Recovering Stereo Depth Maps Using Single Gaussian Blurred Color Structured Light Pattern

Xida Chen Yee Hong Yang
Department of Computing Science, University of Alberta
{xida, yang}@cs.ualberta.ca

Abstract

In this paper, we present a structured light method to recover depth maps. Contrary to most temporal coding methods which require projecting a series of patterns, our method needs one color pattern only. Unlike most spatial coding methods which establish correspondence only along the edges of the captured images, our method produces a dense set of correspondence. Our method is built upon an important observation that a Gaussian blurred De Bruijn pattern preserves the desirable windowed uniqueness property. A Gaussian blurred De Bruijn pattern is used so that the color of every illuminated pixel is used to its fullest advantage. The simulated experiments show that the proposed method establishes a correspondence set whose density and accuracy are close to that of using temporal coding methods. We also demonstrate the robustness of our approach by applying it to several real-world datasets.

1. Introduction

Stereo matching refers to the process of taking two or more images as input, and establishing correspondence among pixels in the images, and finally converting their 2D positions into depths. It has been extensively researched in the computer vision community, and continues to be an active research area. There are many applications for stereo vision in the various fields. For example, it is an important step for a robot to extract information about relative positions of objects in a scene. It can also be used in object recognition and is applied to view morphing, image-based rendering, 3D model building and so on.

Stereo methods can be classified into two major categories, which are *passive* and *active* methods. Passive stereo methods established correspondence among images with no priori information. Passive stereo research has reached a new era since the introduction of a publicly available performance evaluation site called Middlebury Computer Vision Page [14], which allows researchers to compare their methods with other state-of-the-art methods. However, passive stereo methods have limitations when there are occlusions and textureless regions present in the scenes. In contrast, active methods such as laser scanning and structured light, project illumination pattern into the scenes to create identifiable features and hence to minimize the difficulty of establishing correspondence. In fact, the structured light method [22] is used to construct the ground truth for the depth maps used in the Middlebury site [14]. Therefore, the active methods are much more accurate comparing with the passive methods.

In this paper, a structured light method is proposed to recover depth maps. Unlike the method in [22] that requires projecting a series of illumination patterns, our method projects only one single color pattern. At the heart of our approach is a very important but overlooked observation that a Gaussian blurred pattern can be used to construct easily identifiable features, the property of which is proved in Section 3. The usage of a Gaussian blurred pattern provides dense and accurate correspondence. The advantages of using a Gaussian blurred pattern are further discussed in Section 5.

Our major contributions include the following.

- An important observation that a Gaussian blurred De Bruijn pattern can be used in a structured light system.
- A new method that requires only one-shot of the scene to produce dense stereo correspondence.

The organization of this paper is as follows. Section 2 introduces some related work. Our method is proposed in Section 3. Following that, Section 4 shows some experimental results. We discuss the advantages of our method in Section 5. Finally we conclude and present future work in Section 6.

2. Related Work

In the early stage of active stereo vision research, a laser or light stripe is used to scan the scene. This kind of techniques can be time consuming. Later on, much faster methods are developed which project coded patterns into the scene. These methods are referred to as the coded structured light methods, and there are some excellent surveys exist [2] [20] [21]. To provide an exhaustive review of previous work is certainly beyond the capacity

of this paper. Instead we focus on several related methods.

Temporal coding is one of the earliest and most commonly used structured light methods. A series of patterns are projected onto the surface of the scenes or objects, and each pixel is coded by a sequence of illumination values across the projected patterns. A number of temporal coding patterns have been generated, which includes binary codes [12] [15], n-ary codes that reduce the number of patterns required [5], gray codes [3] [10], and sinusoidal patterns [11]. Since the temporal coding techniques can achieve high accuracy, it has been used to generate the ground truth for depth maps [22]. In the experiments, a set of gray code patterns or sine waves are projected onto the scene and two cameras are used to capture the images. The projected intensities give each pixel a unique label which is distinguishable from others. Therefore, it is not difficult to establish the correspondence for every single pixel, and hence high-accuracy stereo depth maps can be generated. In fact, most ground truth of depth maps in the standard stereo evaluation website [14] are generated using this method. Despite their high accuracy and density of the reconstruction results, temporal coding techniques require temporal continuity of the captured scenes.

Spatial coding has become popular lately since it requires only one projected pattern. Therefore, it is possible to be applied to dynamic scenes. The spatial coding techniques encode a spatial location using a pattern that varies spatially instead of temporally. There are various kinds of patterns, e.g. the monochromatic pattern [1], a stripe pattern [25] which is generated based on the De Bruijn sequence. The method detects the edges in the projected pattern as well as in the captured image, and then a multi-pass dynamic programming is developed for matching the edges of the projected pattern with that of the captured image. It is common that a stripe pattern is used [8] [15] [16]. All of these techniques apply an image processing step to detect edges from the captured image which could be difficult and hence the robustness of the techniques is reduced. Grid patterns are also often used [13] [19], which again require an image processing step to detect the grid. Moreover, Young *et al*. [24] develop a theoretical framework to replace time-coded structured light patterns with viewpoint codes. In particular, the method projects only one pattern but uses several cameras to view the scenes or objects from different viewpoints to achieve the same effect as that of using temporal patterns. In general, spatial coding methods are not normally applied to recover depth maps since the established correspondence is only along the edges or the grid corners, which results in a sparse set of correspondence.

Spacetime stereo [7] [26] is a framework developed in order to combine the advantages of both temporal and spatial coding. It reduces the number of patterns required and improves the accuracy of spatial coding by incorporating temporal information. As pointed out in [26], this is a general framework so that almost any existing stereo methods can be extended in such manner.

