

## MAPEAMENTO 3D DE AMBIENTES INTERNOS USANDO DADOS

### RGB-D

#### *3D mapping of indoor environments using RGB-D data*

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

Kourosh Khoshelham<sup>2</sup>

Universidade Federal do Paraná, Programa de Pós-Graduação em Ciências Geodésicas, Curitiba, Brasil<sup>1</sup>.Email:[danielsantos@ufpr.br](mailto:danielsantos@ufpr.br)

University of Melbourne, Department of Infrastructure Engineering, Melbourne 3010, Australia<sup>2</sup>.Email:[k.khoshelham@unimelb.edu.au](mailto:k.khoshelham@unimelb.edu.au)

#### **Resumo:**

Neste trabalho é introduzido um fluxo de mapeamento 3D de ambientes internos usando dados RGB-D. O método explora a integração de imagens RGB e valores de profundidade oriundos do dispositivo Kinect. Cinco etapas principais envolvidas no desenvolvimento do método proposto são discutidas. A primeira etapa trata da detecção de pontos no par de imagens RGB e o estabelecimento automático das correspondências. Na segunda etapa do método é proposto uma normalização das imagens RGB e IR para associar os pontos homólogos encontrados no par de imagens RGB e seus correspondentes na imagem de profundidade. Na terceira etapa as coordenadas XYZ de cada ponto são calculadas. Em seguida, são calculados os parâmetros de transformação entre os pares de nuvem de pontos 3D. Finalmente, é proposto um modelo linear para a análise da consistência global. Para avaliar a eficiência e potencialidade do método proposto foram realizados quatro experimentos em ambientes internos. Uma avaliação da acurácia relativa da trajetória do sensor mostrou erros no registro de pares de nuvens de pontos em torno de 3,0 cm.

**Palavras-chave:** Fotogrametria, Kinect, RGB-D, Mapeamento 3D, Registro de nuvem de pontos 3D, Ajustamento Global.

#### **Abstract:**

In this work a method for 3D mapping of indoor environments using RGB-D data is presented. Our method exploited the RGB and depth images provided by Kinect sensor for pairwise registration. The proposed method contains five main steps. Firstly, keypoints are detected and automatically established their correspondences in the RGB images. In the second step we propose a rectification between RGB and IR images to find the best association of the correspondent keypoints and depth values. Following, the XYZ coordinates can be computed.

Then, a pairwise registration is realized. Finally, we present a linear model for the global consistency. The performance of the method is demonstrated using four datasets of an office environment. An assessment of relative accuracy of the sensor trajectory shows pairwise registration errors around 3,0 cm.

**Keywords:** Photogrammetry, Kinect, RGB-D data, Indoor 3D mapping, Pairwise registration, Global consistency.

## 1. Introdução

O crescimento explosivo de sistemas sensores de imageamento e de navegação tornou o mapeamento tridimensional (3D) de ambientes internos uma tarefa essencial em inúmeras aplicações de engenharia, tais como, para sistema de localização e mapeamento simultâneo conhecido como SLAM (*Simultaneous Localization and Mapping*), para vigilância e gerenciamento de emergências, para navegação e posicionamento, inspeção de infraestruturas, modelagem e reconstrução 3D, Cartografia, entre outras. Em novembro de 2011 o *Google Maps* lançou seu projeto de mapeamento 3D de ambientes internos que permite qualquer usuário de *smartphone* acessar mapas *on-line* de ambientes públicos. Esses mapas podem ser facilmente acessados por tecnologia Wi-Fi, Bluetooth, RFID (*Radio-Frequency Identification*), sinais de televisão, etc. Recentemente, o Google lançou seu projeto mais arrojado em mapeamento 3D de ambientes internos denominado de projeto Tango. Um *smartphone* equipado com sensores de imageamento RGB e de profundidade (D) é base do segredo da empresa. O equipamento permite que usuários detentores da tecnologia possam construir seus próprios mapas tridimensionais em qualquer ambiente internos.

O problema de mapeamento 3D de ambientes internos é tema de pesquisa na comunidade fotogramétrica e da Visão Computacional desde os meados da década de 1990. A maioria dos sistemas desenvolvidos para mapeamento 3D de ambientes internos empregam estéreo ou mono câmeras fotográficas (Debevec et al., 1996; Akbarzadeh et al., 2006; Clemente et al. 2007; Furukawa et al. 2009) e sensores ativos de imageamento, como por exemplo, o sistema laser *scanning* terrestre e câmeras de distância 3D (Masuda e Yokoya, 1994; Zhang, 1994; Weik, 1997; Eggert et al., 1998; Okatani e Deguchi, 2002; Park e Murali, 2003; Stamos e Leordeanu, 2003; Barnea e Filin, 2008; Kim et al., 2009; entre outros). Bendels et al. (2004), Al-Manasir e Fraser (2006), Ellekilde et al. (2007) e May et al. (2009) desenvolveram sistemas de mapeamento 3D de ambientes internos integrando sensores de imageamento ativo e passivo para produzirem dados RGB-D. Apesar dos sensores de imageamento ativo propiciarem dados de profundidade com alta precisão, esses equipamentos são de alto e médio custo, de baixa flexibilidade e relativamente pesados.

Em novembro de 2010 a Microsoft lançou o dispositivo Kinect para fins de entretenimento e jogos computacionais da Microsoft Xbox 360. Este dispositivo foi desenvolvido pela empresa *PrimeSense* em colaboração com a Microsoft. O Kinect é um equipamento leve, flexível e de baixo custo (seu valor comercial gira em torno de R\$ 350,00) quando comparado com sistemas laser *scanning* terrestre e câmeras de distância 3D. O dispositivo Kinect é composto, basicamente, por três sensores: dois sensores CMOS (*Complementary Metal-Oxide-Semiconductor*) que registram energia eletromagnética na faixa do espectro correspondente ao

visível (RGB) e infravermelho próximo (IR); e um emissor LASER (*Light Amplification by Stimulated Emission of Radiation*) infravermelho. Os sensores RGB e IR capturam cenas com 640 x 480 pixels em uma taxa de 30 quadros por segundo (fps), sendo que cada quadro (*frame*) capturado pode conter até 300.000 pontos. Devido a esta característica o Kinect pode propiciar uma nuvem de pontos 3D colorida com alta riqueza de informações visuais. Vale ressaltar que a sinergia provocada pela integração de dados RGB-D pode ser efetivamente empregada no mapeamento 3D tornando-se uma vantagem no processo de registro de nuvens de pontos.

Neste trabalho é introduzido um fluxo de mapeamento 3D de ambientes internos. O método proposto é uma extensão do trabalho apresentado por (Khoshelham et al., 2013). O método é baseado na abordagem ponto-a-ponto e explora a integração de dados RGB-D oriundos do dispositivo Kinect. Cinco etapas principais envolvidas no desenvolvimento do método proposto são discutidas. Primeiramente, pontos são detectados no par de imagens RGB empregando o algoritmo SIFT (*Scale Invariant Feature Transform*), proposto por (Lowe, 2004), e suas correspondências são automaticamente estabelecidas. Na segunda etapa é proposto uma normalização das imagens RGB e IR para associação rigorosa dos pontos homólogos encontrados no par de imagens RGB e seus correspondentes na imagem D. Na terceira etapa as coordenadas XYZ de cada ponto associado são calculadas. A quarta etapa consiste em aplicar a detecção e remoção de *outliers* e calcular os parâmetros de transformação (3 rotações e 3 translações) usando o modelo de corpo rígido 3D e o Método dos Mínimos Quadrados (MMQ). Finalmente, é apresentado um modelo linear para a análise da consistência global.

Este trabalho está organizado da seguinte forma: na seção 2 são discutidos os trabalhos relacionados com o tema proposto; a seção 3 descreve o método proposto; a seção 4 apresenta os experimentos e discute os resultados obtidos; e na seção 5 são apontadas as conclusões e as recomendações futuras.

## 2. Trabalhos relacionados

Basicamente, um sistema especialmente desenvolvido para mapeamento 3D de ambientes internos é dividido em três tarefas: o registro dos pares de nuvem de pontos 3D; a detecção do erro de fechamento; e a análise da consistência global de todos os registros executados. A primeira tarefa consiste em determinar os parâmetros de transformação entre os pares consecutivos de nuvem de pontos 3D. O problema do erro de fechamento consiste em detectar lugares anteriormente visitados pelo sensor, enquanto a análise da consistência global tem como objetivo refinar o conjunto dos parâmetros de transformação obtidos na primeira etapa de mapeamento.

