

Luiz Carlos Pacheco Rodrigues Velho¹ Júlia Rabetti Giannella²

Visualização de coleção de imagens *Visualization of image collections*

RESUMO

A captura ou a elaboração de uma imagem pode ser realizada por diversas técnicas: fotografia, pintura, colagem, computação gráfica ou outra forma artística. No âmbito digital, imagens caracterizam-se por serem matrizes numéricas em pixels que podem ser processadas computacionalmente para análise e, inclusive, síntese em novos objetos de mídia. Também no âmbito digital, a produção imagética multiplica-se e diversifica-se. Neste artigo, apresentamos dois conceitos que inauguram possibilidades para ressignificar a produção e a distribuição de imagens no contexto digital: coleção de imagens e visualização de mídia. Para ilustrar tais conceitos, discutimos duas investigações de natureza teórica-prática conduzidas entre 2016 e 2019 no Laboratório de Visão e Computação Gráfica (Visgraf) do Instituto de Matemática Pura e Aplicada (IMPA).

Palavras-chave: imagem digital; coleção de imagens; visualização; visualização de mídia.

ABSTRACT

The capture or elaboration of an image can be accomplished by several techniques: photography, painting, collage, computer graphics or other artistic form. In the digital realm, images are characterized by being numerical arrays in pixels that can be processed computationally for analysis and even synthesis in new media objects. Also in the digital field, the imagery production is abundant and diverse. In this article, we present two concepts that inaugurate possibilities to resignify the production and distribution of images in the digital context: image collection and media visualization. To illustrate these concepts, we discuss two theoretical and practical investigations conducted between 2016 and 2019 at the Vision and Graphics Laboratory (Visgraf) of the Institute of Pure and Applied Mathematics (IMPA).

Keywords: digital image; image collection; visualization; media visualization.

1 Doutor em Ciência da Computação na Universidade de Toronto e pós-doutorado em Matemática Aplicada no IMPA. Mestre em Animação por Computador no MIT Media Lab. Bacharel em Desenho Industrial na ESDI-UERJ, Professor e pesquisador titular do Instituto Nacional de Matemática Pura e Aplicada (IMPA) e líder do Laboratório VISGRAF que realiza pesquisas nas áreas de Visão, Computação Gráfica e Multimídia. E-mail: lvelho@impa.br

2 Doutoranda do Programa de Pós-Graduação da Escola Superior de Desenho Industrial (ESDI-UERJ) na linha de pesquisa Design e Tecnologia e assistente de pesquisa no Laboratório de Visão e Computação Gráfica (VISGRAF) do Instituto de Matemática Pura e Aplicada (IMPA). E-mail: juliagiannella@gmail.com

I INTRODUÇÃO

Uma imagem é um artefato que pode ser analisado a partir de vários pontos de vista distintos. Do ponto de vista cognitivo, imagens são interpretadas pelos mecanismos da percepção visual humana. Do ponto de vista informacional, imagens traduzem uma representação bidimensional de objetos do mundo real ou de universos simbólicos. Do ponto de vista da mídia, imagens realizam concretamente essa representação, seja através de um dispositivo físico, tais como nas impressões fotográficas, pinturas, desenhos etc., seja por meio de um sistema tecnológico, como a televisão e o cinema. Do ponto de vista cultural, imagens carregam conteúdos simbólicos associados a tradições e experiências sociais.

De todo modo, imagens são objetos de informação, isto é, dados que existem no contexto de um sistema de comunicação. No contexto das imagens digitais, para que elas se tornem recursos disponíveis, recuperáveis e acessíveis, é necessário o desenvolvimento de processos para construção de representações. Um impulso na revolução das imagens digitais foi dado pela expansão da utilização dos computadores que auxiliam o tratamento de imagens a fim de realizar tarefas de processamento, análise e síntese em larga escala. Esses recursos resultam em novas aplicações de exploração visual nas quais a imagem assume um papel central em sistemas inteligentes e interativos.

Neste artigo, estamos interessados em investigar as potencialidades da visualização de coleções de imagens no contexto delineado acima.

2 COLEÇÕES DE IMAGENS

Para poder estudar melhor novas aplicações de visualização de imagens, precisamos de uma conceituação matemática para caracterizar grupos de imagens. Para isso, vamos introduzir o conceito de “Coleção de Imagens” (VELHO; CARVALHO, 2006).

Uma Coleção de Imagens é um conjunto de imagens $\{I^k\}$ com as seguintes propriedades:

- a. Partilham uma *Relação de Afinidade A*
- b. Estão mapeadas em um *Suporte Comum U*
- c. *Possuem um Modelo de Atributos subjacente M*

A Relação de Afinidade pode ser considerada a propriedade mais importante para caracterizar uma coleção de imagens. Ela constitui o elo de ligação partilhado por todas as imagens do conjunto, e que será definido precisamente por meio do Modelo de Atributos. Nesse sentido, a afinidade entre as imagens traduz algum tipo de *coerência* que relaciona as imagens do conjunto umas com as outras. Essa coerência pode ser de natureza diversa, tal como coerência espacial (mesma posição); coerência temporal (mesmo intervalo de tempo); coerência temática (mesmo objeto) etc.

O mapeamento em um Suporte Comum é necessário para estabelecer uma correspondência entre os elementos de uma imagem com os seus equivalentes nas outras imagens - fato essencial para definir um Modelo de Atributos.

Finalmente, o Modelo de Atributos especifica quais aspectos serão explorados em aplicações da coleção de imagens. Matematicamente, ele define as estruturas comuns ao conjunto de imagens e as operações que podem ser realizadas sobre elas.

Intuitivamente, podemos considerar uma Coleção de Imagens como um novo tipo de objeto de mídia, de natureza intermediária entre uma única imagem individual e um vídeo. Com isso, pela instrumentalização de sistemas digitais informatizados podemos criar aplicações midiáticas inéditas, que só agora se tornam possíveis.

2.1 Análise e Mineração de Imagens

Para realizar qualquer trabalho computacional com imagens, precisamos de um modelo matemático que permita definir as propriedades de interesse e as formas de operar com elas (VELHO *et al.*, 2008).

Matematicamente, podemos considerar uma imagem como um objeto de mídia que consiste em uma tupla³ $I = \{U, F\}$, onde:

- **U** é o Suporte Geométrico da imagem **I**. Em geral, um subconjunto retangular do plano (como numa fotografia ou tela de projeção), mas que pode também ser uma outra superfície bidimensional (como a calota esférica da retina do olho humano);
- **F** é a Função de Atributos da imagem **I** que define sobre o suporte **U** o conteúdo de dados associado com a imagem.

Em geral, **F** se refere a atributos visuais da imagem, tais como informações visuais de baixo nível (cor, brilho etc.), mas pode também especificar informações de mais alto nível, como geometria (profundidade), características estruturais, arestas, ou mesmo informações semânticas como a identificação de pessoas ou objetos. Os atributos de baixo nível, em geral, são produzidos por sensores (câmeras) ou processos de síntese de imagem. Já os atributos de alto nível são gerados por operações mais sofisticadas de análise de imagem ou são resultado de anotações, tais como metadados.

A função **F** têm como seu domínio o suporte geométrico **U** e como seu contra-domínio um espaço de atributos **A** (cor etc.). Ou seja, $F : U \Rightarrow A$. Nesse sentido, podemos nos referir a ela como função imagem: $c = F(u,v)$, que especifica para cada ponto com coordenada (u,v) do domínio um valor de atributo **c** (por exemplo, uma cor $c = [R,G,B]$ com componentes espectrais nas faixas do vermelho, verde e azul). Note que neste modelo matemático de imagem, o suporte geométrico **U** é trivial, enquanto a função de atributos **F** é complexa, pois carrega todo o conteúdo informacional da imagem.

Na definição acima, consideramos um *modelo contínuo* de imagem, isto é, **F** toma valores em qualquer ponto (u,v) do domínio. Para fins computacionais, precisamos de um *modelo discreto* de imagem, ou seja, uma *imagem digital*. Nesse modelo, temos que discretizar tanto o domínio quanto o contra-domínio de **F**, de modo que a função toma valores apenas em pontos amostrais (i,j) de uma grade regular,

³ Na matemática, uma tupla é uma sequência finita (também chamada de "lista ordenada") de objetos.

também chamados de *pixels*⁴. Assim, uma imagem digital é representada por uma matriz bidimensional de pixels contendo os valores de atributos $A(i,j)$.

