AUTOMATIC FUSION OF PHOTOGRAMMETRIC IMAGERY AND LASER SCANNER POINT CLOUDS

Eric K Forkuo and Bruce King
Department of Land Surveying & Geo-Informatics
The Hong Kong Polytechnic University
Hung Hom, Hong Kong
eric.forkuo@polyu.edu.hk, bruce.king@polyu.edu.hk

KEY WORDS: Laser scanning, Photogrammetry, Fusion, Matching, Registration, Multisensor, Terrestrial

ABSTRACT

Fusion of close range photogrammery and the relatively new technology of terrestrial laser scanning methods offer new opportunities for photorealistic 3D models presentation, classification of real world objects and virtual reality creation (fly through). Laser scanning technology could be seen as a complement to close-range photogrammetry. For instance, terrestrial laser scanners (TLS) have the ability to rapidly collect high-resolution 3D surface information of an object. The same type of data could be generated using close range photogrammetric (CRP) techniques, but image disparities common to close range scenes makes this an operator intensive task. The imaging systems of some TLSs do not have very high radiometric resolution whereas high-resolution digital cameras used in modern CRP do. Finally, TLSs are essentially Earth-bound whereas cameras can be moved at will around the object being imaged. This paper presents the result of an initial study into the fusion of terrestrial laser scanner generated 3D data and high-resolution digital images. Three approaches for their fusion have been investigated - data fusion which integrates data from the sensors to create synthetic perspective imagery; image fusion (synthetic perspective imagery and the intensity images); and model-based image fusion (2D intensity image and the 3D geometric model). Image registration, which includes feature detection and feature correspondence matching, is performed prior to fusion, to determine the relative rotation and translation of the digital camera relative to the laser scanner. To overcome the differences in datasets, a feature and area based matching algorithm was successfully developed and implemented. Some results of measurements on interest points and correspondence matching are presented. The result of the initial study shows that most promise is offered by model-based approaches.

1. INTRODUCTION

Of recent, close range photogrammery (CRP) and the relatively new technology of terrestrial 3D laser scanning (TLS) are used to automatically, accurately, reliably, and completely measure or map, in three-dimensions, objects, sites, or scenes. Terrestrial 3D laser scanner has the ability to rapidly collect high-resolution 3D surface information of an object or scene. The available scanning systems extend to all objects types, almost regardless of the scale and complexity (Barber *et al*, 2001). The same type of data could be generated using close range photogrammetric (CRP) techniques, but image disparities common to close range scenes makes this an operator intensive task. The imaging systems of some TLSs do not have very high radiometric resolution whereas high-resolution digital cameras used in modern CRP do. Also, TLSs are essentially Earth-bound whereas cameras can be moved at will around the object being imaged. It is intuitive then to consider the fusion of data from the two sensors to represent the objects and scenes, and to create models that are more complete, and thus easier to interpret, than a model created from the 3D point cloud data alone (Elstrom *et al*, 1998). This fusion, which is not application specific, can be useful in: texture-mapping the point cloud to create photo-realistic 3D models which are essential for variety of applications (such as 3D city models, virtual tourist information as well as visualization purposes); extraction of reference targets for registration and calibration purposes (El-Hakim and Beraldin, 1994); automation of 3D measurement (automatic exterior orientation); 3D reconstruction; and if the data is georeferenced, it can be readily incorporated into existing GIS applications.

Fusing data taken from two different sensors requires that the multisensor data have to be correctly registered or relatively aligned and this paper therefore describes an approach to fuse high-resolution perspective 2D imagery and high-resolution 3D point cloud data. Our setup uses 3D point cloud data from 3D laser scanner and 2D intensity image from an independent CCD camera. These equipment provide independent datasets (geometry and intensity) and beg the question as to how can we accurately express these complementary datasets in a single object centred coordinate system. Also, matching features between an intensity image and the geometry automatically in such a multi-sensor environment is not trivial task (Pulli and Shapiro, 2000). It can be close to impossible due to the fact that the datasets are independent, dissimilar (Boughorbal *et al*, 2002), which differ in resolution, field of view, and scale.

