TECHNIQUES FOR IMPROVING STEREO DEPTH MAPS OF FACES

Jason Baker and Vinod Chandran

School of Electrical & Electronic Systems Engineering,
Queensland University of Technology,
GPO Box 2434, Brisbane, QLD 4001, Australia
{jl.baker@qut.edu.au}

ABSTRACT

This paper presents an improved technique for determining the 3D structure of the human face form stereo images. The approach targets specific regions of the face individually. An application of the approach to estimate the 3D structure of the nose is presented. Stereo matching is extended and applied to colour in addition to grayscale, which improves matching in a low texture image like parts of the skin of a human face. Left-Right consistency check is used to determine erroneous disparity estimates. Prior knowledge of the nose structure is used in conjunction to improve the quality of the 3D nose estimates.

1. INTRODUCTION

Face recognition is a method of biometric identification that works by taking a facial image of a person, extracting useful features and comparing these features. A database of face templates for comparison is created during the training process and used in recognition.

The difficulty with image based face recognition is that images used in the recognition process generally vary compared to the images acquired for the recognition database. Thus the face recognition system must be able to cope with numerous possible variations including viewpoint, illumination, expression, age, facial hair, occlusion, spectacles and background changes.

An alternate approach to the face recognition problem is to generate a 3D model of the face. The 3D data of the face is used either singlarly to directly compare to a database of 3D face models, or in conjuction with the 2D images.

Singularly a 3D model can be used to match to a database of 3D face models. This would require the use of 3D object recognition techniques [1]. The success of this approach relies on the level of discrimination between 3D models of different people and the similarity of 3D models of the same person. This would require a very precise and detailed model of a persons face.

In conjuction with a standard 2D image, the 3D model can provide the required facial structure information to correct face alignment problems associated with 2D face recognition tasks. In particular, out of plane head rotations can be partially corrected given the 2D image and the 3D face model.

The generation 3D face models from stereo imagery requires two cameras capturing an image of a scene at the same time from slightly separated viewpoints. The images are first rectified such that the image rows are aligned between the two images. Each pixel is then match from one image to their corresponding equivalent pixel in the alternate stereo image. Generally matching single pixels proves inadequate and a window of surrounding pixels is used. Matching entails finding the equivalent window of pixels across the stereo image pair. A measure of the relationship or degree of correspondence between image windows is used to select the best suited region from a search space. There are several good general stereo matching techniques covered by [2] and [3]. These include techniques based on a Sum of Absolute Differences (SAD), Sum of Squared Differences (SSD) and Normalised Cross Correlation (NCC).

A simplified approach is to select features such as edges or corners and match these features across the images. The result is a sparse group of depth estimates which is inadequate for the 3D face recognition. We choose to perform window based pixel by pixel matching in this work and enhance its capabilities. Given the problem of 3D face modelling from stereo images, the input is no longer arbitrary and it is advantageous to use prior knowledge of the structure of a human face. While faces vary between people there is a general structure for a face, this general structure can be utilized to target the stereo matching techniques and criteria adaptively over specific regions. Different techniques and criteria may suit individual face regions such as the chin and lip, cheek, eyes, forehead or nose regions.

The effectiveness of stereo matching algorithms depends on the variation of pixel values within the template and candidate matching windows. The greater the variation of pixel values within matching windows the better the probability of finding a unique match [4]. This problem can be reduced by using a larger matching window which introduces a greater number of pixels and possibly a unique window. This introduces further problems when considering surfaces containing discontinuities [4]. Unfortunately the texture of the human facial skin is relatively bland in certain regions and increasing the size of the matching windows will reduce the precision with which the face can be modeled in the non-bland regions. Dividing the face into specific regions allows this problem to be considered and individual solutions tailored for each region. The resulting 3D model is expect to provide a more accurate depth model across the entire selected region with fewer significant errors.

This paper is organised into the following sections: Section 2 explains facial region decomposition, the stereo matching algorithm, an extended colour stereo matching algorithm, coarse to fine matching constraint and Left to right consistency check. Section 3 presents the results of the technique applied to the nose region. Section 4concludes with a discussion of the performance.

