Automatic Registration of Laser Reflectance and Colour Intensity Images for 3D Reconstruction

Dias P.[(1)*], Sequeira V.[(1)], Gonçalves J.G.M[(1)], Vaz F. [(2)]

[(1)] *European Commission-Joint Research Centre, TP 270, 21020 Ispra (VA), Italy*
[(2)] *University of Aveiro/IEETA, 3800 Aveiro, Portugal*

Abstract. The objective of the work presented in this paper is to generate complete, high-resolution models of real world scenes from passive intensity images and active range sensors. In previous work, an automatic method has been developed in order to compute 3D models of real world scenes from laser range data. The aim of this project is to improve these existing models by fusing range and intensity data. This paper presents different techniques in order to find correspondences between the different sets of data. Based on these control points, a robust camera calibration is computed with a minimal user intervention in order to avoid the fastidious *point and click* phase that is still necessary in many systems. The intensity images are then reprojected into the laser coordinate frame to produce an image that combines the laser reflectance and the available video intensity images into a colour texture map.

Keywords : 3D Reconstruction; automatic registration; camera calibration; texture map.

1 Introduction

With the increase of the computer capabilities, 3D models are becoming of higher interest as they can be displayed and used in a home PC without requiring dedicated hardware. Commercial products are already available to compute 3D models, e.g. small object for Internet application. On the other hand, real-world scenes are often too complex for realistic model acquisition.

This work tries to combine the information coming from the most typical 3D sensors (Laser Range Finders and Video) in order to overcome these limitations. Almost every laser returns for each measurement, the distance to the point and the reflected light intensity. This leads to two different sets of data. From the distance measurement, the 3D models can be computed, whereas an image can be generated from the reflectance information.

The choice of using the reflectance image rather than the 3D model for the registration was based on several considerations: first, the reflectance image is a 2D (normally, infrared) image such as the video intensity image, which allows the use of already existing techniques for matching intensity images [Heipke96]. In addition, the reprojection of the intensity image into the reflectance image is a very intuitive and fast way of evaluating the registration.

Unfortunately, the images obtained with the laser and the video camera are very different. The camera is a high-resolution passive sensor whereas the laser is an active sensor of lower resolution. To overcome these differences a feature and area based matching algorithm was developed.

Compared to other registration techniques [Elstrom98, Sequeira99], this one is independent from the experimental set-up, does not need any previous calibration and minimises user intervention.

2 Matching laser reflectance with intensity images

The matching algorithm is divided in two main steps: rescaling and matching. Rescaling resizes the video image to match the dimensions of the reflectance image. Matching is used to locate corresponding features.

2.1 Rescaling algorithm

Traditional matching techniques used in photogrammetry or stereo vision [Faugeras93] use images obtained from the same sensors, i.e. same size and same resolution. In our specific case, an intermediate step is required to rescale the two images before using any matching procedure.

* Corresponding author. Tel.: +39-0332-785233; Fax: +39-0332-789185; E-mail: paulo.dias@jrc.it

This rescaling step approximates the planar transformation between the images using five parameters: the planar translation coordinates, the scaling factors for both x and y-axis, and the planar rotation between both images. The result is a coarse approximation of the transformation between the two images but with the main features overlapping enough to allow the use of matching algorithms.

The flowchart in Figure 1 presents the algorithm. In a first step, a Canny edge detector [Canny86] and the distance transform [Borgefors84, Borgefors86] are applied to the video image. The reflectance edges are then iteratively translated, scaled and rotated over the distance transform image until the transformation that leads to a minimum distance between the edges is found.

