Image-Based Motion Compensation for Structured Light Scanning of Dynamic Surfaces

Sören König and Stefan Gumhold

TU Dresden, SMT,
Department for Computer Sciences,
D-01062 Dresden
Fax: +49 (0)351 463 38396
E-mail: Soeren.Koenig@inf.tu-dresden.de
E-mail: Stefan.Gumhold@tu-dresden.de

Abstract: Many structured light scanning systems based on temporal pattern codification produce dense and robust results on static scenes but behave very poorly when applied to dynamic scenes in which objects are allowed to move or to deform during the acquisition process. The main reason for this lies in the wrong combination of encoded correspondence information because the same point in the projector pattern sequence can map to different points within the camera images due to depth changes over time. We present a novel approach suitable for measuring and compensating such kind of pattern motion. The described technique can be combined with existing active range scanning systems designed for static surface reconstruction making them applicable for the dynamic case. We demonstrate the benefits of our method by integrating it into a gray code based structured light scanner, which runs at thirty 3d scans per second.

Keywords: 3d scanning; motion compensation; optical flow; structured light, dynamic surfaces.

1 Introduction

A structured light scanner typically consists of a projector-camera pair. Different light patterns are projected onto the scene such that the projector column can be reconstructed at every pixel in the acquired camera images. Different calibration methods (Zhang, Z., 2000; Rocchini, C., Cignoni, P.,Montani, C. , Pingi, P. and Scopigno, R., 2001) are used to map camera pixels to 3d rays and projector columns to 3d planes. Simple ray-plane intersection finally yields 3d surface points.

The projector column can be coded in one pattern via spatial correspondences (i.e. one-shot approaches), in several patterns multiplexed over time (i.e. gray code), in light intensity (i.e. phase shift) or in a combination of these approaches. The purely spatial coding of projector columns in one-shot approaches is at first sight very attractive because of high frame rates and very simple realization in hardware. On the other hand it is extremely hard to deal with textured surfaces and depth discontinuities. Therefore time-multiplexing of several patterns is necessary for most applications. For static scenes time-multiplexing is a well established approach. Application to dynamic scenes is much more complicated as the temporal correspondence is destroyed by the motion of the projector patterns in the camera images, which is induced by the scene motion.

Partial compensation of the pattern motions are possible by coding the column information in intensity changes as done by the stripe boundary code approach (Hall-Holt, O., Rusinkiewicz, S., 2001). Intensity edges are detected and matched over time. This allows the compensation of pattern motions that are in the order of the stripe widths. But faster motions cannot be compensated.

In this paper we propose a motion compensation scheme that introduces an additional tracking pattern in between the structured light patterns. The special tracking pattern is optimized for maximal tracking performance. Although more patterns have to be acquired in our approach, the improved tracking allows 3d scanning of dynamic scenes with faster motions.

After a discussion of related work, we analyze the different problems introduced by scene motion in multi-pattern structured light methods. Then we design a good tracking pattern and validate its effectiveness. Scanning results are shown in section 4 before the conclusions in section 5.

2 Related Work

A wide variety of techniques has been presented for the problem of 3d shape acquisition (Mouaddib, E., Batlle, J. and Salvia, J., 1997; Salvi, J., Pages, J. and Batlle, J., 2004; Blais, F., 2004). One of the most studied methods in this area is stereovision. Depth estimations are done by triangulation; therefore it is necessary to find spatial correspondences between a pair of stereo images. An overview of many different algorithms for this task can be found in (Scharstein, D. and Szeliski, R., 2002). Especially interesting is the work by (Davis, J., Ramamoorthi, R. and Rusinkiewicz, S. 2003) and (Zhang, L., Curless, B. and Seitz, S. M., 2003) who extent existing stereo-matching techniques into the time domain to increase robustness. In principal stereovision is capable of handling dynamic scenes.

