

# Geração de Malhas Tridimensionais Combinando Dados Visuais e Geométricos

Willian C. S. Martinho, Daniel B. Mesquita, Erickson R. Nascimento  
Department of Computer Science  
Universidade Federal de Minas Gerais  
willianmartinho@gmail.com, {balbino,erickson}@dcc.ufmg.br

Pedro M. M. Castro  
Center for Informatics  
Universidade Federal de Pernambuco  
pmmc@cin.ufpe.br

**Resumo**—Neste trabalho é apresentada uma nova metodologia capaz de criar malhas tridimensionais utilizando características visuais e geométricas. Nossa abordagem é composta por três passos: i) extração de características visuais; ii) alinhamento 3D grosseiro a fino e iii) reconstrução da superfície, por meio do algoritmo de triangulação Delaunay com restrições. A ideia-chave do método é usar características visuais, como quinas e retas, extraídas de uma imagem para definir restrições para a reconstrução da superfície. Nossos experimentos realizados em ambientes internos e desordenados mostram a abordagem como uma promissora técnica para criação de malhas de ambientes.

**Abstract**—In this work we present a novel methodology which is able to create three dimensional meshes based on visual and geometrical features. Our approach is composed of three steps: i) visual features extraction; ii) coarse-to-fine 3D alignment and iii) surface reconstruction, which uses a constrained Delaunay triangulation algorithm. The key idea is to use visual features such as corner and lines extracted from a RGB image to estimate 3D meshes consistent with the regions visualized in images. Our experiments performed in indoor and clutter environments show that our approach is a promising method for creating a detailed 3D meshes.

**Keywords**-rgb-d; reconstrução; malhas; mapeamento.

## I. INTRODUÇÃO

Tarefas como a criação de ambientes realísticos e navegação autônoma de robôs requerem informações detalhadas das cenas, como a estrutura tridimensional com uma precisa descrição das superfícies e suas propriedades fotométricas. Um robô, por exemplo, precisa ter conhecimento de sua posição em relação aos objetos presentes no ambiente, assim como, sua dimensão e resultados de sua interação física para realizar tarefas autônomas com eficiência.

A fusão de dados provenientes de fontes de diferentes natureza, como dados visuais adquiridos com uma câmera e mapas de profundidade de um sensor de distancia, tem alavancado a qualidade de resultados para metodologias de reconhecimento de objetos [1], passando por mapeamento 3D [2] até extração de descritores eficientes e discriminativos [3]. Similar à abordagem desses trabalhos, neste artigo é proposta uma nova metodologia para geração de malhas tridimensionais de ambientes internos combinando informações visuais e geométricas extraídas de dados adquiridos por um dispositivo RGB-D. Além do baixo custo dos sensores RGB-D, esses sensores possuem como vantagem a capacidade de adquirir

informações visuais e de profundidade sincronizadas para cada pixel com alta taxa de aquisição (30 quadros por segundos, por exemplo).

A criação de mapas tridimensionais apresenta diversos problemas e desafios, muitos deles proveniente das limitações impostas pelo sensor de profundidade. Apesar desses sensores fornecerem dados tridimensionais, a informação de vizinhança dos pontos, e por conseguinte, a descrição da superfície que esses pontos pertencem não é fornecida.

Em geral, os sensores RGB-D estimam as informações de profundidade de uma cena por meio da projeção de padrões utilizando iluminação infravermelha. Essa abordagem, apesar de barata, impõe severas limitações aos sensores, uma vez que inviabilizam seu uso em ambientes nos quais há grande incidência de luz infra-vermelha, como ambientes externos. Quanto mais diversificado, complexo, desordenado e dependendo do tipo de material do objetos e superfícies do ambiente mais difícil torna-se extrair o mapa de profundidade. Esses fatores acarretam em perda de informação gerando buracos na nuvem de pontos tridimensional. Essas falhas podem criar grandes problemas no processo de geração das malhas, tais como perda das características geométricas das superfícies e assim dificultar a criação da conectividade entre os pontos.

