

Référencement spatial et sémantisation de sources iconographiques

Dans le chapitre précédent nous nous sommes intéressés à la reconstruction de la morphologie de l'édifice afin de produire une représentation tridimensionnelle enrichie du point de vue sémantique. Dans ce chapitre nous nous concentrerons sur les méthodes permettant de mettre en relation spatiale des sources iconographiques (sous forme d'images 2D) par rapport à la morphologie de l'édifice reconstruite. Comme on le verra dans le chapitre suivant, cette relation spatiale servira aussi pour transférer les attributs sémantiques associés au modèle 3D sur les images 2D. C'est le point central de notre approche : l'utilisation d'une description sémantique de l'édifice (voir section 5.3.2) comme dénominateur commun qui relie la représentation 3D des formes architecturales aux sources documentaires bidimensionnelles.

Ce chapitre se divise en deux parties principales : dans la première partie nous nous intéressons aux méthodes actuelles pour le référencement spatial d'images 2D sur des ensembles de coordonnées 3D, c'est à dire de superposition spatiale entre l'image et le modèle 3D. Dans la deuxième partie de ce chapitre nous nous concentrons sur les modèles géométriques permettant d'exprimer les modèles géométriques des sources iconographiques dans l'espace 3D. Cette opération permet d'exprimer la position et l'orientation d'une image en fonction d'un référentiel, notamment celui du modèle 3D de l'édifice. Une analyse des différents modèles géométriques à utiliser ainsi que des différents niveaux d'approximation à prendre en compte (en fonction du type d'iconographie) permettra enfin d'identifier les méthodes les plus indiquées pour la spatialisation de différents types d'images (photos, vidéos, panoramiques, dessins, peintures, etc.).

6.1. Méthodes de référencement spatial

Afin de s'intéresser à la structuration des bases de données de sources iconographiques en relation à la morphologie d'un édifice, notre travail aborde le problème de l'organisation spatiale de photographies sur une représentation tridimensionnelle de l'édifice issue d'un relevé (tel qu'un balayage laser 3D ou une restitution photogrammétrique). La notion de « résection spatiale » [Hartley R. et al., 2000] (orientation d'une image sur un ensemble de coordonnées 3D connues) appliquée au relevé d'édifice, peut constituer le point de départ pour le développement d'un système de référencement spatial de sources documentaires (photographies, images panoramiques, séquences vidéo, etc.) à l'intérieur d'une scène architecturale.

Le référencement spatial est une opération qui crée une relation spatiale entre l'image 2D et la représentation 3D qui correspond à une superposition. Il s'agit en

quelque sorte de retrouver 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 (c'est le cas d'une photographie, où la cohérence entre la réalité et l'image est forte) ou de façon approximative (c'est le cas des dessins, où la cohérence entre la réalité et l'image est faible).

Les méthodes de référencement spatial permettent d'obtenir différents types d'informations géométriques relatives aux images (position, orientation, paramètres optiques, etc.) et à différents niveaux de précision. Une fois analysées les travaux et les solutions existantes dans ce domaine, on s'intéressera à l'identification des plus adaptées à chaque type de source iconographique à prendre en compte en fonction de ses propriétés. Comme on le verra dans les chapitres suivants, le référencement spatial des sources iconographiques permet d'envisager des modalités d'interrogation des bases de données d'images en fonction de critères spatiaux (à différents niveaux de précision). Nous allons présenter ici un ensemble de méthodes de référencement spatial en les regroupant en trois catégories distinctes : méthodes manuelles, semi-automatiques et automatiques.

6.1.1. Méthodes manuelles

Le référencement manuel des photographies est certainement la solution la plus simple à mettre en place et la plus facile aussi du point de vue de l'utilisateur final. Dans le domaine de la documentation architecturale et archéologique différentes recherches vise aujourd'hui à créer une relation entre les sources 2D et une représentation de la morphologie d'un édifice au travers d'outils permettant d'indiquer de façon approximative la position de la prise de vue d'une photographie. D'habitudes, ces solutions utilisent des représentations cartographiques (les SIG par exemple) permettant d'indiquer sur le plan d'un édifice la position (et parfois la direction de vue) d'une photographie. La position et l'orientation du point de vue de l'image sont généralement présentées par un symbole graphique, comme un point ou une flèche.



Figure 45 L'interface du système de [Kadobayashi R. et al.,2003], capable à référencer manuellement les information 2D et 3D

[Kadobayashi R. et al.,2003] propose un système qui permet d'effectuer des recherches d'images sur la base des informations stockées dans les légendes des photos. En même temps, ce système permet de manipuler les nuages de points et/ou le modèle 3D résultat du relevé et crée par maillage automatique (voir figure 45). Des

cibles de référencement (placées sur différentes parties de l'édifice) sont utilisées comme points de référence pour retrouver des photographies corrélées ; leurs coordonnées sont obtenues par des mesures topographiques avec une station totale. Chaque image a son propre système de coordonnées basé sur la position du scanner au moment de l'acquisition des informations. Ces images sont orientées selon un système de coordonnées géographique. L'interface du système de consultation comporte une vue planimétrique, une élévation et une vue 3D. L'utilisateur peut rechercher des images dans la base de données en cliquant sur les positions relatives indiquées sur les plans de l'édifice ou sur le plan de masse mais aussi en indiquant des mots-clés.

Le système APIS (Architectural Photogrammetry Information System) [Waldhäusl P. et al., 1994] dont certains aspects ont déjà été présenté dans le chapitre 2 se base sur une base de données qui répertorie les bâtiments documentés par des acquisitions photogrammétriques. Le système est basé sur un outil permettant d'enregistrer des informations sur les édifices documentés et pour enregistrer les positions d'où les photographies ont été prises au moment du relevé photogrammétrique. Pour démarrer un projet avec ce système, les photographes doivent préparer leur propre esquisse sous forme de plans et élévations du bâtiment (comme indiqué dans la règle organisationnelle 3x3) pour ensuite y marquer les position et les direction de la caméra de façon approximative (voir figure 46).

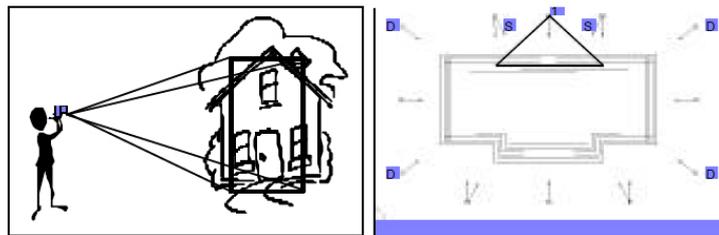


Figure 46 Planification des photographies du règle 3x3

Même si la méthode de référencement manuel des photographies est la méthode la plus simple et la plus compréhensible pour les utilisateurs du système, l'association d'images à leur point de vue nécessite une gestion manuelle de l'administrateur système. Les erreurs de détermination humaines sont toujours possibles, cela peut réduire la précision du système. De plus, c'est une méthode qui a besoin de beaucoup de temps de travail pour arriver à un niveau de précision suffisant pour qu'il soit utilisable dans notre méthodologie.

6.1.2. Méthodes semi-automatiques

Les méthodes semi-automatiques de référencement spatial d'images se basent sur des solutions géométriques appelées de calibration et orientation des caméras. Ceci se réfère au processus d'utilisation des valeurs numériques pour établir des paramètres géométriques et optiques de la caméra virtuelle qui correspondent à un point de vue de l'image par rapport à un référentiel 3D (ensemble de points, modèle 3D, etc.) [Cipolla R. et al, 1999]. Les méthodes de calibration et orientation des caméras peuvent prendre en compte les paramètres intrinsèques (distance focale, distorsion,

etc.) et extrinsèques (translation et rotation) de l'appareil photo (ou plus génériquement de la camera) au moment de la prise de vue rapport à un repère tridimensionnel (voir figure 47). Plusieurs méthodes ont été développées.



Figure 47 Calibration et orientation des caméras

La méthode DLT (Transformation linéaire directe) est une méthode rapportée par [AbdelAziz et al., 1971]. La méthode de DLT emploie un ensemble de points de commande dont les coordonnées d'espace-plan d'objet sont déjà connues. Les points de commande sont normalement fixés à un cadre rigide. Le problème principal de la méthode DLT est que les paramètres du calibrage ne sont pas mutuellement indépendants.

