

CENTRO UNIVERSITÁRIO FEI

FERNANDA SARTORI

RICARDO MORELLO SANTOS

LUCAS BRANDON TENORIO

JOÃO AURÉLIO FRANCISCO JÚNIOR

**3DIR: UMA METODOLOGIA PARA RENDERIZAÇÃO 3D DE AMBIENTES
INTERNAOS COM O USO DO MICROSOFT HOOLENS**

São Bernardo do Campo

2019

SUMÁRIO

1	TRABALHOS RELACIONADOS	3
1.1	OBJETIVO	3
1.2	METODOLOGIA	3
1.3	MÉTRICAS	4
1.4	BASE DE DADOS	7
	REFERÊNCIAS	12

1 TRABALHOS RELACIONADOS

1.1 OBJETIVO

1.2 METODOLOGIA

(introdução de metodologia aqui)

No trabalho de (RUSINKIEWICZ; LEVOY, 2001) foi realizada uma metodologia de comparação para as convergências das variantes do ICP, definindo uma combinação de variantes como *baseline*. Cada variante foi classificada em seis estágios do algoritmo e submetidas à testes de *performance* para identificar a configuração de variantes que otimiza todas as etapas.

O mapeamento de ambientes descrito em (HENRY et al., 2012) utiliza uma versão melhorada do ICP, que envolve informações sobre a cor e a profundidade extraída dos *frames*. A extração ocorre pelo algoritmo FAST, que gera dois conjuntos de características de um par de *frames*, e a melhor transformação rígida entre esses conjuntos é obtida pela técnica do algoritmo RANSAC. Paralelamente, uma etapa de detecção de *loop closure* é realizada, usando pontos característicos do *frame* atual comparados com informações espaciais anteriores. Cada detecção adiciona uma restrição ao *pose graph* que inicia um processo de otimização global. As nuvens de pontos são representadas na forma de *surfels*, uma estrutura que possui dados espaciais dos pontos na superfície e suas cores. Os *surfels* são atualizados e adicionados conforme a posição da câmera se altera, resultando em mapeamentos detalhados em tempo real.

Em (KOLMOGOROV; ZABIH, 2002), a reconstrução de cenas com múltiplas câmeras é feita através de uma função de minimização. Essa função de avaliação considera a luminosidade, a visibilidade e a vizinhança de cada pixel para extrair a profundidade de um conjunto de imagens. A minimização dessa função foi demonstrada como um problema *NP-hard*. Portanto, foi necessário utilizar um algoritmo de aproximação baseado em *graph cuts* para encontrar um mínimo local confiável. No entanto, a metodologia de reconstrução proposta se limita apenas em obter o mapa de profundidade de imagens, que não inclui técnicas para gerar malhas de pontos e posteriormente construir uma superfície 3D.

1.3 MÉTRICAS

Nesta subseção, serão apresentadas as principais métricas encontradas na literatura e como elas foram utilizadas. Ao todo, foram identificados 86 métodos de avaliação. As principais métricas encontradas, ou seja, aquelas mais utilizadas, são apresentadas na Figura 1. Nota-se que pode haver sobreposição de artigos em mais de um setor do gráfico, visto que um dado trabalho pode utilizar mais de uma métrica. Dos trabalhos compreendidos pela revisão bibliográfica, 29 deles apresentam tempo de execução como forma de comparação, ou para indicar diminuição no tempo de processamento em relação ao estado-da-arte. A métrica mais utilizada junto de tempo de execução é o uso de memória, com 8 trabalhos. Sendo assim, 8 trabalhos compararam o método com o estado-da-arte com base em tempo de execução e memória necessários para o processamento. Em segundo lugar, *Root Mean Square Error* (RMSE), com 7 trabalhos, seguido de acurácia, com 4 trabalhos.

