

Active stereo-matching for one-shot dense reconstruction

Sergio Fernandez, Josep Forest and Joaquim Salvi
Institute of Informatics and Applications, University of Girona,
Av. Lluís Santalo S/N, E-17071 Girona (Spain)
{sergiofn, forest, qsalvi}@eia.udg.edu

Keywords: Pattern projection: Stereo matching: Dense reconstruction: 3D measuring devices: Active stereo: Computer vision.

Abstract: Stereo-vision in computer vision represents an important field for 3D reconstruction. Real time dense reconstruction, however, is only achieved for high textured surfaces in passive stereo-matching. In this work an active stereo-matching approach is proposed. A projected pattern is used to artificially increase the texture of the measuring object, thus enabling dense reconstruction for one-shot stereo techniques. Results show that the accuracy is similar to other active techniques, while dense reconstruction is obtained.

1 Introduction

Three dimensional reconstruction using non-contact techniques constitutes an active topic in computer vision (Salvi et al., 2010). Among them, applications requiring real time response have increased during the last years due to the necessity to measure moving objects. Some techniques consist of fast capturing cameras imaging a set of projected patterns of the measuring object. However, the ability to work in real time regardless the motion of the object (up to the acquisition time required by the camera) is only achieved by one-shot techniques. Two main one-shot approaches can be distinguished depending on whether it is passive or active-based approach. In passive approaches, also named as stereo, two or more images are taken from different camera point of view, and the 3D model of the scene is computed by means of matching pixels in the images and capturing their corresponding 3D depth by triangulation (Szeliski, 2010). Area-based and feature-based approaches are proposed for the matching step (Hu and Ahuja, 1994). The former provides a dense reconstruction at the expense of a high error rate for occluded areas or low-textured regions, whereas the latter makes use of a feature extraction algorithm in order to provide a set of reliable points for matching, obtaining better accuracy. However, the density of the reconstruction is directly related to the texture of the object (Batlle et al., 1998). Active approaches, based on structured light (SL), project a pattern that creates artificial texture on the required object (Salvi

et al., 1998). SL techniques are implemented using one or more than one camera (passive device), and one projector (active device). Usually, the projector is modelled as an inverse of the camera, thus being the projector-camera calibration similar to the one used in classical stereo-vision (Falcao et al., 2008). However, this approach suffers from projector distortion, sharp changes in 3D depth and color/albedo when a colored pattern is employed. Another approach uses a pair of calibrated cameras and a non-calibrated projector to increase the texture and ease the feature extraction. Posteriorly, triangulation is done with the two camera images by means of a feature-based stereo matching algorithm. This solution provides lower distortion than the projector calibrated methods, while preserving the texture independence of active approaches.

This paper is structured as follows: section 2 provides some background on stereo-matching, section 3 details our new approach on active dense stereo-matching. Section 4 analyses the results obtained both quantitatively and qualitatively. Finally, section 5 summarizes the conclusions of this work, giving some ideas for further improvements.

2 Brief overview of stereo-matching techniques

Area-based stereo-matching algorithms have been widely studied in the literature (Faugeras and Keriven, 1998) (Kanade and Okutomi, 1994) (Brown et al., 2003). A correlation window is run across the

two images, providing the most likely correspondence between pixels in the two images. The size of the windows determines the number of pixels to consider for the correlation step, directly related to the accuracy of the system. However, errors arise under occluded or poorly textured regions of the images. Moreover, high computation time is required to perform the pixel-to-pixel correspondences map. In the contrary, feature-based algorithms use only relevant points in the image that are match together to create the correspondences. Using the so called Scale Invariant Feature Transform (SIFT) algorithm proposed by Lowe (Lowe, 2004) the image is transformed into a large collection of feature vectors. Key locations are defined as maxima and minima of the difference of Gaussians (DoG) kernel, applied in scale-space to a series of smoothed and resampled images. Other feature detectors have been proposed substituting the DoG by other feature extraction filters. The detectors are classified in two main groups, regarding their ability to detect corners (SIFT, Harris) or blobs (Hessian, MSER). Invariance to image translation, scaling, rotation and partially invariance to illumination changes and robustness to local geometric distortion can be achieved. Finally SIFT descriptor is computed for every keypoint, providing dominant orientations to the localized positions in a 128 gradient length vector. This ensures that the keypoints are more stable for matching and recognition. Another algorithm, called Speeded Up Robust Features (SURF) (Bay et al., 2008), uses a simplified version of SIFT, obtaining slight lower accuracy results whereas the computation time is reduced. More recently, the so called DAISY algorithm proposed by Tola et al. (Tola et al., 2008) has been proposed, performing an area-based features matching for a set of initial point seeds. Local descriptors are used for every pixel, replacing the classical correlation windows with local region information. Therefore, a matching can be pursued pixel by pixel for all positions in the window patch, based on their DAISY descriptors. The descriptor are based on Gaussian filters, giving more importance regarding the distance of the analysed pixel to the initial seed. This algorithm prevents from the errors that occurs in the classical area-based method for low textured images. However, the computation time results drastically increased as not a pixel but a vector must be compared for every position in the image.