A metodologia proposta neste trabalho é composta por três etapas principais. Após adquiridos dados 3D e visuais utilizando um sensor RGB-D que fornece imagens e mapas de profundidade do ambiente, as informações de profundidade são utilizadas para a geração de nuvens de pontos 3D. Em seguida, são extraídas características visuais, como pontos de interesse e quinas nas imagens. Cada ponto de interesse é identificado utilizando um descritor para a realização do casamento e alinhamento grosseiro das nuvens. Posteriormente, a metodologia calcula a triangulação utilizando restrições do domínio geométrico por meio das características visuais, e.g. retas presentes na cena.

Assim, a principal contribuição deste trabalho é a criação de uma metodologia capaz de gerar malhas tridimensionais ricas em informações geométricas e visuais do ambiente. Nossa abordagem provê uma discretização do mapeamento utilizando da triangulação de Delaunay com restrições que respeita a geometria da cena envolvida, bem como as características visuais.

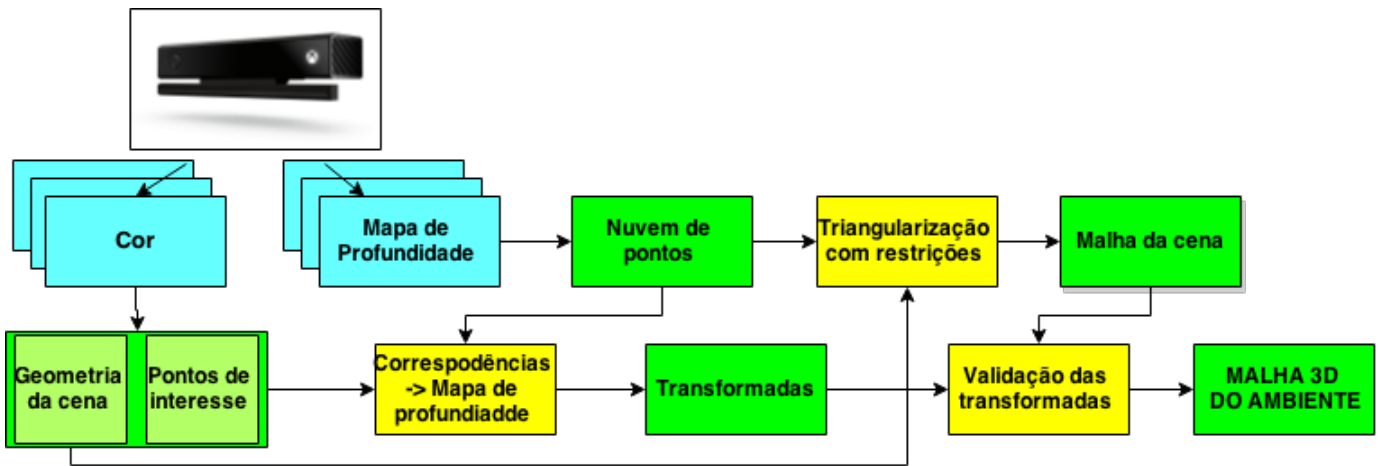


Fig. 1. Diagrama com os principais passos da metodologia. Os dados visuais e de profundidade são obtidos por um sensor RGB-D e posteriormente utilizados para a geração de nuvens de pontos 3D. A correspondência entre dois quadros consecutivos é estimada casando as características extraídas das imagens para a etapa do alinhamento grosseiro para que após o alinhamento fino com o ICP seja calculada a triangulação de Delaunay utilizando características visuais como restrições.

## II. TRABALHOS RELACIONADOS

Ao longo da última década o problema de mapeamento tridimensional de ambientes internos vem recebendo cada vez mais atenção das comunidades de Visão Computacional, Computação Gráfica e Robótica. Sistemas de mapeamento propostos em [4] e [5], apresentam soluções em tempo real utilizando dados geométricos adquiridos com equipamentos do tipo Light Detection And Ranging (LIDAR). Contudo, além do alto custo do equipamento de aquisição dos dados, em geral os dados utilizados não fornecem muitos detalhes do ambiente tendo um uso limitado para estimar a posição em relação a um ambiente.