This paper focuses on three distinct approaches to the multisensor fusion task. The first one is data fusion which integrates data from the two sensors (3D point cloud data and 2D intensity image). The advantage is that the existing traditional image processing algorithms can operate on this generated synthetic image. Also, to register this image to intensity image is much easier task that registering the 2D image into the 3D point clouds directly. The second one, on the other hand, is image fusion which involves feature detection and feature correspondence matching between the generated synthetic image and the intensity image acquired with digital camera. The third one which is the model-based image fusion is to relate each pixel in the 2D intensity image data to its corresponding sampled 3D point on the object surface. The task is to determine the relationship the coordinate systems of the image and the object. The result of this procedure is that the intensity image and the geometric model are positioned and oriented in the same coordinate systems.

In section 2 of this paper, the data multisensor data fusion methodology and integration models are discussed. Section 3 deals with the multisensor image matching procedure. Section 4 describes the mode-based image fusion. The results are discussed in section 5 and conclusions and discussions of future works are outlined in section 6.

2. MULTI-SENSOR DATA FUSION

Multi-sensor data fusion refers to, in our context, the techniques for the combination of datasets from the 3D point cloud data and the 2D image (i.e. intrinsic parameter of the CCD camera) to create a new dataset. The input to this process is 3D data from 3D laser scanner, the 2D intensity image from independent CCD camera, and the intrinsic camera parameters. These sensors are not calibrated. Detailed description of the coordinate systems of the 2D and the 3D sensors, data capture and processing is available in Forkuo and King (2004).

2.1.1 3D Point Cloud

Cyrax 2500 Laser Scanner was used to carry out the laser scanning to acquire a discrete representation of the object. More description of the laser scanner used in this experiment can be found in (Forkuo and King, 2004; CYRA, 2004). Figure 1 shows a screen capture of pseudo-colored 3D point cloud data of the test area. The 3D point cloud allows for the construction of a 3D surface Model of the scene. The resolution of the scan, which controls the number of points recorded in a scene and the level of detail visible in a scan (Barber and Bryan, 2001), is simply the smallest change in distance that the scanner is capable of detecting.

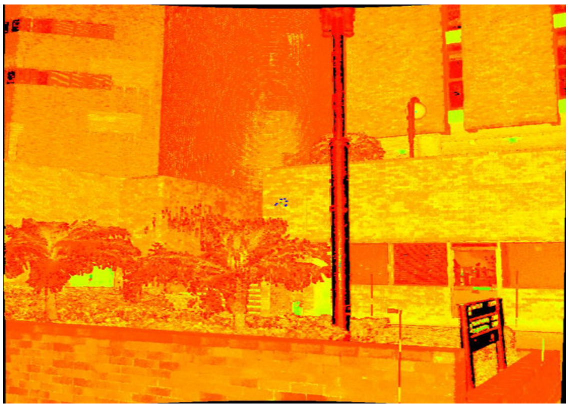

Figure 1 A screen capture of Pseudo-colored 3D point cloud

2.1.2 2D Intensity Images

A series of images were taken at different direction and position (as depicted in figure 2.2) by the digital CCD camera (Nikon D1x), which produces an image of at the size of 23.7mm x 15.6mm width. These images are called real camera images (RCI) and one of these images is represented in figure 2. This camera provides a digital image with the resolution of 3008 by 1960 pixels at true color mode.

2.2 Backprojection of Laser Scanning Data

Multi-sensor mathematical model is a physical model that describes the integration of 3D laser range camera and the CCD camera (Forkuo and King, 2004). We use the photogrammetric principles of collinearity condition with no systematic correction parameters as the basis for the implementation of the transformation of 3D point cloud data to suitable 2D shape information. For details on the collinearity model and the subsequent steps in fusing the dataset, see Forkuo and King (2004).

Figure 2 Real Camera Image (RCI)

2.3 The Synthetic Camera Image

This task is to represent the results of the collinearity equation (discrete x, y) points as image, which could be used in the image-to-image registration task discussed in section 3. A more detailed description can be found in Forkuo and King (2004). . By way of definition, interpolating the backprojected laser point (which contains irregular point spacing) into a regular grid at an even spacing using the intensity values generates what is termed "the Synthetic Camera Image" (SCI). There are two options related to this interpolation. First option is to generate the SCI by keeping the original resolution of the point cloud data and the compute a new pixel size. The second option, on the other hand, is to keep the pixel size of the real camera image and then compute the number of pixels or the resolution. In this paper, this option is used to generate the SCI. The interpolated data was then modeled by f (x, y) = I, where (x, y) is the pixel position and the I , the corresponding intensity value which is mapped to a grayscale. Conventional image processing techniques such contrast stretching and image enhancement were then used to produce the final image in figure 3. It is obvious that the geometric features in the SCI are easier to detect than those in the laser range data. This image offers a major advantage to interactively (controlled by human operator) or automatically conjugate matching with the intensity images produced by digital camera.