Another common way to reduce matching ambiguity and computational complexity is to exchange one camera by a projector, illuminating the scene with light patterns to simplify the search for correspondences. One-shot scanner would produce the smallest error introduced by motion because they reduce the time for measuring on a minimum. Beside this, they are forced with the problem to put all reconstruction information into one pattern leading to a trade off between sampling density and robustness. (Carrihill, B., Hummel, R., 1985) use an intensity ramp to directly encode positions, resulting in noisy scans with limited resolutions. Other methods like (Maruyama, M. and Abe, S., 1989), (Vuylsteke, P. and Oosterlinck, A., 1990) and (Koninckx, T. P., Griesser, A. and Van Gool, L. 2003) rely on neighbourhood coding. In these cases, markers like stripes, or other templates associated with a spatially encoded identifier are used. (Zhang, L., Curless, B. and Seitz, S. M.) use colour codification. (Adan, Antonio, Molina, Fernando, Vazquez, Andres S. and Luis Morena, 2005) apply stereo matching algorithms to find correspondences between a projected noise pattern and the acquired camera image. Most of the multi-pattern scanners are designed for static scenes. Binary and gray code sequences rely on temporal code sequence of on and off pixel states to encode projector columns (Mouaddib, E., Batlle, J. and Salvia, J. 1997).

Another popular method called ”‘phase-shift”’ projects a set of at least three phase-shifted sinusoidal patterns (Wust, C. and Capson, D. W. 1991), (Song Zhang and Peisen Huang, 2004). For reconstruction the global phase positions have to be recovered at each point. This method is very popular because of its high accuracy. To acquire dynamic scenes multi-pattern sequences must be capable of taking motion into account. (Hall-Holt, O., Rusinkiewicz, S. 2001) address this problem by tracking stripe boundaries. (Weise, T., Leibe, B. and Van Gool, L., 2007) propose a real-time hybrid stereo phase-shift method with automatic motion compensation. This is done by analysis of the motion error on pixel level. The proposed approach on the other hand can be combined with all multi-pattern structured light approaches and we demonstrate this at the example of a gray-code scanner.

3 Reconstruction with Motion Compensation

Let us first introduce some notation to describe time-multiplexed structured lights approaches. Let $P_1, \ldots, P_n$ be the n different light patterns and $C_1^s, \ldots, C_n^s$ the acquired camera images of successive sequences indexed through s. Each structured light approach comes with a decoding procedure that allows reconstruction of the projector column j from the different camera images at each pixel location (x, y)

$$j(x, y) = rec\left(C_1^s(x, y), \ldots, C_n^s(x, y)\right)$$

In the next two subsections we analyze the two major sources for reconstruction errors introduced due to motion during the acquisition process.

3.1 Object and Pattern Motion

Figure 1 illustrates the two different kinds of motions that result in the camera images from the motion of the scene. On the left side the plane is moving in tangential direction to the right. The surface motion becomes visible through the motion of the surface texture in the camera image. The corresponding motion is denoted with v_t and illustrated with dotted arrows. The 3d geometry does not change and the illustrated point of the projector pattern stays fixed in the camera

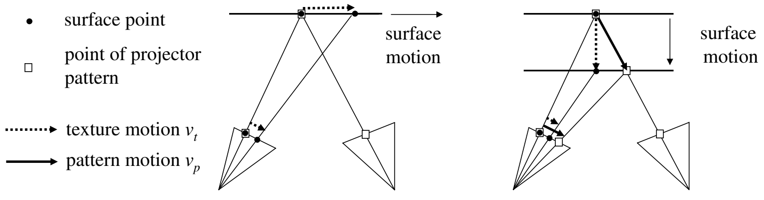

Figure 1 Illustration of texture and pattern motion.

image yielding a pattern motion of $v_p = 0$. On the right side the plane is moving in normal direction. This time both the surface point and the projector point move in the camera image, but with different velocities $v_t \neq v_p$.

While in static scenes information from successive patterns project to the pixel in the camera images, dynamic scenes induce pattern motion that have to be tracked in order to ensure combining the correct reconstruction information from multiple patterns.

3.2 Pattern Separation

Texture motion also complicates the task of separating the structured patterns from the rest of the camera images. This task also involves the removal of influences from surface texture, ambient illuminations, intensity changes due to surface orientation, etc.. Normally this is done by capturing an on- and an off-reference pattern P_{on}/P_{off}. The acquired reference images C_{on}/C_{off} are used to find non-illuminated regions like shadows or background, to estimate per pixel thresholds for binarization and allow normalizing scene-based intensity variations.

