le dossier projet doit pouvoir prendre en compte les différentes données générées. Le dossier projet se constitue donc de cinq sous-dossiers (Figure 149) :

- **Images** : dossier contenant les images brutes et les images redimensionnées pour le calcul de la correspondance 2D/3D
- **MicMac** : dossier contenant les résultats de calcul par photogrammétrie (orientation, calibration et nuage de points)
- **Nuage3D++** : dossier contenant une copie du fichier indexé du nuage de point
- **Calques** : dossier contenant les différents calques classés en sous-dossier. Chaque dossier de calque contient un ensemble d'images correspondant à chacune des images du projet et contenant les annotations projetées sur les images. Ces dossiers contiennent aussi l'export des régions annotées en nuage de points 3D lorsqu'il a été fait.
- **DataBase** : dossier contenant la base de données du projet sous forme de fichiers .mat (format de fichiers MATLAB)

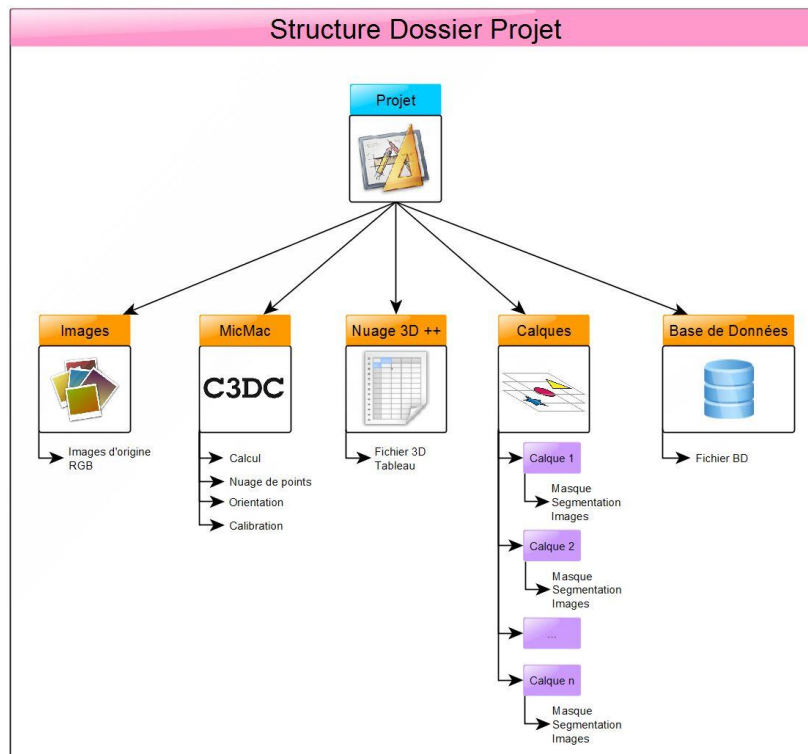


Figure 149 : Structure du dossier projet

La structure ainsi définie des données permet de pouvoir éditer et de mettre à jour les différentes informations liées aux annotations. Le système dialogue donc entre la base de données et les différents fichiers de dossier projet.

Cette structure génère beaucoup de données qui peuvent, suivant la taille du projet, représenter une taille importante en termes de quantité de mémoire utilisée ou d'espace occupé sur le disque. De plus, elle ne permet pas encore d'intégrer tous les aspects de possibilités d'analyses développés dans ce manuscrit. Le système étant amené à évoluer et à être amélioré, cette structure n'est pas figée et est donc aussi amenée à évoluer et à être optimisée en fonction des développements et implémentations futurs du système, notamment au niveau des formats de sauvegardes des différentes informations.

### VIII.B. Interface MATLAB

Les différents modules de calculs ayant été développés sous MATLAB, une première interface a donc été développée sous ce langage. MATLAB permet de créer des GUI (*Graphic User Interface*) pour permettre à un utilisateur d'interagir avec des algorithmes de calculs développés sous MATLAB.

L'interface intègre donc quatre parties (Figure 150) :

- Une partie **Gestion du projet** (création d'un nouveau projet, ouverture d'un projet, informations du projet)
- Une partie **Gestion des données** (création et édition des régions)
- Une partie **Choix des images et des régions**
- Une partie **Aperçu** (aperçu du calque sélectionné sur l'image sélectionnée et informations de la région sélectionnée)

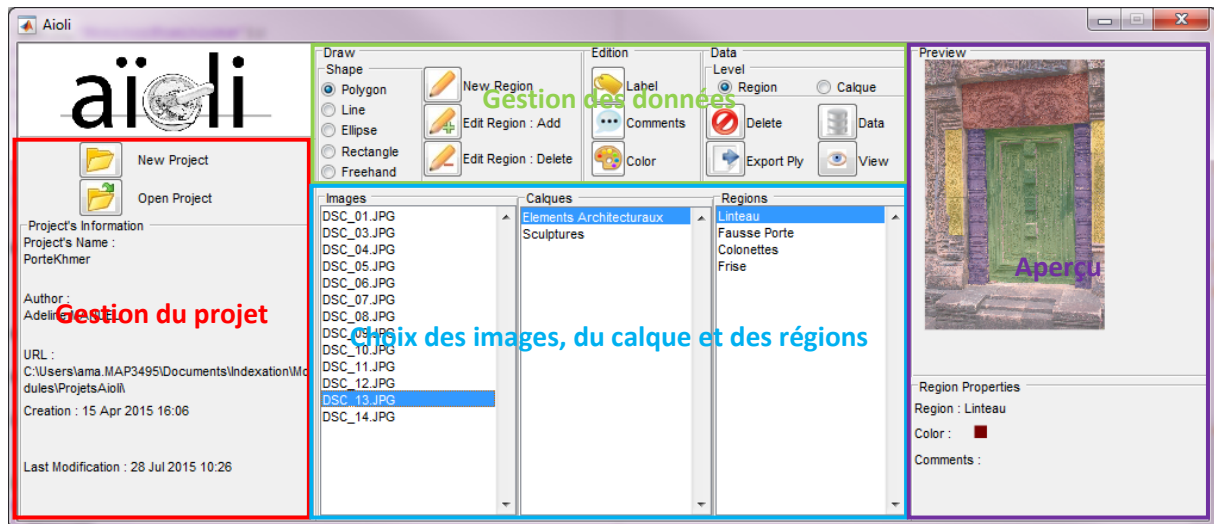


Figure 150 : Aperçu de l'interface développée sous Matlab (pour le détail voir Annexe B)

L'interface reste donc très basique car MATLAB ne permet pas de développer des interfaces avancées. Cette interface communique avec des modules de calculs MATLAB permettant d'effectuer les différentes étapes du processus et avec la structure de base de données et de dossier projet définie au chapitre IV (Figure 151).

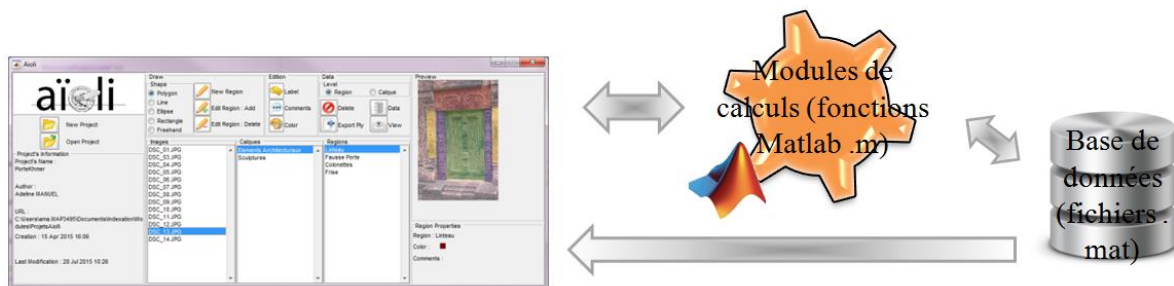


Figure 151 : Fonctionnement général de l'interface Matlab

Lors de la création d'un nouveau projet, l'utilisateur charge les images dans l'interface. Ces images sont alors traitées tout d'abord par MICMAC puis par les modules d'indexation et de recherche de correspondances 3D des images. Ces étapes de prétraitement du projet peuvent être relativement longues. Cependant elles ne sont nécessaires que lors de la création du projet. A la fin de ce prétraitement, l'utilisateur peut commencer son travail d'annotation.

### VIII.C. Interface C++/QT

Les interfaces développées par MATLAB étant relativement limitées, l'utilisation d'un autre langage pour le développement de l'interface a été envisagée. Comme vu dans le Chapitre IV, QT offre des possibilités de développement d'interfaces plus avancées. De plus, MATLAB intègre des fonctions de conversion de son code en bibliothèques C++. Ainsi, cette interface a été développée par un stagiaire afin de pouvoir utiliser les modules développés sous MATLAB converti en C++. D'autre part, des bibliothèques de lecture de fichiers .mat existent sous QT.

Cette interface se décompose en six parties (Figure 152) :

- Une partie **Menu** (création de projet, ouverture de projet)
- Une partie **Outils** (outils pour l'annotation)



- Une partie **Visualisation** (aperçu de l'image sélectionnée et des régions sélectionnées)
- Une partie **Images** (browser des images du projet)
- Une partie **Calques et Régions** (arbre)
- Une partie **Propriétés des régions** (informations liées à la région sélectionnée)

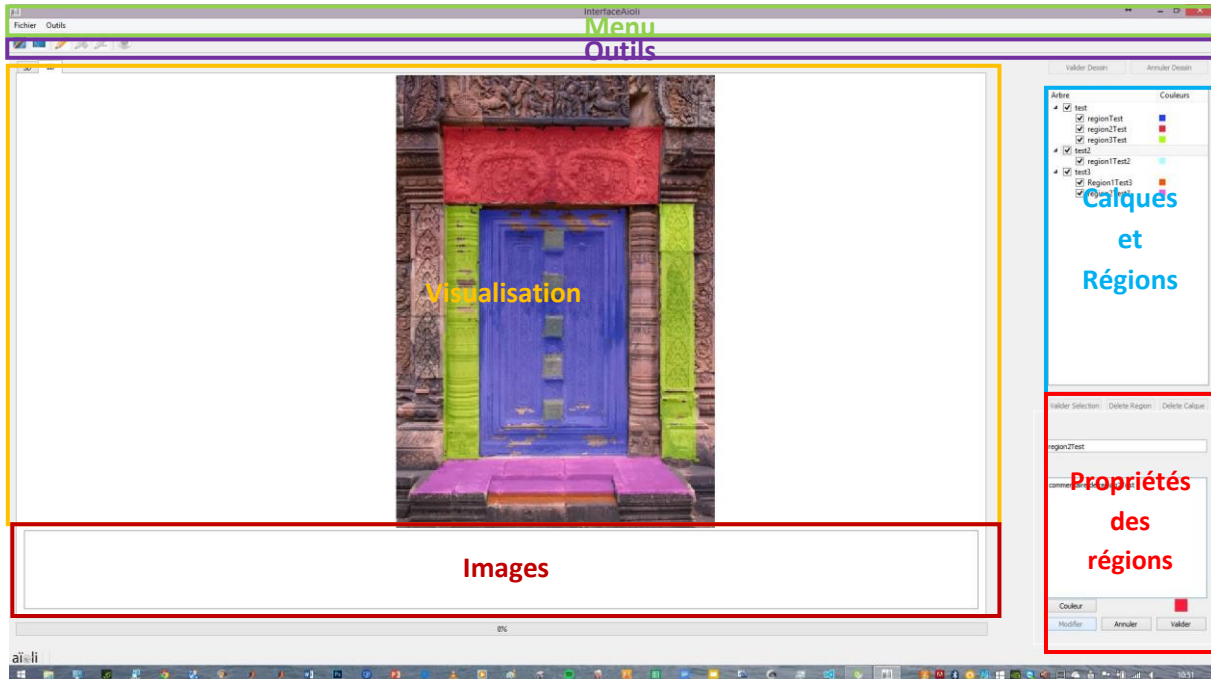


Figure 152 : Aperçu de l'interface Qt (pour le détail voir [Annexe B](#))

L'interface permet donc à l'utilisateur d'interagir avec les données qui sont par la suite traitées à l'aide des modules MATLAB convertis en librairies C++. Le fonctionnement de ces librairies suppose tout de même que le MCR (*Matlab Compiler Runtime*) soit installé sur la machine (Figure 153). Cependant le MCR est disponible librement et peut être intégré au logiciel sous forme d'une partie à installer.

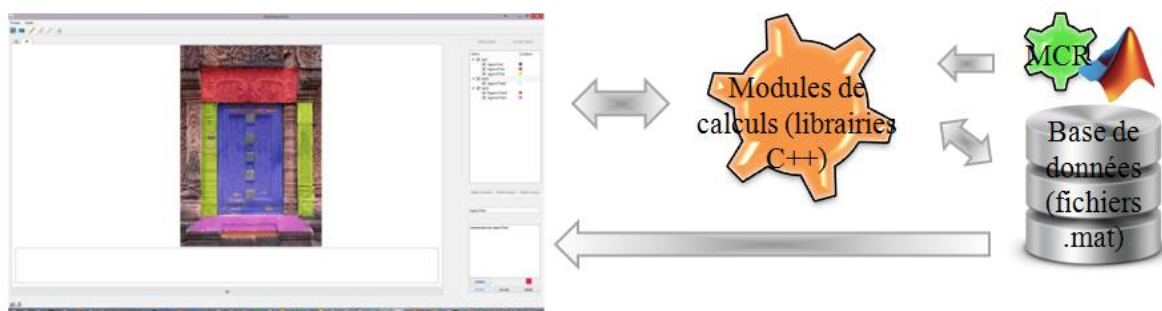


Figure 153 : Fonctionnement général de l'interface Qt

Cette interface est actuellement encore en cours de développement.

#### VIII.D. Interface HTML5/WebGL

Le développement d'une interface pour une utilisation locale suppose dans notre cas que l'utilisateur possède une machine relativement puissante au vu des traitements à effectuer.



Ceci n'étant pas le cas de tous les utilisateurs, le développement d'une interface autorisant le traitement à distance de la partie calcul a été engagé en collaboration avec Christine CHEVRIER et Benjamin LOILLIER du laboratoire MAP-CRAI dans le cadre du projet C3DC [C3DC].

Cette interface suppose donc de pouvoir communiquer avec un serveur dédié au calcul et renvoyant les informations sur l'interface pour l'utilisateur. Elle a donc été développée sous HTML5 associé à WebGL.

Cette interface se décompose en cinq parties (Figure 154) :

- Une partie **Inspecteur** (images, calques et régions associées aux images)
- Une partie **Données** (informations sur la région ou le calque sélectionné)
- Une partie **Visualisation** (aperçu de l'image sélectionnée et des régions sélectionnées)
- Une partie **Outils** (outils de dessin pour l'annotation)

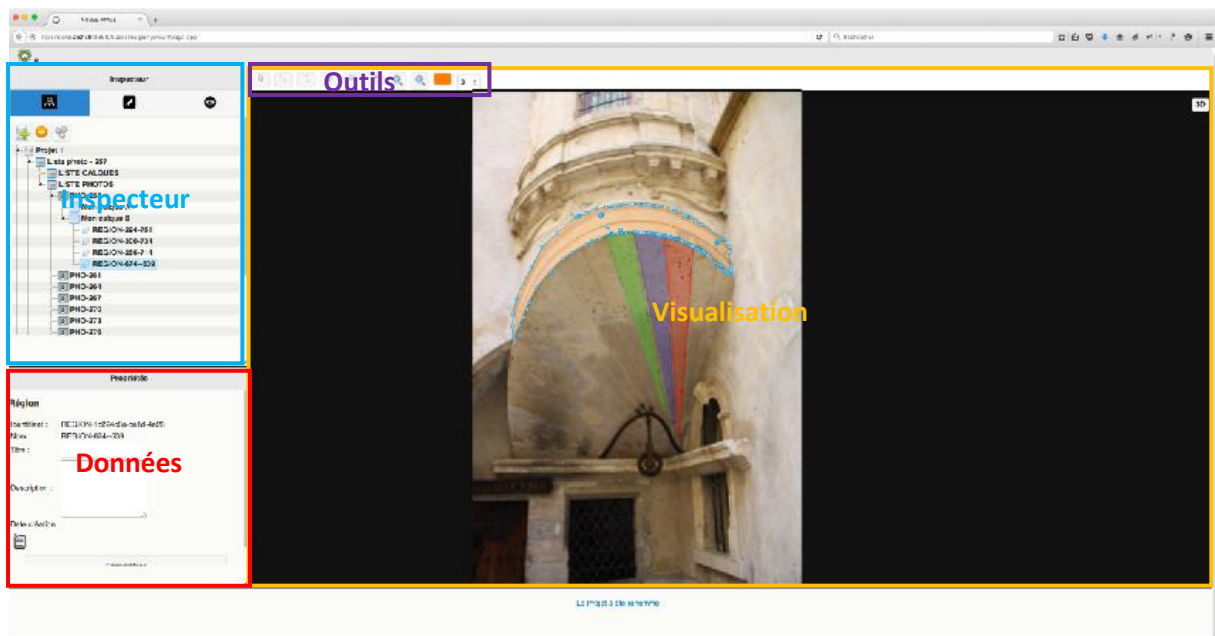


Figure 154 : Aperçu de l'interface web (pour le détail voir [Annexe B](#))

L'interface permet donc à l'utilisateur d'annoter les images puis de traiter les données en les envoyant sur un serveur de calculs où sont stockés les modules de calculs développés sous MATLAB et compilés en exécutables. Afin de pouvoir utiliser ces exécutables, le MCR doit être installé sur le serveur de calcul (Figure 155).

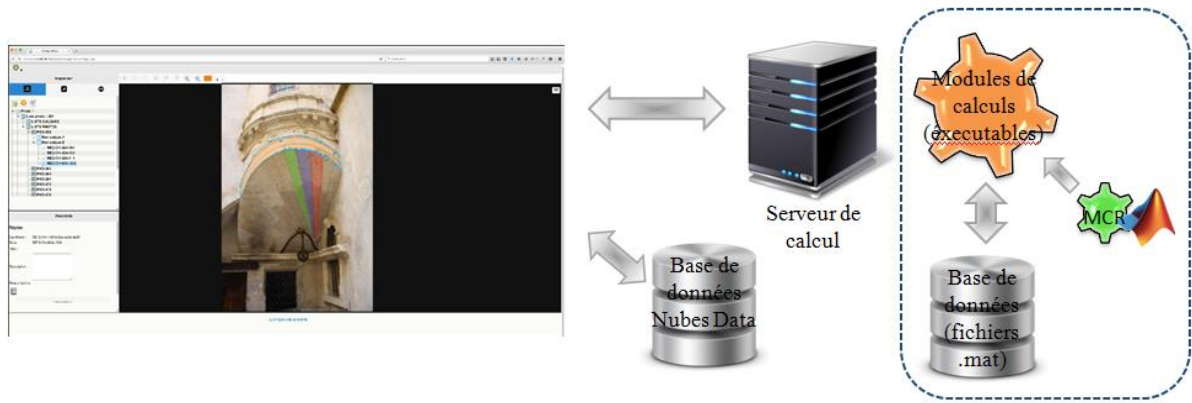
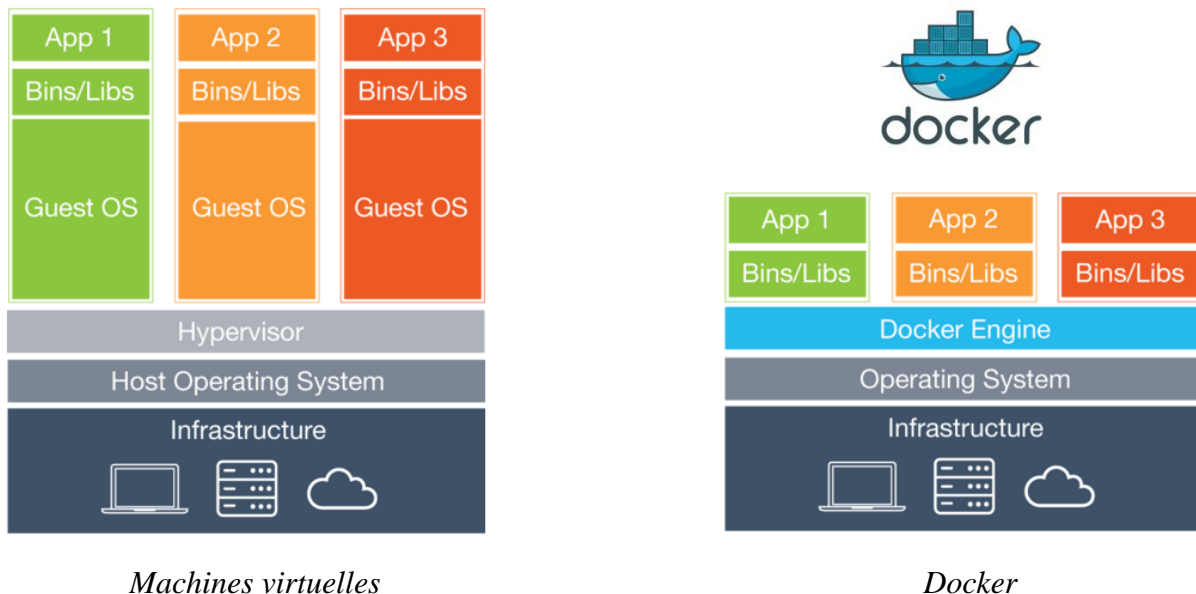


Figure 155 : Fonctionnement général de l'interface web

Le lancement de calculs à distance sur un serveur de calcul passe généralement par le déploiement de machines virtuelles. Cependant une solution est actuellement en train d'émerger afin de faciliter le déploiement d'applications et de gestion du dimensionnement de l'infrastructure sous-jacente. Cette solution appelée Docker [Docker] est une solution open source qui permet d'embarquer une application dans un container virtuel pouvant s'exécuter sur n'importe quel serveur. L'intérêt majeur de Docker, par rapport à l'utilisation de machines virtuelles, est sa légèreté car il n'embarque pas d'OS supplémentaire (Figure 156). Cela se traduit par un lancement plus rapide mais aussi par la capacité de migrer facilement un container d'une machine physique à l'autre. Ainsi l'intégration de Docker pour le lancement des calculs à distance pour cette interface est actuellement en implémentation.



