

BumbAR: Multimedia Authoring Through Augmented Reality

Paulo R. C. Mendes
mendesdpn@gmail.com

Universidade Federal do Maranhão

Carlos de S. Soares Neto
csalles@deinf.ufma.br

Universidade Federal do Maranhão

ABSTRACT

This work describes the evolution of an approach based on augmented reality for authoring multimedia presentations. The BumbAR approach is based on the NCM model and explores the use of augmented reality and real-world objects (markers) as an innovative user interface for describing the behavior and relationships between the media objects that are part of a multimedia presentation. It was evaluated through a qualitative study based on the TAM model. The qualitative study aimed at evaluating users' attitude towards using BumbAR. The results showed that the participants found that the proposed approach is both useful and easy-to-use, but comments made by them showed the need of including new functionalities in the BumbAR tool, which are described in this paper.

KEYWORDS

Augmented Reality, Multimedia, Authoring Tool, User Interface

1 INTRODUÇÃO

A realidade aumentada (RA) é definida por Carmigniani et al. [4] como uma visão direta ou indireta em tempo real de um elemento físico do mundo real que foi aumentado através da adição de informação virtual a ele. Essas informações podem ser imagens, áudios, vídeos, toque, ou sensações hápticas sobrepostas em tempo real [3].

Ferramentas de autoria multimídia suportam o desenvolvimento de produtos multimídia (como vídeos interativos e apresentações) através da integração de objetos de mídia de diferentes tipos (como imagens, sons e vídeos) de uma maneira sincronizada e com significado. Apesar da qualidade de um conteúdo multimídia ser consideravelmente dependente da qualidade individual dos objetos de mídia que o compõe, é somente com uma composição apropriada de todos esses objetos em uma apresentação que os usuários finais irão, verdadeiramente, ter uma experiência de qualidade.

Uma maneira comum de representar apresentações multimídia interativas é através de documentos multimídia, como a linguagem de marcação de hipertexto, do inglês *Hypertext Markup Language* (HTML) [7], a linguagem de integração multimídia sincronizada, do inglês *Synchronized Multimedia Integration Language* (SMIL) [14], e a linguagem de contexto aninhado, do inglês *Nested Context Language* (NCL) [16].

Apesar dos conceitos usados nos modelos de documentos supracitados serem de fácil compreensão, a representação textual deles não é de fácil utilização por não-programadores, que geralmente estão mais familiarizados com representações visuais. Ferramentas de autoria multimídia são programas que permitem a concepção de

apresentações multimídia com o objetivo de simplificar o processo de autoria, permitindo que não-programadores possam criar tais apresentações. De fato, o desenvolvimento de um método que simplifique o processo de autoria deve aumentar a satisfação do usuário em desenvolver apresentações multimídia, abrindo novas possibilidades para pessoas criativas conceberem experiências multimídia. Como, atualmente, a forma mais comum de interação com um computador é ainda através de uma interface gráfica do usuário (do inglês *Graphical User Interface (GUI)*), grande parte das ferramentas de autoria multimídia atuais adotam esse paradigma.

Diferente dessas ferramentas, este trabalho propõe um processo de autoria multimídia baseado em realidade aumentada. Nesse processo, ao invés do uso de uma GUI comum, é utilizada a RA como uma forma inovadora de interface com o usuário.

O restante deste trabalho está dividido da seguinte forma: a Seção 2 detalha o processo de autoria proposto, a Seção 3 mostra os resultados obtidos da avaliação desse processo, e a Seção 4 conclui este trabalho e discute os resultados obtidos.

2 AUTORIA MULTIMÍDIA COM RA

A principal contribuição deste trabalho é a evolução de um processo de autoria baseado em realidade aumentada para a autoria de apresentações multimídia. Tal processo é implementado na ferramenta de autoria *BumbAR tool*, que permite a criação de conteúdo multimídia, bem como a visualização do resultado final. Além disso, a apresentação criada pode ser apresentada em navegadores de internet ou aparelhos de TV digital.

No que se segue, são discutidos o processo de autoria (Seção 2.1) e sua implementação na ferramenta *BumbAR tool* (Seção 2.2).

2.1 Processo de Autoria

O processo de autoria de apresentações multimídia consiste em duas fases distintas: *configuração de mídias* e *criação de elos*.