O método de registro de nuvem de pontos 3D mais popular no meio acadêmico é conhecido como ICP (*Iterative Closest Point*). O método foi originalmente desenvolvido por Besl e Mckay (1992). Os autores apresentaram um algoritmo que registra iterativamente pares de nuvem de pontos usando pontos pseudo conjugados. Os parâmetros de transformação são obtidos através de uma função erro que minimiza o somatório da distância entre os pontos na nuvem de referência e de pesquisa. Apesar da eficiência na determinação dos parâmetros de transformação usando pares de nuvem de pontos 3D oriunda de qualquer sistema de imageamento, o algoritmo é computacionalmente lento, depende da entrada de valores iniciais aproximados e alta sobreposição entre os pares de nuvem de pontos.

A partir do surgimento dos sistemas SLAM diversos autores desenvolveram métodos rápidos e robustos para o emprego do erro de fechamento e da análise de consistência global. Geralmente, os métodos de erro de fechamento e de análise da consistência global empregados em problemas de SLAM são: o método de otimização baseado em árvore de busca (*tree-based network optimizer-TORO*); e o método de otimização baseado em grafos (*general framework for graph optimization-G2O*). Maiores detalhes ver (Grisetti et al., 2007). Também vale citar o método de ajustamento por feixes perspectivos (*SBA Sparse Bundle Adjustment*) proposto por (Lourakis e Argyros, 2009). Esses métodos são rápidos e robustos, mas o conjunto de parâmetros de transformação não pode ser refinado simultaneamente.

Com o surgimento de câmeras RGB-D May et al.(2009), Steinbrucker et al.(2011), Bachrach et al. (2012), Dryanovski et al. (2012), Endres et al. (2012) e Henry et al. (2012) apresentaram abordagens similares para mapeamento 3D de ambientes internos. O ponto mais crítico dos métodos desenvolvidos pelos autores supracitados é a forma como é feita a associação dos pares de pontos RGB (pontos 2D) com seus correspondentes na imagem de profundidade (pontos 3D). Devido ao desalinhamento angular e linear existente entre os sensores RGB e IR, nos dispositivos Kinect, a etapa de associação entre os pontos 2D-3D não pode ser feita de forma direta. A falta de rigidez nesta etapa provoca incertezas no estabelecimento das correspondências e, conseqüentemente, na obtenção da trajetória do sensor. Khoshelham et al. (2013) usaram os conceitos de geometria epipolar e linhas epipolares para associar pontos 2D-3D seguida da determinação dos parâmetros de rotação e translação entre as imagens RGB e IR através de uma resseção espacial de imagens. Os autores mostraram que a qualidade do estabelecimento das correspondências melhora significativamente com o método proposto. Chow et al. (2014) desenvolveram um sistema de mapeamento 3D móvel para ambientes internos. Os autores integraram duas câmeras RGB-D com um sistema *laser scanning* terrestre e propuseram o uso de dados RGB apenas em situações onde valores de profundidade não são capturados pelo sensor Kinect. Segundo os autores, a solução proposta não está sujeita aos erros provocados pela falta de sincronismo no momento da aquisição de dados oriundos dos sensores RGB e IR. No entanto, o sucesso desta abordagem depende do emprego de sensores de navegação IMU (*Inertial Measurement Unit*).

### 3. Método

Considerando que os parâmetros de orientação interior dos sensores RGB e IR foram devidamente calibrados e o desalinhamento angular (*angular misalinhament*), bem como a distância entre as origens dos sistemas referenciais (*lever arm*) de ambas as câmeras também foram previamente determinadas (Khoshelham e Elberink, 2012), o método proposto está dividido em 5 etapas principais, como segue:

1. Extração e detecção das primitivas pontuais no par de imagens RGB e o estabelecimento automático de suas correspondências;
2. Geração dos pares de nuvem de pontos 3D;
3. Associação dos pares de pontos detectados no par de imagens RGB com seus pontos correspondentes na nuvem de pontos 3D;
4. Detecção e remoção de *outliers* e determinação dos parâmetros de transformação entre os pares de nuvem de pontos 3D;

## 5. Análise da consistência global e reconstrução 3D do ambiente mapeado.

**3.1 Extração e detecção das primitivas pontuais nos pares de imagens RGB**

O SIFT é o algoritmo mais empregado em Visão Computacional para detecção de pontos de interesse. É importante ressaltar que o SIFT não é apenas escala invariante, uma vez que mesmo variando a rotação, a iluminação e o ponto de vista da imagem de referência em relação à imagem de pesquisa, ainda são encontrados resultados aceitáveis. A principal ideia deste algoritmo é reduzir o custo computacional através de operações executadas apenas em pontos potenciais. De forma geral, o algoritmo procura pontos invariantes à mudança de escala da imagem utilizando uma função Gaussiana (espaço-escala). Primeiramente, o espaço-escala é definido como uma convolução da imagem  $I(u, v)$  com um operador Gaussiano  $G(u, v, \sigma)$ , como segue (Lowe, 2004):

$$L(u, v, \sigma) = G(u, v, \sigma) * I(x, y) \quad (1)$$

sendo,  $L$  a imagem suavizada,  $u$  e  $v$  as coordenadas do pixel na imagem RGB,  $*$  representa o operador de convolução,  $\sigma$  a variação de escala em  $G$  e  $G(u, v, \sigma) = \frac{1}{2\pi\sigma^2} \cdot e^{-\frac{(u^2+v^2)}{2\sigma^2}}$

O algoritmo proposto por Lowe (2004) realiza cálculos de diferença de Gaussianas (DoG). Esta função é separada por uma escala constante ( $\lambda$ ) e definida pela diferença  $L$  em escalas próximas, como segue:

$$DoG(u, v, \sigma) = L(u, v, \lambda \cdot \sigma) - L(u, v, \sigma) \quad (2)$$

A Equação 2 proporciona um valor aproximado para a escala normalizada Laplaciana-Gaussiana ( $\sigma^2 \nabla^2 G$ ), como segue (Lowe, 2004):

$$\sigma^2 \nabla^2 G = \frac{\partial G}{\partial \sigma} \approx \frac{L(u, v, \lambda \cdot \sigma) - L(u, v, \sigma)}{\lambda \cdot \sigma - \sigma} \quad (3)$$

e, por conseguinte:

$$L(u, v, \lambda \cdot \sigma) - L(u, v, \sigma) \approx (\lambda - 1) \cdot \sigma^2 \nabla^2 G \quad (4)$$

O fator  $\sigma^2 \sigma^2$  faz com que a escala seja invariante. Em seguida, a localização dos pontos de interesse é feita por aproximações de  $\sigma^2 \nabla^2 G$ . Conforme mostrado por Lowe (2004) o fator remanescente  $(\lambda - 1)(\lambda - 1)$  não influencia na posição do pixel detectado. O emprego da função

gaussiana tem o objetivo de obter imagens com ruídos atenuados e bordas realçadas. O resultado esperado é um conjunto de 4 (quatro) oitavas, sendo realizadas 5 (cinco) operações Gaussianas para cada oitava.

Em seguida, os valores de máximo ou mínimo locais (pontos extremos) devem ser detectados para cada DoG obtido. Este procedimento é baseado na comparação de todos os pixels de borda na imagem DoG com seus vizinhos (vizinhança 8), do espaço-escala corrente, em relação aos vizinhos correspondentes no espaço-escala adjacentes superior e inferior. Posteriormente, os pontos de interesse devem ser localizados e se não forem instáveis, devem ser descartados.

A localização dos pontos extremos é realizada através do ajustamento de uma função quadrática 3D. Dado um ponto de amostragem sua localização interpolada do máximo é determinada através da expansão da série de Taylor da função DoG, a saber (Lowe, 2004):

$$DoG(\Delta) = DoG + \frac{\partial DoG}{\partial \Delta} \Delta + \frac{1}{2} \Delta^T \cdot \frac{\partial^2 DoG}{\partial \Delta^2} \Delta \dots \quad (5)$$

sendo,  $\Delta = (u, v, \sigma)^T$

Na Equação 5 a imagem  $DoG(u, v, \sigma)$  é deslocada ( $\Delta$ ) de tal forma que sua origem esteja localizada no ponto de amostragem. Enfim, a localização do extremo pode ser calculada derivando a Equação 5 em função de ( $\Delta$ ) e igualando o resultado a zero, como segue (Lowe, 2004):

$$\Delta = - \frac{\partial^2 DoG^{-1}}{\partial \Delta^2} \cdot \frac{\partial DoG}{\partial \Delta} \quad (6)$$

Para cada ponto máximo e seus pixels vizinhos é calculada sua magnitude ( $mag_{uv}$ ) e orientação ( $\alpha_{uv}$ ), como segue:

$$mag_{uv} = \sqrt{\Delta u^2 + \Delta v^2} \quad (7)$$

$$\alpha_{uv} = \arctan\left(\frac{\Delta v}{\Delta u}\right) \quad (8)$$

Em seguida, para cada ponto extremo e seus pixels vizinhos é gerado um histograma de orientações. Os pixels vizinhos são adicionados ao histograma com pesos correspondentes aos valores de  $mag_{uv}$  e também aos valores determinados por uma função Gaussiana. Esta função Gaussiana é determinada por uma janela de suavização Gaussiana com  $\sigma$  igual à metade da largura da janela do descritor da função. O valor dos pesos atualizados para cada pixel vizinho é dado por Lowe (2004):

$$w' = w + \theta \cdot mag_{uv} \cdot Gs(\Delta u, \Delta v, \sigma') \quad (9)$$

sendo,  $w'w'$  a atualização de  $w$  e  $Gs(\Delta u, \Delta v, \sigma')$  a janela Gaussiana.