Machines virtuelles

Docker

Figure 156 : Différence de structuration entre Docker [Docker] et les machines virtuelles

La base de données utilisée pour cette interface est une base de données SQL basée sur le schéma Nubes Data (Figure 157). Une autre base de données toujours en format Mat est, elle stockée sur le serveur de calcul et sert pour le stockage des données liées au calcul.

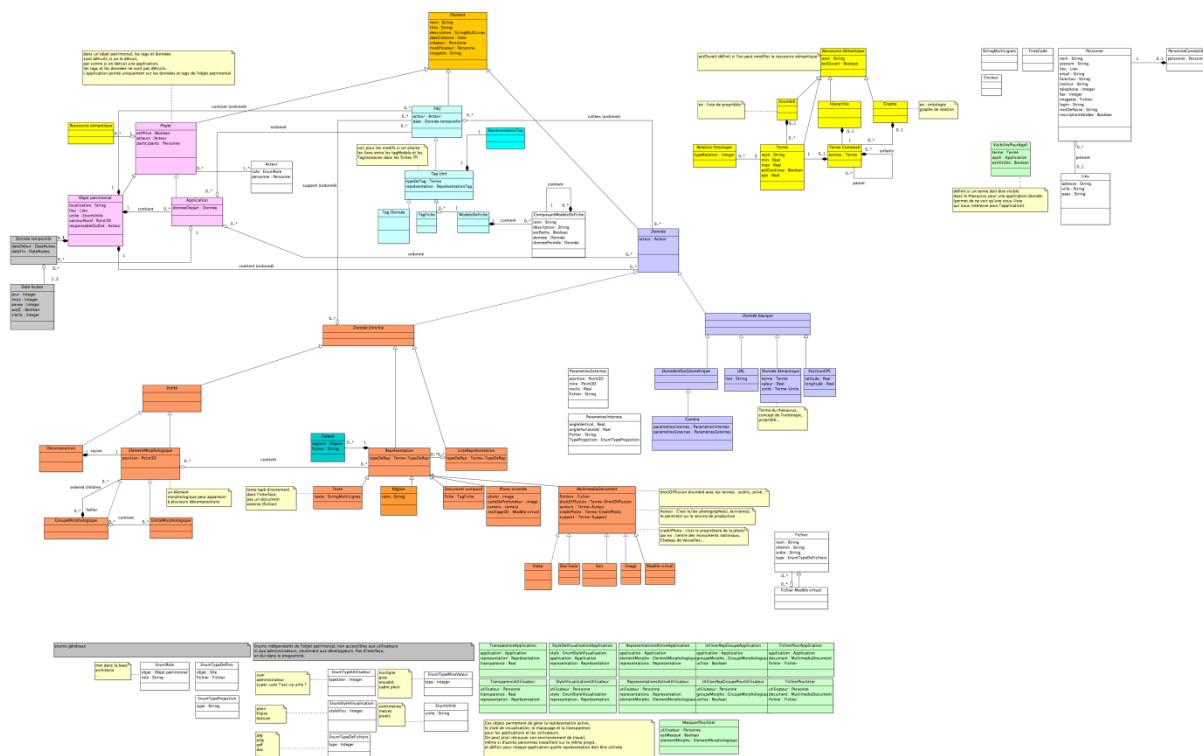


Figure 157 : Schéma global de Nubes Data (pour le détail voir en Annexe C)

L'utilisation de cette base de données permet de pouvoir construire des fiches de descriptions de manière plus précise que simplement par nom et commentaires.

---

## IX. Conclusion

Notre travail s'est concentré sur la conception et le développement d'une méthodologie pour l'annotation sémantique de photographies par référencement spatial. Le point central de ce travail était d'exploiter la relation spatiale entre l'information 3D (nuage de points) et l'information 2D (pixels des photographies). Notre méthode permet de définir et de visualiser des annotations sémantiques pour un objectif de description et d'analyse. La relation spatiale établie entre les informations 2D et 3D est utilisée pour véhiculer les annotations sémantiques entre les photographies de l'ensemble et le nuage de points.

La précision dans le transfert des annotations sémantique dépend principalement du processus de reconstruction du nuage de points et du référencement spatial des photographies. Le niveau de segmentation des images dépend essentiellement de l'utilisateur permettant d'envisager de conduire de multiples segmentations et annotations sémantiques en 2D en fonction des exigences de description et d'analyses. D'autre part, il est possible d'imaginer de projeter du 2D vers le 3D et du 3D vers le 2D, beaucoup d'autres informations qui peuvent servir à enrichir la description sémantique de l'objet.

Plusieurs maquettes informatiques du système de propagation d'annotation ont été développées. Ses fonctionnalités sont disponibles pour les utilisateurs souhaitant commencer à expérimenter le système sur des applications concrètes tout en gardant à l'esprit qu'il s'agit pour l'instant de premières implémentations qui n'ont pas été encore optimisées. Le système donne la possibilité d'accéder à une information riche à propos d'un objet et de manière innovante et pourrait facilement être employé par des chercheurs ou des spécialistes de la documentation et de la conservation.

En comparant notre système avec d'autres travaux de recherche sur l'annotation sémantique d'images, notre système requiert moins de manipulation que des solutions basées sur des méthodes manuelles et reste plus précis que les solutions basées sur les méthodes automatiques ou semi-automatiques. De plus, notre système intégrant l'aspect tridimensionnel de l'objet, il propose une manipulation plus aisée dans la définition des annotations que [Busayarat C. et al, 2010] et offre des résultats de propagation plus précis que [Snaveley N. et al, 2006].

De plus, par la relation spatiale, notre système associe à une image toute une richesse informationnelle contenue dans sa représentation tridimensionnelle (dimension, géométrie, etc.). D'autre part, la solution proposée propose la possibilité de mise à jour de la base d'images orientée autour du nuage de points, ceci intégrant le problème de la gestion de différents états temporels.

### IX.A. Limites

Bien que l'approche proposée réponde à un certain nombre d'exigences du domaine de la représentation et de la documentation du patrimoine, certains aspects comportent aujourd'hui des limitations qui nécessitent encore d'être analysées de façon approfondie.

### IX.A.1. Protocole d'acquisition des photographies

Notre méthode se base sur le référencement spatial autour du nuage de points. La qualité du référencement et du nuage de points est donc vraiment importante. Ceci étant relatif à l'utilisation de MICMAC, il est donc nécessaire que les photographies prises pour la génération du nuage de points suivent un protocole précis inhérent à MICMAC en fonction de l'objet à analyser [Martin-Beaumont N. et al, 2012]. Ce protocole nécessite de posséder un minimum de connaissance en prises photographiques et en photogrammétrie. C'est pourquoi l'acquisition des photographies reste aujourd'hui limitée à un public averti.

### IX.A.2. Intégration d'images de détails

Lors de la recherche de correspondance 3D avec une image, le choix du redimensionnement de l'image pour obtenir une bonne densité de correspondances sur l'image est aisée à définir pour les images ayant servi à la génération du nuage et représentant l'objet à une même échelle. Dans le cas d'images voulant concentrer l'attention sur un détail de l'objet, il est plus difficile d'estimer jusqu'où pousser le redimensionnement de l'image pour la projection tout en maintenant un nombre de pixels suffisant pour obtenir assez de correspondances 3D. La densité du nuage de points doit pouvoir suivre tous les niveaux d'échelle pour les photographies ou au moins intégrer une densité locale plus importante pour des images de détails.

### IX.A.3. Densité du nuage de points et objets cachés

Un autre problème pouvant causer de nombreuses erreurs peut apparaître lors de la recherche des correspondances. Dans le cas où des zones du nuage ne sont pas assez denses, notamment s'il y a des trous dans des régions du nuage de points, la recherche de correspondances 3D sur les images peut, au vu de son implémentation actuelle, trouver une correspondance pour des points situés derrière la partie visible de l'objet. En effet, s'il n'y a pas de points plus proches, c'est le point de la partie arrière qui sera projeté. Ce problème peut par la suite entraîner des erreurs de propagation au niveau des annotations.

### IX.A.4. Gestion d'états temporels

Même si les premières expérimentations menées sur l'utilisation du référencement spatial pour pouvoir mettre en relation différents états temporels sont encourageantes, l'implémentation actuelle reste très limitée pour une utilisation efficace de ce principe.

Il serait aussi intéressant d'intégrer une représentation visuelle de l'évolution de ces états temporels et des annotations liées avec par exemple l'utilisation de graphes de transformations tels que utilisés dans les travaux de [Stefani C. et al, 2010] (Figure 158). Ces travaux n'intègrent qu'une visualisation liée à une représentation évolutive 3D. Cependant l'adaptation de ce type de représentation à notre système est envisageable en prenant en compte les aspects d'évolution du modèle, des images et des annotations.

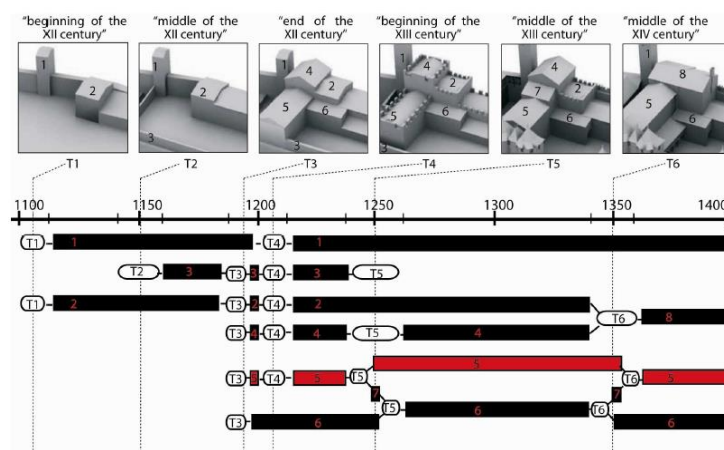


Figure 158 : Graphes de transformations [Stefani C. et al, 2010]

## IX.B. Perspectives de recherche

Les limites identifiées nécessitent donc d'être améliorées afin de pouvoir les intégrer au système de façon efficace et permettre une utilisation aisée de tous ces aspects.

### IX.B.1. Evolutions futures

Dans le cas d'ajout d'images de détails, le problème reste donc de définir un protocole d'acquisition précis afin de définir les conditions de prises de vue et d'intégration prenant en compte les aspects d'orientation et de corrélation locale pour pouvoir permettre d'ajouter de manière efficace ces images de détails.

Les mécanismes d'auto-extension de régions pour la sélection des annotations permettent d'exploiter les informations liées à une image par la relation projective pour définir des solutions de sélection multicritère. De nouveaux mécanismes d'auto-extension pourraient étendre et faciliter les possibilités d'extraction de données. Par exemple, une région peut être étendue à l'aide des normales puis propager sur les autres images. Cependant la projection de la région sur les autres images peut elle aussi être étendue que ce soit à l'aide des normales ou d'autres informations, puis à nouveau propagée. Ainsi en faisant un aller-retour continu entre 2D et 3D, la définition d'une région peut être grandement facilitée et ne pas dépendre d'un seul point de vue.

La prise en compte d'un référencement temporel en plus du référencement spatial ouvrirait des possibilités d'étude de l'évolution d'un objet afin de pouvoir suivre les transformations que l'objet subit dans le temps. Le fait de disposer de plusieurs représentations (2D et 3D) à différents temps permettrait des comparaisons sur plusieurs niveaux pour suivre cette évolution. Ceci nécessite de définir une méthodologie précise pour le référencement spatial de différents états temporels et l'utilisation de la relation spatiale entre les temps mais aussi de réétudier la structuration des données afin de permettre des intégrations efficaces et des possibilités d'interaction innovantes entre différents états d'un bâtiment ou d'un objet.

D'autre part, le système ne permet d'intégrer que des photographies, qui sont des supports possédant une cohérence géométrique très importante par rapport à la réalité. Pour établir une complète documentation d'un objet, il est nécessaire de pouvoir aussi intégrer d'autres types



de sources iconographiques (peinture, dessin, etc.) ayant souvent de grandes incohérences géométriques avec la réalité. Ainsi l'intégration de sources iconographiques autre que les photographies enrichirait les données.

La sélection d'une région sur une image permet l'extraction rapide de régions 3D représentant l'annotation. La facilité par laquelle ces régions sont extraites peut servir de moyen pour la constitution de base de formes 3D. Cette génération de base de références par le cumul d'annotations pourrait servir aux méthodes par apprentissage (*Machine Learning*) pour la reconnaissance automatique de formes.

Ensuite, le système n'envisage pour l'instant que des outils d'analyses quantitatives relativement simples. Cependant au vu du niveau d'informations contenues dans une annotation (représentation 3D et projection 2D), de nombreux autres outils pourraient être développés tels que des outils basés sur la détection de primitives géométriques. Il serait notamment intéressant d'envisager la possibilité de combiner les résultats des différents outils afin d'intégrer plusieurs niveaux d'analyses au sein du même système.

Le système suppose pour l'instant que l'objet étudié soit unique par rapport à des données. Cependant il pourrait être intéressant au vu des possibilités d'analyses fournies par le système de pouvoir gérer des collections d'objets et pouvoir décrire des ensembles d'objets par analyse et comparaison des formes.

Enfin, il reste à développer une interface complète intégrant tous les différents aspects possibles pour permettre à un utilisateur d'établir des analyses à un niveau plus poussé permettant à la fois une gestion multi-support, multi-projection, multi-couche, et multi-temporelle.

#### IX.B.2. Application mobile pour l'annotation sur site

La possibilité de l'ajout d'une nouvelle image qu'elle soit simplement une image couleur ou une image d'un autre type offre des perspectives pour des applications intéressantes. En effet, avec l'évolution des technologies mobiles ces dernières années, des applications mobiles sont envisageables.

En effet, les appareils mobiles tels que les smartphones et les tablettes permettent aujourd'hui de prendre des photographies, de fournir les données GPS de ces images, d'interagir avec les images et de communiquer à distance avec des serveurs. Cependant ce ne sont pas des appareils suffisamment puissants pour pouvoir établir des calculs lourds.

Ainsi ces appareils apparaissent comme des supports mobiles pour permettre l'annotation d'objets directement sur site. Les calculs ne pouvant être réalisés directement sur l'appareil, la possibilité de communiquer avec des serveurs peut combler ce manque en établissant des processus de cloud computing traitant les données à distance et renvoyant les résultats sur l'appareil mobile (Figure 159).

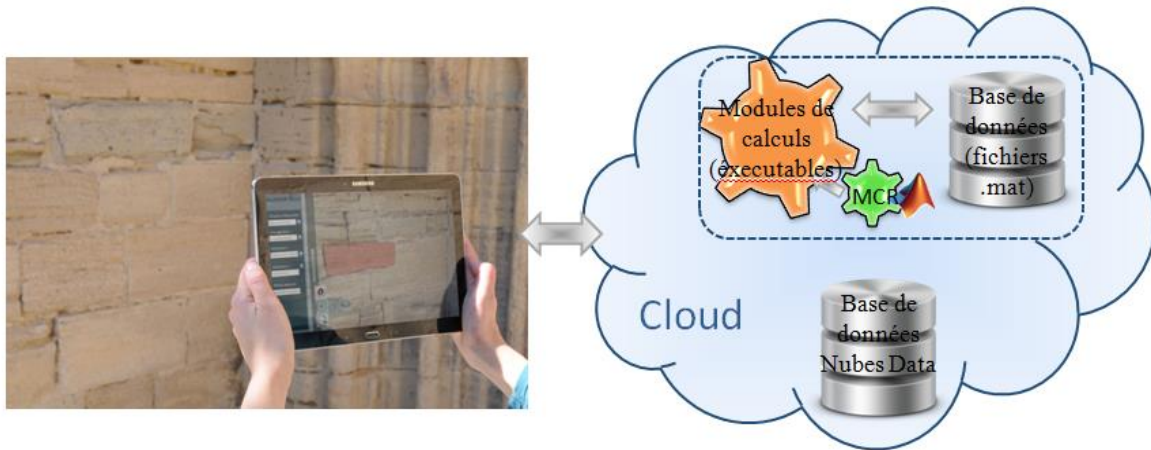


Figure 159 : Principe de l'application mobile

Une telle application supposerait d'avoir déjà mis en place un premier relevé complet de l'objet pour poser les bases du projet. Les photographies prises par l'appareil pourraient permettre de compléter les données sur le site. Ces photographies pourront donc servir à la fois pour voir des annotations déjà faites sur le projet mais pourront aussi servir pour ajouter d'autres annotations aux images du projet.

Si les problèmes géométriques ont déjà été majoritairement résolus, la mise en place d'une structure de traitement à distance nécessite encore quelques approfondissements.



## Bibliographie

- [123D Catch] [www.123dapp.com/catch](http://www.123dapp.com/catch) (Janvier 2016)
- [3DIcons] <http://3dicons.ceti.gr/> (Janvier 2016)
- [3DSA, 2014] <http://3dsa.metadata.net/3dsa/> (Janvier 2016)
- [AbdelAziz Y.I. et al., 1971] AbdelAziz Y.I., Karara H.M., Direct Linear Transformation from comparator coordinates into object space coordinates in close range photogrammetry. Proceedings of the ASP/UI.1971
- [Adrian J. et al., 2014] Adrian J., Lo Buglio D., De Luca L., Evaluating the Curvature Analysis as a Key Feature for the Semantic Description of Architectural Elements, Eurographics Workshop on Graphics and Cultural Heritage - Short Papers / Posters, 2014
- [Ahmadabadian A.H. al. 2013] Ahmadabadian A. H., Robson, S., Boehm, J., Shortis, M., Wenzel, K., Fritsch, D., A comparison of dense matching algorithms for scaled surface reconstruction using stereo camera rigs. In: ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 78, pp. 157-167, 2013
- [Alahi A. et al, 2012] Alahi A., Ortiz R., Vanderghenst P., FREAK: Fast REtinA Keypoints, In Computer Vision and Pattern Recognition, CVPR 2012, 2012 IEEE Internatinal Conference on pp. 510-517
- [American Society of Photogrammetry, 1980] American Society of Photogrammetry, Manual of Photogrammetry, Fourth Édition, 1980
- [Andrews P. et al, 2012] Andrews P., Zaihraye I., Pane J., A Classification of Semantic Annotation Systems, Semantic Web, IOS Press Amsterdam, Vol.3, Issue 3, pp 223-248, August 2012
- [Arbeláez P.A. et al, 2006] Arbeláez P.A., Cohen L.D., A metric approach to vector-valued image segmentation, Int. J. Comp. Vis. 69(1), 119-126, 2006
- [Arbeláez P.A. et al, 2011] Arbeláez P.A., Fowlkes C., Contour detection and hierarchical image segmentation, IEEE Trans. Pattern Anal. Mach. Intel., 33(5), 898-916, 2011
- [Attene M. et al., 2007] Attene M., Robbiano F., Spagnuolo M. et Falcidieno B., Semantic Annotation of 3D Surface Meshes based on Feature Characterization, 2007
- [Attene M. et al., 2009] Attene M., Robbiano F., Spagnuolo M., Falcidieno B., Part-based Annotation of Virtual 3D Shapes, 2009
- [Azzi Y., 2008] Azzi Y. : Images et contenus : classification et sémantisation de sources documentaires visuelles, MIAjournal, 2008
- [Bannour H. et al, 2011] Bannour H., Hudelot C., Towards Ontologies for Image Interpretation and Annotation, IEEE, 2011

