

Autoria Semiautomática de Efeitos Sensoriais em Vídeos 360°

Raphael Abreu
Laboratório MídiaCom
Universidade Federal Fluminense
Niterói, Brasil
raphael.abreu@midia.com.uff.br

Joel A. F. dos Santos
GPM - CEFET/RJ
Rio de Janeiro, Brasil
jsantos@eic.cefet-rj.br

Débora C. Muchaluat-Saade
Laboratório MídiaCom
Universidade Federal Fluminense
Niterói, Brasil
debora@midia.com.uff.br

ABSTRACT

Creating multisensory experiences (mulsemmedia) in 360° videos faces the challenge of synchronizing effects, a process that is laborious when manual and inflexible when fully automatic. This thesis proposes a semi-automatic approach that integrates deep learning (DL) recommendations while maintaining author control. The core proposal is a scalable cloud architecture with multiple recognition modules and an object tracking algorithm for spatial continuity, implemented in the SensorySynch360 web tool. A study with 81 users demonstrated that the approach significantly reduces authoring effort and time, while preserving usability and quality of experience.

KEYWORDS

Autoria 360°, mulsemmedia, autoria semi-automática, efeitos sensoriais, análise de conteúdo

1 INTRODUÇÃO

Mulsemmedia (Multiple Sensorial Media) expande o audiovisual tradicional ao integrar estímulos multissensoriais — como vento, calor, vibração e aromas — sincronizados com o conteúdo de mídia [3]. Vídeos 360° são ideais para essas experiências imersivas, de forma que efeitos sensoriais elevam a Qualidade de Experiência (QoE) do usuário. Porém, a criação de experiências imersivas em vídeos 360° enfrenta um desafio: a sincronização manual de efeitos sensoriais (vento, calor, etc.) com eventos audiovisuais é uma tarefa laboriosa, demorada e propensa a erros, especialmente pela necessidade de definir a localização espacial dos efeitos em todas as direções do espaço 360°. A complexidade da tarefa manual é acentuada em situações como a sincronização de efeitos recorrentes (e.g., vibrações em um filme de ação), a perda de sincronia após edições no conteúdo audiovisual e a animação espaço-temporal de efeitos que devem seguir objetos em movimento na esfera 360°. Abordagens totalmente automáticas, embora rápidas, são rígidas, ignoram a intenção artística do autor e frequentemente falham em generalizar para conteúdos fora de seu domínio de treinamento [1].

Diante deste cenário, a questão de pesquisa central desta tese é: *"Como um método semi-automático afeta a autoria de efeitos sensoriais em vídeos 360°?"*. Para responder a essa questão, a tese avança além da simples criação de uma ferramenta, propondo como contribuição principal uma arquitetura conceitual escalável e uma metodologia de autoria que equacionam os desafios espaço-temporais do

conteúdo 360°. Essa abordagem une a eficiência da automação via IA com a expressividade do controle manual do autor, estabelecendo um fluxo de trabalho validado empiricamente.

2 ARQUITETURA CONCEITUAL

A tese propõe inicialmente um componente para acoplar redes neurais profundas (DNNs) a qualquer ferramenta de autoria mulsemmedia para vídeos 2D. O componente denominado CDC (*Content-Driven Component*) viabiliza essa integração por meio de dicionários configuráveis que associam rótulos — gerados pelos algoritmos de reconhecimento — a efeitos sensoriais definidos pelo autor (e.g., "praia" gera efeito de "vento") conferindo flexibilidade e controle ao processo.