2. RECONSTRUCTION APPROACH

2.1. Facial Feature Extraction

The human face can be considered to be composed of several regions. Each region has a specific feature or unique structural property, for example the cheek region is relatively smooth with a gentle curvature and a consistent texture, while the eye regions have changes in depth and a variety of textures. It is possible to consider each region of the face separately and apply individual matching techniques to various regions of the face depending upon the regions structure and texture. Furthermore prior knowledge of each region can be applied to generate improved 3D models.

The selected face region for this paper is the nose region. The nose is of interest due to the relatively smooth contour which varies at several points across the face. Small discontinuities, located at the base of the nose, allow application specific constraints. The nose also consists of skin with little texture variation.

The nose region needs to be extracted from the two stereo images of the face presented in figure 1. The procedure to extract the selected regions consists of the following steps: Simple skin segmentation, Face contour extraction, Face approximation by an ellipse, Eye localization using ellipse location and dimensions, and Nose region localization based on eye region properties. The resulting nose regions are presented in figure 2.

2.2. Stereo Matching Algorithm

The stereo matching problem requires that for each pixel in the left image a corresponding pixel in the right image be found. We use rectified images [5] which reduce the search space down to a single image line. The matching is generally performed by taking a window around the candidate pixel and comparing it to the candidate matching pixel window.

Fig. 1. The full face image prior to nose extract. The image has a resolution of 1280 x 960 pixels

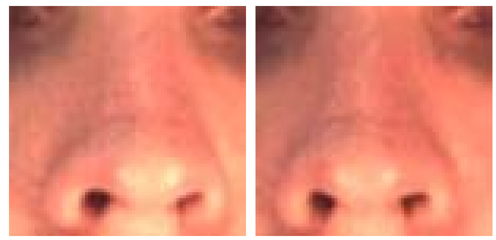

Fig. 2. The selected and extracted (a) left nose image, (b) right nose image.

The matching functions are based upon the afforementioned SAD, SSD or NCC algorithms [2]. We found that both the SAD and SSD algorithms performed extremely poorly in the nose region owing to a lack of intensity variation. The Zero-mean Normalised Cross Correlation (ZNCC) algorithm [6] was also tested using intensity images with a window size of 9 pixels. The resulting surface was very noisy. This is again due to little variation in skin texture across the nose region and the small window size.

2.3. Colour Stereo Matching

The intensity variation across the nose image is low and did not produced adequate quality disparity results. The stereo matching was improved by extending the matching function similarly to the work of Muhlmann *et* al. [7]. Muhlmann *et* al. use the Sum of Absolute Difference (SAD) to calculate a matching measure for each of RGB planes separately. The individual SAD colour planes scores are summed to produce final matching measure. We extended the ZNCC algorithm to utilise the colour information by using the summation of the three ZNCC results for each of the RGB planes presented in equation 1.

$$
\begin{aligned}
I(u,v,y) = {} & \frac{\sum_{(u,v)}(I_{r1}(u,v)) - \bar{I_{r1}}) \cdot (I_{r2}(u,y+v)) - \bar{I_{r2}})}{\sqrt{\sum_{(u,v)}(I_{r1}(u,v)) - \bar{I_{r1}}) \cdot \sum_{(u,v)}(I_{r2}(u,v)) - \bar{I_{r2}})}} \\
& + \frac{\sum_{(u,v)}(I_{g1}(u,v)) - \bar{I_{g1}}) \cdot (I_{g2}(u,y+v)) - \bar{I_{g2}})}{\sqrt{\sum_{(u,v)}(I_{g1}(u,v)) - \bar{I_{g1}}) \cdot \sum_{(u,v)}(I_{g2}(u,v)) - \bar{I_{g2}})}} \\
& + \frac{\sum_{(u,v)}(I_{b1}(u,v)) - \bar{I_{b1}}) \cdot (I_{b2}(u,y+v)) - \bar{I_{b2}})}{\sqrt{\sum_{(u,v)}(I_{b1}(u,v)) - \bar{I_{b1}}) \cdot \sum_{(u,v)}(I_{b2}(u,v)) - \bar{I_{b2}})}} \qquad (1)
\end{aligned}
$$

Utilising colour information in the stereo matching process improves the 3D model by reducing the number of erroneous matches.

