

## SISTEMA E METODOLOGIA SIMPLES DE MESA ROTATÓRIA PARA A CONSTRUÇÃO AUTOMÁTICA DE MODELOS EM 3D

MARCELO LEMOS ROSSI<sup>1</sup>; GILBERTO ARANTES CARRIJO<sup>2</sup>

*Resumo:* Nesse artigo é apresentado uma metodologia para desenvolver um scanner 3D de mesa rotativa, que seja simples, eficiente e de baixo custo, capaz de gerar modelos 3D de objetos que possuam geometria complexa utilizando apenas fotografias tiradas desse objeto. Para descrever todo o processo é apresentado como modelar matematicamente uma câmera para um scanner 3D de mesa rotativa; como calibrar a câmera e quais os dispositivos necessários para ajudar no processo de calibração; como obter os pontos homólogos em imagens; como realizar o processo de associação dos pontos; o processo de triangulação; e finalizando com a obtenção de uma nuvem de pontos representando o objeto em 3D.

*Palavras chaves:* 3D Esperso, Mesa Rotativa, Escaneamento 3D, Visão Computacional, Processamento de Imagem

*Abstract:* - In this paper we present a methodology to develop a simple, efficient and low cost 3D turntable scanner capable of generating 3D models of objects with any arbitrary shapes only using photos from the object. To describe the entire process beginning to how model a stand still camera for a 3D turntable; how to calibrate the camera and the accessories needed to help on the calibration processes; how to obtain the feature points on the image; how to match the feature points; the triangulation processes; and ending with a sparse 3D cloud. Following this methodology is possible obtain a 3D sparse cloud that represents the scanned object.

*Keywords:* - Sparse 3D, Turn Table, 3D Scanning, Computer Vision, Image Processing

### 1 INTRODUÇÃO

---

<sup>1</sup> Faculdade de Engenharia Elétrica, Universidade Federal de Uberlândia, Av. João Naves de Ávila, 2121, Bloco 1E, Uberlândia, MG, CEP: 38400-902. E-mail: marcelo.rossi@ufpel.edu.br

<sup>2</sup> Faculdade de Engenharia Elétrica, Universidade Federal de Uberlândia, Av. João Naves de Ávila, 2121, Bloco 1E, Uberlândia, MG, CEP: 38400-902. E-mail: gilberto@ufu.br

O objetivo desse trabalho é o desenvolvimento de uma mesa digitalizadora 3D, simples, eficiente e de baixo custo, capaz de gerar modelos 3D de objetos com geometria complexa utilizando apenas fotografias do objeto.

Alguns campos de aplicação para o processo pode-se citar a preservação de patrimônio e arqueologia; arte e arquitetura; educação e pesquisa; computação gráfica e efeitos visuais; elementos de jogos e realidade virtual; e impressão 3D. Para todas essas aplicações a geração de cenas virtuais podem ser facilitadas utilizando uma vasta bibliotecas de objetos já modelados em 3D. Dessa forma, um mecanismo simples e eficiente que possa fazer a modelagem dos objetos a serem utilizados pode poupar um tempo razoável. Várias metodologias capazes da reconstrução 3D de cenas têm sido pesquisadas [1 – 4].

## 2 FORMULAÇÃO DO PROBLEMA

Para entender o processo de digitalização do modelo 3D precisa-se compreender o fenômeno físico por trás do processo da fotografia. A Fig.1 apresenta o processo físico por trás da fotografia. Pela Fig.1 pode-se verificar que cada um dos pontos do objeto  $[P_i]$  podem ser representados por três coordenadas  $[X_i \ Y_i \ Z_i]$ , denominadas de coordenadas do espaço objeto. Porém, os pontos da imagem  $[p_i]$  são representados por apenas duas coordenadas  $[u_i \ v_i]$ , denominadas de coordenadas do espaço imagem. Dessa forma, pode-se dizer que o processo de fotografia perde informações da profundidade de uma cena.