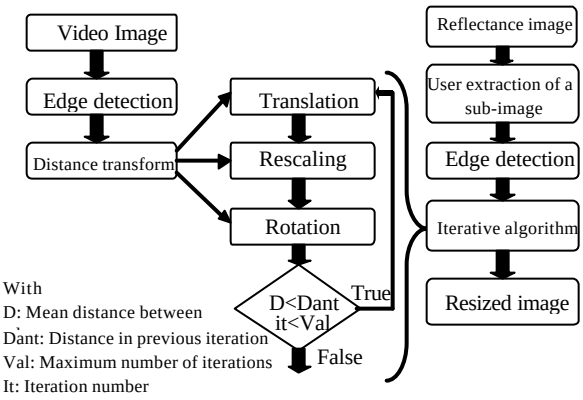
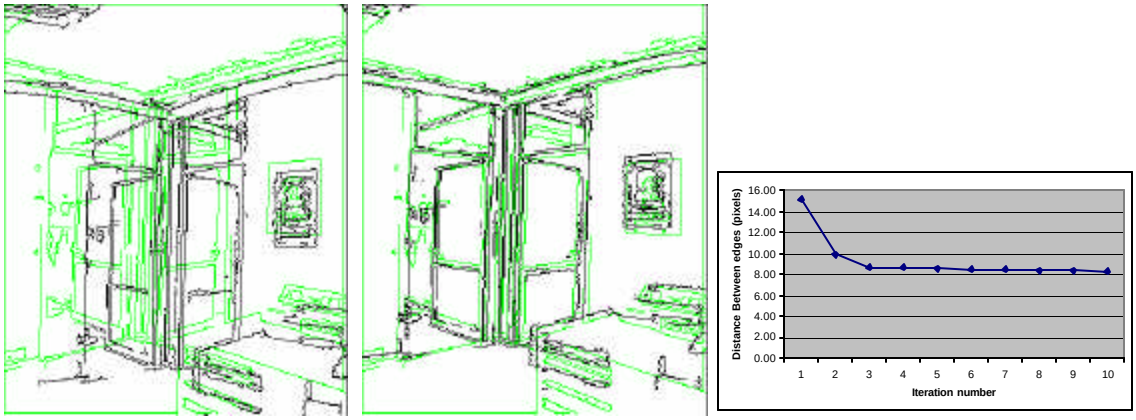


Fig. 1: Flowchart of the rescaling procedure

The results of the rescaling procedure are presented in Figure 2. Figure 2(a) shows the superposition of the edges of the video and the reflectance before running the process. Figure 2(b) presents the results after applying the algorithm with the reflectance edges rescaled. Figure 2(c) is the evolution of the average distance between the edges of both images along the rescaling procedure. Despite the average error of eight pixels, the result of the rescaling is sufficient for the subsequent matching step.



(a) **(b)** **(c)**

Fig. 2: Superposition of edges in a laboratory scene (a) before and (b) after the rescaling and (c) distance between the edges

2.2 Point based matching

The next step of the matching procedure is to find some control points over the images. A Harris corner detector [Harris88] is used to find points of interest based on the following formula:

$$R(x,y) = A(x,y)*B(x,y) - C^2(x,y) - Weight*(A(x,y)+B(x,y))^2$$

A, B: Smoothed direction derivatives;

C: Smoothing of the products of the direction derivatives. The derivatives are calculated with a Sobel operator and smoothed with a Gaussian filter;

Weight:User defined constant. Here, weight=0.04, so grayvalue corners return positive values for R(x,y), while straight edges receive negative values.

For each point of interest, cross correlation is computed over the gradient images (to make the matching more invariant to changes in illumination). As the images are already rescaled, the cross correlation can be computed only in the neighbourhood of each pixel speeding up the process and reducing the number of bad matches.

