PERGAMON

Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT

Pattern Recognition 37 (2004) 827–849

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

Pattern codification strategies in structured light systems

Joaquim Salvi*, Jordi Pagès, Joan Batlle

Institut d'Informàtica i Aplicacions, Universitat de Girona, Av. Lluís Santaló, s/n, E-17071 Girona, Spain

Received 6 March 2003; accepted 2 October 2003

Abstract

Coded structured light is considered one of the most reliable techniques for recovering the surface of objects. This technique is based on projecting a light pattern and viewing the illuminated scene from one or more points of view. Since the pattern is coded, correspondences between image points and points of the projected pattern can be easily found. The decoded points can be triangulated and 3D information is obtained. We present an overview of the existing techniques, as well as a new and definitive classification of patterns for structured light sensors. We have implemented a set of representative techniques in this field and present some comparative results. The advantages and constraints of the different patterns are also discussed.
© 2003 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.

Keywords: Coded patterns; Structured light; 3D measuring devices; Active stereo; Computer vision

1. Introduction

Surface reconstruction is one of the most important topics in computer vision due to its wide field of application. Some examples of applications are range sensoring, industrial inspection of manufactured parts, reverse engineering, object recognition and 3D map building. Amongst all the ranging techniques [1], stereovision is based on viewing the scene from two or more points of view and then finding correspondences between the different images in order to triangulate the 3D position. Triangulation is possible if cameras have been calibrated beforehand [2,3]. However, difficulties in finding the correspondences arise, even when taking into account epipolar constraints. Coded structured light consists of replacing one of the two cameras by a device that projects a light pattern onto the measuring surface. Today the most commonly used devices are LCD video projectors, although previously it was slide projectors. Such devices project an image with a certain structure so that a set of pixels are easily distinguishable by means of a local coding strategy. Therefore, when locating such coded points in the image grabbed by the remaining camera, the correspondence problem is solved with no requirement for geometrical constraints. The projecting images are called patterns, as they are globally structured.

This paper presents a comprehensive survey of coded structured light techniques, which updates the review presented in Ref. [4] and proposes a new, consistent and definitive classification. The paper focuses on the different coding strategies used in the bibliography and reproduces the experimental results of several techniques in order to evaluate and compare their accuracy and analyse their applicability. This article is structured as follows: firstly, the classification is presented in Section 2. Secondly, in Section 3, techniques based on projecting multiple patterns are explained. In Section 4, techniques exploiting the spatial neighbourhood paradigm are presented. Next, in Section 5, coding strategies using direct codification are also explained. In Section 6, the experimental results obtained with a set of implemented techniques are presented. Finally, in Section 7, the conclusion contains a discussion about the advantages and drawbacks of every subgroup of techniques. In addition, general guidelines for choosing the most suitable technique (given the specifications of an application) are proposed.

* Corresponding author. Tel.: +34-972-41-8483; fax: +34-972-41-8098.
E-mail addresses: qsalvi@eia.udg.es (J. Salvi), jpages@eia.udg.es (J. Pagès), jbatlle@eia.udg.es (J. Batlle).

0031-3203/$30.00 © 2003 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2003.10.002

Shenzhen Jimuyida
Exhibit 1009

2. A classification of coding strategies

A coded structured light system is based on the projection of a single pattern or a set of patterns onto the measuring scene which is then viewed by a single camera or a set of cameras. The patterns are specially designed so that codewords are assigned to a set of pixels. Every coded pixel has its own codeword, so there is a direct mapping from the codewords to the corresponding coordinates of the pixel in the pattern. The codewords are simply numbers, which are mapped in the pattern by using grey levels, colour or geometrical representations. The larger the number of points that must be coded, the larger the codewords are and, therefore, the mapping of such codewords to a pattern is more difficult. The aim of this work is to review the available strategies used to represent such codewords.

Pattern projection techniques differ in the way in which every point in the pattern is identified, i.e. what kind of codeword is used, and whether it encodes a single axis or two spatial axis. In reality, it is only necessary to encode a single axis, since a 3D point can be obtained by intersecting two lines (i.e. when both pattern axis are coded) or intersecting one line (the one which contains a pixel of the camera image) with a plane (i.e. when a single pattern axis is coded).

Table 1 shows pattern projection techniques classified according to their coding strategy: time-multiplexing, neighbourhood codification and direct codification. The seven columns on the right of the table indicate whether or not a given pattern is suitable for measuring moving objects, the colour depth used and whether repeated codewords appear (periodic codification) or not (absolute codification). Time-multiplexing techniques generate the codewords by projecting a sequence of patterns along time, so the structure of every pattern can be very simple. Furthermore, in spite of increasing the pattern complexity, neighbourhood codification represents the codewords in a unique pattern. Finally, direct codification techniques define a codeword for every pixel, which is equal to its grey level or colour.

In the following sections, each one of these three classifying groups are explained in detail. Moreover, the different techniques proposed in the bibliography which belong to each subgroup are introduced, and we will show the evolution from the simplest to the most complex technique.

3. Time-multiplexing strategy

One of the most commonly used strategies is based on temporal coding. In this case, a set of patterns are successively projected onto the measuring surface. The codeword for a given pixel is usually formed by the sequence of illumination values for that pixel across the projected patterns. Therefore, the codification is called *temporal* because the bits of the codewords are multiplexed in time. This kind of pattern can achieve high accuracy in the measurements. This is due to two factors: first, as multiple patterns are projected, the codeword basis tends to be small (usually binary) and therefore a small set of primitives is used, which are easily distinguishable among each other; secondly, a coarse-to-fine paradigm is followed, as the position of a pixel is encoded more precisely while the patterns are successively projected.

During the last 20 years several techniques based on time-multiplexing have appeared. We have classified these techniques as follows: (a) techniques based on binary codes: a sequence of binary patterns is used in order to generate binary codewords; (b) techniques based on n-ary codes: a basis of n primitives is used to generate the codewords; (c) Gray code combined with phase shifting: the same pattern is projected several times, shifting it in a certain direction in order to increase resolution; (d) hybrid techniques: a combination of time-multiplexing and neighbourhood strategies. The following sections describe in detail the techniques which can be included in such coding strategies.

3.1. Techniques based on binary codes

In these techniques only two illumination levels are commonly used, which are coded as 0 and 1. Every pixel of the pattern has its own codeword formed by the sequence of 0s and 1s corresponding to its value in every projected pattern. Therefore, a codeword is obtained only when the sequence is completed. An important characteristic of this technique is that only one of the two pattern axis is encoded.

In 1981, Posdamer and Altschuler [5] were the first to propose the projection of a sequence of m patterns to encode 2^m stripes using a plain binary code. Therefore, the codeword associated with each pixel is the sequence of 0s and 1s obtained from the m patterns, the first pattern being the one which contains the most significant bit. The symbol 0 corresponds to black intensity while 1 corresponds to full illuminated white. Therefore, the number of stripes increases by a factor of two for each consecutive pattern. Every stripe of the last pattern has its own binary codeword. The maximum number of patterns that can be projected is the resolution in pixels of the projector device. However, reaching this value is not recommended because the camera cannot always perceive such narrow stripes. It should be noted that all pixels belonging to the same stripe in the highest frequency pattern share the same codeword. Therefore, before triangulating, it is necessary to calculate either the centre of every stripe or the edge between two consecutive stripes. The latter has been shown to be the best choice.

Inokuchi et al. [6] improved the codification scheme of Posdamer and Altschuler by introducing Gray code instead of plain binary. The advantages of Gray code is that consecutive codewords have a Hamming distance of one, being more robust against noise. In Fig. 1a, the corresponding Gray coded patterns can be observed.

In 1981, Minou et al. [7] designed another technique based on time coded parallel stripes. The aim was to create a depth measurement system which was robust in the

Shenzhen Jimuyida
Exhibit 1009

Table 1
The proposed classification

			Static	Moving	Binary	Grey levels	Colour	Periodical	Absolute
Time-multiplexing	Binary codes	Posdamer et al.	√		√				√
		Inokuchi et al.	√		√				√
		Minou et al.	√		√				√
		Trobina	√		√				√
		Valkenburg and McIvor	√		√				√
		Skocaj and Leonardis	√		√				√
		Rocchini et al.	√				√		√
	n-ary codes	Caspi et al.	√				√		√
		Horn and Kiryati	√			√			√
	Gray code + Phase shifting	Bergmann	√			√			√
		Sansoni et al.	√		√				√
		Wiora	√			√			√
		Gühring	√		√				√
	Hybrid methods	Kosuke Sato	√		√	√			√
		Hall-Holt and Rusinkiewicz		√	√				√
Spatial Neighborhood	Non-formal codification	Maruyama and Abe		√	√			√	
		Durdle et al.		√		√		√	
		Ito and Ishii		√		√			√
		Boyer and Kak		√			√		√
		Chen et al.		√			√		√
	De Bruijn sequences	Hügli and Maître		√			√		√
		Monks et al.		√			√		√
		Vuylsteke and Oosterlinck		√	√				√
		Salvi et al.		√			√		√
		Lavoie et al.		√			√		√
		Zhang et al.	√	√			√		√
	M-arrays	Morita et al.	√		√				√
		Petriu et al.		√	√				√
		Kiyasu et al.		√	√				√
		Spoelder et al.		√	√				√
		Griffin and Yee		√	√		√		√
		Davies and Nixon		√			√		√
		Morano et al.	√		√		√		√
Direct coding	Grey levels	Carrihill and Hummel	√			√			√
		Chazan and Kiryati	√			√			√
		Hung		√		√		√	
	Colour	Tajima and Iwakawa	√				√		√
		Smutny and Pajdla	√				√		√
		Geng		√			√		√
		Wust and Capson		√			√	√	
		Tatsuo Sato	√				√	√	

presence of noise. For this purpose, the authors decided to use both binary code and the Hamming error correcting code. The number of coded stripes was only 25, therefore, the plain binary code had length 5 and the correcting code had length 9. It should be noted that the number of coded stripes is very small due to the large amount of bits needed to create a code with a Hamming distance of three, which then allows a single error correction.

Trobina, in 1995 [8], presented an error model of coded light range sensors based on Gray coded patterns. The author demonstrated that the crucial step of these sensors is the accurate location of every stripe in the image. In the finest pattern only half of all the edges can be measured, while the other half can be found in the previous patterns. By simple binarization of the images the stripes can be found. The binary threshold is fixed for every pixel independently. It

Shenzhen Jimuyida
Exhibit 1009

is necessary to acquire images of fully illuminated scenes (white pattern) and non-illuminated scenes (black pattern). The variant threshold is the mean between the grey level of such images as shown in Fig. 1b. Hence, with such binarization, the edges can be detected with *pixel accuracy*. However, the profile of the transition between a white and a black stripe in the images is not a perfect step. It is normally a non linear profile. Two ways of detecting the edges with *sub-pixel accuracy* were proposed.

The first way of detecting stripe edges with sub-pixel accuracy is to find the zero-crossings of the second derivative of the image, orthogonally to the stripes. The problem with this approach is finding the optimal gradient filter size. An alternative way is to project both normal and inverse stripe patterns, i.e. positive and negative patterns. Then, by finding the intersection of both profiles, the stripe edge is located. Since the profiles are non-linear functions, linear interpolation is used among the nearest sample points (grey levels of nearby pixels). As shown in Fig. 1e, by intersecting line AB with line EF, the edge is located. As can be seen in Fig. 1c, the intersection of both inverse and normal profiles do not always coincide with the variant binary threshold, so this method is more accurate. If projecting inverse patterns is not desired, the linearly interpolated normal profile can be intersected with the variant threshold profile. This technique is shown in Fig. 1d, where the segment AB should be intersected with segment CD. After experimental results, Trobina concluded that linear interpolation is more accurate than second derivative and the best results are obtained if both normal and inverse patterns are projected.

Locating the stripes accurately when projecting Gray coded patterns was also the main objective of the work presented by Valkenburg and McIvor in 1998 [9]. In this case, every acquired image is divided into regions of 17×17 pixels. In every region, a 2D third order polynomial is interpolated by means of least square fitting, obtaining a facet that approximates all the stripes in the region of interest. The authors also made experiments fitting sinusoidal functions in the regions, slightly improving the results.