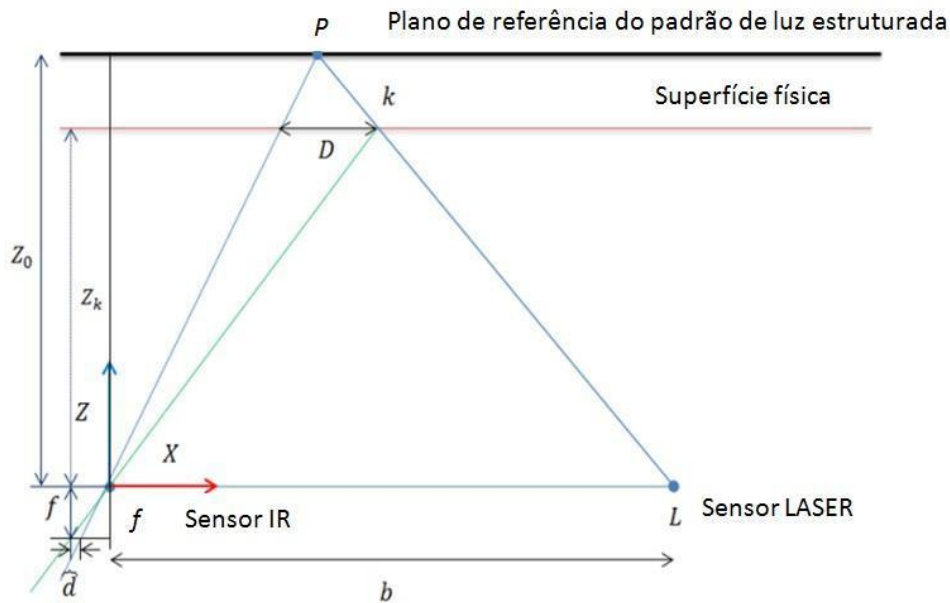
Os valores obtidos são normalizados e para cada ponto de interesse é atribuído um descritor invariante à iluminação, ao ponto de vista e à rotação. O descritor consiste de um vetor contendo os valores do histograma com 128 elementos determinados em função de uma janela de apoio

com um total de 16x16 pixels. O resultado esperado é um conjunto de descritores que podem ser usados no estabelecimento automático de correspondências. 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.

### 3.2 Geração dos pares de nuvens de pontos 3D

De acordo com Freedman et al. (2010) o princípio de funcionamento do dispositivo Kinect para determinação dos valores de profundidade é conhecido como triangulação 3D. Basicamente, o sensor LASER emite um pulso na região do espectro correspondente ao infravermelho e um padrão de luz estruturada é criado a partir de um processo de difração da luz. O padrão de luz estruturada é usado para marcar a distância entre o sensor e a superfície física ( $Z_0$ ) que é, previamente, armazenada na memória do dispositivo. Em outras palavras, é criado um plano de referência. Um ponto ( $P$ ) qualquer do padrão de luz estruturada, contido no plano de referência, é registrado na imagem IR com sua posição deslocada (ver o ponto  $k$  na Fig. 1). Para resolver o problema de correspondência, cada pixel na imagem IR é comparado com os pixels do padrão de luz estruturada (armazenado anteriormente) através de uma convolução em torno da sua vizinhança, cujo tamanho da janela de correlação pode variar de 7x7, 9x9 ou 9x7. O pixel na imagem IR apresentando a maior correspondência dentro da janela de correlação é selecionado como o pixel correspondente. O deslocamento entre o pixel de referência do padrão de luz estruturada e seu correspondente na imagem IR indica o valor de paralaxe ( $d$ ). Desta forma, o valor de profundidade de cada pixel na imagem IR pode ser obtido em função de  $d$ , da distância focal do sensor IR, da linha de base ( $b$ ) e de  $Z_0$ .

A Figura 1 apresenta a geometria do modelo matemático, proposto por Khoshelham e Elberink (2012), para calcular os valores de profundidade ( $Z_K$ ) e as coordenadas  $X_K$ ,  $Y_K$  de cada pixel na imagem IR. Na Figura 1 percebe-se a relação entre a distância de um ponto  $P$  ao sensor, relativo ao plano de referência do padrão de luz estruturada e a medida de paralaxe. O sistema referencial do sensor IR é um sistema tridimensional com origem no centro perspectivo (CP) da câmera IR. O eixo  $Z$  é ortogonal ao plano da imagem apontando para a superfície física, enquanto o eixo  $X$  é perpendicular ao eixo  $Z$  na direção da linha de base determinada entre o sensor IR e o emissor LASER. O eixo  $Y$  é ortogonal aos eixos  $X$  e  $Z$  tornando o sistema de referencial dextrógiro.



**Figura 1:** Geometria do modelo proposto.

Fonte: Khoshelham e Elberink (2012).

Na Figura 1, por semelhança de triângulos, tem-se (Khoshelham e Elberink, 2012):

$$\frac{D}{b} = \frac{Z_0 - Z_k}{Z_0} \quad (10)$$

$$\frac{d}{f} = \frac{D}{Z_k} \quad (11)$$

sendo,  $f$  a distância focal da câmera IR e  $d$  a paralaxe do ponto.

Na Equação 10  $Z_k$  denota a distância (profundidade) do ponto O no espaço-objeto e  $D$  é o deslocamento do ponto  $P$  provocado pela diferença entre a sua posição ideal no plano de referência e sua posição real no espaço-objeto. Substituindo  $D$  da Equação 11 na Equação 10, tem-se:

$$Z_k = \frac{Z_0}{1 + \frac{Z_0}{f \cdot b} \cdot d} \quad (12)$$

Khoshelham e Elberink (2012) expressaram  $d$  em função de uma normalização linear. Neste caso,  $d$  (Equação 12) é reescrito na forma  $m \cdot d' + nm \cdot d' + n$ , sendo  $mm$  e  $nn$  os parâmetros de uma normalização linear (na verdade uma denormalização) e  $d'd'$  a paralaxe normalizada. Substituindo os termos apresentados na Equação 12 e fazendo algumas manipulações matemáticas, tem-se:

$$Z_k = \frac{1}{\left(\frac{m}{f \cdot b}\right) \cdot d' + \left(\frac{Z_0^{-1} \cdot n}{f \cdot b}\right)} \quad (13)$$



A Equação 13 expressa uma relação linear entre o inverso da profundidade de um ponto e seu correspondente valor de paralaxe normalizado, cujos coeficientes podem ser estimados pelo MMQ (Khoshelham e Elberink, 2012). Desta forma, a Equação 13 pode ser usada para calcular os valores de profundidade.

A coordenada  $Z_k Z_k$  do ponto, juntamente com a distância focal calibrada da câmera IR, determinam a escala para o ponto  $k$  no espaço-objeto. As coordenadas planimétricas podem ser obtidas por meio das seguintes equações (Khoshelham e Elberink, 2012):

$$X_k = \frac{Z_k}{f} \cdot x_c' \quad (14)$$

$$Y_k = \frac{Z_k}{f} \cdot y_c' \quad (15)$$

sendo,  $x_c', y_c', x_c', y_c'$  as coordenadas do ponto no espaço-imagem corrigidas das distorções radial simétrica e descentrada do sistema de lentes da câmera IR, cujas equações podem ser encontradas em (Brown, 1971).

### 3.3 Associação entre as primitivas RGB e seus correspondentes na nuvem de pontos 3D

Cada par de ponto detectado na imagem RGB deve ser associado ao seu ponto correspondente na nuvem de pontos 3D. Esta tarefa pode ser realizada a partir dos valores dos parâmetros de orientação interior calibrados dos sensores RGB e IR, bem como dos valores de desalinhamento angular ( $d\omega$ ,  $d\phi$ ,  $d\kappa$ ) e a distância entre as origens ( $dx$ ,  $dy$ ,  $dz$ ) dos sistemas referenciais dos sensores RGB e IR.

