Analysis of a high definition camera-projector video system for geometry reconstruction

José Luiz de Souza Filho, Roger Silva, Dhiego O. Sad,
Renan Dembogurski, Marcelo B. Vieira, Sócrates O. Dantas, and Rodrigo Silva

Universidade Federal de Juiz de Fora - DCC/ICE,
R. Loureno Kelmer, 36036-330, Juiz de Fora-MG, Brazil
{jsouza,correia.roger,dhiegosad,ad.renan,
{marcelo.bernardes,rodrigoluis}@ice.ufjf.br,
dantas@fisica.ufjf.br
http://www.gcg.ufjf.br

Abstract. The present work describes the development of a high definition camera-projector system (with resolutions up to 1920x1080 and 1280x720). The steps and processes that lead to the reconstruction of geometry, from system setup to triangulation, are detailed. A structured light coding scheme that offers a flexible number of stripes for projection was used. One of the objectives of this work is to estimate the limit number of stripes possible within the current resolutions of high definition video. This limit number is the one that provides dense geometry reconstruction with low accuracy and precision errors. To evaluate the reconstructions, we propose a protocol for error measurement. The geometry of general objects is also presented for qualitative evaluation. Our results show that low cost but effective 3D scanners can be built with high definition video devices providing compressed data.

Keywords: camera-projector system, 3D reconstruction, structured light, error evaluation

1 Introduction

A 3d scanner is a device that captures 3D geometric information of objects and scenes from real world. A camera-projector scanner does that by projecting a structured light pattern over an object. The deformation of the projected light pattern allows the geometry reconstruction. Information obtained from high resolution devices provide detailed and dense generation of point clouds.

In this work we describe the development of a 3D scanner with high definition camera and projector. Our scanner uses the $(b, s) - BCSL$ structured light coding [1], which employs the projection of a sequence of colored vertical stripes on the scene. The higher the number of projected stripes is, more details the reconstructed geometry has. A good reconstruction is the one that recovers geometric details of objects or scenes with subpixel accuracy, has a high density number of points and also has a low reconstruction error. Another major advantage of this type of color coding is that only a few number of complementary slides are needed for a complete new geometry shot. Our system provides a set of 3D points for each two consecutive frames.

There are other works that focus on acquiring high density of points as [2] and [3]. An extended overview of structured light reconstruction methods are presented in [4].

2 Structured Light Pattern and camera-projector calibration

In this work, we chose the *(b,s)-BCSL* [1] as structured light pattern. This coding can obtain high resolution geometry without imposing strong restrictions on the scene dealing, for example, with reflection problems. Every transition of adjacent stripes has a unique code defined by the colors on its left and right. The parameter b denotes the number of colors to be used. There are six eligible colors: red (R), green (G), blue (B) and their respective complementary cyan (C), magenta (M) and yellow (Y). The number of stripes is defined by s and it is directly related to the coding and decoding.

The number of different possible codes for transitions is $[b(b-1)]^s$, considering that consecutive stripes cannot have the same colors. The *(b,s)-BCSL* coding uses slides with complementary colors to deal with photometric restrictions of a scene. Each projection of a slide S is followed by a complementary color slide $\bar{S}$. This allows robust transition identification with subpixel precision in $O(1)$.

The triangulation is obtained by calibrating the camera-projector system. There are many methods as seen on [5] and [6], but the one used in this paper is Tsai's with a global optimization [7]. Tsai's is an accurate calibration method and it models the distortion caused by the camera's lens. This turns the linear relations on a non-linear problem due to the radial distortion term in the system. To solve this problem we use an optmization scheme as the Levenberg-Marquadt algorithm [8] [9].

3 Geometry and Photometry Extraction

In this section we describe the steps of the pipeline from getting the initial images and outputting a 3D point cloud. An overview of the whole process is shown on Figure 1.

3.1 Pre-Processing Images

The proposed pipeline has four images I_1, $\bar{I}_1$, I_2 and $\bar{I}_2$ as input. We also acquire a fifth image T lightened with a white projection for further texture mapping. First we use a 5x5 Gaussian low-pass filter to remove noise on each of the four images. Then we apply a color calibration matrix $\mathbf{M}$ to each pixel [10]. As a result there are four color calibrated images C_1, $\bar{C}_1$, C_2 and $\bar{C}_2$.