- [Bay H. et al, 2006] Bay H., Tuytelaars T., Van Gool, L., SURF: Speeded Up Robust Feature, In Computer Vision, ECCV 2006, pp. 404-417, Springer, 2006
- [Bayer B., 1976] US Patent, 3,971,065, <https://www.google.com/patents/US3971065>
- [Bederson B.B., 2001] Benjamin B. Bederson, PhotoMesa: a zoomable image browser using quantum treemaps and bubblemaps, Proceedings of the 14th annual ACM symposium on User interface software and technology, November 11-14, 2001, Orlando, Florida
- [Berkmann J. et al., 1994] Berkman J. and Caelli T., Computation of Surface Geometry and Segmentation Using Covariance Techniques. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 16(11):1114-1116, November 1994.
- [Berners-Lee T. et al, 2001] Berners-Lee T., Hendler J., Lassila O., The semantic Web, Scientific American, Vol. 284, No.5, pp34-43, 2001
- [Bigas M. et al, 2006] Bigas M., Cabruja E., Forest J., Salvi J., Review of CMOS image sensors, Microelectron. J., 37(5), pp433-451, 2006
- [Bilasco I.M. et al, 2005] Bilasco I.M., Gensel J., Villanova-Oliver M., Martin H., 3DSEAM: a model for annotating 3D scenes using MPEG-7, 2005
- [Blaise, 2003] Blaise J-Y. Le formalisme objet appliqué à l'étude de l'édifice patrimonial : Problèmes de modélisation et d'échanges de données sur le réseau Internet. Thèse de doctorat de l'Université d'Aix-Marseille III, Spécialité Automatique et Informatique, 2003.
- [Blaise J.Y. et al, 2004] Blaise J.Y., De Domenico F., De Luca L., Dudek I: Acquisition de données vs gestion de connaissances patrimoniales : le cas des vestiges du théâtre antique d'Arles, 2004
- [Bodin B. et al, 2004] Bodin B., Frich A., Lemoine A., Noël C., Popovic S., Simard P., Thion L., Vidal G., Photos Panoramiques par Assemblage, Eyrolles, 2004
- [Boehler W. et al., 2002a] Boehler W., Marbs A. 3D Scanning Instruments. Proceedings of the CIPA WG 6 International Workshop on Scanning for Cultural Heritage Recording. Corfou, 2002
- [Boehler W. et al., 2002b] Boehler W., Heinz G., Marbs A., Siebold M. 3D Scanning Software: an introduction. Proceedings of the CIPAWG 6 International CIPA WG Workshop on Scanning for Cultural Heritage Recording. Corfu, 2002
- [Bohler W. et al., 2003] Boehler W., Bordas V., Marbs A. Investigating Laser Scanner Accuracy. Proceedings of the International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXIV, Part 5/C15, pp. 696-701, 2003
- [Bonn F. et al, 1992] Bonn F., Rochon G., Précis de Télédétection, Volume 1 : Principes et méthodes, AUPELF-UREF, coll. Presses Universitaires du Québec, Montréal, 1992
- [Bosch A. et al, 2006] Bosch A., Zisserman A., Munoz X., Scene Classification via pls, In Computer Vision, ECCV 2006, pp. 517-530, Springer, 2006



- 
- [Boust C. et al, 2007] Boust C., Ezrati J.J., La mesure de la couleur appliquée à la restauration, à la présentation et à la diffusion des œuvres d'art, *Technè* n°26, 2007
- [Boykov Y.Y. et al, 2006] Boykov Y.Y., Lea G.F., Graph cuts and efficient N-D image segmentation, *Inter. J. Comp., Vis.*, 70(2), 109-131, 2006
- [Brown M. et al, 2011] Brown M., Sussstrunk S., Multi-spectral SIFT for Scene Category Recognition, In *Computer Vision and Pattern Recognition (CVPR)*, 2011 IEEE Conference on pp. 177-184, 2011
- [Bundler] <http://www.cs.cornell.edu/~snavely/bundler/> (Janvier 2016)
- [Burghouts G.J., et al, 2009] Burghouts G.J., Geusebroek J.M., Performance Evaluation of Local Colour Invariants, *Computer Vision and Image Understanding*, 113, pp. 48-62, 2009
- [Busayarat C., 2010] Busayarat C., La maquette numérique comme support pour la recherche visuelle d'informations patrimoniales, 2010
- [C3DC] <http://c3dc.fr/> (Janvier 2016)
- [Calonder M. et al, 2010] Calonder M., Lepetit V., Strecha C., Fua P., BRIEF: Binary Robust Independent Elementary Features, In *Computer Vision, ECCV 2010*, pp. 778-792, Springer, 2010
- [Carbonnell M., 1968] Carbonnell M. L'histoire et la situation présente des applications de la photogrammétrie à l'architecture. Etude sur la photogrammétrie appliquée aux monuments historiques. Application of photogrammetry to historic monuments. ICOMOS, Saint-Mande, France, 4-6, VII, 1968
- [Carneiro G. et al, 2007] Carneiro G., Chan A.B., Moreno P.J., Vasconcelos N., Supervised Learning of Semantic Classes for Image Annotation and Retrieval, *IEEE PAMI* 29(3), pp.394-410, 2007
- [Carson C. et al, 2002] Carson et al, Blobworld: image segmentation using expectation-maximization and its application to image querying, *IEEE Trans. Pattern Anal. Mach. Intel.*, 24(8), 1026-1038, 2002
- [Chapelle O. et al, 1999] Chapelle O., Heffner P., Vapnik V.N., Support Vector Machines for Histogram-Based Image Classification, *IEEE Transactions on Neural Network*, Vol. 10, No. 5, pp. 1055-1064, 1999
- [Chen H.C. et al, 2004] Chen H.C., Chien W.J., Wang S.J., Contrast-based color image segmentation, *IEEE Sign. Proc. Lett.*, 11(7), 641-644, 2004
- [Chen J. et al, 2005] Chen J., Pappas T.N., Mojsilovic A., Rogowitz B., Adaptive perceptual color-texture image segmentation, *IEEE Trans. Image Proc.*, 14(10), 1524-1536, 2005
- [Comaniciu D. et al, 2002] Comaniciu D., Meer P., Mean Shift: a robust approach towards feature space analysis, *IEEE Trans. Pattern Anal. Mach. Intel.*, 24(5), pp. 603-619, 2002
-

- [Conrat J.M. et al, 2010] Conrat J.M., Pajusco P., Dunand A., Utilisation des photos panoramiques pour la compréhension des phénomènes physiques du canal de propagation, *Revue Générale de l'Electricité (REE°*, Vol.2, Février 2010
- [Dalal N. et al, 2005] Dalal N., Triggs B., Histograms of Oriented Gradients for Human Detection, In *Computer Vision and Pattern Recognition, CVPR 2005*, IEEE Computer Society Conference on Vol.1, pp. 886-893, 2005
- [De Luca L., 2009] De Luca L., *La Photomodélisation Architecturale : Relevé, Modélisation et Représentation d'Edifices à partir de Photographies*, Ed. Eyrolles, 264pp, ISBN: 978-2-212-12524-5, 2009
- [De Luca L. et al, 2010] De Luca L. Busayarat C., Stefani C., Renaudin N., Florenzano M., Véron P., An Iconography-Based Modeling Approach for the Spatio-Temporal Analysis of Architectural Heritage, *SMI 2010, Shape Modeling International Conference*, Aix-en-Provence, France, June 21-23, 2010
- [Devaux J.C. et al, 2001] Devaux J.C., Gouton P., Truchetet F., Karhunen-Loève transform applied to region-based segmentation of color aerial images, *Opt. Eng.* 40(7), 1302-1308, 2001
- [Deng Y. et al, 2001] Deng Y., Manjunath B.S., Unsupervised segmentation of color-texture regions in image and video, *IEEE, Trans. Pattern Anal. Mach. Intel.*, 23(8), 800-810, 2001
- [Docker] <https://www.docker.com/> (Janvier 2016)
- [Dumitras A. et al, 2001] Dumitras A., Venetsanopoulos A.N., Angular map-driven snakes with application to object shape description in color image, *IEEE Trans. Image Proc.*, 10(12), 1851-1859, 2001
- [Dyugulu P. et al, 2002] Dyugulu P., Barnard K., De Freitas N., Forsyth D., Object Recognition as Machine Translation: Learning a Lexicon for a Fixed Image Vocabulary, In *7th European Conference on Computer Vision (ECCV)*, Vol. 4, pp. 97-112, 2002
- [El-Hakim S. et al. 2003] El-Hakim, S., Beraldin, J.-A., Blais, F., Critical Factors and Configurations for Practical 3D Image-Based Modeling. *6th Conf. on 3D Measurement Techniques*, Zurich, Switzerland, Vol. II, pp. 159-167., 2003
- [El-Iaithy R.A. et al, 2012] El-Iaithy R.A., Huang J., Yeh M. Study on the use of Microsoft Kinect for Robotics Applications, in *IEEE On Positions Location and Navigation Symposium (PLANS)*, pp 1280-1288, USA, April 2012
- [Elias M. et al., 2008] Elias M., Cotte P., Multispectral camera and radiative transfer equation used to depict Leonardo's sfumato in Mona Lisa, *Applied Optics*, Optical Society of America, 2008, 47 (12), pp.2146-2154
- [Erdogan C. et al, 2012] Erdogan C., Paluri M., Dellaert F., Planar Segmentation of RGBD Images using Fast Linear Fitting and Markov Chain Monte Carlo, *Computer and Robot Vision*, 2012

- 
- [EXIF, 2010] Exif Version 2.3, Japan Electronics and Information Technology Industries Assoc. (JEIDA), [http://www.cipa.jp/std/documents/e/DC-008-2012\\_E.pdf](http://www.cipa.jp/std/documents/e/DC-008-2012_E.pdf) (Janvier 2016), Avril 2010
- [Fan J. et al, 2001] Fan J., Yau D.K.Y., Elmagarmig A.K., Aref W.G., Automatic image segmentation by integrating color-edge extraction and seeded region-growing, *IEEE Trans. Image Proc.* 10 (10), 1454-1466, 2001
- [Faugeras O., 1992] Faugeras O. et al., Camera selfcalibration: theory and experiments, G. Sandini (ed.), *Proc 2nd ECCV*, Vol. 588 of *Lecture Notes in Computer Science*, SpringerVerlag, Santa Margherita Ligure, Italy, pp. 321–334, 1992
- [Feng S.L. et al, 2004] Feng S.L., Manmatha R., Lavrenko V., Multiple Bernoulli Relevance Models for Image and Video Annotations, In *Proceedings of the CVPR04*, pp. 1002-1009, 2004
- [Fitzgibbon A.W. et al., 1998] Fitzgibbon A.W., Zisserman A. Automatic 3D Model Acquisition and Generation of New Images from Video Sequences. *Proceedings of European Signal Processing conference (EUSPICO '98)*. Rhodes, Greece. pp. 1261-1269, 1998
- [Flickner M. et al., 1995] Flickner M., Sawhney H., Niblack W., Ashley J., Qian Huang, Dom B., Gorkani M., Hafner J., Lee D., Petkovic D., Steele D., Yanker P., Query by Image and Video Content. the QBIC system. *IEEE Computer*, 28(9), 23-32, 1995
- [FLICKR] <http://www.flickr.com/> (Janvier 2016)
- [Flusser J., 2000] Flusser J., On the Independence of Rotation Moment Invariants, *Pattern Recognition*, 33(9), pp. 1405-1410, 2000
- [Flusser J., 2002] Flusser J., On the Inverse Problem of Rotation Moment Invariants, *Pattern Recognition*, 35(12), pp. 3015-3017, 2002
- [Fondon I. et al, 2006] Fondon I., Serrano C., Acha B., Color-texture image segmentation based on multistep region growing, *Opt. Eng.*, 45(5), 057002, 2006
- [Förstner W. et al, 1987] Förstner W., Gülch E., A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centres of Circular Features, In *Proc. ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data*, pp. 281-305, 1987
- [Förstner W., 1994] Förstner W., A Framework for Low Level Feature Extraction, In *Computer Vision, ECCV 1994*, pp.383-394, Springer, 1994
- [Fu H. et al, 2015] Fu H., Xu D., Lin S., Liu J., Object-Based RGBD Image Co-Segmentation With Mutex Constraint, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp 4428-4436, 2015
- [Fuchs A. et al. 2004] Fuchs A., Alby E., Réda B., Grussenmeyer P., Perrin J-P. Confrontation du relevé laser 3D aux techniques de relevé conventionnelles et de développement d'outils numériques pour la restitution architecturale. *Revue Française de Photogrammétrie et de Télédétection*, n°173/174 (2004-1/2), pp. 36-47, 2004
-

- [Fukunaga K. et al, 1975] Fukunaga K., Hosteler L.D., The estimation of the gradient of a density function, with applications in pattern recognition, IEEE Trans. Inform. Theor., 21(1), pp. 32-40, 1975
- [Fyfe D., 2007] Fyfe D., La description des formes d'un objet en dessin technique, [www.unites.uqam.ca/enssc/projets/Techno/Boite/Types%20de%20projection.pdf](http://www.unites.uqam.ca/enssc/projets/Techno/Boite/Types%20de%20projection.pdf) (Janvier 2016), 2007
- [Gabriel C., 2015] Gabriel C., Cours d'optique photo : Chapitre 9 La perspective en Photographie, [www.claudegabriel.be](http://www.claudegabriel.be) (Janvier 2016), 2015
- [Gao H. et al, 2001] Gao H., Siu W.C., Hou C.H., Improved techniques for automatic image segmentation, IEEE Tras. Circuits Sys. Video Tech., 11(12), 1273-1280, 2001
- [Gao J. et al 2002] Gao J., Zhang J., Fleming M.G., Novel technique for multiresolution color image segmentation, Opt. Eng., 41(3), 608-614, 2002
- [Gelfand N. et al, 2005] Gelfand N., Mitra N.J., Guibas L.J., Pottmann H., Robust Global Registration, In Symposium on Geometry Processing, Vol. 2, 2005
- [Gong Y. et al., 1994] Gong Y., Zhang H., Chuan H.C., Sakauchi M., Animage database system with content capturing and fast image indexing abilities. In: Proceedings of IEEE Int. Conf. on Multimedia Computing and Systems, 1994
- [GoogleLabResearch, 2014] <http://googleresearch.blogspot.fr/2014/11/a-picture-is-worth-thousand-coherent.html> (Janvier 2016)
- [Grussenmeyer P. et al, 2001] Grussenmeyer P., Hanke K., Streilein A. Photogrammétrie architecturale. Chapitre du livre, Photogrammétrie numérique. Éditions Lavoisier-Hermès, 2001
- [Guarnieri A. et al., 2004] Guarnieri A., Vettore A. El-Hakim S., Gonzo L. Digital photogrammetry and laser scanning in ultural heritage survey. Proceeding of the XXth ISPRS Congress, Istanbul, Turkey, 2004
- [Gupta M. et al, 2011] Gupta M., Agrawal A., Veeraraghavan A., Narasimhan S.G., Structured Light 3D Scanning in the Presence of Global Illumination, IEEE Computer Vision and Pattern Recognition (CVPR), June 2011
- [Halaschek-Wiener C. et al., 2005] Halaschek-Wiener C., Jennifer G., Andrew S., Michael G., Bijan P., Jim H.: PhotoStuff -- An Image Annotation Tool for the Semantic Web, 2005
- [Haralick R.M. et al, 1991] Haralick R.M., Shapiro L.G., Computer and Robot Vision, Addison-Wesley Longman Publishing CO., Inc., 1991
- [Harris C. et al, 1988] Harris C., Stephens M., A Combined Corner and Edge Detector, In Alvey Vision Conference, Vol.15, p. 50, 1988
- [Havemann S. et al., 2008] Havemann S., Settgast V., Berndt R., Eide O. and Fellner D.W. The Arrigo Showcase Reloaded – Towards a sustainable link between 3D and semantics, 2008
- [Heikkila J. et al., 1997] Heikkila J., Silven O. A Four-Step Camera Calibration Procedure with Implicit Image Correction. Proceedings of IEEE Computer Vision and Pattern Recognition, pp. 1106-1112, 1997

- 
- [Helling S. et al, 2004] Helling S., Seidel E., Biehlig W., Algorithms for Spectral Color Stimulus Reconstruction with a Seven-Channel Multispectral Camera, In Proc. 2<sup>nd</sup> European Conference on Color in Graphics, Imaging and Vision CGIV 2004, Aachen, Germany, April 5-8, pp 254-258, 2004
- [Henry P. et al, 2012] Henry P., Krainin M., Herbst E., Ren X., Fox D., RGB-D Mapping: Using Kinect-Style Depth Cameras for Dense 3D Modelling of Indoor Environments, In The International Journal of Robotics Research, Vol. 31, No.5, pp 647-663, April 2012
- [Holtz D. et al, 2012] Holtz D., Holzer S., Rusu R.B., Behnke S., Real-Time Plane Segmentation Using RGB-D Cameras, RoboCup 2011: Robot Soccer World Cup XV, pp 306-317, 2012
- [Hu M.K., 1962] Hu M.K., Visual Pattern Recognition by Moment Invariants, Information Theory, IRE Transactions on, 8(2), pp. 179-187, 1962
- [Hu Z. et al. 2006] Hu Z., Uchimura K., Fusion of Vision, GPS and 3D Gyro Data in Solving Camera Registration Problem for Direct Visual Navigation", International Journal of ITS, 4, 1, pp. 3-12, December 2006
- [Hunter J. et al, 2011] Hunter J., Yu C., Assessing the Value of Semantic Annotation Services for 3D Museum Artefacts, 2011
- [IPTC, 2014] IPTC Photo Metadata Standard, The International Press Telecommunication Council, <https://iptc.org/standards/photo-metadata/iptc-standard/> (Janvier 2015), October 2014
- [Jafar O.A.M., et al, 2012] Jafar O.A.M., Sivakumar R., A Study on Fuzzy and Particule Swarm Optimization Algorithms and their Applications to Clustering Problems, IEEE, International Conference on Advanced Communication Control and Computing Technologies (ICACCCT), 2010
- [Janesick J.R., 2001] Janesick J.R., Scientific Charge-Coupled Devices, SPIE Press, Bellingham, WA., Vol. PM83, pp1-920, 2001
- [Johansson B. et al, 2002] Johansson B., Cipolla R., A System for Automatic Pose-Estimation from a Single Image in a City Scene. In Proc. of International Conference on Signal Processing, Pattern Recognition, and Applications, Crete, Greece, 2002
- [Johnson A.E. et al, 1999] Johnson A.E., Hebert M., Using Spin Images for Efficient Object Recognition in Cluttered 3D Scenes, Pattern Analysis and Machine Intelligence, IEEE Transactions on 21(5), pp. 433-449, 1999
- [Kadobayashi R. et al., 2003] Kadobayashi R., Furukawa R., Kawai Y., Kanjo D., Yoshiomoto J.N., Integrated presentation system for 3D models and image database for a Byzantine ruins, International Archives of Photogrammetry Remote Sensing and Spatial Information Science, Vol.34, no.5/W12, pp.187-192, 2003
- [Kang H. et al., 2000] Kang H., Shneiderman B., Visualization Methods for Personal Photo Collections: Browsing and Searching in the PhotoFinder, Proc. IEEE International Conference on Multimedia and Expo, New York: IEEE, pp. 1539-1542, 2000
-



- [Kanungo T. et al, 2002] Kanungo T., Mount D.M., Netanyahu N.S., Piatko C.D., Silverman R., Wu A.Y., An efficient k-means clustering algorithm: analysis and implementation, *IEEE Trans. Pattern Anal. Mac. Intel.*, 24(7), 881-892, 2002
- [Kass M. Et al, 1988] Kass M., Witkin A., Terzopoulos D., Snakes: active contour models, *Inter. J. Comp. Vis.*, 1(4), pp.321-331, 1988
- [Kazhdan M. et al, 2003] Kazhdan M., Funkhouser T., Rusinkiewicz S., Rotation Invariant Spherical Harmonic Representation of 3D Shape Descriptors, In *Symposium on Geometry Processing*, Vol. 6, pp. 156-164, 2003
- [Kazhdan M. et al., 2006] Kazhdan, M., Bolitho, M., and Hoppe, H. Poisson surface reconstruction. In *Proceedings of the fourth Eurographics symposium on Geometry processing*, volume 7, 2006
- [Kazhdan M. et al., 2013] Kazhdan, M. and Hoppe, H. Screened poisson surface Reconstruction. *ACM Transactions on Graphics (TOG)*, 32(3), 29, 2013
- [Ke Y. Et al, 2004] Ke Y., Sukthankar R., PCA-SIFT: A More Distinctive Representation for Local Image Descriptors, In *Computer Vision and Pattern Recognition, CVPR 2004, Proceedings of the 2004 IEEE Computer Society Conference on* pp. 506-513, 2004
- [Kharkate S.K. et al, 2013] Kharkate S.K., Janwe N.J., Automatic Image Annotation : A Review, *The International Journal of Computer Science & Applications (TIJCSA)*, ISSN – 2278-1080, Vol.1, No.12, February 2013
- [Khan J.F. et al, 2009] Khan J.F., Adhami R.R., Bhuiyan S.M.A., A customized gabor filter for unsupervised color image segmentation, *Elsevier Image Vis. Comp.*, 27(4), 489-501, 2009
- [Kim C., 2005] Kim C., Segmenting a low-depth-of-field image using morphological filters and region merging, *IEEE Trans. Image Proc.*, 14(10), 1503-1511, 2005
- [Kim J.B. et al, 2003] Kim J.B., Kim H.J., Multiresolution-based watersheds for efficient image segmentation, *Elsevier Pattern Recogn. Lett.*, 24(1-3), 473-488, 2003
- [Klasing K. et al., 2009] Klasing K., Althoff D., Wollherr D., and Buss M., Comparison of Surface Normal Estimation Methods for Range Sensing Applications. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Kobe, Japan, May 12-17 2009.
- [Klyne G. et al, 2004] Klyne G., Carroll J.J., Resource Description Framework (RDF): Concept and Abstract Syntax, W3C Recommendation, World Wide Web, <http://www.w3.org/TR/rdf-concepts/> (Janvier 2016), 2004
- [Knopp J. et al, 2010] Knopp J., Prasad M. Willems G., Timofte R., Van Gool L., Hough Transform and 3D Robust Three Dimensional Classification, In *Computer Vision, ECCV 2010*, pp. 589-602, Springer, 2010
- [Kraus K. et al, 1997] Kraus K., Waldhauf P. Manuel de photogrammétrie : principes et procédés fondamentaux. Hermes, 1997

