SUITABILITY ASSESSMENT OF GROUND LEVEL IMAGERY FOR GEOSPATIAL VR MODELING

Anthony Stefanidis   Charalampos Georgiadis   Giorgos Mountrakis

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

Commission V, WG V/4

KEY WORDS: Ground-level Imagery, VR Modeling, Suitability Ranking

ABSTRACT:

Improvements in sensor technology and modeling capabilities have resulted in a trend towards generating large-scale VR models of urban scenes, for use in geospatial applications. Selecting the appropriate imagery to generate VR models from a large collection of available ground-level photos is currently performed through visual inspection. In this paper we introduce well-defined metrics to support the automation of this time-consuming task. We make use of geometric and radiometric suitability metrics, describing image suitability for optimal image selection to use in VR modeling. In the paper we present these metrics, and demonstrate the applicability of our approach through experiments.

1. INTRODUCTION

Recent advancements in sensor availability and computing capabilities have fostered the development and management of large-scale virtual reality (VR) models of urban scenes. Such VR models are becoming increasingly popular in an expanding array of applications, and are expected to substitute traditional geospatial databases (e.g. maps and layers of geographic information systems – GIS) in the near future.

For the generation of massive scale VR models we usually rely on existing GIS information, aerial photogrammetry, airborne or ground-based laser scanning, GPS measurements, and classical surveying techniques to provide the geometric foundation (the equivalent of wireframe models of urban scenes). This information is then augmented by the radiometric content (building façade depictions) of ground level imagery, captured by mobile or static sensors in the area of interest. Processing ground level imagery for façade extraction remains a labor-intensive process, with human operators still heavily involved in it. Automating this process is a crucial task for the image processing community.

For many years now, the main focus of the photogrammetric community in VR applications was mostly on cultural and architectural landmarks (see e.g. [El-Hakim et al., 2004; Pollefeys et al, 2003] for recent relevant publications). Parallel to this work, we can recognize substantial efforts during the last few years towards the generation of VR models of broad urban areas. The group of Seth Teller at MIT worked on the automatic modeling of city areas using specialized equipment like the Argus camera and the Rover roaming platform to collect the required imagery [Coorg & Teller, 1999; Antone & Teller, 2000]. [Neumann et al., 2003] presented a method for integrating data from multiple sensors in real time for the creation of augmented virtual environments. Regarding texture content assessment for VR modeling, [Haala, 2004] presented an approach based on view geometry and image segment resolution, to evaluate suitability without considering the radiometric content of images.

In this paper we present our automated approach for the optimal building façade image selection for VR modeling of urban scenes. Its innovation lies on the simultaneous analysis of radiometric and geometric information. We introduce well-defined metrics and processes in order to evaluate the suitability of extracted facades by considering the corresponding scale and orientation parameters, and radiometric statistics of the images. Furthermore, through its integration with suitability assessment functions we allow users to set their own parameters in this process, thus customizing the results to meet their own patterns of preference.

Our approach is especially suitable for processing of imagery captured at quasi-video rate by ground level sensors roaming a scene, due to the increased availability of potential façade imagery. The structure of the remaining parts of our paper is: section 2 offers an approach overview, in section 3 we analyze the suitability assessment metrics, and section 4 presents experiments leading to our concluding remarks (section 5).

2. APPROACH OVERVIEW

The challenge we address in this paper is to select the best frame to generate the VR model of a façade when we have available numerous frames that depict this facade in various orientation conditions. We make use of suitability metrics that express the degree of each frame's appropriateness, and by quantifying this expression we allow ranking the available frames, from best to worst, for VR modeling purposes.
We assume to have available ground level imagery, with façade outlines approximately delineated in them. Our frame selection approach is based on the analysis of appropriate *radiometric* and *geometric* image characteristics, as visualized in fig. 1. The criteria we consider are as follows:
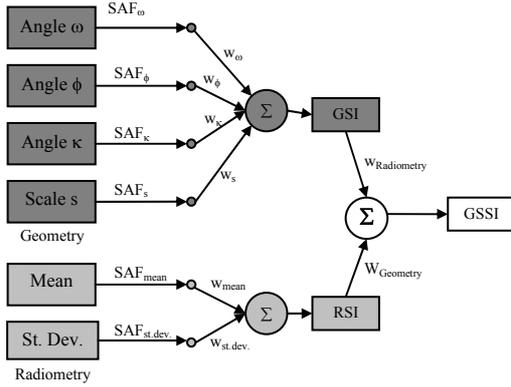
Figure 1: Integrating criteria for suitability

