Automated Texture Extraction from Multiple Images to Support Site Model Refinement and Visualization *

Xiaoguang Wang, Jonathan Lim, Robert T. Collins, and Allen R. Hanson

Department of Computer Science
Box 34610, University of Massachusetts
Amherst, MA. 01003-4610, USA
Email: xwang@cs.umass.edu

Abstract

Texture mapping has wide and important applications in visualization and virtual reality. Surface texture extraction from a single image suffers from perspective distortion, data deficiency, and corruption caused by shadows and occlusions. In this paper, a system is developed for automated acquisition of complete and consistent texture maps from multiple images in order to support subsequent detailed surface analysis and scene rendering. Given camera and light source parameters for each image, and a geometric model of the scene, the textures of object surfaces are systematically collected into an organized orthographic library. Occlusions and shadows caused by objects in the scene are computed and associated with each retrieved surface. A "Best Piece Representation" algorithm is designed to combine intensities from multiple views, resulting in a unique surface intensity representation. Detailed surface structures, such as windows and doors, are extracted from the uniquely represented surface images to refine the geometric model. Experiments show successful applications of this approach to model refinement and scene visualization.

Keywords: texture mapping, CAD and GIS Systems, virtual reality, computer vision

1 Introduction

Texture mapping has become an increasingly important aspect of computer graphics with the wide availability of specialized hardware and software. Visualization and virtual reality applications are now capable of generating high-quality renderings and animations of textured geometric objects. However, in certain applications such as landscape architecture, some level of realism is sacrificed in using artificial textures as opposed to ones synthesized from images of an actual scene or object (for example, using a generic brick texture on a building face in place of an image of the surface itself). Synthesizing surface textures from images of a scene is by no means an easy task; what has not been widely addressed is an automatic and efficient means of texture acquisition and management.

This paper focuses on the automated acquisition of complete and consistent texture maps from multiple images in order to support subsequent detailed surface analysis, scene rendering, and other goals. Given a polygonal model, a camera viewpoint, and a texture map, rendering the texture onto the model surfaces is a well-understood technique in computer graphics [1, 2, 3].

*This work was funded by the RADIUS project under ARPA/Army TEC contract number DACA76-92-C-0041.

However, extraction of a complete and consistent texture map for all surfaces of a 3D object is a challenging task. Figure 1 is an image portion from a set of eight images of a single site. Consider the problems involved in generating texture maps for the largest building in the site, which appears at the top of the image. The textures are not complete: some walls are occluded by the building, some walls are beyond the view of the image. So multiple images must be used. There are only a few pixels in the vertical direction of the building wall; other, more oblique views, may have higher resolution. Buildings are often built fairly close to one another and it may not be possible to get a good view - occlusion and shadowing may break up a texture map and it may be necessary to "piece together" the map from several images in order to remove the occluded or shadowed portions. Since the images may be taken at various times of day, from different positions, and under varied weather conditions, the brightness of surface intensity maps may vary considerably from image to image. All of these factors must be considered when generating a consistent texture map from the image sequence. Traditional graphics algorithms provide no direct way to collect and combine intensities under various conditions for texture consistency and completeness.

