

## UM MÉTODO ADAPTATIVO PARA REGISTRO DE DADOS RGB-D

### *An adaptive method for registration of RGB-D data*

Marcos Aurélio Basso<sup>1</sup>

Daniel Rodrigues dos Santos<sup>1</sup>

Elizeu Martins de Oliveira Junior<sup>1</sup>

Kourosh Khoshelham<sup>2</sup>

<sup>1</sup>Universidade Federal do Paraná. Departamento de Geomática. Programa de Pós Graduação em Ciências Geodésicas. CEP 81531-Curitiba/PR-Brasil. E-mail: marcosbasso@ufpr.br, danielsantos@ufpr.br, elizeu\_filho.jr@hotmail.com

<sup>2</sup>University of Melbourne. Department of Infrastructure Engineering. Melbourne, Australia. Email: k.khoshelham@unimelb.edu.au

#### **Resumo:**

Neste trabalho é apresentado um método adaptativo para registro e mapeamento de ambientes internos com regiões de pouca textura usando dados derivados de câmeras RGB-D. Basicamente, quatro pontos principais envolvidos no desenvolvimento do método são aqui discutidos: a calibração dos sensores integrados no dispositivo Kinect; o registro dos pares de nuvens de pontos 3D adquiridas sucessivamente pelo sensor de imageamento; a inclusão de pesos no algoritmo ICP baseado em abordagem ponto-a-ponto; e o refinamento dos parâmetros de transformação empregando uma análise de consistência global. A calibração dos sensores é feita usando o método de Zhang. O modelo de corpo rígido 3D é empregado para calcular os parâmetros de transformação. Um método adaptativo é proposto para inclusão de pesos baseados na precisão teórica das medidas de profundidade. Um modelo linear capaz de refinar simultaneamente o conjunto de parâmetros de transformação é proposto para a análise da consistência global. Experimentos foram conduzidos para avaliar a eficiência do método proposto e os resultados obtidos mostraram sua potencialidade.

**Palavras chaves:** câmeras RGB-D; calibração de câmeras; mapeamento de ambientes internos; registro de pares de nuvens de pontos; análise de consistência global, ICP.

#### **Abstract:**

In this paper an adaptive method for indoor mapping and registration of environment with texture less regions using RGB-D data is presented. Basically, four main points involved in the method are here discussed: the calibration of the sensors integrated in the Kinect device; the registration of 3D point clouds pairs acquired successively by the imaging sensor; the inclusion of weights in the point-to-point ICP algorithm; and a global adjustment for the refinement of transformation parameters. The sensors calibration is realized using the Zhang method. The 3D rigid body model is employed to calculate the transformation parameters. The weights are defined based on the theoretical accuracy of depth measurements. A linear model is proposed for global adjustment. Experiments were conducted and the results showed the efficiency and potentially of the proposed method.

**Keywords:** RGB-D cameras; camera calibration; indoor mapping; registration of RGB-D data; global adjustment, ICP.

## 1. Introdução

Atualmente, o desenvolvimento de sistemas de mapeamento 3D de ambientes internos (SMAI-3D) é um dos tópicos mais discutidos em Visão Computacional. Um SMAI-3D é importante em diversas aplicações, tais como, localização e mapeamento 3D simultâneos, modelagem e reconstrução 3D de superfícies, navegação, aplicações forenses, realidade aumentada, medicina, planejamento, entre outras. O desenvolvimento de um SMAI-3D rápido e preciso tem como maior desafio a etapa do registro de pares de nuvens de pontos 3D. Isto se deve devido a regiões com padrões repetitivos ou pouco texturizadas, tais como, superfícies planas comumente encontradas em ambientes internos. Besl e McKay (1992) desenvolveram um algoritmo para registro de pares de nuvens de pontos 3D conhecido como ICP (*Iterative Closest Point*). Basicamente, o algoritmo encontra pontos pseudo conjugados e, iterativamente, determina os parâmetros de transformação (3 rotações e 3 translações) entre os pares de nuvens de pontos 3D. Isto é feito por intermédio de um modelo matemático de transformação rígida e uma função que minimiza a distância entre as primitivas correspondentes. A cada iteração são estabelecidos novos pseudo-conjugados e estimados os parâmetros de transformação. Este procedimento é realizado até que as correspondências sejam exatas e o erro calculado consiga atingir um critério de convergência. No entanto, o algoritmo ICP depende da entrada de valores iniciais aproximados e o processamento dos dados é computacionalmente lento.

Chen e Medioni (1992), Zhang (1994), Blais e Levine (1995), Eggert *et al.* (1998), Segal *et al.* (2009), Sande *et al.* (2010), Grant *et al.* (2012), entre outros, propuseram novas alternativas baseadas na minimização da distância ponto-a-plano para melhorar a eficiência computacional do algoritmo ICP. Godin *et al.* (1994), Magee *et al.* (1985), Weik (1997), Roth (1999) apresentaram métodos que exploraram a sinergia entre dados RGB e valores de profundidade (D) extraídos de imagens de intensidade e nuvens de pontos 3D. Os métodos propostos realizam a etapa de registro dos pares de nuvens de pontos 3D em duas partes. Primeiramente, é feito um registro grosseiro dos pares de dados, seguida de um refinamento dos parâmetros de transformação obtidos. No entanto, devido à baixa qualidade geométrica das imagens de intensidade, a etapa do registro de nuvens de pontos é altamente sensível a ruídos, além de influenciar negativamente na detecção de pontos e no estabelecimento correto das correspondências.

Stamos e Leordeau (2003), Bendels *et al.* (2004), Al-Manasir e Fraser (2006), Ellekilde *et al.* (2007), Huhle *et al.* (2008), Barnea e Filin (2008) e Mostofi *et al.* (2014) integraram dados RGB-D derivados de câmeras digitais e sistemas laser *scanning* terrestre. Enquanto os pontos são automaticamente extraídos na imagem RGB através do algoritmo proposto por Lowe (1999) conhecido por SIFT (*Scale Invariant Feature Transform*), o registro entre pares de nuvens de pontos 3D é feito usando somente o conjunto de pontos extraídos na imagem RGB devidamente associado com seus correspondentes na nuvem de pontos 3D. Segundo Nuchter *et al.* (2007) e Greenspan e Yurick (2003) o alto custo computacional do algoritmo ICP pode ser contornado com o emprego da técnica *k-d tree*. Basicamente, os dados são estruturados em um espaço *k*-dimensional. Assim, ao invés de milhares de cálculos de distâncias euclidianas é feita uma pesquisa nos dados usando propriedades que eliminam rapidamente grandes porções do espaço de busca. Esta abordagem reduz drasticamente o esforço computacional. Porém, sensores de imageamento ativo, tais como, o laser *scanning* e as câmeras de distância 3D são de alto custo, de baixa flexibilidade e relativamente pesados.

Devido ao enorme sucesso das câmeras RGB-D (dispositivos Kinect), sistemas baseados em sensores ópticos tem se tornado uma das opções mais econômicas, rápidas e seguras embarcadas em SMAI-3Ds. O Kinect é leve, flexível, de baixo custo, fácil manuseio e propicia uma coesão natural entre dados RGB-D. Basicamente, esses dispositivos propiciam uma nuvem de pontos 3D com precisão de 3 cm a cada 3m e são capazes de fornecer mais 9,2 milhões de pontos por segundo. May *et al.* (2009), Henry *et al.* (2010), Du *et al.* (2011), Engelhard *et al.* (2011), Steinbrucker *et al.* (2011), Bacharach *et al.* (2012), Dryanovski *et al.* (2012) e Endres *et al.* (2012) propuseram métodos de registro de nuvens de pontos 3D usando dados derivados de dispositivos Kinect. Primeiramente, primitivas pontuais detectadas em imagens RGB são empregadas para determinar os parâmetros de transformação entre pares de nuvens de pontos, seguido de um refinamento dos dados usando o algoritmo ICP. Henry *et al.* (2012) propuseram uma nova variação do algoritmo ICP, conhecida como RGB-D ICP. A abordagem proposta também explora as vantagens das informações derivadas das imagens RGB-D. Embora, o emprego do ICP garanta a eficiência do algoritmo em regiões com superfícies de baixa textura ou com padrões repetidos de informações, a associação entre os dados RGB e os valores de profundidade é feita de forma direta. Neste caso, os desalinhamentos entre os sensores são negligenciados. Os resultados encontrados na trajetória do sensor são de baixa qualidade e o método é altamente sensível a *outliers*. Khoshelham *et al.* (2013) propuseram um método para registro de nuvem de pontos 3D obtidos com câmeras RGB-D baseado em abordagem ponto-a-ponto. Neste trabalho, os autores apresentam um modelo de erros ponderados para melhorar a precisão da trajetória do sensor 3D. A ponderação é baseada no modelo teórico do valor de profundidade. Eles também mostraram a necessidade de alinhar perfeitamente as imagens RGB e IR para associar os pontos 2D e 3D correspondentes usando uma resseção espacial de imagens. O método proposto mostrou melhoras significativas na trajetória do sensor e a potencialidade do emprego de pesos nas observações, entretanto, o algoritmo implementado não é robusto para ambientes internos com regiões pouco texturizadas. Chow *et al.* (2014) desenvolveram um sistema híbrido de mapeamento 3D de ambientes internos. Os autores integraram sensores RGB-D em um sistema laser *scanning* terrestre móvel. Pesos foram incluídos no modelo de otimização dos erros e um dispositivo eletrônico IMU (*Inertial Measurement Unit*) garante a eficiência do algoritmo ICP em regiões pouco texturizadas.

Orientado pelos métodos supracitados, neste trabalho é proposto uma abordagem adaptativa para registro de pares de nuvens de pontos 3D. O método se adapta a superfícies pouco texturizadas usando câmeras RGB-D e pondera o modelo de otimização através da precisão teórica da medida de profundidade. Primeiramente, o algoritmo divide a imagem IR (*InfraRed*) de referência em células igualmente espaçadas. Em seguida, calcula os pontos tridimensionais contidos em cada célula e seu centroide é usado como ponto de referência no processo de estabelecimento automático de correspondências. Os parâmetros de transformação são determinados usando o algoritmo ICP. Finalmente, os parâmetros de transformação são globalmente ajustados através de uma análise de consistência global. A principal contribuição do método proposto é a abordagem proposta para reduzir o custo computacional do algoritmo ICP dividindo a nuvem de pontos de referência em células igualmente espaçadas, além da inclusão de valores de peso derivados da precisão teórica da medida de profundidade, ainda não empregada no algoritmo ICP. Também será apresentada a influência negativa da precisão teórica das coordenadas planimétricas da nuvem de pontos 3D e um método adaptativo para detecção e remoção de *outliers* e para inclusão dos pesos, baseado no valor da Raiz quadrada do Erro Médio Quadrático (REMQ) entre os pontos correspondentes, em cada iteração.

Este trabalho está organizado da seguinte forma: na Seção 1 são apresentados os trabalhos relacionados; na Seção 2 a calibração dos sensores, o registro dos pares de nuvens de pontos e o ajustamento global, as quais são detalhadamente descritos; na Seção 3 os experimentos e as

análises dos resultados são apresentados; por fim, na Seção 4 são apresentadas as conclusões e as recomendações para trabalhos futuros.

