Fusing Imagery and 3D Point Clouds for Reconstructing Visible Surfaces of Urban Scenes

(Invited Paper)

Toni Schenk
Dept. of Civil and Environmental Engineering
and Geodetic Science
The Ohio State University
Columbus, OH 43210
Email: schenk.2@osu.edu

Beáta Csathó
Dept. of Geology
University at Buffalo, SUNY
Buffalo, NY 14260
Email: bcsatho@buffalo.edu

***Abstract*— The automatic reconstruction of urban scenes from sensory input data is a daunting task. By and large the task remains unresolved, although a considerable amount of research has been devoted to its solution. Many of the proposed methods are either too application dependent, or address only some aspects of the general problem. Moreover it appears that solutions based on a single sensor source, for example intensity images or laser point clouds, lead to partial solutions. In this paper we propose the reconstruction of visible surfaces from multi-sensor data, embedded in a fusion framework. We postulate that the reconstructed surface is an intermediate and application independent representation of the scene, similar to the 2.5 D sketch proposed by Marr in his vision paradigm. In contrast to the viewer based 2.5 D sketch, our reconstructed surface is represented in a suitable 3D Cartesian reference system. It contains explicit surface information, including shape and surface discontinuities. We argue that such an explicit description greatly benefits applications, such as object recognition, populating or updating GIS, change detection, city modeling, and true orthophoto generation. This is because the 3D object space enables more powerful reasoning methods to aid object recognition and image understanding as opposed to the traditional approach of reasoning in the 2D image space. Another strong motivation for the proposed application independent surface reconstruction scheme is the multi-source scenario with imaging and laser point data, and possibly hyperspectral data. These widely disparate data sets contain common (redundant), complementary and occasionally conflicting information about the surface. The paper discusses the notion of different surfaces and their relationships. Major emphasis is placed on the development of a general, true 3D surface representation scheme that copes with the problem of multi layer surfaces (e.g. multiple overpass).**

I. Introduction

Judged from the number of publications, surface reconstruction and object recognition from sensory input data is a very active research area in photogrammetry. Despite considerable progress, we are still far from a unified and generally accepted methodology that would lend itself to automatic (not even to think of autonomous) reconstruction of urban scenes.

It has long been recognized that surfaces, their properties and characteristics play an important role in image understanding and object recognition. Many other perception tasks, such as navigating a robot, make extensive use of surface properties. Surfaces are intermediate representations in the long processing chain from data to objects.

There is a rich body of literature related to surface reconstruction and extraction of man-made objects from aerial images of urban scenes. Most of the proposed surface reconstruction methods aim at the automatic generation of Digital Surface Models (DSM), for example by computing dense depth maps from stereo (e.g. [2], [10], [5], [4]), or from multiple images ([14]). The complexity of urban scenes, the close proximity of different objects (e.g. trees next to buildings), large elevation differences, occluded regions, shadows, periodic structures, and moving objects pose an almost insurmountable challenge to do surface reconstruction from images only.

The advent of airborne laser scanning (ALS) has shifted research more recently toward surface reconstruction and object extraction from 3D laser point clouds (e.g. [9], [1], [6], [7], [8]). Considerable effort has been devoted to determine the topographic surface, or bare-earth, by filtering out points which have been reflected from objects above the ground. With the ever increasing density of laser points and the availability of additional information, such as the intensity of the returning pulse, multiple returns, or the recording of the entire waveform, there is hope that object extraction may become more successful.

In this paper we combine aerial images, laser point clouds, and possibly multispectral/hyperspectral data for determining the complex surface of urban scenes. Specifically, we propose an intermediate surface representation without thinking of applications. In fact, the surface is reconstructed without relying on any domain knowledge and we see this application-independent surface as the transition between early and late vision processes. The next section provides some background information and motivation for the proposed representation. Sec. 3 introduces the important notion that physical surfaces are topologically complete—in contrast to sensed surfaces that have “holes” (occlusions). The objectives of the proposed

1-4241-4244-0712-5/07/$20.00

surface representation are stated in Sec. 4 and important implementation issues are addressed in Sec. 5. Since this is a conceptual paper, no examples with real data are included. The main purpose of the paper is to stimulate discussions about the fascinating and challenging problem of urban mapping and city modeling.

