On Making Projector Both a Display Device and a 3D Sensor

Jingwen Dai and Ronald Chung

Department of Mechanical and Automation Engineering
The Chinese University of Hong Kong, Hong Kong
{jwdai,rchung}@mae.cuhk.edu.hk

Abstract. We describe a system of embedding codes into projection display for structured light based sensing, with the purpose of letting projector serve as both a display device and a 3D sensor. The challenge is to make the codes imperceptible to human eyes so as not to disrupt the content of the original projection. There is the temporal resolution limit of human vision that one can exploit, by having a higher than necessary frame rate in the projection and stealing some of the frames for code projection. Yet there is still the conflict between imperceptibility of the embedded codes and the robustness of code retrieval that has to be addressed. We introduce noise-tolerant schemes to both the coding and decoding stages. At the coding end, specifically designed primitive shapes and large Hamming distance are employed to enhance tolerance toward noise. At the decoding end, pre-trained primitive shape detectors are used to detect and identify the embedded codes – a task difficult to achieve by segmentation that is used in regular structured light methods, for the weakly embedded information is generally interfered by substantial noise. Extensive experiments including evaluations of both code imperceptibility and decoding accuracy show that the proposed system is effective, even with the prerequisite of incurring minimum disturbance to the original projection.

1 Introduction

The improving performance and declining price of digital video projectors make it possible to use them prevalently. Being able to generate arbitrarily large display is a feature of projectors that makes them exceedingly attractive, especially in applications that demand portability.

On the other hand, the adoption of structured light illumination has been proven to be an effective and accurate means for 3D information perception [1]. Recently, the availability of pico projectors with average dimensions of $4 \times 2 \times 1$ inches has widely extended the application domain of structured light system. There are already pocket DCs, DVs and cellular phones in the consumable market that have both projector and camera built-in, making it possible to implement structured light system in hand-held consumer electronic products.

For these reasons projector-camera (ProCam) system has been actively researched in the last few years. Many research groups apply projectors in unconventional ways to develop new and innovative information displays that go beyond simple screen presentation [2].

G. Bebis et al. (Eds.): ISVC 2012, Part I, LNCS 7431, pp. 654–664, 2012.

Some researchers designed structured light system in the non-visible spectrum [3]. That way the media for regular projection and structure light sensing can be made separate. However, if structured light and regular projection can be achieved through the same projector, additional hardware demand could be reduced and device cost and size could be diminished. This leads to the concept of Imperceptible Structured Light (ISL). ISL makes use of a projection frame rate that is beyond the perceptibility of human vision, so that some of the projected frames can be ”stolen” for structured light use without the user perceiving it. Specifically, it modulates the projected display either spatially or temporally to embed code patterns into the projection for structured light sensing. Due to the limitation of human visual perception, such embedded code patterns can be made largely or entirely unnoticeable to the user (the degree of unnoticeability depends upon how fast is the projection frame rate and how wide is the intensity contrast between the intended projection and the embedded code), but cameras synchronized to the modulation are able to reconstruct the embedded codes for structured light sensing.

There is however challenge in embedding codes into user-specified arbitrary projection. While the codes should be made as undetectable as possible to the user, they have to be decodable to the camera for the purpose of structured light sensing. On top of the dilemma, there is the inevitable fact that the displayed signals are generally corrupted by substantial noise that arises from the nonlinearity of the projector, the sensing resolution and other limits of the camera, and the variation of the ambient illumination. The objective of this work is to deal with the dilemma.

This article describes a novel method of embedding imperceptible structured codes into arbitrarily intended projection. Through precise projector-camera synchronization, structured codes consisting of three primitive shapes are embedded into the projection, in a way that is imperceptible to viewers but extractable from the ”difference” of successive images captured by a camera. To make the decoding process more robust against noise, we do not extract the codes by region segmentation in the image domain. Instead we employ specially trained classifiers to detect and identify the codes. To enhance the error tolerance further, specially designed primitive shapes and large Hamming distance are adopted in the spatial coding. Even with some bits of the codewords missed or wrongly coded, the correct correspondence could still be derived correctly.

