

# PROJECTOR DOMAIN PHASE UNWRAPPING IN A STRUCTURED LIGHT SYSTEM WITH STEREO CAMERAS

*Ricardo R. Garcia and Avideh Zakhor*

Department of Electrical Engineering and Computer Sciences, University of California, Berkeley

## ABSTRACT

Phase-shifted sinusoids are commonly used as projection patterns in structured light systems consisting of projectors and cameras. They require few image captures per 3D reconstruction and have low decoding complexity. Recently, structured light systems with a projector and a pair of stereo cameras have been used in order to overcome the traditional phase discontinuity problem and allow for the reconstruction of scenes with multiple objects. In this paper, we propose a new approach to the phase unwrapping process in such systems. Rather than iterating through all pixels in the two cameras to determine the global phase of each pixel, we iterate through the projector pixels to solve for correspondences between the two camera views. An energy minimization framework is applied to these initial estimated correspondences to enforce smoothness and to fill in missing pixels. Unlike existing approaches, our method allows simultaneous unwrapping of both camera images and enforces consistency across them. We demonstrate the effectiveness of our approach experimentally on a few scenes.

*Index Terms*— Structured light, phase unwrapping, stereo cameras, 3D depth capture

## 1. INTRODUCTION

A wide assortment of projection patterns are used within structured light (SL) systems [1]. In particular, phase-shifted sinusoids are popular since they offer low decoding complexity and only require three captures per 3D reconstruction [2,3]. In most systems with phase shifted sinusoids, multiple periods of the sinusoid are contained within the projected images. As such, only the wrapped phase value, rather than the true absolute phase of each pixel in the camera, is known. In order to determine the true correspondences between the camera and projector pixels, the phase image must be unwrapped.

Two-dimensional phase unwrapping has been extensively studied in the literature [4]. Phase unwrapping algorithms typically assume the majority of points in the scene to have smoothly varying phases/depths and few points with phase/depth discontinuities larger than  $\pi$ . In

practice, these assumptions restrict the types of scenes that can be reconstructed.

In order to overcome the traditional phase discontinuity problem and allow for the reconstruction of scenes with multiple objects, a second camera can be added to the single-camera, single-projector setup typically used in SL systems. The second camera resolves the global phase ambiguity as follows [5]: for each pixel in the first camera, the wrapped phase is determined, and the possible corresponding locations in the projector are determined. Given a projection pattern with  $M$  periods of the sinusoidal pattern, there are  $M$  columns in the projector possibly corresponding to the identified camera pixel. Using these,  $M$  points are triangulated in 3D space, each point is projected onto the image plane of the second camera, and its phase is compared to the original pixel in the first camera. The projected point with the closest phase to the reference pixel is identified as the correct correspondence. To unwrap the phase images of both observing cameras, this algorithm needs to be run twice. Since each run only sets the global phase for a single image at a time, it is possible for corresponding points across views not to be assigned the same global phase resulting in inconsistencies.

In this paper, we propose an alternative approach to using stereo cameras for phase unwrapping in SL systems. Rather than iterating through the pixels in a camera to unwrap the phase, we iterate through the pixels of the projector. There are several potential advantages to such a projector domain phase unwrapping approach. First, in any SL system, only scene points illuminated by the projector can be reconstructed; as such it makes sense for the reconstruction to be projector centric rather than camera centric. Second, in a projector centric approach the phase unwrapping algorithm only needs to be run once regardless of the number of cameras in the system. As such, it is likely to (a) be more computationally efficient, and (b) result in more consistent solutions with fewer discrepancies across cameras. Third, the computational complexity for a projector based solution is proportional to the number of projector pixels rather than the number of camera pixels as is in traditional camera centric reconstruction. Since from a technological point of view, camera resolution is likely to increase at a much faster rate than projector resolution, projector centric solutions are likely to be more computationally efficient.

The outline of this paper is as follows: in Section 2 we describe our basic approach; Section 3 presents results from an experimental data set; Section 4 includes concluding remarks.

