

Revista Brasileira de Cartografia (2015) N<sup>o</sup> 67/7: 1523-1540  
Sociedade Brasileira de Cartografia, Geodésia, Fotogrametria e Sensoriamento Remoto  
ISSN: 1808-0936

## **ANÁLISE DA PONDERAÇÃO DO ERRO RANDÔMICO TEÓRICO DAS COORDENADAS 3D NO REGISTRO DE DADOS RGB-D**

*Analysis of the Weighting 3D Coordinates Based on Theoretical  
Random Errors for RGB-D Data Registration*

**Elizeu Martins de Oliveira Junior<sup>1</sup>, Daniel Rodrigues dos Santos<sup>1</sup>,  
Quintino Dalmolin<sup>1</sup> & Kourosh Khoshelham<sup>2</sup>**

**<sup>1</sup>Universidade Federal do Paraná – UFPR**

**Departamento de Geomática / Programa de Pós Graduação em Ciências Geodésicas**  
Centro Politécnico - Jardim das Américas - Caixa Postal 19001 - CEP 81531-990 - Curitiba PR – Brasil  
elizeuoliveirajunior@gmail.com, {danielsantos; qdalmolin}@ufpr.br

**<sup>2</sup>University of Melbourne**

**Department of Infrastructure Engineering**  
Block B, Melbourne School of Engineering, 403/1-100 Grattan St, Parkville VIC 3010, Melbourne-Australia  
k.khoshelham@unimeb.edu.au

*Recebido em 21 de Setembro, 2015/ Aceito em 1 de Dezembro, 2015*  
*Received on September 21, 2015/ Accepted on December 1, 2015*

### **RESUMO**

Neste trabalho é apresentado um estudo que mostra a influência da precisão teórica das coordenadas 3D no processo de registro de dados RGB-D. Basicamente, três pontos principais são aqui discutidos: o registro dos pares de nuvens de pontos 3D adquiridas sucessivamente pelo sensor de imageamento; a ponderação das coordenadas 3D de pontos individuais; e a análise de consistência global. O registro dos pares de nuvens de pontos é feito em duas etapas. Na primeira etapa, os parâmetros de transformação são obtidos através de um conjunto de pontos, detectados pelo algoritmo SIFT, empregados no modelo de corpo rígido 3D, juntamente com o Método dos Mínimos Quadrados. A segunda etapa consiste em refinar os parâmetros de transformação obtidos na etapa anterior empregando o algoritmo ICP baseado na abordagem ponto-a-ponto. A inclusão dos valores de pesos é feito baseado na propagação de covariâncias das equações de paralaxe. A análise de consistência global é empregada para refinar o conjunto de parâmetros de translação simultaneamente. Experimentos em cenas com distância médias do sensor/alvo de 2 metros foram conduzidos para avaliar a eficiência do método proposto e os resultados indicam a potencialidade do método com o valor médio das discrepâncias abaixo dos 2 centímetros.

**Palavras chaves:** RGB-D, Registro de Nuvem de Pontos 3D, Precisão Teórica de Coordenadas 3D, SIFT, ICP, Análise de Consistência Global.

### **ABSTRACT**

In this paper a weighting scheme based on theoretical random errors of 3D coordinates for RGB-D data registration is introduced. Basically, three main points involved in the method are here discussed: the pairwise registration; the weighting of 3D coordinates; and a global adjustment. The registration of point clouds pairs is done in two steps. In

the first step, the transformation parameters are obtained by a set of points, detected by the SIFT algorithm employed in the 3D rigid body model, beside the Least Squares Method. The second step consists in refine the transformation parameters obtained in the previous step using the ICP algorithm based on point-to-point approach. The inclusion of weight values is made based on the disparity propagation covariance equations. The global consistency analysis is used to refine the set of translation parameters simultaneously. Experiments in scenes with mean distance from the sensor/target of 4 meters were conducted to evaluate the effectiveness of the proposed method and the results indicate the potential of the method to the average value of the discrepancies below 4 centimeters.

**Keywords:** RGB-D, Pairwise Registration, Theoretical Random Errors of 3D Coordinates, SIFT, ICP, Global Adjustment.

## 1. INTRODUÇÃO

De acordo com o relatório de 2010 da Organização Mundial da Saúde (OMS) metade da população mundial reside em áreas urbanizadas. Embora, desde aproximadamente 2400 a.C., o mapa tem sido empregado como uma ferramenta crucial para auxiliar na localização e na navegação, apenas no final do século XX foram desenvolvidos os primeiros Sistemas de Mapeamento 3D (Tridimensionais) de Ambientes Internos (SMAI-3D). Um sistema como este possui diversas vantagens e pode ser empregado em inúmeras aplicações, tais como, em problemas de localização e mapeamento simultâneo (SLAM – *Simultaneous Localization And Mapping*); mapeamento de corredores, de salas, de túneis e de cavernas; modelagem e reconstrução 3D de superfícies; navegação, vigilância e aplicações forenses; realidade aumentada; estabelecimento de rotas de fuga em edifícios em caso de incêndios ou outros incidentes; entre outras.

Usualmente, os sistemas de mapeamento 3D desenvolvidos para ambientes internos são baseados em um conjunto de sensores de imageamento ativo (*LASER - Light Amplification by Stimulated Emission of Radiation SCANNING* e câmeras de distância 3D) e passivo (estéreo câmeras). Porém, são sensores de alto e médio custo, respectivamente, de baixa flexibilidade, relativamente pesados e exigem integração dos sensores. Recentemente, câmeras RGB-D (*Red, Blue e Gree + Depth - Kinect*) tem sido amplamente empregadas em problemas de mapeamento 3D, uma vez que é um dispositivo de baixo custo, leve, flexível, de fácil manuseio e fornece uma sinergia natural em relação à integração de dados RGB e de profundidade. Basicamente, esses dispositivos propiciam uma nuvem de pontos 3D com precisão de 3 cm a cada 3m e são capazes de fornecer mais de 9,2 milhões

de pontos por segundo. Apesar de o Kinect ter sido originalmente construído para finalidades de entretenimento e jogos computacionais, suas características e facilidades de aquisição de dados, bem como o tipo de informação fornecida, tem chamado a atenção de pesquisadores em diversas áreas do conhecimento, como por exemplo, na Fotogrametria, Visão Computacional e Robótica.

Basicamente, quatro pontos principais devem ser discutidos para o desenvolvimento de um SMAI-3D: a calibração dos sensores de imageamento; a extração de feições e o estabelecimento automático de suas correspondências; o registro dos pares de nuvens de pontos 3D adquiridos sucessivamente pelo sensor de imageamento; e, o refinamento dos parâmetros de transformação empregando uma análise de consistência global.

Em Khoshelham e Elberink (2012), Menna *et al.* (2011) e Chow *et al.* (2012), a resolução geométrica e a precisão dos pontos tridimensionais oferecidos pelo dispositivo são tratados como objeto de estudo. Esses trabalhos mostram a importância da calibração dos sensores embarcados no Kinect. Os dois primeiros são baseados na calibração de câmeras com uso de valores de paralaxe fornecidos pelo dispositivo Kinect. Já no trabalho de Chow *et al.* (2012) foram usados valores de profundidade pós-processados e somente a câmera RGB pode ser calibrada.

O desenvolvimento de um SMAI-3D rápido e preciso tem como maior desafio a etapa do registro de pares de nuvens de pontos 3D. Além disso, segundo Werghi (2009), é fundamental estabelecer uma representação que permita a correspondência e o alinhamento efetivo entre as feições nas duas imagens. No presente trabalho o método utilizado para o registro dos pares de nuvens de pontos é o popular ICP (*Iterative Closest Point*) (BESL

& MCKAY, 1992; ZHANG, 1994; MASUDA *et al.* 1996; RUSISKIEWICZ & LEVOY, 2001). O algoritmo baseado no método de Horn (1987), basicamente, encontra pontos pseudo-conjugados (pela maior proximidade dos pontos) 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 as distâncias entre as primitivas correspondentes. No entanto, o algoritmo ICP depende da entrada de valores iniciais aproximados, o processamento dos dados é computacionalmente lento, e, devido à baixa qualidade geométrica das imagens de intensidade, o algoritmo é altamente sensível a ruídos. Para contornar essa questão do estabelecimento das correspondências, as vantagens sinérgicas das informações visuais e de profundidade podem ser exploradas no intuito de estabelecê-las. Neste caso, pressupondo que os sensores embarcados no dispositivo Kinect estão devidamente calibrados utilizando métodos como o de Zhang (2000), por exemplo, pode-se empregar o algoritmo SIFT (*Scale Invariant Feature Transform*) proposto por Lowe (1999) que detecta e extrai pontos chave nas imagens RGB, e através de descritores visuais são estabelecidas as correspondências entre pontos pertencentes a ambas as imagens.