The next step is to subtract each image from its complementary image. For each pair of images we create a difference image D_n and a projected colors image P_n. The images D_1, D_2, P_1 and P_2 are the input of the boundary detection.

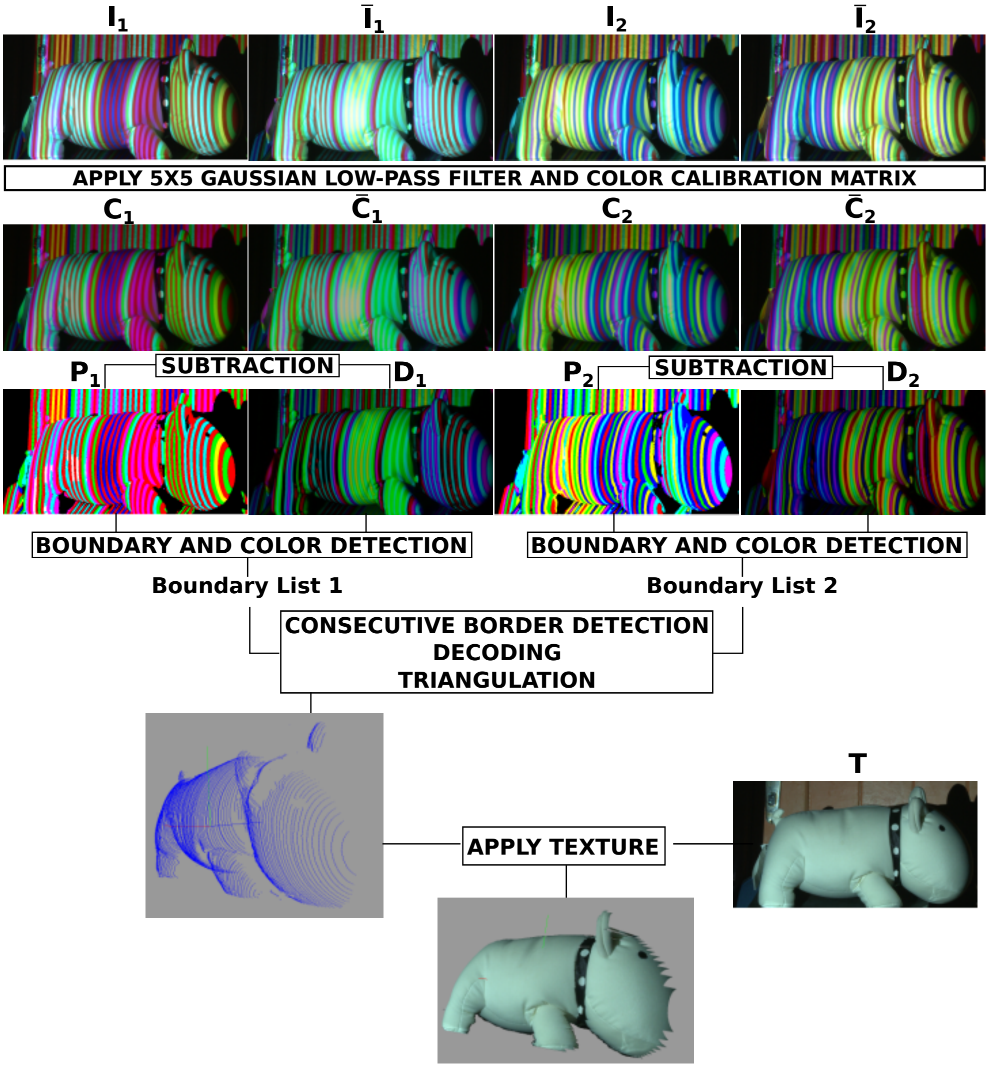


Fig. 1. Pipeline overview.

3.2 Boundary Detection

The boundary positions are identified by looking for zero crossings at any of the Red, Green or Blue channels of the difference images. The left and right colors of a boundary correspond to the colors at that position on the projected colors images.

