Registration and Fusion of Intensity and Range Data for 3D Modelling of Real World Scenes

Paulo Dias
Univ. Aveiro/IEETA
Portugal
paulo.dias@ieeta.pt

Vítor Sequeira
EC, Joint Research Centre, Italy
vitor.sequeira@jrc.it

Francisco Vaz
Univ. Aveiro/IEETA
Portugal
fvaz@ieeta.pt

João G.M. Gonçalves
EC, Joint Research Centre, Italy
joao.goncalves@jrc.it

Abstract

In this paper, a novel technique combining intensity and range data is presented. Passive (intensity based) and active (range based) techniques used for 3D reconstruction have their limitations and separately, none of these techniques can solve all the problems inherent to the modelling of real environments. Our technique aims to demonstrate how both intensity and range data can be registered and combined into a long-range 3D system. The procedure needs an initial estimation for internal and external camera parameters for two or more intensity images. The technique uses passive triangulation of the intensity data to refine the initial camera calibrations and ensure a good registration of range and video data sets. Once a reliable calibration is achieved, corresponding points from the intensity images are triangulated and introduced in the original range data. With our technique, it is possible to complete the models in areas where data is missing or to increase the resolution in areas of high interest and 3D contents.

1. Introduction

3D reconstruction techniques are typically divided into two groups. The first one is composed by active techniques based on range sensors measuring directly the distance between the sensor and points in the real world [1,2,3]. Techniques in the second group are passive and get the 3D information from several 2D digital photographs or video sequences (such as in stereo or photogrammetry) [4,5,6]. A comparative evaluation of the performance of active and passive 3D vision systems appears in [7]. Each one of these techniques has qualities and limitations. Range sensors provide directly precise and accurate 3D points. In addition they are independent from external lighting conditions and do not need any texture to perform well. However, they tend to be expensive, slow to acquire and normally have a limited spatial resolution. On the other side, Intensity images have high resolution that permits very accurate results on well-defined targets. They are easy to acquire and provide texture-maps based on real colour-photographs. Nevertheless, the lack of direct relation between images and depth of the measured object is a major problem for intensity-based techniques. For this reason, the acquired images need to be rich in texture information and the resulting data sets will have varying density depending on the texture of the images. Finally these sensors are affected by changes in the illumination. Table 1 presents the main advantages and limitations of both techniques (advantages are highlighted in shading).

Table 1. Advantages (shaded) and limitations of Range and Intensity based techniques

	Range	*Intensity*
Cost	*Expensive sensors*	*Low cost.*
Acquisition	*Often difficult with large sensors*	*Easy, with a digital camera*
Resolution	*Limited spatial resolution*	*High-resolution digital photos*
Texture map	*Black and white reflectance*	*Provide a realistic colour texture map*
Lighting	*Independent from external lighting*	*Highly dependent on lighting conditions*
Texture relevance	*No need of texture in scene*	*Texture is crucial for good results*
3D processing	*Provide directly 3D measurements*	*Difficult to extract 3D from images*

The above list makes clear how range and intensity images can complete each other. Thus, our aim in this work is to combine them in order to compensate limitations of each sensor with qualities of the other one.

2. Related work

Some work to register range and intensity information exists but it is normally limited to constrained situations. In [8] pose estimation is done by matching control points over similar intensity and range images. Another technique proposed in [3] achieves image to range registration based on vanishing lines and orthogonality conditions but the method is only reliable in scenes with strong geometric contents. More recently, there were some trials using silhouettes for matching intensity and range [9,10,11]. However, most of these methods are optimised for small objects with well-defined contours and not large real world scenes such as the ones considered in this paper.

In previous work a semi automatic registration system was developed [12]. It first aligns the range and intensity data by evaluating the planar transform based on the edges extracted in the reflectance and intensity images. In a second step, correspondences are automatically extracted between the images and feed a Tsai camera calibration algorithm to evaluate the camera internal and external parameters [13]. The method also offers the possibility to add correspondences manually in order to guide and correct the automatic calibration if necessary. More details about the algorithm can be found in [12].