No entanto, ao migrar para vídeos 360°, surgem desafios adicionais relacionados à espacialidade, orientação e persistência temporal de efeitos — como a necessidade de localização do efeito, definição de *keyframes* e rastreamento contínuo de elementos na esfera 360°. Para lidar com essa complexidade, a tese estende o princípio do CDC com uma arquitetura genérica e escalável, apresentada na Figura 1. Ela é composta por cinco estágios: **Map**, que distribui o conteúdo para múltiplos módulos; **Pre-process**, que adapta o formato da mídia (e.g., para projeção cúbica, a fim de minimizar distorções); **Recognition**, que executa os algoritmos; **Post-process**, que converte os resultados para coordenadas esféricas; e **Filter & Combine**, que aplica regras e dicionários definidos pelo autor para fundir as saídas e gerar os efeitos finais. Essa estrutura modular permite a integração de diferentes técnicas de reconhecimento, desde DNNs até visão computacional clássica, de forma cooperativa — como os desenvolvidos nesta tese.

3 IMPLEMENTAÇÃO E FERRAMENTA

A arquitetura foi implementada em um *pipeline* utilizando serviços em nuvem para execução escalável de quatro módulos de reconhecimento. Foi criada a ferramenta SensorySynch360 que oferece uma interface com visualizador 360°, linha do tempo e painéis de propriedades. Ela permite a criação manual e semi-automática de efeitos, suportando animações complexas baseadas em *keyframes*. No fluxo semiautomático, o autor seleciona os módulos de reconhecimento e inicia a análise. A ferramenta aciona a execução escalável dos módulos em nuvem, que processam o vídeo e retornam os rótulos detectados. Por fim, a interface apresenta esses rótulos ao autor para que ele realize a associação com os efeitos sensoriais, mantendo total controle sobre o resultado criativo.

Como os módulos de reconhecimento operam analisando o vídeo a cada segundo, os objetos detectados são independentes em cada instante de tempo. Para criar animações contínuas de efeitos que seguem objetos em movimento, foi necessário desenvolver um algoritmo de rastreamento (*tracking*) de objetos. Este algoritmo

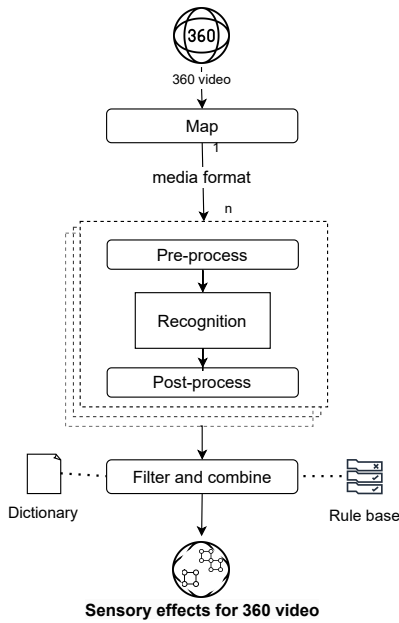


Figura 1: Visão geral da arquitetura de reconhecimento de efeitos sensoriais em 360°.

garante a continuidade espaço-temporal ao associar as *bounding boxes* de um mesmo objeto entre frames consecutivos do vídeo 360°. A associação é feita através de uma abordagem baseada em grafos, onde a similaridade entre anotações de efeitos é calculada em coordenadas esféricas para lidar com as particularidades da geometria 360°.

4 AVALIAÇÕES E RESULTADOS

A avaliação da proposta foi realizada em duas fases. Primeiramente, uma avaliação preliminar focada em vídeos 2D ($n=34$), utilizando o CDC integrado à ferramenta STEVE 2.0 [2], já apontava para os benefícios da abordagem. Os participantes reportaram que a funcionalidade era útil e fácil de usar e, objetivamente, o tempo para realizar a tarefa de autoria foi reduzido em mais de 50%, motivando a investigação em cenários 360° mais complexos.