Parmi les recherches développées pour orienter une image sur un ensemble de coordonnées connues dans le domaine de la photogrammétrie [Hatze, 1988] et de la vision par ordinateur [Faugeras, 1993 ; Fitzgibbon et al., 1998 ; Heikkila et al., 1997 ; Kumar R. et al., 1994], une des méthodes les plus utilisées et celle proposée par [Tsai, 1986]. Son implémentation requiert des correspondances (points) sélectionnées respectivement dans l'espace 3D et sur l'image 2D. Cette méthode est basée sur un calcul en deux phases : premièrement on calcule la position et l'orientation, deuxièmement les paramètres internes de la caméra. Dans plusieurs applications qui nécessitent une bonne précision, on considère la caméra déjà calibrée. Cela veut dire que les paramètres internes sont déjà connus. Dans ce cas, seulement les six paramètres relatifs à la position et à l'orientation doivent être estimés.

6.1.3. Méthodes automatiques

Nous pouvons référencer les images au modèle géométrique tridimensionnel par une technique automatique en deux façons. La première consiste à utiliser un outil pour enregistrer l'information spatial du point de vue d'image au moment qu'elle été prise (hardware). Une autre option est d'utiliser un algorithme de l'ordinateur (software)

pour trouver le point de vue d'image qui a été prise sans enregistrée les données spatiales.

6.1.3.1. Solutions hardware

Dans cette section nous nous intéressons aux solutions hardware existantes pour l'enregistrement automatique des paramètres spatiaux d'un appareil photo au moment de la de prise de vue.

[Hu W. et al. 2006] ont développée une solution d'estimation de la position et de l'orientation d'une camera 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.

Pour la réalisation de représentations tridimensionnelles photoréalistes d'objets architecturaux [Abmayr T. et al., 2004] proposent une solution pour enrichir les nuages de points avec les valeurs colorimétriques acquises par une caméra haute résolution intégrée au système de balayage produit par la société Zoller+Fröhlich. La caméra, développée en collaboration avec la German Aerospace Center (DLR), consiste en une unité pivotante qui intègre une caméra DTC capable d'acquérir des images à 10200 x 500000 pixels. L'acquisition laser et photographique est effectuée l'une après l'autre. Une fois la numérisation laser terminée, la caméra est montée sur le même pied à l'aide d'un adaptateur qui assure le positionnement précis du centre optique de la caméra sur celui du scanner. Cela permet (par la simple détermination du paramètre de rotation horizontale de la caméra) de superposer l'image sur le nuage de points. Cette procédure consiste à associer les attributs RVB extraits de l'image à chaque coordonnée du nuage de points. La figure 48 illustre un nuage de points enrichi par cette technique.



Figure 48 Le hall du château Neuschwanstein : l'image de luminance du scanner est mélangée avec les valeurs RVB correspondantes acquises par la caméra. [Abmayr T. et al., 2004].

[Salemi G. et al., 2005] présentent une expérimentation pour l'intégration d'images panoramiques dans un nuage de points issu d'un relevé laser. L'objectif est d'associer des informations métriques à un système pour l'exploration de panoramas cylindriques. L'intégration est basée sur la fusion des données issues de deux

systèmes d'acquisition. À partir d'une numérisation panoramique (360°) avec un module de balayage temps de vol et d'une acquisition photographique faite par la caméra embarquée dans le scanner (calibrée préalablement par les constructeurs du scanner), les auteurs se focalisent sur l'extraction d'entités caractéristiques sur les images issues de la phase d'acquisition. Le processus d'extraction d'entités se base sur trois phases : la première est l'application d'un filtre « Edge-Sobel » sur les valeurs d'intensité acquises par le scanner (nuances de gris). La seconde est l'application du même filtre en utilisant la modalité « Keep sign of result (dans une seule direction) » sur les valeurs RVB de l'image. De cette manière les arêtes sur l'image sont contournees. Une application de la méthode a été expérimentée à l'occasion du relevé du bâtiment « Suriano Bellotto » à Venise (voir figure 49). Trois images sources ont été utilisées pour le traitement : l'image panoramique en couleurs visible, l'échelle d'intensité en nuances de gris et la carte de profondeur.

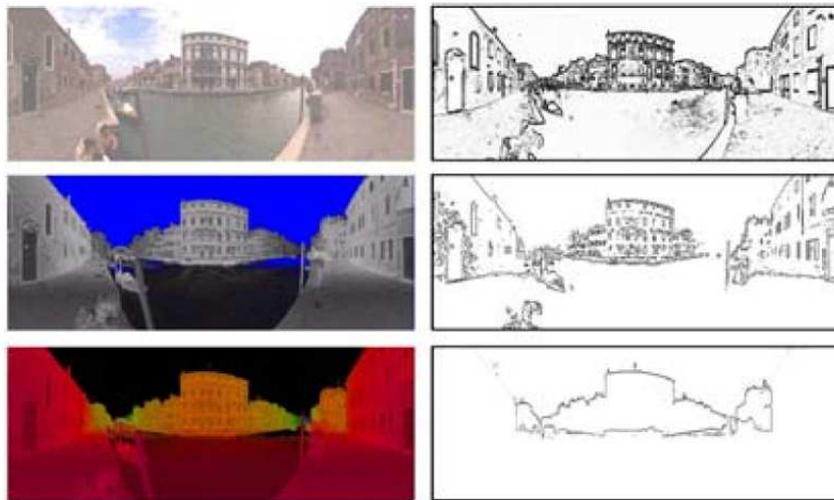


Figure 49 Acquisition laser panoramique à Venise. À gauche, images de l'acquisition laser. À droite, les images filtrées correspondantes. De haut en bas : l'image en vraies couleurs, l'image d'intensité en nuances de gris, la carte de profondeur. [Salemi G et al., 2005].

Le module [Omega] a été développé par le laboratoire MAP-GAMSAU en relation avec notre travail de thèse. Celui-ci permet d'enregistrer, de façon automatique, la position et l'orientation des photographies au moment de la prise de vue. Il utilise une station totale possédant une camera vidéo embarquée. La station total, en mode « tracking », possède un prisme, positionné sur le module OMEGA, ce qui lui permet d'atteindre de très longues portées (jusqu'à 500 mètres) en balayant sur 360° autour de son axe.

La station totale utilise ce qu'on appelle une « méthode télémètre » (en anglais ; range-finder) : un faisceau infra-rouge est intégré dans le processus qui utilise un réflecteur à prisme tétraédrique comme une cible. La station totale est également capable de prendre des photos à partir d'une camera interne dont la position et l'orientation relative sont connues. Le prisme (dont la station totale contrôle la position dans l'espace) et les boules de repères montées sur le module OMEGA sont alors utilisés comme références pour déterminer l'orientation de l'appareil photo au travers d'un algorithme d'analyse d'image (voir figure 50).



Figure 50 Système de OMEGA, une méthode de référencement automatique d'image.

Comme les positions relatives entre le prisme, les boules et l'appareil photo montées sur le module sont connues a priori, la position et l'orientation du module peut être déterminé à partir de la détection de la position des boules (par un algorithme d'analyse d'image) enregistrée par la camera intégrée à la station totale au moment de la prise de vue.

6.1.3.1. Solutions software

Concernant les méthodes algorithmiques basées sur l'analyse et le traitement d'images, [Lee S.C. et al.,2002] proposent une méthode pour estimer la position et l'orientation des caméras à partir d'une extraction automatique des points de fuite dans l'image. Les points de fuite sont extraits à partir d'un ensemble de lignes parallèles dans l'image et sont alors utilisées pour l'estimation des paramètres des caméras.