In dynamic acquisition on- an off-camera-images change over time due to texture motion. Therefore on- and off-references need to be introduced in between each pattern to guarantee good pattern separation, resulting in the pattern sequence $P_{on}, P_1, P_{off}, P_2, ..., P_n, P_{off}$.

3.3 Motion Compensation

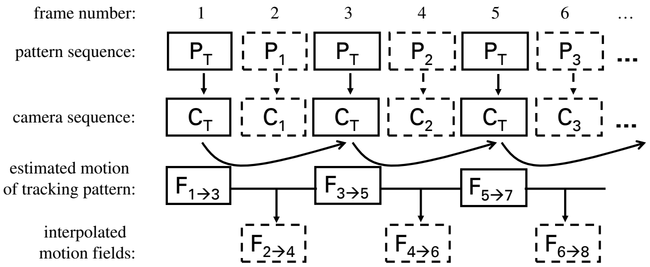

Figure 2 Steps in pattern motion estimation in a sequence of reconstruction patterns.

For the correct 3d reconstruction we only need to know the pattern motion $v_p(x, y)$ for each pixel. We estimate $v_p(x, y)$ by interleaving a tracking pattern P_T after each reference frame resulting in the pattern sequence $P_{on}, P_T, P_1, P_{off}, P_T$, P_2,...,$P_{on}, P_T, P_n, P_{off}$. Although we further increase the number of patterns, the better tracking performance on the optimized tracking pattern P_T allows to acquire faster scene motions. From each successive pair of synchronously captured tracking camera images C_T we use an optical flow algorithm (Bouguet, J.-Y., 2000) to estimate the displacement fields $F_{i \to j}$ from the i-th acquired pattern C_T to the j-th pattern C_T as illustrated in Figure 2 without the on- and off-patterns for briefness. But for reconstruction we are interested in the motion within the acquired reconstruction sequence $C_1, C_2, \ldots, C_n$. Assuming that the motion from one tracking pattern to the next can be approximated as linear enables us to calculate intermediate displacement fields via vector field interpolation. Projector columns can finally be reconstructed via

$$j(x,y) = rec\left(C_1^s(x,y), C_2^s(F_{C_1^s\to C_2^s}(x,y)), \dots, C_n^s(F_{C_1^s\to C_n^s}(x,y))\right).$$

If the optical flow algorithm fails to estimate the motion of the tracking pattern safely due to depth discontinuities no surface point is reconstructed. Typically the order in which the different reconstruction patterns are acquired is not important, if it is possible to reorder them correctly afterwards. This allows a sliding window approach that reconstructs for any n successive coding patterns one 3d scan in the following manner:

$$\begin{aligned}
j_1(x,y) &= rec\left(C_1^s(x,y), C_2^s(F_{C_1^s\to C_2^s}(x,y)), \dots, C_n^s(F_{C_1^s\to C_n^s}(x,y))\right)\\
j_2(x,y) &= rec\left(C_1^{s+1}(F_{C_2^s\to C_1^{s+1}}(x,y)), C_2^s(x,y), \dots, C_n^s(F_{C_2^s\to C_n^s}(x,y))\right)\\
j_3(x,y) &= rec\left(C_1^{s+1}(F_{C_3^s\to C_1^{s+1}}(x,y)), C_2^{s+1}(F_{C_3^s\to C_2^{s+1}}(x,y)), C_3^s(x,y), \dots\right)\\
\vdots \quad &\vdots \quad \vdots
\end{aligned}$$

3.4 Designing a Tracking Pattern

Good tracking via optical flow can only be achieved at positions where changes in image intensity occur and only along the direction of change. Therefore, the intensity gradient field of the tracking pattern should contain many changes in length and direction. Especially corners are good features to track. To make the tracking robust against fast motions it is important to integrate dense features at different scales such that Perlin- or wavelet-noise patterns should be a good choice. To validate our considerations we compared four tracking patterns. Zooms of small regions of these patterns are shown in figure 3 a-d). All the patterns are pure black and white images as they can be most robustly separated with the on- and off-reference images and on the other hand because our high speed DLP-projector only allows projection of black and white images.