## 2. Método

Dado um conjunto de dados adquiridos com uma câmera RGB-D, cujas nuvens de pontos 3D se encontram desalinhadas e transladas uma em relação à outra, é possível reduzir os efeitos do desalinhamento e da translação, e mapear um ambiente interno com superfícies pouco texturizadas de forma rápida, organizada e com precisão usando apenas as informações de profundidade? O método proposto está dividido em quatro etapas, a saber:

- Calibração dos sensores RGB e IR;
- Geração das nuvens de pontos 3D;
- Registro dos pares de nuvens de pontos 3D;
- Análise da consistência global.

### 2.1. Calibração dos Sensores IR e RGB

Neste trabalho o processo de calibração consiste em determinar os parâmetros de orientação interior dos sensores RGB e IR que compõem o dispositivo Kinect, e, o *boresight misalignment* e a distância entre a origem dos sistemas referenciais (*lever arm*) de ambas as câmeras. Para a determinação dos parâmetros de orientação interior do sensor Kinect foi empregado o método de Zhang. Os parâmetros de orientação interior são determinados a partir da seguinte equação (Zhang, 1999):

$$\begin{bmatrix} x'_c \\ y'_c \\ 1 \end{bmatrix} = \begin{bmatrix} fS_x & \tau & x_0 \\ 0 & fS_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (1)$$

sendo,  $x'_c, y'_c$  as coordenadas do ponto no espaço-imagem ( $\mathcal{p}'$ ),  $X, Y, Z$  as coordenadas do ponto correspondente no espaço-objeto ( $\mathcal{P}$ ),  $x_0, y_0$  as coordenadas do ponto principal,  $\tau$  é o fator de não ortogonalidade entre os eixos da câmera,  $f$  a distância focal da câmera,  $S_x, S_y$  os fatores de escala em  $x$  e  $y$ ,  $r_{i,j}$  corresponde aos elementos da matriz de rotação em função dos ângulos de Euler ( $\omega, \phi, \kappa$ ) com  $i=j=1, \dots, 3$ ,  $t_x, t_y, t_z$  são os parâmetros de translação e  $\cdot$  é o operador de multiplicação.

Ainda é necessário considerar os coeficientes de distorção do sistema de lentes da câmera (distorção radial e descentrada), como segue (Brown, 1971):

$$\delta_x = \bar{x}(k_1 r^2 + k_2 r^4 + k_3 r^6) + P_1(r^2 + 2\bar{x}^2) + 2P_2 \bar{x} \bar{y}^2 \quad (2)$$

$$\delta_y = \bar{y}(k_1 r^2 + k_2 r^4 + k_3 r^6) + P_2(r^2 + 2\bar{y}^2) + 2P_2 \bar{x} \bar{y}^2 \quad (3)$$

sendo,  $x'_c$  e  $y'_c$  as coordenadas do ponto no espaço-imagem corrigidas das distorções radiais simétricas e descentradas,  $\bar{x} = x' - x_0$ ,  $\bar{y} = y' - y_0$ ,  $\bar{r} = \sqrt{\bar{x}^2 + \bar{y}^2}$ ,  $k_1, \dots, k_3$  coeficientes de distorção radial das lentes e  $P_1, P_2$  os termos da distorção descentrada.

A relação entre os pontos no espaço-imagem e seus correspondentes no espaço-objeto é obtida usando um plano padrão. Usualmente, um tabuleiro de xadrez é empregado como campo de calibração e, por este motivo, a coordenada Z é inserida como injeção com valor igual a 0 (zero). A injeção imposta define uma transformação projetiva entre o campo de calibração e o plano da imagem. Desta forma, o ponto  $p'(x'_c, y'_c, 1)$  no espaço-imagem e seu correspondente  $P(X, Y, 0, 1)$  no espaço-objeto são representados por coordenadas homogêneas, como segue (Zhang, 1999):

$$s \cdot p' = H \cdot P \quad (4)$$

sendo,  $s$  o fator de escala e  $H = A \cdot [r_1 \ r_2 \ t]$  a matriz homográfica (Hartley e Zisserman, 2003),  $A$  é o primeiro termo do lado direito da Equação 1,  $r_1, r_2, t$  são os elementos da primeira e segunda linhas da matriz de rotação e o vetor de translações em  $x$  e  $y$ , respectivamente.

A partir da Equação 4, os parâmetros de orientação exterior são, simultaneamente, determinados segundo a matriz  $H$ . Fazendo  $H = [h_1, h_2, h_3]$ , obtém-se (Zhang, 1999):

$$H = [h_1, h_2, h_3] = sA[r_1, r_2, t] \quad (5)$$

Como  $r_1$  e  $r_2$  são vetores unitários, tem-se:

$$h_1^T A^{-T} A^{-1} h_1 = h_2^T A^{-T} A^{-1} h_2 \quad (6)$$

Para cada imagem capturada obtêm-se duas Equações lineares, envolvendo elementos da matriz simétrica dada por  $B = A^T \cdot A^{-1}$ , como segue:

$$\begin{bmatrix} v_{12} \\ v_{11} - v_{22} \end{bmatrix} b = 0 \quad (7)$$

sendo,  $b = [B_{11} \ B_{12} \ B_{22} \ B_{13} \ B_{23} \ B_{33}]^T$  os elementos da matriz  $B$  e  $v_{ij} = [h_{i1}h_{j1}, h_{i1}h_{j2}, h_{i2}h_{j1}, h_{i2}h_{j2}, h_{i3}h_{j1} + h_{i1}h_{j3}, h_{i3}h_{j2} + h_{i2}h_{j3}, h_{i3}h_{j3}]^T$ .

De acordo com Zhang (1999) para  $n$  imagens capturadas resolver a Equação 7 significa resolver um sistema de Equações do tipo  $V \cdot b = 0$ , sendo  $V$  uma matriz de dimensões  $(2n_i * 6)$ . A solução para a Equação 5 pode ser encontrada pelo método conhecido como decomposição de valores singulares (*Singular Values Decomposition* - SVD). Desta forma, obtêm-se os parâmetros de orientação exterior da câmera fazendo (Zhang, 1999):

$$r_1 = \lambda A^{-1} h_1 \quad (8)$$

$$r_2 = \lambda A^{-1} h_2 \quad (9)$$

$$r_3 = r_1 \otimes r_2 \quad (10)$$

$$t = \lambda A^{-1} h_3 \quad (11)$$

sendo,  $\lambda = \frac{1}{\|A^{-1}h_1\|} = \frac{1}{\|A^{-1}h_2\|}$  e  $\otimes$  é o operador de produto vetorial. Vale ressaltar que a melhor solução para a matriz de rotação pode ser encontrada em Golub e van Loan (1996).

De acordo com Zhang (1999) o método apresentado encontra a solução minimizando uma distância algébrica que não tem representação física. Para refinar a solução obtida pode-se empregar uma estimativa por máxima verossimilhança. Assumindo que  $p_i'$  esteja isento de erros sistemáticos, cujas coordenadas possuem a mesma incerteza na medida, então, a matriz H pode ser obtida pelo método de Levenberg-Marquardt, como segue (More, 1997):

$$\min_H \sum_{i=1}^n \|p_i - p_i'\|^2 \quad (12)$$

sendo,  $p'_i = \frac{1}{h_3^T P_i} \begin{bmatrix} h_1^T & P_i \\ h_2^T & P_i \end{bmatrix}$ ,  $h_i^T$  representa as linhas de índice  $i$  da matriz homográfica, enquanto  $n$  é o número de imagens,  $p_i$  representa as coordenadas do ponto no espaço-imagem e  $P_i$  representa as coordenadas do ponto correspondente no espaço-objeto.

Segundo Zhang (1999) uma forma natural de determinar simultaneamente todos os parâmetros envolvidos no problema de calibração pode ser feita pelo método de Levenberg-Maquardt, como segue:

$$\min \sum_{i=1}^n \sum_{j=1}^m \|p_{i,j} - p'(A, k_1, k_2, k_3, P_1, P_2, R_i, t_i, P_j)\|^2 \quad (13)$$

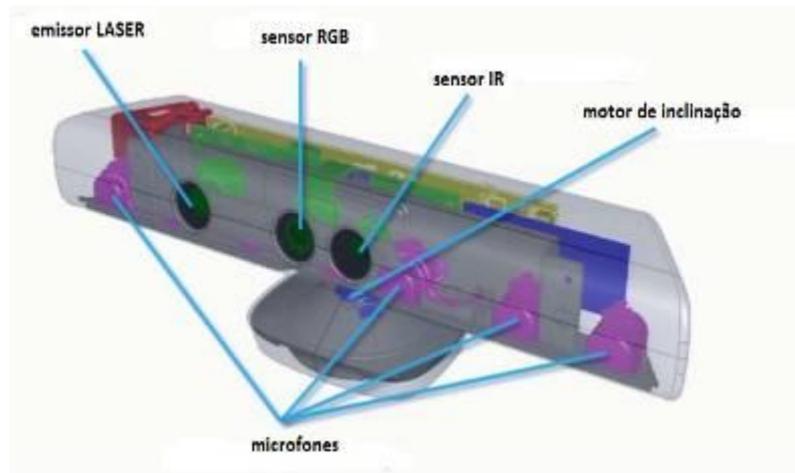
sendo,  $m$  o número de pontos na imagem,  $n$  o número de imagens,  $p'(A, k_1, k_2, k_3, P_1, P_2, R_i, t_i, P_j)$  representa a projeção do ponto  $P_j$  na imagem  $i$ ,  $R_i$  é a matriz das rotações e  $t_i$  é o vetor das translações. Vale lembrar que  $p'$  é função das coordenadas corrigidas das distorções do sistema de lentes como apresentada nas Equações 2 e 3. Haja vista que é um problema não-linear, os valores iniciais podem ser obtidos fazendo  $x = [h_1^T, h_2^T, h_3^T]$ , então a Equação 4 pode ser reescrita como (Zhang, 1999):

$$\begin{bmatrix} P^T & 0^T & -x_0 P^T \\ 0^T & P^T & -y_0 P^T \end{bmatrix} x = 0 \quad (14)$$

Reescrevendo a Equação 14 da forma matricial  $Lx = 0$ , a solução pode ser obtida através do auto vetor de  $L^T L$  associado ao menor auto valor. Maiores detalhes ver Zhang (1999).

## 2.2. Geração da nuvem de pontos 3D

O Kinect é um sensor de movimentos desenvolvido originalmente para o console de videogame Xbox 360 e Xbox One, em parceria com a empresa Prime Sense (Figura 1). O dispositivo permite ao usuário do console interagir com jogos eletrônicos sem a necessidade de um controle físico. É composto por três sensores: dois sensores CMOS (*Complementary metal-oxide-semiconductor*) que registram energia eletromagnética na faixa do espectro correspondente ao visível (câmera RGB) e infravermelho (câmera IR); e um emissor laser infravermelho. Valores de paralaxe ou de profundidade são determinados pelo princípio de triangulação, por intermédio da detecção de alvos (*patterns*) emitidos pelo sensor laser e registrados no sensor da câmera IR, enquanto os valores de cinza dos pixels são medidos de forma independente na câmera RGB. Os sensores dispostos no Kinect podem capturar cenas com 640 x 480 pixels em uma taxa de 30 quadros por segundo. Os dados são adquiridos com precisão da ordem de 3 cm a cada 3 m, com FOV (campo de visão) de ~60 graus para sensores com distância média de aquisição inferior a 7 m.