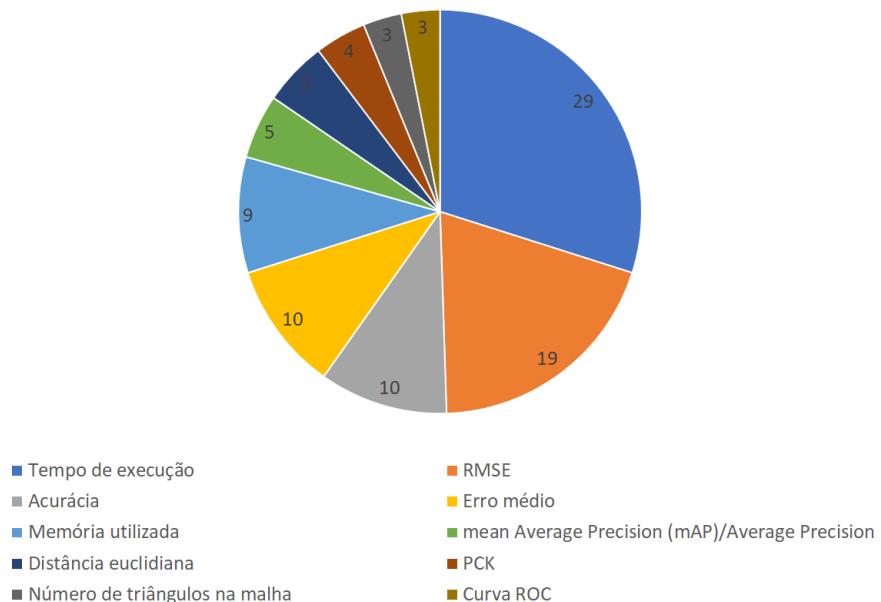


Figura 1 – Métricas mais utilizadas na literatura

Em (BLANZ; VETTER, 1999), foi construído um modelo paramétrico deformável de faces com base em um conjunto de 200 exemplos faciais. Os testes foram feitos utilizando-se diferentes imagens faciais, e o tempo de execução dos testes foi de 50 minutos. Esse tempo inclui a busca da melhor combinação das faces na base de dados com a face apresentada na imagem de entrada. Como trabalhos futuros, objetiva-se diminuir o tempo de execução por meio de um método newtoniano para minimização da função de custo. Também, outro objetivo inclui a expansão da base de dados e a área de reconstrução, que abrange apenas a face.

Como mencionado anteriormente, diversos trabalhos na literatura buscam comparações baseadas em memória e tempo de execução. Métodos que necessitam de reconstrução 3D em tempo real, como SLAM, visam otimizar o processo de mapeamento de ambientes para minimizar o uso de recursos computacionais, como em Grisetti, Stachniss e Burgard (2007), Whelan et al. (2016) e Zhu et al. (2016).

O trabalho de Kazhdan, Bolitho e Hoppe (2006) avalia a performance do algoritmo proposto em termos de velocidade de processamento e exatidão no produto da reconstrução, comparando-se com outros métodos do estado-da-arte. Os resultados mostram que, ao aumentar-se os níveis da *octree* utilizada para resolver o sistema de *Poisson*, maior é a resolução na reconstrução final. Apesar do algoritmo não ser o mais rápido nem o que utiliza menos memória, sua natureza quadrática permite escalá-lo para maiores resoluções sem afetar a performance. Como trabalhos futuros, pode-se incorporar informações secundárias da imagem para aprimorar a reconstrução.

Para avaliar diferentes algoritmos de reconstrução 3D baseados em estereoscopia, Seitz et al. (2006) utilizaram as métricas de acurácia, completude, tempos de execução em diferentes processadores e bases de dados, RMSE e histogramas de distância das reconstruções para o padrão-ouro. O método de Esteban e Schmitt (2004) obteve a melhor acurácia para a reconstrução 3D de um modelo de templo, com 90% dos pontos da malha pertencendo a um intervalo de, no máximo, 36 mm do *ground-truth*. Visto que os algoritmos analisados geram modelos completos na reconstrução, os números para a métrica de completude não foram tão diferentes. Entretanto, os tempos de execução variaram de 30 minutos para mais de um dia.

Um ponto importante a ser considerado para a viabilidade da reconstrução 3D em tempo-real é a busca de pontos fiduciais. Em (GEIGER; ZIEGLER; STILLER, 2011), o tempo de execução para detecção de pontos fiduciais entre duas imagens foi reduzido de uma curva exponencial para linear. Entretanto, assumiu-se um conjunto de imagens obtido a partir de trajetórias suaves da câmera.

Os métodos de reconstrução apresentados até então visam a reconstrução de partes específicas da imagem de entrada. No trabalho de Jancosek e Pajdla (2011), foram realizadas reconstruções de cenas que preservaram regiões descartadas por outros algoritmos, mantendo o tempo de execução similar.