O modelo matemático de imagem permite a realização de operações de natureza diversa com imagens. As operações de *processamento de imagem* transformam uma imagem em outra, enquanto as operações de *análise de imagem* fazem inferências sobre uma imagem para extrair informações estruturais, tais como reconhecimento de padrões.

Recentemente, as áreas de processamento e de análise de imagens tiveram um grande avanço com o desenvolvimento da Visão Computacional e a adoção de técnicas de aprendizado de máquina utilizando redes neurais profundas (*deep learning*). Essencialmente, o *deep learning* consiste na criação de sistemas de processamento e análise de imagens por meio de métodos estatísticos, usando grandes bancos de dados com imagens. Esse mecanismo tem duas etapas: o *treinamento* e a *inferência*, sendo que na primeira etapa os dados são escrutinados para “aprender” uma função inteligente, a qual pode, na segunda etapa, ser aplicada a novas imagens com finalidades diversas.

Na seção de estudos de caso reportada neste artigo, vamos empregar várias técnicas desse repertório de operações com imagens na mineração de imagens em grandes bancos de dados midiáticos para o desenvolvimento de aplicações de visualização de coleção de imagens.

3 VISUALIZAÇÃO E SISTEMAS DE MÍDIA

Atualmente, temos acesso a uma quantidade sem precedentes de mídia - mais do que podemos investigar, ou simplesmente assistir, ou mesmo consultar. Diversos fatores colaboram para tal panorama: o movimento de digitalização de mídias liderado por instituições culturais, museus, bibliotecas etc. preocupadas com a preservação e disseminação de seus acervos; a popularização de câmeras em dispositivos móveis; e a emergência de conteúdo gerado pelos usuários em sites e redes sociais (Facebook, Flickr, Twitter, Instagram, Youtube etc.). A constituição desse *big image data* inaugura um leque de oportunidades e desafios para as disciplinas interessadas na cultura visual. Mas como tirar vantagem desta nova escala de produção midiática? O laboratório Visgraf busca responder a essa e a outras questões por meio da investigação e do desenvolvimento de soluções para visualização de coleção de imagens em diferentes sistemas de mídia.

3.1 Visualização de mídia

Tradicionalmente, a visualização de informação corresponde ao processo de tradução de um determinado fenômeno em números e a subsequente visualização das relações entre esses números. Neste processo, as relações são visualmente implementadas em um suporte geralmente plano, com base em dois princípios. O primeiro princípio é o uso de gráficos primitivos como pontos, linhas e áreas para representar as unidades informativas. O segundo princípio é o uso de variáveis visu-

4 Do inglês *Picture Elements*.

ais (cor, forma, valor, tamanho etc.) para representar semelhanças e diferenças nos dados e revelar padrões. No entanto, como proceder com visualizações cujos dados de *input* são imagens? Lev Manovich, pesquisador na área de novas mídias, mídias digitais, design e estudos do *software* oferece uma perspectiva para refletir sobre tal problema. Em oposição à visualização de informação tradicional, o autor inaugura o conceito de *visualização de mídia*, método que envolve a tradução de um conjunto de imagens em um novo objeto de mídia. Visualização de mídia pode ser formalmente definida como a “criação de novas representações visuais a partir de objetos visuais de uma coleção” (MANOVICH, 2011, p. 41, tradução nossa).

A capacidade de exibir milhares de imagens simultaneamente em um suporte comum permite a visualização da estrutura de uma coleção de imagens. Conceitualmente, visualizar o *todo* pode ser compreendido como a extensão de uma operação intelectual fundamental para análise de mídias: a comparação de um conjunto de imagens e identificação de padrões de coerência.

Atualmente, interfaces de aplicativos em computadores, notebooks, tablets, celulares e *web browsers* permitem a exibição simultânea de dezenas de imagens em formato de grade (*grid*). Essas grades de imagem podem ser organizadas de acordo com metadados conhecidos nas imagens (ano de produção, país, nome do criador, tamanho da imagem, datas de envio) ou atributos processados (cor etc.). Organizadores de imagens comumente usados em aplicativos e editores - incluindo Fotos (da Apple) (Fig.1), Instagram, Adobe Photoshop - disponibilizam esse leiaute de acesso como forma de visualização. Da mesma forma, serviços de pesquisa de imagens na *web* - como o Google Imagens e o Flickr - também recorrem à grade de imagens como interface padrão.

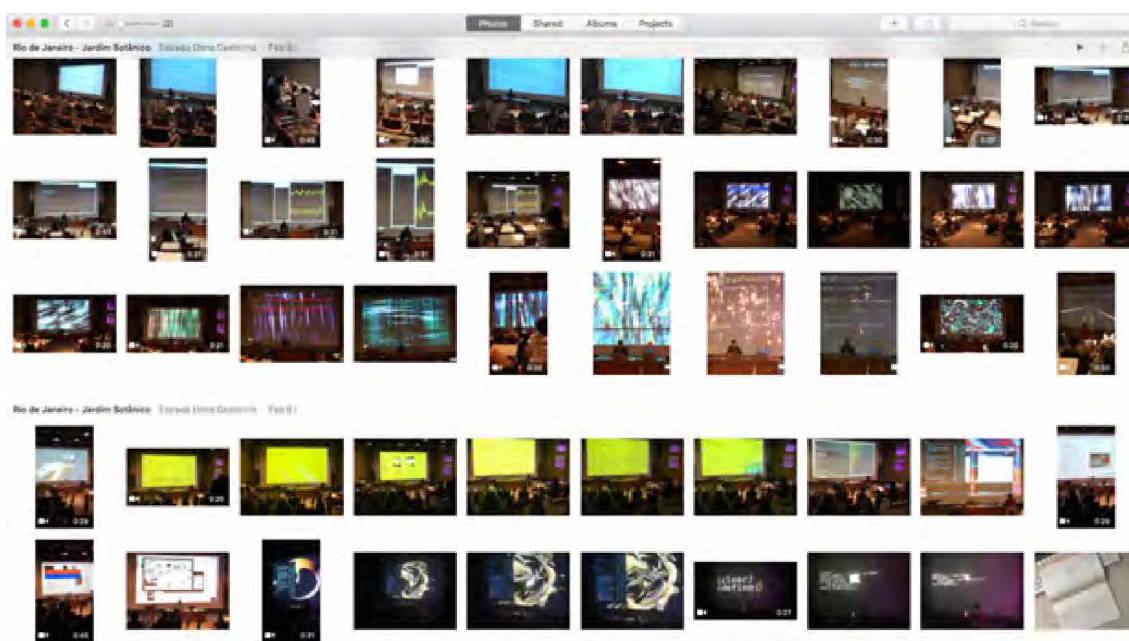


Figura 1 - A interface do aplicativo Fotos é um exemplo de visualização de mídia em grade.

Fonte: Captura de tela (2019).

O método de visualização de mídia explora o fato de que as imagens das coleções vêm, geralmente, acompanhadas de atributos mínimos que podem ser utilizados para definir o mapeamento das imagens na visualização em grades. Dentre as técnicas de visualização de mídia propostas por Manovich (2012a), o mosaico⁵ é uma extensão conceitual e técnica da grade de imagens. O desenvolvimento de mosaicos é baseado em três operações que podem ser implementadas separadamente ou combinadas, a saber:

- a. *Zoom out* - a visualização simultânea de todas as imagens da coleção em um suporte comum e que permite reconhecer a relação de coerência entre as imagens;
- b. Amostragem - recortes (temporais, espaciais ou semânticos) que representam a seleção de um subconjunto de imagens de uma coleção maior;
- c. Remapeamento - rearranjo das imagens em novas configurações.

A Figura 2 é um mosaico contendo 4535 imagens (miniaturas) de capa da revista Time publicadas entre 1923 e 2009. A organização temporal das capas é mapeada no mosaico da esquerda para direita, de cima para baixo. O visualizador do Flickr, no qual uma versão do mosaico é disponibilizado, permite realizar operações de *zoom out*. Nesse exemplo de mosaico não há amostragem, pois todas as capas da coleção original são visualizadas.