i) Geometric criteria. Regarding geometry, we aim to identify facades that are imaged nearly perpendicularly in the corresponding photo, and at maximum scale (i.e. finer resolution). The objects we are considering (i.e. building facades) are mainly planar surfaces. In order to evaluate the geometric suitability of each facade we make use of expressions of its image location as they are conveyed through the three rotation angles (ω, ϕ, κ) and the image scale (s).

ii) Radiometric criteria. Regarding radiometry, the objective is to select facades that are free of occlusions, and have nearly constant illumination throughout. In order to evaluate this, we analyze the chromaticity properties of each image. More specifically we consider the mean value of lightness in an image and the corresponding standard deviation of gray values within a façade rectangle. In our approach we use various color spaces. An optimal facade is one with mean lightness value close to a predefined target value to ensure uniform illumination throughout facades in the VR model, and minimal standard deviation for uniform illumination within each facade.

These criteria are formulated as *suitability assessment functions* (SAF), expressing the effects of parameter variations on the suitability of the corresponding image for use in VR modeling. A weighted aggregation of the SAFs values produces a Geometric Suitability Index (GSI) and a Radiometric Suitability Index (RSI), respectively. At the last processing stage, we combine the radiometric and geometric suitability indices into the final GeoSpectral Suitability Index (GSSI) expressing the overall image suitability.

3. IMAGE SUITABILITY ASSESSMENT

Suitability assessment takes place in two stages, first calculating the individual geometric and radiometric metrics, and then aggregating them into one total metric.

Geometric suitability of imagery for VR modeling is expressed as a function of the corresponding aforementioned geometric parameters (the three rotations ω, ϕ, κ and the single scale parameter s). More specifically, for each parameter individually, we use a suitability assessment function (SAF), expressing the suitability of the corresponding image parameter for use in VR modeling. Images that are suitable receive high SAF values (close to 1), while images with low suitability receive decreased SAF values (close to 0).

Considering the scale parameter for example, as the objective of geometric analysis is to select imagery with the best possible resolution, a SAF value of 1 is assigned to the image with the best resolution (façade covers the complete sensor array) and SAF values decrease linearly until they reach a value of 0 for images where the façade is depicted as a single pixel. Regarding rotation, the objective is to select imagery where the two rotation angles ω, ϕ are minimal, thus ensuring nearly parallel depictions of the object of interest. For these two rotation angles SAF values drop as we move from this ideal towards oblique views (angles approaching 90°). In terms of rotation κ, a primary goal is to select imagery where the object is depicted at minimal rotation angles, as this would minimize the need for resampling (and associated errors) when using this image to generate a VR model. Similar to the other SAFs, values drop as we deviate from this ideal and approach an extreme rotation of 45°.

This is communicated through SAF graphs like the one in Fig. 2. The vertical (Y) axis of these graphs expresses the degree of suitability, ranging from 1 for an ideal situation to 0 for a highly unsuitable one, while the horizontal (X) axis expresses variations of the corresponding geometric parameter (ϕ angle in Fig. 2).