A partir do surgimento de *hardwares* para captura de imagens RGB-D como o *Kinect*, foi possível capturar e reconstruir cenas dinâmicas em tempo-real com rápida movimentação, como em Newcombe et al. (2011). Neste trabalho, obteve-se tempo de execução constante para rastreio e mapeamento denso de *voxels*, apesar do aumento na resolução. Entretanto,

por tratar da reconstrução densa de ambientes, a utilização deste método para reconstrução de grandes ambientes *indoor* é prejudicial em termos de recursos computacionais. Esse problema foi abordado em (WHELAN et al., 2012) e (KELLER et al., 2013). No trabalho de Whelan et al. (2012), foram comparados os tempos de execução para cada uma das quatro bases de dados, utilizando duas odometrias (ICP e FOVIS). O ICP foi o método de odometria mais rápido, devido à sua implementação em GPU. Em (RUSINKIEWICZ; LEVOY, 2001), a variação do ICP proposta atinge tempo constante, mas apenas para alinhamento de um par de malhas. Sendo assim, não aborda o problema de *global registration* do SLAM.

Outro trabalho que aborda o problema da reconstrução 3D em ambientes em grande escala é apresentado em (CHEN; BAUTEMBACH; IZADI, 2013). Para verificar a performance do método, mediu-se o tempo de execução e o consumo de memória na GPU. Nota-se um pico de performance com três níveis na *octree*. Assim, foi possível reconstruir uma livraria de três andares em seis minutos, com taxas de frames por segundo (FPS) variando de 24 a 37, para resolução igual a 1024.

A estrutura de dados *hash* para reconstrução 3D proposta por Nießner et al. (2013) é executada de forma mais rápida em relação à taxa de quadros do Kinect. Assim, utiliza-se mais tempo nas iterações do algoritmo *Iterative closest point* (ICP) para otimizar a estimativa de posição da câmera. Dentre outros dois algoritmos comparados, o método apresentado foi o que obteve maior índice de FPS no processamento. Em relação à memória, a estrutura *hash* ocupa 34 megabytes (MBs), e apresenta índice de *colisão* igual a 0.1%.

Novos trabalhos baseados na reconstrução 3D de ambientes em larga-escala atingiram erros e tempos de processamento significativamente menores. O método proposto em (Steinbrucker et al., 2013) é avaliado dez datasets diferentes, obtendo resultados superiores ao estado-da-arte, incluindo em relação ao *KinectFusion*, em nove deles. Como métrica de avaliação, foi considerado o RMSE do erro de trajetória absoluta (ATE). Enquanto o RMSE médio para os 10 datasets foi de 0.026m para o método apresentado, o *KinectFusion* obteve média de 0.297m, uma ordem de grandeza superior. Por sua vez, o método de reconstrução 3D proposto em (CHOI; ZHOU; KOLTUN, 2015) diminuiu o erro baseado na distância euclidiana do padrão-ouro por um fator de 2 em relação ao estado da arte. O método foi comparado com o *Kintinuous*, uma extensão do *KinectFusion*. Sob a métrica *Balanced rank estimation* (BRE), o método atingiu os maiores índices dentre as técnicas comparadas.

O balanço entre tempo de processamento e detalhes na reconstrução final ainda é um desafio pendente na literatura. Em (ZHOU; MILLER; KOLTUN, 2013), testes qualitativos e quantitativos utilizando a métrica *point-to-plane error* mostram que o método conseguiu re-

construir com mais detalhes do que o estado-da-arte, incluindo o método *Extended KinectFusion* (ROTH; VONA, 2012). Entretanto, a reconstrução de cenas pode levar horas, atingindo 56 horas de processamento em uma cena com mais detalhes.

Métodos paramétricos para reconstrução de faces, como o apresentado no início desta Seção, apresentam falhas na captura de detalhes faciais, ainda mais quando esse processo é feito em tempo real. Em (ZOLLHÖFER et al., 2014a), foram feitas reconstruções de forma qualitativa, mostrando a captura aprofundada de detalhes, e quantitativa, exibindo a métrica de *fitting error* por frame. Uma desvantagem desse método é a perda de rastreio e queda de detalhes na reconstrução quando há aumento na movimentação do objeto reconstruído. Isso é mostrado nos gráficos, com picos de erro na troca de frames ou no aumento da velocidade de movimentação. Entretanto, o método apresentou rápida convergência, necessitando de apenas oito iterações na otimização de Gauss-Newton.

