

AUTOMATIC ADJUSTMENT OF WIDE-BASE GOOGLE STREET VIEW PANORAMAS

E. Boussias-Alexakis^a, V. Tsironis^{a*}, E. Petsa^b, G. Karras^a

^a Laboratory of Photogrammetry, Department of Surveying, National Technical University of Athens, GR-15780 Athens, Greece
(bousias.alexakis@gmail.com, tsironisbi@me.com, gkarras@central.ntua.gr)

^b Laboratory of Photogrammetry, Department of Civil Engineering and Surveying & Geoinformatics Engineering, Technological Educational Institute of Athens, GR-12210 Athens, Greece (petsa@teiath.gr)

Commission I, Working Group ICWG I/Va

KEY WORDS: panoramas, wide baseline, bundle adjustment, image matching, cartographic projection, projective transformation

ABSTRACT:

This paper focuses on the issue of sparse matching in cases of extremely wide-base panoramic images such as those acquired by Google Street View in narrow urban streets. In order to effectively use affine point operators for bundle adjustment, panoramas must be suitably rectified to simulate affinity. To this end, a custom piecewise planar projection (triangular prism projection) is applied. On the assumption that the image baselines run parallel to the street façades, the estimated locations of the vanishing lines of the façade plane allow effectively removing projectivity and applying the ASIFT point operator on panorama pairs. Results from comparisons with multi-panorama adjustment, based on manually measured image points, and ground truth indicate that such an approach, if further elaborated, may well provide a realistic answer to the matching problem in the case of demanding panorama configurations.

1. INTRODUCTION

Thanks to their obvious advantages, spherical panoramic images represent today an increasingly common type of imagery. They provide an omnidirectional field of view, thus potentially reducing the number of required images and also providing far more comprehensive views. They may be generated in various ways, yet it is today rather easy to produce panoramas with low cost-equipment and use of freely available software for automatically stitching together homocentric images onto a sphere, and subsequently mapping them in suitable cartographic projections (Szeliski & Shum, 1997, Szeliski, 2006). Spherical panoramas are thus being exploited in several contexts, including indoor navigation, virtual reality applications and, notably, cultural heritage documentation, where the use of panoramas is now regarded as a ‘natural extension of the standard perspective images’ (Pagani et al., 2011).

Of course, most important is the availability of street-level panoramas, such as those provided by Google. Its popular service Google Street View (GSV) is a vast dataset with regularly updated, geo-tagged panoramic views of most main streets and roads in several parts of the world, typically acquired at a frequency of ~12 m by camera clusters mounted on moving vehicles. Application areas of such pictorial information range, for instance, from space intersection (Tsai & Chang, 2013) to image-based modeling (Torii et al., 2009; Ventura & Höllerer, 2013), vision-based assistance systems (Salmen et al., 2012) and localization or trajectory estimation of a moving camera (Taneja et al., 2014; Agarwal et al., 2015).

A central question regarding the metric exploitation of panoramas is their registration (bundle adjustment). Due to its omnidirectional nature, a spherical panorama has the properties of a sphere, i.e. it defines a bundle of 3D rays. In this sense, the issue

of “interior orientation” (camera geometry) appears in this case to be irrelevant. However, the particular cartographic projection of the panorama on which image measurements will take place must of course be known; this projection in fact represents the interior orientation of a panorama (Tsironis, 2015). Panoramas in a known projection each have, therefore, 6 degrees of freedom. If no ground control is available, the 7 parameters of a 3D similarity transformation need to be fixed.

Thus, for instance, Aly & Bouguet (2012) adjust unordered sets of spherical panoramas to estimate their relative pose up to a global scale. Of course, several simplifications are possible if camera movement is assumed to be somehow constrained (e.g. in Fangi, 2015, small angles are assumed).

A crucial related issue is, of course, automatic point extraction, description and matching. Although spherical operators have indeed been suggested (see Hansen et al., 2010; Cruz-Mota et al., 2012), practically all researchers rely on standard planar point operators such as SIFT, SURF and ASIFT. Several alternatives have been reported. Agarwal et al. (2015) thus use conventional frames (provided by Google when requested for input from a virtual camera) and match them via SIFT to the image sequence. Mi ušík & Košecká (2009) and Zamir & Shah (2011), on the other hand, employ rectilinear (cubic) projections and SURF or SIFT operators for street panoramas. Majdik et al. (2013) generate artificial affine views of the scene in order to overcome the large viewpoint differences between GSV and low altitude images. Others (Torii et al., 2009; Ventura & Höllerer, 2013) match directly on the spherical GSV panoramas but using much denser images than those freely available by Google. Finally, Sato et al. (2011) have suggested the introduction of further constraints into the RANSAC outlier detection process to support automatic establishment of correspondences between wide-base GSV panoramas.

