Range Imaging with Adaptive Color Structured Light

D. Caspi N. Kiryati J. Shamir

Copyright 1998 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

Range Imaging with Adaptive Color Structured Light

D. Caspi N. Kiryati J. Shamir

Abstract

In range sensing with time-multiplexed structured light, there is a trade-off between accuracy, robustness and the acquisition period. The acquisition period is lower bounded by the product of the number of projection patterns and the time needed for acquiring a single image. In this paper a novel structured light method is described. Adaptation of the number and form of the projection patterns to the characteristics of the scene takes place as part of the acquisition process. Noise margins are matched to the actual noise level, thus reducing the number of projection patterns to the necessary minimum. Color is used for light plane labeling. The dimension of the pattern space (and the noise margins) are thus increased without raising the number of projection patterns. It is shown that the color of an impinging light plane can be identified from the image of the illuminated scene, even with colorful scenes. Identification is local and does not rely on spatial color sequences. Therefore, in comparison to other color structured light techniques, assumptions about smoothness and color neutrality of the scene can be relaxed. The suggested approach has been implemented and the theoretical results are supported by experiments.

Index terms: color, computer vision, multilevel Gray code, range sensor, shape from X, structured light, video and data projector.

Affiliation of authors: Department of Electrical Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel.

1 Introduction

A triangulation-based structured light system [3] is similar to a passive stereo vision system with one of the cameras replaced by a projector. In one of its conceptually simplest forms, a light source, normally a laser with a cylindrical lens, projects a plane of light that creates a narrow stripe on the scene [2, 16]. Since the intersection of a known illumination plane and a line of sight uniquely determines a point, the 3-D location of all points along the stripe that are visible by the camera can be simultaneously obtained. For dense reconstruction the scene must be scanned by the plane of light, a fairly slow process. Specialized sensors [1, 12] are a faster alternative to standard CCD cameras in this application.

In order to speed up the range sensing process and eliminate mechanical scanning and laser illumination, various systems that use a spatially modulated white light projector have been suggested. In such systems many light planes or rays can be projected simultaneously as parts of a single illumination pattern. Then, however, their correct identification in the image is a crucial problem that must be addressed.

Several single illumination pattern systems in which light planes or rays are identified by certain markings have been suggested. But, since the markings occupy some space that must remain connected in the image of the illuminated scene, the applicability of these techniques is generally limited to "well-behaved" scenes. Blake *et al* [5] have been able to overcome this limitation, but require a carefully aligned *trinocular* system with two projectors and a camera.

High reliability identification of light planes with minimal assumptions about the nature of the scene can be achieved by time multiplexing, i.e., by sequentially projecting several

patterns. A robust and widely applied structured light system that is based on time multiplexing has been described in [14]. Gray code is used to label the light planes, and each illumination pattern represents one of the bit planes of the Gray code labels. Consider a given location in the image plane. The number of projection patterns needed is in principle equal to the number of bits required for light plane labeling at the desired resolution. Since the power of the light source is distributed across the scene, local image brightness is much smaller than in a single ray or a single plane system. In order to account for variability in the reflectivity of surfaces in the scene, spatially varying bit-decision thresholds must be set. They can be derived using two additional images, one taken under full uniform illumination and the other under just the ambient light.

In the intensity ratio depth sensor [6] grey levels are used for light-plane labeling. Two illumination patterns are required. One is a linear intensity wedge that varies horizontally from half to full intensity. The other illumination is uniform. The image taken under wedge illumination is divided pixel-wise by the reference image that was obtained under uniform illumination. Assuming that the brightness of any point in the scene is proportional to the intensity of the light plane on which it lies, the value at any point in the ratio image identifies the corresponding light-plane. This system is quite sensitive to noise. A *pyramidal* intensity ratio scheme that uses a sequence of multiple-wedge patterns to improve the accuracy in a coarse-to-fine manner is described in [8].

The applicability of range sensing systems that are based on time multiplexing is limited to scenes that are static within the acquisition period. Since an image of the scene must be taken under each illumination pattern, the range acquisition period is lower bounded by the product of the number of illumination patterns and the time needed for acquiring a single

image. Because the latter is normally dictated by technology and standards, the number of projection patterns must be as small as possible subject to the application-specific accuracy requirements.

A theoretical framework for the design of projection patterns for stereometric structured light systems has recently been presented by Horn and Kiryati [10]. The design of M projection patterns for a monochrome structured light system with L light planes has been shown to be equivalent to the placement of L points in a M dimensional space subject to distance constraints imposed by the noise and errors. Performance similar or even better than possible with the Gray code technique has been achieved with fewer projection patterns. The key to this gain is the better matching of the noise margins to the actual level of noise and errors. Since the latter is scene dependent, the projection patterns should be redesigned for each reconstruction task or at least whenever there are significant changes in the characteristics of the scenes or in the imaging setup. This has so far been considered impractical.

In this paper a novel structured light method is described. The projection patterns are automatically adapted to the characteristics of the scene as part of the acquisition process, thus maximizing performance and minimizing the number of projection patterns. Color is used for light plane labeling. The dimension of the code space (and the noise margins) are thus increased without raising the number of projection patterns and the acquisition period. Unlike previous color structured light techniques, few assumptions about smoothness and color neutrality of the scene are necessary and there are no compromises with respect to the spatial resolution and the immunity to noise.

2 Color Structured Light

2.1 Previous approaches

The use of color for light plane labeling seems natural. Suppose that a unique color is given to each light plane. Had it been possible to reliably identify the projected colors in the image of the illuminated scene, range imaging using a single projection pattern would have been accomplished. However, unless the color of a given light plane is monochromatic, the spectral composition of the reflected light will depend on the color of the scene as well as on the projected color. Identification of the projected color will then be possible only if the scene is white or if its spectral response is spatially uniform and well known in advance. These are severely limiting assumptions.

In the rainbow range-finder [17, 15] a bundle of monochromatic light planes is generated by dispersing a white illumination source. The wavelength of light reflected from the scene is then identical to that of the illuminating light plane. In principle, by using a color camera, the reflected colors can be identified and triangulation can be carried out. In practice, since the change of wavelength in the light plane bundle is gradual, the estimate of light plane position is very sensitive to noise.

Even if the projected colors are not pure, they can usually be identified in the image as long as they are very different from one another and the color of the scene is neutral. But then only a few colors can be used. Boyer and Kak [4] assumed that the color of the scene is neutral and suggested using color sequences, i.e. blocks of color light planes, to encode spatial position. This reduces the spatial resolution and requires assumptions about the smoothness of the scene. Improvements in that approach have been proposed by Hügli and

Maître [11]. Recently, Chen *et al* [7] suggested a two camera stereo vision system with an added projector. The purpose of the projector is to create artificial color texture patterns on the scene for easier stereo registration.

We suggest a new approach to the use of color in structured light. We show that the color of an impinging light plane can be identified from the image of the illuminated scene. This can be accomplished even if the scene is colorful, so the assumption about color neutrality of the scene can be relaxed. The identification is local and does not rely on spatial sequences, hence powerful smoothness assumptions are not needed. No compromises are made with respect to the spatial resolution and the immunity to noise. This is achieved by utilizing the color characteristics of the projector and the camera, by projecting more than one pattern and by adapting the patterns to the lightness and color of the scene. Since the properties of the projector and the camera play an important role, we begin with an overview of the experimental system.

