

# Motion Compensation for Structured Light Sensors

Debjani Biswas and Christoph Mertz

Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA, 15289, USA

## ABSTRACT

In order for structured light methods to work outside, the strong background from the sun needs to be suppressed. This can be done with bandpass filters, fast shutters, and background subtraction. In general this last method necessitates the sensor system to be stationary during data taking. The contribution of this paper is a method to compensate for the motion if the system is moving. The key idea is to use video stabilization techniques that work even if the illuminator is switched on and off from one frame to another. We used OpenCV functions and modules to implement a robust and efficient method. We evaluated it under various conditions and tested it on a moving robot outdoors. We will demonstrate that one can not only do 3D reconstruction under strong ambient light, but that it is also possible to observe optical properties of the objects in the environment.

**Keywords:** videostabilization, background subtraction, 3D reconstruction

## 1. INTRODUCTION

Structured light sensing is a common method for short range 3D measurements. It is employed in many industrial and consumer applications (e.g. Kinect). One constraint when these sensors are used around people is that the amount of light they can project is limited by eye-safety concerns. This is particularly relevant in outdoor environments where the ambient sunlight can be overwhelming. In our previous work<sup>1</sup> we describe a low-cost and low-power outdoor scene reconstruction system using the MEMS mirror in a commercially available pico-projector. We use three methods to suppress ambient illumination from the sun: laser as illuminator with a bandpass filter, laser pulsing and synchronizing it with a fast shutter of the camera, and background subtraction. For the last method the sensor is used to produce a thin line-stripe in alternate images when recording a series of subsequent frames in a video. The background is suppressed by subtracting a frame without the line-stripe from the preceding frame with the line-stripe to only retrieve the line-stripe. However, this approach works outside in bright light only while using a static sensor. Our system described in this paper extends it to situations when using a moving sensor and straightforward background subtraction is not sufficient to suppress the background. In order to effectively accomplish background subtraction between two consecutive frames recorded from a moving sensor, we need to estimate and compensate for the movement between these consecutive frames. It is possible to use additional sensors and IMU to measure this motion between frames — but that is an additional expense and may not necessarily be precise. To this effect, we take advantage of the video itself and we employ the procedure of videostabilization for motion estimation and motion compensation between frames.

## 2. BACKGROUND

Videostabilization is the process of generating a new sequence of frames from a video where undesirable image motion between frames caused from handshakes or jitters has been removed. There are usually 3 major stages that constitute a videostabilization process: camera motion estimation, motion smoothing and image warping. We first focus on estimating the inter-frame motion between adjacent frames. The accuracy of this estimation is important to our approach because we want to remove the background as best as possible, and after warping and subtraction, only get the thin line-strip from the laser source in our final images. For global motion estimation there are two major approaches. One is direct pixel based approaches that utilize each pixel in the whole frame to align one frame to another<sup>2</sup>. The displacement of each pixel in the image is determined by the brightness constancy constraint and the “global motion model” that provides either a 2-dimensional or 3-dimensional model to portray the image motion throughout the entire image. The other is feature based methods which usually compute a set of point correspondences between adjacent images and from these, estimate the parameters that describe the global motion<sup>3</sup>. Feature based methods are generally faster but more prone to local effects.

We first attempted a direct pixel based method that uses each pixel in each frame to model the global motion between subsequent frames,  $f(x, y, t)$  and  $f(x, y, t - 1)$ , as a 6 parameter affine transform <sup>4</sup>. These parameters are estimated by setting the derivatives of the inter-frame error to zero. To deal with larger displacements a L-level Gaussian image pyramid based hierarchical technique is applied.

Although this technique can successfully estimate the transformation between successive frame in many cases, it has difficulties making a motion estimation when there are objects in the scene at different distances. It is also very sensitive to illumination changes which is a problem outdoors where ambient light changes are common. Most importantly, we switch a laser illumination on and off between consecutive frames and the method can not account for that.