* Corresponding author

In order to match directly on the spherical panoramas with planar operators, the image base needs to be relatively short, as it is the case in most of the publications cited above. To our knowledge, only Sato et al. (2011) have worked on directly matching between standard wide-base GSV panoramas. Such a solution assumes that tentative matches have already been established (e.g. by SIFT, SURF, ASIFT). The concept “wide-base”, however, does not refer to the absolute size of the image base itself, but rather on the *base-to-distance ratio* which in fact determines the intersection angle on homologue rays. Our contribution focuses on matching standard GSV panoramas of rather narrow streets in densely built urban areas. In this context, a street of ~8 m width recorded from the street center-line at a step of ~12 m produces very unfavourable base-to-distance ratios of about 3:1 with respect to the street façade (in this sense one might speak of ‘ultra wide bases’). Such configurations produce large scale variations and strong incompatibilities between the distortions of projected panoramas (plus more occluded areas). It was thus experienced that even the ASIFT operator could just produce only a few valid matches along the baseline, namely close to the two vanishing points of this direction (when the street ended at streets perpendicular to it).

Mi ušík & Košecká (2009) point out that panorama representation via piecewise perspective, i.e. projection onto a quadrangular prism rather than on a cylinder, permits point matching algorithms to perform better since their assumption of locally affine distortions is expected to be more realistic for perspective images than for cylindrical panoramas. Corresponding tentatively matched 3D rays may then be validated via robust epipolar geometry estimation to produce the essential matrix **E**. However, it would be clearly preferable to create virtual views of panoramas as close as possible to affinity (as did Majdik et al., 2013, in order to register frames to panoramas) and subsequently apply the affine operator (ASIFT) developed by Morel & Yu (2009). Thus, the main purpose of this contribution is to describe, implement and evaluate such an alternative for “ultra wide-base” panoramas. Results will be given and assessed for performed 3D measurements and achieved accuracies.

2. RETRIEVAL AND ADJUSTMENT OF PANORAMAS

2.1 Retrieval of Google Street View panoramic images

In order to retrieve a panoramic image, an algorithm interacting with Google Maps Javascript API and Google Street View API was implemented. The coordinates of the image center in the region of interest need to be specified via the Google Maps Javascript API¹. They serve as input for retrieving frames of a panorama via the Google Street View API². Frame size was set to 640×640, the horizontal field of view to 45°; hence, image resolution is determined by the field of view. Frames were collected with the yaw angle (azimuth) step set to 22.5° (50% horizontal overlap), starting at 0° (direction North), for three consecutive frame strips with respective pitch angles 0°, 22.5° and 45° (roll angle was zero). Thus, 48 frames were retrieved for a panorama.

The individual frames were stitched together to a new spherical panoramic image using the open source photo-stitcher *Hugin*³. The software allows generating several 2D projections of spherical panoramas; here the cylindrical equidistant projection (*plate*

carrée) was used. The size of the projected panoramas at this resolution was around 4.900×1400 (for a coverage of 360°×100°), which is equivalent to an angular resolution $\delta = 0^\circ.073$.

2.2 Bundle adjustment

The geographical (λ, φ) and Cartesian (X, Y, Z) coordinate systems of a sphere with **O** as its center are seen in Fig. 1. Its most usual representation is through a cylindrical equidistant projection (equirectangular projection or *plate carrée*), seen in Fig. 2. A point in the image coordinate system x, y is transformed to λ, φ by means of the angular resolution as $\lambda = x\delta, \varphi = y\delta$. Hence, each image point defines a ray in space; its equation is the basis of bundle adjustment. Since all panoramas had been requested with no rotations, the adjustment can be simplified by assuming that unknowns are only the relative translation parameters.