2.2 The system

Our experimental structured light system consists of a computer-controlled projector, a camera and a computer. The 3-D location of a point in the scene can be measured only if it is visible by both the camera and the projector. Panoramic range imaging could be accomplished by adding projectors and cameras.

The projector is an Epson EMP-3000 video & data color VGA projector designed for presentations. It is connected to the computer in parallel to the monitor via a standard graphics adaptor card. The spatial resolution is 640×480, hence up to 640 distinct light

planes can be projected. The principle of operation of the projector is summarized in Fig. 1. A source of white light is split by dichroic mirrors to three beams, red green and blue. Each beam is modulated by an LCD array. The three beams are then recombined by means of a prism and projected through a common lens. This presentation-type projector proved to be a convenient source of structured light. Its main limitations in this unusual application are the restricted depth of field and geometric lens aberrations, both due to its large aperture.

The system camera is a Sony XC-003P 3-CCD 768×576 analog color camera. The advantage of a 3-CCD camera for color structured light applications is that the full spatial resolution is available in each of the color channels[1]. The spectral response of the camera is shown in Fig. 2. The output of the camera is three analog color channels. The interface with the computer is a Matrox Meteor RGB PCI frame grabber. The computer is a 133MHz Pentium PC clone with the Windows NT 3.51 operating system. The computer generates the projection patterns, acquires the images and carries out the 3-D reconstruction procedure.

2.3 Recovery of projected color

The response of a 3-CCD camera in a given pixel can be modeled as:

$$\begin{aligned} R &= \int_0^\infty f_R(\lambda) I^r(\lambda) d\lambda \\ G &= \int_0^\infty f_G(\lambda) I^r(\lambda) d\lambda \\ B &= \int_0^\infty f_B(\lambda) I^r(\lambda) d\lambda \end{aligned} \tag{1}$$

where

[1]The disadvantage is the slight misalignment between the three arrays.

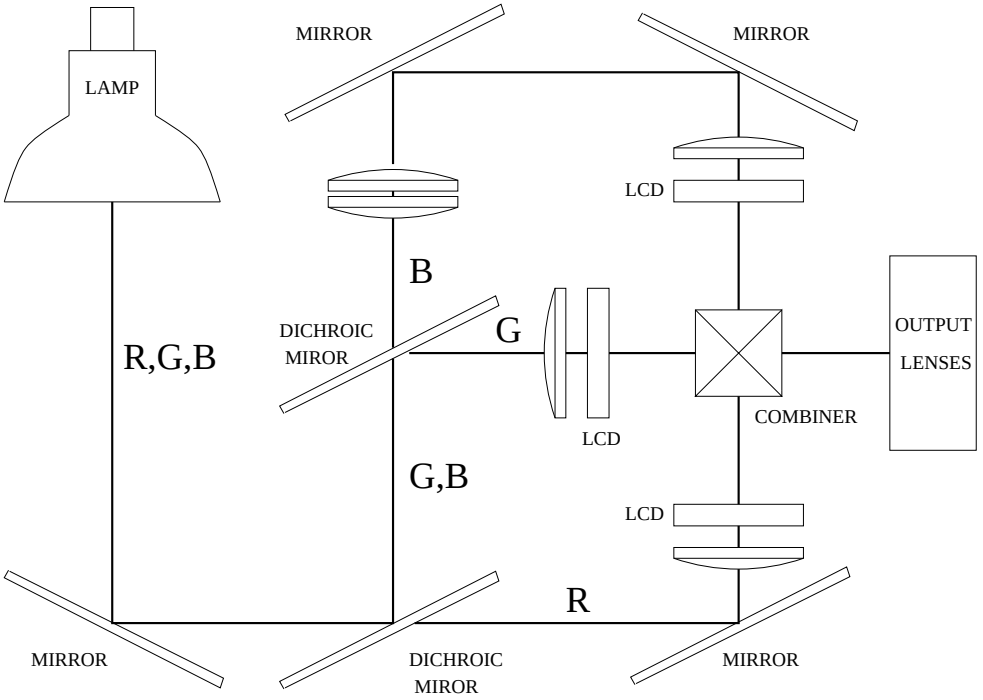

Figure 1: The internal structure of the projector. A source of white light is split by dichroic mirrors to three beams, red green and blue. Each beam is modulated by an LCD array. The three beams are then recombined by means of a prism and projected through a common lens.

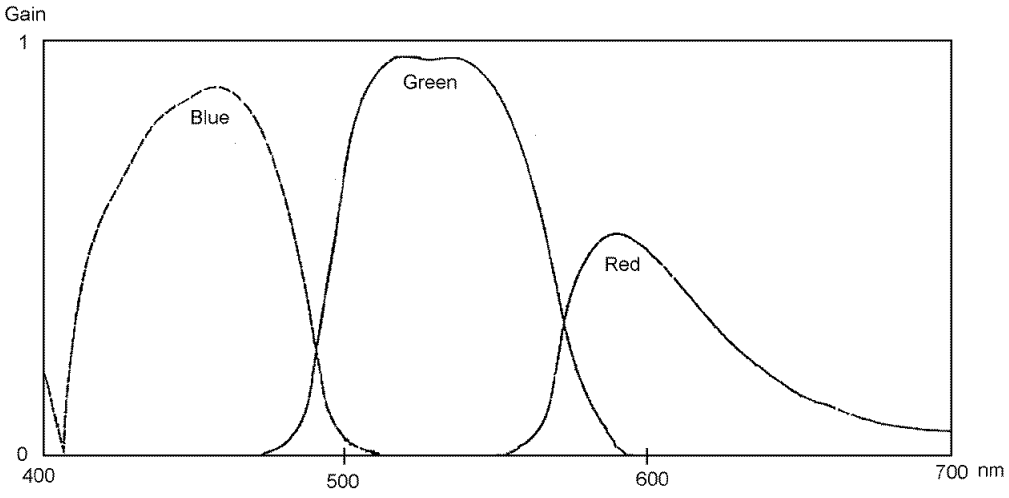

Figure 2: The spectral response of the camera color channels.

R, G, B are the red, green and blue readings respectively.

$f_R(\lambda), f_G(\lambda), f_B(\lambda)$ are the spectral responses of the camera filters in the red, green and blue channels (see Fig. 2).

$I^r(\lambda)$ is the spectrum of light arriving at the camera.

Neglecting the effects of mutual illumination and fluorescence, the relation between the spectral composition of a projected light plane (measured in a reference point on the light plane within the workspace) and that of the light reflected from the scene (measured in the camera) is

$$I^r(\lambda) = I^t(\lambda) \cdot k(\lambda) \tag{2}$$

where

$I^t(\lambda)$ is the projected light spectrum, and

$k(\lambda)$ is proportional to the reflection coefficient of the scene at wavelength λ.

The output of the projector is the sum of a red beam, a green beam and a blue beam. Let the spectrum of the red beam be denoted by $I^t_R(\lambda)$, normalized such that $I^t_R(\lambda_R = 590nm) = 1$, where λ_R corresponds to the peak of the spectral response of the red filter of the camera. Similarly, let $I^t_G(\lambda)$ and $I^t_B(\lambda)$ be the spectra of the green and blue beams, normalized so that $I^t_G(\lambda_G = 530nm) = 1$ and $I^t_B(\lambda_B = 460nm) = 1$.