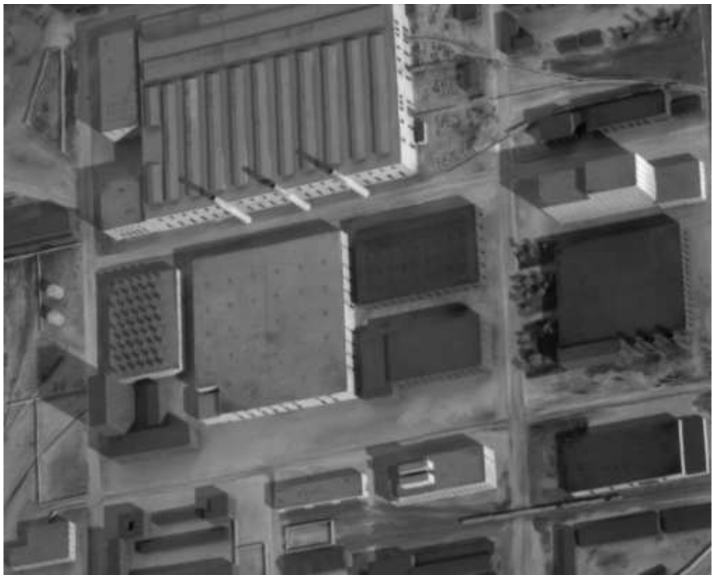

Figure 1: Part of site image J1 from Model Board 1

In the following sections, we describe a multi-image texture mapping approach. The emphasis is on obtaining consistent and complete intensity information with best texture quality from multiple images. The application for this architecture is the ORD/ARPA RADIUS project, whose goal is to develop soft-copy model-based image exploitation tools and infrastructure for image analysts. An important function of the evolving system is the development of a *site model* of an area from multiple images (*site images*). Uses of the site model include 3D site visualization and familiarization, mission planning and assessment, and change detection [4]. Consequently, an important component of the site model is an accurate geometric representation of significant objects in the site, such as buildings, as polygonal models with associated surface texture maps derived from the site images. Experimental results indicate that the techniques presented by the paper provide a convenient way in virtual reality rendering and in the meantime facilitate *model refinement*, where the acquired texture maps drive an analysis of the building surfaces to detect detailed structures such as windows, doors, and roof vents, and enrich the original geometric model. Potential uses of the presented approach are not confined to

RADIUS. It can be applied to any scene containing objects approximated by polygonal facets of arbitrary shape and orientation.

Section 2 describes an orthographic facet image library as an architecture of multi-image texture mapping. Section 3 discusses modeling of occlusions and shadows and how this information is used to determine usable portions of individual texture maps. Section 4 describes how a single, consistent representation of the texture map for a facet is obtained from the tagged library of facets. Section 5 discusses related issues in model refinement and scene rendering, and Section 6 describes future work.

2 Orthographic Facet Image Library

The texture map extraction starts from an initial, coarse understanding of the 3D environment, of which the textures are interested. Collins, et al. [5] describe recent progress in RADIUS site model acquisition. A set of image understanding algorithms have been developed to extract the geometric site model from the RADIUS Model Board 1 site image sequence J1-J8. As a result, 25 buildings that represent most of the structures in the site have been extracted in the form of 3D polygons. Figure 2 is a CAD display of the site model. The site model acquisition module only gives a coarse geometric description of the structures in the site. Due to perspective distortions and corruption caused by occlusions and shadows, detail structures are very difficult to be determined directly from the original site image sequence.