Even in these conditions, the number of outliers is still significant due to the different nature of the images. To discard bad matching pairs, the following criteria are used:

i) If various matches are accumulated on a small area, only the one with the higher cross correlation is considered: a matching pair *i* is considered valid only if:

$$Correlation(x_i, y_i) = \max(correlation(x_j, y_j))$$

$$for\ each\ j:\ \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \leq Threshold$$

where (x_i,y_i) are the coordinates of the video pixel of the matching 'i'.

ii) The average distance between the matching are computed over the images and only the pairs for which distance is lower than a given threshold are considered.

$$\sqrt{(x_v - x_r)^2 + (y_v - y_r)^2} < \frac{Dist}{2}, \text{ with}$$

$$Dist = \frac{\sum_{i=1}^{n} \sqrt{(x_{v_i} - x_{ri})^2 + (y_{v_i} - y_{ri})^2}}{n}$$

(x_v , y_v) : coordinate of pixel in video image;
(x_r , y_r) : coordinate of corresponding pixel in reflectance image;
n : number of matching

The results of this matching procedure are presented in Figure 3 for the laboratory images. In this case, 50 matching are detected and considered valid.

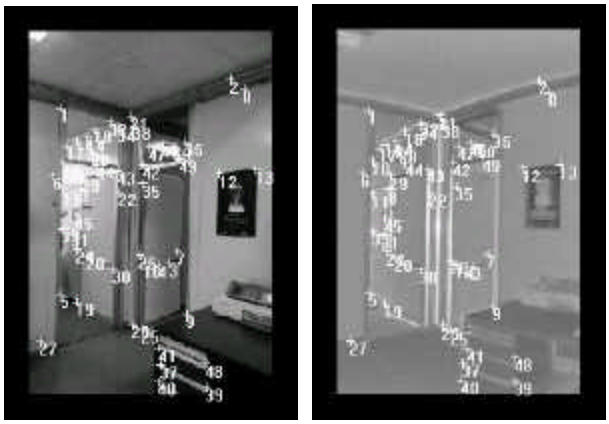


Fig. 3: Point Based Matching in the laboratory image.

3 Camera calibration

3.1 Camera model

A Tsai camera calibration technique [Tsai87] is used to calibrate the camera. This pinhole perspective projection model is based on eleven parameters. Six extrinsic parameters characterised the position and orientation of the camera; the other five describe the camera properties (such as focal length and radial distortion).

3.2 Robust estimation

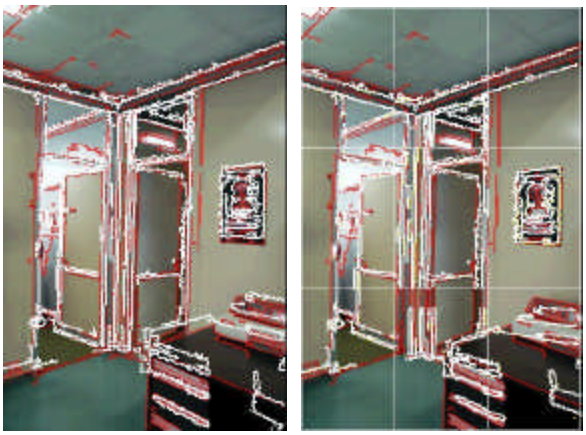
Due to the difficulty of matching the reflectance and the video image, a relatively large number of mismatches will appear during the matching step. Nevertheless, a good calibration for the camera is computed based on a RANSAC estimation technique [Fischler&Bolles81]. A calibration is computed with a sub-set of random points. For each of the "random calibration", the one with the largest support is selected, excluding most of the outliers. The algorithm needs 3 parameters: the number of trials; the sub-set of points used in each trial for the first evaluation, and finally; the threshold to consider a valid trial.

3.3 Iterative calibration

Due to the different nature of the images, the point based matching leads sometimes to a camera calibration that is not satisfactory enough. To improve the reliability of the calibration process, an iterative procedure is used. The point based matching ensures the computation of a first approximation for the camera model. Based on this first calibration, an edge based matching algorithm will track for new matching pairs so that a better calibration can be computed.

Figure 4 illustrates the edge based matching procedure. The edges of the reflectance image are reprojected into the video image using the model of the camera coming from the previous iteration (reprojected points in Figure 4(a)). The rescaling algorithm is then used to decrease the distance between the reflectance reprojected edges and the video edges. Using the whole edges would lead to a very rigid transform and would limit improvement while running the procedure in a grid solved the problem. Each reprojected reflectance edge point is finally associated to the closest video edge point and this new set of correspondences is used in the next iteration.