The shape of the SAF functions may be chosen to reflect the effects of the above-mentioned parameters. Considering scale for example, its effects are linear, and thus can be expressed through a linear function (similar to the straight line in fig. 2). On the other hand, the effects of rotation angles are more complex in their nature (as expressed through the rotation matrix), and thus may be better expressed through more complex functions. An excellent example of such a function is the sigmoidal (dashed line in fig. 2). The sigmoidal offers the advantage of being highly suitable to express very efficiently non-linear user preference patterns, where suitability remains high as long as the actual parameter deviates slightly from the ideal, and then reduces rapidly as soon as we move beyond this range (e.g. 35 degrees in Fig. 2). A sigmoidal function for input X is expressed as:

$$SAF_{Sigmoidal}(X) = \frac{1}{1 + e^{-a(X-c)}} \qquad (1)$$

The c parameter specifies the function's translation along the X axis. The steepness of the slope of the sigmoidal is expressed through parameter a. An important characteristic of the sigmoidal function is the large range of modeling capabilities. Efficient manipulation of the slope can result in representing a variety of cases, ranging from linear to step-like behavior. The diverse capabilities together with the large operational range in the input space and the mathematical continuity of the function (first derivative exists everywhere) establishes the sigmoidal as appropriate solution from a variety of fuzzy functions.

As both approaches (linear and sigmoidal) are monotonic they do not differ in the ranking of numerous frames based on a single geometric parameter, but rather on the actual suitability value that is assessed to a particular frame. This can become quite important when integrating numerous geometric parameters (and corresponding graphs) to a single suitability metric. The actual SAF values may be pre-selected, using expressions like equation 1, or they may be learned through a relevance feedback process (Mountrakis & Agouris 2003; Mountrakis et all 2003).

The four SAF values are integrated in a single *geometric suitability index (GSI)* as the weighted average of the SAF values:

$$GSI = w_sSAF_s + w_\phi SAF_\phi + w_\omega SAF_\omega + w_\kappa SAF_\kappa \quad (2)$$

The weight coefficients w_i in eq. 2 may be selected heuristically, or they may be based on user responses (whereby a user may assign higher weight to the effects of a single parameter as an application may dictate). In our experiments these weight coefficients are selected to be 0.2 for the scale, 0.35 for each of the ϕ, ω angle, and 0.1 for the κ angle. The overall GSI value will be in the range [0,1] similar to the range of each SAF value, as $\Sigma w_i = 1$. Higher GSI values reflect higher suitability of the corresponding imagery for use in VR modeling.

Radiometric suitability assessment involves the analysis of radiometric content in three color spaces: HSV, YIQ and CIELab. The objective is to select imagery with minimal illumination variations, thus representing non-distorted views of the facade of interest. In order to evaluate the degree to which an image satisfies this condition we consider the color parameters of each façade rectangle to identify one that has a mean value close to a predefined target lightness value, and minimal standard deviation.
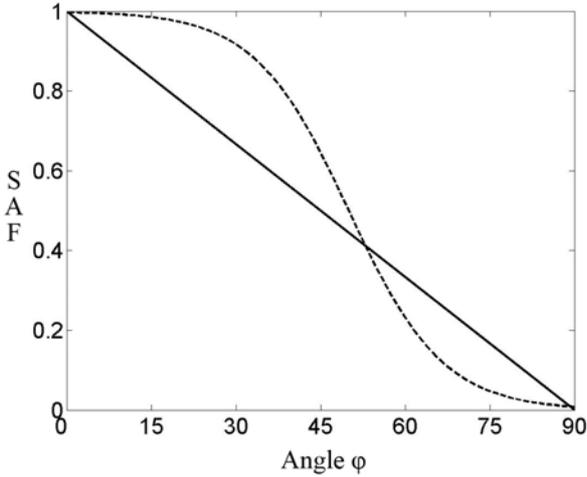

Figure 2: SAF assignment for angle ϕ

The *suitability assessment function* (SAF) values for these parameters is estimated by normalizing the SAF value range [0,1] to [highest, lowest] standard deviation value range for the $SAF_{st.dev.}$. In the case SAF_{mean} we use a symmetric suitability function where the value of 1 corresponds to the predefined target value and decreases gradually to zero. Linear or sigmoidal functions can be used based on preference requirements, following the above-mentioned geometric SAFs paradigm. The weighted average of the two radiometric SAF functions is the radiometric suitability index (RSI):

$$RSI = w_{mean}SAF_{mean} + w_{st.dev.}SAF_{st.dev.} \quad (3)$$

In our applications $w_{mean}=0.25$ and $w_{st.dev.}=0.75$, as we are interested in uniform illumination *within* the façade, which translates in standard deviation values being more important than having the mean close to the pre-defined target. These radiometric indices may be evaluated in either of the aforementioned color spaces independently, or use an average of their results as we see in the experiments section.