**Figura 1:** Sensores integrantes do Kinect.

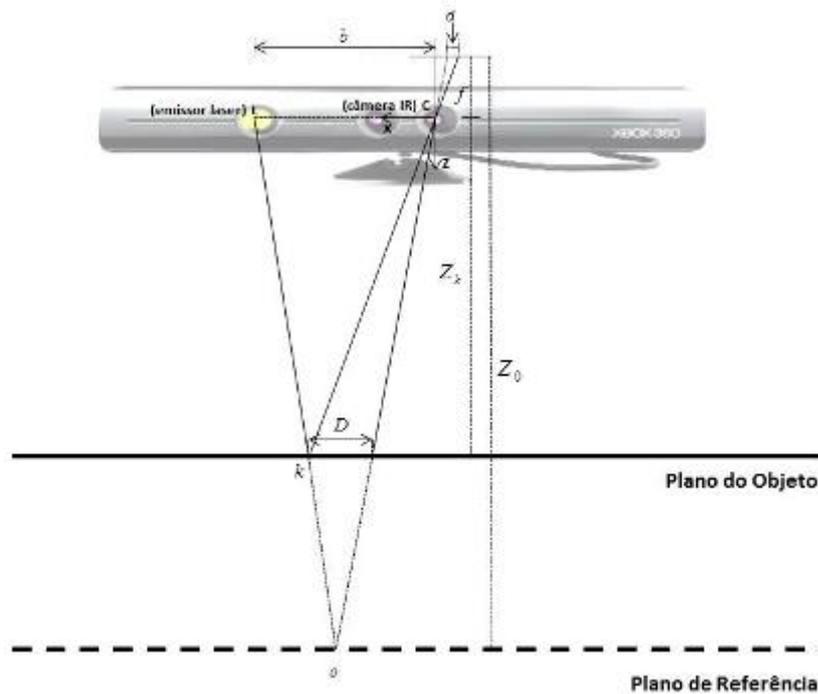
Fonte: <http://kinectando.com.br>.

Para gerar os mapas de profundidade utilizados na reconstrução tridimensional de objetos e ambientes internos o dispositivo utiliza a técnica conhecida como luz estruturada. Basicamente, a técnica de luz estruturada consiste na projeção e captura de padrões de luz IR. Uma fonte de luz emite um feixe de luz que é dividido em múltiplos feixes criando um padrão aleatório (Figura 2), o sensor IR captura o padrão projetado e o correlaciona com um padrão de referência. Este padrão de referência é obtido através da captura de um plano posicionado a uma distância conhecida do sensor e armazenado na memória do dispositivo. A reconstrução é feita a partir da relação geométrica entre o plano de referência e a paralaxe observada em cada ponto capturado.



**Figura 2:** Padrão de luz estruturada gerado pelo Kinect.

A Figura 3 ilustra a geometria do sistema, o deslocamento dos padrões projetados bem como a medida da paralaxe observada.



**Figura 3:** Geometria do modelo matemático.

sendo,  $C$  o centro de perspectiva do sensor IR,  $L$  o emissor IR,  $X$  e  $Z$  são os eixos do sistema de coordenadas centrado no centro de perspectiva do sensor IR,  $k$  um ponto do espaço-objeto,  $o$  ponto no plano de referência correspondente ao ponto no espaço-objeto,  $D$  o deslocamento do ponto  $k$  no espaço-objeto,  $Z_0$  a distância do sensor IR ao plano de referência,  $Z_k$  a distância ou profundidade entre o sensor IR e a superfície física,  $f$  a distância focal do sensor IR,  $d$  a paralaxe observada e  $b$  a linha base (distância entre o sensor e emissor IR).

Assumindo que um objeto esteja posicionado entre o plano de referência e o sensor IR, um ponto capturado será transladado na direção  $X$  em relação ao ponto correspondente armazenado na memória do dispositivo. Como pode ser observado na Figura 03, este deslocamento corresponde à paralaxe observada no posicionamento do ponto  $k$ . A partir de semelhança de triângulos temos (Khoshelham e Elberink, 2012):

$$\frac{D}{b} = \frac{Z_0 - Z_k}{Z_0} \quad (15)$$

$$\frac{d}{f} = \frac{D}{Z_k} \quad (16)$$

Substituindo a Equação (16) na Equação (15), é possível expressar  $Z_k$  da seguinte forma (Khoshelham e Elberink, 2012):

$$Z_k = \frac{Z_0}{1 + \frac{Z_0}{fb}d} \quad (17)$$

A Equação 17 é o modelo matemático básico para a determinação da profundidade de um ponto observado a partir de sua paralaxe e assumindo que os parâmetros  $Z_0$ ,  $f$  e  $b$  são

determinados por técnicas de calibração. A coordenada  $Z$  em conjunto com  $f$  definem a escala da imagem para este ponto. As coordenadas planimétricas para cada ponto do objeto pode então ser calculadas a partir de suas coordenadas na imagem e da escala (Khoshelham e Elberink, 2012):

$$X_k = \frac{Z_k}{f} (x' - x_0 + \delta_x) \quad (18)$$

$$Y_k = \frac{Z_k}{f} (y' - y_0 + \delta_y) \quad (19)$$

De acordo com Khoshelham e Elberink (2012), as medidas de paralaxe advindas do Kinect são normalizadas e quantificadas entre zero e 2.047, assim, fazendo  $d = md' + n$ , o valor da paralaxe  $d$  (Equação 17) pode ser expresso em função de uma normalização linear. Substituindo os termos apresentados na Equação 17, tem-se:

$$Z_k^{-1} = md' + n \quad (20)$$

sendo,  $m$  e  $n$  os parâmetros da normalização linear e  $d'$  a paralaxe normalizada.

A Equação 20 expressa uma relação linear entre o inverso da profundidade de um ponto e seu valor de paralaxe normalizado correspondente. Os coeficientes desta relação linear podem ser estimados pelo Método dos Mínimos Quadrados (MMQ). O método proposto para determinação dos parâmetros de desalinhamento angular e deslocamento da origem entre os sensores RGB e IR, assim como a determinação dos valores  $b$  e  $Z_0$  pode ser encontrado em Khoshelham e Elberink (2012).

### 2.3. Registro de pares de nuvens de pontos 3D usando o ICP

Dada duas nuvens de pontos  $M$  (referência) e  $N$  (pesquisa), o algoritmo ICP consiste em estimar os parâmetros de transformações (3 rotações e 3 translações) entre os pares de nuvens de pontos  $N$  e  $M$ , ambos com pontos definidos em  $\mathbb{R}^3$ . Neste trabalho é proposta uma estratégia adaptativa o registro dos pares de nuvens de pontos 3D. Neste caso, considere um par de imagens IR (imagem de referência e imagem de pesquisa) obtidas com o dispositivo Kinect. A imagem de referência é regularmente dividida em células de dimensão  $dim \times dim$ . Para cada célula existe um conjunto de pixels, cujas coordenadas 3D são calculadas usando as Equações 17-19. Desta forma, é determinado o centroide ( $q$ ) de cada célula, como segue:

$$q = \frac{1}{j} \sum p_j \quad (21)$$

sendo,  $p_j$  o  $j$ -ésimo ponto XYZ contido na célula e  $j$  o número de pontos na célula.

Como resultado, tem-se um conjunto de centroides na nuvem de pontos de referência, cujas correspondências devem ser encontradas na nuvem de pontos de pesquisa. Vale destacar que para cada pixel na imagem de pesquisa também são determinadas suas coordenadas 3D usando as Equações 18-20. Sendo assim, um conjunto de  $p_n$  pontos de pesquisa é então obtido.

Para acelerar o processo de busca pelo vizinho mais próximo foi empregado a estrutura *k-d tree*, proposta por Friedman *et al.* (1977). Basicamente, o algoritmo realiza as seguintes etapas, a saber: Dado um ponto centroide ( $q$ ) e uma raiz da árvore de pesquisa ( $N$ ) faz-se:

1. Para cada  $N$  divide-se todas as folhas da árvore de pesquisa ao longo de  $n$  usando um hiperplano que é perpendicular ao eixo correspondente;
2. Armazena-se dois sub-nós de cada  $N$  interno, um índice  $n$  que especifica a dimensão ao longo do qual  $N$  divide o espaço, e um escalar  $s$  que determina o ponto de divisão entre as duas sub-árvores;
3. Todos os pontos ( $p_n$ ) mais próximos de  $M$  são localizados;
4. O ponto mais próximo ( $p_n$ ) de  $q$  que se intersecta em  $R$  é o correspondente;
5. As etapas 1-4 são repetidas até que seja pesquisado o último centroide.

Encontrado um conjunto de pontos correspondentes, a cada iteração, o modelo de corpo rígido 3D é empregado para determinar os parâmetros de transformação. O modelo de otimização minimiza o somatório do erro de reprojeção ( $E$ ), explicitamente, para reduzir a distância euclidiana de todos os pontos correspondentes nas nuvens de pontos de referência e de pesquisa, como segue:

$$E = \sum_{m=1}^j w_{mn-1} \|\vec{X}_{m,n-1} - (R \cdot \vec{X}_{m,n} + \vec{T})\| \quad (22)$$

sendo,  $\vec{X}_{m,n}$  e  $\vec{X}_{m,n-1}$  os vetores das coordenadas 3D correspondentes nas nuvens de pontos  $M$  e  $N$ , respectivamente,  $w_{mn}$  o peso associado ao par de pontos correspondentes  $m$  e  $n$ ,  $\vec{T}$  o vetor dos parâmetros de translação ( $t_x$ ,  $t_y$  e  $t_z$ ) e  $R$  a matriz de rotação (3x3) em função dos ângulos de Euler ( $\omega$ ,  $\phi$ ,  $\kappa$ ).

Na Equação (22),  $w_{mn}$  é definido em função da precisão teórica das coordenadas de profundidade. As imagens de profundidade são adquiridas em uma taxa de 20 a 30 quadros por segundo (no dispositivo Kinect). Esta frequência de aquisição de dados resulta em pequenas diferenças de rotação e translação entre as imagens consecutivas. Desta forma, se pode aproximar as Equações de observação, como segue (Khoshelham *et al.*, 2013):

$$v_i = \|\vec{X}_{m,n-1} - \vec{X}_{m,n}\| \quad (23)$$

Para Khoshelham *et al.* (2013) os pesos podem ser definidos da seguinte forma:

$$w_{m,n} = \frac{c}{\sigma_{v_i}^2} = \frac{c}{\sigma_{\vec{X}_{m,n-1}}^2 + \sigma_{\vec{X}_{m,n}}^2} \quad (24)$$

sendo,  $\sigma_{\vec{X}}$  a variância de cada coordenada do vetor  $\vec{X}$  e  $c$  uma constante arbitrária.

Aplicando a propagação de covariâncias na Equação (20) e fazendo algumas manipulações matemáticas obtém-se a variância da coordenada  $Z_k$ , como segue (Khoshelham *et al.*, 2013):

$$c_{Z_k}^2 = c_d^2 \sigma_d^2 Z_k^4 \quad (25)$$

sendo,  $\sigma_d$  a precisão da medida de paralaxe.

Substituindo a Equação 23 na Equação 22 obtém-se a ponderação da medida de profundidade, como segue (Khoshelham *et al.*, 2013):

$$w_{Z_k} = \frac{\rho}{Z_{m,n-1}^4 + Z_{m,n}^4} \quad (26)$$

sendo, na prática  $\rho = 1$ .