## 2. PROPOSED METHOD

Our proposed method consists of three steps. First, assuming cameras and projectors are calibrated with respect to each other, for each projector pixel we determine its epipolar line in the two camera views, as shown in Fig.1. By searching along the line in the images, we find correspondences between the camera views. The set of possible correspondences between the two views is treated as the set of labels for that pixel. Second, loopy belief propagation is applied to minimize an energy function that both quantifies the likelihood of each label, and spatially smoothes a set of labels. Finally, the remaining points that are not solved via the stereo approach are unwrapped with respect to their already unwrapped neighboring points. In what follows, we describe each step in detail.

### 2.1. Determining Possible Correspondences

To solve for the global phase of both cameras, we process each pixel in the projected pattern. Given a pixel  $P$  in the projected image, the epipolar line corresponding to  $P$  in each camera view can be determined, illustrated as the lines in Figs. 1(a) and 1(b). In each camera view, we identify the pixels along the epipolar line with the same phase as  $P$ , identified as  $\{A_1, A_2, A_3\}$  and  $\{B_1, B_2\}$  on the epipolar line in Figs. 1(a) and 1(b). Then, the 3D position of each of these points is found by triangulating with point  $P$ , as illustrated in Fig. 2. All of the triangulated points lie along the ray out of the projector intersecting  $P$ .

Once the 3D position of each possible point is determined, the distances between all pairs of candidate points in camera A and B are calculated. For a pair of points corresponding to the true correspondence, the triangulated 3D positions are likely to be near one another, as illustrated by the black dots in Fig. 1. Therefore, we can estimate the likelihood of a correspondence by the Euclidean distance between a pair of points.

### 2.2. Labeling via loopy belief propagation

After the set of possible correspondences is determined for each pixel in the projector, an energy minimization framework is applied to determine the best labeling for each projector pixel. We only include projector pixels in the graphical model for which there is at least one pair of correspondences with distance smaller than a set threshold. If there are no possible correspondences within this threshold, the projector pixel is likely not viewable from both cameras. The set of labels for each projector pixel is all possible pairwise correspondences between the detected

points in each camera view. Specifically, in Fig. 1(a) and 1(b), the set of labels for pixel  $P$  are  $\{(A_1, B_1), (A_1, B_2), (A_2, B_1), (A_2, B_2), (A_3, B_1), (A_3, B_2)\}$ .

Similar to [5], we minimize a cost function consisting of a data and smoothing term:

$$E(f) = E_d(f) + E_s(f) \quad (1)$$

where  $f$  represents the labeling set for the pixels in the projector. The data component of the cost function comes from the likelihood of the correspondence as determined by the distance between the points in 3D space:

$$E_d(f) = \sum_{p_m \in F, p_m = \{A_m, B_m\}} \min(\gamma, \|l_{3D}(A_m) - l_{3D}(B_m)\|) \quad (2)$$

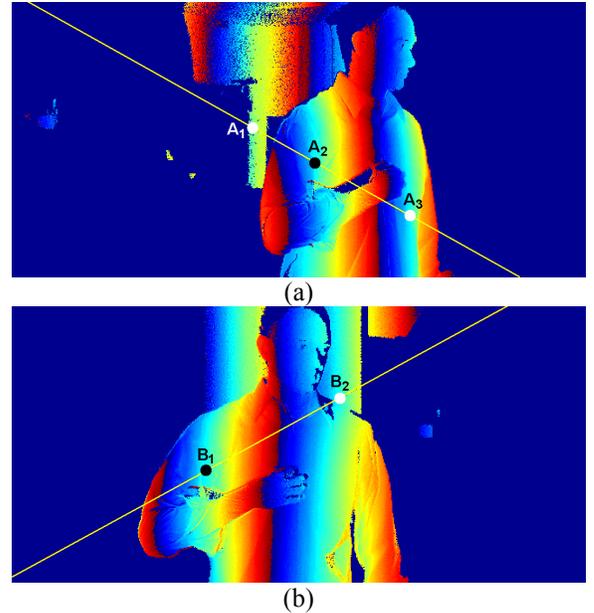


Figure 1. Wrapped phase images with epipolar line plotted. Each circle identifies a pixel with the same phase as the projector pixel; image from (a) camera A; (b) camera B.

