

Storyboard para vídeos pessoais

Lucineide Rodrigues da Silva e Nielsen Cassiano Simões

Resumo—Atualmente, uma grande quantidade de vídeos está acessível em diversas páginas da Internet. Em geral, esses vídeos são de produção pessoal e não possuem um padrão específico. Este artigo trata da utilização de *storyboards* como ferramenta auxiliar na visualização do conteúdo do vídeo, permitindo a um usuário o conhecimento prévio de seu conteúdo antes de sua visualização completa.

Palavras-Chave—Ritmo Visual, *storyboard*, *keyframe*, vídeo pessoal.

Abstract—Nowadays, several videos are available on the Internet. In general, these videos have personal and nonstandard production. This paper deals about the use of storyboards as an alternative viewing for personal videos, making it possible to know the video content without watching it completely.

Keywords—Visual Rhythm, *storyboard*, *keyframe*, personal video.

I. INTRODUÇÃO

Nos últimos anos houve um grande crescimento no número de vídeos publicados na Internet. Como fatores relevantes para este crescimento podem ser mencionados a facilidade que os usuários encontram ao tentar acessar páginas na Internet, assim como disponibilizar seus vídeos nessas páginas e divulgá-los em *sites* e redes de relacionamento. A tarefa de um usuário à procura de um determinado tipo de vídeo não é uma tarefa trivial. Muitos usuários estão interessados apenas em se divertir, sem outro interesse definido. Existe uma grande dificuldade no momento de se optar por assistir a determinado vídeo, uma vez que, em sua maioria, tem-se poucas informações a seu respeito, como seu título e uma única imagem referente ao seu conteúdo.

Para isso, algumas páginas disponibilizam resumos textuais feitos pelo autor do vídeo, de forma que um usuário possa saber sinteticamente sobre o conteúdo abordado. Há ainda páginas que possibilitam aos usuários (que assistiram ao vídeo) deixar comentários a respeito de seu conteúdo, ajudando a outros usuários que porventura vierem a selecioná-lo.

Entretanto, uma forma alternativa para expor um resumo do conteúdo de um vídeo é apresentada pelo *site The Open Video Project* [1]. Neste *site* o usuário encontra um resumo visual para cada vídeo, facilitando no momento de se decidir sobre assisti-lo ou não. Este resumo visual, também chamado de *storyboard*, é obtido retirando do próprio vídeo imagens (quadros) que possam representar, de forma resumida, parte de seu conteúdo. Estas imagens são então apresentadas obedecendo a sequência dos acontecimentos existente no vídeo, dando a impressão de continuidade ao usuário e a percepção de seu conteúdo.

Lucineide Rodrigues da Silva, Universidade Federal do Paraná, Curitiba/PR, Brasil, E-mail: neidynhasilvas@gmail.com; Nielsen Cassiano Simões, Universidade Federal de Mato Grosso, Cuiabá/MT, Brasil, E-mail: nielsen@ic.ufmt.br.

Essa abordagem de se resumir um vídeo utilizando *storyboards* é baseada em técnicas de extração de quadros, que buscam identificar aqueles que demonstrem resumidamente seu conteúdo. Estes quadros extraídos são chamados de quadros-chave ou *keyframes*. Uma grande dificuldade para a seleção de *keyframes* é determinar qual quadro do vídeo é capaz de representar da melhor forma possível um subsegmento de seu conteúdo (podendo inclusive representá-lo por completo). Para isso, encontra-se na literatura muitos métodos para seleção de *keyframes* que, em sua maioria, são direcionados para um tipo de vídeo em específico. Dessa forma, a escolha do método está condicionada a um conhecimento prévio das características do vídeo a ser abordado, o que não acontece com vídeos pessoais.

