

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

**Aluno:** Éverton Cardoso Acchetta, Helmuth August Risch Filho, Lucas Pampolin Laheras e Vinicius Luiz Oliveira Pinheiro Dos Santos (eve.023@hotmail.com, helmuthfilho001@gmail.com, lucaslaheras@hotmail.com, viniciuslops@hotmail.com)

**Orientador:** Prof. Dr. Paulo Sérgio Silva Rodrigues (psergio@fei.edu.br)

**Resumo:** Os investimentos em Realidade Aumentada (RA) vem crescendo consideravelmente nos últimos anos, atingindo US\$ 4.1 bilhões em 2019. Este avanço se deve 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 à seus diversos sensores. Assim, este projeto propõe uma arquitetura de RA que utiliza como sensor somente uma câmera RGB monocular, permitindo que o usuário interaja com a experiência RA utilizando suas próprias mãos para realizar gestos similares à arquitetura do *Microsoft HoloLens 2*, 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.

**Descrição:** Atualmente, a maioria das aplicações de RA interativas tem interação realizada por meio de toques na tela ou gestos pré-definidos. Porém, com o avanço da tecnologia de hardware, novos dispositivos especializados em RA estão surgindo e com isso, novos métodos de interação homem-máquina podem ser necessários. Um exemplo de dispositivo é o *Microsoft HoloLens 2*, que utiliza diversos sensores e câmeras (portanto trata-se de um dispositivo de alto custo) para executar comandos por meio de interações de malhas.

Assim, este trabalho propõe uma arquitetura que permita criar uma experiência em RA utilizando como sensor, somente uma câmera RGB monocular. Esta experiência permitirá que o usuário interaja com o objeto virtual de maneira similar ao *Microsoft HoloLens 2*, onde o usuário poderá mover o objeto virtual com sua própria mão. Esta arquitetura poderá ser utilizada como novo método de interação homem-máquina para hardware de baixo custo.

Os resultados obtidos são promissores, onde a verificação da interação da mão com o objeto virtual funcionou em aproximadamente 86,6% dos testes efetuados. Com relação ao treinamento do YOLO para a identificação dos objetos, foi obtido uma assertividade média de 92,98%.

Este trabalho realizou com sucesso a interação entre o objeto virtual e a malha da mão, onde todo o mapeamento de ambiente, detecção da mão, detecção do objeto e computação da interação são realizados via software de forma fluida (35 fps). Isso é um avanço significativo, visto que pode ser aplicado em diversas áreas de nossa sociedade, tais como medicina e educação.

**Metodologia utilizada:** É composta por 8 etapas principais, ilustradas na Figura 1. O processo se inicia na Etapa (1), onde são obtidas imagens através de uma câmera. O fluxo correspondente à Etapa (2) inicia com um processo de localização da câmera e mapeamento do ambiente (Etapa 2.1), seguido de duas opções: A primeira (Etapa 2.2.1) consiste em identificação de superfícies horizontais e suas utilizações como âncoras; A segunda (Etapa 2.2.2) se trata de identificação de objetos previamente treinados e também suas utilizações como âncoras. Finalizando esta etapa, tem-se o posicionamento do objeto virtual (Etapa 2.3). Iniciando o fluxo da Etapa (3), tem-se a detecção e segmentação de mão (Etapa 3.1). Após este processo, é realizada uma correspondência do esqueleto da mão com o resultado obtido anteriormente (Etapa 3.2). Por último, tem-se a sobreposição do esqueleto no frame (Etapa 3.3). Unindo os dois fluxos paralelos, tem-se a Etapa (4), responsável pela interação do esqueleto obtido na etapa (Etapa 3.2) com o objeto virtual posicionado na cena na etapa (Etapa 2.3).



Figura 1 - Diagrama esquemático da metodologia proposta.

**Resultados:** Com relação aos resultados obtidos verifica-se que, em relação à classificação de superfícies, foi atingida uma média de precisão de 80%, podendo-se então considerar que esse teste é promissor. Por sua vez, para os testes de verificação de reconhecimento de *bounding box*, a menor taxa de assertividade foi do objeto DualShock4, que é de 90,57%. Portanto, considerando que a média de assertividade de todos os objetos foi de 92,98% (Figura 2), pode-se considerar que esse experimento também é promissor. A Tabela 1 representa a quantidade de testes de detecção da angulação dos objetos que obtiveram sucesso, onde a margem de erro é de no máximo 30 graus. O principal teste, sendo o que verifica o objetivo deste trabalho, trata-se da verificação da interação da mão com o objeto virtual, ilustrado na Figura 3, tendo como amostragem 15 (quinze) experimentos, dos quais, em 13 (treze) o objeto foi movido com sucesso até seu destino, respeitando o caminho definido pela movimentação da mão.