Let D_n^R, D_n^G e D_n^B be the color channels of a difference image D_n where $n = \{1, 2\}$. As the *(b,s)-BCSL* projected stripes are vertical we can process each line of D_n individually. The pixels on the right and left of a zero crossing have opposite signs. Between them there is a high intensity variation. A zero crossing should satisfy, at least on one color channel K, the following restriction:

$$|D_n^K(i-\delta_1,j)|+|D_n^K(i+\delta_1,j)|\geq l, \tag{1}$$

where $K=\{R,G,B\}$.

On Equation 1, i and j are the columns and rows of the difference image respectively. The parameter l is a threshold which denotes the minimum intensity variation that characterizes a zero crossing. Empirically we assumed $l=5$ in this paper. δ_1 is a parameter referred to a slightly offset describing the left and right neighbor pixels of a zero crossing. If Equation 1 is satisfied by any channels of D_n at a position (i,j), a boundary is generated. Each boundary is defined as $b=(x,j,a,c_l,c_r)$, where $x\in\mathbb{R}$ is the horizontal coordinate of a zero crossing with subpixel precision, j is the row in which the zero crossing was found, a is the amplitude of the crossing, and c_l and c_r are respectively the left and right colors of the projected stripes where the boundary was identified. The amplitude a and coordinate x are

$$a=|D_n^K(i-\delta_1,j)|+|D_n^K(i+\delta_1,j)| \tag{2}$$

and

$$x=\frac{D_n^K(i+\delta_1,j)\cdot(i-\delta_1)+D_n^K(i-\delta_1,j)\cdot(i+\delta_1)}{a}. \tag{3}$$

e

If more than one boundary is found and the distance between the x coordinate of some of them are lower than a given distance δ_2, these boundaries are linearly combined into a single boundary:

$$x_c=\frac{x_1\cdot a_1+x_2\cdot a_2+...+x_n\cdot a_n}{a_1+a_2+...+a_n}, \tag{4}$$

and the combined amplitude of the new boundary is

$$a_c=\frac{a_1^2+a_2^2+...+a_n^2}{a_1+a_2+...+a_n}. \tag{5}$$

In this paper we defined $\delta_2=3$ meaning there is no redundant transition in a 3 pixel region. High values of δ_2 could filter boundaries of distinct stripes transitions.

The next step is to filter the boundaries detected in all the color channels at once. For each row of the image we traverse the list of boundaries. For each boundary b_i we find the boundary b_h of highest amplitude in a δ_2 distance. The final computed zero crossing position x_f of this D_n region is

$$x_f=\frac{x_i\cdot a_i+x_h\cdot a_h}{a_i+a_h}. \tag{6}$$

After the boundaries are found we need to get the left and right colors to each of them. In order to do that we search the projected colors image P_n. Each color channel K of P_n is $P_n^K(i,j)=0$ if $D_n^K(i,j)>0$ and $P_n^K(i,j)=1$ otherwise.

For a boundary $b = (x, j, a, c_l, c_r)$ the left and right colors are $c_l = P_n(x - \delta_2, j)$ and $c_r = P_n(x + \delta_2, j)$ respectively. If left and right valid colors are found and they are distinct, a complete boundary is formed. If not, this boundary is discarded.

3.3 Boundaries over Time

In the last detection stage we need to relate the boundaries found in both difference images D_1 and D_2 so we can reconstruct the 3D points. For every boundary $b_p = (x_p, j_p, a_p, c_{lp}, c_{rp})$ on the first list we look for a correspondent boundary $b_q = (x_q, j_q, a_q, c_{lq}, c_{rq})$ in a 7x7 region of the second one. If a correspondence is found and the colors c_{lp}, c_{rp}, c_{lq} and c_{rq} denote a valid *(b,s)-BCSL* code, its decoding is possible. Before that, the final x_p must be updated as

$$x_p = \frac{x_p \cdot a_p + x_q \cdot a_q}{a_p + a_q}. \tag{7}$$

With x_p, j_p and the boundary code defined, the depth of a zero crossing point can be calculated using the camera-projector calibration matrices.

3.4 Error Evaluation Protocol

One of the contributions of this work is an error evaluation protocol to analyze the geometry reconstruction. The analysis of the scanner with this protocol presents the experimental limits of reconstruction density with high definition video and stripe based, complementary color codes. High definition off-the-shelf video equipments provide compressed data. Even so, our results show that the reconstruction using these devices has reasonable precision and accuracy.