Usando o mesmo raciocínio realizado anteriormente, mas agora para as coordenadas  $X_k$  e  $Y_k$  (ver Eqs. 18 e 19), tem-se:

$$\sigma_{\vec{X}}^2 = \frac{x_c^2}{f^2} \sigma_{Z_k}^2 + \frac{Z_k^2}{f^2} \sigma_{x_c'}^2 \quad (27)$$

sendo,  $\sigma_{\vec{X}}^2$  as variâncias das coordenadas  $X_k$  e  $Y_k$ ,  $\vec{x}_c'$  o vetor das coordenadas  $x_c'$  e  $y_c'$  no espaço-imagem e  $\sigma_{x_c'}^2$  suas variâncias.

Substituindo as Equações 18 e 25 na Equação 27 tem-se:

$$\sigma_{X_k}^2 = c_1^2 \sigma_d^2 x_c' Z_k^2 + \frac{\sigma_{x_c'}^2}{f^2} Z_k^2 \quad (28)$$

Da mesma forma, substituindo as Equações 19 e 25 na Equação 27 tem-se:

$$\sigma_{Y_k}^2 = c_1^2 \sigma_d^2 y_c' Z_k^2 + \frac{\sigma_{y_c'}^2}{f^2} Z_k^2 \quad (29)$$

Substituindo as Equações 28 e 29 na Equação 24 obtém-se a ponderação das medidas em  $X_k$  e  $Y_k$ , como segue:

$$w_{X_k} = \frac{\rho}{Z_{m,n-1}^2 (X_{k,m,n-1}^2 + q_q) + Z_{m,n}^2 (X_{k,m,n-1}^2 + q_q)} \quad (30)$$

$$w_{Y_k} = \frac{\rho}{Z_{m,n-1}^2 (Y_{k,m,n-1}^2 + q_q) + Z_{m,n}^2 (Y_{k,m,n-1}^2 + q_q)} \quad (31)$$

sendo,  $q_q = c_1^{-2} f^{-2} \sigma_d^{-2} \sigma_{x_c'}^2$  constantes definidas pelos parâmetros  $c_1$  e  $f$ .

A Equação 22 não é linear e envolve 6 parâmetros a serem determinados pelo MMQ. Neste caso, deve ser aplicado o modelo paramétrico não linear (Gemael, 1994). Assim como o ICP, neste trabalho, os parâmetros de transformação são determinados iterativamente. A cada iteração, o desalinhamento angular e o deslocamento linear entre o par de nuvens de pontos 3D (referência e pesquisa) é progressivamente reduzido, e uma nova nuvem de pontos 3D de pesquisa é obtida. O algoritmo implementado trabalha da seguinte forma, a saber:

1. Encontra os pseudo-conjugados usando o método *k-d tree*;
2. Detecta e remove *outliers*;
3. Determina os parâmetros de transformação entre o par de nuvens de pontos 3D;
4. Calcula uma nova nuvem de pontos 3D de pesquisa usando os parâmetros estimados;
5. Calcula a REMQ da média da discrepância dos pontos conjugados (REMQ\_MP) da nuvem de referência e de pesquisa;
6. Se a REMQ\_MP calculada for inferior a um valor pré-estabelecido, atribui pesos aos pontos correspondentes;

7. Os passos 1-6 são repetidos até que o REMQ\_MP seja inferior a um limiar pré-definido ou o número de iterações seja superior ao número máximo de iterações definido.

Neste trabalho, é proposto um método adaptativo para executar as tarefas 2, 6 e 7. Como é conhecido, não se pode admitir correspondências exatas entre pontos em um par de nuvens de pontos 3D obtidas pelo Kinect ou laser *scanning*. Isto faz com que o modelo de otimização (Eq. 22) seja altamente sensível a ruídos, quando são encontrados pseudo-conjugados (processo padrão do ICP nas primeiras iterações). Outro fator que deteriora a qualidade dos dados Kinect é o fato de as imagens RGB e de profundidade não serem obtidas de forma sincronizada, fazendo com que a paralaxe calculada e suas coordenadas ( $x'_c$  e  $y'_c$ ) correspondentes no espaço-imagem não sejam exatamente correspondentes. Isto afeta, principalmente, a qualidade posicional das coordenadas  $X_k$  e  $Y_k$  (Eqs. 18 e 19, que são calculadas em função das coordenadas no espaço-imagem e de  $Z_k$ ). Para reduzir a probabilidade de utilizar *outliers* na etapa de registro de pares de nuvens de pontos, é proposto um método adaptativo para detecção e remoção de *outliers* (tarefa 2). Neste método, são definidos dois critérios: o da distância mínima entre os pontos; e o da variação angular da linha reta que liga dois pontos correspondentes em potencial.

O critério da distância mínima se adapta ao valor da REMQ\_MP. Após cada iteração do algoritmo ICP, a REMQ\_MP calculado é adotado como um novo valor de limiar. Somente pontos potencialmente correspondentes que sejam inferiores ao valor da REMQ\_MP são considerados correspondentes. A medida que as diferenças de desalinhamento angular e de deslocamento linear diminuem, as correspondências entre os pontos tendem a serem mais exatas (apesar de nunca serem exatas). Para evitar ainda mais o uso de falsos positivos, em seguida é empregado o algoritmo RANSAC (*Random Sample Consensus*). Maiores detalhes ver Fischler e Bolles (1981). O critério da variação angular da linha reta que liga dois pontos correspondentes em potencial, também é usado para detectar e remover *outliers*. Neste caso, encontrado um par de pontos correspondentes, é calculado o ângulo entre a linha reta, que liga os pontos correspondentes, e o eixo X do sistema referencial arbitrário do sensor.

Desde que, nas primeiras iterações, o algoritmo ICP encontra somente pseudo-conjugados, os valores de peso não são significativos. Neste trabalho, a tarefa de atribuir pesos às observações (tarefa 3) também se adapta ao valor da REMQ\_MP. Neste caso, somente quando o valor da REMQ\_MP calculada for inferior ao limiar pré-estabelecido (por exemplo, inferior a 1 mm que é quando se admite que estejam sendo estabelecidas correspondências “exatas”), os pesos das observações passam a ser atribuídos (Eqs. 26, 30 e 31) no modelo de otimização. Esta também é uma tarefa importante na etapa de registro dos pares de nuvens de pontos, tendo em vista que a inclusão de pesos em pseudo-conjugados pode influenciar negativamente na convergência do algoritmo.

Finalmente, o critério de parada do algoritmo também está em função do valor da REMQ\_MP. O algoritmo termina o processamento caso a REMQ\_MP calculada não variar abruptamente entre duas e três iterações. Isto significa que a estimativa dos dados está estável e o processamento deve ser finalizado. Esta tarefa também reduz o custo computacional do algoritmo ICP.

## 2.4. Análise da Consistência Global

A etapa de análise da consistência global (ajustamento global) consiste em transformar o conjunto de parâmetros de transformação referenciados num sistema local (determinados na etapa de registro dos pares de nuvens de pontos) em um sistema referencial global. Isto é feito através de um ajustamento simultâneo dos parâmetros de transformação. Neste trabalho, é proposto uma simplificação da técnica de ajustamento global apresentada por Lu e Milios

(1997). A partir de um conjunto de nós e arestas (as relações espaciais entre os sensores) é possível formar uma rede em que cada nó corresponde à posição de cada sensor ( $S_i$ , para  $i = 0, \dots$ , número total de posições) em sua trajetória ( $t_x^{S_{i,i+1}}, t_y^{S_{i,i+1}}, t_z^{S_{i,i+1}}$ ). A simplificação proposta neste trabalho é devido ao emprego direto dos parâmetros de transformação advindos da etapa do registro de forma simultânea. Neste caso, se deve ajustar um conjunto de observações através de um estimador ótimo usando as arestas combinadas em um circuito fechado.

A formulação do algoritmo de estimação ótima, a partir do circuito de medidas conhecidas sobre  $n+1$  observações  $X_0 = [t_x^{S_{1,0}} \ t_y^{S_{1,0}} \ t_z^{S_{1,0}}]^T$ ,  $X_1 = [t_x^{S_{2,1}} \ t_y^{S_{2,1}} \ t_z^{S_{2,1}}]^T$ ,  $X_n = [t_x^{S_{n,n+1}} \ t_y^{S_{n,n+1}} \ t_z^{S_{n,n+1}}]^T$  sugere que cada um deles represente um vetor  $6 \times 1$ , e uma aresta  $D_{ij}$  entre duas observações  $X_i$  e  $X_j$  represente uma diferença mensurável. Neste trabalho, a Equação de medida, referente à função  $D_{ij}$  de  $X_i$  e  $X_j$ , é dada na forma linear:  $D_{ij} = X_i - X_j$ .

Portanto, assumindo que há uma ligação entre todas as observações e para cada  $D_{ij}$  existe uma observação  $\bar{D}_{ij}$  que é assumida portadora de uma distribuição gaussiana com média zero e covariância conhecida  $C_{ij}$ , a estimativa é ótima quando minimizada a seguinte distância de Mahalanobis, como segue (Lu e Milios, 1997):

$$Q = \sum_{0 \leq i < j \leq n} (X_i - X_j - \bar{D}_{ij})^T C_{ij}^{-1} (X_i - X_j - \bar{D}_{ij})^T \quad (32)$$

A partir da concatenação de todas as observações  $X_0, X_1, \dots, X_n$ , ( $X$ ) das diferenças  $D_{ij} = X_i - X_j$ , ( $D$ ) e inserindo uma matriz de incidência ( $H$ ), formada por valores 1, -1, e 0, chega-se a Equação linear de medida em sua forma matricial dada por (Lu e Milios, 1997):

$$D = HX \quad (33)$$

Substituindo a Equação 33 na 32 é possível representar função  $Q$  também em sua forma matricial como segue:

$$Q = (\bar{D} - HX)^T C^{-1} (\bar{D} - HX) \quad (34)$$

Sendo  $\bar{D}$  a concatenação de todas as observações  $\bar{D}_{ij}$  referentes à  $\bar{D}$  e  $C$  a covariância de  $\bar{D}$ , a solução para  $X$  que minimiza  $Q$  é dada por (Lu e Milios, 1997):

$$X = (H^T C^{-1} H)^{-1} (H^T C^{-1} \bar{D}) \quad (35)$$

Para compor a matriz  $C$ , neste trabalho, é proposto o uso da matriz variância covariância (MVC) dos parâmetros de transformação obtidos na etapa de registro dos pares de nuvens de pontos.

### 3. Experimentos e Análise dos Resultados

Para avaliar a potencialidade do método proposto neste trabalho foram conduzidos seis experimentos. A aquisição dos dados foi realizada em diferentes ambientes internos. O dispositivo Kinect usado neste trabalho foi programado para obter dados com uma frequência de 25 quadros por segundo. Para cada quadro é armazenado uma imagem RGB, uma imagem IR e um arquivo contendo os valores de paralaxe discretizados em 11 bits. Para cada experimento foi obtida uma trajetória formando uma volta completa apenas com o movimento circular das mãos

de um operador. O desenvolvimento deste trabalho está baseado no emprego de bibliotecas computacionais, tais como, OpenCV (*Open-Source Computer Vision*), Eigen e FLANN (*Fast Library for Approximate Neighbours*) e VTK (*Visualization Tool Kit*). Os algoritmos foram implementados em linguagem C++ ambiente CodeBlocks para Linux.