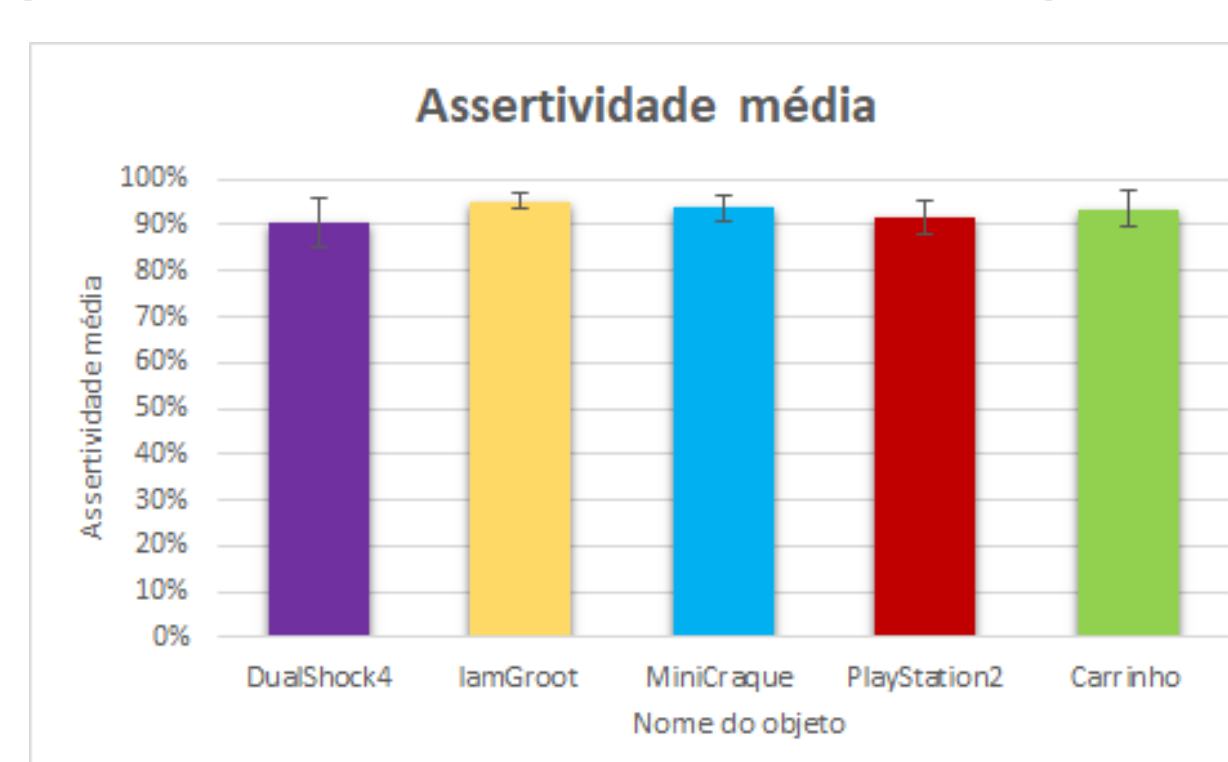


Figura 2 - Gráfico de assertividade média, gerado pelo YOLO, dos objetos, com seus respectivos desvios padrão

	640x480	800x600	960x720	1280x960	1440x1080	1920x1440
Carrinho	51	65	75	82	90	96
DualShock4	8	19	34	68	73	79
IamGroot	32	57	74	97	100	117
MiniCraque	15	25	35	44	66	76
PlayStation2	37	47	61	72	77	100

Tabela 1 - Tabela dos testes qualitativos, relacionando as resoluções das imagens com os objetos, exibindo a quantidade de imagens onde os ângulos foram identificados com sucesso



Figura 3 - Visualização da interação entre a malha da mão real e o objeto virtual, os modelos incluem uma chave inglesa e um carrinho. Quadrados vermelhos indicam contato da mão com o modelo virtual.

**Conclusão:** Este trabalho realizou com sucesso essa interação utilizando uma câmera RGB monocular, onde todo o mapeamento de ambiente, detecção da mão, detecção do objeto e computação da interação são realizados via software de forma fluida (35 fps), porém como pode ser observado em alguns resultados, a precisão com hardware limitado acaba sendo menor se comparado ao *Microsoft HoloLens 2*, que permite inclusive a realização de treinamentos de cirurgias.