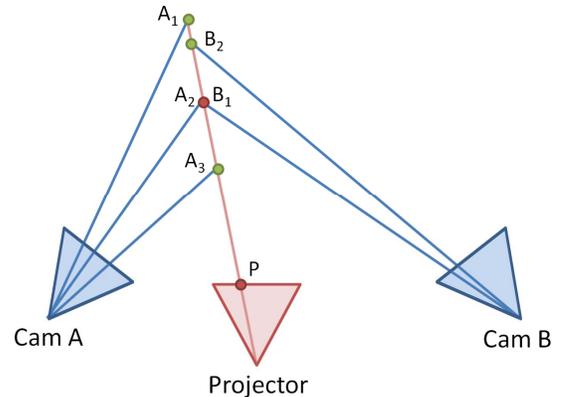


Figure 2. Illustration of the stereo camera geometry used to identify correspondences across camera views.

where  $p_m$  denotes the label of a pixel within the projector image  $F$  corresponding to a pixel  $A_m$  in camera A and  $B_m$  in camera B,  $\gamma$  is a threshold to limit the penalty from occlusions, and  $l_{3D}(A_m)$  denotes the 3D location of the labeled point  $A_m$ . The smoothing component is chosen to enforce a solution that minimizes large discontinuities within the scene. In most SL systems, the depth of the scene is assumed to vary slowly. This means that neighboring pixels within the camera image should have similar depths, and thus similar phase values. Since our method operates on the projector pixels, the smoothness assumption is made on the projector pixels. The basic intuition is that if two neighboring pixels in the projector illuminate the smooth scene surface, the two pixels in the camera image corresponding to these points should also be neighbors. This assumption fails when a scene discontinuity lies between the two examined pixels in the projector. Therefore, we enforce the following smoothness term:

$$E_s(f) = \sum_{\forall p_m, p_n \in F: p_m \in \mathcal{N}(p_n)} \min(\alpha, \|l_{2D}(A_m) - l_{2D}(A_n)\|) + \min(\alpha, \|l_{2D}(B_m) - l_{2D}(B_n)\|) \quad (3)$$

where  $p_m$  and  $p_n$  are neighboring points in the projector image, and  $A_m$  and  $B_m$  are the corresponding points for the pixel  $p_m$  in cameras A and B respectively. The function  $l_{2D}$  provides the 2D coordinates of the identified point, and  $\mathcal{N}(p_n)$  is the 4-connected neighborhood of pixels around  $p_n$ . The threshold  $\alpha$  is set minimize the penalty to occlusions.

To determine the labeling for the projector pixels that minimize (1), we apply loopy belief propagation. The message passed between two neighboring pixels is constructed as:

$$m_{p_m \rightarrow p_n}^i(f_n) = \arg \min_{f_m} \min(\gamma, \|l_{2D}(A_m) - l_{2D}(B_m)\|) + \min(\alpha, \|l_{3D}(A_m) - l_{3D}(A_n)\|) + \min(\alpha, \|l_{3D}(B_m) - l_{3D}(B_n)\|) + \sum_{p_j \in \mathcal{N}(p_m) \setminus p_n} m_{p_j \rightarrow p_m}^{i-1}(f_m) \quad (4)$$

where  $f_n$  is the label assigned to pixel  $p_n$ . To minimize the energy cost function, messages are generated and passed to their neighboring pixels in each iteration. After a predetermined number of iterations is reached, the final label for each individual pixel is chosen by summing the final messages entering that pixel, and choosing the label with the lowest cost.

### 2.3 Filling in missing phases

Unlike the previous two steps where the processing is done in the projector domain, in this step, missing phase values in camera images are filled in. Once the energy minimization is complete, the final labeling is used to assign global phase values to the phase images in the two camera views. For each pixel in the projector, the energy minimization

provides the most probable labeling. The final label of each projector pixel specifies the location where the projected pixel intersects the image plane of cameras A and B. Since we know the global phase for each point in the projected pattern, we can assign the identified image points with the proper global phase. As mentioned previously, our graphical model only provides global phase values for projector pixels that likely have a correspondence. Since not all projector pixels can be labeled, not all the pixels in the camera images can be labeled either. Nonetheless, it is possible to determine the global phase of these remaining pixels by local phase unwrapping with respect to the already unwrapped points.