The projection instructions are given to the graphics adaptor as triplets of RGB values. If the instruction is [r,0,0], then at the reference point the illumination is $r' \cdot I^t_R(\lambda)$, where r' is a monotone nonlinear function of r. Generally, if the projection instruction is [r,g,b], the

illumination at the reference point is

$$I^t(\lambda) = r' \cdot I_R^t(\lambda) + g' \cdot I_G^t(\lambda) + b' \cdot I_B^t(\lambda) \tag{3}$$

where

$$\begin{bmatrix} r' \\ g' \\ b' \end{bmatrix} = P \left\{ \begin{bmatrix} r \\ g \\ b \end{bmatrix} \right\} \tag{4}$$

and the operator P is the transformation from the projection instruction to actual illumination. Substituting Eq. 3 in Eq. 1 we obtain for the red channel

$$R = r' \int_0^\infty f_R(\lambda) I_R^t(\lambda) k(\lambda) d\lambda + g' \int_0^\infty f_R(\lambda) I_G^t(\lambda) k(\lambda) d\lambda + b' \int_0^\infty f_R(\lambda) I_B^t(\lambda) k(\lambda) d\lambda \tag{5}$$

and similar expressions for the green and blue channels.

Let Λ_R, Λ_G and Λ_B respectively denote the supports of the projector filters $I_R^t(\lambda)$, $I_G^t(\lambda)$ and $I_B^t(\lambda)$. We introduce the approximation that the reflectance of the scene is constant within each of these bands, i.e., $\forall \lambda \in \Lambda_i \quad k(\lambda) \cong const = k_i \quad i \in \{R, G, B\}$. This approximation is much weaker, and much more realistic, than the color neutrality assumption in previous works on color structured light. Now, for the red channel,

$$R = r' k_R \int f_R(\lambda) I_R^t(\lambda) d\lambda + g' k_G \int f_R(\lambda) I_G^t(\lambda) d\lambda + b' k_B \int f_R(\lambda) I_B^t(\lambda) d\lambda \tag{6}$$

and in a matrix form for all color channels:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \underbrace{\begin{bmatrix} \int f_R(\lambda) I_R^t(\lambda) d\lambda & \int f_R(\lambda) I_G^t(\lambda) d\lambda & \int f_R(\lambda) I_B^t(\lambda) d\lambda \\ \int f_G(\lambda) I_R^t(\lambda) d\lambda & \int f_G(\lambda) I_G^t(\lambda) d\lambda & \int f_G(\lambda) I_B^t(\lambda) d\lambda \\ \int f_B(\lambda) I_R^t(\lambda) d\lambda & \int f_B(\lambda) I_G^t(\lambda) d\lambda & \int f_B(\lambda) I_B^t(\lambda) d\lambda \end{bmatrix}}_{A} \cdot \underbrace{\begin{bmatrix} k_R & 0 & 0 \\ 0 & k_G & 0 \\ 0 & 0 & k_B \end{bmatrix}}_{K} \begin{bmatrix} r' \\ g' \\ b' \end{bmatrix} \tag{7}$$

Abbreviating the components of the matrix A, substituting Eq. 4 and taking into account the ambient camera readings $[R_0, G_0, B_0]^T$, we obtain in each pixel

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \underbrace{\begin{bmatrix} a_{RR} & a_{RG} & a_{RB} \\ a_{GR} & a_{GG} & a_{GB} \\ a_{BR} & a_{BG} & a_{BB} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} k_R & 0 & 0 \\ 0 & k_G & 0 \\ 0 & 0 & k_B \end{bmatrix}}_{K} P \left\{ \begin{bmatrix} r \\ g \\ b \end{bmatrix} \right\} + \begin{bmatrix} R_0 \\ G_0 \\ B_0 \end{bmatrix} \tag{8}$$

Given the camera readings $[R, G, B]^T$ in a given pixel, would it be possible to recover the projection instruction $[r, g, b]^T$ associated with the illumination plane?

- The vector $[R_0, G_0, B_0]^T$ can be obtained by acquiring an image of the scene under just the ambient light.

- The vector transformation P can be obtained off-line as part of a colorimetric calibration procedure. Since the color channels are decoupled, P actually consists of three scalar transformations. All three are nonlinear but monotone, hence invertible. Furthermore, P is essentially independent of location, so the inverse P^{-1} can be represented by three small look-up tables.

- The projector-camera coupling matrix A can also be measured in advance within the colorimetric calibration procedure. We obtained

$$A = \begin{bmatrix} 1.004 & 0.044 & 0.026 \\ 0.003 & 1.004 & 0.108 \\ 0.003 & 0.004 & 1.145 \end{bmatrix} . \tag{9}$$

A is nearly diagonal. This indicates that the color filters in the camera and the dichroic mirrors in the projector are fairly well matched and that the color crosstalk is small. Therefore, its inverse A^{-1} exists and is numerically stable. A is not very sensitive to spatial position, hence we take it as constant and regard the small local variations as noise.

- The reflectance matrix K is location dependent and unknown. Once it is obtained for any pixel in the image, the recovery of the corresponding instruction $[r, g, b]^T$ is possible. K can be determined for all pixels by acquiring an additional image of the scene under uniform full white illumination, i.e., with $[r, g, b]^T = [255, 255, 255]^T$ for all light planes. In this case, denote the camera readings at a certain pixel by $[R_w, G_w, B_w]^T$. Then

$$\begin{bmatrix} R_w \\ G_w \\ B_w \end{bmatrix} = \underbrace{\begin{bmatrix} a_{RR} & a_{RG} & a_{RB} \\ a_{GR} & a_{GG} & a_{GB} \\ a_{BR} & a_{BG} & a_{BB} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} k_R & 0 & 0 \\ 0 & k_G & 0 \\ 0 & 0 & k_B \end{bmatrix}}_{K} P \left\{ \begin{bmatrix} 255 \\ 255 \\ 255 \end{bmatrix} \right\} + \begin{bmatrix} R_0 \\ G_0 \\ B_0 \end{bmatrix} \tag{10}$$

and the matrix K can easily be calculated.

Note that the intensity ratio range sensor [6] is a special case within this framework. Suppose that

- A single color channel is used, say the green channel. Then the matrices A and K degenerate to a scalars.
- P is a scalar linear relation of the form $g' = P \cdot g$, where P is a positive constant.
- There is no ambient light, i.e., $G_0 = 0$.

Then $G = AKP \cdot g$ and $G_w = AKP \cdot 255$, hence

$$g = 255 \frac{G}{G_w} \ , \tag{11}$$

which is the intensity ratio indexing rule.

To summarize, the original color of an impinging light plane can be identified from the image of the illuminated scene. To achieve this, two additional reference images are needed, one taken under the ambient light and one under uniform white illumination. It is assumed that the reflectivity of the scene is approximately constant within each of the red, green and blue color bands, but color neutrality is not required and saturated scene colors can be accommodated.

In principle, since the projected color can be recovered, it is possible to create a structured light system with color based light-plane indexing along these lines. In practice, to accomplish reliable accurate operation in the presence of noise, errors and deviations from the model, careful attention must be given to the color coding of the light planes. This is considered in the next section.

3 Color Coding

3.1 Color selection

Consider a tiny patch in the scene and the corresponding pixel in the camera. Once the two reference images have been taken, the camera readings $[R_0, G_0, B_0]^T$ and $[R_w, G_w, B_w]^T$ for that pixel are known. Suppose that the depth of the tiny patch should be obtained by projecting a single additional pattern with L color-coded light planes. The $[R, G, B]$ camera readings that correspond to these L colors can be represented by L points within the *color box* shown in Fig. 3.

