IMAGE ORIENTATION FOR INTERACTIVE TOURS OF VIRTUALLY-MODELED SITES [1]

Charalampos Georgiadis, Anthony Stefanidis, Sotirios Gyftakis, Peggy Agouris

Dept. of Spatial Information Science & Engineering, University of Maine, 348 Boardman Hall, Orono, ME 04469-5711, USA – {harris, tony, sotiris, peggy}@spatial.maine.edu

Commission V, WG V/4

KEY WORDS: Automated Orientation, Ground-level Imagery, VR Sites

ABSTRACT:

This paper addresses the issue of image orientation for interactive tours. We consider situations where image sequences are captured in an urban site for which a virtual-model is already available. We present an approach that allows us to capture the orientation of these images using as control information building façades (instead of traditional points). Our approach allows us to orient image sequences relative to few absolutely oriented frames, devising an innovative version of dependent orientation for ground-level motion imagery. We present theoretical issues behind our approach and experimental results that demonstrate accuracies of 0.1 degrees in rotation estimation, and 0.15 meters in camera position determination.

1. INTRODUCTION

Close range photogrammetry is rapidly evolving, aided by advancements in sensor technology, wireless communications, and location-based computing. Whereas image capture and data processing used to be two clearly distinct processes, one performed in the field and the other in the lab, often separated by few days, these advancements now allow us to combine them sequentially in the field. The main premise behind location-based services is the knowledge of one's position (e.g. captured through GPS) and the access to a record of services available in this area (e.g. through a GIS). Thus, through a spatial query we can provide for example a user with a list and map of the ATM locations nearest to his/her present location.

Advancements in large-scale 3D virtual reconstruction of architectural sites, ranging from architectural monuments to regular urban environments are introducing an interesting extension to location-based services: instead of using a GIS as the reference database from which we recover the information of interest, we can use a VR model as the reference information. This offers a substantial advantage as the accessible information includes radiometry (e.g. façade images) in addition to the positional and thematic information that is typically found in a GIS. In this paper we address this emerging capability.

We consider situations where a user roams an area with a hand-held camera, and has wireless access to a VR model of this area. Our objective is to support the automated orientation of imagery collected at quasi-video rates by this user, using façade information from the VR database as control. Images are captured on the ground, depicting the façade of a building in an area of interest, and by accessing the VR database we can position the user's location. In doing so we use facades (instead of traditional control points) as the control information.

We can recognize a variety of applications that would make use of this capability, with intelligent VR-guided tourist tours as the most immediate: users equipped with handheld cameras move around, point at objects of interest, and their position is estimated, and used to provide them with relevant information, e.g. information about the building they point to. This would substitute traditional paper-based guides as a novel, interactive tour guide, with substantial potential to generate multimedia records from such visits (e.g. an annotated video of the tour). Beyond tourists, such an approach may also be used to support a variety of intelligent navigation applications, especially in environments where GPS access is limited (e.g. in urban neighbourhoods where tall structures limit GPS accessibility), and within the concept of geosensor networks [Stefanidis & Nittel, 2004]. Lastly, this work can be used to orient imagery which in turn can be used to create 3D virtual models of additional structures not included in our VR database.

In this paper we outline our approach and present some critical components of it, especially related to façade extraction and orientation recovery. The paper is organized as follows. In Section 2 we present relevant work on VR modelling. In Section 3 we give an overview of our approach, and follow with relevant details in Sections 4 and 5. Experimental results in Section 6 demonstrate the performance of our approach.

2. RELATED WORK

The 3D virtual reconstruction and visualization of building landscapes has received substantial emphasis during the last few years. We can recognize two clusters of relevant work:

- progress in complex building reconstruction, especially important for cultural monuments, and
- progress in the virtual reconstruction of large-scale urban environments, especially crucial for filed computing.

During the last few years we have had substantial progress in the 3D reconstruction of archaeological monuments. The state-

[1] This paper is a modified version of an upcoming publication to appear in the journal *Photogrammetric Engineering & Remote Sensing*

of-the-art approaches for the reconstruction of large-scale heritage and architectural sites use laser scanning complemented by close-range image captures [El Hakim et al, 2003,2004; Addison & Gaiani, 2000; Kanaya et all, 2001]. [Pollefeys et all, 2003] present a method for the reconstruction of architectural monuments using close range imagery from a hand held camera, while [Gruen et al, 2003] presented their experiments with the precise modeling of the complex statues of the Buddha of Bamiyan. [Georgiadis et al, 2000] present a method for on-site photogrammetric reconstruction techniques, while [Sechidis et al, 2001] developed a method for the creation of 3D georeferenced video for visualization and measurements.

