

Um estudo bibliométrico e analítico sobre a produção científica relacionada a modelos computacionais de saliência visual (2010–2025)

Felipe Mattei Ximenes¹, Aline Farias Gomes de Sousa¹

¹Universidade Federal do Sul e Sudeste do Pará (UNIFESSPA)
Faculdade de Engenharia da Computação (FEC)
Marabá-PA – Brasil

felipemattei@unifesspa.edu.br, alinefarias@unifesspa.edu.br

Resumo. Esta pesquisa apresenta uma análise bibliométrica de modelos de saliência visual entre 2010 e 2025, com base nas plataformas IEEE Xplore e SBC OpenLib. Empregou-se um pipeline em Python, com pandas e matplotlib, para analisar tendências temporais, principais veículos e autores, bem como a co-ocorrência de palavras-chave. Observou-se pico de 35 publicações em 2018, média de 4.43 autores por artigo e alta recorrência dos termos “video saliency”, “visual attention”, “deep learning” e “saliency detection”. Destacam-se frentes emergentes associadas a vídeo 360° e atenção multimodal. A IEEE Xplore apresentou 291 publicações, no entanto não foram localizados estudos na SBC OpenLib dentro dos critérios adotados, sinalizando lacuna de cobertura nacional e oportunidade de pesquisa.

1. Introdução

Saliência visual descreve regiões de um estímulo com maior probabilidade de atrair a atenção humana. Trabalhos seminais desta área [Itti et al. 1998] inspiraram novas direções no estudo da atenção visual humana. Em vídeo, a saliência audiovisual [Zhu et al. 2025] resulta da combinação de pistas visuais (cor, contraste, movimento) e auditivas (fala, eventos sonoros), com aplicações em compressão, Qualidade de Experiência (QoE), edição e ambientes imersivos. Avanços em *deep learning* e em *transformers* multimodais [Li and Liu 2025] consolidaram o campo, enquanto métricas como Área sob a Curva (AuC), Saliência do *Scanpath* Normalizado (NSS) e Coeficiente de Correlação (CC), aliadas a bases com *eye-tracking*, viabilizam avaliação comparável entre métodos.

Este estudo apresenta um panorama de publicações em saliência visual no período 2010–2025, com foco nas bases IEEE Xplore e SBC OpenLib, destacando evolução temporal, veículos (periódicos e eventos), autores e palavras-chave. Contribuindo assim com um mapeamento quantitativo atualizado, identificando lacunas temáticas e metodológicas.

2. Metodologia

Foram consultadas as bases IEEE Xplore e SBC OpenLib, considerando publicações no período de 2010 a 2025. A estratégia de busca empregou a expressão: ("audiovisual saliency"OR "audio-visual saliency"OR "audiovisual saliency model"OR "modelo de saliência audiovisual"OR "audiovisual

```
attention"OR "audio-visual attention"OR "multimodal saliency"OR "video saliency")
```

Foram considerados resultados provenientes de periódicos e eventos. Os registros foram exportados em formato tabular (csv), contendo título, autores, ano, veículo de publicação, DOI, palavras-chave e resumo.

O processamento foi realizado em Python, utilizando *pandas* e *matplotlib*. O *pipeline* compreendeu: (i) contagem de publicações por ano, (ii) identificação dos veículos, autores e coautorias mais recorrentes, e (iii) análise de frequência e coocorrência de palavras-chave. Os totais por base e estatísticas descritivas (por exemplo, número de publicações e média de autores por artigo) são apresentados na seção 3 por meio de gráficos analíticos e grafos construídos no *Gephi*.

3. Resultados

3.1. Evolução temporal

O gráfico da Figura 1 apresenta a evolução anual de publicações entre 2010 e 2025, distinguindo artigos de periódicos (linha vermelha) e eventos (linha laranja), além do total anual (barras azuis). Observa-se crescimento expressivo a partir de 2015, com pico em 2018 (35 artigos) e manutenção de um volume elevado de publicações até 2024. Nenhum artigo relacionado ao tema foi encontrado na base SBC OpenLib, enquanto na IEEE Xplore foram identificadas 291 publicações.

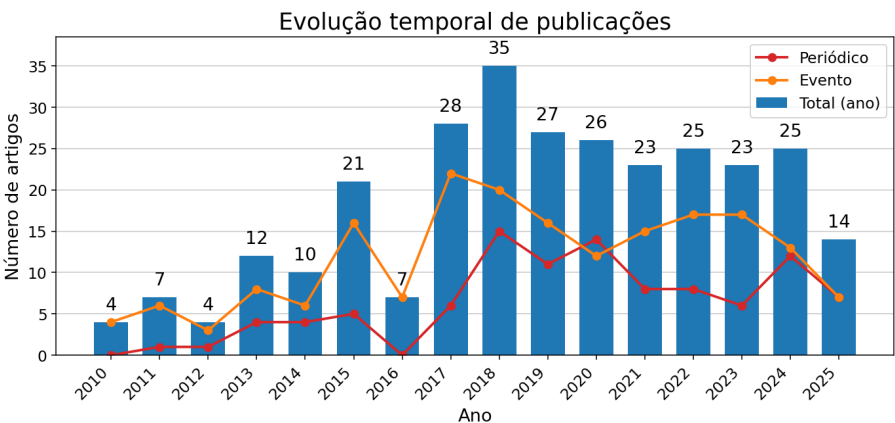


Figura 1: Publicações por ano (IEEE Xplore).

3.2. Veículos e autores

A Tabela 1 apresenta os cinco principais veículos de publicação, com destaque para periódicos IEEE voltados ao processamento de imagem e vídeo, além de eventos de visão computacional e multimídia [Li and Liu 2025]. A Figura 2 exhibe o grafo de coautoria, evidenciando os autores mais recorrentes e suas conexões, com colaboração média de 4.43 autores por artigo.