In order to counter these effects, we have investigated feature based methods for motion estimation such as the one OpenCV employs<sup>5-7</sup>. The transformation matrix  $M$  represents the motion between consecutive frames  $I_0(x)$  and  $I_1(x')$ .

$$M = \begin{bmatrix} m_0 & m_1 & m_2 \\ m_3 & m_4 & m_5 \\ m_6 & m_7 & m_8 \end{bmatrix} \quad (2.1)$$

An image is warped using  $\mathbf{x}' \sim M\mathbf{x}$  where  $\mathbf{x} = (x, y, 1)$  and  $\mathbf{x}' = (x', y', 1)$  are pixel locations in homogeneous co-ordinates and  $\sim$  denotes equality up to a scale. We assume an affine model, i.e.  $m_6 = m_7 = 0$  and  $m_8 = 1$ .

To recover the parameters in equation 2.1,  $M$  is iteratively updated using gradient descent after re-sampling frame  $I_1$  also bringing it in closer registration with frame  $I_0$  <sup>6</sup>. A hierarchical motion estimation framework is used, where an image pyramid is first constructed to reduce the area of search, starting at the coarsest level<sup>8</sup>. A global transformation chain is thus obtained by applying this parameter estimation for every pair of adjacent frames<sup>5</sup>. If the scene is non-planar, the motion is estimated by computing the optical flow between frames after applying the above method of global transformation. This is done by obtaining a local motion field after applying a pyramidal version of Lucas-Kanade optical flow computation<sup>9,10</sup>.

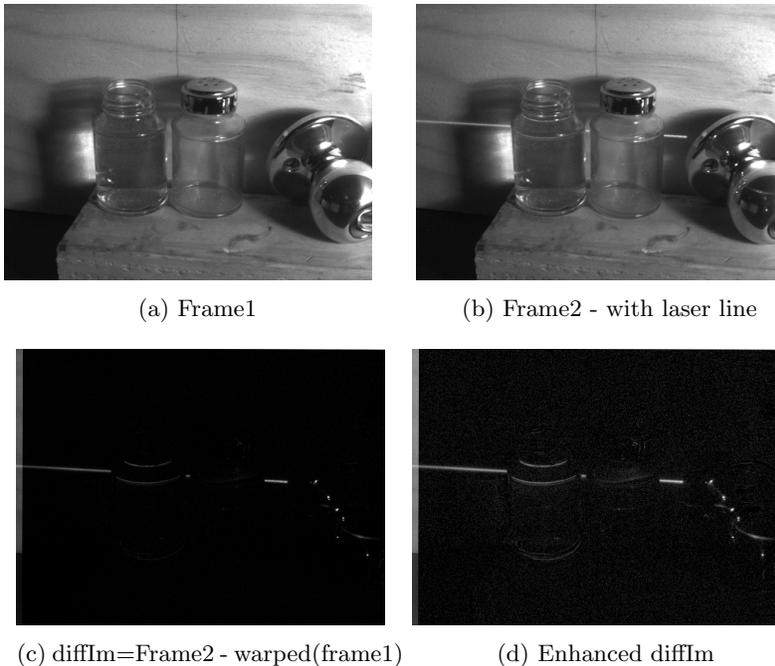


Figure 1: 1a, 1b: The 2 consecutive frames with slightly different view points due to the motion of the sensor. Figures 1c, 1d: Result of background subtraction after motion estimation and warping of Frame1. 1d is an enhanced image, it shows that the background is reduced to a small noise level.

### 3. METHOD

We use the videostabilization module of OpenCV to estimate the affine transformation between 2 frames. First, motion is estimated from the source frame (**frame1**) to the destination frame (**frame2**). Then **frame1** is warped to **frame2** using this estimated affine transformation. After the motion has been compensated between these two frames, the warped source frame (**warped(frame1)**) is subtracted from the original destination frame (**frame2**) to achieve a background subtraction and the line-stripe projected from the laser becomes apparent. Figure 1 illustrates this process, where only the difference, the projected laser line, and a small amount of noise remains. For these test images, the scene has been illuminated with a strong lamp and a weak laser light so that the signal and background are about equal.

