

## 4 Resultados

Este capítulo apresenta os resultados de segmentação usando o algoritmo proposto. Os resultados apresentados neste capítulo foram obtidos pela análise de um acervo de vídeos de origem e gêneros diversos. Esses vídeos foram obtidos em acervos múltiplos, acessíveis através da Internet. O intuito de criar esse acervo a partir de fontes diversificadas é não influenciar a análise dos resultados por comportamentos padrões de um determinado codificador.



a.



b.



c.

d.



e.

Figura 22 – Amostras de vídeos do acervo: a. Vídeo Clipe b. Documentário ecológico c. Partida de Futebol (preto e branco) d. Partida de Futebol (cores) e. Animação

#### 4.1. Critérios e Métricas

A Figura 22 apresenta amostras da coleção de vídeos utilizados na análise de resultados. Entre os gêneros contidos nos vídeos do acervo de referência, encontram-se vídeos de esportes, como futebol, atletismo e dança; documentários culturais e ecológicos; programas de entrevistas; vídeo-clipes; animações e outros. A diversidade dos gêneros utilizados na construção do acervo está associada à busca da independência da análise dos resultados em relação à linguagem visual de produção dos vídeos.

A classificação dos cortes encontrados, para obtenção das taxas de desempenho do algoritmo proposto neste trabalho, é feita usando os critérios de classificação dos cortes por ele detectados, segundo o padrão TREC (Smeaton et al., 2002).

A conferência Text Retrieval Conference (TREC) é um evento direcionado ao estudo de problemas de descrição de informação (TREC, 2005). Em 2001, essa conferência passou a tratar também de problemas relacionados à organização da informação contida em acervos de vídeos digitais. Foram estabelecidos critérios de classificação padronizados, para permitir uma avaliação comparativa dos resultados dos vários grupos de pesquisa em segmentação de vídeo em tomadas de câmera, conforme apresentados a seguir:

- As formas de transição foram divididas em apenas duas categorias, cortes secos e transições graduais (agrupando *dissolves*, *fades*, *wipes*, entre outras).
- As transições graduais compostas por menos de 6 quadros são classificadas como cortes secos.
- A correspondência entre duas tomadas de câmera, a referencial e a obtida, é aceita caso essas tomadas possuam intervalos equivalentes pela extensão de até cinco quadros em suas fronteiras. O intervalo de uma tomada de câmera é definido por sua posição e duração.
- A correspondência entre duas transições graduais é estabelecida se o intervalo de interseção entre elas for, pelo menos, 0.333 da maior e 0.499 da menor.

O desempenho do algoritmo proposto por esta pesquisa é medido por três métricas (Lienhart, 2001): a taxa de acertos, a taxa de cortes não detectados e a taxa de cortes falsos detectados.

A taxa de acertos estabelece a razão entre o número de tomadas encontradas e o número real de tomadas (Equação 9). A taxa de cortes não detectados (considerada taxa de perda), estabelece a razão entre o número de tomadas de câmera não detectadas e o número real de tomadas (Equação 10). A terceira métrica utilizada, taxa de cortes falsos, estabelece a razão entre o número total de falsas tomadas de câmera detectadas e o número real de tomadas. (Equação 11)

Dado o número total de transições, suas posições e duração no vídeo, essas métricas são obtidas por:

$$\text{Taxa de acertos} = \frac{\text{total de cortes corretos detectados}}{\text{total de cortes corretos}} \quad (9)$$

$$\text{Taxa de perda} = \frac{\text{total de cortes não detectados}}{\text{total de cortes corretos}} \quad (10)$$

$$\text{Taxa de cortes falsos} = \frac{\text{total de cortes falsos detectados}}{\text{total de cortes corretos}} \quad (11)$$

## 4.2. Gráficos de Desempenho

Os gráficos apresentados a seguir são resultantes da análise do acervo por variações do parâmetro de vizinhança, considerando o conjunto ordenado {0, 4, 15, 30, 60} e do parâmetro de sensibilidade, considerando o conjunto ordenado {0,015, 0,030, 0,060, 0,100, 0,300}, resultando em 25 possibilidades.

Considerando a taxa de exibição de 30 quadros por segundo, o conjunto ordenado adotado para a variação do parâmetro de vizinhança equivale, respectivamente, à não utilização da comparação da vizinhança, aos intervalos de décimos de segundos, meio segundo, um segundo, ou ao intervalo de dois segundos.

Os gráficos apresentados nas Figuras 23 e 24 representam a variação das médias das taxas de acerto e cortes falsos, fixando o parâmetro de sensibilidades pelos valores acima definidos. É possível observar, em ambos os casos, o comportamento inversamente proporcional entre essas taxas e a vizinhança mínima.