Segundo Wang *et al.* (2012) o SIFT é considerado o descritor mais robusto para o registro de imagens, entretanto, apresenta dificuldades diante imagens extremamente grandes e fornece pontos chave com uma distribuição irregular na imagem. Porém, no presente trabalho, as imagens são relativamente pequenas (640x480) e essa irregularidade na distribuição, na verdade, contribui para uma melhor estimativa dos parâmetros. Já em Amintoosi *et al.* (2012), os parâmetros de rotação e translação do registro das imagens são estimados por uma média das diferenças dos locais dos pares de pontos chave obtidos pelo SIFT. Para compensar essa generalização, módulos de ajustes como modificação e compensação da orientação e do deslocamento, entre outros, são aplicados. Em Zhang *et al.* (2012), uma variação do SIFT chamada Hams-SIFT é utilizada em um algoritmo de localização em ambientes internos baseado nas características visuais da

seqüência de imagens do piso. Essa variação acontece na extração de pontos chave em cada camada, obtidas pelas Diferenças de Gaussianas de primeira ordem, que é feita, nesse caso, com o detector de Hams. Desta forma, o resultado é um algoritmo que descreve a trajetória do movimento do alvo obtendo, portanto, uma estimativa da posição do sensor.

Uma vez que cada par de nuvens de pontos 3D é registrada em um sistema referencial local e os erros obtidos no processo de registro são exponencialmente propagados é necessário empregar um ajustamento global para refinar os parâmetros de transformação obtidos em cada registro. Lu e Milios (1997) propuseram o método de registro global conhecido como LUM, que considera cada par de nuvens de pontos 3D, bem como as relações espaciais entre elas, como variáveis randômicas derivadas dos parâmetros de transformação obtidos na etapa de registro, anteriormente executada. Assim, são refinados os parâmetros de transformação entre todos os pares de nuvens de pontos baseado num critério de probabilidade máxima para otimamente combinar todas as relações espaciais.

O primeiro SMAI-3D usando câmeras RGB-D foi proposto por Henry *et al.* (2012). Embora a abordagem proposta por Henry *et al.* (2012), conhecida como RGB-D ICP, melhore a eficiência do algoritmo em regiões com superfícies de baixa textura ou com padrões repetidos de informações e diminua o custo computacional do ICP, a associação entre os pontos correspondentes na RGB e os valores de profundidade é feita de forma aproximada e os autores também não admitiram, no processo de registro dos pares de nuvens de pontos, as incertezas provenientes das coordenadas tridimensionais, e, portanto, o resultado esperado é uma trajetória do sensor de baixa qualidade. Neste trabalho, visando solucionar os problemas apontados, além da realização da correta associação dos pontos na imagem RGB àqueles na nuvem de pontos, é proposta a inclusão de pesos no processo de registro dos pares de nuvens de pontos 3D para obtenção de uma trajetória do sensor mais robusta e de melhor acurácia relativa. Também, é proposto um modelo linear para o ajustamento simultâneo em bloco do conjunto de pares de nuvens de pontos capaz de refinar os parâmetros de translação do

registro baseado no método de LUM (LU & MILLIOS, 1997).

## 2. MÉTODO

A Figura 1 mostra um diagrama do método proposto neste trabalho. Basicamente, o método está dividido em 4 etapas. Primeiramente, é realizada a calibração dos sensores embarcados no Kinect (RGB e IR) e determinados os parâmetros de montagem dos sensores. Em seguida, feito o registro de cada par de nuvem de pontos. Esta etapa é dividida em duas partes: 1) Detecta e extrai pontos correspondentes no par de imagens RGB empregando o algoritmo SIFT. Remove falsas correspondências através da Matriz Fundamental e do RANSAC. Gera as nuvens de pontos 3D para cada par de imagens processadas. Normaliza as imagens IR em relação as imagens RGB e associa os pontos chave correspondentes com as nuvens de pontos 3D por meio de uma geometria epipolar. E finalmente, estima os parâmetros de transformação (Rotação e Translação) do registro grosseiro usando um modelo matemático rigoroso e o Método dos Mínimos Quadrados (MMQ); 2) Divide a imagem IR de referência em células igualmente espaçadas. Calcula o centroide de cada célula e o ponto 3D e estabelece as correspondências entre o centroide de cada uma das células com os vizinhos mais próximos na nuvem de pesquisa usando o método *k-d tree*. Aplica o ICP e refina os parâmetros de transformação do registro grosseiro. Finalmente, é aplicado o ajustamento por pares de nuvens de pontos independentes.



Fig. 1 - Diagrama do método proposto.

A seguir, será apresentada cada etapa do método proposto descrito no diagrama da Figura 1.

### 2.1 Calibração do Sistema RGB-D

O Kinect é um dispositivo 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. A etapa de calibração é feita na forma estéreo e consiste em determinar os parâmetros de orientação interior dos sensores RGB e IR (infravermelho), bem como os parâmetros do sistema estéreo (desalinhamento angular  $d\omega$ ,  $d\phi$ ,  $d\kappa$  e a distância entre as origens dos sistemas  $dx$ ,  $dy$ ,  $dz$ ). A determinação dos parâmetros de orientação interior é feita através do método de Zhang (2000). O método encontra a solução minimizando uma distância algébrica que não tem representação física. Para isto, primeiramente, é empregada uma estimativa por máxima verossimilhança. Assumindo que esteja isento de erros sistemáticos, cujas coordenadas possuem a mesma incerteza na medida, a matriz homográfica pode ser obtida da forma, a saber (MORE, 1997):

$$\min_H \sum_{i=1}^n \|P_i - P'_i\|^2 \quad (1)$$

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

A matriz homográfica  $H = A \cdot [r_1 \ r_2 \ t]$  é composta pelos termos,  $A$  uma matriz  $3 \times 3$  com os coeficientes:  $\tau$  de não-ortogonalidade entre os eixos da câmera,  $f$  a distância focal da câmera,  $S_x$  e  $S_y$  os fatores de escala em  $x$  e  $y$ , e, o segundo termo do lado direito da equação da matriz homográfica sendo  $r_1$  e  $r_2$  os elementos da primeira e segunda coluna da matriz de rotação, e,  $t$  o vetor de translação (HARTLEY & ZISSERMAN, 2003).

Segundo Zhang (2000) a determinação simultânea dos Parâmetros de Orientação Interior (POI) e Exterior (POE) é feita pelo método de Levenberg-Maquardt, como segue:

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

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  $i$ ,  $R_i$  a matriz das rotações e  $t_i$  o vetor das translações. Vale lembrar que  $p'$  é função das coordenadas corrigidas das distorções do sistema de lentes como apresentada na Equação (BROWN, 1971):

$$\delta x = \bar{x}(k_1 r^2 + k_2 r^4 + k_3 r^6) + P_1(r^2 + 2\bar{x}^2) + 2P_2 \bar{x} \bar{y}^2 \quad (3)$$

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

sendo,  $x'_c$  e  $y'_c$  as coordenadas do ponto no espaço-imagem corrigidas das distorções radiais simétricas e descentradas,  $\bar{x} = x' - x_0$ ,  $\bar{y} = y' - y_0$ ,  $r = \sqrt{\bar{x}^2 + \bar{y}^2}$ ,  $k_1, \dots, k_3$  os coeficientes de distorção radial das lentes e os termos da distorção descentrada. Maiores detalhes ver Zhang (2000).

## 2.2 Registro de Pares de Nuvens de Pontos 3D

Esta etapa é dividida em duas partes. Primeiramente, os parâmetros de transformação iniciais são determinados e, em seguida, é feito um refinamento dos dados.

### 2.2.1 Registro Grosseiro dos Pares de Nuvens de Pontos 3D

Este processo consiste em determinar os parâmetros de transformação iniciais para melhorar a eficiência computacional do ICP. O registro grosseiro é dividido em quatro tarefas, a saber:

- Extração e detecção de pontos chave empregando o algoritmo SIFT e a associação desses pontos na imagem de profundidade;
- Cálculo das coordenadas dos pontos 3D usando as equações de paralaxe;
- Ponderação das coordenadas de acordo com a precisão teórica das coordenadas 3D;
- Determinação dos parâmetros de transformação iniciais entre os pares de nuvem de pontos 3D com a transformação de corpo rígido solucionada por MMQ;

#### 2.2.1.1 Detecção de Pontos Chave com o SIFT

Primeiramente é empregado o SIFT para detectar pontos chave na imagem RGB e a geometria epipolar para encontrar as

correspondências em potencial. Basicamente, o algoritmo SIFT é dividido em duas partes: detecção de extremos locais; e geração de descritores locais. Dado um par de imagens RGB, o objetivo é detectar pontos chave e calcular seus descritores de forma que possam ser estabelecidas suas correspondências. O algoritmo SIFT trabalha da seguinte forma, a saber (LOWE, 2004):

1. Detecção de extremos no espaço escala através de cálculos de Diferenças de Gaussianas;
2. Definição da localização exata dos pontos extremos através do ajustamento de uma função quadrática 3D;
3. Atribuição de uma ou mais orientações para cada localização dos pontos chave baseadas em direções do gradiente local da imagem;
4. Composição dos descritores dos pontos chave através de histogramas de orientação de regiões ao redor de cada ponto chave;

Finalmente, a correspondência entre os pontos de interesse pode ser efetuada por intermédio do cálculo de alguma medida de similaridade que descreve a relação entre os descritores em uma dada vizinhança. Maiores detalhes o leitor deve consultar Lowe (1999).

