Realidade Virtual - T2

Professor: Márcio Sarroglia Pinho Aluno: Felipe Freitas Silva Data de apresentação: 18/10/2024

Trends in Virtual and Augmented Reality Research: A Review of Latest Eye Tracking Research Papers and Beyond

Publicação: Agosto 2019

Local: University of Delaware

Autores: Jicheng Li and Roghayeh Barmaki

Li, J.; Barmaki, R. Trends in Virtual and Augmented Reality Research: A Review of Latest Eye Tracking Research Papers and Beyond. Preprints 2019, 2019090019.

<u> https://doi.org/10.20944/preprints201909.0019.v</u>

Abstract

Although proposition of "Virtual Reality" (VR) and "Augment Reality" (AR) can be traced back to the 60s, both areas are actually blooming in recent decades. Thanks to the latest deep learning techniques, an enormous advance in computer vision research community has taken place. Since VR and AR are highly related to computer vision tasks, these areas have enjoyed the benefits as well. Yet there is no sufficient survey on such impact and new research areas arising from it. This paper mainly focuses on the latest research progress in ACM Symposium on Eye Tracking Research & Applications (ETRA) 2019, as well as several recent representative paper works. It aims to figure out the influence of deep learning techniques on latest VR/AR research. Meanwhile, new issues have popped up with the development of VR and AR technology, such as privacy and computation efficiency. This paper draws attention to such newly produced topics as well. In addition, this paper also investigates on the effect of latest VR and AR techniques on people, such as level of teamwork in collaborative tasks, assistance and treatment to patients and the disabled, etc.

1. Introdução

Definição de Realidade Virtual (RV)

- Simulação gerada por computador de ambiente 3D
- Requerem equipamentos extras
 - HMD (Head-Mounted Displays)
 - Controladores



Sistemas Imersivos de RV x Não Imersivos

Sistemas não imersivos normalmente são obtidos por meio de um computador e não obstruem a vista do usuário, mas uma experiência sub-par. O paper foca em sistemas imersivos, um exemplo conhecido citado foi o jogo Beat Saber.



Realidade Aumentada (AR)

Versão "aprimorada" da realidade, sobrepõe informações digitais sobre o mundo real.

- Filtros do Snapchat
- Pokemon GO

Pode ser construtiva ou destrutiva, isto é, acrescentar algo no mundo ou ocultar. É interativo em tempo real; mais "responsivo".



VR x AR e Deep Learning

- AR: Adiciona elementos digitais ao mundo real visíveis com dispositivos como smartphones.
- VR: Cria um mundo virtual imersivo, normalmente ocultando o mundo real.

- Deep Learning: Métodos de aprendizado de representação com múltiplos níveis de abstração.
- Redes Neurais Convolucionais (CNNs):
 Superam o desempenho humano em tarefas de visão, como classificação.
- Aplicação em VR/AR: Melhora o rastreamento ocular, um elemento essencial para VR/AR.

2. Método

Objetivo do Método:

Revisar artigos publicados em conferências e workshops especializados em rastreamento ocular (eye tracking) no contexto de Realidade Virtual (VR) e Realidade Aumentada (AR).

Fontes Principais de Pesquisa:

ERTA 2019 (incluindo eventos co-localizados e workshops):

COGAIN (Communication by Gaze

ET4S (Eye Tracking For Spatial Research)

ETVIS (Eye Tracking And Visualization)

ETWEB (Eye Tracking For The Web)

Critérios de Seleção de Artigos:

Pesquisa de artigos com as palavras-chave: "Virtual Reality", "Augmented Reality" e "Eye Tracking".

Artigos publicados nos últimos 3 anos foram pré-selecionados.