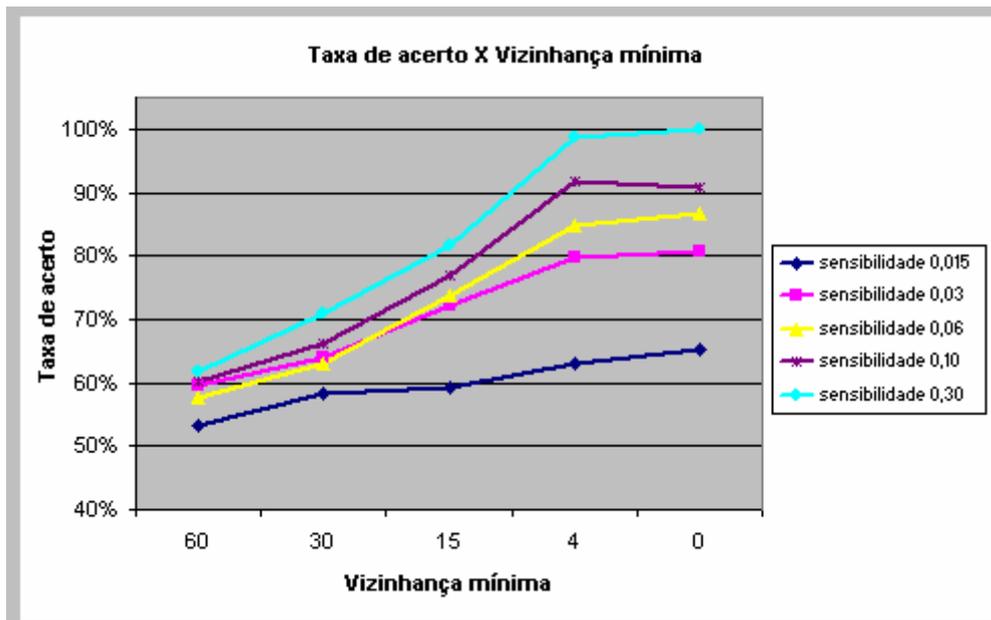


Figura 23 - Taxa de acerto X Vizinhaça mínima

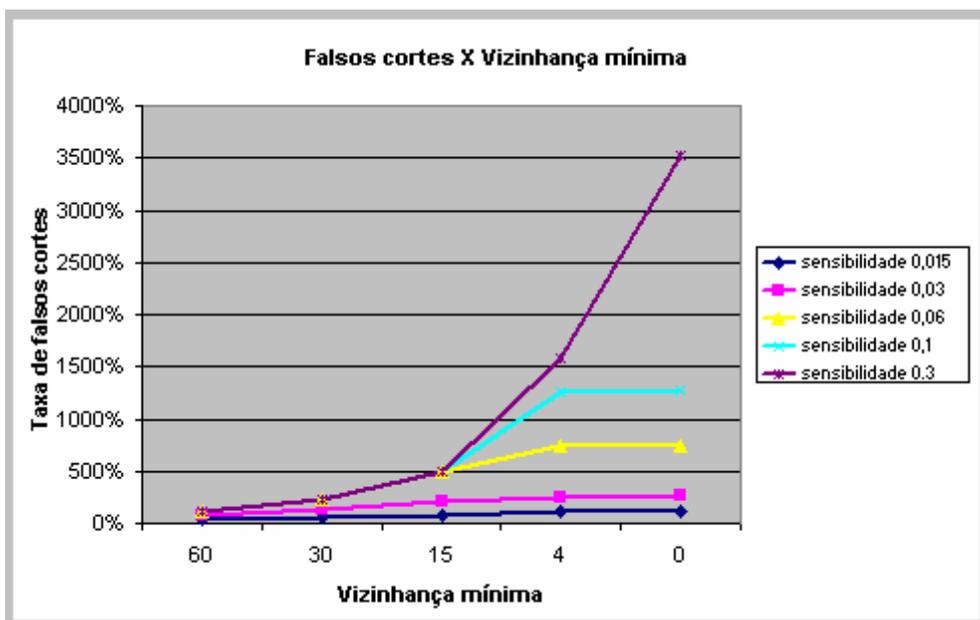


Figura 24 - Taxa de falsos cortes X Vizinhaça mínima.

Os gráficos apresentados nas Figuras 25 e 26 representam a variação das médias das taxas de acerto e cortes falsos, fixando o parâmetro de vizinhaça, utilizando os valores previamente indicados. As curvas encontradas demonstram o comportamento diretamente proporcional entre essas taxas e o parâmetro de sensibilidade.

