

Méthodes d'annotation sémantique d'images

Dans le chapitre précédent, nous nous sommes intéressés à une classification, basées sur différents critères, des différents types d'images que l'on peut utiliser pour documenter un édifice patrimonial. Dans ce chapitre, nous nous intéressons aux méthodes et aux solutions existantes pour la sémantisation d'images.

L'étiquetage du contenu sémantique des images au travers d'un ensemble de mots clés est une problématique connue sous le nom d'« annotation d'image ». L'annotation est principalement utilisée pour la gestion de base de données d'images, notamment afin d'améliorer les possibilités de la recherche. Des images annotées peuvent habituellement être retrouvées en utilisant une recherche par mots-clés, alors que des images non-annotées peuvent être extrêmement difficiles à trouver dans une base de données. L'annotation d'images par mots clés permet également d'améliorer les processus d'indexation d'informations utilisés par les moteurs de recherche Web [Rodden K.,1999].

Aujourd'hui, Il y a approximativement trois manières pour enrichir d'un point de vue sémantique des images [Wenyin L. et al., 2001] :

- l'annotation manuelle ;
- annotation semi-automatique ;
- annotation automatique.

L'annotation sémantique n'est pas encore utilisé dans le domaine spécifique de l'architecture patrimoniale, par conséquent, nous présentons ici les travaux de recherche et les développements relatifs à un domaine d'application beaucoup plus large.

3.1. Méthodes manuelles

La plupart des systèmes de base de données d'images utilisent aujourd'hui l'annotation manuelle [Gong Y. et al., 1994]. Il s'agit d'une méthode d'annotation qui permette aux utilisateurs d'associer des mots-clés descriptifs aux images conservées au sein d'une base de données. Deux types d'annotation différentes existent : celle basée sur les mots-clés et celle basée sur les ontologies. Dans le premier cas les utilisateurs sont libres d'étiqueter les ressources considérées avec n'importe quel mot-clé qu'ils retiennent, adéquat à la description du contenu de l'image ; dans le deuxième cas les

mots à utiliser pour la description du contenu des images sont reliés à une structure de concept représentant un domaine de connaissances précis.

Dans le cas de l'annotation libre par mots-clés les utilisateurs ne sont pas forcés de suivre un arrangement formalisé. Puisque aucune conceptualisation partagée n'est prise en considération, l'association de l'étiquette à une interprétation sémantique précise peut seulement être accidentellement réalisée. Les exemples bien connus de ce genre d'annotation pour les images 2D sont [Flickr] et [Riya]. Ces systèmes se basent sur des interfaces permettant aux utilisateurs de tracer un rectangle autour d'un objet dans une photographie puis d'introduire quelques termes descriptifs.

Différents travaux de recherche ont exploré des techniques pour améliorer le processus d'annotation sémantique manuelle. [Shneiderman B. et al., 2000] ont développé une méthode d'étiquetage des noms des personnes dans un ensemble de photos dans le cadre du projet «*PhotoFinder*» (voir figure 19). Par cette méthode, l'utilisateur peut simplement choisir manuellement un mot clé à partir d'une liste de noms et se servir d'une manipulation de type « Glisser - déposer » (*drag & drop*) sur les images afin d'en indiquer la personne correspondante. Bien qu'elle évite une grosse partie de travail de dactylographie, cette méthode reste toujours manuelle et requiert beaucoup de manipulations.

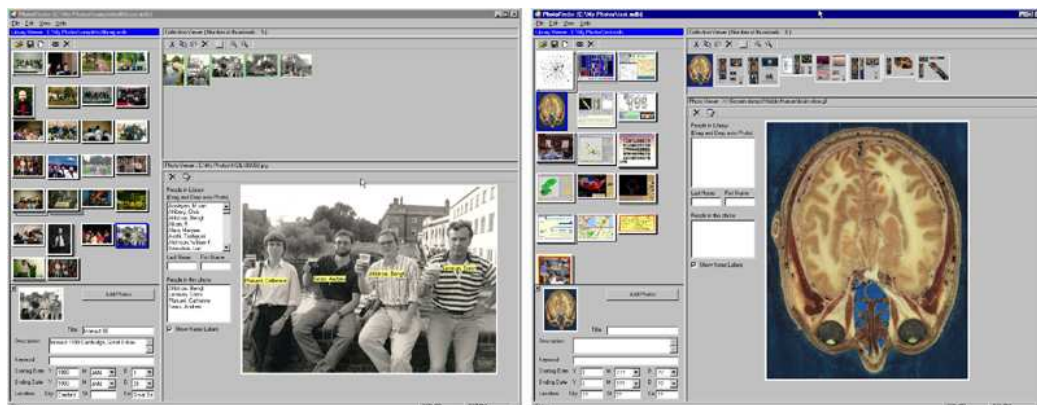


Figure 19 L'interface de PhotoFinder

Une technique permettant d'accélérer cette tâche lourde est l'annotation par lots : un group d'images sont sélectionnées, puis le même mot-clé est appliqué à chaque photo avec une seule action [Kang H. et al., 2000; Kuchinsky A. et al., 2002].

Du côté de l'annotation basée sur l'ontologie, les étiquettes sont définies par des spécifications formelles et explicites d'une conceptualisation partagée d'un domaine de connaissances [Gruber T.R. 1993], [Guarino N., 1998].

Puisque la conceptualisation est partagée, il n'y a aucune liberté dans le choix des noms d'étiquettes, mais cette contrainte est par contre récompensée par un arrangement commun des étiquettes. D'ailleurs, la conceptualisation partagée peut

également être traitée par des applications informatiques, fournissant des possibilités particulièrement probantes en terme de recherche de contenu [Saathoff C., 2006].

Photostuff [Halaschek-Wiener C. et al., 2005] donne aux utilisateurs la possibilité d'annoter des régions d'images en utilisant une ontologie ainsi que d'éditer les métadonnées de l'image dans le Web. Dans M-OntoMat-Annotizer [Petridis K. et al., 2006], l'utilisateur peut dessiner et accentuer des segments (i.e. régions) d'une image, visualiser (browse) des ontologies spécifiques à différents domaines, puis annoter les régions identifiées (voir figure 20).