3 A novel proposal on active stereo-matching

The proposed technique employs a pair of calibrated cameras and a non calibrated projector. The projector is used to project a previously designed pattern onto the surface to reconstruct, so as to increase the texture provided by the surface. The textured surface is then captured by both cameras, and a feature-based matching algorithm is applied. Thanks to the rich texture present in the images, a dense 3D reconstruction of the scene can be obtained. The flow diagram of the proposed method is shown in Fig. 1:

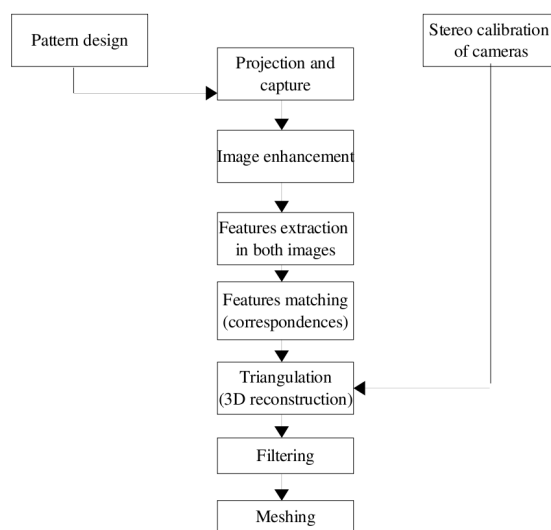


Figure 1: Flow diagram of the proposed algorithm.

3.1 Pattern design

The design of an optimal pattern for increasing the texture in stereo-vision has been previously studied in the literature. For instance, Tehrani et al. (Tehrani et al., 2008), proposed a trinocular system with a color slit based pattern that maximizes the distance in the Hue channel between adjacent slits. However, it is reasonable to employ both coding axis in order to increase the features in the recovered images. This is the case of M-arrays, like the ones proposed by Griffin et al. (Griffin et al., 1992) and Morano et al. (Morano et al., 1998). Moreover, it is important to consider the need to minimize the error against changes in color and albedo. This leads to a maximization of the hamming distance between adjacent areas. In order to submit these requirements, we proposed a zero mean greyscale squared random pattern, as can be observed in Fig. 2. The size of the squares must be set as

the minimum of the joint resolution of the projector-camera system, so as to obtain dense reconstruction. In our case this corresponds to 4 pixels per square, as the density of the reconstruction is limited by the resolution of the camera.

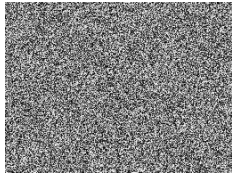


Figure 2: Random squared greyscale pattern proposed for projection.

3.2 Stereo camera calibration

Cameras calibration has been performed using an adapted version of the stereo-calibration Matlab toolbox provided by DeBouquet (Bouquet, 2004). The cameras are calibrated using a checkerboard pattern, capturing different positions and angles (from 15 up to 20 images were employed). It is important to mention that both radial and tangential distortion of both cameras were considered for calibration. For more information about this algorithm see the work of DeBouquet (Bouquet, 2004).

3.3 Projection and capture

The projector is placed between the two cameras, so as to have the most similar projection pattern seen from the cameras and avoid effects caused by defocusing. The cameras (which must have identical characteristics) capture the same artificially textured scene seen from two different points of view, in a one-shot image.

3.4 Image enhancement

An algorithm to increase the contrast between adjacent pixels is applied to the captured images. This reduces the effect of different albedo and different illumination. To this end, a DC subtraction algorithm followed by an local histogram equalization is pursued. This provides good filtering results at a very low computational cost.

3.5 Features extraction

In this work, the feature-based approach is combined with the artificially textured images. Therefore dense matching can be achieved while preserving a low error rate as no textureless areas are presented. The

proposed technique was implemented with a combination of Harris corners detector and MSER blobs detector. SIFT algorithm was selected to create the vector descriptors.

3.6 Features matching

The algorithm used to find correspondences is based on the similarity between two descriptors corresponding to the same textured feature in the object shape. That is, two descriptors coming from the 2D images correspond to the same 3D point if their module and orientations are similar. This similarity is measured based on the Euclidean distance of their feature vectors.

3.7 Triangulation

Every pair of (x,y) coordinates given by the matching step are for the triangulation, using the extrinsic and extrinsic parameters of the cameras provided by the calibration step. The algorithm returns a point in 3D space corresponding to the 2D projections in image planes provided as input. The output is a cloud of points in (x,y,z) representing the shape of the reconstructed object.