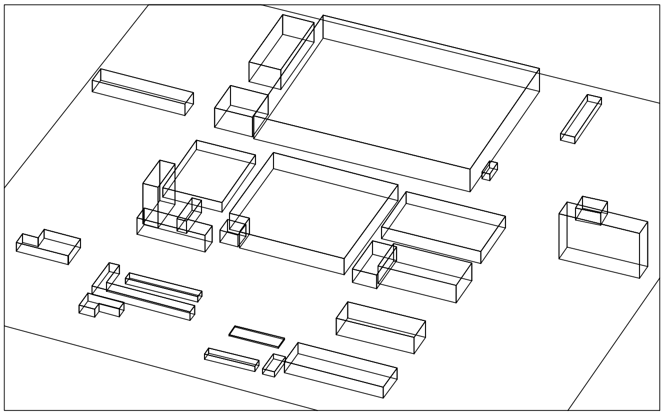

Figure 2: A CAD display of the acquired site model

The acquisition of the site model provides a natural and convenient way of texture extraction and management for the goals of geometric model refinement and textured model visualization. The architecture of our automated texture extraction and management system is an *orthographic facet image library* (OFIL). A *facet* in the site model is a modeled polygonal surface of a building, such as a wall or a roof. A *facet image* is an intensity image of the facet as seen under orthographic projection. An OFIL for a site model stores indexed orthographic images of all the polygonal building facets that have been modeled in the site. The intensity values of each facet image are sampled from a site image using the traditional inverse texture mapping algorithm. If the facet appears in more than one site image, the library will hold all the facet images (*versions*) for the facet. These multiple versions are well-indexed to facilitate library access. For example, a horizontal roof facet usually appears in all the aerial site images and thus has a complete set of orthographic versions in the library, whereas other facets like vertical walls only appear in some of the site images. Thus the availability of a facet version is an important piece of information to be indexed in the library. Other information like localization (how the facet is aligned in the orthographic image) and visibility (the obliqueness and lighting conditions of the facet in the site image) need to be recorded as well. In summary, an OFIL is an image database whose records are orthographically projected facet images together with relevant information to aid retrieval and analysis of these images.

Construction of an OFIL as an intermediate-level representation has advantages over the mechanism of storing raw site image intensities. First, individual facets are stored separately so that specific surface structure extraction techniques can be applied only to relevant surfaces: window extraction on wall images, roof vent computations on roof images, etc. Second, many man-made structures related to buildings have rectilinear, repetitive patterns, like the lattices of windows on building walls. The orthographic facet image provides a view that is free from perspective distortion, which is critical to the development of efficient techniques for extracting these patterns. Third, the collection and alignment of all the visible versions of a building facet provides a mechanism for comparing and combining intensities from multiple views to produce a better, or clearer, view of each facet. Finally, a set of separately stored facet images is a natural and convenient component of a system for rendering texture-mapped 3D perspective views.

3 Occlusion and Shadow Modeling

One important feature of the OFIL architecture is its ability to handle *occlusions* and *shadows* that arise on the object surfaces in a scene. Occlusions in a site image occur when objects stand between the camera and the object surface of interest. During texture mapping, intensity values from the occluding objects get mapped onto the orthographic facet image as well. Shadow areas occur on the surface due to objects standing between the light source and the surface of interest. Generally speaking, occlusions do not provide useful intensity information for surface texture analysis, while shadow areas may still be useful, provided that enough dynamic range exists in the shadow area to reconstruct the texture of the surface. Typically, unocclude, sunlit parts of the surface are the best sources of intensity information.

To avoid the negative effects of occlusions and shadows on subsequent facet image analysis, an extra record is associated with each facet image, explicitly indicating which pixels in the image are occluded and which are in shadow. In the OFIL, each orthographic facet image version is associated with an orthographic *labeling image* of the same size, in which each pixel is composed of a number of "attribute" bits that record whether the corresponding pixel on the facet image is occluded or in shadow. Figure 3 shows an example of this kind of labeling.

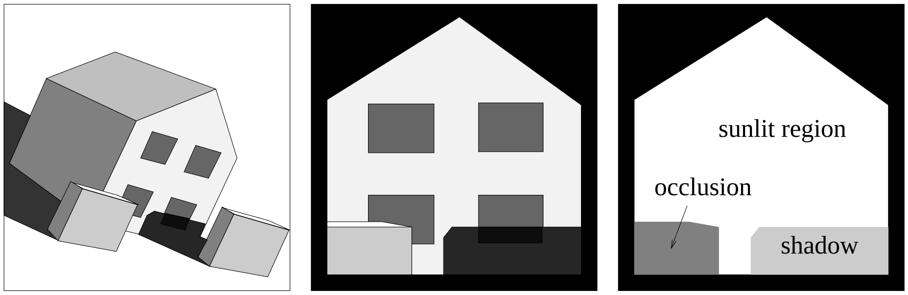

Figure 3: Occlusion and shadow labeling (left: site image, middle: facet image, right: facet labeling image)

The computation of occlusions and shadows in the current OFIL system is performed in a model-driven way, using the geometric data contained in the site model, the camera parameters of the site image, and light source parameters. This is a classic problem of hidden surface and shadow computation in computer graphics [1].

