On-line Dense 3D Reconstruction Method based on ANN using a Single Structured Light Pattern

Jinbo Liu*, Xiaoliang Sun2, Jinshui Wu1
1Hypervelocity Aerodynamics Institute, China Aerodynamics Research and Development Center, Mianyang 621000, China
2College of Aerospace Science and Engineering, National University of Defense Technology, Changsha 410000, China
Email: liujinbo2088@163.com, liujinbo@cardc.cn, alexander_sxl@nudt.edu.cn, wjscardc@126.com

Abstract: Dense 3D reconstruction techniques to measuring dynamic scenes and deformable objects with little texture have been widely studied for various applications like surface deformation measurement. We assume that the actual situation of scanning is acquiring sequential shape and measuring surface deformation on-line at the moment when an object deforms under the action of aerodynamic forces. This paper presents such a technique for on-line acquisition of 3D surface data based on a one-shot scanning method that reconstructs 3D shape from a single image where a simple color-coded pattern using de Bruijn sequence is projected onto an object. The proposed approach has advantages that it requires no assumption of global smoothness and continuous surface. To realize 3D reconstruction from a single image, there are several issues to be solved, for example, difficulty on decoding structured light pattern because of influence of chromatic aberration in both projection systems and imaging systems, and difficulty on establishing accurate correspondences between projector and camera pixels. This paper describes the solutions of the issues by combining two methods, that is (1) an efficient and robust pattern decoding method based on artificial neural networks (ANN), and (2) a sub-pixel matching method using phase map and multiple view geometry. Furthermore, we adopt some effective strategies to improve algorithm robustness. Practical experiments were carried out to test the accuracy and efficiency of the scanning system in typical configuration. The results show that the scanning system can reconstruct 3D shape on-line in high-resolution with the accuracy of 0.09 mm and efficiency of 10 fps.

1. Introduction
Dense and precise shape on-line acquisition of deformable objects with little texture is strongly required in wide fields [1]. For example, if changing surfaces of an object being deformed under the action of aerodynamic forces are acquired on-line, it is expected to make a great contribution to human facial expressions and body motions.

For acquiring dense and accurate 3D shape, many sorts of scanning systems which are based on time-of-flight lasers [2], laser scanning [3], stereovision [4] and pattern projection [5-9] have been developed.

Structured light is one of a variety of methods for acquiring 3D surface data of objects. Shape reconstruction techniques with a structured light system, which encodes positional information of a projector into temporal or special changes in a projected pattern, have been investigated in Ref. [10].
system using structured light with only temporal coding is easy to implement, accurate, dense and robust, so it has been widely used for real applications [11]. However, such systems are not suitable for dynamic scenes and surface deformation measurement because they usually need to capture more than three pattern images and can be degraded by disturbance, such as vibration, deformation and movement between gaps of image shot [12]. Recently, several techniques using only spatial encoding of a pattern were presented, called "one-shot scanning".

Zhang et al. present a color structured light technique for recovering object shape from one or more images. The technique works by projecting a pattern of stripes of alternating colors and matching the projected color transitions with observed edges in the image. The algorithm eliminates global smoothness assumptions and strict ordering constraints. One obvious disadvantage is that each range map takes about 1 minute to compute using a 900 MHz Pentium PC [14]. H. Kawasaki et al. propose a single scanning technique using a simple grid pattern formed by a number of straight lines distinguishable only as vertical or horizontal lines. Their method allows robust processing because it requires only local information of connectivity between adjacent grid points. However, the processing time for shape reconstruction on a PC with a CPU of 3.8 GHz is about 1.6s [17]. Ulusoy et al. propose a more stable solution with the help of special spaced grids, which are grid patterns with spacings that follow a de Bruijn sequence. But it is a sparse reconstruction algorithm and the running time for a thousand grid intersections is about three minutes [13, 22]. Ref. [23] presents a 3D scanning method from a single image using single-colored static pattern. The method utilizes topology information to achieve dense and precise shape reconstruction. The drawback is that its calculation time is around one minute to five minutes. All the methods above, called "off-line methods", require high computational costs, which cannot be used to measure surface deformation on-line.