#### 2.2.1.2 Remoção de *Outliers*

Após serem encontradas as correspondências entre os pontos chave, para cada par de imagens RGB é realizada a detecção e remoção de falsas correspondências. Esta tarefa é feita em dois passos, a saber:

1. Geração da geometria epipolar com o cálculo da Matriz Fundamental (MF) através da associação entre os pontos chave detectados com o RANSAC e o MMQ;
2. Remoção de falsas correspondências usando um limiar de distância perpendicular da reta epipolar e o ponto correspondente na imagem de pesquisa;

Para calcular a matriz de transformação é necessário empregar a Equação:

$$A \cdot cf = 0 \quad (5)$$

sendo,  $cf = (F_{11}, F_{12}, F_{13}, F_{21}, F_{22}, F_{23}, F_{31}, F_{32}, F_{33})$  os coeficientes da Matriz Fundamental  $F$ , e  $A$ , uma matriz de dimensão  $n \times 9$ , onde  $n$  representa o número de correspondências, a saber:

$$A = \begin{bmatrix} x'_{x1}x_{x1} & x'_{x1}y_{x1} & x'_{x1} & y'_{x1}x_{x1} & y'_{x1}y_{x1} & y'_{x1} & x_{x1} & y_{x1} & 1 \\ x'_{x2}x_{x2} & x'_{x2}y_{x2} & x'_{x2} & y'_{x2}x_{x2} & y'_{x2}y_{x2} & y'_{x2} & x_{x2} & y_{x2} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x'_{xn}x_{xn} & x'_{xn}y_{xn} & x'_{xn} & y'_{xn}x_{xn} & y'_{xn}y_{xn} & y'_{xn} & x_{xn} & y_{xn} & 1 \end{bmatrix} \quad (6)$$

Para solucionar  $cf$ , é aplicado o MMQ. Entretanto, como há muitas correspondências entre pontos chave, é necessário verificar quais devem ser usadas para gerar a melhor MF. Assim, são geradas diversas MFs, a partir de diferentes combinações, e então escolhida a combinação ótima. Como esse processo é caro computacionalmente, o RANSAC é empregado. Segundo Fischler e Bolles (1981), o RANSAC gera modelos a partir de um conjunto mínimo de dados. Basicamente, o algoritmo se divide em: 1), estabelecer uma hipótese (um conjunto mínimo de correspondências é aleatoriamente selecionado a partir de todo conjunto de dados disponível); 2) num teste (o algoritmo verifica quais elementos de todo conjunto de dados que são consistentes com o modelo estabelecido pelo conjunto mínimo). Os elementos considerados consistentes formarão um conjunto chamado de consenso. Esses dois processos ocorrem iterativamente, e, ao final deles, o conjunto que será adotado como modelo corresponderá ao conjunto mínimo que apresentar o maior conjunto consenso. A partir destes modelos e utilizando o critério de parada com 30% de pontos remanescentes (*inliers*), ou seja, na iteração que se alcança 30% ou menos do número total de correspondências como *inliers*, é obtido o modelo mais representativo entre os avaliados.

Sendo assim, dado um ponto chave  $x = [x_p, y_p, 1]$  na imagem de referência (esquerda) e  $x' = [x'_p, y'_p, 1]$  na imagem de pesquisa (direita), é possível eliminar falsas correspondências, desde que a distância  $\delta$  perpendicular entre a reta epipolar originada de um ponto na imagem de referência e o detectado na imagem de pesquisa seja menor que um determinado limiar  $\tau$  (ver Fig. 2). Caso contrário, essa correspondência será considerada uma *outlier*. Portanto, para verificar se o ponto é um *outlier* é realizado um teste da distância através da equação:

$$l'_x = F \cdot x \quad (7)$$

sendo que,  $F \cdot x$  define a linha epipolar  $l'_x$  do ponto na imagem de referência e “ $\cdot$ ” representa o operador de multiplicação (ZHANG, 1998).

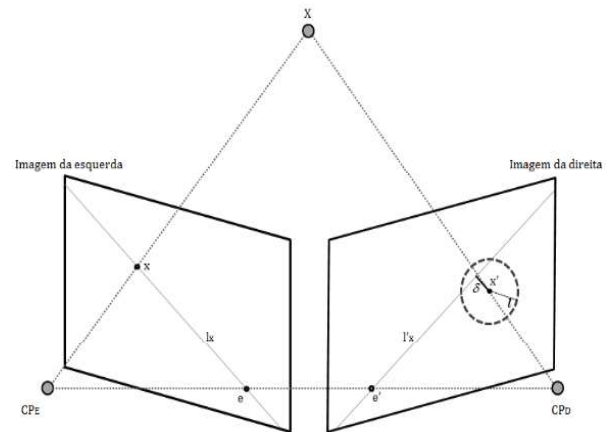


Fig. 2 - Processo de remoção de outliers com o RANSAC. Fonte: Adaptado de Zhang (1998).

Na Figura 2, é um ponto no espaço objeto, e são os epipolos,  $CP_E$  e  $CP_D$  os respectivos centros perspectivos das imagens de referência (esquerda) e de pesquisa (direita). Outro critério que será usado neste trabalho para a remoção de falsas correspondências é o ângulo formado (no plano das abcissas) pela linha imaginária que liga os pontos correspondentes nos pares de nuvens e outra que conecta os dois Centros Perspectivos, assim, se este ângulo for maior que 45 graus, esta correspondência deverá ser descartada.

Se um conjunto de pontos chave com correspondências apropriadamente estabelecidas é obtido deve-se associar esses pontos com seus correspondentes na nuvem de pontos 3D gerada com as equações de paralaxe.

### 2.2.1.3 Cálculo das Coordenadas 3D

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 pela câmera. Uma fonte de luz emite um feixe de luz que é dividido em múltiplos feixes criando um padrão aleatório (Figura 3), 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 (<http://kinectando.com.br>).



Fig. 3 - Padrão de luz estruturada gerado pelo Kinect.

A Figura 4 ilustra a geometria do sistema, o deslocamento dos padrões projetados bem como a medida da paralaxe observada. 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,  $o$  o ponto no plano de referência correspondente ao ponto no espaço-objeto,  $D$  o deslocamento do ponto  $k$  no espaço-objeto,  $Z_k$  a distância do sensor IR ao plano de referência, 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).

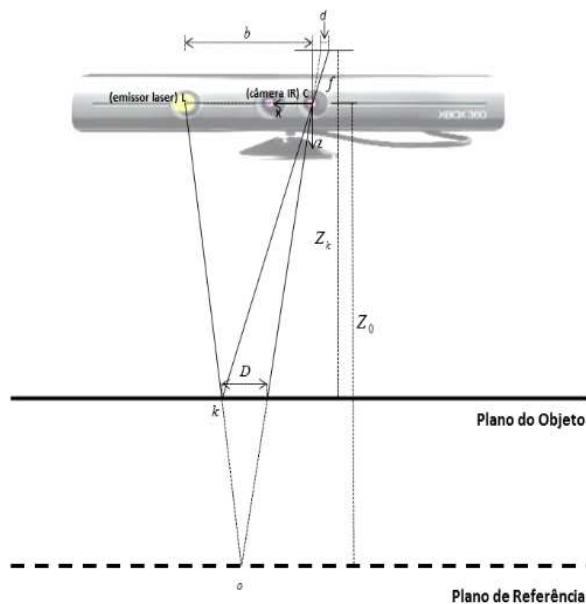


Fig. 4 – Relação entre profundidade relativa e paralaxe medido.

Assumindo que um objeto esteja posicionado entre o plano de referência e o sensor IR, um ponto capturado será transladado na direção em relação ao ponto correspondente armazenado na memória do dispositivo. Como pode ser observado na Figura 4, este deslocamento corresponde à paralaxe observada no posicionamento do ponto. A partir de semelhança de triângulos se obtém (KHOSHELHAM & ELBERINK, 2012):

$$\frac{D}{b} = \frac{Z_0 - Z_k}{Z_0} \quad (8)$$

$$\frac{D}{d} = \frac{Z_k}{f} \quad (9)$$

Substituindo a equação (9) na equação (8), é possível expressar  $Z_k$  da seguinte forma (KHOSHELHAM & ELBERINK, 2012):

$$Z_k = \frac{Z_0}{1 + \frac{Z_0}{fb}d} \quad (10)$$

A Equação (10) é o modelo matemático básico para a determinação da profundidade de um ponto pertencente à imagem, observando sua paralaxe e assumindo que os parâmetros  $Z_0$ ,  $f$  e  $b$  são determinados por técnicas de calibração. Khoshelham e Elberink (2012) expressaram em função de uma normalização linear. Neste caso, (na Equação 10) é reescrito na forma  $d = c_0 \cdot d' + c_1$ , sendo  $c_0$  e  $c_1$  os parâmetros de uma normalização linear (na verdade uma denormalização) e a paralaxe normalizada  $d'$ . Substituindo os termos apresentados na Equação (10) e fazendo as devidas manipulações matemáticas, tem-se:

$$Z_k^{-1} = \left( \frac{c_0}{fb} \right) d' + \left( Z_0^{-1} \frac{c_1}{fb} \right) \quad (11)$$

A Equação (11) expressa uma relação linear entre o inverso da profundidade de um ponto e seu correspondente valor de paralaxe normalizado (KHOSHELHAM e ELBERINK, 2012). A coordenada em conjunto com 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):

$$X_k = \frac{Z_k}{f}(x' - x_0 + \delta x) \quad (12)$$

$$Y_k = \frac{Z_k}{f}(y' - y_0 + \delta y) \quad (13)$$

onde,  $x'$  e  $y'$  são as coordenadas do ponto na imagem (observadas no plano focal da câmera IR),  $x_0$  e  $y_0$  são as coordenadas do ponto principal e  $\delta x$ ,  $\delta y$  são as correções de distorção das lentes.