Como existe um desalinhamento angular entre as imagens RGB e IR, neste trabalho, é proposto uma normalização da imagem IR. Esta etapa do método é sumamente importante, tendo em vista que, a falta de correção dos desalinhamentos existentes entre os sensores RGB e IR provoca incertezas no estabelecimento das correspondências e, conseqüentemente, na determinação dos parâmetros de transformação. A normalização das imagens consiste em 7 etapas, a saber:

1. Primeiramente, a normalização da imagem é feita usando as Equações 16 e 17, como segue (Mikhail et al., 2001):

$$x_n = -f \cdot \frac{(r_{11}x_c' + r_{12}y_c' - r_{13}f)}{(r_{21}x_c' + r_{22}y_c' - r_{23}f)} \quad (16)$$

$$y_n = -f \cdot \frac{(r_{21}x_c' + r_{22}y_c' - r_{23}f)}{(r_{31}x_c' + r_{32}y_c' - r_{33}f)} \quad (17)$$

sendo,  $x_n, y_n$  e  $x'_n, y'_n$  as coordenadas normalizadas no sistema referencial com origem no centro da imagem IR,  $f$  e  $f'$  as distâncias focais calibradas dos sensores IR e RGB, respectivamente,  $r_{ij}$  os elementos da matriz de rotação composta pelos ângulos  $d\omega, d\phi, d\kappa$ , para  $i=j=1, \dots, 3$ ;

2. 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 na imagem IR. A projeção da linha é feita em função do valor de profundidade máximo e mínimo de alcance do dispositivo ( $Z_{min} = 40cm$  e  $Z_{max} = 7m$ ), do valor da linha base ( $b$ -calculado usando  $dx, dy, dz$ ) e o valor de profundidade  $Z_k$  do ponto (calculado usando a Eq. 13);
3. Para todos os pixels dentro do intervalo da linha epipolar são calculadas suas coordenadas 3D (usando as Eqs. 13-15) e reprojetaos para o sistema referencial digital da imagem RGB através das Equações 18 e 19;
4. Então, é verificado se o pixel transformado na etapa 3 corresponde ao pixel extraído pelo algoritmo SIFT na imagem RGB;
5. 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;
6. As etapas 1-5 são repetidas para todos os pontos detectados na imagem RGB.

$$u = f' \cdot \left( \frac{x_k}{z_k} \right) - \left( \frac{x'_a}{r_p} \right) \quad (18)$$

$$v = f' \cdot \left( \frac{y_k}{z_k} \right) - \left( \frac{y'_a}{r_p} \right) \quad (19)$$

sendo,  $u$  e  $v$  as coordenadas dos pixels reprojetaos para a imagem IR normalizada,  $T_p$  o tamanho do pixel do sensor IR e  $x'_a, y'_a$  as coordenadas no sistema referencial do centro da imagem IR adicionadas dos erros sistemáticos (ponto principal e distorções das lentes da câmera IR).

Finalmente, é obtido um conjunto de pontos associados à nuvem de pontos 3D. Vale lembrar que falsos positivos são estabelecidos na etapa de correspondência automática. Esses pontos são conhecidos como *outliers* e devem ser detectados e removidos do processamento. Esta tarefa é importante, uma vez que devido as características de aquisição de dados do dispositivo Kinect, a etapa de determinação dos parâmetros de transformação é extremamente sensível a *outliers*.

Neste trabalho, a etapa de detecção e remoção de *outliers* é realizada estatisticamente através do método denominado *interquartile range* (IQR). O IQR tem como objetivo encontrar uma medida de variação de um dado conjunto de  $m_i$  pontos 3D baseado em  $\frac{1}{4}$  da amostra. Basicamente, o IQR divide a amostra de dados em quatro partes iguais, sendo  $Q1$  e  $Q3$  os valores de mediana das metades inferior (25%) e superior (75%) da amostra, respectivamente, e  $Q2$  o valor de mediana calculada (50%). Os coeficientes supracitados podem ser calculados como segue:

$$Q1 = \frac{1}{4} \cdot (n + 1) \quad (20)$$

$$Q2 = \frac{1}{2} \cdot (n + 1) \quad (21)$$

$$Q3 = \frac{3}{4} \cdot (n + 1) \quad (22)$$

sendo,  $n$  a quantidade de dados da amostra.

Usando as Equações 20-23 podem ser calculados os valores máximo ( $Outlier_{max}$ ) e mínimo ( $Outlier_{min}$ ) na amostra, como segue:

$$Outlier_{max} = Q3 + Coef_{IQR} \cdot IQR \quad (24)$$

$$Outlier_{min} = Q1 - Coef_{IQR} \cdot IQR \quad (25)$$

sendo,  $Coef_{IQR}$  o valor da metade superior da amostra admitida.

Finalmente, os pontos 3D correspondentes com valores calculados entre  $Outlier_{max}$  e  $Outlier_{min}$  são considerados *inliers* (correspondências verdadeiras) os demais valores são considerados *outliers* (falsos positivos) e devem ser removidos do conjunto de dados.

### 3.4 Associação entre as primitivas RGB e seus correspondentes na nuvem de pontos 3D

Dado um conjunto de pontos 3D correspondentes um modelo matemático rigoroso deve ser elaborado para determinar os parâmetros de transformação entre cada par de nuvens de pontos 3D. Neste trabalho, o modelo empregado minimiza o somatório dos erros ( $E$ ), como segue:

$$E_j = \sum_{i=1}^n \|X_{i,j-1} - (R \cdot X_{i,j} + T)\| \quad (26)$$

sendo,  $X_{i,j-1}$  e  $X_{i,j}$  o par de coordenadas de pontos 3D correspondentes nas nuvens de pontos de pesquisa ( $i$ ) e de referência ( $j$ ),  $T$  o vetor dos parâmetros de translação ( $t_x, t_y, t_z$ ) e  $R$  a matriz dos parâmetros de rotação em função dos ângulos de Euler ( $\omega, \phi, \kappa$ ).

A Equação (26) não é linear e envolve 6 parâmetros a serem determinados pelo MMQ. Neste caso, deve ser aplicado o modelo paramétrico não linear (Wolf e Ghiliani, 1997). Valores iniciais são obtidos antes da detecção de *outliers* através de um conjunto de 6 pontos 3D, aleatoriamente sorteados na amostra dos dados.

### 3.5 Associação entre as primitivas RGB e seus correspondentes na nuvem de pontos 3D

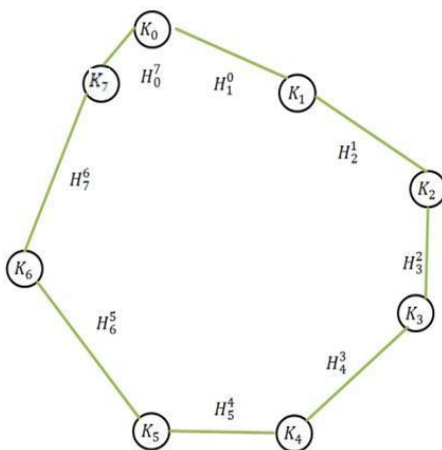
Neste trabalho é proposta uma abordagem baseada no ajustamento simultâneo dos parâmetros de transformação para refinar o conjunto dos valores de rotação e translação relativos, obtidos na etapa do registro de pares de nuvem de pontos 3D. É importante destacar que, como o processamento não é feito em tempo real, a tarefa de detecção de lugares revisitados (erro de fechamento) é negligenciada da etapa de análise da consistência global.

Dada uma trajetória do sensor formada por um conjunto de  $K$  estações ocupadas pelo Kinect ( $K_0, \dots, K_n$ ), como mostra a Figura 2, o novo conjunto de parâmetros é estimado simultaneamente baseado na seguinte equação proposta, a saber:

$$F: H_{\text{nov}}^i = H_{j+1}^{i+1} - H_j^i \tag{27}$$

sendo,  $FF$  o modelo funcional,  $H_{\text{nov}}^i H_{\text{nov}}^i$  o novo conjunto de parâmetros de transformação ajustado,  $H_{j+1}^{i+1} H_{j+1}^{i+1}$  a matriz homogenia dos parâmetros de transformação da estação posterior e  $H_j^i H_j^i$  a matriz homogenia dos parâmetros de transformação da estação anterior.

Neste caso, as observações são os valores de rotação e translação determinados na etapa de registro entre as nuvens de pontos 3D. Na Equação 27 cada estação ocupada pelo sensor está representada pelos parâmetros determinados em relação à estação atual e anterior, ou seja,  $K_{ij} = [\omega_{ij}, \phi_{ij}, \kappa_{ij}, tx_{ij}, ty_{ij}, tz_{ij}]^T$   $K_{ij} = [\omega_{ij}, \phi_{ij}, \kappa_{ij}, tx_{ij}, ty_{ij}, tz_{ij}]^T$ , sendo  $i = 0, \dots, K$  e  $j = i+1$ .