Triagem mais refinada focou em trabalhos relacionados a:

Deep Learning

Aprendizado colaborativo

Multimodalidade

Interaction)

3. Tendências de Pesquisa e Direções Futuras

3.1 Atenção Visual

Atenção Visual e Predição de Saliência

Orientação Espacial em Ambientes Virtuais

J. L. Louedec et al. [11] propuseram uma rede neural profunda que combina abordagens top-down e bottom-up para prever a atenção visual de jogadores de xadrez, gerando mapas de saliência significativos a partir de dados de rastreamento ocular.





Figure 1: Saliency map from a chess game. At left, input chess board image. At centre, eye tracking results (points represent eye fixations, lines are scan path between fixations). At right, saliency map computed from eye tracking. For each pixel, a probability between 0 (black) and 1 (white) is computed. Adopted from [11, Fig. 1].

Em [16], foi comparada a orientação da atenção espacial em ambientes virtuais, dada por dicas visuais e auditivas. Descobriu-se que pistas válidas tornavam os participantes mais rápidos nas tarefas, com diferentes modalidades de pistas (auditiva, visual) orientando a atenção global ou localmente.

Atenção Visual e Predição de 3000 ms Fixation Endogenous Exogenous Auditory Visual Auditory Visual J. L. Louedec et al. [11] propus neural profunda que combina (g e bottom-up para prever a ater jogadores de xadrez, gerando significativos a partir de dados ocular. **Ipsilateral** Contralateral Valid Invalid

Figure 1: Saliency map from a chess Figure 2: From top to bottom the time course of a trial: fixation of 3 board image. At centre, eye tracking reseconds, cueing of 300 ms according to the block modality and type, fixations, lines are scan path between fixa300 ms inter-stimulus interval, and 300 ms target presentation with a computed from eye tracking. For each psmall red ball. Adopted from [16, Fig. 2]. (black) and 1 (white) is computed. Adopted monitoring research psmall red ball.

m Ambientes Virtuais

a a orientação da atenção s virtuais, dada por dicas scobriu-se que pistas válidas ntes mais rápidos nas tarefas, idades de pistas (auditiva, enção global ou localmente.

Treinamento de Atenção Conjunta com Realidade Virtual

Simulação de Respostas em Treinamentos de Enfermagem

O trabalho de [12] apresentou uma abordagem de treinamento de atenção conjunta com um professor virtual customizável, que guiava o usuário em um jogo de tambores, utilizando feedbacks multimodais para aumentar o engajamento e melhorar o foco nas áreas relevantes.



Figure 3: The User Interface for customizing the CVH Teacher. Adopted from [12, Fig. 3].

Participantes interagiram com um paciente virtual em um ambiente de treinamento, e foi demonstrado que animações conversacionais e não conversacionais conseguiram capturar a atenção visual dos participantes [18].



Figure 4: Screenshot shows a participant interacting with the virtual patient in the RRTS, and recording his vitals in the EHR screen. Adopted from [18, Fig. 1].

Otimização de Elementos Visuais em Ambientes Virtuais

Predição de Olhar Sem Hardware Externo

Rawan et al. [14] propuseram um método de otimização de layout baseado em dados de atenção visual, colocando elementos em posições que maximizam o tempo de visualização, conforme ilustrado no layout de um museu virtual.

O modelo SGaze [8] utilizou correlações entre o movimento da cabeça e dos olhos para prever o olhar em tempo real, sem a necessidade de rastreadores oculares externos, sendo útil para tarefas de exploração passiva em cenas visuais.





Figure 5: Left: an input 3D scene with its corresponding layout. Right: the optimal placement of visual elements that will attain the target gaze duration. The eyes depict the camera location and angle in taking the screenshots. Adopted from [14, Fig. 1].

Otimização de Elementos Visuais em Ambientes

Predição de Olhar Sem Hardware Externo

Virtuais

Rawan e otimizaç visual, co maximiz ilustrado





e o er o

ara Jais.

Figure 5: Left: an input 3D scene with its corresponding layout. Right: the optimal placement of visual elements that will attain the target gaze duration. The eyes depict the camera location and angle in taking the screenshots. Adopted from [14, Fig. 1].