Cada vídeo possui uma característica diferente decorrente tanto do conteúdo abordado quanto de seu processo de edição. O tipo de vídeo mais frequentemente encontrado nas páginas da Internet é denominado vídeo pessoal, ou seja, vídeos de produção audiovisual ou edições não profissionais, cuja preocupação está apenas no registro do acontecimento.

Não há na literatura nenhum método específico de seleção de *keyframes* para esse tipo de vídeo. Isso se dá devido à grande heterogeneidade encontrada, pois boa parte dos métodos de seleção de *keyframes* consideram características muito comuns em outros tipos de vídeo, tais como documentários, reportagens, noticiários, e outros. Uma das principais características está relacionada a ocorrência de transições entre tomadas do vídeo, além de sua duração e da estabilidade da câmera nas filmagens. Nenhuma dessas características pode ser determinante em um vídeo pessoal, pois não se pode precisar a ocorrência de alguma delas, dificultando o processo de seleção de *keyframes*.

Este trabalho apresenta um método de extração automática de *keyframes* para vídeos pessoais baseado em uma imagem representativa (Ritmo Visual) do vídeo. A partir dos quadros extraídos, são criados os *storyboards* de cada vídeo. Durante a execução deste trabalho foram selecionados e estudados os principais métodos de extração de *keyframes* encontrados na literatura. Cada método foi implementado para a geração dos *storyboards*, que foram utilizados em uma avaliação comparativa com os *storyboards* resultantes da utilização do método proposto. Todos os vídeos pessoais analisados durante a execução do trabalho foram encontrados em *sites* especializados em visualização e publicação desse tipo de vídeo.

II. MÉTODOS DE EXTRAÇÃO DE KEYFRAMES

A maior dificuldade na extração de *keyframes* é selecionar um quadro que melhor represente o conteúdo de um segmento do vídeo sem redundâncias. Isso torna a extração um método muito específico para cada tipo de vídeo.

Uma das primeiras formas de se extrair *keyframes* foi proposta por Zhang et al. em [2], em que tomadas de vídeo não

são consideradas. A extração dos *keyframes* é então realizada a cada subsegmento pré-definido de quadros. Entretanto, essa abordagem não reduz significativamente as redundâncias para segmentos de vídeo cujas tomadas são, em geral, muito maiores que o subsegmento considerado.

Zhuang et al. apresentam em [3] um método de extração de *keyframes* que primeiramente utiliza um algoritmo de detecção de bordas para determinar tomadas. Para cada tomada, é realizada uma classificação não supervisionada baseada no histograma dos quadros utilizando um limiar previamente definido. Ao final, é definido um *keyframe* para cada classe. Semelhantemente, Chang e Chen propõem em [4] uma divisão em quatro etapas para a extração de *keyframes*. Primeiro, identifica-se tomadas utilizando semelhança de cores e detecção de bordas. Em seguida, é feita a extração de três quadros para cada tomada. Então, estes quadros são classificados de acordo com a similaridade de cena e conteúdo. Por último, para cada classe, apenas um *keyframe* é selecionado considerando a sua importância relativa às informações de energia e cor. Neste caso, para vídeos que possuem tomadas de grande duração, os *keyframes* extraídos dessa tomada podem não representar adequadamente seu conteúdo. Além do mais, a classificação realizada no segundo método vincula o *keyframe* ao conceito de cena, eliminando a dinâmica atribuída na produção do vídeo.

O método proposto por de Avila et al. em [5] baseia-se em uma sub-amostragem do vídeo e na classificação dos quadros, analisando o histograma de cor e perfis de linha. Na sequência, é selecionado um quadro para cada classe baseado em sua representatividade e, em seguida, os *keyframes* são comparados de acordo com o histograma para evitar que *keyframes* com conteúdo semelhantes sejam exibidos. Uma desvantagem dessa abordagem está relacionada ao fato de ser necessário estabelecer um valor máximo de *keyframes* a serem extraídos. Dessa forma, o conhecimento prévio sobre o vídeo, como tamanho ou número de quadros existentes, acaba sendo primordial para se obter um bom resultado. A filtragem dos *keyframes* selecionados é outro ponto que também pode distorcer o resultado, desconsiderando a dinâmica atribuída na produção do vídeo.

