A novel 3D torso image reconstruction procedure using a pair of digital stereo back images

A. Kumar & N. Durdle
Department of Electrical & Computer Engineering,
University of Alberta, Canada

Abstract

This paper presents a novel procedure for creating a 3D torso image from a pair of stereo digital 2D back images. The aim of this procedure is to obtain 3D images that can be used for assessment of external spinal deformities in scoliosis. Scoliosis is a condition characterized by lateral deviation of the spine coupled with rotation of individual vertebra resulting in visible torso asymmetries. The procedure provides clinicians with a cost effective and mobile setup of acquiring 3D images. To improve the registration process, a novel approach combining tree weighted colour based image segmentation and differential geometry was developed. Image reconstruction involved pre-processing, triangulation and texture application to obtain a 3D image. Analysis was performed using human subjects and objects of known dimension. Evaluation of system performance was done against existing stereovision procedures and range scanning systems. The final 3D image was compared to that obtained from the Konica Minolta Vivid 700 laser scanner. Each image was divided into 360 cross sections for evaluation against size and shape. The 3D image reconstructed from this novel procedure was 75–100% accurate when compared against the 3D image from the laser scanner. The results demonstrate that the procedure is a cost effective clinical tool for assessing torso shape and symmetry.

Keywords: scoliosis, registration, image reconstruction, belief propagation, differential geometry, disparity map.

WIT Transactions on Biomedicine and Health, Vol 13,
www.witpress.com, ISSN 1743-3525 (on-line)
doi:10.2495/BIO090241

1 Introduction

Scoliosis is a condition characterized by lateral deviation of the spine coupled with rotation of individual vertebra resulting in visible torso asymmetries [1]. The assessment of severity of scoliosis is traditionally done using radiographs of the spine. However, radiographs do not describe the visible torso deformity associated with scoliosis [2]. Many three dimensional data acquisition techniques have been investigated for developing a system to assist clinicians in the evaluation of external scoliosis deformities. This is because most scoliosis patients and their families are more concerned with the shape of the torso than the internal alignment of the spine. Traditional procedures for assessment of torso shape are based on landmarks. Since the back surface is smooth and featureless, it becomes very difficult to locate these landmarks in real time. Other techniques such as difference mapping, Moiré topography, ISIS scanning, Quantec system scanning and laser scanning have been developed over the years for assessment of torso shape [2]. Disadvantages in these methods range from poor resolution images to expensive processes.

In light of these problems and due to advancements in stereo computer vision, there is a need to develop a cost effective, accurate technique that can be clinically used for assessing scoliosis.

Progress in computer vision and availability of faster computer processors has lead to development of stereo vision algorithms for simultaneous stereo camera capture, calibration and reconstruction. Stereo capture systems consist of digital or TV cameras positioned with known geometry. Significant advances have been recently made in the area of computer vision, as a result of publically available performance testing such as the Middlebury data set [3], which has allowed researchers to compare their algorithms against all state-of-the-art algorithms [4]. Stereo correspondence or registration of stereo images is one of the most active research areas in computer vision [3]. In the area of stereo vision research, stereo images refer to images captured at different viewpoints using cameras with known geometry [5]. In order to register the stereo images, we need to determine the closest (least error) or best point-to-point correspondence between the two images.

Stereo registration algorithms can be used on stereo images captured using calibrated digital or TV cameras to obtain three dimensional (3D) point set data. A triangulation algorithm is primarily applied on point set data to obtain a 3D surface. 3D surface reconstructions of scenes or localized objects in a scene using stereo vision has modern applications in 3D modelling, computer graphics, facial expression recognition, surgical planning, architectural structural design etc. [3,6]. The 3D scene geometry established from this process is used to reconstruct 3D torso images using stereo reconstruction to study scoliosis.