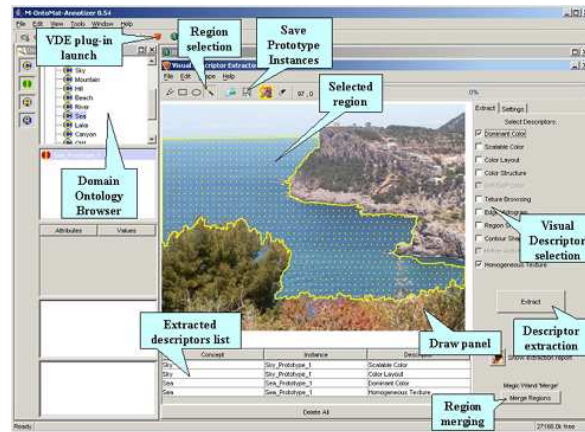


Figure 20 L'interface de M-OntoMat-Annotizer

Dans le domaine de l'architecture archéologique [Blaise J.Y. et al., 2007] et [Azzi Y., 2008] proposent un système qui permet à l'utilisateur d'identifier des régions des images en traçant la silhouette des objets à décrire puis en associant à chaque région créée un terme du vocabulaire architectural faisant partie d'une base de données ontologique (voir figure 21).

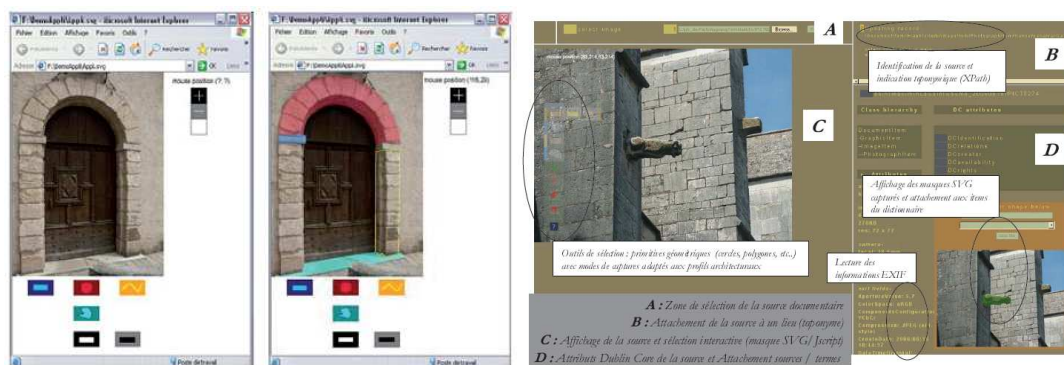


Figure 21 Annotation sémantique d'images d'architecture patrimoniale
[Azzi Y., 2008; Blaise J.Y. et al., 2007]

[Cross J. et al.] proposent un dispositif de collecte et d'annotation d'images pour les applications de terrain dans les chantiers de fouilles archéologiques. Les images sont recueillies par webcam et géo-rectifiées. L'utilisateur peut annoter les images en utilisant un système de couches d'annotations. Plusieurs objets peuvent

être attachés à une annotation, et les annotations peut être présentées individuellement (en une seule couche), ou comme un composite.

Analyse : Bien que l'annotation manuelle du contenu d'image soit considérée un meilleur cas en termes d'exactitude du contenu sémantique des images, puisque des mots-clés choisis sont basé sur la détermination humaine, c'est un processus qui a besoin de beaucoup de travail. En outre, l'annotation manuelle peut également présenter des erreurs de récupération à cause de oubli d'utilisateurs ou descripteurs après une longue période de travail.

Les critères pour saisir un mot-clé au travers de cette méthode sont complètement subjectifs. Les termes ne sont pas référencés par le même standard. Même si certains systèmes sont basés sur des mots-clés structurés dans une ontologie, le nombre et l'association des termes est toujours dépendante d'une décision personnelle. Une personne peut entrer un seul mot-clé par image, alors que une autre personne pourrait en utiliser plusieurs. Certaines annotations jugées inutiles pour certains pourraient être importantes pour d'autres.

En outre, dans les solutions basées sur cette méthode il est extrêmement difficile de conduire de mise à jour de la structure de la base de données. Car, comme les mots-clés sont entrés manuellement, lorsqu'on modifie un mot-clef ou l'organisation des mots-clés, la seule manière de mettre à jour la base de données est de modifier les termes associés image par image.

Aujourd'hui, de nombreuses expériences du domaine de la documentation du patrimoine ont adoptée cette méthode de sémantisation des images grâce à sa simplicité d'utilisation. Si on prend en considération cette méthode dans une utilisation basique, elle peut aujourd'hui répondre aux besoins exprimés par les communautés scientifiques et professionnelles concernées. Mais si l'on s'intéresse à l'apport potentiel de ces technologies dans le domaine de la documentation patrimoniale, il est certain qu'il y aura à terme un besoin plus important de manipuler des collections d'images beaucoup plus importantes et un besoin complémentaire d'augmenter la complexité des descriptions (plusieurs niveaux de détails). Au vu des limites actuelles de ces méthodes manuelles, il est d'hors et déjà évident que ces méthodes ne pourrons pas répondre aux besoins potentiels des utilisateurs de cette communauté.

3.2. Méthodes automatiques

Les méthodes automatiques d'annotation sémantique d'images se basent sur des techniques d'analyse d'image basées sur le contenu (*content-based image analysis techniques*) [Flickner M. et al., 1995]. Normalement, ces processus sont composés de deux tâches, la segmentation d'image (détection automatique des formes d'éléments dans l'image) et l'identification d'objets (la comparaison entre les formes détectées et des objets cohérents dans une bibliothèque).

L'identification des classes d'objets peut être réalisée en utilisant une combinaison des modèles particuliers [Winn J. et al., 2005 ; Fergus R. et al., 2003] (voir figure 22). Plusieurs chercheurs ont considéré ces deux tâches séparément. Par exemple [Kumar S. et al., 2004] se sont concentrés seulement sur le problème de la segmentation.