II. Background

A. Human visual system and surfaces

Humans are remarkably adept in organizing visual stimuli, imposing meaningful structure for interpreting a scene. Marr ([11]) approaches vision as a complex information processing task that has to be understood on the three levels of computational theory, representation and algorithms, and hardware implementation. Human visual processes are organized in a modular fashion and the representational framework includes the retinal images, the primal sketch, the 2.5 D sketch, and the 3D model representation. The primal sketch makes image information explicit, for example intensity changes and their organization.

The 2.5 D sketch is a pivotal point in Marr's vision theory. It is the end result of early vision processes that aim at extracting information about the visible surface of a scene without special knowledge what is in the scene. This viewer centered representation objectively describes the physical reality of the scene *before* it is decomposed into objects. The 2.5 D sketch consists of the two quantities depth and surface orientation, and their discontinuities. Discontinuities of surface orientations are manifest by contours, for example. We emphasize the important notion that surface information in the 2.5 D sketch is independent of objects. It does not matter if the reconstructed visible surface is in fact the surface of object X, or Y, or Z. That is, no semantic meaning is associated. The *construction* of the 2.5 D sketch as the result of early vision processes is the end of pure perception and solely data-driven, requiring only general knowledge about the physical world. In contrast, the *interpretation* of the 2.5 D sketch, or late vision processes for that matter, are no longer data-driven and require special knowledge about the nature of objects, their use or functionality.

The notion that vision is modular is a widely adopted paradigm in cognitive science and computer vision. It is inconceivable to perform such complex tasks as recognizing objects in one process that operates directly on the raw visual stimuli. The strongest motivation for our proposed application-independent surface reconstruction is the 2.5 D sketch.

B. Perceptual organization

Perception is a fundamental process that allows organisms to interpret sensory stimuli to create a meaningful description of the world. Such descriptions are the starting point for making decisions and executing actions. Perceptual organization is concerned with imposing structure on primitives that are obtained from sensory input data. It is commonly agreed in the computer vision community that sensory input data should first be organized into perceptually meaningful groups before higher-level processes begin. For example, image interpretation, object recognition, robot navigation, image understanding do not operate directly on the sensed signals, rather on more abstract descriptions of the physical world.

In [12], the authors propose to classify perceptual processes at the signal, primitive, and structural levels. Another distinction is the spatial and temporal domain perceptual organization is performed, e.g. 2D or 3D, with or without time domain. Perceptual organization at the signal level does not require domain-specific, higher-order knowledge. It is purely data-driven, relying only on the general principles. In contrast, perceptual processes on the structural level are application dependent, requiring specific information. Organizing structures to higher order entities that correspond to structures of real world objects is partially driven by information about objects. This is an important point because our proposed surface representation is application independent; processes that use domain specific knowledge are not allowed in the construction.

III. Surfaces

This section describes different surfaces in preparation for the definition of our proposed application-independent surface reconstruction (AISR) scheme.

A. Physical surfaces

The physical world consists of objects. Disregarding flying objects we note that the surfaces of all objects are connected and we call this joint surface the *physical surface*. The physical surface is continuous everywhere; the only discontinuity is in the surface orientation that may occasionally change abruptly. Continuity implies that the physical surface has no ends, referred to as topologically complete in this paper.

Imagine an ant with its capability to walk on the top and—defying gravity—on the bottom surface of horizontal objects, on vertical surfaces, and on overhangs. Assuming no hostile surface material (e.g. a sticky surface), the ant would be able to reach every place of the physical surface without ever meeting an end that would force it to turn around or to jump. It would climb up a tree, follow twigs, walk on top of leaves and back on the bottom surface of the leaf (which is a different physical surface, connected to the top with a tiny, elongated surface, reflecting the thickness of the leaf). The ant's only problem is to master the changes in surface orientation which occasionally may pose a challenge if you think of the end of a leaf with the rapid transition from top to bottom surface, amounting to a change of 180^o in one step.