Considering large-scale environments, substantial work has addressed the development of 3-D virtual models of large-scale complex urban scenes. Pioneering efforts include the development of Virtual Los Angeles, a large scale virtual model of the LA metropolitan area [Jepson et al., 1996], the virtual model of the city of Stuttgart in Germany, with more than 5000 buildings in an area of 2km x 3km [Haala & Brenner, 1999], and similar efforts in Bath, UK [Day et al., 1996] and Vienna, Austria [Ranzinger & Gleixner, 1997]. These models are typically generated using ground-level imagery and 3D terrain information derived from aerial sensors.

Regarding the use of ground level image sequences, [Zisserman et al, 1999] addressed the use of epipolar geometry and trifocal tensors to automatically retrieve 3d scene structure. [Baillard & Zisserman, 2000; Werner & Zisserman, 2002] use inter-image homographies to validate planar facets for the reconstruction of 3D piecewise planar elements from multiple images. [Cipolla & Boyer, 1998; Broadhurst & Cipolla, 1999] use vanishing points to recover interior orientation, rotation and translation.

Regarding image orientation itself, work relevant to our approach includes the efforts of [Chia et al, 2002] to compute the relative orientation of images using one or two reference frames and epipolar geometry, and the approach of [Simon & Berger, 2002] to estimate the orientation of an image using another image of known orientation and the planar homography between these two images. [Pollefeys et al, 2002] analyze points observed in the overlapping area of image sequences to determine the planes to which these points belong, and subsequently solve for the exterior orientation parameters.

The approach we introduce in this paper benefits from and complements work on the 3D virtual reconstruction of large spaces. Our orientation approach complements the above-mentioned existing orientation solutions to form a novel method that accommodates the particular needs of the roaming application mentioned in section 1. Our primary innovation is the development of a novel strategy to orient ground level image sequences using building facades as control information, as it is outlined in section 3 and presented in more detail in the subsequent sections 4 and 5.

3. APPROACH OVERVIEW

We address the orientation of sequences of ground-level still imagery captured at quasi-video by sensors roaming in an urban area. We assume a situation where few images in our sequence are already absolutely oriented and introduce a novel approach to orient all remaining in-between images relative to them, thus propagating orientation information in image sequences. For applications where a user is roaming an urban area and captures image sequences with a handheld camera, this entails orientation information available e.g. every few minutes or every few city blocks. The algorithm we present here propagates this information between these instances to derive orientation information for all intermediate frames. Our two-step approach is an innovative variation of the photogrammetric method of dependent orientation, to take advantage of advances in modern geospatial databases, especially wireless access and VR models.

Motivated by MPEG video compression, we proceed with a technique where few absolutely oriented frames serve as anchor frames (*A-frames*) to support the orientation of intermediate frames (*I-frames*) relative to them as visualized in Fig. 1. The orientation of individual A-frames may be determined using traditional photogrammetric techniques, or may be provided by GPS and gyroscope during image capture. Alternatively, we have presented in another publication an approach based on image queries to register imagery to a VR database and thus recover the absolute orientation of individual image frames [Georgiadis et al, 2002]. That technique makes use of approximate object location information (e.g. approximations on the order of 5-10 meters that may be provided by an inexpensive hand-held GPS, or even by associating a street address to camera location) and proceeds first with radiometric and geometric queries of individual objects [Stefanidis et al, 2003] and then with object configuration [Stefanidis A. et al, 2002] to register an image to a VR database.

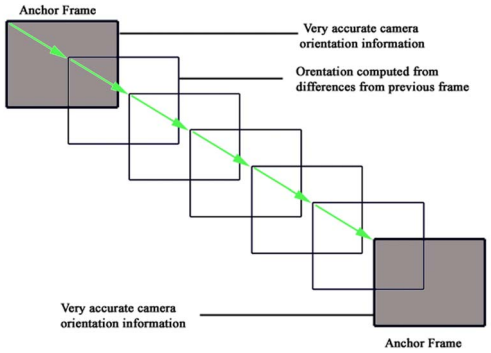

Figure 1: Proposed two-step approach scheme.

In this paper we focus on the development of a process and relevant algorithms to recover the orientation of I-frames by orienting them relative to the nearest A- or (already oriented) I-frame. Thus in the sequence of frames in Fig. 1 the orientation of the first I-frame will be accomplished by orienting it relative to its preceding A-frame. Then, the second I-frame will be oriented relative to the first I-frame, and this will continue until all I-frames are oriented. In this manner we devise a novel variant of the dependent orientation approach to propagate orientation information along sequences of photos, addressing close range motion image sequences: images are analyzed in pairs, where the first image has known parameters, and the second is oriented relative to it. This dependent orientation proceeds by analyzing differences in image content (location, rotation, and size of object facades in them) between the two images.

