Multi-View 3D Data Acquisition Using a Single Uncoded Light Pattern

Conference Paper · July 2011

DOI: 10.13140/2.1.4434.4963

3 authors, including:

Stefania Cristina
University of Malta

Kenneth P Camilleri
University of Malta

MULTI-VIEW 3D DATA ACQUISITION USING A SINGLE UNCODED LIGHT PATTERN

Stefania Cristina[1], Kenneth P. Camilleri[1] and Thomas Galea[2]
[1]*Department of Systems and Control Engineering, University of Malta, Msida, Malta*
[2]*Megabyte Ltd, Mosta, Malta*

Keywords: 3D Reconstruction, Machine vision, Stereo, One-shot, Uncoded light pattern, Data fusion. Multiple views.

Abstract: This research concerns the acquisition of 3-dimensional data from images for the purpose of modeling a person's head. This paper proposes an approach for acquiring the 3-dimensional reconstruction using a multiple stereo camera vision platform and a combination of passive and active lighting techniques. The proposed one-shot active lighting method projects a single, binary dot pattern, hence ensuring the suitability of the method to reconstruct dynamic scenes. Contrary to the conventional spatial neighborhood coding techniques, this approach matches corresponding spots between image pairs by exploiting solely the redundant data available in the multiple camera images. This produces an initial, sparse reconstruction, which is then used to guide a passive lighting technique to obtain a dense 3-dimensional representation of the object of interest. The results obtained reveal the robustness of the projected pattern and the spot matching algorithm, and a decrease in the number of false matches in the 3-dimensional dense reconstructions, particularly in smooth and textureless regions on the human face.

1 INTRODUCTION

The acquisition of 3-dimensional data by stereovision techniques is a very widely researched area in the field of computer vision due to its wide range of applications. However, the range finders proposed to date still face major challenges such as the reconstruction of textureless surfaces.

Passive range finders depend only on stereo images acquired in ambient lighting to operate. Multi-view passive systems (Okutomi and Kanade, 1993; Kang et al., 2001; Gallup et al., 2008) provide larger surface coverage and higher depth accuracy when compared to passive two-camera systems. However, their performance is still challenged by texturless surfaces. Active range finders code the viewed scene with a light pattern to address the issues in reconstructing textureless surfaces. While temporal coding (Chang, 2003; Zhang et al., 2002) and direct codification (Miyasaka et al., 2000; Liang et al., 2007) techniques are not suitable for reconstructing non-static objects since they project multiple patterns, spatial neighbourhood coding techniques (Shi et al., 2005; Song and Chung, 2008) project only a single pattern but they are not robust to surface depth discontinuities.

The object of interest in this work that is sought to be 3-dimensionally reconstructed is an ear-to-ear representation of a person's head. A multi-view approach consisting of passive and active lighting techniques is used due to the textureless nature of the human face. The active technique projects a single, binary and uncoded dot pattern, and an algorithm is proposed which matches corresponding spots by exploiting the redundant data in multiple stereo images. This matching algorithm is not susceptible to surface depth discontinuities and results in an initial, sparse reconstruction. A dense 3-dimensional reconstruction is then obtained using a passive lighting technique, which is guided by the initial reconstruction to reduce the likelihood of occurrence of false matches.

This paper is organized as follows. Section 2 details the various stages of the proposed approach. Experimental results are discussed in Section 3. Section 4 presents the concluding remarks and suggests further work.

2 METHODS

The concept behind the proposed spot matching algorithm is that for every 3-dimensional point on the

Cristina S., P. Camilleri K. and Galea T. (2011).
MULTI-VIEW 3D DATA ACQUISITION USING A SINGLE UNCODED LIGHT PATTERN.
In *Proceedings of the 8th International Conference on Informatics in Control, Automation and Robotics*, pages 317-320
DOI: 10.5220/0003545203170320

surface of interest there is only one consistent and physical depth value on which every stereo pair in the multi-camera setup must agree, and this permits the identification of corresponding spots between multiple images. The multi-camera setup in this work comprises 4 cameras, which are calibrated using Bouguet's *Camera Calibration Toolbox for Matlab* (Bouguet, 2010).

Spot Pattern Extraction: A simple, single pattern composed of centered horizontal and vertical stripes, spanning from one border to the other, and binary dots is projected on the object of interest, as shown in Figure 1. The stripes delineate quadrant sets of dots allowing each quadrant to be processed separately and inhibiting every dot matching error from propagating into other quadrants.