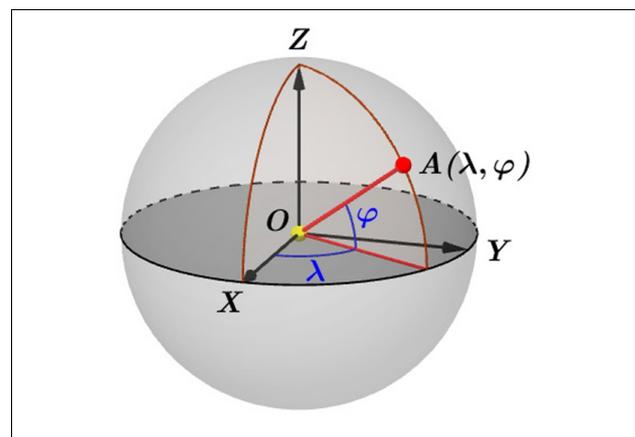


Figure 1. Geographical (λ, φ) and Cartesian (X, Y, Z) coordinate systems of the sphere.

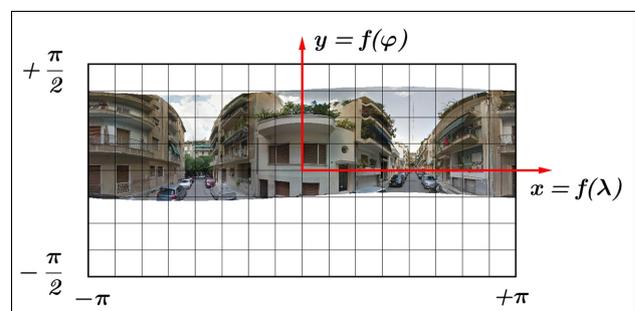


Figure 2. Equirectangular projection of a panorama.

To optimize stitching (and to refine orientation), *Hugin* exploits verticality of automatically extracted straight lines for leveling the panorama, i.e. ensuring that pitch and roll angles equal zero. This “upright constraint” of a common vertical orientation may well serve as a simplifying constraint on pose estimation (Ventura & Höllerer, 2013). Indeed, we adjusted with manual image point measurements several pairs of GSV panoramas stitched by *Hugin* and found negligible differences in their relative rotation angles. Finally, this was also verified by similarity transformations between known points on a house façade and their reconstruction from panoramas with rotation angles assumed to be zero (see Section 4).

Hence, only 4 (rather than 7) degrees of freedom must be fixed in GSV multi-panorama or stereo-panorama bundle adjustments, i.e. three translations and scale (as provided by the GPS data).

¹ <https://developers.google.com/maps/documentation/javascript/>

² <https://developers.google.com/maps/documentation/streetview/>

³ <http://hugin.sourceforge.net/>

3. AUTOMATIC POINT MATCHING AND FILTERING

As mentioned, a usual technique for automatic key-point extraction and matching on panorama pairs is to change from the standard equirectangular projection to a cubic one. The latter consists of six typical planar central projections, hence point operators such as SIFT, or ASIFT, may be used for feature extraction and matching. However, due to the “ultra wide” baseline conditions of this project, such a projection is not very efficient as no sufficient overlapping areas exist. An approach adopted here is to generate a custom projection, namely a “triangular prism projection” (TPP). This consists of central projections on three vertical planes which form a triangular prism in 3D space. Each projection, which is actually a conventional perspective image, has here a field of view of 120°, both vertically and horizontally. TPP plots on the horizontal plane as an equilateral triangle, with one of its vertices lying on the projection of the baseline of the stereo pair on this plane (point V in Fig. 3). Only two of the three TPP images are used here (4 TPP image components per panorama, i.e. as many as in standard cubic projections), as seen in Fig. 3. Overall, TPP represents a generalization of the cubic projection with adaptive FOV angle per image component, designed to behave optimally in similar wide baseline conditions.

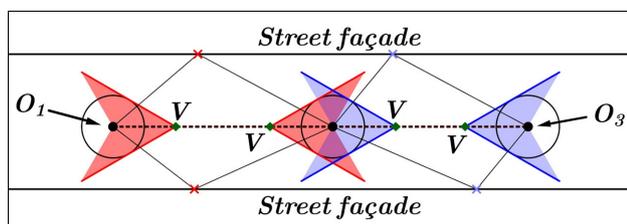


Figure 3. Triangular prism projection of spheres. Red and blue are the projections of the panoramic pairs (the vanishing lines of the vertical façade plane are projected in points V).