Our method is a major extension of a standard spatial coding technique using De Bruijn pattern. It requires no temporal information which is the main benefit of spatial coding techniques, and it produces dense correspondence which is the main advantage of temporal coding methods.

3. Proposed Method

The pattern used in our method is introduced first in this section, and following that the method to establish the correspondence is presented.

3.1. Gaussian Blurred De Bruijn Pattern

The pattern used by our method is based on a color stripe pattern generated by Zhang *et al*. [25], which is shown in Figure 1. The pattern consists of 125 vertical

Figure 1: Color stripe pattern used in Zhang *et al*. [25].

stripes using eight different colors. It is generated by the generator that is publically available online [18] and is based on the De Bruijn sequence [9]. A De Bruijn sequence of order o over an alphabet of a symbols is a cyclic string of length a^o that contains each substring of length o exactly once. An example of De Bruijn sequence is $S = 00101110$ which has an order $o = 3$. It means that any substring of S with length 3 appears exactly once. A De Bruijn pattern has a good *windowed uniqueness* property. Because of such property, global optimization methods such as dynamic programming (DP) can be applied to establish the correspondence.

In our method, Gaussian blur is applied to the pattern shown in Figure 1, which results in a blurred pattern shown in Figure 2. We made an important observation that

Figure 2: Blurred pattern after Gaussian blur is applied to the pattern shown in Figure 1.

after applying Gaussian blur, the resulting pattern has the windowed uniqueness property as long as the original pattern has that property. Such observation has never been studied or proved before. Based on this observation, we show that a blurred De Bruijn pattern can be used in a structured light system, which can produce a denser set of correspondence.

The windowed uniqueness property can be described as follows. In the pattern without applying Gaussian blur, a set of n contiguous pixels appear exactly once. Suppose

there are two sets of n contiguous pixels that are different from each other, then there exists at least one color channel that is different and we only focus on that channel. Denote the pixel value of that channel to be P and Q for the two sets. To be more specific, $P = p_1 p_2 \dots p_n$ and $Q = q_1 q_2 \dots q_n$. $P \neq Q$ represents that there exists one or more i such that $p_i \neq q_i$ where $i \in \{1, 2, \dots, n\}$. After applying Gaussian blur to P and Q, the results are denoted as $P' = p'_1 p'_2 \dots p'_n$ and $Q' = q'_1 q'_2 \dots q'_n$, respectively. We need to prove that $P \neq Q \rightarrow P' \neq Q'$. It states that if two sets of n contiguous pixels are different, then the results of applying Gaussian blur to them are different, which means that the result has the windowed uniqueness property.

Theorem. Given the above assumptions, $P \neq Q \rightarrow P' \neq Q'$.

Proof. According to the definition of Gaussian blur, p'_i is the weighted sum of the neighboring pixels of p_i. We denote the neighboring pixels of p_i as a vector $\boldsymbol{\mathcal{N}_{p_i}} = \{p_{i-m}, \dots, p_{i-1}, p_i, p_{i+1}, \dots, p_{i+m}\}$. Similarly, $\boldsymbol{\mathcal{N}_{q_i}} = \{q_{i-m}, \dots, q_{i-1}, q_i, q_{i+1}, \dots, q_{i+m}\}$ is the vector denoting the neighboring pixels of q_i. Since $P \neq Q$, we can always find an i such that $\boldsymbol{\mathcal{N}_{p_i}} \neq \boldsymbol{\mathcal{N}_{q_i}}$. We divide the relationship between $\boldsymbol{\mathcal{N}_{p_i}}$ and $\boldsymbol{\mathcal{N}_{q_i}}$ into two scenarios.

1) $\boldsymbol{\mathcal{N}_{p_i}}$ is the mirror of $\boldsymbol{\mathcal{N}_{q_i}}$ and $\boldsymbol{\mathcal{N}_{p_i}} \neq \boldsymbol{\mathcal{N}_{q_i}}$. In this case $p_{i-m} = q_{i+m}, \dots, p_{i-1} = q_{i+1}$, $p_i = q_i$, $p_{i+1} = q_{i-1}$, $\dots$, $p_{i+m} = q_{i-m}$.

2) All the other possibilities where $\boldsymbol{\mathcal{N}_{p_i}} \neq \boldsymbol{\mathcal{N}_{q_i}}$.

● We prove the second scenarios first. We denote the weight of the Gaussian kernel as a vector $\boldsymbol{w}$. Then $p'_i = \boldsymbol{w^T \mathcal{N}_{p_i}}$ and $q'_i = \boldsymbol{w^T \mathcal{N}_{q_i}}$ where $\boldsymbol{w^T}$ is the transpose of $\boldsymbol{w}$. We need to prove that $\boldsymbol{\mathcal{N}_{p_i}} \neq \boldsymbol{\mathcal{N}_{q_i}} \rightarrow p'_i \neq q'_i$. The reason is that if there exists an i where $p'_i \neq q'_i$, then $P' \neq Q'$.