Objects containing regions with different reflective properties are difficult to reconstruct. When projecting patterns at low illumination intensities, the signal-to-noise ratio of the system decreases and, therefore, depth from low reflective regions cannot be obtained. On the other hand, when projecting high illumination intensity patterns, depth from regions with high reflectance cannot be recovered due to pixel saturation. So, most binary coded techniques assume that the objects have uniform albedo, otherwise, the whole surface cannot be reconstructed. In 2000, Skocaj and Leonardis [10] proposed a new strategy to overcome these limitations by increasing the number of projected patterns. Projecting multiple images at different illumination intensities of a given stripe pattern allows each view of such patterns to be combined into a single *radiance map*. A radiance map contains for, each pixel, the relative reflective factor of the corresponding surface points. In general, we have that

$$g \alpha r \cdot l, \tag{1}$$

where g is the pixel value, r is the reflective of the corresponding surface point and l is the illumination intensity incident to the surface point. If relative radiance values are considered, then for every pixel i in the image j the relation expressed in Eq. (2) exists.

$$r_{ij} = g_i / l_j. \tag{2}$$

The variation of the illumination in the scene should not affect the reflective value of the surface point. However, the relationship is non-linear due to the distortions introduced by both the projector and the camera. In order to eliminate this non-linearity, an equation system is defined which considers all the pixels under different illumination values. The overdetermined system is minimised by using least-squares. Then, the best fitting reflectance value of every corresponding surface point is obtained. Hereafter, a global radiance map can be defined containing the reflectance values relating to every pixel of the image. Using this radiance map, the projected intensity l_j can be inversely recovered for every pixel value g_{ij} and its associate reflectance value r_i. The minimum number of illumination intensities to be projected is two (binary). However, by projecting more intensities, better results are obtained. In order to calculate the range image, the sub-pixel localisation of the stripe edges proposed by Trobina [8] was applied. The contribution of this work was a simultaneous reconstruction of both high and low reflective surfaces.

During the last few years, most of the work dealing with binary coded patterns has been aimed at improving the sub-pixel localisation of stripe edges. In 2001, Rocchini et al. [11], introduced a slight change in the typical Gray coded patterns in order to ease the localisation of stripe transitions. For this purpose, they proposed encoding the stripes with blue and red instead of black and white. Moreover, a green slit of a pixel width was introduced between every stripe. Then, the stripe transitions of the higher resolution pattern are reconstructed by locating the green slit with sub-pixel accuracy.

3.2. Techniques based on n-ary codes

The main drawback of the schemes based on binary codes is the large number of patterns which need to be projected. However, the fact that only two intensities are projected eases the segmentation of the viewed patterns. There are some works which consider the problem of reducing the number of patterns by means of increasing the number of intensity levels used to encode the stripes. There now follows an explanation of the two formal schemes of increasing the coding basis.

In 1998 Caspi et al. [12] proposed a multilevel Gray code based on colour. The extension of the Gray code is based on an alphabet of n symbols, where every symbol is associated

Shenzhen Jimuyida
Exhibit 1009

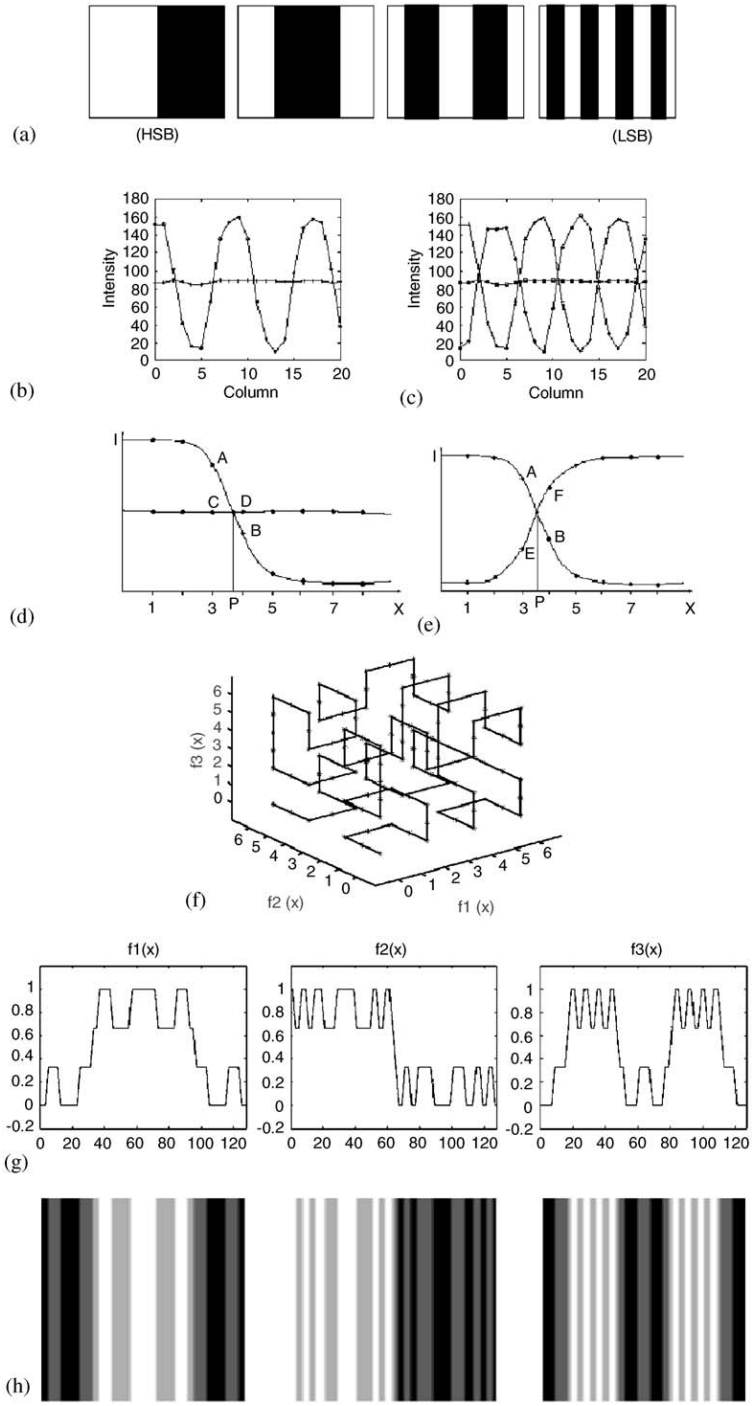

Fig. 1. Stripe patterns coded with binary and *n*-ary codes: (a) patterns coded with Gray code; (b) variant binary threshold of normal stripe pattern; (c) variant binary threshold of both normal and inverse stripe patterns; (d) stripe position by normal stripe pattern and binary threshold; (e) stripe position using normal and inverse patterns. (f) 3-D second order Hillbert curve with 128 codewords placed on it for *n*-ary codification; (g) intensity profiles of the patterns encoded with the codewords extracted from the Hillbert curve; (h) the resulting *n*-ary patterns.

Shenzhen Jimuyida
Exhibit 1009

to a certain RGB colour. This extended alphabet makes it possible to reduce the number of patterns. For example, with binary Gray code, m patterns are necessary to encode 2^m stripes. With an n-Gray code, n^m stripes can be coded with the same number of patterns.

This work was very important since it is the generalisation of the most widely used coding strategy in the time-multiplexing paradigm. The n-ary code shares the same characteristics of a binary Gray code by fixing a Hamming distance of one between consecutive codewords. The work by Caspi et al. not only develops the mathematical basis for generating n-ary codes, but also analyses the illumination model of a structured light system. This model takes into account the light spectrum of the LCD projector, the spectral response of a 3-CCD camera and the surface reflectance. The whole model is presented in Eq. (3),

$$\underbrace{\begin{bmatrix} R \\ G \\ B \end{bmatrix}}_{\vec{C}} = \underbrace{\begin{bmatrix} a_{rr} & a_{rg} & a_{rb} \\ a_{gr} & a_{gg} & a_{gb} \\ a_{br} & a_{bg} & a_{bb} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} k_r & 0 & 0 \\ 0 & k_g & 0 \\ 0 & 0 & k_b \end{bmatrix}}_{K} \vec{P} \underbrace{\begin{Bmatrix} r \\ g \\ b \end{Bmatrix}}_{\vec{c}} + \underbrace{\begin{bmatrix} R_0 \\ G_0 \\ B_0 \end{bmatrix}}_{\vec{C_0}} \tag{3}$$

where $\vec{c}$ is the projection instruction for a given stripe, i.e. $[0\ 0\ 0]^{\mathrm{T}}$ represents black and $[255\ 255\ 255]^{\mathrm{T}}$ is white; $\vec{P}$ is the non-linear transformation from projection instruction to actually projected intensities for every RGB channel; and A is the projector–camera coupling matrix. Every element a_{ij} of matrix A is the convolution of the camera spectral response for channel i with the spectrum of the light projected in channel j. Matrix A shows the crosstalk between colour channels. $\vec{C}$ is the vector containing the RGB camera readings of a certain pixel. $\vec{C_0}$ is the camera readings corresponding to the scene under ambient lighting. Finally, K is the surface reflectance matrix specific to every scene point projected into a camera pixel. This matrix contains a reflectance constant for every RGB channel.

The main benefit of this model is that it considers a constant reflectance for every scene point in the three RGB channels. This is much more realistic than considering colour neutrality of the scene, which is commonly assumed in most systems dealing with colour coding schemes. In the case of colour neutrality, matrix K is the identity for every pixel.

In order to calculate the different terms in Eq. (3) it is necessary to fulfill a colourimetric calibration. With this procedure, A and $\vec{P}$ are obtained. $\vec{P}$ is a non-linear function, but invertible, so it can be implemented in three look up tables (one for every colour channel). The colourimetric calibration is only necessary once. Then, matrix K is obtained by just taking a reference image under white illumination ($\vec{c_{\mathrm{w}}} = [255\ 255\ 255]^{\mathrm{T}}$) and solving Eq. (3). Some approximations of the model can be done in order to avoid the whole colourimetric procedure, as explained in Ref. [13].

The illumination model proposed allows the projected colour to be estimated from camera readings. This is very important when working with colour encoded stripes, since correct identification of colours leads to correct codewords. Therefore, such a model can be applied to any system dealing with colour.

Another important aspect of Caspi's work is its adaptation to the environment. This means that the system can be configured with different parameters. In this case, the parameters are the number of patterns to be projected M, the number of colours used L, and the noise immunity factor α. If M and L are chosen, then α is fixed. Otherwise, if α and L are chosen, then M is fixed. The noise immunity factor α determines the distance between adjacent consecutive codewords in every RGB channel. The higher α is, the more robust is the colour identification and the higher the number of patterns M since less colours can be used or fewer stripes need to be coded. After experiments, Caspi et al. determined that the n-ary Gray codes achieve the accuracy and robustness of the binary Gray code technique using fewer patterns.

Another technique that encodes adjacent stripes with n-ary codewords is the one proposed by Horn and Kiryati [14]. The alphabet of the codes is based on multiple grey levels instead of a binary alphabet. The aim of the work was to find the smallest set of patterns that meet the accuracy requirements of a certain application producing the best performance under certain noise conditions.

Given an alphabet of n symbols, a code is created so that consecutive codewords have a Hamming distance of one. Every element of the alphabet is mapped to a certain grey level. When expressing the differing elements of two consecutive codewords in terms of the associated grey levels, the difference is constant for all pairs of consecutive codewords. For example, if 256 grey levels are available, when using a binary Gray code the distance of consecutive codewords in terms of the grey levels is 255 or 100%. By increasing the basis of the code and maintaining the length, more codewords can be generated in spite of decreasing the distance in terms of % of available grey levels between consecutive ones. The authors proposed the use of space filling curves such as Hillbert or Peano curves [15] for defining the code. Such curves represent a path in an n-dimensional space, passing through a set of points so that consecutive points are joined by straight segments. The distance between consecutive points is constant along all the curve. Like the system proposed by Caspi et al., there are several parameters which can be tuned, so that L is the number of stripes to encode; K the number of projecting patterns, and therefore the length of the codewords; m the order of the curve used to place the codewords; and r the desired distance between consecutive codewords (% of the total of grey levels), which is proportional to the noise immunity factor of the system.

Shenzhen Jimuyida
Exhibit 1009

The K parameter is also the dimension of the space filling curve that will be used to generate the code. If parameters K and m are fixed, then the larger the parameter r is, the more robust the resulting codification is against noise, but the number of stripes, L, decreases. Therefore, there is a trade off between noise immunity, number of patterns, distance between codewords and number of encoded stripes. For a given r, if the number of stripes L must be increased it means that the length of the space filling curve must also grow. To achieve this, there are two possible solutions: increasing the curve order, which has the problem of reducing the distance between consecutive codewords (since the distance between consecutive points of the curve also reduces); or increasing the dimension of the curve, i.e. increasing the number of projecting patterns.

In Fig. 1f, a 3D Hillbert curve of the second order is shown. Every dimension of the curve is associated with one of the patterns to be projected. The number of stripes to encode has been fixed at $L = 128$, so a total number of 128 3D points have been placed equidistantly along the curve. Consecutive points along the curve correspond to adjacent stripes in the patterns. The value of every point component is the grey level associated with the stripe in one of the patterns. Therefore, every point in the curve produces the codeword of grey levels for the corresponding stripe. The number of grey levels used in the example is seven. The extracted intensity profiles of every three patterns are shown in Fig. 1g, while the resulting patterns are shown in Fig. 1h. Every intensity profile is the projection of the points in the curve in one of the three axis: $f1(x)$, $f2(x)$ and $f3(x)$ are the intensity profiles of patterns 1, 2 and 3, respectively.

Horn and Kiryati tested their system with a 3D Hillbert second order curve where 256 codewords were placed. Therefore, the number of patterns was three and the number of grey levels 13. Such configuration produced better performances than a binary Gray coding scheme based on the projection of nine patterns (512 stripes encoded). The design of coded patterns proposed by Horn and Kiryati produces more accurate results and reduces the number of projecting patterns.