Na fase de *configuração de mídias*, os autores devem definir os objetos de mídia que fazem parte da apresentação e suas propriedades. Além disso, cada objeto de mídia é associado a um dos marcadores de realidade aumentada. Um *objeto de mídia* é identificado por seu nome e pelo seu localizador de recurso uniforme, do inglês *Uniform Resource Locator* (URL), que aponta para o arquivo de mídia correspondente. No contexto desta proposta, objetos de mídia englobam vídeos, imagens e áudios. Tais objetos possuem propriedades que definem como as mídias a eles associadas são apresentadas. Na abordagem *BumbAR*, as propriedades suportadas são as comumente utilizadas em apresentações multimídia em duas dimensões, como posição na tela (x e y), tamanho (largura e altura), transparência, e volume.

Na fase de *criação de elos*, é definida a evolução da apresentação multimídia no decorrer do tempo, e como ela reage a interações do usuário. Para fazer isso, é utilizada uma interface com o usuário baseada em RA, em que marcadores de RA pré-configurados para











In: XVI Workshop de Trabalhos de Iniciação Científica (WTIC 2019), Rio de Janeiro, Brasil. Anais Estendidos do Simpósio Brasileiro de Sistemas Multimídia e Web (Web-Media). Porto Alegre: Sociedade Brasileira de Computação, 2019.
ISSN 2596-1683

os *objetos de mídia* são dinamicamente combinados a marcadores de RA relacionados ao comportamento dos *objetos de mídia* da apresentação. Tal combinação é feita por meio da colisão, no mundo real, de marcadores relacionados à *objetos de mídia* com marcadores relacionados a comportamento, resultando na criação de relacionamentos entre objetos de mídia. Adicionalmente, para enfatizar os diferentes marcadores e dar retorno às ações do usuário, cada marcador de realidade aumentada é renderizado como um cubo 3D, que é referenciado como *Bloco RA* no restante deste trabalho.

Os relacionamentos definidos entre *objetos de mídia* seguem o modelo NCM [15]. Elos (mais especificamente, elos causais) são criados a partir da colisão de *blocos RA* relacionados a *objetos de mídia e comportamentos*. Um elo causal determina que, quando uma condição é satisfeita, uma ou mais ações são executadas. Cada ação e condição suportada é representada por um marcador RA e um *bloco RA* específicos.

A Tabela 1 mostra exemplos de cada um dos tipos de *Blocos RA* suportados. Os símbolos usados para os marcadores e *Blocos RA*, entretanto, variam de acordo com a condição, ação ou *objeto de mídia* específicos. Por exemplo, apesar de seguirem o mesmo modelo, o marcador e *Bloco RA* de uma condição *onEnd* tem um nome e símbolo diferentes de um marcador e *Bloco RA* de uma condição *onBegin*, mostrados na primeira linha da Tabela 1. A Tabela 2 mostra as condições e ações individuais a serem suportadas pela ferramenta *BumBAR*. Os ícones usados para cada uma delas são baseados na notação visual criada por Laiola Guimarães et al. [10].

Tabela 1: Os modelos de blocos de realidade aumentada da proposta BumBAR

Bloco	Uso	Marcador RA	Bloco RA
Condição	É usado para representar condições.		
Ação	É usado para representar ações.		
Mídia	É usado para representar objetos de mídia.		
Inicial	É usado para definir objetos de mídia que começam a tocar assim que a apresentação multimídia começa.		
Deleção	É usado para remover elos previamente criados entre objetos de mídia.		

Para criar um elo causal entre dois *objetos de mídia*, o usuário deve seguir os seguintes passos:

- (1) Marcar o *bloco de mídia* que é parte da condição do elo, criando um *bloco de mídia-com-condição*;
- (2) Marcar o *bloco de mídia* que é parte da ação do elo, criando um *bloco de mídia-com-ação*;
- (3) Associar o *bloco de mídia-com-condição* com o *bloco de mídia-com-ação*, criando um elo causal;

Tabela 2: Condições e ações usadas da proposta BumBAR.