De acordo com Khoshelham e Elberink (2012) o valor  $d$  pode ser expresso em função de uma normalização linear. Substituindo os termos apresentados na Equação (11), tem-se:

$$Z_k = 1 / c_2 d' + c_3 \quad (14)$$

sendo,  $c_2$  e  $c_3$  os parâmetros da normalização linear e  $d'$  a paralaxe normalizada.

#### 2.2.1.4 Associação dos Pontos Chave à Nuvem de Pontos

A associação dos pontos chave, obtidos com o algoritmo SIFT, e seus correspondentes na imagem de profundidade é feito a partir dos valores dos parâmetros de orientação interior calibrados dos sensores RGB e IR, bem como dos valores de  $d\omega$ ,  $d\phi$ ,  $d\kappa$  e  $dx$ ,  $dy$ ,  $dz$  dos sistemas referenciais dos sensores RGB e IR. O método empregado para esta tarefa foi proposto por Dos Santos e Khoshelham (2015). Primeiramente, é feita uma normalização da imagem IR. Apesar da imagem IR estar alinhada com a imagem RGB a diferença entre as origens dos sistemas referenciais não está compensada. Para tratar este problema uma linha epipolar é projetada no sistema de referência da nuvem de pontos. A projeção da linha é feita em função do valor de profundidade máximo e mínimo de alcance do dispositivo ( $Z_{min} = 40cm$  e  $Z_{max} = 7m$ ) e, do valor da linha base ( $b$ -calculado usando  $dx$ ,  $dy$ ,  $dz$ ) e o valor de profundidade  $Z_k$  do ponto. Em seguida, para todos os pixels dentro do intervalo da linha epipolar são calculadas suas coordenadas 3D e reprojetaos para o sistema referencial digital da imagem RGB. Então, é verificado se o pixel reprojetao corresponde ao pixel extraído pelo

algoritmo SIFT na imagem RGB. Para que suas coordenadas do ponto 3D correspondente sejam associadas ao ponto detectado na imagem RGB é necessário que a diferença entre eles seja igual a zero. Esse processo é então repetido para todos os pontos detectados na imagem RGB. Maiores detalhes ver Dos Santos e Khoshelham (2015).

#### 2.2.1.5 Modelo Matemático Funcional

Desta forma, dado o conjunto de pontos 3D correspondentes, é empregado o modelo de transformação de corpo rígido na minimização do somatório dos erros ( $\epsilon$ ) para a determinação dos parâmetros de transformação do registro grosseiro de cada par de nuvens de pontos 3D, como mostra a equação a seguir:

$$E = \sum_{m=1}^n W_{mp} \left\| X_{m,p-1} - (R \cdot X_{m,p} + \vec{T}) \right\| \quad (15)$$

sendo,  $X_{m,p}$  e  $X_{m,p-1}$  as coordenadas 3D correspondentes na nuvem de pontos de Referência ( $M$ ) e de Pesquisa ( $P$ ), respectivamente,  $W_{mp}$  o peso associado ao par de pontos correspondentes  $m$  e  $p$ ,  $\vec{T}$  o vetor dos parâmetros de translação ( $t_x, t_y, t_z$ ) e  $R$  a matriz de rotação em função dos ângulos de Euler ( $\varpi, \phi, \kappa$ ).

Como a Equação (15) é não-linear deve ser aplicado um ajustamento pelo MMQ com o modelo paramétrico com iteração. Os pesos  $W_{mp}$  são encontrados em função da precisão teórica das coordenadas tridimensionais e formarão a matriz dos pesos no MMQ.

#### 2.2.1.6 Ponderação das Coordenadas 3D

Os valores dos pesos são estimados para todo par de pontos correspondentes (*inliers*) baseados na precisão teórica das coordenadas 3D ( $X_k$ ,  $Y_k$  e  $Z_k$ ). Sabendo que as imagens de profundidade são adquiridas em uma taxa de 20 a 30 quadros por segundo, resultando em pequenas diferenças de rotação e translação entre as imagens consecutivas, pode-se aproximar as equações de observação, como segue:

$$v_i = \left\| X_{m,p-1} - X_{m,p} \right\| \quad (16)$$

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



$$w_{mp} = \frac{c}{\sigma_v^2} = \frac{c}{\sigma_{X_{m,p-1}}^2 + \sigma_{X_{m,p}}^2} \quad (17)$$

sendo,  $\sigma_X^2$  a variância de cada ponto  $X$  e  $c$  uma constante arbitrária.

Segundo Kholhelham e Elberink (2012), assumindo que na Equação (11) os parâmetros de calibração estejam determinados acuradamente e  $d'$  seja uma variável randômica com uma distribuição normal, é possível propagar a variância da medida de paralaxe para obter a variância da medida de profundidade como segue:

$$\sigma_{Z_k}^2 = \left( \frac{\partial Z_k}{\partial d} \right) \sigma_{d'}^2 \quad (18)$$

Fazendo as devidas manipulações matemáticas e a normalização  $d = c_0 d' + c_1$ , se chega à expressão para o desvio padrão da profundidade (KHOSHELHAM e ELBERINK, 2012):

$$\sigma_{Z_k} = \left( \frac{c_0}{fb} \right) Z_k \sigma_{d'} \quad (19)$$

sendo,  $\sigma_{d'}$  e  $\sigma_Z$  o desvio padrão da medida de paralaxe normalizada e da profundidade calculada, respectivamente.

Portanto, a variância da profundidade  $\sigma_{Z_k}^2$  tem a seguinte relação com a variância da paralaxe normalizada medida  $\sigma_{d'}^2$  (KHOSHELHAM *et al.*, 2013):

$$\sigma_{Z_k}^2 = c_2^2 \sigma_{d'}^2 Z_k^4 \quad (20)$$

Substituindo a Equação (20) na (17) obtém-se a ponderação da medida de profundidade, como segue (KHOSHELHAM *et al.*, 2013):

$$w_{Z_k} = \frac{\Psi}{z_{i,j-1}^4 + z_{i,j}^4} \quad (21)$$

sendo, na prática  $\Psi = 1$ .

Neste trabalho também é proposto a inclusão dos valores de peso para as coordenadas em  $X_k$  e  $Y_k$ , assim, propagando os erros nas

equações 12 e 13. Assumindo que o erro randômico das coordenadas de imagem  $x'_c$  e  $y'_c$  possa ser ignorado, obtém-se o erro randômico das coordenadas  $X_k$  e  $Y_k$ , como segue:

$$\sigma_{X_k} = \left( \frac{c_0 x'_c}{f^2 b} \right) Z_k^2 \sigma_{d'} \quad (22)$$

$$\sigma_{Y_k} = \left( \frac{c_0 y'_c}{f^2 b} \right) Z_k^2 \sigma_{d'} \quad (23)$$

Usando o mesmo raciocínio realizado anteriormente, mas agora para as coordenadas  $X_k$  e  $Y_k$ , tem-se:

$$\sigma_{\bar{X}}^2 = \frac{\bar{x}_c^2}{f^2} \sigma_{Z_k}^2 + \frac{Z_k^2}{f^2} \sigma_{x'_c}^2 \quad (24)$$

sendo,  $\sigma_{\bar{X}}^2$  as variâncias das coordenadas  $X_k$  e  $Y_k$ ,  $\bar{x}_c$  as coordenadas  $x'_c$  e  $y'_c$  no espaço-imagem e  $\sigma_{x'_c}^2$  suas variâncias.

Substituindo as equações (12) e (20) na Equação (24) tem-se:

$$\sigma_{X_k}^2 = c_2^2 \sigma_{d'}^2 x'_c Z_k^2 + \frac{\sigma_{x'_c}^2}{f^2} Z_k^2 \quad (25)$$

Da mesma forma, substituindo as equações (13) e (20) na equação (24) tem-se:

$$\sigma_{Y_k}^2 = c_2^2 \sigma_{d'}^2 y'_c Z_k^2 + \frac{\sigma_{y'_c}^2}{f^2} Z_k^2 \quad (26)$$

Substituindo as equações (26) e (25) na equação (21) obtém-se a ponderação das medidas em  $X_k$  e  $Y_k$ , como segue:

$$w_{X_k} = \frac{\Psi}{Z_{m,p-1}^2 (X_{k,m,p-1}^2 + \mu) + Z_{m,p}^2 (X_{k,m,p-1}^2 + \mu)} \quad (27)$$

$$w_{Y_k} = \frac{\Psi}{Z_{m,p-1}^2 (Y_{k,m,p-1}^2 + \mu) + Z_{m,p}^2 (Y_{k,m,p-1}^2 + \mu)} \quad (28)$$

sendo,  $\Psi = c c_2^{-2} \sigma_{d'}^{-2}$ , embora na prática adotado como 1 e  $\mu = c_2^{-2} f^{-2} \sigma_{d'}^{-2} \sigma_{x'y'}^2$ , constantes definidas pelos parâmetros  $c_2$  e  $f$ , e,  $\sigma_{x'y'}^2 = \sigma_{x'}^2 = \sigma_{y'}^2$ .

Obtidos os valores dos pesos é feito o ajustamento por MMQ calculando os parâmetros de transformação na etapa do registro grosseiro. Em seguida, esses parâmetros de transformação são aplicados à nuvem de pontos de pesquisa esperando que o desalinhamento inicial entre as nuvens de pontos de referência e de pesquisa seja minimizado.