2.4. Coarse to Fine Matching

A coarse to fine matching strategy is implemented to constrain the matching region. This is achieved by selecting a large matching window, in this case a window size of 19 pixels. The largest disparity is determined and the search space for all subsequent matching windows in constrained to the largest disparity value (a small variation is added to allow for variances at the largest disparity values).

The coarse to fine matching strategy is updated for each reduced matching window size (from 19 pixels to 7 pixels in 2 pixel decrements). This constraint has the advantage of reducing the search space and thus reducing the processing time.

Given that correspondence is determined for 7 different window sizes each producing two disparity maps with varying accuracy and noise properties, these can be combined to produce a single estimate. The larger window correspondence results produced a stepped result with minimal noise, while the smaller window size results produced a smoother response but with considerably more noise. Simply taking the mean of the estimates produced a disparity map with sub-pixel disparity estimates although with substantial noise due to outliers. A median estimate produced a stepped disparity map with minimal visual errors. A third option is to discard the largest and smallest disparity estimates, assuming that these estimates are outliers. The mean of the remaining estimates produces a similar noise-reduced but stepped-discontinuous disparity estimate in the nose region.

2.5. Left to Right Consistency Check

A common check for correctness of disparity estimates is to perform correspondence from the left to right image and also from the right to left image. If the disparities don't match then the estimate is marked as bad. While this approach may be suitable for certain stereo matching applications, the resulting 3D model would contain insufficient disparity estimates.

The unique approach is to consider the error between the left and right disparity estimates. The result of calculating the error is presented in figure 3. It is evident that the greatest error is along the slope of the nose while the least error is present in the nose/cheek convergency region and the ridge of the nose.

The disparity estimate error is calculated separately for each window size. If the left-right disparity error is one or less pixels, the disparity estimate for that pixel is labelled as acceptable. If the number of pixels with acceptable disparity estimates is greater than a threshold (in this case 5 out of 7), the acceptable estimates are averaged to produce a final disparity estimate. The unacceptable disparities are mark as such and processed further.

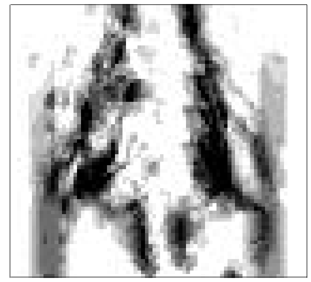

Fig. 3. The accumilated error image from Left-Right consistency check. The darker regions indicate that more of the window sizes has Left-Right consistency errors.

2.6. Using Prior Knowledge of the Nose

Selecting a particular region of the face allows certain assumptions and constraints to be applied to achieve improved results. These assumptions and constraints may not work across an entire face but achieve imrpoved results when localised to a face region.

The error plot of figure 3 illustrates that the largest left-right consistency error (the darkest region) is down the sloping side of the nose. Combining the Left-Right and Right-Left disparity estimates using an average, results in a broad nose tip and broad nose ridge shown in figure 4. These results arise from the averaging of erroneous disparity estimates with good disparity estimates. It is not possible to determine which disparity estimates are correct and which are incorrect without prior knowledge of the structure of the underlying depth data, in this case the nose. By analysing the error values we have determined that the sign of the error indicates which group of disparity estimates (Left-Right or Right-Left) provide the best disparity estimates. Thus for a negative error we select the Left-Right disparity estimates and the Right-Left disparity estimates for the positive error. Compare the resulting nose shapes from the standard averaging in figure 4 to the results from using the prior knowledge in figure 5.