3.2 Aprendizagem Colaborativa

Colaboração em Jogos: Efeitos do Compartilhamento de Atenção Visual

Efeitos de Sistemas AR com HMD em Colaborações Profissionais

Spakov et al. [19] investigaram o impacto do compartilhamento de informações sobre a atenção visual em um jogo colaborativo. Os jogadores exploravam uma casa escura para encontrar chaves, utilizando o olhar da cabeça (direção que a pessoa está encarando) ou o olhar dos olhos (onde a pessoa está focando). Comparando versões de baixa imersão (desktop) e alta imersão (VR-HMD), os resultados indicaram que o compartilhamento da informação do olhar na condição de alta imersão melhorou o desempenho geral dos jogadores.



Figure 7: A screenshot from the game. Highlighted areas represent visual attention of both players. Adopted from [19, Fig. 2].

Em [13], os efeitos do uso de sistemas AR com HMDs foram analisados em tarefas colaborativas de resposta a crises simuladas. Profissionais da polícia, corpo de bombeiros e força aérea alternaram entre HMDs e mapas de papel para coordenar as operações. Embora o contato visual tenha sido baixo em ambas as condições, os participantes relataram maior confiança e confiança nas ferramentas com HMDs, sugerindo que a falta de contato visual não afetou negativamente a colaboração no contexto profissional.



Figure 8: Participants using HMD-based AR system as auxiliary technical tools. Adopted from [13, Fig. 2].

Colaboração er de Atenção Vis

Spakov et al. [1] compartilhame visual em um jo exploravam um utilizando o olh encarando) ou ofocando). Comp (desktop) e alta indicaram que olhar na condiç desempenho que



laborações

R com HMDs
as de resposta
ia, corpo de
HMDs e
ções. Embora o
s as condições,
ça e confiança
ue a falta de
a colaboração

Figure 7: A screenshot from the game. Highlighted areas represent visual attention of both players. Adopted from [19, Fig. 2].

Colaboração em Jogos: Efeitos do Compartilhamento de Atenção Visual

Efeitos de Sistemas AR com HMD em Colaborações Profissionais

Spakov et al. [19] investigaram o impacto do compartilhamento de informações sobre a atenção visual em um jogo colaborativo. Os jogadores exploravam uma casa escura para encontrar chaves, utilizando o olhar da cabeça (direção que a pessoa está encarando) ou o olhar dos olhos (onde a pessoa está focando). Comparando versões de baixa imersão (desktop) e alta imersão (VR-HMD), os resultados indicaram que o compartilhamento da informação do olhar na condição de alta imersão melhorou o desempenho geral dos jogadores.



Figure 7: A screenshot from the game. Highlighted areas represent visual attention of both players. Adopted from [19, Fig. 2].

Em [13], os efeitos do uso de sistemas AR com HMDs foram analisados em tarefas colaborativas de resposta a crises simuladas. Profissionais da polícia, corpo de bombeiros e força aérea alternaram entre HMDs e mapas de papel para coordenar as operações. Embora o contato visual tenha sido baixo em ambas as condições, os participantes relataram maior confiança e confiança nas ferramentas com HMDs, sugerindo que a falta de contato visual não afetou negativamente a colaboração no contexto profissional.



Figure 8: Participants using HMD-based AR system as auxiliar technical tools. Adopted from [13, Fig. 2].

Colaboração em Jogos de Atenção Visual

Spakov et al. [19] inves compartilhamento de i visual em um jogo cola exploravam uma casa utilizando o olhar da ca encarando) ou o olhar focando). Comparando (desktop) e alta imersa indicaram que o comp olhar na condição de a desempenho geral dos



em Colaborações

mas AR com HMDs porativas de resposta la polícia, corpo de m entre HMDs e poperações. Embora o a ambas as condições, confiança e confiança rindo que a falta de mente a colaboração

Figure 8: Participants using HMD-based AR system as auxiliary technical tools. Adopted from [13, Fig. 2].