Koninckx et al. propose a technique allowing dense shape reconstruction based on a single image using a simple pattern, i.e. a set of stripes. This was achieved by combining dense unidentified stripes and several identified stripes. Their pipeline can run at approximately 20 Hz. But the method depends on relative numbering of dense patterns, which may be disturbed by shape discontinuities and line detection failures [15]. Chen et al. present a principle of uniquely color-coded pattern projection. A color matrix has been designed for improving reconstruction efficiency. The matrix is produced by a special code sequence and a number of state transitions. It only takes about 100ms for shape reconstruction but in low-level resolution [16]. Sagawa et al. propose an efficient dense 3D reconstruction method using single-colored grid pattern which consists of sinusoidal curves [24]. With the special pattern, irregularity of pattern is increased and solution becomes stable. The method can realize on-line processing at about 10 fps but not able to reach the high accuracy required by the class of applications we are dealing with.

Each of the color pattern methods above has an additional constraint that the surface does not change the reflected color too much. For example, chromatic aberration may lead to failure shape reconstruction. Therefore, color recognition is of critical importance for stability of methods. In addition, most real-time methods typically face a trade-off between resolution and efficiency. Especially for on-line methods, many researchers adopt the strategy of reducing resolution to improve efficiency, which may be not suitable for surface deformation measurement. In this paper, we present a method for dense and precise shape on-line acquisition of deformable objects with little texture. The main contributions of the paper are as follows: (1) Dense and precise shape on-line acquisition of deformable objects with little texture from a single image using a simple color-coded pattern is realized, (2) an efficient and robust pattern decoding algorithm based machine learning to allow a quick identification of pattern using just 6 colors of de Bruijn sequence is presented, (3) an actual scanning system that can be used to measure surface deformation on-line is constructed to show the performance of the method.

2. Material and Methods

The proposed one-shot scanning system consists of an array CCD camera with a resolution of 1280×1024 pixels and a DLP data projector with a resolution of 854×480 pixels as shown in Figure 1-a. Using a de Bruijn sequence, we can generate color stripes that combine to make a sequence for
which each three consecutive stripes are unique. In practice, we only need 106 stripes and thus work with 6 different colors as shown in Figure 1-b. The projector pattern is fixed and does not change, so no synchronization is required. A simple colored-coded pattern is projected from the projector and captured by the camera. Procedures of 3D shape measurement are presented by a flowchart as shown in Figure 2.

![Figure 1](image1.png)

Figure 1. (a) Scanning system. (b) Projected pattern: each stripe’s band is 8 pixels.

![Figure 2](image2.png)

Figure 2. Flowchart of 3D shape measurement by a one-shot acquisition.

2.1. System Calibration

In order to guarantee the best result, structured light system must be calibrated before each use. Although projectors are modeled as inverse cameras in such systems, the calibration procedure must be adapted to the fact that projectors cannot directly measure the pixel coordinates of 3D points projected onto the projector image plane as camera do, which is similar to Ref. [18]. Solve distortion coefficients, intrinsic and extrinsic parameters of the camera and projector using Zhang’s stereo calibration algorithm [19].

Decoding color codes is to obtain unwrapped phase map from one captured pattern image, which is of critical importance for establishing accurate correspondences between projector and camera pixels. However, it is difficult to decode structured light pattern accurately because of influence of chromatic aberration in both projection systems and imaging systems. This Section presents an efficient and robust pattern decoding method based machine learning with advantages of requiring no assumption of global smoothness and continuous surface and supporting parallel operation. Procedures of decoding color codes are presented by a flowchart as shown in Figure 3. Each part will be introduced in detail below.

![Figure 3](image3.png)

Figure 3. Flowchart of decoding color codes
2.2. Image Preprocessing
Image noise caused by electronic error of components and external environmental disturbance is unavoidable. Since gray intensity of the pattern image changes continuously, mean filter with a window size of 3×3 pixels can be used to remove image noise and improve image quality without serious blurring effect as shown in Figure 4.

