Correction of Motion Artifacts for Real-Time Structured Light

Jakob Wilm[1,2(✉)], Oline V. Olesen[1,2], Rasmus R. Paulsen[1], and Rasmus Larsen[1]

[1] Department of Applied Mathematics and Computer Science, Technical University of Denmark, Richard Petersens Plads, Building 324, DK-2800 Kgs. Lyngby, Denmark
jakw@dtu.dk
http://compute.dtu.dk/

[2] Department of Clinical Physiology, Nuclear Medicine and PET, Rigshospitalet, Copenhagen University Hospital, University of Copenhagen, Copenhagen, Denmark

Abstract. While the problem of motion is often mentioned in conjunction with structured light imaging, few solutions have thus far been proposed. A method is demonstrated to correct for object or camera motion during structured light 3D scene acquisition. The method is based on the combination of a suitable pattern strategy with fast phase correlation image registration. The effectiveness of this approach is demonstrated on motion corrupted data of a real-time structured light system, and it is shown that it improves the quality of surface reconstructions visually and quantitively.

Keywords: 3D vision · Structured light · Motion correction · Registration

1 Introduction

Structured light techniques are very popular for 3D scene capture, and a large variety of variants have been proposed. See [5] for an overview. The general idea is to aid in the detection of stereo correspondences by active projection of light onto the scene, followed by triangulation, to yield a surface reconstruction.

In time-multiplexed or multi-pattern structured light [13], a sequence of patterns is projected to generate one surface reconstruction, which encodes scene points, thereby disambiguating the matching process. The flexible nature of of structured light allows to choose a trade-off between number of patterns and accuracy, and to optimise for the radiometric properties of the scene.

With advances in light projection technology and computing power, it is today possible to perform structured light surface scanning with multiple patterns in real time. This enables accurate object tracking, dynamic deformation studies, fast object digitisation and many other applications. A general limitation with any multi-pattern method however is the underlying assumption of no motion between the patterns of a single sequence. A violation of this assumption

R.R. Paulsen and K.S. Pedersen (Eds.): SCIA 2015, LNCS 9127, pp. 142–151, 2015.
DOI: 10.1007/978-3-319-19665-7_12

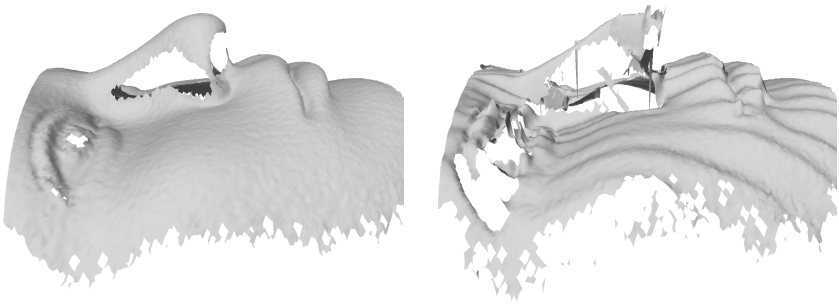

Fig. 1. Reconstructions of a phantom head using our structured light setup, showing the effects of motion during time-multiplexing. Left: static scene. Right: slight movement during acquisition.

often leads to large artifacts in the reconstructed surface. Such artifacts are shown as an example in figure 1.

By abandoning time multiplexing, single-shot structured light can be realised [14][6]. In the case of Microsoft Kinect 1, a static pseudorandom pattern is used. With single-shot techniques, the motion problem is avoided, but the reconstruction quality is also not very high or lateral resolution is lowered. Multiplexing by means of wavelength is also possible, i.e. by placing different patterns in the red, green and blue channels of the projection. This is not as robust to surface reflectance properties and spectral bleeding/crosstalk may occur. A review of single-shoot methods is provided in [20].

Time multiplexing remains the most robust technique, and is used in many commercial products, mainly for industrial inspection and reverse engineering. The chosen pattern strategy is important for the performance of the system, and two classes of methods remain very popular: binary Gray-coding and phase shifting profilometry (PSP). The former tends to be robust, but generally requires relatively many patterns, worsening the effects of motion. PSP can encode the scene unambiguously with only three patterns of a single sinusoid phase, shifted by 0 deg, 120 deg and 240 deg respectively. Depth resolution can be improved by using more shifts, or multiple phases, which gives a limited ambiguity that can be resolved using a phase-unwrapping algorithm.