Também surgiram trabalhos que começaram a utilizar grandes conjuntos de informações visuais a fim de gerar uma representação geométrica mais precisa [6], [7]. Esses métodos procuram estimar a posição da câmera e calcular o mapa de profundidade da cena por meio de um conjunto de pontos de interesses. Para aplicações de realidade virtual, de rastreamento e mapeamento na robótica o método é empregado obtendo bons resultados. Porém, existem limitações quando uma descrição detalhada do ambiente é requerida e é necessário alta precisão do rastreamento e localização de objetos.

A necessidade da representação geométrica do ambiente com precisão levou a trabalhos como [8] que propõe novas metodologias para estimar o mapa de profundidade por meio de múltiplas imagens em tempo real. Os resultados demonstram que a integração de inúmeras informações visuais pode trazer significativo aumento na precisão do mapa de profundidade da cena, bem como prover maior riqueza de detalhes do ambiente.

Nos últimos anos houve um rápido avanço no desenvolvimento de sensores de profundidade. O surgimento de dispositivos de baixo custo do tipo RGB-D, como Microsoft Kinect, Asus Xtion PRO LIVE e Intel RealSense 3D Camera,

vem proporcionando o surgimento de várias aplicações e novos algoritmos.

Em [9] é proposto um novo sistema de mapeamento em tempo real para complexos e arbitrários ambientes internos. A principal desvantagem desse método é seu alto custo computacional, o que torna seu uso inviável para o mapeamento de grandes áreas. Uma solução para esse problema é apresentando em [10], no qual os autores propõem uma alteração para mapear as regiões da cena dinamicamente, retirando pontos que já foram atualizados incluindo outros pontos no resultado final. O algoritmo é implementado utilizando uma hierarquia *multithread* com componentes capazes de operar em tempo real. Essa arquitetura facilita a criação e integração de novos módulos sem impactar no tempo de processamento. Exemplos de outros trabalhos que utilizam sensores RGB-D são [11], [12], [13]. Todos esses trabalhos compartilham o objetivo de utilizar as informações contidas nas imagens visuais e nos mapas de profundidade para a reconstrução das superfícies presentes no ambiente.

### A. Metodologia

O primeiro passo de nossa metodologia, a informação visual obtida por meio das imagens RGB é utilizada para a detecção dos pontos de interesses. Cada ponto de interesse é identificado por um descriptor que é utilizado para determinar a correspondência dos pontos entre as imagens adquiridas em momentos distintos. Uma vez que as imagens e os mapas de profundidades estão registrados, a correspondência estimada utilizando os dados visuais (pontos de interesses e descritores) serve para realizar o alinhamento grosseiro das nuvens tridimensionais. Dessa maneira, o método alinha nuvens de pontos de uma cena adquirida de diferentes pontos de vista em um sistema de coordenadas global.

No segundo passo, o alinhamento das nuvens é refinado para a obtenção da transformação final de cada nuvem. No terceiro

passo, cada nuvem é triangularizada separadamente utilizando como restrição a permanência das características visuais (pontos de interesse e retas) obtidas nos passos anteriores por meio dos dados visuais. Dessa maneira os pontos pertencentes às retas detectadas nas imagens, bem como vértices de triângulos correspondentes a pontos de interesse, são preservados na triangulação. Assim, as características visuais coincidentes com a malha tridimensional do ambiente permanecem nas malhas geradas de nuvens de pontos tridimensionais obtidas de pontos de vista distintos. Os principais passos e o fluxo de dados da metodologia proposta neste trabalho são descritos no diagrama da Fig. 1.

### B. Sensor de Aquisição de Dados

O sensor Kinect One é um sensor do tipo RGB-D. Ele é capaz de fornecer, simultaneamente e com frequência de aquisição de 30 Hz, informações de cor e profundidade de uma cena. Esse sensor permite obter mapas de profundidade de  $512 \times 424$  pixels, medindo na faixa entre 0,5 a 4,5 metros. Ele também possui uma câmera de cor 1080p que captura a uma taxa de 30 Hz ou 15 Hz com pouca luz.