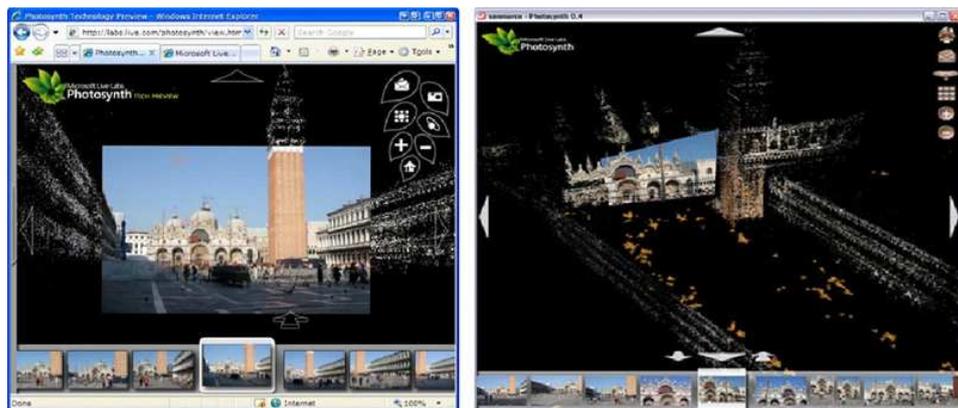


Figure 51 L'interface de Photosynth

Photosynth [Snavely N. et al.,2006 ; Snavely N. et al.,2007] est un logiciel développé par Microsoft Live Labs et l'Université de Washington. Ce logiciel analyse des photographies numériques et génère un nuage de points correspondants aux coordonnées extraites à partir des points homologues repères de façon automatique

dans une collection d'images (voir figure 51). Le système permet des mises à jour des collections d'images au travers d'une procédure de superposition manuelle de nouvelles images sur les images déjà orientées dans l'espace. Une fois une solution approchée est indiquée par l'utilisateur, l'algorithme d'extraction de points homologues [Lowe D., 1999] crée des correspondances entre la nouvelle image et une ou plusieurs images orientées, ces correspondances sont ensuite utilisées pour l'orientation de la nouvelle image par un calcul d'ajustement des faisceaux.

[Johansson B., 2002b] décrit un système automatique pour l'estimation de la position et de l'orientation d'une image simple représentant une scène urbaine, dans le cas où les modèles des édifices présents dans l'image sont disponibles. Chaque édifice a un modèle qui consiste en un certain nombre de plans parallèles associés. Une homographie (un plan dans la scène superposé à une image) est estimée de façon automatique pour chaque édifice présent dans l'image. Le modèle ayant le meilleur score de superposition est choisi pour déterminer la position et l'orientation de la caméra. Cette solution s'appuie sur l'hypothèse que les édifices sont composés principalement par des plans et que ces plans ont des arêtes verticales et horizontales bien visibles (fenêtres et/ou portes des édifices)(voir figure 52).

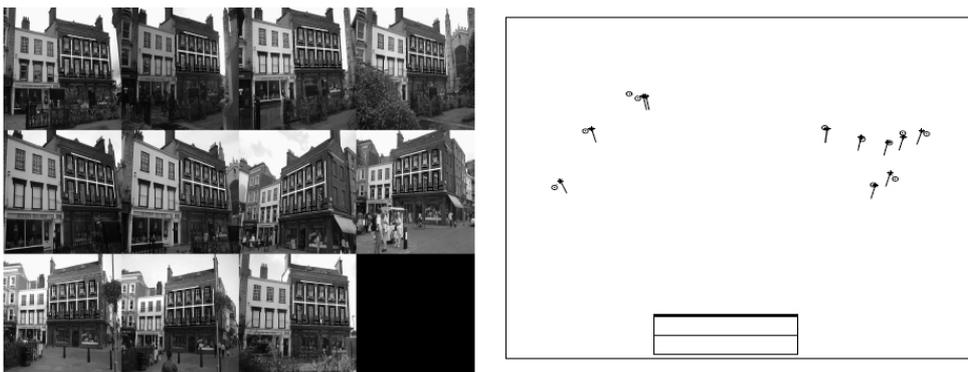


Figure 52 À gauche, images de différents points de vue avec le modèle adapté à un bâtiment. À droite, Le résultat de l'estimation de la position

Pour chaque édifice dans l'image, un modèle 3D est généré. Ce modèle consiste en un certain nombre de plans en 3D en fonction de la détection de la position de leurs arêtes. Pour chacun des édifices candidats, le modèle correspondant est superposé à l'image. Dans une deuxième phase, l'estimation de la position et de l'orientation de la caméra associées à l'image sont calculées. Ceci est possible en ayant des informations métriques concernant le modèle. Dans ce cas la méthode utilisée est l'auto-calibration qui fonctionne avec un simple plan dans une seule image. Comme la caméra a 11 paramètres, 6 externes et 5 internes, quand la géométrie du plan est connue, l'homographie donne deux contraintes dans la calibration.

6.2. Analyse des solutions existantes

A partir de l'analyse des solutions de référencement spatial d'images étudiés, un certain nombre de considérations sont à prendre en compte afin d'établir leur degré d'adaptabilité au problème spécifique que nous souhaitons traiter.

GPS, tout en étant particulièrement flexibles et faciles à mettre en place dans leur utilisation, ne peuvent pas vraiment être prise en compte pour nos exigences à cause de leur faible niveau de précision. En effet, pour les utilisations que nous envisageons de mettre en place au sein de notre système d'informations, nous nécessitons d'un degré de cohérence géométrique très fin, de l'ordre du décimètre en termes de positionnement de l'appareil photo, ce niveau de précision ne pouvant pas être obtenu par la technologie GPS. Ces technologies ne sont pas prises en compte non plus à cause du fait qu'au delà de la position de l'appareil photo, nous devons également connaître son orientation (toujours de façon assez précise) dans l'espace. En revanche, les solutions basées sur une superposition fine des images panoramiques sur un nuage de points, ou encore la solution OMEGA permettent un référencement en termes de positionnement et d'orientation à la fois tout à fait compatible avec les niveaux de cohérences nécessaires à une superposition précise des images photographiques sur le modèle 3D résultant du relevé architectural. Toujours en prenant en considération la précision, il est certes que toutes les solutions logiciels basées sur un calcul de calibration, puis orientation d'ensembles d'images (manuel ou automatique) restent les plus adaptées à l'exploitation des relations projectives entre les images photographiques et les modèles 3D, car toutes les coordonnées utilisées pour la construction de ce modèle est directement extraites des images. Cela veut dire que le modèle 3D issu d'une reconstruction à partir d'images se trouvera toujours en cohérence géométrique avec les images photographiques.

En ce qui concerne la flexibilité d'application, il est évident que l'utilisation de dispositifs hardware contraint énormément le champ d'application aux seules photographies ou images vidéos, de plus acquises dans des conditions particulières de prise de vue (notamment dans des campagnes de relevé métrique). Dans ce sens, les solutions hardware ne seront pas une solution effective pour toutes les sources iconographiques collectées par étude documentaire et provenant d'archives historiques. Concernant la flexibilité d'acquisition, les solutions basées sur un calcul de résection spatiale (points 2D saisis sur l'image et points 3D correspondants saisis sur le modèle 3D) s'avèrent les plus adaptées à la grande hétérogénéité de formats, de niveaux de résolutions, d'état de conservation caractéristique des fonds photographiques des monuments historiques. En effet, ces solutions déterminent la superposition image/modèle à partir d'une estimation à la fois des paramètres intrinsèques et extrinsèques en ayant ainsi la possibilité de s'adapter (avec des niveaux d'approximation conséquents) à différents types de photographies. De plus, ces solutions sont aussi adaptées au référencement spatial de représentations artistiques (dessins, peintures, gravures, etc.) dont la construction géométrique se base sur la perspective. En effet, malgré les incohérences (parfois particulièrement accentuées) existantes entre un dessin en perspective et la réalité, le calcul d'ajustement de faisceaux permet d'obtenir une estimation de la position et de l'orientation de l'image convenable, tout en restant dans la forte approximation.

6.3. Solution retenue

Afin d'implémenter dans notre système une solution pour le référencement spatial de sources iconographiques, nous avons pris en considération deux aspects principaux.

Le premier aspect concerne la grande hétérogénéité d'images qui constituent l'iconographie relative à un édifice historique. Comme expliqué dans le chapitre de cette thèse, plusieurs supports (dessins, photos, peintures), plusieurs formats et résolutions, plusieurs états de conservation et surtout différents degrés de certitude peuvent être attribués aux sources iconographiques. Dans ce sens, une première préoccupation a été de concevoir une solution particulièrement facile à gérer par les utilisateurs finaux dont les exigences sont limitées à une spatialisation approximative des documents graphiques.

Le deuxième aspect concerne le fait que, comme nous l'avons expliqué dans le chapitre 4 présentant le schéma général de notre approche, nous nous appuyons sur une représentation 3D de l'édifice conformé à la réalité (toujours avec un niveau d'approximation) en termes métriques et géométriques. C'est la raison pour laquelle nous n'avons pas besoin de mettre en place de solutions qui créent des représentations tridimensionnelles directement à partir des images.

A partir de ces deux constants nous avons pris en compte l'implémentation de trois différentes solutions de référencement spatial pouvant représenter trois différents degrés de précision.