The remainder of this paper is structured as follows. In Section 2, related works on imperceptible structured light sensing are briefly reviewed. The principle of embedding imperceptible codes along with robust coding and a noise-tolerant decoding mechanism are described in Section 3. In Section 4, system setup and experimental results are shown. Conclusion and possible future work are offered in Section 5.

2 Related Works

A proof of concept for embedding invisible structured light patterns into DLP (Digital Light Processing) projections first appeared in the ”Office of the Future” project [4]. In this work, binary codes are embedded by projecting temporally alternating code images and their complements. Provided that the frequency of projection reaches the *flicker fusion threshold* ($\geq 75Hz$), the pattern and the inverse pattern are visually integrated over time in human perception, and the illumination has the appearance of a flat field (”white” light) to humans. However, the demonstration required significant modification effort on the projection hardware and firmware, including removal of the color wheel and reprogramming of the controller. The resulting images were also in greyscale only. The implementation of such a setting was impossible without mastering and full access to the projection hardware.

Cotting et. al. introduced a coding scheme [5] that synchronizes a camera to a specific time slot of a DLP micro-mirror flipping sequence in which imperceptible binary patterns are embedded. However, not all mirror states are available for all possible intensities, and the additional hardware, DVI repeater with tapped vertical sync signal, is not an off-the-shelf instrument.

However, with the development of digital projection technology, some so-called 3D compatible DLP projectors with fresh rate of $120Hz$ or higher emerged recently. This makes it possible to implement imperceptible structured light without any hardware modification or extra assisting hardware. Many researcher began to study how to determine the embedded intensity properly to guarantee code imperceptibility.

In [6], subjective evaluation results and their statistical analysis on the visual perceptibility of embedded codes in different ways were reported. The factors affecting code visibility are also concluded. Park et al. [7] presented a technology for adaptively adjusting the intensity of the embedded code with the goal of minimizing its visibility. It was regionally adapted depending on the spatial variation of neighboring pixels and their color distribution in the YIQ color space. The final code intensity was then weighted by the estimated local spatial variation. Since two manually defined parameters adjusted the overall strength of the integrated code, the system was not able to automatically calculate an optimized intensity. Grundhofer et al. [8] proposed a method considering the capabilities and limitations of human visual perception for embedding codes. It estimated the Just Noticeable Differences (JND) based on the human contrast sensitivity function and adapted the code intensity on the fly through regional properties of the projected image and code, such as luminance and spatial frequencies. The shortcoming of this method was that some parameters need be pre-measured using some optical devices (e.g. photometer), which were not accessible to nonprofessional users.

To the best of our knowledge, up to now, few works focus on the decoding method in imperceptible code embedding configuration, especially when huge external noise could exist.

3 Methodology

3.1 Principle of Embedding Imperceptible Codes

The fundamental principle behind imperceptible structured code embedding is the temporal integration achieved by projecting each image twice at high frequency: a first image containing actual code information (e.g., by adding or subtracting a certain amount (Δ) to or from the pixels of the original image, depending upon the code) and a second image that compensates for the distortion in the first image. The vital aspects of ISL sensing are code embedding and projector-camera synchronization.

Since projection is generally in color, it is possible to embed color code through three different channels theoretically. However, to enhance code robustness toward noise, we use binary code and embed it into all three color channels simultaneously. Let B, O, I and I' be the binary code image, the original image, the projected image, and the complementary image (that is also projected) respectively. Then the projected image and complementary image could be formulated as

$$I_i(x,y) = O_i(x,y) + P(x,y), \tag{1}$$

$$I'_i(x,y) = O_i(x,y) - P(x,y), \tag{2}$$

$$P(x,y) = \begin{cases} \mathit{\Delta}, & when \quad B(x,y) = 1; \\ 0, & when \quad B(x,y) = 0. \end{cases} \tag{3}$$

where i = {R,G,B} indicates red, green and blue channels, $\mathit{\Delta}$ is the embedded intensity.

To avoid intensity saturation at lower and higher intensity levels when adding or subtracting $\mathit{\Delta}$, the original image needs to have the intensity range in each color channel compressed to between $\mathit{\Delta}$ to $255 - \mathit{\Delta}$. Since the embedded intensity required in the coding is small enough, the visual degradation due to contrast reduction is generally negligible.