### 3.1. EXPERIMENTOS

Os parâmetros de orientação interior dos sensores RGB e IR foram obtidos por um processo de calibração de câmeras usando o método de Zhang. Para a calibração dos sensores foram tomadas, simultaneamente, 16 imagens RGB e IR sobre um tabuleiro de xadrez padrão. Os valores de  $Z_0$ ,  $b$ ,  $m$ ,  $n$  e os valores de *boresight misalignment* ( $\Delta\omega$ ,  $\Delta\varphi$ ,  $\Delta\kappa$ ) e *lever arm* ( $\Delta T_x$ ,  $\Delta T_y$ ,  $\Delta T_z$ ) entre a origem dos sistemas referenciais dos sensores foram obtidos usando o método proposto por Khoshelham e Elberink (2012). A Figura 4 mostra o conjunto de imagens RGB e IR obtidas pelo dispositivo Kinect.

Na Figura 4 as imagens RGB são aquelas com alto contraste e as imagens IR representam as imagens de baixo contraste. Os parâmetros de orientação interior de ambos os sensores e os valores de desalinhamento angular e o deslocamento entre a origem linear entre os sistemas referenciais dos sensores podem ser encontrados na Tabela 1.

**Tabela 1:** Parâmetros de orientação interior dos sensores RGB e IR, seus desvios padrão e os valores de desalinhamento angular e deslocamento das origens.

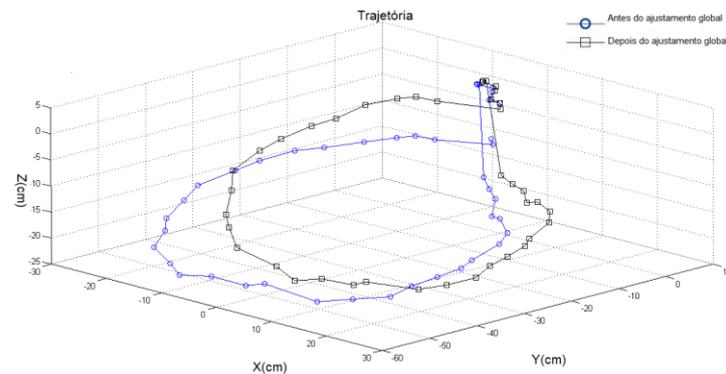
Parâmetros	RGB	IR			
$f$	$4,09 \pm 0,00036 \text{ mm}$	$4,55 \pm 0,00036 \text{ mm}$			
$x_0$	$-0,0351 \pm 0,00008 \text{ mm}$	$-0,0273 \pm 0,00008 \text{ mm}$			
$y_0$	$-0,1521 \pm 0,00008 \text{ mm}$	$-0,0897 \pm 0,00008 \text{ mm}$			
$k_1$	$2,36 \times 10^{-1} \pm 8,0 \times 10^{-6} \text{ mm}^2$	$-7,76 \times 10^{-2} \pm 8,5 \times 10^{-6} \text{ mm}^2$			
$k_2$	$-6,85 \times 10^{-1} \pm 7,5 \times 10^{-7} \text{ mm}^2$	$3,37 \times 10^{-1} \pm 8,0 \times 10^{-7} \text{ mm}^2$			
$k_3$	$6,44 \times 10^{-3} \pm 8,0 \times 10^{-6} \text{ mm}^2$	$-4,68 \times 10^{-1} \pm 8,0 \times 10^{-7} \text{ mm}^2$			
$p_1$	$1,24 \times 10^{-3} \pm 8,0 \times 10^{-6} \text{ mm}^2$	$-2,38 \times 10^{-3} \pm 8,0 \times 10^{-7} \text{ mm}^2$			
$p_2$	$2,82 \times 10^{-3} \pm 8,0 \times 10^{-6} \text{ mm}^2$	$1,37 \times 10^{-3} \pm 8,0 \times 10^{-7} \text{ mm}^2$			
$m$	$n$				
	$-3,07 \times 10^{-5}$	$0,0334$			
$\Delta\omega$	$\Delta\varphi$	$\Delta\kappa$	$\Delta T_x$	$\Delta T_y$	$\Delta T_z$
$0,0240^\circ$	$-0,0001^\circ$	$-0,0011^\circ$	$25,18 \text{ mm}$	$-4,1 \times 10^{-3} \text{ mm}$	$-9,3 \times 10^{-3} \text{ mm}$

A Tabela 2 mostra as características de cada experimento realizado para avaliar o método proposto.

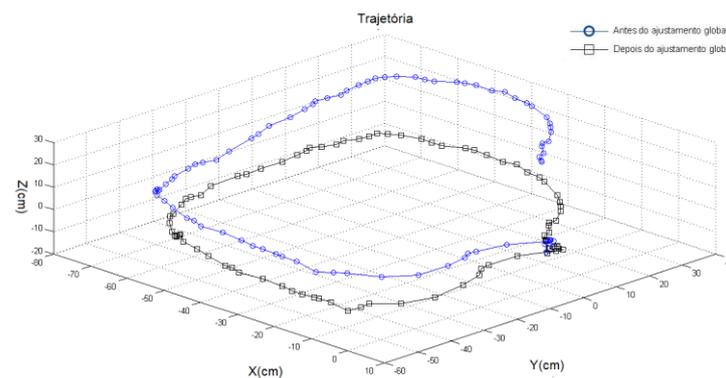
**Tabela 2:** Características dos experimentos realizados.

Experimento	Textura do ambiente	Número de frames	Ponderação das observações
Cenário A	Alta	45	XYZ
Cenário B	Alta	83	XYZ
Cenário C	Alta	95	XYZ
Cenário D	Baixa	78	Apenas em Z
Cenário E	Baixa	205	Apenas em Z
Cenário F	Baixa	124	Apenas em Z

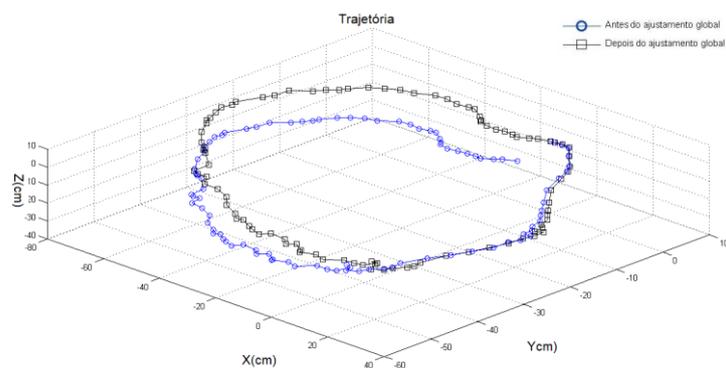
A Figura 5 mostra a trajetória do sensor antes (azul) e depois da análise de consistência global (preto) obtidas com o método proposto para os Cenários A-C.



(a)



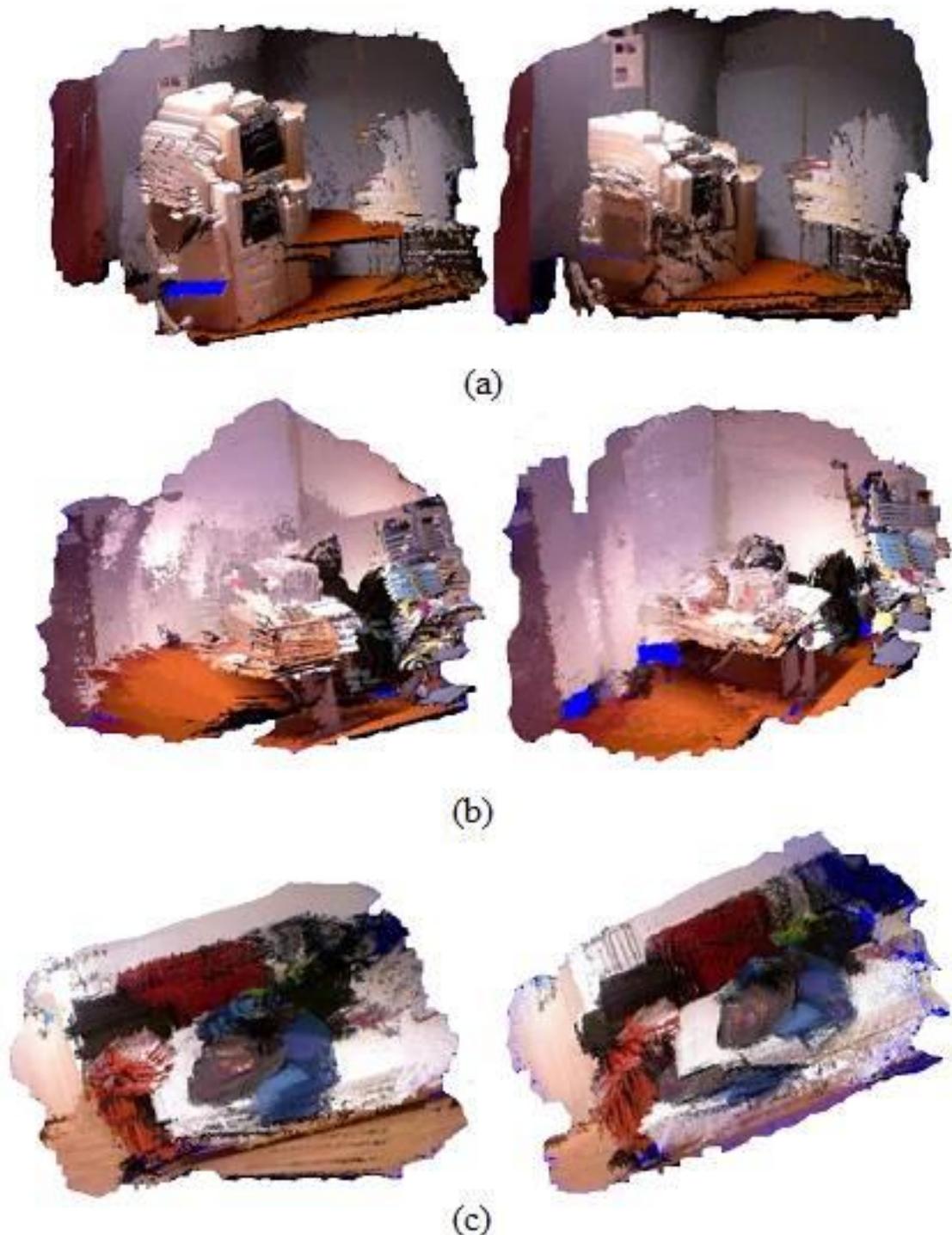
(b)



(c)

**Figura 5:** Trajetória do sensor antes e depois da análise da consistência global. (a) Cenário A; (b) Cenário B; (c) Cenário C.