Apesar das estimativas de profundidade do sensor possuir baixo nível de ruídos, permitindo representar a cena de maneira fiel, em algumas regiões podem ocorrer falhas de captura. Tais regiões correspondem a "buracos" no mapa de profundidade onde não há informação disponível. Outro fator que leva a perda de dados, mas dessa vez visuais, é o movimento muito rápido do dispositivo o que produz um efeito de borrramento no mapa.

### C. Detecção e descrição de características visuais

As informações de cor adquiridas pelo sensor são utilizadas essencialmente para extrair características e detectar regiões de interesse. Neste trabalhos foram utilizadas as seguintes metodologias:

*Detecção de Retas:* Detecção de retas utilizando a transformada de Hough (HT). A transformada é um método de detecção de padrões no espaço de parâmetros [14], [15]. A HT converte um problema de detecção que pode ser difícil no espaço imagem em um problema de detecção de picos locais mais facilmente resolvidos no espaço de parâmetros.

*Detector e descritor ORB:* Para detectar e descrever regiões de interesse utilizamos o método Oriented FAST and Rotated BRIEF (ORB) [16]. Esse método apresenta diversos melhoramentos na metodologia do BRIEF [17] para descrição de pontos de interesse e do FAST [18] para a detecção desses pontos. Ele provê invariância à rotação e escala na detecção e descrição do ponto de interesse. Para aumentar o desempenho de detecção, os autores aplicam um limiar de intensidade entre o pixel central e um anel circular em torno do centro o FAST-9 (raio circular de 9), além de empregar o Harris corner [19], como medida, para determinar quinas. Com principais vantagens, o descritor ORB é invariante à rotação e escala e rápido de ser calculado e tem baixo custo de armazenamento na memória, uma vez que o descritor final é representado por uma string de bits.

### D. Alinhamento Fino

Após o alinhamento grosseiro por meio das características visuais, a metodologia refina o alinhamento utilizando o algoritmo Iterative Closest Points (ICP) [20]. O algoritmo procura pela transformação que minimiza a distância Euclidiana entre um par de nuvens 3D. Dado um par de nuvens (entrada e referência), o algoritmo associa cada ponto da nuvem de entrada com o ponto mais próximo na nuvem de referência. Onde em seguida estima-se a translação e rotação que minimiza uma função de erro entre os pontos correspondentes.

### E. Triangularização de Delaunay com restrições

Com as nuvens em um mesmo sistema de coordenadas e alinhadas, o passo seguinte é gerar a malha tridimensional para as superfícies das cenas. Dentre os métodos mais consolidados e estudados para geração de malhas não estruturadas encontra-se a triangulação de Delaunay [21]. A triangulação de Delaunay de um conjunto de pontos, é uma triangulação tal que nenhum triângulo tem um círculo circunscrito que envolve qualquer ponto do conjunto de pontos. A triangulação de Delaunay cria triângulos de forma a maximizar o menor ângulo para todo o conjunto de triângulos, o que tende a evitar ângulos internos muito pequenos. Dessa forma é gerada uma malha de qualidade, sem triângulos degenerados.

Neste trabalho, é imposto que a malha gerada respeite características visuais e limites geométricos do ambiente. Para tal emprega-se a triangulação de Delaunay com restrições. Esse método restringe que determinadas arestas estejam contidas na triangulação. Particularmente, essas arestas representam delimitações da superfície devido a algum fator da cena, como por exemplo objetos e obstáculos, que se assemelham a uma reta. Dessa maneira produzimos uma malha com qualidade que representa detalhadamente a cena tanto em sua dimensão quanto geometricamente.