An outline of our approach with its major algorithmic components is shown in Fig. 2. We assume that blobs representing coarse approximations of building facades are available (e.g. provided by a user, extracted using coarse edge detection, or even predicted using information on sensor

movement). Using this input we proceed by precisely delineating these facades through an automated line fitting process, and identify their corners as intersections of these lines. Façade lines and corners are used to orient an I-frame to the nearest A- or I-frame through an innovative two-step orientation process. Rotation differences between two successive frames are computed first using vanishing points of planar elements in them (e.g. facades in our case). Then we estimate the shift and scale components of orientation by rectifying these façade elements and analyzing their differences. Our only requirement for this process is that differences in content among successive frames are relatively small so that they share some common façade components. This is a reasonable assumption when processing ground level motion imagery at quasi-video rates. In the next two sections we will present in detail the major theoretical issues behind our approach.

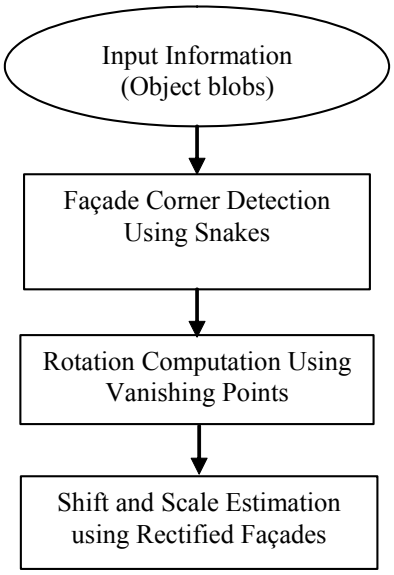

Figure 2: Approach outline for façade-based image orientation.

Our process is visualized in Fig. 3 where we see a portion of a 3-D VR model of an urban scene and two locations, A and B, from which we have captured two images. These two images are shown in Fig. 4, and we can see a common façade identified and delineated in them. The orientation of image A is assumed to be known (oriented directly as an A-frame, or indirectly as an I-frame in a previous sequence). Using the technique outlined in this paper we will determine sensor orientation for image B, thus propagating orientation from image A to image B. Façade information (lines, corners, and corresponding vanishing points) will be used for this orientation.

4. FAÇADE DELINEATION

4.1 Obtaining approximations

As mentioned in the previous section our approach makes use of approximate delineations of façade outlines in the form of blobs in their vicinity. Such approximations may be provided by a human operator (using e.g. an on-line display or other annotation techniques), or they may be determined using an automated approach. Here we present a Kalman-filtering-based approach as an example of how to obtain these approximations. As a person roams in an area with a camera we can identify two distinct trajectories:

- the trajectory of the camera carrier (blue line in Fig. 5); it is safe to assume that this movement is not random, but rather follows certain trends regarding its velocity and translation.
- the trajectory of the camera's view: in order to visualize the second trajectory we consider the intersection of a cylinder surrounding the trajectory by the optical axis (red line in Fig. 5).

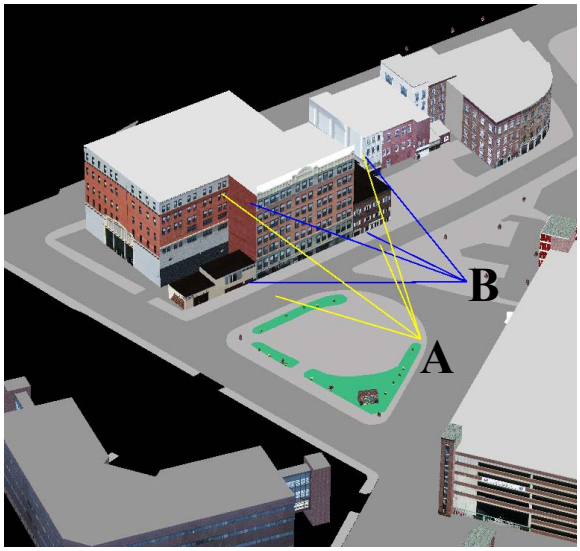

Figure 3: Portion of a 3-D virtual model showing the locations of two camera locations.

Figure 4: Images A (left) and B (right) with a common façade delineated in them.

We can then use Kalman filtering for the red and blue trajectories separately to predict their state. The reader is referred to [Maybech, 1979] for the classic formulation of Kalman filtering, and [Vasquez & Mayora, 1991; Roumeliotis & Bekey, 2000] for its use in robotics for mobile localization. Using a simple Kalman filtering technique based on a simple motion model we are able to predict the next location of the camera's trajectory and the projection of the optical axis in the cylinder, using as input the (X,Y,Z) coordinates of each trajectory. Finally we compute the predicted rotation angles.