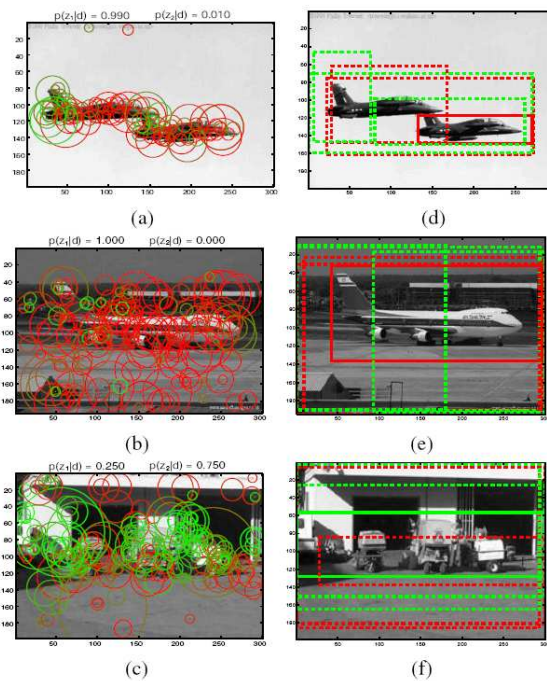


Figure 22 Les résultats d'une détection sémantique proposée par le modèle TSI-pLSA de [Fergus R. et al., 2005]

L'outil TextonBoost, proposé par [Shotton J. et al., 2006] combine la reconnaissance de formes et la segmentation d'images. Ils utilisent une combinaison de caractéristiques spécifiques appelé *Texton* pour encoder la forme, la texture et l'apparence de différentes classes d'objets. Bien que leur méthode a produit des résultats significatifs, les formes bruitées et les textures causent son échec sur les frontières des objets (voir figure 23).

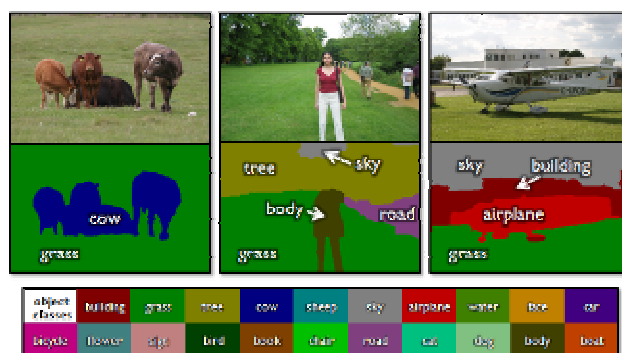


Figure 23 Les résultats de la détection sémantique de TextonBoost [Shotton J. et al., 2006]

Dans le domaine de l'architecture patrimoniale, il n'y a pas beaucoup de recherches s'appuyant sur l'analyse automatique de la sémantique d'image. Par contre, dans un domaine proche, certains chercheurs ont développé des méthodes et des outils pour analyser l'architecture urbaine. [Johansson B. et al., 2002a] proposent une méthode pour détecter un élément architectural simple dans la scène d'une ville, comme par exemple la détection automatique des fenêtres des bâtiments.

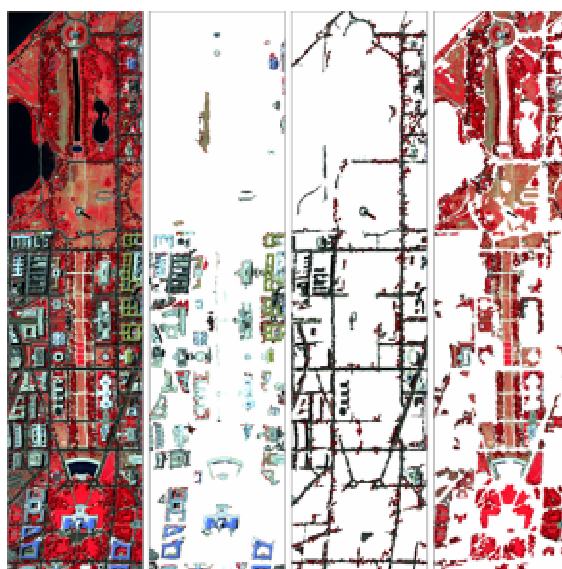


Figure 24 Un exemple de détection sémantique à partir d'images aériennes [Akçay H.G. et al., 2007]

À l'échelle urbaine, la sémantisation automatique d'images aériennes peut produire beaucoup plus de résultats utilisables. Des techniques spécifiquement développées pour la détection de bâtiments peuvent être trouvées dans la littérature de [Mayer H., 1999]. La détection automatique d'objets géographiques comme les structures urbaines se fondent sur la position, l'échelle et l'orientation d'éléments urbains dans l'image [Akçay H.G. et al., 2007]. (voir figure 24).

La plupart des recherches sur l'analyse de l'image urbaine se concentrent sur la classification basée sur pixel. Les textures peuvent également être utilisées comme méthode alternative [Karathanassi V. et al., 2000 ; Bhagavathy S. et al., 2007]. Des éléments architecturaux ont également été utilisés pour classer les régions d'images

de zones rurales, résidentielles ou urbaines [Unsalan C. et al., 2005 ; Aksoy S., 2006] (voir figure 25).

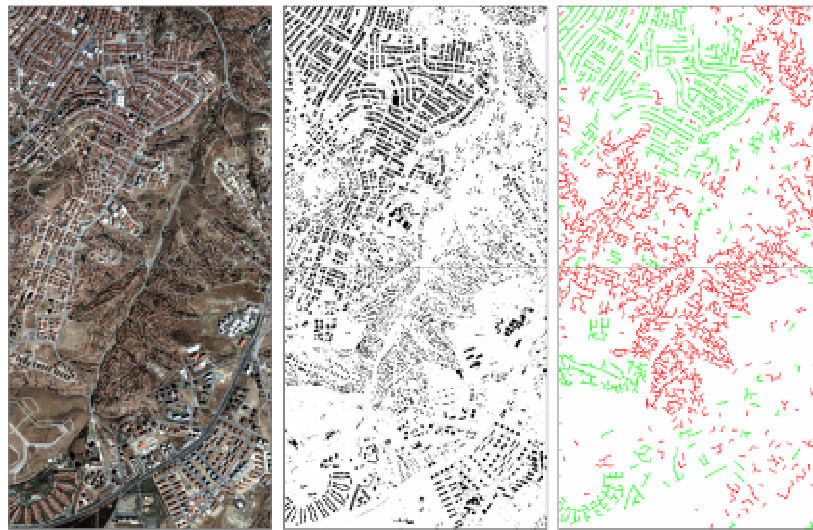


Figure 25 Les résultats de la détection sémantique d'images aériennes à l'échelle urbaine utilisant des patterns de bâtiment [Aksoy S., 2006]

Analyse La sémantisation automatique par analyse d'image est en général très rapide et nécessite de moins de travail humain (par rapport à l'annotation manuelle). Malgré cela, les solutions actuelles, ne sont pas encore suffisamment précises et fiables. Au-delà du problème de la précision, les limites de cette méthode sont que le champ d'application possible est très restrictif. En effet, les deux étapes d'analyse d'image (segmentation et identification d'objets) doivent avoir les données de base appropriées.