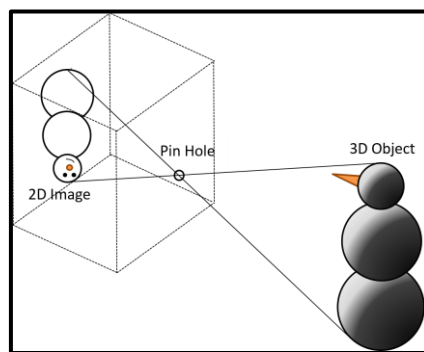


Fig.1: Perda da profundidade no processo da fotografia

De acordo com [5] o modelo matemático para uma câmera de orifício, que é o modelo mais simples de câmera, é definido de acordo com a Equação (1) e simplificado como (2). Uma câmera de orifício é composta por uma câmara opaca em que há um orifício em uma face e isso faz com que a imagem da cena seja projetada na face oposta à do orifício, como apresentada na Fig.1.

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f & s & x_0 & 0 \\ 0 & f & y_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (1)$$

$$[p_i] = [C] \cdot [P_i] \quad (2)$$

Para as Equações (1) e (2) as coordenadas dos pontos  $[P_i]$  e  $[p_i]$  foram escritas no sistema de coordenadas homogêneas [6, 7]. A matriz  $[C]$  é denominada de matriz da câmera, pois ela carrega as características da câmera como:  $f$  corresponde à distância focal;  $s$  corresponde à distorção do pixel; e  $x_0$  e  $y_0$  definem a posição do centro focal da cena no espaço imagem [3]. A matriz  $[C]$  também carrega informações das distorções que o sistema óptico causa ao obter a imagem. Dessa o conhecimento da matriz  $[C]$  é importante para desfazer tais distorções.

Devido ao seu formato, a matriz  $[C]$  é irreversível, e assim, pode-se afirmar que é impossível recuperar a profundidade da cena conhecendo-se apenas uma imagem e a matriz da câmera. Para contornar esse problema deve-se utilizar mais de uma imagem e, então, fazer a triangulação dos pontos [8].

O processo de triangulação consiste em relacionar o mesmo ponto da cena em duas imagens, nas quais já se conhecem as posições das câmeras, e, dessa forma, recuperar a informação sobre coordenada perdida durante a aquisição das imagens.

O processo de triangulação é apresentado na Fig.2 de forma simplificada. Como apresentado pela Fig.2 quando um ponto no espaço objeto é capturado por duas câmeras em posições diferentes esse ponto é observado nas duas imagens obtidas (imagem da esquerda e imagem da direita). Os pontos  $u$  e  $u'$  correspondem às projeções do ponto no espaço objeto nas imagens, respectivamente, da esquerda e da direita. Os pontos  $C$  e  $C'$  representam os centros óptico de cada câmera, ponto no qual os raios de luz passam para formar a imagem em uma câmera, enquanto que os pontos  $e$  e  $e'$ , denominados de epípolos, representam as projeções dos centros ópticos ( $C$  e  $C'$ ) de uma câmera na imagem da outra. Por fim, as linhas  $l$  e  $l'$ , denominadas de linhas epipolares, correspondem a interseção de um plano que passa pelo ponto no espaço objeto e os centros ópticos ( $C$  e  $C'$ ).