**Figura 2:** Trajetória e estações ocupadas pelo sensor Kinect.

O modelo proposto (ver Equação 27) está em função apenas de observações. Deste modo, deve ser empregado o modelo paramétrico linear do MMQ adaptado para o método de ajustamento livre. A matriz das derivadas parciais ( $AA$ ) é obtida em relação à  $K_{isen}$  e  $K_{isen}$ , enquanto a matriz dos pesos ( $PP$ ) corresponde à matriz variância covariância (MVC) dos parâmetros de transformação determinados, anteriormente, na etapa do registro de pares de nuvem de pontos 3D.

Uma vez que as estações ocupadas pelo dispositivo são globalmente ajustadas, pode ser realizada a reconstrução 3D da cena. Para cada frame é gerada uma nuvem de pontos 3D a partir dos dados RGB-D. 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 ( $isen$ ), deve-se montar uma matriz homogênea que contém os valores ajustados de rotação e translação ( $H_{isen}^{isen} H_{isen}^{isen}$ ), como segue:

$$H_{K_{isen}}^{K_{isen}} = \begin{bmatrix} R_{ij}^K & T_{xyz}^K \\ 0 & 1 \end{bmatrix} \tag{28}$$

sendo,  $H_{K_{j\text{sen}}}^{K_{i\text{sen}}} H_{K_{j\text{sen}}}^{K_{i\text{sen}}}$  a matriz de transformação relativa de cada par de nuvem de pontos devidamente ajustada,  $R_{ij}^K R_{ij}^K$  e  $T_{xyz}^K T_{xyz}^K$  a matriz de rotação e o vetor de translação relativa de cada posição do sensor, respectivamente.

A posição do dispositivo Kinect ( $Pr$ ) pode ser calculada combinando a seqüência de transformações, a saber:

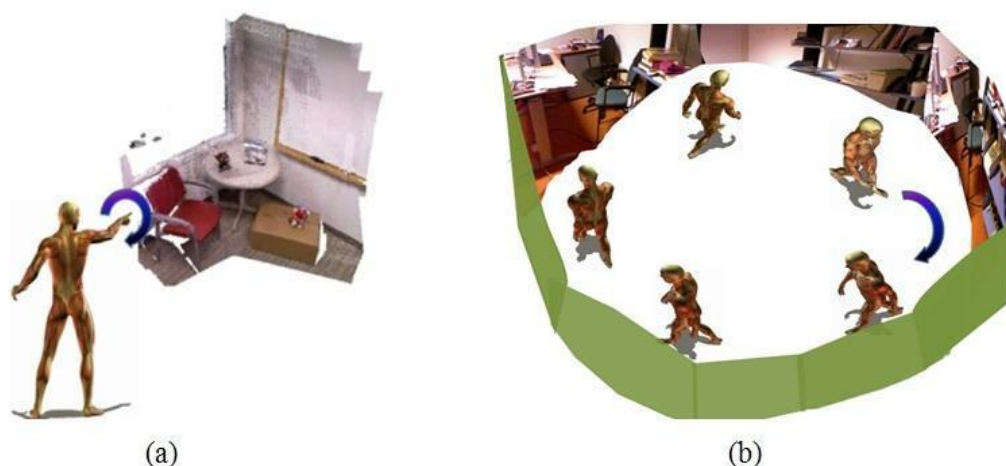
$$P_r = \prod_{i\text{sen}}^0 H_{K_{j\text{sen}}}^{K_{i\text{sen}}} \cdot H_{K_{i\text{sen}}}^{K_{i\text{sen}}} \quad (29)$$

sendo,  $H_{K_{i\text{sen}}}^{K_{i\text{sen}}} H_{K_{i\text{sen}}}^{K_{i\text{sen}}}$  a matriz homogênea que representa a posição inicial do sensor com valores de rotação e translação iguais a 0 (zero).

Finalmente, todos os pontos transformados são concatenados para reconstrução 3D do ambiente mapeado. O método é simples, uma vez que não é levada em consideração a sobreposição dos pontos e também as variações de iluminação.

## 4. Experimentos e análise dos resultados

Para o desenvolvimento deste trabalho foram empregadas diversas bibliotecas computacionais de código aberto (*open-source*), tais como, o OpenCV (*Open-Source Computer Vision*), as bibliotecas Eigen e FLANN (*FastLibrary for Approximate Neighbours*) e o VTK (*Visualization Tool Kit*). Os algoritmos foram implementados em linguagem de programação GCC ambiente *CodeBlocks* para Linux. A aquisição dos dados foi feita de duas formas diferentes. A primeira forma de aquisição dos dados é denominada *hand-motion*. Neste caso, o operador se mantém numa posição fixa e a trajetória do sensor é obtida através do movimento de uma das mãos formando um círculo completo. Na segunda forma de aquisição dos dados, o operador percorre um ambiente interno carregando nas mãos ou em qualquer outra plataforma o dispositivo Kinect, sendo este processo conhecido como *stop-and-go*. A Figura 3 mostra os processos de aquisição de dados conhecidos como *hand-motion* e *stop-and-go*.



**Figura 3:** Processos de aquisição de dados para mapeamento 3D de ambientes internos. (a) Processo de aquisição *hand-motion*; (b) Processo de aquisição *stop-and-go*.

Em geral, o processo de aquisição de dados conhecido como *hand-motion* é usado em métodos para preenchimento de dados tridimensionais em ambientes internos que não foram completamente mapeados, seja usando sistemas *laser scanning* terrestre ou qualquer outro dispositivo. Já o processo conhecido como *stop-and-go* é empregado em todos os métodos para mapeamento 3D de ambientes internos.

## 4.1 Experimentos

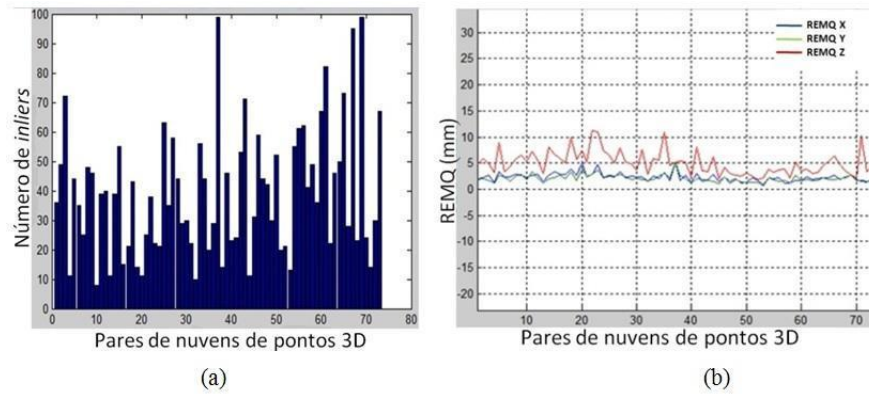
Neste trabalho foram conduzidos 4 experimentos para avaliar o método proposto. Os parâmetros de orientação interior das câmeras e os valores dos parâmetros  $mm$ ,  $nn$ ,  $Z_0Z_0$ ,  $bb$ , bem como dos valores do desalinhamento angular e o deslocamento entre as origens dos sistemas referenciais dos sensores RGB e IR podem ser encontrados em Khoshelham e Elberink (2012). A Tabela 1 mostra as características de cada um dos ensaios produzidos.

**Tabela 1:** Características dos experimentos produzidos.

Experimento	Processo de aquisição	Número de frames	Taxa de aquisição
1	<i>Hand-motion</i>	74	25 fps
2	<i>Hand-motion</i>	60	20 fps
3	<i>Hand-motion</i>	124	30 fps
4	<i>Hand-motion</i>	324	30 fps

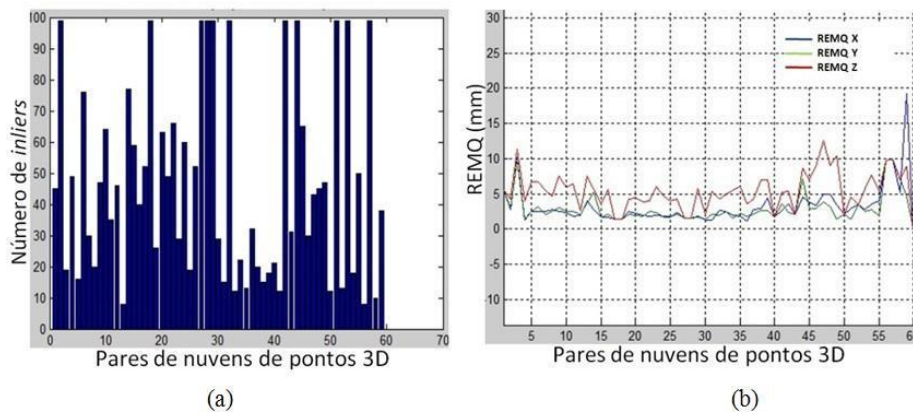
De acordo com a Tabela 1, no primeiro, segundo e terceiro experimentos os dados foram adquiridos empregando o processo *hand-motion*. O processo de aquisição de dados conhecido como *stop-and-go* foi usado apenas no quarto experimento. Cada frame adquirido propicia uma imagem RGB e outra IR (dimensões de 640 x 480 pixels) e um arquivo contendo os valores de paralaxe discretizados em 11 bits.