In the presented method, we use the modified "2+1" phase shifting method of Zhang and Yau [19], which according to the authors reduces the effects of motion, and perform fast image registration on the acquired camera frames, to correct for the synchronisation error, and vastly improve the quality of scene reconstructions. By employing phase-based image registration, the motion correction is efficient enough to run in real-time in our 20 Hz surface scanning pipeline.

2 Previous Work

Employing more than three patterns in phase shifting profilometry allows for sanity checking, and masking corrupted output [9]. This is at the expense of

additional patterns in the sequence, generally making it slower (but more accurate), but it is questionable to what degree the unmasked output can be trusted.

In [7], a motion corrected real-time method is presented, for one-dimensional known motion.

The consequences of motion can be alleviated to some extend by running at high frequencies. Zhang et al. have presented structured light with dithered binary sinusoid patterns running at 1000 Hz reconstructions per second [15]. Their method requires strong projector lighting and high-speed cameras due to the short integration time of any single camera frame, and involves quality trade-offs making it unsuitable for many applications.

Liu et al. [10] propose motion corrected light with binary patterns and estimation of a global velocity vector based on the reconstructions, while Lu et al. show the theoretical feasibility of motion correction by means of image alignment [11].

3 Structured Light and PSP

We use direct codification structured light, i.e. using a single camera-projector pair as described in [16]. Projector pixel coordinates are denoted (u_p, v_p), while camera pixel coordinates are (u_c, v_c). Employing a vertical baseline, the standard 3-step PSP algorithm encodes the projected images as

$$I_n^p(u_p, v_p) = \frac{1}{2} + \frac{1}{2}\cos\left(2\pi\left(\frac{u_p}{N_p} - \frac{n}{3}\right)\right),$$

where n indicates the pattern index, $n \in 1 \ldots 3$ and N_p the number of projector columns. The camera captures the n'th pattern as

$$I_n^c(u_c, v_c) = A^c + B^c\cos\left(2\pi\left(\frac{u_p}{N_p} - \frac{n}{3}\right)\right).$$

Note the dependency on the projectors horizontal coordinate, u_p. A^c is the scene intensity including ambient contributions, while B^c is the modulation by projector light. A^c can be considered the magnitude of the Fourier DC component, while B^c is the magnitude at the principle frequency, and $\frac{u_p}{N_p}$ its normalised phase. In order to create correspondences, $\frac{u_p}{N_p}$ is extracted with the Fourier transform and scaled by N_p.

Our camera and projector pair are calibrated according to the method described in [18]. Corresponding sets of u_p, u_c, v_c-coordinates are used to triangulate object surface points as shown in figure 2. We denote one collection of projected patterns a pattern sequence, and the corresponding camera frames a frame set.

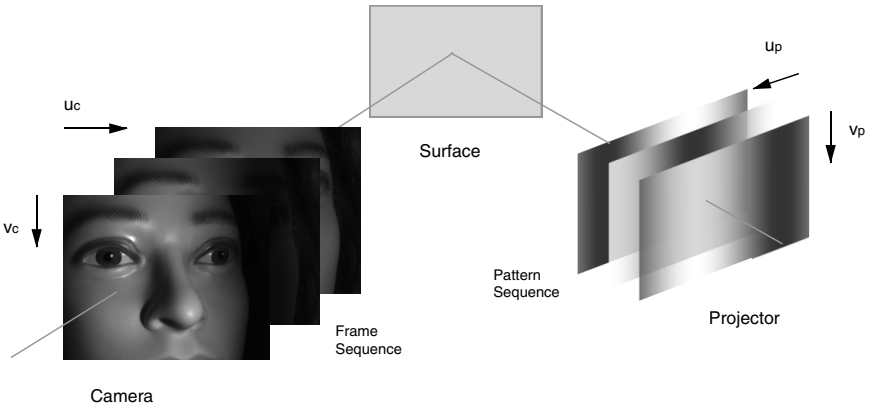

Fig. 2. Principle of structured light 3d scanning shown with the 3-step PSP pattern strategy. For each image point, the corresponding projector column coordinate, u_p, is extracted, which allows for triangulation of points on the object surface.