Figure 3. The generated Synthetic Camera Image

3. MULTISENSOR IMAGE FUSION

In many image processing applications it is necessary to compare multiple images of the same scene acquired by different sensors or image taken by the same sensor but at different times or from different locations. This section describes the process of matching multi-source data of the same scene acquired from different viewpoints and by different. The purpose of multi-sensor image matching in this paper is to establish the correspondence between RCI and SCI, and to determine the geometric transformation that aligns one image with the other. Existing multi-sensor image matching techniques fall into two broad categories: manual image and automatic image matching the implementation and results of manual multi-sensor image matching, which includes, interior, relative and absolute orientations using two different types of software for comparison purposes, has been discussed in Forkuo and King (2003). The manual measurement was necessary to understand the key issues such as geometric quality, both spatial and geometric resolutions of the generated synthetic camera image.

3.1 Automatic Multisensor Image matching

Once the 2D intensity image has been generated from the 3D point cloud, the location of corresponding feature in the Synthetic Camera Image (SCI) and the Real Camera Image (RCI) is determined. The most difficult part of the automatic registration is essentially the correspondence matching: Given a point in one image, find the corresponding point in each of the other image(s). Although the automatic correspondence is not a problem for vertically oriented images, it is still a problem in the terrestrial case and it is even much complex in terrestrial multi-sensor case. It can be observed that, since both image types are formed using similar mechanisms, the location of many objects are identifiable in each image. However, there are differences in illumination, perspective, reflectance as well as lack of appropriate texture (Milian et al, 2002) between these images. Also, images from different sensors usually have their own inherent noise (Habib and Alruzouq, 2004). Furthermore, the automatic registration problem can be complicated, in our case, by differences in image resolution and scale, and low image quality (especially with the SCI). One approach to automatically overcome the correspondence problem is both area and feature based approach was used (Dias et al, 2002). The first step for correspondence matching or simply pairwise matching is the extraction of features, generally interest points from both images using Harris corner detector. Initial correspondence between these points is then established by correlating regions around the features. The similarity is then judged by the accumulated development of corresponding interest points in the two images (Rothfeder et al, 2003). We have discussed the matching algorithm which consists of feature extraction process followed by the cross correlation matching in Forkuo and Bruce (2004).

3.1.1 Automatic Feature Detection and Extraction

The automatic registration problem requires finding features (edges, corners) in one image and correlates them in another. For this paper, Harris corner detector as proposed in Harris and Stephens (1988) is used detect and extract corners in both images. This operator has been widely used and it has been shown to be robust to viewpoint changes (i.e. image rotations and translations) and illumination changes (Dufournaud *et al*, 2004; Rothfede*r et al*, 2003). However, the Harris point detector is not invariant to changes in scale (Dufournaud *et al*, 2004. It uses a threshold on the number of corner extracted based on the image size. The number of corners detected in images is variable (Rothfede*r et al*, 2003) and in figure 4, the two images are shown with the detected corners features. Once feature points are extracted from image pair, correspondence matching can be performed.

3.1.2 Correspondence matching

This section concentrates on determining corresponding between two sets of extracted interest points that were detected with Harris corner operator. To match these features automatically, the zero mean normalized cross correlation (ZNCC) measure, which is invariant to varying lighting conditions (Lhaullier and Quan, 2000) is used. This method uses a small window around each point to matched (this point becomes the center of a small window of gray level intensities), and this window (template) is compared with similarly sized regions (neighborhood) in the other image (Rothfede*r et al*, 2003). In other words, the ZNCC method is based on the analysis of the gray level pattern around the detected point of interest and on the search for the most similar pattern in the successive image (Giachetti, 2000). Each comparison yields a score, *a measure of similarity*. The match is assigned to the corner with highest matching score (Smith *et al*, 1998).