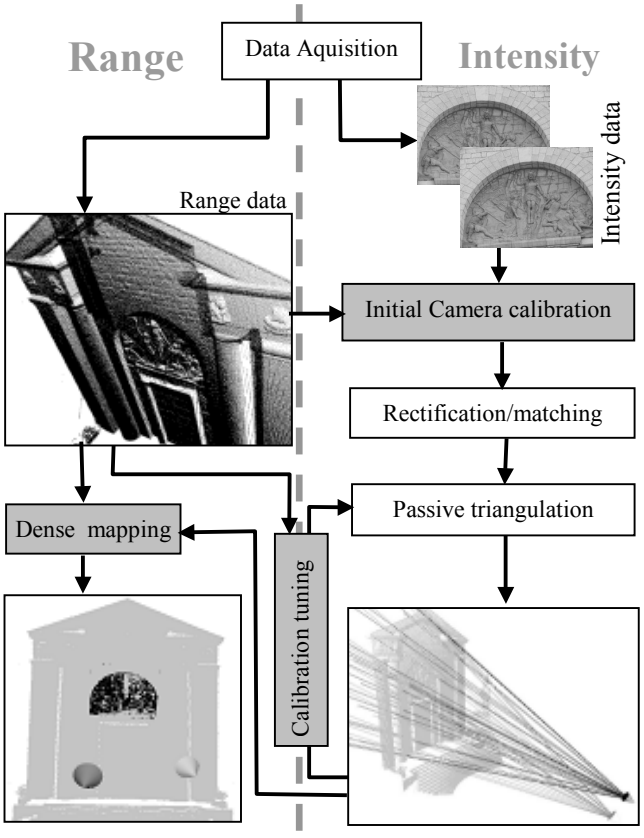

Figure 1. The passive triangulation process and its applications

The quality of the initial registration obtained with this semi-automatic process relies mainly on the precision of the correspondences, for this reason, additional tools to refine the initial calibration are valuable. The work in this paper combine several registered digital photographs with traditional photogrammetry tools in order to compute triangulated points from intensity images. These triangulated points can be useful first, to refine the original camera calibration and secondly to add 3D points into the range data giving the possibility to increase 3D information density or fill holes in the model.

Figure 1 gives an overview of the whole process, in shaded are the steps where range and intensity data are compared and/or combined. Section 3 presents with more details the rectification and matching of the image and the passive triangulation process. The calibration tuning process is described in section 4, and section 5 present the dense mapping step. For the initial calibration, more details can be found in previous work [12].

3. The passive triangulation procedure

3.1. Fundamental matrix estimation

The main difficulty to overcome when working with intensity images in 3D reconstruction is the matching problem: how to get accurate corresponding points over different images? The technique we adopted uses the fundamental matrix to link intensity images through epipolar geometry, according to the following equation:

$$\boldsymbol{x}_i'^{T} F \boldsymbol{x}_i = 0 \qquad (1)$$

where, $\boldsymbol{x}=(x, y, \zeta)$ is a homogeneous co-ordinate in the first image, $\boldsymbol{x}'$ is the corresponding co-ordinate in the second image and F is the 3×3 Fundamental matrix [4,14].

Our technique involves the computation of an initial estimation for the camera internal and external parameters based on the automatic registration of laser reflectance and colour intensity images [12].

The initial camera calibrations are used to guide a cross correlation-matching algorithm over the intensity images. The correspondences feed a Ransac 8-point algorithm that computes the fundamental matrix between images [14,15]. This matrix is used to find corresponding epipolar lines and build the rectified images where rows correspond to the same epipolar lines. Using rectified images to compute correspondences between images reduces the matching problem from two to one dimension.

Figure 2 present two digital photographs from a church in Laveno/Italy used to test our algorithm. In this figure, we also display the epipolar lines obtained through the fundamental matrix estimation step. The rectified images computed with these epipolar lines appear in Figure 3.

Figure 2. Original images of the Laveno Church and epipolar lines

Figure 3. Rectified images from the Laveno church model

3.2. Passive triangulation

The rectified images computed in 3.1 are used to find corresponding points. A canny edge detector is applied in one of the intensity images to detect points of interest. The output of the edge detector is then sub-sampled to reduce the number of matching candidates. For each selected point of interest, cross correlation is computed along the epipolar line to find the best possible matching. Once more, initial calibrations are used to estimate the probable position of the matching pixel and reduce the search area.

Sub-pixel accuracy is obtained by fitting a second-degree curve to the correlation coefficients in the neighbourhood of the disparity. The maximum of the function occurs where the gradient of the function is null, and thus its position is computed with the following formula [16]:

$$x = d + \frac{1}{2} \times \frac{C(d-1) - C(d+1)}{C(d-1) - 2C(d) + C(d+1)} \qquad (2)$$

where $C(d)$ is cross correlation coefficient at position d.
x is the sub-pixel position of the maximum.

Two additional conditions are considered when seeking for correspondences over the intensity images:

- A threshold is used to avoid the consideration of points with a low cross correlation.
- The matching must be symmetric. If a point p_1 has the highest cross correlation with point p_2, then p_2 must also have its maximum cross correlation with p_1. A maximum disparity between the two symmetric matches is considered.