5 No original, em inglês, essa técnica de visualização de mídia é denominada *montage*.

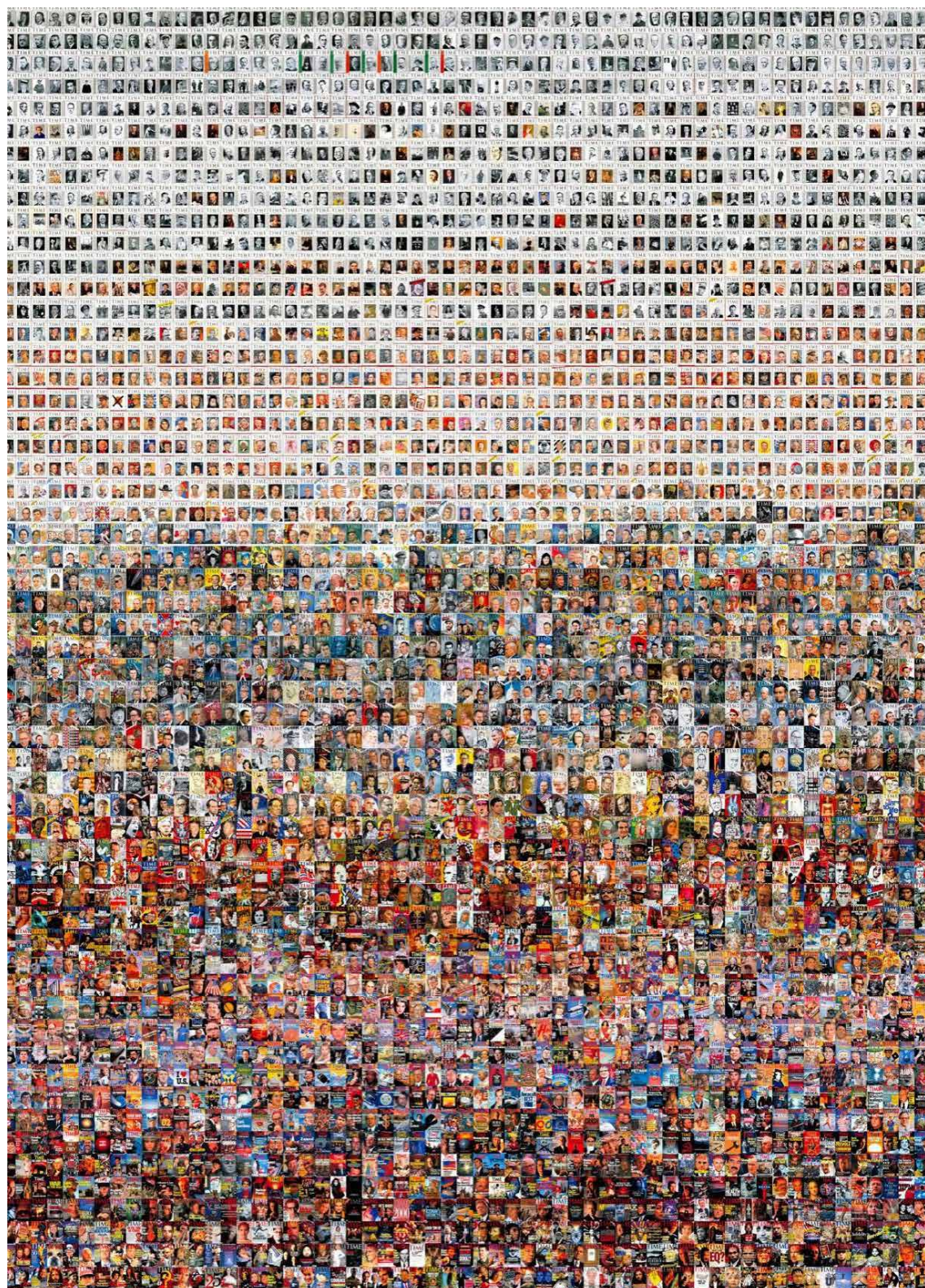


Figura 2 - Mosaico contendo 4535 miniaturas de capa da revista Time.

Fonte: Manovich (2012b).

A organização por data de publicação revela vários padrões históricos. A visualização em mosaico mostra a era da pré-impressão colorida na década de 1920, seguida de um período de experimentos iniciais em impressão colorida (com aplica-

ção de faixas de cor na margem esquerda da capa) e depois a mudança gradual das capas monocromáticas para as capas coloridas. Na década de 1930, a fotografia é a principal técnica de representação de imagens utilizada nas capas. Depois de 1940, a pintura prevalece por um determinado período. Finalmente, nas décadas mais recentes, a fotografia gradualmente volta ao domínio.

3.2 Sistemas de mídia

O que hoje chamamos de *novas mídias* pode ser entendido como um campo acadêmico e artístico dedicado à investigação e difusão de objetos culturais a partir do uso de computadores e tecnologias computacionais. Seu histórico pode ser reconstituído desde o fim da década de 1970, com a criação de relevantes conferências internacionais na área de Computação Gráfica, como o Siggraph, nos Estados Unidos, e o Ars Electronica, na Áustria (MANOVICH, 2005). Mas foi somente no fim da década de 1980 que o campo das novas mídias começou a amadurecer, com o surgimento de instituições dedicadas a sua pesquisa e produção. Na Europa, podemos citar o ZKM, em Karlsruhe (1989); New Media Institute, em Frankfurt (1990); e Inter-Society for the Electronic Arts (Isea), na Holanda (1990).

Mas o que são exatamente as novas mídias? E quais são as características que as definem como um sistema digital? As quase quatro décadas de existência do campo parecem não ser suficientes para responder precisamente a essas questões. O laboratório Visgraf, no entanto, busca refletir sobre essas questões por meio de projetos práticos orientados à pesquisa. Ao longo dos últimos 30 anos, o Visgraf conduziu trabalhos na área de novas mídias a partir da definição de aspectos que podem ser distribuídos em três pilares, a saber:

Aspectos tecnológicos: diz respeito ao conjunto de objetos que constitui a coleção. No caso, cada objeto é uma imagem. Diz respeito também aos atributos (metadados ou *features* visuais) de cada imagem que podem ser adicionadas manualmente ou extraídas computacionalmente;

Aspectos produtivos: diz respeito às decisões projetuais e metodológicas que podem ser orientadas tanto por um objetivo de pesquisa mais específico como por um estudo mais exploratório.

Aspectos distributivos: diz respeito às características do meio em que a interação com visualização de mídia acontecerá e aspectos de implementação do sistema de mídia.

4 ESTUDOS DE CASO

Nesta seção, apresentamos dois projetos conduzidos no laboratório Visgraf entre 2016 e 2019 e que investigam e desenvolvem aplicações para visualização de coleções de imagens.

4.1 Visualização das imagens da Rio-2016

Em 2016, o laboratório Visgraf iniciou o projeto OBSERVATÓRIO2016 (GIANNELLA; VELHO, 2016), que se caracteriza por ser uma plataforma *online* para monitoramento, organização e visualização do debate sobre as Olimpíadas Rio-2016 no Twitter. A plataforma coletou, entre 18 de abril e 25 de agosto de 2016, aproximadamente 1 milhão de *tweets*. De forma a dar visibilidade às diferentes opiniões sobre as Olimpíadas, estabelecemos sete procedimentos para coleta de dados do Twitter por meio de seu API público. Os dados são apresentados em oito visualizações interativas em um portal (Fig. 3).