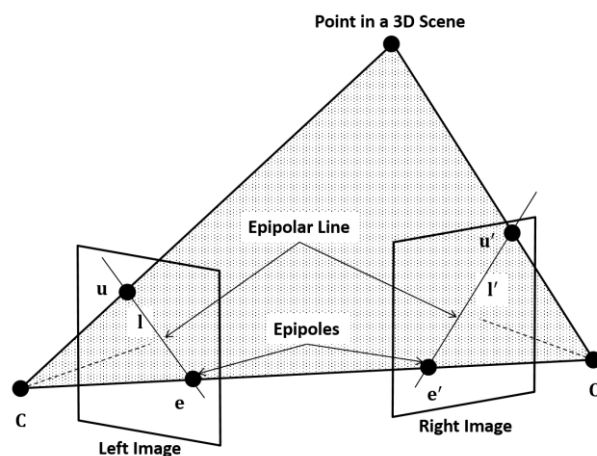


Fig.2: Processo de triangulação

Conhecido todos os elementos apresentados na Fig.2 e os seus ângulos é possível, através da relação de triângulos, obter a distância de um ponto no espaço objeto até as câmeras e obter a coordenada do ponto no espaço. Dessa forma, para que a triangulação seja bem-sucedida é necessário identificar o mesmo ponto da cena em diferentes imagens e, também, conhecer as posições das câmeras.

Para simplificar o cálculo das posições das câmeras sugere-se, nesta metodologia simples para a construção automática de modelos em 3D utilizando mesa rotatória, o uso de um motor de passo para girar o objeto de interesse. Com isso as posições das câmeras serão fornecidas pelo usuário/software ao definir a distância da câmera até o objeto, resolução do motor de passo e a quantidade de passos entre que o motor avançará entre as fotografias.

Os pontos a serem correlacionados não podem ser escolhidos ao acaso, pois isso pode levar a correlações inválidas. Os pontos a serem escolhidos devem possuir características as quais permitam que eles sejam identificados em imagens distintas, evitando pontos alinhados e de forma que eles estejam bem distribuídos ao longo da imagem. Esses pontos são, então, conhecidos como pontos homólogos e existem vários algoritmos que permitem identificar pontos homólogos nas imagens como o FAST [9], Harris [10], SURF [11], BRISK [12] e o MSER [13].

Conhecido os vários pontos homólogos em duas imagens distintas basta correlacioná-los, ou seja, deve-se encontrar os pares correspondentes. Esse processo deve ser feito de forma que o conjunto apresente o menor erro possível. Técnicas como a dos mínimos quadrados podem levar a erros como apresentados na Fig.3 pois ela pode obter pontos trocados, como no caso do ponto 1 da Fig.3 ter sido escolhido de forma correta, então os pontos 4 e 3 foram escolhidos de forma errada. Entretanto, se os pontos 3 e 4, da Fig.3, estiverem sido escolhidos de forma correta então o ponto 1 estaria trocado.

Esse erro se dá porque as técnicas baseadas – mínimos quadrados tendem a obter uma resposta baseado na maior probabilidade de o conjunto ocorrer, sem levar em consideração os efeitos individuais de cada ponto. Na tentativa de contornar esses erros desenvolveu-se algoritmos mais robustos capazes de superar essa dificuldade. Estimadores Robustos bem conhecidos são o de Mediana e o M-estimadores. Entretanto, para o condicionamento robusto dos modelos paramétricos utilizados em visão computacional o algoritmo RANSAC tornou-se o método padrão [14].

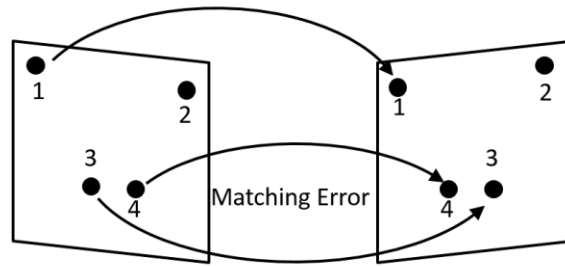


Fig.3: Erro no processo de correlação

Para finalizar o processo de triangulação precisa-se conhecer a posição de cada câmera. De forma a obter um sistema simples será utilizado um motor de passo. Esse tipo de motor possui avanços discretos e fixos. Assim, as posições de câmera são conhecidas *a priori* através da escolha do usuário/software.

