

Automatic Registration of Color Images to 3D Geometry

Yunzhen Li*

National University of Singapore

Kok-Lim Low†

National University of Singapore

Abstract

We present an approach to automatically register a large set of color images to a 3D geometric model. The problem arises from the modeling of real-world environments, where surface geometry is acquired using range scanners whereas the color information is separately acquired using untracked and uncalibrated cameras. Our approach constructs a sparse 3D model from the color images using a multiview geometry technique. We project special light patterns onto the scene surfaces to increase the robustness of the multiview geometry reconstruction. The sparse model is then approximately aligned with the detailed model. Planes found in the detailed model are exploited to refine the registration. Finally, the registered color images are mapped to the detailed model using weighted blending, with careful consideration of occlusion and the preservation of image details.

CR Categories: I.4.1 [Image Processing and Computer Vision]: Digitization and Image Capture—Imaging geometry, Scanning; I.4.8 [Image Processing and Computer Vision]: Scene Analysis—Range data, Sensor fusion

Keywords: Image-to-geometry registration 2D-to-3D registration range scanning multiview geometry SIFT image blending

1 Introduction

Active range sensing has been used to reconstruct high-quality 3D digital models of complex real-world objects and environments, and has found many practical applications in areas ranging from manufacturing to cultural heritage modeling [Levoy et al. 2000] [Stumpfel et al. 2003] and urban scene modeling [Liu and Stamos 2005].

Besides the geometric information, photometric properties of the object or environment surfaces have to be acquired as well in order to reconstruct a visually-realistic digital model. Some range scanners are capable of acquiring color information together with the range data, while some lack the capability. Nevertheless, for realistic display of a 3D digital model, view-dependent surface reflection must be captured. This requires the color of each surface point to be sampled from many directions, and therefore many more color images are usually needed than range images. In other applications, very often, the geometric model is to be mapped with very detailed color images much higher in resolution than that of the range images. In this case, even if the range scanner can capture color, the color may not have the required resolution. Many close-up or zoomed-in pictures have to be acquired in order to achieve the surface coverage as well as the required resolution.

Our current work focuses on the modeling of indoor environments. To allow greater maneuverability and flexibility, we use an ordinary digital color camera that is untracked and uncalibrated to take photographs of the indoor scene. Each color image is acquired with an unknown camera pose and unknown camera intrinsic parameters. These images must then be registered with the geometry of the 3D model constructed from the range images. Current registration methods require human users to manually identify corresponding points on each image with those on the range images or the 3D model. This approach is error-prone and becomes extremely tedious when there is a large number of images to register. There are automated registration methods [Liu and Stamos 2005] [Stamos and Allen 2001], but these methods are designed to exploit specific features unique to outdoor urban environments. Some approaches avoid, or partially avoid, the registration problem by using tracked calibrated camera or by co-locating the camera and the scanner [Williams et al. 2004].

We present a novel and practical approach to registering a large set of color images to their corresponding 3D geometric model. The approach is almost automatic and requires very little manual input. The approach is more general than previous methods, in that it can be applied not only to most indoor environments, but to other environments and objects.

Instead of trying to find corresponding features between color images and the 3D model, our method takes the approach of establishing correspondences among the color images using a robust image feature detection technique, and uses a multiview geometry technique to construct a sparse 3D model from the color images. To improve the robustness of the multiview geometry reconstruction for scenes with large surfaces that lack features, we use light projectors to project special light patterns onto the scene surfaces to artificially introduce image features. The sparse model is then approximately aligned to the 3D detailed model using only a few pairs of corresponding points between one of the color images and the 3D detailed model. Our method then extracts planes in the detailed model and exploits these to improve the sparse model and refine the registration.

1.1 Contributions

The main contribution of our work is the idea of taking the approach of establishing correspondences among the color images instead of directly finding corresponding features between the 2D and 3D spaces [Liu and Stamos 2005] [Stamos and Allen 2001]. The latter approach works well only for higher-level features, such as parallel straight lines, and this imposes assumptions and restrictions on the types of scenes the method can handle. For most indoor environments, these higher-level features usually exist, but they are often too few or do not appear in most of the color images due to small field of view and short shooting distance. Our approach works for more types of scenes and even for objects.

The main problem of feature correspondence is the lack of features on large uniform surfaces. This occurs a lot in indoor environments where large plain walls, ceiling and floor are common. We get around this problem by using light projectors to project special light patterns onto the scene surfaces to artificially introduce image features.