We capture planes with different orientations to evaluate and compare their reconstructions. By using least squares on each reconstructed plane set of points we get their equations. To compare the similarity between two planes $\pi_1 : a_1x + b_1y + c_1z = d_1$ and $\pi_2 : a_2x + b_2y + c_2z = d_2$ the following metric based on the distance in the projective space is proposed:

$$d = \sqrt{(\frac{a_1}{d_1} - \frac{a_2}{d_2})^2 + (\frac{b_1}{d_1} - \frac{b_2}{d_2})^2 + (\frac{c_1}{d_1} - \frac{c_2}{d_2})^2}, \tag{8}$$

where each plane $\pi : ax + by + cz = d$ is considered a point $P = (a, b, c, d) \in \mathbb{RP}^3$ and $[a, b, c]$ is a unit vector in $\mathbb{R}^3$.

3.5 Quantitative Analysis

The objective of this protocol is to measure the accuracy and precision of the developed scanner. Accuracy is the exactness of distance measures and precision is related to repeatability and its high values indicate low variations between the calculated distances.

A sequence of five reconstructions of a known planar object is performed for each fixed number of stripes evaluated (Fig. 6 and 7). A mean plane is estimated uniting all five point clouds. This plane minimizes the mean square error between all scanned points for a fixed number of stripes. The precision of the reconstruction is calculated as the standard deviation of the five planes' distances in relation to the mean plane. The accuracy is given by the correspondent mean distance.

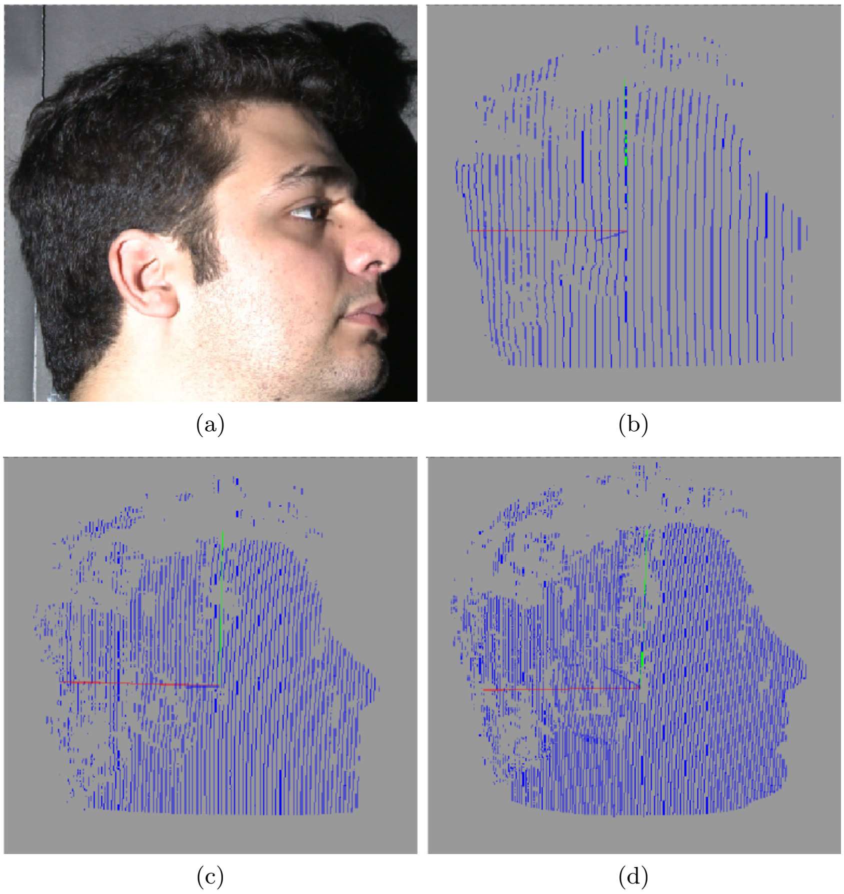
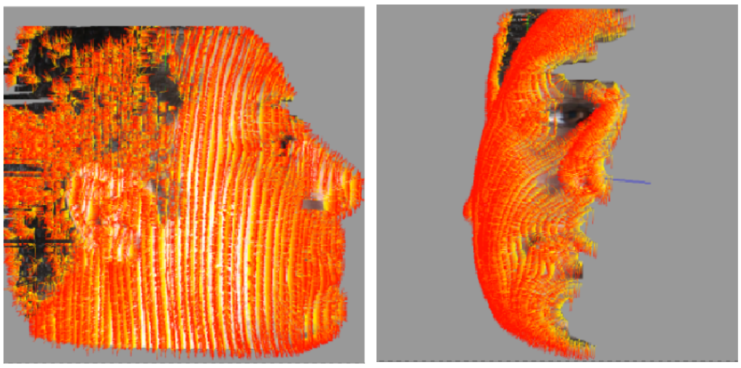