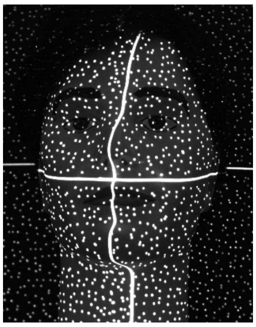

Figure 1: Projection of the light pattern onto the object of interest, as acquired from one of the cameras.

The spot pattern is extracted by binarising the images of the object and the projected light pattern, and filling any holes in the spots. The centroid of each extracted spot is then found and used to represent the approximate location of each projected spot. This yields the images $f_i^{(u)}$ for any spot quadrant, where camera index $i = 1,\ldots,n$, and superscript u denotes unrectified images.

Spot Matching: In order to generate a sparse reconstruction of the object of interest, corresponding spots between multiple stereo pairs need to be identified. Contrary to the active techniques reviewed in Section 1, which rely on coded patterns to identify correct pixel correspondences, the proposed algorithm exploits the redundant data in multiple stereo images to match each spot in the uncoded binary pattern. This ensures that the method is unaffected by the underlying surface colour and is not susceptible to surface depth discontinuities. The spot matching procedure first involves identifying candidate matches and then identifies the most likely match on the basis of range consistency.

In order to identify the possible matching spots in the pattern, the binary images are first rectified with respect to a reference camera i according to (Fusiello et al., 2000), in order to reduce the search problem for correspondences to a 1-dimensional search along epipolar lines. This yields the pairs of images, $f_i^{(r(i,j))}$ and $f_j^{(r(i,j))}$ for cameras indexed i and j respectively, where $j = 1,\ldots,n,\ j \neq i$, with a particular rectification $r(i,j)$ between camera pair i, j. The spots in images $f_i^{(r(i,j))}$ and $f_j^{(r(i,j))}$ are also arbitrarily assigned a unique index value $k^{(i)}$ and $k^{(j)}$ respectively, for identification purposes.

Now, consider the spot indexed $k^{(i)}$. Its true match must theoretically lie on the epipolar line in image $f_j^{(r(i,j))}$. Therefore, the indices of the spots residing on the epipolar line are included in the set, $c_j(k^{(i)}) = \{k_1^{(j)}, k_2^{(j)}, \ldots, k_N^{(j)}\} = \{k_\eta^{(j)}, 1 \leq \eta \leq N\}$, of candidate matching dots for the spot with index, $k^{(i)}$. In addition, spots residing on a number of rows above and below the epipolar line are also included in the set of candidate matches, since the estimation of each spot location may shift the true match off the epipolar line.

Now, each candidate match is assigned a score and a depth value. The score value was chosen to be a linearly decreasing function of distance from the epipolar line. The depth value $Z^{(r(i,j))}(k^{(i)}; k_\eta^{(j)})$ is calculated by triangulation between the reference spot with index value, $k^{(i)}$, and each candidate match, $k_\eta^{(j)}$, in $c_j(k^{(i)})$ as described by Equation 1.

$$Z^{(r(i,j))}(k^{(i)}; k_\eta^{(j)}) = \frac{B_{i,j} F^{(r(i,j))}}{d(k^{(i)}, k_\eta^{(j)})} \tag{1}$$

where, $\eta = 1,\ldots,N$, $B_{i,j}$ denotes the baseline length of the stereo pair (i,j), $F^{(r(i,j))}$ denotes the common focal length of cameras i and j, and $d(k^{(i)}, k_\eta^{(j)})$ is the disparity value in pixel units between spot index $k^{(i)}$ in $f_i^{(r(i,j))}$ and the candidate matching spot index $k_\eta^{(j)}$ in $f_j^{(r(i,j))}$.