The geometric and radiometric SAFs are integrated in a single **GeoSpectral Suitability** index (GSSI), calculated as the weighted aggregation of GSI and RSI:

$$GSSI = w_{GSI}GSI + w_{RSI}RSI \quad (4)$$

In our applications we assign equal weight to each component ($w_{GSI}=w_{RSI}=0.5$), but these values may be altered as necessary. Depending on user preferences higher weight can be given to the geometric part (e.g. when images are used for 3d modeling or scene reconstruction) or the radiometric part (e.g. to eliminate shadows from our VR model).

4. EXPERIMENTS

In order to demonstrate the performance of our approach we present here experiments with synthetic and real datasets. Firstly, in order to evaluate the effects of illumination variations in the *radiometric* SAF values we use the dataset of Fig. 3. The top left image (1p) is the original image. All others are artificially created by altering the original image to simulate the effects of various illumination conditions and occlusions. We generated the various versions (1a-1n) by altering the radiometric content either:

- locally (second row) through the introduction of artificial occlusions and/or shadows, or
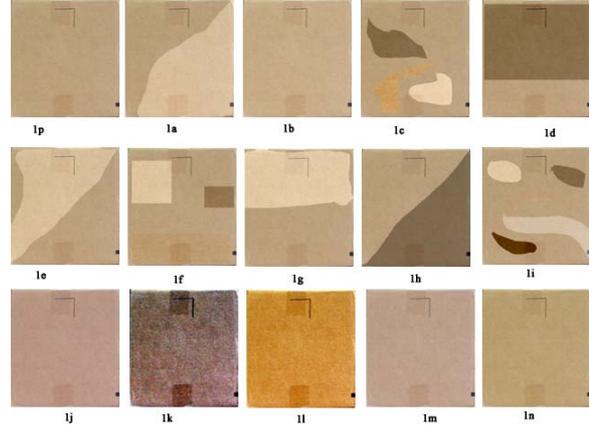- globally (third row).

Figure 3: Indoor Synthetic Dataset

We applied our linear metric in three different color spaces and integrated the mean and standard deviation SAF values as presented in the previous section. For our RSI metric we used the average value of the results in all three color spaces. As we can see in table 1 all three color spaces give similar results for the top four and last four candidates, with minor variations in the intermediate ordering. Thus ranking images radiometrically using only one of the three color spaces is adequate for our purposes.

In order to demonstrate the evaluation of *geometric* suitability we used three different datasets. We captured several images of building facades in our Campus and present here three representative samples. For the first façade we captured 10 different images depicting it, for the second façade we captured 9 images, and for the third façade we captured 11 images. In figure 4 we can see the images of the first dataset, in figure 5 the images of the second dataset and in figure 6 the images of the third dataset.

	YIQ		HSV		CIELab		AVERAGE	
		Rank		Rank		Rank		Rank
1b	0.97	1	0.97	1	0.98	1	0.98	1
1n	0.93	2	0.94	2	0.94	2	0.94	2
1p	0.93	3	0.93	4	0.93	3	0.93	3
1m	0.92	4	0.93	5	0.93	4	0.93	4
1j	0.91	5	0.93	3	0.9	5	0.91	5
1e	0.68	6	0.7	7	0.81	6	0.73	6
1a	0.64	7	0.66	8	0.8	7	0.7	7
1l	0.59	8	0.73	6	0.53	11	0.62	9
1f	0.57	9	0.61	9	0.72	8	0.64	8
1g	0.51	10	0.56	10	0.71	9	0.59	10
1c	0.37	11	0.39	11	0.56	10	0.44	11
1i	0.18	12	0.31	12	0.34	12	0.28	12
1d	0.16	13	0.19	13	0.3	13	0.22	13
1h	0.1	14	0.08	14	0.29	14	0.15	14
1k	0.06	15	0.03	15	0	15	0.03	15