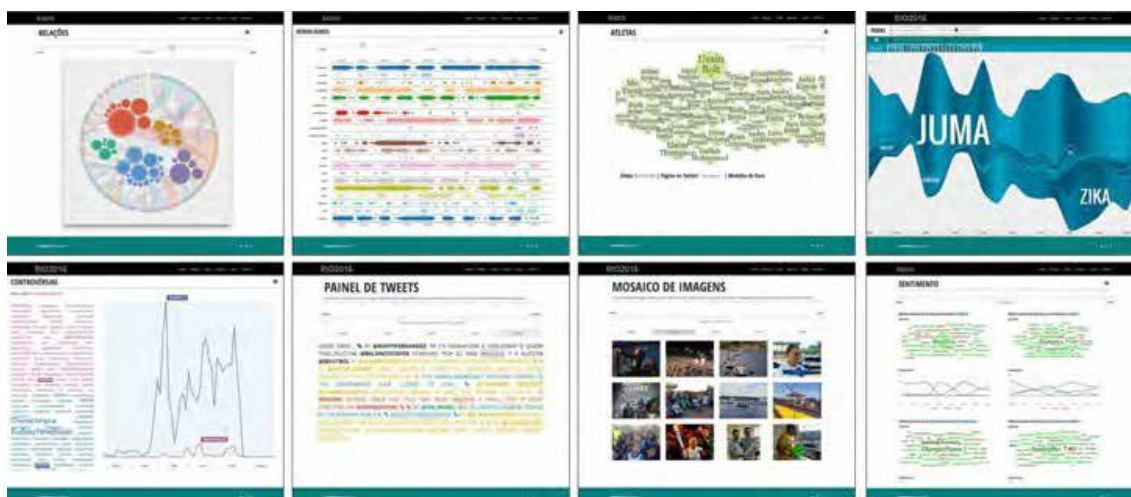


Figura 3 - Visualizações interativas do portal OBSERVATÓRIO2016. Fonte: Captura de tela (2016).

Aproximadamente 180.000 dos *tweets* coletados incluíram imagens únicas (fotografias, ilustrações, montagens etc.)⁶. A produção e circulação de grandes volumes de imagens digitais nas redes sociais desperta a atenção de pesquisadores interessados na prática criativa humana. A investigação de Lev Manovich em métodos digitais para estudar a manifestação da cultura visual é um trabalho pioneiro que busca valer-se de tal oportunidade. O desafio que se coloca, neste cenário, é entre a abundância de dados visuais e a não trivialidade de seu tratamento. Se, por um lado, o acesso ao vasto repertório imagético é facilitado pelos avanços em tecnologias de coleta e armazenamento de dados visuais, por outro lado, o acesso em si não é suficiente para propor problemas de análise e síntese de imagens. Volumosas coleções de artefatos visuais justificam a urgência de se explorarem mecanismos inteligentes capazes de identificar e agrupar imagens não apenas a partir de seus metadados textuais, mas principalmente a partir de metodologias de análise e pro-

6 As imagens podem ser consultadas por dia e categoria em <http://oo.impa.br/dimagens/>. As imagens também estão armazenadas em um banco de dados do Visgraf. Para acesso, através de um API Rest, entre em contato com lvelho@impa.br.

cessamento de dados visuais. O laboratório Visgraf desenvolve, entre outras atividades, pesquisa e projetos em visão computacional, área da computação que lida com maneiras de analisar e sintetizar imagens a partir do reconhecimento de formas, estilos e propriedades visuais em imagens e vídeos. Nesse sentido, o laboratório identificou a oportunidade de inaugurar uma nova fase de pesquisa para o projeto OBSERVATÓRIO2016, voltada para análise e visualização das imagens coletadas a partir técnicas computacionais.

4.1.1 Aspectos tecnológicos

A partir da identificação da oportunidade de pesquisa, o projeto original desdobrou-se em uma investigação exploratória que teve como objetivo inicial a familiarização com recursos avançados, oriundos da Inteligência Artificial, para análise e mineração de dados. Os resultados dessa segunda fase de pesquisa foram o *website* "OBSERVATÓRIO2016: Experiências em *Deep learning*" e um poster apresentado em conferência internacional (GIANNELLA;VELHO, 2017). Nesse trabalho, análise e síntese de imagens - dois problemas tradicionais do campo da visão computacional - são exploradas a partir de técnicas *deep learning*, uma técnica de implementação de aprendizagem por máquina (*machine learning*). As seguintes tarefas, respectivamente referentes à análise e síntese de imagens, foram realizadas:

- a. Classificação de imagens: no nosso trabalho, realizamos dois tipos de classificação:
 - **Tocha olímpica:** analisamos computacionalmente as imagens para classificar a probabilidade delas conterem o objeto tocha olímpica. Como resultado, identificamos 4094 imagens com mais de 83% de probabilidade de conter o objeto;
 - **Modalidades olímpicas:** analisamos computacionalmente as imagens para classificar a probabilidade de elas representarem determinadas modalidades olímpicas⁷. Assim como para tocha olímpica, estabelecemos o limiar de 83%.
- b. Geração automática de uma nova imagem e vídeos: no nosso trabalho, as imagens analisadas e classificadas dentro de um limiar satisfatório foram usadas para gerar novas mídias. Essas mídias representam novas maneiras de visualizar as imagens:
 - **Mosaico da tocha:** coleção de imagens do tour da tocha olímpica visualizada na forma de um mosaico interativo;
 - **Globo de vídeos:** *slideshows* musicais que agrupam imagens por modalidades olímpicas e atletas.

Durante as pré-Olimpíadas, tornou-se evidente que muitas das imagens coletadas por nossos *scripts* de consulta ao API do Twitter eram relacionadas ao reveza-

⁷ As modalidades olímpicas escolhidas foram, em ordem alfabética: atletismo, badminton, basquete, canoagem slalom, ciclismo, esgrima, futebol, ginástica, golfe, hipismo, hóquei sobre grama, judô, levantamento de peso, nado sincronizado, polo Aquático, tênis tiro com arco, vela e vôlei de praia.

mento da tocha olímpica e representavam este objeto icônico. Uma parcela dessas imagens veio acompanhada por palavras ou *tags* que remetem à tocha, mas outra parcela não faz nenhuma menção textual. Além disso, uma grande parcela de tweets que faz menção ao termo “tocha”, e a outros termos similares, vem acompanhada de imagens que nada tem a ver com a representação visual do objeto. Em outras palavras, a análise textual por si só não é suficiente para a seleção de um conjunto de imagens contendo a tocha.

Recorremos, então, à técnica de *deep learning* - mais especificamente à prática de treinamento supervisionado utilizando redes neurais convolucionais - para reconhecer visualmente a presença da tocha olímpica nas imagens do nosso banco de dados. O campo do reconhecimento de padrões visuais foi recentemente aprimorado pelo eficiente desempenho das chamadas Redes Neurais Convolucionais⁸ (CNN). Em 2012, o trabalho de Krizhevsky *et al.* (2012) no treinamento de uma rede neural convolucional profunda para classificar 1,2 milhão de imagens em 1000 diferentes classes durante o Concurso ImageNet LSVRC-2010 impactou substancialmente a comunidade de visão computacional.

Mais recentemente e em virtude dos esforços da Google, a tarefa de classificação de imagens tornou-se relativamente mais fácil de ser resolvida. Isso porque a empresa liberou, em 2015, o uso *opensource* da rede neural Inception-v3 - capaz de classificar imagens a partir de 1000 categorias fornecidas pela ImageNet - e da biblioteca de códigos TensorFlow - que auxilia o desenvolvimento de um sistema de classificação de imagens.

Além disso, a rede Inception-v3 pode ser retreinada para detectar uma nova classe de objetos, qualidade que é muitas vezes referida como transferência de aprendizagem e tem contribuído para disseminar, ainda mais, o potencial das CNNs no contexto da visão computacional. Dito isso, para retreinar a rede a fim de reconhecer a tocha olímpica nas imagens, precisamos, além de uma infraestrutura computacional adequada para otimizar o processo, de uma amostra de cem imagens da tocha olímpica manualmente rotuladas. Cabe aqui dizer que o sucesso de uma nova categoria de classificação depende dos exemplos de imagens fornecidos à máquina.

Após ser retreinada, a rede neural reconheceu a presença da tocha olímpica nas imagens do banco de dados segundo uma lógica de probabilidade. Até o dia 25 de junho de 2016 (quando decidimos criar o mosaico de visualização da tocha), cerca de 1500 imagens com mais de 85% de probabilidade de conter a tocha olímpica tinham sido classificadas pela rede. Usamos essas imagens para criar a visualização de mídia *mosaico da tocha*. Para a seleção das imagens das modalidades olímpicas, recorreremos ao mesmo procedimento (Fig. 4).

