

Reducing Interference Between Multiple Structured Light Depth Sensors Using Motion

Andrew Maimone*

Henry Fuchs†

Department of Computer Science
University of North Carolina at Chapel Hill



Figure 1: Left to Right: A) Depth map of room using a single Kinect. B) Depth map with 5 additional Kinects overlapping the view, causing prominent interference holes. C) Depth map with 5 additional overlapping Kinects while applying our interference reduction technique.

ABSTRACT

We present a method for reducing interference between multiple structured light-based depth sensors operating in the same spectrum with rigidly attached projectors and cameras. A small amount of motion is applied to a subset of the sensors so that each unit sees its own projected pattern sharply, but sees a blurred version of the patterns of other units. If high spacial frequency patterns are used, each sensor sees its own pattern with higher contrast than the patterns of other units, resulting in simplified pattern disambiguation.

An analysis of this method is presented for a group of commodity Microsoft Kinect color-plus-depth sensors with overlapping views. We demonstrate that applying a small vibration with a simple motor to a subset of the Kinect sensors results in reduced interference, as manifested as holes and noise in the depth maps. Using an array of six Kinects, our system reduced interference-related missing data from from 16.6% to 1.4% of the total pixels. Another experiment with three Kinects showed an 82.2% percent reduction in the measurement error introduced by interference.

A side-effect is blurring in the color images of the moving units, which is mitigated with post-processing. We believe our technique will allow inexpensive commodity depth sensors to form the basis of dense large-scale capture systems.

Keywords: motion, input devices, sharpening and deblurring

Index Terms: I.4.3 [Computing Methodologies]: Image Processing and Computer Vision—Image Enhancement

1 INTRODUCTION

The recent availability of the inexpensive Microsoft Kinect™ depth cameras has spurred development of 3D scene capture and motion tracking. However, as multiple units with overlapping views cause

prominent interference, resulting in holes and noise in the computed depth maps, the Kinect’s usefulness for building large dense real-time capture systems has been limited.

Previous attempts to resolve this inference problem have been effective, but do not scale past a few overlapping units. Maimone and Fuchs [5] and Kuster et al. [4] used hole filling and filtering techniques to fill interference holes and smooth noise for a small number of Kinect units overlapping in view. However, these approaches are designed for filling small isolated holes characteristic of limited interference. Berger et al. [1] installed a rotating shutter over each unit’s projector so that the devices could be time multiplexed. They found a reduction in missing data if a single unit was enabled per time-slice; however, for real-time applications the approach does not practically scale beyond a few units.

In this paper, we aim to overcome the limitations of past interference reduction techniques for a class of structured light cameras. Our solution is scalable to many overlapping units and does not cause a reduction in frame rate. The primary detriment is a small blurring of the color camera (if the device is so equipped) that can be mitigated with software post-processing. We demonstrate the effectiveness of the technique on an array of Kinect depth sensors and believe our method is applicable to other devices that are similar in operation. We believe our technique is the first¹ to allow simultaneous effective operation of a dense array of Kinect sensors.

2 BACKGROUND AND CONTRIBUTIONS

Structured light scanners project a known pattern into a scene and calculate depth by measuring the deformity of the pattern as seen by a camera. If multiple devices with overlapping views operate in the same spectrum at the same time, they will see the patterns of other units and may have trouble disambiguating them from their own.

¹After submission of this paper, we learned that a similar technique was developed independently at Microsoft Research Cambridge and will appear in an upcoming publication:

Butler, A., Izadi, S., Hilliges O., Molyneaux, D., Kim, D., Hodges, S. Shake ‘n’ Sense: Reducing Interference for Overlapping Structured Light Depth Cameras. To Appear In Proc. ACM CHI Conference on Human Factors in Computing Systems, May 2012.

*e-mail: maimone@cs.unc.edu

†e-mail: fuchs@cs.unc.edu

This problem is exacerbated as more units overlap and if similar patterns are used.

The Kinect sensor, as an inexpensive device designed to be used individually in the home, constantly projects the same fixed pattern of dots for all units on the same wavelength. Depth values are computed by determining the disparity between the observed projected pattern and a reference pattern for a known depth [2]. When correlating the known and observed pattern, the unit can fail to find a match in areas where there are overlapping dots from another device, resulting in missing data, or “holes”, in the depth map.

Since the Kinect hardware returns no data when it does not have confidence in pattern correlation, software options are limited. If interference is minimal, holes tend to be small and isolated. If one assumes surfaces are smooth and continuous, the holes can be filled as in [4] or [5]. As interference increases and holes become larger (as in Figure 1B), the surface can no longer be inferred unless strong assumptions are made about the scene content. Therefore, a hardware solution becomes necessary.