By selecting a suitable patch size (correlation window) and threshold for the matching process reduces the number of detection of false correspondence pairs. However, in our case, the number of mismatches (referred to as outliers) may be quite large (as can be observed in figure 5). This occurs in particular when some corners cannot be matched. Also, there are likely to be several candidates matches for some corners which are very similar (Smith *et al*, 1998). These correspondences are refined using a robust search procedure such as the RANdom SAmple Consensus (RANSAC) algorithm (Capel and Zisserman, 2003; Fischler and R. C. Bolles, 1981). This algorithm allows the user to define in advance the number of potential outliers through the selection of a threshold. The best solution is that which maximizes the number of points whose residuals are below a given threshold. Details can be found in Fischler and R. C. Bolles (1981). Once outliers are removed, the set of points identified as inlers may be combined to give the final solution (RANSAC inliers) and the result is shown in figure 6. These *inlying* correspondences are used in the model-based image fusion.

4. MODEL-BASED IMAGE FUSION

In this context, model-based fusion is the process of establishing a link between each pixel in the 2D intensity image data to its corresponding sampled 3D point on the object surface. The task is to determine the relationship the coordinate systems of the image and the object by photogrammetric process of exterior orientation. The exterior orientation process is achieved in two steps. For the first step, we relate each matched pixel of the extracted feature in the SCI data to its corresponding 3D point from the point cloud data using interpolation constants. That is, the automatic link between the object coordinate system and the image coordinate system has been established. This means that the image coordinate, object coordinates and the returned laser intensity for the centre of each pixel are generated.

This 3D point is used as ground control point for the automatic exterior orientation solution. In the second step, the matches between the RCI and SCL, and their corresponding object coordinates are used for the exterior orientation computation with simultaneous bundle adjustment approach. This computation, which is control point-point-free method, has important applications in terrestrial photogrammetric engineering (Styliadis et all, 2003). Also, solving the camera positions and orientations, the RCI can be reprojected into the point cloud surface to produce the photorealistic model.

5. RESULTS AND ANALYSIS

Figure 4 shows the results of the several hundreds "interest points" detected (denoted with asterisks) automatically using the Harris feature detector. As can be observe on both images, most of the points of interest found in two images have correspondences. The ZNCC has been implemented to match the corners in SCI, with the ones in RCI and the results of them are superimposed on the images (figure 5). The matches are shown by the line linking matched points to their position in the other images. The feature point selection found approximately 800 points of interest and with the ZNCC measure, using a matching patch size of (17 x 17) pixels, using integer pixel locations, and correlation threshold of 0.8, there were 300 correspondences. As can be observed in figure 5, a relatively large number of mismatches occurred. These correspondences were refined with RANSAC algorithm and out of the 300 correspondences, about 160 points were discovered as inliers. As can be seen in figure 6, there are large number good corresponding sets of points for the orientation procedure. It should be noted that the size of the matching window has a significant impact on the quality of the matches. Also, the quality of the digital images, particularly the SCI influences, the accuracy and the success of the matching process. However, the initial results demonstrate the ability of the ZNCC algorithm to match automatically measured points of interest.

Tables 1 and 2 present the values of the exterior orientation and the accuracy of the measurements for two real scenes. The initial results of the first scene which includes, feature detection and correspondence matching, are presented in Forkuo and King (2004). To verify the validity of the matching algorithm, the result of the second real scene is also presented in table 2. Both scenes were acquired with the same laser scanner and digital camera. It could be seen that the camera position for the SCI for both scenes has zero coordinates (i.e. $X_o = Y_o = Z_o = 0$), with angular rotation parameter also equal to zero $(\omega = \varphi = \kappa = 0)$. These exterior orientation parameters of laser scanner do confirm the assumption already discussed in Forkuo and King (2004). The same table also contains the accuracy of the bundle adjustment in terms of the root mean square error of the object point coordinates and of the image measurements. The accuracy in X (MX= 0.001m), in Y (MY= 0.001m) and in Z (MZ=0) for both scenes and the overall accuracy in the object space coordinate determination for both scenes (MXYZ) is within 0.001m. Also, the accuracy for the image measurement for both x (Mx = 0.017) and y (My = 0.010) for real scene one and for x(Mx =0.014) and y(My=0.018) for real scene two is within two pixels. However, the accuracy of the measurement can vary significantly by looking into important factors such as the resolutions and the quality of the images, employing sub-pixel processing techniques, camera calibration and possibly number of images.