The matching points in the images are used to compute the camera projective rays by inverting the perspective projection equations from Tsai model. The result is the parametric equation of the 2 rays:

$$\vec{L}_0(s) = \vec{B}_0 + s\vec{M}_0 \qquad (3)$$
$$\vec{L}_1(t) = \vec{B}_1 + t\vec{M}_1 \qquad (4)$$

The squared-distance function is then:

$$Q(s,t) = \left|\vec{L}_0(s) - \vec{L}_1(t)\right|^2 \qquad (5)$$
$$Q(s,t) = as^2 + 2bst + ct^2 + 2ds + 2et + f \qquad (6)$$

where,

$$a = \vec{M}_0 \bullet \vec{M}_0, b = -\vec{M}_0 \bullet \vec{M}_1, c = \vec{M}_1 \bullet \vec{M}_1$$
$$d = \vec{M}_0 \bullet (\vec{B}_0 - \vec{B}_1), e = -\vec{M}_1 \bullet (\vec{B}_0 - \vec{B}_1)$$
$$f = (\vec{B}_0 - \vec{B}_1) \bullet (\vec{B}_0 - \vec{B}_1)$$

The minimum distance between rays occurs at the point where the gradient of the function is null:

$$\vec{\nabla} Q = 2(as + bt + d, bs + ct + e) = (0,0) \qquad (7)$$

A user threshold defines the maximum distance between two rays above which triangulated points are not considered. Otherwise, the triangulated point is computed as the centre of the segment between the two closest points in the rays [17].

Figure 4 illustrates the triangulation process. It presents the range cloud of points, the position of the camera for the two considered images (as two cones) and the projective rays for a few points.

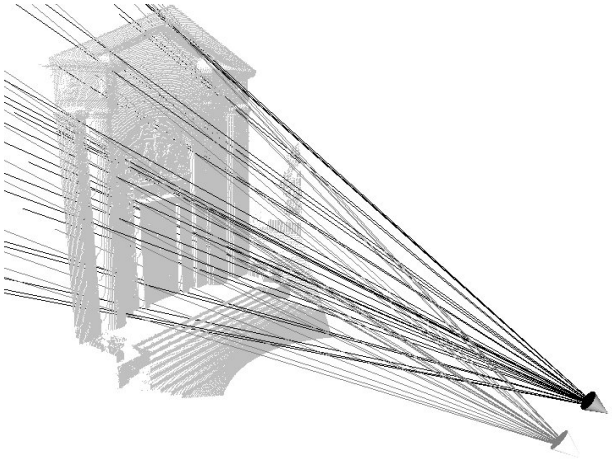

Figure 4. The passive triangulation process

4. Refining the calibrations

4.1. Association triangulated/range 3D points

The process presented in 3.2 gives, for each point of interest, the 3D position of the triangulated point. It is possible to use this information to evaluate the quality of the camera calibration, by measuring the distance from the triangulated points to the 3D cloud acquired with the laser. To optimise this computation, the range cloud of points is bucketed into referenced small cubes. This allows for a fast navigation inside the cloud of points and a fast computation of the closest range points. In our process, we only consider 3D discontinuities, corresponding to areas of the range data rich in geometrical contents and thus easier to match precisely. The discontinuities are detected by analysing each 3D range point neighbourhood. 3D points where high variations in the depth occur are marked as discontinuities [18]. Figure 5 presents the detected discontinuities in the range image of the Laveno church.

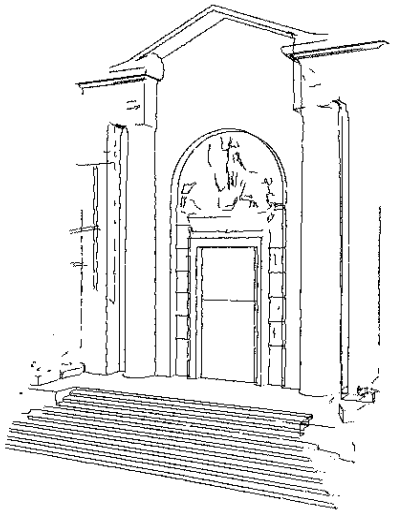

Figure 5. Orientation discontinuities

For each triangulated point, the closest orientation discontinuity point in the range image is found. Figure 6 presents, in the same image, the cloud of points from the range image and the 3D triangulated points obtained from the initial calibrations. The segments in the figure indicate the triangulated points and their closest orientation discontinuity in the range data. Points for which the distance between the triangulated and range discontinuity is larger than a given distance are not considered in the iteration to compute the new camera model.