The "2+1" phase shifting method is a slight modification of the 3-step method, in which one of the patterns is a constant lit image:

$$I_1^p(u_p, v_p) = \frac{1}{2} + \frac{1}{2}\cos\left(2\pi\frac{u_p}{N_p}\right) \tag{1}$$

$$I_2^p(u_p, v_p) = \frac{1}{2} + \frac{1}{2}\cos\left(2\pi\frac{u_p}{N_p} - \frac{\pi}{2}\right) \tag{2}$$

$$I_3^p(u_p, v_p) = \frac{1}{2} \tag{3}$$

From the corresponding camera frame set, the horizontal projector coordinate can be extracted from camera frames as

$$u_p = \frac{N_p}{2\pi}\,\mathrm{atan2}\left(I_2^c - I_3^c, I_1^c - I_3^c\right),$$

while intensity information is extracted simply as $A^c = I_3^c$.

The availability of these direct intensity frames, which are not corrupted by inter-frame motion, allows us to estimate the motion, and perform a correction. Specifically, we register intensity frames between frame sets, to estimate in 2D a global motion trend, $\boldsymbol{s} = [\Delta x, \Delta y]$. All frames of a set are then corrected by interpolating from the motion trend. This removes most of motion induced artifacts from single object sources that undergo rigid motion.

4 Phase Correlation

In order estimate inter-frame motion, intensity frames are registered to each other. While the image deformation is most accurately described by a dense warp

field, any estimation method for such would be computationally prohibitive in a real-time context. We therefore utilise phase correlation, also called the Fourier-Mellin method, which can estimate a rigid transformation very quickly. It is a long well known method that is very noise-robust and particularly efficient [8][3], lending itself to real-time processing such as video stabilisation [4]. We use it here to register images of two successive frame sets, in order to correct for the misalignment of frames and the resulting artifacts in the 3D reconstruction.

In many applications, such as pose tracking, the tracked object is considered to undergo rigid 6 DOF motion between frames. In such scenarios, fast phase correlation registration can account for most of the misalignment, and reduce artifacts considerably.

With two images, f_1, f_2, of dimensions $N_x \times N_y$ the phase correlation technique first applies a spatial discrete Fourier transform to both images. The determination of translational shift is based on the Fourier shift theorem.

Assuming only a circular translational shift, the images are given as

$$f_2(x, y) = f_1(\mathrm{mod}(x - \Delta x,\, N_x), \mathrm{mod}(y - \Delta y,\, N_y)),$$

and their respective Fourier spectra are related by

$$\mathcal{F}_2(u, v) = \mathrm{e}^{-2\pi i(u\Delta x/N_x + v\Delta y/N_y)}\mathcal{F}_1(u, v),$$

with (u, v) denoting frequency components.

The normalised cross-power spectrum is given by

$$P = \frac{\mathcal{F}_1 \odot \mathcal{F}_2^*}{\|\mathcal{F}_1 \odot \mathcal{F}_2^*\|} = \mathrm{e}^{2\pi i(u\Delta x/N_x + v\Delta y/N_y)},$$

with $\mathcal{F}^*$ denoting the complex conjugate, and $\odot$ the elementwise/Hadamard product. The cross correlation of f_1 and f_2 is now calculated as the inverse Fourier transform of P, and the shift estimated as its peak position

$$\boldsymbol{s} = \arg\max_{x,y} \mathcal{F}^{-1}(P).$$

Since the discrete Fourier shift theorem holds only for circular shifts, a window function is used on the Fourier transforms. In fact, this is usually necessary, as the edges on tiled input images provide strong high-frequency features in Fourier domain.

To obtain the peak position with sub-pixel precision, the centroid of the cross-correlation peak is computed by also including values from a small neighbourhood around $\boldsymbol{s}$.

It is also possible to recover scale and rotation parameters between images in an analogue way by converting magnitude spectra to the log-polar domain before computing the cross-power spectrum [3].

5 Experimental Setup

Our structured light setup consists of a single camera-projector pair. The camera integration time is a multiple of the projector refresh period for truthful

gray-value reproduction, and a hardware trigger signal ensures accurate synchronisation of camera and projector. The projector update frequency is 120 Hz, and the camera integration time 8.333 ms. Due to trigger latency, we are able to capture every other projector image, resulting in $60\,\mathrm{s}^{-1}$ camera frames and 20 Hz surface reconstructions per second using the 2+1 phase shifting method.