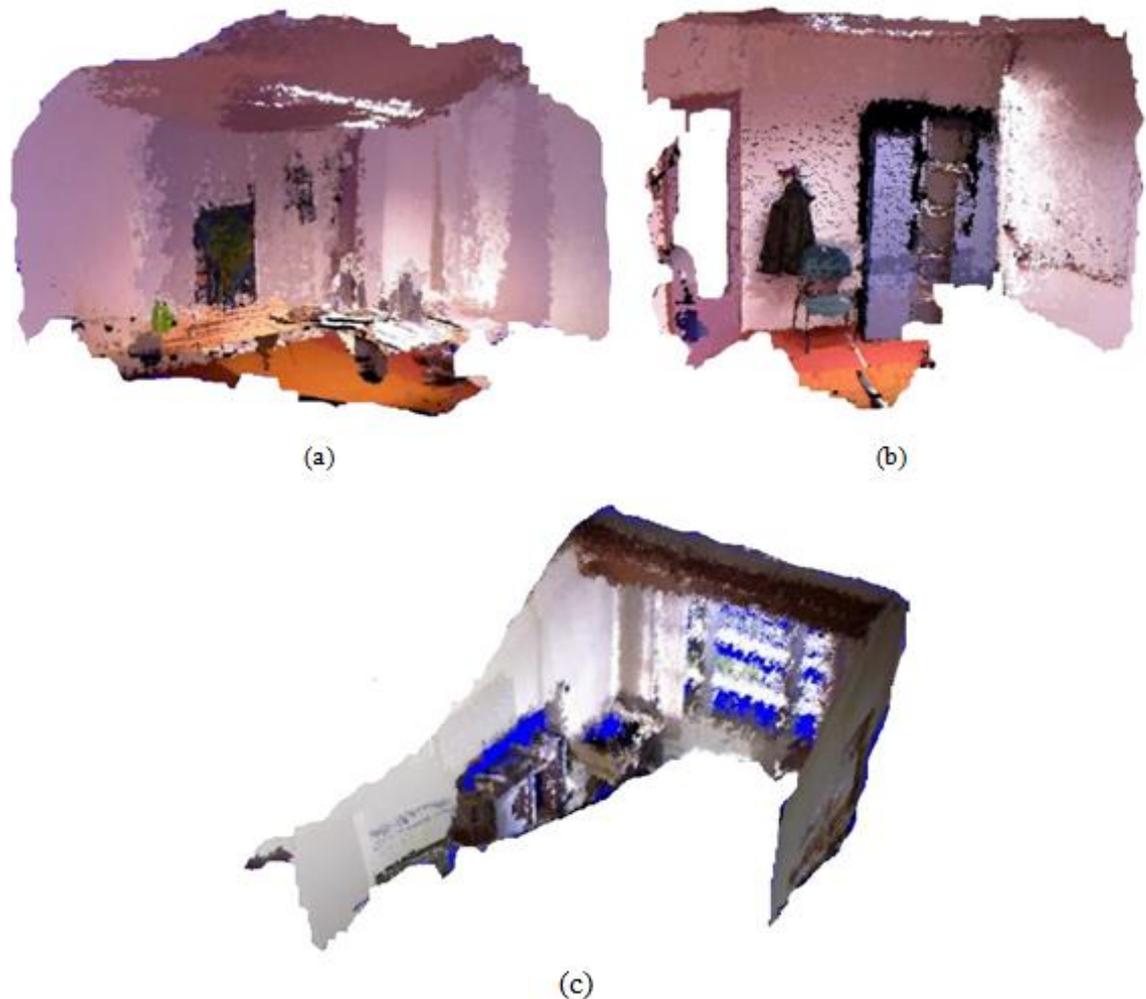
A Figura 6 mostra o ambiente interno tridimensionalmente mapeado antes e depois da análise de consistência global dos Cenários A-C.



**Figura 6:** Nuvem de pontos 3D de ambientes internos mapeados. (a) Cenário A; (b) Cenário B; (c) Cenário C.

Como pode ser visto na Tabela 2 os ambientes internos dos Cenários A-C são compostos por regiões de alta textura com diversos tipos de objetos presentes na cena, enquanto os Cenários D-F são formados por ambientes internos com regiões pouco texturizadas. Também é importante notar que para os experimentos correspondentes aos Cenários A-C foram incluídos pesos das

observações em  $X$ ,  $Y$  e  $Z$ , enquanto para os demais experimentos (Cenários D-F), apenas foram incluídos os pesos das observações em  $Z$ , sendo adotado o valor unitário para a ponderação das demais coordenadas. A Figura 7 mostra o ambiente interno tridimensionalmente mapeado depois da análise de consistência global dos Cenários D-F.



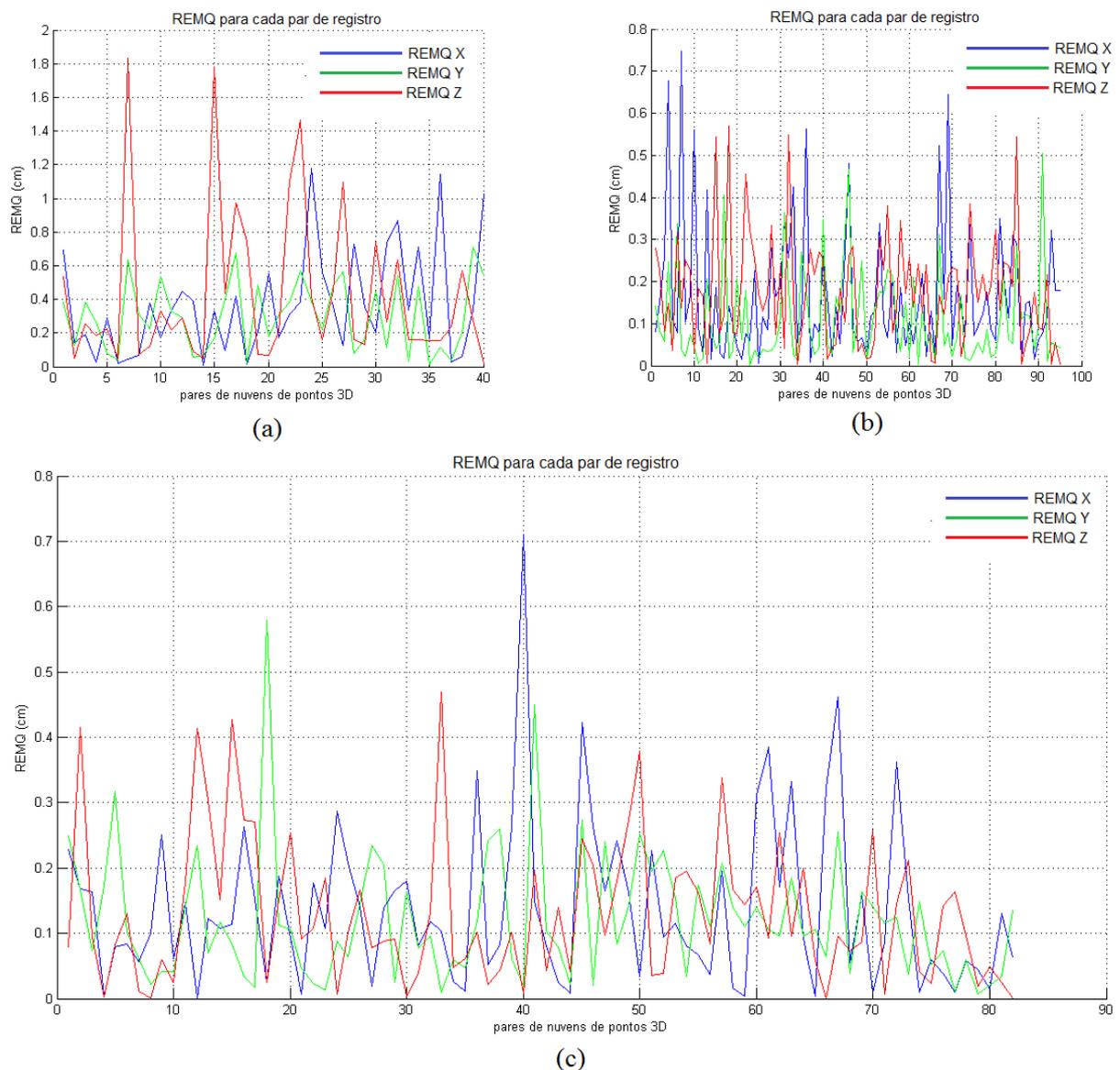
**Figura 7:** Nuvem de pontos 3D dos ambientes internos. (a) Cenário D; (b) Cenário E; (c) Cenário F.

A seguir são apresentadas as discussões dos resultados obtidos com o método proposto.

## 4.2. Análise dos Resultados

Para analisar quantitativamente os resultados obtidos foi calculada a REMQ das discrepâncias entre os pontos correspondentes encontrados na última iteração do algoritmo ICP (neste caso, os pontos são considerados correspondentes). A REMQ das discrepâncias de cada experimento é exibida na Figura 8. Os resultados mostram que o menor e o maior REMQ das coordenadas  $X$ ,  $Y$  e  $Z$  calculado no experimento relativo ao Cenário A (Figura 8a) foram 0,01 e 1,19 cm para a coordenada em  $X$ , para a coordenada em  $Y$  foram obtidos 0,02 e 0,7 cm, enquanto para a coordenada em  $Z$  foram encontrados valores de 0,1 e 1,81 cm, respectivamente. Em relação ao

Cenário C (Figura 8b) foram obtidos valores da REMQ de 0,01 e 0,75 cm para a coordenada em  $X$ , para a coordenada em  $Y$  foram determinados valores iguais a 0,05 e 0,5 cm, enquanto para a coordenada em  $Z$  foram encontrados valores de 0,2 e 0,57 cm, respectivamente. Por fim, para o Cenário B (Figura 8c) foram obtidos valores da REMQ de 0,02 e 0,71 cm para a coordenada em  $X$ , para a coordenada em  $Y$  foram determinados valores iguais a 0,1 e 0,59 cm, enquanto para a coordenada em  $Z$  foram encontrados valores de 0,05 e 0,47 cm, respectivamente. Para todos os experimentos o valor da média das discrepâncias ficou próximo de zero, mostrando que nenhuma tendência significativa influenciou o modelo matemático. O algoritmo se comportou de forma adequada, uma vez que era esperada uma redução nos valores de REMQ conforme a convergência do método. No entanto, como mostra a Figura 8a, em alguns casos percebem-se valores espúrios da REMQ para as coordenadas tridimensionais. Isto pode ser explicado pelo emprego dos valores de peso das observações oriundas das coordenadas  $X$  e  $Y$ .



**Figura 8:** REMQ calculada em cada registro de pares de nuvens de pontos. (a) Cenário A; (b) Cenário C; (c) Cenário B.

Na Figura 5 a trajetória do sensor, para todos os experimentos, é obtida como o esperado. Antes da análise da consistência global foram encontrados valores da média das discrepâncias em torno de 10 cm, sendo obtidos valores médios em torno de 3 cm após a análise da consistência global, sendo que as imagens foram coletadas a uma distância aproximada de 3 metros. No entanto, como pode ser observado na Figura 6, o alinhamento obtido no mapeamento 3D dos ambientes relacionados aos Cenários A-C não são perfeitos. A Tabela 3 mostra os valores máximo e mínimo da média das discrepâncias para cada experimento (A-C). Vale lembrar que para os experimentos A-C foram incluídos valores dos pesos das observações em  $X$ ,  $Y$  e  $Z$ .