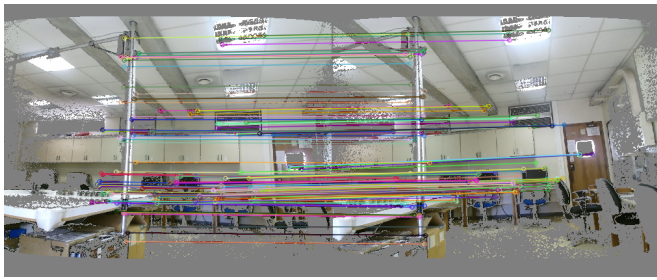
## III. IMPLEMENTAÇÃO

### A. Extração das Características Visuais

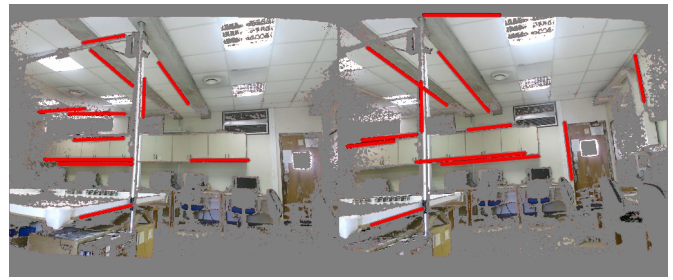
O processamento da imagem é feito por meio da detecção de pontos de interesse e descrição das características de sua vizinhança utilizando o detector e descritor ORB. Tais pontos são classificados em relação a sua intensidade e de sua vizinhança. Dessa forma é possível corresponder, com precisão, imagens da mesma cena que apresentam mesma característica. Também é realizado a extração de retas presentes imagens. Essas retas na imagens são importantes pois em geral poderão ser a projeção de qualquer aresta de algum objeto ou obstáculo do ambiente. O processo é detalhado na Fig. 2.

### B. Alinhamento Tridimensional

Cada *frame* adquirido representa a cena com certa complementaridade em relação ao *frame* anterior, assim, quando houver inconsistência em determinada região de um *frame* é possível corrigi-la através da sobreposição de outros *frames*. O mapa de profundidade capturado pelo sensor RGB-D é convertido em uma nuvem de pontos tridimensional, onde cada



(a)



(b)

Fig. 2. Extração de características visuais. Observa-se que dois *frames* subsequentes apresentam características semelhantes. A cor em cinza claro representa os buracos presentes no mapa de profundidade. As linhas em diversas cores identificam as correspondências realizadas e as retas em vermelho representam as características visuais detectadas na cena. Em geral essas características são projeções das fronteiras de objetos ou obstáculos.

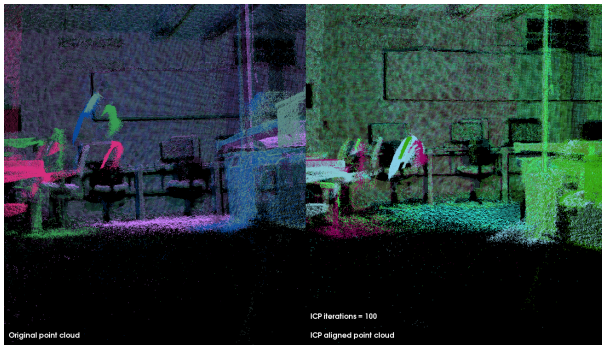


Fig. 3. Resultado final após o alinhamento fino por meio do ICP. À esquerda é apresentado o resultado do alinhamento grosseiro de duas nuvens e à direita as nuvens alinhadas após o refinamento com ICP.

pixel do mapa é transformado em um ponto com coordenadas  $(x,y,z)$ .

As correspondências apresentadas entre as imagens também estão presentes nas nuvens de pontos. Por meio delas é possível obter uma transformada rígida tridimensional para um alinhamento grosseiro. Para essa abordagem baseamos no cálculo da decomposição em valores singulares (SVD) [22]. As correspondências entre as imagens e nuvens de pontos são sempre feitas a partir da imagem atual com a sua posterior. Dessa forma os *frames* devem ser ordenados em sequência, sem grandes translações, para garantir que sempre há superposição.