Métodos de reconstrução 3D em tempo-real baseados em CPU permitem a portabilidade para dispositivos que não dispõem de GPUs para processamento, como drones ou robôs mais simples. Em (STEINBRUCKER; STURM; CREMERS, 2014), a utilização de uma estrutura *octree* limitada permitiu a otimização do algoritmo de reconstrução, reduzindo 55% do tempo gasto por frame RGB-D.

1.4 BASE DE DADOS

Nesta seção, serão apresentadas as principais bases de dados encontradas na literatura. No gráfico da Figura 2, percebe-se que as bases RGB-D SLAM (Benchmark of Sturm), ICL-NUIM, NYU Depth v2 e SUN3D são as mais utilizadas.

Muitos dos trabalhos publicados na área avaliaram suas metodologias reconstruindo modelos sintéticos de ambientes internos ou através de sequências de frames RGB-D. As bases de dados ICL-NUIM e SUN3D são utilizadas em trabalhos como (DAI et al., 2017b), (CHOI; ZHOU; KOLTUN, 2015), (WHELAN et al., 2016) e (ZOLLHÖFER et al., 2018). No trabalho de (DAI et al., 2017b), foram utilizados dois ambientes sintéticos, uma sala e um escritório. Por sua vez, em (CHOI; ZHOU; KOLTUN, 2015) e (ZOLLHÖFER et al., 2018), utilizou-se as cenas internas da base ICL-NUIM. Essas cenas incluem um sistema *off-line* para a reconstrução automática baseado em ajuste de pacote.

A base de dados SUN3D¹ proposta por (XIAO; OWENS; TORRALBA, 2013), fornece um número de reconstruções que foram criadas com assistência manual usando uma interface

¹sun3d.com

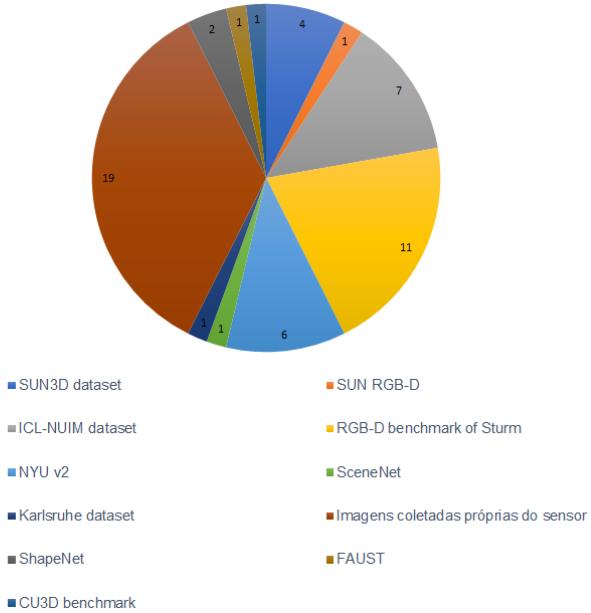


Figura 2 – Base de dados mais utilizadas na literatura

interativa na qual permite o usuário estabelecer correspondências em nível de objeto com um conjunto de frames como entrada. Ela contém 415 sequências RGB-D com anotações de polígonos em 2D, mesmo havendo uma nuvem de pontos em 3D a base ainda se encontra no domínio 2D(SONG; LICHTENBERG; XIAO, 2015).

Em contraste com a SUN3D, a base de dados SUN RGB-D² (SONG; LICHTENBERG; XIAO, 2015), muito utilizada para o reconhecimento de cenas, contém 10335 imagens RGB-D com anotações densas em 2D e 3D para objetos e salas capturadas. Os sensores utilizados foram, *Intel RealSense*, *Asus Xtion*, *Kinect V1* e *Kinect V2*. A base de dados ainda conta com métricas 3D para avaliação de algoritmos com ênfase em reconhecimento de cenas.

O *Imperial College London and National University of Ireland Maynooth* (ICL-NUIM)³ proposto por (HANDA et al., 2014), é uma base de dados para avaliação de odometria visual, reconstrução 3D e algoritmos SLAM. Ela possui sequências RGB-D de ambientes gerados sinteticamente com padrões ouro para diferentes perspectivas e para um modelo de superfície que permite um método de avaliar o mapa final ou a precisão da reconstrução de superfície.