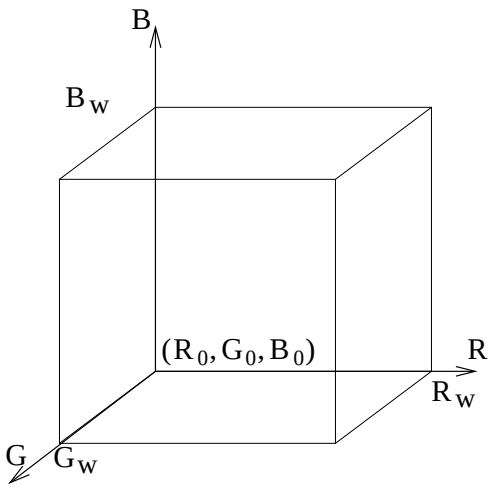

Figure 3: The color box.

For good noise immunity, the colors should be selected such that the distance between any two points in the color box would be sufficiently large (in terms of some metric that depends on the noise distribution). Once the placement of points in the color box has been decided, the corresponding projection colors can be obtained via the matrices A and K. The required instruction can then be calculated using the transformation P^{-1}.

Optimal placement (in some sense) of the L points in the color box is nontrivial. In particular, notice that the noise is the combined outcome of several sources. One is imaging noise (camera and frame grabber) and the other is projection noise (graphics adaptor and projector). The projection noise is not uniform within the camera color box. Since online design of projection patterns is an important aspect of our approach, simplicity is crucial. The projector-camera coupling matrix A is nearly diagonal, so we decided to consider crosstalk between color channels as *noise* and treat each color channel separately. Furthermore, the noise level throughout the color box is assumed to be equal to the worst-case actual noise level. This means that the L points are positioned in the color box on a 3-D rectangular lattice. The distance between any two points should provide a sufficient noise margin.

Assume that the overall noise distribution in each of the color channels is Gaussian and let its standard deviation be denoted by $\sigma_i \quad i = R, G, B$. In our system, the measurement standard deviation values (over time and space) were $\sigma_R = 3.0$, $\sigma_G = 1.9$ and $\sigma_B = 2.4$ (in units of grey levels at the frame grabber output). We set the distance between adjacent points in the color box, parallel to the color axes, to

$$D_i = \alpha \cdot \sigma_i \quad i \in \{R, G, B\} \tag{12}$$

where α is a noise immunity parameter. The number of levels in each channel is therefore

$$n_R = \frac{R_w - R_0}{D_R} \qquad n_G = \frac{G_w - G_0}{D_G} \qquad n_B = \frac{B_w - B_0}{D_B} \tag{13}$$

and the total number of colors that can be reliably identified is $L_{max} = n_R \times n_G \times n_B$.

If the scene is light and the noise immunity requirements moderate, L_{max} can be sufficiently large to allow reasonable range imaging with a single projection pattern. However, in most realistic applications $L_{max} < 640$ and the full spatial resolution of our projector cannot be exploited by means of a single pattern. In that case, one or more additional color projection patterns should be used. With M patterns, the maximal number of light planes that can be uniquely labeled is

$$L_{max} = (n_R \times n_G \times n_B)^M \ . \tag{14}$$

In terms of M and the noise immunity parameter α, the number of distinct light planes is at most

$$\mathrm{L} \leq \left[\frac{R_w - R_0}{\sigma_R} \times \frac{G_w - G_0}{\sigma_G} \times \frac{B_w - B_0}{\sigma_B} \times \frac{1}{\alpha^3} \right]^M \quad . \tag{15}$$

Similarly, when M and L are fixed, the noise immunity parameter α is upper bounded; when α and L are set, the number of projection patterns is lower bounded.

Once the points have been placed in the color box, the corresponding projection colors are calculated. Neglecting color crosstalk, i.e., assuming that A is nearly diagonal, the distances between projection intensity levels are

$$d_i' = \frac{D_i}{a_{ii} k_i} \quad i \in \{R, G, B\} \tag{16}$$

where a_{ii} and k_i are the diagonal elements of the matrices A and K. The projection instructions are then obtained via P^{-1}.

Color selection has so far been based on the camera readings $[R_0, G_0, B_0]^T$ and $[R_w, G_w, B_w]^T$

in a specific pixel in the reference images, hence on a tiny patch in the scene. However, a single set of colors must suit the whole scene, so the values of all pixels in the reference images must be taken into consideration. While various averaging strategies can be suggested, we follow a worst-case approach. This means that the selected colors should allow reliable range imaging even in darker parts of the scene.

Extraction of depth information is however impossible in shadow regions or in black parts of the scene, from which hardly any light is reflected. A "blind" worst-case strategy will thus lead to a fruitless attempt to select colors for operation in these hopeless regions. Noise margins will be maximized, the number of levels in each primary color will be minimized and the number of projection patterns will be uselessly increased.

Shadow and black regions are fairly easy to detect given the two reference images, say by thresholding in the three color channels. In practice, slight spatial dilation of the regions tagged as shadow/black proved to be beneficial. Our selection of colors is based on the lowest values of $R_w - R_0$, $G_w - G_0$ and $B_w - B_0$ observed outside the black and shadow regions.

3.2 Codeword allocation

Multi-pattern color coding of light planes (see Fig. 4) consists of the following layers:

- Code alphabet: Each code letter corresponds to one of the primary colors. Let n_R, n_G, n_B denote the number of levels used in the red, green and blue channels respectively. Then each code letter is an integer in the range $0...n_R - 1$ or $0...n_G - 1$ or $0...n_B - 1$ according to its color.

- Symbol: A triplet of letters, one from each primary color. A symbol corresponds to

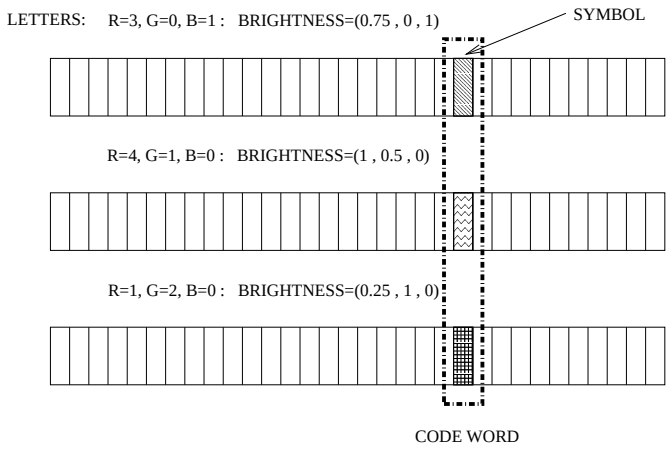

Figure 4: An example of a code word that consists of three symbols, each a combination of three letters. Here $n_R = 5, n_G = 3, n_B = 2$. The letters correspond to primary brightness levels, and a combination of three letters, one in each primary color, is the color of a specific light plane in one of the three projection patterns.

the color of a specific light plane in a certain illumination pattern.

- Code word: Each light plane is generally represented by a code word that is a combination of symbols. When using M color illumination patterns, the code word is M symbols long ($3 \cdot M$ letters).

- Projection patterns: A single projection pattern carries a symbol for each light plane. The set of projection patterns carries the complete code word for each light plane.