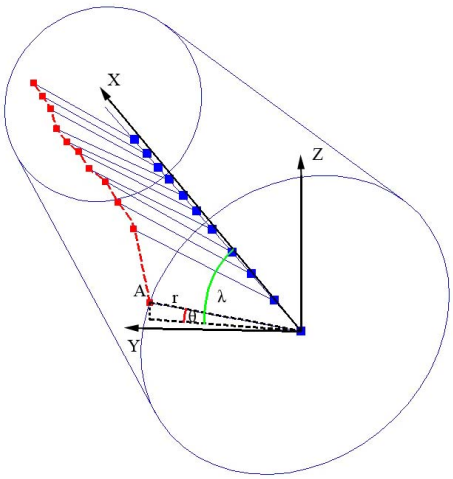

Figure 5: Trajectory of the camera carrier (blue dots) and of the camera view (red dots)

The computation of the two rotation angles from a camera station and the respective projection point A following the same notation shown in figure 5 with respect to the angles is done as follows:

$$\lambda = \arctan(\frac{Dy}{Dx}) \quad (1)$$

$$\theta = \arctan(\frac{Dz}{\sqrt{Dx^2 + Dy^2}}) \quad (2)$$

where the λ angle corresponds to the ϕ angle and the θ angle corresponds to the ω angle of our coordinate system, and Dx, Dy, and Dz are the coordinate differences between the predicted camera station and the predicted projection of the optical axis in the cylinder.

4.2 Façade Corner Estimation Using Snakes

As our approach determines orientation using vanishing points which in turn are estimated using corners of facades we use snakes to determine these corners. The snakes model is a controlled continuity spline under the influence of image forces as well as internal and external constraint forces [Kass *et al.*, 1988].

Snakes match a deformable model to an image object by means of energy minimization, which energy terms are included in the following active contour model:

$$E_{snake} = \alpha \cdot E_{cont} + \beta \cdot E_{curv} + \gamma \cdot E_{edge} \quad (3)$$

where E_{cont} is the continuity energy, E_{curv} is the curvature energy, E_{edge} is the photometric energy, and α, β, γ are the corresponding weights associated. The first two terms are the first and second order derivatives of the contour (internal forces), while the third term is the image gradient (external force) that attracts the contour to image edges.

For initialization of the snakes model we assume that we are given the corners of the façade as they are predicted from the Kalman filtering. Next, the snakes model interpolates more points between each corner point and the energy minimization starts. During optimization the snake points move to new locations according to the values of the continuity, curvature, and photometric energy terms of their neighbourhood points.

Next, we fit straight lines to the snake solution to precisely delineate the façade outlines. It is easily understood that we consider facades with linear edges in our applications a9rather than complex arcs etc.). Lines are fit in a least squares manner,

$$y = a_i x + b_i \quad (4)$$

and the façade corner points (X_{ij}, Y_{ij}) are estimated as intersections of these lines:

$$\hat{X}_{ij} = \frac{b_j - b_i}{a_i - a_j} \quad (5)$$

$$\hat{Y}_{ij} = a_i \hat{X}_{ij} + b_j \quad (6)$$

for $i=1,..,3$, $j=i+1$. For $i=4$, $j=1$.

5. ORIENTATION RECOVERY USING PLANAR ELEMENTS

The line fitting process presented in the previous section allows us to precisely delineate façades in the sequence of imagery that is being analyzed. In this section we present how this information can be used in an innovative orientation solution, orienting one image in our sequence relative to the previous image in the same sequence.

The problem we address is the orientation of two images using as conjugate feature the outline of the same façade in their overlapping area. The relative orientation solution would fail if all points belonged to the same plane (as is the case with our façade). In order to bypass this problem, the computer vision community makes use of a planar homography solution. However, this solution would provide us with a set of projective orientation parameters, but not the well-known exterior orientation parameters directly. In order to overcome this problem we introduce here a two-step process that allows first the direct computation of the rotation angles and then proceeds with the computation of translation parameters. Thus, our approach decomposes the orientation parameters of a camera into two sets, rotation and translation components, and determines them separately.

5.1 Rotations

For the computation of rotation angles we use vanishing points. They offer the great advantage of working even in situations where only a portion of a façade is visible, e.g. when a corner of the façade is outside the image frame but some portions of the corresponding sides are partially visible. In extreme situations it can also work when a complete side is missing, provided that we can identify entities (e.g. windows) that have edges parallel to the missing side.

In order to demonstrate our approach let us consider a local right-handed coordinate system with origin at the bottom left corner of the building façade. The X axis is along the width of the façade, the Y axis is perpendicular to the X axis along the vertical dimension of the façade, and the Z axis is perpendicular to the façade plane, pointing away from the building. In the image space, the x and y axes are parallel to their object space counterparts. The rotation angles are defined in the traditional photogrammetric manner: ω around the X axis, φ around Y, and κ around Z.