Fig. 2. Reconstruction of a face. (a) Captured scene. (b) 100 stripes; 16860 points. (c) 200 stripes; 29128 points. (d) 250 stripes; 56334 points.

To act as a reference, we create a base plane which is the mean plane of the reconstruction using 50 stripes. We have manually checked the distance given by our scanner for the marked points of the calibration pattern. Several measures were made using a ruler to validate this base plane.

(a)



(b)

Fig. 3. Reconstruction of a face. (a) Textured surface and (b) its normals.

4 Experimental Results

In this section we show the results of the 3D reconstructions of our scanner. we also present the results that relate image's resolutions and number of projected stripes, realized with the developed error protocol.

Figure 2 shows three point clouds of a scanned face. Each of the clouds was generated with a different number of projected stripes: Figures 2 (b) and (c) have 100 and 200 stripes with a 1280x720 resolution. The one with 100 stripes generated 16860 points and the other with 200 stripes has 29128 points. Over Figure 2 (d), there are 250 projected stripes in a 1920x1080 resolution, resulting in 56334 points. Projecting more stripes results in more detailed acquired geometry. With 250 projected stripes we can see a better contour of the chin, mouth and details of the ear. One may see that there are only a few points acquired over the hair. It is a limitation of all active light reconstruction systems: the quality of the recovered geometry is sensitive to the materials that composes the scene and their interaction with projected light.

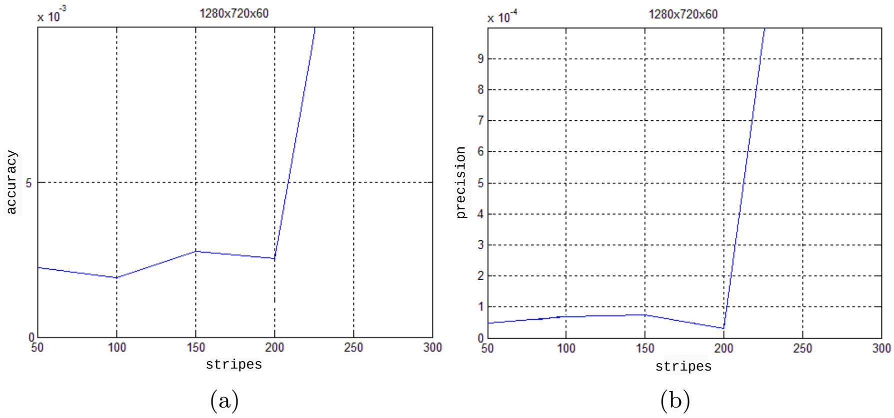


Fig. 4. (a) Accuracy and (b) precision for 1280x720x60.

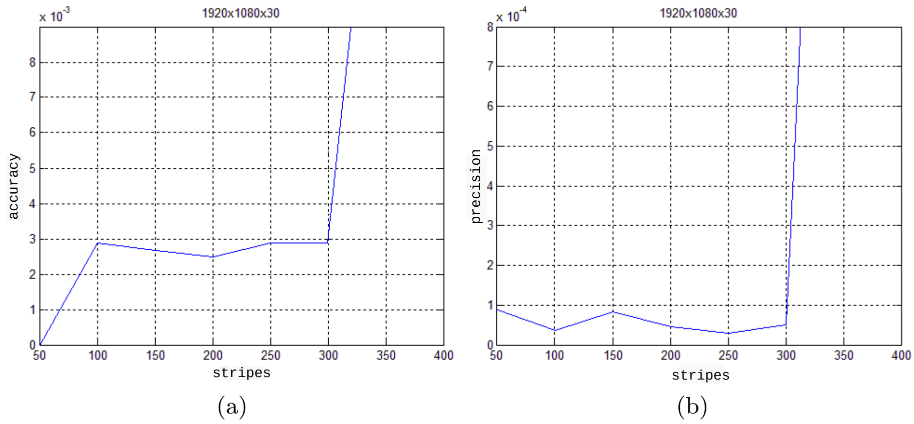


Fig. 5. (a) Accuracy and (b) precision for 1920x1080x30.