Para avaliar a acurácia relativa da trajetória do sensor foram calculadas as discrepâncias entre os pontos 3D na nuvem de pontos de pesquisa, obtida após a etapa de registro dos dados, e seus correspondentes na nuvem de pontos de referência (conjunto de *inliers*), a média das discrepâncias e sua raiz do erro médio quadrático (REMQ). A Figura 4 mostra o número total de *inliers* e a raiz do erro médio quadrático da média das discrepâncias, obtidos na etapa de registro de todos os pares de nuvens de pontos 3D para o experimento 1.



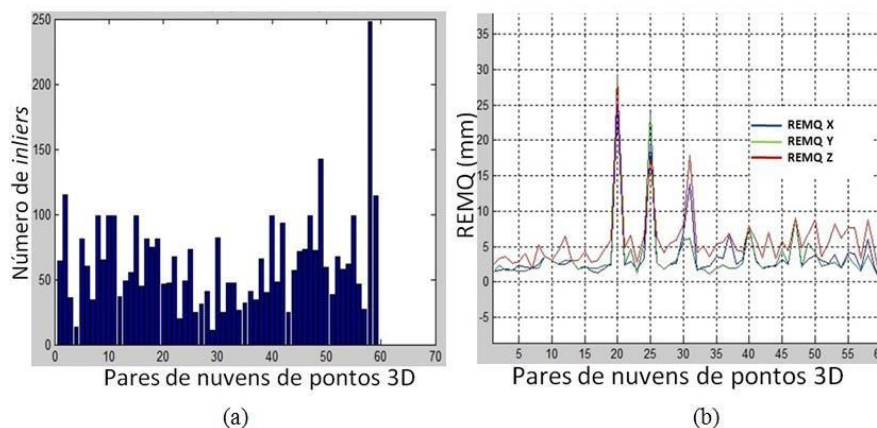
**Figura 4:** Estatísticas obtidas na etapa de registro dos pares de nuvens de pontos 3D para o experimento 1. (a) Número de *inliers*; (b) REMQ da média das discrepâncias.

A Figura 5 mostra o número total de *inliers* e a REMQ da média das discrepâncias, obtidos na etapa de registro de todos os pares de nuvens de pontos 3D para o experimento 2.



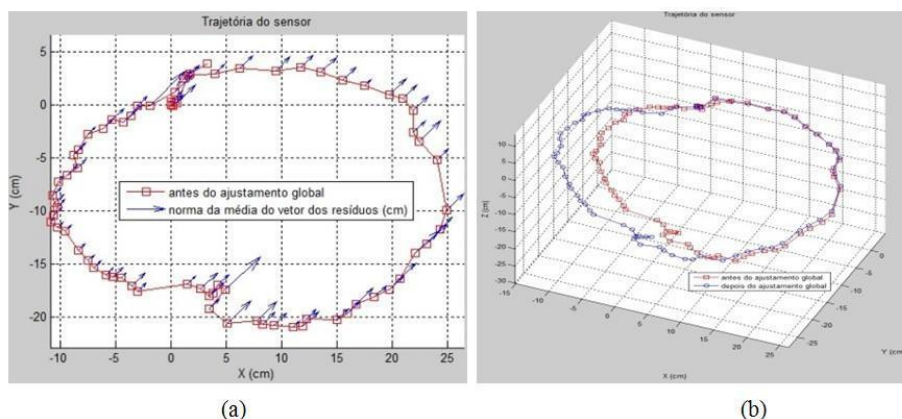
**Figura 5:** Estatísticas obtidas na etapa de registro dos pares de nuvens de pontos 3D para o experimento 2. (a) Número de *inliers*; (b) REMQ da média das discrepâncias.

A Figura 6 mostra o número total de *inliers* e a REMQ da média das discrepâncias, obtidos na etapa de registro de todos os pares de nuvens de pontos 3D para o experimento 3.



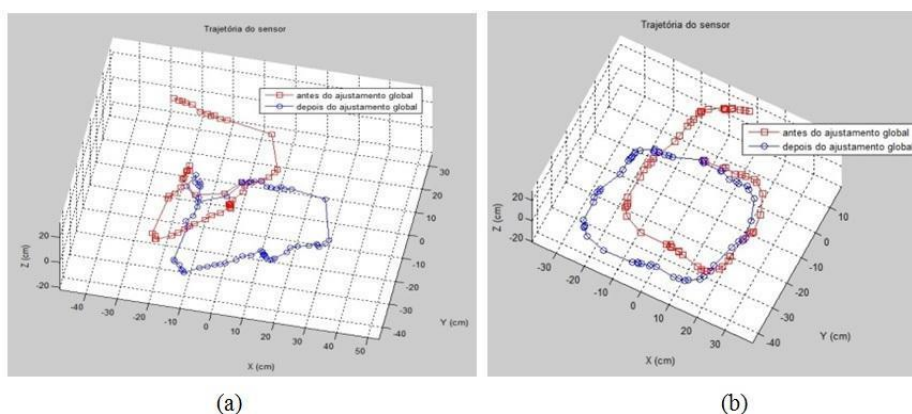
**Figura 6:** Estatísticas obtidas na etapa de registro dos pares de nuvens de pontos 3D para o experimento 3. (a) Número de *inliers*; (b) REMQ da média das discrepâncias.

A Figura 7 mostra a trajetória do sensor, obtida no experimento 1, antes e depois da análise da consistência global e o vetor dos resíduos correspondentes para cada registro dos pares de nuvens de pontos realizado com o método proposto.



**Figura 7:** Trajetória do sensor no experimento 1. (a) trajetória do sensor antes e depois do ajustamento global; (b) norma do vetor dos resíduos.

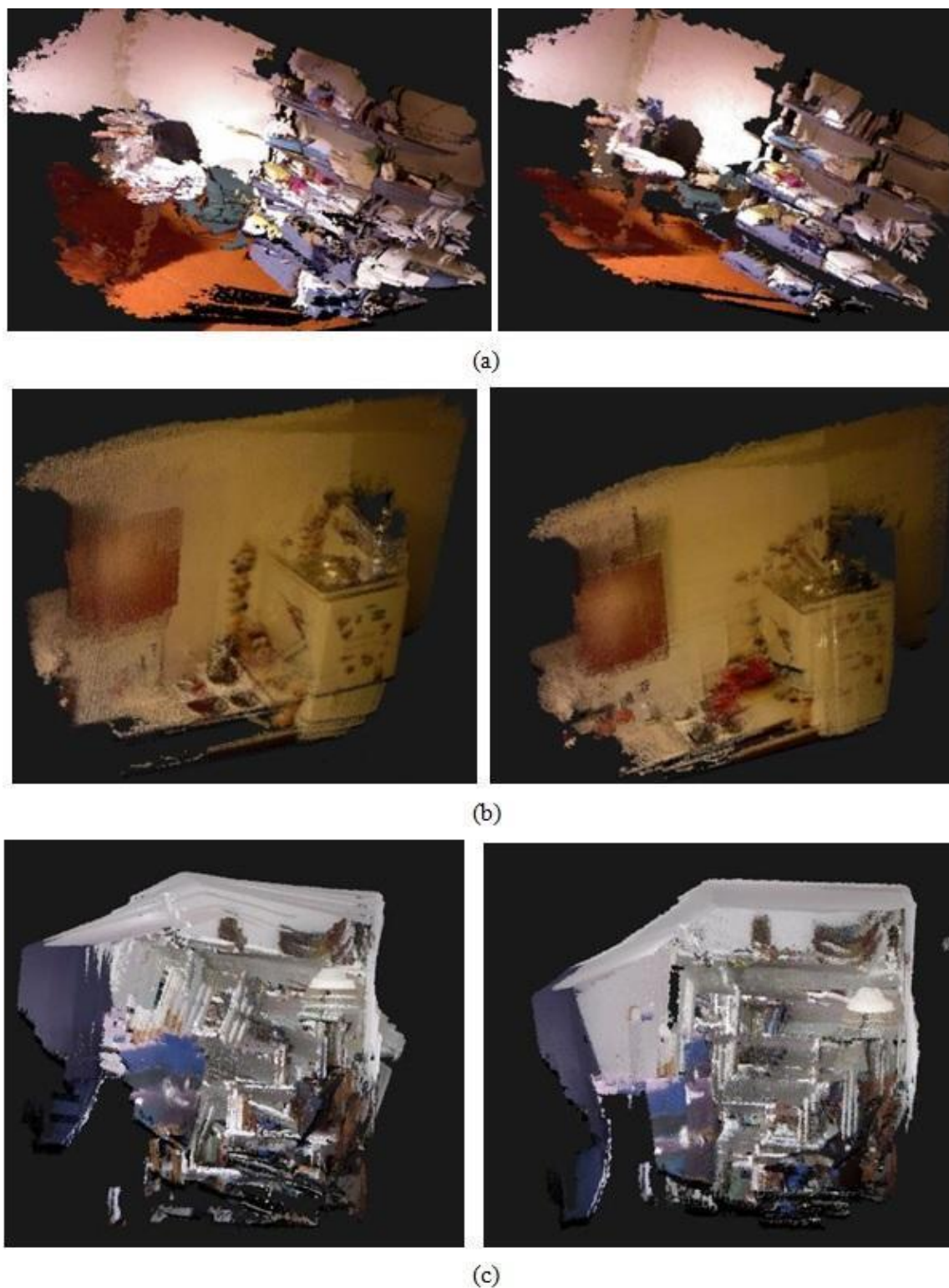
A Figura 8 mostra a trajetória do sensor, obtida nos experimentos 2 e 3 antes e depois da análise da consistência global.