*liyuz@comp.nus.edu.sg

†lowkl@comp.nus.edu.sg

Our method requires the user to manually input only six pairs of correspondences between one of the color images and the 3D model. This allows the sparse model to be approximately aligned with the detailed model. We detect planes in the detailed model, and by minimizing the distances between some of the points in the sparse model and these planes, we are able to refine the multiview geometry and the registration as a whole using sparse bundle adjustment (SBA) [Lourakis and Argyros 2004]. This approach is able to achieve better registration accuracy in the face of non-uniform spatial distortion in the geometric model.

Our goal here is not to render the completed model with view-dependent reflection. Instead, we assign each point on the surface of the 3D model a single color by carefully blending colors from multiple overlapping color images. Our method takes into consideration the different exposures of the color images and the occlusion of surfaces in the 3D model. It produces a colored model with very smooth color transitions and yet preserves fine details. Even though we did not demonstrate it, our method of registering a large set of color images to the geometric model naturally applies to creating models with view-dependent reflection [Chen et al. 2002] and models with very high-resolution texture maps.

In the next section, we review related work for similar registration problems, and also related work in multiview geometry. Section 3 gives a detailed description of the main steps in our method. We discuss our experiment results in Section 4 and then conclude the paper with discussion of some future work in the next section.

2 Related Work

We first review previous work on automatic registration of color images to 3D geometric model or range images. Since multiview geometry is central to the working of our method, we therefore review work related to it, with emphasis on the particular technique used in the Photo Tourism work [Snavely et al. 2006], which is very similar to our implementation.

2.1 Automatic Image-to-Geometry Registration

There are two major classes of algorithms for automatic registration of color images to 3D geometric model or range images — feature based and statistics based. Feature-based methods attempt to match corresponding features found in both the color images and the 3D model. For the modeling of buildings and urban scenes, Stamos and Allen [Stamos and Allen 2001], and Liu and Stamos [Liu and Stamos 2005] detect and match line features between the images and the 3D models. The latter method further exploits line parallelism and orthogonality that naturally exist in urban scenes. These constraints can greatly improve the reliability of their automatic registration. Zhao et al. [Zhao et al. 2005] constructs dense point-cloud model from video, and uses the ICP (iterative closest point) algorithm [Besl and McKay 1992] to align the point-cloud model to a detailed geometric model obtained using range scanning. The registration is then used to texture map the video frames onto the geometric model.

Statistics-based algorithms exploit the statistical dependence between color images and 3D models. Some range sensors are capable of recording the intensity of the reflected sensing light at each range sample, and to reduce noise, the sensing light is often not in the visible light spectrum. Williams et al. [Williams et al. 2004] and Hantak and Lastra [Hantak and Lastra 2006] use the dependence between color and the infra-red intensity at each range sample, and several similarity measures, such as mutual information and chi-square statistics, are used to search for the best match. For this approach, good camera pose initialization is crucial to de-

rive the correct camera pose. The registration results of different information-theoretic metrics are compared in [Hantak and Lastra 2006]. Also, Pong and Cham [Pong and Cham 2006] have explored mutual information between image intensity and object surface normals for the alignment of 3D objects to their 2D images.

2.2 Multiview Geometry

The success of a multiview geometry reconstruction relies on the robustness of the detected image features. The Scale Invariant Feature Transform (SIFT) [Lowe 2004], or variants of it, has become popular in applications that require robust image features, such as in panorama reconstruction [Brown and Lowe 2007] and in Photo Tourism [Snavely et al. 2006]. SIFT keypoints and descriptors are invariant to scale and rotational transformations, and to other affine transformations and perspective transformations to a certain extent. A recent study [Szeliski 2006] shows that SIFT generally provides the most robust features for most applications.

The Photo Tourism system [Snavely et al. 2006] uses a multiview geometry technique to reconstruct sparse 3D point-cloud models of tourist locations from a large collection of photographs obtained from the Internet or from personal photo collections. The system allows users to explore the location and browse the photo collection in 3D. The system builds the sparse model by first detecting and matching SIFT features in the images. Given the feature correspondences and the approximate initial camera parameters stored in the EXIF tags of the images, the multiview geometry, which consists of the sparse 3D point-cloud model and the intrinsic and extrinsic parameters of each view, is computed using a structure-from-motion (SfM) technique [Hartley and Zisserman 2004]. The multiview geometry computation repeatedly applies a two-view geometry algorithm. The geometry expands by adding new views incrementally. To guarantee correctness, after each view is added, sparse bundle adjustment (SBA) [Lourakis and Argyros 2004] is applied to refine the overall structure.

