

5 Conclusão

Este capítulo faz o fechamento da dissertação, apresentando as principais contribuições, comparando os conceitos e algoritmos propostos com trabalhos correlatos em segmentação de vídeo e propondo trabalhos futuros.

5.1. O Problema do Limiar

O desenvolvimento de algoritmos para a detecção das tomadas de câmera de vídeos em grandes acervos busca a identificação dos cortes entre duas tomadas consecutivas de forma automática. As diversas abordagens propõem métodos menos custosos do que a identificação manual dos cortes por visualização de todo o conteúdo dos vídeos usando as formas tradicionais de *Video Tracking*.

Um desafio apontado pelas diversas abordagens encontradas na literatura para o problema da detecção das tomada de câmera é a configuração do limiar para análise das funções características de corte. Por não ser conhecida uma função característica ideal, não é possível garantir a escolha de um limiar para separação perfeita de todos os cortes dos vídeos sem a inclusão de cortes falsos.

Não é conhecido um sistema de equações que conduza a uma solução matematicamente ótima para a detecção de cortes, visto que a caracterização de cortes depende de fatores que, além de serem numerosos, variam pela influência de características como o gênero dos vídeos (esporte, jornalismo, ação, animação, etc...), a natureza das imagens (variações de intensidade e frequências) e as escolhas do processo de compactação (pois o padrão MPEG deixa livre diversas escolhas para os compactadores).

Os diversos tipos de algoritmo para detecção das fronteiras entre duas tomadas de câmera se diferenciam principalmente pela métrica usada para a identificação dos quadro-corte e pela natureza dos dados analisados por essas métricas.

Pelo o exposto acima, os métodos que propõem uma única métrica para a classificação das tomadas de câmera tendem a apresentar erros de detecção, amenizados pela heurística mista proposta nesta dissertação. A união das diversas métricas compensa os pontos fracos apresentados pela utilização isolada de uma única métrica, permitindo o crescimento do número de acertos e uma adaptação de vídeos de natureza genérica.

5.2. Segmentação Baseada na História da Codificação

A abordagem proposta por este trabalho ameniza a tarefa de encontrar um limiar ótimo para a caracterização dos cortes, propondo uma nova abordagem através da configuração de diversos fatores mais suaves, partindo da idéia de descartar quadros com conteúdo similar em refinamentos sucessivos, de forma a eliminar os quadros não-cortes.

A idéia básica do algoritmo proposto é identificar padrões falsos de corte pelo levantamento da “história” da codificação, através de uma análise comparativa do comportamento dos quadros de todo o vídeo. O presente trabalho propõe que esses padrões falsos de corte detectados sejam denominados de "assinatura do codificador". Para distinguir os quadros com características de corte real dos quadros com características influenciadas pelo codificador, é proposta uma sequência de filtragens nos dados analisados, de forma a suavizar a influência dessas assinaturas nos valores obtidos pelas métricas de caracterização de similaridade. Este trabalho também propõe a denominação de "segmentação baseada em história da codificação" para essa nova classe de métodos de segmentação.

No algoritmo proposto, diversas métricas de similaridade são aplicadas procurando descartar quadros semelhantes de forma adaptável aos gêneros dos vídeos e aos padrões de compactação distintos, inicialmente desconhecidos.

As etapas de refinamento são planejadas de forma a considerar menos problemática a existência de cortes falsos no conjunto final de solução, do que a procura por cortes perdidos. A correção manual da perda de cortes no conjunto de quadros apresentado, como solução pelo algoritmo, exige uma nova análise de

todo o conteúdo do vídeo pela visualização, usando métodos tradicionais de acesso, enquanto que a correção manual da apresentação dos quadros chaves excedentes, representativos das tomadas falsas de câmera, pode ser feita analisando apenas o próprio conjunto solução, sendo sua correção mais fácil por causa do menor volume de dados manipulados.

As contribuições mais significativas do presente trabalho, direcionada à segmentação de vídeo no domínio comprimido, são a identificação da existência de padrões de assinatura do codificador e o desenvolvimento de heurísticas adaptadas para a distinção dessas assinaturas. Essa identificação procura características da história da compactação para distinguir os padrões dos quadros localizados na fronteira entre duas tomadas de câmera, de padrões dos quadros em seu interior.

