

6 Conclusão

Neste capítulo, conclui-se a dissertação, apresentam-se suas principais contribuições e propõem-se trabalhos futuros baseados no que foi elaborado.

6.1. Resumo

A meta principal deste trabalho foi elaborar um algoritmo para a classificação e segmentação do áudio MPEG-1 Layer 2 que tivesse o menor tempo de execução possível sem que fosse perdida a precisão no resultado da sua análise.

Foram analisados diversos métodos para a realização dessas tarefas sugeridos na literatura, ilustrando os três domínios empregados nesses trabalhos: não-comprimido, comprimido e nos fatores de escala. Como ilustrado, há poucos trabalhos que se baseiam nos fatores de escala, que é o domínio mais rápido, porém onde se trabalha com menos informações. Em particular, não foi encontrada nenhuma menção na literatura da segmentação sendo feita nesse domínio.

Esta dissertação apresentou métodos para as soluções de cada um dos problemas, sendo que a segmentação se utiliza das informações geradas na classificação como parte importante de sua base de informações.

O método de classificação baseia-se no método apresentado pelo grupo da Universidade de Dublin, porém com diversos aperfeiçoamentos. Como inicialmente os resultados encontrados a partir desse método não foram os esperados, foram então desenvolvidas novas análises que foram integradas ao método, em busca de resultados mais precisos. Além disso, foi integrada ao algoritmo uma nova classificação, a de aplausos. Essa classificação é muito importante, principalmente como indício para a segmentação e como um mecanismo para análise do tipo de vídeo como um todo. Mesmo considerando

todos esses aperfeiçoamentos na classificação, a precisão do algoritmo ainda foi melhorada, atingindo uma excelente média de acertos.

Apesar de já haver um trabalho no qual o algoritmo de classificação desenvolvido se baseou, o que foi proposto nesta dissertação apresenta uma clara evolução. Sua principal contribuição é a adição da detecção de aplausos, além do grau de precisão não apenas ter sido mantido, e sim aumentado, mesmo com essa nova categoria.

A segmentação feita a partir dos fatores de escala é a outra grande contribuição apresentada. Para seu desenvolvimento, buscou-se inspiração no artigo de J. Foote (2000) para a etapa de detecção de possíveis cortes. A partir dessa idéia, desenvolveu-se um algoritmo completamente novo para a realização dessa tarefa.

Esse algoritmo se baseia na detecção dos cortes e na aplicação de filtros para estabelecer os candidatos que representam mudanças reais de cena. O resultado apresentado foi excelente, com um alto grau de detecção de mudanças reais de cena. A maneira como o algoritmo foi estruturado permite flexibilidade, pois diversos aspectos são parametrizados, o que possibilita seu ajuste fino (“tuning”), se necessário, para um acervo de vídeos de um tipo com características específicas e semelhantes. A taxa alta de cortes falsos que ocorre em alguns tipos de vídeos não se mostra um problema estrutural do algoritmo. Uma análise detalhada dos eventos que ocorrem nesses vídeos e causam os cortes falsos poderá gerar um novo filtro que resolva isso, conforme será mencionado no sub-capítulo seguinte, de trabalhos futuros.

O desenvolvimento de ambos os algoritmos, trabalhando somente a partir dos fatores de escala, permite uma eficiência muito grande. Considerando que a maior parte do tempo gasto na execução dos algoritmos usuais ocorre na recuperação das informações do vídeo, a minimização da quantidade de informações que são lidas permite garantir que o tempo gasto para a execução do algoritmo desenvolvido será menor do que o de outros trabalhos.

A maior dificuldade de trabalhar nesse domínio é a falta de precisão das informações dos fatores de escala. Durante esta dissertação, mostrou-se que os métodos desenvolvidos superaram essa barreira, e obtiveram um grau de precisão excelente. Ambos os algoritmos trabalham com uma taxa de acerto média na casa

dos 90%, e ainda permitem evoluções com a integração de novas análises às que já são realizadas.

6.2. Trabalhos futuros

Esta dissertação concentrou-se na análise das informações do áudio de um vídeo isoladamente. Porém, para a segmentação de um vídeo, a análise das imagens é primordial. Na maioria dos estudos realizados que unem as duas fontes, o resultado da análise da imagem é usado como base, e as informações do áudio são utilizadas para qualificar o resultado obtido a partir das imagens. Assim, a incorporação do algoritmo desenvolvido nesta dissertação a um algoritmo que trabalhe a partir das imagens se torna essencial, e possibilitaria um aumento da precisão de ambos.