Após a estimativa inicial da transformação, as nuvens de pontos não estão totalmente alinhadas, já que o algoritmo é afetado por falha nas correspondências. Para que ocorra um alinhamento mais preciso, utiliza-se o algoritmo ICP, obtendo o resultado mostrado na Fig. 3. Os algoritmos descritos possuem implementações disponíveis na biblioteca Point Cloud Library (PCL) [23].

### C. Reconstrução da Superfície

Após estimar todas as matrizes de transformação e restrições associadas às características visuais da cena, inicia-se o processo de triangulação das nuvens de pontos. Para isso foram utilizadas as estruturas de dados disponíveis na biblioteca CGAL [24] e a triangulação Delaunay com restrições.

A triangulação é feita em um plano bi-dimensional definido por um grid de mesma dimensão da nuvem de pontos ( $424 \times 512$  pixels). Todos os vértices da nuvem de pontos estão representados por uma coordenada cartesiana  $(x,y)$  projetada nesse plano. Declaramos na triangulação as restrições das arestas de interesse encontradas previamente nas imagens e mapeadas na retas detectadas (retas vermelhas na Fig.2 (b)).

A premissa da imposição das restrições nas malhas é a de que restrições são características inerentes ao ambiente sendo necessário preservá-las nas triangulações. Dessa forma, cada triangulação na nuvem de pontos dos *frames*, apesar de terem poses diferentes, apresentarão as mesmas características quando essas forem detectadas nas imagens.

Por fim, é gerado um grande conjunto de malhas e suas respectivas transformações rígidas associadas ao número *frames* adquiridos do ambiente. Cada matriz de transformação contém informações de translação e rotação que alinham a malha atual em relação a anterior. O procedimento final da metodologia consiste em aplicar as transformações para alterar os valores dos vértices das nuvens de pontos mantendo a relação de vizinhança dos vértices.

## IV. EXPERIMENTOS

O método proposto foi avaliado utilizando um conjunto de dados RGB-D adquiridos de um ambiente interno de um laboratório de pesquisa. O ambiente é composto majoritariamente por mesas, cadeiras e monitores de computador. Todos os dados foram coletados utilizando o Kinect One controlado por uma pessoa fazendo movimentos laterais de maneira lenta, a uma distância média de 2,64 m. A malha final gerada para um conjunto de seis *frames* é mostrada na Fig. 4.

### A. Performance

O experimento foi realizado em um computador com as seguintes especificações:

- Processador Intel Core i5-480M de 2.66GHz;
- Memória 4GB DDR3 SDRAM.

A etapa de detecção e descrição dos pontos de interesse e correspondências entre as imagens executou em 5.494 ms. O alinhamento das nuvens de pontos foi a etapa que demandou a maior parcela do tempo de processamento. O algoritmo ICP realizou um total de 5 ciclos com 100 iterações entre cada

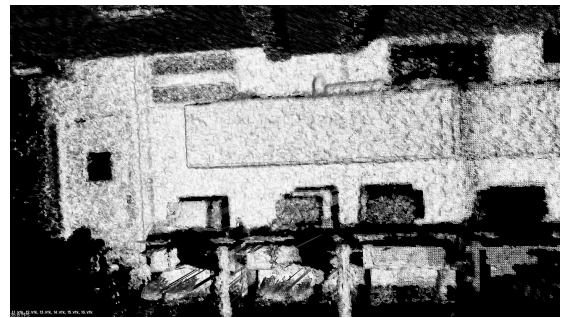




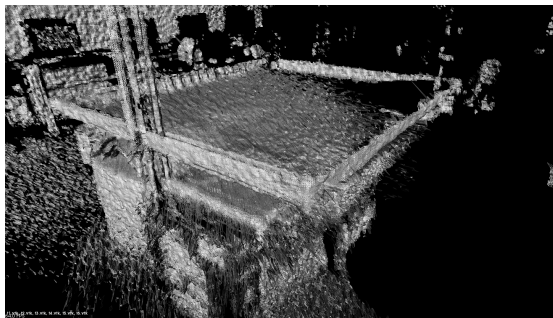
(a)