En ce qui concerne la segmentation d'image, le problème de la détection des formes est particulièrement difficile à traiter en particulier dans le cas d'images qui présentent beaucoup d'éléments hétérogènes à leur intérieur. D'autres problèmes sont dus à la présence de plusieurs objets similaires ou quand la couleur et le contraste de fond est proche de celui du sujet principal de l'image.

En ce qui concerne l'identification d'objets, les solutions actuelles nécessitent d'une très grande base de données afin de permettre l'identification de plusieurs typologies d'objets. Si la base de données ne contient suffisamment de modèles pouvant être superposés aux formes détectées dans l'image, ces systèmes ne peuvent pas reconnaître leur sémantique. Et même dans le cas où le modèle de l'objet détecté dans l'image est présent dans la base de données, il est aussi nécessaire que le point de vue (et donc la projection) de cet objet coïncide avec le ou les points de vue associés au modèle. C'est la raison pour laquelle l'utilisation de cette méthode dans les images aériennes est beaucoup plus fiable, car elle bénéficie d'un point de vue zénithal.

En revenant à notre domaine d'intérêt, l'architecture patrimoniale, il est sûr que les limitations actuelles des méthodes automatiques rendent ces méthodes particulièrement inadaptées aux besoins des scientifiques et des professionnels de ce

domaine. En effet, en termes de complexité d'image, les photographies d'édifices architecturaux sont souvent assez complexes d'un point de vue de la quantité d'éléments présents dans la scène (souvent de différentes tailles) et rarement bien identifiables. En ce qui concerne la constitution de bases de données d'images correspondantes aux formes architecturales, la richesse et la diversité du patrimoine architecturale empêchent d'imaginer d'identifier des classes d'éléments pouvant être superposés correctement aux éléments photographiés. Enfin, en ce qui concerne la variété de point de vue (ou de points de prise photographique), les photos acquises pour des besoins d'étude patrimonial ou de documentation sont souvent assez hétérogènes en termes de paramétrage et de configuration de prise de vue.

3.3. Méthodes semi-automatiques

Les méthodes semi-automatiques d'annotation sémantique d'images se basent sur des techniques combinant les solutions présentées précédemment (automatiques et manuelles). Des algorithmes sont utilisés pour le filtrage de la sémantique appropriée en utilisant les informations textuelles liées à l'image ou par l'identification des contenus d'image (comme dans les méthodes complètement automatiques), puis l'utilisateur peut choisir manuellement les mots-clés qui se rapportent à l'objet dans l'image. En employant le contrôle des utilisateurs, la sémantique obtenue est mise à jour de façon incrémentale. Intégrées à des interfaces appropriées, les méthodes d'annotation semi-automatique ont un grand potentiel en terme d'application à plusieurs domaines [Suh B. et al., 2004].

[Shen H.T. et al. 2000] ont employé une méthode dans laquelle les images sont utilisées pour indexer automatiquement d'autres images dans la même page Web. Ils emploient l'information textuelle riche de la page pour extraire des éléments de description potentielle des images. [Srihari R. et al., 2000] extraient du texte des entités textuelles (par exemple, les noms des gens, les noms des endroits et des choses, etc.) pour indexer automatiquement les images qui accompagnent le texte des pages web.

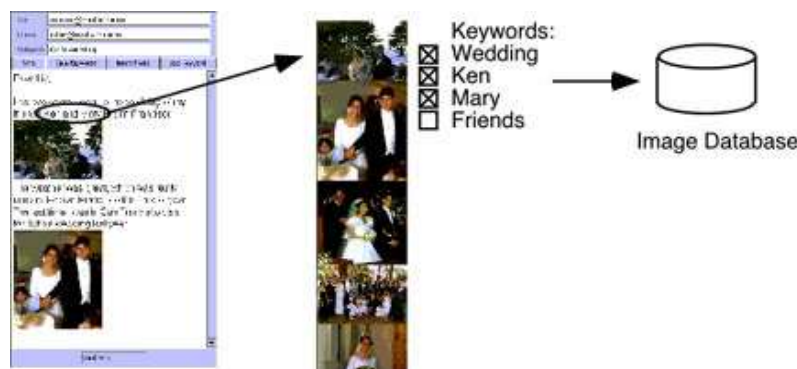


Figure 26 Le processus de la détection sémantique du ARIA

[Lieberman H., 2000] décrit le système ARIA (Annotation and Retrieval Integration Agent) qui intègre la récupération et l'utilisation d'image. Le système emploie l'extraction des mots présents dans les messages électroniques (dans lesquels des images sont incluses) pour indexer ces images (voir figure 26). Cette méthode semble fournir des résultats raisonnables, mais la précision du contexte textuel n'est pas aussi haute que l'indexation manuelle. [Wenyan L. et al., 2001] développent un système capable de générer et d'affiner progressivement des annotations sémantiques d'images de façon automatique en demandant à l'utilisateur la validation et l'organisation (ou la réorganisation) du résultat final (voir figure 27).

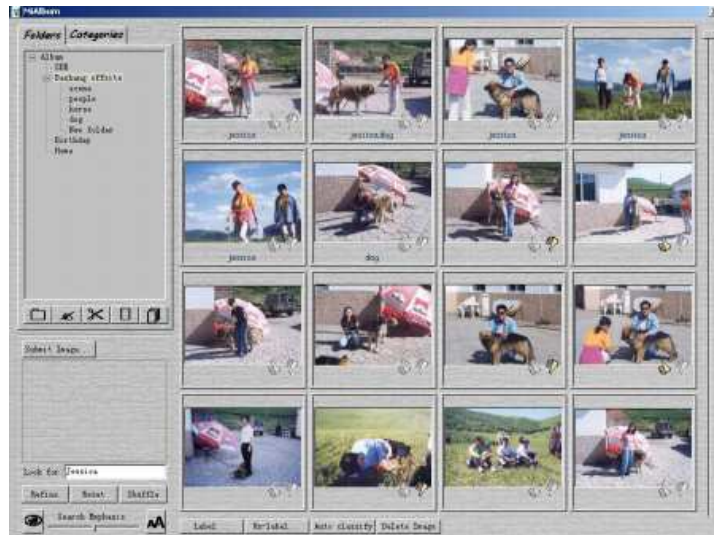


Figure 27 L'interface de *MiAlbum* pour la recherche basée sur texte, pertinence de résultat et d'annotation semi-automatique

[Ono et al. 1996] ont essayé d'employer des techniques d'identification d'objets pour choisir automatiquement des mots-clés descriptifs appropriés (dans un ensemble prédéfini) pour chaque image. Cependant, ils ont seulement examiné leur système avec des mots-clés et des images limités, ainsi la possibilité d'employer ce système dans un contexte plus général est peu claire. D'ailleurs, puisque les techniques d'identification d'image ne sont pas complètement fiables, les utilisateurs doivent en outre confirmer ou vérifier des mots-clés produits automatiquement par le système.