- Refinamento dos Parâmetros de Transformação Iniciais
- O refinamento dos parâmetros de transformação iniciais é feito empregando as seguintes tarefas, a saber:
- Divisão do quadro IR em células de tamanhos iguais;
- Determinação do vizinho mais próximo pela técnica *k-d tree*;
- Refinamento dos parâmetros de transformação usando o algoritmo ICP e o MMQ com os valores iniciais advindos do registro grosseiro;

### 2.2.2.1 Divisão do Quadro IR em Células

Dadas as nuvens de pontos  $M$  e  $P$ , o algoritmo ICP consiste em estimar os parâmetros de transformações (3 rotações e 3 translações) entre os pares de nuvens de pontos  $P$  e  $M$  no  $\mathbb{R}^3$ . A nuvem de referência é regularmente dividida em células de dimensão  $10 \times 10$  centímetros. Para cada célula é determinado o seu centroide ( $q$ ), como segue:

$$q_k = \left( \frac{1}{k} \sum X_k, \frac{1}{k} \sum Y_k, \frac{1}{k} \sum Z_k \right) \quad (29)$$

sendo,  $X_k$ ,  $Y_k$  e  $Z_k$  o  $k$ -ésimo ponto contido na célula e 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 através das equações (12)-(14), sendo obtido um conjunto de  $P_i$  pontos de pesquisa.

### 2.2.2.2 Busca do Vizinho mais Próximo pela Técnica *K-d tree*

Para acelerar o processo de busca pelo vizinho mais próximo foi empregada a estrutura *k-d tree*. Basicamente, o algoritmo realiza

as seguintes tarefas, a saber: Dado um ponto centroide ( $q$ ) e uma raiz da árvore de pesquisa ( $P$ ) faz-se:

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

### 2.2.2.3 Solução do Modelo para a Etapa do Refinamento dos Parâmetros de Transformação

Encontrado um conjunto de pontos correspondentes, a cada iteração, o modelo de corpo rígido 3D é empregado para refinar os parâmetros de transformação do registro grosseiro. 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 mostra a Equação (15). Como descrito anteriormente, a Equação (15) não é linear e envolve 6 parâmetros a serem determinados por MMQ. Neste caso, deve ser aplicado o modelo paramétrico com iteração (sem pesos neste caso). A cada iteração, a distância 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.

Uma vez que não se pode admitir correspondências exatas entre pontos em um par de nuvens de pontos 3D obtidas pelo Kinect, isto faz com que o modelo de otimização (Eq. 15) seja altamente sensível a ruídos, dada a forma como são atribuídos os 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. 12 e 13), que são calculadas em função das coordenadas no espaço-imagem e de  $Z_k$ . Para reduzir a probabilidade de utilizar *outliers* na etapa do registro de pares de nuvens de pontos são definidos dois critérios: o da distância mínima entre os pontos; e o da variação angular (no plano das abcissas) da linha imaginária que liga os pontos correspondentes em potencial nos pares de nuvens.

O critério da distância mínima se adapta ao valor da Raiz do Erro Médio Quadrático (REMQ) das nuvens de pontos de Referência e de Pesquisa (REMQ\_MP). Após cada iteração do algoritmo ICP, a REMQ\_MP calculada é adotado como um novo valor de limiar. Somente pontos potencialmente correspondentes que sejam inferiores ao valor da REMQ\_MP são considerados correspondentes. À medida que as diferenças angulares e lineares do registro diminuem, as correspondências entre os pontos tendem a serem mais exatas (apesar de nunca serem exatas). 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.

Finalmente, o critério de parada do algoritmo também está em função do valor da REMQ\_MP. O algoritmo termina o processamento caso a diferença da REMQ\_MP calculada daquela iteração com a anterior seja menor que um determinado limiar. Isto significa que a estimativa dos dados está estável e o processamento deve ser finalizado. Este passo também reduz o tempo de processamento do algoritmo ICP.

A etapa do registro entre pares de nuvens de pontos deve ser aplicada enquanto o número de imagens for menor que  $num + 1$ .

### 2.3 Análise da consistência global

Para este trabalho, a etapa de análise da consistência global (ajustamento global) consiste

em transformar o conjunto de parâmetros de translação 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 translação. É importante destacar que, como o processamento não é feito em tempo real, a tarefa de detecção de lugares revisitados (erro de fechamento) é negligenciada da etapa de análise da consistência global.

Neste trabalho, a análise de consistência global dos parâmetros de translação obtidos no registro dos pares de nuvens de pontos utiliza de uma variação no método 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  $i = 0$  número total de posições) em sua trajetória  $t_x^{S_{i,i+1}}, t_y^{S_{i,i+1}}, t_z^{S_{i,i+1}}$ . Diferente de Lu e Milios (1997), neste trabalho será feito o emprego direto dos parâmetros de translação 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.

Portanto, dada uma trajetória do sensor formada por um conjunto de  $K$  estações ocupadas pelo Kinect ( $K_0, \dots, K_n$ ), como mostra a Figura 5, o novo conjunto de parâmetros de translação é estimado simultaneamente baseado na seguinte equação, a saber:

$$F X_{novo}^i = X_j^i - X_j^i \quad (30)$$

sendo,  $F$  o modelo funcional,  $X_{novo}^i$  o novo conjunto de parâmetros de translação ajustado,  $X_{j+1}^i$  a matriz homogênea dos parâmetros de translação da estação posterior e  $X_j^i$  a matriz homogênea dos parâmetros de translação da estação anterior.

A formulação do algoritmo de estimação ótima, a partir do circuito de medidas conhecidas sobre  $n + 1$  observações  $X_0 = [t_x^{S_{10}}, t_y^{S_{10}}, t_z^{S_{10}}]^T$ ,  $X_1 = [t_x^{S_{21}}, t_y^{S_{21}}, t_z^{S_{21}}]^T$ ,  $X_n = [t_x^{S_{n,n+1}}, t_y^{S_{n,n+1}}, t_z^{S_{n,n+1}}]^T$  sugere que cada um deles represente um vetor  $6 \times 1$ , e uma aresta  $L_{ij}$  entre duas observações  $X_i$  e  $X_j$

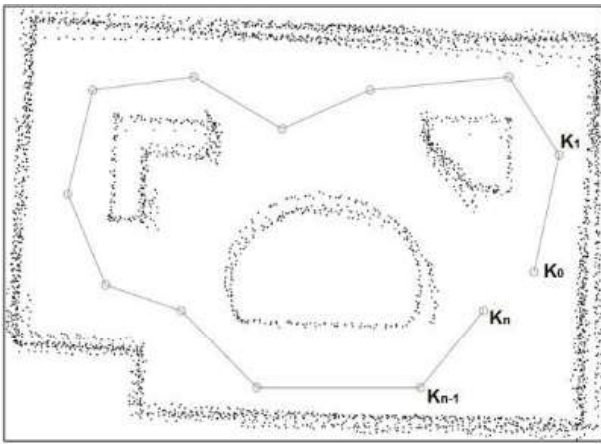


Fig. 5 - Exemplo de trajetória e estações ocupadas pelo sensor

represente uma diferença mensurável. Neste trabalho, a equação de medida, referente à função  $L_{ij}$  de  $X_i$  e  $X_j$ , é dada na forma linear, como segue:

$$L_{ij} = X_i - X_j \quad (31)$$

Portanto, assumindo que há uma ligação entre todas as observações e para cada  $L_{ij}$  existe uma observação  $\bar{L}_{ij}$  que é admitida 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):

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

Apartir da concatenação de todas as observações  $X_0, X_1, \dots, X_n$ , ( $X$ ) das diferenças  $L_{ij} = X_i - X_j$ , ( $L$ ) e inserindo uma matriz de incidência ( $A$ ), formada por valores 1, -1, e 0, chega-se a equação linear de medida em sua forma matricial dada por (LU e MILIOS, 1997):

$$L = AX \quad (33)$$

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

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

sendo,  $\bar{L}$  a concatenação de todas as observações  $\bar{L}_{ij}$  referentes à  $L_{ij}$  e  $C$  a covariância de  $\bar{L}$ . A solução para  $X$  que minimiza  $Q$  é dada por (LU e MILIOS, 1997):

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

Para compor a matriz  $C$ , neste trabalho, é proposto o uso da MVC dos parâmetros de translação do registro obtidos na etapa de refinamento dos pares de nuvens de pontos. Assim, um ajustamento pelo MMQ é realizado e o novo conjunto de parâmetros é estimado simultaneamente baseado na equação (30). Como foi dito, as observações neste ajustamento são os valores de translação determinados na etapa do registro entre os pares de nuvens de pontos 3D.

### 2.3.1 Reconstrução do Cenário 3D

Uma vez que as posições das estações ocupadas pelo dispositivo são globalmente ajustadas, pode ser realizada a reconstrução 3D da cena. Para cada quadro é gerada uma nuvem de pontos 3D. A fim de reconstruir todo o ambiente as informações devem ser combinadas a partir de um único ponto de vista. A priori, cada nuvem de pontos 3D é transformada, projetando todos os seus pontos de acordo com a posição do sensor correspondente. Considerando uma sequência finita de estações ocupadas pelo dispositivo Kinect, essa posição é encontrada combinando a sequência de transformações através de um produtório das matrizes homogênea, que são formadas pelas matrizes de rotação e os vetores de translação relativa de cada posição do sensor.