$\boldsymbol{w^T \mathcal{N}_{p_i}}$ and $\boldsymbol{w^T \mathcal{N}_{q_i}}$ can be written as:

$$\boldsymbol{w^T \mathcal{N}_{p_i}} = w_{i-m} p_{i-m} + \dots + w_i p_i + \dots + w_{i+m} p_{i+m} \quad (1)$$

$$\boldsymbol{w^T \mathcal{N}_{q_i}} = w_{i-m} q_{i-m} + \dots + w_i q_i + \dots + w_{i+m} q_{i+m} \quad (2)$$

Notice that there could exist some p_j and q_j such that $p_j = q_j$ and $w_j p_j = w_j q_j$. All the terms with the same $w_j p_j$ and $w_j q_j$ are cancelled once the following operation is performed: $\boldsymbol{w^T \mathcal{N}_{p_i}} - \boldsymbol{w^T \mathcal{N}_{q_i}}$. Since we assume that $\boldsymbol{\mathcal{N}_{p_i}} \neq \boldsymbol{\mathcal{N}_{q_i}}$, we can always find an i such that $p_i \neq q_i$. Therefore the equation can be re-written as in the following equation.

$$p'_i - q'_i = \boldsymbol{w^T \mathcal{N}_{p_i}} - \boldsymbol{w^T \mathcal{N}_{q_i}} =$$

$$\frac{1}{\sqrt{2\pi\sigma^2}} \left(c_0 e^{-\frac{k_0^2}{2\sigma^2}} + c_1 e^{-\frac{k_1^2}{2\sigma^2}} + \dots + c_l e^{-\frac{k_l^2}{2\sigma^2}} \right) \quad (3)$$

Here k_i denotes all the remaining indices that are not cancelled. None of the items in (3) is 0, i.e. $c_0 \neq 0, c_1 \neq 0, \dots, c_l \neq 0$. Denote $a = e^{-\frac{1}{2\sigma^2}}$. Assume that σ is a rational number. Since e is a transcendental number, it follows that a is also transcendental. On the other hand, (3) can be re-written as:

$$p'_i - q'_i = \boldsymbol{w^T \mathcal{N}_{p_i}} - \boldsymbol{w^T \mathcal{N}_{q_i}} =$$

$$\frac{1}{\sqrt{2\pi\sigma^2}} \left(c_0 a^{k_0^2} + c_1 a^{k_1^2} + \dots + c_l a^{k_l^2} \right) \quad (4)$$

Since $c_0 \neq 0, c_1 \neq 0, \dots, c_l \neq 0$, we conclude that (4), which is $p'_i - q'_i$, can never equal to 0. It can be seen by Reductio ad absurdum, if $p'_i - q'_i = 0$, then a is algebraic, which is clearly a contradiction. Therefore, $\boldsymbol{\mathcal{N}_{p_i}} \neq \boldsymbol{\mathcal{N}_{q_i}} \rightarrow p'_i \neq q'_i$.

● Next we prove the first scenario when $\boldsymbol{\mathcal{N}_{p_i}}$ is the mirror of $\boldsymbol{\mathcal{N}_{q_i}}$ and $\boldsymbol{\mathcal{N}_{p_i}} \neq \boldsymbol{\mathcal{N}_{q_i}}$. Consider (1) and (2), the weight of the Gaussian kernel is symmetric with respect to its center. That is, $w_{i-m} = w_{i+m}, \dots, w_{i-1} = w_{i+1}$. Since $\boldsymbol{\mathcal{N}_{p_i}}$ is the mirror of $\boldsymbol{\mathcal{N}_{q_i}}$, which means that $p_{i-m} = q_{i+m}, \dots, p_{i-1} = q_{i+1}$, $p_i = q_i$, we can easily see that $\boldsymbol{w^T \mathcal{N}_{p_i}} = \boldsymbol{w^T \mathcal{N}_{q_i}}$. This is the only case when $\boldsymbol{\mathcal{N}_{p_i}} \neq \boldsymbol{\mathcal{N}_{q_i}}$ yet $\boldsymbol{w^T \mathcal{N}_{p_i}} = \boldsymbol{w^T \mathcal{N}_{q_i}}$ because $\boldsymbol{w^T \mathcal{N}_{p_i}} - \boldsymbol{w^T \mathcal{N}_{q_i}} = 0$, which means that $p'_i = q'_i$. In order to prove the theorem, we need to show that if $\boldsymbol{\mathcal{N}_{p_i}}$ is the mirror of $\boldsymbol{\mathcal{N}_{q_i}}$, then $\boldsymbol{\mathcal{N}_{p_{i+1}}}$ is not the mirror of $\boldsymbol{\mathcal{N}_{q_{i+1}}}$. If that is true, the relationship between $\boldsymbol{\mathcal{N}_{p_{i+1}}}$ and $\boldsymbol{\mathcal{N}_{q_{i+1}}}$ belongs to the second scenario which we have proved already.

The basic procedure of proving that $\boldsymbol{\mathcal{N}_{p_{i+1}}}$ is not the mirror of $\boldsymbol{\mathcal{N}_{q_{i+1}}}$ is by using Reductio ad absurdum and can be described as follows. From the definition of $\boldsymbol{\mathcal{N}_{p_i}}$ and $\boldsymbol{\mathcal{N}_{q_i}}$, we have:

$$\boldsymbol{\mathcal{N}_{p_{i+1}}} = p_{i-m+1} \dots p_i p_{i+1} \dots p_{i+m+1}$$

$$\boldsymbol{\mathcal{N}_{q_{i+1}}} = q_{i-m+1} \dots q_i q_{i+1} \dots q_{i+m+1}$$

Suppose $\boldsymbol{\mathcal{N}_{p_{i+1}}}$ is the mirror of $\boldsymbol{\mathcal{N}_{q_{i+1}}}$, then we can prove that $\boldsymbol{\mathcal{N}_{q_i}}$ is a mirror of itself. Since $\boldsymbol{\mathcal{N}_{p_i}}$ is the mirror of $\boldsymbol{\mathcal{N}_{q_i}}$, we conclude that $\boldsymbol{\mathcal{N}_{p_i}} = \boldsymbol{\mathcal{N}_{q_i}}$, which is clearly a contradiction to our assumption.