Para técnicas de reconstrução que necessitam avaliar o SLAM ou o sistema de odometria visual em dados RGB-D, a base proposta por (STURM et al., 2012) é uma base de dados na qual consiste em 39 sequências que foram gravadas em dois ambientes internos diferentes. Eles foram registrados por um *Kinect* de alta acurácia. A base utilizou um sistema de captura de movimentos para obter um padrão ouro das diferentes trajetórias da camera. Também, foram

²<http://rgbd.cs.princeton.edu/>

³www.googl.com

	NYU Depth V2	SUN3D
Vídeos RAW	Sim	Sim
Área coberta	parte de um quarto	vários quartos
Comprimento típico	Centenas de frames	Milhares de Frames
Rótulo de Objeto	Frames esparços	Vídeo inteiro
Perspectivas de câmera	Não	Sim
Rótulo de instância	Dentro de um quadro	Dentro do vídeo inteiro
Melhoria de profundidade	filtragem bilateral cruzada	enchimento de multi-frames TSDF

Tabela 1 – Comparação entre as bases de dados NYU Depth v2 e SUN3D.

propostas ferramentas para avaliar um sistema SLAM comparando o movimento estimado da câmera com a verdadeira trajetória. Tais ferramentas empregam dois métodos para realizar essa tarefa, o erro de perspectiva relativo (RPE) e o erro absoluto da trajetória (ATE).

O trabalho realizado em (HU, 2017) ,que reconstrói objetos utilizando um cubóide, propôs uma base de dados denominada CU3D com três sequências sintéticas sem ruídos e padrões ouro para trajetórias de câmera e modelos de malha. Seis sequências escaneadas do mundo real com ruído e apenas um padrão ouro dos modelos escaneados. Por último foi escaneado cinco sequências do mundo real com ruído porém sem padrão ouro dos modelos. Em (HU, 2017) utilizaram essa base de dados para avaliar e comparar sua metodologia com outros trabalhos relacionados, como o KinectFusion(NEWCOMBE et al., 2011) e ElasticFusion(WHELAN et al., 2016).

No artigo(DAI et al., 2017a) foi proposto o ScanNet⁴, uma base de dados com ricos ambientes escaneados do mundo real contendo 2.5M imagens RGB-D e 1513 *scans* de 707 lugares distintos, com o objetivo de entender melhor os modelos 3D para fazer reconhecimento de cena, que também pode ser utilizado para reconstrução de ambientes internos. Ainda em reconhecimento de cenas, a base de dados NYU Depth v2⁵ proposta por (SILBERMAN DE-REK HOIEM; FERGUS, 2012), contendo 464 sequências RGB-D curtas nos quais 1449 *frames* foram anotados com polígonos 2D denotando segmentação semântica. Os autores dessa base visitaram um grande número de ambientes internos, capturando vídeos com o *kinect* de salas, quartos, banheiros e cozinhas, totalizando em 464 ambientes distintos classificados em 26 classes de cenas.Para uma breve comparação da base de dados NYU Depth V2, na Tabela 1, podemos observar suas diferenças com a SUN3D mostradas em (XIAO; OWENS; TORRALBA, 2013).

⁴<http://www.scan-net.org/>

⁵https://cs.nyu.edu/~silberman/datasets/nyu_depth_v2.html

Em (GEIGER; ZIEGLER; STILLER, 2011) é utilizado uma base de dados para experimentos de odometria visual chamada Karlsruhe⁶ com sequências estéreo de alta qualidade gravadas por um veículo em movimento. Nela contém dados padrão ouro de GPS e IMU assim como sequências estéreo com resolução $1344x391\ pixels$ e 10 FPS (*Frame Per Seconds*).

Na área de reconstrução de objetos utilizando algoritmos de treinamento(LUN et al., 2017), (FAN; SU; GUIBAS, 2017), é utilizado a base de dados ShapeNet proposta por (CHANG et al., 2015), que contém modelos 3D de uma infinidade de categorias semânticas e organizadas sob a taxonomia *WordNet*. É uma coleção de conjuntos de dados que fornece muitas anotações semânticas para cada modelo 3D, como alinhamentos rígidos consistentes, simetria de partes e planos, tamanhos físicos, palavras-chave e outras anotações planejadas.

Em um âmbito diferente mas ainda sobre reconstrução 3D, há a preocupação em desenvolver uma boa técnica de registro de superfície para a criação de modelos estatísticos, recuperação de forma e rastreamento. O problema é ainda mais difícil para objetos articulados como o corpo humano. Em (BOGO et al., 2014), uma nova técnica de registro de malha é proposta juntamente com a FAUST⁷, uma base de dados que contém 300 escaneamentos de 10 pessoas em uma ampla gama de posições juntos com uma metodologia de avaliação.