Os métodos de extração de *keyframes* até aqui considerados levam em conta as características do vídeo, como por exemplo, a existência de tomadas ou informações gerais do vídeo. Entretanto, vídeos pessoais não possuem padrões estabelecidos, uma vez que os usuários fazem seu registro de forma amadora e sem preocupações com as técnicas de produção de vídeo.

III. MÉTODO DESENVOLVIDO

Não há na literatura um método de extração de *keyframes* específico para vídeos pessoais. Este é um dos objetivos deste trabalho, no qual se apresenta uma abordagem específica de extração de *keyframes* para este tipo de vídeo.

Algumas particularidades do vídeo pessoal estão relacionadas ao fato de não se ter informações sobre a existência de transições (tomadas), pois os usuários podem publicar esses vídeos na Internet sem a necessidade de editá-los.

Além disso, em um vídeo pessoal não se tem informações sobre o volume de movimentos de seu conteúdo. Estes vídeos

podem conter muitos ou poucos movimentos internos, não sendo interessante definir esta característica como fundamental, pois o que é importante em um vídeo movimentado se diferencia do que é importante em um vídeo 'calmo'. Dessa forma, não foram consideradas as técnicas baseadas em movimentos internos do vídeo.

A maioria dos vídeos pessoais possui apenas uma tomada e, em geral, são relativamente curtas. Abordagens que utilizam detecção de tomada para extração de *keyframes* são pouco eficientes neste caso, cujos resultados não representam adequadamente o conteúdo do vídeo.

A análise de um vídeo digital pode ser realizada de duas formas: percorrendo-se o vídeo analisando as informações quadro a quadro, ou utilizando-se uma imagem representativa do vídeo, também chamada de Ritmo Visual. Essa imagem de Ritmo Visual é uma imagem resultante da aplicação de uma função de transformação em um vídeo digital, como apresentado na Definição 1 [6].

Definição 1 (Ritmo Visual: definição geral): Seja V um segmento de vídeo digital de N quadros, no domínio $2D + t$. O Ritmo Visual é a transformação de cada quadro f_t em uma coluna da imagem I_{GVR} definida como:

$$I_{GVR}(t, z) = \tau(f_t, z)$$

em que f_t representa o t -ésimo quadro do vídeo V e z é definido conforme a função de transformação τ .

A imagem de Ritmo Visual permite reduzir significativamente a análise do vídeo, pois todos os quadros são analisados apenas uma vez, transformando o problema de análise de vídeo em um problema de segmentação de imagem, preservando a característica temporal presente no vídeo.

Simoies et al. apresentam em [6] uma abordagem para a extração de *keyframes* baseada em uma imagem de Ritmo Visual específica para jogos de futebol, utilizando algum algoritmo de detecção de tomadas. Entretanto, como esta abordagem considera a presença de gramado como cor predominante no vídeo, ela não é recomendada para outros tipos de vídeos para os quais não se pode atribuir alguma cor predominante.

Inicialmente a proposta para a extração de *keyframes* em vídeos pessoais foi utilizar uma imagem de Ritmo Visual considerando alguma transformação já definida anteriormente, como as apresentadas em [7] e [8]. Após algumas análises, as informações espaciais obtidas nessa imagem não eram suficientes para caracterizar vídeos pessoais, uma vez que são muito genéricos e nem sempre apresentam eventos temporais, como as transições entre tomadas.