The particular geometry of this projection was designed for establishing correspondences on adjacent panoramas; the issue of finding matches on panorama triplets has not been addressed. In spite of the use of a robust point operator such as ASIFT, however, no sufficient matches in any pair emerged on these projections. Hence, a 2D projective transformation is applied to each image of the TPP (which has a wider FOV than the standard cubic projection, and thus allows more extended reconstruction), under the hypothesis of the planar nature of façades in urban environments as well as the near-parallelism of the baselines of the stereo pairs with the basic plane of the street façade. This transformation maps the vanishing line of the façade plane back to the line at infinity (Hartley & Zisserman, 2003). In our particular case, the vanishing lines do not need to be computed using image features (e.g. by line extraction to find orthogonal vanishing points and the vanishing line) as their position on these projections can be safely predicted. Assuming vertical façade plane and baselines which ideally run parallel to it, the vanishing line emerges as intersection of the image plane with a vertical plane containing the vector of the baseline (its trace is V in Fig. 3). A typical 2D transformation matrix in such a case is of the type of Eq. (1):

$$\mathbf{H} = \begin{bmatrix} s & 0 & 0 \\ 0 & s & 0 \\ l_1 & l_2 & l_3 \end{bmatrix} \quad (1)$$

where $\mathbf{l} = (l_1, l_2, l_3)^T$ is the vanishing line in homogeneous coordinates and s an isotropic scaling factor to compensate for any scaling issues that might arise under the transformation.

Theoretically, under this projective transformation a façade plane, and all planes parallel to it, should be reconstructed up to affinities. In practice, however, such rectified images are not totally free of projective distortions due to the uncertain location of the vanishing line; yet this does not impede the use of standard point operators like the ASIFT on such “quasi-affine” views. In fact, ASIFT provided 75-250 initial matches for each stereo pair in our tests. Evidently, not all initial matches represent authentic correspondences. For filtering out false matches, a common approach relies on robust estimation of the essential matrix \mathbf{E} for each stereo pair. But before this, a more heuristic technique has been applied to remove obvious blunders. For all matches, the slope of the line connecting the two points, i.e. the ratio of Δy to the sum of Δx plus the width of the rectified image, is calculated; median m and standard deviation σ for these slope values are computed, and only matches with values in the range $m \pm k\sigma$ ($k = 2$ or 3) are kept. In Fig. 4 a simple example is seen. Finally, all paired points remaining after RANSAC filtering are expressed in the geographical system (λ, ϕ) on the pansphere by successively inverting the projective and TPP transformations.

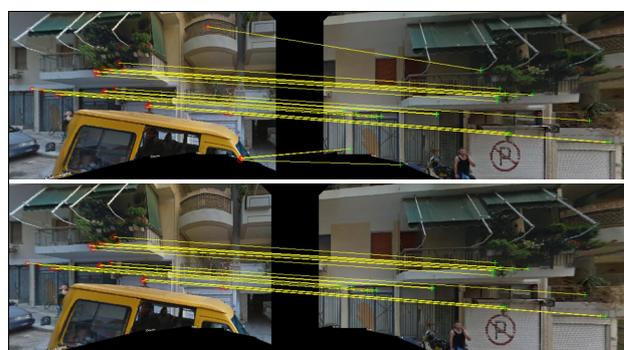


Figure 4. Outlier removal by filtering out deviating line slopes.

But even then erroneous points still manage to survive, namely wrong matches which, nevertheless, satisfy the tolerance of the epipolar constraint. Such points may be detected after a (stereo or multi-image) bundle adjustment. Here, a vertical plane was fitted to all reconstructed points of the pair using RANSAC with a tolerance of ± 1 m. Thus, point pairs with wrong disparities are filtered out (at the cost of sacrificing some valid matches). A further measure was to discard points intersected with standard errors above a limit (here 20 cm). An example is seen in Fig. 5.