Strictly speaking, the ant cannot reach every point on the physical surface; for example, it will walk across tiny cracks. A larger ant will cross even wider cracks and will walk over obstacles its little brother would have experienced as a series of surface orientation changes. Let us equip the ant with a 3D positioning system that will send a signal whenever the ant moves by a distance equal to its size. This will result in a discrete sampling of the physical surface with the level of detail proportional to the ant's size. Obviously, this analogy raises the question about scale. We will address the scale problem in Sec. IV-D. In conclusion, the physical surface is continuous everywhere with the exception of the surface orientation that occasionally changes abruptly. We also consider it opaque. As a consequence, the opaque physical surface has no direct information what sort of objects it envelops.

B. Sensed surfaces

In order to approximate (reconstruct) the physical surface, various types of sensors are employed. However, the sensors will only cover a part, called the *sensed surface*. A camera, for example, records intensity values that are proportional to the reflectance properties, the illumination, and the inclination angle (angle between surface normal and illumination direction). Since the camera is sensitive to the visible portion of the spectrum, the sensed surface is what we perceive as the *visible surface*. Airborne laser scanning systems (ALS) record the travel time of a laser pulse from the sensor to a spot on the surface where it is reflected (back scattered), partially back to the sensor. Whereas imaging sensors provide complete coverage of the sensed surface, ALS systems sample it discretely. Sensors such as ground penetrating radar or geophysical instruments, do not sense the physical surface. Hence, they are not useful in the surface reconstruction as discussed here.

C. Occluded surfaces

Considering the complexity of the physical surface it is nearly impossible to sense it everywhere. For example, airborne sensors will only record data from portions of the physical surface that is visible from the sensor's position. We call the portions of the physical surface that is not sensed *occluded surface* (not to be confused with the part that is invisible from a viewer's position). Occluded surfaces are simply the difference between physical and sensed surface. Unlike physical surfaces, sensed surfaces are disconnected and may have abrupt depth changes at their boundaries. Adding the occluded surfaces avoids this problem: sensed *and* occluded surfaces together are closed. We also believe that the explicit representation of occluded surfaces is useful in object recognition. For example, a surveillance task benefits from the knowledge about occluded surfaces in that it may deploy sensors to cover missing areas.

D. Virtual surfaces

A major problem in segmenting 3D point clouds into spatially coherent surface patches is the selection of seed regions. If we are lucky then the selected points belong to a smooth portion of the sensed surface. It cannot be avoided that the seed region straddles discontinuities (in surface depth and/or orientation), however. Fitting a model to noisy data from unknown populations is a challenge. Fig. 1 illustrates the problem. All points selected for the fitting process pass the threshold criteria, but the fitted surface is *not* a part of the physical surface. Such surfaces are adequately called *virtual surfaces* and should be avoided since they are likely to confuse subsequent interpretations.

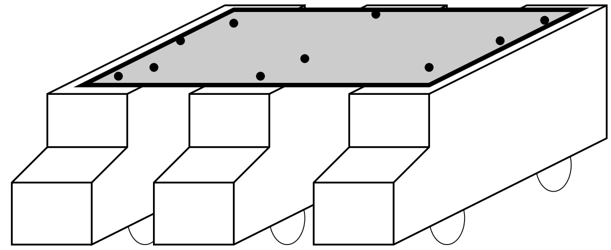

Fig. 1. Example of virtual surface. Points belonging to different portions of the physical surface are wrongly grouped together.

Segmentation by region growing suffers from similar problems. The process starts with a seed region and adds neighboring points as long as they fit a specified tolerance. Even if the points of the seed region are on a physical surface patch, points of other patches may be added, depending on the geometry and the adjustment procedure.

E. Surface descriptions

Fan [3] divides surface descriptions into two classes: global and segmented descriptions. We briefly comment on a few popular schemes in use.