3.3. Combination of Gray code and Phase shifting

Patterns based on Gray code, as well as binary and n-ary codes, have the advantage that the pixel codification is made punctually and no spatial neighbourhood has to be considered in the codification. However, the discrete nature of such patterns limits the range resolution. Furthermore, phase shifting methods exploit higher spatial resolution as they project a periodic intensity pattern several times by shifting it in every projection. The drawback of these methods is the periodic nature of the patterns, which introduces ambiguity in the determination of the signal periods in the camera images. The integration of Gray code methods (GCM) with phase shift methods (PSM) brings together the advantages of both strategies, i.e. the unambiguity and robust codification of pattern stripes of GCM, plus the high resolution of PSM. The combination of both techniques leads to highly accurate 3D reconstruction. However, the number of projecting patterns increases considerably. We will now discuss some of the coded structured light systems that use this approach.

In Ref. [16], Bergmann designed a technique where some Gray coded patterns are projected in order to label the measuring surface regions where every period of a sinusoidal intensity pattern will be projected. Therefore, the ambiguity problem between signal periods is resolved. The sinusoidal patterns are represented by grey levels. A total number of four Gray patterns are projected in order to label 16 different regions on the measuring surface. Then, the periodic intensity pattern is projected four times by shifting 1/4 of the period, each time. For every given pixel (x, y) of the camera image, the phase of the first periodic pattern projected to the corresponding surface point must be found. For this purpose, a classic four-step phase shift is applied in Eq. (4). I_1, I_2, I_3 and I_4 are the grey levels of pixel (x, y) from camera images corresponding to every one of the four projected shifted patterns.Once the phase of a given pixel is known, the period of the sinus where the pixel lies is obtained with the Gray code labelling. Therefore, the pattern stripe projected to a certain surface point can be precisely calculated.

$$\phi(x, y) = \arctan\left(\frac{I_2 - I_4}{I_3 - I_1}\right) \tag{4}$$

Sansoni et al. compared the accuracy of GCM and PSM separately [17]. After experiments, they realised that both PSM and GCM obtain similar precision in their measurements (about 0.18 mm). However, the resolution of PSM showed to be about 0.01 mm in front of the 0.22 mm of GCM. However, PSM failed in the steep slope changes at the measuring surface borders due to the occlusion of some periods of the PSM patterns. One interesting feature of the PSM patterns used by Sansoni et al. is that they are discrete stripe patterns with rectangular profiles. The sinusoidal profile is achieved by defocusing the LCD projector. When combining both methods, the mean error of the measurements was about 40 μm with a standard deviation of ±35 μm. Furthermore, Georg Wiora discussed the suitability of using LCD projectors when applying PSM alone or combined with GCM [18]. He argued that such devices do not have enough contrast and radiant flux for PSM patterns with large resolution. According to the author, the best devices to use are special slide projectors, which allow 26,000 black and white stripes to be projected on a 13 mm slide (for more details refer to the article). Moreover, Wiora's article discusses the problems of mechanical misalignment of slides for these devices as well as problems of non-sinusoidal phase shift patterns.

In 2001, Gühring proposed substituting the PSM for a new method called *line shifting* [19], which was also combined with GCM. Gühring pointed out that *phase shifting* has a series of drawbacks. For example, when reconstructing surfaces with non-uniform albedo (with sharp changes from black to white) the phase cannot be determined precisely.

Shenzhen Jimuyida
Exhibit 1009

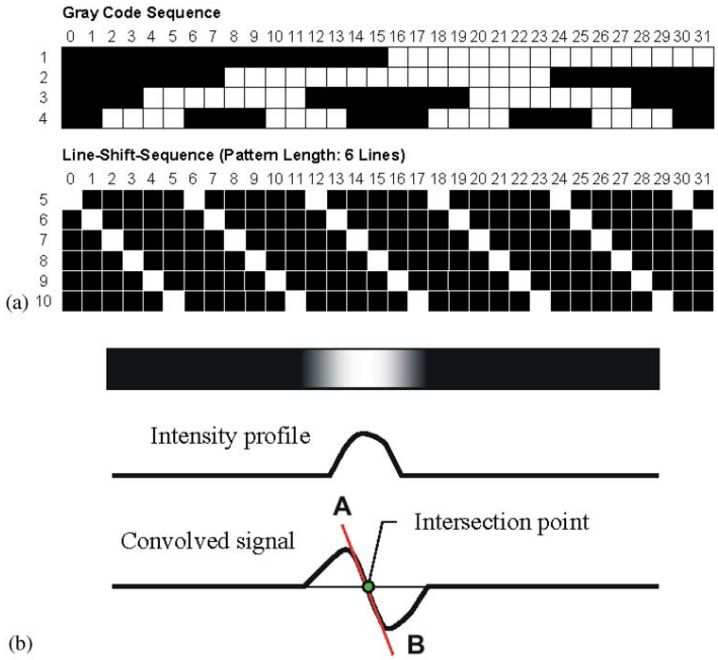

Fig. 2. Line shifting technique: (a) Gray code and line shifting patterns for 32-pixel-wide patterns; (b) Intensity profile of a camera illuminated line and the peak detection with sub-pixel accuracy.

Moreover, camera devices tend to integrate over a certain area so that pixel values are affected by its neighbours. To avoid this problem, camera resolution must be sufficiently higher than projector resolution. In order to avoid the problems of *phase shifting*, the author proposed substituting the sinusoidal periodic profile of such methods by a *multistripe* pattern, shifted several times. Gühring designed a 640×640 pattern for LCD projectors, where every 6th column is white and the remaining ones are black. By consecutively shifting the pattern 6 times, the whole resolution for every row of the pattern is covered. While repeating the process with row-encoded patterns, the entire resolution of the pattern is used. Since a multistripe pattern is also a periodic pattern (of discrete nature), the projection of Gray code patterns are also required in order to solve the ambiguities that arise.

To summarise, 32 patterns were projected onto the measuring surface: nine vertical Gray codes; six vertical multistripe patterns, each one shifted a column towards the right; nine horizontal Gray coded patterns; six horizontal multistripe patterns each one shifted one row downwards; two additional patterns for grey level normalisation (one fully illuminated and the other with the lamp switched off). With regard to the Gray coded patterns, a total number of n regions should be labelled, being n the number of lines projected in every *line shifting* pattern. Therefore, every pattern area where a line is shifted has its own label. For example, if patterns of 32 columns were projected, six lines would be shifted and, therefore, six bands of six pixels width should be labelled by the Gray code. Therefore, three patterns of Gray code should be projected. However, in the transitions of each region, a decoding error implies a large measuring error of around one period. That is why Gühring decided to introduce an oversampling technique consisting of projecting an additional Gray coded pattern. In this way, thinner bands of pixels are labelled, and more robust decodification of the regions is obtained. The maximum error when calculating the global position of an illuminated line due to transition of a Gray codeword was two pixels. Without the redundancy introduced by this oversampling, the same error rose to six pixels. In Fig. 2a, the case of a row of a 32-pixel-wide pattern is shown, with the four patterns of Gray code and the six patterns containing line shifting. As can be seen, every region where a line is shifted contains three Gray codewords.

As for the patterns containing *line shifting*, the peak position of every illuminated line was intended to be located in the camera images with sub-pixel accuracy. Since the intensity profile of such perceived lines presented a gaussian distribution, the peak detector proposed by Blais and Rious [20] was used. This detector applies a linear derivative filter (higher order filters can be applied) at each pixel of every image row (when treating vertical multistripe patterns). The result obtained for each row is a set of local maxima and minima indicating the transitions from black to white

Shenzhen Jimuyida
Exhibit 1009

regions and vice versa. Afterwards, for each pair of consecutive maximum and minimum, the zerocrossing of the linear interpolation between them is calculated and the sub-pixel position of the intensity peak is obtained. The detection process of a peak for a certain image row is shown in Fig. 2b.

Gühring's line shifting method had a similar or even better resolution than techniques based on PSM and more accurate measurements. Note that this approach was inspired by traditional laser scanner techniques, which have been shown as the most accurate 3D profilers. The author developed a system set up based both on LCD and DMD projectors, obtaining similar results with an average error of 30 μm and a maximum deviation of 0.281 mm for both devices.

3.4. Hybrid methods

In the bibliography, there are some methods which are based on multiple pattern projection, so they use time-multiplexing, but also take into account spatial neighbourhood information in the decoding process. For example, the idea of Kosuke Sato [21] consisted of designing a certain binary pattern whose rows have a sharp impulse on its auto-correlation function. The pattern is projected several times by shifting it horizontally several times (the more times the pattern is shifted, the greater the resolution obtained). For every projection, an image is grabbed, in which the maximum autocorrelation peak of every row is computed. Then, as the phase shift of the corresponding projected pattern is known, the pixels containing such peaks can be reconstructed by triangulation. According to the author, this strategy achieves better accuracy than projecting and shifting a single slit, since the peak of cross-correlation shows sharper impulse and can be located more precisely.

In 2001, Hall-Holt and Rusinkiewicz [22] divided four patterns into a total of 111 vertical stripes that were painted in white and black. Codification is located at the boundaries of each pair of stripes. The codeword of each boundary is formed by 8 bits. Every pattern gives 2 of these bits, representing the value of the bounding stripes. The most interesting aspect about this method is that it supports smooth moving scenes, something unusual in the time-multiplexing paradigm. This capability is due to a stage of boundary tracking along the patterns.

4. Spatial neighbourhood

The techniques in this group tend to concentrate all the coding scheme in a unique pattern. The codeword that labels a certain point of the pattern is obtained from a neighbourhood of the points around it. However, the decoding stage becomes more difficult as the spatial neighbourhood cannot always be identified and 3D errors can arise. Normally, the visual features gathered in a neighbourhood are the intensity or colour of the pixels or groups of adjacent pixels around it.

These spatial neighbourhood techniques can be classified as follows: (a) strategies based on non-formal codification: the neighbourhoods are generated intuitively; (b) strategies based on De Bruijn sequences: the neighbourhoods are defined using pseudorandom sequences; or (c) strategies based on M-arrays: extension of the pseudorandom theory to the 2-D case. In the following subsections some techniques from these three subgroups are summarised.

4.1. Strategies based on non-formal codification

Some authors have proposed techniques based on patterns designed so that it is divided into a certain number of regions, in which some information generates a different codeword, without using any mathematical coding theory. For instance, in 1993, Maruyama and Abe [23] designed a binary pattern coded with vertical slits containing randomly distributed cuts. The system was designed for measuring surfaces with smooth depth changes. The random cuts generate a set of linear segments so that the position of a segment in the pattern is determined by its own length and the lengths of the 6 adjacent segments. The decoding stage starts by matching every segment of the pattern with the observed slits of equal length. Multiple matchings can be found for every segment. In order to find out the correct matching, the lengths of the six adjacent segments must be considered. Once all the perfect matchings have been found, a region growing algorithm is applied in order to identify unmatched segments. The main drawback of this technique is that segment lengths can vary depending on the distance between the camera and the surface and the optics of both the camera and projector. All these factors considerably limit the robustness and the reliability of the system.

Some years later, in 1998, a periodic pattern composed of the horizontal slits encoded with three grey levels was proposed by Durdle et al. [24]. The pattern is formed by the sequence $BWGWBGWGBGWBGBWBGW$, where B is a band of four black pixels, W is a band of four white pixels and G is a band of four half bright pixels. This sequence is repeated in the pattern until it covers all the vertical resolution. Due to the periodicity of the pattern, discontinuities larger than the pattern period are not permitted. The decoding stage of the system is composed of two steps: firstly, the starting point of every pattern period is searched in the grabbed image for every column, by finding the correlation peaks between the image column and a template of the projected period. Secondly, for every period, another template matching is made in order to find the subsequences WBG, WBG and GWB. Repeating these processes for every image column, a large set of correspondences is found.

Ito and Ishii presented a *three-level checkerboard* pattern in 1994 [25]. The proposed pattern is a unique grid where each square is painted with one out of three possible intensity levels. The intensity level of a cell is chosen so that it is different from its four immediate neighbours. A *node* is defined as an intersection between four cells of

Shenzhen Jimuyida
Exhibit 1009

the checkerboard. The *main code* of a node is identified by the intensity levels of its four adjacent cells. Since three intensity levels are used, the number of different main codes is 18 ($3 \times 2 \times 2 \times 2$). The *subcode* of a node is defined as the clockwise combination of the main codes of the four adjacent nodes. The subcode constitutes the codeword for every edge intersection of the pattern. Therefore, in order to decode the position of an observed node, it is necessary to analyse a spatial neighbourhood of 12 cells of the grid. Note also that both spatial coordinates of the nodes are encoded. In order to get a robust coding scheme, every possible subcode should appear only once in the pattern. However, Ito and Ishii were not studying the automatic generation of such a pattern, so they decided to allow repetitions of subcodes. The authors used epipolar restrictions between the camera and the projector in order to differentiate between nodes which share the same subcode.