3.3 Detecção de Emoções

Reconstrução Facial em Tempo Real para Emoções em VR

S. Chen et al. [2] desenvolveram um sistema que captura e reconstrói rostos 3D em tempo real enquanto os usuários estão usando headsets de realidade virtual (HMDs). Como parte do rosto é ocultada pelo HMD, o sistema utiliza duas câmeras infravermelhas embutidas nos óculos de VR para capturar imagens dos olhos e uma câmera adicional externa para capturar as expressões faciais. Sensores de postura na cabeça, integrados a um celular, fornecem dados para ajustar a pose. Esse conjunto de informações alimenta um avatar, que reflete com precisão as emoções do usuário, conforme



mostrado no Figura 9.

Figure 9: 3D facial expression reconstruction and eye gaze tracking, (a) The picture captured by an extra RGB camera to show the setup., (b) The three captured IR images. (c) The reconstructed 3D face and eye gaze. (d) An avatar driven by the captured 3D face. Adopted from [2, Fig. 3)

Avatares Expressivos Utilizando Redes Neurais Convolucionais (CNNs)

Hickson et al. [7] propuseram um algoritmo que infere expressões faciais, mesmo quando parte do rosto está oculta pelo HMD, gerando avatares dinâmicos e expressivos em tempo real. Para aumentar a precisão da rede neural convolucional, eles introduziram a "personalização", que ajusta o modelo com base nas características faciais do usuário. Utilizando apenas as imagens dos olhos capturadas por uma câmera infravermelha integrada no headset VR, o sistema é capaz de inferir um subconjunto de expressões faciais, sem a necessidade de câmeras externas adicionais (Figura 10).

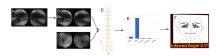
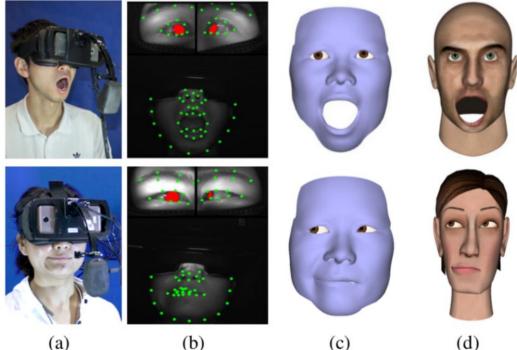


Figure 10: A: Raw eye images from the HMD. B: Rectified eye images C: The average neutral image for this user session, used for personalization. D: The difference between the rectified headset image and the mean neutral image is the input to a deep neural network. In the non personalization case, the mean neutral image is not subtracted from the rectified image. E: Output takes the form of a distribution over expressions. F: This distribution is used to generate an expressive avatar. Adopted from [7, Fig. 3].

Reconstrução Facial em T

S. Chen et al. [2] desenvolve reconstrói rostos 3D em te estão usando headsets de parte do rosto é ocultada câmeras infravermelhas e capturar imagens dos olho para capturar as expressõ cabeça, integrados a um ca pose. Esse conjunto de i que reflete com precisão a mostrado no Figura 9.



des Neurais

goritmo que infere parte do rosto está dinâmicos e expressivos cisão da rede neural personalização", que cterísticas faciais do ens dos olhos capturadas grada no headset VR, o onjunto de expressões eras externas adicionais

Figure 9: 3D facial expression reconstruction and eye gaze tracking. (a) The picture captured by an extra RGB camera to show the setup. (b) The three captured IR images. (c) The reconstructed 3D face and eye gaze. (d) An avatar driven by the captured 3D face. Adopted from [2, Fig. 3].