- 
- [Koutsoudis A. et al. 2013] Koutsoudis, A., Vidmar, B., Ioannakis, G., Arnaoutoglou F., Pavlidis, G., Chamzas, C., Multi-image 3D reconstruction data evaluation. In: Journal of Cultural Heritage, Vol. 15, Issue 1, pp. 73–79, 2013
- [Kowalski P., 1990] Kowalski P. Vision et mesure de la couleur, Masson, Seconde Edition, 1990
- [Kumar R. et al., 1994] Kumar R., Hanson A. Robust Methods for Estimating Pose and a Sensitivity Analysis. CVGIP-Image Understanding. Vol. 60, No. 3, 1994. pp 313-342, 1994
- [Lanh T.V. et al, 2007] Lanh T.V., Chong K.S., Emmanuel S., Kankanhalli S., A Survey on Digital Camera Image Forensic Methods, Proc. IEEE, Int. Conf. Multimedia Expo, pp. 16-19, 2007
- [Lee S.C. et al., 2002] Lee S.C., Jung S.K., Nevatia R., "Automatic Pose Estimation of Complex 3D Building Models", [PDF], IEEE Workshop on Application of Computer Vision, 148-152, 2002
- [Lee S.H., et al, 2010] Lee S.H., Koo H.I., Cho N.I., Image segmentation algorithms based on the machine learning of features, Elsevier Pattern Recogn. Lett., 31(14), 2325-2336, 2010
- [Leutenegger S. et al, 2011] Leutenegger S., Chli M., Siegwart R.Y., BRISK: Binary Robust Invariant Scalable Keypoints, In Computer Vision, ICCV 2011, 2011 IEEE International Conference on pp. 2548-2555
- [Li J. et al, 2006] Li J., Wang J.Z., Real-Time Computerized Annotation of Pictures, Proc. ACM Multimédia, pp. 911-920, 2006
- [Li Y., 1992] Li Y., Reforming the Theory of Invariants Moments for Pattern Recognition, Pattern Recognition, 25(7), pp. 723-730, 1992
- [Lin C-Y et al. 2003] Lin C-Y, Tseng B.L., Smith J.R., VideoAnnEx : IBM MPEG-7 annotation tool for multimedia indexing and concept learning, in IEEE International Conference on Multimedia and Expo, 2003
- [Liu J. et al, 2009] Liu J., Li M., Liu Q., Lu H., Ma S., Image Annotation via Graph Learning, Pattern Recognition, 42(2), pp. 218-228, 2009
- [Liu Z. et al, 2011] Liu Z., Shen L., Zhang Z., Unsupervised image segmentation based on analysis of binary partition tree for salient object extraction, Elsevier Sign. Proc., 91(2), 290-299, 2011
- [Lowe D.G., 1999] Lowe D.G., Object recognition from local scale-invariant features, Proceedings of the International Conference on Computer Vision, 1999
- [Ma Y. et al, 2013] Ma Y., Lévy F., Nazarenko A., Annotation Sémantique pour des Domaines Spécialisés et des Ontologies Riches, TALN-RECITAL 2013, 17-21 Juin, Les Sables d'Olonne, 2013
- [Mair E. et al, 2010] Mair E., Hager G.D., Burschka D., Suppa M., Hirzinger G., Adaptive and Generic Corner Detection Based on the Accelerated Segment Test, In Computer Vision, ECCV 2010, pp. 183-196, Springer, 2010
- [Malik J. et al, 2001] Malik J., Belongie S., Leung T., Shi J., Contour and texture analysis for image segmentation, Inter. J. Comp. Vis., 43(1), 7-27, 2001
-

- [Malladi R. et al, 1995] Malladi R., Sethian J.A., Vemuri B.C., Shape modelling with front propagation: a level set approach, *IEEE Trans. Pattern Anal. Mach. Intel.*, 17(2), 158-175, 1995
- [Manferdini A.M. et al., 2010] Manferdini A.M, Remondino F., Reality-Based 3D Modeling, Segmentation and Web-Based Visualization, 2010
- [Marbs A. et al., 2001] Marbs A., Heinz G., Boehler W. The potential of non-contact close range laser scanners for cultural heritage recording. *Proceedings of CIPA International Symposium*. Potsdam, Germany, 2001
- [Martin-Beaumont N. et al, 2012] Martin-Beaumont N., Nony N., Deshayes B., Pierrot-Deseilligny M., De Luca L., Photographer-Friendly Work-Flows for Image-Based Modelling of Heritage Artefacts, *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XL-5/W2, 2013 XXIV International CIPA Symposium, 2 – 6 September 2013, Strasbourg, France
- [Marqueed, 2014] [www.marqueed.com](http://www.marqueed.com) (Janvier 2016)
- [Messaoudi T. et al, 2014] T. Messaoudi, A. Manuel, E. Gattet, L. De Luca, P. Véron, Laying the foundation for an Information System Dedicated to Heritage Building Degradation Monitoring Based on the 2D/3D Semantic Annotation of Photographs, *Eurographics Workshop on Graphics and Cultural Heritage - Short Papers / Posters*, Darmstadt, Germany, 6-8 October 2014
- [Messaoudi T. et al, 2015] Messaoudi T, De Luca L., Véron P., Towards an ontology for annotating degradation phenomena, *Digital Heritage International Congress (DigitalHeritage)*, Sep 29-Oct 2, Granada, Spain, 2015
- [McQueen J., 1967] McQueen J., Some methods for classification and analysis of multivariate observations, in *Proc. 5<sup>th</sup> Berkeley Symp. Math. Statisti. Prob.*, Vol.1, pp. 281-296, University of California Press, Berkeley, California, 1967
- [Mian A.S. et al, 2006] Mian A., Bennamoun M., Owens R., Three-Dimensional Model-Based Object Recognition and Segmentation in Cluttered Scene, *Pattern Analysis and Machine Intelligence*, *IEEE Transactions on* 28(10), pp. 1584-1601, 2006
- [Michailovich O. et al, 2007] Michailovich O., Rathi Y., Tannenbaum A., Image segmentation using active contours driven by the Bhattacharyya gradient flow, *IEEE Trans. Image Proc.* 16(11), 2787-2801, 2007
- [Mikolajczyk K. et al, 2004] Mikolajczyk K., Schmid C., Scale & Affine Invariant Interest Point Detectors, *International Journal of Computer Vision*, 60(1), pp. 63-86, 2004
- [Mikolajczyk K. et al, 2005] Mikolajczyk K., Schmid C., A Performance Evaluation of Local Descriptors, *Pattern Analysis and Machine Intelligence*, *IEEE Transaction on* 27(10), pp. 1615-1630, 2005
- [Moffitt, F.H. et al., 1980] Moffitt, F.H., Mikhail E.M., *Photogrammetry*, 3rd Ed. Harper & Row, Inc. N.Y., U.S.A., 1980

- 
- [Moravec H.P., 1977] Moravec H.P., Towards Automatic Visual Obstacle Avoidance, In Artificial Intelligence (IJCAI), 5<sup>th</sup> International Joint Conference on p. 584, 1977
- [Mori Y. et al, 1999] Mori Y., Takahashi H., Oka R., Image-to-word Transformation Based on Dividing and Vector Quantizing Images with Words, In MISRM'99 1st International Workshop on Multimedia Intelligent Storage and Retrieval Management, 1999
- [Mortara M. et al, 2006] Mortara M., Patané G., Spagnuolo M. From geometric to semantic human body models. In: Computers & Graphics, vol. 30 (2) pp. 185-196. Elsevier, 2006.
- [Mukherjee J., 2002] Mukherjee J., MRF clustering for segmentation of color images, Elsevier Pattern Recogn. Lett., 23(8), 917-929, 2002
- [Nespeca R., 2015] Nespeca R., Report Stage: Analysis, Cartography and Information Extraction from Point Cloud to Ontology for Software Development, 2015
- [Nguyen H.T. et al, 2003] Nguyen H.T., Worring M., Boomgaard R., Watersnakes: energy-driven watershed segmentation, IEEE Trans. Pattern Anal. Mach. Intel., 25(3), 330-342, 2003
- [Niso, 2004] National Information Standard Organization, Understanding Metadata, [www.niso.org/publications/press/UnderstandingMetadata.pdf](http://www.niso.org/publications/press/UnderstandingMetadata.pdf) (Janvier 2016), 2004
- [Novatnack J. et al, 2008] Novatnack J., Nishino K., Sacle-Dependent/Invariant Local 3D Shape Descriptors for Fully Automatic Registration of Multiple Sets of Range Images, In Computer Vision, ECCV 2008, pp. 440-453, 2008
- [Oliver A. et al, 2012] Oliver A., Kang S., Wünsche B.C., MacDonald B., Using the Kinect as a Navigation Sensor for Mobile Robotics, IVCNZ'12 Proceedings of the 27<sup>th</sup> Conference on Image and Vision Computing New Zealand, pp 509-514, 2012
- [OMEGA] Déclaration d'invention présentée au nom du CNRS et faisant l'objet d'une demande de brevet déposée par le Cabinet Plasseraud le 16 juin 2009 sous le numéro FR 09 54 054 : OMEGA – ORIENTATION MODULE (inventeurs : Livio De Luca, Matthieu Deveau), 2009
- [Osada R. et al, 2002] Osada R., Funkhouser T., Chazelle B., Dokin D., Shape Distribution, ACM Transactions on Graphics (TOG), 21(4), pp. 807-832, 2002
- [Palma G. et al, 2010] Palma G., Corsini M., Dellepiane M., Scopigno R., Improving 2D-3D Registration by Mutual Information Using Gradient Maps, Puppo E., Brogni A., De Floriani L., Eds, Eurographics Italian Chapter Conference, pp.89-94, Genova, Italy, 2010
- [Pamart A. et al, 2015] Pamart A., Lo Bulgio D., De Luca L., Morphological Analysis of Shape Semantics from Curvature-Based Signature, Digital Heritage International Congress (DigitalHeritage), Sep 29-Oct 2, Granada, Spain, 2015
-

- [Pandya D. et al, 2014] Pandya D., Shah B., Comparative Study on Automatic Image Annotation, International Journal of Emerging Technology and Advanced Engineering, ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 3, March 2014
- [Pérouse de Montclos J.M., 2004] Pérouse de Montclos J.M., Architecture : méthode et vocabulaire, Editions du Patrimoine, Centre des Monuments Nationaux, 2011
- [Peteler F. et al, 2015] Peteler F., Gattet E., Bromblet P., Guillon O., Vallet J.M., De Luca L., Analyzing the Evolution of Deterioration Pattern : A First Step of an Image-Based Approach for Comparing Multitemporal Data Sets, Digital Heritage International Congress (DigitalHeritage), Sep 29-Oct 2, Granada, Spain, 2015
- [Petridis K et al., 2006] Petridis K, Anastasopoulos D, Saathoff C, Timmermann N, Kompatsiaris I, Staab S. M-OntoMat-Annotizer: Image annotation linking ontologies and multimedia low-level features. In: KES 2006 10th Intl. conf. on knowledge based, intelligent information and engineering systems, 2006
- [PhotoScan] <http://www.agisoft.com/> (Janvier 2016)
- [PHOTOSYNTH, 2008] <https://photosynth.net/default.aspx> (Janvier 2016)
- [Pierrot-Desseiligny M. et al., 2006] Pierrot-Desseiligny M., Paparoditis N., A Multi-Resolution and Optimization-Based Image Matching Approach: An Application to Surface Reconstruction from SPOT-HRS Stereo Imagery, In: IAPRS, Vol. XXXVI-1/W41 in ISPRS Workshop on Topographic Mapping from Space (With Special Emphasis on Small Satellites, Ankara, Turkey, 2006
- [Pierrot-Desseiligny M. et al., 2011a] Pierrot-Desseiligny M., De Luca L., Remondino F., Automated image-based procedures for accurate artifacts 3D modeling and orthopictures generation, Geoinformatics FCE CTU Journal, Vol.6, pp. 291-299, Prague, Czech Republic, 2011
- [Pierrot-Desseiligny M. et al., 2011b] Pierrot-Desseiligny M., Clery I., APERO, an open source bundle adjustment software for automatic calibration and orientation of set of images, 2011
- [Pottmann H. et al, 2009] Pottmann H., Wallner J., Huang Q.X., Yang Y.L., Integral Invariants for Robust Geometry Processing, Computer Aided Geometric Design, 26(1), pp. 37-60, 2009
- [Precioso F et al, 2005] Precioso F. et al, Robust real-time segmentation of images and videos using a smooth-spline snake-based algorithm, IEEE Trans. Image Proc., 14(7), 910-924, 2005
- [Ramanath R. et al, 2005] Ramanath R., Snyder W.E., Yoo Y., Drew M.S., Color Image Processing Pipeline, IEEE Singal Processing Magazine, 22(1), pp34-43, 2005
- [Remondino F. et al, 2006] Remondino F., El-Hakim S., Image-based 3D Modelling: a Review, The Photogrammetric Record, 21(115), pp269-291, 2006



- 
- [Remondino F. et al., 2012] Remondino, F., Del Pizzo, S., Kersten, T.P., Troisi, S., Low-cost and open-source solutions for automated image orientation – A critical overview. Proc. EuroMed 2012 Conference, M. Ioannides et al. (Eds.), LNCS 7616, Springer, Heidelberg, pp. 40-54, 2012
- [Rodden K., 1999] Rodden, K., How do people organise their photographs? In BCS IRSG 21st Ann. Colloq. on Info. Retrieval Research, 1999
- [Rosten E. et al, 2006] Rosten E., Drummond T., Machine Learning for High-Speed Corner Detection, In Computer Vision, ECCV 2006, pp. 430-443, Springer, 2006
- [Rublee E. et al, 2011] Rublee E., Rabaud V., Konolige K., Bradski G., ORB: An Efficient Alternative to SIFT or SURF, In Computer Vision, ICCV 2011, 2011 International Conference on pp.2564-2571
- [Rusu R.B., 2009] Rusu R.B., Semantic 3D Object Map for Everyday Manipulation in Human Living Environments, Technische University at Munchen, PhD Thesis, 2009
- [Saleri R. et al, 2013] Saleri R., Erica N., Remondino F., Menna F., Accuracy and Block Deformation Analysis in Automatic UAV and Terrestrial Photogrammetry-Lesson Learnt, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Septembre 2013, Strasbourg, France, II-5/W1, pp.203-208, 2013
- [Sarkar S. et al, 2000] Sarkar S., Sounddararajan P., Supervised learning of large perceptual organization: graph spectral partitioning and learning automata, IEEE Trans. Pattern Anal. Mach. Intel., 22(5), pp.504-525, 2000
- [Saupe D. et al, 2001] Saupe D., Vranic D.V., 3D Model Retrieval with Spherical Harmonics and Moments, Springer, 2001
- [Schmid C. et al, 2000] Schmid C., Mohr R., Bauckhage C., Evaluation of Interest Point Detectors, International Journal of Computer Vision, 37(2), pp. 151-172, 2000
- [Sezgin M. et al, 2004] Sezgin M. Sankur B. Survey over image thresholding techniques and quantitative performance evaluation, J ; Elec. Imag., 13(1), 146-165, 2004
- [Shakarji C., 1998] Shakarji C. Least-Squares Fitting Algorithms of the NIST Algorithm Testing System. Journal of Research of the National Institute of Standards and Technology, 103(6):633–641, November-December 1998.
- [Shi J. et al, 1994] Shi J., Tomasi C., Good Features to Track, In Computer Vision and Pattern Recognition, Proceedings CVPR'94, IEEE Computer Society Conference on pp. 593-600, 1994
- [Sicilia M.A., 2006] Sicilia M.A., Metadata, semantics and ontology: providing meaning to information resources, International Journal of Metadata, Semantics and Ontologies, Vol.1, No1, 2006
-

- [Silberman N. et al, 2012] Silberman N., Hoiem D., Kohli P., Fergus R., Indoor Segmentation and Support Inference from RGBD Images, Proc. ECCV'12 Proceedings of the 12th European Conference on Computer Vision – Volume Part V, pp 746-760, 2012
- [Slimani T., 2013] Slimani T., Semantic Annotation: The Mainstream of Semantic Web, International Journal of Computer Applications Technology and Research, Vol.2, Issue 6, pp 763-770, 2013
- [Smith S.M. et al, 1997] Smith S.M., Brady J.M., SUSAN: a New Approach to Low Level Image Processing, International Journal of Computer Vision, 23(1), pp. 45-78, 1997
- [Snavely N. et al., 2006] Snavely N., Seitz S., Szeliski R., Photo Tourism: Exploring Photo Collections in 3D, 2006
- [Socher R. et al, 2011] Socher R., Chiung-Yu Lin C., Ng A.Y., Manning C.D., Parsing Natural Scenes and Natural Language with Recursive Neural Network, Proceedings of the 28<sup>th</sup> International Conference on Machine Learning, ICML 2011, Bellevue, Washington, USA, June 28 – July 2, 2011, p129-136.
- [Song F., et al, 2010] Song F., Guo Z., Mei D., Feature Selection Using Principal Component Analysis, System Science, Engineering Design and Manufacturing Informatization (ICSEM), Vol.1, pp. 27-30, 2010
- [Stefani C. et al, 2010] Stefani C., De Luca L., Véron P., Florenzano M., Time indeterminacy and spatio-temporal building transformations: an approach for architectural heritage understanding, International Journal on Interactive Design and Manufacturing (IJIDeM), February 2010, Volume 4, Issue 1n, PP. 61-74, 2010
- [Stefani C. et al, 2012] Stefani C., Brunetaud X., Janvier-Badosa S., Beck K., De Luca L., Al-Mukhtar M., 3D Informations System for the Digital Documentation and the Monitoring of Stone Alteration, Progress in Cultural Heritage Preservation, 4th International Conference, EuroMed 2012, Limassol, Cyprus, October 29 – November 3, 2012. Proceedings, pp 330-339, 2012
- [Studer R. et al, 1998] Studer R., Benjamins R., Fensel D., Knowledge Engineering: Principles and Methods, Data & Knowledge Engineering, 25(1-2), pp 161-198, 1998
- [Sturm P. et al, 2011] Sturm P., Ramalingam S., Tardif J.P., Gasparini S., Barreto J., Camera Models and Fundamental Concepts Used in Geometric Computer Vision. Foundations and Trends in Computer Graphics and Vision, 6(1-2), pp.1-183, 2011
- [Sun J. et al, 2009] Sun J., Ovsjanikov M., Guibas L., A Concise and Provably Informative Multi-Scale Signature Based on Heat Diffusion, In Computer Graphics Forum, Vol. 28, pp. 1383-1392, 2009
- [Taillet R. et al, 2009] Taillet R., Febvre P., Villain L., Dictionnaire de Physique, De Boeck, Coll. De Boeck Supérieur, novembre 2009

- 
- [Tao W. et al, 2007] Tao W., Jin H., Zhang Y., Color image segmentation based on mean-shift and normalized cuts, *IEEE Trnas., Syst. Man Cyber.*, 37(5), 1382-1389, 2007
- [Teague M.R., 1980] Teague M.R., Image Analysis vie the General Theory of Moments, *JOSA*, 70(8), pp. 920-930
- [Tola E. et al, 2010] Tola E. Lepetit V., Fua P., DAISY: An Efficient Dense Descriptor Applied to Wide-Baseline Stereo, *Pattern Analysis and Machine Intelligence*, *IEEE Transactions on* 32(5) , pp. 815-830
- [Toschi I. et al. 2014] Toschi, I., Capra, A., De Luca, L., Beraldin, J. A., & Cournoyer, L., On the evaluation of photogrammetric methods for dense 3D surface reconstruction in a metrological context. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2(5), 371–378, 2014
- [Trzcinski T. et al, 2012] Trzcinski T., Lepetit V., Efficient Discriminative Projections for Compact Binary Descriptors, In *Computer Vision and Pattern Recognition, CVPR 2012*, 2012 IEEE Conference on pp. 228-242, 2012
- [Trzcinski T. et al, 2013] Trzcinski T., Christoudias M., Fua P., Lepetit V., Boosting Binary Keypoints Descriptors, In *Computer Vision and Pattern Recognition, CVPR 2013*, 2013 IEEE Conference on pp. 2874-2881, 2013
- [Tsai R.Y., 1986] Tsai R.Y., A versatile camera calibration technique for high accuracy 3D Machine vision metrology using off-the-shelf TV cameras and lenses ,*IEEE Journal of Robotics and Automation*, Vol. RA3, No. 4, August 1987
- [Ugarizza L.G. et al, 2009] Ugarizza L.G., Saber E., Vantaram S.R., Amuso V., Shaw M., Bhaskar R., Automatic image segmentation by dynamic growth and multiresolution merging, *IEEE Trans. Image Proc.*, 18(10), 2275-2288, 2009
- [Unnikrishnan R. et al, 2007] Unnikrishnan R., Pantofaru C., Hebert M., Towards objective evaluation of image segmentation algorithms, *IEEE Trans. Pattern Anal. Mach. Intel.*, 29(6), 929-944, 2007
- [Uren V.S. et al, 2006] Uren V.S., Cimiano P. Iria J., Handschuh S., Vargas-Vera M., Mota E., Ciravegna F. Semantic Annotation for Knowledge Management: Requirements and a Survey of the State of the Art, *J. Web Sem.*, 4(1), pp 14-28, 2006
- [Van De Weijer J. et al, 2006] Van Der Weijer J., Schmid C., Coloring Local Feature Extraction, In *Computer Vision, ECCV 2006*, pp. 334-348, Springer, 2006
- [Van Gool L. et al, 1996] Van Gool L., Moons T., Ungureanu D., Affine/Photometric Invariants for Planar Intensity Patterns, In *Computer Vision, ECCV'96*, pp. 642-651, Springer, 1996
-