Since the depth value of the true match must be consistent between all stereo pairs, this true match can be identified by seeking that candidate match that has the same depth value in all stereo pairs. However, any inaccuracies in the camera calibration parameters and the inaccurate approximation of each spot location may cause the depth values of the true match to vary by some value between different stereo pairs. To counteract this discrepancy and identify the true match, a weighted histogram of the depth values of each candidate match is generated, where each candidate match is weighted by its score value forming a histogram such as shown in Figure 2. A separate mapping table is also used to retain the relationship be-

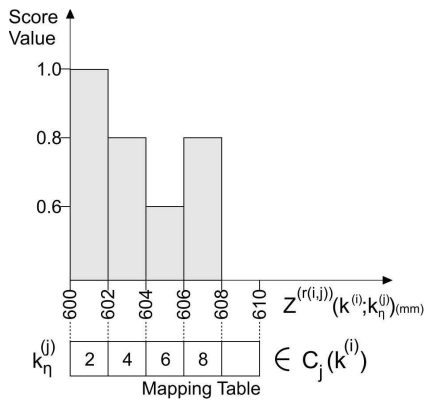

Figure 2: A weighted histogram of the depth values of each candidate match is generated for each stereo pair, and a separate mapping table retains the relation between the index value and depth value of each candidate match.

tween the index, $k_\eta^{(j)}$, of a particular candidate match and its depth value. The histogram of each stereo pair is aligned and summed to generate a single global histogram. The mode depth interval is then chosen to represent the range, and the corresponding spots are identified and taken to represent the same dot in the pattern.

Post-Processing of the Matched Spot Pattern: Theoretically, true spot matches have similar intensity values in stereo image pairs acquired in ambient lighting. Ignoring any non-linearities in camera sensitivity, all correctly matched spots should fit a straight line. Therefore, the intensities of all matching spot pairs are plotted in a 2-dimensional plot where each axis represents the intensity of each camera in the pair, as shown in Figure 3. The best straight line is then fit to the distribution by linear regression, and those points whose distance from the line is above a certain tolerance are taken to represent mismatches and are discarded.

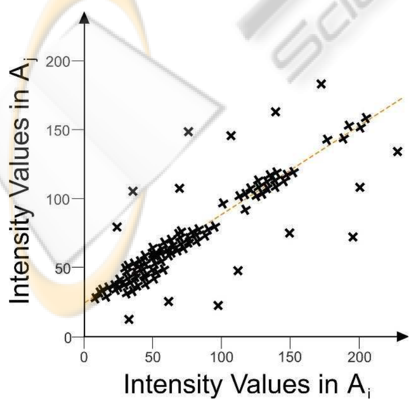

Figure 3: The intensities of all matching spot pairs are plotted, where A_i and A_j denote the ambient lighting images acquired from cameras *i* and *j* respectively.

Localized Correspondence and 3D Reconstruction: The initial, sparse representation of the object of interest is used to guide a passive correspondence technique in order to obtain a dense 3-dimensional reconstruction. All stereo image pairs are being intensity calibrated by adopting a modified version of (Kawai and Tomita, 1998).

The passive correspondence method used to obtain a dense 3-dimensional representation of the object of interest, is based on the Sum of Sum of Square Differences (SSSD) in-inverse-distance technique by Okutomi and Kanade (1993). In the proposed approach, the correspondence search range is constrained for each individual pixel using the set of matched spots. This increased the robustness of the correspondence algorithm, as detailed in Section 3.

3 RESULTS AND DISCUSSION

The proposed approach was tested by projecting a random dot pattern on a test mannequin, as shown in Figure 1. From the results obtained it was found that on average the spot matching algorithm yielded 97.53% of correctly matched spots per pattern quadrant.

For comparison purposes, the reconstruction in Figure 4(a) was generated from intensity calibrated images using the method in (Okutomi and Kanade, 1993). This result shows several false matches on surface regions which mostly lack any discernible texture such as the forehead, the cheeks and the chin, when compared to the result obtained using the proposed approach in Figure 4(b). The occurrence of these false matches is a direct consequence of the large search range considered in (Okutomi and Kanade, 1993). The SSSD-in-inverse-distance function in Figure 5(a) of a false match in Figure 4(a) shows multiple minima whose SSSD value is very close, and a global minimum at the wrong inverse-distance value. On the other hand, constraining the search range in the proposed approach results in the SSSD-in-inverse-distance function in Figure 5(b), which contains a single minimum and correctly identifies the inverse-distance value of the true match.

4 CONCLUSIONS

This work has dealt with the 3-dimensional reconstruction of a person's head using a multi-camera setup and a combination of active and passive lighting methods, to combine their merits and obtain a more reliable 3-dimensional reconstruction.