The next issue is the allocation of code words to light planes. We first consider the simple case in which a single projection pattern is used. In a structured light system, transitions between light planes are gradual due to the finite spatial bandwidth of the optical system and light plane blending in the elements of the camera CCD array. The advantage of the Gray code in black & white structured light [14] is that the effect of a detection error due to a transition is limited to an error between the two adjacent light planes. The same reasoning

B	0			1		
G	0	1	2	2	1	0
R	0 1 2 3 4	4 3 2 1 0	0 1 2 3 4	4 3 2 1 0	0 1 2 3 4	4 3 2 1 0

Table 1: An example of code word allocation according to the generalized Gray code, with blue as the most significant digit (MSD) and red as the least significant digit (LSD).

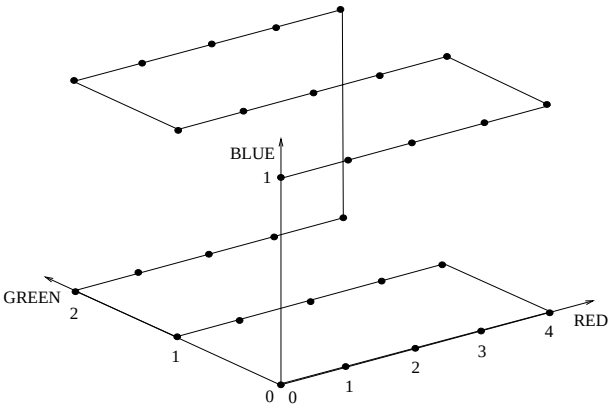

Figure 5: The allocation of symbols to light planes defines a scanning order of the points in the color space and an implicit 3-D digital curve.

holds here, but a multilevel code is needed, with each digit in a different base. Suitable generalizations of the Gray code are indeed available [9].

If the blue channel represents the most significant digit (MSD) and the red channel the least significant digit (LSD), the generalized Gray code is as shown in Table 1. Once the symbols, i.e., the set of points in the color box, have been allocated to light planes, they can be regarded as a digital curve. This is illustrated, for our example, in Fig. 5.

The generalized Gray code shown in Table 1 and in Fig. 5 is not the only possible one. For example, by letting the red channel represent the MSD and letting the blue channel represent the LSD, the code shown in Table 2 is obtained. Which encoding should be preferred? We suggest to base the selection on the effective spatial bandwidth of the resulting patterns.

R	0			1			2			3			4		
G	0	1	2	2	1	0	0	1	2	2	1	0	0	1	2
B	01	10	01	10	01	10	01	10	01	10	01	10	01	10	01

Table 2: An alternative code word allocation according to the generalized Gray code, with red as the most significant digit (MSD) and blue as the least significant digit (LSD).

MSD								LSD
B^{M-1}	...	B^0	G^{M-1}	...	G^0	R^{M-1}	...	R^0

Table 3: Example of code generation in the multi-pattern case. Denote the color components of the i-th projection pattern by $R^i, G^i, B^i \quad i = 0 \ldots M-1$. The association of color components with digits is shown for the case of $n_R > n_G > n_B$.

Since the system is bandlimited due to depth of field and other factors, minimizing the effective bandwidth of the patterns can be expected to lead to better performance.

The largest number of transitions in the code is in the color channel that represents the least significant digit (LSD) in the pattern. Therefore, its effective bandwidth is normally the largest. If the number of levels used in that color channel is small, then the ratio of transition magnitude to channel amplitude is large and the effective bandwidth is high and vice versa. Thus, the color channel with the largest number of levels should be assigned to the LSD and the channel with the smallest number of levels to the MSD. In our example, the encoding shown in Table 1 is compatible with this guideline and should be preferred.

We have so far concentrated on the case of a single projection pattern. Consider now the M pattern case, where a code word consists of $3M$ digits. We suggest to associate the M least significant digits with the color channel that has the largest number of levels and the M most significant digits with the color channel that has the smallest number of levels. See Table 3.

4 Experiments

The suggested approach has been implemented. The hardware components of the experimental system were already described in section 2. In the beginning of a range imaging session, the user sets two of the following three parameters: L – the number of light planes, M – the number of projection patterns and α – the noise immunity parameter. From this point on, the following procedure is automatically carried out:

1. A first reference image is taken under ambient light only. The camera readings $[R_0, G_0, B_0]^T$ in all pixels are recorded.

2. A second reference image is taken under uniform white illumination. The camera readings $[R_w, G_w, B_w]^T$ in all pixels are recorded.

3. Black and shadow regions are identified by thresholding and dilation. They are marked and excluded from further consideration.

4. For each pixel, the matrix K is computed by solving Eq. 10. (The matrix A^{-1} and the inverse operator P^{-1} are known in advance and are the same for all pixels).

5. The lowest values of $R_w - R_0$, $G_w - G_0$ and $B_w - B_0$ outside the black and shadow regions are found and the respective values of k_R, k_G and k_B in the corresponding pixels are stored. Among the parameters L, M and α, the one that was *not* set by the user is now automatically adapted by solving Eq. 15. The color box is sampled on a rectangular lattice with the appropriate intervals. The sampling points are transformed to projection instruction via Eq. 16 and the operator P^{-1}. Code words are built according to the generalized Gray code.

6. M patterns are projected and M images are acquired.

7. For each pixel in each of the M images, $[r, g, b]^T$ is obtained by solving Eq. 8. (At this stage the diagonal matrix K is known for every pixel).

8. For each pixel, the code word nearest to the M $[r, g, b]^T$ vectors is determined. It corresponds to a light plane.

9. The 3-D location of the scene point that corresponds to each pixel is obtained by triangulation between the light plane and the line of sight.

Currently, the whole procedure takes about 30 seconds that are spent mostly on hard disk access. Once this is eliminated, computing time will be reduced to a few seconds. We believe that with hardware acceleration, acquisition time approaching the limiting value of M/R can be achieved, where R denotes the frame rate.

A geometric calibration procedure should be carried out whenever the projector or camera are moved. The calibration of our system follows the principles outlined in [18] and relies on software developed in the Image Science Laboratory, ETH, Zurich, Switzerland. Colorimetric calibration, i.e., measurement of the matrix A and the operator P, is needed only when the color characteristics of the projector or camera are changed, in particular when the projector bulb is replaced.

The typical distances of the camera and the projector from the scene in our setup are 1.5m and 1.0m respectively. The optical axes of the camera and the projector roughly intersect in the workspace, creating an angle of about 20^o. Given the geometric setup, a bound on the accuracy of the system can be derived as follows. Ideally, a projected light plane and a

line of sight intersect at a point. In practice, due to finite cell size in the LCD panels light planes are thick. Similarly, due to finite pixel size in the CCD array, the line of sight becomes a cone. Their intersection is therefore a volume rather than a geometric point, leading to ambiguity that sets an upper bound on accuracy [13, 10]. Rough calculations show that in our system the standard deviation of the depth error cannot be smaller than about 0.5mm. Better accuracy could be achieved by using a higher resolution projector and camera.

The flexibility of the projector allows easy implementation of various light plane coding schemes by software modification. Experimental comparison of various approaches is straightforward, since they can all be tested in the same setup, with the same hardware, and in a similar software environment. In addition to the algorithm presented here, we implemented the standard (black & white) Gray code scheme [14].