A análise das funções características encontradas na literatura das abordagens com análise de informações codificadas demonstrou o aparecimento de padrões de corte falsos relativos a decisões de codificação. Esse conjunto significativo de cortes falsos é identificado neste trabalho como resultante da influência das assinaturas encontradas durante o levantamento da história da codificação.

O uso de árvores de ordenação para comparação dos valores obtidos pelas métricas, em relação ao comportamento de todo o vídeo, permite uma maleabilidade da heurística proposta e a identificação de características da história da codificação dos vídeos. A alteração do comportamento esperado de um quadro pela comparação com a história observada de codificação é usada como indicativo de corte.

Em especial, a busca pela existência de um padrão majoritário para cada vídeo, em relação ao tamanho dos GOPs e ao tipo de quadro ocupado por posição no interior dos GOPs, permitiu a identificação da assinatura do codificador em relação à correção do erro de propagação entre quadros intercodificados.

Os quadros que fogem ao padrão majoritário detectado foram utilizados como indicativos da existência de um corte em sua vizinhança entre duas tomadas de câmera.

A ponderação das probabilidades de corte dos quadros, de acordo com o canal ocupado, permite a comparação entre os valores da Função de Probabilidade Adaptada de forma a distinguir a influência da assinatura do codificador.

A heurística desenvolvida propõe a detecção das tomadas de câmera pela escolha dos quadros que fogem aos padrões observados na história da compactação e não apresentam indicativos de similaridade relevantes. Para um aprimoramento da heurística proposta, novas métricas de similaridade, bem como, novos refinamentos baseados na história da codificação dos vídeos podem ser agregados às atuais etapas de refinamento.

6.3 Trabalhos Correlatos

O presente trabalho teve uma influência inicial das abordagens de Tobin (1999) e de Chen & Kuo (2002). A análise do comportamento dos algoritmos propostos nesses trabalhos e a identificação de suas restrições fez com que a pesquisa apresentada nessa dissertação extrapola por um caminho diferente.

Até o momento de fechamento da presente dissertação, não é de conhecimento da autora a existência de trabalhos similares. O trabalho mais próximo (Ewerth & Freisleben, 2004) apresenta uma métrica baseada na posição do quadro dentro do GOP e sugere, brevemente, a adição desta métrica em outros algoritmos da literatura. Entretanto, aquele trabalho está longe de se assemelhar à abordagem proposta nesta dissertação. Ewerth & Freisleben (2004) usam uma única métrica, enquanto que o presente trabalho usa várias métricas dentro de uma concepção de refinamento progressivo. A semelhança entre o trabalho de Ewerth & Freisleben (2004) é a análise feita por canal, entretanto usa histogramas que requerem semi-descompressão, enquanto que o presente trabalho faz uma análise por canal através da Função de Probabilidade Adaptada totalmente dentro do domínio compactado. Na presente dissertação, histogramas são usados para permitir uma etapa de análise hierárquica. De qualquer maneira, o trabalho de Ewerth & Freisleben (2004) não teve qualquer influência na presente pesquisa, visto que os principais resultados desta dissertação já haviam sido alcançados antes da publicação desses autores.

Uma comparação numérica com outros algoritmos e com produtos do mercado não foi possível de ser feita, pois não se teve acesso às implementações divulgadas em trabalhos da literatura.

5.3. Trabalhos Futuros

O algoritmo desenvolvido neste trabalho detecta as transições graduais pelo aparecimento de sucessivos quadros no conjunto solução em um pequeno intervalo de tempo onde ocorre a transição gradual. Entretanto, não foi desenvolvido um mecanismo de classificação automática dos quadros obtidos no conjunto solução, entre os diversos tipos de transição (Drew et. al, 2000; Jun et al. 2000; Kim et al., 2001; Ngo et. al, 1999; Wu et al., 1998). Para essa classificação é necessária a análise de características únicas existentes em cada tipo, de forma a distinguir cortes secos, *dissolves*, *fades*, *wipes*, entre outras.