However a problem associated with the existing stereo registration algorithms such as that developed by Klauss et al. [7] is that it assumes frontal parallel plane geometry. This means that it assumes depth is constant (with respect to the rectified stereo pair) over a region under consideration [8]. Reconstruction of smooth and curved surfaces where depth is constantly changing violates this

www.witpress.com, ISSN 1743-3525 (on-line)

assumption. Li and Zucker [6] have tried to solve this problem using differential geometry but they have not applied their algorithm to the one developed by Klauss et al. [7]. Belief Propagation and Graph Cuts are the most commonly used methods for refinement of the registration process [3]. Tree Re-weighted Message Passing, which might become a serious rival to Belief Propagation and Graph Cuts [9], has not been used with differential geometry. Tree Reweighted Message Passing's improvement over Belief Propagation and Graph Cuts becomes significant for more difficult functions [9].

Stereo reconstruction algorithm is applied to registered images. Since, the registration process leaves stray points due to errors, this need to be removed. Triangulation to connect the 3D points obtained through registration into polygons and texture application is the last step to obtain a fully reconstructed 3D object.

2 Objective

The objective of this paper is to present a procedure to reconstruct a 3D image from 2D stereo images of the torso using a stereo camera setup. The 3D image can facilitate the assessment of scoliosis clinically. This requires the process to require minimal user input in order to prevent errors; to be error correcting in order to prevent stray data points; to be cheaper and more accurate than existing methods. The aim of this paper is two-fold. Firstly, to create a novel procedure by investigating and improving on existing methods in computer vision for stereo image registration (correspondence matching in a pair of stereo digital images). Finally, the registered image is preprocessed, leading to a reconstructed 3D image.

3 Materials

A stereo digital camera setup is required to obtain 2D images of the torso. The cameras used for the setup are two 3 Megapixel Nikon digital cameras. The acquired images are processed using an Intel® Core 2 Duo 2.4 GHz, 4GB RAM PC to obtain the reconstructed 3D torso image. The stereo digital camera setup is shown in fig. 1. The vertical camera setup is used opposed to the horizontal camera setup primarily because of the shape and size of the torso. Since, the torso is longer (in length) than it is wider, the vertical camera setup allows for images to be captured in portrait orientation. This increases the total number of pixels in the digital images that represent the torso in the vertical setup as opposed to the horizontal setup.

4 Methods

The 3D torso image reconstruction procedure can be divided into 3 stages as shown in fig. 2

www.witpress.com, ISSN 1743-3525 (on-line)

Figure 1: Vertical stereo digital camera setup.

4.1 2D Stereo back image acquisition

Stereo image acquisition is performed using the vertical stereo camera setup. Image calibration and rectification are not required in this process. This saves the total time required for 3D image reconstruction procedure. Image calibration and rectification defines the relationship between pixels on a particular image to 3D coordinate in world space. In this reconstruction procedure, the relative position of each of the 3D points in the final reconstructed torso image is what defines the size and shape of the torso. Therefore, absolute positions of the 3D point in world space are not relevant. Fig. 3 shows images taken using the vertical camera setup.

Figure 2: Stages of torso image reconstruction.

4.2 Stereo back image registration

Stereo registration methods are assessed using univalued disparity function of one image with respect to the other (referred to as reference image). When first introduced in human vision literature, disparity was used to describe the difference in location of corresponding features seen by the left and right eye [3].

The procedure for obtaining disparity is divided into three stages: mean shift colour segmentation, adaptive local pixel matching, and differential geometry in a tree reweighted belief propagation procedure.

The x, y spatial coordinates of the disparity space are taken to be coincident with the pixel coordinates of a reference image. The correspondence between a pixel in a reference image and pixel in the matching image is linearly related to each other by the disparity function. The disparity function obtained over the 2D images is known as the disparity space image (DSI). DSI gives a 3D point set data used in the final stage for 3D Torso Point generation. It represents the confidence or log likelihood of a match implied by the disparity function. The values of DSI are converted into z spatial coordinate values using the focal length of the cameras. We use 128 levels of disparity (which implies the same number of distinct z values). We find using these many disparity levels provides accuracy and speed in the reconstruction procedure.