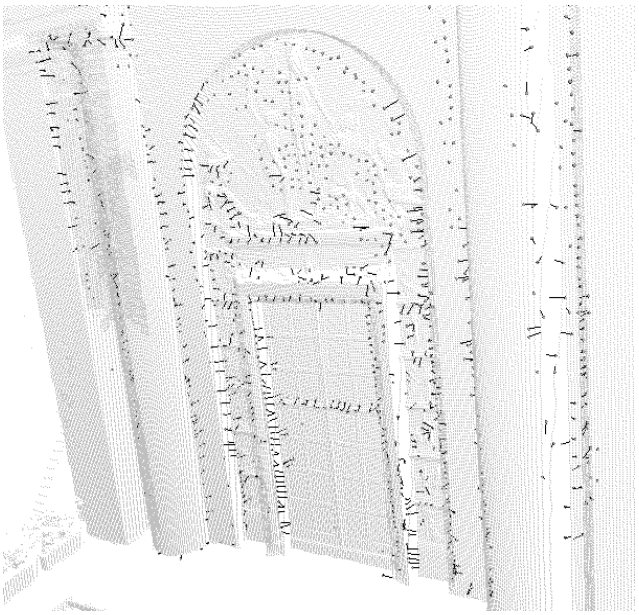

Figure 6. Range points (light grey) and intensity triangulated points (black) with the initial camera model estimations

4.2. Iterative calibration process

To improve the calibration and force triangulated points to converge into the 3D cloud, the closest orientation discontinuity in the range image is used with the matching points to perform a new Tsai camera calibration for each image. The process is iterative and points are triangulated and compared with the range points in the 3D cloud, in each loop of the cycle.

This process continues as long as the average distance between triangulated and range points decreases. At this stage an additional optimisation is introduced in the loop to correct matching errors in the intensity images. In these new cycles of the process, pixel positions of the correspondences are updated using the current camera model. The closest 3D range orientation discontinuity (used to calibrate the cameras) are re-projected into the images, and the new match coordinates in every intensity

image are computed as the centre between the re-projected point and the original position of the matching. This process allows for a fine-tuning of the camera calibration parameters by updating the correspondences co-ordinates but it is introduced only in a final step of the process when the cameras models are already close to a final stable solution. Figure 7 presents the results of the triangulation at the end of the process using the optimised camera parameters. The graphic in Figure 8 illustrates the evolution of the average distance between the range points and the intensity based triangulated points. In this example, the additional optimisation in the matching begins at iteration number ten.

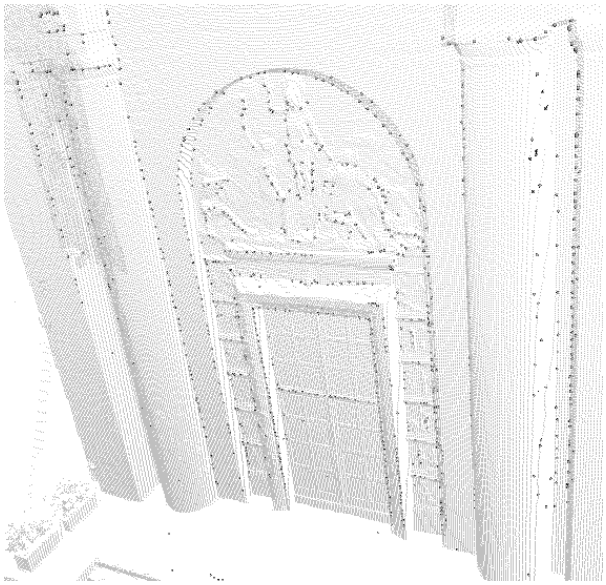

Figure 7. Range points (light grey) and the intensity triangulated points (black) after the optimisation process

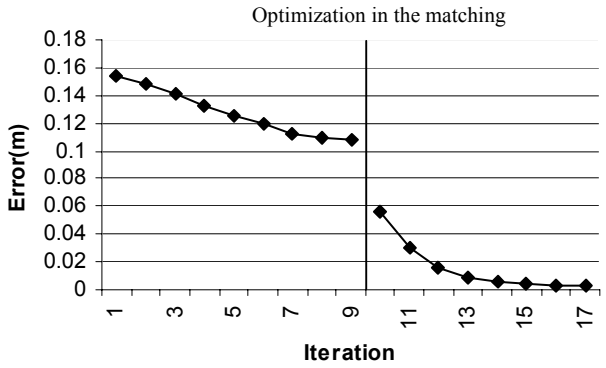

Figure 8. Evolution of the error (m) along the optimisation process

Regarding processing times, it takes approximately 8 minutes on a PIV 2GHz to perform the whole optimisation loop (17 iterations with 3006 triangulated points).

The example presented here uses two intensity images, but the passive triangulation can be extended to as many images as desired: the process presented in 3.2 is applied to each pair of images and the final triangulated point is the centre of gravity between all the triangulated points. In this case, in the calibration-tuning loop, the model of all the cameras is updated in each iteration. The process to compute the triangulated point with three cameras is illustrated in Figure 9.