The technique proposed by Boyer and Kak [26] uses a pattern formed by vertical slits coded with the three basic colours (red, green and blue) and separated by black bands. The sequence of coloured slits was designed so that if the pattern is divided into subpatterns of a certain length, none is repeated. The most interesting thing about this work is the decoding stage. Boyer and Kak realised that the morphology of the measuring surface acts as a perturbation applied to the projected pattern (which acts like a signal), so the received pattern can contain disorders or even deletions of the slits. In order to match each received slit with the corresponding projected slit, a four-step algorithm was designed called *stripe indexing process*.

The first step is called *correlation*. Each unique subpattern of the original pattern is sliced along the received pattern to find all the positions where a perfect match takes place. Secondly, a region growing process of the matched subpatterns is carried out and tries to cover as many correspondences of slits as possible. This subprocess was called *crystal growing*. Thirdly, a fitting process is applied in order to remove erroneous matchings. When two subpatterns overlap, the thinnest is cut so that the shared slits are only associated with the largest one. Finally, the matched slits are indexed. The whole process must be done for every epipolar line of the camera image. The authors did not take into account the information obtained from the previous epipolar line in order to validate the current one. The drawback of this method is the complex algorithms involved to recover the pattern. Moreover, the crystal growing procedure does not always lead to the correct identification of the slits, so uncertainty should be considered. The advantage of the method is the possibility of obtaining shape from moving objects. This work is of great value, since it inspired a series of works which dealt with more evolved techniques the problem of lost slits or disorders among slits.

In 1997, Chen et al. presented a range sensor based on two cameras and an LCD projector [27]. The latter projected a unique pattern in order to ease the search for correspondences along pairs of epipolar lines. However, the technique can also be applied when only one camera is used. The pattern consisted of a series of vertical coloured slits separated by black bands. The slit colours were chosen using a trial and error algorithm in order to find a sequence with low autocorrelation in the Hue component of the slits. The decoding method proposed by Chen et al. was the most developed part of the system. The decoding stage was divided into two parts: the *intra-scanline* search and the *inter-scanline* consistency. Since Chen et al. used two cameras, all the points laying on a line in the first image have their correspondences in a line of the second image. Both lines constitute a pair of epipolar lines. The *intra-scanline* process tries to match every edge of every pair of epipolar lines, in order to triangulate their 3D position at a later stage. Every epipolar line contains a sequence of colours separated by black gaps corresponding to the projected coloured slits that are visible from the point of view of the camera. Since the cameras in this system have different points of view, for a given pair of epipolar lines, the observed sequences of colours can differ. Therefore, usually one of the cameras perceives more slits than the other one and, therefore, the number of edges differs. In order to match the edges observed by both cameras *dynamic programming* was used. Dynamic programming is capable of obtaining the optimal set of insertions, deletions and substitutions that must be applied to the perceived sequence in order to obtain the original projected sequence. Nevertheless, this algorithm is only robust against deletions and erroneous identification of the colour of some slits, but does not ensure a good solution if disorders among slits have occurred. Therefore, the measuring surface must be monotonic. Furthermore, in the *inter-scanline consistency* stage, an attempt is made to match the edges that have not been matched in the pair of images by using the information of adjacent epipolar lines.

The weak point of the work from Chen et al. is the lack of robustness of the pattern used. Although the sequence of colours is generated accomplishing some constraints, it is not optimal. Furthermore, the authors assumed that the observed slits cannot be reordered with respect to the projected ones, but this is not always true.

4.2. Strategies based on De Bruijn sequences

The pattern projection techniques presented in the previous subsection were usually generated by brute-force algorithms in order to obtain some desirable characteristics. In this subsection, a group of patterns encoded with a well-defined type of sequences called *De Bruijn sequences* is presented. First, a theoretical introduction into the field is given in order to explain why these sequences are suitable for encoding patterns.

A *De Bruijn sequence* of order m over an alphabet of n symbols is a circular string of length n^m that contains each substring of length m exactly once. Similarly, a *pseudo-random sequence* or a *m-sequence* has a length of $n^m - 1$ because it does not contain the substring formed by 0's [28].

Shenzhen Jimuyida
Exhibit 1009

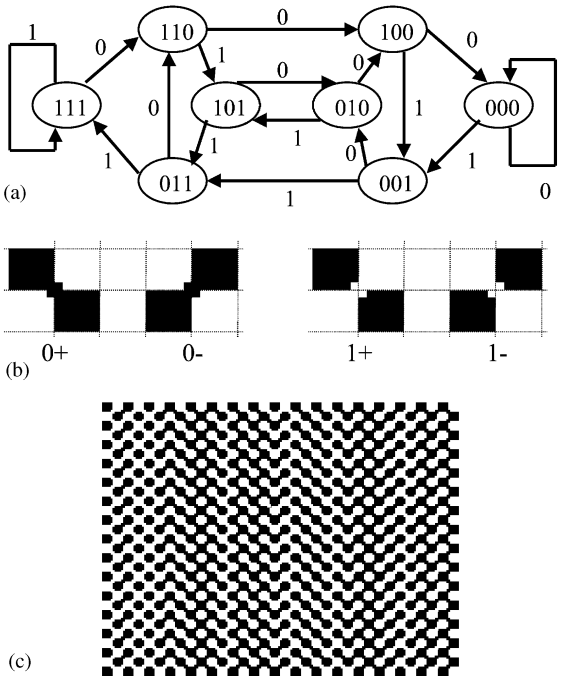

Fig. 3. De Bruijn codification strategy: (a) Example of a De Bruijn graph to construct De Bruijn sequences; (b) primitives proposed by Vuylsteke and Ooesterlick to represent the binary values; (c) the resulting pattern using the primitives.

This sort of sequence can be obtained by searching *Eulerian circuits* or *Hamiltonian circuits* over different kinds of *De Bruijn graphs* [29]. For example, in the graph shown in Fig. 3a, all the words of length $m-1$ (with m equal to 4) are included in the vertices. An Eulerian circuit is a path starting and ending in the same vertex and passing through all the edges exactly once. Gathering the edge labels of such circuit, a De Bruijn sequence of order m is obtained, i.e. $S=1000010111101001$. If a Hamiltonian circuit is searched over the same graph (a path which passes through all the vertices only once and starts and ends in the same vertex), a De Bruijn sequence of order $m-1$ is obtained ($S=00101110$). An interesting property of a De Bruijn sequence is that it presents a flat autocorrelation function with a unique peak at moment 0. It can be shown that this is the best autocorrelation function that can be achieved and this means that it is clearly uncorrelated. Pseudorandom sequences have been used to encode patterns based on column or row lines and grid patterns. In the following paragraphs, some relevant works using De Bruijn sequences to encode patterns are explained.

In 1998, Hügli and Maitre [30] improved the pattern proposed by Boyer and Kak [26]. In this case, a pattern composed of horizontal coloured slits was also projected. However, the sequence of colours was chosen using a pseudorandom sequence. The authors studied sequences where two consecutive slits with the same colours were not allowed. Therefore, the length of a sequence accomplishing this constraint is $Q(Q-1)^{N-1}$, where Q is the number of colours used and N the window size.

Monks et al. [31] designed a pattern based on horizontal coloured stripes in order to reconstruct dynamic scenes. A total number of 6 colours were used to paint the slits, separated by black bands. The colouring of the slits was chosen so that every subsequence of three colours appeared only once. Therefore, a De Bruijn sequence of order three based on an alphabet of six symbols was used. The given sequence was taken from the article published by Hügli and Maitre [30].

In this technique, the camera image is thresholded in the *HSV* colour space, using the *Hue* component to distinguish among the six colours used, i.e. red, green, blue, yellow, magenta and cyan, which are equally spaced with respect to this component. In the decoding stage, Monks et al. faced the problem of loss of slits for every column of the camera image. Due to surface discontinuities, some of the slits were not observed by the camera. To recover the position in the pattern of a given slit is necessary to correctly identify the colours of the slit itself and the slits above and below it. The authors decided to build a graph for the whole camera image, where every node represents an image edge between a coloured slit and the black band above it. In every node, the colour of the corresponding slit is stored. Two nodes are connected if the corresponding slits are consecutive in an image column and the distance between them is not very long (otherwise, an occlusion had occurred and some slits may have been deleted). Every column in the image produces a new path in the graph. All the nodes of the graph corresponding to a set of at least three consecutive slits are shared for all those image columns which detected the same subsequence. These nodes are shared since their position in the pattern is directly found thanks to the window property of the De Bruijn sequence. For all the other slits detected in an image column, a new node is inserted in the graph. A minimum-cost, matching algorithm is used to match the original projected sequence and the graph. The match algorithm minimises a cost function based on the costs associated with a node insertion and deletion and the cost of replacing the colour of a node.

The system was applied to speech recognition, projecting the pattern to the speakers's face and reconstructing the mouth pose in order to recover the pronounced letters. The work by Monks et al. has a robust decoding stage, which can be applied to other systems based on De Bruijn sequences.

Vuylsteke and Oosterlinck [32] presented, in 1990, a binary encoded pattern by means of De Bruijn sequences. A total number of 63 columns were encoded in a unique pattern, so the system is suitable for moving scenes. The pattern structure is a checkerboard where the column of every grid point is encoded. The encoding system is based on two binary *pseudorandom sequences* of order 6 and length 63,

Shenzhen Jimuyida
Exhibit 1009

shown in Eq. (5).

$$\{c_k\} = 1111110000010000110001010011110100011100100101101110110011011010 10,$$
$$\{b_k\} = \{c_{k-17}\}, \tag{5}$$

where $\{b_k\}$ is the same sequence c_k shifted by $\varphi = 17$ positions. Both sequences have a window property of length 6. When combined in a bitmap, as shown in Eq. (6) for every row, the obtained code assignment has the window property of 2×3. Every element of the sequences represents the individual representation of each grid point of the pattern.

$$\begin{matrix} c_0 & b_1 & c_2 & b_3 & c_4 & b_5 & c_6 & b_7 & c_8 & b_9 & \cdots \\ b_0 & c_1 & b_2 & c_3 & b_4 & c_5 & b_6 & c_7 & b_8 & c_9 & \cdots \\ c_0 & b_1 & c_2 & b_3 & c_4 & b_5 & c_6 & b_7 & c_8 & b_9 & \cdots \end{matrix} \tag{6}$$

In fact, since only column codification is desired, the windows in Eq. (7) should produce the same codeword.

$$\begin{bmatrix} b_3 & c_4 & b_5 \\ c_3 & b_4 & c_5 \end{bmatrix} \quad \begin{bmatrix} c_3 & b_4 & c_5 \\ b_3 & c_4 & b_5 \end{bmatrix}. \tag{7}$$

In order to do this, the binary codewords are generated by reading first the elements of c_k and then the elements of b_k. In the example given, the codeword for both windows would be $((c_3, c_4, c_5), (b_3, b_4, b_5))$. To be able to distinguish among elements of c_k and elements of b_k, it is necessary to chose different representations in the pattern. For this purpose, every grid point is marked with a bright or a dark spot representing the binary states 1 and 0. Then, the four neighbouring squares of the grid point in the checkerboard are painted depending on whether the corresponding sequence element belongs to c_k or to b_k. Both representations are shown in Fig. 3b. The binary states for sequence c_k are labelled $0+$ and $1+$, while the ones for sequence b_k are $0-$ and $1-$. The whole pattern representation is shown in Fig. 3c. The segmentation of the pattern in the camera images is easily done as only two intensity levels are used and also there is symmetry around every grid point. The decoding stage consists of recovering every 2×3 window and obtaining its codeword, which leads to the column position in the pattern. According to the authors, using rectangular windows is more robust than using single row windows, as the neighbourhood involved is more compacted and less sensitive to surface discontinuities. In 1995, Pajdla reimplemented the whole method, improving the calibration process [33].

Later, Salvi et al. proposed a pattern consisting of a grid of thin vertical and horizontal slits [34]. The authors argued that grid coding is a better solution because the grid points can be easily segmented. Furthermore, the neighbourhood around a grid point can be found by just tracking the edges of the grid. The pattern was designed as a 29×29 grid using three different colours for horizontal slits and three more colours for vertical slits. The colour assignment was made by using the same De Bruijn sequence of order 3 (size of window property) for both rows and columns of the grid. The grid intersection points are reconstructed after decoding their position in the pattern using the window property. Some years later, in 2000, Petriu et al. [35] proposed a similar pattern of 15×15 grid points. However, they did not propose reconstructing the grid crossing points, but reconstructing the four corner points of every intersection. Therefore, the resolution of the system is larger. The only requirement is to increase the thickness of the slits, so that the four corners of a crossing do not fall in the same pixel of the image. Moreover, in 1999, Lavoie et al. [36] proposed a similar pattern to the one proposed by Salvi et al. A pseudorandom sequence of third order based on 5 colours was used, obtaining a sequence of length 124 where every subsequence of length 3 is unique. Both vertical and horizontal slits of the grid are coded with the same pseudorandom sequence. The most interesting aspect of the technique proposed by Lavoie et al. is that they do not reconstruct the crossing points, but rather, the curves. For this purpose, *non uniform rational Bézier splines* (NURBS) were used. The NURBS have some nice properties under affine transformations as they are invariant under scaling, rotation, translation, shear and parallel and perspective projection. Once the grid is segmented in the camera image, the recovered grid points are used as control points to interpolate 2D NURBS in the image for both rows and columns of the grid. Due to the projection invariance property of a NURBS curve, a reverse projection transform can be performed in order to obtain the 3D NURBS that fit the measuring surface.