With two consecutive frame sets $I_t^c = I_{1,t}^c, I_{2,t}^c, I_{3,t}^c$ and $I_{t+1}^c = I_{1,t+1}^c$, $I_{2,t+1}^c, I_{3,t+1}^c$ at times t and $t+1$ respectively, we perform phase correlation based registration between the flat intensity frames $I_{3,t+1}^c$ and $I_{3,t}^c$. This shift, $\boldsymbol{s}$, serves as the global motion estimate at that time point. Before reconstructing a surface from frame set I_{t+1}^c, its frames are corrected to:

$$I_{1,t}^c\left(\boldsymbol{x} - \frac{2}{3}\boldsymbol{s}\right) \quad I_{2,t}^c\left(\boldsymbol{x} - \frac{1}{3}\boldsymbol{s}\right) \quad I_{3,t}^c\left(\boldsymbol{x}\right)$$

As noted above, phase correlation in the log-polar domain allows for determination of global scale and rotation. However, in most cases of small-scale misalignment caused by camera or object movement, a translation registration is adequate and preferred by us due to the lower computational demands.

We carry out experiments by scanning a phantom head that is moving in a controlled manner. It is mounted on a stepper motor that rotates back and forth on a 36 deg arc at a constant speed of $11,4\,\mathrm{deg/s}$ at a distance of approximately 20 cm to the camera. In our camera, this results in per-coordinate pixel shifts of up to 5 px between to consecutive frames. This scenario is modelled after our head-tracking approach used for medical imaging motion correction [12].

To quantify the quality of our motion corrected structured light approach, we perform a tracking experiment on the phantom head. The head is held stationary to capture an uncorrupted reference scene. The motor is then started, and the object surface is scanned and registered to the reference by means of the iterative closest point algorithm. We use the point-to-plane error metric [2], as it generally converges best on this kind of data [17]. Correspondences are found by means of back-projecting data into the camera frame [1]. The so-obtained tracking data is used to evaluate the quality of the reconstructed object surface by means of the root-mean-square (RMS) error of the alignment to the reference surface. Partial overlap of surfaces is handled by removing those point correspondences that match to the border of the reference point cloud.

6 Results

The alignment of image frames I_1^c and I_3^c during the motor controlled motion scene is shown in figure 3. It is seen, that the global shift was estimated correctly (see e.g. edge of iris, corners of the eye), and a large amount of misalignment between the images is accounted for.

The effect of motion correction on the resulting surface reconstruction is shown in figure 4.

The correction shows reduction of artifacts and distortion, especially around the nose area.

Fig. 3. Registration of camera frames I_1^c (green colour channel) and I_3^c (magenta colour channel) belonging to the same frame sequence. Left: before registration. Right: after registration. The misalignment is most easily seen at the edge of the iris and corners of the eye.

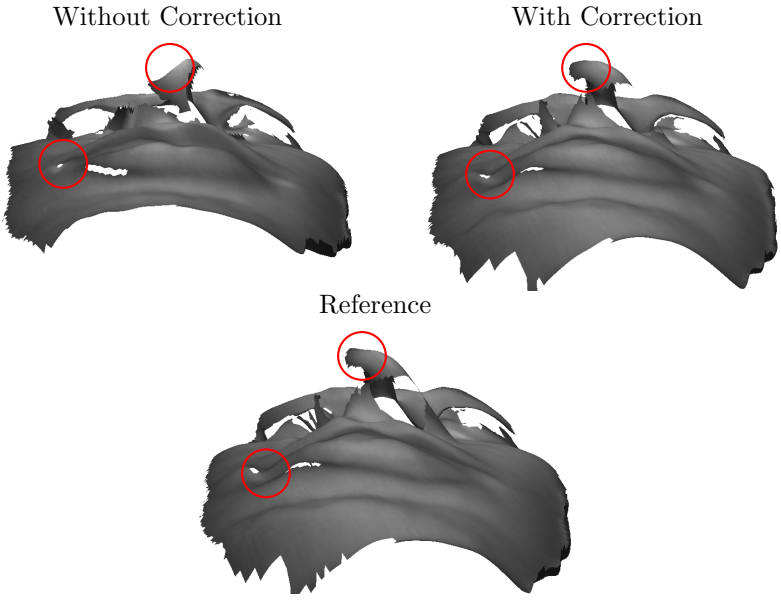

Fig. 4. Results of performing phase correlation based motion correction on an object moving at constant speed. Two aliased regions are highlighted in red circles. Top left: without correction. Top right: with correction. Bottom: stationary reference scan, e.g. no motion.