Nome	Papel	Significado
onBegin	condição	Ativado quando o <i>objeto de mídia</i> começa a tocar.
onEnd	condição	Ativado quando o <i>objeto de mídia</i> para ou acaba.
onPause	condição	Ativado quando o <i>objeto de mídia</i> é pausado.
onResume	condição	Ativado quando o <i>objeto de mídia</i> continua a ser executado.
start	ação	<i>Objeto de mídia</i> inicia
stop	ação	<i>Objeto de mídia</i> para.
pause	ação	<i>Objeto de mídia</i> é pausa
resume	ação	<i>Objeto de mídia</i> continua.

O passo 1 acima é realizado arrastando e colidindo um *bloco de condição* com um *bloco de mídia*. Fazendo-se isso, um ícone que representa a condição usada é adicionado ao canto superior direito do *bloco de mídia*, e este torna-se um *bloco de mídia-com-condição*. Um *bloco de mídia-com-condição* é representado pelo mesmo *bloco de mídia* que foi usado para criá-lo, mas marcado com uma condição (para fazer parte do elo). A Figura 1 mostra esse processo (é possível observar que o ícone aparece no canto superior direito do *bloco de mídia*). Os passos 2 e 3 são feitos de modo similar.

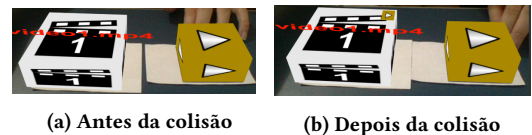


Figura 1: Bloco de mídia se tornando um bloco de mídia-com-condição

A abordagem proposta também possui o *bloco inicial*. Tal bloco é usado para marcar um *bloco de mídia* informando que o *objeto de mídia* associado a ele deve ser iniciado assim que a apresentação multimídia sendo criada começa. De modo semelhante aos processos descritos anteriormente, deve-se arrastar e colidir um *bloco de mídia* com o *bloco inicial*. Dessa forma, um ícone com o símbolo do *bloco inicial* é adicionado ao canto superior esquerdo do *bloco de mídia*.

Além dos marcadores relacionados aos blocos, o processo proposto possui dois marcadores especiais, o *marcador de elos* e o *marcador de visão estrutural*. Esses marcadores foram adicionados à proposta com base nos resultados colhidos do experimento que é descrito na Seção 3.

O *marcador de elos* mostra, empilhados, todos os elos criados pelo usuário utilizando o processo de criação de elos descrito anteriormente. Quando um *bloco de elo* é renderizado, é colocado um texto sobreposto a ele representando o elo que o bloco representa, seguindo o padrão <objeto de mídia 1> <ação> <objeto de mídia 2>, de modo que quando a condição associada ao primeiro

objeto de mídia for satisfeita, a ação relacionada ao segundo *objeto de mídia* será executada. Caso deseje, o usuário pode remover um elo criado. Isso evita, por exemplo, que caso um erro seja cometido durante o processo de criação da apresentação multimídia, esta não precise ser refeita do começo. Para a remoção de um elo, são usados o *marcador de elos* e de *deleção* em conjunto. Para efetuar a remoção do elo, o usuário deve colidir o *bloco de deleção* com o *bloco de elo* correspondente ao elo que deseja remover. A Figura 2 mostra a remoção do elo "onBegin media2 Stop media1".



(a) Antes da colisão (b) Depois da colisão

Figura 2: Deleção de elo.

Com o objetivo de fornecer uma visão geral da apresentação sendo criada, para que o usuário possa visualizar as relações entre *objetos de mídia*, foi criado o *marcador de visão estrutural*. Tal necessidade foi evidenciada a partir dos comentários dos participantes do estudo descrito na Seção 3. Tal marcador mostra em RA os *blocos de mídia* e os elos criados entre eles de modo semelhante a um grafo, com os *blocos de mídia* sendo os nós e os elos sendo as arestas. Essa visão estrutural foi inspirada na visão estrutural do NCL Composer [2]. A Figura 3 mostra a visão estrutural de uma apresentação que possui os seguintes elos: "onBegin media2 Stop media1" e "onBegin media3 Start media4".

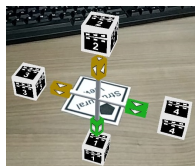


Figura 3: Exemplo de visão estrutural.