Recently, in 2002, Zhang et al. [13] developed a technique based on De Bruijn codification that achieves excellent performance. The proposed pattern consisted of 125 vertical slits coloured by using a De Bruijn sequence of third order and 8 colors (equal adjacent colors in the resulting sequence were eliminated). Zhang et al. studied the problems that occlusions and discontinuities in the measuring surface can produce when observing the projected pattern. As other authors had previously pointed out, they agreed that deletions and disorders among the slits may appear in the observed sequence. In order to match the observed sequence with the projected one, dynamic programming was adopted as in the work by Chen et al. However, Zhang et al. pointed out that simple dynamic programming is only successful when the measuring surface is monotonic, i.e. disorders among slits cannot appear. In order to eliminate such a limitation, they invented the multi-pass dynamic programming. Zhang et al. also extended their technique to the time-multiplexing paradigm by projecting the pattern several times by shifting it consecutively and locating the slits with sub-pixel

Shenzhen Jimuyida
Exhibit 1009

accuracy in each iteration. Therefore, the total resolution is increased.

4.3. Strategies based on M-arrays

There is a set of authors who have adopted the theory of *perfect maps* in order to encode a unique pattern, taking advantage of the interesting mathematical properties of these matrices. In the following paragraphs we give an introduction to this mathematical theory.

Let M be a matrix of dimensions $r \times v$ where each element is taken from an alphabet of k symbols. If M has the window property, i.e. each different submatrix of dimensions $n \times m$ appears exactly once, then M is a *perfect map*. If M contains all submatrices of $n \times m$ except the one filled by 0's, then M is called an *M-array* or *pseudorandom array*. For more information see Refs. [28,37]. This kind of array has been widely used in pattern codification because the window property allows every different submatrix to be associated with an absolute position in the array.

M-arrays can be constructed by folding a pseudorandom sequence. In order to create an M-array of dimensions $n_1 \times n_2$, a pseudorandom sequence of length $n = 2^{k_1 k_2} - 1$ is required, where $n = n_1 \cdot n_2$, $n_1 = 2^{k_1} - 1$ and $n_2 = n/n_1$. The resulting array has a window property of $k_1 \times k_2$. The procedure is as follows: the first element of the sequence is placed in the north-west vertex of the array. Successive elements of the sequence are written in the array following the main diagonal and continuing from the opposite side whenever an edge is reached. For example, given the binary pseudorandom sequence of the fourth order 010011000101111 whose length is $n = 15$, good parameters for constructing an *M-array* are $n_1 = 2^2 - 1 = 3$ and $n_2 = 15/3 = 5$. Then, an *M-array* of 3×5 with window property 2×2 is obtained. It must be noted that *M-arrays* also share the *circular* property of the pseudorandom sequences, so it is necessary to complete the array by adding to the right the firsts $k_2 - 1$ columns and the firsts $k_1 - 1$ rows below it. The complete array is shown in Eq. (8)

$$\begin{bmatrix} 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix} \tag{8}$$

The main differences between the techniques included in this group is the way in which the elements of the array are represented in the pattern. Some authors prefer to define the pattern as an array of coloured spots, where each colour represents one of the symbols of the coding alphabet. Other authors prefer to define different shapes for each symbol. When perceiving the projected pattern, an algorithm to recover the maximum number of visible windows must be fulfilled. This is the crucial step of these systems. Since spatial neighbourhood is used, not all the pattern will be visible from the camera's point of view, due to shadows and occlusions. The robustness of these methods depends on the correct decodification of the visible parts, taking advantage of the properties of the M-arrays.

Using arrays to codify a pattern means that a bidimensional coding scheme is being used, because every point of the pattern has a different codeword which encodes both vertical and horizontal coordinates. Since the codification is concentrated in one pattern, these techniques are suitable for measuring dynamic scenes. However, some authors prefer to project additional patterns in order to ease the segmentation part of the system or to carry out an intensity or colour normalisation. In this case, the system is limited to static scenes. In any case, the number of projected patterns is always lower when compared to time-multiplexing methods. In the following paragraphs, the existing patterns based on M-arrays are addressed and the most relevant are briefed.

A binary M-array of 24×24 was proposed in 1988 by Morita et al. [38]. This array has the window property of 3×4. The M-array representation is made by painting black dots on a white background, for the array elements corresponding to symbol 1. Two patterns are projected on the measuring surface: the first one contains all the possible black dots in order to locate their centers in the camera image. The second pattern is the M-array representation. Therefore, the method is restricted to static scenes. However, it can be adapted to moving scenes by only projecting the M-array pattern and making the segmentation and decoding algorithm more robust.

In 1992, Petriu et al. [39] used an M-array to encode a grid pattern where each cross-point represents an element of the M-array. The binary state of every cross-point is represented with the presence or absence of a square painted on it. The system was intended for object recognition, based on a database containing previously reconstructed surfaces.

Some years later, Kiyasu presented an interesting study [40]. The aim of the work was to obtain the shape of specular polyhedrons, i.e. objects composed of flat surfaces with high reflectance. Normally, most coded structured light systems are not intended to reconstruct specular surfaces but lambertian ones. A binary M-array represented with a grid of 18×18 circular spots with a window property of 4×2 was used.

In 1997, Spoelder et al. began to develop a prototype to measure the shape of the cornea [41], by means of projecting a binary M-array of 65×63 elements. Cyan and yellow were used to encode the binary values of the M-array. The structure of the designed pattern is a checkerboard, where the white squares are used to place the elements of the M-array. In Fig. 4a a portion of the pattern can be observed. The black squares were introduced in order to ease the pattern segmentation in the camera images. Due to the complex reflectance characteristics of the cornea, the recovered pattern from the camera images has a lot of data loss. This required the design of a complex segmentation algorithm, which we will now summarise. Firstly, the cross-points of the

Shenzhen Jimuyida
Exhibit 1009

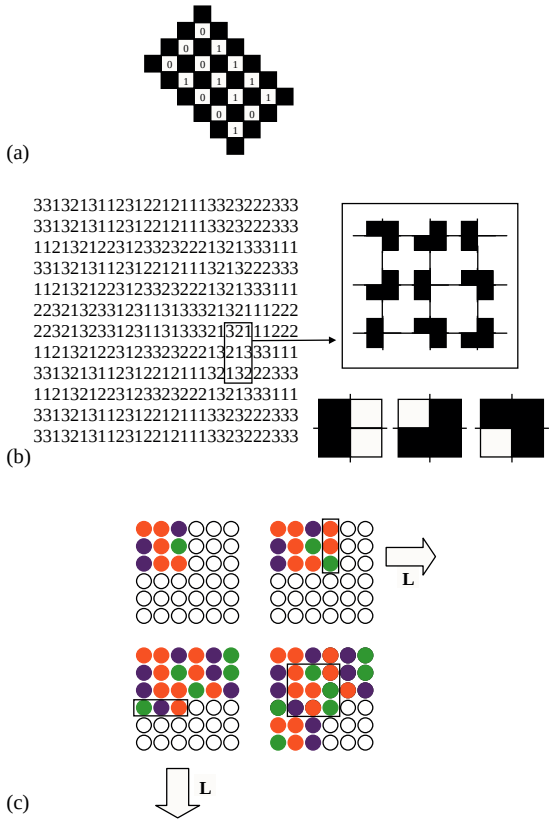

Fig. 4. M-array based patterns examples: (a) Binary M-array located in a checkerboard the 0 and 1 symbols are replaced with two different filling colours; (b) Example of M-array based on three symbols proposed by Griffin et al. Three shape primitives were proposed to represent the symbols of the alphabet $\{1,2,3\}$; (c) Morano et al. algorithm to generate M-arrays with coloured spots representation.

checkerboard are located by mask filtering. Secondly, every detected cross-point is labelled by observing the colours of the adjacent non-black squares and using the window property. Then, a graph is constructed by linking the neighbours. This step leads to a series of disconnected *sub patterns* that must be matched to the original projected pattern. Each subpattern is positioned on the projected pattern in the position where the minimum Hamming distance is achieved. The elements which do not fit in the original pattern are intended to be corrected.

One of the most famous works of this group was presented by Griffin et al. [42] in 1992. The authors defined a systematic process for constructing a maximum size array of $n \times m$ based on an alphabet of b symbols with certain restrictions: every element of the array has a unique codeword formed by its own value and the values corresponding to its four neighbours (north, south, east and west). As can be seen, such an array is a special case of *perfect maps*, since it has window property of 3×3, but not all the possible windows appear. Some authors call these arrays *perfect submaps*. The construction process of such matrices is as follows: first, let Vhm be the sequence based on alphabet b containing all the possible triplets of symbols (a *De Bruijn sequence*). Let Vvm be the vector made by the sequence of all the pairs of symbols of alphabet b. Consequently, the first row of the matrix is $f_{0i} = Vhm_i$.

The rest of the matrix elements are calculated using Eq. (9). The row index is indicated with i and varies from 0 to the length of Vhm, while j is the column index varying from 0 to Vvm length.

$$f_{ij} = 1 + ((f_{i-1j} + Vvm_j)\, mod\, b). \tag{9}$$

For example, if an alphabet $b = \{1,2,3\}$ is taken, then the following vectors are obtained:

$$Vhm = (3313213112312212111332322233 3),$$

$$Vvm = (3121132233). \tag{10}$$

Then, applying the algorithm the matrix shown in Fig. 4b is generated.

For their experiments, Griffin and Yee generated an array of 18×66 using the alphabet of four symbols $\{1,2,3,4\}$. In order to project this array, two strategies were adopted. The first consisted of representing each alphabet element with a different colour. Then the projected pattern was defined as an array of coloured dots. The second approach consisted of defining a set of shape primitives for every element of the alphabet. An example of such primitives is shown in Fig. 4b. Then the background of the pattern is painted black and white forming a grid of 18×66 with one of the primitives at every crossing point. The representation of a window of 3×3 of the M-array is shown in Fig. 4b. This second representation of the M-array is much more robust in the presence of coloured objects. Some years later, in 2001, Hsieh presented an analytical method for decoding the position of a given codeword of Griffin's array [43], using just simple arithmetic operations with the elements of the window and decoding the pattern quickly.

In 1998, Davies and Nixon [44] proposed a unique pattern of coloured spots for obtaining shapes from dynamic scenes. Specifically, the system was applied for automatic speech identification by projecting the pattern onto the speakers' face at video rate. The spots are coded by following Griffin's method. Cyan, yellow and magenta colours were chosen to paint the spots, which are placed hexagonally in the pattern. In this technique a segmentation algorithm is applied for obtaining the image coordinates of the visible dots. First, an edge detector filter is used in order to find the contours of the perceived ellipses corresponding to the projected circular dots. For every epipolar line in the camera image, all the ellipses nearly positioned onto the line are searched. Then accurate positions of the ellipses are found using an adapted formulation of the Hough Transform. When all possible ellipses have been located, the decoding process using

Shenzhen Jimuyida
Exhibit 1009

the window property leads to correspondences between the camera image and the projected pattern.

One of the most interesting techniques from this group was given by Morano et al. [45]. The authors proposed an algorithm for constructing an M-array, fixing the length of the alphabet, the window property size, the dimensions of the array and the Hamming distance between every window. Previously all the methods worked with a Hamming distance of one, which did not allow error correction. In fact, the arrays used by Morano et al. are simply *perfect submaps* since not all the possible windows are included.

The algorithm used to generate an array with fixed properties is based on a brute-force approach. For example, when constructing an M-array based on three colours with window property of 3×3 the following steps are taken: first, a subarray of 3×3 is chosen randomly and is placed in the north-west vertex of the M-array that is being built. Then consecutive random columns of 1×3 are added to the right of this initial subarray, maintaining the integrity of the window property of the array and the Hamming distance between windows. Afterwards, rows of 3×1 are added beneath the initial subarray in a similar way. Then, both horizontal and vertical processes are repeated by incrementing the starting coordinates by one, until the whole array is filled. Whenever the process reaches a state where no possible elements can be placed, while accomplishing the global window property, the array is cleared and the algorithm starts again with another initial subarray. The basic steps of the algorithm are represented in Fig. 4c. The study of the performance of this algorithm showed that using M-arrays of 45×45 pixels with window property of 3×3 using 3 or more colours, fixed Hamming distances between windows from the typical 1 up to 4 can be generated. Moreover, in most cases, multiple solutions can be found.