The degree of imperceptibility thus depends upon the embedded intensity. A larger intensity ensures that the code be more tolerant toward noise and more readable in the image of the projection, whilst a smaller intensity makes the embedded codes more invisible. In our design, code imperceptibility has higher priority, and thus embedded intensity is set to a very small value.

In order to achieve imperceptible structured light projection, the frequency of projection must exceed the flicker fusion threshold, which is $75Hz$ for most of the people. The embedded codes could be internally and simply extracted from the ”subtraction image”[1] between consecutively captured images as

$$S(x,y) = \max_i [C_i(x,y) - C'_i(x,y)], \quad i = \{R, G, B\}. \tag{4}$$

In principle, the subtraction image should be a binary image that has intensity values between $2\mathit{\Delta}$ and 0. However, the subtraction image in reality is generally disturbed by rather substantial external noise. Since the embedded intensity is always small, the subtraction image has low signal-to-noise ratio. It is generally nontrivial to retrieve the embedded codes. In the rest of this section, we describe how robust coding and noise-tolerant decoding approaches can help tackle the issue.

3.2 Design of Embedded Pattern

Considering the constraints of imperceptible code embedding, we employ the spatial multiplex scheme to design our pattern. Due to the choice of using binary code for robust code embedding, the symbols cannot be coded with different colors. We thus use an alphabet set comprising three different geometrical primitives: cross, sandglass, and rhombus, as shown in Fig. 1. There are three advantages of this configuration. First, all the shapes own a natural center point, which simplifies the shape identification process in the decoding stage. Then, there are sufficient variations between different

[1] All the subtraction images in this article are scaled to [0, 255] for illustration purpose.

shapes; even with large disturbance from noise on the shapes, the decoding method could distinguish them. Moreover, the directional information carried by the cross shape could rectify the observation window during the step of neighborhood detection without enforcing any other constraint.

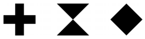


Fig. 1. The primitive shapes: cross, sandglass and rhombus

In the decoding stage, the centroid of each detected primitive is considered as the feature point position, and the 9-bit codeword associated to each feature point is composed of the elements in the 3×3 window centered on it. In traditional structured light methods, the uniqueness of the codeword is usually assured by M-arrays (perfect maps), which are random arrays of dimensions $r \times v$ in which a sub-matrix of dimensions $n \times m$ appears only once in the whole pattern [9]. The M-arrays give a total of $rv = 2^{nm} - 1$ unique sub-matrices in the pattern and a window property of $n \times m$. However, the Hamming distance between the codewords is 1, which is generally too small for our code embedding scenario in which the codeword retrieval error could be large due to noise. In our system, we generate a matrix of dimensions 27×29 using the method proposed by Albitar [10], in which 95.97% of the codewords have a Hamming distance higher than 3 and the average Hamming distance is $\bar{H} = 6.0084$, so that even some bits in the codeword are missed or incorrectly coded, the codeword is still distinguishable. On the basis of this matrix, the binary code image composed of the primitive shapes appears like the one illustrated in Fig. 2, in which the size of each primitive shape is a collection of 11×11 pixels while the interval between each shape is 11 pixels. The total number of feature points is 783.

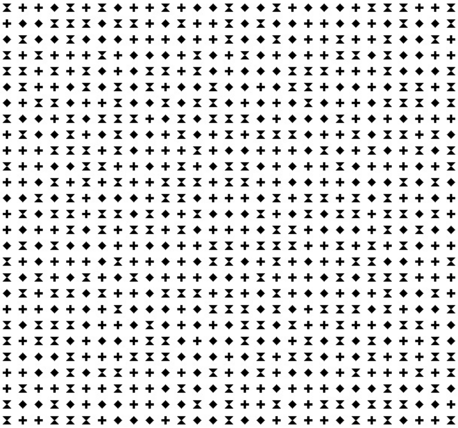


Fig. 2. The embedded binary code image

3.3 Primitive Shape Identification and Decoding

In the decoding stage, the existence of intense noise (due to influence from projector projection, camera sensing, ambient illumination and object surface reflection) makes it impossible to segment the primitive shape by the integrated use of region segmentation and edge or contour detection as in ordinary structured light methods. Here, we regard the primitive shapes as objects to ”identify” and ”detect” rather than ”segment”.