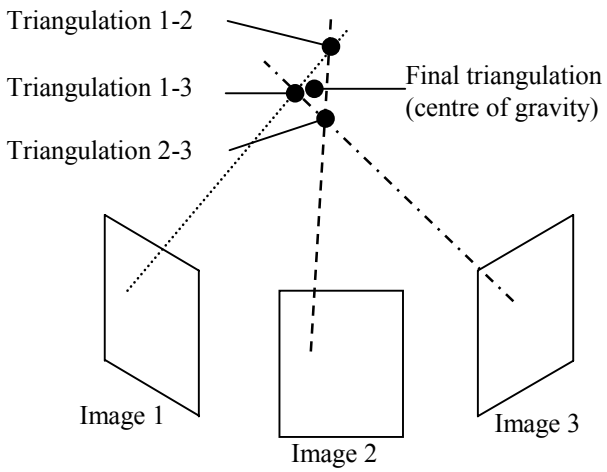

Figure 9: Passive triangulation with multiple images

5. Adding points to the 3D range cloud

The tools developed in the previous sections ensure that the cameras are well registered with the range data. These tools offer a new possibility when computing our 3D models: they permit to add new 3D points to the original range cloud. This can be particularly useful in areas where range data is missing (non reflective areas, occlusions, missing scans, etc.) or in parts of the models highly textured and rich in 3D content. In these situations, additional 3D data can be computed from pairs of intensity images to compensate for the lack of range data or to increase the 3D point density. Intensity images can be a valuable source of data since it is possible to acquire them easily, fast and with a high resolution.

The process here is approximately the same as the one presented in section 2 but applied to as many points as possible. This corresponds to the dense depth-mapping step in stereo/photogrammetry techniques. The main difference is due to the fact that the range is not used to guide the matching anymore since we are trying to add

data in region where range information may not be available.

The matching is done along epipolar lines considering that the images were taken from close viewpoints (such as in stereo techniques) meaning that matching points are spatially close in both images. As in section 2, thresholds for cross correlation as well as the symmetry condition are used during the matching phase. In addition, we also consider an ordering condition to guarantee that the order of detected correspondences in the two images is the same. These conditions are widely used in dynamic matching techniques [19].

Practically, all points where the variation of the gradient is significant are triangulated. Only uniform areas where the matching is not reliable are not considered (e.g. large areas of same colour such as walls).

The triangulated points can then be introduced into the range cloud of points before entering a 3D reconstruction process, but with data coming from both range and intensity sensors

In Figure 10 we present some results obtained on the Laveno church. This example is the one used along this document to illustrate our methods, and the intensity images are the ones presented in Figure 2.

In this example, the range data was acquired with a Zoller and Fröhlich IMAGER 5003 laser scanner [20]. The resolution of the obtained range image is 906x1036 measured points. The photographs were acquired with a Canon PowerShot Pro70 digital camera and have a final resolution of 1024x1536 pixels. Our algorithm introduced 608695 additional 3D points from passive triangulation of intensity images into the range cloud of points.

Figure 11 presents similar results with the same range scan of the Laveno church. The two digital photographs were acquired with a Canon Powershot Pro90 camera with a resolution of 1856x1392 pixels. In this example the images cover only a reduced part of the range data and the high resolution of the photographs is used to increase significantly the number of 3D points in the area above the main door of the church. 483387 points were introduced into the original range cloud of points in this case.

Figure 12 presents the results obtained with a model of the San Stefano church in Verona, Italy. In this case, the range image was acquired with a Cyrax 2500 laser scanner. Digital photographs were taken with a Canon ProShot G2 digital camera. The resulting resolutions were 850x850 for the range image and 2272x1704 pixels for the digital photographs. Thanks to our method, 743661 additional points were added in the range data.

These examples show clearly that our processes are completely independent from the sensors since different cameras and laser were used in the three examples presented.