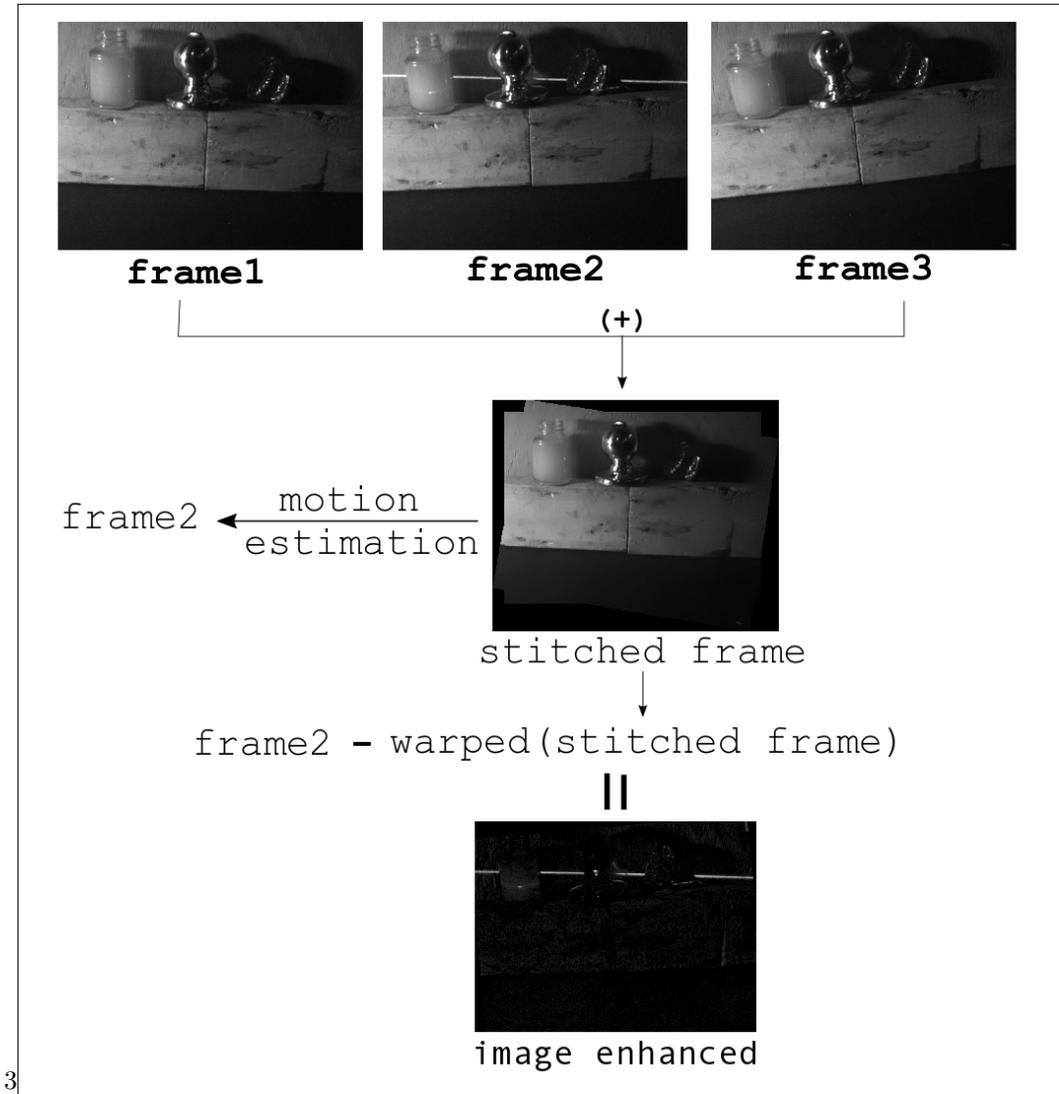


Figure 2: Result of side border removal to perform complete background subtraction using stitching before motion estimation, warping and subtraction.