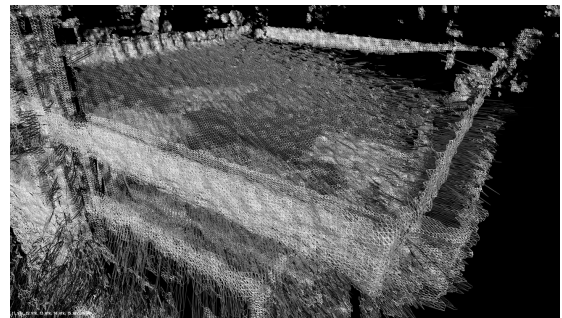
(b)



(c)



(d)



(e)

Fig. 4. Malha gerada de um ambiente interno de um laboratório de pesquisa (a). Todo conjunto de dados se encontra alinhado e a malha final respeita as características geométricas do ambiente detectadas nas imagens visuais. As imagens em (b) e (c) representam a aproximação da malha. Podemos ver a densidade de pontos e os inúmeros objetos que compõem a malha, como cadeiras e monitores. Em (d) e (e) temos em detalhe a representação de objetos.

nuvem de pontos e um total de 399.562 ms de execução. A triangulação de todas as nuvens de pontos gastou 5.360 ms para finalizar o processamento. O tempo total de processamento para os experimentos foi aproximadamente de 7 minutos.

### B. Aspecto qualitativo

Como pode ser observado nas imagens mostradas na Fig. 4, o método proposto foi capaz de gerar uma malha tridimensional final para o ambiente coerente com as características visuais extraídas das imagens. O mapa de profundidade capturado pelo sensor RGB-D apresentava regiões de falha e com ruídos que foram corrigidos. A Fig. 4 (e) e a Fig. 4 (d) apresenta a malha gerada em detalhe. Observa-se que há alta densidade de pontos o que corrige regiões com falhas. O resultado final mostrado na Fig. 4 é a representação tridimensional do laboratório por meio de uma malha densa, as Fig. 4 (b) e a Fig. 4 (c) mostram detalhes de objetos compõem o laboratório.

### C. Limitação

Apesar da metodologia ter sido capaz de gerar uma malha coerente com o ambiente mapeado, ela é limitada para ambientes que apresentam pouca iluminação, pois nesse caso não existe informação visuais do ambiente e assim o sistema não consegue detectar características visuais. Um outro problema ocorre quando há falhas de captura produzindo "buracos" nas nuvens de pontos, impossibilitando o uso das informações visuais nessas regiões.

## V. CONCLUSÃO

O mapeamento tridimensional de ambientes internos para aplicações na robótica, entretenimento e realidade aumentada é de suma importância. Com o potencial dos sensores RGB-D de baixo custo e boa precisão esse campo demandará cada vez mais estruturas de dados e algoritmos capazes de trabalhar com dados visuais e geométricas de maneira sinérgica. Nesse trabalho apresentamos uma forma de gerar um mapeamento com triangulação de nuvens de pontos provenientes das sensores RGB-D combinando dados visuais e de profundidades. O cerne deste projeto é a utilização de informações visuais do ambiente em conjunto com as de profundidade e a busca por restrições que caracterizam a geometria do ambiente o e atribuem vínculo entre os dados, tornando-se assim uma mapeamento mais robusto.

Os resultados mostraram que é possível utilizar essa abordagem por meio de poucos dados do ambiente, tendo a metodologia se mostrado promissora para o problema. No entanto, faz-se necessário realizar uma avaliação mais completa com outros tipos de ambientes, bem como o uso de outras característica visual, como regiões curvas, por exemplo.

## REFERÊNCIAS

[1] K. Lai, L. Bo, X. Ren, and D. Fox, "A Large-Scale Hierarchical Multi-View RGB-D Object Dataset," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, May 2011.

[2] P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox, "RGB-D Mapping: Using Depth Cameras for Dense 3D Modeling of Indoor Environments," in *International Symposium on Experimental Robotics (ISER)*, 2010.

[3] E. R. do Nascimento, G. L. Oliveira, A. W. Vieira, and M. F. Campos, "On the development of a robust, fast and lightweight keypoint descriptor," *Neurocomputing*, vol. 120, 2013.