Analyse Il est certes qu'en combinant les techniques automatiques et manuelles, les méthodes semi-automatiques d'annotation sémantique d'images permettent d'obtenir des résultats meilleurs en termes de précision et de productivité effective.

Tout en restant probablement les méthodes les plus performantes, dans le domaine spécifique de la documentation architectural le problème de l'incohérence des formes persiste. Car les édifices historiques sont généralement très complexes et composés de nombreux éléments hétérogènes difficiles à classer de façon préalable dans le dessein de constituer une base de modèles à utiliser pour la reconnaissance.

4. Une approche pour la caractérisation sémantique 2D/3D d'édifices patrimoniaux

Suite à l'analyse des méthodes actuelles d'annotation sémantique d'images, on peut retenir que les solutions actuelles, tout en donnant des résultats applicables à certains contextes, ne sont certainement pas encore au point pour aborder le problème plus complexe de la segmentation et annotation précise d'artefacts patrimoniaux. En effet, même si les méthodes semi-automatiques semblent offrir des résultats convenables, le domaine de la documentation du patrimoine reste particulièrement difficile à traiter au vu de l'élévée complexité et hétérogénéité des formes architecturales.

Les vrais problèmes dérivent probablement du fait que cette complexité et cette hétérogénéité (typique du contexte de l'architecture patrimoniale) oblige à traiter les problèmes de classification des formes non pas à partir d'un simple aperçu bidimensionnelle de leur aspect (sur les images), mais à partir d'une véritable analyse de leur forme (sur des modèles 3D).

Pour donner un exemple, la documentation photographique des artefacts se base, le plus souvent, sur la prise de multiples photographies, à partir de différents points de vue, avec différents niveaux de zoom des mêmes formes. Or, si l'on imagine d'utiliser des solutions d'annotation sémantique d'images telles que l'on vient d'analyser dans le chapitre précédent, il est évident qu'il serait particulièrement problématique de garder (sauf dans le cas de l'utilisation d'une solutions complètement manuelle) sur toutes les images un niveau cohérent de précision à la fois en termes de segmentation et en termes d'annotation sémantique. De plus, si l'on s'intéresse à une annotation particulièrement fine d'édifices patrimoniaux (c'est le cas dans les travaux menés pour des finalités de conservation et/ou de restauration), il sera probablement nécessaire de conduire l'annotation sémantique non pas seulement au niveau d'une segmentation générale du bâti (colonne, porte, mur, etc..), mais au niveau des éléments architecturaux (base, fut, chapiteau, etc.) ainsi que des sous éléments (tore, baquette, scotie, etc.).

Ceci indique bien, que dans ce domaine, l'étude et la description des éléments, doit nécessairement s'appuyer sur un support de représentation plus sophistiqué, capable d'explicitier la complexité morphologique d'un édifice (relation entre ses parties et l'ensemble) ainsi que les caractères formels spécifiques aux éléments qui le composent (aspects spécifiques liés à l'apparence visuelle des surface).

Or, dans le domaine du relevé et de la documentation graphique du patrimoine architectural, des techniques et des outils (scanner laser 3D, photogrammétrie,

photomodélisation, etc.) permettent aujourd'hui de collecter, de structurer et de traiter des informations acquises (réseaux de points, nuages de points, photographies, etc.) afin de construire des représentation tridimensionnelles d'édifices patrimoniaux. Ceci correspond à créer un modèle 3D des artefacts étudiés contenant le niveau d'information (morphologie, apparence visuelle) nécessaire à une analyse complète.

Une hypothèse, introduite par [Blaise J.Y. et al., 2004] considère que le modèle 3D interactif peut être considéré comme un moyen d'accès privilégié à l'information patrimoniale. En s'inscrivant au sein de ce concept, nous avons alors la possibilité de traiter notre problème dans un contexte bien plus large : **on peut utiliser la représentation de la morphologie d'un édifice comme source principale pour distribuer / propager les attributs sémantiques sur l'ensemble des sources iconographiques (le représentant en deux dimensions) qui se trouvent dans une condition de cohérence géométrique/spatiale avec son modèle tridimensionnel.**

En lisant le problème dans cette nouvelle dimension, notre approche peut être ainsi décrite : au lieu de créer une relation directe entre chaque région de chaque image 2D et une annotation sémantique, nous pouvons utiliser un modèle de description sémantique (structurant l'ensemble de termes pouvant décrire la morphologie d'un édifice) comme dénominateur commun permettant d'établir des relations entre la représentation 3D complète des formes architecturales qui composent l'édifice et l'ensemble des sources iconographiques (collections d'images 2D segmentées) correspondantes (voir figure 28).

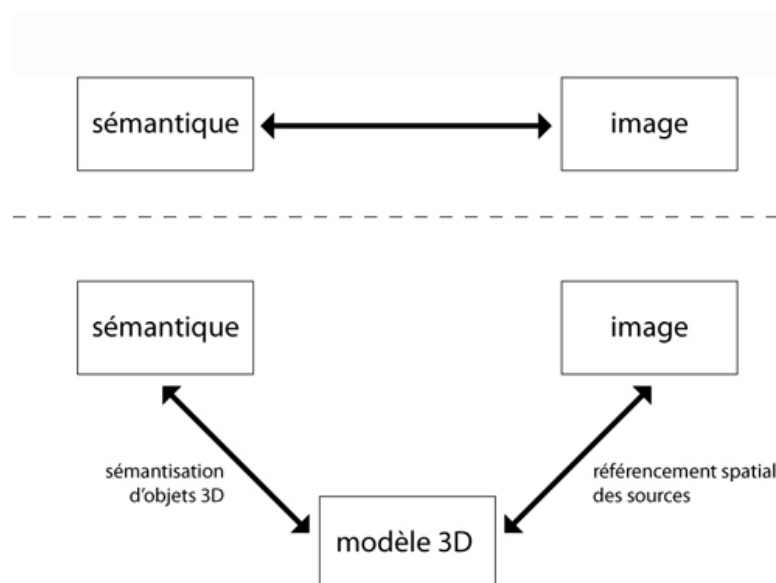


Figure 28 Différence entre les approches actuelles de sémantisation d'image (en haut) et l'approche proposée (bas)

Ceci correspond à mettre en place une méthodologie complète de caractérisation sémantique de représentations 2D et 3D d'édifices en s'appuyant sur la notion de référencement spatial : la détermination des relations projectives existantes entre la morphologie d'un artefact (modèle 3D) et l'ensemble des ses représentations

iconographiques (images 2D). La relation projective établie sera ensuite exploitée pour véhiculer des annotations sémantiques de la représentation 3D à l'ensemble de représentations 2D.