The vanishing points in the image domain are the intersections of image lines that correspond to elements that were parallel to the X and Y axes in the object space. We would use a two vanishing points configuration: a vertical vanishing point *VV(xV,yV*) is the intersection of lines that define the two vertical edges of the façade, while a horizontal vanishing point *VH(xH,yH)* is the intersection of the two lines that define the horizontal edges of the façade. In our approach vanishing points can be determined as intersections of corresponding façade lines. Regarding rotation recovery, based on the formulation of [Karras et al, 1993] and modifying them for our reference system we can use these two vanishing points to calculate the three rotation angles and focal length c as follows:

$$c = \sqrt{-x_H x_V - y_H y_V} \tag{7}$$

$$\tan\kappa = \frac{x_V}{y_V} \tag{8}$$

$$\tan\omega = \frac{c}{x_V \sin\kappa + y_V \cos\kappa} \tag{9}$$

$$\tan\phi = \frac{c}{\cos\omega(y_H \sin\kappa - x_H \cos\kappa)} \tag{10}$$

The computation of focal length using vanishing points is more sensitive to noise than the computation of angles, so in our approach we considered the camera to be calibrated and assume the focal length to be known. Using vanishing points in the above formulation allows us to estimate directly the rotation parameters at each location. Accordingly this eliminates the propagation of errors that we encounter in dependent relative orientation solutions. The accuracy of these rotation parameters is affected solely by the accuracy in our façade line fitting process.

5.2 Translation

For the computation of the relative translation parameters we present a process based on the rectification of a stereopair to façade-normalized views. This entails the rectification of our images to a plane parallel to the specific façade, thus simulating an image taken without rotations in the relative coordinate system that we defined in section 5.1. By creating these rotation-free versions of our images the computation of the relative translation parameters is trivial. In this section we present the process we follow to create these façade-normalized views.

We proceed by recognizing that the effects of rotation on our original imagery may be grouped into two different groups: one related to the effects a κ rotation, and the other related to the combined effect of the other two rotations (ϕ and ω). The difference between them is that κ rotation (unlike ϕ and ω) does not affect the actual shape of the object, just rotates its position in the image plane. From traditional photogrammetry we know that a ϕ rotation would result in an orthogonal façade (in the object space) imaged as a trapezoid with the vertical lines remaining parallel. An ω rotation on the other hand would result in an orthogonal façade imaged as a trapezoid with horizontal lines remaining parallel.

In order to rectify the image of a façade we have to re-project it onto an image plane parallel to the façade in the object space. Thus, we rectify a façade image first by a simple rotation to remove the effects of κ and then with a colinearity-based projection to accommodate the combined effect of ϕ and ωangles. It should be noted here that for our mobile orientation application we do not actually need to resample the complete façade, we simply have to correct its four corner coordinates. Thus, time requirements remain minimal for this task.

Through this process we can generate two façade-normalized views for the same object in two successive images (i and j). If the two images had the same scale (which would mean that they were taken from the same distance from the object), the relative translation along the Z axis would be zero. Furthermore we would be able to compute the translation along the two other axes by simply computing the difference between the corner points in the images and multiplying this difference by the image scale factor.

In Fig. 6 we can see the setup when we have two normalized views of different scale for the same object. The image scale can be defined as:

$$\text{Scale}_i = \frac{c_i}{H_i} = \frac{d_i}{D} \tag{11}$$

where c_i is the focal length of the camera, Hi is the perpendicular distance from the building façade, D is the length of a façade edge in reality, and di is the length of the same façade edge in image i.

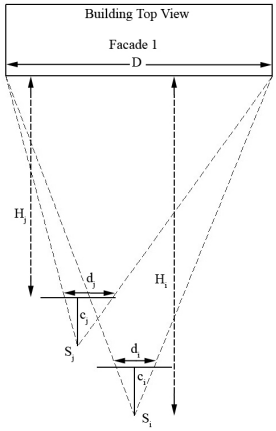

Figure 6: Image scale for rectification

The scale difference between these two images is directly estimated as the ratio of conjugate sides in them:

$$Scale_{ij} = \frac{d_i}{d_j} \tag{12}$$

Thus the distance Hj (in Fig. 6) along the Z axis between the second image (j) and the façade is given as function of the corresponding distance between the first image (i) and the façade as when the two images have the same focal length c:

$$\frac{d_i}{d_j} = \text{Scale}_{ij} \Leftrightarrow \frac{\frac{D \cdot c_i}{H_i}}{\frac{D \cdot c_j}{H_j}} = \text{Scale}_{ij} \Leftrightarrow \frac{H_i}{H_j} = \text{Scale}_{ij} \Leftrightarrow H_j = H_i \frac{1}{\text{Scale}_{ij}} \tag{13}$$

Using this equation we can compute the range of the second image knowing the range of the first image, the coordinates of the corners for a known line segment in both images, and the rotation angles of the images.

The next step is to apply a scale correction in image j so we can have the same scale in both images. This is a straightforward

procedure and it is accomplished by simply multiplying the corner point image coordinates with the scale factor Scaleij.
The remaining two shifts can be estimated through the following two equations:

$$DX = dx_{ij} * Scale_{i,object} \quad (14)$$

$$DY = dy_{ij} * Scale_{i,object} \quad (15)$$

where dx_{ij}, dy_{ij} are the coordinate differences in the image system after the scale correction, and $Scale_{i,object}$ is the scale between the first image and the object.