For the Kinect, we do not have control over how the device *interprets* what it sees, so we must change *what* it sees. Hardware solutions that utilize time multiplexing, such as shuttering in [1], must necessarily reduce that frame rate or exposure time by a factor inversely proportional to the number of overlapping devices, and thus are not practical for more than a few devices. Thus a hardware solution must cope with all devices projecting simultaneously.

Under these constraints, we contribute a novel solution to the interference problem between multiple structured light sensors that involves the introduction of relative motion between units.

3 TECHNIQUE

3.1 Theory of Operation

If a structured light scanner with a rigidly attached camera and projector is moved relative to another unit with overlapping views, and the amount of movement of its projected pattern is significant over the camera integration time, then each unit will see a sharp version of its own pattern and a blurred version of the other unit’s pattern (see Figure 2). Units in motion relative to the scene may also blur their own patterns slightly if depths change significantly over the integration time.

The blurring of the other unit’s pattern will help disambiguate the overlapping patterns if the following conditions are met:

1. The pattern is high frequency, i.e. on the scale of the amount of blur incurred, and blurring occurs in a meaningful direction, for example: not parallel to lines in the pattern. In this case, blurring results in a significant contrast loss. For the Kinect, this condition is met: the pattern consists of small dots that are only a few pixels across as seen by its camera.
2. The initial patterns of the two units are close in intensity as seen by each camera. A significant contrast loss in the interfering pattern should make a unit’s own pattern more prominent. For the Kinect, this condition is sometimes met: the pattern intensity varies with the distance to the camera and the reflectance functions of the surfaces in the scene.
3. Blurring of a unit’s own pattern (if depths change significantly during camera integration time) should occur at a smaller scale than spacing between pattern features. That is, features of a unit’s own pattern should remain distinguishable. For the Kinect, this condition is usually met, although the spacing between pattern features increases with depth, so the amount of blur should be limited for close objects.

3.2 Motion Patterns

The motion described in Section 3.1 has a significant impact on the design of our interference reduction system. We considered the following parameters:

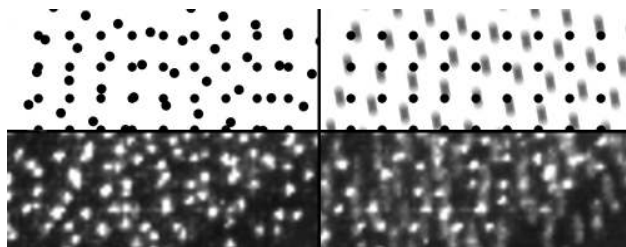


Figure 2: Pattern disambiguation. Left: Two similar overlapping patterns of similar intensity are difficult to disambiguate. Right: Introducing blur into one of the high frequency patterns significantly reduces its contrast, simplifying separation. Top: Idealized images. Bottom: Overlapping IR dot patterns of two Kinets.

1. Motion speed: controls the magnitude of the blur
2. Motion pattern: controls the blur shape. Desirable traits are rotational motion, which allows a constant amount of blur to be applied to the entire image, and cyclical motion, which allows the speed (and blur magnitude) to remain constant over time.
3. Total extent of motion from reference position: if motion oscillates a minimum amount around reference position, continuous camera registration is not required.
4. Degree of motion control: if camera motion is precisely controlled (and known), blur present in the color camera can be more effectively canceled.
5. Motion relationship to image sensor: if camera has a rolling shutter, certain motion patterns can cause undesirable effects (ex: “rolling” images).

Note that is only necessary to move enough cameras so that no two cameras with overlapping views are both motionless.

3.3 Effect on Color Image

If our interference reduction technique is applied to a depth camera that has a rigidly attached RGB camera (used to apply a color texture to the computed depth map), as does the Kinect, a negative side-effect is that it may produce slightly blurred images. Techniques for mitigating these effects are listed in Section 4.2.

4 IMPLEMENTATION

4.1 Motion Control

We created three motion control prototypes to test our motion-based interference reduction technique with Kinect sensors.

For the first prototype, a Kinect unit was mounted to a stepper motor in such a way that its pitch could be controlled (see left of Figure 3). The device was swept up and down several degrees, resulting in multiple frames per sweep cycle. Although this one-axis motion platform was effective at reducing multi-Kinect interference, it had several disadvantages. As the motion extent was substantial, integrating the data with other units would require continuous camera re-registration. The motion pattern also required a frequent change in direction, resulting in a brief period of no motion that caused interference to reappear. The platform also required a custom mount and approximately \$100 USD in parts, two-thirds of the cost of a Kinect.

For the second prototype, the stepper motor controller was re-programmed to oscillate up and down the small angle necessary to induce the desired blur, eliminating the need to continuous re-registration. The oscillation frequency was matched to multiples of the Kinect’s frame rate to prevent undesirable aliasing effects with the rolling shutter and was high enough so that changes in direction did not cause interference to reappear. This prototype was also effective but still expensive.