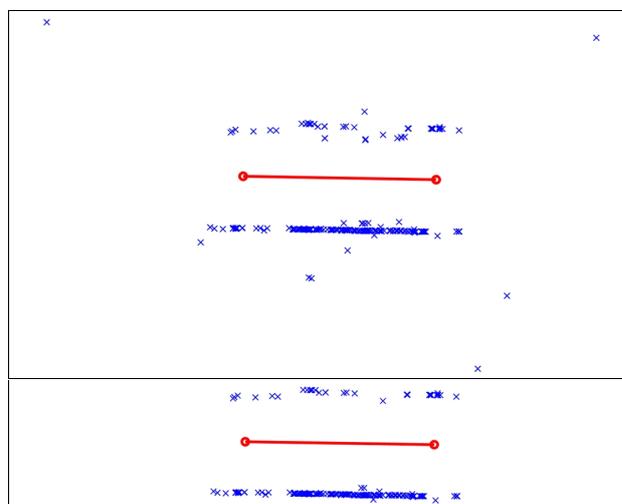


Figure 5. Plan view of automatically reconstructed points of a stereo pair: initial points (above) and after 3D filtering (below).

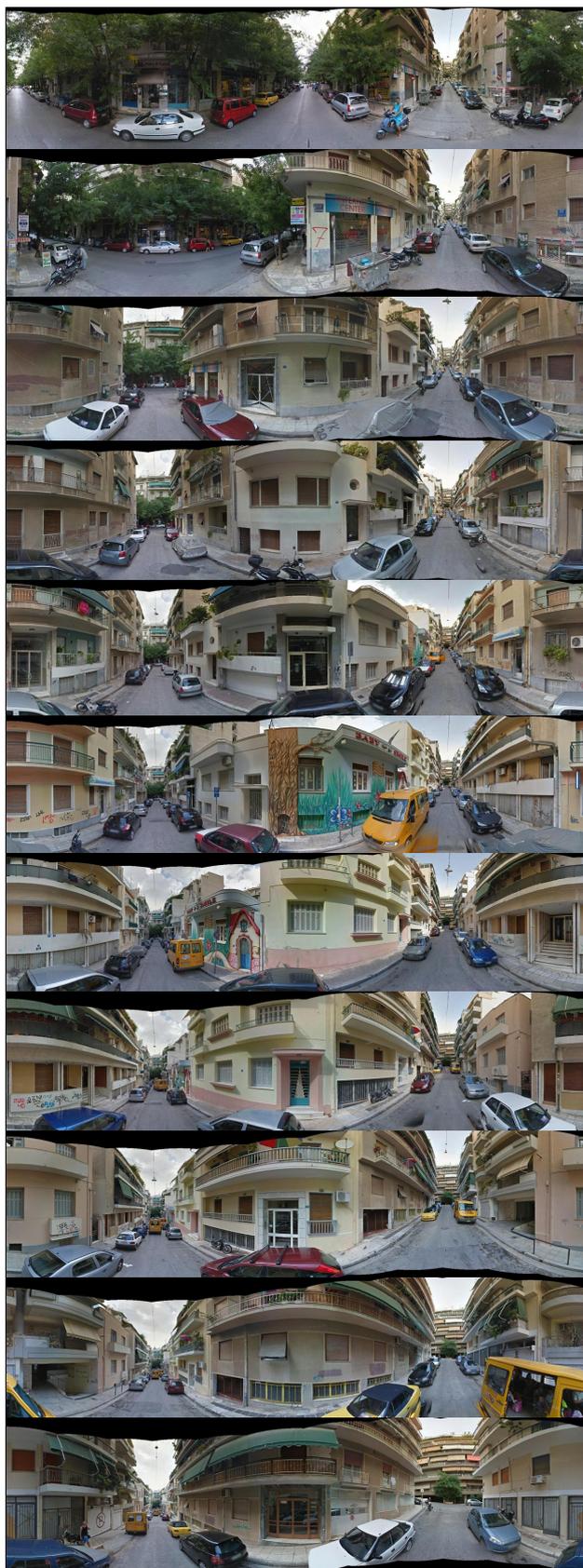


Figure 6. The 11 panoramas of test data set.

Concluding, it is stressed that in this specific case (assumed parallelism of baselines to the street façade) one might, of course, simply project the stereo panoramas onto a plane parallel to the

façade and perform point matching on these projections; thus, no need for two transformations (TPP projection and projective transformation) would exist. Nonetheless we chose to adopt this more general two-step approach which might also be applicable in instances where parallelism of baselines to street façades cannot be assumed. Vanishing points would then have to be identified on rectilinear projections, such as TPP, with automatic techniques (e.g. Rother, 2002). It is noted that alternatives for automatically extracting vanishing points directly on equirectangular projections have also been reported (Oh & Jung, 2012).