Au premier niveau nous trouvons une solution pour la superposition manuelle de sources à projection en perspective (ou pseudo-perspective) et orthogonale directement sur le modèle 3D au travers du contrôle du déplacement et de l'orientation de la caméra de navigation à l'intérieur de la scène 3D de l'édifice. Cette solution est particulièrement flexible et adaptée au référencement spatial de sources iconographiques comme les dessins, les peintures et les gravures à projection perspective (ou presque) et orthogonale (plans, élévations et coupes).

Au deuxième niveau nous trouvons une solution pour la superposition géométrique d'images photographiques basée sur un calcul de résection spatiale. Cette solution est suffisamment flexible pour prendre en compte la grande variété (en termes de résolutions, de format et de précision) des fonds de photographies concernant les édifices patrimoniaux (photos d'archive, cartes postales, photos de tourisme et photos de relevé).

Au troisième niveau nous trouvons une solution pour la récupération directe des paramètres de calibration et d'orientation de photographies de relevé calculés dans des solutions de photogrammétrie pure ou dans le cadre de l'utilisation complémentaire et intégrée de balayage laser 3D et photogrammétrie.

L'aspect intéressant de notre solution est que toute source iconographique spatialisée par une solution approximative, peut toujours faire l'objet d'un nouveau référencement spatial conduit par une méthode plus précise. Il est par exemple possible de superposer rapidement un jeu de photographies avec la méthode de

référencement manuel pour ensuite affiner leur orientation via la méthode d'orientation géométrique.

Dans les paragraphes suivants nous présentons les solutions adoptées en fournissant aussi des détails sur leur implémentation informatique au sein de notre système.

6.3.1. Téléchargement sur serveur des sources iconographiques

La procédure de référencement spatial implémentée démarre par le téléchargement des images sur le serveur du système. Ce processus se base sur un script PHP permettant de sélectionner une image parmi les répertoires de l'ordinateur du côté client. Les images sont enregistrées sur le serveur en passant par un formulaire permettant de leur associer des attributs descriptifs tel que le nom de l'auteur, le type de sources, le format, la propriété, etc (voir figure 53). Ces attributs sont organisés conformément au modèle de description des sources iconographiques de la base de données officielle du ministère de la Culture et de la Communication français.

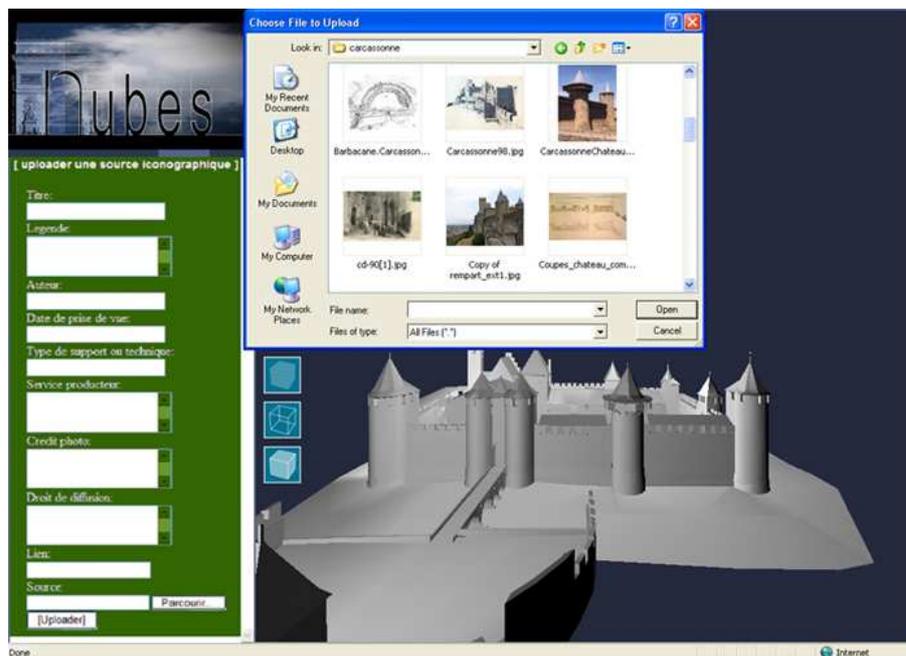


Figure 53 L'interface pour le téléchargement d'images sur serveur

Aux informations descriptives rentrées par l'utilisateur s'ajoutent des informations directement extraites de la lecture des paramètres de l'image (dimensions en pixels) et l'url du fichier image stocké dans la base de données. Un script PHP envoie ensuite une commande à la scène 3D temps réel afin que l'image soit téléchargée et visualisée dans la scène 3D. Un plan (d'un niveau de transparence configurable de façon interactive) est alors placé dans la scène 3D de manière orthogonale à la camera de navigation. La taille du plan est calculée en fonction des proportions de l'image (à partir des informations dimensionnelles stockées dans la base de données).

6.3.1. Superposition manuelle de sources à projection orthographique et pseudo-perspective

Pour la superposition manuelle de sources iconographiques à projection perspective et pseudo perspective nous avons développé une interface interactive 3D. L'objectif de cet outil est de fournir une voie très rapide et facile aux utilisateurs pour positionner de façon approximative des images comme des dessins, des peintures et des gravures dont la construction graphique se base sur la perspective aussi bien que des photographies. L'interactivité programmée permet aux utilisateurs de manipuler un plan 2D (sur lequel l'image est plaquée) en le contrôlant dans l'espace 3D en liaison directe avec la camera de navigation temps réel.

Au travers de contrôles de mouvement et de rotation, l'utilisateur peut déplacer la camera de navigation (toujours en ayant l'image à superposer comme plan d'image de la camera) afin de retrouver le point de vue correspondant dans la scène 3D (voir figure 54)



Figure 54 L'interface pour le référencement manuel.

Au-delà de la position et de l'orientation du point de vue, un contrôle permet de modifier (toujours en temps réel) l'angle de champ de la camera (distance focale) afin d'adapter la vision de la scène 3D à la perspective de l'image (voir figure 55).



Figure 55 L'interface pour le référencement manuel. Variation de l'angle de champ de la caméra.

Afin d'offrir une manipulation plus simple en phase de superposition, d'autres contrôles permettent de pivoter l'image (en sens horaire et antihoraire) autour de l'axe principal de la camera (voir figure 56).

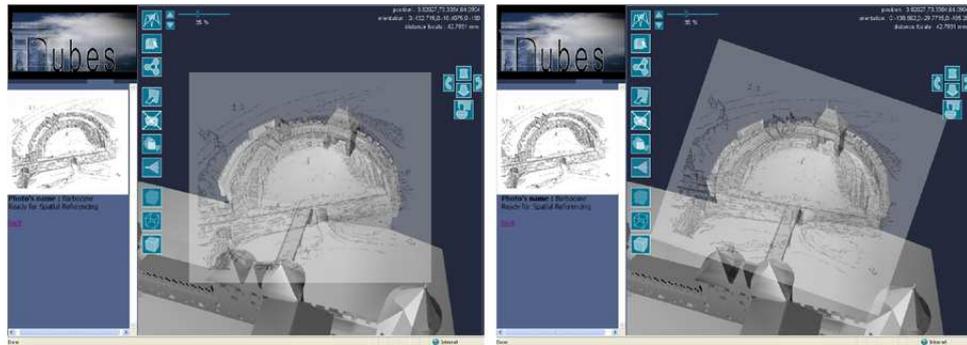


Figure 56 L'interface pour le référencement manuel. Rotation du plan image.

Afin de prendre en compte la superposition de plans, de coupes et d'élévations, un contrôle permet de basculer entre la vision de la scène 3D au travers d'une camera à projection perspective à une vision à projection orthogonale (voir figure 57). Dans ce cas, le contrôle de la distance focale est remplacé par un contrôle de zoom.