Reconstrução Facial em Tempo Real para Emoções em VR

S. Chen et al. [2] desenvolveram um sistema que captura e reconstrói rostos 3D em tempo real enquanto os usuários estão usando headsets de realidade virtual (HMDs). Como parte do rosto é ocultada pelo HMD, o sistema utiliza duas câmeras infravermelhas embutidas nos óculos de VR para capturar imagens dos olhos e uma câmera adicional externa para capturar as expressões faciais. Sensores de postura na cabeça, integrados a um celular, fornecem dados para ajustar a pose. Esse conjunto de informações alimenta um avatar, que reflete com precisão as emoções do usuário, conforme mostrado no Figura 9.

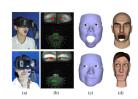


Figure 9: 3D facial expression reconstruction and eye gaze tracking, (a) The picture captured by an extra RGB camera to show the setup., (b) The three captured IR images. (c) The reconstructed 3D face and eye gaze. (d) An avatar driven by the captured 3D face. Adopted from [2, Fig. 3)

Avatares Expressivos Utilizando Redes Neurais Convolucionais (CNNs)

Hickson et al. [7] propuseram um algoritmo que infere expressões faciais, mesmo quando parte do rosto está oculta pelo HMD, gerando avatares dinâmicos e expressivos em tempo real. Para aumentar a precisão da rede neural convolucional, eles introduziram a "personalização", que ajusta o modelo com base nas características faciais do usuário. Utilizando apenas as imagens dos olhos capturadas por uma câmera infravermelha integrada no headset VR, o sistema é capaz de inferir um subconjunto de expressões faciais, sem a necessidade de câmeras externas adicionais (Figura 10).

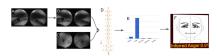
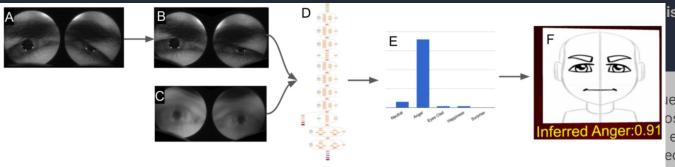


Figure 10: A: Raw eye images from the HMD. B: Rectified eye images C: The average neutral image for this user session, used for personalization. D: The difference between the rectified headset image and the mean neutral image is the input to a deep neural network. In the non personalization case, the mean neutral image is not subtracted from the rectified image. E: Output takes the form of a distribution over expressions. F: This distribution is used to generate an expressive avatar. Adopted from [7, Fig. 3].

Reconstrução F

S. Chen et al. [2] reconstrói rosto estão usando he parte do rosto é câmeras infrave capturar imager a pose. Esse co



para capturar as Figure 10: A: Raw eye images from the HMD. B: Rectified eye images los capturadas cabeça, integrac C: The average neutral image for this user session, used for personal-expressões que reflete com ization. D: The difference between the rectified headset image and as adicionais mostrado no Figthe mean neutral image is the input to a deep neural network. In the non personalization case, the mean neutral image is not subtracted from the rectified image. E: Output takes the form of a distribution over expressions. F: This distribution is used to generate an expressive avatar. Adopted from [7, Fig. 3].

ie infere osto está e expressivos ede neural ação", que faciais do

3.4 Acessibilidade e Inclusão

Treinamento de Atenção Conjunta para Crianças com Transtorno do Espectro Autista (TEA)

J. Liu et al. [12] desenvolveram o jogo de realidade virtual Imagination Drum, utilizando um Humano Virtual Customizável (CVH) como professor virtual. Este jogo foi projetado para treinar a atenção conjunta de crianças com Transtorno do Espectro Autista (TEA). A pesquisa mostrou que o Imagination Drum aumentou a capacidade de comunicação das crianças, capturou sua atenção, e melhorou seu engajamento em atividades sociais, como aulas de música na escola.

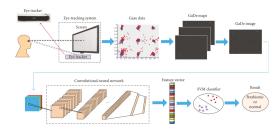


Figure 11: The proposed strabismus recognition framework. Adopted from [3, Fig. 1].