Labeling images play an important role in indexing the pixel attributes of orthographic facet images. In the current system, labeling images also provide other information besides occlusion and shadow. The complete set of attribute bits in a labeling image is:

- *Facet Bit.* The system is able to handle arbitrary polygonal facets. This bit tells whether a pixel in the rectangular facet image is contained in the facet polygon.
- *Presence Bit.* A building surface may partly lie outside of the boundaries of a site image. This bit labels whether a pixel's intensity is present in the site image.
- *Occlusion Bit.* Tells whether the pixel is occluded.
- *Shadow Bit.* Tells whether the pixel is in shadow.

To provide a glimpse of the OFIL, Figure 4(a) shows a set of orthographic facet images, with labeling, for a particular building facet in Model Board 1. This rectangular facet is the right wall of the largest building shown on top of site image J1 in Figure 1. This wall appears only in site images J1, J2, J6, and J8, and thus only these four versions are available. In site image J1, part of the wall is cut by the image border, as is marked in the labeling image for the version from J1. Facet versions from J6 and J8 look darker because they are self-shadowed, i.e. oriented away from the light source. In site image J2 and J6, this wall is viewed from such an oblique angle that the textures mapped from these two images provide very little additional information over much of the wall surface. However, near the lower left of the wall there is another small building that occludes the wall in versions J1 and J8, but not in J2 and J6 due to the extreme obliqueness of the viewing angle. From this example we can see that multiple images are necessary to see all the portions of this particular building face, and that the OFIL has collected and organized the available information about this wall facet.

4 Unique Intensity Representation

The OFIL collects all the intensity information about each building facet in the form of separate, but aligned, orthographic image versions. For many tasks it is desirable to produce a single, unique intensity representation of the facet. A simple approach is to select one "good" version from the facet images as the unique representation. The drawback of this approach is that any occlusions or shadows in that facet version will be included as artifacts in the resulting representative texture map. In addition, the version that is least corrupted by occlusions and shadows is not necessarily the clearest one. In Figure 4, the version that contains the least occlusions and shadows is the one from site image J2, but that version is too blurred to be a good representation due to the obliqueness of the viewing angle.

In this section we present a *best piece representation* (BPR) method for combining intensities into a unique representation of a facet. This method is based on the observation that different regions on a facet may only have good visibility in different image versions due to the existence of occlusions and shadows. The final representation is a combined image whose pixel intensities are selected from among multiple image versions in the OFIL. A representative facet image synthesized in this way is called a *BPR image*.