The procedure stops when no further improvement appears (no significant changes are detected in the distance between edges) or when a maximum number of iteration is reached (typically two or three). Results are presented in Figure 4(b) with a 3*3 grid.



(a) **(b)**

Fig. 4: Edge based matching results for the laboratory image. (a) Reprojection of the reflectance edges using the initial camera calibration. (b) Results after applying the rescaling algorithm within each cell of the grid.

4 Reprojection

Once a model for the camera has been obtained, it is possible to associate to each 3D point its projection into the video image; this is done in the reprojection step. The aim is to obtain a texture map ready to apply to the reconstructed surfaces making the whole 3D model richer and more realistic.

In most cases, different camera views overlap in parts of the models. To avoid rough transitions in these border areas, colour blending is applied.

4.1 Reprojection and Z-buffering

Projecting the video image pixels into the reflectance image can lead to errors of projection due to occluded points as shown in Figure 5. To avoid this problem, a Z-buffer is filled with the Z coordinate of the closest 3D points for each pixel of the video image. Using this Z-buffer to handle occlusions, it is possible to select the correct video points when fusing the data.

(a) **(b)**

Fig. 5: The occlusion problem, (a) direct reprojection, (b) reprojection using Z-buffer to reject occluded points.

By just back-projecting the known 3D position, the final texture map resolution will be limited by the range image size. To overcome this limitation, the reflectance can be "zoomed" to the size of the video image, and a linear interpolation used to compute the "extras" 3D positions. Using this interpolation process it is possible to exploit the full resolution of the intensity image for the texture map.

4.2 Colour Blending

In most cases, the field of view of the laser image is much larger than the one from a normal camera. Nowadays it is possible to find lasers capable of acquiring 270° or even 360° images; thus, several video images are necessary to cover the whole reflectance image. To solve this problem, each image is back-projected into the laser coordinate frame followed by a colour blending technique to smooth the transitions between images.

The colour blending process computes for each point the number of video images that have a pixel projected in this position and the distance to the closest contour of each image. The final value of the pixel will be a weighted combination of the intensity values in each image according to the distance to the closest contour. The formula used to compute the final intensity value in each RGB image is:

$$\text{Intensity}(x, y) = \frac{\sum_{i=1}^{n} \text{Intensity}_i(x, y) * dist_i(x, y)}{DISTANCE\_TOTAL}, \text{ with}$$

$$DISTANCE\_TOTAL = \sum_{i=1}^{n} dist_i(x, y)$$

$Intensity_i(x,y)$:	*intensity of pixel (x,y) in image i.*
$dist_i(x,y)$:	*distance of pixel (x,y) to the closest border in image i.*
n:	*number of images with a pixel projected in (x,y)*

5 Experimental results

Figure 6 shows a table with the processing times for the different steps of the whole calibration/reprojection procedure for two images of the laboratory scene and the blending of the 12 images in Fig 7(b). The resolution of the images is 1850*444 for the reflectance and 1024*1536 for the video. The processing times depend on many factors, mainly, the image resolution and the number of images to merge. Blending is the more time consuming process as it is not optimised yet. Some tests have been done with simpler techniques (using averaging or rectangular borders instead of a weighted blending) and these decreases already significantly the processing time even if the blending results are of lower quality (see Figure 6(b))

	Rescaling	Matching	Calibration	Z-buffering	Reprojection
Time (sec)	7	6	20	24	6
Time(sec)	16	5	16	24	6

(a)

Average blending	42 s
Rectangular blending	114 s
Weighted blending	819 s

(b)