A seguir serão apresentados os ensaios realizados e a discussão dos resultados obtidos.

## 3. EXPERIMENTOS E RESULTADOS

Para avaliar o método proposto neste trabalho foram realizados 4 ensaios em cenários (A-D) que representam ambientes internos, tais como, quartos, salas, escritórios etc. 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. Primeiramente foi realizada a calibração do Kinect bem como de seus sensores RGB e IR. Essa etapa pode ser dividida em duas tarefas: 1) calibração dos POI dos sensores RGB e IR embarcados no Kinect, bem como a

determinação dos ângulos de desalinhamento e a distância entre as origens dos sistemas referenciais de cada sensor (POE); 2) a determinação dos parâmetros de normalização dos valores de profundidade  $c_2$  e  $c_3$ .

Os POI dos sensores RGB e IR foram obtidos por um processo de calibração de câmeras usando o método de Zhang. Esse método foi escolhido devido à facilidade que oferece no processo de calibração do Kinect. Para essa calibração foi utilizado o programa ‘Kinect-stereo-calib’ disponível gratuitamente pela MRPT. Portanto, foram tomadas, simultaneamente, 16 imagens RGB e IR sobre um tabuleiro de xadrez padrão em diferentes pontos de vista.

Os POI de ambos os sensores e os POE podem ser encontrados na Tabela 1. Em relação à calibração estéreo, os parâmetros de translação e rotação são calculados da câmera da esquerda para a câmera da direita, ou seja, da IR para a RGB.

Tabela 1: POI dos Sensores RGB e IR, seus Desvios Padrão e os Valores dos POE

Parâmetros	RGB	IR
$f$ (mm)	$4,09 \pm 0,00036$	$4,55 \pm 0,00036$
$x_0$ (mm) (offset)	$-0,0351 \pm 0,00008$	$-0,0273 \pm 0,00008$
$y_0$ (mm) (offset)	$-0,1521 \pm 0,00008$	$-0,0897 \pm 0,00008$
$k_1$ (mm <sup>2</sup> )	$2,36 \times 10^{-1} \pm 8,0 \times 10^{-6}$	$-7,76 \times 10^{-2} \pm 8,5 \times 10^{-6}$
$k_2$ (mm <sup>2</sup> )	$-6,85 \times 10^{-1} \pm 7,5 \times 10^{-7}$	$3,37 \times 10^{-1} \pm 8,0 \times 10^{-7}$
$k_3$ (mm <sup>2</sup> )	$6,44 \times 10^{-3} \pm 8,0 \times 10^{-6}$	$-4,68 \times 10^{-1} \pm 8,0 \times 10^{-7}$
$p_1$ (mm <sup>2</sup> )	$1,24 \times 10^{-3} \pm 8,0 \times 10^{-6}$	$-2,38 \times 10^{-3} \pm 8,0 \times 10^{-7}$
$p_2$ (mm <sup>2</sup> )	$2,82 \times 10^{-3} \pm 8,0 \times 10^{-6}$	$1,37 \times 10^{-3} \pm 8,0 \times 10^{-7}$
$\Delta\omega$	$\Delta\varphi$	$\Delta\kappa$ (°)
0,0240	-0,0001	-0,0011
$\Delta T_x$	$\Delta T_y$	$\Delta T_z$ (mm)
25,18	-0,0041	-0,0093

Como podem ser verificados na Tabela 1, os POE indicam que o centro do sensor RGB está localizado aproximadamente na mesma linha de base entre o sensor IR e o emissor LASER. Os POE encontrados são os esperados dada a integração dos sensores IR e RGB. A acurácia absoluta da calibração foi de 0,55 pixels para ambas as imagens, conforme os dados da REMQ do resíduo de reprojeção do ponto no espaço objeto.

Já os valores de (declive) e (intersecção

da linha) foram obtidos usando o método proposto por Khoshelham e Elberink (2012). Ao contrário do que foi feito em Khoshelham e Elberink (2012), em que é usada uma fita métrica convencional para medir os deslocamentos do sensor frente a seu alvo, neste trabalho foi usado um interferômetro LASER disponível no Laboratório de Instrumentação Geodésica (LAIG) marca Hewlett Packard modelo 5508<sup>a</sup> que propicia medidas com precisão em torno 0,01 mm.

Para isto, foram obtidos os valores de paralaxe no centro da placa (ver Figura 6) para cada imagem de profundidade e plotados em relação ao inverso das distâncias do sensor para placa conforme se fazia os precisos deslocamentos com o interferômetro.



Fig. 6 - Experimento realizado com o interferômetro e com o Kinect para a normalização dos valores de profundidade. (a) imagem RGB; (b) imagem de profundidade.

A Figura 7 ilustra um gráfico que mostra a relação profundidade-paralaxe onde é possível perceber a relação linear expressada pela Equação (4) resultando nos valores para  $c_2$  de  $-3,07 \times 10^{-5}$  e para  $c_3$  de 0,0334.

A partir da obtenção dos parâmetros da calibração foi possível conduzir os 4 ensaios propostos. Para este trabalho não serão usadas medidas verdadeiras das cenas, ou seja, pontos de controle com coordenadas absolutamente definidas, portanto, em cada experimento foi obtida uma trajetória formando uma volta completa apenas com o movimento circular das mãos para que, de certa forma, o método consiga registrar essas nuvens deixando a primeira e a última o mais coincidente possível. A Tabela 2 mostra o número de quadros para cada experimento realizado para avaliar o método proposto.

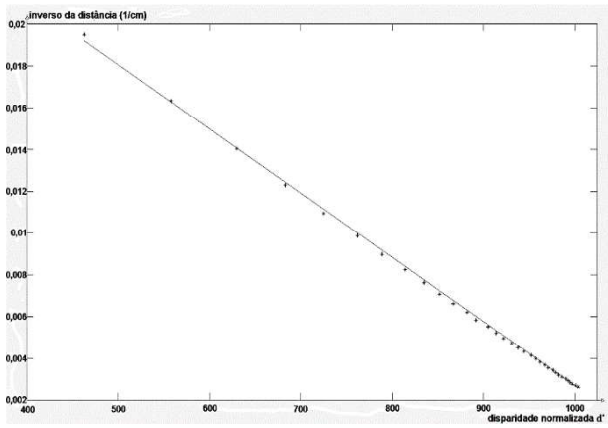


Fig. 7 – Relação linear da paralaxe normalizada com o inverso da profundidade.

Tabela 2: Experimentos Realizados

Experimento	Número de quadros
Cenário A	74
Cenário B	41
Cenário C	63
Cenário D	75

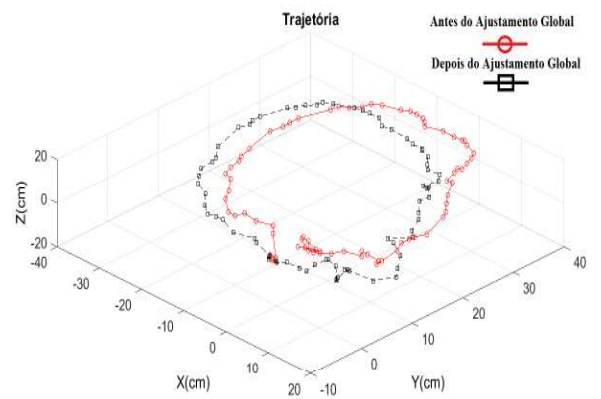
As Figuras 8 e 9 mostram as trajetórias do sensor sem o uso de pesos (c, verde), com pesos em (a, vermelha), e em XYZ (b, azul) para os cenários A e D, respectivamente. Nas mesmas figuras são plotadas as trajetórias (a, b e c, preta) ajustadas (sem ponderação) desses ensaios.

A Figura 10 mostra o ambiente interno tridimensionalmente mapeado sem pesos e com pesos em Z dos cenários A e B. Na Figura é possível perceber uma melhora na reconstrução com o uso de pesos quando comparado ao não uso de pesos nas coordenadas 3D.

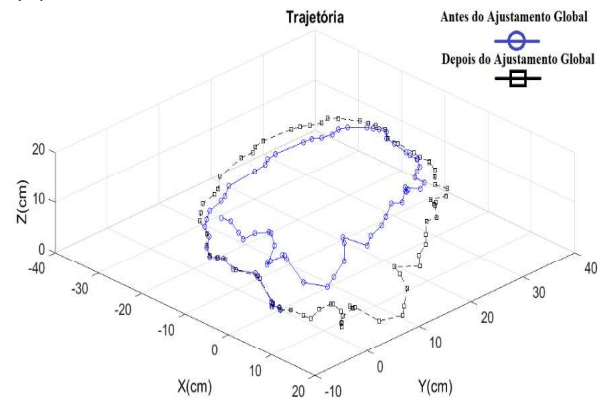
A seguir são apresentadas as discussões dos resultados obtidos com o método proposto.