3.8 Filtering

A post-triangulation filtering step reveals necessary due to some erroneous matchings that originate outliers in the 3D cloud of points. Three different filtering steps are applied regarding the constraints given by epipolar geometry, the statistical 3D positions and the bilateral-based filter.

Epipolar geometry based filtering: according to the epipolar geometry, all pixel in the left image create an epipolar line in the right image (Faugeras and Toscani, 1986). The actual corresponding pixel in the right image should pass through this line, or remain closer than a given tolerance value. If the pixel lies too far from its corresponding epipolar line the correspondence is considered as wrong, and the pair of pixels is discarded.

3D statistical filtering: in the 3D space, the outliers are characterised by their extremely different 3D coordinates regarding the surrounding points. Therefore, pixels having a 3D coordinates different than the 95% of points distribution are considered for suppression. This is done in two steps for all the points in the 3D cloud. First the distance to the centroid of the cloud of points is computed, for every pixel. Afterwards, those pixels having a distance to the centroid

greater than two times the standard deviation of the cloud of points are considered as outliers.

Bilateral filtering: an extension to 3D data of the 2D bilateral filter proposed by Tomati and Manduchi (Tomasi and Manduchi, 1998) was implemented. The algorithm smoothes the 3D cloud of points up to a given value, while the discontinuities are preserved. As result, the variance due to little misalignment in the matching process is reduced, decreasing the error in the reconstructed surface.

3.9 Meshing

Finally, a meshing step is applied to obtain a surface from the 3D cloud of points. To do so, a bidimensional Delaunay meshing algorithm is applied on coordinates (x,y).

4 Experimental results

The proposed algorithm has been tested under real conditions. To this end, a pair of cameras Canon EOS50D 10Mpixels have been employed. The projector is a Epson EMP400W 768x1024 resolution.

4.1 Quantitative results

In order to evaluate the system in terms of accuracy, a flat plane has been reconstructed using the proposed algorithm, at a distance of 50cm. At this point, it is important to remark that other stereo dense reconstruction algorithms are texture dependent in terms of accuracy. Therefore, the comparison has been done with other active techniques, which do not depend on the texture of the given surfaces. The proposed technique has been compared with a one-shot technique (proposed by Monks et al. (Monks et al., 1992)) providing sparse reconstruction, and with a time-multiplexing technique (Guhring (Guhring, 2001)) which provides dense reconstruction for static scenarios. Principle Component Analysis (PCA) was applied to obtain the equation of the 3D plane for every technique and for every reconstruction. This technique is used to span the 3D cloud of points onto a 2D plane defined by the two eigenvectors corresponding to the two largest eigenvalues. The results of mean error and standard deviation of the 3D cloud of points with respect to the reconstructed plane are shown in table 1.

As can be observed, the proposed technique performs better in terms of standard deviation than the other active techniques, as the same time dense re-

| Technique | Average (mm) | Stdev (mm) |
|---------------|--------------|------------|
| Monks et al. | 1.21 | 1.20 |
| Guhring | 3.96 | 2.46 |
| Active stereo | 1.30 | 0.75 |

Table 1: Results for a flat plane (proposed calibration technique; Average deviation of the reconstructing error; Standard deviation of the reconstructing error).

construction is provided regardless the texture of the plane.

4.2 Qualitative results

The algorithm has been tested with a real object, as can be observed in Fig. 3. The 3D shape was reconstructed optimally, obtaining high density in the 3D cloud of points. The little holes at the base of the object correspond to the suboptimal projection of the pattern, which leads to blurred regions that are filtered out after the 3D reconstruction. Nevertheless, dense reconstruction is achieved for optimal projection of the pattern onto the object.

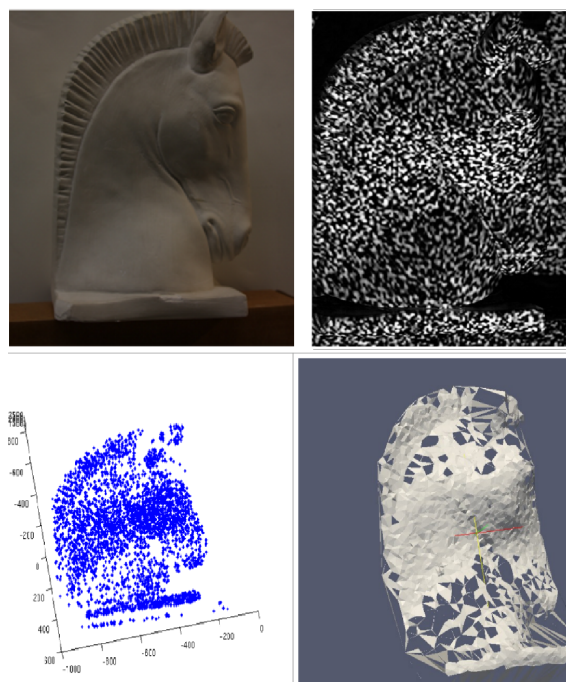


Figure 3: Reconstruction of a ceramic horse of dimensions 20x30cm. The figures are: original surface. Captured imaged after pattern projection. Filtered 3D cloud of points. Mesh.