Suppose $\boldsymbol{\mathcal{N}_{p_{i+1}}}$ is the mirror of $\boldsymbol{\mathcal{N}_{q_{i+1}}}$, we infer that:

$$p_{i-m+1} = q_{i+m+1}, p_{i-m+2} = q_{i+m}, \dots, p_i = q_{i+2},$$

$$p_{i+1} = q_{i+1}, \dots, p_{i+m} = q_{i-m+2}, p_{i+m+1} = q_{i-m+1} \quad (5)$$

We can see that p_{i+1} and q_{i+1} are the centers of $\boldsymbol{\mathcal{N}_{p_{i+1}}}$ and $\boldsymbol{\mathcal{N}_{q_{i+1}}}$, respectively. Since $\boldsymbol{\mathcal{N}_{p_i}}$ is the mirror of $\boldsymbol{\mathcal{N}_{q_i}}$, then:

$$p_{i-m+1} = q_{i+m-1}, p_{i-m+2} = q_{i+m-2}, \dots, p_i = q_i,$$

$$p_{i+1} = q_{i-1}, \dots, p_{i+m-1} = q_{i-m+1}, p_{i+m} = q_{i-m} \quad (6)$$

Putting (5) and (6) together, $\boldsymbol{\mathcal{N}_{q_i}}$ has the following property.

$$q_{i+m} = q_{i+m-2}, q_{i+m-1} = q_{i+m-3}, q_{i+m-2} = q_{i+m-4}, \cdots$$

$$q_{i+3} = q_{i+1}, q_{i+2} = q_i, q_{i+1} = q_{i-1}, q_i = q_{i-2}, \cdots$$

$$q_{i-m+4} = q_{i-m+2}, q_{i-m+3} = q_{i-m+1}, q_{i-m+2} = q_{i-m}$$

The above property can be re-written as follows.

$$q_{i+m} = q_{i+m-2} = q_{i+m-4} = \cdots = q_{i-m-4} = q_{i-m-2} = q_{i-m}$$

$$q_{i+m-1} = q_{i+m-3} = q_{i+m-5} = \cdots q_{i-m+5} = q_{i-m+3} = q_{i-m+1}$$

This property indicates that $\boldsymbol{\mathcal{N}_{q_i}}$ is a set of numbers that is a repetition of its first two numbers q_{i-m} and q_{i-m+1}, and the last number of $\boldsymbol{\mathcal{N}_{q_i}}$, which is q_{i+m}, is the same as its first number q_{i-m}. That is, $\boldsymbol{\mathcal{N}_{q_i}}$ has the pattern $xyxyxy \ldots xyxyx$ and its length is $2m + 1$, where x and y represent the two repeated numbers. Therefore, the vector $\boldsymbol{\mathcal{N}_{q_i}}$ is symmetric with respect to its center, which means that $\boldsymbol{\mathcal{N}_{q_i}}$ is a mirror of itself. Since $\boldsymbol{\mathcal{N}_{p_i}}$ is the mirror of $\boldsymbol{\mathcal{N}_{q_i}}$, hence $\boldsymbol{\mathcal{N}_{p_i}} = \boldsymbol{\mathcal{N}_{q_i}}$, which contradicts our assumption that $\boldsymbol{\mathcal{N}_{p_i}} \neq \boldsymbol{\mathcal{N}_{q_i}}$.

Therefore, $\boldsymbol{\mathcal{N}_{p_{i+1}}}$ cannot be the mirror of $\boldsymbol{\mathcal{N}_{q_{i+1}}}$. And this belongs to our second case, which has already been proved that $p'_{i+1} = \boldsymbol{w^T \mathcal{N}_{p_{i+1}}} \neq \boldsymbol{w^T \mathcal{N}_{q_{i+1}}} = q'_{i+1}$. Therefore, $P' \neq Q'$. ■

This property of a Gaussian blurred De Bruijn (GBDB) pattern has never been studied to our best knowledge. It states that a GBDB pattern has the windowed uniqueness property as well so that global minimization methods can be applied to establish correspondence.

3.2. Decoding

While could be extended to multi-view stereo, our method currently focuses on two-view stereo. The decoding process is to establish the correspondence between the two captured images. The above theorem states that every set of n contiguous pixels in a GBDB pattern appears only once. Based on that, we can use DP to establish the correspondence. One assumption here is that the two captured images are under similar illumination. This can be easily achieved since a projector is used as an active illumination source.

Traditional spatial coding methods normally have two steps. The first is to detect edges from the captured images and the projected pattern. The second is to match the detected edges using edge gradients. Therefore, these methods only recover matching pixels along edges, which are very sparse. As a result, these methods are rarely used in depth map recovery. One major advantage of our method is that it provides a dense set of correspondence. We have discovered that although pixels belong to the same color stripe have the same color in the pattern, their color can be different when the pattern is blurred and projected onto an object. This can be seen in Figure 3, which is a cropped image from one of our experimental datasets. We can see that surface points that are

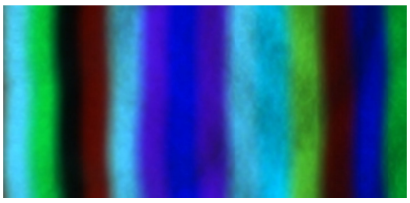


Figure 3: A sub-region of a captured image in our experiment.

illuminated by the same color stripe are likely to have different colors. Combining this observation with the windowed uniqueness property of a GBDB pattern, we can apply DP to establish correspondence for almost every illuminated pixel that is not occluded.