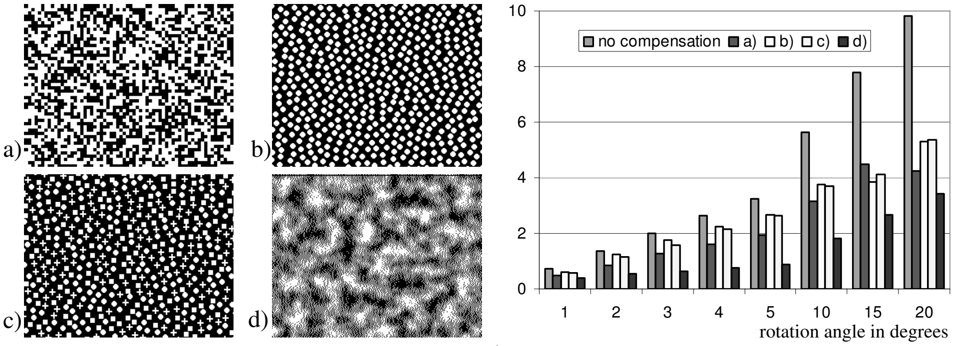

Figure 3 Different tracking patterns: a) binary noise on a coarse pixel grid, b) densely non-uniform distributed circles, c) multiple densely non-uniform distributed primitives, d) Binarized Perlin-noise using an error-diffusion dithering method. Right: Mean tracking error measured as distance between tracked and ground-truth pattern position in pixel.

We compared the tracking quality on a synthetic scene of a rotating bunny simulating the projector with a projective texture and the camera with a ray tracer that accounts for lambertian illumination, depth-of-field and shadows. The diagram on the right of Figure 3 shows the tracking error for different rotation angles between tracked images. We chose a rotation to produce many different kinds of pattern velocities and accelerations. Tracking errors are measured as follows: For each camera pixel in the first frame, we looked up the ground-truth pattern position and calculate the distances to the projector position in the second frame at the tracked position. Then we calculate the mean over all error. The errors in the row 'no compensation' are measured without tracking. We can see that all patterns are able to reduce the errors introduced by motion. The best results are achieved with the dithered Perlin-noise pattern. It is also remarkable that the error using motion compensation is increasing much slower than it is introduced by faster rotations.

4 Experimental Results

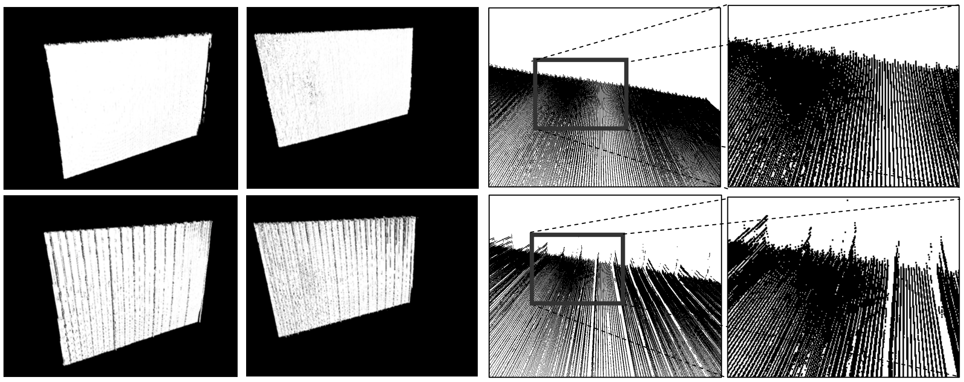

Figure 4 An acquired moving plane with (top row) and without (bottom row) motion compensation. Right: Zoomed views of the second reconstruction.