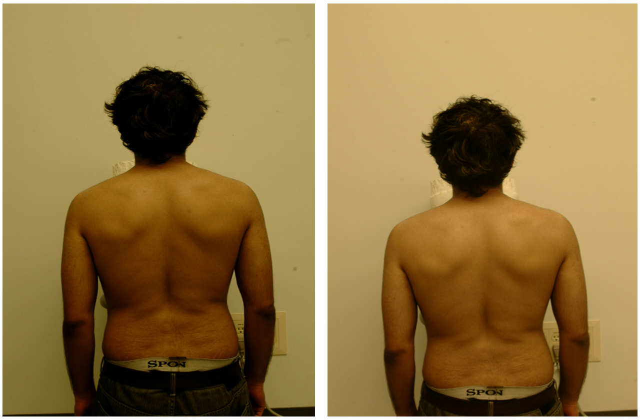

Figure 3: Images acquired from the top and bottom cameras respectively of a vertical digital camera setup.

The process of colour segmentation is to decompose the reference image into regions of colour or greyscale [7]. The mean-shift analysis approach is essentially defined as a gradient ascent search for maxima in a density function defined over a high dimensional feature space [7]. Comaniciu and Meer's [9] mean shift segmentation is insensitive to differences in camera gain. Fig. 4 shows the mean shift colour segmented image obtained using Comaniciu and Meer's method.

The next step involving local pixel matching is an essential step for defining a disparity plane. The aim of this step is to provide an initial estimate of the disparity space image (DSI). The disparity plane is based on 3D x-y-d space supporting slanted and curved surfaces where x, y are spatial coordinates and d is the inverse depth or disparity [10]. The disparity planes are calculated using local pixel matching. Local pixel matching requires calculation of a matching score and an aggregation window [7]. Matching score is obtained using a self-adapting dissimilarity measure that combines the sum of absolute pixel intensity differences (SAD) and a gradient-based measure as implemented by Klauss et al. [7].

www.witpress.com, ISSN 1743-3525 (on-line)

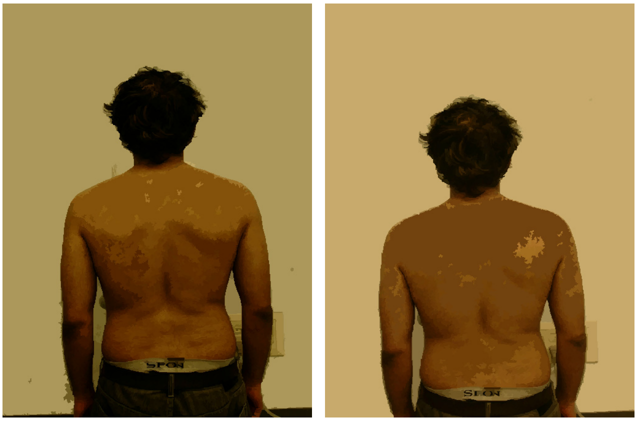

Figure 4: Mean shift segmented images of top and bottom cameras respectively.

Finally, the calculation of the final disparities is performed. The algorithms that perform well in this stage are based on an energy minimization framework. This means that we need to choose at each pixel the disparity associated with the minimum cost value [3].

$$E(d) = Edata(d) + \lambda Esmooth(d). \quad (1)$$

It involves minimizing two separate energy functions that are summed together to calculate the final energy minimization term as given by the eqn (1) where d represents disparity. The symbol Edata(d) in eqn (1) measures how well the disparity function agrees with the input image pair and is given by the summation of matching score over the spatial coordinates. The formulation of Edata(d) follows in eqn (2) where C is the matching score.

$$Edata(d) = \Sigma\, C(x, y, d(x, y)). \quad (2)$$

The symbol Esmooth(d) in eqn (2) encodes smoothness in the image by measuring the differences between the neighbouring pixels' disparities [3]. Esmooth(d) can be described by eqn (3) where ρ is some monotonically increasing function of disparity difference.