- [Vandenbroucke N., 2000] Vandenbroucke N., Segmentation d'images couleurs par classification de pixels dans des espaces d'attributs colorimétriques adaptés. Application à l'analyse d'images de football, 2000
- [Vantaram S.R., et al, 2010] Vantaram S.R.,Saber E., Dianat S.A., Shaw M., Bhaskar R., Multiresolution adaptative and progressive gradient-based color image SEGmentation, J. Elect. Image, 19(1), 013001, 2010
- [Vantaram S.R. et al, 2012] Vantaram S. R., Saber E., Survey of contemporary trends in color image segmentation. Journal of Electronic Imaging, 21(4):1-28, 2012.
- [Vedaldi A., 2006] Vedaldi A., An open implementation of the SIFT detector and descriptors, UCLA CSD Tech Report 070012, 2006
- [Verdie Y. et al, 2014] Verdie Y., Yi K., Fua P., Lepetit V., TILDE: A Temporally Invariant Learned Detector, arXiv preprint arXiv :1411.4568, 2014
- [Viola P. et al, 2001] Viola P., Jones M., Rapid Object Detection Using a Boosted Cascade of Simple Feature, In Computer Vision and Pattern Recognition, CVPR 2001, 2001 IEEE Conference on Vol.1, pp. 511-518, 2001
- [Vosselmant G. et al, 2010] Vosselmant G., Maas H.G., Airbone and Terrestrial Laser Scanning, CRC, Boca Raton, 318 pp., ISBN: 978-1904445-87-6, 2010
- [Waldhäusl P. et al., 1994] Waldhäusl P., Ogleby C. 3 x 3 Rules for simple photogrammetric documentation of architecture. In: J.G.Fryer (Editor), International Archives of Photogrammetry and Remote Sensing, Vol.XXX, Part5, pp 426 – 429, 1994
- [Wallin A. et al, 1995] Wallin A., Kübler O., Complete Sets of Complex Zernike Moment Invariants and the Role of the PSeudoinvariants, IEEE Transactions on Pattern Analysis and Machine Intelligence, 17(11), pp. 1106-1110, 1995
- [Wan S., 2011] Wan S., Image Annotation Using the Simple Decision Tree, Management and e-Commerce and e-Government (ICMeCG), 2011 5<sup>th</sup> International Conference on pp. 141-146
- [Wang B. et al, 2007] Wang B., Li Z., Yu N., Li M., Image Annotation in a Progressive Way, IEEE International Conference on Multimedia and Expo, pp. 811-814, 2007
- [Wang J. et al, 2009] Wang J., Ju L., Wang X., An edge-weighted centroidal voronoi tessellation model for image segmentation, IEEE Trans. Image Proc. 18(8), 1844-1858, 2009
- [Wang Y. et al, 2006] Wang Y., Yang J., Peng N., Unsupervised color-texture segmentation based on soft criterion with adaptative mean-shift clustering, Elsevir Pattern Recogn. Lett., 27(5), 386-392, 2006
- [Willems G. et al, 2008] Willems G., Tuytelaars T., Van Gool L., An Efficient Dens and Scale-Invariant Spatio-Temporal Interest Point Detector, In Computer Vision, ECCV 2008, pp. 650-663, Springer, 2008

- [Wu Z. et al, 1993] Wu Z., Leahy R., An Optimal Graph Theoretic Approach to Data Clustering: Theory and its Application to Image Segmentation, IEEE Trans. Pattern Anal. Mach. Intel., 15(11), 1101-1113, 1993
- [Zhang L. et al, 2010] Zhang L., Ji Q., Image Segmentation with a Unified Graphical Model, IEEE Trans. Pattern Anal. Mach. Intel., 32(8), 1406-1425, 2010
- [Zhang D. et al, 2012] Zhang D., Islam MD.M., Lu G., A Review on Automatic Image Annotation Techniques, Pattern Recognition, Vol. 45, No. 1, pp. 346-362, 2012
- [Zhang Z., 2012] Zhang Z., Microsoft Kinect and Its Effect, Multimedia, IEEE Vol.19, No.2, pp. 4-10, Février 2012
- [Zhao Y. et al, 2008] Zhao Y., Zhao Y., Zhu Z., Pan J-S., A Novel Image Annotation Scheme Based on Neural Network, Intelligent Systems Design and Applications, ISDA'08, 8th International Conference on Vol. 3, pp. 644-647, 2008
- [Zogg H.M., 2008] Zogg H.M., Investigations of High Precision Terrestrial Laser Scanning with Emphasis on the Development of a Robust Close-Range 3D-Laser Scanning System, Institute of Geodesy and Photogrammetry, ETH Zurich, 2008





## Liste des publications

### **Publications dans des revues avec comité de lecture**

A. Manuel, L. De Luca, P. Véron, A hybrid approach for the semantic annotation of spatially oriented images, *IJHDE (International Journal of Heritage in the Digital Era)*, Volume 3 Number 2 p. 305-320, 2014

### **Publications dans des actes de congrès avec comité de lecture**

T. Messaoudi, A. Manuel, E. Gattet, L. De Luca, P. Véron, Laying the foundation for an Information System Dedicated to Heritage Building Degradation Monitoring Based on the 2D/3D Semantic Annotation of Photographs, *Eurographics Workshop on Graphics and Cultural Heritage - Short Papers / Posters*, Darmstadt, Germany, 6-8 October 2014

A. Manuel, E. Gattet, L. De Luca, P. Véron, An approach for precise 2D/3D semantic annotation of spatially-oriented images for in situ visualization applications, 2013 Digital Heritage International Congress (DigitalHeritage), Vol. 1, pp. 289-296, IEEE, 2013.

A. Manuel, C. Stefani, L. De Luca, P. Véron, 2D/3D semantic annotations towards a set of spatially oriented photographs, *Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXIV International CIPA Symposium, Strasbourg, France, 2-6 September 2013

A. Manuel, C. Stefani, L. De Luca, P. Véron, Distribution of semantic annotations towards a set of spatially oriented photographs, *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XL-5/W1, 2013; 3D-ARCH 2013 - 3D Virtual Reconstruction and Visualization of Complex Architectures, Trento, Italy, 25 – 26 February 2013



## Table des illustrations

Figure 1 : Schéma de fonctionnement de la plateforme C3DC .....	17
Figure 2 : Contributions scientifiques du projet Monumentum .....	18
Figure 3 : Disposition de Bayer des filtres de couleur .....	22
Figure 4 : Spectre électromagnétique et classes de rayonnement .....	22
Figure 5 : Les différentes tailles de capteurs photographiques .....	23
Figure 6 : Influence de la focale sur l'angle de champ pour un même capteur .....	23
Figure 7 : Principe de fonctionnement des scanners longue portée [Marbs A. et al, 2001] .....	24
Figure 8 : Exemple de scanner à décalage de phase : Faro Focus 3D .....	25
Figure 9: Principe de fonctionnement des scanners par triangulation [Marbs A. et al., 2001] .....	25
Figure 10 : Motif projeté sur l'objet [Gupta M. et al, 2011] .....	26
Figure 11 : Principe de la technologie Kinect .....	26
Figure 12 : Dispositif de perspective de Filippo Brunelleschi .....	28
Figure 13 : Les différentes projections perspectives avec un, deux ou trois points de fuite .....	28
Figure 14 : Distorsions .....	29
Figure 15 : Principe de la projection orthogonale .....	29
Figure 16 : Exemple d'orthophotographie [Saleri R. et al, 2013] .....	30
Figure 17 : Assemblage de deux images [De Luca, 2009] .....	31
Figure 18 : Projection cylindrique d'une scène [De Luca, 2009] .....	31
Figure 19 : Projection sphérique d'une scène [De Luca, 2009] .....	32
Figure 20 : Projection cubique d'une scène [De Luca, 2009] .....	32
Figure 21 : Répartition des systèmes d'acquisition et des projections par rapport au type de données enregistrées et pouvant être générées .....	33
Figure 22 : Histogramme d'une image en niveaux de gris .....	35
Figure 23 : Gradients d'images .....	37
Figure 24 : Exemples de descripteurs locaux .....	38
Figure 25 : Exemple de cadastre .....	39
Figure 26 : Paramètres de calibration interne et externe .....	40
Figure 27 : Les différentes techniques de restitution photogrammétrique. De gauche à droite : mono, stéréo, multi-images .....	40
Figure 28 : Distribution des formes et comparaison [Osada R. et al, 2002] .....	42
Figure 29 : Composant d'une représentation par SI .....	43
Figure 30 : Exemple de métadonnées associées à une même image numérique .....	45
Figure 31 : Modèles d'annotations [Andrews P. et al, 2012] .....	49
Figure 32 : Exemple d'annotation avec [FLICKR] à gauche et avec [Marqueed, 2014] à droite .....	51
Figure 33 : Interface de Photomesa [Bederson B.B., 2001] .....	51
Figure 34 : Interface de VideoAnnEx [Lin C-Y et al. 2003] .....	52
Figure 35 : Interface de M-OntoMat-Annotizer [Petridis K et al, 2006] .....	53
Figure 36 : Annotation sémantique d'images d'architecture patrimoniale [Azzi Y., 2008] .....	53
Figure 37 : Concept de l'annotation automatique d'images [Pandya D. et al, 2014] .....	55
Figure 38 : Exemples de résultats de détection par réseaux neuronaux pour [Socher R. et al, 2011] .....	57
Figure 39 : Exemples de résultats obtenus par le système du Google Lab Research [GoogleLabResearch, 2014] classés en fonction de l'évaluation de la description .....	58
Figure 40 : Comparaison des approches d'annotation automatique d'images [Pandya D. et al, 2014] .....	60
Figure 41 : Exemples d'objets patrimoniaux complexes [3DIcons] .....	60
Figure 42 : Résultats de détection d'ALIPR [Li J. et al, 2006] .....	61
Figure 43 : Exemples de reconstruction surfaciques pour différentes méthodes [Kazhdan M. et al., 2013] .....	63
Figure 44 : Schéma global du modèle 3DSEAM [Bilasco I.M. et al, 2005] .....	63
Figure 45 : Une mesh de base (a) est segmentée en utilisant plusieurs algorithmes (b), (c) et (d) puis seules les parties appropriées des chaque segmentation (b), (c) ou (d) sont sélectionnées et annotées (e) [Attene M. et al, 2009] .....	64

Figure 46 : De gauche à droite : (1) Modèle de base, (2) segmentation de la géométrie en sous éléments, (3) définition de l'organisation hiérarchique et (4) visualisation sur un outil open-source 3D [Manferdini A.M. et al, 2010] .....	64
Figure 47 : Aperçu de l'interface du projet Arrigo [Havemann S. et al, 2008].....	64
Figure 48 : Exemple d'objets sur [3DSA, 2014] .....	65
Figure 49 : Etapes de générations : (a) collections d'images, (b) génération du nuage de points, (c) interface de Photosynth [Snavely N. et al, 2006] .....	66
Figure 50 : Annotations sur Phototourism [Snavely N. et al, 2006] .....	66
Figure 51 : Maquette numérique et référencement spatial des images [Busayarat C., 2010].....	67
Figure 52 : Projection d'annotations du modèle 3D sur les images 2D [Busayarat C., 2010] .....	67
Figure 53 : Représentation spatio-temporelle des transformations d'un bâtiment [De Luca L. et al, 2010].....	68
Figure 54 : Différence entre les approches actuelles de sémantisation .....	72
Figure 55 : Objet patrimonial et acquisitions multiples .....	73
Figure 56 : Les trois premières étapes de l'approche : un nuage point est réalisé à partir d'un ensemble de photographies ; les photographies sont spatialement référencées autour du nuage ; la relation projective entre 2D et 3D permet de transférer les annotations entre les images et le nuage de points .....	74
Figure 57 : Différence entre données réelles et représentation géométrique des données .....	79
Figure 58 : Le système OMEGA [OMEGA] pour le référencement automatique d'images .....	81
Figure 59 : L'interface du système de [Kadobayashi R. et al, 2003] .....	82
Figure 60 : Interface pour le référencement manuel de [Busayarat C., 2010].....	82
Figure 61 : A gauche, images de différents points de vue, à droite, le résultat de l'estimation des positions [Johansson B., 2002] .....	83
Figure 62 : Principe de l'algorithme défini par [Palma G. et al, 2010].....	84
Figure 63 : Recherche des points SIFT entre deux images .....	86
Figure 64 : Orientation des images autour des points SIFT .....	87
Figure 65 : Approche pyramidale : Evolution des cartes de profondeur durant l'appariement dense .....	87
Figure 66 : Nuage de points final .....	88
Figure 67 : Exemple de lignes de commandes pour la génération d'un nuage de points à partir d'images au format JPG .....	89
Figure 68 : Extrait du fichier XML d'orientation d'une image donné par TAPAS.....	90
Figure 69 : Interprétation géométrique de la matrice d'orientation .....	90
Figure 70 : Extrait du fichier XML de calibration d'une image donné par TAPAS.....	91
Figure 71 : Exemples de distorsion .....	91
Figure 72 : Les différentes transformations du modèle sténopé.....	92
Figure 73 : Elimination de points multiples sur un pixel : seul le plus proche du centre optique (CO) est conservé, les autres points sont considérés comme non visibles sur cette image. ....	94
Figure 74 : Projection du nuage de points sur différentes dimensions d'une image, en blanc les pixels où une correspondance 3D a été trouvée. Plus l'image est redimensionnée pour la projection, plus la densité de correspondances 3D est élevée.....	95
Figure 75 : Sur ce nuage de points, les trous dans les parties des contreforts peuvent provoquer des erreurs de correspondances .....	95
Figure 76 : Correction des points par l'étude de la carte de profondeur .....	96
Figure 77 : Masques de corrélation et image associée .....	97
Figure 78 : Orthophotographie.....	98
Figure 79 : Fichier de repère .....	98
Figure 80 : Fichier MTD.....	98
Figure 81 : Position de l'orthophotographie dans l'espace par rapport au nuage de points .....	99
Figure 82 : Image (luminance) extraite d'un nuage de point relevé par scanner laser 3D Faro (projection sphérique).....	99
Figure 83 : Projections inverses .....	100
Figure 84 : Décomposition d'une image TIFF en trois couches : Rouge, Vert, Bleu (niveaux de gris colorés) .	101
Figure 85 : Décomposition d'un fichier XYZ TIFF en trois couches : X, Y, Z (valeurs des composantes des coordonnées 3D) .....	102

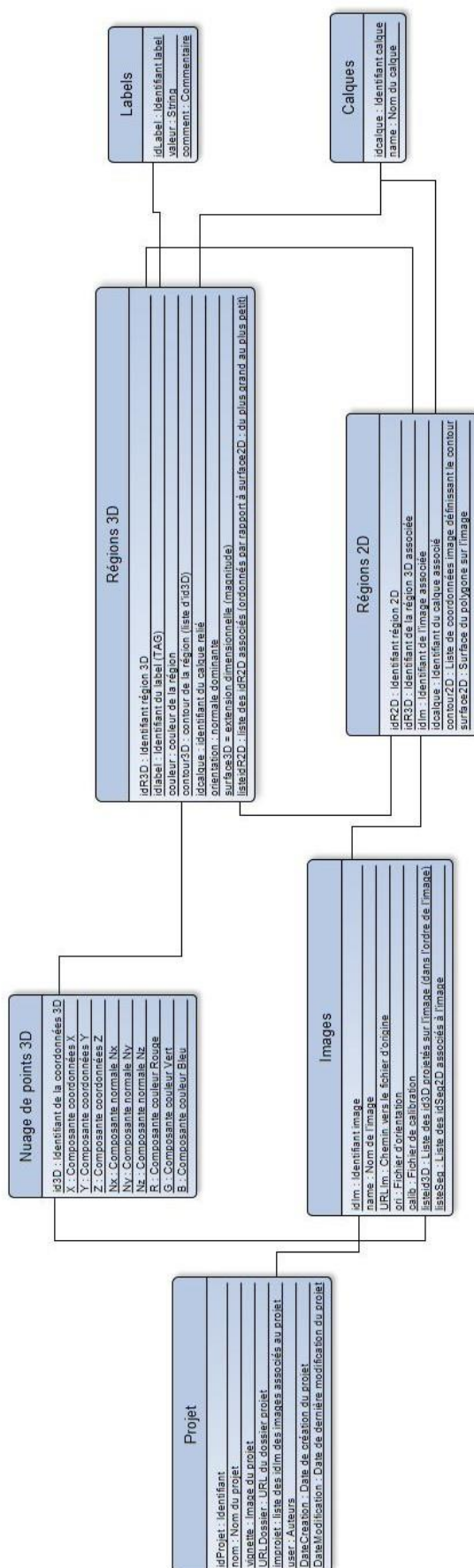


Figure 86 : Tableaux stockant les index des points 3D associés aux données 3D .....	103
Figure 87 : Différents types d'images pouvant être acquis pour l'analyse (campagne de relevé du CICRP) .....	104
Figure 88 : Orientation d'une nouvelle image parmi les images déjà orientées .....	105
Figure 89 : Différences au niveau de la diachronie pour plusieurs photographies d'un même objet .....	105
Figure 90 : Orientation d'une nouvelle image par l'intermédiaire d'une autre image : la caméra bleue est orientée par rapport aux caméras vertes, et la caméra rouge est orientée par rapport à la caméra bleue .....	106
Figure 91 : Images acquises par une caméra thermique .....	106
Figure 92 : Images d'un même objet prises à deux niveaux de détails .....	107
Figure 93 : Images intermédiaires pour l'orientation d'images de différents niveaux de détails .....	107
Figure 94 : Orientation des images constituant la base .....	108
Figure 95 : Orientation des images intermédiaires .....	108
Figure 96 : Orientation des images de détails par rapport aux images intermédiaires .....	109
Figure 97 : Orientation des images de détails parmi toutes les images .....	109
Figure 98 : Génération d'une zone plus dense dans le nuage de points pour des images de détails .....	110
Figure 99 : Nuages de points pour différents temps d'un même objet .....	111
Figure 100 : Etapes du processus d'orientation multitemporelle .....	111
Figure 101 : Orientation d'images de différentes acquisitions autour du nuage de points d'un des temps .....	112
Figure 102 : Correspondances 3D multi-temporelles .....	112
Figure 103 : Etapes du transfert d'une annotation .....	114
Figure 104 : Sélection d'une colonne complète à l'aide d'images .....	115
Figure 105 : Définition d'une annotation à partir de plusieurs vues .....	116
Figure 106 : Définition de l'annotation et masque extrait .....	116
Figure 107 : Extraction des positions des pixels blancs .....	117
Figure 108 : Extraction des données 3D en fonction du type de stockage .....	118
Figure 109 : Représentation 3D de l'annotation extraite à partir de la sélection sur une image .....	118
Figure 110 : Echantillonnage des pixels dans le masque .....	119
Figure 111 : Etape de projection vers les autres images .....	119
Figure 112 : Résultat de projection : masques obtenus et superposition avec l'image associée .....	120
Figure 113 : Résultats du transfert avec ou sans échantillonnage, et avec ou sans correction .....	121
Figure 114 : Orientation d'images de différentes acquisitions autour des nuages de points à différents temps ..	122
Figure 115 : Propagation multitemporelle .....	122
Figure 116 : Comparaison d'une altération sélectionnée au temps T1 (bleu) et au temps T2 (rouge) sur le temps T0 (image) à l'aide de la propagation multi-temporelle (les images n'ont pas exactement la même orientation) .....	123
Figure 117 : Ajustement du résultat en utilisant deux temps comme intermédiaire .....	123
Figure 118 : Informations pouvant être liées aux points du nuage et aux pixels des images .....	125
Figure 119 : Taxonomie de haut niveau des méthodes de segmentation d'images selon [Vantaram S.R. et al, 2012] .....	126
Figure 120 : Taxonomie de haut niveau des méthodes de segmentation d'images selon [Vantaram S.R. et al, 2012] .....	127
Figure 121 : Nuage de points représentant une image couleur dans l'espace des couleurs RGB .....	128
Figure 122 : Principe de la croissance de région : (a) en gris les pixels respectant le critère, en rouge le pixel seed, (b) les pixels autour sont testés sur le critère, (c) les pixels autour respectant le critère sont ajoutés à la région, (d) et (e) le procédé est répété jusqu'à ce que tous les pixels autour ne respectent plus le critère, (f) région finale ..	129
Figure 123 : Résultats de gPb-OWT-UCM [Arbeláez P.A. et al, 2011] .....	130
Figure 124 : Comparaison entre la co-segmentation d'image RGB et la co-segmentation d'image RGBD [Fu H. et al, 2015] .....	130
Figure 125 : Résultats de l'évaluation par l'index NPR de la segmentation Mean-Shift [Unnikrishnan R. et al, 2007] .....	131
Figure 126 : Couches additionnelles d'informations .....	132
Figure 127 : Sélection des pixels répondant au critère de tolérance dans l'espace des couleurs autour du point de référence .....	133