This set of equations, together with the rotation angles recovered using vanishing points express the orientation parameters of the second image relative to the first one. These parameters are defined in the local coordinate system. In a long sequence of images this would imply that all images would be oriented relative to each other, and the orientation parameters would refer to a coordinate system with origin at the lower left point of the first façade. If we want to express the same orientation parameters in an absolute coordinate system (in the traditional photogrammetric manner) we can transform this façade-centered system to an absolute one using the positional information available for the reference façade.

6. EXPERIMENTS

The method outlined in the previous sections has been implemented in Matlab, and its various components tested. Here we present sample experiments that demonstrate the performance of our line fitting and orientation estimation.

6.1 Corner Estimation through Snakes

In our experiments the size of the processed image is 640 x 480 pixels, in order to approximate frames captured with a camcorder. The images were acquired using a Minolta digital camera. The size of the CCD sensor of the camera is 6.6 x 8.8 mm which corresponds to a pixel size of 13.75 microns for a 640x480 pixels image.

We consider rather smooth changes in the orientation parameters over time, and thus the Kalman predicted outlines were of good accuracy. Errors in the predicted façade corner points were in the neighbourhood of 5-15 pixels. These initial approximations are refined using our snakes fitting. A representative experiment is shown in Fig. 7. The solution in this set-up is obtained through the solution of equation (1), using as weights: $\alpha = 2$, $\beta = 3$, and $\gamma = 4$. In the optimization part, the search is performed in a neighbourhood width of 15x15 pixels.

The resulting approximation of façades is displayed in Figure 7, the initial approximation (in red) together with the estimated one (in green). A zoom-in section of the estimated corner points (green squares) are displayed in Figure 8, together with the initial approximation (red triangles) and the actual corner points (blue circles). The actual corner points are located manually on the picture. The accuracies of the initial and estimated corner points are shown in Table 1. The results show that the snakes and line fitting improve the Kalman approximation to offer corner determination with an accuracy of 2-5 pixels.

Figure 7: First experiment (red: initial approximation, green: snakes and line fitting solution).

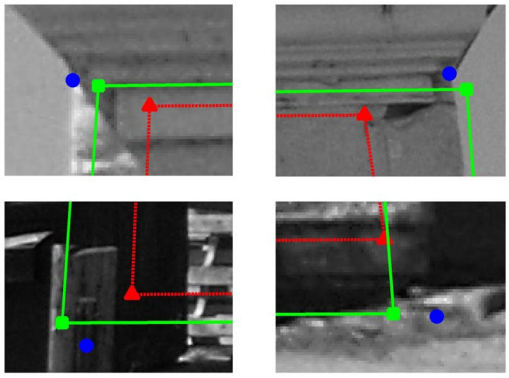

Figure 8: Corner points (up to bottom, left to right), circle: actual, triangle: initial, square: estimated.

Corner Points	Top Left (pixels)	Top Right (pixels)	Bottom Left (pixels)	Bottom Right (pixels)
Initial	4.31	6.37	12.04	8.11
Estimated	2.32	4.05	5.75	3.74
Difference	-46.23%	-36.41%	-52.24%	-53.90%

Table 1. Accuracy of corner points (experiment 1).

As shown in Table 1, the average accuracy of our corner determination is approximately 4 pixels, an improvement of more than 45% compared to the approximate values.

6.2 Orientation Estimation

In order to evaluate the accuracy of the orientation approach introduced in Section 5 we used a sequence of real images captured in an urban environment (University Campus). A VR model of this area was available to support our experiments. The length of the depicted trajectory is approximately 120 meters. Assuming that a human walks with an average velocity of approximately 1.2 m/sec, we captured images at a distance of 1-1.4 meters between successive stations, corresponding to a situation where images are captured at an approximate rate of 1 frame per second as the camera carrier walks within this area. Under this scenario the dataset processed here corresponds to approximately 2 minutes of roaming video in this area. We used traditional photogrammetric techniques and control information in this area to establish the orientation of each frame, and consider these values as the reference ones. We then used our approach to process the imagery, and compared the orientation

results to the reference values. The results of this comparison are shown in Figures 9 and 10.
The first graph (Fig. 9) shows the accuracy of the recovered rotation angles using vanishing points. The next graph (Fig. 10) shows the errors in the absolute position information in the three axes.

As we can see in Fig 9, the typical rotation errors using our approach remained below 0.4 degrees. More specifically, as shown in table 3, the mean error values in all 3 angles approach 0 degrees, indicating the lack of a bias, while the mean absolute error is 0.07 degrees for ω, 0.08 degrees for φ, and 0.10 degrees for κ.