**Tabela 3:** Média das discrepâncias após a análise da consistência global.

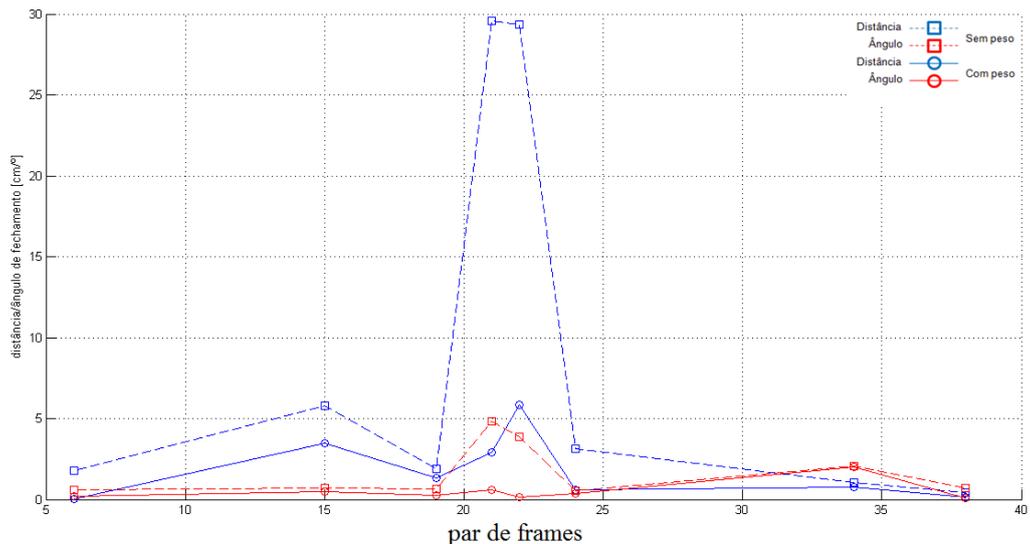
Experimento	Média das discrepâncias (cm)
Cenário A	Min: 0,51    Max: 3,84
Cenário B	Min: 0,21    Max: 1,60
Cenário C	Min: 0,11    Max: 3,00

Para testar a adequação do método proposto incluindo apenas valores de peso das observações em  $Z$ , foram executados os três últimos experimentos (Cenário D-F). Como pode ser observado na Figura 7, os ambientes mapeados são compostos por regiões pouco texturizadas (paredes e muros brancos). Os ambientes mapeados apresentam menores distorções em relação aos Cenários A-C. Isto se deve, principalmente, devido ao emprego dos valores de peso das observações apenas em  $Z$ , mostrando que a inclusão de pesos nas observações em  $X$  e  $Y$  influencia negativamente o resultado da etapa de registro dos pares de nuvens de pontos e, conseqüentemente, na análise de consistência global. A REMQ de cada um dos experimentos referentes aos Cenários D-F é apresentada na Tabela 4.

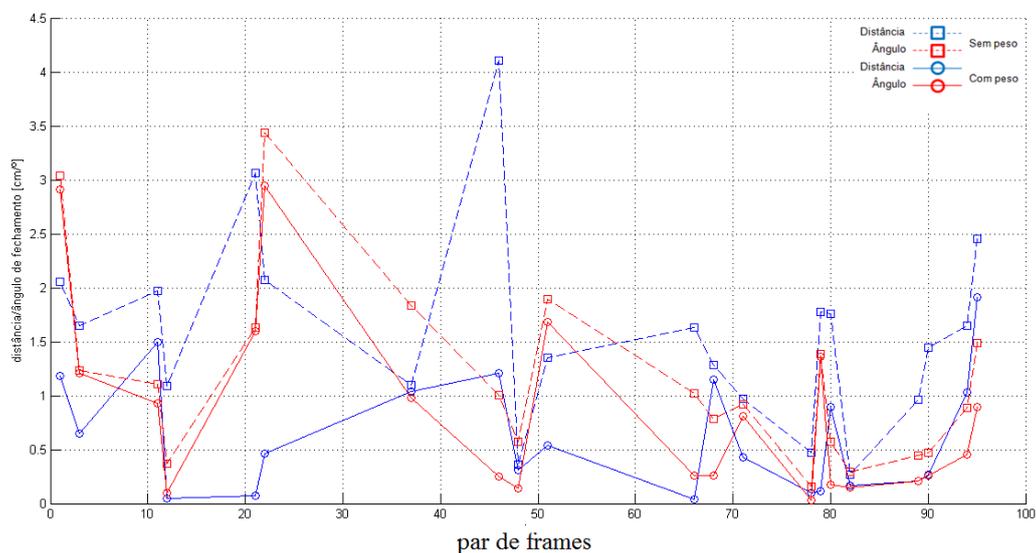
**Tabela 4:** Valores máximo e mínimo da REMQ.

Experimento	REMQ (cm)
Cenário D	Min: 0,19    Max: 2,50
Cenário E	Min: 0,35    Max: 3,52
Cenário F	Min: 1,21    Max: 2,91

A Figura 9 contém dois gráficos que apresentam as distâncias e ângulos de fechamento para sequências de pares de nuvens de pontos sorteadas aleatoriamente dos Cenários A e C após o registro dos pares com e sem pesos. As sequências foram ordenadas seguindo a sucessão dos pares. Essas distâncias e ângulos de fechamento representam as diferenças de translação e rotação (em módulo) de registros com e sem peso. Como pode ser observado, ambas as distâncias e os ângulos de fechamento são melhoradas como resultado da utilização de pesos.



(a)



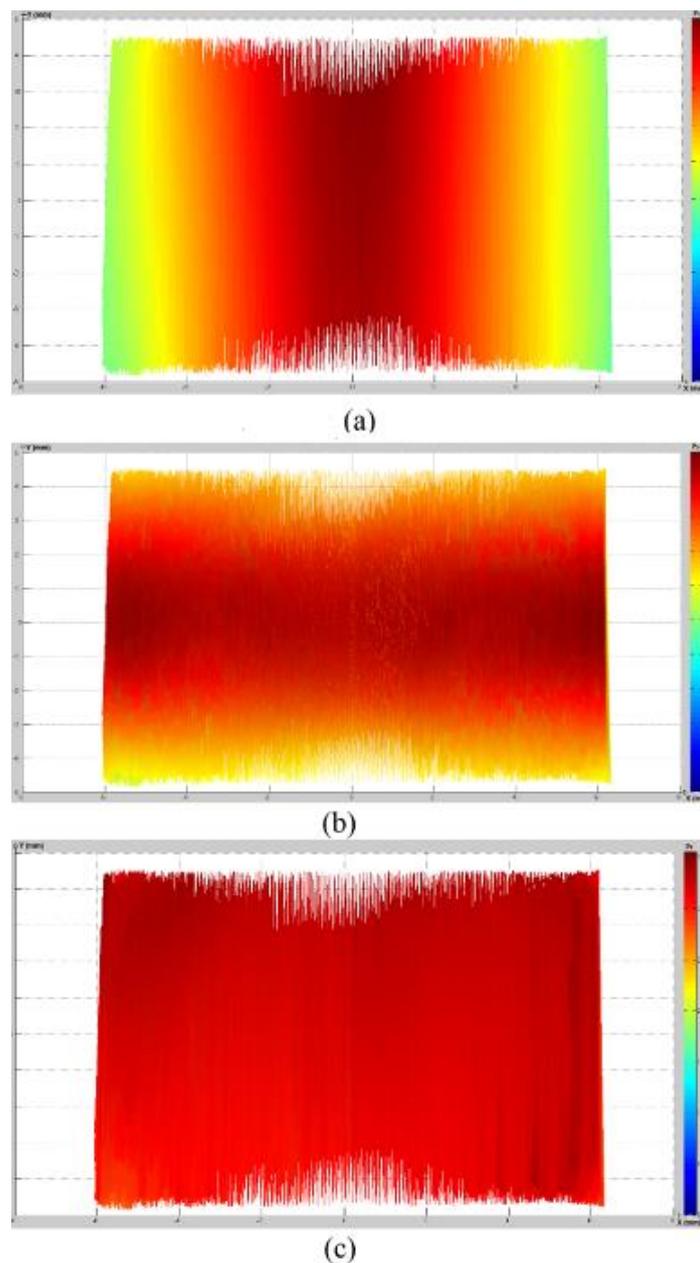
(b)

**Figura 9:** Distância (azul) e ângulo (vermelho) de fechamento para sequência de registros entre pares de nuvens de pontos com e sem peso. (a) Cenário A; (b) Cenário C;

Embora a distância mapeada seja curta, o método proposto trabalhou moderadamente em ambientes pouco texturizados e com inclusão dos valores de peso das observações apenas em  $Z$ , ao contrário do que foi obtido nos experimentos A-C. Além disso, há situações em que a falha do método proposto. Para algumas regiões, os recursos visuais são inexistentes em função de regiões especulares ou com materiais de alta absorção da energia eletromagnética no comprimento de onda do IR.

Como a precisão teórica das medidas em  $X$  e  $Y$  é calculada em função das coordenadas no espaço imagem, corrigidas dos erros sistemáticos, os valores obtidos são piores à medida que o ponto se afasta do centro geométrico da imagem. Com a finalidade de mostrar esse efeito foi feito um experimento usando como referência uma nuvem de pontos 3D obtida pelo sensor Kinect em uma parede plana homogênea. A Figura 10 mostra a distribuição das medidas de precisão das coordenadas  $X$ ,  $Y$  e  $Z$ .

Como é mostrado na Figura 10, se pode observar a influência dos pesos baseados no erro randômico teórico das coordenadas  $X$ ,  $Y$  e  $Z$ . Nas Figuras 10a e 10b, conforme o ponto se afasta do centro geométrico da nuvem de pontos 3D menor é a contribuição dos pesos das observações nas coordenadas em  $X$  e  $Y$  (do azul ao vermelho e variando de 0 a algo próximo de 10, quanto mais escura a cor maior a contribuição do peso). Esta redução na contribuição de pontos periféricos causa um efeito negativo no registro de pares de nuvens de pontos. Já a Figura 10c revela a contribuição homogênea dos pesos em  $Z$  para qualquer região da nuvem de pontos 3D (também variando do azul ao vermelho e de 0 a aproximadamente 3,5). É inevitável destacar que no caso dos pesos em  $Z$  quanto mais próximo o objeto do sensor, melhor sua contribuição no processo de ajustamento de observações.



**Figura 10:** Distribuição dos valores de peso para as observações. (a) Influência dos valores de pesos em  $X$ ; (b) Influência dos valores de pesos em  $Y$ ; (c) Influência dos valores de pesos em  $Z$ .

A seguir serão apresentadas as conclusões e recomendações para trabalhos futuros.

## 4. Conclusões e Recomendações para Trabalhos Futuros

Este trabalho apresentou um método adaptativo para registro e mapeamento de dados derivados de câmeras RGB-D em ambientes internos. Para investigar a potencialidade do método proposto foram realizados experimentos em ambientes internos com regiões pouco e moderadamente texturizadas. O método proposto utiliza apenas as informações de profundidade do sensor, sem se beneficiar de qualquer informação visual. As principais características do método proposto é a subdivisão da nuvem de pontos de referência em células, que, além de organizar o processo de estabelecimento de correspondências, reduz o custo computacional do algoritmo ICP, e, a inclusão dos valores de profundidade baseada na precisão teórica da medida na matriz dos pesos no algoritmo ICP que é uma estratégia ainda não explorada pela comunidade. O método usado neste trabalho para a calibração dos sensores RGB e IR, bem como do *boresight* e *lever-arm* entre os sensores, é o método de Zhang, que não inclui o polinômio de distorção descentrada. Esta simplificação pode inclusive não modelar a correlação entre alguns parâmetros de orientação existentes. No entanto, o foco principal deste trabalho não é a calibração e, por isso, resolvemos empregar o método de Zhang (2000) por ser totalmente automático e flexível para nossa aplicação. Em trabalhos futuros pretendemos empregar modelos mais rigorosos, inclusive implementar um modelo que considere no processo de calibração o dispositivo LASER do Kinect.