3 Proposed Method

The following subsections describe our approach in details. The major steps are

1. data acquisition,
2. multiview geometry reconstruction,
3. approximate registration of the sparse model to the detailed model,
4. registration refinement, and
5. mapping color to the model surfaces.

3.1 Data Acquisition

During data acquisition, range images are captured using a range scanner, and color images are captured using an untracked and uncalibrated ordinary digital camera. The zoom setting of the camera is not fixed, and auto-focus and auto-exposure are enabled. If the exposure time is too long, we mount the camera on a lightweight tripod.

Very often in indoor environment, there are some large surfaces, such as walls and ceiling, that have very few image features required for the multiview geometry reconstruction. For these surfaces, we use light projectors to project a static light pattern on them (see Figure 1). For views that can see this light pattern, we

mount the camera on the tripod, lock the focus and the zoom setting (but not the exposure), and take one image with and another without the projected light pattern. We use aperture-priority auto-exposure to ensure that the depth of field does not change. Of each of these image pairs, only the one with light pattern is used in the multiview geometry reconstruction, and the other one is used for applying color to the 3D model.



Figure 1: Plain surfaces in the scene are projected with special static light pattern to improve feature detection.

The light pattern actually comes from an ordinary high-resolution image. We converted it to a grayscale image, tuned its contrast, uniformly subdivided it into 16 pixels-by-16 pixels blocks, and randomly shuffled the blocks to get the final image. The random shuffling of the blocks introduces additional high-frequency structures at the block boundaries, and at the same time increases the uniqueness of the low-frequency structures at each point. We were careful to choose an original image that has very few large uniform regions (for copyrights reason, we are unable to show the original image here). Our experiments have shown that the light pattern can greatly increase the number of correctly-matched SIFT keypoints.

3.2 Multiview Geometry Reconstruction

We run the SIFT algorithm [Lowe 2004] to detect SIFT keypoints in the color images. For efficiency, we use the approximate nearest neighbors kd-tree algorithm (ANN) [Arya et al. 1998] to match the descriptors of the keypoints for every pair of images. Similar to the Photo Tourism system [Snavely et al. 2006], we remove incorrect matches (outliers) using the following steps. First, for each image pair, we use an RANSAC approach [Fischler and Bolles 1981] to compute a fundamental matrix [Hartley 1995]. Those matches that cannot satisfy the fundamental matrix equation are removed. Next, the geometric consistency of each keypoint is checked across all images. Matches are connected across multiple images to form tracks. A track is inconsistent if it contains two different keypoints from any single image. Matches that form inconsistent tracks are removed.

The multiview geometry computation starts with the two-view geometry of the image pair that contains the largest number of consistent matches. Degenerate cases, where (1) the two images were taken from the same location, or (2) the matches in the two images are all on a single plane, are detected by first computing a homography transformation between the matched keypoints in the two images using the RANSAC approach. If more than 95% of the matched keypoints satisfy the homography transformation, we consider it a degenerate case. We discard the image pair and use another pair that contains the next largest number of consistent

matches. Otherwise, we just proceed normally to estimate the essential matrix and extract the relative camera pose from the essential matrix. The approximate intrinsic parameters used in the essential matrix computation are obtained from the image EXIF tags. There are four possible relative poses, but we choose the one that makes the associated reconstructed 3D points appear in front of the two cameras. We then apply bundle adjustment to refine the camera parameters and the 3D point locations.

We add a new view that has the largest number of matches with the existing 3D points. Using the correspondences between the 2D keypoints and the associated 3D points, we use a simple linear optimization method to initialize the camera parameters of the new view, and then use a nonlinear optimization to refine them. New 3D points are generated from the new matches between the new view and the existing views. The new view, together with the newly generated 3D points, are added into the multiview geometry, and sparse bundle adjustment [Lourakis and Argyros 2004] is applied to refine the whole structure. The process stops when no more view can be added. Figure 2 shows a multiview geometry reconstruction of the office scene shown in Figure 1. One can easily see the relatively denser point clouds on the plain walls projected with the special light pattern.