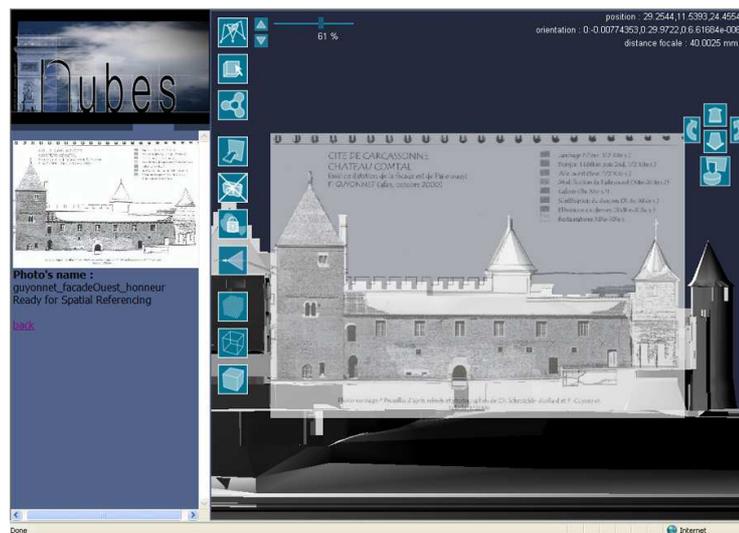


Figure 57 L'interface pour le référencement manuel. Projection orthographique.

Pour prendre en compte la superposition de sources présentant des sections (plans et coupes par exemple), une fonction supplémentaire permet de contrôler les limites de la pyramide visuelle de la camera (frustum) afin de limiter la vision aux seuls éléments représentés dans la source.

6.3.2. Résection spatiale de sources à projection perspective

Dans ce cas de figure nous prenons en compte un modèle géométrique de camera à projection en perspective défini par des paramètres intrinsèques et extrinsèques.

Les paramètres intrinsèques expriment les propriétés projectives de l'appareil photo (ou d'une perspective dessinée). Il s'agit de la distance principale (ou focale de l'objectif dans le cas d'une photographie), des coordonnées du point principal (centre perspectif projeté sur le plan image) et de la distorsion qui correspond aux erreurs de représentation introduites par la forme et les imperfections de l'objectif (toujours dans le cas d'une photographie). La distorsion est par un simple coefficient de distorsion radial. En effet, les corrections de la distorsion se caractérisent par un déplacement quasiment nul au point principal et graduellement accentué vers les extrémités de l'image. Ce modèle, beaucoup plus simple d'un modèle polynomial généralement utilisé en photogrammétrie, offre des avantages en termes de flexibilité, notamment en présence d'images avec une faible cohérence géométrique avec le réel (photos déformées, dessins en perspective, etc.)

Les paramètres extrinsèques représentent la position et l'orientation de l'appareil photographique au moment de la prise de vue. La position et l'orientation de la caméra par rapport à ce repère sont déterminées par une translation et une rotation au moyen de six composantes au total. Une translation est définie par les composantes (T_x , T_y , T_z) relatives aux axes. Une rotation est décrite par trois angles (R_x , R_y , R_z) autour de chacun des axes.

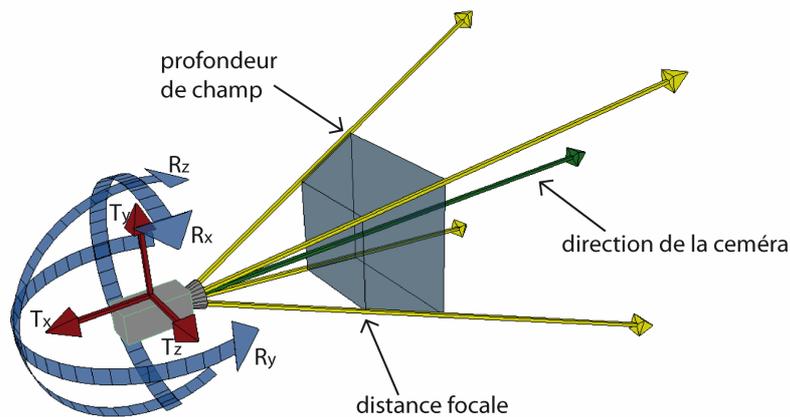


Figure 58 Le modèle géométrique d'une caméra à projection perspective.

Pour réaliser le calcul de résection spatiale nous utilisons l'algorithme de [Tsai R.Y., 1986] qui est basé sur un modèle de projection perspective et permet d'estimer les onze paramètres suivants : distance focale de la caméra ; coefficient de distorsion radiale de l'objectif ; coordonnées du centre de distorsion radiale ; facteur d'échelle pour prendre en considération l'incertitude due aux imperfections en phase de numérisation de l'image; angles de rotation pour la transformation entre les coordonnées de l'objet et de la caméra; composantes de translation pour la transformation entre les coordonnées de l'objet et de la caméra.

La calibration de Tsai consiste en la transformation des coordonnées 3D d'un point en coordonnées correspondantes dans l'image. Une calibration coplanaire requière que la composante Z des coordonnées 3D soit fixée à 0 et au moins 5 points

d'entrée. Tandis qu'une calibration basique non coplanaire requiert au moins 7 points d'entrée.

Malgré cela nous avons pu constater qu'une calibration optimisée requiert 11 points minimum pour les deux modalités. D'autres considérations autour de la qualité de la calibration concernent la stratégie de saisie des correspondances : les points choisis doivent en effet être distribués largement dans le champ visuel de la caméra et couvrir toute la gamme de profondeur présente sur l'image (ou dans la géométrie que l'on veut restituer).

Cette solution offre l'avantage d'estimer les paramètres internes et externes séparément. Il s'agit d'une option particulièrement utile car elle donne la possibilité de fixer les paramètres internes de la caméra calculés sur la première image d'un bloc puis de faire seulement l'estimation des paramètres externes des autres images de la même série (prises avec le même appareil photographique et avec le même niveau de zoom par exemple).

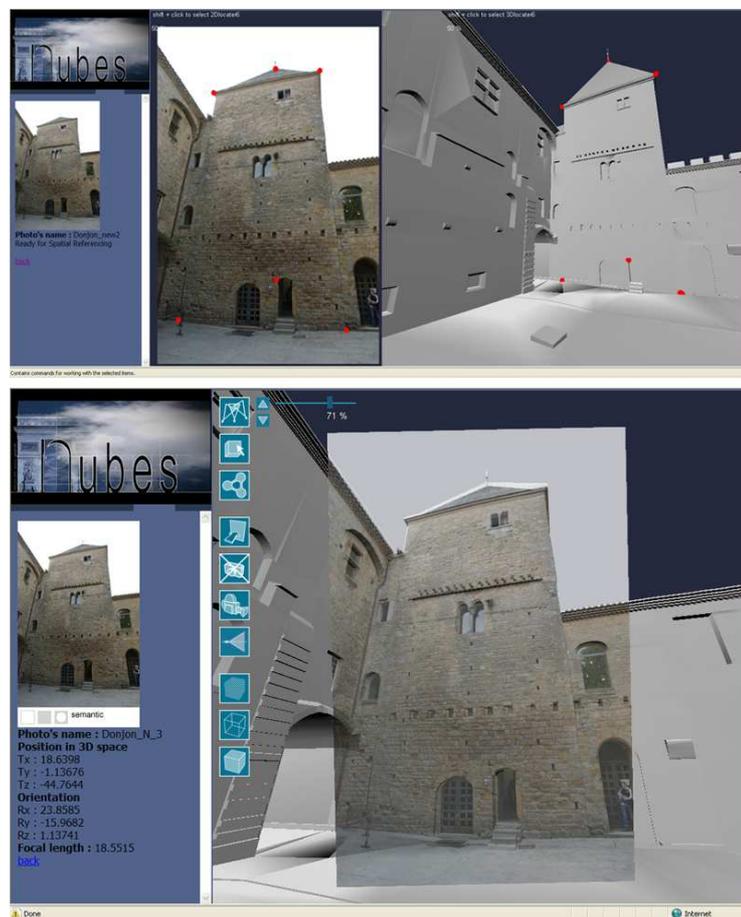


Figure 59 L'interface pour la résection spatiale d'images. En haut : le processus de sélection des points d'entrée sur le modèle 3D et sur l'image, en bas : le résultat du référencement.

A partir de nombreuses applications nous avons pu apprécier un niveau de précision acceptable pour notre usage (avec des erreurs de re-projection de l'ordre de 4 ou 5 pixels) sur des photographies issues d'appareils photo numériques conventionnels (voir figure 59). Elle est également une solution particulièrement flexible

car elle ne nécessite pas d'informations sur l'appareil photographique utilisé pour la prise de vue (c'est bien le cas des photographies historiques par exemple). Malgré cela, elle demeure une technique manuelle importante du côté de l'utilisateur (saisie d'au moins 7 points dans la scène 3D et dans l'image 2D).