Figure 3: Motion platforms. Left: Motion platforms used for prototypes 1-2 of Section 4.1. Right: Motion platform used for prototype 3 of Section 4.1.

Table 1: Test Case 1 - Two Adjacent Close Range Overlapping Units

		Reference		Motion Speed							
		1	2	No Motion		Low		Medium		High	
		1	2	1	2	1	2	1	2	1	2
Additional Values (% Total Pixels)	Missing	-	-	14.1	9.5	3.1	4.9	1.5	3.4	1.4	1.8
RMS Instability (mm)		1.8	1.6	2.0	1.7	1.9	1.7	1.7	1.6	2.0	1.6

For the third prototype, a small DC motor was attached to the bottom of the Kinect with an eccentric mass on its shaft so that rapid motor rotations induce vibrations in the Kinect (see right of Figure 3). The motor was mounted with its shaft parallel to the camera’s line of sight so that oscillations primarily occur in the camera’s X and Y axis. The amount of vibration was controlled by modifying the DC motor voltage; this very simple implementation was also effective at reducing interference and was very inexpensive (\$5 USD per Kinect plus a shared power supply). However, motion patterns were more difficult to predict and control as the cameras vibrated about their mounted positions. This prototype was used for all results listed in Section 5.

4.2 Color Image De-Blurring

To reduce the effects of blurring, we estimated the point spread function of the blur using a high frequency, high contrast test target and deconvolved the image using the Lucy-Richardson method [6] as implemented in MATLAB. In the future, we plan to incorporate more sophisticated deblurring techniques, such as those utilizing inertial measurements of the moving camera [3], and implement a real-time solution. Another alternative is a reduction of the integration time of the camera, if such controls are available. Blur could also be eliminated completely by utilizing an external camera that is not rigidly mounted to the moving depth sensor.

5 RESULTS

5.1 Interference Reduction

To evaluate our method in terms of ground truth (interference-free data), we compared an array of Kinects operating simultaneously with our interference reduction technique applied to data captured from each camera individually while all others were turned off. We also compared the performance of the simultaneously operating Kinect array with and without our technique applied.

We characterized performance in terms of measurement error, measurement instability, and the amount of additional missing depth values in the depth map. Error was measured by placing a planar target in the scene and comparing individual measurements to a plane fit to the data and was computed as the average RMS point-plane distance over 15 frames. (Note that we did not use the interference-free data as ground truth for error measurement because it contains a reasonable amount of error itself). Instability was measured by comparing corresponding measurements in the depth maps across multiple frames of a static scene and was computed as the RMS error between all pairs of images over 15 frames. Pixels were considered in the calculations only if they did not cor-



Figure 4: Results for Test Case 2. Rows: Views from six cameras with overlapping views. Left Column: Reference images with no interference, each unit turned on individually. Middle Column: All units operating simultaneously. Right Column: All units operating simultaneously with our interference reduction technique applied.

respond to missing data in either image and the error was below a threshold. (The threshold prevents large errors caused by noise at depth discontinuities from skewing the results). The amount of additional missing depth values was computed as the number of missing data values above the amount found in the ground truth data set, expressed as a percent of total image area and averaged over a 15 frame sequence. Note that the ground truth data set also has missing data – the result of surfaces that were occluded, did not reflect infrared light, or were highly specular.

We applied our evaluation to three different Kinect configurations. The amount of missing values was evaluated in Test Cases 1 and 2, and error was measured in Test Case 3. Instability was also measured in all cases.

Test Case 1, quantified in Table 1, presents a near worst case pattern overlap scenario to show the robustness of our method. Two cameras were placed in very close proximity to each other (nearly touching), were aimed at slightly different angles, and were very close to objects in the scene (near the Kinect’s minimum depth range of approx. 50cm). The scene used is the same as shown in Figure 6. Camera 1 was placed in motion while camera 2 stayed a rest, resulting in very significant interference. As motion speeds increased, the amount of missing depth values decreased by a minimum factor of 3 to a maximum factor of 10. Changes in measurement instability were negligible.