Before the decoding process, we assume that the captured images are rectified so that the correspondence can be established one scanline at a time. DP is applied in our method in order to find the matching and the process is similar to the one-pass DP described in [25]. However, instead of matching the gradient of the detected edges, our method matches the color of every illuminated pixel along each scanline. This is valid due to two reasons. First, the windowed uniqueness property of a GBDB pattern guarantees that a set of contiguous pixels along each scanline appears only once. Second, the color within one color stripe will be different in the captured images. The scoring function used in our method is different from that of [25] and is defined as follows.

$$score(I_1, I_2) = 1 - \frac{\sum_{k=R,G,B} \left(\frac{I_1(k) - I_2(k)}{255} \right)^2}{3} \qquad (7)$$

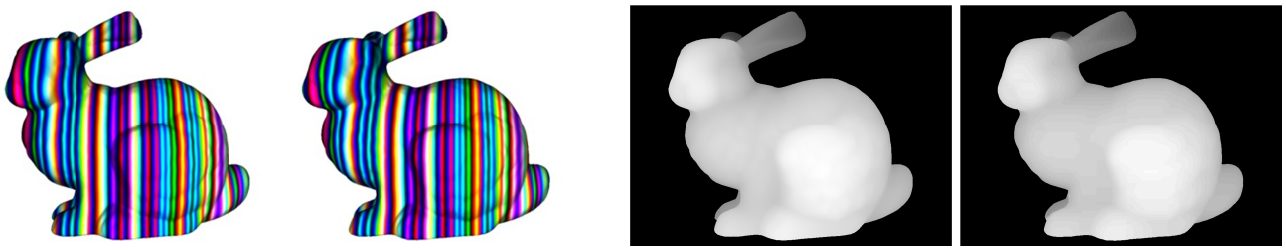
Here I_1, I_2 are the colors of two pixels from the two captured images along the same scanline. The scoring function measures the similarity between two colors. It normalizes the colors so that it is invariant to the magnitude of the colors. The pseudo-code for establishing correspondence can be found in [6].

4. Experimental Results

We apply our method to several simulated datasets to demonstrate its correctness, and to provide quantitative evaluation. Besides that, we apply our method to many real-world datasets including static and dynamic scenes to demonstrate the robustness, and to provide qualitative evaluation.

4.1. Simulated Experiments

We download the Stanford Bunny model and use Maya to project the GBDB pattern onto the bunny. The un-blurred pattern is shown in Figure 1, whose order $o = 3$. Then two virtual cameras are used to capture the scene from different viewpoints. The captured images are shown in Figure 4(a). The blurred pattern used is the result of applying a Gaussian blur with kernel size of 15*15. In the captured images, we can see that the edges of the color



(a). Captured images by using Maya (b). Simulated ground truth (c). Result by our method

Figure 4: Result of simulated experiment. (a). Captured images from two different viewpoints by using virtual cameras in Maya. (b). Simulated ground truth of the depth map from one viewpoint by Maya. (c). Depth map recovered by applying our method.

stripes in the pattern are blurred. The ground truth of the depth map from one viewpoint is shown in Figure 4(b) and is obtained using Maya. Figure 4(c) shows the result when applying our method. Most part of the result is correct except where occlusions are presented. As we know, the occluded parts cannot be recovered explicitly. Therefore, we apply a simple interpolation technique to fill in those holes due to the occlusion.

In order to evaluate our method quantitatively, we use two measurement metrics.

1. Percentage: An important advantage of our method is to provide dense correspondence while using only one-shot. Suppose the number of pixels whose depth are recovered by our method is M, and the number of pixels whose depth value are valid in simulated ground truth is N, then the percentage is defined to be $\frac{M}{N} * 100\%$.

2. Normalized RMS (NRMS): Another advantage of our method is the accuracy. In order to measure that, we apply an error metric that is similar to the one applied in stereo vision [23]. For every pixel whose disparity is valid in the simulated ground truth, the disparity is normalized and denoted as $d_T(x, y)$. The recovered depth by our method for a certain pixel is denoted as $d_C(x, y)$ which is also normalized. Then the normalized RMS (root-mean-squared) error is defined as follows.

$$NRMS = \sqrt{\frac{1}{N}\sum_{(x,y)} |d_C(x,y) - d_T(x,y)|^2}$$

Comparison with Temporal Coding

Figure 4(c) shows the recovered depth map by our method. And the percentage is 96.39% with NRMS 0.011. To better understand the numbers, we simulate another experiment. This experiment is the same as the one shown in [22], which serves as an experiment to obtain the ground truth depth maps for the datasets used in [14]. It is a temporal coding technique and can be briefly described as follows. A set of 40 code patterns are projected onto the scene and the codewords for each pixel from the captured images are obtained so that the correspondence can be established via codewords. Applying this method, every pixel is able to find its correspondence except if it is occluded. This experiment is employed in order to find the percentage of the occluded pixels from the captured images. The percentage of pixels that are not occluded is 98.11%. Therefore, only a small part (1.72%) of the un-occluded pixels is not recovered by our method. The NRMS for the temporal coding method is 0.0092. Comparing to this number, the accuracy of our method is close, and our method requires projecting only one pattern instead of 40 in the temporal coding method.