É importante neste caso frisar que há mudanças de cenas em diferentes tipos de programas que são mais facilmente detectadas a partir do áudio. Portanto, não se deve considerar o resultado dos algoritmos que se baseiam no áudio, como o proposto nesta dissertação, apenas como qualificadores. O ideal para um algoritmo final seria utilizar as informações das duas fontes combinadas, para atingir a maior precisão possível, sem que se percam cortes que podem ser vistos em apenas um caso.

A união da análise do áudio com o vídeo permitirá uma melhora significativa na taxa de cortes falsos detectados, já que se torna possível a observação dos cortes detectados pelo vídeo como confirmação ou não de candidatos gerados pelo áudio, e vice-versa. Porém, há outras pesquisas que podem possibilitar uma melhora nessa taxa.

O algoritmo de segmentação possui alguns parâmetros que foram padronizados durante os testes realizados nesta dissertação, que teve como objetivo testar a funcionalidade do algoritmo sem informações prévias dos vídeos. Porém, a existência de tais informações pode permitir variar esses parâmetros em busca de maior efetividade. O tamanho da janela de análise e o tamanho mínimo de um segmento são dois bons exemplos disso. Um vídeo com cenas longas como característica, por exemplo, pode ter sua análise melhorada a partir do aumento do valor desses dois parâmetros.

Nesses casos, deve-se sempre estabelecer uma conexão com o limite do grau de inovação. Esse limite é um parâmetro que estabelece o quão sensível o algoritmo é a candidatos a cortes. Assim, ele também pode ser alterado a partir tanto de informações do vídeo como dos valores estabelecidos para os outros parâmetros, de acordo com tais informações.

A obtenção de informações gerais sobre o tipo de vídeo está geralmente ligada a dicas dos usuários. A classificação, porém, pode ter uma participação significativa nessa tarefa. É possível identificar padrões para diferentes tipos de vídeos, em termos da variação da classificação dos seus envelopes. Um vídeo que contém longos períodos de música cortados por períodos breves de falas e aplausos, por exemplo, pode ser identificado a partir desse padrão como uma apresentação musical longa. Em um caso como esse, o programa já poderia se ajustar para detectar os intervalos entre músicas como as principais mudanças de cena existentes, aumentando a efetividade da segmentação.

Outro fator que pode aumentar a efetividade dos algoritmos é a incorporação de novas técnicas. No caso da classificação, onde já temos uma precisão muito alta, um foco possível é a resolução dos problemas detectados. O caso da música clássica, por exemplo, já foi pesquisado e teve seu desempenho melhorado. Mas mesmo esse caso ainda não atingiu a precisão dos tipos de vídeos onde o algoritmo se sai melhor. Assim, o estudo de casos específicos que causam erros pode aprimorar o algoritmo.

Outro meio de se conseguir isso é a inclusão de novos parâmetros. Neste caso, seria estudado como diferentes parâmetros utilizados por outros algoritmos, como o centróide de frequências, se aplicam no domínio dos fatores de escala, para que pudessem ser incluídos no algoritmo como um novo fator a ser considerado na classificação.

Para a segmentação, o desenvolvimento deve concentrar-se na incorporação de novos filtros. Através deles é possível diminuir a taxa de cortes falsos obtidos e até possibilitar um relaxamento dos limites para a detecção de candidatos, aumentando também a taxa de acertos. Um exemplo citado durante esta dissertação é um filtro a partir de pequenos intervalos de fala. Esses intervalos, que em alguns casos podem causar a detecção de um corte falso, ainda não foram tratados especificamente para a composição de um filtro.

Outra possibilidade é o ranqueamento. Para alguns tipos de vídeos, um filtro baseado simplesmente no valor do grau de novidade encontrado pode reduzir drasticamente a taxa de cortes falsos. Nesse caso, ordenam-se os candidatos a mudanças de cenas a partir do grau de novidade, e são selecionados apenas os melhores candidatos, a partir dessa característica. Esse filtro se mistura com o parâmetro que limita o grau de novidade mínimo para determinar um candidato, porém ele pode servir como um facilitador até para o usuário, de modo que não seria necessário alterar esse parâmetro caso o ranqueamento já estabelecesse um corte dos candidatos mais fracos. Além disso, foi observado que em alguns vídeos os candidatos onde o grau de novidade é mais alto são exatamente as mudanças de cena verdadeiras – o que indica que esse filtro pode ser muito útil.

O último tópico interessante encontrado é a detecção de interlocutores. Um estudo de métodos que possibilitem essa detecção no domínio dos fatores de escala pode ser muito útil para a segmentação de alguns tipos específicos de vídeos, como telejornais. Nesse caso específico, a mudança do apresentador geralmente indica que uma notícia diferente está sendo apresentada e é portanto um forte indicativo de mudanças de cena.