6.3.3. Récupération directe des paramètres de photographies orientées

Afin d'intégrer des orientations de photographies provenant des solutions de photogrammétrie, notre système permet de rentrer directement les valeurs relatifs aux paramètres intrinsèques et extrinsèques des cameras au travers d'un formulaire. Ceci permet de bénéficier d'un degré de cohérence géométrique maximale, notamment dans le cas où la morphologie de l'édifice est directement reconstruite à partir d'images. C'est le cas des représentation 3D élaborées par la technique de la photo-modélisation [De Luca L., 2009] qui se sert de points extraits des photographies pour conduire la reconstruction géométrique des éléments architecturaux en utilisant des fonctions de modélisation interactive.

6.4. Annotation sémantique des images par projection

Comme décrit dans la présentation générale de notre approche de spatialisation et de sémantisation de sources iconographiques (chapitre 4), nous utilisons la relation spatiale entre la représentation 3D et les images 2D. L'utilisation de l'information 3D comme solution pour créer une relation entre une description sémantique et des images corrélées est originale dans le domaine de l'annotation sémantique d'images. Une fonction dans Photosynth [Snavely N. et al., 2006] utilise un concept similaire pour explorer une relation bidirectionnelle entre l'image et la sémantique en utilisant un nuage de points (extrait directement des images) comme véhicule de distribution/propagation. Cette fonction permet de transférer de façon automatique la sémantique d'une image à une autre que contient la(es) même(s) région(s) de la scène.

L'utilisateur de ce système peut sélectionner manuellement la région d'une image et entrer une annotation textuelle. L'annotation est alors enregistrée et reliée à la portion de nuage de points sélectionnée et apparaît comme une enveloppe (rectangulaire dans l'image). Une fois l'annotation créée elle est automatiquement transférée sur toutes les autres images concernées. Pour déterminer si une annotation est appropriée pour une camera donnée, le système teste sa visibilité dans l'image et adapte l'enveloppe. Au-delà de l'enrichissement sémantique des images, l'habilité de transférer les annotations de ce système offre une solution très portable pour le développement de systèmes de réalité augmentée basé sur la vision par ordinateur [Feiner S. et al. 1997].

Comme nous l'avons déjà dit, dans notre travail, nous nous concentrons sur la relations entre une représentation 3D détaillée (et non pas un nuage de points approximatif de la scène) et les images 2D. Cette relation est réalisée par la projection des contours des objets géométriques (en utilisant spécifiquement une représentation polygonale) sur les images 2D référencées spatialement dans le but de générer de

façon automatique l'annotation sémantique de ces images. Par rapport à la méthode développée au sein de Photosynth, la différence la plus importante consiste en le fait que nous utilisons un vrai modèle géométrique 3D comme distributeur / propagateur, à la place d'un nuage de points. Ceci nous permet de gérer une complexité, une précision et une lisibilité majeure des régions projetées.

En utilisation les résultats des méthodes de référencement spatial décrites dans les paragraphes précédents, le point de vue d'une camera virtuelle relative à chaque image correspond au point de vue dans le monde réel. De cette manière les contours des objets de la scène sont projetés sous forme d'une couche vectorielle additionnelle s'associant à l'image originelle. Dans l'image vectorielle projetée, chaque segment de polygone correspond parfaitement à la forme des objets 3D de la scène en fonction au point de vue de l'image et associe aussi tout les attributs sémantiques qui qualifient les éléments 3D (voir section 5.3.2). Ce processus est complètement automatique.

6.4.1. Présentation de l'environnement du développement

Pour réaliser la projection des couches sémantiques du modèle 3D aux images 2D spatialisées nous avons développée un processus en s'appuyant sur les langages PHP/MySQL (voir section 4.3.2.2), MEL et SVG.

Langage MEL (Maya Embedded Language). : est un langage de script de Maya, un logiciel de modélisation 3D. Le langage MEL permet d'avoir accès à toutes les fonctions du logiciel concernant la modélisation, le rendu, l'animation ainsi que la création d'interfaces customisées. Maya a une architecture basée sur les commandes. Il est possible de rentrer explicitement des commandes dans les langages de script. Chaque commande a plusieurs arguments (flags) qui établissent la façon dont la commande est exécutée. Les arguments et leurs valeurs implicites ou explicites sont requis pour exécuter une commande.

SVG est un format de données conçu pour décrire des représentations graphiques vectorielles basé sur XML. Ce format est spécifié par le World Wide Web Consortium. Les coordonnées, les dimensions et les structures des objets vectoriels sont indiquées sous forme numérique dans le document XML. Un système spécifique de style (CSS ou XSL) permet d'indiquer les couleurs et les polices de caractères à utiliser. Ce format gère les formes géométriques de base, les parcours (qui utilisent les courbes de Bézier), le remplissage (qui peuvent être créés à l'aide de gradients, des patterns ou de filtres), les markers (des motifs distribués le long des chemins) et le canal alpha pour la transparence. Comme dans tout document XML, les objets sont organisés sous forme d'arbre et de balises. Le format SVG, accessible depuis JavaScript, permet l'intégration d'animations et la manipulation des objets graphiques grâce à des scripts, à des événements et par l'intermédiaire du modèle Document Object Model (DOM). SVG peut être visualisé nativement dans la plupart des navigateurs web ou via un plug-in.

6.4.2. Développement de la procédure de sémantisation d'images

Du point de vue du développement, le système de projection sémantique est basé sur la communication entre la base de données MySQL et Maya. Cette communication permet de collecter les informations nécessaires depuis la base de données et générer les modèles géométriques des cameras associées aux sources iconographiques spatialisées dans Maya (dans lequel la représentation polygonale de la scène 3D est aussi chargée). Le moteur de rendu de Maya est ensuite utilisé pour calculer un rendu vectoriel de la scène de chaque image 2D (à partir de chaque point de vue). La dernière phase consiste à associer chaque région projetée (polygone en SVG) aux attributs sémantiques stockés dans la base de données.

Ce processus est donc composé de quatre étapes : collecte des données depuis la base de données, reconstruction des points de vue image, projection des couches sémantiques, ajout d'attributs supplémentaires et de comportements interactifs dans les images sémantisées. Le schéma présenté dans la figure 60 explique ce processus.

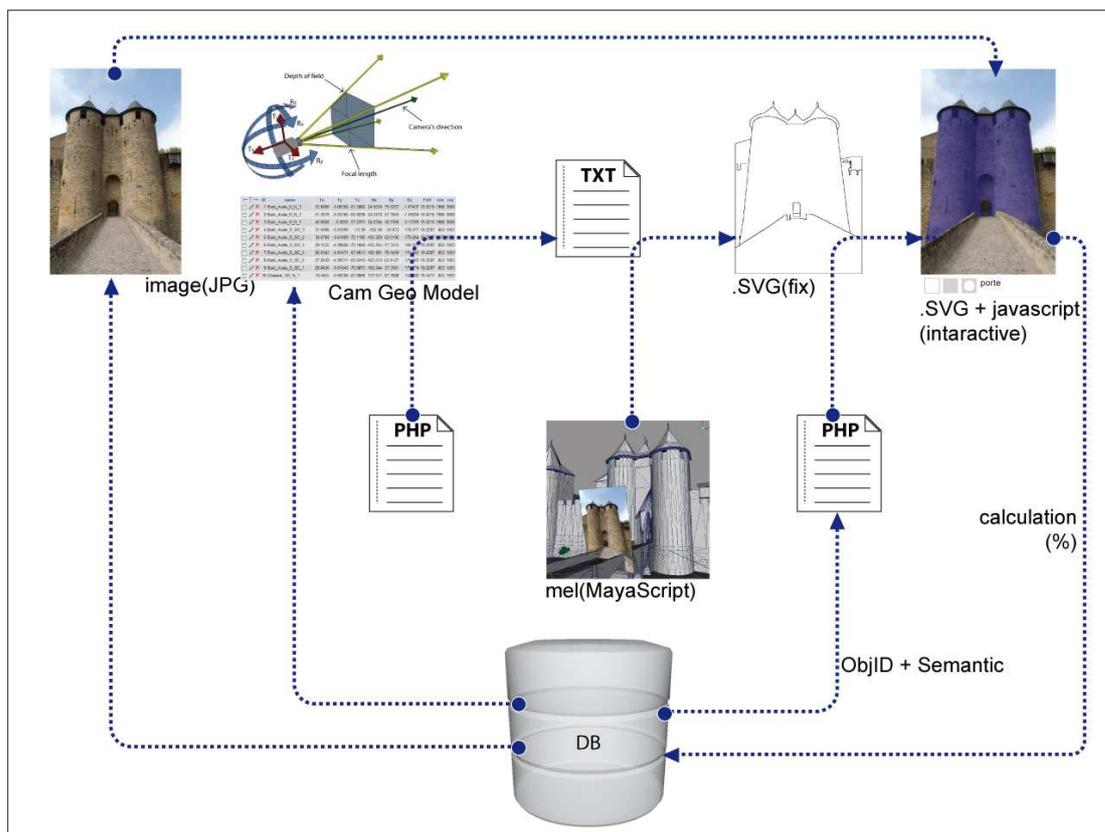


Figure 60 Schéma de la sémantisation d'images.