![Figure 4. Image preprocessing](image_url)

2.3. Color Identification
Color identification is of critical importance for stripe indexing. However, it’s difficult to get accurate stripe color identification results because of chromatic aberration caused by projection systems and imaging systems. Artificial neural networks (ANN) based on supervised learning can solve this problem well [20]. Multi-layer perceptron (MLP) is the most commonly used type of ANN as shown in Figure 5. The ability to identify color can be trained by labeled data. Figure 5 represents a 3-layer perceptron with 17 inputs, 7 outputs and the hidden layer including 40 neurons:

![Figure 5. A 3-layer perceptron consists of the input layer, output layer and hidden layers.](image_url)

(a) Input layer
Inputs are notated as \( x_i \) \((i = 1, 2, \ldots, 16, 17)\) representing the color feature of each pixel in the captured pattern image. \( x_1, x_2 \) and \( x_3 \) are equal to the values of red, green and blue channel of each pixel and scaled to fit 0 to 1 range. \( x_i \) \((i = 4, 5, \ldots, 16, 17)\) can be computed out according to \( x_1, x_2 \) and \( x_3 \).

(b) Hidden layer
Neurons of hidden layer are notated as \( z_h \) \((h = 1, 2, \ldots, 39, 40)\), which can be computed as:

\[
z_h = f \left( \sum_{i=1}^{17} w_{hi} x_i + w_{h0} \right)
\]

Where \( w_{h0} \) is the weight of bias term of input layer and \( w_{hi} \) \((i = 1, 2, \ldots, 16, 17)\) are the weights of other neurons of input layer, \( f \) is the activation function.

\[
f(x) = \frac{1 - e^{-x}}{1 + e^{-x}}
\]

The values retrieved from the input layer are summed up with certain weights, individual for each neuron, plus the bias term and then the sum is transformed using the activation function. In our MLP, all the neurons have the same activation functions.

(c) Output layer
Outputs are notated as \( y_j \) \((j = 0, 1, 2, 3, 4, 5, 6)\) representing the color identification results of each pixel in the captured pattern image, which can be computed as:

\[
y_j = f \left( \sum_{h=1}^{40} v_{jh} z_h + v_{j0} \right)
\]
Where $v_{j0}$ is the weight of bias term of hidden layer and $v_{jh}$ ($h = 1, 2, \ldots, 39, 40$) are the weights of other neurons of other neurons of hidden layer, $f$ is the same activation function as equation (1). Once all the weights are known, the network can be used to identify pixels’ color. The weights can be obtained through back propagation algorithm (BPROP) [21]. BPROP takes a training set, multiple input vectors with corresponding output vectors, and iteratively adjusts the weights to enable the network to give the desired response to the provided input vectors.

We divide the data set into 2 parts, one is taken as a training set to compute all the weights of the network and the other one is used to test the performance of the network. The experimental results show that the accuracy rate of the network is up to 99.83%.

2.4. Color Filtering

Although ANN has a high accuracy rate of color identification, there are still some pixels with a wrong color on the edge of stripes, which will certainly lead to decoding errors. In this section, a strategy for color filtering used to delete error points is presented. Error points always appear in the edge transitions between adjacent stripes. Row-scan method is used to get the error points as shown in Figure 6. For $i$-th row, detect each stripe’s edge in order and calculate the stripe’s width. If the width of the current stripe is less than half of the width of its adjacent stripe, the color of the current stripe will be set to black. Since the phase-unwrapping algorithm is based on the peak of stripe gray intensity, the color filtering will not reduce the precision of unwrapped phase map while ensuring the correct position of the stripe in de Bruijn sequence.

![Color filtering diagram]

Figure 6. The basic principle of color filtering

2.5. Peak Detection

The peak of stripe gray intensity is the basis of the phase-unwrapping algorithm. However, for the surface whose curvature changes greatly, small light spots often appear in the captured pattern image and influence the peak detection of stripe gray intensity. In this section, a strategy for accurate and efficient peak detection is presented as shown in Figure 7. The strategy can be executed as follows:

1. Transform the image of color identification into a binary image, where only the pixels matching the specified color are set to white while other pixels are set to black.
2. Use median filter with a window size of 3x3 pixels to process the binary image.
3. Search for the brightest pixels in each row of every stripe.
4. Calculate coordinates of the pixels obtained from step (3) using gray centroid method.
5. Establish peak lines according to the distance threshold.
6. Store all the lines for the stripe of the specified color.
7. Change the specified color and repeat step (1) ~ step (6).