Compared with other object identification or recognition methods, the machine learning approach proposed by P. Viola [11] has been shown to be capable of processing images rapidly with high detection rates for visual object detection. The approach is adopted here for training a detector that identifies the three primitive shapes. Below we use cross shape as an example to describe the procedure of detector training.

The performance of any training-based detector has a great deal to do with the availability of training samples. Unlike generic objects like human face, body, or vehicle, which have a large number of samples in a great many of public databases, we have to collect the specific training samples ourselves in the required configuration. 500 color images with different content were collected from Internet, and 40 cross shapes were embedded in those images at different positions to generate 500 pairs of projected images and complementary images. By projecting them to a locally smooth textureless surface with orientation variations, 500 subtraction images could be derived from image capture exercises. The sub-images containing cross shapes were then segmented by manual labeling, which were considered as positive training samples. The background images with holes filled by random noise were divided into small patches to generate negative training samples. The training sample preparation process is shown in Fig. 3.

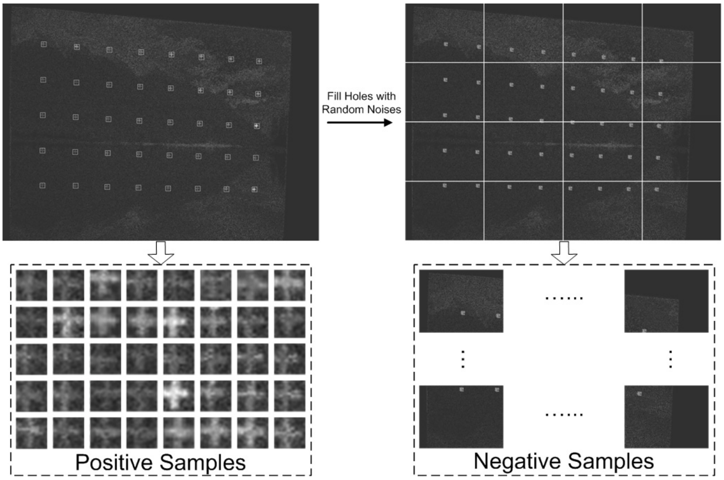


Fig. 3. Training sample preparation

To obtain the optimal performance, the positive samples were resized to 20 × 20, and the extended haar-like features and Gentle Adaboost algorithm were employed, following the suggestion in [12]. Eventually, from over 7000 positive samples and 3000 negative samples, a 16-stage cascade classifier for cross detection was trained. Following the same procedure, the detectors for sandglass and rhombus shapes were derived as well.

By using the pre-trained primitive shape detectors, the centroid of each primitive, i.e., the position of each feature point, can be determined. Once a feature point is extracted from the image, its codeword can be produced from the associated 3×3 intensity window centered on the feature point. Its corresponding point on the projector's display panel is known *a priori*. This way 3D position on the object surface can be determined via triangulation. The above is the 3D sensing step we use in the system.

4 Experiments

To assess the feasibility of the proposed method for embedding imperceptible codes in regular projection, we conducted experiments on both imperceptibility evaluation and accuracy evaluation.

The projector-camera system we used consisted of a DLP projector (Mitsubishi EX240U projector) of 1024×768 resolution and $120Hz$ refresh rate, and a camera (Adimec OPAL-1000 CCD camera with Myutron FV1520 f15mm lens) of 1024×1024 resolution and $123fps$ frame rate, both being off-the-shelf equipments. The focal length of the camera was fixed ar $15mm$, while that of the projector was in the range of $25 - 31mm$. The ILS was configured for a working distance (the distance from the camera to the mean position of the working area) of about $800mm$.