Using the rotating phantom head described, we use phase correlation to estimate the translational shift between intensity frames over time. This is shown in figure 5 with the raw shifts, and their cumulative sums. From these plots,

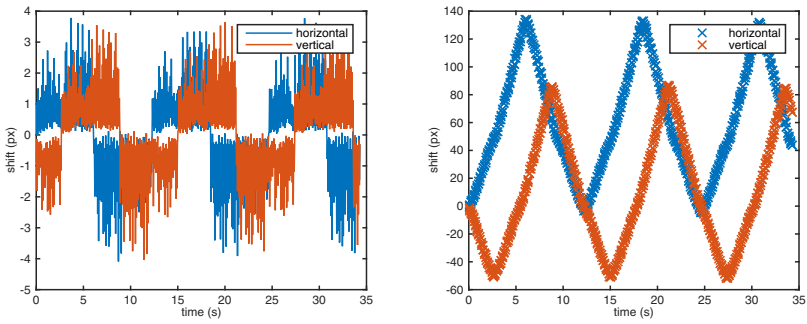

Fig. 5. Left: Translational shift estimates in horizontal (blue) and vertical (red) directions. Right: The cumulative sum of shift estimates.

it becomes apparent that phase correlation faithfully and reproducibly captures the global object movement over time.

RMS errors for the moving scene are shown in figure 6.

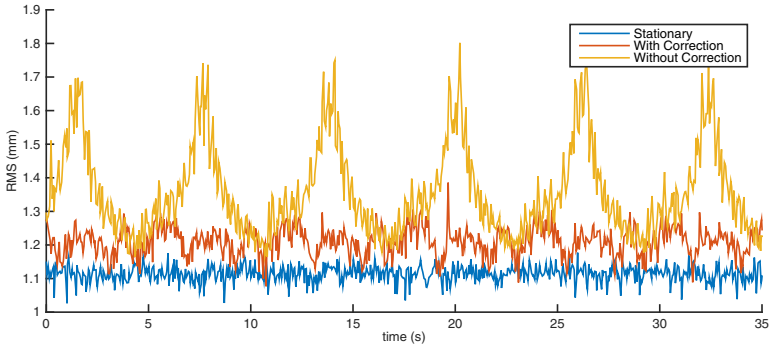

Fig. 6. RMS error of surface alignment to stationary scene at $t = 0$ using ICP. Variation in the stationary case is due to image noise. It is seen that motion corrected frame sets result in markedly lower RMS. In the uncorrected case, RMS peaks when the object is farthest from stationary situation. Mean RMS over time: stationary (1.12 mm), with correction (1.20 mm), without correction (1.34 mm).

The static scene trace is influenced solely by image noise in the camera frames. With corrected motion, the error increases by approximately 7% (time-averaged RMS), while in the uncorrected case, the increase in error is over 19%. It is noted that the increase in RMS is mainly due to large artifacts in high-frequency regions of the camera frame. It is seen that our motion correction approach effectively decreases these artifacts, and lowers the RMS alignment error.

The average processing time for the alignment of two camera intensity frames ($300 \times 200\,\mathrm{px}$) was around 5 ms.

7 Conclusion and Discussion

We have proposed a fast alignment strategy to reduce motion artifacts in time-multiplexed structured light. While the method is limited to recovering a global translational shift, this was shown to explain a large amount of the misalignment, and hence reduce artifacts in the resulting surface reconstructions. The used method was used for its robustness and computational efficiency. With its fixed-time performance, the phase correlation technique is suitable for real-time processing, as done in our implementation.