2.6. Stripe Indexing

Stripe indexing is of critical importance for phase-unwrapping algorithms. As mentioned above, for a de Bruijn sequence, each three consecutive color stripes are unique, which is the key idea of stripe in-
dexing methods. In order to improve stripe indexing methods’ robustness, a strategy is presented in this section. Three scenarios that might occur during image processing to eliminate the assumption of continuous surface are shown in Figure 8.

![Figure 8](image_url)

Figure 8. (a) Stripes are continuous and color identification results are accurate. (b) Color identification results are accurate but Stripes are discontinuous. (c) Stripes are discontinuous and one stripe’s color identification result is wrong.

2.7. Phase Wrapping

Phase unwrapping is crucial here for establishing accurate correspondences between projector and camera pixels. Given the unwrapped phase map, we can obtain the mapping relationship between columns of the projected pattern and columns of the captured image. A phase retrieval method is presented in this section to calculate the absolute phase of each pixel in the captured image as shown in Figure 9.

Arbitrary two adjacent peak lines are notated as \( l_i \) and \( l_{i+1} \), where \( i \) and \( i+1 \) represent the positions of the corresponding stripes in the de Bruijn sequence. The phase retrieval method is implemented by the way of row-scanning. For \( j \)-th row, notate the points in \( l_i \) and \( l_{i+1} \) as \( p_i(x', y') \) and \( p_{i+1}(x', y') \). \( I(x, y) \) represents the intensity of the pixel \( p(x, y) \) in the image.

\[
I(x-2, y) \geq I(x-1, y) \geq I(x, y) \leq I(x+1, y) \leq I(x+2, y)
\]  

(4)

If a pixel satisfies the above constraint, notate the pixel as \( v_i(x'_i, y'_i) \) representing the valley of intensity between \( p_i \) and \( p_{i+1} \). Once the scan is complete, the unwrapped phase map can be obtained.

![Figure 9](image_url)

Figure 9. The basic principle of phase unwrapping.

2.8. Sub-pixel Matching

Image matching is an intermediate but crucial step inside the triangulation-based structured light approach.

From section 2.7, we can obtain the mapping relationship between columns of the projected pattern and columns of the captured image, which is a point-to-line correspondence. In this section, a sub-pixel matching method using unwrapped phase map and multiple view geometry is introduced to establish point-to-point correspondences between projector and camera pixels as shown in Figure 10. \( p_i(x, y) \) represents an arbitrary point in the captured pattern image. According to the unwrapped phase map obtained from section 2.7, the point \( p_i(x_p, y_p) \) corresponding to \( p_c \) must be on the line \( x = x_p \) in the projected pattern image. \( x_p \) can be calculated as follows:
\[ x_p = P(x_c, y_c) T \pi \]

(5)

Where \( P(x_c, y_c) \) represents the unwrapped phase of \( p_c(x_c, y_c) \), \( T \) represents the period of stripes.

Moreover, \( p_p \) must be in the epipolar line. Therefore, once the epipolar line is identified in the projected pattern image, the intersection of two straight lines can be regarded as the corresponding point \( p_p \).

Since the system calibration has been completed in section 2, the epipolar line can be calculated as follows:

\[ l_e = [a \ b \ c]^T = F [p_c \ 1]^T \]

(6)

\( F \) is fundamental matrix. Thus the coordinates of \( p_p \) can be written as follows:

\[ p_p = \left( x_p - \frac{ax_p - c}{b} \right) \]

(7)

Then, we can obtain all the corresponding points using equation (5)~(7). Based on optical triangulation, on-line dense 3D shape reconstruction can be achieved. In order to improve the precision, BA is considered [9].

3. Results

To confirm the advantages of our method, we conduct a dense 3D shape reconstruction using an object which is similar to paper but made of different materials. The goal is to acquire sequential shape and measure surface deformation on-line at the moment when the object deforms under the action of aerodynamic forces.