Como alternativa, procurou-se então utilizar uma função de transformação para a imagem de Ritmo Visual que considerasse as características globais de cada quadro. Dessa forma, optou-se por utilizar o histograma de cor [9] de cada quadro como função de transformação (Definição 2), construindo uma imagem denominada Ritmo Visual por histograma [6] com níveis de cinza variando de 0 a 255. A partir das Definições 1 e 2, tem-se a imagem de Ritmo Visual por Histograma. Na Fig. 1 é apresentada a Imagem de

Ritmo Visual por Histograma de um vídeo. Para a eliminação de alguns ruídos da imagem de Ritmo Visual por Histograma, que por ventura apareçam devido a qualidade do vídeo, por exemplo, é necessária a utilização de uma suavização pelo filtro da mediana [10].

Definição 2 (Imagem de Ritmo Visual por Histograma):

Seja VR_{hist} a imagem de Ritmo Visual, conforme Definição 1, cuja função de transformação τ é definida como uma função global do quadro da forma:

$$\tau(f_t, z) = Hist_t[z],$$

em que $Hist_t$ é o histograma do quadro f_t , e $z \in [0, 255]$.

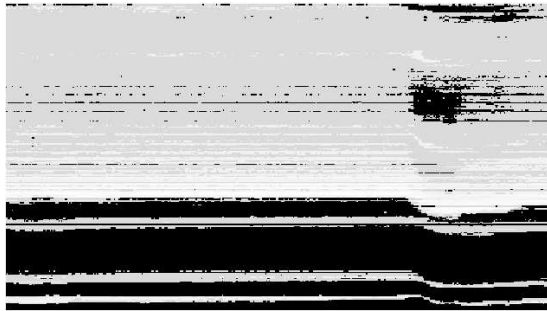


Fig. 1. Imagem de Ritmo Visual por Histograma de um vídeo.

O processo de análise da imagem de Ritmo Visual por Histograma inicia-se com a determinação da região de interesse na imagem. O que se deseja determinar nessa imagem é o comportamento dos pixels de maior ocorrência. Essa região é representada na imagem de Ritmo Visual por Histograma pelos níveis de cinza mais claros (bem próximos do branco). A seleção dessa região de interesse é obtida por meio de uma limiarização utilizando um parâmetro T , de forma que os pixels da imagem de Ritmo Visual por Histograma cujo valor seja inferior a T são substituídos por zero.

Na sequência, como o que se deseja é verificar o comportamento dos pixels na imagem, deve-se então realizar uma Projeção Vertical de forma a obter uma topologia e observar seu comportamento no tempo (eixo x da imagem).

A Projeção Vertical é calculada somando-se todos os valores de cada coluna da imagem [6] de Ritmo Visual por Histograma. Dessa forma, cria-se um vetor com o resultado de cada coluna.

Os valores resultantes da Projeção Vertical em sua maioria são bem próximos, o que dificulta a identificação de alterações no decorrer do vídeo. Entretanto, essas variações podem ser realçadas calculando-se o Laplaciano [6], conforme Definição 3, de forma que elementos vizinhos com valores muito próximos sejam menos significativos que aqueles com valores mais distantes.

Definição 3 (Laplaciano): Seja P um vetor de tamanho W . O vetor laplaciano L_{ap} , definido sobre P , pode ser representado por:

$$L_{ap}(t) = |2 * P(t) - P(t - 1) - P(t + 1)|,$$

para $t \in [1, W - 2]$, sendo W a largura da imagem.

Com as variações realçadas pela aplicação do Laplaciano, pretende-se determinar quais os instantes do vídeo em que essas variações foram mais intensas, já que podem representar algo a ser observado pelo usuário. Para tanto, é necessário determinar a maior variação presente no vetor do Laplaciano.

Nem sempre o ponto de maior variação é o ponto chave para resumir o conteúdo do vídeo. Por exemplo, se tivermos um vídeo sem nenhuma atividade inicial e de repente um objeto entra em cena, provavelmente o ponto de maior variação será a entrada do objeto em cena. Entretanto, se extrairmos somente o quadro onde ocorre a entrada do objeto, não mostraremos os momentos iniciais do vídeo, em que outros tipos de variações podem estar acontecendo. Certamente este resumo não seria adequado para este tipo de vídeo.