Table 1: Radiometry Suitability Results

These images were captured at a camera-façade distance ranging between 5 and 15 meters for these datasets. The images (3072x2048 pixels) were captured using a Canon Digital Rebel uncalibrated camera. The camera pixel size was 7.4 microns and since the camera is uncalibrated we used as focal length the nominal value provided by the manufacturer.

Using vanishing points we computed the three rotation angles for each view, and then we selected the image with the lower distortions (rotation angles) to serve as a reference frame for the computation of the relative scale between the images.

In order to consider delineation errors in the assessment of suitability using our metrics we created three error-affected versions of each original façade image by introducing random errors in the corresponding corner determination. The magnitude of these errors ranged from a maximum of 2 pixels for version 2 to a maximum of 5 pixels for version 3, and 10 pixels for the fourth version. We used the results of these three versions in addition to the original errorless results (version 1).

Errors in corner determination affect the accuracy with which we compute the corresponding rotation angles and scale parameters; and accordingly may affect the evaluation of our suitability metrics. By considering all four versions of our imagery we investigated the robustness of our metrics to such errors.

Figure 4: Facade 1 Dataset

Image	GSI	GSI 2p	GSI 5p	GSI 10p
1	0.7254	0.7258	0.7189	0.7309
2	0.7617	0.7639	0.7563	0.7658
3	0.8064	0.8076	0.8106	0.8104
4	**0.8668**	**0.8710**	**0.8711**	**0.8807**
5	0.8437	0.8460	0.8421	0.8512
6	0.7920	0.7938	0.7835	0.7948
7	0.7627	0.7633	0.7635	0.7761
8	0.7380	0.7388	0.7340	0.7457
9	0.7209	0.7245	0.7169	0.7244
10	0.7622	0.7653	0.7567	0.7665

Table 2: Dataset 1 Results

The results of our metric for façade 1 are presented in Table 2. The images of fig. 4 are numbered left-to-right, top-to-bottom (i.e. image 4 is the fourth image in the first row, 7 is the second in the second row). The façade of interest is marked by a red box in image 3 of Fig. 4. Using our metrics we can see that image 4 is ranked as best for this façade. Visual inspection verifies the suitability of this façade for VR rendering. It is also interesting to observe that image 4 was ranked first in all 3 error-affected versions (columns headed GSI 2p to indicate error of 2 pixels, GSI 5p to indicate error of 5 pixels, and GSI 10p to indicate error of 10 pixels). Thus our metrics display robustness with respect to delineation errors.

Figure 5: Façade 2 Dataset

Name	GSI	GSI 2p	GSI 5p	GSI 10p
1	0.7056	0.6910	0.6797	0.6643
2	0.6625	0.6476	0.6363	0.6198
3	0.6709	0.6665	0.6526	0.6566
4	0.6433	0.6348	0.6433	0.5819
5	0.6694	0.6639	0.6616	0.6595
6	0.6930	0.6888	0.6730	0.6697
7	0.7481	0.7425	0.7436	0.7270
8	**0.7755**	**0.7677**	**0.7737**	**0.7434**
9	0.7245	0.7065	0.7384	0.6407

Table 3: Dataset 2 Results

We performed similar experiments with two other facades, with the corresponding datasets shown in Figs. 5 and 6, with the façade of interest delineated by a red box in one image in each of these datasets. The corresponding ranked results are tabulated in Tables 3 and 4 respectively. Visual inspection verifies the suitability of the best ranked images. Furthermore, we see that the best candidate remains ranked at the top in the presence of noise for these facades as well. The only exception is for façade 3, where the best match (image 6) drops to second

choice, as it is overtaken by image 4 in the presence of gross delineation errors (10 pixel error). Nevertheless, the difference in suitability between images 4 and 6 for façade 3 was minimal, so this minor reversal does not affect the validity of our approach. These representative experiments demonstrate the overall robustness of our approach even in the presence of delineation errors.