	Exterior Orientation Parameters					
	X_o	Y_o	Z_o	ω	φ	κ
SCI	0	0	0	0	0	0
RCI	0.249	0.345	1.610	-4.108	2.943	-0.607
RMS Residual						
Object Space Coordinates				**image coordinates**		
MX	MY	MZ	MXYZ		Mx	My
0.001	0.001	0	0.001		0.017	0.010

Table 1: Exterior Orientation Parameters of real scene 1 (Positional unit: meter; Angular unit: degree) and Measurement Accuracy

	Exterior Orientation Parameters					
	X_o	Y_o	Z_o	ω	φ	κ
SCI	0	0	0	0	0	0
RCI	**0.021**	**0.238**	**1.151**	**-1.633**	**-0.212**	**-1 153**
RMS Residuals						
Object Space Coordinates (m)				**image coordinates(mm)**		
MX	MY	MZ	MXYZ		Mx	My
0.001	0.001	0	0.001		0.014	0.018

Table 2: Exterior Orientation Parameters of real scene 2 (Positional unit: meter; Angular unit: degree) and Measurement Accuracy

6. CONCLUSIONS

The fusion of the 2D images and 3D point cloud has been assessed, and a synthetic image has been generated by integrating information from the two sensors. Features have detected, extracted and matched to develop geometric relationship between the digital camera and laser scanner. The initial results show that we have successfully obtained corresponding points in both images. Bundle adjustment is used to reconstruct the 3D object space coordinates and to recover camera positions. The accuracy of the object coordinate determination is with 0.001m and for the image coordinate measurement; the measurement error is within two pixels. However, future research will concentrate on investigating the effect of resampling the RCI to a smaller size and the use of combined edge and corners approach instead of only corners. Also, the impact of camera calibration, particularly, lens distortion, on the matching results will be investigated. The RANSAC algorithm has been implemented to filter false correspondences. However some further developments of the algorithm are still required.