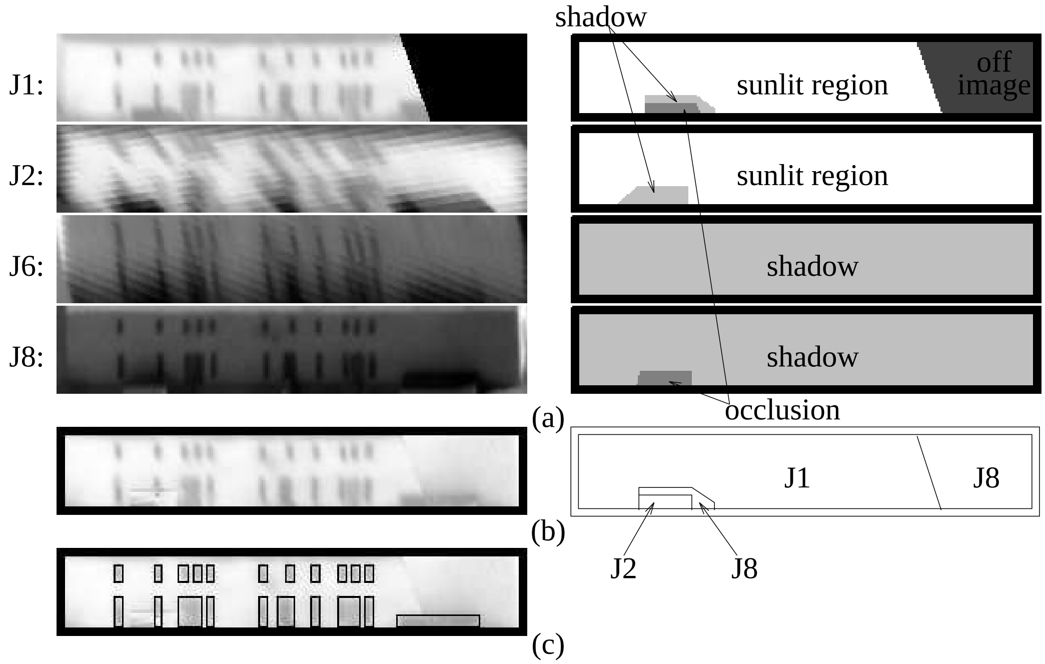

Figure 4: The application of OFIL architecture to a building facet
(a) left: orthographic facet image versions, right: their labeling
(b) left: the BPR image of the facet, right: regions of the intensity sources
(c) the result of symbolic rectangle extraction toward model refinement

The orthographic alignment of the facet image versions and the occlusion and shadow labeling make the BPR method very easy to compute nder the OFIL architecture. We first define the term *piece*. Each facet image version is generally partitioned into three pieces: *sunlit piece*, *shadow piece* and *useless piece*. Sunlit pieces contain all the pixels that are labeled by Facet and Presence bits and not labeled by Occlusion and Shadow bits. They represent the sunlit portions of the surface. Shadow pieces, representing the shadowed part on the surface, contain all the pixels that are labeled by Facet, Presence and Shadow bits. All the other pixels are considered part of a useless piece (typically an occlusion) that provides no intensity information for the facet. Any particular pixel in the facet falls into a piece of one of these three kinds in each of its versions. To synthesize a representative facet image, the BPR algorithm runs through the pixels of the representative image, determines for each pixel which piece it falls into in each available facet version, evaluates the quality of each piece based on a criterion described below, and finally picks as the representative intensity of the facet pixel the corresponding pixel value from the highest quality piece.

Some issues arise in implementing the BPR method. One of them is how to evaluate the quality of a piece. Generally speaking, a good piece is a piece that reveals a clear, high resolution look at the surface detail. The detail-revealing ability of a piece is evaluated by a heuristic measure that takes into account such factors as the distance between the camera and the surface, the obliqueness of the viewing angle, and the lighting condition on the surface. In the BPR algorithm, for each piece p on facet image version v, we evaluate the piece using the value given by the function

$$f(p) = \alpha(p)A(v),$$

in which $A(v)$ is the area that the facet occupies in the site image from which version v was obtained, and

$$\alpha(p) = \begin{cases} 1, & \text{if piece } p \text{ is a sunlit piece} \\ a, & \text{if piece } p \text{ is a shadow piece } (0 \leq a \leq 1) \\ 0, & \text{if piece } p \text{ is a useless piece.} \end{cases}$$

The area factor $A(v)$ reflects the combined effects of surface-camera distance and viewing obliqueness. The weighting factor $\alpha(p)$ is set according to the attribute bits of the piece. Shadow pieces have a smaller range of intensity values, and are assumed to reveal less information than sunlit pieces. In our system we lower the heuristic value of shadow pieces by letting $a = 0.5$, an empirical constant. With the heuristic function defined in this way, every pixel in the BPR image comes from the associated piece with the highest value of $f(p)$.