Figure 128 : Résultats de la "baguette magique" sur les couleurs d'une image pour une tolérance de 22 avec ou sans contiguïté.....	133
Figure 129 : Sélection des normales répondant au critère de tolérance autour de la normale de référence .....	134
Figure 130 : Résultats de la "baguette magique" sur les normales d'une image pour une tolérance de 22 degrés avec ou sans contiguïté.....	134
Figure 131 : Représentation des espaces 3D de chacun des types d'images .....	135
Figure 132 : Résultats de segmentation d'une même image sur les couleurs et sur les normales.....	136
Figure 133 : Projection d'une segmentation 3D (obtenue avec ShapeAnnotator [Attene M. et al. 2007] sur les images .....	137
Figure 134 : Projection d'une segmentation 3D (obtenue avec Plumber [Mortara M. et al, 2006] sur les images .....	137
Figure 135 : A droite, sélection sur l'image, à gauche extension à l'aide de la segmentation recouverte .....	138
Figure 136 : A droite, sélection sur l'image, à gauche, extension à l'aide de la segmentation et avec critère de recouvrement.....	138
Figure 137 : Extension de sélection à l'aide d'une segmentation.....	139
Figure 138 : Intégration d'un diagramme de recouvrement dans l'interface de Nubes [Stefani C. et al, 2012] ..	140
Figure 139 : Superposition de trois annotations pour la recherche de corrélation [Messaoudi T. et al, 2015] : en rouge, élément architectural : contrefort, en jaune, matériau : pierre jaune de Caromb, en blanc, altération : alvéolisation .....	141
Figure 140 : Nuage de points des annotations.....	142
Figure 141 : Annotation et profil extrait .....	142
Figure 142 : Courbures extraites à partir du nuage de points d'une annotation .....	143
Figure 143 : Occlusion ambiante .....	143
Figure 144 : Rugosité.....	144
Figure 145 : Carte de profondeur à partir d'un plan par moindres carrés et reprojection sur une image.....	144
Figure 146 : Définition de surface et volume élémentaires pour l'estimation de surface et de volume d'un nuage de points [Nespeca R., 2015] .....	145
Figure 147 : Informations extraites d'une annotation .....	146
Figure 148 : Structuration des données au sein de la base de données (voir Annexe A) .....	148
Figure 149 : Structure du dossier projet .....	149
Figure 150 : Aperçu de l'interface développées sous Matlab (pour le détail voir Annexe B) .....	150
Figure 151 : Fonctionnement général de l'interface Matlab .....	150
Figure 152 : Aperçu de l'interface Qt (pour le détail voir Annexe B) .....	151
Figure 153 : Fonctionnement général de l'interface Qt.....	151
Figure 154 : Aperçu de l'interface web (pour le détail voir Annexe B) .....	152
Figure 155 : Fonctionnement général de l'interface web.....	153
Figure 156 : Différence de structuration entre Docker [Docker] et les machines virtuelles .....	153
Figure 157 : Schéma global de Nubes Data (pour le détail voir en Annexe C).....	154
Figure 158 : Graphes de transformations [Stefani C. et al, 2010] .....	157
Figure 159 : Principe de l'application mobile.....	159

## Annexe A : Structuration des Données

## Schéma global de la structuration des données :



## Détails de la structuration des données:

### Données Projet

Projet
idProjet : Identifiant
nom : Nom du projet
vignette : Image du projet
URLDossier : URL du dossier projet
improjet : liste des idIm des images associés au projet
user : Auteurs
DateCreation : Date de création du projet
DateModification : Date de dernière modification du projet

### Données Nuage de Points 3D

Nuage de points 3D
id3D : Identifiant de la coordonnées 3D
X : Composante coordonnées X
Y : Composante coordonnées Y
Z : Composante coordonnées Z
Nx : Composante normale Nx
Ny : Composante normale Ny
Nz : Composante normale Nz
R : Composante couleur Rouge
G : Composante couleur Vert
B : Composante couleur Bleu

### Données Images

Images
idIm : Identifiant image
name : Nom de l'image
URLIm : Chemin vers le fichier d'origine
ori : Fichier d'orientation
calib : Fichier de calibration
listeid3D : Liste des id3D projetés sur l'image (dans l'ordre de l'image)
listeSeq : Liste des idSeq2D associés à l'image

### Données Calques

Calques
idcalque : Identifiant calque
name : Nom du calque

### Données Régions 3D

Régions 3D
idR3D : Identifiant région 3D
idlabel : Identifiant du label (TAG)
couleur : couleur de la région
contour3D : contour de la région (liste d'id3D)
idcalque : identifiant du calque relié
orientation : normale dominante
surface3D = extension dimensionnelle (magnitude)
listeidR2D : liste des idR2D associés (ordonnés par rapport à surface2D : du plus grand au plus petit)

## Données Régions 2D



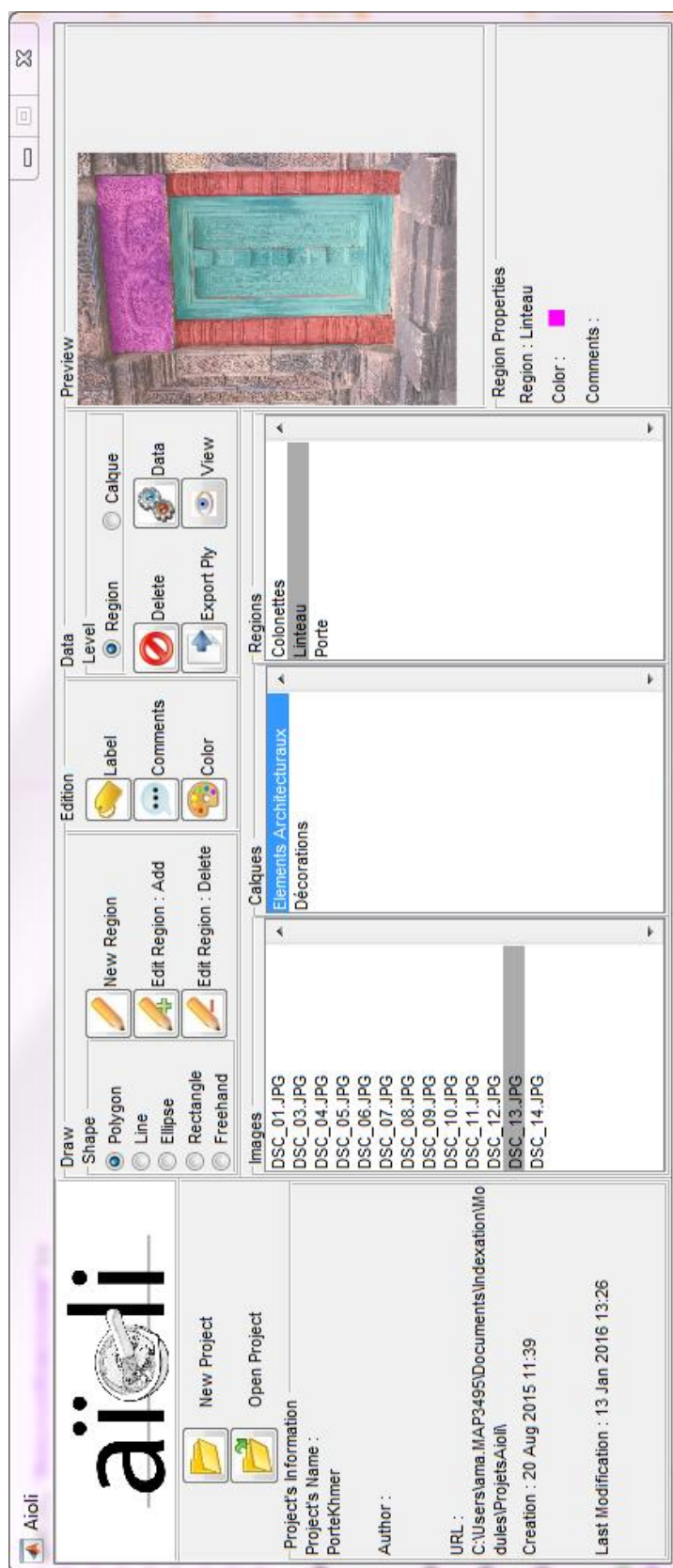
## Données Labels



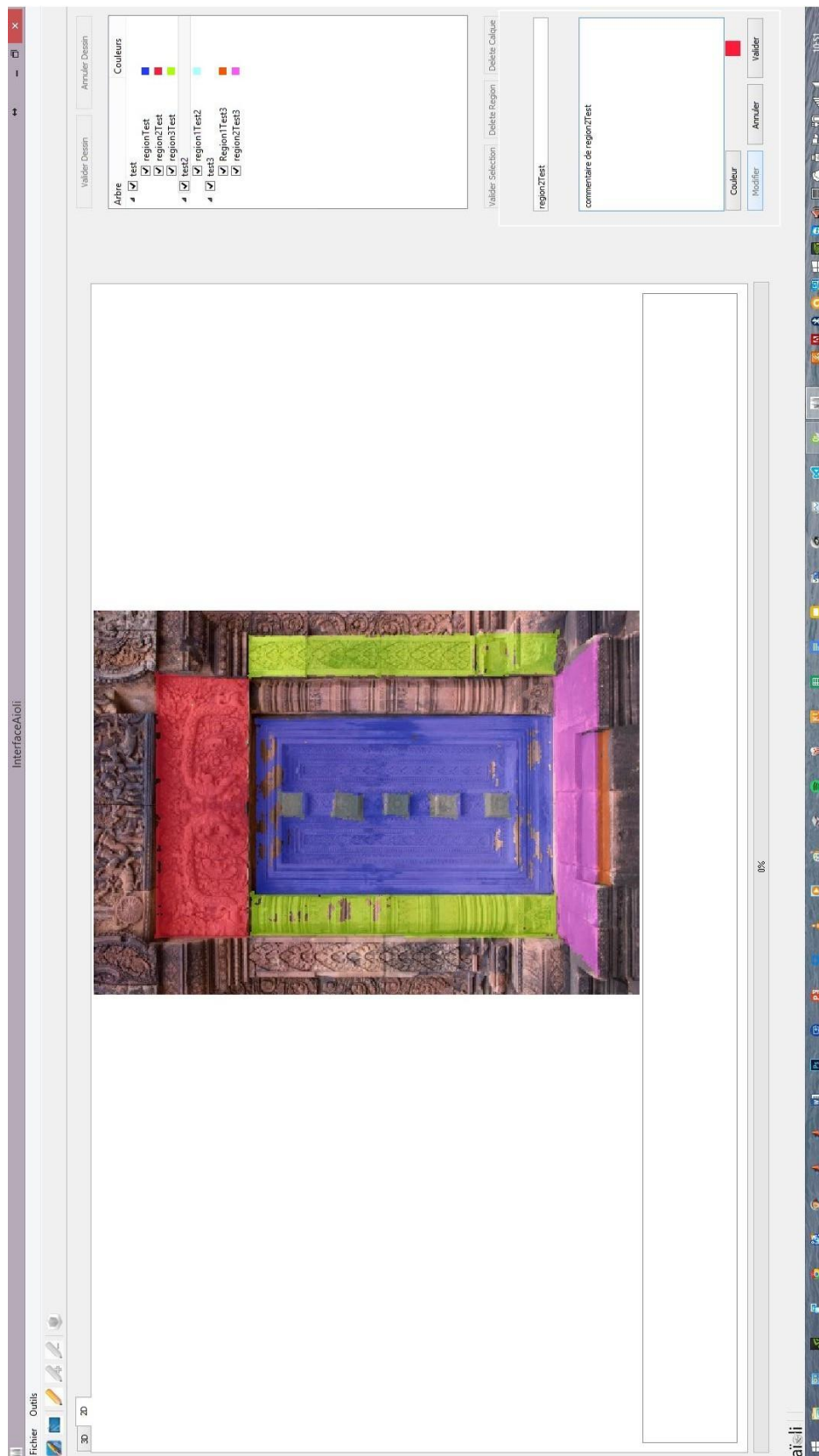


## Annexe B : Interfaces d'Annotations

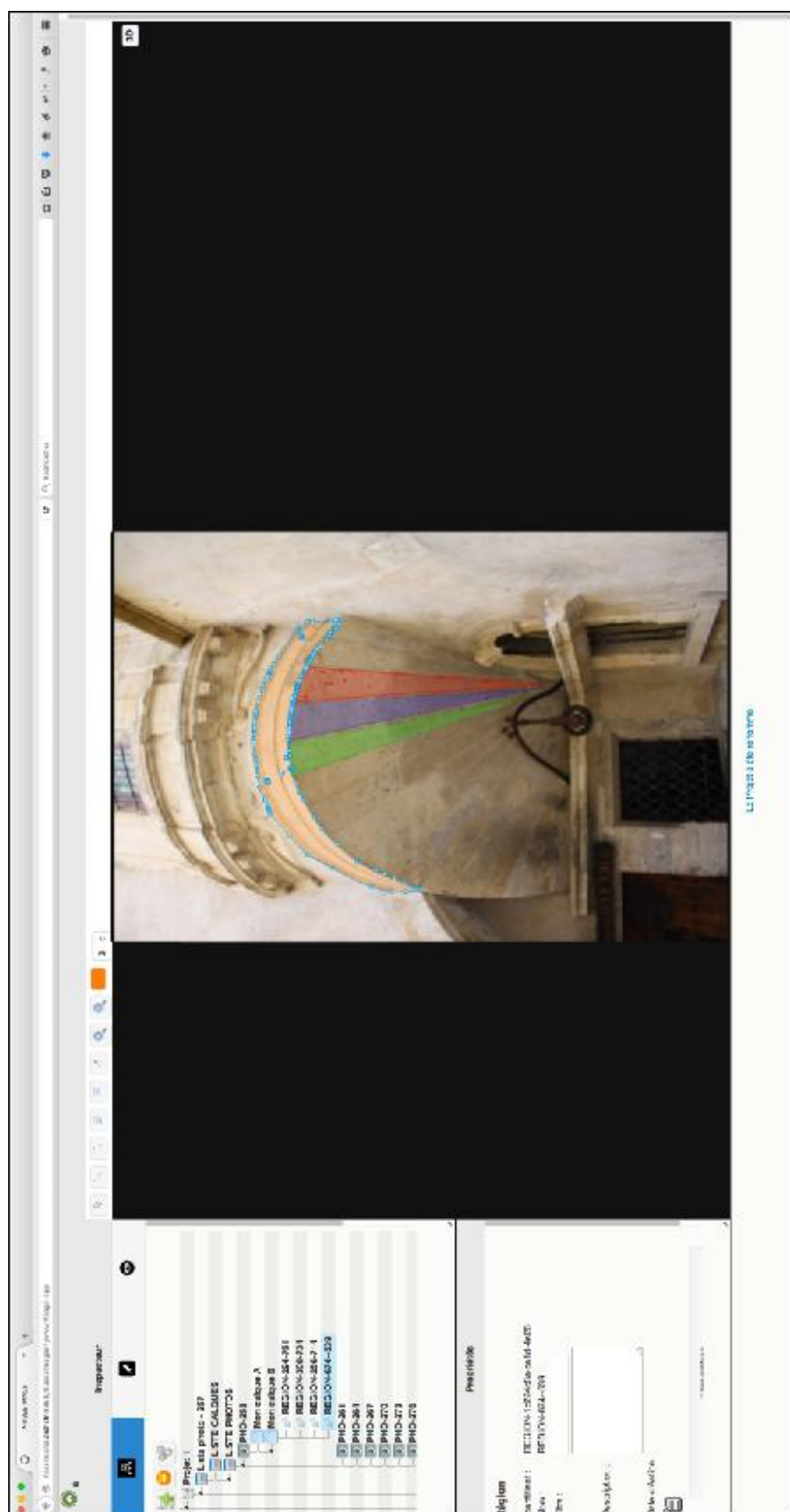
## Interface Matlab



## Interface QT

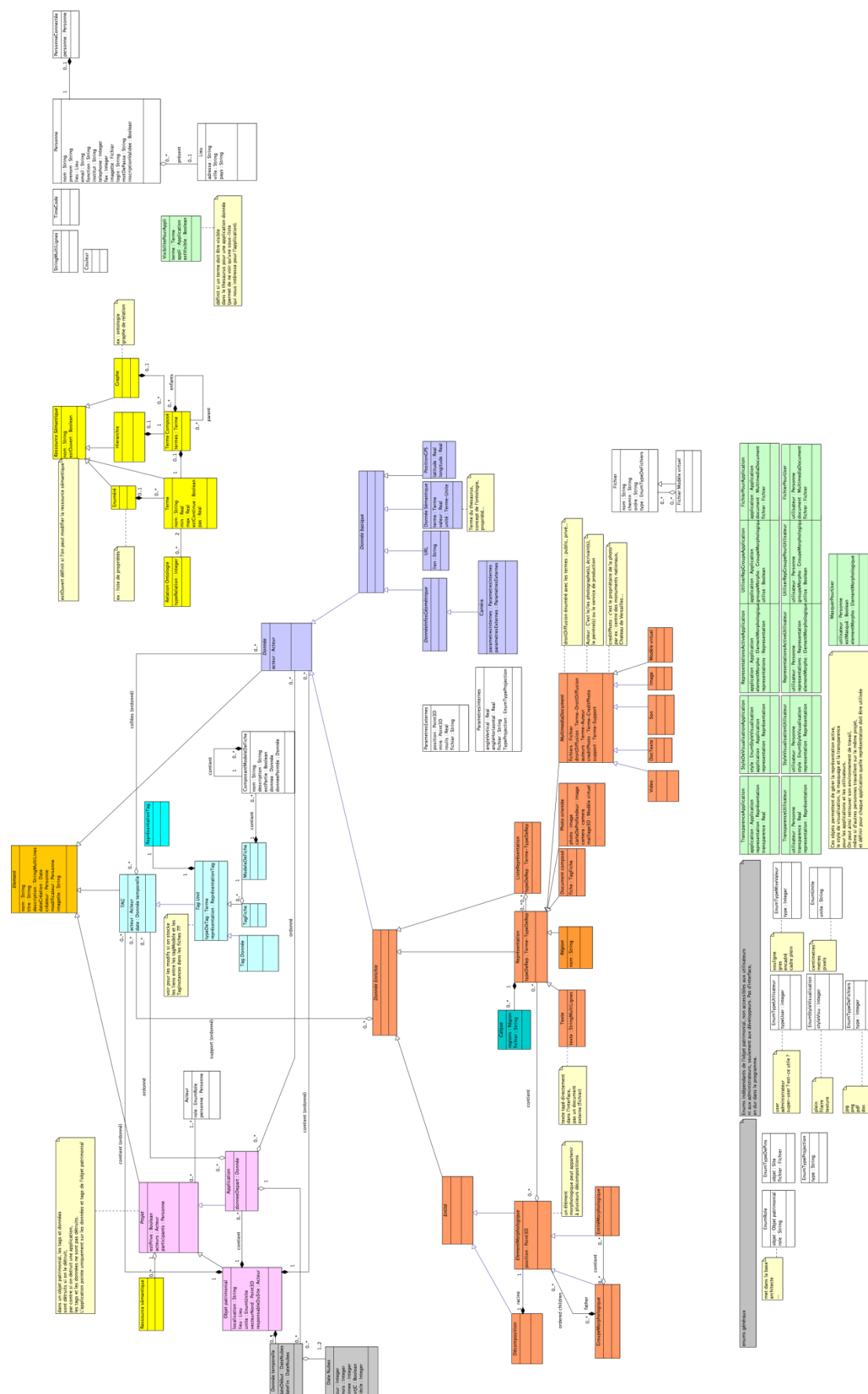


## Interface WebGL



## Annexe C : Modèle de Données Nubes Data

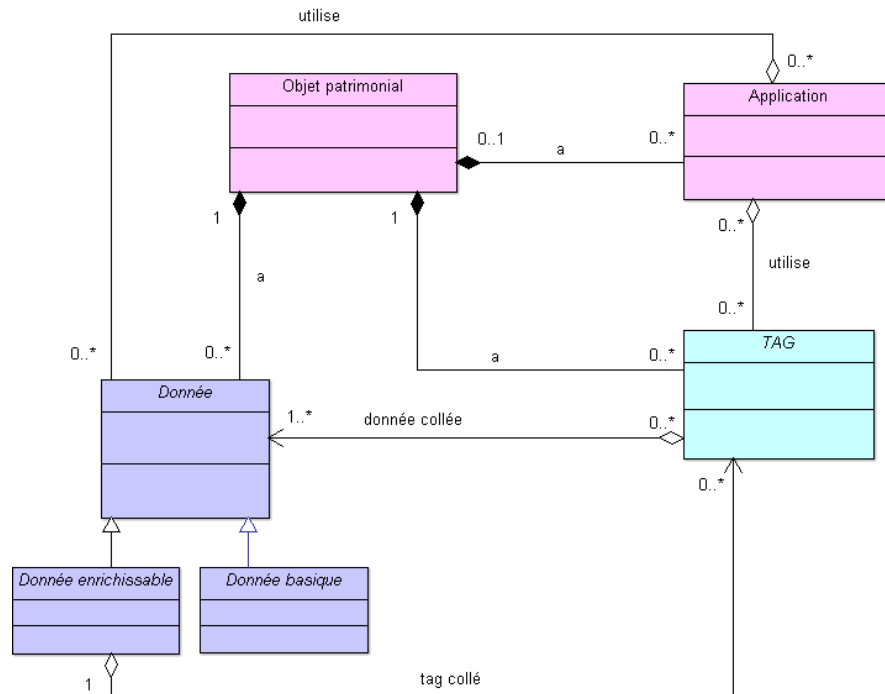
### Schéma global du modèle de données :