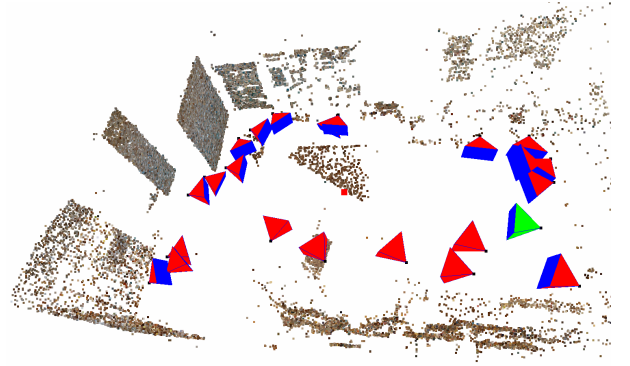


Figure 2: A multiview geometry reconstruction of an office scene. Each red pyramid represents a camera pose.

3.3 Approximate Registration

The sparse model constructed by the multiview geometry is up to a very approximate projective transformation, and has unknown scale and pose relative to the detailed model created from the range images. In this step, we get some inputs from the user to register the multiview system to the detailed model.

The user first chooses one of the color images in the multiview system and then selects six or more correspondence points on this image and on the 3D detailed model. Using the correspondences between the 2D and 3D points on the detailed model, we use a simple linear optimization method to initialize the camera parameters of the selected view, and then use a nonlinear optimization to refine them. The estimated intrinsic and extrinsic parameters, with respect to the detailed model, of the selected view are then propagated to the other views in the multiview system to update their intrinsic and extrinsic parameters accordingly and to also update the locations of the 3D points in the sparse model.

This is the only part of our method that requires manual inputs. However, the amount of inputs is very small when compared to many existing methods where the user has to input many more point correspondences for every color image. At this stage, the registration is still not accurate enough due to the error in the point corre-

spondences input by the user. The next step describes one approach to refine the registration.

In our implementation, our system actually lets the user select the correspondence points between one of the color images and one of the range images of the scene. These range images are assumed to have been registered with each other (using technique such as [Pulli 1999]). Therefore, the approximate registration of the sparse model with one range image can be automatically propagated to all other range images. In the registration refinement step, our method makes use of data in all the range images.

3.4 Registration Refinement

There is a need to further refine the registration. One source of error is from the point correspondences input by the user, another source is the non-uniform spatial distortion in the detailed model, which is mainly caused by calibration errors in the range scanners. We would like to allow the multiview system to be aligned with the detailed model in a non-rigid manner. For this, we make a small addition to our sparse bundle adjustment of the multiview system.

We first detect planes in the detailed model using the PCA method [Pearson 1901]. Next, for each plane, we identify 3D points in the multiview system that are close to it within a threshold and associate these points to the plane. The sum of the squared distances between the 3D points and their associated planes is added as a new term to the error function of the original sparse bundle adjustment. A constant coefficient is multiplied to the new term so that it would not dominate the error function. We then run the sparse bundle adjustment on the new system.

Our registration refinement approach is more appropriate than using the ICP algorithm [Besl and Mckay 1992]. The ICP algorithm treats the two models as rigid shapes, so it is not able to adjust the registration to adapt to the distortion in the detailed model. Moreover, the intrinsic and extrinsic parameters of the views still need to be further tuned, which cannot be achieved using the ICP algorithm. The bundle adjustment approach we are taking is powerful enough to address all these issues.

Certainly, this approach works well only if planes exist in the scene. Our method can be extended to deal with scenes that have very little planar surfaces. The idea is very similar to the ICP algorithm, in which we associate each point in the multiview system with its nearest point in the detailed model, and use the distance between them as part of the error metric in the bundle adjustment. However, this approach requires more changes to the original sparse bundle adjustment implementation, unlike in the planar case, in which each plane can be set up as a “view” and the distances between it and the associated 3D points can be treated as pixel errors in the “view”.

Figures 3 and 4 show an example office scene and a laboratory scene, respectively, where their geometric models have been partially mapped with the color images registered using our method. Figure 5 shows the result of our registration refinement. One can see that the registration errors among the multiple images and the geometric model have been greatly reduced by the registration refinement.