2.2 Implementação da Ferramenta

A ferramenta de autoria *BumbAR tool* foi implementada como uma cena da *engine* de jogos Unity 3D [17]. Através da reutilização de funcionalidades e objetos do Unity, a ferramenta BumbAR apresenta um ambiente tanto para a autoria, com o uso de realidade aumentada, de apresentações multimídia, quanto para tocá-las. Uma demonstração pode ser vista em <https://youtu.be/GoLuOYLlJaQ>.

Tal ferramenta possui dois modos: *modo de edição* e *modo de apresentação*. O primeiro é o modo em que as apresentações multimídia são criadas usando o processo de autoria descrito anteriormente, e o segundo modo toca as apresentações desenvolvidas.

O desenvolvimento da ferramenta foi feito em C# [8], uma das linguagens suportadas pelo Unity 3D. Para o reconhecimento de marcadores de realidade aumentada, foi utilizado o pacote Vuforia [13] para Unity 3D. Os *Blocos RA*, a implementação de colisões, e controle de *objetos de mídia* são feitos usando componentes nativos do Unity 3D. Eventos e métodos da linguagem C# são usados para implementar o sistema de *elos causais* utilizados na ferramenta; *condições* são implementadas usando eventos da linguagem C#, e *ações* são implementadas utilizando os métodos da linguagem.

Os *marcadores RA* discutidos na Seção 2.1 são suportados na ferramenta. A adição de novos marcadores (por exemplo, para suportar todos os eventos do modelo NCM) é possível e relativamente simples para pessoas com conhecimento do funcionamento do Unity 3D. Marcadores extras podem ser criados utilizando o portal de desenvolvimento do Vuforia. Para implementar novas ações e condições, é necessário implementar seus comportamentos usando eventos e métodos da linguagem C#.

Como mencionado anteriormente, além de tocar a apresentação multimídia criada, a ferramenta *BumbAR* também suporta a exportação da apresentação para a linguagem NCL. NCL é a linguagem padrão do Sistema Brasileiro de TV Digital Terrestre [1] e recomendação ITU-T H.761 da União Internacional de Telecomunicações [9]. Existem tocadores NCL gratuitos e comerciais tanto para TV [6] quanto para a Web [11].

3 RESULTADOS

Com o objetivo de avaliar a *BumbAR tool*, foi desenvolvido um estudo empírico qualitativo no qual 15 participantes, com idades entre 20 e 27 anos, desenvolveram uma apresentação multimídia utilizando a ferramenta e preencheram um questionário com suas percepções sobre ela. O principal objetivo desse estudo é analisar a atitude dos usuários em relação à criação de apresentações multimídia utilizando a ferramenta desenvolvida e, de maneira mais geral, uma interface baseada em realidade aumentada para a criação de tais apresentações.

Para desenvolver a apresentação, os participantes utilizaram uma instalação de realidade aumentada composta por um notebook, com a ferramenta *BumbAR tool*, uma webcam, e os marcadores de realidade aumentada utilizados pela ferramenta.

O questionário foi baseado no modelo de aceitação de tecnologias, do inglês *Technology Acceptance Model* (TAM) [5], que é amplamente utilizado para avaliar a aceitação de novas tecnologias. O modelo TAM propõe sentenças baseadas em dois conceitos: utilidade percebida, do inglês *perceived usefulness* (PU), e facilidade de uso percebida, do inglês *perceived ease-of-use* (PEOU). Davis et al. [5] definem PU como o nível em que uma pessoa percebe que o uso de um sistema melhoraria sua performance em fazer seu trabalho, e PEOU como o nível em que uma pessoa acredita que o uso de um sistema seria livre de esforço. O questionário completo pode ser visto em <https://bit.ly/2Z3qrbs>.

A Figura 4 mostra as respostas para cada sentença em valores absolutos e em percentuais. As sentenças elaboradas são afirmativas positivas em relação à ferramenta, e todas elas obtiveram a maioria de suas respostas positivas. A sentença PU6, que diz "eu acredito que a ferramenta é útil", e PEOU5, que diz "na minha opinião, a ferramenta é fácil de usar", obtiveram um nível significativo de concordâncias, com respostas 5 (concordo um pouco) ou superior. Esses resultados revelam que, na opinião dos participantes, a ferramenta que utiliza o processo proposto é útil e fácil de usar.