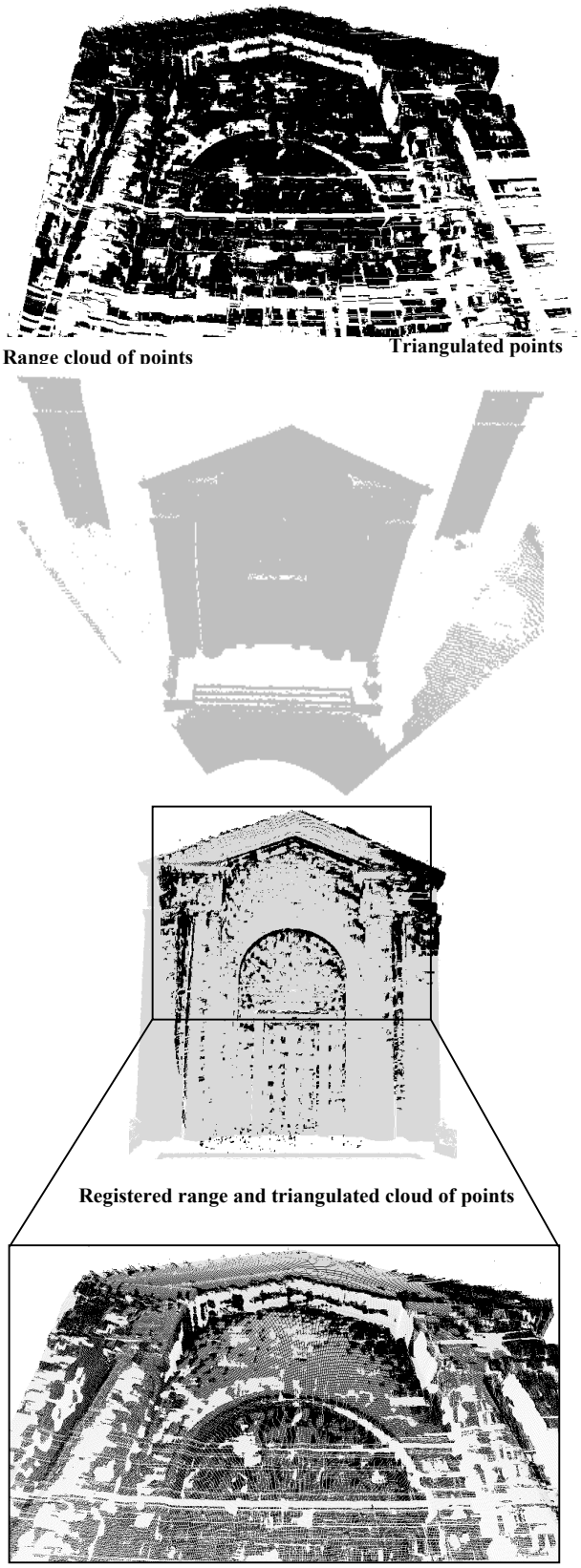

Figure 10. Addition of points in the Laveno church

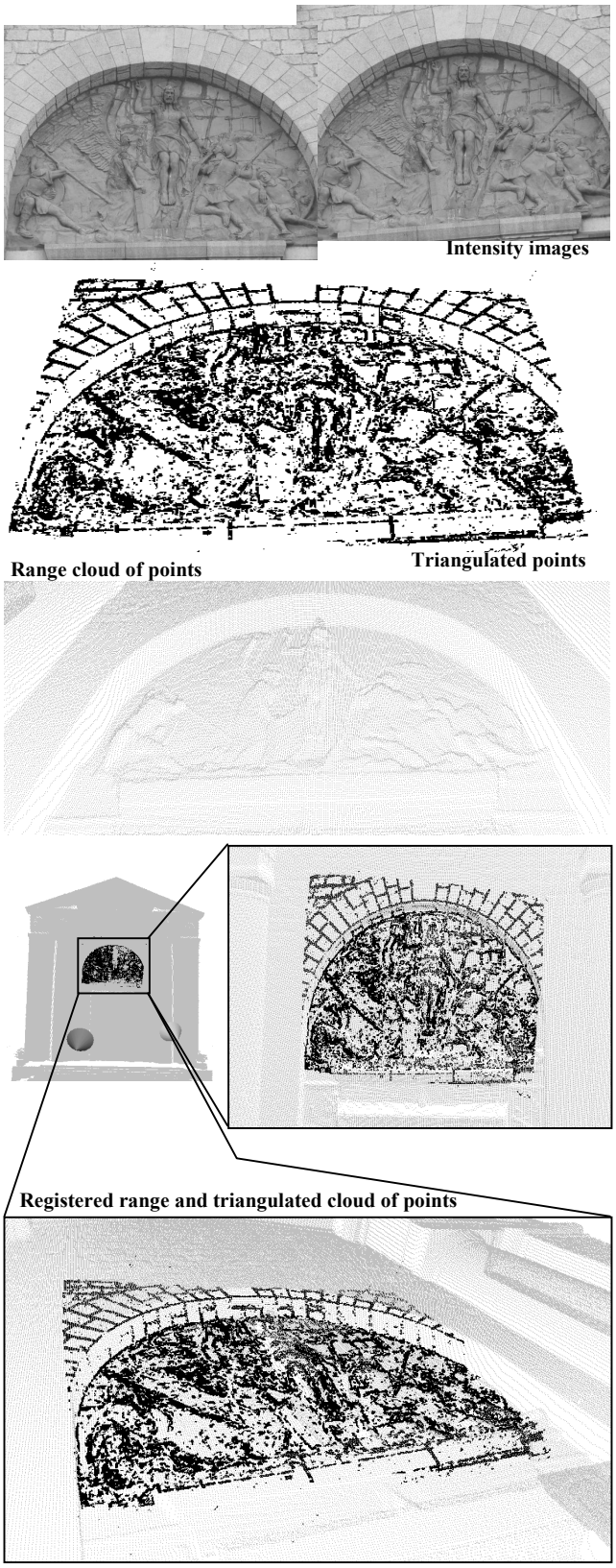

Figure 11. Second example with the Laveno church

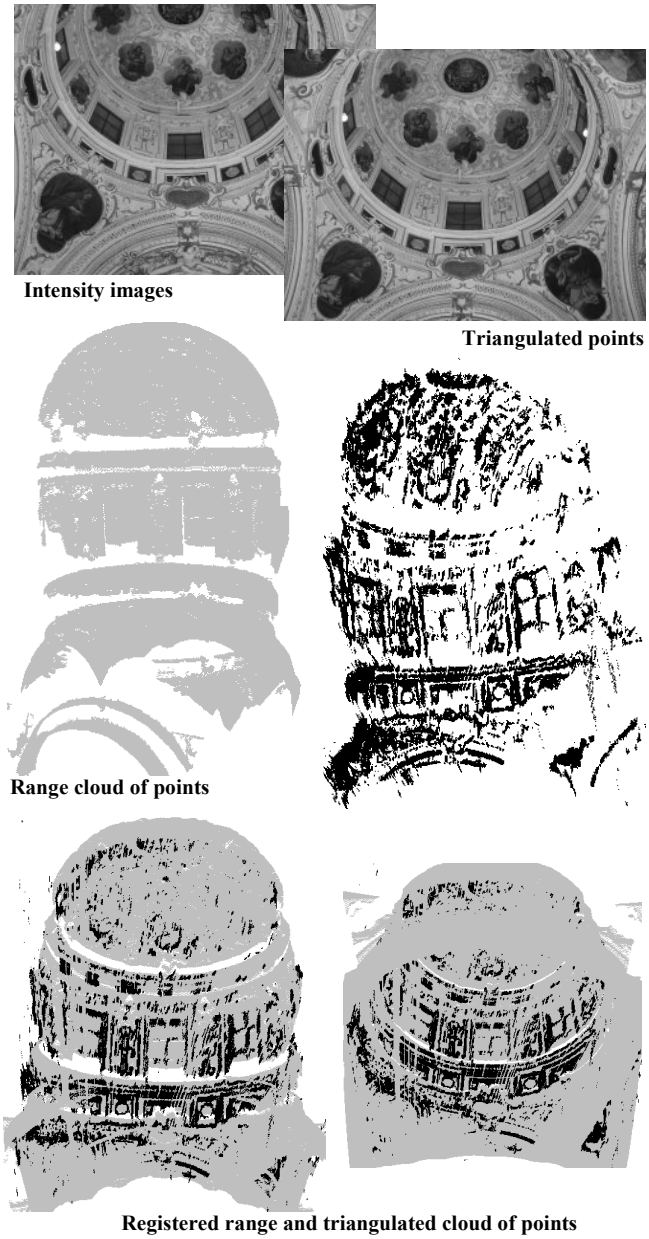

Figure 12. Addition of points in San Stefano church (Verona, Italy)

6. Discussion and conclusions

This paper demonstrates the possibility to use passive triangulation of intensity images to improve 3D models computed from range data. A method was presented to improve initial camera calibrations by measuring the error between range data and points that are triangulated from two or more intensity images. Once the data (range and intensity) is fully registered, the 3D information can come from the two sources allowing to select the best data, for a

given area of the model or for a given application. The intensity data can be used to add 3D points in some areas of the model, as shown with several examples acquired with different sensors. In these examples, the density of 3D points has been significantly increased in some parts of the final model using digital photographs.

The quality of the process depends mainly on two factors. The original parameters of the cameras can influence the process used to refine the camera calibrations since they are used as initial estimation. This means that if images are badly registered initially, the algorithm cannot compensate and the result is of poor quality. The other important factor on the process is the quality of the 3D edge detection in range data, since orientation discontinuities in the cloud of points are used in the calibration refining process to select the closest 3D point for the next calibration. Fortunately the resolution of new lasers has increased making detection of discontinuities easier and much more reliable.

Finally, the acquisition of intensity data is also an important matter. Because of the dense mapping step, it is important to acquire data from close viewpoints and rich in texture information, in order to ensure a reliable matching between features. On the other side if the photographs are acquired from too close, the passive triangulation will result more inaccurate.

