

Metodologia e Implementação de uma Arquitetura para Interatividade Manual Egocêntrica em Realidade Aumentada Monocular

Lucas Pampolin Laheras¹, Éverton Cardoso Acchetta², Vinicius Luiz Oliveira Pinheiro dos Santos³,
Helmuth August Risch Filho⁴, Paulo Sérgio Silva Rodrigues⁵
¹⁻⁵ Ciência da Computação, Centro Universitário FEI
lucaslaheras@hotmail.com, psergio@fei.edu.br

Resumo: Este artigo propõe uma arquitetura de RA que utiliza somente uma câmera RGB monocular, permitindo que o usuário interaja com a experiência RA utilizando suas próprias mãos, onde é possível manusear o objeto virtual da mesma maneira que um objeto real seria manipulado. Os resultados obtidos são promissores, onde a verificação da interação da mão com o objeto virtual funcionou em aproximadamente 80% dos testes efetuados, respeitando o caminho definido pela movimentação da mão.

1. Introdução

Os investimentos em Realidade Aumentada (RA) vêm crescendo consideravelmente nos últimos anos, atingindo US\$ 4.1 bilhões em 2019. Este avanço é devido ao aumento da utilização da RA em áreas como educação, treinamentos, jogos e medicina. Somado a isso, os avanços tecnológicos em *hardware* possibilitam dispositivos que a poucos anos atrás, eram impensáveis. Um exemplo popular é o *Microsoft HoloLens 2*, que permite que o usuário utilize as próprias mãos como meio de interagir com a experiência em RA. Porém, a desvantagem deste dispositivo é seu alto custo devido a seus diversos sensores.

Estudos, como o apresentado em [1] mostram que a utilização de RA na educação resulta em maior engajamento dos alunos, além de permitir melhor entendimento de assuntos que exijam visualização de conceitos. Também na área de educação, a RA tem sido utilizada para treinamentos especializados, permitindo que mais pessoas tenham um treinamento apropriado e em qualquer local, sem a necessidade de contato real com o objeto de treinamento.

Na atualidade, de acordo com a pesquisa bibliográfica feita para o desenvolvimento deste trabalho, existem três tipos de interação com objeto virtual. Interação na tela, onde para a movimentação do objeto virtual é executada a partir de um mouse ou uma tela tátil [2]. Interação com gestos específicos, onde são selecionados alguns gestos que ao serem reconhecidos ativam uma função [3]. Interação com a mão, fazendo com que a mão consiga interagir como se o objeto virtual fosse real [4], mas com a tecnologia atual esse tipo de interação acaba sendo dependente de muitos sensores. Assim, este trabalho propõe e implementa uma metodologia computacional de RA para criar interatividade entre gestos manuais e objetos virtuais com câmera monocular sem marcadores artificiais.

2. Metodologia



Figura 1 – Diagrama esquemático da metodologia proposta.

A arquitetura, ilustrada na Figura 1, utilizada neste trabalho utilizou algoritmos de mapeamento de ambiente e posicionamento de câmera, como o *Oriented FAST and Rotated BRIEF - Simultaneous Localization And Mapping* (ORB-SLAM), seguido por técnicas de detecção de superfícies, e reconhecimento de objetos como o *You Look Only Once* (YOLO). Em paralelo, é usado um sistema chamado *GANerated Hands* [5], que detecta mãos humanas e fornece as coordenadas espaciais de suas articulações, para que, ao final utilizando uma *engine* 3D chamada Panda3D, seja possível realizar a manipulação de um objeto virtual com uma mão real.

3. Resultados

A fim de medir a confiabilidade deste projeto, foram propostas cinco categorias de experimentos, com o objetivo de validar as etapas da metodologia proposta. Essas categorias são divididas em dois estágios: experimentos com módulos isolados (classificação de superfícies horizontais, detecção de objetos definidos e detecção do ângulo dos objetos) e experimentos com implementação completa (definição de resolução da câmera utilizada e interação do objeto virtual com os movimentos de mão).

Para os resultados da classificação de superfície foram feitos 15 experimentos. Dessa amostragem, 13 experimentos foram considerados como sucesso e os outros 2 experimentos foram considerados como falha, pois não atenderam os requisitos necessários. Portanto, como dos 15 experimentos 13 foram considerados como sucesso, esse teste ficou com precisão de aproximadamente 86,67%.

No experimento de detecção de superfícies foi verificada a precisão do *bounding box* e a intersecção da

união (IOU) para cada uma das 150 imagens que foram capturadas posteriormente, enviadas ao YOLO e feito o recorte manual, foram geradas porcentagens de assertividades e IOU após o treinamento. Todos os objetos no experimento foram classificados corretamente com taxa 100% de verdadeiros positivos. A Tabela 1 apresenta os resultados referentes a IOU médio, desvio padrão e variância para cada objeto e para todos os objetos considerados.