Robustness Test

We generate more datasets using Maya to demonstrate that our method is robust. To be more specific, we generate datasets by applying Gaussian blur to Figure 1 with different kernel size. The Gaussian kernel size varies from 3*3 to 39*39. Besides, the width of each color stripe in Figure 1 is 7 pixels, and we generate more datasets with width of 8 pixels and 9 pixels. We plot the graph of both percentage and NRMS and show them in Figure 5. We

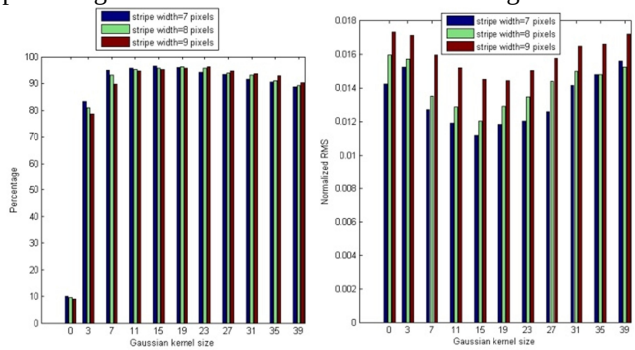


Figure 5: Left: percentage graph by our method with Gaussian kernels of different sizes. Right: Normalized RMS graph.

implement a traditional spatial coding method and its percentage and error are shown at the point when the kernel size is 0. Since the traditional spatial coding method requires edge detection, we only apply it to the dataset when projecting a non-blurred pattern. We expect a U-shape or V-shape bar graph for the error graph, in which case the accuracy first improves and then degrades as the kernel size increases. The percentage graph is expected to be inverted U-shape or inverted V-shape. The shape of the graph can be explained by the nature of Gaussian blur and DP. When the size of Gaussian kernel is small, the color within one color stripe will be similar in the captured images. Hence, the matching percentage is smaller and there exist ambiguities which cause larger error when

applying DP. When the kernel is increased to a certain size, the color within one color stripe varies significantly in the captured images. Therefore, more correspondence can be established and there are fewer ambiguities. However, when the size of Gaussian kernel keeps increasing, more neighboring pixels are mixed and hence more pixels appear similar again. Therefore, when the kernel size is too big, the number of matching pixels decreases and the error increases. Through this figure, we show that the density of the recovered depth increases dramatically by applying our method when comparing to the traditional spatial coding method. The best accuracy of the results is higher than that of applying traditional spatial coding method. We can see that the shapes of both graphs are similar to what we expected. All the graphs reach their peaks when the kernel size is around $((o-1)m+1)$, where o is the order of the De Bruijn pattern and m the width of each color stripe. In the original pattern, a set of $((o-1)m+1)$ consecutive pixels includes o colors and this set appears exactly once in the pattern. The experiment shows that the performance of our method is best when the Gaussian kernel is mixing exactly o colors of the original pattern.

4.2. Real-world Experiments

Our method is applied to many real-world datasets including static and dynamic scenes to demonstrate its robustness. In these experiments, the blurred pattern is not used anymore. Instead we apply a simpler idea which is to set the projector out of focus to achieve a similar effect of projecting a Gaussian blurred pattern. Indeed, it has been pointed out by [4] that when the projector is out-of-focus, the display is uniformly blurred by a PSF (point-spread function), which can be modeled as a 2D Gaussian. Therefore, when the projector is out-of-focus, it is similar to projecting a Gaussian blurred pattern. In the real world, projector may not be able to focus on every part of the object, in which case the traditional spatial coding methods would have trouble with detecting edges in the parts that are out of focus. It can also cause problem to temporal coding methods when identifying the codewords. We show that even if the projector is out of focus, our method can provide dense correspondence and accurate results. Throughout all the experiments, two PointGrey Flea2 CCD cameras are used to capture images and a Telex P600 LCD projector is used to project patterns. The resolution of all images is 1024*768. The projector allows users to change the focus to different levels. We adjust it purposely from level 1 to level 10 where level 1 is focused on the object and level 10 is severely out of focus. Figure 6 shows two captured images with the blurred level of the projector at 1 and 5. When the traditional spatial coding methods are applied to datasets such as the one with blur level at 5, the edge detection process could fail and result in low accuracy.

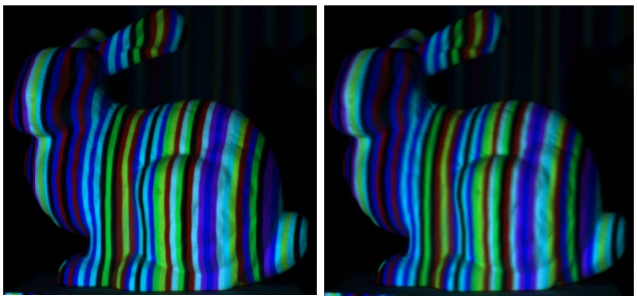
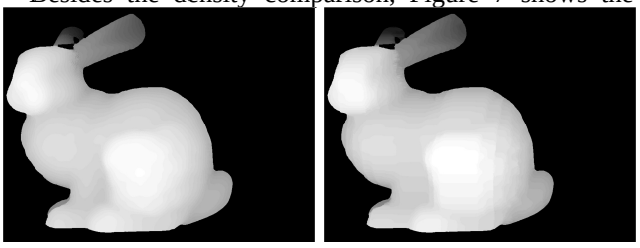


Figure 6: Captured images with projector under different blur level. Left: blur level = 1. Right: blur level = 5.

Static Scenes. Our method is applied to three static scenes: Bunny, Lid and Toy. Similar to the simulated experiments, the number of pixels whose depth is recovered grows and then decreases as the blur level increases. Besides applying our method, a traditional spatial coding method [25] is also applied only when the projector is set focused on the object. However, the number for the traditional coding method is less than 10% of the number by applying our method. The reason is that most traditional coding methods only establish correspondence along the edges of captured images.