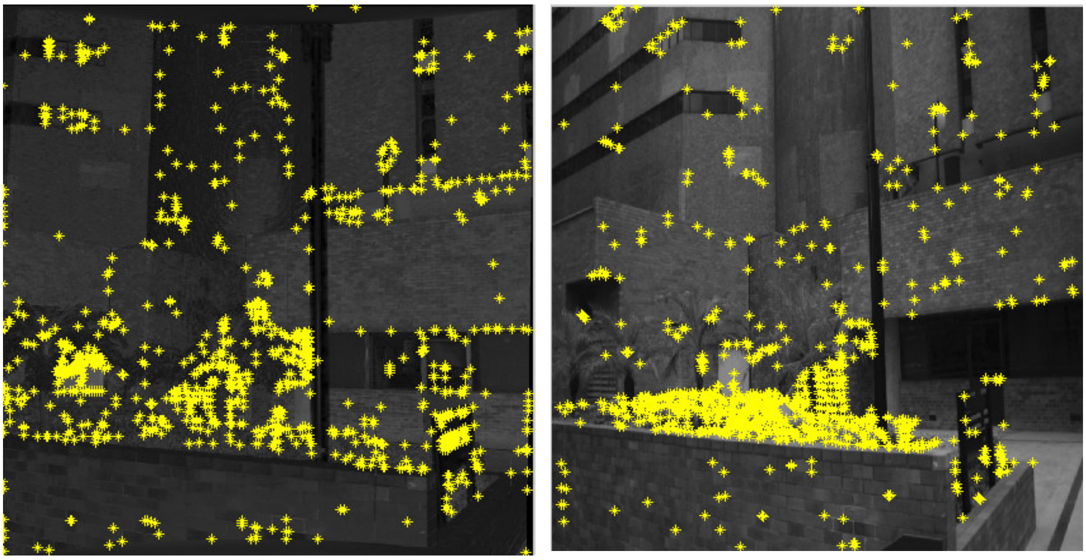

Figure 4: Detected Corner Features superimposed on the images

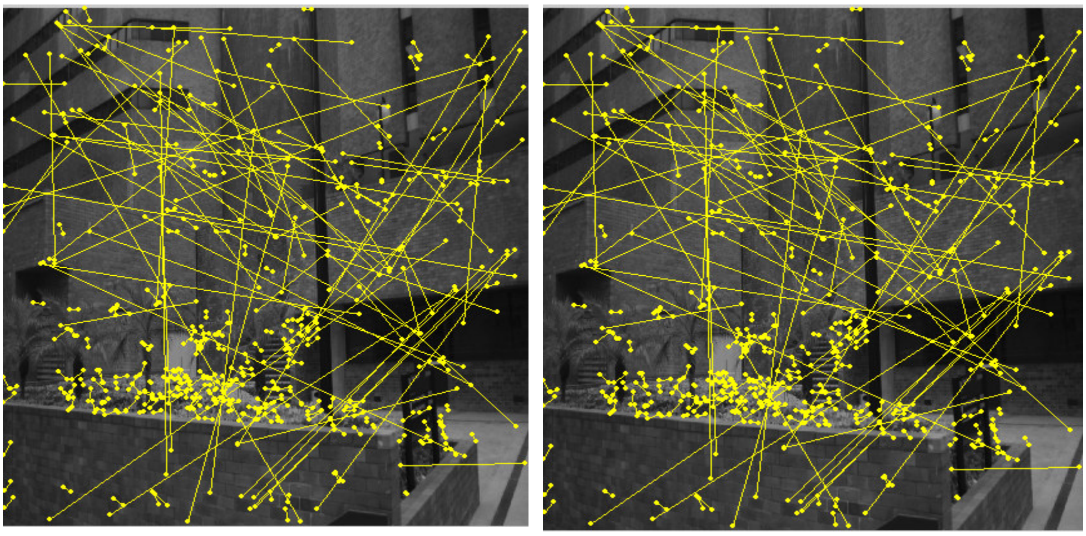

Figure 5: The Detected Correspondences of both pairs

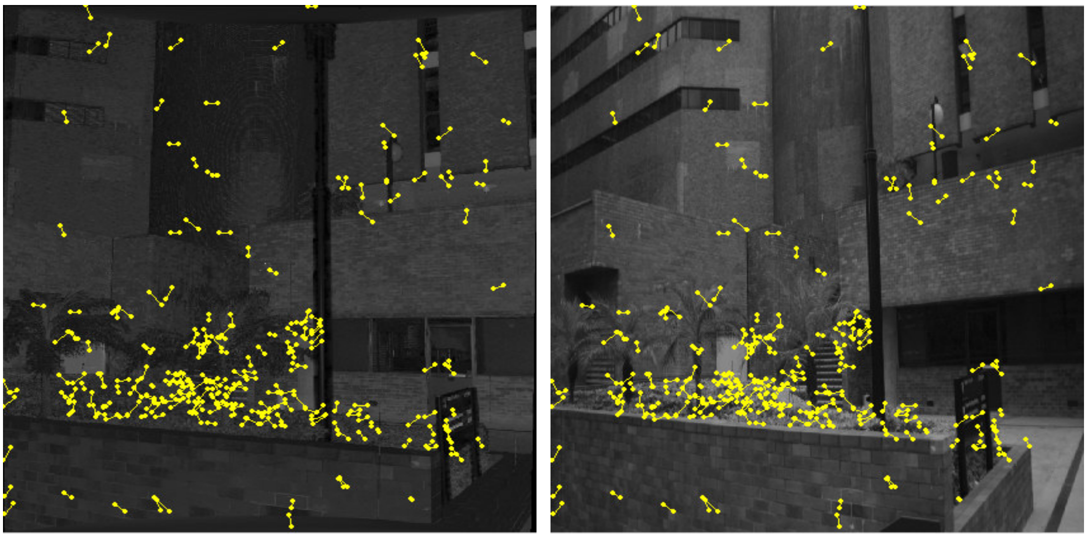

Figure 6: The Final Verified Detected Correspondences of both pairs (RANSAC inliers)

REFERENCES

Barber, D., Mills, J., Bryan., (2001). “Laser Scanning and Photogrammetry-21st Century Metrology”. CIPA 2001 International Symposium, University of Postdam, Germany, 18-21 September 2001.