In general, translucent, specular, and low reflection intensity materials are problematic. Figure 3 (b) shows the estimated normals for the 250 projected stripes reconstruction of a face and (a) its final result already with texture.

In this section we evaluate the relationship between the number of stripes and the resolution of the captures, using our error protocol. The objective is to estimate, for each resolution, a limit for the number of projected stripes.

The result of accuracy and precision for the configuration 1280x720, with a capture rate of 60Hz is shown on Figure 4. The measurements show a low error between 50 and 200 projected stripes. For 250 stripes though, this error significantly increases, showing that the safe limit of projected stripes for the 1280x720x60 configuration was exceeded. The precision measurement is similar to the accuracy. Low oscillations between 50 and 200 stripes and a high step for the 250 error. These results indicate that similar reconstructions of the plane are repeated until 200 stripes. The accuracy and precision results shows that reliable reconstructions stands until projecting 200 stripes. This may be noticed on Figure 6. For 250 there are decoding errors and miscalculated depths.

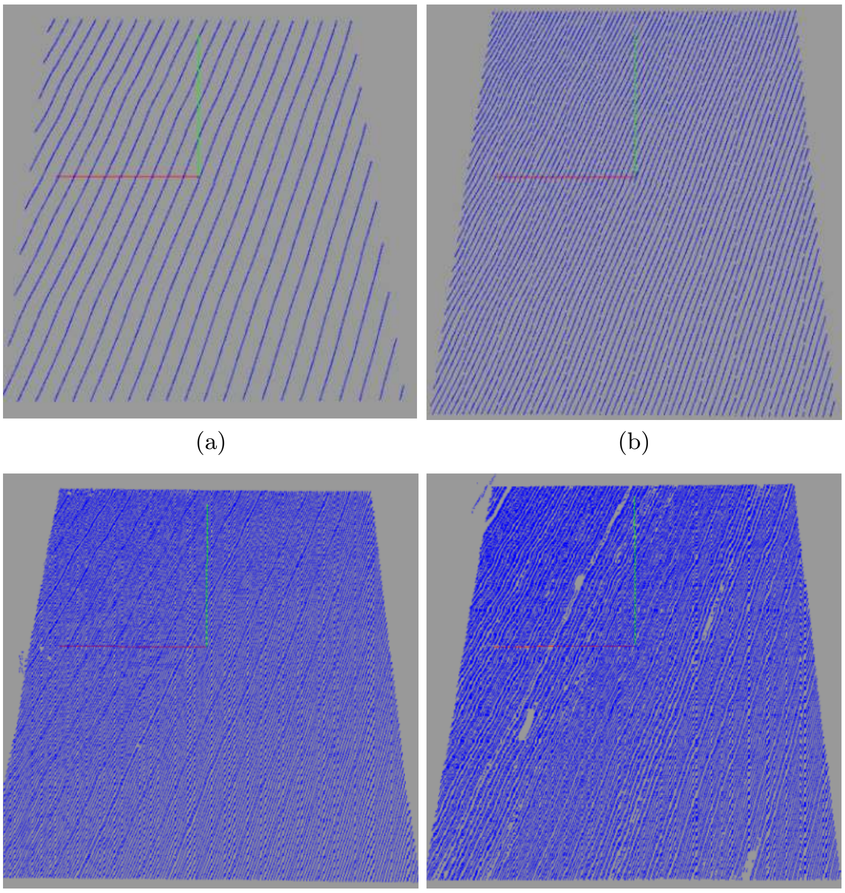
(a) (b)

(c) (d)

Fig. 6. Reconstructed planes for 1280x720x60. (a) 50 stripes; 13694 points. (b) 150 stripes; 40750 points. (c) 250 stripes; 67905 points. (d) 300 stripes; 78668 points.