### 3 MESA DIGITALIZADORA 3D

Para cumprir com o objetivo desse trabalho, sistema e metodologia simples de mesa rotatória para a construção automática de modelos em 3D, criou-se a estrutura de uma mesa digitalizadora 3D, controlada pelo computador, que permita realizar os cálculos geométricos com facilidade e capaz de obter resultados eficientes. A Fig.4 apresenta uma visão geral dessa mesa digitalizadora 3D.

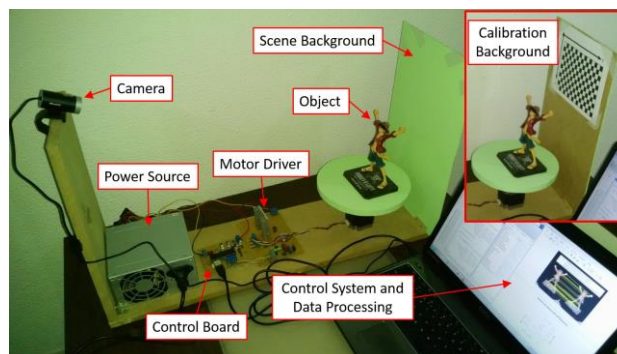


Fig.4: Mesa digitalizadora 3D e seus acessórios

Como pode ser visto na Fig.4, a mesa digitalizadora 3D é composta por uma câmera fixa; um motor de passo capaz de fornecer o movimento preciso e, também, permitir conhecer *a priori* o posicionamento das câmeras; uma base rotatória para apoiar o objeto a ser modelado na cor verde limão, de forma a permitir utilizar a técnica de chroma-key; fundo de cena com um lado verde limão (chroma-key) e do outro com um mecanismo de calibração da câmera; circuito de acionamento do motor; conexão USB para ligar o motor ao sistema de controle e processamento de dados; fonte de potência para alimentação do motor; e um computador que fará o papel do sistema de controle e processamento de dados.

### 3.1 Câmera fixa

Para adquirir as imagens a escolheu-se a câmera Microsoft® Life Cam Cinema. Ela foi escolhida devido a sua alta resolução, de até 5 megapixel (2880X1620 pixels). Dessa forma a câmera consegue capturar muitos detalhes da cena. Esses detalhes serão importantes na hora de detectar pontos homólogos da cena. Outras características da câmera são sistema de estabilização da imagem e zoom automático. A câmera é apresentada na Fig.5.



Fig.5: Câmera da mesa digitalizadora 3D

### 3.2 Mesa giratória, motor e seus acessórios

A mesa giratória consiste de um apoio circular feito de compensado com 20 cm de diâmetro e 15 cm de espessura.

Para girar o apoio circular é utilizado o motor de passo. O motor de passo apresenta como vantagem o controle preciso do ângulo de giro do apoio circular e, dessa forma, tem-se um bom controle da posição do objeto que será digitalizado. O ângulo de passo desse motor é de  $1,8^\circ$  por passo, porém utilizou-se a técnica de meio passo para o controle do motor melhorando, assim, a resolução do movimento para  $0,9^\circ$  por pulso de controle.

O controle do motor é feito por uma interface USB de forma a permitir um computador orientar o motor e, assim, o software de modelagem 3D terá o conhecimento *a priori* da posição da câmera para o processo de triangulação.

### 3.3 Fundo da cena

O fundo da cena é composto por um anteparo que desempenha duas funções. Um dos seus lados é de cor verde limão, de forma a permitir o uso de técnicas de chroma-key e, assim, remover o fundo da cena deixando apenas o objeto em evidência.

Do outro lado do anteparo foi colocado um padrão quadriculado cuja função é a calibração da posição da câmera e a obtenção dos coeficientes da matriz da câmera. Como esse padrão possui forma conhecida e bem definida, ao ser fotografado em diversas posições diferentes, ele sofrerá diversas deformações causadas pelo sistema óptico da câmera. Assim, ao comparar as imagens distorcidas do padrão quadriculado com a geometria real do padrão calcula-se as distorções causadas pelo sistema óptico da câmera e obtém-se, assim, a matriz  $[C]$ , também conhecida como matriz característica da câmera.