7. Acknowledgments

This research was partly funded by the EC CAMERA TMR (Training and Mobility of Researchers) network, contract number ERBFMRXCT970127 and by the Portuguese Foundation for Science and Technology through the Ph.D. grant PRAXIS XXI/BD/19555/99.

8. References

[1] V. Sequeira, K. Ng, E. Wolfart, J.G.M. Gonçalves, D. Hogg, "Automated Reconstruction of 3D Models from Real Environments", *ISPRS Journal of Photogrammetry and Remote Sensing (Elsevier)*, vol. 54, pp. 1-22, 1999.

[2] S. F. El-Hakim, C. Brenner, G. Roth, "A multi-sensor approach to creating accurate virtual environments", *ISPRS Journal for Photogrammetry and Remote Sensing*, Vol. 53(6), pp. 379-391, December 1998.

[3] I. Stamos, P. K. Allen, "3-D Model Construction Using Range and Image Data", IEEE International Conference on Computer Vision and Pattern Recognition, pages 531-536 (volume I), South Carolina, June 2000.

[4] Faugeras O.D., *Three-Dimensional Computer Vision: A Geometric Viewpoint,* Cambridge, MA: Mit Press, 1993.

[5] M. Pollefeys, *Self-Calibration and Metric 3D Reconstruction from Uncalibrated Image Sequences,* Ph.D. Thesis, Dept. of Electrical, Katholieke Universiteit Leuven, 1999.

[6] P. Debevec, "Reconstructing and Augmenting Architecture with Image-Based Modelling, Rendering and Lighthing", Proceedings of the International Symposium on Virtual Architecture (VAA'01), pp. 1-10, Dublin 21-22 June 2001.

[7] S.F. El-Hakim, J.A. Beraldin, F. Blais, "A Comparative Evaluation of the Performance of Passive and Active 3-D Vision Systems", *SPIE Proc. 2646*, *St.Petersburg Conf. on Digital Photogrammetry*, pp. 14-25, June 1995.

[8] P. W. Smith, M. D. Elstrom, "Stereo Based Registration of Multi-Sensor Imagery for Enhanced Visualization of Remote Environments", Proceedings of the IEEE International Conference on Robotics and Automation, pp. 1948-1953, May 1999.

[9] P. Neugebauer, K. Klein, "Texturing 3D Models of Real World objects from multiple unregistered photographic views", *Eurographics*, Vol. 18(3), pp. 245-256, September 1999

[10] H. Lensch, W. Heidrich, H.P. Seidel, "Automated Texture Registration and Stitching for Real World Models", Pacific Graphics 2000, Hong Kong, pp 317-327, October 2000.

[11] R. Kurazume, K. Nishino, Z. Zhang, and K. Ikeuchi, "Simultaneous 2D images and 3D geometric model registration for texture mapping utilizing reflectance attribute", Proceedings of 5th Asian Conference on Computer Vision (ACCV), Vol. I, pp. 99-106, January 2002.

[12] P. Dias, V. Sequeira, J.G.M. Gonçalves, F. Vaz, "Automatic Registration of Laser Reflectance and Colour Intensity Images for 3D Reconstruction", *Robotics and Autonomous Systems*, Vol. 39 (3-4), pp. 157-168, June 2002.

[13] R.Y. Tsai, "A versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses", *IEEE Journal of Robotics and automation*, Vol. RA-3, No. 4, pp. 323-344, August 1987.

[14] P.H.S. Torr, D.W. Murray, "The Development and Comparison of Robust Methods for Estimating the Fundamental Matrix", *International Journal on Computer Vision*, Vol. 24, No. 3, pp. 271-300, September 1997.

[15] M.A. Fischler, R.C. Bolles, "Random Sample Consensus: a paradigm for model fitting with application to image analysis and automated cartography", *Communications of the ACM 24 (6)*, pp. 381-395, 1981.

[16] C. Sun, "Fast Stereo Matching Using Rectangular Subregioning and 3D Maximum-Surface Techniques", *International Journal of Computer Vision*, No.1, Vol 47, pp. 99-117, May 2002.

[17] Eberly D.H., *3D Game Engine Design: A Practical Approach to Real-Time Computer Graphics,* Morgan Kaufmann Publishers, September 2000.

[18] V. Sequeira, *Active Range Sensing for Three-Dimensional Environment Reconstruction,* PhD Thesis, Dept. of Electrical and Computer Engineering, IST-Technical University of Lisbon, Portugal, 1996.

[19] I. Cox, S. Hingorani, and S. Rao, "A Maximum Likelihood Stereo Algorithm", Computer Vision and Image Understanding, Vol. 63, No.3, pp. 542-567, May 1996.

[20] C. Fröhlich C., M. Mettenleiter, F. Haertl, "Imaging Laser Radar for High-Speed Monitoring of the Environment", Zoller&Fröhlich technical report 1998.