The results for 1920x1080 with 30Hz configuration is presented on Figure 5. The accuracy error is stable, with low variations, up to 350 projected stripes. While for precision, an error increase is noted after 300 stripes. A safe reconstruction should have a maximum number of 300 projected stripes. Some reconstructions with different numbers of stripes for this resolution are shown on Figure 7. There, one may notice that for 350 stripes there are some miscalculated points at the middle left of the plane. For 400 stripes, in addition to the points reconstructed outside the plane, there are lots of empty areas. Thus, this visual analysis accords with the graphics presented on Figure 5, indicating that for those number of stripes the precision and accuracy of reconstructions are compromised.

The use of planes with different orientations in this error analysis is sufficient to infer the maximum number of stripes for each resolution. Planes with more depth helps to estimate, as specified in Section 3.4, the method's behavior reconstructing more complex objects. With 1920x1080 pixels, 300 stripes is a safe limit for objects with non-null curvatures.



(a) (b)

(c) (d)

Fig. 7. Reconstructed planes for 1920x1080x30. (a) 50 stripes; 20383 points. (b) 150 stripes; 60963 points. (c) 350 stripes; 142225 points. (d) 400 stripes; 156776 points.

5 Conclusion and Future Work

A high resolution scanner, requiring few parameters, and a protocol for reconstruction error evaluation were shown in this paper. Objects of different shapes and materials were reconstructed generating dense point clouds. It was possible to identify points even in specular and black objects. However, active light scanners have limitations. Shadowed areas and some surfaces that have low reflection properties might result in reconstruction imperfections (Fig. 2).

The use of complementary slide patterns provide a robust identification of transition boundaries. Combined with the high speed and high image quality of off-the-shelf camera-projectors, the reconstruction results have reasonable precision and accuracy. Low cost but effective 3D scanners are possible to be constructed with high-definition cameras and projectors.

An error evaluation protocol to analyze precision and accuracy based on projective space distances was proposed. The objective of this protocol was to define an upper bound for the number of projected stripes, assuring a high number of points acquired with a minimal desired error.

A possible application could generate complete 3D models of objects or scenes, aligning point clouds acquired from different views [11, 12].

Acknowledgment

The authors thank to FAPEMIG (*Fundação de Amparo à Pesquisa do Estado de Minas Gerais*), CAPES (*Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior*) and UFJF for funding this research.

References

1. Sa, A., Carvalho, P.C., Velho, L.: (b, s)-bcsl : Structured light color boundary coding for 3d photography. In: 7th Int. Fall Workshop on Vision, Modeling, and Visualization. (2002)
2. Guhring, J.: Dense 3-d surface acquisition by structured light using off-the-shelf components. In: Proc. Videometrics and Optical Methods for 3D Shape Measurement. (2001) 220–231
3. Fechteler, P., Eisert, P., Rurainsky, J.: Fast and high resolution 3d face scanning. In: Proceedings of the International Conference on Image Processing, ICIP 2007, September 16-19, 2007, San Antonio, Texas, USA, IEEE (2007) 81–84
4. Salvi, J., Fernandez, S., Pribanic, T., Llado, X.: A state of the art in structured light patterns for surface profilometry. Pattern Recogn. **43** (August 2010) 2666–2680
5. Zhang, Z.: A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence **22** (2000) 1330–1334
6. Coda, C., Velho, L., Lewiner, T., Peixoto, A.: Um sistema generico de calibracao de camera. In: Proceedings of SIBGRAPI. (2007)
7. Tsai, R.: A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. Robotics and Automation, IEEE Journal of **3**(4) (1987) 323 –344

8. Levenberg, K.: A method for the solution of certain nonlinear problems in least squares. Quart. Appl. Math **2** (1944) 164–168
9. Marquadt, D.: An algorithm for the least-squares estimation of nonlinear parameters. SIAM J. Appl. Math. **11** (1963) 431–441
10. B., V.M., S, A.M., Velho, L., Carvalho, P.C.: A camera-projector system for real-time 3d video. In: IEEE Int. Workshop on Projector-Camera Systems. Volume 3. (2005) 96–100
11. Rusinkiewicz, S., Hall-Holt, O., Levoy, M.: Real-time 3D model acquisition. ACM Trans. on Graphics (Proc. SIGGRAPH) **21**(3) (July 2002) 438–446
12. Besl, P., McKay, N.: A method for registration of 3-d shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence **14** (1992) 239–256