Scene #1: A planar color picture

A planar color picture was positioned on a plane parallel to the camera plane, at a distance of about 154.7cm. (The deviation from parallelism was smaller than 0.5mm, the lower bound on the standard deviation of the depth error. This was verified by analysis of local averages of depth estimates in various regions of the scene.) An image of the scene taken under uniform white illumination is shown in Fig. 6. Range imaging was performed in three different conditions:

1. Adaptive color coding with $L = 640$ light planes and a noise immunity factor of $\alpha = 5$. The number of projection patterns was automatically set to $M = 3$, with two brightness levels in red and blue, and three in green.

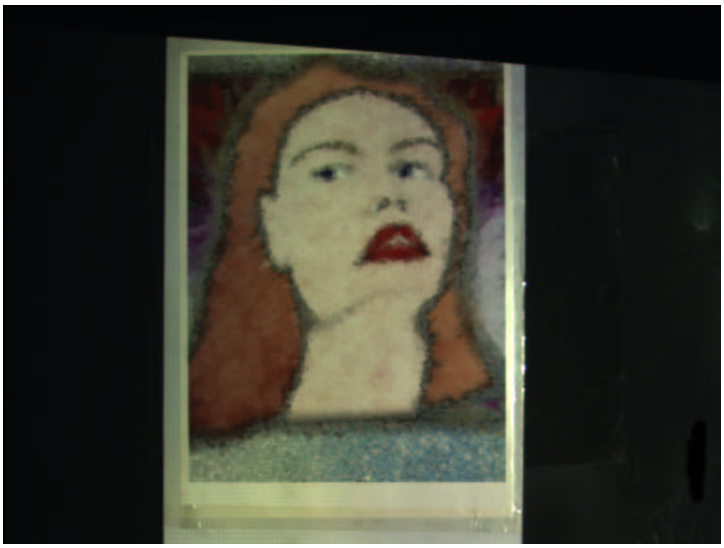

Figure 6: An image of the color picture scene taken under uniform white illumination. The dominant colors in the actual picture are orange, red, beige and blue.

2. Adaptive color coding with $L = 640$ light planes and a noise immunity factor of $\alpha = 9$.

   The number of projection patterns was automatically set to $M = 5$, with two brightness levels in red and green, without using blue.

3. Black & white Gray coding with $M = 10$ illumination patterns.

The results are summarized in Table 4. The average distance from the camera plane and the standard deviation of the depth values for each of the three conditions are presented. The top rows refer to averaging over the whole color picture, the bottom to averaging only in the center. As expected, results are better in the center since it is lighter and since lens aberrations are smaller in the center of the field. It can be seen that good performance is obtained even with only three color projection patterns, and that the accuracy of the $M = 10$ black & white Gray code method can be achieved with $M = 5$ color patterns.

	Color $M = 3$	Color $M = 5$	Gray $M = 10$
Average depth [cm]	154.59	154.56	154.60
Standard deviation [cm]	0.10	0.09	0.09
Average depth [cm]	154.56	154.52	154.58
Standard deviation [cm]	0.09	0.06	0.05

Table 4: Depth imaging results for the planar color picture scene. The average distance from the camera plane and the standard deviation of the depth values for each of the three conditions are presented. The top rows refer to averaging over the whole color picture, the bottom to averaging only in the center.

Scene #2: Exponential staircase

As shown in Fig. 7, this scene is a staircase in which the steps are roughly a geometric series: 1mm, 2mm, 4mm, 8mm and 16mm. The staircase was positioned such that its baseplane was parallel to the camera plane, at a distance of about 1.5m. Range imaging was performed in two different conditions:

1. Adaptive color coding with $L = 640$ light planes and a noise immunity factor of $\alpha = 12$. The number of projection patterns was automatically set to $M = 4$, with two brightness levels in red, green and blue.

2. Black & white Gray coding with $M = 10$ illumination patterns.

The depth values along a single row are shown in Fig. 8 for the two experimental conditions. The results with color coding with $M = 4$ projection patterns (left) and black & white Gray coding with $M = 10$ patterns (right) are similar. The periodic errors are due to the moiré effect caused by the grid structure of the CCD arrays in the camera and the LCD panels in the projector.

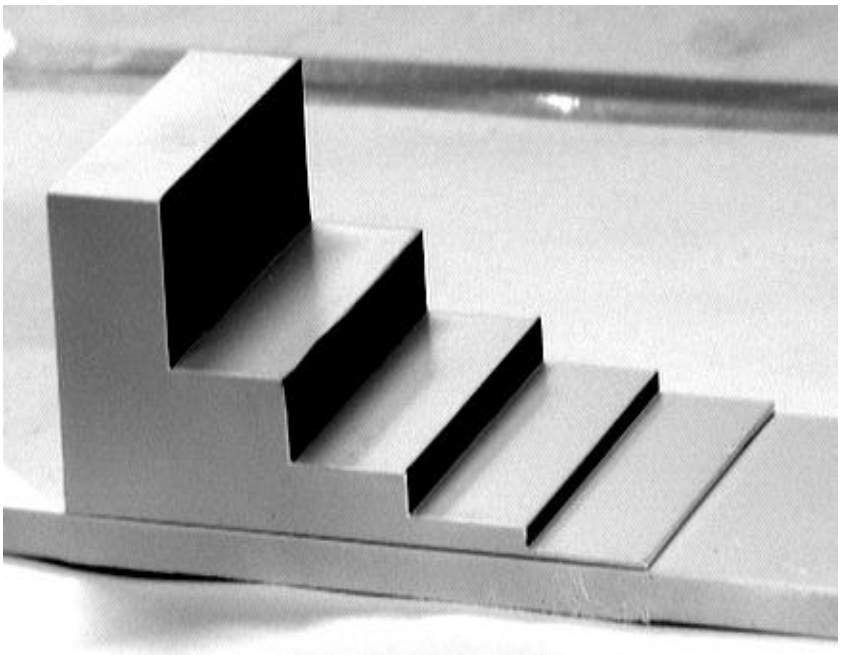

Figure 7: The exponential staircase. The steps are approximately 16mm, 8mm, 4mm, 2mm and 1mm.

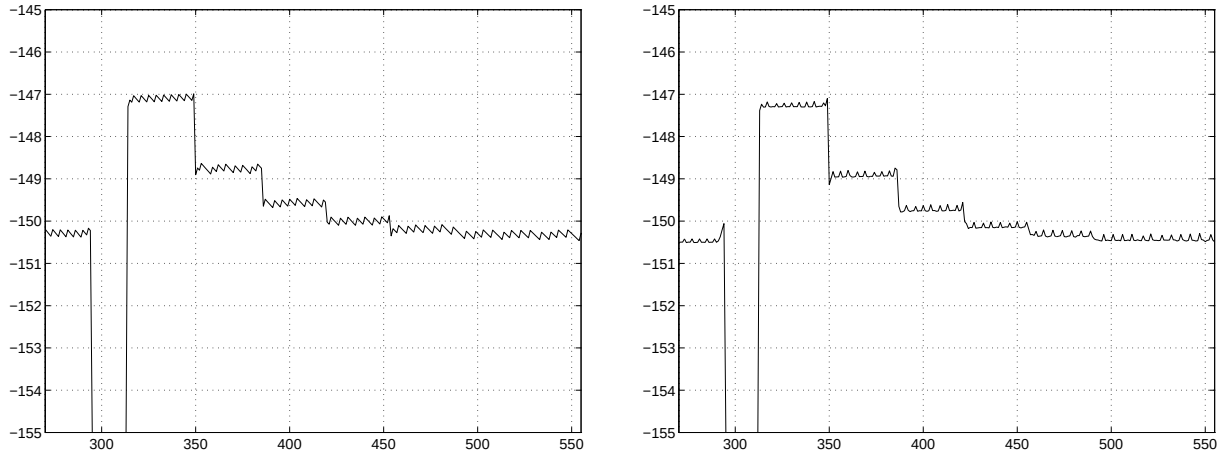

Figure 8: Reconstructed depth values (cm) in a single row of the exponential staircase scene. The horizontal axis is in pixel units. *Left:* Color coding with $M = 4$ projection patterns. *Right:* Black & white Gray coding with $M = 10$ patterns. The periodic errors are due to moiré effects.