$$Esmooth(d) = \Sigma\, \rho(d(x, y) - d(x+1, y)) + \rho(d(x, y) - d(x, y+1)). \quad (3)$$

The Esmooth(d) operates on the frontal parallel assumption that is altered in this procedure as suggested by Li and Zucker using "floating" disparities [11]. Li and Zucker apply "Floating disparities" on the Max-Product Belief Propagation framework. Tree-Reweighted Message Passing as defined by Kolmogorov [12] is used in this procedure as the energy minimization framework. The key subroutine of the Tree-Reweighted Message Passing algorithm is Max-Product Belief Propagation [12]. The "floating" disparities [11] are therefore added to the Max-Product Belief Propagation component of Tree-Reweighted Belief Propagation. The Tree-Reweighted Belief Propagation is advantageous because messages are passed in a sequential order rather than a parallel order requiring half the space. Convergence is reached in two passes rather than having a convergence condition as in the case of Max-Product Belief Propagation. Therefore applying differential geometry to Tree-Reweighted Belief Propagation produces better results. The result of the image registration is the DSI shown in fig. 5 where the disparity levels on the top image are shown using the bottom image as a reference image. We can see that using the above technique to acquire DSI leads to a smooth variation in disparity across the back image.

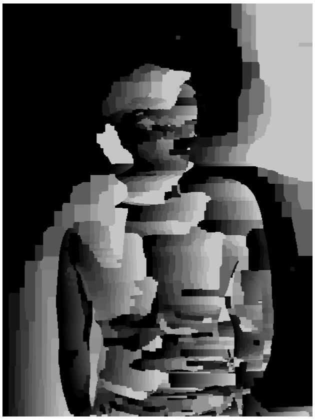

Figure 5: DSI of top camera image using bottom camera image as reference.

4.3 3D torso point generation and triangulation

This is the last step of the image reconstruction process and involves using eqn (4) to obtain the z spatial coordinate using the above-calculated DSI at a known x, y spatial coordinate.

$$z = \frac{b \times f}{d(x,y)}. \tag{4}$$

In eqn (4), b (baseline) represents the distance from the optical centre of the top camera to that of the bottom camera, f is the focal length of the cameras and d(x,y) is the disparity at that x, y location on the image. The x, y, z values are plotted using Visualization Toolkit (VTK) software [14]. The stray points are removed since they either comprise of errors in calculation of the DSI, errors in calculation of z spatial coordinate or belong to regions of the image that do not represent the torso. There are also holes and occlusions in the image due to missing z data points caused to errors noted above. Pre-processing the 3D point set needs is done to remove stray data points and fill in occlusions. Triangulation is also done to join the 3D points into polygons that represent the 3D surface of the torso.

www.witpress.com, ISSN 1743-3525 (on-line)

Firstly, using an edge detection algorithm in VTK on the DSI we determine the region that represents the torso. All the 3D points outside this region are discarded. This helps us to eliminate most of the stray points. Now all the 3D points are joined together by connecting lines between points of nearest Euclidean distance. This gives us a 3D image of the torso, which has a few stray points and occlusions. The image is of the same format and characteristics as that obtained from range scanning systems. The task of further processing and triangulating the 3D image is implemented in VTK using the technique described by Kumar et al. [14]. The resultant 3D torso image is shown in fig. 6.

5 Results

The output of image registration stage was compared to 2 existing registration procedures, segment-based adaptive belief propagation (adaptive BP) and colour-weighted hierarchical belief propagation (hierarchical BP). It outperformed existing methods, particularly for high curvature regions and significantly large cross sections. Its accuracy of reconstruction ranged from 85–100% compared to 75-100% for existing methods.

Figure 6: Reconstructed 3D torso image.