### 3.1 Análise dos Resultados

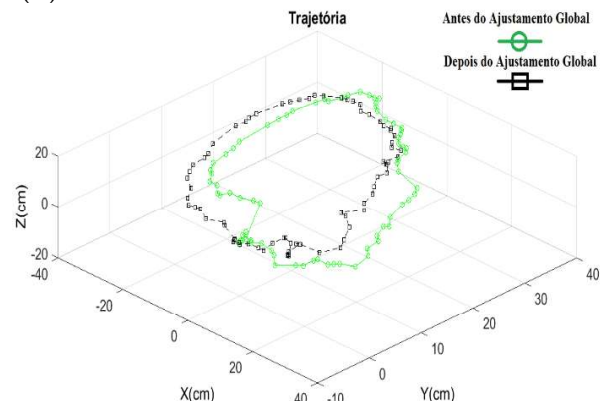
Nas Figuras 8 e 9, as trajetórias do sensor, para os experimentos A e D, são obtidas como o esperado. Em relação ao comparativo do não uso de pesos, do uso em e em , no registro dos pares de nuvens de pontos 3D, as Figuras 8 e 9 mostram que para os cenários A e D, respectivamente, a trajetória que mais se parece com a ajustada é aquela aplicando peso somente em (Figuras 8 e 9; a). Além do mais, por meio desse experimento é possível perceber como a



(a)

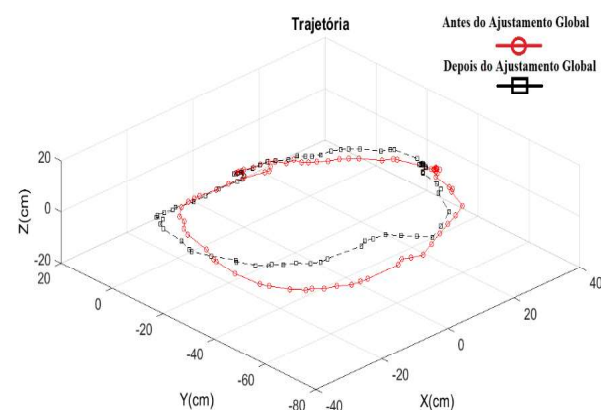


(b)



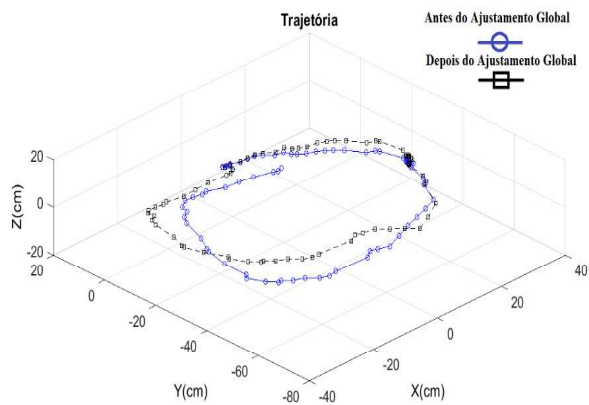
(c)

Fig. 8 - Trajetória do sensor no cenário A depois da análise de consistência global e antes com peso em (a), em (b) e sem pesos (c)

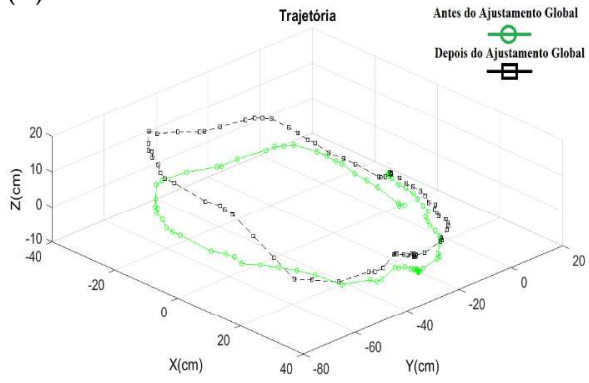


(a)



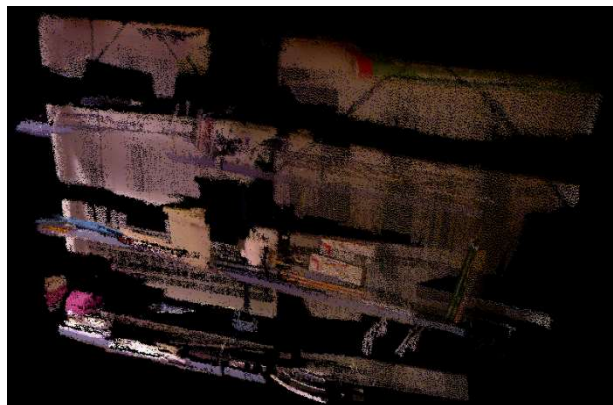


(b)



(c)

Fig. 9 - Trajetória do sensor no cenário D depois da análise de consistência global e antes com peso em (a), em (b) e sem pesos (c)



(c)



(d)

Fig. 10 – Nuvem de pontos 3D de ambientes internos mapeados. Cenário B: (a) não usando pesos e (b) usando pesos; Cenário C: (c) não usando pesos (d) usando pesos.



(a)



(b)

propagação de erros age no registro dos pares de nuvens de pontos 3D quando comparado às trajetórias antes e depois do ajustamento global. Portanto, as figuras reafirmam a dependência de uma análise de consistência global consistente, que ajusta os parâmetros de transformação do registro fazendo com que o *loop* feche no final da trajetória, e, assim, gere nuvens de pontos visual e consistentemente melhores.

Para analisar quantitativamente os resultados obtidos foi calculada a REMQ das discrepâncias e a média das discrepâncias entre os pontos correspondentes encontrados na última iteração do refinamento com o algoritmo ICP (neste caso, os pontos considerados correspondentes). A REMQ das discrepâncias dos experimentos A e B é exibida na Figura 11.

As Tabelas 3, 4 e 5 mostram os valores máximos e mínimos da REMQ para os experimentos A, B e C, respectivamente.

Tabela 3: REMQ das Coordenadas X, Y e Z Relativo ao Cenário A

Coordenada	REMQ (cm)
X	Min: 0,01 Max: 0,9
Y	Min: 0,02 Max: 0,46
Z	Min: 0,05 Max: 0,54

Tabela 4: REMQ das Coordenadas X, Y e Z Relativo ao Cenário B

Coordenada	REMQ (cm)
X	Min: 0,03 Max: 0,7
Y	Min: 0,01 Max: 0,6
Z	Min: 0,08 Max: 0,43

Tabela 5: REMQ das Coordenadas X, Y e Z Relativo ao Cenário C

Coordenada	REMQ (cm)
X	Min: 0,03 Max: 0,89
Y	Min: 0,02 Max: 0,25
Z	Min: 0,06 Max: 0,75

Para todos os experimentos o valor da média das discrepâncias ficou próximo de zero, mostrando que nenhuma tendência significativa influenciou o modelo matemático.

Em ambientes considerados pouco texturizados, mesmo em um mapeamento de curta distância, o método de registro grosseiro pode falhar, uma vez que menos de 3 pontos sejam detectados pelo algoritmo SIFT, ou, em casos que, após a remoção de *outliers*, restarem menos de 3 correspondências. A Figura 12 mostra um gráfico de *inliers* que foram encontrados no registro grosseiro dos pares de nuvens de pontos dos cenários C e D. O gráfico mostrado na Figura 12a revela que conforme o sensor percorre sua trajetória o número de *inliers* cai chegando a apenas 5 correspondências, e então, o número volta a subir assim que o *loop* retorna, aproximadamente, ao seu ponto de origem. Diante de situações como essa, o método sugerido neste trabalho se comporta bem. Já o gráfico mostrado na Figura 12b ilustra um caso extremo, como citado anteriormente, em que o SIFT falha e somente o algoritmo ICP é utilizado.

Essa é a maior vantagem na utilização dessa metodologia fino-para-grosseiro no registro de nuvens de pontos.

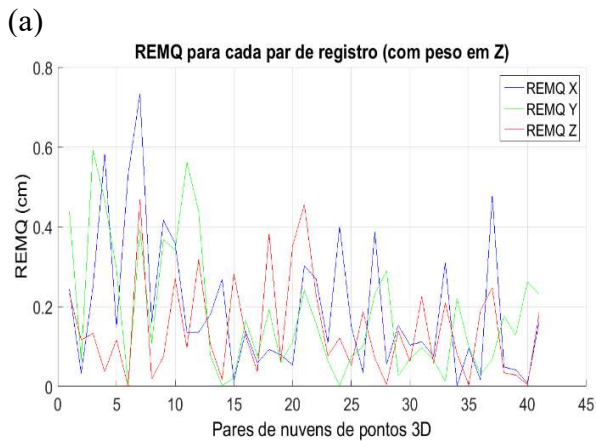
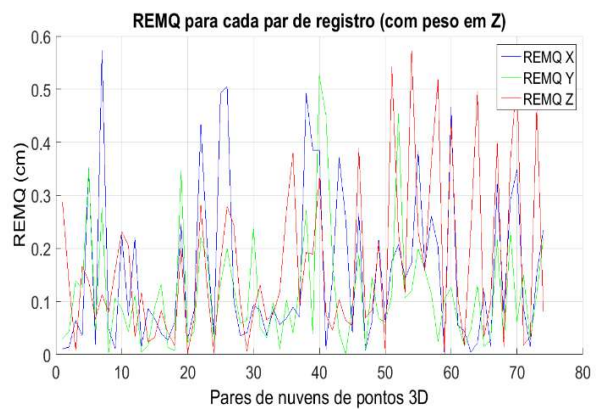


Fig. 11 - REMQ da média das discrepâncias calculada em cada registro de pares de nuvens de pontos. (a) cenário A; (b) cenário B.