As we can see in Fig 10, errors in the determination of camera coordinate differences between successive frames using our approach are typically below 0.3 meters. More specifically, as summarized in table 3, the mean error is practically 0, indicating the lack of bias. The mean absolute error was 0.06m in X, 0.6m in Y, and 0.11m in Z.

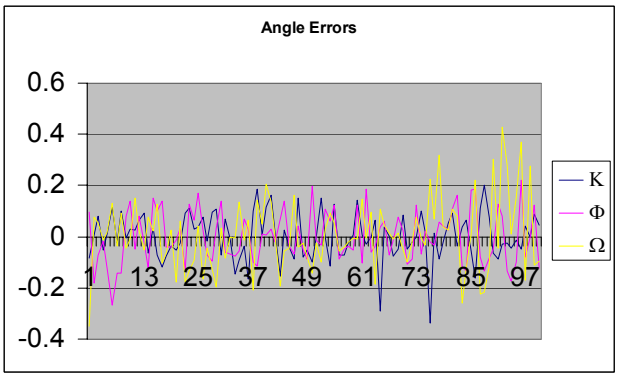

Figure 9: Recovered Angles Errors

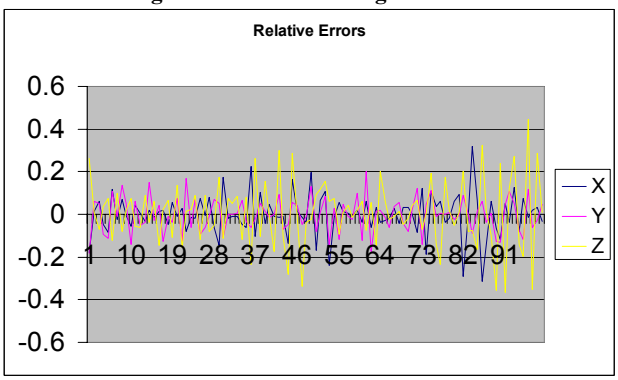

Figure 10: Recovered Coordinate Errors

	ω angle (deg)	φ angle (deg)	κ angle (deg)	ΔX (m)	ΔY (m)	ΔZ (m)
Mean error	0.0002	-0.012	0.007	0.002	-0.003	0.006
Mean absolute error	0.069	0.085	0.101	0.066	0.063	0.11

Table 2: Accuracy metrics

7. CONCLUSIONS

In this paper we presented a novel approach for the propagation of orientation information along sequences of ground-level imagery captured by a camera that is roaming through an urban scene. We use building facades as conjugate features for orientation determination, and employ vanishing points to determine the orientation parameters. We presented an approach to determine these vanishing points using coarse approximations of facades, and a methodology to orient images by determining separately their orientation and shift parameters. Experimental results demonstrated the performance of our techniques. Facades can be delineated precisely, allowing us to locate their corners with an average accuracy of 4 pixels. This allows us to recover orientation with accuracies on the order of 0.07-0.10 degrees in rotation estimation, and 0.06-0.11 meters in camera position determination. Thus our technique offers a promising alternative to traditional orientation schemes that involve time-consuming photogrammetric techniques, or expensive differential GPS, gyroscope, and INS systems.

ACKNOWLEDGEMENTS

This work was supported by the National Geospatial-Intelligence Agency (NMA 401-02-1-2008), and by NSF (ITR-0121269).

References

Addison A. C. and M. Gaiani Virtualized Architectural Heritage: New Tools and Techniques IEEE MultiMedia vol 07(2), 2000 pp. 26-31

Baillard C. & A. Zisserman, 2000. A Plane-Sweep Strategy for the 3D Reconstruction of Buildings from Multiple Images. Int. Archives of Photogrammetry & Remote Sensing, 33(B2), pp. 56–62.

Broadhurst A. & R. Cipolla, 1999. Calibration of Image Sequences for Model Visualisation. IEEE Conf. on Computer Vision and Pattern Recognition, Fort Collins, Vol. 1, pp. 100-105.

Chia K., A. Cheok & S. Prince 2002. Augmented Reality Registration from Natural Features, International Symposium on Mixed and Augmented Reality (ISMAR'02), Darmstadt, pp. 305-313.

Cipolla R. & E.G. Boyer, 1998. 3D Model Acquisition from Uncalibrated Images. Proc. IAPR Workshop on Machine Vision Applications, Chiba, pp. 559-568.

Day A., V. Bourdakis & J. Robson, 1996. Living with a Virtual City, Architectural Research Quarterly, Vol. 2, pp. 84-91.

El-Hakim, S.F., J.-A Beraldin, M. Picard Detailed 3D reconstruction of large-scale heritage sites with integrated techniques. IEEE Computer Graphics & Applications, May/June 2004, 23(3), pp. 21-29.

El-Hakim, S.F., J.-A Beraldin, M. Picard., A. Vettore Effective 3D Modeling of Heritage Sites 4th International Conference of 3D Imaging and Modeling (3DIM'03). Banff, Alberta, Canada. October 6-10, 2003. pp. 302-309.