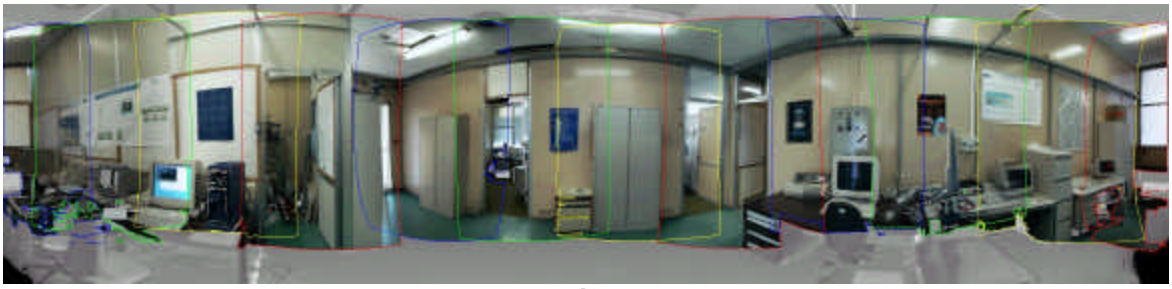
Fig. 6: Processing times of the different steps for two images of the laboratory scene

Figure 7 and Figure 8 present the different steps of the reprojection process for two real scenes. The first one is a typical laboratory scene acquired with RIEGL LMS-Z210 laser range scanner [Riegl99] and a Canon digital camera. The second one is the model of a church in the Bornholm Island (DK) acquired with the Zoller&Fröhlich LARA laser range finder [Fröhlich98] and a Minolta digital camera.

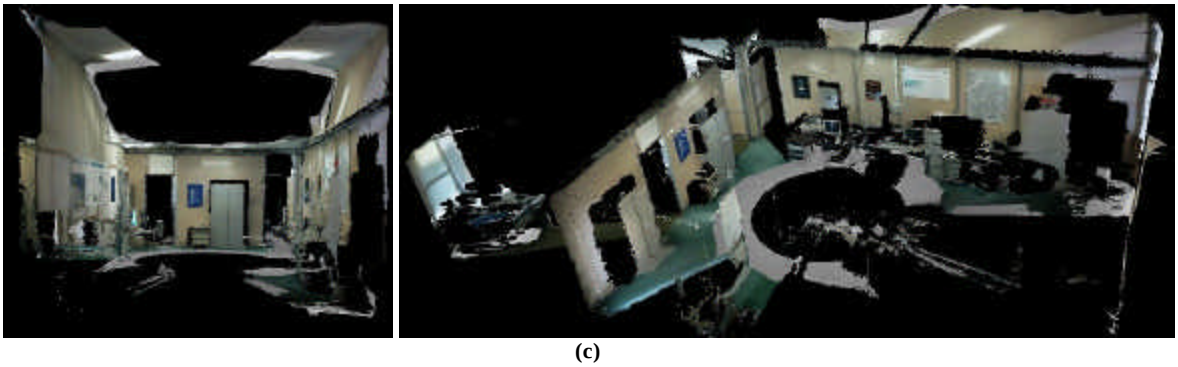
Figure 7(a) shows the reprojection of the test image, used in previous sections to illustrate the automatic matching, into the reflectance. Figure 7(b) is the final texture map obtained from the reprojection and blending of twelve images (for illustration, the borders of each reprojected image are also displayed). Figure 7(c) presents two snap-shots of the model after applying the texture map. The only intervention of the user was the selection of a rectangle in the reflectance image corresponding approximately to the same field of view of the different video images as input for the rescaling procedure. The rest of the procedure was fully automatic.

Figure 8 presents similar results for the Bornholm church model. Figure 8(a) shows the texture map obtained after the reprojection of seven video images into the reflectance image. In Figure 8(b) three snap-shots of the textured 3D model are shown. In this case some control points have been manually chosen since the video and the reflectance images were taken from very different viewpoints.