1) Global surface descriptions: Digital Elevation Models (DEM) and Digital Surface Models (DSM) are undoubtedly the most widely used descriptions of surfaces in mapping and engineering applications. A DEM contains elevations of the topographic surface (bare earth) at regularly spaced grid posts. Since it is void of any objects with a vertical dimension it is not suitable for urban applications. In contrast to the DEM, a DSM includes elevations on the top surface of objects. If explicit information about surface discontinuities is added, such as breaklines (elevation discontinuities) and formlines (surface orientation discontinuities), the term DTM is used (Digital Terrain Model). Triangulated Irregular Nets (TIN) are popular for modeling surfaces that are sampled at arbitrary locations (e.g. by ALS systems). The major disadvantage of all these popular surface description methods is the lack of explicit surface shape information. Moreover, they cannot describe vertical surfaces nor multiple surfaces at the same location (e.g. bridge and surface below it).

2) Segmented surface descriptions: Here an attempt is made to approximate a set of spatially coherent 3D points by a mathematical surface, with a planar surface the most popular one. Many different approaches have been proposed in computer vision (segmenting range images) and mapping (segmenting laser point clouds). Fitting a 2D polynomial to a point cloud is trivial as long as the points belong to the same surface. The crux is to find these points prior to fitting and this requires the determination of surface discontinuities.

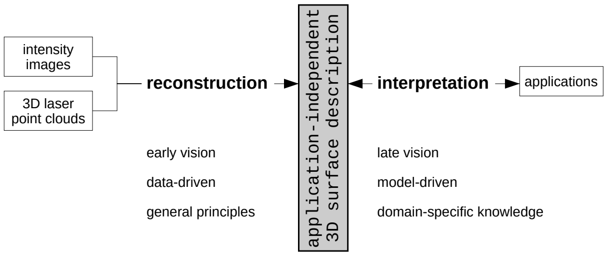

Fig. 2. Reconstructed surface as transition between early vision and late vision processes.

IV. Proposed application-independent surface reconstruction

A. Objective

Simply stated, the objective of our proposed application-independent surface reconstruction (AISR) scheme is to approximate (reconstruct) the physical surface from intensity images and 3D laser point clouds as faithfully and as rich in explicit surface information as possible, without the use of domain-specific knowledge. Sensors can be on airborne and/or terrestrial platforms (e.g. mobile mapping systems, terrestrial laser scanning systems). Imaging systems include multispectral/hyperspectral scanners.

The reconstruction is purely data-driven but the description is on highest possible level. AIRS is a platform where surface information extracted from different sensors is posted, combined, and analyzed. Only basic knowledge about the physical surface is allowed in the analysis whose major task is to resolve conflicts, identify ambiguities, and assign confidence levels to the surface description. In this sense, AISR will be the best explanation for data. As illustrated in Fig. 2, the reconstructed surface is the transition from early vision to late vision processes.

B. Reference system

The reconstruction is performed in a 3D Cartesian coordinate system with the $x, y-$plane tangent to the ellipsoid at a point within the project area. The positive $y-$axis points toward north, the $x-$axis toward east, and the $z-$axis completes the right-handed system. This choice of the coordinate system offers several advantages. First it provides a suitable, common reference for combining surface information extracted from the different sensors involved. Secondly it introduces intrinsic meaning to some surface patches. For example, a *horizontal* surface patch is on a equipotential surface, a *vertical* patch is aligned with the local vertical to the ellipsoid, and the azimuth of a surface boundary is the angle between north and the boundary. This intrinsic knowledge may be useful when it comes to the interpretation of AIRS.

C. Surface description

AIRS is the result of segmenting the sensed surface into spatially coherent regions (surface patches) that are bounded by surface discontinuities. The boundaries of the surface patches are represented as 3D polylines. The surface patch boundary must be closed.

The shape of the surface patches is made explicit by approximating it with a second degree polynomial. Higher degree polynomials are known for introducing artificial oscillations and should be avoided. Many segmentation methods are based on fitting planes. While planar surface patches are abundant in scenes that contain many man-made objects, the restriction to plane fitting may introduce problems for approximating topographic surfaces.

In case the fitting of a polynomial to a designated surface patch fails (e.g. fit error exceeds a threshold value), the volume of the point cloud and low order moments are computed and stored instead of the polynomial coefficients.

A key feature of AISR is the explicit representation of occluded surfaces. As elaborated in Sec. III-A, the physical surface is topological closed. If it were segmented into surface patches, then all the boundaries belong to two patches. That is, every patch of the physical surface is completely surrounded by other patches. If the sensed surface does not completely cover the physical surface (hardly possibly in urban scenes) then the completeness property does not hold. To satisfy it, small gaps are filled and occluded surfaces are added.