4.1 Imperceptibility Evaluation

Embedded code imperceptibility and user satisfaction are of the first priority in the system design. We conducted a subjective evaluation based on a questionnaire. Ten persons were invited to participate in this experiment, of which six were male and four were female, and seven wearing glasses. 500 images were collected from Google Image randomly, in which our proposed pattern was embedded with different intensities. The viewers were seated in front of a white planar screen at a distance of about $2m$, and asked to comment on the images projected to the screen. The questions asked were simplified from the questionnaire in [6], focusing on the feeling of flickering, the recognition of image deterioration, and the overall satisfaction for projection quality. The score for each question was divided into 10 levels.

The average scores of the subjective evaluation are illustrated in Fig. 4. When the embedded intensity is small, i.e., $\Delta = 5, 10$, the viewer could rarely notice the embedded codes and were satisfied with the projection quality. With the increase of the embedded intensity, the viewers' sense of flickering and image degradation became stronger. When $\Delta = 25$, almost every viewer was not satisfied with the projection quality.

In practice, because it was difficult to retrieve weakly embedded codes with the standard commercial cameras, we choose $\Delta = 10$ in our configuration, striking a compromise between user satisfaction and code imperceptibility.

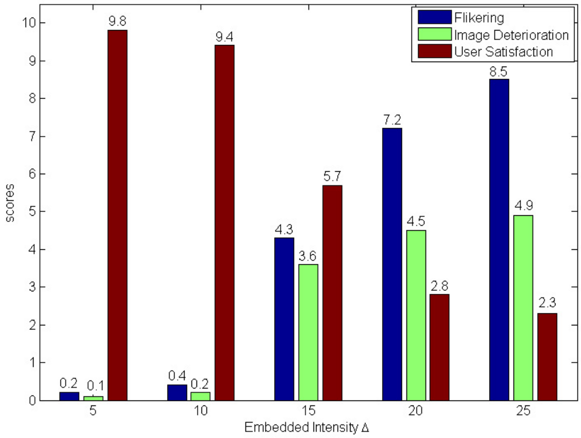
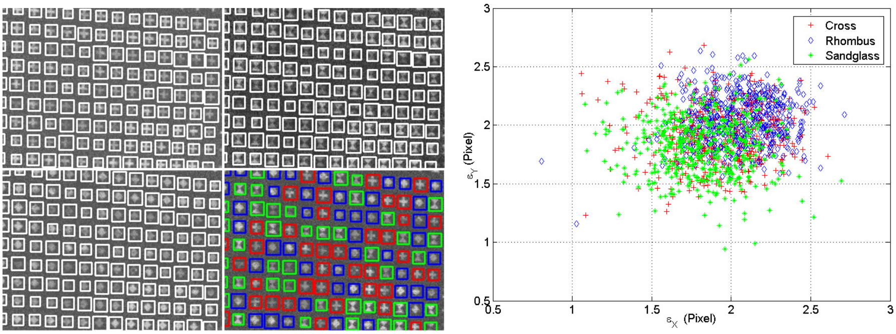


Fig. 4. Subjective evaluation result on code imperceptibility

4.2 Accuracy Evaluation

After code imperceptibility evaluation, the experiments on code retrieval accuracy were carried out. To assess accuracy, experimental data with ground-truth were required. Three different primitives and the spatially coded pattern image were embedded into the 500 images used for imperceptibility evaluation respectively, with intensity $\varDelta = 10$. Then the projected and complementary images were projected successively to a smooth surface, while the camera conducted synchronized capture. The surface was adjusted to different positions and orientations with respect to the camera to involve sufficient shape distortion in the test data. Then the subtraction images embracing embedded codes information were derived for accuracy evaluation. The ground-truth was obtained by manual labeling in the image data captured under binary pattern illumination.