8 Do inglês, *Convolutional Neural Network*.

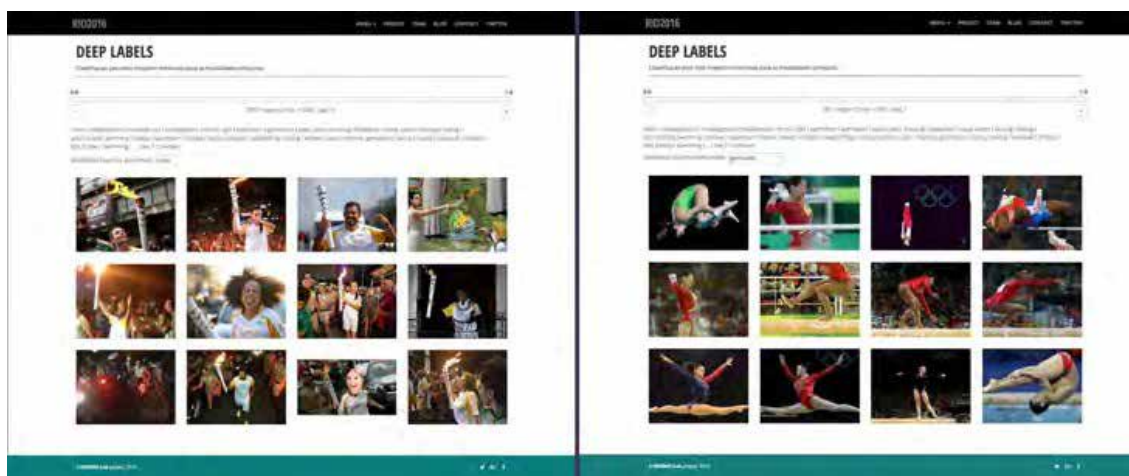


Figura 4 - À esquerda, imagens classificadas com mais de 83% de probabilidade de conter a tocha olímpica. À direita, imagens classificadas com mais de 83% de probabilidade de representar a modalidade Ginástica.

Fonte: Captura de tela (2016).

4.1.2 Aspectos produtivos

O sucesso da etapa de análise e mineração dos dados nos motivou a desenvolver duas visualizações de mídia: o *mosaico da tocha* e o *globo de vídeos*. Cada uma é resultado de uma diferente técnica de síntese de imagens.

Conceitualmente, o *mosaico da tocha* é produzido a partir das três operações (*zoom out*, amostragem e remapeamento) descritas previamente. O *zoom out* permite a visualização interativa das imagens em diferentes granularidades (da visão geral do todo até a visão detalhada de um item) e a amostragem se dá pelo recorte semântico (as imagens visualizadas são as que contêm a tocha olímpica). O remapeamento talvez seja a operação mais inusitada desta visualização de mídia. Ao invés de utilizar um parâmetro racional para o remapeamento das imagens (como a sequência temporal ilustrada na Figura 2), as miniaturas são remapeadas de modo que a imagem final (a visualização de mídia propriamente dita) sugira, de um ponto de observação afastado, a forma de outra imagem (um objeto, uma paisagem, um retrato etc.). A ideia central deste tipo de remapeamento deriva do trabalho de síntese de imagens de Finkelstein e Range (1998), em que dada uma determinada imagem (imagem alvo), outra imagem (mosaico) é automaticamente construída a partir de várias imagens menores (imagens em miniatura). Para sugerir visualmente a forma do mosaico, as imagens em miniatura são computacionalmente analisadas e organizadas para corresponder a uma imagem grande e, em seguida, suas cores são ajustadas para melhor sugerir a forma. A Figura 5, a seguir, ilustra esse tipo de remapeamento:



Figura 5 - À esquerda, a imagem alvo e, à direita, o mosaico. Fonte: elaborado pelo autor.

Já o desenvolvimento do *globo de vídeos* engloba a noção de remixagem das imagens em *slideshows* musicais. E para sintetizar cada um dos *slideshows* apresentados no *globo de vídeos*, recorreremos a recentes recursos automatizados da Apple para produção de filmes.

A versão do sistema operacional da Apple lançada em setembro de 2016 (iOS 10) trouxe uma forte presença de inteligência artificial em seus aplicativos nativos. O aplicativo Fotos, por exemplo, utiliza *deep learning* para reconhecer rostos, objetos e cenas nas fotografias e vídeos armazenados no álbum de fotos do usuário. No iOS 10, o aplicativo *Photos* oferece a funcionalidade “Memórias”, que gera automaticamente *slideshows* musicais com base no reconhecimento de pessoas, lugares, cenas e datas. Depois de pronto, o aplicativo permite personalizar algumas propriedades (duração, número de imagens etc.) de *slideshow* por meio de seu ambiente de autoria.

Para que *Photos* criasse automaticamente os *slideshows* musicais com as imagens da Rio-2016, importamos, progressivamente, amostras de imagens no aplicativo com base nas classificações previamente feitas pela rede neural (no caso, amostras temáticas referentes às modalidades esportivas). Na ocasião, escolhemos trabalhar com o aplicativo Fotos para iPad. A Figura 6 contém reproduções de telas do aplicativo Fotos.

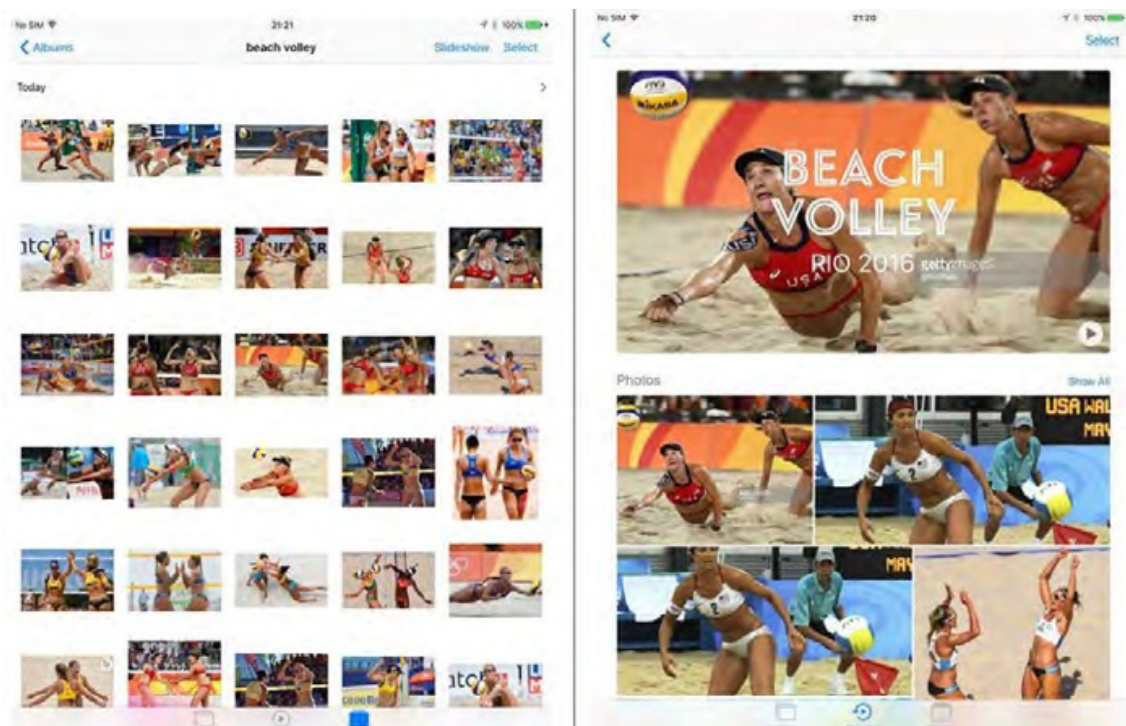


Figura 6 - Telas do aplicativo Fotos. A tela à esquerda apresenta parcialmente a coleção com todas as imagens (miniaturas) da categoria “vôlei de praia” importadas pelo aplicativo. A tela à direita mostra a aba “Memórias” com o *slideshow* criado pelo aplicativo, na parte superior, e algumas das imagens individuais, na parte inferior.

Fonte: Captura de tela (2017).