Em outros artigos como (IZADI et al., 2011), (NEWCOMBE et al., 2011), (SAXENA; CHUNG; NG, 2007), (HENRY et al., 2010) e (ZOLLHÖFER et al., 2014b), nenhuma base de dados é utilizada para a validação da metodologia, no entanto foram utilizados modelos reconstruídos com sensores RGB-D. A Figura 3 mostra os dados das bases apresentadas nessa seção, como os modelos recriados do CU3D, objetos do banco ShapeNet, imagens de profundidade e RGB-D da SUN3D.

Na literatura, essas são as bases de dados mais utilizadas no âmbito da reconstrução 3D de cenas *indoor*, reconstrução de objetos e reconhecimento de ambientes. Na Figura 2 é possível observar a quantidade de artigos estado-da-arte das áreas citadas acima, que utilizaram determinada base de dados para a valiação de suas metodologias. A maioria dos artigos utilizaram suas próprias imagens, na maioria das vezes, para reconstrução de ambientes internos basta apenas capturar o mapa de profundidade de um ambiente qualquer para observar os resultados da reconstrução. Quando surge a necessidade de avaliar o processo de forma quantitativa, bases de dados com padrões ouro são necessárias.

⁶http://www.cvlbs.net/datasets/karlsruhe_sequences/

⁷<http://faust.is.tue.mpg.de/>

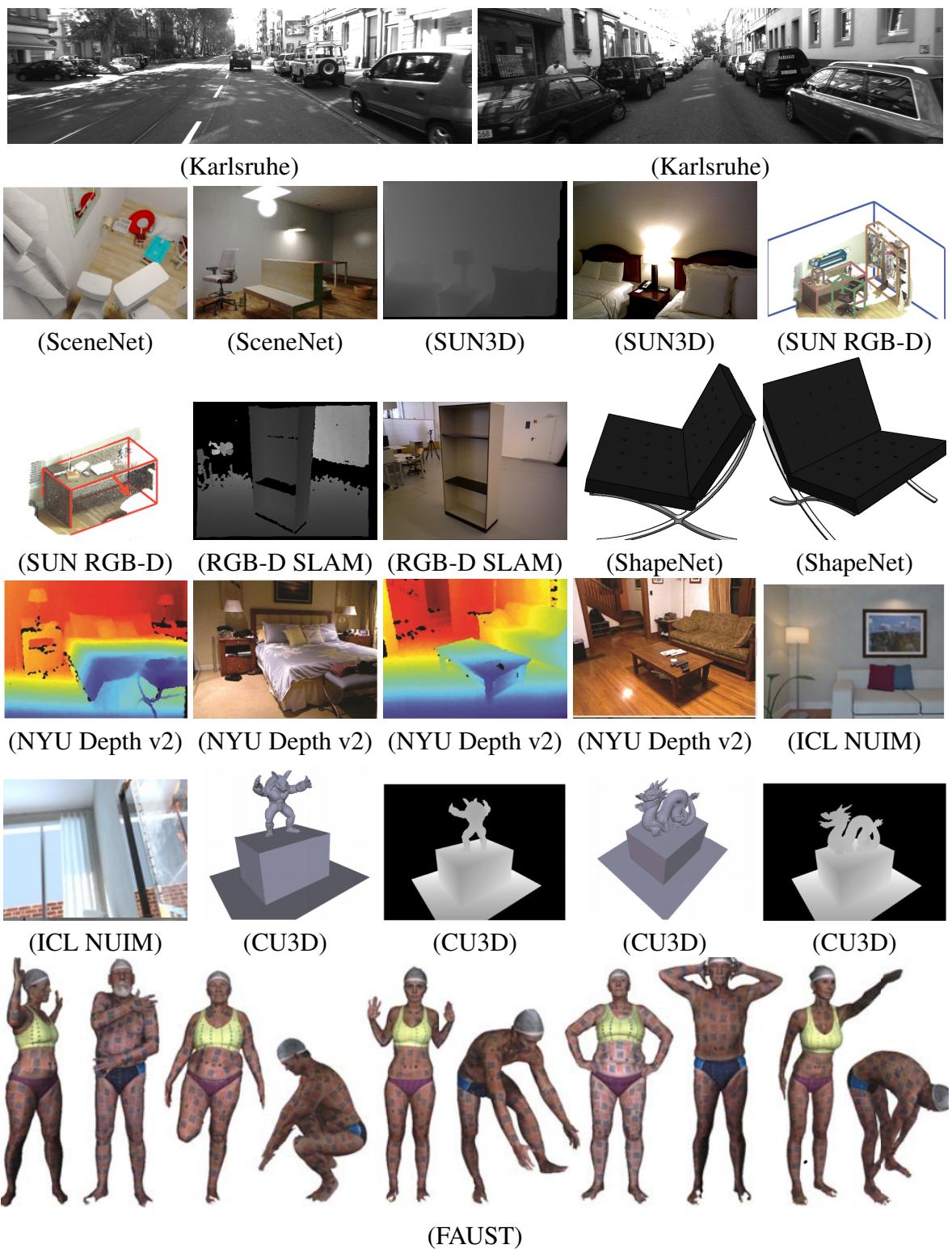


Figura 3 – Exemplos das imagens das bases de dados buscadas na literatura

REFERÊNCIAS

- BLANZ, V.; VETTER, T. A morphable model for the synthesis of 3d faces. In: **SIGGRAPH**. [S.l.: s.n.], 1999.
- BOGO, F. et al. Faust: Dataset and evaluation for 3d mesh registration. **2014 IEEE Conference on Computer Vision and Pattern Recognition**, p. 3794–3801, 2014.
- CHANG, A. X. et al. Shapenet: An information-rich 3d model repository. **CoRR**, abs/1512.03012, 2015.