Experimental results in some subtraction images are presented in Fig. 5(a). The four sub-figures display the cross (top-left), sandglass (top-right), rhombus (bottom-left) shapes, and the spatially coded pattern (bottom-right) respectively. For qualitative evaluation, the detected features are indicated by rectangles, and in bottom-right sub-figure, the cross,sandglass and rhombus shapes are separately marked by red, green and blue rectangles. The average feature point detection errors along the x-axis and y-axis (as shown in Fig. 5 (b)) are formulated as $\epsilon_X = \frac{1}{N}\sum_{i=1}^{N} |X_d - X_g|_i$, $\epsilon_Y = \frac{1}{N}\sum_{i=1}^{N} |Y_d - Y_g|_i$, where N is the total number of embedded shapes, (X_d, Y_d) and (X_g, Y_g) are the detected position coordinates and ground-truth respectively. More detailed quantitative testing results are listed in Table 1. Through the proposed method, 91.23% of the embedded feature points could have their correspondences found correctly. By analyzing the missed and false detection cases, we find that the mistakes were mainly caused by large noise that occludes the embedded codes, implying that external noise has the greatest influence on the decoding process.



(a) (b)

Fig. 5. (a) Some qualitative experiment results on (embedded) code detection accuracy. (b) The average detection error upon the three primitive shapes.

Table 1. The quantitative experiment results on (embedded) code detection accuracy

	Hits(%)	Missed(%)	False(%)	$[\epsilon_X, \epsilon_Y]$(pixel)	Corr. Acc.(%)
Cross	86.21	11.63	2.16	[1.931, 1.927]	—
Rhombus	85.83	12.57	1.60	[2.056, 2.051]	—
Sandglass	87.49	11.64	0.87	[1.816, 1.821]	—
Whole Pattern	86.33	11.06	2.61	[2.013, 2.043]	91.23

4.3 3D Reconstruction Accuracy Evaluation

To evaluate the accuracy of the proposed method in the 3D reconstruction task, we conducted an experiment to compare the performance of our method with that of a classical structured light method using visible patterns. As shown in Fig. 6-(a1)(b1)(c1) and Fig. 6-(a2)(b2)(c2), three objects (sphere, cone and cylinder) with known dimensions were illuminated by visible binary pattern image (the same as Fig. 2) and code embedded normal projection respectively.

In the classical structured light scenario, some feature points were extracted by segmentation and shape identification using the method proposed in [10]; whilst in our code embedded normal projection scenario, the feature points were detected and classified through the pre-trained primitive shape detectors. The depth value of each feature point was calculated through triangulation using the intrinsic and extrinsic parameters of the projector and camera. Then on the basis of point clouds calculated through our method, surfaces were rendered as illustrated in Fig. 6-(a3)(b3)(c3). Since the dimensions of the objects are known, we could conduct quantitative accuracy assessment. The residual mean error E_μ and standard deviation E_σ of the calculated 3D points with respect to the ground-truth were listed in Table 2. It is evident that our method has almost the same performance as that of the classical structured light method in 3D reconstruction. By the reason that the textures on the cylindrical object obstruct code retrieval, the reconstruction error on the particular object is greater than those of the other two objects. It is worth pointing out that in our method the decoding process was conducted in the subtraction image, which would alleviate the texture's influence to a certain extent.

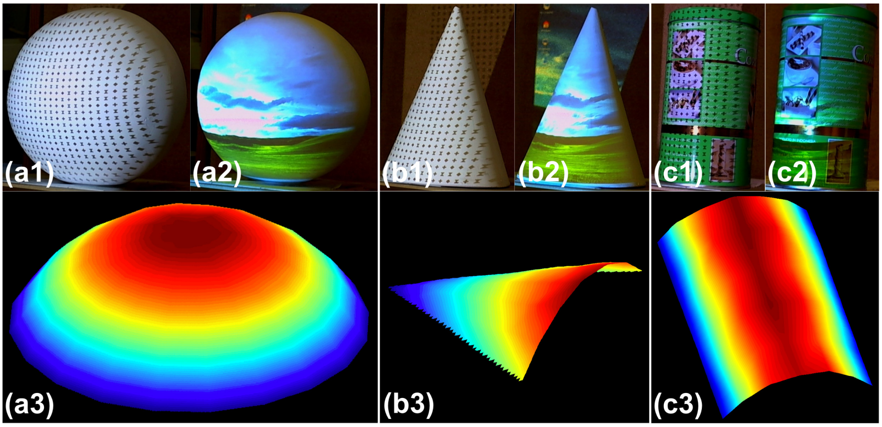


Fig. 6. Some results of 3D reconstruction

Table 2. 3D reconstruction accuracies on a variety of shapes

Object	General SL [10]		Our Method	
	$E_\mu(mm)$	$E_\sigma(mm)$	$E_\mu(mm)$	$E_\sigma(mm)$
Sphere	1.502	0.576	1.410	0.587
Cylinder	2.054	0.824	1.939	0.762
Cone	1.383	0.557	1.391	0.564