6.4.2.1. Collecte de données depuis la base de données

Trois types d'informations sont collectées pour ce processus provenant des phases antérieures de notre approche (annotation sémantique des formes 3D, sources iconographiques et leur point de vue) :

- l'information 2D des images originales (largeur et hauteur) à partir de la table des sources iconographiques ;
- l'information spatiale relative au modèle géométrique des caméras à partir de la table des référencements spatiaux;
- les identifiants des éléments 3D et leurs attributs sémantiques à partir de la table de la description sémantique de la morphologie de l'édifice.

Ces informations sont nécessaires pour recréer les points de vue des caméras dans Maya. Nous devons les transférer depuis la base de données MySQL dans une forme lisible et représentable par Maya. Les deux premières catégories de données sont exportées depuis la base de données en utilisant un script PHP et réécrite sous forme d'un tableau textuel configuré en fonction d'un parseur MEL.

6.4.2.2. Reconstruction des points de vues des sources iconographiques

Un script MEL lie le tableau généré et utilise les valeurs pour créer et paramétrer des caméras virtuelles dans la scène 3D afin de reproduire la position et l'orientation (ainsi que les paramètres intrinsèques) des images spatialisées.

Les valeurs nécessaires pour la reconstruction des points de vue des images sont :

- Le vecteur de position et d'orientation, pour déterminer le point de vue dans l'espace 3D ;
- Une valeur décimale pour la distance focale, pour déterminer l'angle de champ de la caméra ;
- Les proportions de l'image originale, pour déterminer les dimensions du plan d'image.

Chaque caméra est créée automatiquement, et nommée par l'identifiant de l'image originale, ensuite ses paramètres sont ajustés aux valeurs lus dans les tableau.

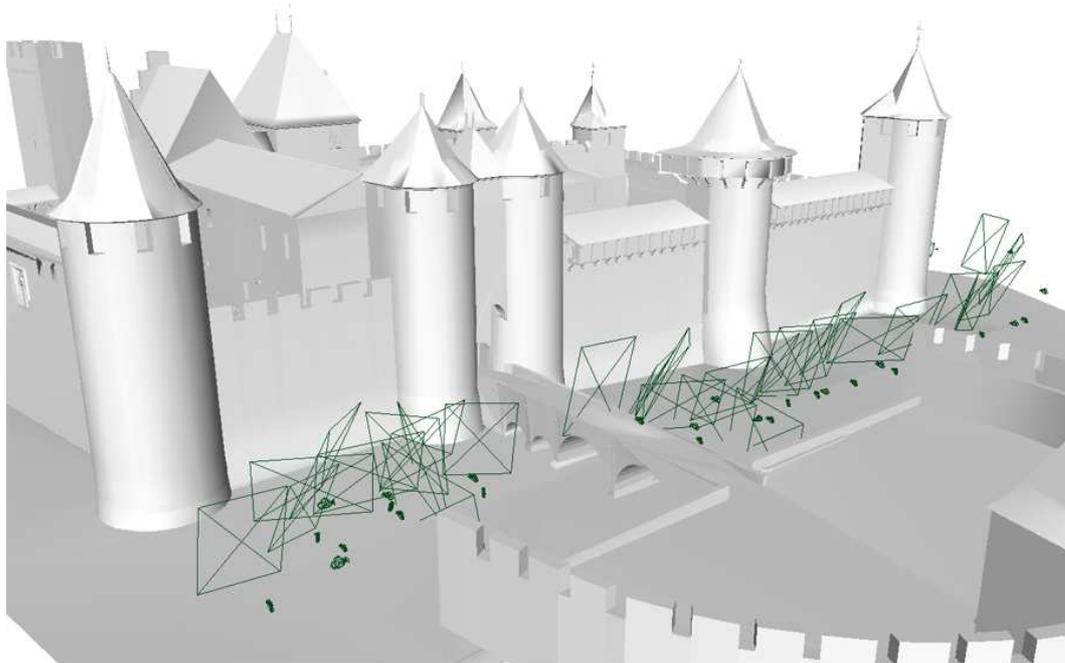


Figure 61 Reconstruction des points de vues de images dans une scène 3D (Maya).

Une fois les caméras virtuelles créées dans la scène 3D, Maya télécharge automatiquement les éléments 3D de la scène à partir de la liste des représentations polygonales présentes dans la base de données (voir figure 61). Cette liste de représentation fait référence au modèle de description sémantique de la morphologie de l'édifice et elle véhicule les identifiants des entités morphologiques reliées aux annotations sémantiques. Une fois ce téléchargement terminé, la scène 3D est prête au calcul de rendu vectoriel.

6.4.2.3. Rendu vectoriel des couches sémantiques

Afin de créer une image vectorielle de la scène 3D, nous utilisons une librairie de fonctions de rendu appelée MayaVector. La procédure MEL développée lance un rendu automatique de la scène 3D à partir de chaque caméra virtuelle reconstruite (relative aux points de vue des images spatialisées). Cette librairie nous permet de projeter facilement la silhouette des éléments 3D sur l'image entière dans la base de données. Cela produit une image vectorielle en format SVG ayant un point de vue correspondant à celui de l'image originale. La procédure de rendu est programmée de façon à calculer seulement les arêtes de contours des objets dans la scène (en incluant les arêtes d'intersection entre les différents objets) et d'ignorer les autres arêtes polygonales (voir figure 62).

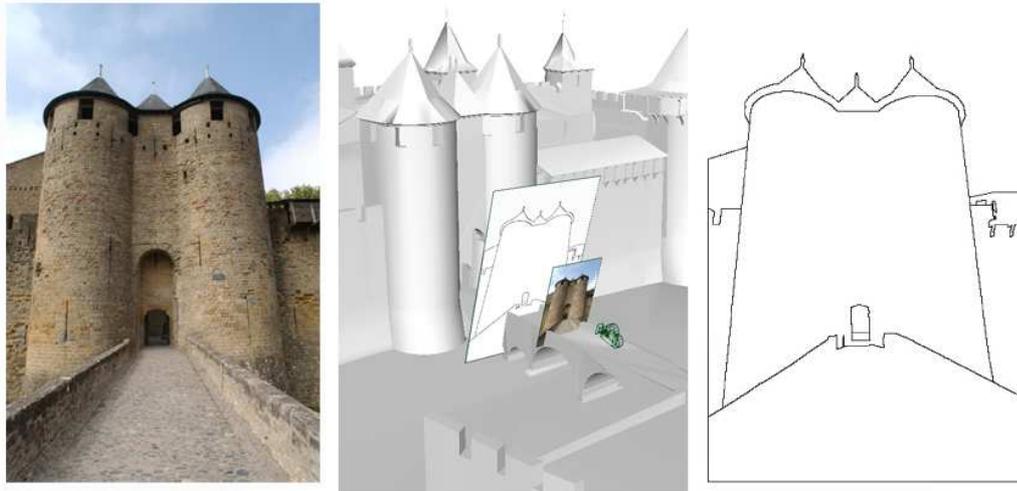


Figure 62 Projection des contours des éléments 3D sur l'image. A gauche, l'image originale, à droite l'image vectorielle SVG.

Au moment du calcul du rendu vectoriel nous associons l'identifiant de l'élément 3D (entité morphologique de la description sémantique) à la région polygonale 2D correspondante dans l'image vectorielle. L'image SVG est ainsi enregistrée dans la base de données (à côté de l'originale) et prête à être enrichie par des attributs supplémentaires.