Dessa forma, considerando que no vídeo poderemos ter grandes e pequenas variações, o maior valor encontrado no vetor do Laplaciano é então utilizado para uma limiarização. Esta limiarização elimina aquelas variações que representam menos de 50% da maior variação existente.

Incluindo-se as posições iniciais e finais do vídeo, que podem representar sempre um ponto de variação, são então calculados os pontos para extração de *keyframes*. Esse cálculo considera todos os pontos de variação encontrados no vídeo com pelo menos seis quadros de distância um do outro. São calculadas as posições centrais entre dois pontos consecutivos de variação considerados, determinando os pontos a serem extraídos os *keyframes*. O vídeo é então percorrido a fim de se extrair os quadros que estejam em cada posição central calculada.

Como resultado, os *keyframes* extraídos são utilizados para a composição do *storyboard* do vídeo analisado.

IV. RESULTADOS

Os vídeos utilizados neste trabalho são vídeos pessoais retirados de *sites* especializados em publicação e visualização de vídeos na Internet. Foram utilizados ao todo vinte vídeos pessoais, cujas características são diferentes entre si, tais como: vídeos curtos, longos, com imagens muito repetitivas, com muitos movimentos durante a filmagem, com pouca iluminação, entre outras.

Para a geração de *storyboards* foram considerados os seguintes métodos de extração de *keyframes* baseados em algumas características comuns em vídeos:

- **Método 1:** seleciona 1(um) *keyframe* dentro do vídeo a cada n quadros;
- **Método 2:** seleciona *keyframes* considerando as tomadas existentes no vídeo;
- **Método 3:** seleciona *keyframes* utilizando agrupamento dos quadros de acordo com a similaridade entre eles;
- **Método 4:** Os *keyframes* gerados utilizando o IMARS [11], um *software* desenvolvido para esta finalidade;
- **Método 5:** O método proposto.

Para a implementação destes métodos utilizou-se a linguagem C e a biblioteca OpenCV ¹ em um sistema operacional GNU/Linux. A descrição de como cada um desses métodos

¹<http://opencv.willowgarage.com/wiki/>

foi implementado é dada na sequência. Utilizou-se também um *software* para indexação automática, classificação e busca em grandes coleções de imagem e vídeo digital (IMARS), onde se pode realizar a extração de *keyframes* em vídeos digitais [11].

O método 1 utiliza um parâmetro n para que se possa extrair os *keyframes*. Ele percorre o vídeo considerando uma quantidade fixa de n quadros entre cada *keyframe* extraído. Para tornar esse método automático, utilizou-se um valor percentual relativo à quantidade de quadros existentes no vídeo, da forma apresentada na Tabela I. De acordo com o tamanho do vídeo, o valor de n é definido de modo a resultar um *storyboard* com no máximo $x\%$ dos quadros do vídeo. Essa definição considera que vídeos pessoais muito longos são, em sua maioria, vídeos com quadros bastante repetitivos que poderiam gerar *storyboards* muito redundantes.

TABELA I

DEFINIÇÃO DO VALOR DE n PARA CADA *storyboard* PELO MÉTODO 1.

Quantidade de quadros	% de quadros no <i>storyboard</i>
até 400	4%
de 401 até 999	2%
acima de 999	1,5%

Durante o desenvolvimento do método 2 foi necessário o uso de alguma forma de detecção de transições dentro do vídeo. O χ^2 foi então utilizado para a comparação de histogramas de cor dos quadros do vídeo para a detecção de tomadas, visto que este apresentou melhor desempenho em comparação a outros métodos tradicionais [12]. Terminada a etapa de detecção de tomadas no vídeo é necessário determinar o ponto a ser extraído o *keyframe* de cada tomada. Foi determinada a posição central de cada tomada para a extração. Em algumas tomadas observou-se que as mesmas eram muito próximas umas das outras. Para que isso não atrapalhasse a extração de *keyframes*, ficou determinado que cada tomada deveria conter no mínimo seis quadros, o equivalente a 0,2 segundos.