O estudo principal avaliou a proposta com 81 participantes. Inicialmente, todos foram instruídos sobre os objetivos da pesquisa e, na sequência, concordaram com os termos do Termo de Consentimento Livre e Esclarecido (TCLE) antes de iniciar as atividades. Após essa etapa, os participantes compararam a autoria manual com a semi-automática na ferramenta SensorySynch360. Foram coletados dados objetivos (e.g., ‘totalClicks’) e subjetivos (questionários SUS, NASA-TLX, UEQ-s). Os resultados objetivos mostram que o modo semi-automático reduziu significativamente o tempo de tarefa, o número de cliques e a quantidade de modificações manuais nos efeitos. A Figura 2 apresenta a métrica de tempo de duração de autoria entre diferentes vídeos e métodos de autoria. Os *box-plots*

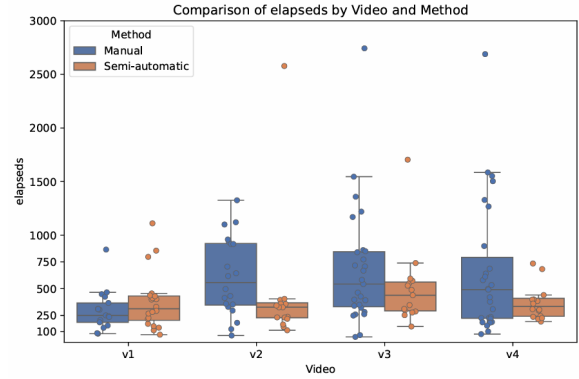


Figura 2: Tempo de execução de tarefas para cada vídeo no método manual e semi-automático.

exibem a distribuição de duração para os vídeos v1 e v4. Os gráficos demonstram que as tarefas de autoria manual de efeitos, em geral, demandaram mais tempo quando comparadas às tarefas usando autoria semi-automática.

A análise do NASA-TLX indicou que o modo semi-automático foi percebido como levemente menos demandante mentalmente e exigindo menor esforço, sem aumentar a frustração do usuário. A usabilidade do sistema foi avaliada como “boa” (SUS Score de 75.5) e a experiência do usuário (UEQ-s) manteve-se alta e comparável em ambos os modos, validando que a eficiência foi obtida sem sacrificar a qualidade da experiência do autor. O feedback qualitativo revelou que, embora a ferramenta seja intuitiva, a principal dificuldade da autoria manual é a manipulação de efeitos no espaço 360°, uma tarefa que é acelerada pelo modo semi-automático.

5 CONCLUSÃO

Esta tese apresentou contribuições fundamentais para a autoria mulsemídia em 360°, formalizadas em uma arquitetura de software modular e em um fluxo de trabalho semi-automático que equilibram automação e controle criativo. A implementação e avaliação empírica dessa proposta, materializada na ferramenta SensorySynch360, demonstraram sua viabilidade e eficácia. O trabalho contribui com a arquitetura, a ferramenta e um rico conjunto de dados de avaliação que podem guiar futuras pesquisas na área de sistemas multimídia inteligentes e imersivos. Como trabalhos futuros, vislumbra-se a expansão dos módulos de reconhecimento, a criação de um dataset público para autoria 360°, bem como a exploração de mecanismos para autoria colaborativa.

REFERÊNCIAS

- [1] Raphael Silva de Abreu, Douglas Mattos, Joel dos Santos, Gheorghita Ghinea, and Débora Christina Muchaluat-Saade. 2021. Toward Content-Driven Intelligent Authoring of Mulsemedia Applications. *IEEE MultiMedia* 28, 1 (2021), 7–16. DOI: <http://dx.doi.org/10.1109/MMUL.2020.3011383>
- [2] Douglas Paulo de Mattos, Débora C Muchaluat-Saade, and Gheorghita Ghinea. 2020. An Approach for Authoring Mulsemedia Documents Based on Events. In *International Conference on Computing, Networking and Communications, 2020*. IEEE, 7.
- [3] Gheorghita Ghinea, Christian Timmerer, Weisi Lin, and Stephen R. Gulliver. 2014. Mulsemedia : State of the Art, Perspectives, and Challenges. *ACM Transactions on Multimedia Computing, Communications, and Applications* 11, 1s (2014), 1–23. DOI: <http://dx.doi.org/10.1145/2617994>