Another issue is the consistency of the intensity data. A BPR image is a synthesized image whose intensities are selected from different facet image versions. These intensities cannot be juxtaposed directly because they often come from different pieces captured under different lighting conditions. We solve this problem by making two assumptions. One is that every local piece on a surface has a similar intensity histogram distribution to the whole surface when seen under the same lighting conditions. This is true when the texture is fairly uniform on the surface, like on a wall where the windows are aligned evenly. The other assumption is that the intensities in a piece never reverse their order under any lighting conditions. This assumption asserts that if windows are darker than the wall under sunlight, they will remain darker than the wall even when seen in shadows. Under these assumptions, we use a histogram adjustment algorithm, prior to running the BPR algorithm, to make the intensities from different facet image versions consistent. The algorithm has two steps. First, it chooses a useful piece (a sunlit piece or a shadow piece) as an *exemplar piece*, and computes its intensity histogram distribution. Second, the intensities of all the other pieces for that surface are adjusted to have the same histogram distribution as the exemplar piece. Another heuristic function, g, is designed for selecting the exemplar piece. For any piece p on the facet image version v,

$$g(p) = \alpha(p)A(v)S(p) = f(p)S(p),$$

where $\alpha(p)$, $A(v)$ and $f(p)$ are as described above, and $S(p)$ is the area percentage of piece p to the whole facet. The meaning of $A(v)S(p)$ is the area of piece p in the site image from which facet image version v is texture mapped. The bigger the area a piece occupies, the richer the texture it contains, and the more qualified it is to be chosen as the exemplar piece.

Figure 4(b) shows the BPR image synthesized using the facet images in Figure 4(a). The sunlit piece from the J1 version is chosen as the exemplar piece for histogram adjustment. We can see that some regions in the BPR image contain intensities from version J1, some others from J8 and J2. The intensities from different sources are consistent as well.

A complete orthographic facet image library has been built up in the RADIUS texture extraction system from J1-J8 of Model Board 1 imagery. For the 25 modeled buildings, 133 facet images are created automatically using the BPR algorithm as surface texture maps. Among them, there are 108 rectangular walls, 21 rectangular roofs and 4 L-shaped roofs.

5 Model Visualization and Refinement

The purpose of model visualization is to provide a visually realistic rendering of the scene. Given a desired camera pose, we want to render a 3D view of the site with textures mapped onto the modeled facets of the objects. In our visualization system, the textures of the object surfaces are mapped from the BPR facet images, so that the rendered scenes are free of defects, like shadows and occlusions, caused in the original perspective images. A sample 3D site rendering is shown in Figure 5. The big building on top of site image J1 in Figure 1 is at the right side of the new rendered image, as seen from this angle, and clearly shows the BPR synthesized wall from Figure 4(b). Sequences of rendered 3D views are constructed in this way for site visualization and familiarization. They are used for generating animated, virtual fly-throughs, when played back at video speeds.

Detail building structures such as windows and doors may not be easily visible, but are useful in providing symbolic and geometric data for applications such as landmark recognition and mission planning and assessment. Automatic detailed structure extraction is usually a difficult process in aerial image understanding because of noise and lack of detail caused by the relatively small fields of views on these structures. High-level knowledge, such as that of rectilinear and repetitive patterns in man-made structures, is always necessary for detailed structure extraction. However, high-level knowledge is not easily applicable to the site images due to data corruption caused by perspective distortion, occlusion, and lighting condition variations. Texture maps automatically obtained from the previous sections are free from these problems, and thus provide an environment for knowledge-based detailed structure extraction. The symbolically extracted structures are then incorporated into the original site model to form an extended, refined model.