4. EXPERIMENTAL RESULTS

Our test data consisted of the 11 successive GSV spherical panoramas (i.e. 10 stereo pairs) seen in Fig. 6. These had been acquired at a step of ~12 m, which covered fully a short straight street (~120 m long, ~8 m wide). Initially, a multi-panorama adjustment was performed involving all images, with tie points measured manually directly on equirectangular projections. About 25-50 points were measured on each panorama. It was possible (though tiresome) to carefully select a few tie points common to each panorama triplet. By fixing the projection center of one panorama and scaling the model via the GPS data, unknowns were the model coordinates of the projection centers of all remaining panoramas and those of the tie points. The street runs almost parallel to the West-East direction; its axis is thus close to the X axis and perpendicular to the Y axis (pointing North). The RMS standard errors of the estimated locations of the projective centers (X_0 , Y_0 , Z_0) in this system were:

$$\sigma_{X_0} = \pm 25 \text{ cm} \quad \sigma_{Y_0} = \pm 5 \text{ cm} \quad \sigma_{Z_0} = \pm 3 \text{ cm}$$

The uncertainty of camera localization emerges as significantly larger along the street axis, since it essentially depends on the limited number of triple intersections.



Figure 7. Common areas on the *plate carrée* of a panorama pair.



Figure 8. Triangular prism projections of the images of Fig. 7.



Figure 9. "Quasi-affine" projections of the images of Fig. 8.



Figure 11. Final matches on two adjacent panoramas.



Figure 10. Tentative matches obtained by the ASIFT operator.

As already mentioned, automatic point matching could be performed here only pairwise. To illustrate this process, Fig. 7 presents a part of the common area of two adjacent panoramas. The corresponding piecewise perspectives of the TPP projections are seen in Fig. 8, whereas the “quasi-affine” transformations are shown in Fig. 9. In Fig. 10 the tentative matches on these images obtained by the ASIFT operator are presented. Finally, Fig. 11 shows all final matches on the two panoramas.

Here, unknowns for each panorama pair were the B_Y and B_Z base components (for B_X the value obtained from the multi-panorama adjustment for the corresponding pair was used) and the model coordinates of the automatically matched tie points.

A first evaluation involves the comparison of base components B_Y , B_Z obtained by both approaches (Table 1). The results are regarded as satisfactory, taking into account that the stereo adjustments rely on a geometrically weaker configuration. If one model is excluded (model 2 for B_Y , model 8 for B_Z), the respective values then become 1 ± 10 cm and 6 ± 4 cm. It is observed, however, that the stereo solution underestimates systematically the B_Z component.

model	multi-panorama adjustment (<i>manual</i>)		stereo-panorama adjustment (<i>automatic</i>)	
	B_Y (cm)	B_Z (cm)	B_Y (cm)	B_Z (cm)
1	98	75	91	62
2	42	52	70	49
3	44	59	45	49
4	68	68	70	66
5	60	92	59	84
6	46	84	51	80
7	46	108	64	98
8	32	121	16	101
9	85	138	95	137
10	50	101	48	99
$\Delta B_Y: \mu = 4$ cm, $\sigma = \pm 12$ cm				
$\Delta B_Z: \mu = 7$ cm, $\sigma = \pm 6$ cm				

Table 1. Model base components B_Y , B_Z from multi-panorama and stereo-panorama adjustments, with mean μ and standard deviation σ of their differences.

A similar evaluation involves estimations of the azimuth of the straight street axis (referred to the X-direction), as obtained per model from their B_X , B_Y components and seen in Table 2.

model	multi-panorama adjustment (<i>manual</i>)	stereo-panorama adjustment (<i>automatic</i>)
1	5°.0	4°.6
2	2°.7	4°.5
3	2°.7	2°.8
4	3°.2	3°.4
5	2°.7	2°.6
6	2°.7	3°.0
7	2°.5	3°.4
8	1°.8	0°.9
9	3°.9	4°.4
10	2°.9	2°.8
$\mu \pm \sigma$	$3.0^\circ \pm 0.9^\circ$	$3.2^\circ \pm 1.1^\circ$
RMS Difference: 0°.7		

Table 2. Azimuth of street axis per model from multi-panorama and stereo-panorama adjustments, with mean μ and standard deviation σ along with RMS differences.

Here again, the model-wise automatic estimation compares well with the results from the manual multi-image solution.