No método 3 considerou-se o algoritmo *k-means* para a clusterização, aplicada sobre os histogramas de cor de todos os quadros do vídeo, como proposto pelos autores em [5]. Entretanto, como o algoritmo *k-means* necessita que seja definida uma quantidade k de *clusters* a ser considerada, o valor k foi parametrizado de modo que k corresponda a uma porcentagem da quantidade de quadros do vídeo conforme apresentado na Tabela II.

TABELA II

DEFINIÇÃO DO VALOR DE n PARA CADA *storyboard* PELO MÉTODO 3.

Quantidade de quadros	Valor de k
até 400	4%
de 401 até 4000	1%
acima de 4000	0,5%

A extração realizada pelo método 4 utiliza os parâmetros padrões do *software* considerado. Este método realiza a extração de *keyframes* baseando-se na diferença visual entre os quadros do vídeo, tais como cor, textura, forma e disposição espacial [11].

Já o método 5, implementado seguindo os passos descritos na seção anterior, considerou o valor 170 para o parâmetro de limiarização T , representando 67% no histograma normalizado.

Os *storyboards* gerados utilizando os métodos descritos acima foram disponibilizados em uma página na Internet para que os usuários pudessem avaliá-los e opinar a seu respeito. Para cada vídeo, foram então gerados cinco *storyboards*, um para cada método proposto.

A validação feita pelos usuários é uma etapa muito importante no processo de avaliação dos *storyboards* gerados durante o trabalho. Essa avaliação foi disponibilizada aos usuários na Internet. Os avaliadores foram, em geral, alunos de vários cursos da Universidade Estadual de Mato Grosso do Sul - UEMS, totalizando vinte e cinco, e que já tinham o costume de assistir vídeos pessoais em diversos *sites* da Internet. Cada aluno avaliou os cinco *storyboards* para um conjunto de no máximo 3 vídeos, e para cada vídeo, escolheram dentre os métodos o que gerou o melhor *storyboard*. Para que a avaliação pudesse ser iniciada, era obrigatório assistir ao vídeo por completo.

Analisando as respostas fornecidas pelos usuários, concluiu-se que 78,34% consideraram que o uso de *storyboards* para resumir um vídeo é uma forma eficiente para ajudar na tomada de decisão por assistir ou não a determinado vídeo. Dentre os avaliadores 71,97% afirmaram que o *storyboard* mostrado resumia de forma adequada o conteúdo do vídeo.

A Tabela III apresenta o resultado da avaliação dos usuários quanto à quantidade de imagens exibidas em cada *storyboard* gerado. De forma geral, essa foi a característica utilizada pelos usuários para escolher qual método de criação de *storyboard* foi o melhor. Analisando as respostas, concluímos que o método 3 foi o melhor avaliado, seguido pelo método 5 e, na sequência, pelo método 4.

TABELA III

AVALIAÇÃO DOS USUÁRIOS QUANTO À QUANTIDADE DE IMAGENS APRESENTADA EM CADA *storyboard*.

Quantidade de Imagens	Método 1	Método 2	Método 3	Método 4	Método 5
Pequena	6,25%	56,25%	9,67%	9,67%	51,61%
Boa	28,15%	31,25%	61,29%	38,70%	41,93%
Grande	65,62%	12,50%	29,03%	51,61%	6,45%

Analisando os dois melhores métodos de acordo com a opinião dos usuários e as características dos vídeos, é possível concluir que quando o vídeo tem pouca alteração em seu conteúdo o usuário preferiu ver imagens muito parecidas a uma quantidade pequena de imagens. Um exemplo deste caso pode ser visto nas Fig. 2 e 3.