Based on the described ideas we built up a scanner system consisting of a high speed projector and a high speed camera. We integrate a 10 bit gray code method into our compensation framework. Because our optical flow estimator is not able to do its calculation in real-time we have to store the cameras images and perform offline calculations. This may be improved by using faster algorithms (Bruhn, A., Weickert, J., Feddern, C., Kohlberger, T. , and Schnörr, C., 2003), (Jose L. Martin and Aitzol Zuloaga and Carlos Cuadrado and Jesus Laizaro and Unai Bidarte, 2005). We do several scans on comparable moving scenes once measured with and once measured without using our motion compensation scheme. We use setups constructed with a small step motor for being able to reproduce scenes with equal motions. The first comparison depicted in figure 4 shows reconstructions of a moving plane. Camera frames are acquired at a rate of 90 fps. This leads to a reconstruction speed of 30 scans per second in the motion compensation scheme because every third pattern is a reconstruction pattern. Because the tracking pattern is not used in the version without compensation this sequence produces 60 scans per second. Figure 5 contains the result of scanning a rotating bust placed on a turntable. Like before we measure two equal movements one with and one without motion compensation. In both scans without motion compensation the effects of wrong decoded projector column positions can be recognized. Enabling

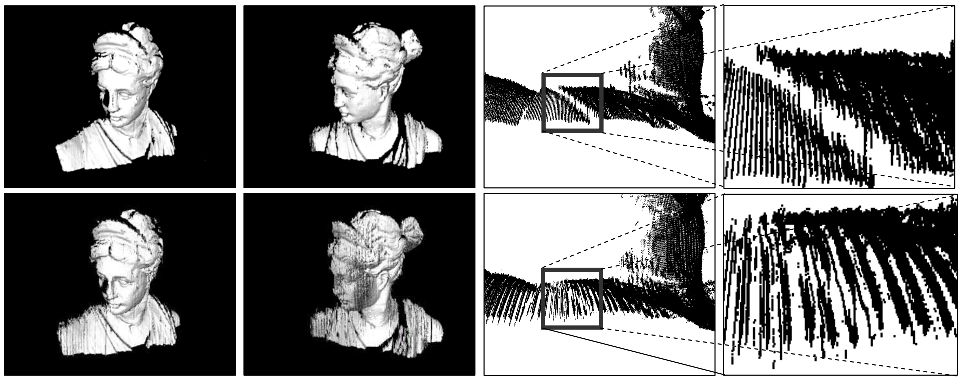

Figure 5 An acquired rotating bust with (top row) and without (bottom row) motion compensation. Right: Zoomed views of the second reconstruction.

motion compensation increases the quality of the reconstruction by avoiding or at least reducing the effects of motion. Although the acquisition of one complete sequence takes longer time, because of the additional tracking pattern, the results are significantly better than without compensation. We also noticed that scanning with motion compensation often increases the local sampling density. This can be explained by the fact, that our gray code reconstructor discards points if bits of the binary code can not be classified clearly as on or off. This normally happens on blurry stripe borders. When such borders start to move more code words are affected by these regions.

5 Conclusion and future work

In this work we presented a motion compensation technique, which makes it possible to use existing multi shot methods for the acquisition of dynamic scenes. This is done by correcting the changes in pattern positions estimated with the help of an interleaved tracking pattern. We demonstrate the benefits of our method by enabling the gray code method to scan moving objects at a rate of 30 reconstructions per second acquiring images at the rate of 90 fps. Future work will investigate the improvements which can be achieved by using more advanced optical flow algorithms to increase calculation speed and tracking accuracy and handling of flow discontinuities. It would be useful to integrate the proposed tracking into other sequences and also into one shot scanner to make their measurements more stable. Instead of interpolating the pattern flow fields, they can be projected and measured at the same time using different colours. This would reduce the number of needed patterns but may also introduce problems for textured objects.

References and Notes

1 Zhang, Z. (2000) 'A Flexible New Technique for Camera Calibration', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 11, pp. 1330–1334.

2 Rocchini, C., Cignoni, P.,Montani, C. ,Pingi, P. and Scopigno,R. (2001) 'A low cost 3D scanner based on structured light', *EG Proceedings*,Vol. 20, No. 3.,pp. 299–308

3 Mouaddib, E. , Batlle, J. and Salvia, J. (1997) 'Recent Progress in Structured Light in order to Solve the Correspondence Problem in Stereovision', *IEEE International Conference on Robotics and Automation Albuquerque*, Vol. 1, pp. 130–136.