3.5 Mapping Color to Model

When the views are well-registered to the 3D model, the color information extracted from the images at each point on the 3D model surface can be used to construct a representation for view-dependent rendering using surface light field techniques [Chen et al. 2002]. However, as rendering is not the main focus of this work, we just want to simply assign each point on the surface of the 3D



Figure 3: A geometric model of an office scene, partially mapped with the color images registered using our method.



Figure 4: A geometric model of a laboratory scene, partially mapped with the color images registered using our method.

model a single color. At each surface point, we carefully blend colors from multiple overlapping color images so that the colored model has smooth color and intensity transitions and yet preserves fine details. To achieve that, there are a few issues that we need to consider:

1. The color images were taken using different exposures. To make the exposure consistent, we use the method by Debevec and Malik [Debevec and Malik 1997] to find the transfer function of our camera, and we use that to set all the images to the same exposure.
2. Some of the rays emanating from the views actually intersect more than one surface in the detailed model. If this is not handled correctly, the surface that is occluded from a view may still get color contribution from it.
3. The effect of inaccurate registration is often magnified near large depth discontinuities, where the color “spills” over to the wrong side of the depth boundaries.
4. Even when the images have been adjusted to the same exposure, their brightness at overlapping regions are often quite different, due to vignetting and view-dependent reflection.
5. Slight mis-registration of the views can easily cause loss of fine details in the image after blending.

To deal with Issue 2, we first render a depth map of the detailed model from each view. Then, to check whether a surface point should get contribution from a color image, we project it to the corresponding view of the color image and compare its depth value with the value in the view’s pre-rendered depth map. If the former is greater (by a threshold), then it is considered occluded from the view.

To deal with Issue 3, we use an edge detection algorithm to detect large depth boundaries in the depth maps. We then create a depth boundary mask image (DBMI) by dilating the edge image so that pixels near to or on the depth boundaries have value of 1, and the rest of the pixels have value 0. Figure 6 shows a depth map and its corresponding DBMI. The DBMIs are used in the blending process where we also handle Issue 4.



Figure 5: Registration results before registration refinement (top) and after registration refinement (bottom).



Figure 6: (Left) A depth map of the detailed model rendered from a view. (Right) Its corresponding depth boundary mask image.

We deal with Issue 4 by using weighted blending where we assign higher weight to the center of the image and decrease the weight towards the edges of the image. For a surface point on the detailed model, the color contribution from the images is defined by the weight function

$$W_i(X, Y, Z) = \left(1 - \sqrt{\frac{\left(\frac{x-w}{w}\right)^2 + \left(\frac{y-h}{h}\right)^2}{2}}\right) \cdot (1 - B_i(x, y))$$

where $W_i(X, Y, Z)$ is the weight of image i at surface point (X, Y, Z) , x and y are the projection of (X, Y, Z) in image i , w and h are half the image width and height, and $B_i(x, y)$ is the depth boundary mask image of image i .

Given that $M_i(x, y)$ is the color of pixel (x, y) of image i , the color assigned to the surface point (X, Y, Z) is

$$M(X, Y, Z) = \frac{\sum_{i=1}^n W_i(X, Y, Z) M_i(x, y)}{\sum_{i=1}^n W_i(X, Y, Z)}$$

where n is the number of views that can see the surface point (X, Y, Z) . Figure 7 shows the results with and without applying the weighted blending. One can easily see the “seams” at the image boundaries.

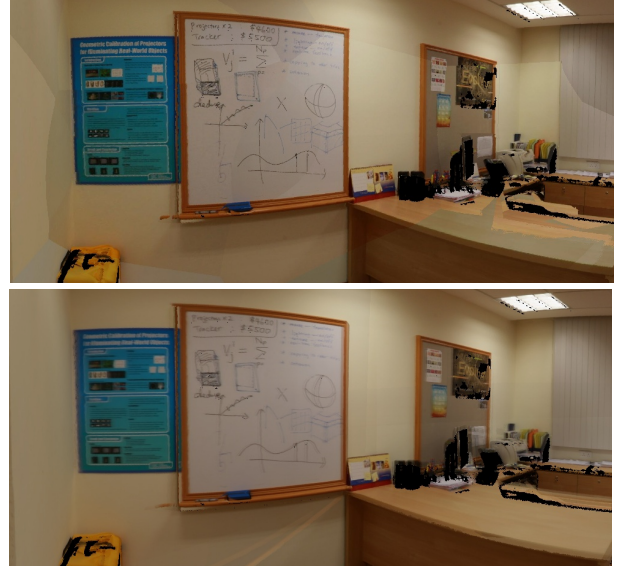


Figure 7: Results without weighted blending (top) and with weighted blending (bottom).