A scanning system that can be used to measure surface deformation on-line has been actually developed by using off-the-shelf components, a DLP data projector with a resolution of 854×480 pixels and an array CCD camera with a resolution of 1280×1024 pixels. Processing is performed by a PC with an Intel Core i7-6700 CPU @3.4 GHz. The projected pattern consists of de Bruijn generated sequences with 8 pixels wide stripes (approximately 106 stripes total).

As shown in Figure 11, several markers distributed in rectangular area are used to evaluate the precision of the proposed method. The distance between markers remains the same during the experiment, which can be used for surface registration. The object deforms under the action of aerodynamic forces. By comparing the reconstructed shape before and after deformation, the whole surface deformation of the object is calculated. \( d_i \) and \( d_i \) represent the true distance and the reconstructed distance between two markers.
The precision of the proposed method is defined as follows:

\[
err = \frac{1}{C_{18}} \sum_{i=1}^{C_{18}} (d_i - d) \]

(8)

Figure 12 shows the target object, the captured pattern image, the result of reconstruction and the surface deformation report at different states. The arrow represents the direction of aerodynamic forces. As is apparent, detailed shapes are successfully recovered with the proposed method. Table 1 displays the analysis of execution performance, which indicates that the proposed method is dense, accurate (0.09mm) and efficiency (100ms). That speed and precision are adequate for most applications. Moreover, the proposed method requires no manual process to remove many outliers. In the operation, only the process of color identification is made in parallel. So with good engineering skills for system implementation in both software and hardware, it will not be difficult to achieve the speed of about 20fps for the scanning system.

| States | Precision   | Actual time | Number of reconstructed points |
|--------|-------------|-------------|--------------------------------|
| (a)    | 0.084 mm    | 109 ms      | 1,220,007                      |
| (b)    | 0.095 mm    | 125 ms      | 1,221,974                      |
| (c)    | 0.102 mm    | 93 ms       | 1,157,778                      |

4. Conclusions
In this paper, a one-shot active stereo system that can reconstruct dynamic scenes and deformable objects with little texture on-line is proposed. In the proposed approach, a color-coded pattern is emitted from a video projector and the pattern projected onto the scene is captured by a camera. To decode the structured light pattern efficiently and robustly, the color of each pixel is identified based on machine learning. Moreover, some effective strategies are adopted to improve approach robustness. To establish accurate correspondences between projector and camera pixels, a sub-pixel matching method using unwrapped phase map and multiple view geometry is introduced. The proposed approach has advantages that it requires no assumption of global smoothness and continuous surface. Finally, practical experiments are carried out to test the accuracy and efficiency of the scanning system in typical con-
The results show that the performance of the scanning system is adequate for most applications.

Although this work corroborates that such systems are feasible, the current implementation remains to be improved in several ways: (1) Crosstalk may occur on the edge of stripes. To improve the accuracy of color identification, the adjacent stripes’ color should have clear difference with each other. (2) In the operation, only the process of color identification is made in parallel. Some software and hardware skills may be applied to further improve the speed to above 20 fps.

Acknowledgements
This research was financially supported by the National Natural Science Foundation of China (Grant No.11802321).