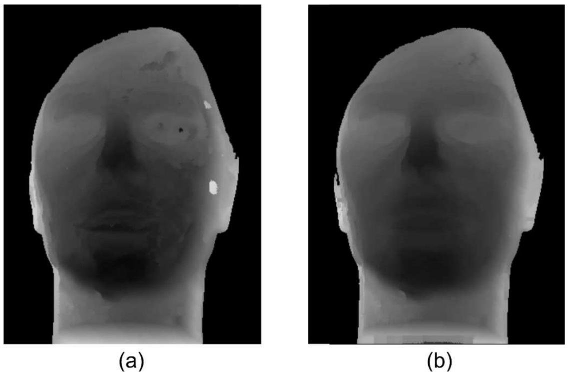

Figure 4: The depth map computed using the approach in (Okutomi and Kanade, 1993) (a), compared to the depth map generated using the proposed approach (b).

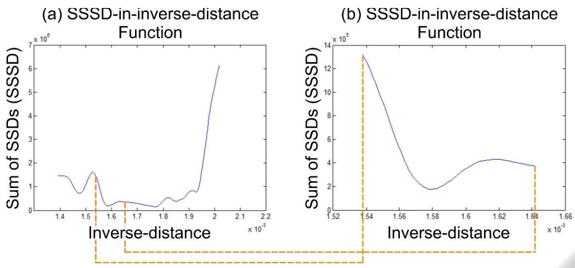

Figure 5: The large search range considered in (Okutomi and Kanade, 1993) and the lack of surface texture result in a global minimum at the wrong inverse-distance value (a), while constraining the search range in the proposed approach ensures the SSSD-in-inverse-distance function to contain a single minimum at the true match (b).

The projection of a single, uncoded binary pattern and the proposed spot matching algorithm are applicable in reconstructing non-static surfaces which contain depth discontinuities and are not colour-neutral. This generates a sparse but reliable 3-dimensional reconstruction, which is then used to constrain the search range of a passive correspondence technique to produce a dense reconstruction of the object of interest.

Further work includes scaling up of the system to include a larger number of cameras, in order to enhance the performance of the spot matching and correspondence algorithms due to the availability of a larger amount of multi-view redundant data.

ACKNOWLEDGEMENTS

This work is part of the project '3D-Head' funded by the Malta Council for Science and Technology under Research Grant No. RTDI-2004-034.

REFERENCES

Bouguet, J. Y. (2010). Camera calibration toolbox for matlab. Available at: http://www.vision.caltech.edu/bouguetj/calib_doc/. [Accessed January 2010].

Chang, N. L. (2003). Efficient dense correspondences using temporally encoded light patterns. In *IEEE International Workshop on Projector-Camera Systems*.

Fusiello, A., Trucco, E., and Verri, A. (2000). A compact algorithm for rectification of stereo pairs. In *Machine Vision and Applications*. Springer-Verlag.

Gallup, D., Frahm, J. M., Mordohai, P., and Pollefeys, M. (2008). Variable baseline/resolution stereo. In *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society.

Kang, S. B., Szeliski, R., and Chai, J. (2001). Handling occlusions in dense multi-view stereo. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society.

Kawai, Y. and Tomita, F. (1998). Intensity calibration for stereo images based on segment correspondence. In *International Association of Pattern Recognition, Workshop on Machine Vision Applications*.

Liang, Z., Gao, H., Nie, L., and Wu, L. (2007). 3d reconstruction for telerobotic welding. In *Proceedings of the 2007 IEEE International Conference on Mechatronics and Automation*.

Miyasaka, T., Kuroda, K., Hirose, M., and Araki, K. (2000). High speed 3-d measurement system using incoherent light source for human performance analysis. In *Proceedings of the 19th Congress of The International Society for Photogrammetry and Remote Sensing*.

Okutomi, M. and Kanade, T. (1993). A multiple-baseline stereo. In *IEEE Transactions of Pattern Analysis and Machine Intelligence*. IEEE Computer Society.

Shi, L., Yang, X., and Pan, H. (2005). Face modeling using grid light and feature point extraction. In *Proceedings of the International Conference on Computational Science and its Applications*. Springer Berlin/Heidelberg.

Song, Z. and Chung, R. (2008). Grid point extraction exploiting point symmetry in a pseudo-random color pattern. In *Proceedings of the 15th IEEE International Conference on Image Processing*.

Zhang, L., Curless, B., and Seitz, S. M. (2002). Rapid shape acquisition using color structured light and multi-pass dynamic programming. In *Proceedings of the First International Symposium on 3D Data Processing Vizualization and Transmission*. IEEE Computer Society.