Este padrão quadriculado pode ser facilmente identificado através do algoritmo ROCHADE (Robust Checkerboard Advanced Detection for Camera Calibration) [15]. A Fig.6 mostra a detecção do padrão quadriculado pelo algoritmo ROCHADE independente da orientação do padrão quadriculado.

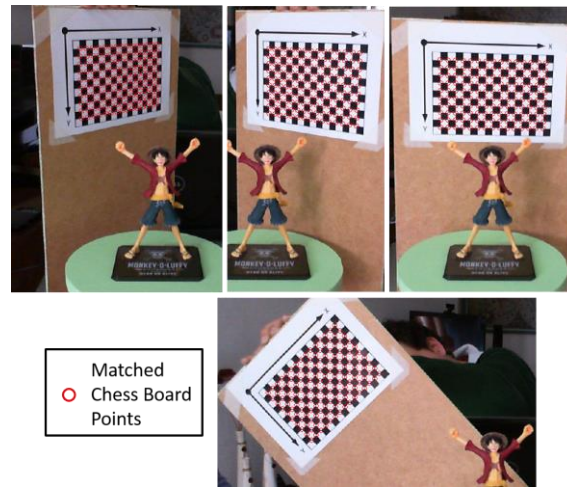


Fig.6: Resultado do algoritmo ROCHADE

Com a capacidade de detecção do padrão quadriculado é possível utilizar o algoritmo de quatro passos definido por Heikkalä e Silvén [16] para a obtenção da matriz característica da câmera.

### 3.4 Movimento da Câmera

Conhecendo matriz característica da câmera precisa-se, agora, conhecer as posições da câmera para cada imagem.

A bancada foi projetada de forma que a câmera fica parada enquanto o objeto é racionado sobre a mesa giratória. Entretanto, o equacionamento necessário para descrever o objeto girando torna essa aproximação difícil.

Para facilitar a abordagem matemática foi utilizada a inversão de mecanismo. Com a inversão de mecanismo o centro de coordenadas global foi definido como o centro da base giratória e, por essa técnica, a câmera é visualizada girando ao redor do objeto, como apresentado na Fig. 7.

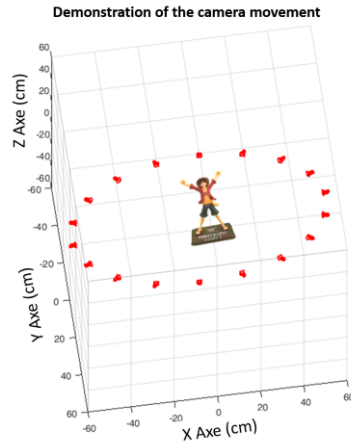


Fig.7: Inversão de mecanismo para a bancada

Definindo-se o eixo vertical do objeto como o eixo Y do sistema de coordenadas espaciais tem-se que o movimento da câmera será no plano ZX. Assim, a rotação da câmera pode ser descrita pela Equação (3) derivada da robótica [17].

$$R = \begin{bmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) \end{bmatrix} \quad (3)$$

Na Equação (3) o ângulo  $\theta$  é o ângulo de giro definido escolhido para o motor de passo girar. Por fim, a posição E da câmera para cada imagem pode ser definida pela Equação (4).

$$E = [T] \times [R] \quad (4)$$

Em que [T] é o vetor que define a posição da câmera ao objeto e pode ser obtido através do algoritmo de calibração definido em [16].

### 3.5 Associação de pontos homólogos em duas imagens

Após identificados os pontos homólogos em duas imagens distintas, utilizando os algoritmos propostos por [9-13], resta fazer-se a associação dos mesmos com os seus pares homólogos, para que seja possível realizar a triangulação e obter a posição do ponto no espaço.

A associação dos pontos homólogos é feita através de testes de compatibilidade dos pontos nas duas imagens, sendo esse teste de compatibilidade feito utilizando o RANSAC. A Fig. 8 apresenta o resultado da associação de pontos homólogos.