**Figura 8:** Trajetória do sensor antes e depois do ajustamento global. (a) Experimento 2; (b) Experimento 3.

A Figura 9 mostra visualmente os 3 ambientes internos tridimensionalmente modelados antes e após a análise da consistência global.

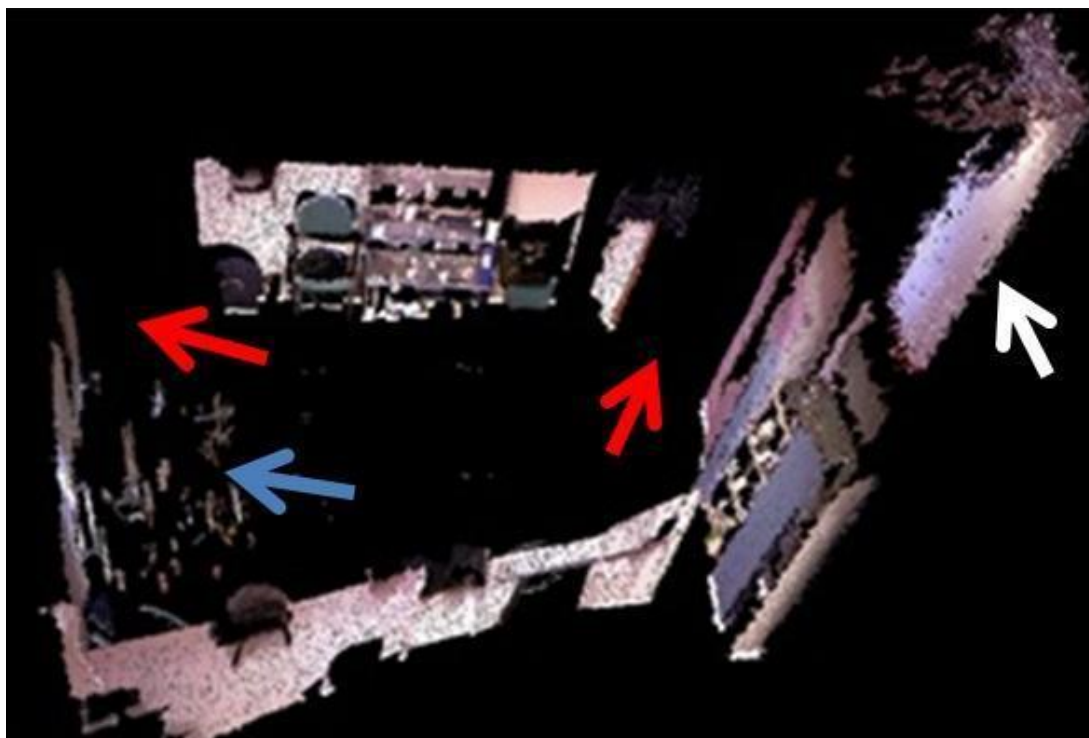




**Figura 9:** Ambientes internos tridimensionalmente modelados antes (à esquerda) e depois (à direita) do ajustamento global. (a) Experimento 1; (b) Experimento 2; (c) Experimento 3.

Com o objetivo de avaliar a eficiência do método proposto em um ambiente interno complexo foi empregado o processo de aquisição de dados *stop-and-go*. A média das discrepâncias e sua

REMQ também foram calculadas. A Figura 10 mostra um ambiente interno complexo mapeado com o método proposto.



**Figura 10:** Ambiente interno tridimensionalmente reconstruído usando o processo de aquisição de dados *stop-and-go*.

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

#### 4.1 Análise dos resultados

Como mostram as Figuras 4-6 quanto maior o número de *inliers* melhor o REMQ, mostrando que a qualidade da acurácia relativa da trajetória do sensor depende do número de *inliers* empregado na etapa de registro dos pares de nuvens de pontos 3D. Especificamente, a Figura 4 mostra o menor valor da REMQ na coordenada X e maior na coordenada Z (6,5 mm e 11 mm, respectivamente). A Figura 5 apresenta o menor valor da REMQ na coordenada Y e maior na coordenada X (2,3 mm e 19 mm) e na Figura 6 é encontrado o menor valor de REMQ na coordenada X e maior na coordenada Z (1,2 mm e 27 mm, respectivamente). A Tabela 2 mostra a média das discrepâncias em XYZ para cada um dos 3 experimentos acima discutidos.

**Tabela 2:** Média das discrepâncias em XYZ.

Experimento	X (mm)	Y (mm)	Z (mm)
1	0,000067	-0,000048	0,00829
2	0,000262	0,000946	0,004700
3	-0,000018	-0,000017	0,037963

Como pode ser observado na Tabela 2, em todos os experimentos foi encontrados valores da média das discrepâncias em XYZ muito próximo de zero (restando apenas erros numéricos), não sendo encontradas tendências significativas, dado que o valor do REMQ é igual ao desvio padrão calculado.

Nos experimentos realizados o maior valor da norma do vetor dos resíduos para o ensaio 1 é de 1,65 cm, enquanto para o ensaio 2 é de 2,72 cm e 2,25 cm para o ensaio 3. Antes do ajustamento global a trajetória do sensor é afetada por erros acumulativos derivados da etapa de registro dos pares de nuvens de pontos 3D. Como pode ser observado na Figura 7b e nas Figuras 8a e 8b, a trajetória do sensor é significativamente melhorada após a análise da consistência global. A Figura 9 mostrou que a qualidade visual da modelagem 3D também depende do ajustamento global, sendo todos os pares de nuvem de pontos ajustados para um mesmo sistema referencial.

Na Figura 10 a seta em azul indica um conjunto de objetos que não foram completamente modelados devido ao tipo de material dos objetos, tais como, material plástico e superfícies especulares. A seta vermelha indica parte do ambiente que não foi modelado por se tratar de uma porta aberta do local mapeado. A seta em branco mostra uma porta do corredor externo ao local mapeado. O ambiente modelado neste experimento obteve uma REMQ da média das discrepâncias em XYZ em torno de 2,5 cm.

## 5. Conclusões e recomendações

Neste trabalho foi apresentada uma proposta para mapeamento 3D de ambientes internos baseado em uma abordagem ponto-a-ponto que explora a integração de dados RGB-D. Cinco aspectos principais devem ser discutidos: a detecção e correspondência automática de pontos nos pares de imagens RGB; a associação entre os pontos RGB e seus correspondentes na nuvem de pontos 3D; o registro dos pares de nuvem de pontos 3D; e a análise da consistência global. Para investigar a potencialidade do método proposto foram realizados 4 experimentos.

Admitindo que a taxa média de aquisição dos dados é em torno de 25 fps a detecção e correspondência automática nos pares de imagens RGB usando o algoritmo SIFT apresentou-se extremamente eficiente, pois praticamente não existe variações em rotação e mudança de ponto de vista. Quando é efetuada a associação entre os pontos SIFT 2D e seus correspondentes na nuvem de pontos 3D usando o método proposto neste trabalho, a correspondência é exata, uma vez que a associação entre os pontos na imagem e na nuvem de pontos 3D depende das compensações angulares e lineares entre as imagens RGB e IR.

Como a aquisição das imagens RGB e IR é feita em tempos ligeiramente diferentes existe um pequeno deslocamento entre as imagens RGB e IR. Desta forma, a equação que otimiza o problema de registro dos pares de nuvem de pontos 3D empregada neste trabalho é altamente

sensível aos *outliers*. A adaptação do método estatístico *interquartile range* para detecção de *outliers* mostrou-se eficiente. No entanto, a detecção e remoção de *outliers* ainda necessitam de maiores investigações.