Although this approach works well, our results from the above method can be improved. For example, in Figure 1 one can notice the presence of a side border in the final subtracted image. This is because of the slight difference in the field of view between the 2 subsequent frames. Since we would like to accomplish background subtraction for the entire image we use 3 consecutive frames. The three images used are, the image with the

laser line (**frame2**), the preceding (**frame1**) and following (**frame3**) images without the line. First, we stitch the first and the thirds to obtain a stitched frame (**frame1 + frame3 = stitchedframe**). Next we adjust the size of resulting stitched frame to the size of the image (**frame2**) from which the motion model is going to be estimated. From then on we treat the two frames as before: we estimate the affine transformation from the stitched frame to the original frame (**frame2**) and using this estimated motion we warp the stitched frame to **frame2**. Finally, we subtract the stitched image from **frame2** to get the result with only the line stripe.

Figure 2 illustrates one of the experiments analyzed with this approach. The images show that our technique was able to take out the border in the resulting background subtracted images.

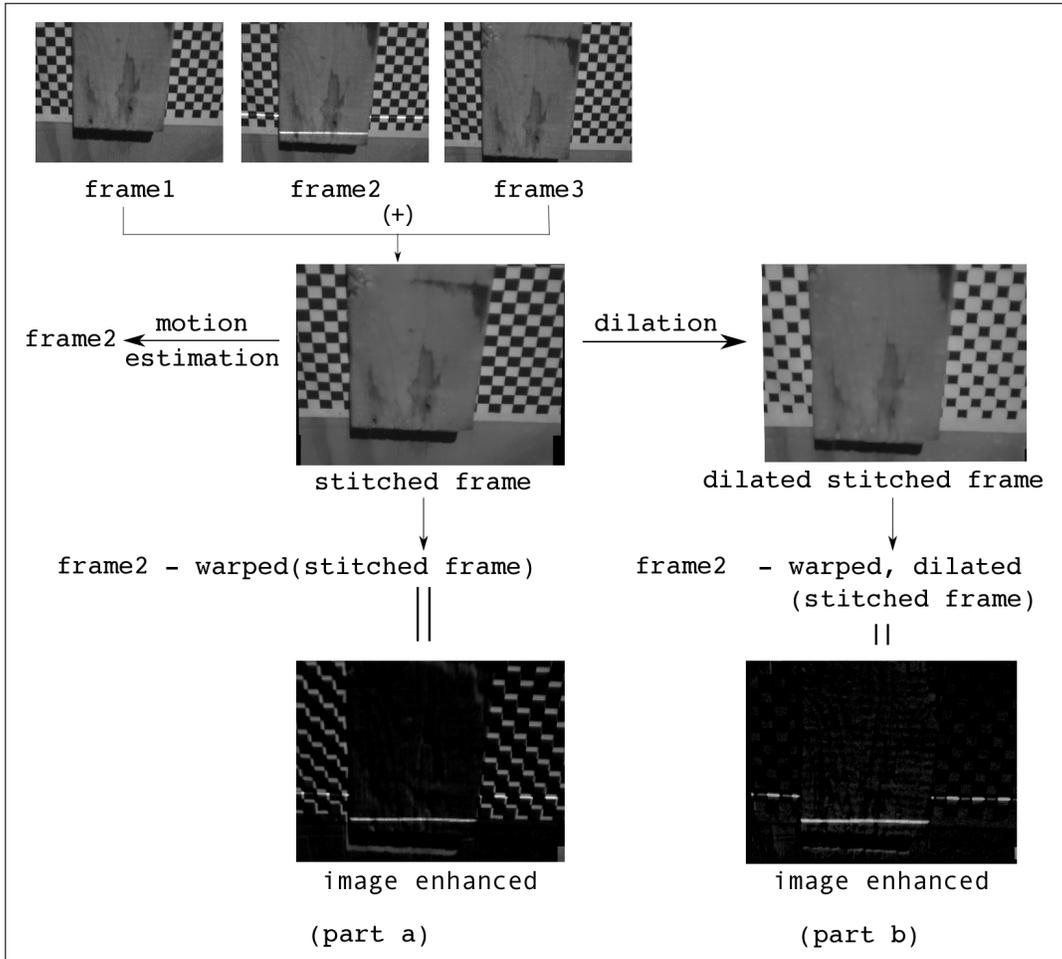


Figure 3: Effect of applying Dilation. Left (part a): Without applying dilation, the resulting subtracted image has some residuals. Right (part b): After dilation is applied to the warped image and then subtracted, we achieve an improved background subtracted result. The data was collected in bright sunlight, the setup is shown in (Figure 4 left).