As one example of detailed structure extraction, we have developed a generic algorithm for detecting dark, oriented rectangular patterns to extract windows and doors on wall surfaces. Figure 4(c) shows the results of applying this algorithm to the BPR image in Figure 4(b). The algorithm is performed on a wall facet image in three steps: seeding, region growing, and lattice adjustment. Seeding is to look for local "intensity dips" or blobs by checking the intensity variations. To avoid an explosion of seed numbers, seeding is restricted only to those repetitive intensity variations since it is known that windows and doors only appear in that way in this site. Then, starting from each seed, a region growing procedure is called to mark the blobs as window and door candidates. Regions are forced to grow as rectangles because windows and doors are always in this shape in the data set. The growing is stopped by local intensity criteria. The rectangles so far obtained might not be aligned regularly due to intensity noise. The knowledge of global pattern repetition is again applied to remove unqualified candidates that are in bad posisions, and to adjust the positions and sizes of the rectangles that are believed to be windows or doors. 21 dark regions in Figure 4(c) are extracted as hypothesized windows and doors and written into the refined model. The rightmost flat rectangle and the second-left rectangle on the bottom of the image cross over two or more image pieces whose intensities come from different image versions, suggesting that our BPR algorithm maintains good consistency of image intensities from different sources. It is worth noting that high-level knowledge is critical in this kind of detailed structure extraction. The OFIL architecture simplifies the process substantially as the high-level knowledge can be applied very conveniently under this architechture.

Refined models also provide better model visualization. Figure 6 shows a stereo view of the big building on top of site image J1. The windows and doors were extracted using the algorithm described above. Two levels of floors inside the building are added into the model with synthesized textures. We can see the floor textures through the windows and doors, which are designed to be transparent. The refined model clearly provides added realism over the flat, 2D textures extracted from the original site images.