A principal característica do modelo de otimização proposto neste trabalho é a integração dos dados RGB-D. Uma avaliação da acurácia relativa das trajetórias dos sensores apresentou um REMQ em torno de 3,0 cm para todos os experimentos. O resultado é aceitável, tendo em vista que a precisão do equipamento Kinect é de 3,0 cm a cada 3,0 m.

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 a medida que o erro de fechamento é detectado.

De forma geral, o método proposto apresenta como desvantagens a necessidade de um número mínimo necessário de *inliers* para determinar os parâmetros de transformação. Este método não 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. O método também não está adaptado para situações de SLAM.

Como recomendações para trabalhos futuros serão testados novos algoritmos para detecção e remoção de *outliers*, serão implementados algoritmos para tratar ambientes com regiões pouco texturizadas e adaptar o método para situações de SLAM, além de integrar sensores de navegação inercial e realizar experimentos em ambientes com maior nível de complexidade.

## AGRADECIMENTOS

O primeiro autor agradece ao CNPq (Conselho Nacional de Pesquisa e Desenvolvimento) pelo apoio concedido (processo: 473567/2013-0).

## REFERÊNCIAS BIBLIOGRÁFICAS

Al-Manasir, K., e Fraser, C.S., Registration of terrestrial laser scanner data using imagery, *The Photogrammetric Record*, v. 21, p. 255-268, 2006.

Akbarzadeh, A., Frahm, J.-M., Mordohai, P., Clipp, B., Engels, C., Gallup, D., Merrell, P., Phelps, M., Sinha, S., Talton, B., Wang, L., Yang, Q., Stewenius, H., Yang, R., Welch, G., Towles, H., Nister, D., Pollefeys, M., Towards Urban 3D Reconstruction from Video, *3D Data Processing, Visualization, and Transmission, Third International Symposium on*, v., no., p.1-8, 2006.

Bachrach, A., Prentice, S., He, R., Henry, P., Huang, A.S., Krainin, M., Maturana, D., Fox, D., e Roy, N., Estimation, planning, and mapping for autonomous flight using an RGB-D camera in GPS-denied environments: *The International Journal of Robotics Research*, v. 31, p. 1320-1343, 2012.

Barnea, S., e Filin, S., Keypoint based autonomous registration of terrestrial laser point-clouds, *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 63, p. 19-35, 2008.

Bendels, G.H., Degener, P., Wahl, R. e Klein, R., Image-based registration of 3D-range data using feature surface elements, *In Proceedings of the 5th International conference on Virtual Reality, Archaeology and Intelligent Cultural Heritage: Oudenaarde, Belgium, Eurographics Association*, p. 115-124, 2004.

Besl, P.J., e McKay, H.D., A method for registration of 3-D shapes, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, v. 14, p. 239-256, 1992.

Brown, D.C., Close-range camera calibration, *Photogrammetric Engineering*, v. 37, no. 8, p. 855-866, 1971.

Chow, J.C.; Lichti, D.D.; Hol, J.D.; Bellusci, G.; Luinge, H., IMU and Multiple RGB-D Camera Fusion for Assisting Internos Stop-and-Go 3D Terrestrial Laser Scanning. *Robotics*, v. 3, p. 247-280, 2014.

Clemente, L., Davison, A., Reid, I., Neira, J. e Tárδος, J., Mapping large loops with a single hand-held camera, *In Proceedings of Robotics: Science and Systems*, 2007.

Debevec, P.E., Taylor, C.J. e Malik, J., Modeling and rendering architecture from photographs, *In Proceedings of SIGGRAPH '96*, 1996.

Dryanovski, I., Jaramillo, C., e Jizhong, X., Incremental registration of RGB-D images, *Robotics and Automation (ICRA), IEEE International Conference on*, p. 1685-1690, 2012.

Eggert, D.W., Fitzgibbon, A.W., e Fisher, R.B., Simultaneous registration of multiple range views for use in reverse engineering of CAD models, *Computer Vision Image Understanding*, v. 69, p. 253-272, 1998.

Ellekilde, L.-P., Huang, S., VallsMiro, J., e Dissanayake, G., Dense 3D Map Construction for Internos Search and Rescue, *Journal of Field Robotics*, v. 24, p. 71-89, 2007.

Endres, F., Hess, J., Engelhard, N., Sturm, J., Cremers, D., e Burgard, W., An evaluation of the RGB-D SLAM system, *Robotics and Automation (ICRA), IEEE International Conference on*, p. 1691-1696, 2012.

Fischler, M. A. e Bolles, R. C., Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography, *Comm. of the ACM* 24, 1981.

Freedman, B., Shpunt, A., Machline, M. e Arieli, Y., U.S. Patent 0018123, 2010.

Furukawa, Y., Curless, B., Seitz, S., Szeliski, S., Reconstructing building interior from images, *In Proceedings of the International Conference on Computer Vision (ICCV)*, 2009.

Henry, P., Krainin, M., Herbst, E., Ren, X., e Fox, D., RGB-D mapping: Using Kinect-style depth cameras for dense 3D modeling of internos environments, *The International Journal of Robotics Research*, v. 31, p. 647-663, 2012.

Grisetti, G., Stachniss, C., Grzonka, S., Burgard, W., A tree parameterization for efficiently computing maximum likelihood maps using gradient descents, *In Proceedings of Robotics: Science and Systems (RSS)*, 2007.

Kim, Y.M., Theobalt, C., Diebel, J., Kosecka, J., Micusik, B. e Thrun, S., Multi-view image and ToF sensor for dense 3D reconstruction, *In Workshop on 3-D Digital Imaging and Modeling (3DIM)*, 2009.

Khoshelham, K., e Elberink, S.O., Accuracy and Resolution of Kinect Depth Data for Internos Mapping Applications, *Sensors*, v. 12, p. 1437-1454, 2012.

Khoshelham, K., dos Santos, D.R. e Vosselman, G., Generation and weighting of 3D point correspondences for improved registration of RGB-D data. In: *ISPRS Annals V.ume II-5/W2:ISPRS Workshop laser scanning*, Antalya, Turkey / ed by M. Scaioni et al. Antalya: ISPRS, 2013, ISSN: 2194-9050, p. 127-132, 11-13 November, 2013.

Lourakis, M.I.A., e Argyros, A.A., SBA: A software package for generic sparse bundle adjustment, *ACM Transactions on Mathematic Software*, v. 36, p. 1-30, 2009.

Lowe, D., Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision*, v. 60, no.2, p. 91-110, 2004.

Masuda, T., e Yokoya, N., A robust method for registration and segmentation of multiple range images, *In Proceedings of the Second CAD-Based Vision Workshop*, p. 106-113, 1994.

May, S., Droschel, D., Holz, D., Fuchs, S., Malis, E., Nuchter, A. e Hertzberg, J., Three-dimensional mapping with time-of-flight cameras, *Journal of Field Robotics*, v. 26, no. 11-12, p. 934-965, 2009.

Mikhail, E. M., Bethel, J. S. e Mcglone J. C., *Introduction to modern photogrammetry*, Inc. New York: John Wiley & Sons, 2001, 479 p.

More, J., *The levenberg-marquardt algorithm: Implementation and theory*, In G.A. Watson, editor, Numerical Analysis, Lecture Notes in Mathematics, Springer-Verlag, 1997, 630 p.

Okatani, I.S., e Deguchi, K., A method for fine registration of multiple view range images considering the measurement error properties, *Computer Vision and Image Understanding*, v. 87, p. 66-77, 2002.

Park, S.Y., e Murali, S., A fast point-to-tangent plane technique for multi-view registration, *3-D Digital Imaging and Modeling (3DIM), Proceedings Fourth International Conference on*, p. 276-283, 2003.

Stamos, I., e Leordeanu, M., Automated feature-based range registration of urban scenes of large scale, *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, v. 2, p. 555-561, 2003.

Steinbrucker, F., Sturm, J., e Cremers, D., Real-time visual odometry from dense RGB-D images, *Computer Vision Workshops (ICCV Workshops), IEEE International Conference on*, p. 719-722, 2011.

Wei, G. e Ma, S., Implicit and explicit camera calibration: Theory and experiments, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 16, no. 5, p. 469-480, 1994.

Weik, S., Registration of 3-D partial surface models using luminance and depth information, *In Proceedings of International Conference on Recent Advances*, p. 93-100, 1997.

Wolf, P. R. e Ghilani, C. D., *Adjustment Computations: Statistics and Least Squares in Surveying and GIS*, Third edition John Wiley, New York, USA, 1997, 564 p.

Zhang, Z., Iterative point matching for registration of free-form curves and surfaces, *International Journal of Computer Vision*, v. 13, p. 119-152, 1994.

Recebido em outubro de 2014. Aceito em maio de 2015.