4.1. Approche proposé

Notre approche est composée de cinq étapes.

La première étape se concentre sur l'obtention d'une représentation tridimensionnelle complète de l'édifice à partir d'un relevé de terrain. Ce processus peut être divisé en sous étapes : l'acquisition des données à référence spatiale sur l'édifice réel et la construction de son modèle 3D à partir de ces données.

La deuxième étape consiste en la caractérisation sémantique du modèle 3D. Cette phase consiste à une décomposition de la représentation tridimensionnelle de l'édifice afin d'en isoler les éléments caractéristiques et de les qualifier par des attributs sémantiques. Dans cette phase on décrit la morphologie du bâti en fonction du vocabulaire de l'architecte (chapiteau, base, fenêtre, colonne, porte, etc.) en utilisant des thesaurus spécialisés.

La troisième étape se concentre sur l'établissement de la relation entre le modèle 3D et les images 2D, en s'appuyant sur le référencement spatial. Cette étape permet de retrouver le point de vue exacte (ou approximatif) de l'image dans l'espace 3D (voir figure 29). Différentes méthodes seront étudiées pour cette étape.

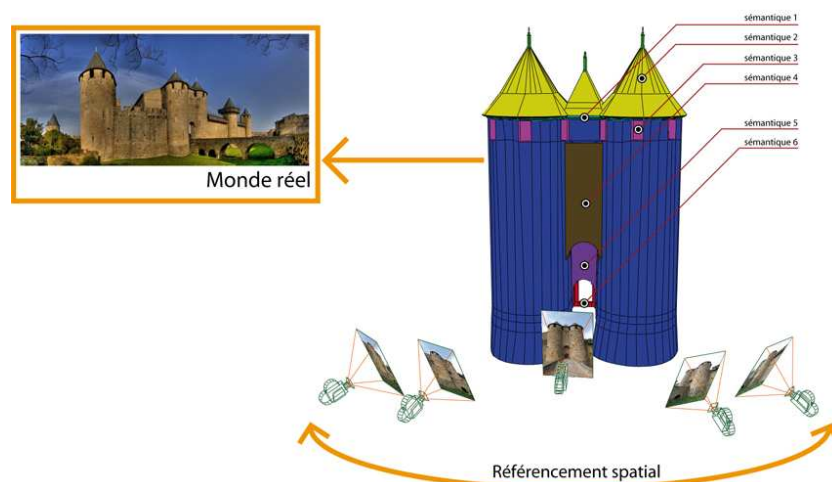


Figure 29 Les trois premières étapes de l'approches proposée : un modèle 3D est réalisé à partir d'une acquisition de l'édifice réel ; le modèle 3D est caractérisé du point de vue sémantique ; les images sont référencées spatialement par rapport au modèle 3D.

La quatrième étape consiste en la projection de la description sémantique du modèle 3D (décomposition) vers les images référencées spatialement (segmentation). Le résultat de cette étape se concentre sur en l'ajout d'une couche sémantique (projections des contours des formes 3D) sur les images (voir figure 30). Cette couche est réalisée par une représentation vectorielle (coïncidente avec l'image originelle) composée de polygones reliés à des attributs sémantiques.

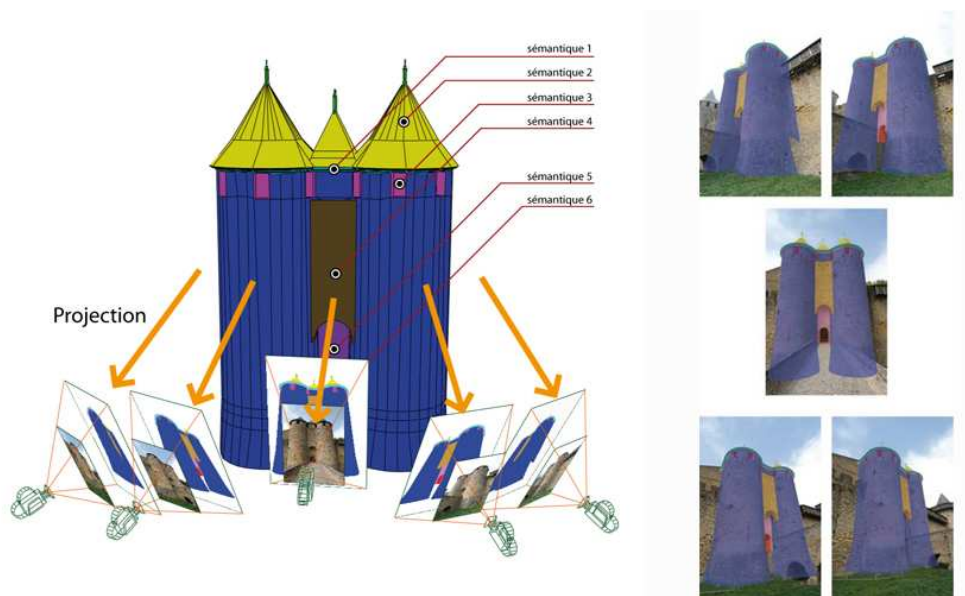


Figure 30 La quatrième étape de l'approche proposée : projection (du modèle 3D vers les images 2D) d'une couche sémantique.

Le grand avantage de notre approche consiste en la capacité à gérer des descriptions très complexes (c'est le cas de l'architecture patrimoniale) et surtout en fonction de multiples stratégies de description. L'annotation sémantique est en effet conduite directement sur la représentation tridimensionnelle, puis elle est systématiquement et automatiquement transférée dans les images référencées spatialement. La cohérence établie entre l'édifice réel (relevé 3D), sa morphologie (reconstruction 3D et structuration sémantique) et les images (référencement spatial sur le modèle 3D) permet la génération de contours particulièrement précis de formes (parfois très complexes) pouvant être projetées sur les images afin de les segmenter en fonction de la structuration du modèle 3D. De plus, la structure de description sémantique de chaque image serait toujours basée sur la même norme.