4 Blais, F. (2004) 'Review of 20 Years of Range Sensor Development', *Journal of Electronic Imaging*, Vol. 13, No. 1, pp. 231–240.

5 Salvi, J., Pags, J. and Batlle, J. (2004) 'Pattern codification strategies in structured light systems', *Pattern Recognition*, Vol. 37, No. 4, pp. 827–849. Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms

6 Scharstein, D. and Szeliski, R. (2002) 'A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms', *International Journal of Computer Vision*, Vol. 47, No. 1-3, pp. 7–42.

7 Davis, J., Ramamoorthi, R. and Rusinkiewicz, S. (2003) 'Spacetime stereo: A unifying framework for depth from triangulation',*IEEE Conf. on Computer Vision and Pattern Recognition* pp.II: 359–366

8 Zhang, L., Curless, B. and Seitz, S. M. (2003) 'Spacetime Stereo: Shape Recovery for Dynamic Scenes', *Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. II–367-74.

9 Scharstein, D. and Szeliski, R. (2003) 'High-Accuracy Stereo Depth Maps Using Structured Light', *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 1, pp. 195–202.

10 Wust, C. and Capson, D. W. (1991) 'Surface profile measurement using fringe projection', *MVA*,4, pp.193–203

11 Zhang, L., Curless, B. and Seitz, S. M. (2002) 'Rapid Shape Acquisition Using Color Structured Light and Multi-pass Dynamic Programming', *Proceedings First International Symposium on 3D Data Processing Visualization and Transmission*, pp. 24–36.

12 Song Zhang, Peisen Huang (2004) 'High-resolution, Real-time 3D Shape Acquisition', *Proc. of Conf. on Comp. Vis. and Pat. Recog. Workshop (CVPRW)*, Vol. 3, pp.28.

13 Hall-Holt, O., Rusinkiewicz, S. (2001), 'Stripe Boundary Codes for Real-Time Structured-Light Range Scanning of Moving Objects', ICCV 2001, pp. 359–366

14 Carrihill, B., Hummel, R. (1985), 'Experiments with the intensity ratio depth sensor', *Comp. Vis., Graphics and Image Processing*, Vol. 32, Academic Press, pp. 337-358.

15 Maruyama, M., Abe, S. (1989) 'Range Sensing by Projecting Multi-Slits with Random Cuts', *Int. Workshop on Ind. Appl. of Mach. Int. and Vis. (MIV)*, pp. 163–168.

16 Vuylsteke, P. and Oosterlinck, A. (1990) 'Range Image Acquisition with a Single Binary-Encoded Light Pattern', *IEEE Transactions on Pattern Analysis ans Machine Intelligence*, Vol. 12, No. 2, pp. 148–164.

17 Koninckx, T. P., Griesser, A. and Van Gool, L. (2003) 'Real-time Range Scanning of Deformable Surfaces by Adaptively Coded Structured Light', *Proceedings Fourth Int. Conf. on 3-D Digital Imaging and Modeling, 3DIM*, pp. 293–300.

18 Adan, Antonio, Molina, Fernando, Vazquez, Andres S. and Luis Morena (2005) '3D Feature Tracking Using a Dynamic Structured Light System', *Proceedings of the 2nd Canadian conference on Computer and Robot Vision*, pp. 168–175.

19 Weise, T., Leibe, B. and Van Gool, L. (2007) 'Fast 3D Scanning with Automatic Motion Compensation', to appear in *IEEE Conference on Comp. Vis. and Pattern Recognition (CVPR).*

20 Bouguet, J.-Y. (2000) ‘Pyramidal Implementation of the Lucas Kanade Feature Tracker’, Intel Corp., Microprocessor Research Labs

21 Bruhn, A., Weickert, J., Feddern, C., Kohlberger, T. , and Schnörr, C. (2003) ‘Real-time optical flow with variational method’, *CAIP*, LNCS 2756, pp. 222–229

22 Martin, J. L., Zuloaga, A., Cuadrado, C., Laizaro, J. and Bidarte, U. (2005) ‘Hardware implementation of optical flow constraint equation using FPGAs’,*Comput. Vis. Image Underst.*, Vol. 98, No. 3, pp. 462–490