The limitation of this method is that it only accurately explains single object translational movement. With offline-processing, a large variety of video stabilisation techniques can be combined with the 2+1 phase shifting method. These include, but are not limited to subframe correlation methods, feature-based parametric image registration and optical flow based methods.

For many situations in which structured light is used for tracking or object digitisation, our method promises to significantly improve reconstruction results.

References

1. Blais, G., Levine, M.D.: Registering multiview range data to create 3D computer objects. IEEE PAMI **17**(8), 820–824 (1995)
2. Chen, Y., Medioni, G.: Object Modeling by Registration of Multiple Range Images. Image and vision computing (1992)
3. De Castro, E., Morandi, C.: Registration of translated and rotated images using finite fourier transforms. IEEE PAMI **9**(5), 700–703 (1987)
4. Erturk, S.: Digital image stabilization with sub-image phase correlation based global motion estimation. IEEE Transactions on Consumer Electronics **49** (2003)
5. Geng, J.: Structured-light 3D surface imaging: a tutorial. Advances in Optics and Photonics **160**(2), 128–160 (2011)
6. Guan, C., Hassebrook, L., Lau, D.: Composite structured light pattern for three-dimensional video. Optics express **11**(5), 406–417 (2003)
7. Ishii, I., Koike, T., Takaki, T.: Fast 3D shape measurement using structured light projection for a one-directionally moving object. In: IECON 2011, pp. 135–140 (2011)
8. Kuglin, C.D., Hines, D.C.: The phase correlation image alignment method. In: IEEE International Conference on Cybernetics and Society, pp. 163–165 (1975)
9. Lau, D., Liu, K., Hassebrook, L.G.: Real-time three-dimensional shape measurement of moving objects without edge errors by time-synchronized structured illumination. Optics letters **35**(14), 2487–2489 (2010)
10. Liu, Y., Gao, H., Gu, Q., Aoyama, T., Takaki, T., Ishii, I.: A fast 3-d shape measurement method for moving object. In: Int. Conf. Progress in Informatics and Computing (PIC), pp. 219–223 (2014)
11. Lu, L., Xi, J., Yu, Y., Guo, Q.: New approach to improve the accuracy of 3-D shape measurement of moving object using phase shifting profilometry. Optics Express **21**(25), 30610–30622 (2013)

12. Olesen, O.V., Paulsen, R.R., Hjgaard, L., Roed, B., Larsen, R.: Motion tracking for medical imaging: a nonvisible structured light tracking approach. IEEE Transactions on Medical Imaging **31**(1), 79–87 (2012)
13. Posdamer, J.L., Altschuler, M.D.: Surface measurement by space-encoded projected beam systems. Computer Graphics and Image Processing **18**(1), 1–17 (1982)
14. Sakashita, K., Yagi, Y., Sagawa, R., Furukawaa, R., Kawasaki, H.: A system for capturing textured 3D shapes based on one-shot grid pattern with multi-band camera and infrared projector. In: Proc. 3DIMPVT, pp. 49–56 (2011)
15. Wang, Y., Laughner, J.I., Efimov, I.R., Zhang, S.: 3D absolute shape measurement of live rabbit hearts with a superfast two-frequency phase-shifting technique. Optics express **21**(5), 6631–6636 (2013)
16. Wilm, J., Olesen, O.V., Larsen, R.: SLStudio : open-source framework for real-time structured light. In: Proc IPTA 2014, pp. 1–4. IEEE Xplore (2014)
17. Wilm, J., Olesen, O.V., Paulsen, R.R., Højgaard, L., Roed, B., Larsen, R.: Real time surface registration for PET motion tracking. In: Heyden, A., Kahl, F. (eds.) SCIA 2011. LNCS, vol. 6688, pp. 166–175. Springer, Heidelberg (2011)
18. Wilm, J., Olesen, O.V., Larsen, R.: Accurate and Simple Calibration of DLP Projector Systems. SPIE Photonics West 2014, 897–909 (2014)
19. Zhang, S., Yau, S.T.: High-speed three-dimensional shape measurement system using a modified two-plus-one phase-shifting algorithm. Optical Engineering **46**, 1–6 (2007)
20. Zhang, Z.H.: Review of single-shot 3D shape measurement by phase calculation-based fringe projection techniques. Optics and Lasers in Engineering **50**(8), 1097–1106 (2012)