4.1.3 Aspectos distributivos

O meio de acesso às visualizações elaboradas para a coleção de imagens da Rio-2016 é uma aplicação *web* para desktops. Para implementar o *mosaico da tocha*⁹, usamos um visualizador *web* em código aberto chamado OpenSeadragon. Esse visualizador, escrito em linguagem JavaScript, possibilita um ambiente de imersão para imagens em alta resolução. Os elementos de interface padrão incluem recursos como *zoom in*, *zoom out*, ajustar ao tamanho de tela e alternar para tela cheia (Fig. 7).

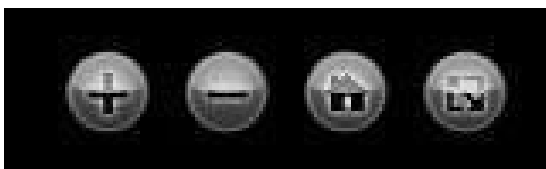


Figura 7 - Recursos do OpenSeadragon. Fonte: captura de tela (2017).

9 O *mosaico da tocha* está disponível em: http://velho.impa.br/dl_rio2016/mosaico.html



Figura 8 - Quatro níveis de zoom do *Mosaico da tocha*. Fonte: captura de tela (2017).

Para implementar o *globo de vídeo*¹⁰, usamos uma interface web para visualização e navegação de vídeos em ambiente 3D chamada Goo Video Sphere. Criada em WebGL¹¹ e disponibilizada em código aberto pela Goo Engine, essa interface permite a incorporação de vídeos em stream em telas que preenchem uma esfera tridimensional (Fig. 9). Recursos como *zoom*, *pan* e clique podem ser usados para navegar pelo globo - podendo inclusive rotacioná-lo e entrar nele - e acessar os *slideshows* em tela cheia.



Figura 9 - *Globo de vídeos*. Fonte: captura de tela (2017).

10 O *globo de vídeos* está disponível em: http://velho.impa.br/dl_rio2016/videos.php.

11 WebGL (*Web Graphics Library*) é uma API em JavaScript, disponível a partir do novo elemento canvas da HTML5, que oferece suporte para renderização de gráficos 2D e gráficos 3D.

4.2 Visualização do acervo fotográfico do Instituto Moreira Salles

Em junho de 2018, o Instituto de Matemática Pura e Aplicada (IMPA), através do Laboratório Visgraf, e o Instituto Moreira Salles (IMS), através da Coordenação de Fotografia, assinaram um acordo de cooperação (IMPA, 2018) com foco na investigação e no desenvolvimento de aplicações para difusão do acervo fotográfico do IMS. Acreditamos que tanto a experiência vivenciada quanto os resultados obtidos nesse processo vislumbram uma forma inédita de integração entre duas instituições destacadas em suas respectivas áreas de excelência – científica e cultural, com benefício mútuo no contexto das complexas transformações tecnológicas e sociais da atualidade.

O plano de trabalho centrou em três principais linhas de pesquisa e atuação:

- a. Georreferenciamento de fotografias urbanas: foco em imagens urbanas do Brasil dos séculos XIX e XX, que se destacam em quantidade e relevância no acervo fotográfico do IMS. Este eixo inclui os seguintes tópicos de interesse: exploração de tecnologias e metodologias para georreferenciar fotografias urbanas, tendo como referência a posição, precisa ou aproximada, de câmera; visualização e justaposição entre fotografias e modelos tridimensionais do espaço urbano; interface entre acervo fotográfico e plataformas de visualização cartográfica;
- b. Tecnologias e mídias para bancos de imagens: investigação de temas emergentes da Tecnologia de Informação para o acesso e a difusão de acervos fotográficos. Este eixo inclui os seguintes tópicos de interesse: visualização de dados semânticos estruturados no Wikidata; visualização de imagens fotográficas em multirresolução; identificação de protocolos abertos e bem documentados que permitam a comparação de imagens entre diferentes instituições;
- c. Operacionalização da plataforma Liquid Galaxy: Estudo da plataforma de visualização Liquid Galaxy (LG) e desenvolvimento de protótipos. Este eixo inclui os seguintes tópicos de interesse: colaboração com a empresa End Point¹², doadora do equipamento, para suporte e treinamento; operacionalização do LG nas instalações do IMS; elaboração de metodologias para publicação de conteúdo e construção de narrativas panorâmicas; e identificação de oportunidades para desenvolvimento de projetos.

Um dos resultados da cooperação de pesquisa foi o desenvolvimento de demos para a plataforma LG. Esses demos são formas de georreferenciar e visualizar o acervo fotográfico do IMS em interfaces panorâmicas, imersivas e dotadas de uma série recursos que serão comentados nas próximas seções.

4.2.1 Aspectos tecnológicos

No período em que acordo vigorou (julho de 2018 a janeiro de 2019), foram desenvolvidos dois demos para a plataforma LG: Centros Culturais do IMS e Obras Seleccionadas de Marc Ferrez no Rio de Janeiro. Enquanto o primeiro demo tem

12 End Point é uma empresa consultora de *software* com sede em Nova York.

como objetivo apresentar uma visão geral e histórica dos três centros culturais do instituto (Poço de Caldas, Rio de Janeiro e São Paulo), o segundo demo encarrega-se da visualização de itens de uma coleção fotográfica. No presente artigo, nos concentramos sobre o segundo demo.

O IMS é uma importante instituição na paisagem cultural brasileira. Possui importantes patrimônios em quatro áreas: fotografia, em mais larga escala, música, literatura e iconografia. O acervo fotográfico de cerca de 2 milhões de imagens é distribuído entre aproximadamente 52 coleções - muitas vezes a obra completa de seus fotógrafos -, representando, portanto, um expressivo conjunto da fotografia do século XIX e XX no Brasil.

Marc Ferrez é o mais importante fotógrafo do século XIX no acervo do IMS. Sua obra compreende uma coleção de aproximadamente sete mil fotografias. Nascido no Rio de Janeiro em 1843, descendente de uma família francesa, Ferrez é popularmente conhecido pelas fotografias de paisagens naturais e de cenas urbanas no cidade carioca. Seu trabalho documenta a transformação do Rio de Janeiro ao longo do tempo e representa algumas evidências de urbanização e desenvolvimento industrial.

A presença de marcos naturais e arquitetônicos conhecidos em suas fotografias (caracterizadas pelas vistas externas) (Fig. 10) torna possível estimar a posição da câmera para anotação de metadados geográficos.



Figura 10 - Panorama do Rio de Janeiro, vendo-se a Glória e o Catete em fotografia tirada de Santa Teresa, c. 1885. Marc Ferrez.

Fonte: Coleção Gilberto Ferrez / IMS (2018).

Para entender como a estimativa e a anotação dos metadados geográficos foram realizadas no nosso trabalho, é necessário, antes, descrever a plataforma LG. LG é uma plataforma de visualização de dados *multi-display* que permite uma experiência panorâmica imersiva por meio de apresentações interativas, usando recursos

como vídeos, fotos, gráficos 2D e 3D e arquivos específicos para anotação geográfica. Seu *hardware* padrão inclui sete telas HDTVs posicionadas lado a lado com um ângulo de quinze graus, resultando em um amplo *display* FoV com uma resolução de definição alta (7056 x 1080 pixels). O equipamento ainda inclui uma tela *touchscreen*, montada sobre um pódio, que serve à seleção de cenas dentro de uma apresentação e um controle, chamado de *SpaceNav 3-dimensional joystick*, que permite navegação em 6 eixos (Fig. 11). As aplicações do LG abrangem uma grande variedade de contextos de uso, da consultoria em sistemas de informação geográfica (GIS) ao mercado imobiliário, passando por instalações em museus, hotéis e escolas (LIQUID GALAXY, 2019). Nosso interesse investigativo está no LG como suporte panorâmico para consulta das coleções fotográficas, com particular atenção em como seus recursos podem ser explorados para visualização geoespacial e histórica¹³ do acervo.



Fig. 11 - *Hardware* do LG. Da esquerda para direita: tela multi-*display*, tela multitouch, controle de navegação. Fonte: GIANNELLA; VELHO, 2018.