Também foram considerados comentários dos participantes sobre pontos positivos e negativos da ferramenta. De fato, os comentários dos participantes se mostraram relevantes, e seriam perdidos em um experimento puramente quantitativo. Do lado positivo, a maioria dos participantes mencionou a intuitividade da ferramenta BumbAR e que ela torna o processo de autoria agradável. Do lado negativo, a maioria dos participantes enfatizou a necessidade de

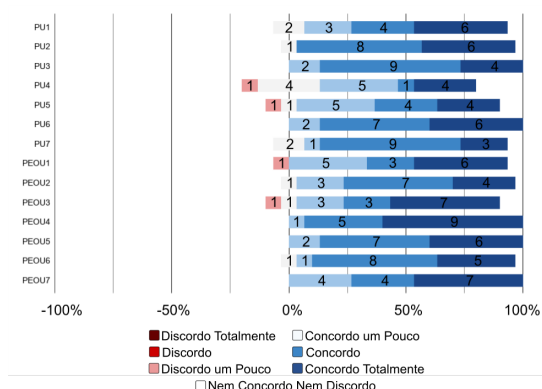


Figura 4: Respostas dos participantes sobre a ferramenta BumbAR

mostrar o histórico de elos criados, o que motivou a criação dos marcadores *de visão estrutural* e *de elos*, descritos na Seção 2.1.

4 CONCLUSÕES E DISCUSSÃO

Este trabalho de iniciação científica propôs uma abordagem baseada em realidade aumentada para o desenvolvimento de apresentações multimídia. A abordagem proposta é baseada no modelo NCM, mais especificamente em seus elos causais, condições e ações. Além disso, foi apresentada a ferramenta BumbAR, que implementa a abordagem proposta com uma implementação baseada no Unity 3D, permitindo que usuários criem, visualizem, e gerem um documento NCL de suas apresentações multimídia. Para avaliar a proposta deste trabalho, foi desenvolvido um estudo qualitativo baseado no modelo TAM. Os participantes do estudo consideraram que a proposta desenvolvida é tanto útil quanto fácil de usar. Tal estudo também permitiu a evolução da proposta, evidenciando a necessidade da criação de algo que permitisse a visualização da apresentação sendo criada como um todo.

É importante ressaltar que, neste trabalho, a realidade aumentada foi utilizada em apenas uma das etapas do processo da criação de apresentações multimídia: a criação de relações causais entre os objetos de mídia. É válido questionar a viabilidade da utilização desse tipo de abordagem na especificação espacial desses objetos de mídia. Nesse contexto, por exemplo, a especificação de que um objeto de mídia que ocupa um terço da altura da tela e dois quintos da largura não seria simples com a realidade aumentada. Isso se deve, principalmente, à falta de uma alta precisão na detecção de marcadores. Uma dificuldade motora do usuário também poderia prejudicar consideravelmente tal processo de especificação de objetos de mídia.

Uma possível solução para esse tipo de dificuldade seria o uso de descritores pré-definidos para a especificação de regiões na tela. A linguagem NCL permite a criação de descritores que podem ser utilizados por diversos objetos de mídia, facilitando o reúso. Esses descritores definem a configuração espacial de objetos de mídia. No caso da realidade aumentada, poderiam existir marcadores de descritores pré-definidos que seriam, em tempo de execução, associados a objetos de mídia pelo usuário. Entretanto, a discussão ressurgiu no que diz respeito à criação de tais descritores. Caso os descritores fossem impostos pela ferramenta, não seria necessária

uma preocupação de como os usuários iriam defini-los, pois já estariam pré-estabelecidos. Essa imposição, contudo, diminuiria a variedade de posições que objetos de mídia poderiam ocupar na tela. Por essas razões, a realidade aumentada pode não substituir completamente todas as etapas do processo de autoria multimídia. Entretanto, este trabalho demonstra que a RA pode trazer mais simplicidade e conveniência em algumas etapas desse processo.

Como resultado deste trabalho de iniciação científica, foi publicado o artigo "Exploring an AR-based User Interface for Authoring Multimedia Presentations" [12]. A principal evolução em relação a tal artigo é a inclusão dos marcadores de *visão estrutural* e de *elos*, cuja necessidade foi evidenciada durante a avaliação realizada.