O processo de correspondência ponto-a-ponto é feito de forma rápida e organizada, uma vez que apenas os centroides das células das nuvens de pontos de referência e seus correspondentes nas nuvens de pontos atuais são empregados no processo de registro dos dados, além de ser empregado o método *k-d tree*. Duas restrições geométricas foram propostas para diminuir o número de falsos positivos, a distância mínima e a variação angular das linhas retas que ligam os pontos correspondentes adaptados ao valor da REMQ. O método também apresentou uma técnica adaptativa para o critério de convergência baseado no inverso da distância do REMQ. Esta tarefa também é importante para reduzir o tempo de processamento computacional. A detecção e remoção de *outliers* é uma tarefa complexa e exige maiores investigações, uma vez que eles afetam a estimativa dos parâmetros e, conseqüentemente, o cálculo do REMQ da média das discrepâncias. De forma geral, o algoritmo ICP se comporta adequadamente e se pode esperar uma melhora na acurácia relativa da trajetória do sensor, apenas quando se emprega valores de peso relativos às coordenadas em Z. Em Endres *et al.* (2012), em um ensaio com 687 frames, as acurácias relativas encontradas foram de 4,2 cm e 2,5° para a translação e rotação, respectivamente, enquanto que com o nosso método, o maior valor encontrado foi de 3,52 cm para o Cenário E. A inclusão dos pesos baseado na precisão teórica do valor de profundidade foi usada pela primeira vez no algoritmo ICP, quando comparado ao estado da arte. No entanto, maior quantidade de experimentos deve ser realizada para afirmar a robustez do método proposto para fins de registro de nuvem de pontos 3D. O método também não requer aproximações iniciais, visto a alta taxa de aquisição das informações (em torno de 25fps).

Como trabalhos futuros serão conduzidas investigações sobre abordagens ICP ponto-a-plano no processo de registro dos pares de nuvens de pontos 3D, a inclusão de novas restrições geométricas no processo de estabelecimento de correspondências, uma solução global para o método e aplicar técnicas de filtragem robustas que reduzem o número de *outliers* presentes na nuvem de pontos 3D.

## AGRADECIMENTOS

Os autores externam seus agradecimentos a CAPES e ao CNPq (processo no. 302644/2013-0) pelo apoio concedido ao projeto de pesquisa.

## REFERÊNCIAS BIBLIOGRÁFICAS

- Al-Manasir, K, and Fraser, C. S. "Registration of Terrestrial Laser Scanner Data Using Imagery." *The Photogrammetric Record* 21 (2006): 255-268.
- Bachrach, A.; Prentice, S.; He, R.; Henry, P., Huang, S.A; Krainin, M.; Maturana, D.; Fox, D. and Roy, N. "Estimation, Planning, and Mapping for Autonomous Flight Using an RGB-D Camera in GPS-denied Environments." *The International Journal of Robotics Research* 31 (2012): 1320-1343.
- Barnea, S. and Filin, S. "Keypoint Based Autonomous Registration of Terrestrial Laser Point-clouds." *ISPRS Journal of Photogrammetry and Remote Sensing* 63 (2008): 19-35.
- Bendels, G. H.; Degener, P.; Wahl, R.; Körtgen, M. and Klein, R. "Image-based Registration of 3D-range Data Using Feature Surface Elements." *Proceedings of the 5th International Conference on Virtual Reality, Archaeology and Intelligent Cultural Heritage, Oudenaarde, Belgium, Eurographics Association, 2004, 115-24.*
- Besl, P.J., and McKay, H. D. "A Method for Registration of 3-D Shapes." *IEEE Transactions on Pattern Analysis and Machine Intelligence* *IEEE Trans. Pattern Anal. Machine Intell.* 14 (1992): 239-256.
- Blais, G., and Levine M. D. "Registering Multiview Range Data to Create 3D Computer Objects." *IEEE Transactions on Pattern Analysis and Machine Intelligence* *IEEE Trans. Pattern Anal. Machine Intell.*, 1995, 820-824.
- Brown, D. C. "Close-range Camera Calibration." *Photogrammetric Engineering* 37 (1971): 855-866.
- Chen, Y. and Medioni, G. "Object Modelling by Registration of Multiple Range Images." *Image and Vision Computing* 10 (1992): 145-155.
- Chow, J.; Lichti, D.; Hol, J. Giovanni Bellusci, and Henk Luinge. "IMU and Multiple RGB-D Camera Fusion for Assisting Indoor Stop-and-Go 3D Terrestrial Laser Scanning." *Robotics* 3 (2014): 247-280.
- Dryanovski, I.; Jaramillo, C. and Xiao J. "Incremental Registration of RGB-D Images." 2012 *IEEE International Conference on Robotics and Automation*, 2012, 1685-1690.
- Du, H.; Henry, P.; Ren, X.; Cheng, M. Dan B. Goldman, Steven M. Seitz, and Dieter Fox. "Interactive 3D Modeling of Indoor Environments with a Consumer Depth Camera." *Proceedings of the 13th International Conference on Ubiquitous Computing - UbiComp '11*, 2011.
- Eggert, D. W.; Fitzgibbon, A. W. and Fisher, R.B. "Simultaneous Registration of Multiple Range Views for Use in Reverse Engineering of CAD Models." *Computer Vision and Image Understanding* 69 (1998): 253-272.

- Ellekilde, L. P.; Huang, S.; Valls, J. M. and Dissanayake, G. "Dense 3D Map Construction for Indoor Search and Rescue." *Journal of Field Robotics* J. Field Robotics 24 (2007): 71-89.
- Endres, F.; Hess, J.; Engelhard, N.; Sturm, J.; Cremers, D. and Burgard, W. "An Evaluation of the RGB-D SLAM System." 2012 IEEE International Conference on Robotics and Automation, 2012, 1691-1696.
- Engelhard, N.; Endres, F.; Hess, J.; Sturm, J. and Burgard, W. "Realtime 3D Visual SLAM with a Hand-held RGB-D Camera." *RGB-D Workshop on 3D Perception in Robotics at the European Robotics Forum*, Vasteras, Sweden, 2011.
- Fischler, M. A. and Bolles, R. C. "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography." *Communications of the ACM Commun.* ACM 24, no. 6 (1981): 381-395.
- Freidman, J. H.; Bentley, J. L. and Finkel, R. A. "An Algorithm for Finding Best Matches in Logarithmic Expected Time." *ACM Transactions on Mathematical Software* ACM Trans. Math. Softw. TOMS 3 (1977): 209-226.
- Gemael, C. *Introdução Ao Ajustamento De Observações: Aplicações Geodésicas*. Curitiba: Ed. UFPR, 1994.
- Godin, G.; Rioux, M. and Baribeau, R. "Three-dimensional Registration Using Range and Intensity Information." *Videometrics III*, 1994, 279-290.
- Golub, G. H. and Van Loan, C. F. *Matrix Computation*. 3rd ed. Maryland, Baltimore: John Hopkins University Press, 1996.
- Grant, D.; Bethel, J. and Crawford, M. "Point-to-plane Registration of Terrestrial Laser Scans." *ISPRS Journal of Photogrammetry and Remote Sensing* 72 (2012): 16-26.
- Greenspan, M. and Yurick, M. "Approximate K-D Tree Search for Efficient ICP." *Fourth International Conference on 3-D Digital Imaging and Modeling*, 2003. 3DIM 2003. Proceedings. 3 (2003): 247-280.
- Hartley, R. and Zisserman, A. *Multiple View Geometry in Computer Vision*. 2nd ed. Cambridge, UK: Cambridge University Press, 2003.
- Henry, P.; Krainin, M.; Herbst, E.; Ren, X. and Fox, D. "RGB-D Mapping: Using Kinect-style Depth Cameras for Dense 3D Modeling of Indoor Environments." *The International Journal of Robotics Research* 31 (2012): 647-663.
- Henry, P.; Krainin, M.; Herbst, E.; Ren, X. and Fox, D. "RGB-D Mapping: Using Depth Cameras for Dense 3D Modeling of Indoor Environments." *Experimental Robotics Springer Tracts in Advanced Robotics*, 2010, 477-491.
- Huhle, B.; Jenke, P. and Strasser, W. "On-the-fly Scene Acquisition with a Handy Multi-sensor System." *IJISTA International Journal of Intelligent Systems Technologies and Applications* 5 (2008): 255-263.
- Khoshelham, K.; Santos, D. R. and Vosselman, G. "Generation and Weighting of 3D Point Correspondences for Improved Registration of RGB-D Data." *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences II, no. 5/W2 (2013): 127-132.
- Khoshelham, K. and Elberink, S. O. "Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications." *Sensors* 12 (2012): 1437-1454.

- Lowe, D. G. "Distinctive Image Features from Scale-Invariant Keypoints." *International Journal of Computer Vision* 60 (2004): 91-110.
- Lu, F. and Milios, E. "Consistent Range Scan Alignment for Environment Mapping." *Autonomous Robots* 4, no. 4 (1997): 333-349.
- Magee, M. J.; Boyter, B. A.; Chien, C. H. and Aggarwal, J. K. "Experiments in Intensity Guided Range Sensing Recognition of Three-Dimensional Objects." *IEEE Transactions on Pattern Analysis and Machine Intelligence IEEE Trans. Pattern Anal. Mach. Intell. PAMI-7* (1985): 629-637.
- May, S.; Droschel, D.; Holz, D.; Fuchs, S.; Malis, E.; Nüchter, A. and Hertzberg, J. "Three-dimensional Mapping with Time-of-flight Cameras." *Journal of Field Robotics J. Field Robotics* 26 (2009): 934-965.
- Moré, J. J. "The Levenberg-Marquardt Algorithm: Implementation and Theory." *Lecture Notes in Mathematics Numerical Analysis*, 1997, 105-116.
- Mostofi, N., A. Moussa, M. Elhabiby, and N. El-Sheimy. "RGB-D Indoor Plane-based 3D-Modeling Using Autonomous Robot." *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XL-1* (2014): 301-308.
- Nuchter, A.; Lingemann, K. and Hertzberg, J. "Cached K-d Tree Search for ICP Algorithms." *Sixth International Conference on 3-D Digital Imaging and Modeling (3DIM 2007)*, 2007, 419-426.
- Roth, G. "Registering Two Overlapping Range Images." *Second International Conference on 3-D Digital Imaging and Modeling (Cat. No.PR00062)*, 1999, 191-200.
- Sande, C. D.; Soudarissanane, S. and Khoshelham, K. "Assessment of Relative Accuracy of AHN-2 Laser Scanning Data Using Planar Features." *Sensors* 10, no. 9 (2010): 8198-8214.
- Segal, A. V.; Haehnel, D. and Thrun, S. "Generalized-ICP." In *Proc. of Robotics: Science and Systems (RSS)* 25 (2009).
- Stamos, I., and Leordeanu, M. "Automated Feature-based Range Registration of Urban Scenes of Large Scale." *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2003. *Proceedings. 2* (2003): P. II-561.
- Steinbrucker, F.; Sturm, J. and Cremers, D. "Real-time Visual Odometry from Dense RGB-D Images." *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, 2011, 719-722.
- Weik, S. "Registration of 3-D Partial Surface Models Using Luminance and Depth Information." *Proceedings. International Conference on Recent Advances in 3-D Digital Imaging and Modeling (Cat. No.97TB100134)*, 1997, 93-100.
- Zhang, Z. "Iterative Point Matching for Registration of Free-form Curves and Surfaces." *Int J Comput Vision International Journal of Computer Vision* 13 (1994): 119-152.
- Zhang, Z. "Flexible Camera Calibration by Viewing a Plane from Unknown Orientations." *Proceedings of the Seventh IEEE International Conference on Computer Vision*, 1999, 666-673.

Recebido em abril de 2015.

Aceito em setembro de 2015.