Os casos onde os usuários consideraram a quantidade de imagens exibidas no *storyboard* pequena foram os vídeos com pouco conteúdo visual, ou seja, vídeos sem grandes alterações em seu conteúdo. Como o método proposto avalia as alterações ocorridas no decorrer do vídeo, se um vídeo tem poucas variações os pontos de extração de *keyframes* serão distantes, diminuindo a redundância e consequentemente o *storyboard* apresentará uma quantidade pequena de imagens.

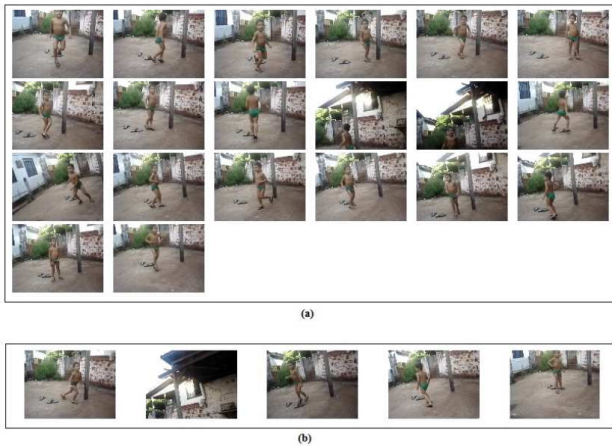


Fig. 2. Storyboard de um vídeo. Na parte (a) gerado pelo método 3. Na parte (b) gerado pelo método 5.

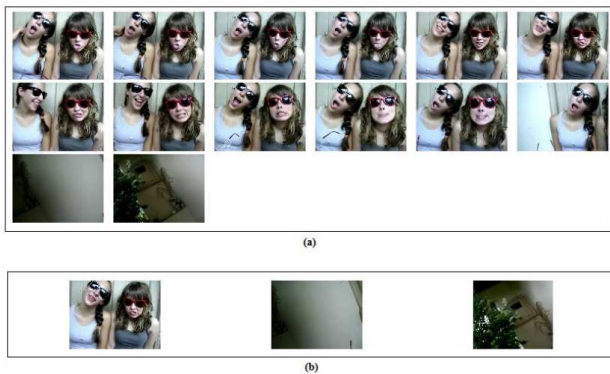


Fig. 3. Storyboard de um vídeo. Na parte (a) gerado pelo método 3. Na parte (b) gerado pelo método 5.

Por outro lado, observou-se que nos casos onde os usuários disseram ter uma quantidade grande de imagens, o que ocorre é que o vídeo tem muitas alterações em seu conteúdo (seja pela movimentação da câmera ou dos objetos em cena) gerando muitos pontos de extração de *keyframes*. Esse problema pode ser resolvido inserindo ao método uma etapa de filtragem dos quadros escolhidos como *keyframes* para evitar a redundância de informações nos *storyboards*, mas preservando a dinâmica do vídeo.

V. CONCLUSÕES

O método proposto para a extração de *keyframes* baseado em uma imagem representativa do vídeo (Ritmo Visual) foi desenvolvido voltado ao resumo de vídeos pessoais e considera as particularidades desse tipo de vídeo. Durante o estudo, foram pesquisados e implementados outros métodos de extração de *keyframes* tradicionais, baseados na análise quadro a quadro. Ao final, todos os métodos implementados, incluindo o método proposto e o resultado obtido por meio de um *software* específico para extração de *keyframes*, foram utilizados para a criação de *storyboards* de vídeos pessoais. Esses *storyboards* foram, então, submetidos à avaliação de usuários para a comparação entre os métodos considerados neste trabalho.

Com os resultados da pesquisa foi possível concluir que 78,34% dos usuários consideraram que o uso de *storyboards* é uma ferramenta eficiente para ajudar na tomada de decisão sobre assistir ou não a um vídeo, e 71,97% afirmaram que a quantidade de quadros apresentado em cada *storyboard* foi suficiente para o resumo dos vídeos assistidos.