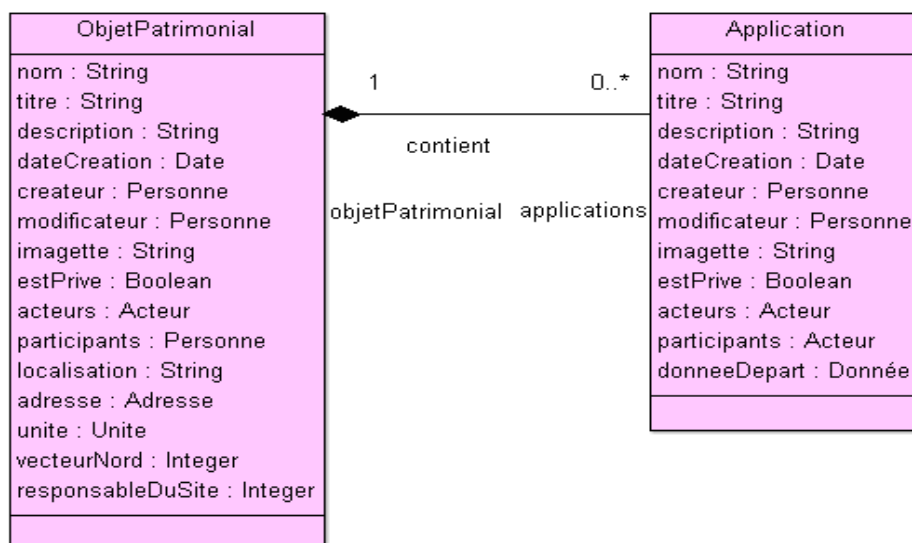
## Détails du modèle de données :

### Données Générales



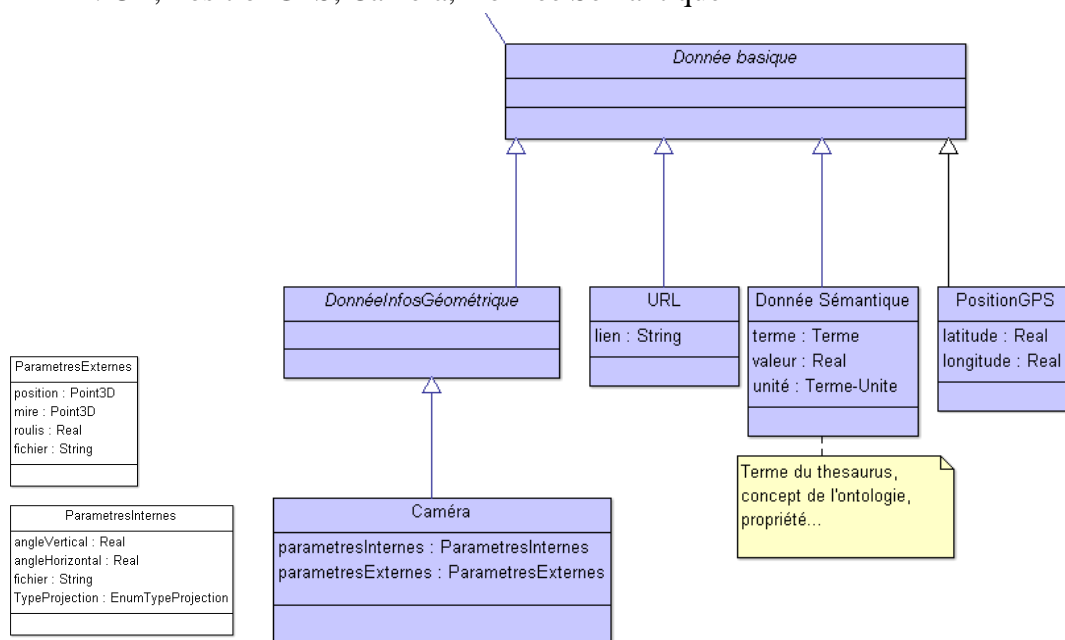
### Objet Patrimonial et Applications

- Objet patrimonial : monument, statue, site archéologique...
- Application : Site web, Pdf 3D, Jeu sérieux unity, etc...



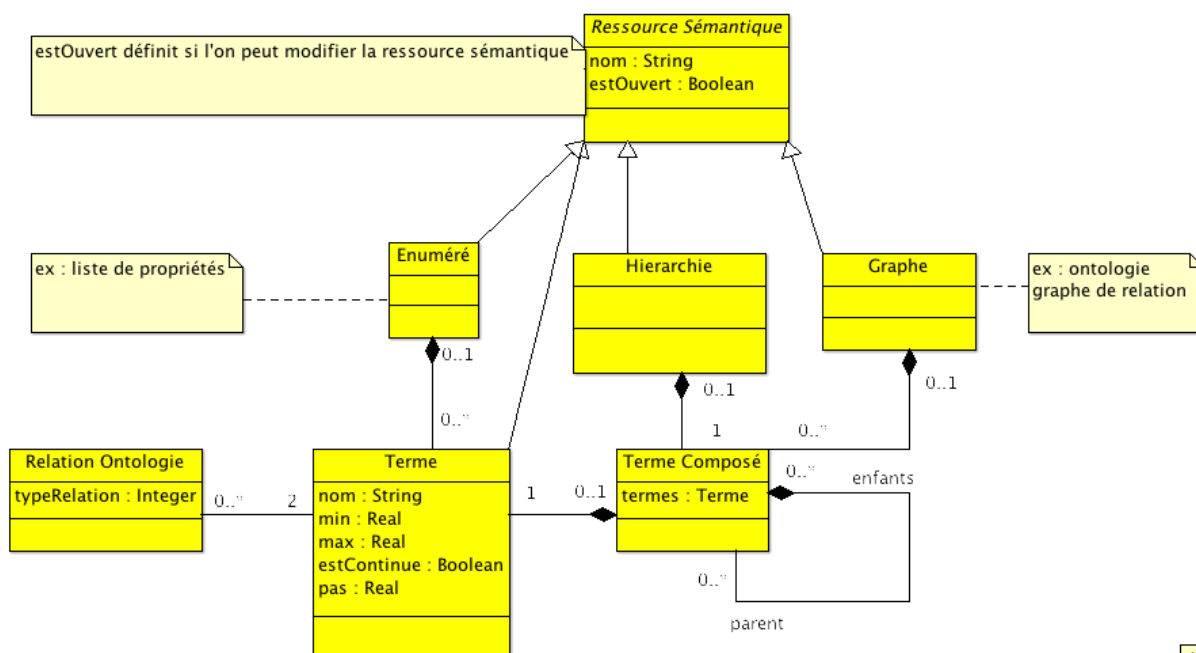
## Données Basiques

- Ces données ne peuvent pas être enrichies
- Ex : Url, PositionGPS, Caméra, Donnée Sémantique



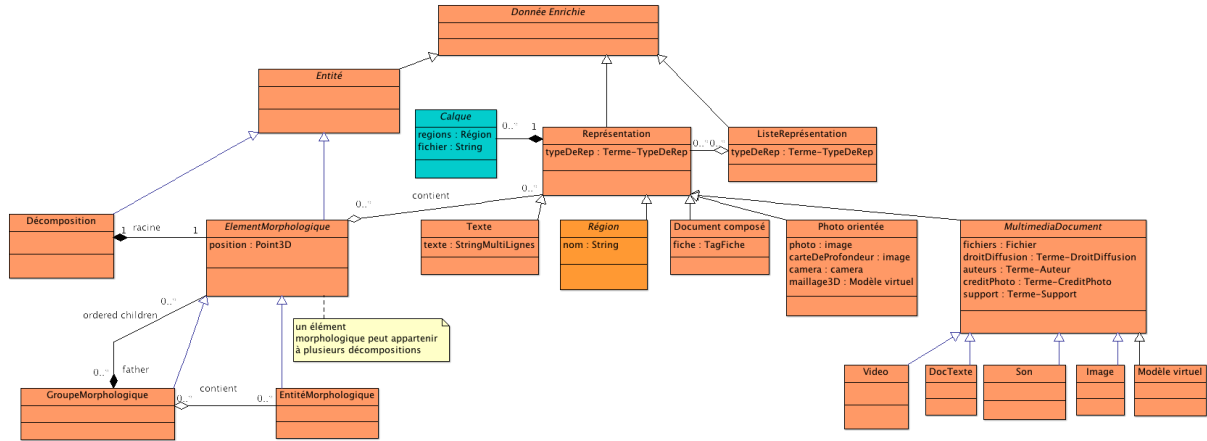
## Ressources Sémantiques

- Permettent d'enrichir sémantiquement une donnée
- Modèle simplifié en attendant une véritable structure adaptée aux ressources sémantiques

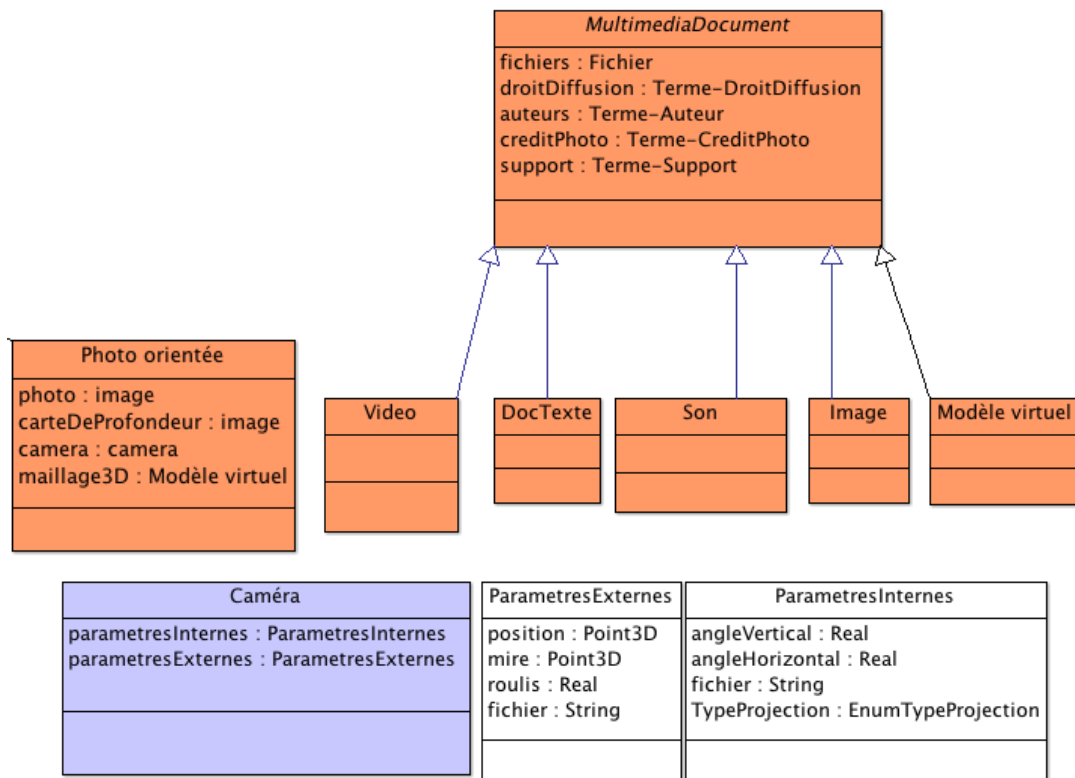


## Données Enrichissables

- Ces données peuvent être enrichies par le biais d'un tag
- Ex : Photo orientées, Document multimédia, Texte, etc...

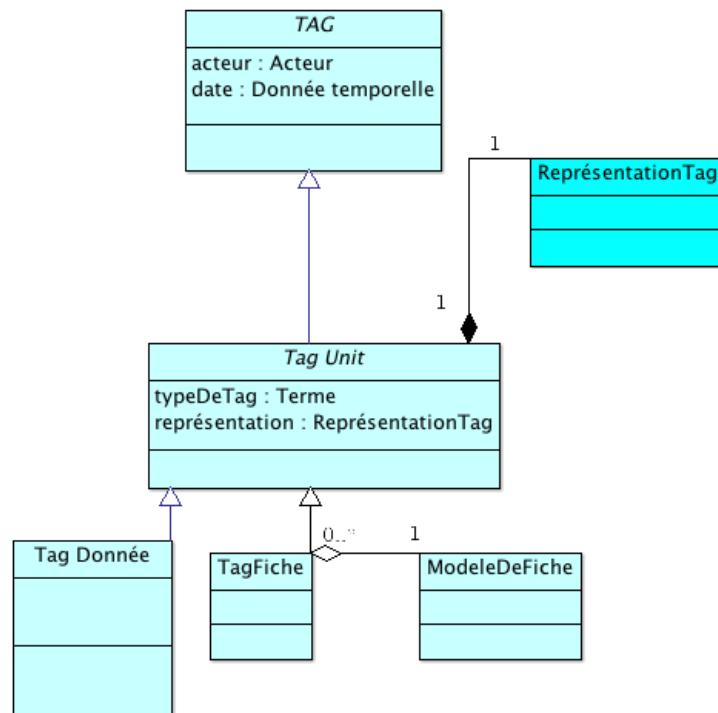


## Exemple de Structure de Données



## Tags

- C'est l'intermédiaire entre deux données
- On peut y coller n'importe quelle donnée (enrichissable ou basique)
- Un tag ne peut être collé que sur une donnée enrichissable

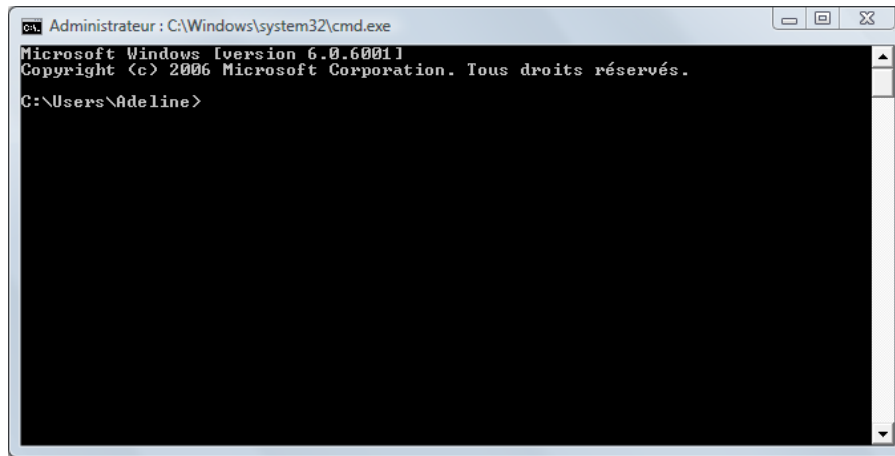


## Annexe D : Fonctionnement de Aïoli

Le fonctionnement des différentes interfaces d'annotation étant similaire, seul l'interface Matlab sera décrite.

## Interface à l'ouverture

Pour lancer Aioli sous sa version Matlab avec les messages de retour, il faut passer par l'invite de commande.

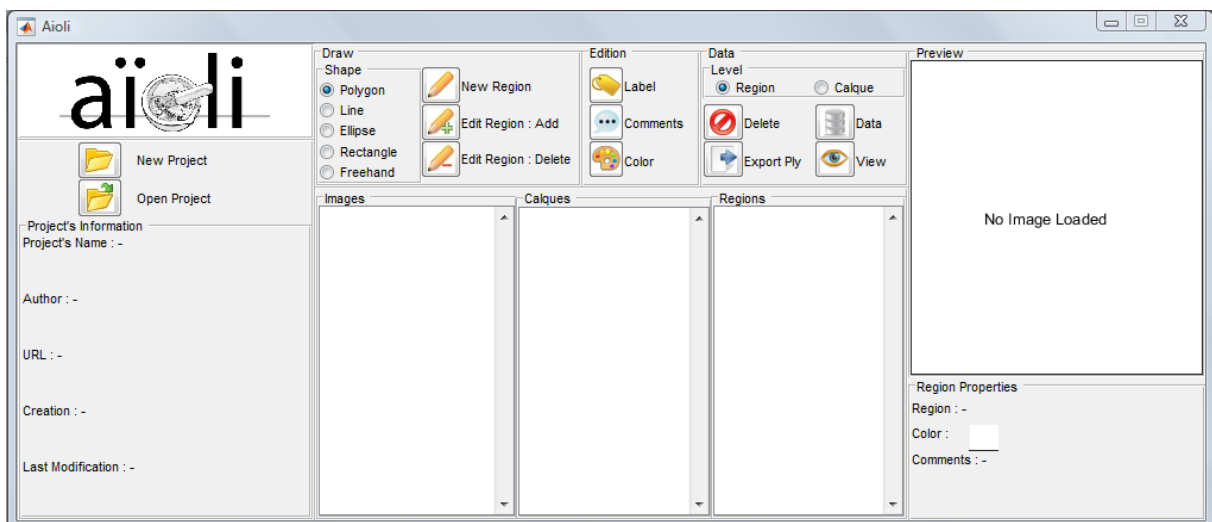


Tout d'abord, lancer la commande :

```
@echo on
```

Puis glisser/déposer l'exécutable nommé Aioli.exe ou taper le chemin complet de l'application et appuyer sur Entrée.

Si les commandes ont bien été prises en compte, le logiciel doit se lancer et le mot « Aioli » doit s'afficher dans l'invite de commande.



Ne pas fermer l'invite de commande pour avoir les messages d'avancement des processus.

Pour les autres interfaces, cette étape n'est pas nécessaire.

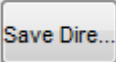
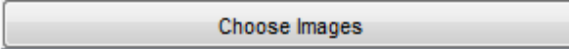


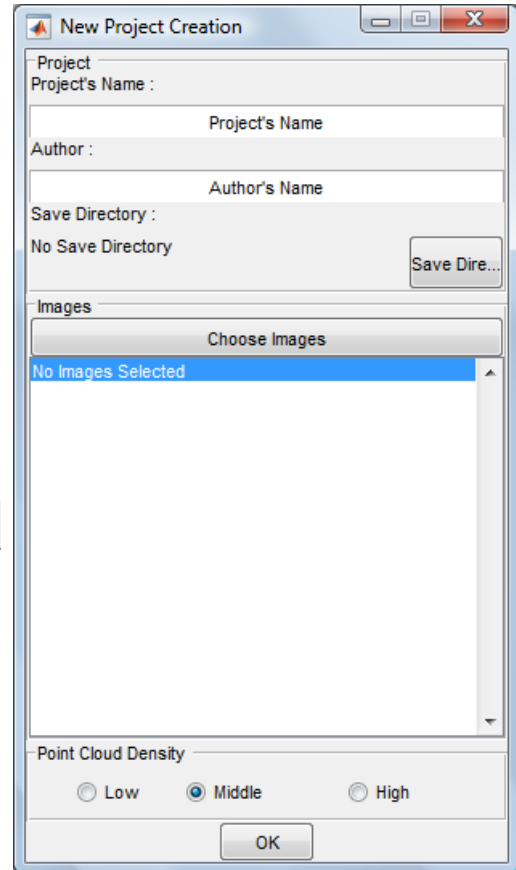
## Création d'un projet

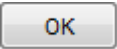


New Project

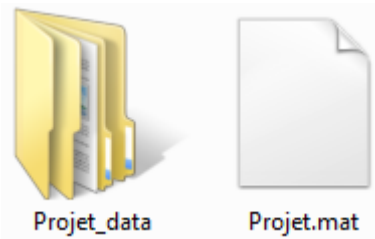
*Informations Projet* : La création d'un nouveau projet commence par le renseignement des informations du projet

- **Project's Name** : Nom du projet. Ce nom sera ensuite utilisé pour nommer le fichier à sélectionner pour la réouverture du projet
- **Author** : Auteurs du projet
- **Save Directory** : Sélection du dossier où le projet sera sauvegardé. Sélection par parcours de dossiers en choisissant .
- **Images** : Liste des images du projet. Sélection des images par parcours dans les dossiers en appuyant sur .
- **Point Cloud Density** : Choix de la densité du nuage de points pour le traitement par MICMAC. Plus la densité est élevée plus le temps de calcul est long, et moins la densité est élevée, moins le nuage de points aura de points :
  - **Low** : Basse qualité, Densité = 8
  - **Middle** : Qualité moyenne, Densité = 4
  - **High** : Qualité haute, Densité = 2



La validation  des informations lance le calcul de MICMAC, puis l'indexation du nuage de points.

A la fin du traitement, un fichier MAT portant le nom du projet et un dossier portant le nom du projet suivi de « \_data » sont créés dans la direction de sauvegarde choisie.