AISR also contains the results of classifying multi/hyperspectral images but this requires purely data-driven classification methods that do not require training, for example.

We summarize the surface description of AIRS as follows:

1) Boundaries of surface patches: 3D polylines with confidence level.
2) Surface patches:
   - coefficients of second order polynomial
   - fit error (surface roughness)
   - clusters of points not belonging to the fitted surface

   –or–

   - volume and moments of points within patch
3) Occluded surfaces (label, no surface properties).
4) Connectedness graph (direct neighbors of surface patches).
5) Class labels.
6) Confidence level.

D. Scale

In a strict sense, the scale at which the segmentation of the sensed surface is performed is application dependent. While some applications operate on a rather coarse representation, others may depend on as a fine resolution as possible. Since coarse representations can be obtained from a detailed one (but not the other way around), we perform the reconstruction at a detail level. The level of details is limited by two factors, however. For one, a surface patch must have a minimum number of surface points (e.g. laser points) for making the surface fitting process robust. Therefore the smallest surface patches are defined by the density of the laser point cloud. The sampling density of today's ALS systems may be as high

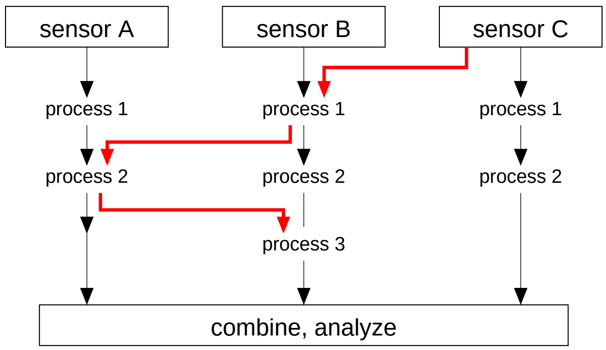

Fig. 3. Traditionally, processing sensor data is performed independently and extracted information is combined at a later step. As illustrated by red lines, some processes may benefit from information gained by other sensors.

as a few points per square meter. Terrestrial laser scanning systems have even higher point densities.

Discrete surface points can also be obtained from intensity images, for example by extracting and matching interest points. Obviously, the point density depends on the gray level distribution within a surface patch, and the occurance of corresponding points in overlapping images.

The boundaries of surface patches are determined by surface discontinuities (abrupt changes in surface orientation, boundary between sensed and occluded surface). Hence, boundaries are topographic properties of the physical surface. If these boundaries are obtained from intensity images then it is possible, and indeed recommended, to employ a scale-space approach, considering only boundaries that persist through a range of scales. In this fashion a scale is imposed on the segmentation.

V. Realization

This section describes some key elements of the realization of the application-independent surface reconstruction scheme. First and foremost the implementation is embedded in a fusion framework that exploits the synergism of multiple sensors to the maximum possible extent. Fig. 3 depicts the typical scenario of processing sensory input data independently from each other before the results are combined in a later step. Fusion can take place at an earlier stage, however. For example, some processes may benefit from intermediate results of other sensors. A point in case is the matching of extracted features from intensity images—a difficult and often ill-posed process, particularly if images widely differ in viewing geometry and radiometry. Matching in object space with the support of surface information obtained from the laser point cloud makes it a much more reliable process.

A. Stratgegy

The reconstruction of the surface from multiple sensors is performed in a fusion framework. Segmenting the sensed surface is driven by determining surface patch boundaries that are features of the physical surface. These boundaries are obtained from overlapping intensity images and independently checked with the point cloud. We pursue a strategy that determines first patches of high confidence before proceeding with patches of lesser confidence. Our rigid sensor orientation (sensor alignment) approach establishes a transformation between the sensor systems and global reference frame. This greatly facilitates the transformation of sensor information to the 3D reference system and vice versa.