3. RESULTS

Utilising the prior knowledge of the nose shape and the errors between the Left-Right and Right-Left disparity esti-

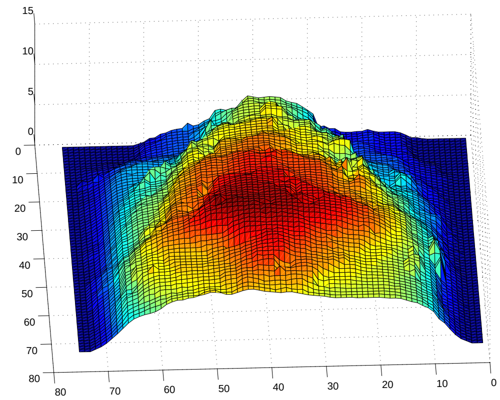

Fig. 4. The resulting 3D shape of the nose using Left-Right and Right-Left averaged disparity estimates. The nose has a broader shape due to the variations in disparity estimates.

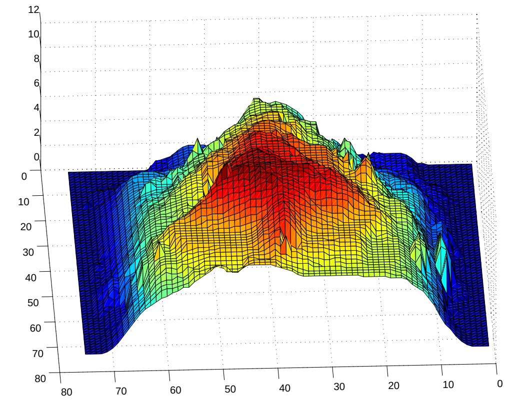

Fig. 5. The final 3D shape of the nose using prior knowledge regarding the results of Left-Right and Right-Left disparity estimates. The nose has a narrow shape due to the correct selection of disparity estimates.

mates, we can improve the shape of the nose by narrowing the nose width. Comparing a direct averaging of the disparity estimates from both the Left and Right images in figure 4 to those of a nose estimate utilising only the appropriate viewpoint and Left-Right consistent disparity, shown in figure 5.

4. CONCLUSION

This paper has presented a few improvements to stereo matching techniques for calculating disparity and hence the 3D structure specifically for the human face.

Using a colour matching algorithm improves the stereo matching for the human face regions containing minimal texture variation.

A coarse to fine matching criteria helps reduce the error associated with smaller stereo matching windows by constraining the matching region to previous determined disparity ranges. This also reduces the computational cost by reducing the search region.

The Left-Right consistency check identifies the disparity estimates which differ between the left and the right match persepectives. Normally these are discarded, however we have been able to apply this information in conjunction with prior knowledge of the nose structure to provide denser depth maps and to improve the nose shape.

This framework for 3D face reconstruction allows individual face regions to be considered and specific improvements made to each region. Its application to other parts of the face will be explored in future work.

Acknowledgements

This project was supported in part by a grant from the Office of Naval Research, USA, Award No: N000140310663.

5. REFERENCES

[1] G.J. Mamic and M. Bennamoun, *Object Recognition - Fundamentals and Case Study*, Springer Verlag, New York, 2002.

[2] J. Banks, *Reliability Analysis of Transform-Based Stereo Matching Techniques, and a New Matching Constraint*, Ph.D. thesis, SCSN, School of EESE, QUT, September 1999.

[3] R. Zabih D. Scharstein, R. Szeliski, “A taxonomy and evaluation of dense two-frame stereo correspondence algorithms,” in *Proceedings of Workshop on Stereo and Multi-Baseline Vision*, Kauai, Hawaii, December 2001, pp. 131–140.

[4] T. Kanade and M. Okutomi, “A stereo matching algorithm with an adaptive window: Theory and experiment,” *IEEE trans. on Pattern Analysis and Machine Intelligence*, pp. 920–932, 1994.

[5] E. Trucco A. Fusiello and A. Verri, “Rectification with unconstrained stereo geometry,” *Proceedings of the British Machine Vision Conference*, pp. 400–409, 1997.

[6] J. Banks and P. Corke, “Quantative evaluation of matching methods and validity measures for stereo vision,” *International Journal of Robotics Research*, vol. 20, no. 07, pp. 400–409, July 2001.

[7] Jrgen Hesser Karsten Mhlmann, Dennis Maier and Reinhard Mnner, “Calculating dense disparity maps from color stereo images, an efficient implementation,” *International Journal of Computer Vision*, vol. 47, no. 1-3, pp. 79–88, April - June 2002.