Once the generated pattern is projected onto the measuring surface, it must be recovered and the dots must be well labelled in order to find correspondences between the camera and the projector. As every dot is contained in nine windows, the authors applied a voting algorithm where every window proposes a codeword of length 9 (which indicates its position in the pattern) for every one of its elements. Then every observed dot has up to nine codewords proposed by every window to which it belongs. The codeword with the maximum number of votes is the more reliable, so it is used to label the dot. The results showed that when using a Hamming distance of 3 instead of 1, the number of dot mislabellings decreases, due to the possibility of correcting one error per window.

Another interesting contribution made by Morano et al. was to note that a system based on M-arrays can also be used when colour cannot be projected (because the scene is too colourful or because a colour camera is not available). If N colours are used to encode the M-array, $\log_2(N) + 1$ patterns can be projected, encoding every colour with a binary codeword. The system becomes more robust since only two intensity levels are used but it is limited to static scenes.

5. Direct codification

There are certain ways of creating a pattern so that every pixel can be labelled by the information represented on it. Therefore, the entire codeword for a given point is contained in a unique pixel. In order to achieve this, it is necessary to use either a large range of colour values or introduce periodicity. In theory, a high resolution of 3D information can be obtained. However, the sensitivity to noise is very high because the "distance" between "codewords", i.e. the colours used, is nearly zero. Moreover, the perceived colours depend not only on the projected colours, but also on the intrinsic colour of the measuring surface. This means, in most cases, that one or more reference images must be taken. Therefore, these techniques are not typically suitable for dynamic scenes. Direct codification is usually constrained to neutral colour or pale objects. For this reason, it is necessary to perceive and identify the whole spectrum of colours, which requires a "tuning" stage that is not always easy to achieve (depending on the devices used).

We shall now discuss two groups of methods using direct codification: (a) codification based on grey levels: a spectrum of grey levels is used to encode the points of the pattern; (b) codification based on colour: these techniques take advantage of a large spectrum of colours.

5.1. Codification based on grey levels

Carrihill and Hummel [46] developed a system called *intensity ratio depth sensor*. It consists of a linear wedge spread along vertical columns containing a scale of grey levels. A ratio is calculated between every pixel of the perceived wedge and the same pixel value under constant illumination. This ratio is related to the column of the pattern that has been projected in the pixel. Since two patterns must be projected dynamic scenes are not considered. The authors used a slide projector and a monochrome camera with 8 bits of intensity per pixel. The authors aimed to tune the setup so that the relationship between the ratio and the image column number was nearly linear. However, Carrihill and Hummel achieved poor accuracy in their measurements, with a mean error of about 1 cm. This was due to the high sensitivity to noise and non-linearities of the projector device.

In 2000, Miyasaka et al. [47] reproduced the intensity ratio depth sensor by using an LCD projector and a 3CCD camera. With this setup, more accurate results were obtained. The authors took into account that the reflectance of the surface points is not constant for all the light frequencies and each RGB channel of the camera was treated independently. Furthermore, only a narrower band of light frequencies was considered.

In 1995, Chazan and Kiryati [48] carried out experiments using an extension of the Carrihil and Hummel method called *pyramidal intensity-ratio depth sensor*, also known as the *sawtooth sensor*. The motivation behind this new approach was the high sensitivity to noise of the original

Shenzhen Jimuyida
Exhibit 1009

method. As a wide intensity spectrum is projected in only one shot, the camera must be able to perceive such a spectrum nearly linearly, which is very difficult to achieve using an LCD projector. The new method consisted of consecutively projecting the linear wedge by increasing its period. Therefore, the first pattern is a simple-period wedge from black to white. The second contains two linear wedges, the third contains four wedges and so on. At the end, the last pattern contains 2^n linear wedges. Since every period is a linear wedge from black to white, the last pattern uses less grey levels in each period. This means that adjacent grey levels in the last pattern are less similar and easily distinguishable. However, since periodicity is present, the grey level of a certain pixel in the last perceived pattern is not enough to decode its position. To resolve the ambiguity the previously viewed patterns are used. This strategy is quite similar to the time-multiplexing techniques, but in this case, the exact codewords are not recovered. Moreover, since the sharp transitions between periods can lead to high errors, every periodic pattern is projected twice, shifting it by half a period. Then when reading grey levels close to a period transition (black or white), the corresponding shifted pattern is used to avoid the period transition. For every projected pattern an image of the scene is grabbed. Then an intensity ratio is calculated for every image with respect to a constant light image. The *sawtooth sensor* is more accurate than the classic *intensity ratio depth sensor*. Experiments made by the authors over distances of about 80 cm show that the typical errors of 1 cm of the *intensity ratio depth sensor* can be reduced to 1 mm with the *pyramidal intensity ratio depth sensor*. However, the number of patterns increases substantially.

Prior to Chazan and Kiryati's work, in 1993, Hung proposed a grey level sinusoidal pattern [49]. The author pointed out that the period of the observed pattern increases proportionally with the distance between the projector and the object, and therefore, the frequency decreases. The idea was to estimate the instant frequency in every pixel of the camera image and then depth can be calculated for every pixel. The system was tested with synthetic images with gaussian noise. Although the results were good, real experiments should be completed taking into consideration the non-linear behaviour of the devices. Furthermore, since the pattern is periodic, ambiguity problems can arise.

5.2. Codification based on colour

The methods belonging to this group use the same principle as the ones discussed in Section 5.1. However, colour is used to encode pixels instead of using grey levels. For instance, Tajima and Iwakawa [50] presented the rainbow pattern. A large set of vertical narrow slits were encoded with different wavelengths, so that a large sampling of the spectrum from red to blue was projected. In order to project this spectrum, a nematic liquid crystal was used to diffract white light. The images were grabbed by a monochromatic camera with 11 bits of intensity depth. Two images of the scene were taken through two different colour filters. By calculating the ratio between both images an index for every pixel is obtained that does not depend on illumination, nor on the scene colour. Geng [51] improved on this approach by using a CCD camera and a linear variable wavelength filter in front of it. Hence, only a single image of the measuring surface had to be captured from the scene.

In 1999, Sato presented the *multispectral pattern projection range finder* [52]. In this work, the author discussed the complicated optical system required for the rainbow range finder of Tajima. Sato proposed a new technique that only needs an LCD projector and a CCD camera. Moreover, the new technique could eliminate the colour of the measuring surface, so the results were not affected by the spectral reflectance of the surface. The technique consisted of projecting a periodic rainbow pattern 3 times, shifting the hue phase 1/3 of its period in every projection. An extra image was synthesised by a certain linear combination of the three grabbed images. Afterwards, Sato demonstrated that the Hue value of every pixel of the synthesised image is equal to the projected Hue value in the first pattern. Therefore, correspondences between synthesised image pixels and projected rainbow columns can be done. In order to get a good resolution, the pattern had to be periodic, so the identification of the periods is a key point in the decoding stage.

Wust and Capson, in 1991, presented a technique based on a three-step phase shift method [53]. However, instead of projecting three times a periodic pattern shifted in every projection, a single pattern was used. The pattern was designed with three overlapping sinusoids shifted between them, in order to encode the columns of every row. The first sinusoid is represented with red, the second is shifted 90° and is represented with green, and the third, which is shifted 90° with respect to the green one, is represented with blue. Then, once the pattern is projected and an image is grabbed, the phase shift can be calculated for every pixel using the following equation:

$$\Phi(x, y) = \arctan\left(\frac{I_r - I_g}{I_g - I_b}\right), \tag{11}$$

$\Phi(x, y)$ is the phase in a given pixel where the intensities of the red, green and blue are denoted as I_r, I_g and I_b respectively. The technique of Wust and Capson requires only a unique pattern, so moving surfaces can be measured. However, the surface must be predominantly colour neutral and must not contain large discontinuities.

6. Experimental results

We implemented a set of seven representative techniques taken from the proposed classification groups. All the techniques have been tested under the same conditions in order to evaluate their advantages and constraints.

Shenzhen Jimuyida
Exhibit 1009

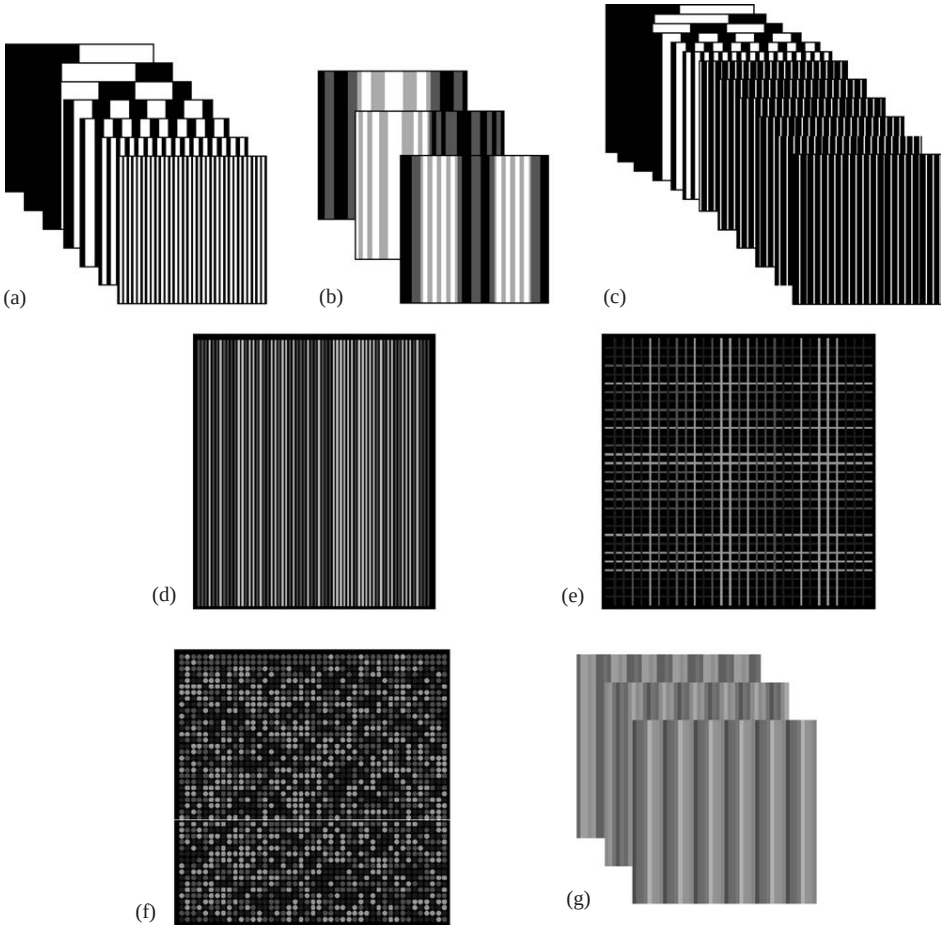

Fig. 5. Patterns corresponding to the implemented techniques: (a) Posdamer; (b) Horn and Kiryati; (c) Gühring; (d) De Bruijn; (e) Salvi; (f) Morano; (g) Sato.

A low-cost structured light system was used. It is composed of an LCD video projector (Mitsubishi XL1U) working at 1024×768 pixels, a camera (Sony 3CCD) and a frame grabber (Matrox Meteor-II) digitizing images at 768×576 pixels and 3×8 bits per pixel. A standard PC was used for implementing the algorithms. Both camera and video projector were calibrated with a linear pinhole model [3].

The patterns chosen are determined from a set of parameters (number of colours or grey levels, number of projecting patterns and spatial resolution) constituting just a sample of every technique. The purpose of the experimental results is mainly to compare every technique with respect to the others rather than going deeper into the complexity of fixing the optimal parameters depending on the measuring surface. Therefore, the pattern parameters we have used are those that maximise the resolution of the vision system and minimise the pattern segmentation complexity for every technique. Moreover, the variety of resolutions that have been programmed permits appreciation of how the reconstruction results change according to this parameter. The implemented techniques are listed below and are represented in Fig. 5.

- Time-multiplexing:
  - Posdamer: stripe patterns encoded with Gray code of 7 bits so that 128 stripes are encoded [5].
  - Gühring: the *line shifting* technique using six Gray coded patterns and 21 slits shifted six times [19].
  - Horn: three patterns encoding 64 stripes by using four grey levels [14].
- Spatial neighbourhood:
  - De Bruijn: a pattern with 64 vertical slits encoded with a *De Bruijn* sequence of third order and four colours.
  - Salvi: a grid pattern of 29×29 slits encoded with a De Bruijn sequence of third order and three colours [34].
  - Morano: pattern consisting of colour dots encoded with an M-array of 45×45 elements and three colours [45].

Shenzhen Jimuyida
Exhibit 1009

Table 2
Quantitative results

Technique	Stdev (μm)	3D Points	Resolution (%)	patterns
Posdamer	37.6	25213	21.67	9
Horn	9.6	12988	11.17	5
Gühring	4.9	27214	23.38	14
De Bruijn	13.1	13899	11.94	1
Salvi	72.3	372	0.32	1
Morano	23.6	926	0.80	1
Sato	11.9	10204	8.77	3

The headings are: author's name of the technique; standard deviation of the reconstructing error; average number of 3D points; % of pixels from images that have been reconstructed in average; number of projected patterns.