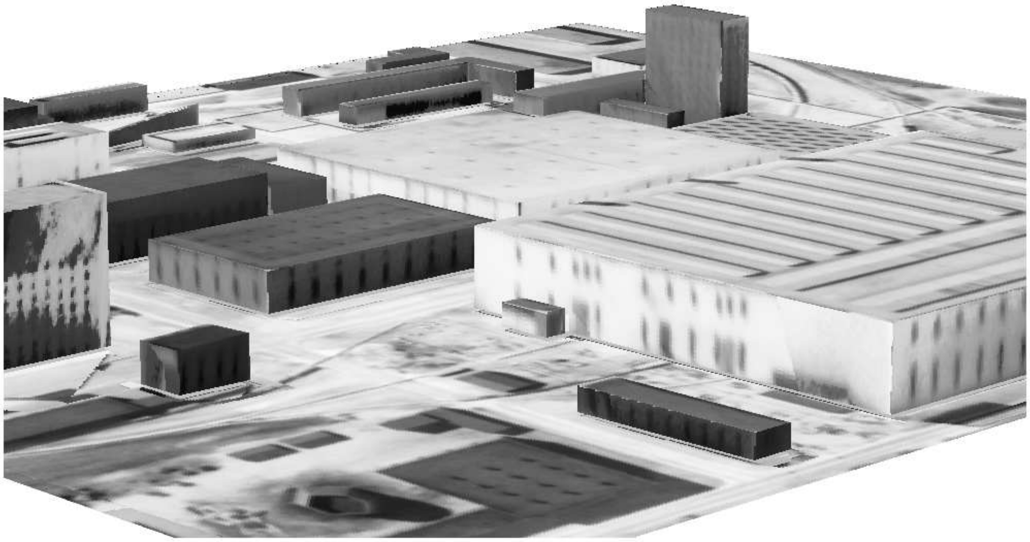

Figure 5: Model visualization using BPR texture maps

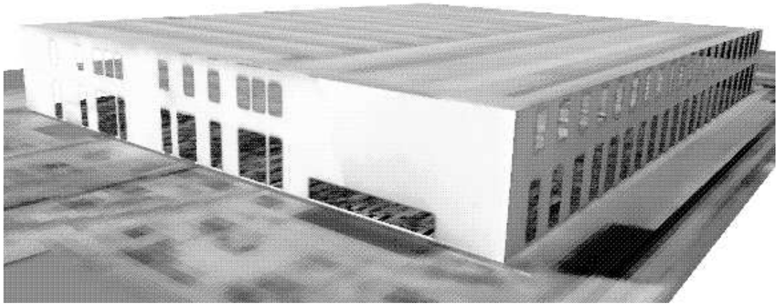

Figure 6: Visualization of the big building with the refined model

6 Discussion and Future Work

Efficient intermediate representations play an important role in complex image understanding. In this paper we have proposed an orthographic facet image library architecture to aid image understanding of buildings from aerial site images. With the assumption that camera parameters, lighting conditions, and a geometric site model are known, facet intensity information is collected systematically from the original set of perspective images into an organized orthographic library. Occlusions and shadows on each facet are recorded in the library. A best piece representation algorithm is proposed to combine intensity information from multiple image versions to give a complete and consistent intensity representation of each facet, which substantially simplifies later work in model refinement and visualization. This paper has presented experimental results of the application of the proposed architecture to model refinement and scene rendering.

High-end graphics engines are now capable of generating texture-mapped, real-time, scene fly(walk)-throughs suitable for use in flight simulation, site visualization, and mission planning. These virtual reality environments are currently built by hand, and are expensive and time consuming to produce. The techniques presented in this paper form one component of a system for automatically acquiring environmental models of great geometric and photometric realism.

In addition, generation of a viewpoint-independent surface intensity map yields a rich surface description that can greatly simplify vision tasks such as object recognition and the analysis of surface markings.

The drawback of the proposed architecture is that it is almost purely top-down, or model-driven. Inaccuracy and incompleteness in the 3D model can cause problems. For example, undetected objects (such as smoke stacks) in the site will not only be missed in the library, hence in any reconstructed view, but their absence also leads to misidentification of the occlusions and shadows caused by them. Inaccuracy of the camera model and/or camera parameters can cause mis-alignment of the different versions of a facet. Inaccuracy of sun angle parameters can cause erroneous shadow labeling and bad data fusion. One possible way to mend these deficiencies, which is currently under study, is to introduce bottom-up, or data-driven, processing modules into the library. This includes shadow detection on the facet image and model/camera/sun parameter refinement as directed by the comparison between the model-driven predicted and data-driven extracted shadows. The boundaries of shadows in images are often blurred, while model-driven shadow computation always gives sharp edges. This phenomenon further corrupts the quality of intensity fusion by the BPR algorithm. An enhanced BPR method using antialiasing techniques is being developed to avoid this problem.

The BPR algorithm selects pixel intensity values from the most competitive pieces in the OFIL, but it ignores completely the less competitive versions of a facet. Another possible approach is to generate a super-resolution representation of the facet images, which combines intensities from all facet image versions and is expected to utilize as much information as possible from the OFIL. Some of the requirements of a super-resolution method are: a more accurate registration of the facet image versions than required by BPR, a more strict comparability of the intensities from different versions than required by BPR, and a mechanism at the sub-pixel level to fuse the intensities from different versions into one image.

Acknowledgements

We would like to thank Professor Edward Riseman for his discussions and comments, and Fred Weiss for his technical support.

References

[1] A. Watt and M. Watt, *Advanced Animation and Rendering Techniques: Theory and Practice*, ACM Press, New York, NY, 1992.

[2] P. Heckbert, "Survey of Texture Mapping," *IEEE Computer Graphics and Applications*, vol. 6, no. 11, pp. 56-67, November 1986.

[3] D. Forsyth and C. Rothwell, "Representations of 3D Objects that Incorporate Surface Markings," *Applications of Invariance in Computer Vision*, Springer-Verlag Lecture Notes in Computer Science no. 825, pp. 341-357.

[4] D. Climenson and T. Strat "RADIUS: Site Model Content," *Proc. Arpa Image Understanding Workshop*, Monterey, CA, 1994, pp. 277-286.

[5] R. Collins, Y. Cheng, C. Jaynes, F. Stolle, X. Wang, A. Hanson, E. Riseman, "Site Model Acquisition and Extension from Aerial Images", *Fifth International Conference on Computer Vision*, Cambridge, MA, 1995, pp. 888-893.