- CHEN, J.; BAUTEMBACH, D.; IZADI, S. Scalable real-time volumetric surface reconstruction. **ACM Trans. Graph.**, ACM, New York, NY, USA, v. 32, n. 4, p. 113:1–113:16, jul. 2013. ISSN 0730-0301. Disponível em: <<http://doi.acm.org/10.1145/2461912.2461940>>.
- CHOI, S.; ZHOU, Q.-Y.; KOLTUN, V. Robust reconstruction of indoor scenes. **2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**, p. 5556–5565, 2015.
- DAI, A. et al. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In: **Proc. Computer Vision and Pattern Recognition (CVPR)**, IEEE. [S.l.: s.n.], 2017.
- _____. Bundlefusion: Real-time globally consistent 3d reconstruction using on-the-fly surface re-integration. **ACM Trans. Graph.**, v. 36, 2017.
- ESTEBAN, C. H.; SCHMITT, F. Silhouette and stereo fusion for 3d object modeling. **Computer Vision and Image Understanding**, v. 96, n. 3, p. 367 – 392, 2004. ISSN 1077-3142. Special issue on model-based and image-based 3D scene representation for interactive visualization. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1077314204000542>>.
- FAN, H.; SU, H.; GUIBAS, L. J. A point set generation network for 3d object reconstruction from a single image. **2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**, p. 2463–2471, 2017.
- GEIGER, A.; ZIEGLER, J.; STILLER, C. Stereoscan: Dense 3d reconstruction in real-time. **2011 IEEE Intelligent Vehicles Symposium (IV)**, p. 963–968, 2011.
- Grisetti, G.; Stachniss, C.; Burgard, W. Improved techniques for grid mapping with rao-blackwellized particle filters. **IEEE Transactions on Robotics**, v. 23, n. 1, p. 34–46, Feb 2007. ISSN 1552-3098.
- HANDA, A. et al. A benchmark for rgb-d visual odometry, 3d reconstruction and slam. **2014 IEEE International Conference on Robotics and Automation (ICRA)**, p. 1524–1531, 2014.