Besides the density comparison, Figure 7 shows the accuracy comparison between a temporal coding method and our method. The input is the Bunny dataset. We use a 3D printer to print out the Stanford Bunny model, and the printed out object is placed in front of the cameras and patterns are projected onto the object. Figure 7(a) shows the result by using the temporal coding method proposed in [22]. Our result is shown in Figure 7(b). Although our method cannot produce exactly the same result as applying the temporal method, it is very comparable.

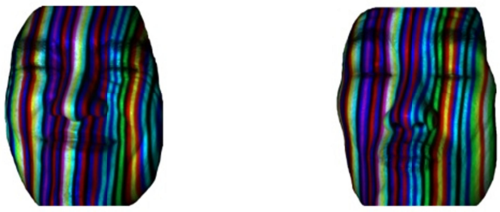


(a) (b)

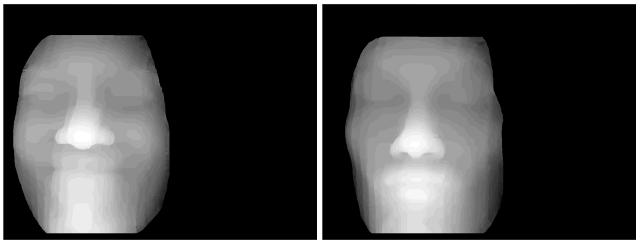
Figure 7: Accuracy comparison between our method and a temporal coding method. See text for more details.

Dynamic Scenes. Since our method requires no temporal information, it can be applied to dynamic scenes as well. The dynamic scenes in our experiments are Fist and Face. Since temporal coding methods cannot be applied to dynamic scenes, comparison cannot be provided. In these datasets, we mainly provide the recovered depth maps for qualitatively visual evaluation.

Figure 8 shows the results of applying our method to the Face dataset. The top row shows two continuous images with different facial expressions. We only show



(a). Masked images from one viewpoint



(b). Corresponding depth maps

Figure 8: Applying our method to a dynamic scene.

images from one viewpoint and the regions that include human hair or are not illuminated are masked out. Figure 8(b) shows the recovered depth maps. Although we cannot measure the accuracy quantatively, we can see that they are visually pleasant.

5. Discussion

First we compare our method to traditional temporal and spatial coding methods. Comparing to the temporal coding methods, ours is not limited to static scenes, and still maintains similar accuracy as a temporal method which has been demonstrated in simulated experiments. Comparing to the spatial coding methods, ours can produce dense depth maps, while the former only recover depth along edges.

Secondly, we justify the usage of a De Bruijn pattern. A De Bruijn pattern is used because of its windowed uniqueness property so that global minimization can be applied. In this experiment, we generate a pattern which is not a De Bruijn pattern, and then redo the simulated experiment. The pattern is generated by repeating a set of eight color stripes, and the set is {black, red, green, blue, yellow, magenta, cyan, white}. The pattern is then blurred by applying Gaussian blur. The applied pattern is shown in Figure 9. We show that the result of applying this pattern

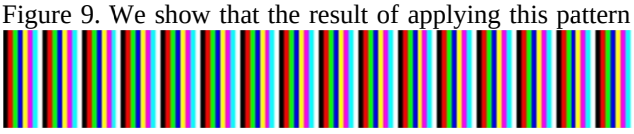


Figure 9: Blurred repeated pattern

is worse than that of applying a De Bruijn pattern.

The computed depth map is shown in Figure 10, and the ground truth is the one shown in Figure 4(b). We can see that the result is not even correct anymore. Therefore, taking advantage of windowed uniquess property is crucial

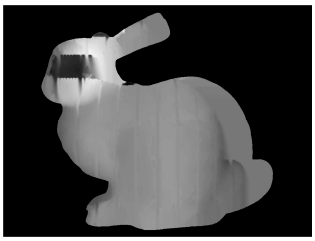


Figure 10: Computed depth map by applying our method with the pattern shown in Figure 9.

to our method.

Another major contribution is that of using a Gaussian blurred pattern and there are four reasons. First, we made an important observation that a GBDB pattern also maintains the windowed uniqueness property, which is proved in Section 3.1, so that DP can be applied to established correspondence. Second, with the GBDB pattern, the color information of the illuminated pixels is fully used to provide much denser correspondence comparing to the traditional spatial coding methods. It has been illustrated in both simulated and real-world experiments. Third, traditional spatial coding methods usually require the detection of edges in the captured image. This step may not be robust enough due to the image processing techniques that are applied. However, our method does not require edge detection, and hence increases the robustness. Last but not least, since projectors normally have large apertures, thus the in-focus region is very limited. Therefore, the patterns can be blurred when projected onto the scene if the depth range of the scene is big. In this case, the edge detection in the traditional spatial coding method would fail. However, our method can still establish correspondence since it works with blurred pattern.

We design one final experiment to compare our method with other methods. The traditional spatial coding methods are edge-based while ours is intensity-based. Moreover, the pattern can be blurred or non-blurred. Therefore, there are four possible combinations. We apply these four combinations to simulated datasets. In particular, a non-blurred pattern and a Gaussian blurred pattern with kernel size of 7*7 are used. Quantitative evaluation is provided and the percentage and NRMS are used as measurement metrics. Table 1 shows the evaluation results. From the table we can see that using intensity-based methods provides better results constantly. Moreover, the combination of intensity-based with blurred pattern provides the best result.

	Blurred pattern	Non-blurred pattern
Edge-based	11.34% 0.03207	13.44% 0.02822
Intensity-based	95.01% 0.01069	76.00% 0.01702

Table 1: Evaluation of four possible methods.

6. Conclusion and Future Work

We have presented a depth recovery method which requires only one shot and yet provides accurate depth map whose density is much higher than traditional spatial coding methods. Moreover, the projector needs not be focused on the objects, which is more realistic in real world applications.