The final 3D image was compared to that obtained from the Konica Minolta Vivid 700 laser scanner. The image of a human subject was divided into 360 lateral cross sections for evaluation against size and shape. The 3D image reconstructed from this novel procedure was 75–100% accurate when compared against the 3D image from the Konica Minolta® Vivid 700 laser scanner. The results demonstrate that the procedure is a cost effective clinical tool for assessing torso shape and symmetry.

6 Discussion

The 3D torso reconstruction procedure provides a cost effective alternative for assessing torso shape and symmetry. Future work will focus on testing with more human subjects and with optical imaging methods other than the laser scanner. It will also consist of more comprehensive testing of the vertical camera setup under varying light conditions.

References

[1] Ajemba, P.O, Durdle, N.G, Hill, D.L & Raso, V.J., Re-positioning Effects of a Full Torso Imaging System for the Assessment of Scoliosis. *Canadian Conf on Electrical and Computer Engg*. 2004, Vol: 3, pp. 1483- 1486.

[2] Ajemba, P.O, Durdle, N.G, Hill, D.L & Raso, V.J., A Torso Imaging System for Quantifying the Deformity Associated with Scoliosis. *Instrumentation and Measurement, IEEE Transactions on,* Vol. 56, Issue 5, Oct. 2007, pp. 520–1526.

[3] Scharstein, D. & Szeliski, R., A Taxonomy and Evaluation of Dense Two-Frame Stereo Algorithms. *International Journal of Computer Vision*, 47(1/2/3), pp. 7-42, April-June 2002.

[4] Yang, Q., Wang, L., Yang, R., Stewenius, H. & Nister, D., Stereo Matching with Colour-Weighted Correlation, Hierarchical Belief Propagation and Occlusion Handling. *Computer Vision and Pattern Recognition*, 2006 IEEE Computer Society Conference on, Vol. 2, pp. 2347- 2354

[5] Sun, J., Zheng, N. & Shum, H., Stereo Matching Using Belief Propagation. *Pattern Analysis and Machine Intelligence*, IEEE Transactions on, Vol 25, Issue 7, July 2003, pp. 787–800

[6] Li, G. & Zucker, S.W., Stereo for Slanted Surfaces: First Order Disparities and Normal Consistency. *Proc. of EMMCVPR*, LNCS, 2005 - Springer

[7] Klaus, A., Sormann, M. & Karner, K., Segment-Based Stereo Matching Using Belief Propagation and a Self-Adapting Dissimilarity Measure. *Pattern Recognition*, ICPR 2006, Vol. 3, pp. 15-18

[8] Li, G. and Zucker, S.W., Differential Geometric Consistency Extends Stereo to Curved Surfaces. *Proc. of 9-th European Conference on Computer Vision,* ECCV (3) 2006, pp. 44-57

[9] Comaniciu, D. & Meer, P., Mean Shift: A Robust Approach toward Feature Space Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* May 2002 (Vol. 24, No. 5), pp. 603-619.

[10] Bleyer, M. & Gelautz, M., Graph-based surface reconstruction from stereo pairs using image segmentation. *Proc. of the SPIE*, Volume 5665, pp. 288-299, 2004

[11] Li, G. & Zucker, S.W., Surface Geometric Constraints for Stereo in Belief Propagation. *Computer Vision and Pattern Recognition*, 2006 IEEE Computer Society Conference on, Vol 2, 2006, pp. 2355–2362

www.witpress.com, ISSN 1743-3525 (on-line)

[12] Kolmogorov, V., Convergent Tree-Reweighted Message Passing for Energy Minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 28, No. 10, October 2006.

[13] Visualization Toolkit, www.vtk.org

[14] Kumar A., Ajemba P., Durdle N. & Raso J., Pre-processing Range Data for the Analysis of Torso Shape and Symmetry of Scoliosis Patients. Studies in health technology and informatics, 123(), pp. 483-487, 2006.

www.witpress.com, ISSN 1743-3525 (on-line)