Table 2: Test Case 2 - Six Overlapping Units

		Reference						No Motion						Medium Motion					
		1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6
Additional Missing Values (% Total Pixels)		-	-	-	-	-	-	17.2	15.7	18.3	17.2	15.2	16.2	0.9	0.8	2.8	1.5	-0.0	2.2
RMS Instability (mm)		18.1	24.9	18.4	18.7	13.1	21.5	19.3	26.8	18.1	19.9	13.5	23.1	20.4	26.4	20.7	20.3	14.6	23.3

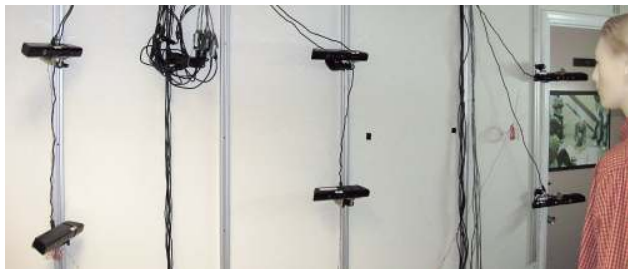


Figure 5: Six Kinect unit placement for Test Case 2 of Section 5.1

Table 3: Test Case 3 - Three Overlapping Units

		Reference			No Motion			Medium Motion		
		1	2	3	1	2	3	1	2	3
RMS Error (mm)		10.7	8.7	14.0	21.6	15.6	29.4	12.6	9.3	17.4
RMS Instability (mm)		6.9	5.3	6.9	9.7	7.3	12.2	12.8	7.8	15.4

Our second test case, shown in Figure 4 and quantified in Table 2, presents a near-worst-case scenario in the number of cameras that overlap a single nearby surface. The views of six densely placed Kinects converge on a nearby subject (a mannequin placed approximately 1.3 m from the central cameras) as shown in Figure 5. All cameras except camera 2 were placed in motion. Using our motion technique, the amount of additional missing depth values decreased by an average factor of 12. Few additional missing depth values remained and in one case fewer missing values were found than in the ground truth image. Measurement instability generally increased slightly for units in motion.

Our third test case, quantified in Table 3, provides an evaluation of measurement error. Three cameras were placed in an arc approx. 2 m from a 43 cm \times 56 cm planar target placed roughly parallel to the image plane of the central camera. Measurements corresponding to the planar target were segmented from the rest of the scene by depth values and were used for the previously described measurement error computation. Without our motion technique applied, the increase in error due to multi-camera interference was nearly 100%; with our interference technique applied, this figure dropped to approximately 18%, but an increase in measurement instability was noted.

5.2 Image Blurring

The three motion amplitudes (“low”, “medium”, and “high”) of Table 1 correspond approximately to motion blur with lengths of 1.5, 2.5, and 4.5 pixels respectively. Figure 6 shows a frame from a capture session with motion amplitude set to medium and deblurring applied as described in Section 4.2. The test pattern in the background shows that the image remains relatively sharp.

6 CONCLUSIONS AND FUTURE WORK

We have presented a solution to the problem of interference between multiple structured light depth cameras that is scalable and allows all devices to run at full frame rate. Our simple hardware solution uses parts costing only a few US dollars and can be installed



Figure 6: Deblurred color image from camera in motion at “Medium” setting of Table 1.

in minutes.

We have demonstrated that our system is effective at reducing interference holes and measurement error and scales to a large number of densely placed Kinect sensors with prominently overlapping views. We have also shown that a side-effect of our method, blurring of the color image, can be mitigated with image deconvolution.

We intend to test our technique on an expanded sensor array that can provide dense scene capture of an entire small room. Additionally, we plan to implement more sophisticated image deblurring techniques and implement them in real-time.

ACKNOWLEDGEMENTS

The authors would like to thank Kurtis Keller, Andrei State, and Jonathan Bidwell for technical discussions and advice. This work was supported in part by the National Science Foundation (award CNS-0751187) and by the BeingThere Centre, a collaboration of UNC Chapel Hill, ETH Zurich, NTU Singapore, and the Media Development Authority of Singapore.

REFERENCES

- [1] K. Berger, K. Ruhl, C. Brümmer, Y. Schröder, A. Scholz, and M. Magnor. Markerless motion capture using multiple color-depth sensors. In *Proc. Vision, Modeling and Visualization (VMV)*, 2011.
- [2] B. Freedman, A. Shpunt, M. Machline, and Y. Arieli. Depth mapping using projected patterns. Patent Application, 05 2010. US 2010/0118123 A1.
- [3] N. Joshi, S. B. Kang, C. L. Zitnick, and R. Szeliski. Image deblurring using inertial measurement sensors. In *ACM SIGGRAPH 2010 papers, SIGGRAPH '10*, pages 30:1–30:9, New York, NY, USA, 2010. ACM.
- [4] C. Kuster, T. Popa, C. Zach, C. Gotsman, and M. Gross. Freecam: A hybrid camera system for interactive free-viewpoint video. In *Proceedings of Vision, Modeling, and Visualization (VMV)*, 2011.
- [5] A. Maimone and H. Fuchs. Encumbrance-free telepresence system with real-time 3d capture and display using commodity depth cameras. In *Mixed and Augmented Reality (ISMAR), 2011 10th IEEE International Symposium on*, pages 137–146, oct. 2011.
- [6] W. H. Richardson. Bayesian-based iterative method of image restoration. *J. Opt. Soc. Am.*, 62(1):55–59, Jan 1972.