For both methods, the average and the standard deviation of the depth values were computed in each stair. These results are summarized in Table 5(a). It can be seen that the accuracy obtained with $M = 4$ color projection patterns approaches that of the black & white Gray code method with $M = 10$ patterns. Note that both are close to the 0.5mm bound. Based on these measurements, we have also estimated the step sizes and compared them with mechanical measurements carried out with a Mitutoyo CD-6"CR digimatic caliper, see Table 5(b).

Scene #3: Head

The scene is a manikin head, shown under uniform white illumination in Fig. 9. Range imaging was carried out in the following conditions:

1. Adaptive color coding with $L = 640$ light planes and a noise immunity factor of $\alpha = 5$. The number of projection patterns was automatically set to $M = 4$, with two brightness levels in each primary color.

2. Black & white Gray coding with $M = 10$ illumination patterns.

In both cases the reconstructed range data was written in VRML 2.0 and visualized with a suitable browser. The results, shown in Fig. 10 are quite similar. Projector occlusion (shadow) behind the ear prevents depth recovery there.

Scene #4: Office equipment

An arrangement of office equipment is shown in Fig. 11 (left). Range imaging was carried out using adaptive color coding with $L = 640$ light planes and a noise immunity factor of

		Color $M = 4$	Gray $M = 10$
Top stair	Average depth [cm]	147.13	147.30
	Standard deviation [cm]	0.06	0.05
2nd stair	Average depth [cm]	148.78	148.94
	Standard deviation [cm]	0.06	0.05
3rd stair	Average depth [cm]	149.61	149.76
	Standard deviation [cm]	0.07	0.06
4th stair	Average depth [cm]	150.03	150.16
	Standard deviation [cm]	0.07	0.06
Bottom stair	Average depth [cm]	150.23	150.36
	Standard deviation [cm]	0.07	0.06
Base plane	Average depth [cm]	150.35	150.46
	Standard deviation [cm]	0.07	0.06

(a)

Step sizes [cm]		
Color $M = 4$	Gray $M = 10$	Caliper
1.65	1.64	1.59
0.82	0.81	0.79
0.41	0.40	0.39
0.20	0.20	0.20
0.12	0.10	0.10

(b)

Table 5: (a) Average and standard deviation of the depth measurements in each stair of the exponential staircase, with $M = 10$ pattern Gray coding and with $M = 4$ color coding. (b) Average step sizes obtained with both methods, compared with mechanical measurements.

Figure 9: An image of the head scene taken under uniform white illumination.

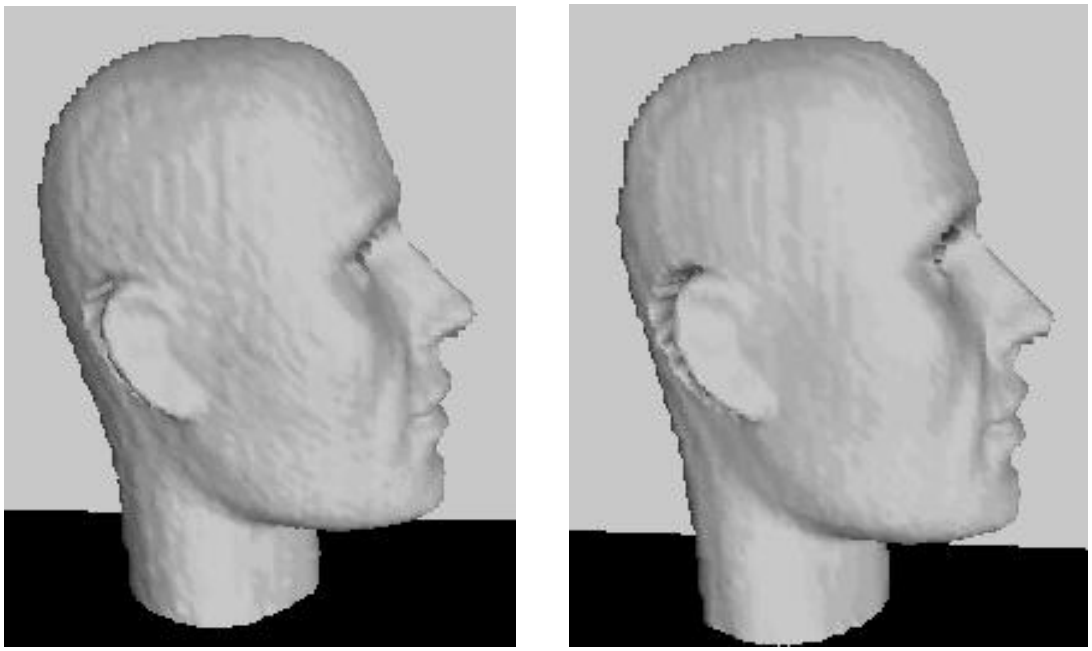

Figure 10: Visualization of the range estimation of the head scene. *Left:* Color structured light with four projection patterns. (b) Black & white Gray structured light with 10 projection patterns.

$\alpha = 6$. The number of projection patterns was automatically set to $M = 2$, with three brightness levels in each primary color. The reconstruction is shown in Fig. 11 (right) as a range image: lighter regions are closer to the camera plane. In several parts of the scene depth cannot be obtained due to various reasons; all these areas are automatically detected and marked black. In particular,

- Due to the specularity of the metallic part of the floppy disk, the light that it reflects does not reach the camera, so triangulation is impossible.
- In shadow regions, such as on the telephone and near the mouse, occlusion precludes triangulation.
- Almost no light is reflected from the keys of the telephone and from the black marking on the mouse, so their depth is not calculated.
- Because of the specularity of the cup, high intensity reflected light saturates the camera (along a stripe). This nonlinear distortion leads to color decoding failure. A possible solution is to use a camera with a higher dynamic range.

5 Conclusions

We presented a structured light method that reaches the accuracy and robustness of the Gray code technique [14] while using fewer projection patterns. The time interval during which the scene must remain static can therefore be reduced, thus widening the range of potential applications. This is accomplished by two new aspects of the suggested approach.