A further evaluation refers to the estimation of street inclination per panorama model. Street slopes between the approximate locations of panorama projection centers were directly measured with an optical inclinometer of an assumed precision of 0.75%. Slope estimates were also derived from the differences in elevation B_Z between neighbouring camera positions obtained by the manual (multi-image) and automatic (stereo) adjustments. Their values are tabulated in Table 3.

reference (<i>inclinometer</i>)	multi-panorama adjustment (<i>manual</i>)	stereo-panorama adjustment (<i>automatic</i>)
5.9	6.6	5.5
5.5	5.9	5.6
5.9	6.3	5.2
6.3	5.7	5.4
7.9	7.1	6.5
9.5	8.6	8.2
10.6	10.1	9.1
11.4	11.7	9.7
11.4	11.2	11.1
11.0	10.3	10.1
RMS difference: 0.6		RMS difference: 1.1

Table 3. Slope (%) of the 10 successive street segments.

The automatic model-wise adjustment estimates street slopes with an RMS uncertainty of 1.1%. This value is roughly equivalent to an uncertainty in elevation of 12 cm between camera stations. Considering the inherent uncertainty of the reference data, this comparison is satisfactory. It is noted that the multi-panorama solution compares even better, indicating that, in principle, street slopes might be reliably estimated from GSV imagery.

A final comparison uses the check points of Fig. 12 (measured by tape on a house façade), which had also been included and reconstructed as tie points in the multi-panorama adjustment.



Figure 12. Ground check points.

These points, which appeared in two models (A, B), were subsequently intersected using their manually measured image coordinates and the corresponding values for B_x (obtained by the multi-panorama adjustment) and B_y , B_z (from the two automatic stereo solutions). The check points thus reconstructed were compared with the field measurements via 7-parameter 3D similarity transformations. Results for scale differences and rotation angles are presented in Table 4 (translations are irrelevant).

results	multi-panorama adjustment (manual)	stereo-panorama adjustment (automatic)	
		model A	model B
scale $\Delta\lambda$	4.1%	4.0%	3.3%
pitch (X axis)	0°.6	0°.6	-0°.1
roll (Y axis)	0°.2	-1°.1	0°.5
yaw (Z axis)	1°.9	2°.0	3°.1
standard error	4.0 cm	5.1 cm	5.0 cm

Table 4. Results of 3D similarity transformations between reference and reconstructed points.

The standard error of the adjustment implies that the reconstructed points fit well the ground data in all instances, namely even in the case of the automatic adjustment of independent models. The very small pitch and roll angles indicate that the “upright” assumption is realistic, in the sense that for many practical purposes the Z axis of suitably retrieved and projected GSV panoramas may indeed be considered as vertical. On the other hand, a clear scaling problem is, of course, observed between the provided GPS information (which served for scaling the panorama models) and the field measurements.

5. CONCLUDING REMARKS

The information potential of street-level panoramic images, like those of Google Street View, is widely acknowledged. A variety of approaches and applications with differing degrees of automation, such as those cited here, are being constantly reported. A crucial factor regarding automation is, obviously, the image acquisition geometry. In cases of densely built areas and relatively narrow streets, the current standard recording step of GSV imagery gives rise to strongly unfavourable base-to-distance ratios, which further aggravate inherent distortions of panoramic configurations and representations. This puts matching among panoramas to the test.

In this contribution the possibility of synthesizing ‘quasi-affine’ views of successive panoramas of street façades, which can be handled more efficiently by affine point operators, has been in-

vestigated. A suitable piecewise (rectilinear) perspective projection has been used, followed by a projective transformation. Alternatively, one might consider the generation of ‘quasi-similar’ projections by combining the vanishing lines of the street façade and the known internal geometry of the panoramas, and then apply the standard SIFT or SURF operators instead of ASIFT.

Our first trials indicate that sufficient matches on adjacent panoramas may be obtained thus, allowing their successful pairwise adjustment. The described approach has produced satisfactory results. However, for the image geometry studied no matches on more than two adjacent panoramas are apparently possible, i.e. it is not feasible to perform automatic multi-panorama adjustments (which, of course, represent a more robust configuration against the geometry of independent panoramic stereo models). In this direction, the elaboration of representations suitable for multi-panorama matching is a topic of future research.