A empresa End Point fornece um *Content Management System*¹⁴ (CMS), disponibilizado através de um navegador da *web*, para criação de conteúdo para as apresentações interativas. Esse CMS constitui-se como o principal ambiente de autoria para a plataforma LG à medida que é responsável por criar a estrutura narrativa da apresentação e organizar os diferentes recursos da apresentação. Ainda assim, alguns recursos precisam ser previamente preparados antes de serem importados no CMS. Arquivos contendo a anotação geográfica de objetos 2D (fotografias, por exemplo) são exemplos de recursos a serem desenvolvidos fora do ambiente do CMS.

Keyhole Markup Language (KML) é um formato de arquivo para anotação e exibição de dados geográficos em um navegador da Terra, como o Google Earth. O KML possui uma sintaxe baseada em *tags* com aninhamento de elementos e atributos e é baseado no padrão XML. No nosso trabalho, o desenvolvimento em KML foi realizado combinando a manipulação da interface gráfica do Google Earth (GE) e a escrita do código usando um *software* de edição. Essa combinação permite criar arquivos contendo instruções para justaposição de imagens no GE, isto é, a sobreposição de uma fotografia bidimensional em um modelo tridimensional do espaço urbano. Em termos técnicos, essa fotografia é referida como um elemento de <photoOverlay> na linguagem em KML.

¹³ O termo histórico está sendo usado no sentido de permitir uma comparação histórica entre o local registrado na fotografia original e o mesmo local nos dias de hoje, representado pelo modelo 3D.

¹⁴ Sistema de gerenciamento de conteúdo.

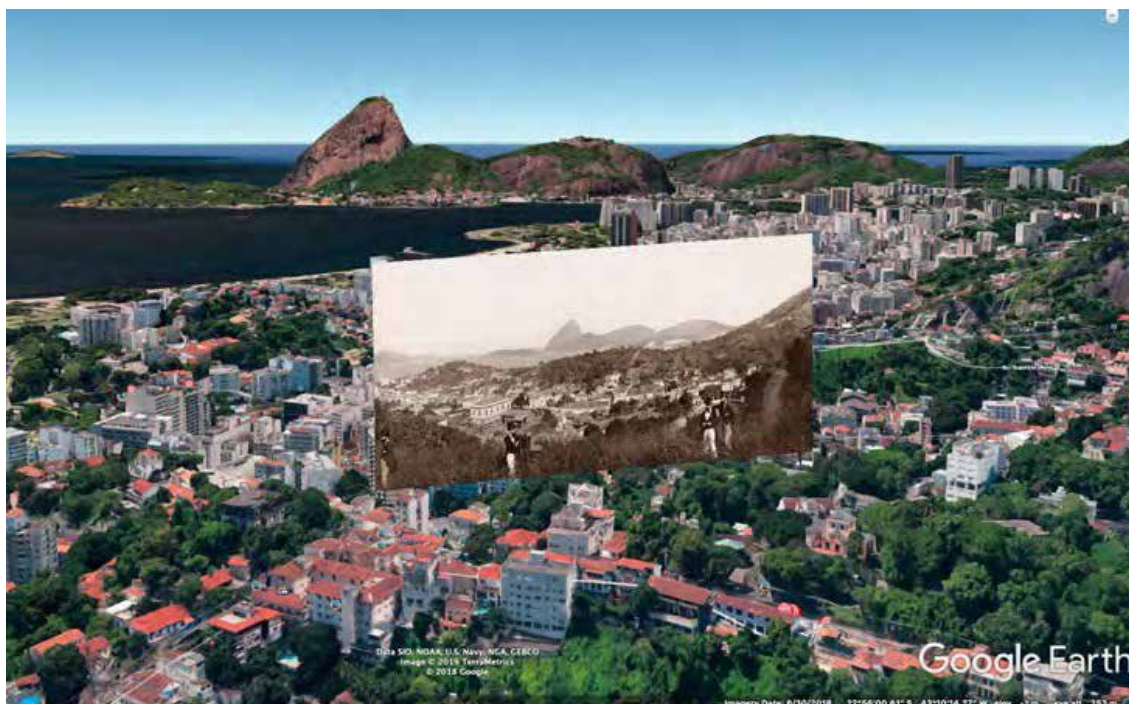


Figura 12 - Justaposição da fotografia bidimensional no modelo tridimensional do espaço urbano.
Fonte: captura de tela (2019).

4.2.2 Aspectos produtivos

Selecionamos dez fotografias de Marc Ferrez captadas no Rio de Janeiro que são passíveis de terem sua posição de câmera estimadas no GE. O princípio norteador do demo é que a apresentação permita comparar o Rio antigo com o Rio atual. Nosso critério para escolha das fotografias levou em consideração a distribuição espacial das imagens no terreno carioca. Escolhemos imagens que, em uma determinada vista aérea, não se sobrepusessem nem ficassem muito distante umas das outras.



Figura 13 - Distribuição das fotografias selecionadas no modelo 3D do espaço urbano.
Fonte: captura de tela (2019).

Usando o *software* GE para desktop, definimos os parâmetros¹⁵ de <Photo Overlay> a serem exportados para cada imagem como arquivos KML. Então, usando o CMS da LG, desenvolvemos uma apresentação para visualizar fotografias registradas por Marc Ferrez no século XIX sobrepostas à paisagem urbana atual, representada pelo modelo tridimensional do GE em segundo plano. A demonstração cumpre a tarefa de localizar espacialmente as fotografias de Ferrez e comparar uma mesma vista em diferentes períodos de tempo.

As funcionalidades do CMS do LG permitem diferentes estruturas narrativas com diferentes níveis de interatividade para as apresentações criadas. Um dos escopos do plano de trabalho instituído entre IMPA e IMS foi o estudo da plataforma LG e o mapeamento das diferentes formas de produzir conteúdo dentro do contexto de acervos museológicos. Em relatório técnico (GIANNELLA; VELHO, 2018), classificamos as estruturas narrativas em dois níveis que se desdobram em sub níveis, a saber:

1. Não-interativa
 - a. Linear com *looping*
2. Interativa
 - a. Linear
 - b. Não-linear
 - i. Dinâmica
 - ii. Exploratória

Para desenvolvimento do demo, optamos pela estrutura interativa não-linear dinâmica, que pode ser compreendida como uma estrutura em que a duração da apresentação, assim como a concatenação das cenas que a constituem, dependem da interação do usuário. Essa interação se dá dinamicamente de duas formas: por meio do controle de navegação 3D, que permite uma visualização interativa a partir das funcionalidades de *pan*, *zoom* e *rotação*, e por meio da tela *touchscreen*, que permite a seleção de cenas de modo não sequencial. As figuras a seguir ilustram, respectivamente, a cena inicial do demo e uma cena secundária.

15 Mais especificamente parâmetros de coordenada geográfica, altitude e ângulo de câmera.

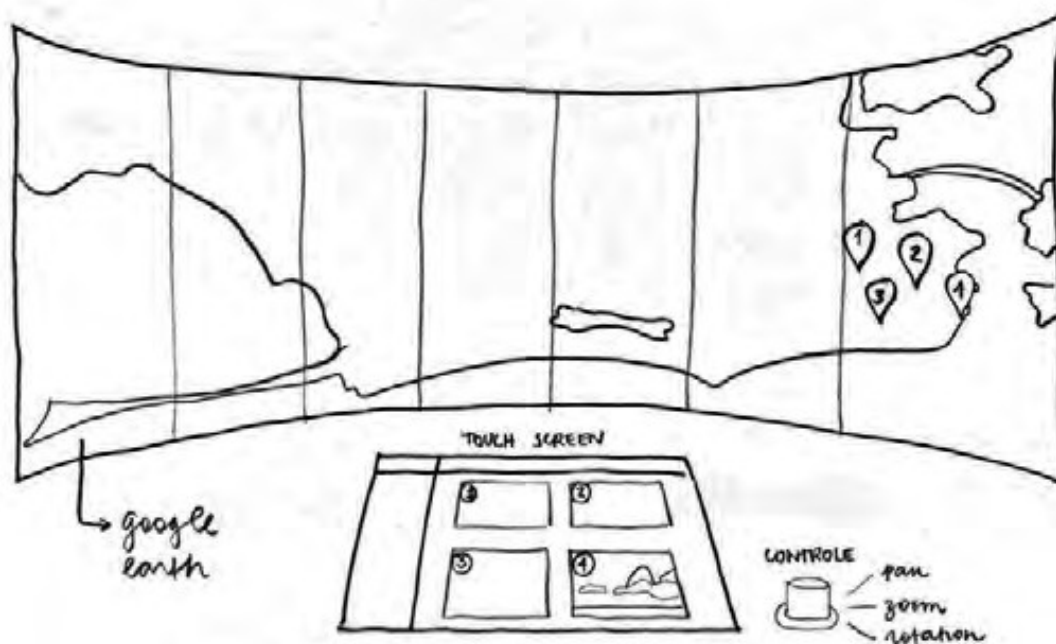


Figura 14 - Ilustração conceitual do demo: a cena inicial da apresentação é uma visualização ortogonal do modelo tridimensional do Rio de Janeiro no qual elementos visuais (no caso ícones) são georreferenciados no local aproximado de onde as fotografias correspondentes foram capturadas.
 Fonte: elaborado pelo autor.