## Ouverture d'un projet

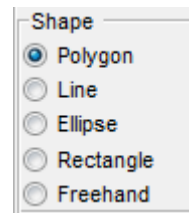



Open Project

L'ouverture d'un projet passe simplement par la sélection du fichier MAT portant le nom du projet.

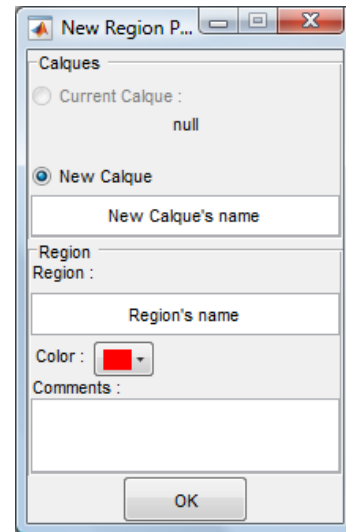
## Nouvelle Annotation New Region

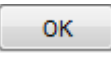
Pour définir une nouvelle annotation, plusieurs formes peuvent être choisies pour le dessin (polygone, ligne, ellipse, rectangle, libre). Le choix s'effectue avant de cliquer sur  New Region dans la partie Draw de l'interface.

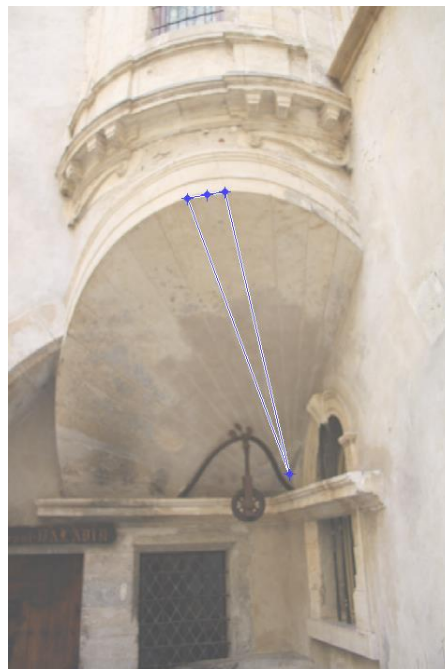


 New Region ouvre tout d'abord un menu pour pouvoir indiquer les informations de l'annotation :

- Choix du calque :
  - **Current Calque** : Calque sélectionné dans la liste des calques.
  - **New Calque** : Permet d'ajouter la région à un nouveau calque. Valeur par défaut si aucun calque n'existe.
- Informations Région :
  - **Region's name** : Nom de la région
  - **Color** : Choix de la couleur de visualisation de l'annotation
  - **Comments** : Commentaires de l'annotation



La validation  affiche l'image sélectionnée dans la liste des images pour permettre à l'utilisateur d'effectuer son tracé. Une fois le tracé effectué, le dessin doit être confirmé en double cliquant sur les bords du tracé.



La validation du tracé lance la propagation sur les autres images.

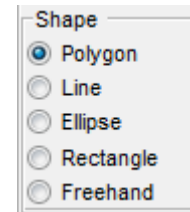
## Edition d'Annotation

### Extension de l'annotation



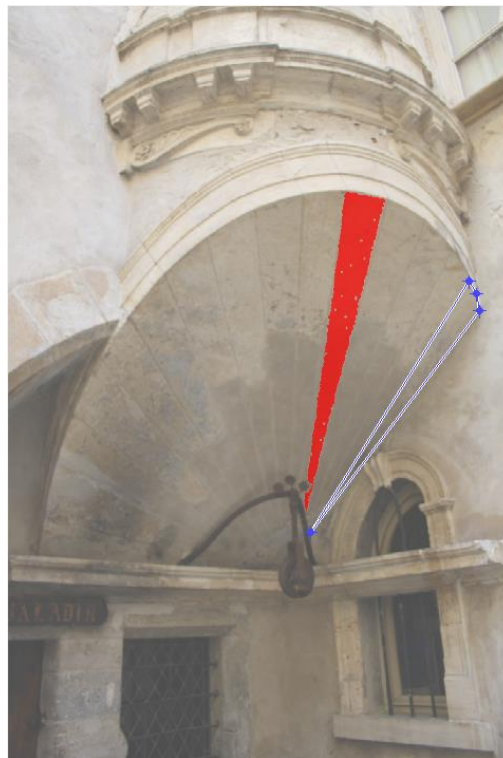
Edit Region : Add

La forme du dessin peut aussi être choisie parmi les formes disponibles (Polygone, Ligne, Ellipse, Rectangle, Libre).



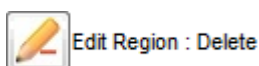
Edit Region : Add

affiche l'image sélectionnée dans la liste des images pour permettre à l'utilisateur d'effectuer son tracé. N'importe quelle image peut être choisie pour compléter l'annotation. La partie de la région déjà existante est affichée. Si d'autres régions existent dans le calque, elles sont affichées avec une transparence plus importante.

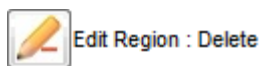
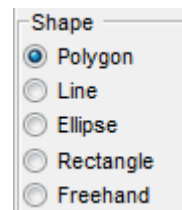


La sélection du tracé et sa validation fonctionnent de la même manière que pour une nouvelle annotation. Les points à l'intérieur de la sélection seront ajoutés à la région. La validation lance la mise à jour de la région sur toutes les images.

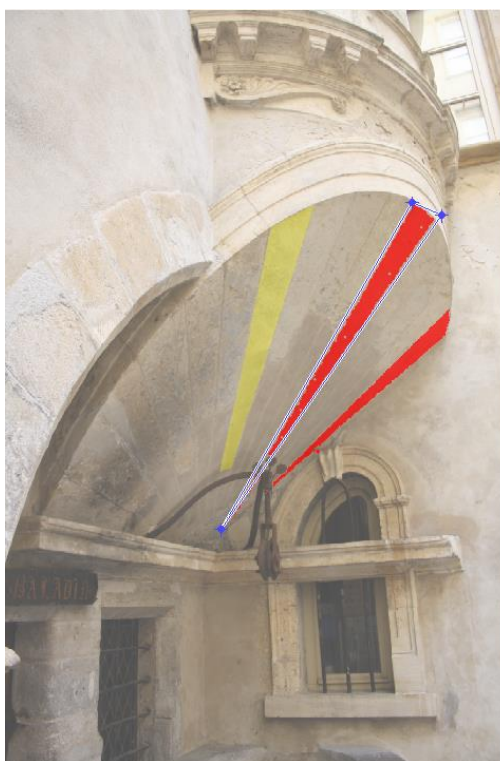
## Réduction de l'annotation



La forme du dessin peut aussi être choisie parmi les formes disponibles (Polygone, Ligne, Ellipse, Rectangle, Libre).



affiche l'image sélectionnée dans la liste des images pour permettre à l'utilisateur d'effectuer son tracé. N'importe quelle image peut être choisie pour éditer l'annotation. La partie de la région déjà existante est affichée. Si d'autres régions existent dans le calque, elles sont affichées avec une transparence plus importante.



La sélection du tracé et sa validation fonctionnent de la même manière que pour une nouvelle annotation. Les points à l'intérieur de la sélection seront supprimés de la région. La validation lance la mise à jour de la région sur toutes les images.

## Données de description

Les données liées à la région peuvent être éditées indépendamment les unes des autres.



Label

permet de modifier le nom de la région.



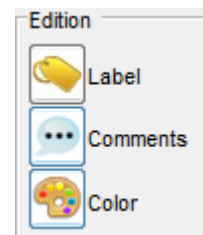
Comments

permet de modifier le commentaire de la région



Color

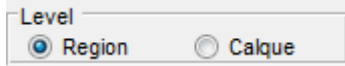
permet de modifier la couleur de visualisation de la région.



## Suppression de région ou de calque



permet de supprimer la région sélectionnée ou le calque sélectionné. Le choix entre suppression de calque ou suppression de région se fait selon la sélection



Une confirmation est demandée avant la suppression d'un calque ou d'une région.

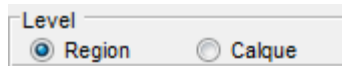
**!/ Attention :** La suppression d'un calque entraîne la suppression de toutes les régions contenues dans ce calque.

## Visualisation



permet de visualiser en plus grand une région ou un calque sur l'image sélectionnée dans la liste d'images. Le choix de visualisation d'une région ou d'un calque se

fait selon la sélection



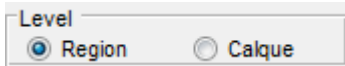
et en fonction du calque ou de la région sélectionnés dans les listes.

## Export Ply

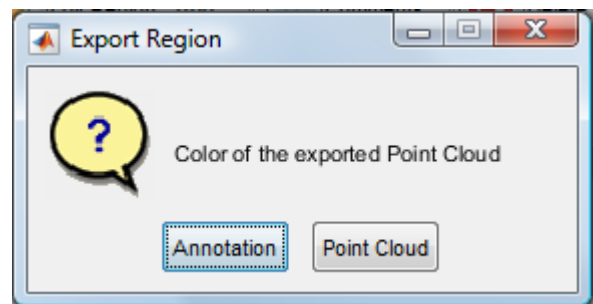


permet d'exporter la région ou le calque sélectionnés sous forme de nuage de points au format PLY. Le choix de l'exportation d'une région ou d'un calque se fait selon la

sélection



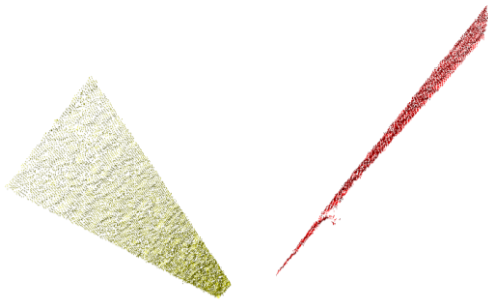
L'export permet de choisir la couleur de visualisation des régions soit en fonction de la couleur choisie pour la visualisation, soit avec la couleur du nuage de points.



Le nuage de points est sauvegardé dans le dossier projet finissant par « \_data », dans le sous-dossier « Calques », puis dans le dossier correspondant au calque. Le nom du fichier PLY est nommé en fonction de l'identifiant du calque, de la région (si seule une région est exportée) et du type de couleur choisie.

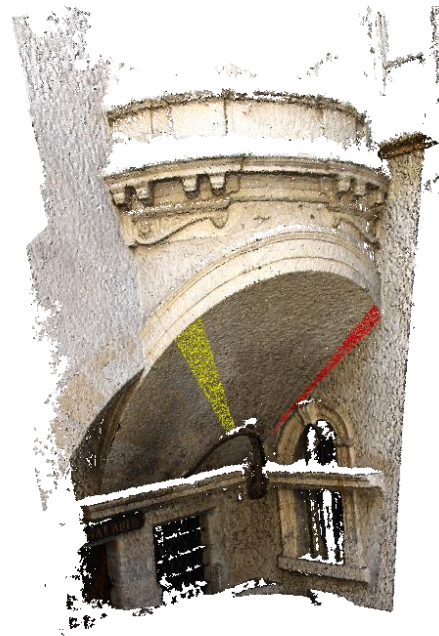
Le nuage de point exporté peut simplement être visualisé sous Meshlab ou directement sur les interfaces QT et WebGL. En superposant le nuage de points final de MicMac et le nuage de points des régions aux couleurs de visualisation, la position des régions dans le nuage de points est possible.

**!/ Attention :** Le fichier PLY n'est pas mis à jour automatiquement. Il faut le régénérer à chaque changement sur les régions.



*Nuage de points d'un calque avec les différentes régions colorées en fonction de la couleur de l'annotation.*

*Superposition entre le nuage de point de MicMac et le nuage de points extrait des annotations*



## Utilisation en lignes de commandes

Toutes les étapes de création du projet et d'annotation peuvent être utilisées en dehors de l'interface Matlab en lignes de commandes à l'aide des modules NewProject.exe, MicMac.exe, Data.exe et Annotation.exe. Ce sont exactement ces modules qui sont appelées avec les paramètres bien définis dans l'interface.

- *Création Dossier Projet*

Ouvrir l'invite de commande

Lancer la commande `@echo on` (pour avoir le retour sur l'invite de commande)

Glisser-déposer NewProject.exe puis ajouter les arguments entre guillemets " à la suite (l'ordre des arguments entre les guillemets n'a pas d'importance) :

- ProjName : Nom du Projet
- Author : Auteurs du Projet
- SaveDir : Emplacement où le projet sera créé



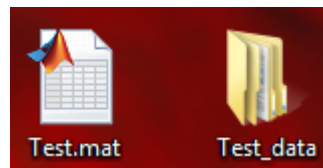
Lancer la commande

Le module est terminé lorsque End Project Creation est affiché

```
C:\Users\ama.MAP3495>J:\Aioli\NewProject.exe "ProjName=Test Author=Adeline Savedir=C:\Users\ama.MAP3495\Desktop"

C:\Users\ama.MAP3495>New Project Test in C:\Users\ama.MAP3495\Desktop
Project File Creation
DataBase Creation
Project Information
Project's Name: Test
Project's Directory: C:\Users\ama.MAP3495\Desktop\
Project's Author: Adeline
Project's Creation Time: 19-Nov-2015 13:55:53
End Project Creation
```

A la fin du module, un fichier NomduProjet.mat et un dossier NomduProjet\_data doivent avoir dans le dossier choisi



- *Traitement MicMac*
  - *Traitement par MicMac.exe*

Se placer dans le dossier NomduProjet\_data avec la commande cd :

```
C:\Users\ama.MAP3495\Desktop>cd C:\Users\ama.MAP3495\Desktop\Test_data
C:\Users\ama.MAP3495\Desktop\Test_data>
```

Glisser-déposer MicMac.exe puis ajouter les arguments entre guillemets " à la suite (l'ordre des arguments entre les guillemets n'a pas d'importance) :

- FormatImg : Format des images
- Quality : Densité du nuage de points (valeurs possibles : High, Middle, Low)

Le module est terminé lorsque End MicMac est affiché

```
C:\Users\ama.MAP3495\Desktop\Test_data>J:\Aioli\MicMac.exe "FormatImg=JPG Quality=Middle"
```

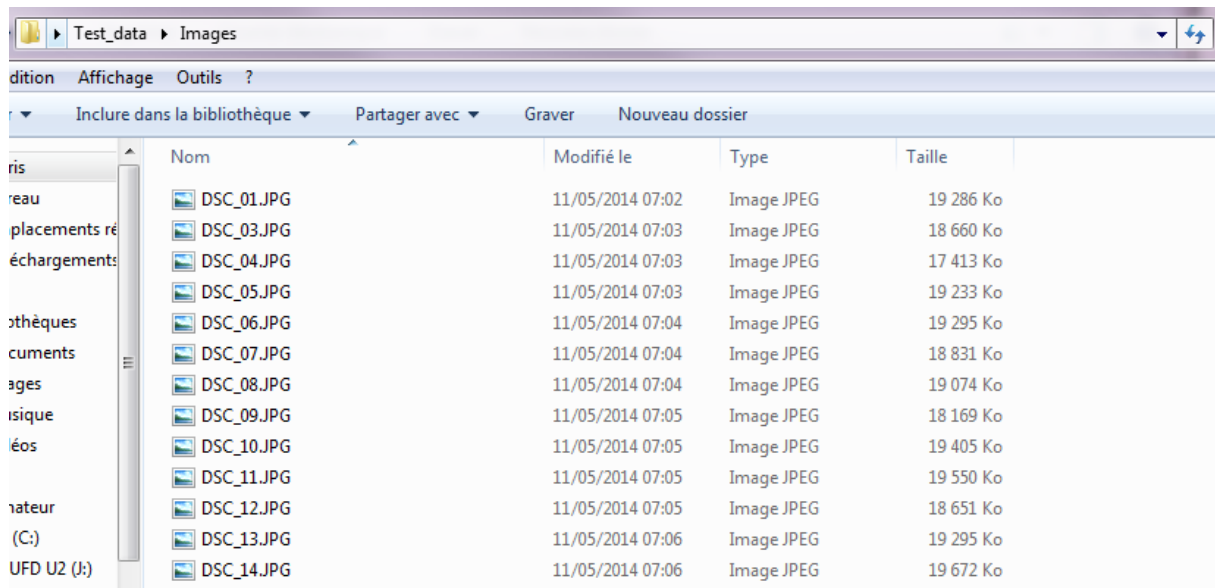
A la fin du module, les données générées par MicMac sont stockées dans le dossier MicMac du dossier projet.

- *Importation de données MicMac déjà générées en dehors de Aioli*

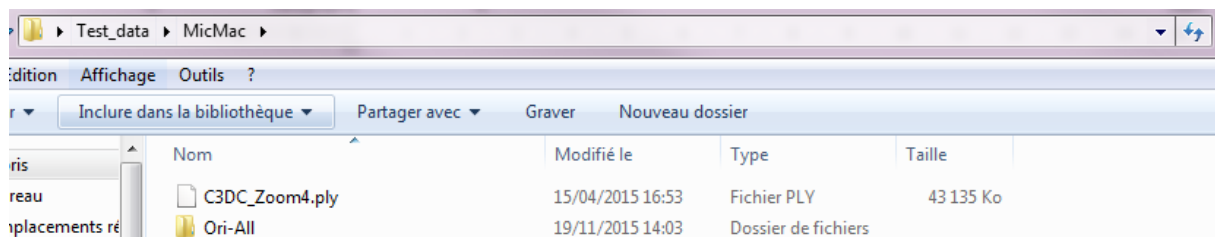
Il est possible de ne pas passer par MicMac.exe pour générer le nuage de point, les orientations et les calibrations des images. Cela peut être utile notamment dans le cas de projet complexe où les commandes basiques de traitement par MicMac se révèlent insuffisantes pour la génération des données.

Dans le dossier NomduProjet\_data, aller dans le dossier Images

Copier, dans ce dossier, les images ayant servie à la génération du nuage de points



Dans le dossier NomduProjet\_data, aller dans le dossier MicMac  
Copier, dans ce dossier, le dossier d'orientation de MicMac, qui doit être nommé Ori-All, et copier le fichier PLY du nuage de points, qui doit être nommé C3DC\_Zoom[ValeurdeZoomF].ply



- *Indexation*

Se placer dans le dossier NomduProjet\_data avec la commande `cd` (si cela n'a pas déjà été fait à l'étape précédente) :

```
C:\Users\ama.MAP3495\Desktop>cd C:\Users\ama.MAP3495\Desktop\Test_data
C:\Users\ama.MAP3495\Desktop\Test_data>
```

Glisser-déposer Data.exe et lancer la commande.

Le module est terminé lorsque `End Data` est affiché

Un fois Data.exe terminé le fichier peut être ouvert dans Aioli.exe comme n'importe quel autre projet

- *Annotation*

Se placer dans le dossier NomduProjet\_data avec la commande `cd` (si cela n'a pas déjà été fait à l'étape précédente) :

```
C:\Users\ama.MAP3495\Desktop>cd C:\Users\ama.MAP3495\Desktop\Test_data
C:\Users\ama.MAP3495\Desktop\Test_data>
```

Glisser-déposer Annotation.exe puis ajouter les arguments entre guillemets " à la suite (l'ordre des arguments entre les guillemets n'a pas d'importance) :

- 
- Mode : Mode de sélection (valeurs possibles : New, Edit+, Edit-, EditLabel, EditColor, Delete, Export, Data, Tree, View)
  - Img : Nom de l'image sur laquelle la sélection doit être faite
  - Shape : Forme de la sélection (valeurs possibles : Line, Polygon, Ellipse, Rectangle, Freehand)
  - Calque : Nom du calque de la région en cours de sélection
  - Region : Nom de la région en cours de sélection
  - NewLabel : Nouveau nom pour la région (pour le Mode EditLabel)
  - Comments : Nouveau commentaire pour la région (pour le Mode EditLabel)
  - RGB : Nouvelle couleur pour la région (pour le Mode EditColor) sous la forme [R G B]
  - PtCloudAttr : Attribut de la couleur du nuage de point extrait de l'annotation (pour le Mode Export) (valeurs possibles : Annotation ou PtCloud)
    - *Création d'une nouvelle annotation (Mode New)*  
 Le Mode New est le mode par défaut.  
 Arguments nécessaires : Img, Shape, Calque, Region
    - *Edition d'une annotation*
      - *Forme (Mode Edit+ ou Edit-)*  
 Arguments nécessaires : Img, Shape, Calque, Region
      - *Attributs (Mode EditLabel ou EditColor)*  
 Arguments nécessaires pour le Mode EditLabel : Calque, Region, NewLabel  
 Arguments nécessaires pour le Mode EditLabel : Calque, Region, RGB
    - *Suppression d'un calque ou d'une annotation (Mode Delete)*  
 Arguments nécessaires pour la suppression d'un calque : Calque  
 Arguments nécessaires pour la suppression d'une région : Calque, Region
    - *Visualisation des annotations (Mode Data, Tree ou View)*  
 Arguments nécessaires pour le Mode Data : Calque, Region  
 Arguments nécessaires pour le Mode Tree : Aucun  
 Arguments nécessaires pour le Mode View : Img, Calque, Region
    - *Export du nuage de points (Mode Export)*  
 Arguments nécessaires pour le Mode Export : Calque, Region, PtCloudAttr
-