5 Conclusions

In this work an active feature-based stereo-matching algorithm is proposed. The projected random pattern provides artificial textured images not depending on the nature of the scene. The selected black and white pattern maximises the hamming distance between adjacent points. The features extraction and matching was pursued using Harris and MSER detector, and SIFT descriptors. Posteriorly, a filtering algorithm based on three different approaches was implemented after triangulation in order to suppress outliers, thus giving reliable results at the expense of losing some density of reconstruction in occluded or blurred areas of the object. The quantitative results show the good performance of the proposed system in terms of accuracy of the reconstructed cloud of points. Moreover, qualitative results show the performance of the system under real shaped objects and normal light conditions. The use of a trinocular, where the projector is calibrated with respect to the cameras, is considered for future work. This, jointly with the use of structured light, constrains the matching process providing more accurate results. However, the complexity of finding correspondences and performing the triangulation of the extracted features in the images increases by a factor of three, providing much higher computation time.

Acknowledgments

This work is supported by the FP7-ICT-2011-7 project PANDORA: Persistent Autonomy through Learning, Adaptation, Observation and Re-planning funded by the European Commission and the project RAIMON: Autonomous Underwater Robot for Marine Fish Farms Inspection and Monitoring (CTM2011-29691-C02-02). S. Fernandez is supported by the Spanish government scholarship FPU.

REFERENCES

- Batlle, J., Mouaddib, E., and Salvi, J. (1998). Recent progress in coded structured light as a technique to solve the correspondence problem: a survey. *Pattern Recognition*, 31(7):963–982.
- Bay, H., Tuytelaars, T., and Van Gool, L. (2008). Surf: speeded-up robust features. In *9th European Conference on Computer vision*, volume 110, pages 346–359.
- Bouguet, J. (2004). Camera calibration toolbox for matlab.
- Brown, M., Burschka, D., and Hager, G. (2003). Advances in computational stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 993–1008.
- Falcao, G., Hurtos, T., and Massich, J. (2008). Plane-based calibration of a projector camera system. *VIBOT master*, (1):1–12.
- Faugeras, O. and Keriven, R. (1998). Complete dense stereovision using level set methods. *Computer VisionECCV'98*, pages 379–393.
- Faugeras, O. and Toscani, G. (1986). The calibration problem for stereo. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 86, pages 15–20.
- Griffin, P., Narasimhan, L., and Yee, S. (1992). Generation of uniquely encoded light patterns for range data acquisition. *Pattern Recognition*, 25(6):609–616.
- Gühring, J. (2001). Dense 3-D surface acquisition by structured light using off-the-shelf components. *Videometrics and Optical Methods for 3D Shape Measurement*, 4309:220–231.
- Hu, X. and Ahuja, N. (1994). Matching point features with ordered geometric, rigidity, and disparity constraints. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1041–1049.
- Kanade, T. and Okutomi, M. (1994). A stereo matching algorithm with an adaptive window: Theory and experiment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 920–932.
- Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110.
- Monks, T., Carter, J., and Shadle, C. (1992). Colour-encoded structured light for digitisation of real-time 3D data. In *Proceedings of the IEE 4th International Conference on Image Processing*, pages 327–30.
- Morano, R., Ozturk, C., Conn, R., Dubin, S., Zietz, S., and Nissano, J. (1998). Structured light using pseudorandom codes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):322–327.
- Salvi, J., Batlle, J., and Mouaddib, E. (1998). A robust-coded pattern projection for dynamic 3d scene measurement. *Pattern Recognition Letters*, 19(11):1055–1065.
- Salvi, J., Fernandez, S., Pribanic, T., and Llado, X. (2010). A state of the art in structured light patterns for surface profilometry. *Pattern recognition*, 43(8):2666–2680.
- Szeliski, R. (2010). *Computer vision: Algorithms and applications*. Springer-Verlag New York Inc.
- Tehrani, M., Saghaeian, A., and Mohajerani, O. (2008). A New Approach to 3D Modeling Using Structured Light Pattern. In *Information and Communication Technologies: From Theory to Applications, 2008. ICTTA 2008.*, pages 1–5.
- Tola, E., Lepetit, V., and Fua, P. (2008). A fast local descriptor for dense matching. In *Proc. CVPR*, volume 2. Citeseer.
- Tomasi, C. and Manduchi, R. (1998). Bilateral filtering for gray and color images. In *Computer Vision, 1998. Sixth International Conference on*, pages 839–846. IEEE.