Before unwrapping the remaining points, a quality map is derived from the camera's phase images. In two-dimensional phase unwrapping, quality maps quantify the confidence in unwrapping a local region of an image and are often based on local rates of change in the phase [4]. Regions of the phase image with slowly varying phase are unwrapped first since they can be unwrapped with fewer errors. In our implementation, the quality map is generated by computing local derivatives in the original wrapped phase image, and also by calculating the density of points with assigned phases via stereo unwrapping. Fig. 3(a) illustrates the points with phases from the stereo observations described in this paper. Fig. 3(b) illustrates the local density of global pixels from Fig. 3(a). Fig. 3(c) illustrates the local derivatives, and Fig. 3(d) illustrates the final quality map. The final quality map is generated by point wise multiplying the density image and the local derivative image.

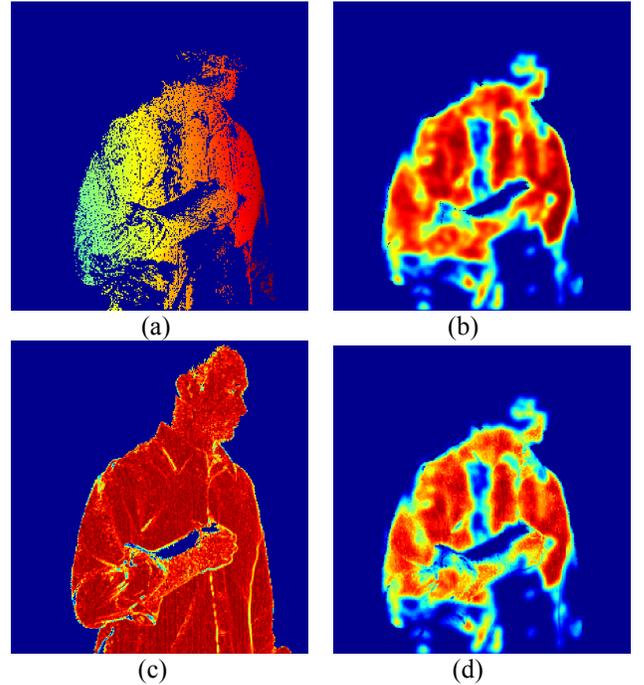


Figure 3. Quality map components; (a) pixels unwrapped using the approach of Section 2.1 and 2.2; (b) density of stereo unwrapped points; (c) local derivative measure; (d) final quality map.

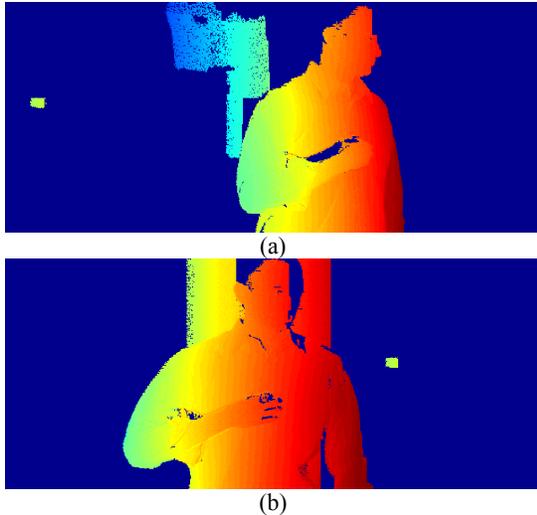


Figure 4. Final unwrapped images of (a) camera A, corresponding to Fig. 1(a); (b) camera B, corresponding to Fig. 1(b).

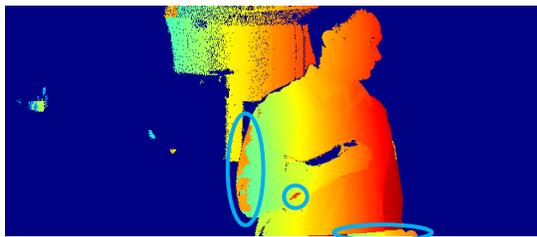


Figure 5. Unwrapped image for camera A, corresponding to Fig. 1(a), using the method in [5]. Incorrectly unwrapped regions are circled.