B. Procedure

The major tasks of the surface reconstruction scheme include

- sensor registration
- boundary detection
- surface patch analysis
- detecting and adding occluded surfaces
- completeness check

C. Sensor registration

This first step is also known as *sensor alignment* or *sensor orientation*. It entails a transformation, forward and backward, between sensor data and reference frame. The point clouds originating from ALS systems are usually recorded in the WGS84 reference frame and it is straight forward to transform them to the proposed local project system (Sec. IV-B). The registration of intensity images requires the determination of the exterior orientation parameters. If a platform orientation system was used then the orientation parameters can be inferred from the navigation data. This data may not be precise enough or not available at all. Therefore we perform an independent orientation that will register intensity images to the 3D point cloud, based on sensor invariant features. Such features are present if the sensors respond to the same phenomena of the physical surface. For example, rapid changes in surface orientation may be extracted from the 3D point cloud by way of analyzing local surface curvatures. In intensity images, the same feature of the physical surface may be manifest as a gray-level discontinuity. Algorithms of feature-based photogrammetry permit the orientation of images with linear and 2D features (see [13] for more details).

D. Boundary detection

The boundaries of surface patches are distinct features of the physical surface, with discontinuities in the surface orientation the most prominent one. The boundaries are “visible” in intensity images, at least partially, and can be detected by an edge operator. However, it is very unlikely that the surface patch boundaries would neatly show up as closed contours in an image. Rather, only segments will be detected as well as many other edges that are not related to surface patch boundaries at all. The challenging problem of detecting surface patch boundaries is approached by the following steps:

1) Detect edges at multiple image scales. Only edges that occur over a range of scales will be considered.
2) Transform edges from all images to reference system.
3) Determine corresponding edges in the 3D reference system based on spatial proximity.

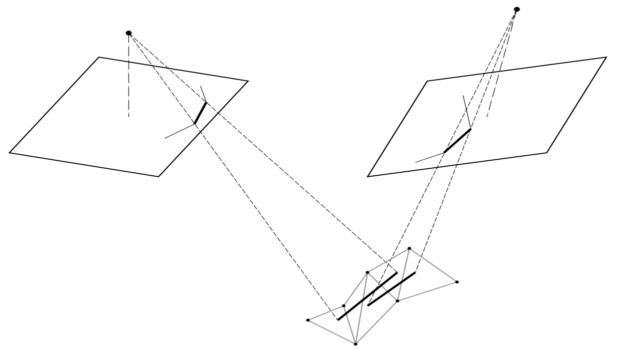

Fig. 4. Illustration of projecting image edges onto the 3D point cloud.

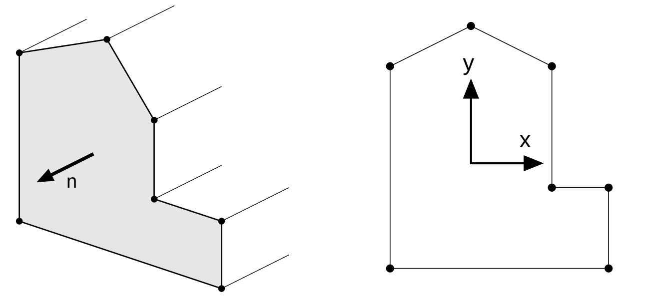

Fig. 5. Left panel shows boundary points and a fitted plane through the points with the plane normal $\mathbf{n}$. The right panel shows the local coordinate system and the transformed boundary points.

4) Edges that have been matched on multiple images (minimum two) in previous step will be intersected, resulting in 3D boundaries.
5) The 3D boundaries will be checked for closeness, small gaps will be filled. Closed boundaries are candidates for surface patch boundaries.

Edge detection at multiple scales is a standard procedure in computer vision. Increasing the scale increases the signal-to-noise ratio; edges that persist over several scales are more reliable and are more likely to correspond to features of the physical surface. To preserve the geometrical accuracy (localization) of edges obtained at larger scales, scale-space tracking is necessary.

The 2D image edges are projected to the 3D reference system by intersecting the projection rays with the 3D point cloud. As depicted in Fig. 4, the projection ray of an edge point (edge pixel or vertex of a segmented edge) is defined by the exterior orientation parameters of the image. This ray intersects the surface, for example the triangle of a TIN model of the discrete 3D laser point cloud.