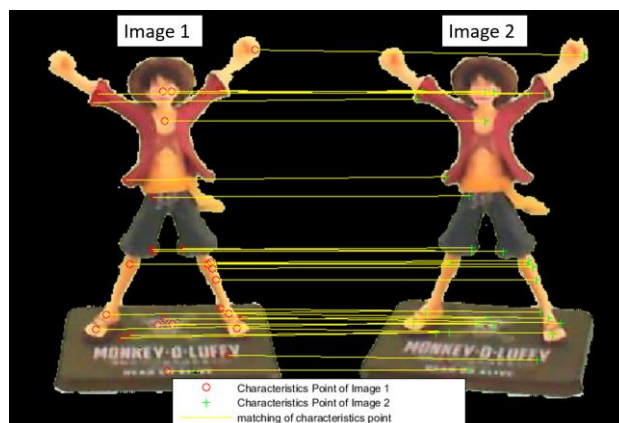


Fig.8: Associação de pontos em duas imagens

### 3.6 Triangulação dos pontos homólogos e o 3D esparso

Conhecido os pares de pontos homólogos e a matriz de características da câmera e as suas posições é possível fazer a triangulação desses pontos casados e, dessa forma, encontrar a distância dos pontos às câmeras, como apresentado na Fig. 2. Conhecida a distância dos pontos e as posições da câmera torna-se possível o posicionamento dos pontos no espaço criando, assim, uma nuvem de pontos no espaço que representa o objeto em análise. Essa nuvem é denominada de 3D Esparso. A Fig. 9 apresenta o resultado do procedimento para a obtenção da forma do objeto analisado em 3D.

### 3.7 Avaliação dos resultados

De forma a comparar os resultados obtidos neste trabalho com o objeto real mediu-se algumas dimensões do objeto para compará-las com as distâncias obtidas pela nuvem de pontos. A Fig.10 apresenta a comparação das medidas ao longo do eixo X. De forma a obter uma medição mais precisa utilizou-se, na Fig.10, dois pontos da nuvem de pontos que representem os extremos da base do objeto utilizado para obter uma medida apenas na direção de X. De acordo com a Fig.10 a base do objeto mede 10,3 cm e a distância dos pontos obtidos que representam os extremos da base, medido apenas da direção do eixo X ( $\Delta X$ ), é de 3,8177 cm. Assim, tem-se um fator de redução de 2,698 entre o objeto real e as dimensões obtidas pela nuvem de pontos na direção do eixo X.

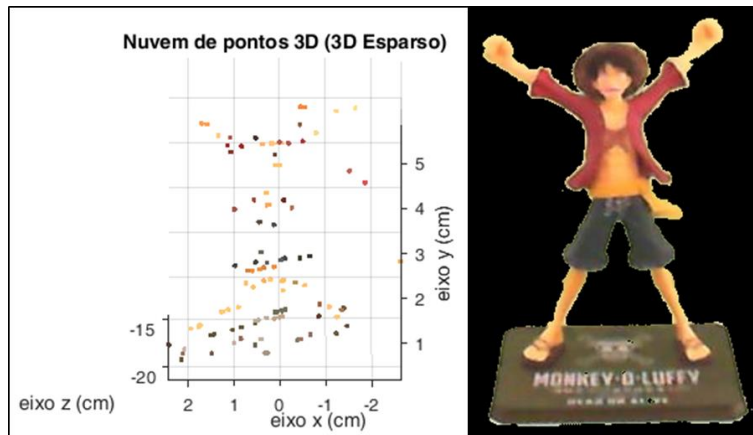


Fig.9: 3D Esparso e o objeto utilizado

A mesma análise feita na direção do eixo X pode ser feita para a direção do eixo Y ao analisar a Fig.11. Pela Fig.11 temos que a distância da parte superior da base do objeto até a parte superior do chapéu do boneco é de 14 cm. Ainda pela Fig.11, ao analisar a nuvem de pontos, temos que a distância na direção do eixo Y ( $\Delta Y$ ) de um ponto da base do objeto até o ponto mais alto do chapéu mede 4,0643 cm. Por fim, obteve-se um fator de redução de 3,377 entre o objeto real e as dimensões obtidas pela nuvem de pontos na direção do eixo Y.