[4] J.-S. Gutmann and K. Konolige, "Incremental mapping of large cyclic environments," *Proceedings 1999 IEEE International Symposium on Computational Intelligence in Robotics and Automation. CIRA'99 (Cat. No.99EX375)*, 1999.

[5] D. Hifnel, W. Burgard, D. Fox, and S. Thrun, "An Efficient FastSLAM Algorithm for Generating Maps," *Intl. Conference on Intelligent Robot and Systems*, no. October, 2003.

[6] G. Klein and D. Murray, "Parallel tracking and mapping for small AR workspaces," *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, ISMAR*, 2007.

[7] R. a. Newcombe, S. J. Lovegrove, and A. J. Davison, "DTAM: Dense tracking and mapping in real-time," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2320–2327, 2011.

[8] R. a. Newcombe and A. J. Davison, "Live Dense Reconstruction Using a Single Passive Camera," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011.

[9] R. a. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohli, J. Shotton, S. Hodges, and A. Fitzgibbon, "KinectFusion: Real-time dense surface mapping and tracking," *2011 10th IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2011*, pp. 127–136, 2011.

[10] T. Whelan, M. Kaess, and M. Fallon, "Kintinuous: Spatially extended kinectfusion," *RSS Workshop on RGB-D: Advanced Reasoning with Depth Cameras*, 2012.

[11] P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox, "RGB-D mapping: Using Kinect-style depth cameras for dense 3D modeling of indoor environments," *The International Journal of Robotics Research*, vol. 31, no. 5, pp. 647–663, 2012.

[12] F. Endres, J. Hess, N. Engelhard, J. Sturm, D. Cremers, and W. Burgard, "An evaluation of the RGB-D SLAM system," *Robotics and Automation (ICRA), 2012 IEEE International Conference*, vol. 1, no. c, pp. 1691–1696, 2012.

[13] F. Endres, J. Hess, J. Sturm, D. Cremers, and W. Burgard, "3-D Mapping with an RGB-D camera," *IEEE Transactions on Robotics*, vol. 30, no. 1, pp. 177–187, 2014.

[14] J. Illingworth and J. Kittler, "A survey of the hough transform," *Computer Vision, Graphics, and Image Processing*, vol. 44, no. 1, pp. 87–116, 1988.

[15] J. Matas, C. Galambos, and J. Kittler, "Robust Detection of Lines Using the Progressive Probabilistic Hough Transform," *Computer Vision and Image Understanding*, vol. 78, no. 1, pp. 119–137, 2000.

[16] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: an efficient alternative to SIFT or SURF."

[17] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "BRIEF : Binary Robust Independent Elementary Features," *Proceedings of the European Conference on Computer Vision. ECCV'10*, pp. 778–792, 2010.

[18] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," pp. 430–443, 2006.

[19] C. Harris and M. Stephens, "A Combined Corner and Edge Detector," *Proceedings of the Alvey Vision Conference 1988*, pp. 147–151, 1988.

[20] P. J. Besl and N. D. McKay, "A method for registration of 3-D shapes," pp. 239–256, 1992.

[21] M. d. Berg, O. Cheong, M. v. Kreveld, and M. Overmars, *Computational Geometry: Algorithms and Applications*, 3rd ed. Santa Clara, CA, USA: Springer-Verlag TELOS, 2008.

[22] D. Eggert, a. Lorusso, and R. Fisher, "Estimating 3-D rigid body transformations: a comparison of four major algorithms," *Machine Vision and Applications*, vol. 9, no. 5-6, pp. 272–290, 1997.

[23] R. B. Rusu and S. Cousins, "3D is here: Point Cloud Library (PCL)," *Proceedings - IEEE International Conference on Robotics and Automation*, 2011.

[24] J.-D. Boissonnat, O. Devillers, S. Pion, M. Teillaud, and M. Yvinec, "Triangulations in CGAL," *Computational Geometry*, vol. 22, no. 1-3, pp. 5–19, 2002.