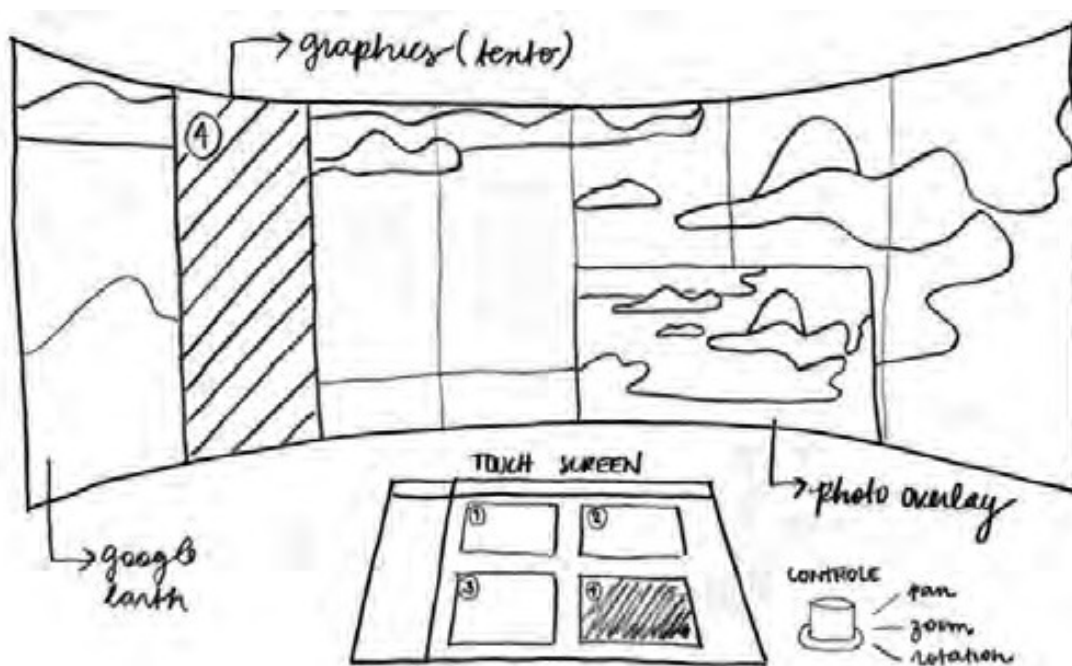


Figura 15 - Ilustração conceitual do demo: cena secundária Ao interagir com a tela *touchscreen* e/ou com o controle de navegação, o usuário imerge para uma cena de visualização oblíqua do modelo tridimensional do Rio de Janeiro. Nessa cena, o usuário vê a fotografia selecionada de Marc Ferrez (*photo overlay*) de um ponto de vista em que pode comparar a foto original com a vista atual da cidade (Google Earth). O recurso *graphics* é utilizado para oferecer informação complementar.
 Fonte: elaborado pelo autor.

4.2.3 Aspectos distributivos

Uma vez finalizada, a apresentação é disponibilizada para acesso via tela *multitouch* da plataforma LG. A tela mostra uma grade de cenas em miniatura (Fig. 16), cada uma correspondente a uma fotografia de Ferrez registrada em um lugar diferente do Rio de Janeiro. A tela *multi-display* (Fig. 17), por sua vez, apresenta a visualização das fotografias justapostas no modelo 3D do GE e outros elementos gráficos, como infoboxes e uma versão da fotografia em maior resolução¹⁶. O usuário pode selecionar as cenas na ordem que preferir e também pode navegar pelo espaço urbano usando o controle joystick. Nesse demo, os usuários são convidados a usar o controle para tentar posicionar seu ponto de observação no mesmo ângulo de visão no qual a fotografia foi registrada, comparando, dessa forma, a vista atual da cidade com a vista passada capturada pelo fotógrafo. Um vídeo da demonstração pode ser visto em <<https://youtu.be/yZpTpdq-j14>>.



Figura 16 - Tela *multitouch*. Fonte: acervo pessoal.

16 A necessidade de inserir uma nova versão da fotografia surgiu de uma limitação do elemento <*photo overlay*>, que reduz a resolução da imagem.



Figura 17 - A tela *multi-display* do LG apresenta uma fotografia de Ferrez justaposta ao modelo 3D do GE. Fonte: acervo pessoal.

REFERÊNCIAS

- FINDELSTEIN, A.; RANGE, M. Image Mosaics. In: 7th International Conference on Electronic Publishing, 1998. London, UK. Proceedings [...]. London, UK, 1998. Disponível em: <<http://dl.acm.org/citation.cfm?id=647506.725891>>. Acesso em: 10 jun. 2019
- GIANNELLA, J. R.; VELHO, L. Liquid Galaxy: a multi-display platform for panoramic geographic-based presentations. Technical Report. VISGRAF Lab. IMPA, 2018.
- GIANNELLA, J. R.; VELHO, L. Observatório2016. In: Digital Humanities, 2017, Montreal. Proceedings [...]. Montreal: McGill University, 2017. Disponível em: <<https://dh2017.adho.org/abstracts/558/558.pdf>>. Acesso em: 17 jun. 2018.
- GIANNELLA, J. R.; VELHO, L. Observatório2016 - Technical Report. IMPA, dez. 2016. Disponível em: <https://www.visgrafimpa.br/Data/RefBib/PS_PDF/tr-08-2016/tr-08-2016.pdf>. Acesso em: 27 jan. 2018
- IMPA. IMPA e IMS firmam parceria com foco em acervo fotográfico. 2018. Disponível em: <<https://impa.br/noticias/impa-e-ims-firmam-parceria-com-foco-em-acervo-fotografico/>>. Acesso em: 10 jun. 2019
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: 25th International Conference on neural information Processing Systems. Proceedings [...]. Lake Tahoe, Nevada, 2012.
- LIQUID GALAXY. Disponível em <<https://liquidgalaxy.endpoint.com/>>. 2019
- MANOVICH, L. Novas Mídias como Tecnologia e Ideia: dez definições. In: LEÃO, L. (Ed.). O Chip e o Caleidoscópio: Reflexões sobre as novas mídias. São Paulo: Senac, 2005.
- MANOVICH, L. What is visualisation? Visual Studies, v. 26, n. 1, p. 36-49, 15 mar. 2011.
- MANOVICH, L.; GATES, K. Media visualization: Visual techniques for exploring large media collections. In: Media Studies Futures. Malden, MA: Wiley-Blackwell, 2012a.
- MANOVICH, L. Media visualization: Visual techniques for exploring large media collections. In: GATES, K. (Ed.). Media Studies Futures. Malden, MA: Wiley-Blackwell, 2012.
- MANOVICH, L. Visualizing image and video collections: Examples, 2012b. Disponível em: <<https://www.slideshare.net/formalist/visualizing-image-and-video-collections-examples>>. Acesso em: jun. 2019

VELHO, L.; CARVALHO, P. Analysis and Synthesis from Image Collections, 2006. IMPA Summer Short Course. Disponível em: <<http://lvelho.impa.br/icol06/>>. Acesso em: jun. 2019

VELHO, L. ; FRERY, A.; GOMES, J. Image Processing for Computer Graphics and Vision. New York: Springer-Verlag, 2008.