(a)

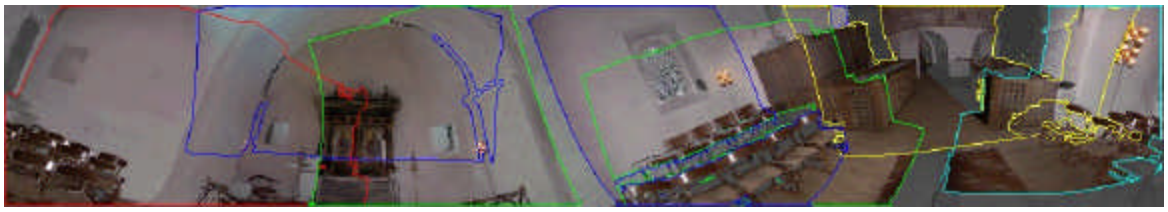


(b)



(c)

Fig. 7: Some examples of the reprojection process (a,b) and two snap-shots of the final textured model of the laboratory(c).



(a)

(b)

Fig. 8: The texture map (a) and three snap-shots (b) of the textured 3D model of the Bornholm church.

6 Conclusions

The main characteristics of the algorithm presented in this paper are: i) independence from the acquisition sensors. The results presented (laboratory and Bornholm church) were acquired with completely different experimental set-ups (lasers and cameras from two different manufacturers with different resolutions and options); ii) simple visual evaluation of the quality of the registration by superimposing the visual and range edges; iii) it enables the user to interact with the system and, when necessary, guide the process to ensure a good and reliable calibration even with images taken from very different viewpoints. The algorithm performance depends on the quantity and quality of the 2D features and on the similitude between the images (better results are obtained with images taken from close viewpoints).

The main innovations introduced are the adaptive rescaling of the datasets to locate reliable registration points, the colour blending of the merged video images and the occlusion handling for selecting the video points to be used in the final texture map.

Throughout this project, the main topic of investigation was the registration of 3D sensor data taken from a range sensor providing range and intensity data with video-based colour data. Based on this registration, a texture map for the 3D models is computed. This increases significantly the quality and the realism of the models, while keeping the initial geometry.

Acknowledgements

This research was primarily 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.

References

[Borgefors84] Borgefors G., *An improved version of the Chamfer matching algorithm.* 7th international conference on pattern recognition 1984.

[Borgefors86] Borgefors G., *A new distance transformation aproximating the Euclidean distance.* 8th international conference on pattern recognition 1986.

[Canny86] Canny J.F., *A Computacional Approach to Edge Detection* IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-8, No. 6 - pp. 679-698, 1986.

[Elstrom98] Elstrom M.D., *A stereo-based technique for the registration of color and ladar images* Master Degree thesis, University of Tennessee, Knoxville, August 1998.

[Faugeras93] Faugeras O., *Three-Dimensional Computer Vision*, MIT Press, 1993.

[Fischler&Bolles81] Fischler M.A., Bolles R.C., *Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography*. Communications of the ACM 24 (6), pp. 381-395, 1981.

[Fröhlich98] Fröhlich C., Mettenleiter M., Haertl F., *Imaging Laser Radar for High-Speed Monitoring of the Environment*. Zoller&Fröhlich technical report 1998.

[Harris88] Harris C.G., Stephens M., *A combined corner and edge detector.* In 4th Alvey Vision Conference, pages 147-151, 1988.

[Heipke96] Heipke C., *Overview of image matching techniques*. OEEPE Workshop on the application of Digital Photogrammetric Workstations, 1996.

[Riegl99] RIEGL GmbH, *Laser Mirror Scanner LMS-Z210- Technical document & user's instruction manual*. November 1999.

[Sequeira99] Sequeira V., Ng K., Wolfart E., Gonçalves J.G.M. and Hogg D., *Automated Reconstruction of 3D Models from Real Environments.* ISPRS Journal of Photogrammetry and Remote Sensing (Elsevier), vol. 54, pp. 1-22, 1999.

[Tsai87] Tsai R.Y., *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, August 1987, pages 323-344.