O método melhor avaliado foi o que utiliza clusterização para a extração de *keyframes*, mesmo em alguns casos gerando *storyboards* redundantes. Já o método desenvolvido neste trabalho obteve a segunda melhor avaliação dentre os demais métodos.

Foi possível analisar que 51,61% dos usuários afirmaram que o método desenvolvido utilizava uma quantidade pequena de imagens nos *storyboards* gerados. Dessa forma, conclui-se que quando os vídeos têm pouco conteúdo visual o usuário prefere ver *storyboards* redundantes à um *storyboard* com uma pequena quantidade de imagens. Isso ocorreu principalmente em vídeos cujo conteúdo visual é muito semelhante.

Para vídeos com pouca atividade, talvez fosse interessante adicionar mais *keyframes* intermediários para evitar um *storyboard* com uma quantidade muito reduzida de imagens, mesmo que semelhantes, melhorando a satisfação dos usuários. Entretanto, o *storyboard* apresentado pelo método proposto é eficiente para resumir o vídeo, sendo interessante principalmente para a exibição alternativa em dispositivos portáteis e para indexação.

REFERÊNCIAS

- [1] Interaction Design Laboratory, "The open video project: a shared digital video collection." <http://www.open-video.org/>, acessado em julho de 2011, 2000.
- [2] H. J. Zhang, C. Y. Low, S. W. Smoliar, and J. H. Wu, "Video parsing, retrieval and browsing: an integrated and content-based solution," in *MULTIMEDIA '95: Proceedings of the third ACM international conference on Multimedia*, (New York, NY, USA), pp. 15–24, ACM, 1995.
- [3] Y. Zhuang, Y. Rui, T. S. Huang, and S. Mehrotra, "Adaptive key frame extraction using unsupervised clustering," *Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on*, vol. 1, pp. 866 – 870, 1998.
- [4] I.-C. Chang and K.-Y. Chen, "Content-selection based video summarization," in *Digest of Technical Papers International Conference on Consumer Electronics (ICCE)*, pp. 1–2, 2007.
- [5] S. E. F. de Avila, A. da Luz Jr., A. de A. Araújo, and M. Cord, "Vsumm: An approach for automatic video summarization and quantitative evaluation," in *XXI Brazilian Symposium on Computer Graphics and Image Processing*, pp. 103–110, 2008.
- [6] N. C. Simões, N. J. Leite, and B. Marcotegui, "Automatic key-frame extraction from broadcast soccer videos," in *International Conference on Computer Vision Theory and Applications*, 2009.
- [7] S. J. F. Guimarães, M. Couprie, N. J. Leite, and A. de Albuquerque Araújo, "Análise do vídeo a partir da sua simplificação," in *Anais do II Workshop em Tratamento de Imagens*, 2001.
- [8] N. C. Simões, "Detecção de algumas transições abruptas em sequências de imagens," Master's thesis, Instituto de Computação, Universidade Estadual de Campinas, 2004.
- [9] E. M. R. Silva, "Desenvolvimento de um sistema de busca de imagens na Internet por similaridade de cor baseado em histograma," tech. rep., Pontifícia Universidade Católica de Minas Gerais, 2006.
- [10] R. C. Gonzalez and R. E. Woods, *Processamento de Imagens Digitais*. Edgard Blücher Ltda, 2000.
- [11] IBM T. J. Watson Research Center, "Imars - ibm multimedia analysis and retrieval system." Disponível em <http://www.alphaworks.ibm.com/tech/imars>. Último acesso em julho de 2011, 199-?.
- [12] L. R. da Silva, "Avaliação de técnicas de extração automática de keyframes para visualização de vídeos pessoais," tech. rep., Universidade Estadual de Mato Grosso do Sul, 2010.