Modeling the motion as an affine transformation is only strictly correct when all the objects are in one plane. However, sometimes objects in the scene are at significantly different distances and there are still residuals after the background subtraction. (Part a) of Figure 3 shows an example where this effect is especially pronounced at sharp intensity edges. In order to compensate for this and for any other kinds of errors or noise in the motion estimation, we apply a sliding max filter to the warped image before subtraction. We typically set each pixel in the warped image to the maximum value of its  $3 \times 3$  pixel neighborhood. This way we bias the background subtraction to subtract more than the background itself. The trade off is the possibility that we may not be able

to see the laser line in some parts of the resulting subtracted image. But this will only happen if the laser line projects on a dark part (e.g. black square on the checkerboard) when there is bright background nearby (e.g. white square of checker board). (Part b) of Figure 3 illustrates the effect of applying dilation to the warped image before subtraction. We get a much better background subtraction using this approach. There are no more residuals from objects at different distances or from sharp intensity edges.

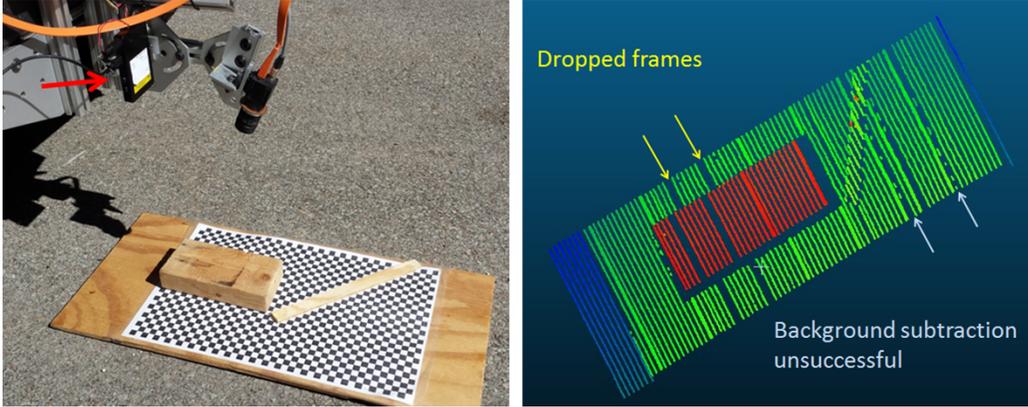


Figure 4: Left: Laser pico-projector (red arrow) and camera mounted on the front of a vehicle. Right: 3D scan of the objects. Color indicates height. Two places are shown where frames were dropped and two places are shown where the background subtraction was unsuccessful.

#### 4. EXPERIMENTS

As we would like to extend our previous work<sup>1</sup> from a stationary sensor to a moving sensor platform, an important test is if this method works outside in bright sunlight. We use a similar setup: A pico-projector (Figure 4 left) produces the laser line. A moving camera captures the frames in the scene and has a bandpass filter and a fast shutter to suppress ambient light. Figure 3 shows images taken by this moving camera in bright sunlight and the results after background subtraction. The data was collected and analyzed live on a MacMini. The camera frame rate was 30Hz, but some frames were dropped. There are still possibilities to speed up the system and the full 30Hz update rate should be achievable. Figure 4 shows the 3D reconstruction of the scene. One can see that some scan lines are missing points. This is because some motion estimations weren't quite accurate leading to unsuccessful background subtractions.