6.4.2.4. Ajout d'attributs sémantiques et de comportements interactifs

La procédure pour l'ajout d'attributs sémantiques est réalisée par un script PHP. Ce script utilise des commandes pour déstructurer, restructurer et éditer l'image SCG en fonction des informations stockées dans la base de données et relatives à la description sémantique de la morphologie de l'édifice. Chaque région polygonale de l'image est associée à une annotation sémantique au travers de l'identifiant du polygone SVG provenant de la description de la morphologie de l'édifice et véhiculé de la scène 3D à l'image 2D par la procédure que l'on vient de décrire précédemment.

Le dernier processus se concentre sur l'ajout d'un comportement interactif de l'image en utilisant javascript. Cette interactivité est utilisée dans notre système de recherche visuelle (voir détails dans la section 7.1.3). Le résultat de ce processus est une image interactive (sensible au mouvement du curseur) affichant le nom et la sémantique associée aux régions de l'image. Cette image possède l'image d'original comme un fond. Quand l'utilisateur passe sur un objet avec le curseur, la région polygonale relative est colorisée et un texte dynamique affiche l'annotation sémantique correspondante (voir figure 63).

Le niveau de précision de la segmentation et de l'annotation sémantique des images dépend des résultats du référencement spatial.

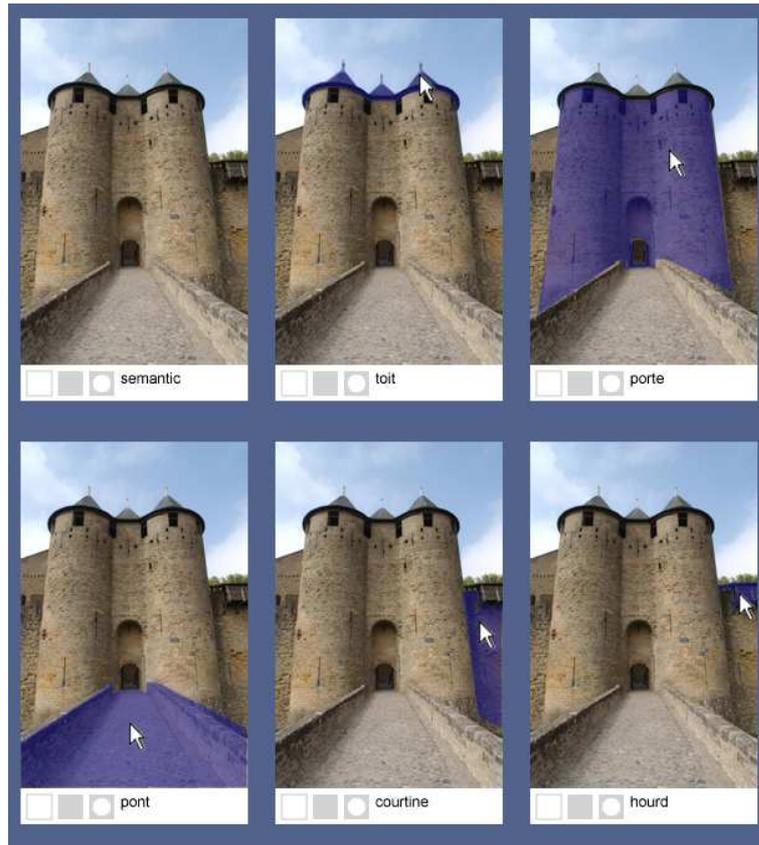


Figure 63 Image sémantisée SVG associant des comportements interactifs au passage de la souris

Une valeur relative à la surface de chaque région de l'image est calculée en fonction de la position des sommets du polygone. Cette information est utilisée dans les procédures de recherche (voir section 7.3.1), par exemple en utilisant un pourcentage de surface comme critère pour rechercher des images contenant un élément spécifique ou encore pour trier les résultats d'une recherche.

La phase finale du processus de sémantisation des images est la mise à jour de la base de données. Un script PHP collecte l'information ajoutée à l'image SVG (l'identifiant des régions présentes dans l'image, la surface de chaque élément sur l'image) et les enregistre dans la base de données.

Ces images interactives SVG sont des représentations hybrides combinant trois types d'informations : l'image 2D, les éléments 3D représentés à son intérieure et leurs attributs sémantiques. Ces représentations permettent de voir les sources iconographiques d'une manière inédite. La visualisation de ces informations peut en effet assister les spécialistes de la documentation architecturale pour étudier les sources iconographiques, et peuvent aussi permettre au grand public de mieux comprendre l'architecture patrimoniale.

Le diagramme dans la figure 64 représente les dialogues entre les parties du processus de téléchargement, le référencement spatial et la projection de couche sémantique.

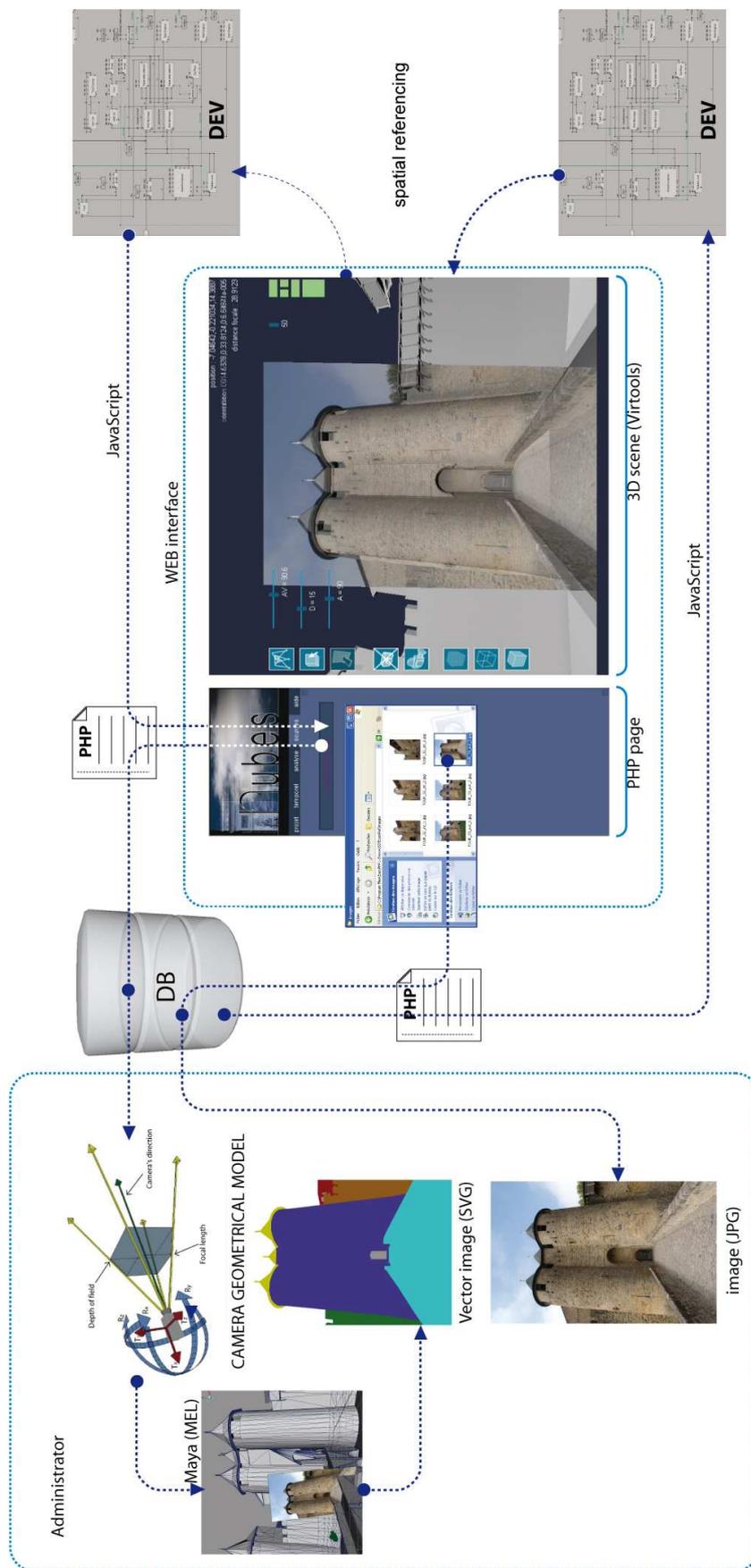


Figure 64 Dialogue entre les parties du système pour les processus de téléchargement, de référencement spatial et de projection des couches sémantiques