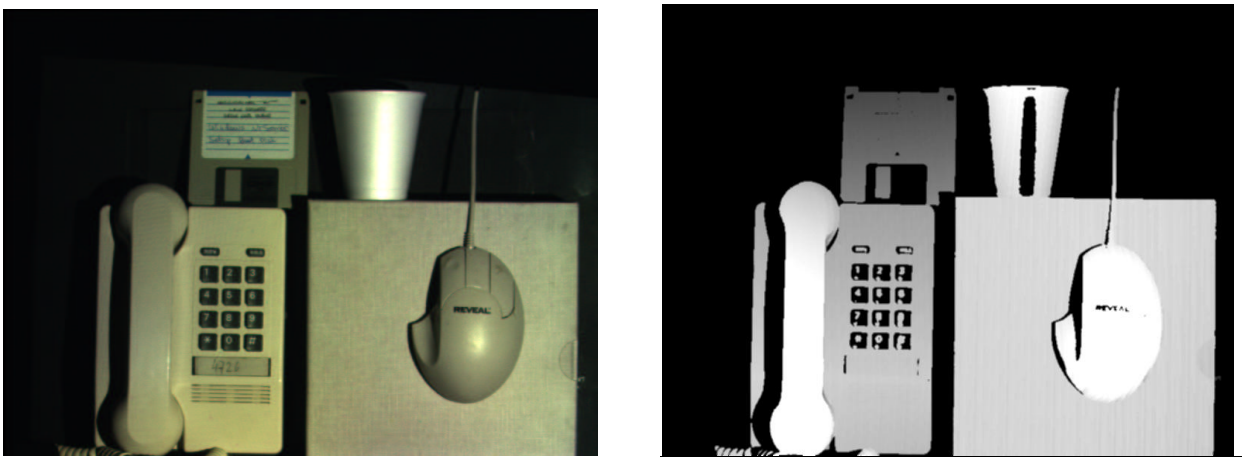

Figure 11: Office equipment. *Left:* An image of office equipment taken under uniform white illumination. *Right:* A range image obtained using color structured light with two projection patterns. Lighter regions are closer to the camera. Areas in which depth cannot be obtained (due to various reasons) are automatically detected and marked black.

One is a new approach for the use of color in structured light. It has been previously believed that color neutrality of the scene is a prerequisite for successful color labeling of light planes. We have shown that even with colorful scenes, the color of an impinging light plane can be identified from the image of the illuminated scene. Identification is carried out in each pixel separately and does not rely on information from neighbors, hence powerful smoothness assumptions are not needed and the spatial resolution is not compromised.

Both the Gray code method [14] and the intensity ratio technique [6] rely on two reference images for operation in the presence of ambient light and albedo variations. In our approach the information extracted from the reference images also provides compensation for the color content of the scene. Even though the suggested surface color recovery algorithm degenerates in the monochromatic case to the intensity ratio scheme, there is an important difference. In the intensity ratio range sensor [6], light plane labeling is analog hence susceptible to noise. Our coding scheme, based on N-ary Gray coding, is digital and provides the necessary noise

immunity.

The other new aspect is the online adaptation of the number and form of the projection patterns to the scene, as part of the acquisition process. Since the signal to noise ratio in each color channel eventually depends on the reflection from the scene in that band, adaptation eliminates the need for wasteful worst case design of the set of projection patterns. Instead, noise margins are matched to the actual noise levels, thus keeping the number of projection patterns at the necessary minimum. With darker scenes, the adapted projected patterns converge to the standard Gray code method.

The fundamental assumption in structured light methods that are based on intensity ratios is linear dependence of the intensity of light reflected from a point in the scene on the intensity of the light plane projected onto that point. Violations of this assumption, e.g., due to mutual illumination or fluorescence, can lead to errors. In our method, as long as these effects are small, they can be absorbed in the noise margins. Beyond a certain level, performance will deteriorate.

The progress reported here was made possible by the recent introduction of single-beam color video & data projectors. In our projector, light is modulated by transmission through LCD panels. With this technology, the intensity of the projected patterns is suitable for short distance indoor applications. At the time of writing, much more powerful projectors are being introduced, that could allow limited outdoor use of structured light methods.

Acknowledgments

We are grateful to the Image Science Laboratory of the ETH in Zurich, Switzerland, and in particular to Dr. Marjan Trobina, for the system calibration software. We thank Dr. Boris Spektor, Technion, for his help and advice. This research was supported in part by the fund for the promotion of research at the Technion and by the Smoler research fund.

References

[1] A. Åstrom, *Smart Image Sensors*, Ph.D. Thesis, Linköping University, Sweden, 1993.

[2] G.J. Agin and T.O. Binford, "Computer Description of Curved Objects", *Proc. 3rd IJCAI*, pp. 629-640, Stanford, 1973.

[3] P. Besl, "Active Optical Range Imaging Sensors", *Machine Vision and Applications*, Vol. 1, pp. 127-152, 1988.

[4] K.L. Boyer and A.C. Kak, "Color-Coded Structured Light for Rapid Image Ranging", *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 9, pp. 14-28, 1987.

[5] A. Blake, D. McCowen, H.R. Lo and P.J. Lindsey, "Trinocular Active Range Sensing", *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 15, pp. 477-483, 1993.

[6] B. Carrihill and R. Hummel, "Experiments with the Intensity Ratio Depth Sensor", *Comput. Vis. Graph. and Image Process.*, Vol. 32, pp. 337-358, 1985.

[7] C.S. Chen, Y.P. Hung, C.C. Chiang and J.L. Wu, "Range Data Acquisition Using Color Structured Lighting and Stereo Vision", *Image and Vision Computing*, Vol. 15, pp. 445-456, 1997.

[8] G. Chazan and N. Kiryati, *Pyramidal Intensity Ratio Depth Sensor*, Technical Report No. 121, Center for Communication and Information Technologies, Department of Electrical Engineering, Technion, Haifa, Israel, October 1995.

[9] M.C. Er, "On Generating the N-ary Reflected Gray Codes", *IEEE Trans. Computers*, Vol. 33, pp. 739-741, 1984.

[10] E. Horn and N. Kiryati, *Towards Optimal Structured Light Patterns*, Technical report No. 194, Center for Communication and Information Technologies, Department of Electrical Engineering, Technion, Haifa, Israel, June 1997. A short version appeared in *Proc. Int. Conf. on Recent Advances in 3-D Digital Imaging and Modeling*, pp. 28-37, Ottawa, Canada, May 1997.

[11] H. Hügli and G. Maître, "Generation and Use of Color Pseudo Random Sequences for Coding Structured Light in Active Ranging", *Proc. SPIE Industrial Inspection*, Vol. 1010, pp. 75-82, 1988.

[12] T. Kanade, A. Gruss and L.R. Carley, "A Very Fast VLSI Rangefinder", *Proc. Int. Conf. on Robotics and Automation*, pp. 1322-1329, April 1991.

[13] A.M. McIvor, *The Accuracy of Range Data from Structured Light System*, Report No. 190, Industrial Research Limited, Auckland, New Zealand, 1994.

[14] K. Sato and S. Inokuchi, "Three-Dimensional Surface Measurement by Space Encoding Range Imaging", *J. of Robotic Systems*, Vol. 2, pp. 27-39, 1985.

[15] V. Smutný, "Analysis of Rainbow Range Finder Errors", in V. Hlavác and T. Pajdla (eds.), *Proc. 1st Czech Pattern Recognition Workshop*, pp. 59-66, November 1993.

[16] Y. Shirai and M. Suwa, "Recognition of Polyhedrons with a Range Finder", *Proc. 2nd IJCAI*, pp. 80-87, London, 1971.

[17] J. Tajima and M. Iwakawa, "3-D Data Acquisition by Rainbow Range Finder", *Proc. 10th ICPR*, pp. 309-313, Atlantic City, 1990.

[18] M. Trobina, *Error Model of a Coded-Light Range Sensor*, Technical Report BIWI-TR-164, ETH Zurich, Switzerland, 1995.