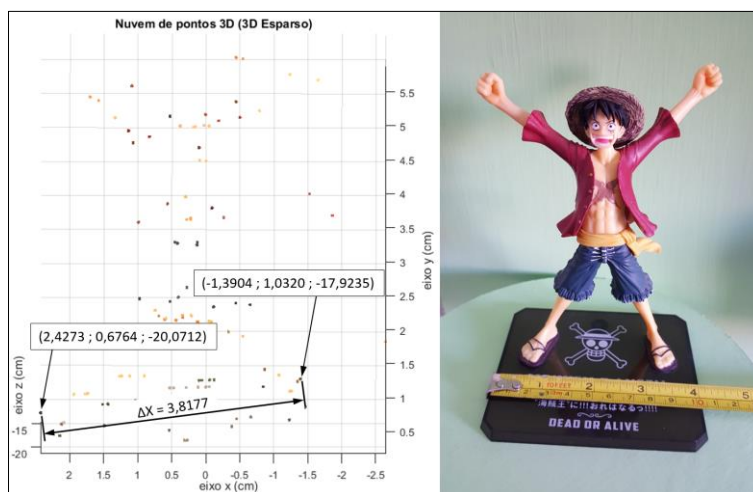


Fig.10: 3D Esparso e o objeto utilizado

Devido pequena quantidade de pontos, na nuvem de pontos, em direção do eixo Z que representem a base do objeto não foi possível, neste momento, avaliar as distâncias nesta direção como foi feito nas Fig.10 e Fig.11.

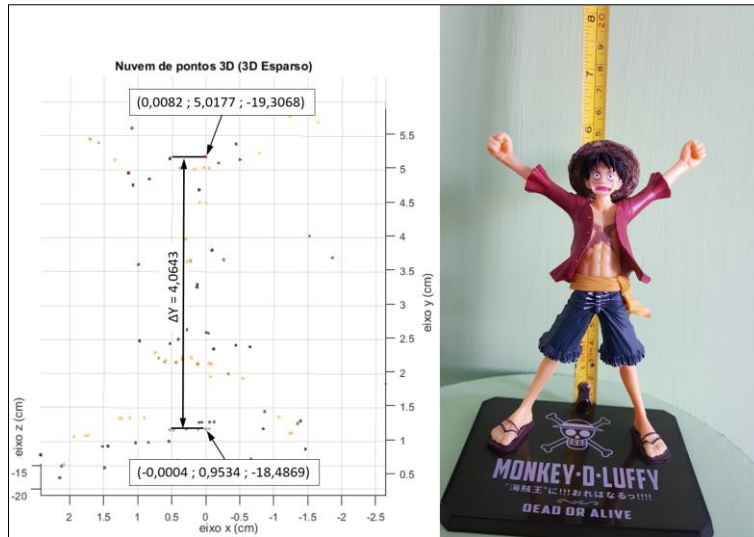


Fig.11: 3D Esperso e o objeto utilizado

## 4 CONCLUSÃO

A metodologia proposta aqui descreve a criação de um instrumento simples, eficiente e de baixo custo, capaz de gerar modelos 3D de objetos com formas complexas utilizando apenas fotografias do objeto. Os resultados iniciais desta pesquisa apresentam que é possível obter uma nuvem de pontos 3D que representa o objeto apenas através de análises das fotografias desse objeto. Entretanto, verifica-se que a nuvem de pontos obtida possui dimensão diferente ao compará-la com o objeto. Nota-se, também, uma diferença de proporção entre o eixo X e o eixo Y, o que indica, a pesar do bom resultado, a necessidade de mais investigações das técnicas de calibração e triangulação.