Georgiadis Ch., A. Stefanidis, and P. Agouris, 2002 Fast Recovery of Image Orientation using Virtual Urban Models, International Archives of Photogrammetry & Remote Sensing, 34 (5), Corfu, pp. 161-166.

Georgiadis, Ch., V. Tsioukas, L. Sechidis, E. Stylianidis, and P. Patias, 2000, Fast and accurate documentation of archaeological sites using in-the-field photogrammetric techniques, Proc. and

CD of ISPRS XIX Congress, Supplement B5,Amsterdam, July 16-23, 2000.

Gruen A., F. Remondino, L. Zhang IEEE Conference on Computer Vision and Pattern Recognition (CVPR) - Workshop on Applications of Computer Vision in Archeology, 16-22 June 2003, Madison, Wisconsin, USA, on CD-ROM

Haala N. & C. Brenner, 1999. Virtual City Models from Laser Altimeter and 2D Map Data, Photogrammetric Engineering & Remote Sensing, 65(7), 787–795.

Jepson W., R. Liggett & S. Friedman, 1996. Virtual Modeling of Urban Environments, Presence, 5(1).

Kanaya, I., Kadobayashi, R., and Chihara, K. Three-dimensional modelling of Shofukuji burial chamber, Proceedings of the Seventh International Conference on Virtual Systems and Multimedia (VSMM2001), pp. 113-120 (October 2001).

Karras G., P. Patias & E. Petsa, 1993. Experiences with Rectification of Non-Metric Digital Images when Ground Control is not Available. Proc. of CIPA XV International Symposium, Bucharest, September 22-26.

Kass, M., A. Witkin, and D. Terzopoulos (1988). Snakes: Active contour models. *International Journal of Computer Vision*, **1**(4): 321-331.

Lawn J.M. & R. Cipolla, 1996. Reliable Extraction of the Camera Motion using Constraints on the Epipole, European Conf. on Computer Vision, Lecture Notes in Computer Science, Vol. 1065, Vol. 2, pp. 161-173.

Maybech, 1979. Stochastic Models, Estimation, and Control, Academic Press.

M. Pollefeys, L. Van Gool, M. Vergauwen, K. Cornelis, F. Verbiest, J. Tops, 3D Capture of Archaeology and Architecture with a Hand-Held Camera, The International Archive of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXIV, Part 5/W12 Ancona, pp. 262-267, 2003.

Pollefeys M., F. Verbiest & L. Van Gool, 2002. Surviving Dominant Planes in Uncalibrated Structure and Motion Recovery, European Conf. On Computer Vision, Lecture Notes in Computer Science, Vol. 2351, pp. 837-851.

Ranziger M. & G. Gleixner, 1997. GIS-Datasets for 3D Urban Planning, Computers, Environments & Urban Systems, 21(2), pp. 159-173.

Roumeliotis S. & G. Bekey, 2000. Bayesian Estimation and Kalman Filtering: A Unified Framework for Mobile Robot Localization, Proc. IEEE Int. Conference on Robotics and Automation, San Francisco, pp. 2985-2992.

Sechidis L., V. Tsioukas, P. Patias, “Geo-referenced 3D Video as visualization and measurement tool for Cultural Heritage”, International Archives of CIPA, vol. XVIII-2001, ISSN 0256-1840, pp. 293-299

Simon G. & M.-O Berger. 2002. Pose Estimation for Planar Structures. IEEE Computer Graphics and Applications, 22 (2), pp. 46-53.

Stefanidis A., P. Agouris, M. Bertoloto, J. Carswell, and Ch. Georgiadis, 2002. Scale- and Orientation-Invariant Scene Similarity Metrics for Image Queries, *International Journal of Geographical Information Science*, 16(8), pp. 749-772.

Stefanidis A., Ch Georgiadis., P. Agouris 2003. Registration of Urban Imagery to VR Models Through Radiometric Queries Videometrics VII, SPIE Proceedings Vol. 5013, Santa Clara, CA, pp. 176-185.

Stefanidis A. & S. Nittel, 2004. *GeoSensor Networks,* CRC Press, Orland, FL.

Vasquez R. & J. Mayora, 1991. Estimation of Motion and Position of a Rigid Object using a Sequence of Images, Proc. 19th Conference on Computer Science, San Antonio, pp. 540-550.

Werner T. & A. Zisserman, 2002. New Techniques for Automated Architectural Reconstruction from Photographs, European Conf. On Computer Vision, Lecture Notes in Computer Scince Vol. 2351, pp. 541-555.

Zisserman, A., A.W. Fitzgibbon & G. Cross, 1999. VHS to VRML: 3D Graphical Models from Video Sequences. IEEE International Conference on Multimedia & Systems, Florence, pp.51-57.