Boughorbal, F., Page, D., and Abidi, M., (2000). “Automatic Reconstruction of Large 3D Models of Real Environments from Unregistered Data-sets”. *SPIE Conf. on 3D Image Capture and Applications III*, Vol. 3958, p. 234-243.

Capel, D., and Zisserman, A., (2003). “Computer Vision Applied to Suoer-Resolution”. IEEE Signal Processing Magazine. pp 75-86.
http://www.robots.ox.ac.uk/~vgg/publications/papers/capel03.pdf

CYRA., (2004). CYRAX 2500 Product Specification, Cyra Technologies Inc. http://www.cyra.com

Dias, P., Sequeira, V., Gonçalves, J. G. M., Vaz, F., (2002). “Automatic Registration of Reflectance and Color Intensity Image for 3D Reconstruction*”. Robotics and Autonomous*, Vol 39, pp 157-168.

Dufournaud, Y., Schmid, C.,and Horaud, R., ((2004).”Image Matching Scale adjustment”.*Computer Vision and Image Understanding*, 93 (2), February 2004. pp 175-194.

El-Hakim, S. F., and Beraldin, J.-A., (1994).”On the Integration of Range and Intensity Data to Improve Vision-Based Three-Dimensional Measurement”.In *Videometric III*, proceeding of SPIE, Vol.2350, pp 306-327.

Elstrom, M. D., Smith, P. W., and. Abidi,. M. A., (1998). "Stereo-Based Registration of LADAR and Color Imagery", *SPIE Conference on Intelligent Robots and Computer Vision XVII: Algorithms, Techniques, and Active Vision*, 3522, pp. 343-354.

Fischler, M. A., and Bolles, R. C.,”Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography”. *Comm. ACM*, 24(6):381–395, 1981.

Forkuo, E. K., King. B. A., (2003). “Fusing of Photogrammetric Imagery and Laser Scanner Point Clouds”. Accepted for publication on the “Proceedings, ASPRS 2004 Annual Conference”, Denver, Colorado, 23-28 May 2004.

Forkuo, E. K., King. B., (2003). “Registration of Photogrammetric Imagery and Laser Scanner Point Clouds”. Proceedings, 7th South East Asian Surveying Congress, Hong Kong, 3-7 November 2003.

Giachetti, A., (2000). “Matching Techniques to Compute Image Motion”. *Image and Vision Computing*, 18, pp 247-260.

Habib, A. F., Alruzouq, R.I., (2004). “Line-based Modified Iterative Hough Transform for Automatic Registration of Multi-source Imagery”. *The Photogrammetric Record*, 19(105), pp 5-21.

Harris, C., and Stephens, M., (1988). “A combined corner and edge detector”. Proceedings of the 4th ALVEY vision conference, University of Manchester, England, September 1988, pp 147—151.

Lhuillier, M., and Quan, L., (2000). “Robust Dense Matching Using Local and Global Geometric Constraints”. ICPR 2000, 1968-1972.

Malian, A., Heuvel van den, F.A., and Azizi, A., (2002). “A Robust Photogrammetric System for Wound Measurement*”. International Archives of Photogrammetry and Remote Sensing*, Vol. 34, Part 5, pp 264 – 269.

Pulli, K., and Shapiro, L., (2000). “Surface Reconstruction and Display from Range and Color Data”. *Graphical Models and Image Processing,* 62, pp. 165-201.

Rothfeder, J. L., Feng, S., and Rath, T. M., (2003). “Using Corner Feature Correspondences to Rank Word Images by Similarity”. : Proceeding. of the Workshop on Document Image Analysis and Retrieval (DIAR), June 21, 2003.

Smith, P., Sinclair D., Cipolla R., and Wood K., (1998).”Effective Corner Matching”.On-Line Proceedings of the Ninth British Machine Vision Conference. http://www.bmva.ac.uk/bmvc/1998/papers/d125/h125.htm.

Styladis, A. D., Karagiannidis C. J., Zestas, N.C., Chatzara, K. N., (2003). “Pose Determination from a Single Image in a Controlled CAD Environment”, *Journal of WSCG*, Vol. 11 (1).