Assim, este equipamento apresenta-se promissor em diversas áreas que demandam uma metodologia para criação de objetos em 3D como preservação de patrimônio e arqueologia; arte e arquitetura; educação e pesquisa; computação gráfica e efeitos visuais; elementos de jogos e realidade virtual; e impressão 3D.

## 5 AGRADECIMENTOS

Os autores agradecem a Universidade Federal de Uberlândia e a Universidade Federal de Pelotas pela infraestrutura laboratorial. Eles também agradecem ao CNPQ, FAPEMIG e CAPES pelo apoio para a publicação desse artigo.

## REFERÊNCIAS BIBLIOGRÁFICAS:

- [1] H. Cornelius, R. Šára, D. Martinec, T. Pajdla, O. Chum, J. Matas, Towards Complete Free-Form Reconstruction of Complex 3D Scenes from an Unordered Set of Uncalibrated Images, *Lecture*

*Notes in Computer Science*, Springer Science + Business Media, 2004, DOI: 10.1007/978-3-540-30212-4\_1

- [2] O. Faugeras, *Three-dimensional Computer Vision: A Geometric Viewpoint*, The MIT Press, 1993.
- [3] R. Hartley, A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2004.
- [4] Y. Ma, *An Invitation to 3-D Vision: From Images to Geometric Models*, Springer, 2004.
- [5] D. Forsyth, J. Ponce, *Computer Vision: A Modern Approach*, Pearson, 2002;
- [6] J. G. Semple, G. T. Kneebone, *Algebraic Projective Geometry*, Clarendon Press, 1998.
- [7] R. Mohr, Projective Geometry and Computer Vision. In C. H. Chen, L. F. Pau, Wang, P. S. Patrick, *Handbook of Pattern Recognition & Computer Vision*, River Edge, 1993.
- [8] R. I. Hartley, In defense of the eight-point algorithm, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol 19, n 6, pp. 580-593, DOI: 10.1109/34.601246.
- [9] E. Rosten, T. Drummond, Machine Learning for High-Speed Corner Detection, *9<sup>th</sup> European Conference on Computer Vision*, 2006, pp. 430-443.
- [10] C. Harris, M. Stephens, A Combined Corner and Edge Detector, *4<sup>th</sup> Alvey Vision Conference*, 1988, pp. 147-151.
- [11] H. Bay, T. Tuytelaars, L. V. Gool, A Combined Corner and Edge Detector, *4<sup>th</sup> Alvey Vision Conference*, 1988, pp. 147-151.
- [12] S. Leutenegger, M. Chli, R. Siegwart, BRISK: Binary Robust Invariant Scalable Keypoints, *International Conference on Computer Vision (ICCV)*, 2011, pp. 2548-2555.
- [13] J. Matas, O. Chum, M. Urban, T. Pajdla, Robust Wide Baseline Stereo from Maximally Stable External Regions, *The British Machine Vision Conference*, 2002, pp. 1-10.
- [14] M. A. Fischler, R. C. Bolles, Random Sample Consensus: a Paradigm for model Fitting with applications to image analysis and automated cartography, *Communications of the ACM*, 1981, pp. 381-395, DOI: 10.1145/358669.358692.
- [15] S. Placht, P. Fürsattel, E. A. Mengue, H. Hofmann, C. Schaller, M. Balda, E. Angelopoulou, ROCHADE: Robust Checkboard Advanced Detection for Camera Calibration, *Computer Vision – ECCV 2014*, 2014, pp. 766-779, DOI: 10.1007/978-3-319-10593-2\_50.
- [16] J. Heikkilä, O. Silvén, A four-step camera calibration procedure with implicit image correction, *IEEE Conference on Computer Vision and Pattern Recognition*, 1997, pp. 1106-1113, DOI: 10.1109/CVPR.1997.609468.