- Direct codification:
  - Sato: the three periodic patterns proposed by Sato [52] were employed.

The performance of the techniques was evaluated by means of quantitative and qualitative tests. In the following subsections the experiments and their results are presented.

6.1. Quantitative evaluation

A white plain (flat surface) at a distance of about 120 cm to the camera was reconstructed 30 times using all the implemented techniques. A multiple regression was applied in order to obtain the equation of the 3D plain for every technique and for every reconstruction. The same experiment was repeated by bringing the plain closer to the camera by about 40 mm. Then the average and the standard deviation of the distance between both plains was calculated for every technique. The results of the experiment are shown in Table 2. The table includes the standard deviation, in μm, of the average distance between both parallel plains, the average number of 3D points that were reconstructed, the % of image pixels inside a region of 515 × 226 pixels that were decoded, and the total number of projected patterns for every technique (including white and black patterns for intensity normalisation when needed).

6.2. Qualitative evaluation

In order to evaluate the performance of the techniques, it is also useful to observe the reconstruction of certain surfaces and analyse them from a qualitative point of view. For this purpose, two surfaces were reconstructed.

The first surface was a statue of a white horse of dimensions $21 \times 15 \times 10$ cm. The statue and the reconstructions obtained are shown in Fig. 6. The reconstructions are presented both as clouds of points and rendered surfaces. Techniques with higher resolution (time-multiplexing techniques and De Bruijn patterns based on a single axis codification) enable

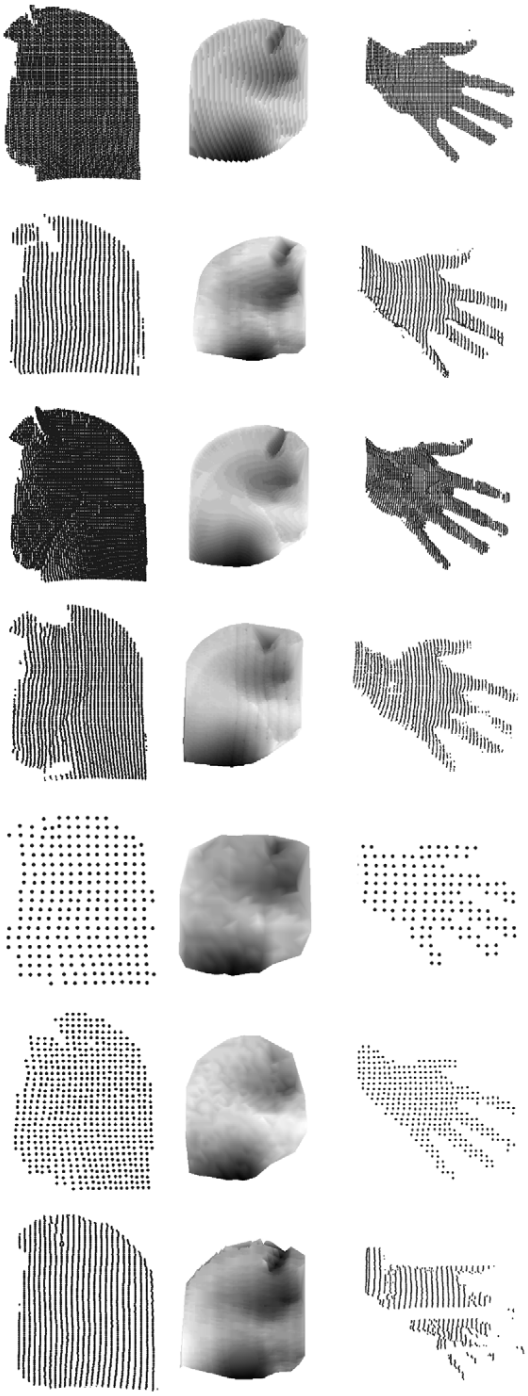

Fig. 6. Reconstruction results for every one of the seven implemented techniques. Top to bottom: Posdamer, Gühring, Horn, De Bruijn, Salvi, Morano and Sato. At left, the cloud of points corresponding to the horse statue reconstruction. In the middle, the corresponding rendered surface from another view point. At right, the cloud of points from reconstruction of a human hand.

Shenzhen Jimuyida
Exhibit 1009

details of the horse's profile to be distinguished, while other techniques with lower resolution (mainly based on spatial neighbourhood) obtain basically the global profile.

The second test consisted of reconstructing a human hand. This surface is useful for evaluating the performance of the techniques when the surface violates monotonicity, i.e. it contains discontinuities. In this case, the discontinuities are produced by the gaps between the fingers. Results are shown in Fig. 6. Techniques based on time-multiplexing are not affected since for recovering the codewords of a pixel, it is only necessary to gather its value along the projected patterns. Techniques based on spatial neighbourhood using a single axis codification (De Bruijn) suffer large amounts of data loss as the local smoothness assumption of the measuring surface is violated. Nevertheless, techniques that encode both pattern axis (Morano and Salvi) can identify some regions near discontinuities due to the propagation of codewords among adjacent points. Direct coding techniques should be robust against discontinuities if no periodicity is used in the patterns. Since the technique proposed by Sato exploits periodicity, it fails when reconstructing the fingers. Furthermore, periodicity is required for such techniques in order to reduce the number of colours in the pattern as it is very difficult to correctly differentiate among the emitted colours if a large spectrum is used.

The experiments that have been carried out allow comparison of the different groups of techniques classified in the paper. It has been shown that techniques based on time-multiplexing achieve the most accurate results. Moreover, line-shifting combined with Gray Code permits exploitation of the whole theoretical resolution of patterns. The results also demonstrate that locating the pattern stripes with sub-pixel accuracy (in the case of Gühring [19] and Horn [14] implementations), leads to better results than using pixel accuracy (in the case of Posdamer [5] current implementation). Techniques based on spatial neighbourhood have also obtained satisfactory results. For example, the pattern consisting of vertical slits coded with a De Bruijn sequence has obtained very accurate measurements as the slits are also detected with sub-pixel accuracy. However, it has failed when measuring discontinuities. Such problems could be partially solved by using dynamic programming [13]. Furthermore, techniques based on both axis codification, i.e. the grid by Salvi et al. [34] and the array of dots by Morano et al. [45], are more robust against discontinuities as redundancy in the coding strategy permits extension of decoded regions to contiguous non-decoded regions. Finally, the direct coded pattern presented by Sato [52] has obtained very accurate results (also locating the stripes with sub-pixel accuracy) and robustness against colourful surfaces. However, this technique has the problem of stripe decodification among the pattern due to its periodic structure when a surface containing discontinuities is measured. Such problems could be overcome by projecting some Gray patterns to remove the ambiguity between periods.

7. Conclusions

We have presented a comprehensive survey of coded structured light techniques. A new classification of the reviewed techniques has been proposed from the point of view of the coding strategies used to generate the projected patterns.

Time-multiplexing was the first paradigm of coded structured light used to obtain 3D data from an unknown surface. The advantages of these techniques are the easy implementation, the high spatial resolution and the accurate 3D measurements that can be achieved. The main drawback is their inapplicability to moving surfaces since multiple patterns must be projected. Techniques based on projecting stripe patterns encoded with Gray code can obtain very good accuracy, but maximum resolution cannot be achieved. In order to obtain maximum resolution, a technique based on a combination of Gray code and Phase shifting must be used. In this subgroup, the technique proposed by Gühring [19] must be highlighted. Its drawback however, is the large number of projecting patterns (32 patterns when using maximum resolution). If maximum spatial resolution is not the principal aim of the application, but rather the minimisation of the number of projecting patterns, a technique based on n-ary codes is appropriate. Such methods obtain an accuracy equal to or even better than a Gray code approach, reducing exponentially the number of projecting patterns. For example, a Gray code technique based on the projection of 8 patterns can encode 256 stripes, while an n-ary technique only requires three patterns and 13 grey levels or colours to obtain such resolution, for $n = 13$. However, the system using n-ary codes must be calibrated in order to correctly differentiate among the set of grey levels or colours used. If a good calibration cannot be achieved, then it is recommended to reduce the number of grey levels or colours by projecting more patterns.

Spatial neighbourhood coding is the second big group of coded structured light techniques. The advantage compared with time-multiplexing is that such strategy permits, in most cases, moving surfaces to be measured. However, since the codification must be condensed in a unique pattern, the spatial resolution is lower. Moreover, local smoothness of the measuring surface is assumed in order to correctly decode the pixel neighbourhoods. Since this local smoothness is not always accomplished, errors in the decoding stage can arise producing errors in the 3D measurements. In order to minimise such errors, the algorithms of the decoding stage must be more robust, resulting in an increase in the overall complexity of the technique. Techniques which define the neighbourhoods empirically usually present pattern periodicity or repetition of neighbourhoods, which is not recommended. Such problems have been eliminated by strategies based on De Bruijn sequences and M-arrays.

Techniques based on a unique pattern coded using a De Bruijn sequence have a trade off between the length of the sequence, i.e. the resolution of the system, the number of

Shenzhen Jimuyida
Exhibit 1009

colours involved and the size of the window property. Most of these methods use either horizontal or vertical windows with a limited size in order to preserve the local smoothness assumption of the measuring surface. If the window size is not too big (in our opinion a good limit is about 10% of the sequence length), more than two colours must be used with the aim of preserving a good resolution. The number of colours used increases the noise sensitivity when measuring colourful scenes. Using up to six colours is not very problematic. With five colours and a window size of 3, a resolution of 125 slits per pattern (similar to a Gray code system based on seven patterns) can easily be achieved with a robust decoding stage. The most complete technique that can be found in the bibliography is the one proposed by Zhang et al. [13]. This technique takes into account that disorders between elements of the sequence can occur when projecting the pattern. The solution proposed is based on multi-pass dynamic programming, which seems the most robust way to recover the original sequence. Furthermore, techniques based on *M-arrays* are more difficult to generate. However, since every coded point has both row and column codewords, a higher degree of redundancy is included. In order to take advantage of this redundancy, an additional step must be programmed in the decoding stage for validating the codeword of every coded point. Similar trade offs to the ones involved when using De Bruijn sequences also appear with M-arrays-based patterns. The segmentation complexity of the observed patterns in such techniques must also be addressed. The most typical representations of an M-array in a pattern are the grid representation and the array of dots. In our opinion, the grid representation can be segmented more easily by edge detection. The encoded points are the intersection of edges, so they can be found very accurately. In addition, when projecting dots, their mass centres must be located. So it is important to detect when a dot appears only partially in the image, since its mass centre will be incorrect. Moreover, the grid techniques allow adjacent cross-points to be located by tracking the edges, while with the dot representation, some sort of euclidian distance must be used to locate the neighbours of a given dot.

It should be noted that a technique based on spatial neighbourhood can always be translated to a time-multiplexed technique by expressing the colours in binary intensity levels distributed over a sequence of patterns.

Direct coding techniques are useful for achieving large spatial resolution and few projecting patterns. However, these techniques present a lot of drawbacks. Firstly, the limited bandwidth of LCD projectors provokes integration of intensities over adjacent pixels. Secondly, variations of light intensities due to the different colours and depths of the measuring surface. Finally, the error quantisation introduced by the camera, which is very sensitive to noise. Therefore, the correct identification of every projected intensity or colour is not easy to achieve. In most cases, the use of such techniques requires a device that projects a unique wavelength for every grey level or colour. Therefore, LCD projectors are not suitable for such a purpose. Some authors use non-standard optical devices to decompose white light, producing monochromatic light planes. Furthermore, since a large spectrum of wavelengths are used, cameras with large depth-per-pixel must be considered (about 11 bits per pixel) for accurate quantisation. It should be noted that most of these techniques cannot measure moving scenes because they need additional patterns to normalise intensity or colour. In addition, these techniques are usually limited to colour-neutral surfaces. Nevertheless, some techniques that can be implemented with an LCD projector and a standard CCD camera were proposed by Wust and Capson [53] and Sato [52]. Moreover, both techniques are theoretically capable of reconstructing colourful surfaces, and the technique by Wust and Capson can also measure moving surfaces. Accordingly, experimental results given by the technique by Sato showed that, in the synthesised image, most part of the surface colours are eliminated.

Finally, the depth per pixel used by a coded structured light technique is also an important parameter. The noisier the application environment in which the technique will be applied, the smaller the number of grey levels or colours used should be. Therefore, time-multiplexing techniques based on binary patterns are the most robust against noise. However, when increasing the number of grey levels or colours, differentiating the slits becomes more difficult. The non-linearities of the light spectrum of the projector and the spectral response of the cameras, and the non-uniform albedo of the measuring surface mean that the read colours hardly match with the projected ones. In order to overcome this problem, a full colourimetric calibration procedure should be considered. The illumination model proposed by Caspi et al. [12] or even a simple linear normalisation may be a good solution.

8. Summary

This article is an exhaustive and comprehensive survey of coded structured light techniques used for 3D measurement of surfaces. Coded structured light belongs to the group of active stereo vision systems and it is considered an important and widely used active shape acquisition technique.