REFERÊNCIAS

- [1] NBR ABNT. 2011. 15606-2, 2011. Digital Terrestrial Television-Data Coding and Transmission Specification for Digital Broadcasting-Part 2: Ginga-NCL for fixed and mobile receivers-XML application language for application coding.
- [2] Roberto Gerson A Azevedo, Eduardo Cruz Araújo, Bruno Lima, Luiz Fernando G Soares, and Marcelo F Moreno. 2014. Composer: meeting non-functional aspects of hypermedia authoring environment. *Multimedia tools and applications* 70, 2 (2014), 1199–1228.
- [3] Volkert Buchmann, Stephen Violich, Mark Billinghurst, and Andy Cockburn. 2004. FingARtips: gesture based direct manipulation in Augmented Reality. In *Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Australasia and South East Asia*. ACM, Suntec City, Singapore, 212–221.
- [4] Julie Carmigniani, Borko Furht, Marco Anisetti, Paolo Ceravolo, Ernesto Damiani, and Misa Ivkovic. 2011. Augmented reality technologies, systems and applications. *Multimedia tools and applications* 51, 1 (2011), 341–377.
- [5] Fred D Davis, Richard P Bagozzi, and Paul R Warshaw. 1989. User acceptance of computer technology: a comparison of two theoretical models. *Management science* 35, 8 (1989), 982–1003.
- [6] Luiz Gomes Soares, Marcio Moreno, Carlos Salles Soares Neto, and Marcelo Moreno. 2010. Ginga-NCL: Declarative middleware for multimedia IPTV services. *IEEE Communications Magazine* 48, 6 (June 2010), 74–81. <https://doi.org/10.1109/MCOM.2010.5473867>
- [7] Ian S Graham. 1995. *The HTML sourcebook*. John Wiley & Sons, Inc., New York, USA.
- [8] Anders Hejlsberg, Scott Wiltamuth, and Peter Golde. 2006. *The C# programming language*. Adobe Press, New York, USA.
- [9] ITU-T. 2009. H. 761, Nested Context Language (NCL) and Ginga-NCL for IPTV Services, Geneva, Apr. 2009.
- [10] Rodrigo Laiola Guimarães, Carlos de Salles Soares Neto, and Luiz Fernando Gomes Soares. 2008. A visual approach for modeling spatiotemporal relations. In *Proceedings of the eighth ACM symposium on Document engineering*. ACM, São Paulo, Brasil, 285–288.
- [11] Erick Lazaro Melo, Caio César Viel, Cesar Augusto Camillo Teixeira, Alexandre Coelho Rondon, Daniel de Paula Silva, Danilo Gasques Rodrigues, and Endril Capelli Silva. 2012. WebNCL: A Web-based Presentation Machine for Multimedia Documents. In *Proceedings of the 18th Brazilian Symposium on Multimedia and the Web (WebMedia '12)*. ACM, New York, NY, USA, 403–410. <https://doi.org/10.1145/2382636.2382719>
- [12] Paulo Renato Conceição Mendes, Roberto Gerson de Albuquerque Azevedo, Ruy Guilherme Silva Gomes de Oliveira, and Carlos de Salles Soares Neto. 2018. Exploring an AR-based User Interface for Authoring Multimedia Presentations. In *Proceedings of the ACM Symposium on Document Engineering 2018 (DocEng '18)*. ACM, New York, NY, USA, Article 9, 9 pages. <https://doi.org/10.1145/3209280.3209534>
- [13] PTCInc. 2018. Vuforia Augmented Reality SDK. <https://www.vuforia.com/>
- [14] Lloyd Rutledge. 2001. SMIL 2.0: XML for Web multimedia. *IEEE Internet Computing* 5, 5 (2001), 78–84.
- [15] Luiz Fernando Gomes Soares, Rogério Ferreira Rodrigues, and Romualdo Monteiro de Resende Costa. 2005. Nested Context Model 3.0. *Monographs in Computer Sciences* 18/05 (2005), 35.
- [16] LUIZ FERNANDO GOMES SOARES Soares. 2009. *Programando em NCL 3.0: desenvolvimento de aplicações para middleware Ginga: TV digital e Web*. Elsevier, Rio de Janeiro, Brasil.
- [17] UnityTechnologies. 2018. Unity 3D. <https://unity3d.com/>