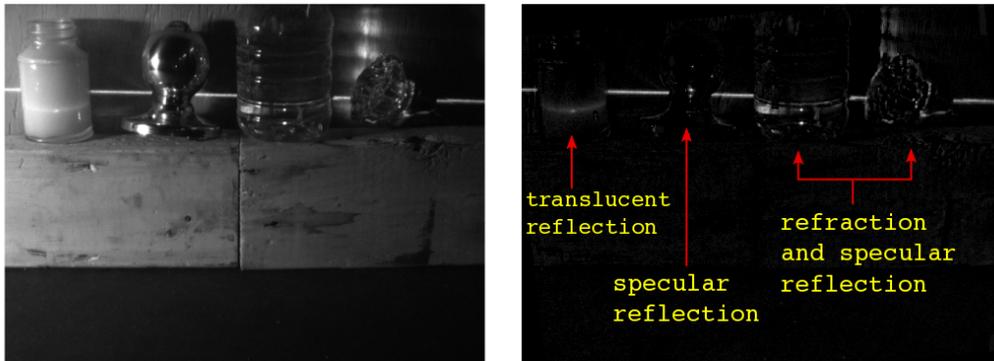


Figure 5: Left: Original image of the frame with the laser line from the pico-projector. The objects are milky water in a glass bottle, polished metal, bottle with clear water and a glass door knob. Right: Image after motion compensated background subtraction. The subtle optical effects are visible.

In the former work<sup>1</sup> the system could not only do 3D reconstruction, but also identify optical difficult objects like glass and metal. The key for dealing with these objects is to have the background almost completely removed so that the subtle optical effects become apparent. Our background subtraction method described is able to do this. It can be seen in Figure 5 where the laser light interacts with translucent material (milky water), glass, and polished metal.

## 5. CONCLUSION

We have shown in this paper that it is possible to do background subtraction for structured light sensing with images captured by a moving sensor. It can be done by using the images themselves to estimate and then compensate for the motion of the sensor. The method has been evaluated and successfully tested indoor and outdoors in bright sunlight. The background is sufficiently removed and we demonstrate the 3D reconstruction of a scene. One can also observe subtle optical effects caused by the interaction of the laser line with certain types of material such as glass or metal. Future improvements to our system include speeding up the full system and making the motion estimation more robust for some special cases.

## 6. ACKNOWLEDGMENTS

This work was conducted through collaborative participation in the Robotics Consortium sponsored by the US Army Research Laboratory (ARL) under the Collaborative Technology Alliance Program, Cooperative Agreement W911NF-10-2-0016.

## REFERENCES

- [1] Mertz, C., Koppal, S. J., Sia, S., and Narasimhan, S. G., “A low-power structured light sensor for outdoor scene reconstruction and dominant material identification,” in [*9th IEEE International Workshop on ProjectorCamera Systems*], (June 2012).
- [2] Irani, M. and Anandan, P., “About direct methods,” in [*Proceedings of the International Workshop on Vision Algorithms: Theory and Practice*], *ICCV '99*, 267–277, Springer-Verlag, London, UK, UK (2000).
- [3] Torr, P. H. S. and Zisserman, A., “Feature based methods for structure and motion estimation,” in [*Proceedings of the International Workshop on Vision Algorithms: Theory and Practice*], *ICCV '99*, 278–294, Springer-Verlag, London, UK, UK (2000).
- [4] Farid, H. and Woodward, J. B., “Video stabilization and enhancement,” *TR2007-605, Dartmouth College, Computer Science* (1997).
- [5] Matsushita, Y., Ofek, E., Ge, W., Tang, X., and Shum, H.-Y., “Full-frame video stabilization with motion inpainting,” *IEEE Trans. Pattern Anal. Mach. Intell.* **28**, 1150–1163 (July 2006).
- [6] Shum, H.-Y. and Szeliski, R., “Systems and experiment paper: Construction of panoramic image mosaics with global and local alignment,” *International Journal of Computer Vision* **36**(2), 101–130 (2000).
- [7] Grundmann, M., Kwatra, V., and Essa, I., “Auto-directed video stabilization with robust l1 optimal camera paths,” in [*IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*], (2011).
- [8] Bergen, J. R., Anandan, P., Hanna, K. J., and Hingorani, R., “Hierarchical model-based motion estimation,” in [*Computer VisionECCV'92*], 237–252, Springer (1992).
- [9] Lucas, B. D., Kanade, T., et al., “An iterative image registration technique with an application to stereo vision.,” in [*IJCAI*], **81**, 674–679 (1981).
- [10] Bouguet, J.-Y., “Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm,” *Intel Corporation, Microprocessor Research Labs* **5**, 1–10 (2001).