Un autre grand avantage de notre approche consiste en la facilité à modifier la structure de la description sémantique. Compte tenu du fait que chaque image est référencée spatialement par rapport au modèle 3D, à chaque changement au niveau de la décomposition de la morphologie de l'édifice ou au niveau de l'annotation sémantique des parties isolées, correspondra une mise à jour automatique de la segmentation et de l'annotation sémantique des images corrélées. Cette possibilité est particulièrement importante dans le domaine de la documentation architecturale, en effet, si d'un côté il y a souvent l'exigence de faire évoluer la description de l'édifice (soit en termes de granularité spatiale, soit en termes de choix des termes de vocabulaire), de l'autre côté plusieurs descriptions des mêmes formes (décompositions et descriptions par termes) peuvent être conduites sur le même édifice en fonction de différentes exigences d'analyse (conservation, restauration, aménagement, valorisation, etc.). Dans ce cas, il sera alors possible d'envisager des descriptions sémantiques multicouches (voir figure 31).

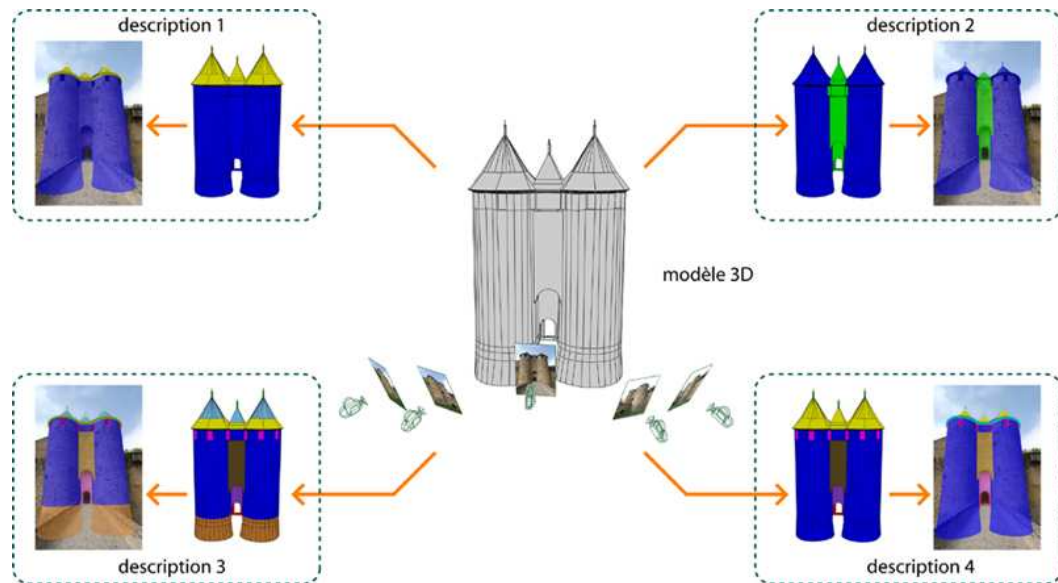


Figure 31 Un exemple de description sémantique multicouche : à chaque décomposition du modèle 3D correspond une projection de couche sémantique différente sur les images.

La dernière étape de notre approche, se concentrera sur la définition et le développement informatique d'un système d'informations capable de rechercher visuellement les résultats du processus entier ci-dessus décrit en fonction de critères spatiaux, morphologiques et sémantiques. Il s'agit d'une application Web utilisable par des spécialistes du domaine de la documentation architecturale comme par le grand public.

4.2. Problèmes à traiter

A partir de l'approche proposée qui contient cinq étapes de démarche, le travail effectué prend en compte également cinq préoccupations distinctes.

4.2.1. Acquisition de données 3D et reconstruction géométrique basés sur la réalité (chapitre 5)

Notre approche utilise une acquisition de données 3D comme base de départ. Ces données doivent nécessairement être produite à partir du relevé d'un édifice réel afin de garantir la cohérence spatiale (avec différents niveaux d'approximation) avec les sources iconographiques (photographies, dessin, peintures, etc..) le représentant. La reconstruction géométrique à partir de ces données doit elle aussi être le plus cohérent possible avec la réalité afin de garantir (une fois structurée d'un point de vue sémantique) une projection correcte des couches sémantiques sur les images. Plus le modèle 3D sera approximé, plus la segmentation des images corrélées sera imprécise et incohérente.

Pour l'étape de relevé, l'approche devra alors prendre en considération l'utilisation (individuelle et ou complémentaire) des moyens actuels d'acquisition de données 3D (balayage laser et photogrammétrie) afin de définir leur niveau d'adaptation à notre objective.

Même pour l'étape de reconstruction géométrique, il sera nécessaire d'étudier les moyens existants le plus appropriés à notre problème spécifique. L'objectif de cette étape consiste en effet à identifier les moyens permettant d'élaborer une représentation 3D la plus cohérente possible avec la réalité et en même temps permettant une structuration et une annotation sémantique adaptée aux besoins spécifiques du domaine. Un dernier aspect concerne également l'évaluation des possibilités effectives d'exploitation de cette représentation, que ce soit en termes de manipulation, stockage et/ou visualisation.

4.2.2. Caractérisation sémantique de la représentation 3D (chapitre 5)

La logique des décompositions sémantiques d'un édifice est toujours contrainte à la description de modèles théoriques. Ces approches sont en effet fondées sur la structuration de formes symboliques. L'objectif de cette partie de notre travail consiste à ajouter une couche sémantique à la description de la morphologie d'édifices existants.

Notre approche devra alors prendre en compte l'étude des méthodologies existantes de caractérisation sémantique des représentations 3D en fonction des véritables exigences de description que la morphologie d'un édifice requiert.

4.2.3. Référencement spatial de sources iconographiques (chapitre 6)

L'approche devra prendre en compte l'utilisation d'espaces 3D comme dénominateur commun pouvant établir des relations projectives entre la représentation 3D et les représentation 2D de l'édifice étudié. Ceci permettra d'associer à chaque image 2D un modèle géométrique exprimant la position et l'orientation de son point de vue dans le référentiel du modèle 3D.

A ce fin, nous avons besoin d'étudier les méthodes existantes afin de choisir la ou les plus indiquées en fonction de différents contextes d'utilisation (professionnel, amateur, etc.). Par ailleurs, il sera nécessaire de s'intéresser à la définition d'un modèle générique de représentation des points de vue des images en fonction de la diversité des sources iconographiques prises en considération (types de projection, niveaux de cohérence avec le modèle, etc.).

Enfin, on évaluera également le niveau d'approximation intrinsèque aux différentes techniques et/ou outils de référencement spatial d'images.