Once the quality map is generated, the points without an assigned global phase are unwrapped. Specifically, each pixel in the quality map is assigned a value from 0-1, and the quality values are placed into bins that span 0-1. Starting with the highest quality bin, we iterate through the non-unwrapped pixels in that bin and unwrap camera pixels that are neighbors to already globally unwrapped pixels. After all pixels in a bin are unwrapped, the pixels in the next quality bin are unwrapped. A list of the remaining unwrapped pixels is maintained, and the unwrapping process continues until the list of remaining pixels is empty.

### 3. RESULTS

Figs. 4(a) and 4(b) show the unwrapped results of the camera images shown in Figs. 1(a) and 1(b), respectively. The algorithm is clearly effective in properly unwrapping the illustrated phase images. Fig. 5 shows the unwrapping results for Fig. 1(a) using the approach in [5], but without using the left to right consistency check. The consistency is likely to remove some outlying pixels, but would not correct the mislabeled segments encircled in Fig. 5. Clearly, the reconstruction quality of Fig. 4(a) is superior to that of Fig. 5. In a sample sequence of 700 frames, our method provides more accurate results than [5] in 80% of frames, and performance comparable to or better than [5] in 96% of

frames [6]. The merged point cloud corresponding to the two cameras resulting from the unwrapped phase images in Figs. 4(a) and 4(b) is shown in Fig. 6. As seen, the merged point cloud is denser than either component; furthermore, the two components are reasonably aligned, and provide increased scene coverage by alleviating occlusion.

Our method is effective for scenes in which the majority of the points in the scene can be viewed by both cameras. In situations where the two camera views are significantly different, the stereo assisted method is unable to correctly identify many points in the scene

### 4. CONCLUSIONS AND FUTURE WORK

We have demonstrated a projector centric method to unwrap stereo images of a phase-shifted SL system. In contrast to [5], our method is able to simultaneously unwrap the phase of both cameras. In the future, we plan to modify our method for dynamic scenes by extending phase unwrapping to the temporal domain [7].

### 11. REFERENCES

- [1] J. Salvi, J. Pages, J. Battle, "Pattern codification strategies in structured light systems", *Pattern Recognition*, Volume 37, Issue 4, *Agent Based Computer Vision*, April 2004, Pages 827-849.
- [2] S. Zhang, P. Huang, "High-resolution, real-time 3-D shape acquisition," *IEEE Computer Vision and Pattern Recognition Workshop*, 03(3), 28-35, 2004.
- [3] S. Zhang, S. Yau, "Absolute phase assisted three-dimensional data registration for a dual-camera structured light system", *Appl. Opt.*, 47(17), 3134-3142, 2008.
- [4] D.C. Ghiglia, M.D. Pritt, *Two-Dimensional Phase Unwrapping: Theory, Algorithms, and Software* (Wiley, 1998).
- [5] T. Weise, B. Leibe, L. Van Gool, "Fast 3D Scanning with Automatic Motion Compensation," *Computer Vision and Pattern Recognition*, 2007. *CVPR '07. IEEE Conference on*, vol., no., pp.1-8, 17-22 June 2007.
- [6] [http://www-video.eecs.berkeley.edu/research/4D\\_SL/](http://www-video.eecs.berkeley.edu/research/4D_SL/)
- [7] R. Garcia, A. Zakhor, "Temporally-Consistent Phase Unwrapping for a Stereo-Assisted Structured Light System" *3DIMPVT 2011, Hangzhou, China*, May 16-19, 2011.

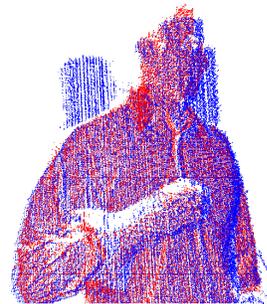


Figure 6. Merged camera A (red) and B (blue) points clouds.