Figure 6: Façade 3 Dataset

Name	GSI	GSI 2p	GSI 5p	GSI 10p
1	0.6733	0.6777	0.6688	0.6916
2	0.6905	0.6926	0.6895	0.7100
3	0.7424	0.7445	0.7402	0.7594
4	0.8011	0.8026	0.8014	**0.8206**
5	0.7897	0.7920	0.7898	0.8097
6	**0.8039**	**0.8045**	**0.8033**	0.8145
7	0.7343	0.7358	0.7349	0.7591
8	0.6996	0.7021	0.6967	0.7185
9	0.6650	0.6676	0.6623	0.6928
10	0.6181	0.6211	0.6152	0.6307
11	0.6653	0.6674	0.6625	0.6936

Table 4: Dataset 3 Results

Regarding time performance we should point out that the most time consuming components of our suitability evaluation process are object delineation and the computation of mean and standard deviation values for each façade rectangle. In the worst scenario (using a full resolution image and a façade covering 80% of a frame), calculations took 1.22 sec per image to run on a Pentium4 2.8 GHz. This is perfectly acceptable for off-line image processing, and can even meet real-time in-field analysis when images are captured at a rate of approximately 1 frame per second.

5. CONCLUSIONS

In this paper we addressed the optimal selection of close range imagery for VR modeling within geospatial environments. The inputs for our analysis are geometric and radiometric properties like scale, the three rotation angles, brightness, and illumination variations. We introduced a set of suitability assessment functions to express image fitness for the task at hand. More specifically, the flexibility of SAFs ranging from linear to complex sigmoidal functions, together with our weighting scheme, support advanced customization of the results.

Experiments demonstrated the robustness of radiometric ranking in various color spaces. Another important finding is that ranking images according to our geometric suitability method does not require the precise delineation of objects in all datasets. Using imprecise object blobs is adequate for the correct ranking of candidate imagery. To summarize, our metrics allow for an automated ranking of incoming imagery, thus eliminating the need for visual inspection and selection. By automating this process, we alleviate a cumbersome task, and enable the exploitation of large datasets of imagery to support VR modeling.

References

Antone M. & Teller S., 2000. Automatic Recovery of Relative Camera Rotations for Urban Scenes, *CVPR* II: 282-289.

Coorg S. & Teller S., 1999. Extracting Textured Vertical Facades from Controlled Close-Range Imagery, *CVPR,* 625-632.

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.

Haala N. 2004. On The Refinement of Urban Models by Terrestrial Data Collection, *Intern. Arch. of Photogrammetry & Remote Sensin*g, 35(B3): 564-569.

Neumann U., You S, Hu J., Jiang B., Lee J. W., 2003. Augmented Virtual Environments (AVE): Dynamic Fusion of Imagery and 3D Models, *IEEE Virtual Reality'03*, 61-67.

Mountrakis G., Agouris P., 2003. Learning Similarity with Fuzzy Functions of Adaptable Complexity, *Symp. on Spatial and Temporal Databases (SSTD)*, LNCS Vol. 2750, 412-429.

Mountrakis G., P. Agouris, A. Stefanidis, 2003. Multitemporal Geospatial Query Grouping Using Correlation Signatures, *IEEE ICIP'03,* Barcelona, Vol. III, 545-548.

Pollefeys M., L. Van Gool, M. Vergauwen, K. Cornelis, F. Verbiest, J. Tops, 3D Capture of Archaeology and Architecture with a Hand-Held Camera, IAPRS, Vol. XXXIV, Part 5/W12 Ancona, pp. 262-267, 2003