Reconhecimento de Estrabismo com Eye Tracking

Z. Chen et al. [3] propuseram uma abordagem de reconhecimento de estrabismo utilizando dados de rastreamento ocular e redes neurais convolucionais profundas. Eles converteram os dados brutos de rastreamento ocular em GaDe maps (Mapas de Desvio do Olhar), representando o padrão de desvio ocular e alimentando redes neurais para classificar diferentes tipos de estrabismo. A técnica demonstrou eficiência na identificação de estrabismos recessivos, intermitentes e manifestos, conforme mostrado nas Figuras 11 e 12.

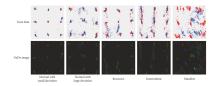


Figure 12: Examples of gaze data and corresponding GaDe images.Red *: left gaze data. Blue x: right gaze data. Colors in the second row represent the R, G, and B channels of GaDe images. The The first two columns represent normal data with small deviation and large deviation. The third, fourth, and fifth columns represent data of, respectively, recessive strabismus, intermittent strabismus, and manifest strabismus. Adopted from [3, Fig. 4].

Treinament com Transto

Gaze data

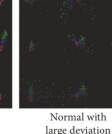


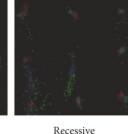


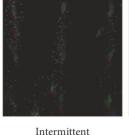
Tracking

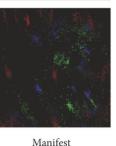
J. Liu et al. |
virtual Imag
Virtual Cust GaDe image
Este jogo fo
conjunta de
Autista (TE/
Imagination
comunicaçã
melhorou se

Normal with









volucionais rutos de pas de o de desvio classificar

agem de

dados de

manifestos,

Figure 12: Examples of gaze data and corresponding GaDe images.Red *: left gaze data. Blue ×: right gaze data. Colors in the second row represent the R, G, and B channels of GaDe images. The The first two columns represent normal data with small deviation and large deviation. The third, fourth, and fifth columns represent data of, respectively, recessive strabismus, intermittent strabismus, and manifest strabismus. Adopted from [3, Fig. 4].

3.5 Privacidade&3.6 Eficiência

Privacidade em Rastreamento Ocular

Eficiência Computacional no Rastreamento Ocular

Steil et al. [17] abordaram preocupações de privacidade relacionadas a câmeras de primeira pessoa usadas em dispositivos de rastreamento ocular. Para mitigar essas preocupações, eles propuseram o sistema PrivacEye, que utiliza uma rede neural profunda pré-treinada para detectar cenas sensíveis. Quando uma cena sensível é detectada, o obturador da câmera é acionado automaticamente, bloqueando a captura de imagens. Figura 13 mostra o funcionamento do PrivacEye.

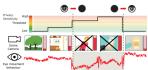


Figure 13: PrivacEye uses a mechanical camera shutter (top) to preserve users 'and pystanders' privacy with head-mounted eye trackers. Privacy-sensitive situations are detected by combining deep scene image and eye movement features (midelle) while changes in eye movement behaviour alone trigger the reopening of the camera shutter (bottom). Adopted from [17, Fig. 1].

Katrychuk et al. [9] propuseram uma solução para melhorar a eficiência computacional de sensores de rastreamento ocular em plataformas de realidade aumentada e virtual. Eles utilizaram photosensor oculography (PSOG), que capta reflexos de infravermelho dos olhos. O desafio é que os dispositivos PSOG sofrem com a degradação de desempenho devido ao deslocamento de sensores. A pesquisa aplicou uma solução baseada em aprendizado de máquina para corrigir esse problema, conforme ilustrado na Figura 14.

Privacidade er

Steil et al. [17] privacidade rel pessoa usadas ocular. Para m propuseram o rede neural procenas sensíve detectada, o ol automaticame imagens. Figur PrivacEye.