References
[1] Sagawa, R., Ota, Y., Yagi, Y., Furukawa, R., Asada, N. (2009) Dense 3D reconstruction method using a single pattern for fast moving object. In: ICCV. Kyoto. pp. 1779-1786.
[2] Ulrich, A., Studnicka, N., Riegler, J., Orlandini, S. (2002) Long-range high-performance time-offlight-based 3D imaging sensors. In: 3D Data Processing Visualization and Transmission. Padova. pp. 852-855.
[3] Forest, J., Salvi, J. (2002) An overview of laser slit 3D digitizers. In: International Conference on Robots and Systems. Hungary. pp. 73-78.
[4] Matabosch, C., Salvi, J., Forest, J. (2003) Stereo rig geometry determination by fundamental matrix decomposition. In: Workshop on European Scientific and Industrial Collaboration. Boston. pp. 405-412.
[5] Salvi, J., Pages, J., Batlle J. (2004) Pattern codification strategies in structured light systems. Pattern Recognition, 37 (4): 827-849.
[6] Zhang, L., Snavely, N., Curless, B., Seitz, S. M. (2004) Spacetime faces: High-resolution capture for modeling and animation. In: ACM Annual Conference on Computer Graphics. New York. pp. 548-558.
[7] Hall-Holt, O., Rusinkiewicz, S. (2001) Stripe boundary codes for real-time structured-light range scanning of moving objects. In: ICCV. Seoul. pp. 359-366.
[8] Davis, J., Nehab, D., Ramamoorthi, R., Rusinkiewicz, S. (2005) Spacetime stereo: A unifying framework for depth from triangulation. IEEE Transaction on Pattern Analysis and Machine Intelligence, 27(2): 296-302.
[9] Triggs, B., McIshauchlan, P., Hartley, R., Fitzgibbon, A. (1999) Bundle adjustment – a modern synthesis. In: Workshop on Vision Algorithms: Theory & Practice. Corfu. pp. 298-372.
[10] Batlle, J., Mouaddib, E., Salvi, J. (1998) Recent progress in coded structured light as a technique to solve the correspondence problem: a survey. Pattern Recognition, 31(7): 963-982.
[11] Inoukuchi, S., Sato, K., Matsuda, F. (1984) Range imaging system for 3-D object recognition. In: ICPR. Stockholm. pp. 806-808.
[12] Zhang, Z. (2012) Review of single-shot 3D shape measurement by phase calculation-based fringe projection techniques. Optics and Lasers in Engineering, 50: 1097-1106.
[13] Ulusoy, A. O., Calakli, F., Taubin, G. (2010) One-shot scanning using de Bruijn spaced grids. In: International Conference on Computer Vision Workshops. Kyoto. pp. 1786-1792.
[14] Li, Z., Curless, B., Seitz, S. (2002) Rapid shape acquisition using color structured light and multi-pass dynamic programming. In: International Symposium on 3D Data Processing Visualization & Transmission. Thessaloniki. pp. 24-36.
[15] Koninckx, T. P., Gool, L. V. (2003) Real-time range scanning of deformable surfaces by adaptively coded structured light. In: International Conference on 3-d Digital Imaging & Modelling. Alberta. pp. 293-300.
[16] Chen, S. Y., Li, Y. F., Zhang, J. W. (2008) Vision processing for realtime 3-D data acquisition based on coded structured light. IEEE Transaction on Image Processing, 17(2): 167-176.
[17] Kawasaki, H., Furukawa, R., Sagawa, R., Yagi, Y. (2008) Dynamic scene shape reconstruction using a single structured light pattern. In: CVPR. Alaska. pp. 1-8.
[18] Moreno, D., Taubin, G. (2012) Simple, accurate, and robust projector-camera calibration. In: International Conference on 3d Imaging. Belgium. pp. 464-471.
[19] Zhang, Z. Y. (2000) A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis & Machine Intelligence, 22(11): 1330-1334.
[20] Riedmiller, M. (1994) Advanced supervised learning in multi-layer perceptrons – From backpropagation to adaptive learning algorithms. Computer Standards & Interfaces, 16(3): 265-278.
[21] Ngolediage, J., Naguib, R., Dlay, S. (1993) Fast backpropagation for supervised learning. In: International Joint Conference on Neural Networks. Kangas. pp. 2591-2594.
[22] Ulusoy, A. O., Calakli, F., Taubin, G. (2010) Robust one-shot 3D scanning using loopy belief propagation. In: Computer Vision and Pattern Recognition Workshops. Kyoto. pp. 15-22.
[23] Kawasaki, H., Masuyama, H., Sagawa, R., Furukawa, R. (2012) Single color one-shot scan using topology information. In: ECCV. Florence. pp. 486-495.
[24] Sagawa, R., Sakashita, K., Kasuya, N., Kawasaki, H., Furukawa R. (2012) Grid-based active stereo with single-colored wave pattern for dense one-shot 3D scan. In: Second International Conference on 3D Imaging. Zurich. pp. 363-370.