- HENRY, P. et al. Rgb-d mapping: Using depth cameras for dense 3d modeling of indoor environments. In: **ISER**. [S.l.: s.n.], 2010.
- _____. Rgb-d mapping: Using kinect-style depth cameras for dense 3d modeling of indoor environments. **I. J. Robotics Res.**, v. 31, p. 647–663, 2012.
- HU, Y. Cufusion: Accurate real-time camera tracking and volumetric scene reconstruction with a cuboid. In: **Sensors**. [S.l.: s.n.], 2017.
- IZADI, S. et al. Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera. In: **UIST**. [S.l.: s.n.], 2011.
- JANCOSEK, M.; PAJDLA, T. Multi-view reconstruction preserving weakly-supported surfaces. **CVPR 2011**, p. 3121–3128, 2011.
- KAZHDAN, M. M.; BOLITHO, M.; HOPPE, H. Poisson surface reconstruction. In: **Symposium on Geometry Processing**. [S.l.: s.n.], 2006.
- KELLER, M. et al. Real-time 3d reconstruction in dynamic scenes using point-based fusion. In: . [S.l.: s.n.], 2013. p. 8.
- KOLMOGOROV, V.; ZABIH, R. Multi-camera scene reconstruction via graph cuts. In: **Proceedings of the 7th European Conference on Computer Vision-Part III**. Berlin, Heidelberg: Springer-Verlag, 2002. (ECCV '02), p. 82–96. Disponível em: <<http://dl.acm.org/citation.cfm?id=645317.756415>>.
- LUN, Z. et al. 3d shape reconstruction from sketches via multi-view convolutional networks. **2017 International Conference on 3D Vision (3DV)**, p. 67–77, 2017.
- NEWCOMBE, R. A. et al. Kinectfusion: Real-time dense surface mapping and tracking. **2011 10th IEEE International Symposium on Mixed and Augmented Reality**, p. 127–136, 2011.
- NIESSNER, M. et al. Real-time 3d reconstruction at scale using voxel hashing. **ACM Trans. Graph.**, v. 32, p. 169:1–169:11, 2013.
- ROTH, H.; VONA, M. Moving volume kinectfusion. In: **BMVC**. [S.l.: s.n.], 2012.
- RUSINKIEWICZ, S.; LEVOY, M. Efficient variants of the icp algorithm. In: **3DIM**. [S.l.: s.n.], 2001.
- SAXENA, A.; CHUNG, S. H.; NG, A. Y. 3-d depth reconstruction from a single still image. **International Journal of Computer Vision**, v. 76, p. 53–69, 2007.
- SEITZ, S. M. et al. A comparison and evaluation of multi-view stereo reconstruction algorithms. **2006 IEEE Computer Society Conference on Computer Vision and Pattern**

Recognition (CVPR'06), v. 1, p. 519–528, 2006.

SILBERMAN DEREK HOIEM, P. K. N.; FERGUS, R. Indoor segmentation and support inference from rgbd images. In: **ECCV**. [S.l.: s.n.], 2012.

SONG, S.; LICHTENBERG, S. P.; XIAO, J. Sun rgbd: A rgbd scene understanding benchmark suite. **2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**, p. 567–576, 2015.

Steinbrucker, F. et al. Large-scale multi-resolution surface reconstruction from rgbd sequences. In: **2013 IEEE International Conference on Computer Vision**. [S.l.: s.n.], 2013. p. 3264–3271. ISSN 1550-5499.

STEINBRUCKER, F.; STURM, J.; CREMERS, D. Volumetric 3d mapping in real-time on a cpu. **2014 IEEE International Conference on Robotics and Automation (ICRA)**, p. 2021–2028, 2014.

STURM, J. et al. A benchmark for the evaluation of rgbd slam systems. **2012 IEEE/RSJ International Conference on Intelligent Robots and Systems**, p. 573–580, 2012.

WHELAN, K. F. et al. Kintinuous: Spatially extended kinectfusion. In: **AAAI 2012**. [S.l.: s.n.], 2012.

WHELAN, T. et al. Elasticfusion: Real-time dense slam and light source estimation. **I. J. Robotics Res.**, v. 35, p. 1697–1716, 2016.

XIAO, J.; OWENS, A.; TORRALBA, A. Sun3d: A database of big spaces reconstructed using sfm and object labels. **2013 IEEE International Conference on Computer Vision**, p. 1625–1632, 2013.

ZHOU, Q.-Y.; MILLER, S.; KOLTUN, V. Elastic fragments for dense scene reconstruction. **2013 IEEE International Conference on Computer Vision**, p. 473–480, 2013.

ZHU, Y. et al. Target-driven visual navigation in indoor scenes using deep reinforcement learning. **CoRR**, abs/1609.05143, 2016. Disponível em: <<http://arxiv.org/abs/1609.05143>>.

ZOLLHÖFER, M. et al. Real-time non-rigid reconstruction using an rgbd camera. **ACM Trans. Graph.**, v. 33, p. 156:1–156:12, 2014.

_____. Real-time non-rigid reconstruction using an rgbd camera. **ACM Trans. Graph.**, v. 33, p. 156:1–156:12, 2014.

_____. State of the art on 3d reconstruction with rgbd cameras. **Comput. Graph. Forum**, v. 37, p. 625–652, 2018.