However, the weighted blending still does not very well preserve fine details (Issue 5). For this problem, we first extract details from the color images and add the details to the weighted blending result. More specifically, for each surface point (X, Y, Z) , we find the image that has the highest weight at the surface point:

$$I(X, Y, Z) = \arg \max_i W_i(X, Y, Z)$$

where $I(X, Y, Z)$ is the index of the image. Then the image detail at (X, Y, Z) is computed as follows:

$$H(X, Y, Z) = M_{I(X, Y, Z)}(x, y) - M_{I(X, Y, Z)}^{blur}(x, y)$$

where $M_{I(X, Y, Z)}^{blur}(x, y)$ is obtained by blurring the image $M_{I(X, Y, Z)}(x, y)$ with a Gaussian filter. The final color value assigned to surface point (X, Y, Z) is therefore

$$M(X, Y, Z) + H(X, Y, Z).$$

Figure 8 shows a close-up view of a poster in the office scene. With our method, the improvement in details is very evident.

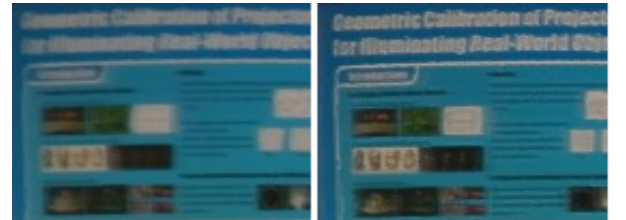


Figure 8: Results of weighted blending without preservation of details (left) and with preservation of details (right).

4 Implementation and Experiment Results

We have tested our method on several indoor scenes. We used a DeltaSphere-3000 laser scanner to make panoramic scans of the room. The range accuracy of the scanner is about 1 centimeter, and

the range image resolution is 10 samples per degree. In the case that there are not enough features, we also used projectors to project the special light pattern on the plain walls in the office.

For the office scene shown in Figure 3, we use 30 color images and 7 of them contain projected pattern. The feature detection and multiview geometry reconstruction takes about 10 minutes to compute. The user input for the approximate registration typically takes about one minute. The registration refinement takes about two minutes and the final color mapping about one minute.

For the laboratory scene shown in Figure 4, we use 47 color images. Projected pattern is not required. The feature detection and multiview geometry reconstruction takes about 55 minutes to compute. The registration refinement takes about three minutes and the final color mapping about two minutes.

In all our results, we can achieve accuracies within three pixels everywhere on each image when the geometric model is projected into the respective camera view. These accuracies are similar to those achieved by manual registration, in which the user clicks about a dozen correspondence points on each image and the geometric model.



Figure 9: 3D renderings of the final colored model.

5 Conclusion and Future Work

We have presented a practical and effective approach to register a large set of color images to a 3D geometric model. Our approach does not rely on the existence of any high-level feature between the images and the geometric model, therefore it is more general than previous methods. In the case where there is very little image features in the scene, our approach allows the use of projectors to project special light patterns onto the scene surfaces to greatly increase the number of usable image features. To refine the registration, we use planes (or any surfaces) in the geometric model to constrain the sparse bundle adjustment. This approach is able to achieve better registration accuracy in the face of non-uniform spatial distortion in the geometric model. We have also described a way to blend images on the geometric model surface so that we obtain a colored model with very smooth color and intensity transitions and the fine details are preserved.

In our current implementation, the feature detection and multiview geometry reconstruction take the longest processing time. A large part of this can be implemented in GPU. Our future implementation will also demonstrate the use of our registration method for the creation of a view-dependent reflection model. For this application, our weighted blending step will not be required.

Acknowledgements

We thank the Photo Tourism team for making their bundle adjustment implementation publicly available. We have used the non-

commercial version of HDR Shop for adjusting the image exposure. We thank Prashast Khandelwal for his contribution to the preliminary work of this project. This project is supported by the Singapore Ministry of Education Academic Research Fund (WBS: R-252-050-241-112 & R-252-050-241-133).