	IOU Média	Desvio Padrão	Variância
Carrinho	93,56%	3,85%	0,1480%
DualShock4	90,57%	5,27%	0,2776%
IamGroot	95,23%	1,87%	0,0349%
MiniCraque	93,67%	3,02%	0,0911%
PlayStation2	91,86%	3,66%	0,1341%
Todos os objetos	92,98%	4,04%	0,1630%

Tabela 1 – Tabela relacionando a média de assertividade, desvio padrão e variância de cada objeto.

Nos experimentos de detecção do ângulo dos objetos quantitativos, onde a quantidade de pontos fiduciais encontrados dentro do *bounding box* fornecido pelo YOLO foi comparada com os pontos de referência da base utilizada para treinamento do YOLO, com ângulos pré-definidos. Os resultados deste teste estão representados na Tabela 2.

	640x480	800x600	960x720	1280x960	1440x1080	1920x1440
Carrinho (150 imagens)	102	124	145	147	150	150
DualShock4 (150 imagens)	20	37	66	127	139	150
IamGroot (150 imagens)	37	64	82	112	115	138
MiniCraque (150 imagens)	53	71	94	130	136	150
PlayStation2 (150 imagens)	80	115	138	149	149	150

Tabela 2 – Tabela relacionando a média de assertividade, desvio padrão e variância de cada objeto

Para os testes de resolução, foram utilizadas as resoluções 1080p, 720p e 640p. Após os testes, calculou-se a média da taxa de atualização de *frames* do sistema. Os resultados foram: 47 *fps* para resolução 640x480, 35 *fps* para resolução 1280x720 e 18 *fps* para resolução 1920x1080.

Nos experimentos realizados para validar a interação da mão com objeto (ilustrada na Figura 2), foi levado em consideração a precisão da interação entre a malha da mão e objeto virtual projetado na cena. Assim, foram definidos pontos iniciais e finais para que o objeto seja empurrado. Foram realizados 15 testes ao total, onde 12 vezes em que foi feita essa simulação, o objeto chegou ao ponto final percorrendo os 40 centímetros pré-definidos, respeitando o caminho definido pela movimentação da mão. Em 3 dessas situações, o objeto não atingiu com sucesso ao destino. Com isso, a precisão deste teste foi de 80%.

4. Conclusões

Analisando o principal experimento, a verificação da interação da mão com o objeto virtual onde obteve 80% de assertividade, os três resultados considerados como negativos, tiveram como justificativa o fato do *GanHands* apresentar a falha onde o polegar gera movimentos não reais, movendo o objeto de maneira não desejada. Portanto, pode-se considerar que este trabalho atingiu com sucesso o objetivo proposto. Para trabalhos futuros constam algumas melhorias e até novas implementações, tais como: Aprimoramento

do algoritmo para reconhecimento da mão e estruturação da malha no ambiente 3D; Adição de objetos virtuais que permitam uma interação mais imersiva com o usuário; Aplicação deste trabalho para novos contextos de estudo, como por exemplo medicina, treinamento, suporte remoto e escolarização.

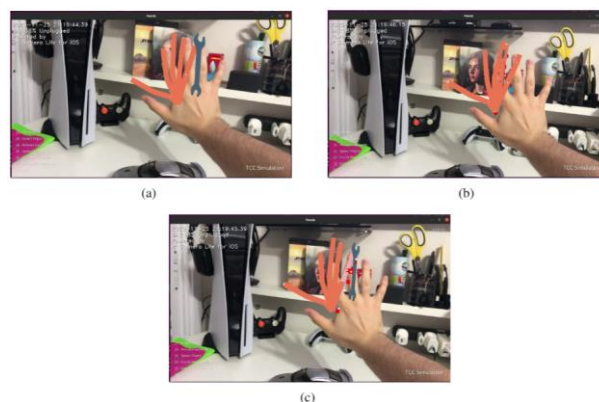


Figura 2 – Visualização da interação (a) Mão quase encostando no objeto. (b) Mão tangenciando o objeto. (c) Mão empurrando o objeto.

5. Referências

- [1] WU, H.-K. et al. Current status, opportunities and challenges of augmented reality in education. *Computers Education*, v. 62, p. 41 – 49, 2013. ISSN 0360-1315.
- [2] BALDAUF, M.; LASINGER, K.; FRÖHLICH, P. Private public screens: Detached multi-user interaction with large displays through mobile augmented reality. In: *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia*. New York, NY, USA: Association for Computing Machinery, 2012.
- [3] KARAMBAKHS, A. et al. Deep gesture interaction for augmented anatomy learning. *International Journal of Information Management*, Elsevier, v. 45, p. 328–336, 2019.
- [4] O-larnnithipong, N. et al. Hand tracking interface for virtual reality interaction based on marg sensors. In: *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. [S.l.: s.n.], 2019. p. 1717–1722.
- [5] MUELLER, F. et al. Generated hands for real-time 3d hand tracking from monocular rgb. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2018. p. 49–59

Agradecimentos

Este projeto foi construído como o Trabalho de Conclusão de Curso dos autores dentro do Laboratório de Sistemas Imersivos (immersivelab.fei.edu.br), apoiado pelo Centro Universitário FEI dentro do Projeto VRFEI.