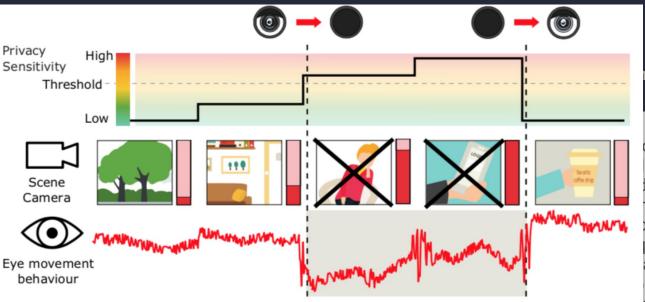


Figure 13: PrivacEye uses a mechanical camera shutter (top) to preserve users 'and bystanders 'privacy with head-mounted eye trackers. Privacy-sensitive situations are detected by combining deep scene image and eye movement features (middle) while changes in eye movement behaviour alone trigger the reopening of the camera shutter (bottom). Adopted from [17, Fig. 1].

nento Ocular

olução para e sensores de le realidade notosensor os de jue os adação de de sensores. ada em esse a 14.

4. Conclusão

 Este artigo revisou os avanços recentes em pesquisas de Realidade Virtual (VR) e Realidade Aumentada (AR), com ênfase em rastreamento ocular, aprendizado profundo e colaboração em ambientes virtuais.

 Avanços Significativos: Há grandes avanços no uso de VR e AR, especialmente no mercado de jogos, com essas tecnologias sendo cada vez mais aplicadas fora de ambientes laboratoriais. Potencial Futuro: O progresso em hardware e inteligência artificial (IA) oferece um enorme potencial para a aplicação de VR e AR em uma ampla gama de áreas, com futuras inovações aguardadas à medida que essas tecnologias continuam a evoluir.

Trabalhos Relacionados

- Moreno-Arjonilla, J., López-Ruiz, A., Jiménez-Pérez, J.R. et al. Eye-tracking on virtual reality: a survey. Virtual Reality 28, 38 (2024). https://doi.org/10.1007/s10055-023-00903-y
- Zhou, Feng & Duh, Henry & Billinghurst, Mark. (2008). <u>Trends in Augmented Reality Tracking, Interaction and Display: A Review of Ten Years of ISMAR.</u> 2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality. 2. 193-202. 10.1109/ISMAR.2008.4637362.
- Abdullah M. Al-Ansi, Mohammed Jaboob, Askar Garad, Ahmed Al-Ansi, Analyzing augmented reality (AR) and virtual reality (VR) recent development in education, Social Sciences & Humanities Open, Volume 8, Issue 1, 2023, 100532, ISSN 2590-2911, https://doi.org/10.1016/j.ssaho.2023.100532.

- Adhanom IB, MacNeilage P, Folmer E. Eye Tracking in Virtual Reality: a Broad Review of Applications and Challenges. Virtual Real. 2023 Jun;27(2):1481-1505. doi: 10.1007/s10055-022-00738-z. Epub 2023 Jan 18. PMID: 37621305; PMCID: PMC10449001.

Lamb Maurice, Brundin Malin, Perez Luque Estela, Billing Erik. Eye-Tracking Beyond Peripersonal Space in Virtual Reality: Validation and Best Practices. Frontiers in Virtual Reality. 3. 2022. https://www.frontiersin.org/journals/virtual-reality/articles/10.3389/frvir.2022.864653. 2673-4192

Eye Tracking Research in Virtual Reality

by LEAD-ME COST Action

10.3389/11VII.2022.804033. 20/3-4192

Moreno-Ario Eye-tracking (2024). http

> Zhou, Feng Augmented of Ten Years Symposium 10.1109/ISN

Abdullah M. Al-Ansi, Ana recent devel Open, Volun

https://doi.org/10.1016/j.ssano.2023.100532.

Captions Object

. Caption Emitter . Caption Manager lications and oub 2023 Jan 449001. personal Space Practices. virtual-reality/ar

Eye Tracking in

(2):1481-1505.

Luque Estela,

Dúvidas?

Obrigado:)