References

- ARYA, S., MOUNT, D. M., NETANYAHU, N. S., SILVERMAN, R., AND WU, A. Y. 1998. An optimal algorithm for approximate nearest neighbor searching fixed dimensions. *J. ACM* 45, 6, 891–923.
- BESL, P. J., AND MCKAY, H. D. 1992. A method for registration of 3-D shapes. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 14, 2, 239–256.
- BROWN, M., AND LOWE, D. G. 2007. Automatic panoramic image stitching using invariant features. *Int. J. Comput. Vision* 74, 1, 59–73.
- CHEN, W.-C., BOUGUET, J.-Y., CHU, M. H., AND GRZESZCZUK, R. 2002. Light field mapping: efficient representation and hardware rendering of surface light fields. *ACM Trans. Graph.* 21, 3, 447–456.
- DEBEVEC, P. E., AND MALIK, J. 1997. Recovering high dynamic range radiance maps from photographs. *Computer Graphics 31*, Annual Conference Series, 369–378.
- FISCHLER, M., AND BOLLES, R. 1981. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* 24, 6 (June), 381–395.
- HANTAK, C., AND LASTRA, A. 2006. Metrics and optimization techniques for registration of color to laser range scans. In *3DPVT*, 551–558.
- HARTLEY, R. I., AND ZISSERMAN, A. 2004. *Multiple View Geometry in Computer Vision*, second ed. Cambridge University Press, ISBN: 0521540518.
- HARTLEY, R. I. 1995. In defence of the 8-point algorithm. In *ICCV '95: Proceedings of the Fifth International Conference on Computer Vision*, IEEE Computer Society, Washington, DC, USA, 1064.
- LEVOY, M., PULLI, K., CURLESS, B., RUSINKIEWICZ, S., KOLLER, D., PEREIRA, L., GINZTON, M., ANDERSON, S., DAVIS, J., GINSBERG, J., SHADE, J., AND FULK, D. 2000. The digital michelangelo project: 3D scanning of large statues. In *Siggraph 2000, Computer Graphics Proceedings*, ACM Press / ACM SIGGRAPH / Addison Wesley Longman, K. Akeley, Ed., 131–144.
- LIU, L., AND STAMOS, I. 2005. Automatic 3D to 2D registration for the photorealistic rendering of urban scenes. In *2005 Conference on Computer Vision and Pattern Recognition (CVPR 2005)*, 137–143.
- LOURAKIS, M., AND ARGYROS, A. 2004. The design and implementation of a generic sparse bundle adjustment software package based on the levenberg-marquardt algorithm. Tech. Rep. 340, Institute of Computer Science - FORTH, Heraklion, Crete, Greece, Aug.
- LOWE, D. G. 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60, 91–110.

- PEARSON, K. 1901. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, London, UK, 559–572.
- PONG, H., AND CHAM, T. 2006. Alignment of 3D models to images using region-based mutual information and neighborhood extended gaussian images. In *ACCV'06*, I:60–69.
- PULLI, K. 1999. Multiview registration for large data sets. In *International Conference on 3D Digital Imaging and Modeling*, 160–168.
- SNAVELY, N., SEITZ, S. M., AND SZELISKI, R. 2006. Photo tourism: Exploring photo collections in 3D. *ACM Transactions on Graphics (SIGGRAPH Proceedings)* 25(3), 835–846.
- STAMOS, I., AND ALLEN, P. 2001. Automatic registration of 2-D with 3-D imagery in urban environments. In *ICCV'01*, II: 731–736.
- STUMPFEL, J., TCHOU, C., YUN, N., MARTINEZ, P., HAWKINS, T., JONES, A., EMERSON, B., AND DEBEVEC, P. 2003. Digital reunification of the parthenon and its sculptures. In *VAST 2003: 4th International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage*, 41–50.
- SZELISKI, R. 2006. Image alignment and stitching: A tutorial. *Foundations and Trends in Computer Graphics and Vision* 2, 1.
- WILLIAMS, N., LOW, K.-L., HANTAK, C., POLLEFEYS, M., AND LASTRA, A. 2004. Automatic image alignment for 3D environment modeling. *17th ACM Brazilian Symposium on Computer Graphics and Image Processing*, 388–395.
- ZHAO, W., NISTER, D., AND HSU, S. 2005. Alignment of continuous video onto 3D point clouds. *IEEE Trans. Pattern Anal. Mach. Intell.* 27, 8, 1305–1318.