Corresponding edges of several overlapping images must be clustered in the 3D reference system. Hence, finding corresponding edges amounts to analyzing line clusters. This approach has several advantages over the more traditional methods that seek correspondences in images. For one, edges from all images are simultaneously analyzed, thus supporting each other. This is yet another example of fusion that makes weak processes more robust. The transformation to the common reference system corrects the shape distortions of image edges induced by viewing geometry and topography. It is therefore possible to match edges from images with widely different viewing geometries (e.g. oblique aerial images)—an almost insurmountable problem when attempted in image space.

Once the correspondence of edges is established, the precise location in the 3D reference system is found by simultaneous, multiple intersection. Note that the resulting 3D edges are obtained solely from images, *independent* of the point cloud.

E. Surface patch analysis

1) Point selection: The selection of points of the 3D point cloud that are within a surface patch boundary appears trivial. Indeed this is true for horizontal surface patches and single layer surfaces. Here, we select all points that are within the surface patch boundary. But consider vertical surfaces, or several layers of horizontal surfaces (e.g. multiple overpasses), structures that occur frequently in urban scenes. The selection of points is now more involved and we proceed by fitting a plane through the boundary points, followed by establishing a rigid body transformation between the reference system and a local coordinate system that has its origin at the centroid of the boundary points and is oriented such that the $x, y-$plane is coplanar to the fitted plane. That is, the $z-$axis is parallel to the plane normal. A subset of the 3D point cloud (candidates of surface patch points) is transformed to the local surface patch coordinate system. Every point within the (transformed) boundary and within a distance threshold is now considered a patch point.

Fig. 5, left panel, shows boundary points and a fitted plane through the points with the plane normal $\mathbf{n}$. The right panel shows the local surface patch coordinate system and the transformed boundary points.

2) Surface fitting: To make the shape properties of a surface patch explicit, the patches are approximated by a second degree polynomial. Surface fitting is performed in the local patch coordinate system, introduced in the previous section. A robust estimation method is used because it is important to exclude points that do not belong to the fitted surface. Imagine a surface patch whose boundary delineates a parking lot. Some of the extracted patch points may have been reflected from small objects on the parking lot, for example, cars, shrubs, parking meters, and so on. Although small objects may well have generated edges in the images, they have been discarded (insignificant) during the scale-space edge detection approach.

Surface patch points that have been left out in the fitting process are further analyzed. If they form spatial clusters, the location and the number of points becomes a part of the surface patch description, together with the coefficients of the polynomial fit and the fit error.

If the the surface fitting process fails, for example fewer than half of the patch points are on a second degree surface, the volume of the convex hull of the patch points is computed and stored instead of the coefficients.

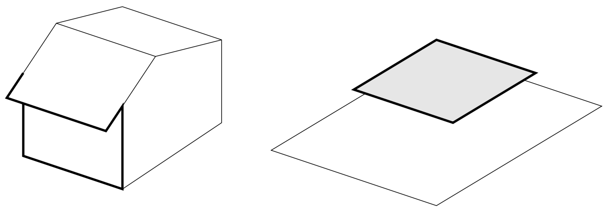

Fig. 6. Left panel shows surface patches that are only partially connected with another patch. The heavy lines mark boundary segments that have only one patch, but they are connected and define an occluded surface patch. The right panel shows a surface patch without any connections to other patches (free floating).

3) Spectral properties of surface patches: If multi or hyperspectral data is available then their image pixels within the surface patch are extracted and geometrically corrected with the surface parameters determined in the surface fitting process. The spectra of all patch pixels are now compared for similarity and an average spectrum is computed and stored together with the other surface patch description. In case the surface patch is spectrally not homogenous, the average of all populations is stored, but no attempt is made to subdivide the surface patch into spectrally homogenous regions.

After processing all surface patches, an unsupervised classification with the average spectra is performed and the class labels are added to the patches.

F. Detecting occluded surfaces

During the surface patch segmentation a list is generated that identifies for every segment of a patch boundary the neighboring surface patches. After processing all patches, the list is checked for boundary segments that connect only one patch—an indication for an occluded surface.