Essa classificação, bem como, o agrupamento de quadros do conjunto solução de uma mesma transição gradual são sugeridos como trabalhos futuros.

Seguindo a abordagem proposta neste trabalho, a melhoria dos resultados obtidos está relacionada a dois fatores principais: diminuição das taxas de falsos cortes detectados e automação da escolha dos parâmetros de sensibilidade e vizinhança.

Um primeiro fator para diminuição das taxas de falsos cortes é o uso adequado dos parâmetros. Vídeos com tomadas de câmera longas devem ser associados a parâmetros de vizinhança mínima grande e pouca sensibilidade, sendo a recíproca verdadeira.

Outro fator que contribui para melhores resultados dessa taxa é o desenvolvimento de novos critérios de refinamento, que podem ser obtidos tanto pela observação de comportamentos de codificação, como os propostos por esta dissertação, quanto por critérios de análise das imagens.

A obtenção dos quadros-chaves (*keyframes*) é apontado como um dos possíveis trabalho futuros da presente pesquisa. Esses quadros estabelecem a representação do conteúdo de toda a tomada de câmera. Diversos estudos sugerem abordagens para a obtenção desses quadros (Ceccarelli et al., 1997; Chen et al., 1999; Lew et al., 2000; Tam et al., 2000; Zhang, 2003).

A longo prazo, a continuação desta pesquisa aponta para a construção de unidades superiores à tomada de câmera, semanticamente organizados. Esses níveis podem ser obtidos pela construção de unidades de cenas, agrupando as tomadas de câmera com uma identidade espaço-temporal em comum (Aygün,

2003; Vasconcelos & Lippman, 1997; Lee et al., 2000; Zhang, 2003). O aumento da relação semântica entre as unidades detectadas é feito pela análise conjunta entre o canal de vídeo e o canal de áudio, permitindo a caracterização do gênero dos vídeos e a construção do nível da unidade de cena.

Outra direção apontada para a continuação da pesquisa iniciada por esta dissertação é a segmentação do canal de áudio dos vídeos. Esta segmentação permite a extração de informações semânticas tais como:

- a distinção entre música e fala;
- a segmentação do locutor, pela detecção da frequência fundamental da voz;
- a busca automática por palavras chaves, através do processamento da fala; entre outras.

Sinais musicais possuem uma estrutura regular periódica característica denominada ritmo, enquanto sinais de diálogos não possuem estrutura periódica em um período grande de tempo. A detecção de intensidade deste ritmo é usada como fator discriminante para a distinção entre intervalos dos vídeos com trilhas musicais de intervalos com diálogos.

O trabalho mais recente de segmentação de áudio propõe o processamento desse canal diretamente no fluxo de dados codificado, sem decodificá-lo, motivados por razões de eficiência (Jarina et al., 2002^a; Jarina et al., 2002b.).

A segmentação de vídeo em tomadas de câmera é o primeiro passo necessário para estruturar a informação contida em bancos de vídeos. Existem diversas metodologias de extração de informações semânticas direcionadas a temas específicos de vídeos.

As pesquisas em indexação de bancos de vídeos de noticiários apontam para a extração de informações tais como a segmentação e classificação entre: tomadas do apresentador, associadas à identificação do locutor por algoritmos de reconhecimento de faces; tomadas de reportagem e tomadas de intervalo comercial (Hanjalic et al., 1998; Iyengar & Neti, 2000).

As propostas relacionadas à organização de acervos de vídeos esportivos buscam a detecção dos eventos mais importantes na modalidade esportiva analisada tais como o gol e a cesta para o futebol e o basquete respectivamente (Ekin, 2003; Kobla et al., 1999; Marlow et al., 2003; Pan et al., 2002)

Outra proposta interessante encontrada, de alto valor comercial, é a pesquisa de Hoad & Zobel (2003). Neste trabalho, os autores propõem a identificação vídeos similares em grandes coleções de vídeo para tarefas de assegurar direitos autorais e verificação de autoria.