Tabela 1: Periódicos/Eventos mais frequentes.

Periódico	Nº de artigos
<i>IEEE Transactions on Image Processing</i>	26
<i>IEEE Transactions on Circuits and Systems for Video Technology</i>	24
<i>IEEE Transactions on Multimedia</i>	13
<i>IEEE Access</i>	9
<i>IEEE Transactions on Pattern Analysis and Machine Intelligence</i>	6

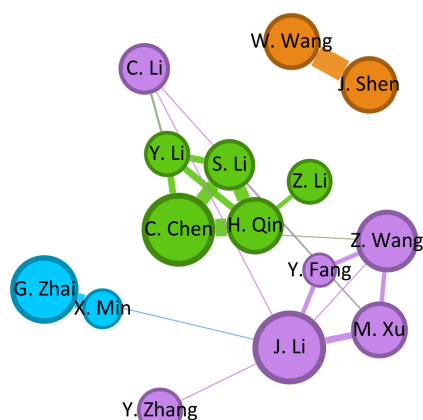


Figura 2: Grafo de coautoria (Top-N). Tamanho do nó proporcional ao número de publicações; cor por comunidade; espessura da aresta proporcional ao número de coautorias.

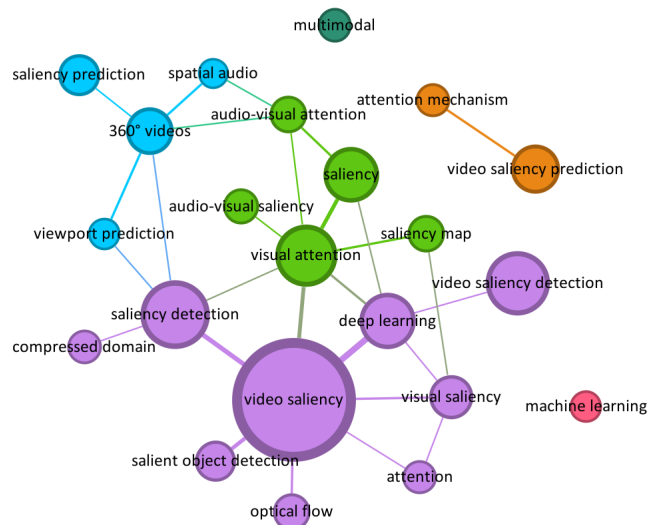


Figura 3: Rede de frequência e pares de coocorrência de palavras-chave (Top-N). Tamanho do nó proporcional à frequência; cor por comunidade; espessura da aresta proporcional à coocorrência.

3.3. Palavras-chave e coocorrência

A Figura 3 mostra a rede de frequência e coocorrência das palavras-chave. O tamanho dos nós reflete a frequência dos termos, as cores indicam grupos temáticos (palavras-chave mais relacionadas entre si) e a espessura das arestas indica a intensidade da coocorrência. Destacam-se os termos *video saliency*, *saliency detection*, *visual attention* e *deep learning*, além de tópicos emergentes como *360° videos* e *spatial audio*, com conexões fortes, por exemplo, entre *deep learning* e *video saliency* [Chen et al. 2025, Girmaji et al. 2025].

4. Discussão e conclusão

Na base IEEE Xplore, a produção científica sobre modelos de saliência visual apresentou crescimento contínuo entre 2010 e 2017, com pico em 2018 (35 artigos), acompanhando a popularização das abordagens baseadas em *deep learning*. A colaboração média de 4.43 autores por artigo reflete o caráter interdisciplinar do tema, que envolve visão computacional e aprendizado de máquina. Já a busca na base da SBC não retornou resultados, sugerindo baixa representatividade nacional nesse domínio específico.

O grafo de coautoria (Figura 2) evidencia grupos produtivos e colaborações concentradas, enquanto a rede de palavras-chave (Figura 3) revela um núcleo conceitual formado por *video/visual saliency*, *visual attention* e *deep learning*, além de clusters temáticos voltados a conteúdo imersivo (vídeo 360°/VR, *viewport prediction*) e à multimodalidade (áudio espacial/*ambisonics*). Esses padrões sugerem consolidação do eixo supervisionado com mecanismos de atenção e adoção crescente de arquiteturas modernas em cenários imersivos.

Em síntese, a área mostra-se em consolidação e diversificação, com tendência de avanço em aplicações audiovisuais complexas e ambientes imersivos, abrindo espaço para novas pesquisas voltadas a modelos computacionais de saliência visual e à ampliação da produção científica em contextos brasileiros.

Referências

- [Chen et al. 2025] Chen, C., Ma, G., Song, W., Li, S., Hao, A., and Qin, H. (2025). Saliency-free and aesthetic-aware panoramic video navigation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- [Girmaji et al. 2025] Girmaji, R., Jain, S., Beri, B., Bansal, S., and Gandhi, V. (2025). Minimalistic video saliency prediction via efficient decoder spatio temporal action cues. In *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.
- [Itti et al. 1998] Itti, L., Koch, C., and Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 20(11):1254–1259.
- [Li and Liu 2025] Li, C. and Liu, S. (2025). Tm2sp: A transformer-based multi-level spatiotemporal feature pyramid network for video saliency prediction. *IEEE Transactions on Circuits and Systems for Video Technology*.
- [Zhu et al. 2025] Zhu, Y., Duan, H., Zhang, K., Zhu, Y., Zhu, X., Teng, L., Min, X., and Zhai, G. (2025). How does audio influence visual attention in omnidirectional videos? database and model. *IEEE Transactions on Image Processing*.