Fig. 6 shows two examples of occluded surfaces. The first case is simple since all boundary segments with one connected surface patch are connected and closed. Thus, an occluded surface is added and labeled accordingly. However, it consists of two planes (roof overhang and vertical wall), which violates the surface patch condition. This and the next example demonstrate that occluded surfaces are abstract quantities that are introduced to topologically close the surface segmentation.

The second case shows a surface patch whose entire boundary connects only one surface patch. Unlike in the previous case, there is no other single boundary in the neighborhood with which one can connect to form an occluded surface. The surface patch is free floating—a physical impossibility—but there is no clue from the data on how to connect. Free floating surface patches are labeled accordingly, but further interpretation is left to the application as this requires domain knowledge.

VI. CONCLUSION

We have presented the conceptual framework of an application-independent surface reconstruction scheme, suitable to deal with the complexities of urban scenes. Explicit surface information, such as shape and boundary of surface patches and spectral properties, is represented in a 3D object space coordinate system. This information-rich, abstract representation is obtained by fusing multi-sensor data, such as intensity and spectral images, and laser point clouds, without domain knowledge. The true 3D representation of surfaces copes with such complex situations as bridges above ground surfaces and multiple freeway overpasses.

The purpose of the surface reconstruction is to act as an intermediate representation for applications, such as bold-earth determination, generation of true orthophotos, autonomous vehicle navigation (terrestrial and airborne), object recognition, change detection, site modeling, and city modeling. The abstract description does not only make relevant surface information explicit but is in itself the most effective data compression method.

Future work will concentrate on the implementation and on the refinement of the design. Case studies with airborne and terrestrial sensors will demonstrate the feasibility of the proposed system, particularly for the generation of true 3D city models.

REFERENCES

[1] P.E. Axelsson, Processing of laser scanner data—algorithms and applications. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54(2-3), 138-147, 1999.

[2] C. Baillard, O. Dissard, O. Jamet and H. Maître, Extraction and textural characterization of above-ground areas from aerial stero pairs: A quality assessment. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54(2-3), 83-94, 1999.

[3] T.J. Fan, *Describing and recognizing 3-D Objects Using Surface Properties*. Springer-Verlag, 1990.

[4] M. Fradkin, M. Roux, H. Maître and U.M. Leloglu, Surface reconstruction from multiple aerial images in dense urban areas. In: *Proc. IEEE Conf. on Computer Visionamd Pattern Recognition*, vol. 2, Fort Collins, CO, USA, 262-267, 1999.

[5] L. Gabet, G. Giraudon and L. Renouard, Automatic generation of high resolution urban zone digital elevation models. *ISPRS Journal of Photogrammetry and Remote Sensing*, 52(1), 33-47, 1997.

[6] P. Gamba, B. Houshmand, Digital surface models and building extraction: a comparison of IFSAR and LIDAR data. *IEEE Trans. on Geoscience and Remote Sensing*, 38-2(4), 1959-1968, 2000.

[7] J. Hyyppa, O. Kelle, M. Lehikoinen and M. Inkinen, A segmentation-based method to retrieve stem volume estimates from 3D tree height models produced by laser scanners. *IEEE Trans. on Geoscience and Remote Sensing*, 39(5), 969-975, 2001.

[8] I. Lee, *Perceptual Organization of Surfaces*, Diss., The Ohio State University, Dept. of Civil and Environmental Eng. and Geodetic Science, 159 p., 2002.

[9] H.G. Maas and G. Vosselman, Two algorithms for extracting building models from raw laser altimetry data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54(2-3), 153-163, 1999.

[10] H. Maître and W. Luo, Using models to improve stereo reconstruction. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 14(2), 269-277, 1992.

[11] D. Marr, *Vision*. W.H. Freeman and Company, 1982.

[12] S. Sarkar and K.L. Boyer, Perceptual orgaqnization in computer vision: a review and a proposal for a classificatory structure. *IEEE Trans. on Systems, Man and Cybernetics*, 24(2), 246–267, 1993.

[13] T. Schenk, From point-based to feature-based aerial triangulation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58(2004), 315-329, 2004.

[14] H. Wiman, Automatic generation of of digital surface models through matching in object space. *Photogrammetric Record*, 16(91), 83-91, 1998.