REFERENCES

- Agarwal P., Burgard W., Spinello L., 2015. Metric localization using Google Street View. *Preprint arXiv:1503.04287v1 [cs.RO]*, 14 March.
- Aly M., Bouguet J.-Y., 2012. Street View goes indoors: automatic pose estimation from uncalibrated unordered spherical panoramas. *IEEE Workshop on Applications of Computer Vision (WACV)*, pp. 1-8.
- Cruz-Mota J., Bogdanova I., Paquier B., Bierlaire M., Thiran J.-P., 2012. Scale invariant feature transform on the sphere: theory and applications. *International Journal of Computer Vision*, 98(2), pp. 217-241.
- Fangi G., 2015. Towards an easier orientation for spherical photogrammetry. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XL-5/W4, pp.279-283
- Hansen P., Corke P., Boles W., 2010. Wide-angle visual feature matching for outdoor localization. *International Journal of Robotics Research*, 29(2-3), pp. 267-297.
- Hartley R., Zisserman A., 2003. *Multiple View Geometry in Computer Vision*. 2nd edition, Cambridge University Press.
- Majdik A.L., Albers-Schoenberg Y., Scaramuzza D., 2013. MAV urban localization from Google Street View data. *Proc. IEEE/RSJ International Conference on Intelligent Robots & Systems (IROS)*, pp. 3979-3986.
- Mi ušík B., Košecká J., 2009. Piecewise planar city 3D modeling from street view panoramic sequences. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2009)*, pp. 2906-2912.
- Morel J., Yu G., 2009. ASIFT: A new framework for fully affine invariant image comparison. *SIAM Journal on Imaging Sciences*, 2, pp. 438-446.
- Oh S.H., Jung S.K., 2012. Vanishing point estimation in equi-rectangular images. *Proc. 8th International Conference on Multimedia Information Technology and Applications*, pp. 45-47.
- Pagani A., Gava C., Cui Y., Krolla B., Hengen J.-M., Stricker

D., 2011. Dense 3D point cloud generation from multiple high-resolution spherical images. *12th International Symposium on Virtual Reality, Archaeology & Cultural Heritag (VAST 2011)*, pp. 17-24.

Rother, C., 2002. A new approach to vanishing point detection in architectural environments. *Image and Vision Computing*, 20 (9-10), pp. 647-655.

Salmen J., Houben S., Schlipfing M., 2012. Google street view images support the development of vision-based driver assistance systems. *IEEE Symposium on Intelligent Vehicles (IV)*, pp. 891-895.

Sato T., Pajdla T., Yokoya N., 2011. Epipolar geometry estimation for wide-baseline omnidirectional street view images. *IEEE International Conference on Computer Vision Workshops*, pp. 56-63.

Szeliski R., 2006. Image alignment and stitching: A tutorial. *Foundations and Trends in Computer Graphics & Vision*, 2(1), pp. 1-104.

Szeliski R., Shum H.-Y., 1997. Creating full view panoramic image mosaics and environment maps. *Proc. 24th Annual Conference on Computer Graphics & Interactive Techniques, SIGGRAPH '97*, pp. 251-258.

Taneja A., Ballan L., Pollefeys , 2014. Never get lost again: vision based navigation using StreetView Images. *Asian Conference on Computer Vision (ACCV 2014)*, Lecture Notes in Computer Science, Springer, vol. 9007, pp. 99-114.

Torii A., Havlena M., Pajdla T., 2009. From Google Street View to 3D city models. *IEEE International Conference on Computer Vision Workshops*, pp. 2188-2195.

Tsai V.J.D., Chang C.-T., 2013. Three-dimensional positioning from Google street view panoramas. *IET Image Processing*, 7(3), pp. 229-239.

Tsironis V., 2015. *Study of the Geometry of Spherical Panoramas*. Diploma Thesis, Faculty of Rural & Surveying Engineering, National Technical University of Athens, Greece.

Ventura J., Höllerer T., 2013. Structure and motion in urban environments using upright panoramas. *Virtual Reality*, 17(2), pp. 147-156.

Yu G., Morel J.M., 2011. ASIFT: An algorithm for fully affine invariant comparison, image processing on line.
DOI:10.5201/ipol.2011.my-asift.

Zamir A.R., Shah M., 2011. Accurate image localization based on Google Maps Street View. *11th European Conference on Computer Vision (ECCV 2010)*, Part IV, pp. 255-268.