4.2.4. Indexation du contenu des sources iconographiques (chapitre 6)

L'approche devra définir en détail le processus de création de couches sémantiques à projeter sur les sources iconographiques à partir de la caractérisation sémantique du modèle 3D et en s'appuyant sur le référencement spatial. La projection de la silhouette bidimensionnelle de la structuration sémantique du modèle 3D se base en effet sur le point de vue (déterminé par la phase de référencement spatial) des sources iconographiques.

On pourra traiter ce problème de deux étapes. Dans la première étape on s'intéressera à la création d'une image vectorielle à partir de la décomposition de la représentation 3D et d'une projection des contours de chaque élément isolée. Dans ce contexte, nous étudierons une technique de rendu 2D permettant le transfert et la conservation d'attributs relevant de la structure de description morphologique de la représentation 3D.

Dans la deuxième étape, on se concentrera sur l'association entre la sémantique associée à la représentation 3D (reliée à des termes structurés dans des thésaurus spécialisés) et les régions polygonales projetés sur les images.

4.2.5. Recherche visuelle de sources iconographiques basée sur critères spatiaux, morphologiques et sémantiques (chapitre 7)

En ce qui concerne la conception et le développement d'un système d'informations capable d'intégrer, de gérer et d'exploiter l'enrichissement sémantique des sources iconographiques qu'il est possible de conduire au travers de l'approche que nous venons de décrire, nous nous intéresserons à l'identification des moyens qui permettent de profiter pleinement des apports produits par la sémantisation 2D/3D d'images.

Dans ce sens on se concentrera sur la création d'une interface permettant la recherche de sources iconographiques à partir de trois critères différents.

- Critère spatial : l'interface devra permettre de retrouver des sources iconographiques en fonction du point d'observation de l'utilisateur à l'intérieur d'une scène 3D.
- Critère morphologique : l'interface devra permettre de retrouver des sources iconographiques en fonction de la sélection d'un ou plusieurs éléments de la représentation 3D de l'édifice.
- Critère sémantique : l'interface devra permettre de retrouver des sources iconographiques contenant des éléments reliés à un ou plusieurs attributs sémantiques spécifiques.

L'interface devra également être d'utilisation facile afin de permettre un enrichissement effectif de la base de données iconographiques relative à l'édifice (par les spécialistes de la documentation comme par le grand public).

4.3. Développement informatique

Les apports méthodologiques de ce travail de thèse sont expérimentés et validés au sein de différents développements informatiques intégrés au projet NUBES [Nubes] mené au sein du laboratoire MAP-Gamsau.

4.3.1. Présentation du projet NUBES

Le projet NUBES concerne le développement d'un système d'informations à l'échelle architecturale qui prend en compte les relations que l'on peut établir entre la représentation de l'édifice (forme, dimensions, état de conservation, restitution hypothétique, etc.) et des informations hétérogènes concernant différents domaines (technique, documentaire, historique, etc.). La démarche proposée vise à organiser autour d'un modèle de description sémantique de multiples représentations (et informations associées) dans le but de définir un système pour l'observation pluridisciplinaire de l'édifice. La représentation numérique de la morphologie des édifices étudiés a pour vocation d'être exploitée pour la constitution de banques de données patrimoniales et devient alors le support privilégié de navigation et d'interactions avec les sources documentaires qui lui sont associées. Ces informations et les représentations géométriques multiples des objets constituant l'édifice peuvent être restitués de manière dynamique en fonction de différents points de vue utilisateurs.

Dans le cadre du projet NUBES, les développements informatiques abordés dans le cadre de notre thèse concernent spécifiquement le module « Nubes Imago », un système pour la recherche visuelle de sources iconographiques en fonction de critères spatiales, morphologiques et sémantiques. Les fonctionnalités implémentées dans cette application concernent :

- le téléchargement sur serveur (upload) de sources iconographiques ;
- le référencement spatial de sources iconographiques ;
- la manipulation dans l'espace de représentations tridimensionnelles d'édifices et des sources iconographiques superposées ;
- la recherche de sources iconographiques en fonction de critères spatiaux, morphologiques et sémantiques.

4.3.2. Présentation de l'environnement de développement

4.3.2.1. PHP/MySQL

PHP est un langage interprété exécuté du côté serveur. La syntaxe du langage provient de celles du langage C, du Perl et de Java. Ses caractéristiques principales concernent la possibilité de d'inclure le script PHP au sein d'une page HTML et l'interfaçage avec MySQL, un SGBD (Système de Gestion de Bases de Données) disponible sur de nombreuses plateformes : Unix, Linux, Windows, MacOS X, Solaris, etc.

4.3.2.2. Virtools DEV

Le paradigme de développement spécifique de Virtools considère les objets 3D comme des composants individuels, indépendants des données associées. L'architecture résultante permet d'attribuer des comportements modulaires aux objets qui composent la scène. Les applications créées avec ce système sont directement exploitables comme exécutables ou comme plugin Internet.

Virtools DEV répond à tous les besoins de notre système car il peut afficher un rendu basé sur les points pour les nuages de points, un rendu filaire ou ombragé pour la maquette d'un édifice, mais aussi un rendu texturé pour l'affichage des photos dans l'espace. Virtools DEV est composé de 5 composants clefs : l'interface Graphique, le moteur à comportement, le SDK (un ensemble d'outils de développement) et le langage de script (VLS).

4.3.3. L'architecture du système

En relation à la structure générale de la plateforme NUBES, notre système se base sur une architecture en trois parties :

- une base de données développées en MySQL contenant les représentations 3D des édifices, les nomenclatures utilisées pour l'annotation sémantique et les sources iconographiques ;
- une scène 3D interactive pour l'affichage et la manipulation des représentations 3D et des superpositions image/modèle, développée en Virtools DEV ;
- une application web en PHP qui constitue l'accès utilisateur aux données stockées dans la base et fournit les passerelles nécessaires pour le dialogue entre la scène 3D et la base de donnée.

La scène 3D est intégrée à l'application Web sous forme d'un plug-in qui communique avec l'application en PHP/MySQL par des fonctions javascript.

L'application Web est basée sur le dialogue entre une base relationnelle SQL et la scène 3D par le biais de script php et java. En ce qui concerne l'envoi de commande de la partie Web vers la scène 3D nous utilisons des fonctions javascript spécifiquement programmées. Dans le sens inverse un bloc DEV (GetWebData) permet de formuler des requêtes (URL avec des variables) dans la base de données en passant par une page PHP. Pour gérer la synchronisation entre les représentations dans l'espace et les données stockées dans la base, nous dupliquons de façon dynamique les tableaux relatifs aux entités manipulées dans une base de données interne à la scène 3D.