5 Conclusion and Future Work

We have described a novel system of embedding imperceptible structured codes into user-define arbitrary projection, that strikes the balance between imperceptibility and detectability of the codes. Through precise projector-camera synchronization, structured codes consisting of three primitive shapes are embedded into the regular projection, in a way that is imperceptible to the user but extractable by a camera (via the difference of successive images). Disturbance from various external sources makes it difficult to retrieve the codes by the region segmentation approaches adopted in general structured light systems. Instead of segmenting the codes, specially trained classifiers are employed to detect and identify them. To increase the robustness of code extraction, large Hamming distance are adopted in spatial coding. Even if some bits are missed or wrongly decoded, the correct correspondence between the projection panel and the image plane could still be arrived at correctly for structured light sensing. Extensive experimentation shows that the method is a promising one.

In the current system, the image capture interval is $10ms$. In sensing object that moves fast, the substantial displacement between successive images will result in blur or destruction of the embedded codes in the difference image. Some compensation methods need be in place to deal with the problem. In addition, the embedded code could be denser for more precise 3D sensing. New coding scheme capable of generating denser patterns should be used. The proposed method enables a regular projector to serve the dual role of a display device as well as a 3D sensor. That provides a platform for more natural user interface schemes. Our future work will lie on these directions.

References

1. Salvi, J., Fernandez, S., Pribanic, T., Llado, X.: A state of the art in structured light patterns for surface profilometry. Pattern Recognition 43, 2666–2680 (2010)
2. Bimber, O., Iwai, D., Wetzstein, G., Grundhöfer, A.: The visual computing of projector-camera systems. In: ACM SIGGRAPH 2008 Classes. SIGGRAPH 2008, pp. 1–25 (2008)
3. Fofi, D., Sliwa, T., Voisin, Y.: A comparative survey on invisible structured light. In: Proc. of Machine Vision Applications in Industrial Inspection XII, pp. 90–98 (2004)
4. Raskar, R., Welch, G., Cutts, M., Lake, A., Stesin, L., Fuchs, H.: The office of the future: A unified approach to image-based modeling and spatially immersive displays. In: Proc. of SIGGRAPH 1998, pp. 179–188 (1998)
5. Cotting, D., Naef, M., Cross, M., Fuchs, H.: Embedding imperceptible patterns into projected images for simultaneous acquisition and display. In: Proc. of IEEE and ACM ISMAR, pp. 100–109 (2004)
6. Park, H., Seo, B.-K., Park, J.-I.: Subjective evaluation on visual perceptibility of embedding complementary patterns for nonintrusive projection-based augmented reality. IEEE Trans. Circuits Syst. Video Technol. 20(5), 687–696 (2010)
7. Park, H., et al.: Content adaptive embedding of complementary patterns for nonintrusive direct-projected augmented reality. In: HCI International, pp. 132–141 (2007)
8. Grundhofer, A., Seeger, M., Hantsch, F., Bimber, O.: Dynamic adaptation of projected imperceptible codes. In: Proc. of IEEE and ACM ISMAR, pp. 1–10 (2007)
9. Etzion, T.: Constructions for perfect maps and pseudorandom arrays. IEEE Transactions on Information Theory 34, 1308–1316 (1988)
10. Graebling, P.: Robust structured light coding for 3d reconstruction. In: Proc. of IEEE ICCV, pp. 1–6 (2007)
11. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proc. of IEEE CVPR, pp. 511–518 (2001)
12. Lienhart, R., Kuranov, A., Pisarevsky, V.: Empirical analysis of detection cascades of boosted classifiers for rapid object detection. In: Proc. of DAGM PRS, pp. 297–304 (2003)