3D shape acquisition techniques have a large number of applications since permit to measure the profile of a given surface. For instance, in order to manufacture an existing industrial part whose model is not available, a method of 3D measurement can be applied for obtaining the desired CAD model leading to fast prototyping. Another industrial application is the inspection of manufactured parts. Therefore, a quality control can be fulfilled by comparing the 3D measured surface of every manufactured part obtained by visual inspection with the original model of the desired prototype. 3D shape digitisers have also been used for reproducing museum artifacts for souvenir industries, digital museums and scientific studies. Moreover, the field of application of 3D measurement is also useful in robotics. For example, object

Shenzhen Jimuyida
Exhibit 1009

recognition in robot navigation and 3D map building are some of the potential applications in this field.

Among all the shape acquisition techniques, active optical stereo is typically based on an emitter which projects some sort of structured light on the measuring surface which is viewed by a CCD camera. Correspondences between the camera image and the projected pattern can easily be found. These correspondences lead to obtain a 3D measurement of the surface by means of triangulation. The number of correspondences that can be recovered in a single image depends basically on the sort of pattern projected and, therefore, on the light emitter device chosen. Laser scanners have been widely used for projecting monochromatic plains. However, the main disadvantage of these systems is the large number of images that must be acquired during the scanning process. The inclusion of some sort of codification in the illumination patterns permits to project a dense pattern on the measuring surface. Although slide projectors were initially used for projecting coded patterns, nowadays they have been replaced by video projectors. The advantage of video projectors are the feasibility to implement low-cost 3D acquisition systems and the flexibility for changing the projecting patterns depending on the measuring surface easily.

This article describes in detail a large number of coded structured light techniques for implementing flexible active stereo systems. The existing techniques are distributed in a new classification according to the principle of codification used to design the illumination patterns. A representative set of techniques from the proposed classification have been implemented in the paper. The experimental results allows us to discuss the advantages and constraints of every coding strategy from the point of view of accuracy, resolution in the 3D measurements and qualitative results of the 3D surface obtained.

This article is a useful guide for easily finding the most suitable pattern codification technique for a given application.

References

[1] R. Jarvis, Range sensing for computer vision, Advances in Image Communications, Elsevier Science Publishers, Amsterdam, 1993, pp. 17–56.

[2] O. Faugeras, Three-Dimensional Computer Vision, MIT Press, Cambridge, MA, 1993.

[3] J. Salvi, X. Armangué, J. Batlle, A comparative review of camera calibrating methods with accuracy evaluation, Pattern Recognition 35 (7) (2002) 1617–1635.

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

[5] J.L. Posdamer, M.D. Altschuler, Surface measurement by space-encoded projected beam systems, Comput. Graph. Image Process. 18 (1) (1982) 1–17.

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

[7] M. Minou, T. Kanade, T. Sakai, A method of time-coded parallel planes of light for depth measurement, Trans. IECE Jpn. 64 (1981) 521–528.

[8] M. Trobina, Error model of a coded-light range sensor, Technical Report, Communication Technology Laboratory, ETH Zentrum, Zurich, 1995.

[9] R.J. Valkenburg, A.M. McIvor, Accurate 3d measurement using a structured light system, Image Vision Comput. 16 (2) (1998) 99–110.

[10] D. Skocaj, A. Leonardis, Range image acquisition of objects with non-uniform albedo using structured light range sensor, in: Proceedings of the 15th International Conference on Pattern Recognition, Vol. 1, 2000, pp. 778–781.

[11] C. Rocchini, P. Cignoni, C. Montani, P. Pingi, R. Scopigno, A low cost 3D scanner based on structured light, in: A. Chalmers, T.-M. Rhyne (Eds.), EG 2001 Proceedings, Vol. 20 (3), Blackwell Publishing, Oxford, 2001, pp. 299–308.

[12] D. Caspi, N. Kiryati, J. Shamir, Range imaging with adaptive color structured light, Pattern Anal. Mach. Intell. 20 (5) (1998) 470–480.

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

[14] E. Horn, N. Kiryati, Toward optimal structured light patterns, Image Vision Comput. 17 (2) (1999) 87–97.

[15] H. Sagan, Space Filling Curves, Springer, New York, 1994.

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

[17] G. Sansoni, S. Lazzari, S. Peli, F. Docchio, 3d imager for dimensional gauging of industrial workpieces: state of the art of the development of a robust and versatile system, in: International Conference on Recent Advances in 3-D Digital Imaging and Modeling, Ottawa, Ont., Canada, 1997, pp. 19–26.

[18] G. Wiora, High resolution measurement of phase-shift amplitude and numeric object phase calculation, in: Proceedings Vision Geometry IX, Vol. 4117, SPIE, Bellingham, Washington, USA, 2000, pp. 289–299.

[19] J. Gühring, Dense 3-d surface acquisition by structured light using off-the-shelf components, Videometrics Opt. Meth. 3D Shape Meas. 4309 (2001) 220–231.

[20] E. Trucco, R.B. Fisher, A.W. Fitzgibbon, D.K. Naidu, Calibration, data consistency and model acquisition with laser stripers, Int. J. Comput. Intregrated Manufactur. 11 (4) (1998) 293–310.

[21] K. Sato, Range imaging based on moving pattern light and spatio-temporal matched filter, in: IEEE International Conference on Image Processing, Vol. 1, 1996, pp. 33–36.

[22] O. Hall-Holt, S. Rusinkiewicz, Stripe boundary codes for real-time structured-light range scanning of moving objects, in: The Eighth IEEE International Conference on Computer Vision, 2001, pp. II: 359–366.

[23] M. Maruyama, S. Abe, Range sensing by projecting multiple slits with random cuts, Pattern Analy. Mach. Intell. 15 (6) (1993) 647–651.

[24] N.G. Durdle, J. Thayyoor, V.J. Raso, An improved structured light technique for surface reconstruction of the human trunk,

Shenzhen Jimuyida
Exhibit 1009

in: IEEE Canadian Conference on Electrical and Computer Engineering, Vol. 2, 1998, pp. 874–877.
[25] M. Ito, A. Ishii, A three-level checkerboard pattern (tcp) projection method for curved surface measurement, Pattern Recognition 28 (1) (1995) 27–40.
[26] K.L. Boyer, A.C. Kak, Color-encoded structured light for rapid active ranging, IEEE Trans. Pattern Analy. Mach. Intell. 9 (1) (1987) 14–28.
[27] C. Chen, Y. Hung, C. Chiang, J. Wu, Range data acquisition using color structured lighting and stereo vision, Image Vision Comput. 15 (1997) 445–456.
[28] F.J. MacWilliams, N.J.A. Sloane, Pseudorandom sequences and arrays, Proc. IEEE 64 (12) (1976) 1715–1729.
[29] H. Fredricksen, A survey of full length nonlinear shift register cycle algorithms, Soc. Ind. Appl. Math. Rev. 24 (2) (1982) 195–221.
[30] H. Hügli, G. Maître, Generation and use of color pseudo random sequences for coding structured light in active ranging, in: Proceedings of Industrial Inspection, Vol. 1010, 1989, pp. 75–82.
[31] T. Monks, J. Carter, Improved stripe matching for colour encoded structured light, in: Fifth International Conference on Computer Analysis of Images and Patterns, 1993, pp. 476–485.
[32] P. Vuylsteke, A. Oosterlinck, Range image acquisition with a single binary-encoded light pattern, Pattern Analy. Mach. Intell. 12 (2) (1990) 148–163.
[33] T. Pajdla, Bcrf—binary-coded illumination range finder reimplementation, Technical Report KUL/ESAT/MI2/9502, Katholieke Universiteit Leuven, ESAT, Kardinaal Mercierlaan 94, B-3001 Leuven, April 1995.
[34] J. Salvi, J. Batlle, E. Mouaddib, A robust-coded pattern projection for dynamic 3d scene measurement, Pattern Recognition Lett. (19) (1998) 1055–1065.
[35] E.M. Petriu, Z. Sakr, S.H.J.W., A. Moica, Object recognition using pseudo-random color encoded structured light, in: Proceedings of the 17th IEEE Instrumentation and Measurement technology Conference, Vol. 3, 2000, pp. 1237–1241.
[36] P. Lavoie, D. Ionescu, E. Petriu, A high precision 3D object reconstruction method using a color coded grid and nurbs, in: Proceedings of the International Conference on Image Analysis and Processing, Venice, Italy, 1999, pp. 370–375.
[37] T. Etzion, Constructions for perfect maps and pseudorandom arrays, IEEE Trans. Inform. Theory 34 (5) (1988) 1308–1316.
[38] H. Morita, K. Yajima, S. Sakata, Reconstruction of surfaces of 3-d objects by *m*-array pattern projection method, in: IEEE International Conference on Computer Vision, 1988, pp. 468–473.
[39] E.M. Petriu, T. Bieseman, N. Trif, W.S. McMath, S.K. Yeung, Visual object recognition using pseudo-random grid encoding, in: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, 1992, pp. 1617–1624.
[40] S. Kiyasu, H. Hoshino, K. Yano, S. Fujimura, Measurement of the 3-D shape of specular polyhedrons using an *m*-array coded light source, IEEE Trans. Instrum. Meas. 44 (3) (1995) 775–778.
[41] H.J.W. Spoelder, F.M. Vos, E.M. Petriu, F.C.A. Groen, Some aspects of pseudo random binary array-based surface characterization, IEEE Trans. Instrum. Meas. 49 (6) (2000) 1331–1336.
[42] P. Griffin, L. Narasimhan, S. Yee, Generation of uniquely encoded light patterns for range data acquisition, Pattern Recognition 25 (6) (1992) 609–616.
[43] Y.C. Hsieh, Decoding structured light patterns for three-dimensional imaging systems, Pattern Recognition 34 (2001) 343–349.
[44] C.J. Davies, M.S. Nixon, A hough transform for detecting the location and orientation of 3-dimensional surfaces via color encoded spots, IEEE Trans. systems, man cybernet. 28 (1) (1998) 90–95.
[45] R.A. Morano, C. Ozturk, R. Conn, S. Dubin, S. Zietz, J. Nissanov, Structured light using pseudorandom codes, Pattern Analy. Mach. Intell. 20 (3) (1998) 322–327.
[46] B. Carrihill, R. Hummel, Experiments with the intensity ratio depth sensor, in: Computer Vision, Graphics and Image Processing, Vol. 32, Academic Press, New York, 1985, pp. 337–358.
[47] T. Miyasaka, K. Kuroda, M. Hirose, K. Araki, High speed 3-D measurement system using incoherent light source for human performance analysis, in: Proceedings of the 19th Congress of The International Society for Photogrammetry and Remote Sensing, The Netherlands, Amsterdam, 2000, pp. 65–69.
[48] G. Chazan, N. Kiryati, Pyramidal intensity-ratio depth sensor, Technical Report 121, Center for Communication and Information Technologies, Department of Electrical Engineering, Technion, Haifa, Israel, October 1995.
[49] D. Hung, 3d scene modelling by sinusoid encoded illumination, Image Vision Comput. 11 (1993) 251–256.
[50] J. Tajima, M. Iwakawa, 3-D data acquisition by rainbow range finder, in: International Conference on Pattern Recognition, 1990, pp. 309–313.
[51] Z.J. Geng, Rainbow 3-dimensional camera: new concept of high-speed 3-dimensional vision systems, Opt. Eng. 35 (2) (1996) 376–383.
[52] T. Sato, Multispectral pattern projection range finder, in: Proceedings of the Conference on Three-Dimensional Image Capture and Applications II, Vol. 3640, SPIE, San Jose, California, 1999, pp. 28–37.
[53] C. Wust, D.W. Capson, Surface profile measurement using color fringe projection, Mach. Vision Appl. 4 (1991) 193–203.

About the Author—JOAQUIM SALVI graduated in Computer Science in the Polytechnical University of Catalunya in 1993, he received the M.S. degree in Computer Science in July 1996 and the Ph.D. in Industrial Engineering in January 1998 in the University of Girona. He received the best thesis award in Industrial Engineering of the University of Girona. At present, he is an associate professor in the Electronics, Computer Engineering and Automation Department of the University of Girona. His current interests are in the field of computer vision and mobile robotics, focused on structured light, stereovision and camera calibration.

Shenzhen Jimuyida
Exhibit 1009

About the Author—JORDI PAGÈS received the B.S. degree in Computer Science in the University of Girona in 2001 before joining the Computer Vision and Robotics Group. In 2003 he received the M.S. degree in Computer Science. At preset he is involved in the study of active stereovision applied to the visual Servoing field and he is working for his Ph.D. in the Computer Vision and Robotics Group in the University of Girona.

About the Author—JOAN BATLLE graduated in Physics at the Autonomous University of Barcelona, received the Ph.D. degree in Computer Science at the Polytechnical University of Catalunya. At present, he is the rector of the University of Girona and the leader of the Computer Vision and Robotics Group. His research activity is mainly focused on real-time vision and autonomous robots, and he is involved in some governmental projects about underwater robots and technology transference to industry.

Shenzhen Jimuyida
Exhibit 1009