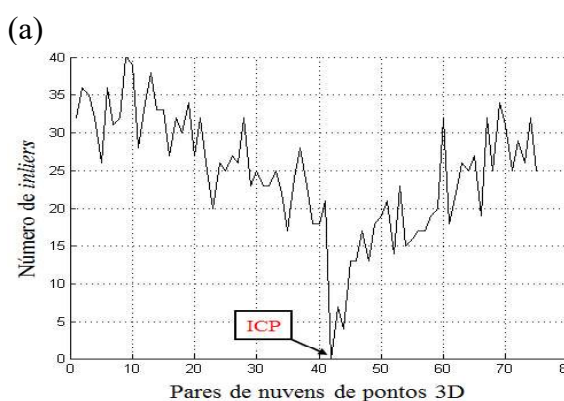


Fig. 12 - Inliers advindos do registro grosseiro com o método SIFT: (a) cenário C; (b) cenário D.



A Figura 13 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 B 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.

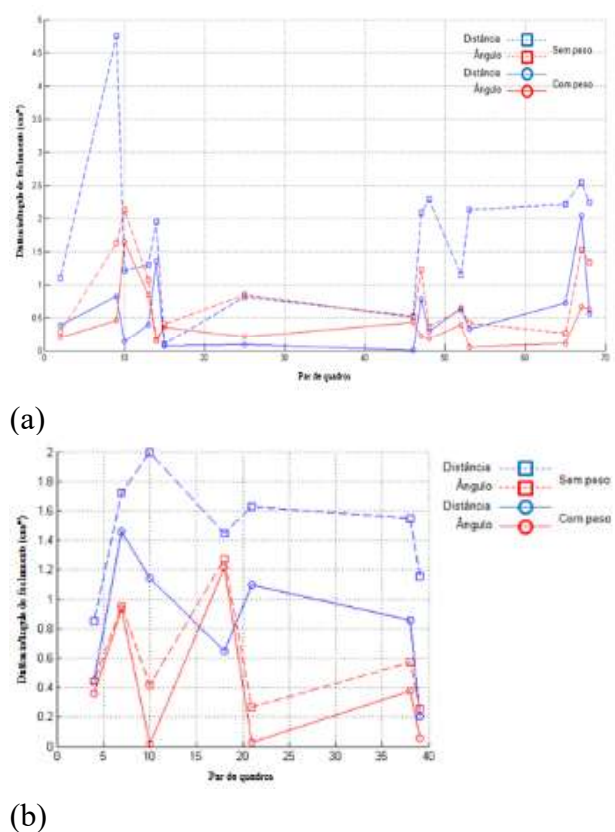


Fig. 13 - Distância (azul) e ângulo (vermelho) de fechamento para sequência de registros entre pares de nuvens de pontos com e sem peso. (a) cenário A; (b) cenário B.

A explicação do não uso de pesos em XYZ se baseia em como a precisão teórica das medidas em X e Y são calculadas. Elas são feitas em função das coordenadas no espaço imagem, corrigidas dos erros sistemáticos, assim, os valores obtidos são piores à medida que o ponto se afasta do centro geométrico da imagem. Esta redução na contribuição de pontos periféricos causa um efeito negativo no registro dos pares de nuvens de pontos, uma vez que são esses pontos

periféricos que mais contribuem para o registro. Essa constatação pode ser observada nas Figuras 8 e 9, as quais ilustram essa maior deterioração da qualidade dos parâmetros conforme o desenrolar da tomada dos quadros, até mesmo em comparação àqueles processamentos sem o uso de pesos. 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.

### 3. CONCLUSÕES E RECOMENDAÇÕES

Este trabalho apresentou uma análise da ponderação do erro randômico teórico das coordenadas 3D no registro de dados RGB-D. As principais características do método são: a inclusão de pesos baseado no erro randômico teórico das medidas de profundidade; e a determinação dos parâmetros de transformação em duas etapas, além do desenvolvimento de um modelo linear para ajustamento simultâneo dos pares de nuvens de pontos 3D tratando o problema de detecção de lugares revisitados em modo pós-processamento a partir da Matriz Variância Covariância dos parâmetros de transformação. O método de Zhang foi usado neste trabalho para a calibração dos sensores RGB e IR, bem como para a determinação dos parâmetros de montagem do sistema. Este método foi empregado por ser totalmente automático e flexível para a nossa aplicação.

De forma geral, o método proposto apresenta como maior vantagem não necessitar de um número mínimo de *inliers* detectados pelo SIFT para determinar os parâmetros de transformação do registro, sendo que neste caso, o algoritmo ICP por si próprio deve ser capaz de realizar o registro das nuvens de pontos juntamente com o ajustamento global. Portanto, este método pode ser empregado em ambientes com regiões pouco texturizadas ou composta por muitos objetos com material especular ou de alta absorção de energia radiante na região do infravermelho próximo.

De acordo com o que foi observado por Khoshelham *et al.*, (2013) e com base nas Figuras 8 e 9, na qual é possível ver que a trajetória que mais se aproxima da ajustada é a com o uso de pesos na coordenada Z, conclui-se que houve uma melhoria significativa no registro do conjunto de nuvens de pontos 3D obtidas

com o Kinect com ponderação na coordenada . É importante frisar que o método se comportou de forma adequada, uma vez que era esperada uma redução nos valores de REMQ conforme a convergência do algoritmo.

A critério de comparação, as nuvens de pontos de todos os cenários foram obtidas a uma distância média de 2 metros, e 98,88% dos registros apresentaram uma acurácia relativa inferior a 11mm, que é aproximadamente a resolução da profundidade à essa distância, segundo Khoshelham e Elberink (2012).

O método proposto para análise da consistência global é bastante simples, uma vez que emprega uma função linear e permite que as observações possam ser ajustadas simultaneamente, propiciando melhor precisão no ajustamento das observações, ao contrário dos métodos propostos na literatura que, geralmente, descartam as informações à medida que o erro de fechamento é detectado.

Para a calibração do sistema e dos sensores, em trabalhos futuros pretende-se empregar modelos mais rigorosos, inclusive implementar um modelo que considere no processo de calibração o dispositivo LASER do Kinect. Também é recomendado, como trabalhos futuros, investigações sobre abordagens ICP ponto-a-plano no processo de registro dos pares de nuvens de pontos 3D, além da inclusão de novas restrições geométricas no processo de estabelecimento de correspondências e aplicar técnicas de filtragem robustas que reduzam o número de *outliers* presentes na nuvem de pontos 3D.

## AGRADECIMENTOS

À CAPES e ao CNPq (processo: 473567/2013-0) pelo apoio concedido.

## REFERÊNCIAS BIBLIOGRÁFICAS

AL-MANASIR, K.; FRASER, C. S. Registration of terrestrial laser scanner data using imagery. **The Photogrammetric Record**, v. 21, 2006, p. 255-268.

AMINTOOSI, M.; FATHY, M.; MOZAYANI N. A Fast Image Registration Approach Based on SIFT Key-Points Applied to Super-Resolution, **The Imaging Science Journal**, vol. 60, ed. 4, 2012, p. 185-201.

BARNEA, S.; FILIN, S. Keypoint Based Autonomous Registration of Terrestrial Laser Point-Clouds. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 63, 2008, p. 19-35.

BELS, P. J.; MCKAY, H. D. A method for registration of 3-D shapes. **IEEE Trans. Pattern Anal. Mach. Intell.**, v. 14, 1992, p. 239–256.

BENDELS, G. H.; DEGENER, P.; WAHL, R., K, M.; KLEIN, R. Image-based Registration of 3D-Range Data Using Feature Surface Elements. In: **5ª International Conference on Virtual Reality, Archaeology and Intelligent Cultural Heritage**, Oudenaarde, Belgium, Eurographics Association, 2004, p. 115-124.

BROWN, D. C. Close-range camera calibration. **Photogramm**, v. 37, Eng. 1971, p. 855-866.

CHOW J. C. K.; ANG, K. D.; LICHTI, D. D.; TESKEY, W. F. Performance Analysis of a Low-Cost Triangulation-Based 3D Camera: Microsoft Kinect System. In: **XXII ISPRS Congress, Melbourne, Australia, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, v. XXXIX-B5, 2012, p. 175 – 180.

DOS SANTOS, D. R.; KOSHELHAM, K.; Mapeamento 3D de Ambientes Internos usando Dados RGB-D, **Boletim de Ciências Geodésicas**, v. 21, ed. 3, Curitiba, July/Sept 2015, p. 442-464. Disponível em: <http://dx.doi.org/10.1590/S1982-21702015000300025>.

ELLEKILDE, L.-P.; HUANG, S.; VALLSMIRO, J.; DISSANAYAKE, G. Dense 3D Map Construction for Indoor Search and Rescue. In: **Journal of Field robotics**, v. 24, 2007, p. 71-89.

HORN, B. K. P. Closed-form solution of absolute orientation using unit quaternions, **Journal Optical Society of America**, v. 4, Abril de 1987, p. 629-642. Disponível em: [http://people.csail.mit.edu/bkph/papers/Absolute\\_Orientation.pdf](http://people.csail.mit.edu/bkph/papers/Absolute_Orientation.pdf). Acesso em 08/08/2014.

HUHLE, B.; JENKE, P.; STRASSER, W. On-the-Fly Scene Acquisition with a Handy Multi-Sensor System, **Journal Intelligent Systems Technologies and Applications**, v. 5, 2008, p. 255-263. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/>