To our best knowledge, the proposed method is the first one that requires absolutely no temporal information, and yet provides dense recovery that are close to applying temporal coding methods. We believe that with a proper design of the color pattern and the configuration of the cameras, this method could be used to recover the ground truth of the depth of a scene, which achieves the same accuracy of temporal coding methods [22]. The limitation of [22] is that it uses a series of patterns and therefore cannot be applied to dynamic scenes. However, our method has no such limitation.

An interesting future direction is to investigate on the blurred De Bruijn pattern. We believe that it could be useful in some other areas if they are properly applied.

References

[1] C. Albitar, P. Graebling, and C. Doignon. Robust structured light coding for 3D reconstruction. *Proceedings of International Conference of Computer Vision*. Rio de Janeiro, Brazil, pp. 1-6, 2007.

[2] J. Batlle, E. Mouaddib, and J. Salvi. Recent progress in coded structured light as a technique to solve the correspondence problem: A survey. *Pattern Recognition*, 31(7), pp. 963-982, 1998.

[3] D. Bergmann. New approach for automatic surface reconstruction with coded light. *T.F. Schenck(Ed.), Proceedings of Remote Sensing and Reconstruction for Three-Dimensional Objects and Scenes*. Vol. 2572, SPIE, San Diego, CA, pp. 2-9, 1995.

[4] M. Brown, P. Song and T. Cham. Image pre-conditioning for out-of-focus projector blur. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. vol. 2, Washington DC, USA, 2006.

[5] D. Caspi, N. Kiryati, and J. Shamir. Range imaging with adaptive color structured light. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(5), 1998.

[6] I. J. Cox, S. L. Hingorani, S. B. Rao, and B. M. Maggs. A maximum likelihood stereo algorithm. *Computer Vision and Image Understanding*, 63(3), pp. 542-567, 1996.

[7] J. Davis, R. Ramamoorthi, and S. Rusinkiewicz. Spacetime stereo: a unifying framework for depth from triangulation. . *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. vol. 2, pp. 359-366. Madison, WI, 2003.

[8] P. Fechteler, P. Eisert. Adaptive color classification for structured light systems. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. pp. 1-7, 2008.

[9] H. Fredricksen. The lexicographically least debruijn cycle. *Journal of Combinatorial Theory*. 9(1), pp. 509-510, 1970.

[10] J. Guhring. Dense 3-D surface acquisition by structured light using off-the-shelf components. *Proceedings of Videometrics and Optical Methods for 3D Shape Measuring*, SPIE, vol. 4309, pp. 220-231, 2001.

[11] P. S. Huang, C. Zhang, and F.-P. Chiang. High-speed 3-D shape measurement based on digital fringe projection. *Optical Engineering*. 42(1), pp. 163-168, 2003.

[12] S. Inokuchi, K. Sato and F. Matsuda. Range imaging system for 3-D object recognition. *Proceedings of the International Conference on Pattern Recognition*. pp. 806-808, 1984.

[13] H. Kawasaki, R. Furukawa, R. Sagawa, and Y. Yagi. Dynamic scene shape reconstruction using a single structured light pattern. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. pp. 1-8. Anchorage, AK, 2008.

[14] Middlebury stereo vision page. http://vision.middlebury.edu/stereo/

[15] J. Pages, J. Salvi, and J. Forest. Optimized De Bruijn patterns for one-shot shape acquisition. Image and Video Computing. 23(8), pp. 707-720, 2005.

[16] J. Pages, J. Salvi, and J. Forest. A new optimised De Bruijn coding strategy for structured light patterns. *17th International Conference on Pattern Recognition*. vol. 4, pp. 284-287, Cambridge, 2004.

[17] J. L. Posdamer, M. D. Altschuler. Surface measurement by space-encoded projected beam systems. *Computer Graphics and Image Processing*. 18(1), pp. 1-17, 1982.

[18] F. Ruskey. *The Combinatorial Object Server*. http://www.theory.csc.uvic.ca/~cos/, 2000.

[19] R. Sagawa, Y. Ota, Y. Yagi, R. Furukawa, N. Asada, and H. Kawasaki. Dense 3D reconstruction method using a single pattern for fast moving object. *Proceedings of International Conference on Computer Vision*. Kyoto, Japan, 2009.

[20] J. Salvi, J. Pages, and J. Batlle. Pattern codification strategies in structured light systems. *Pattern Recognition*, 37(4), pp. 827-849, 2004.

[21] J. Salvi, S. Fenandez, T. Pribanic, X. Llado. A state of the art in structured light patterns for surface profilometry. *Pattern Recognition*. 43(8), pp. 2666-2680, 2010.

[22] D. Scharstein, and R. Szeliski. High-Accuracy stereo depth maps using structured light. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. vol. 1, pp. 195-202. Madison, WI, 2003.

[23] D. Scharstein, and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*. 47(1/2/3), pp. 7-42, 2002.

[24] M. Young, E. Beeson, J. Davis, S. Rusinkiewicz, and R. Ramamoorthi. Viewpoint-coded structured light. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. pp. 1-8. Minneapolis, MN, 2007.

[25] L. Zhang, B. Curless, and S. M. Seitz. Rapid shape acquisition using color structured light and multi-pass dynamic programming. *International Symposium on 3D Data Processing Visualization and Transmission*. Padova, Italy, pp. 24-26, 2002.

[26] L. Zhang, B. Curless, and S. M. Seitz. Spacetime stereo: shape recovery for dynamic scenes. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. vol. 1, Madison, WI, 2003.