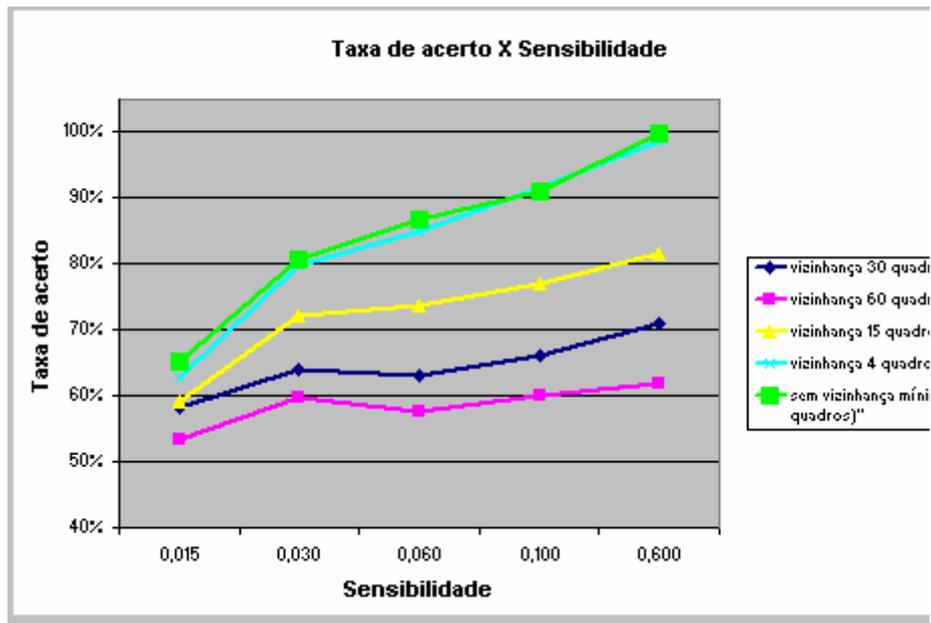


Figura 25 – Taxa de acerto X Sensibilidade

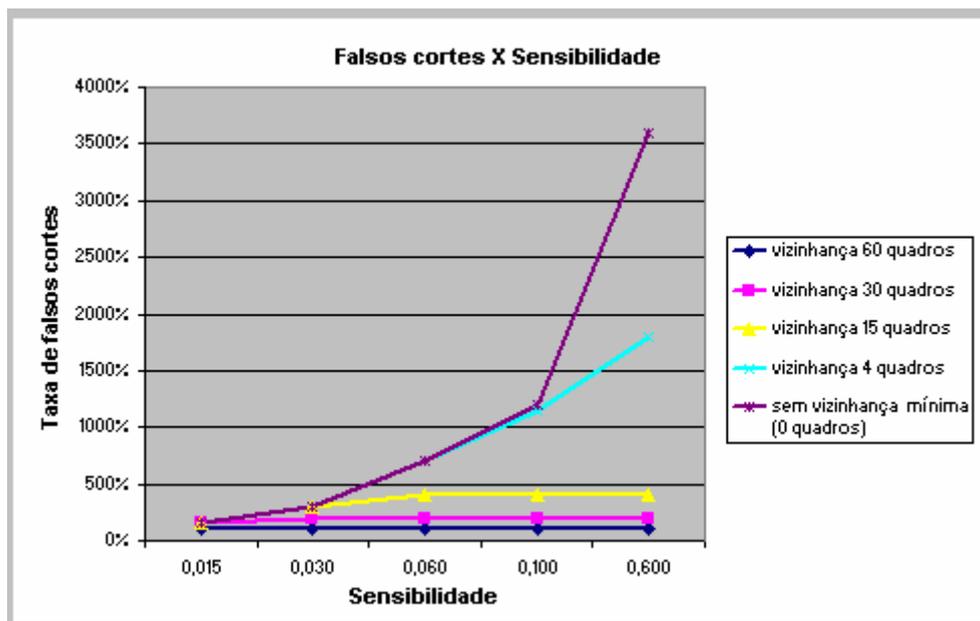


Figura 26 – Taxa de falsos cortes X Sensibilidade

Para a construção do gráfico ilustrado pela Figura 27 foram coletados os piores e os melhores desempenhos do algoritmo, considerando todos os tipos de transição, pelo ajustamento dos parâmetros de vizinhança e sensibilidade, com base em valores dos conjuntos indicados, em relação a cada vídeo do acervo. Apenas para maior clareza dos resultados, as taxas de acerto das transições de cortes secos e graduais são ilustradas separadamente. Cabe ressaltar que o

principal objetivo das heurísticas desenvolvidas nesta pesquisa é a detecção de cortes secos entre duas tomadas de câmera, uma vez que esse tipo de transição representa a grande maioria das transições entre tomadas de câmera dos acervos de vídeo, representando entre 60% e 70% do total (Smeaton et al., 2002).

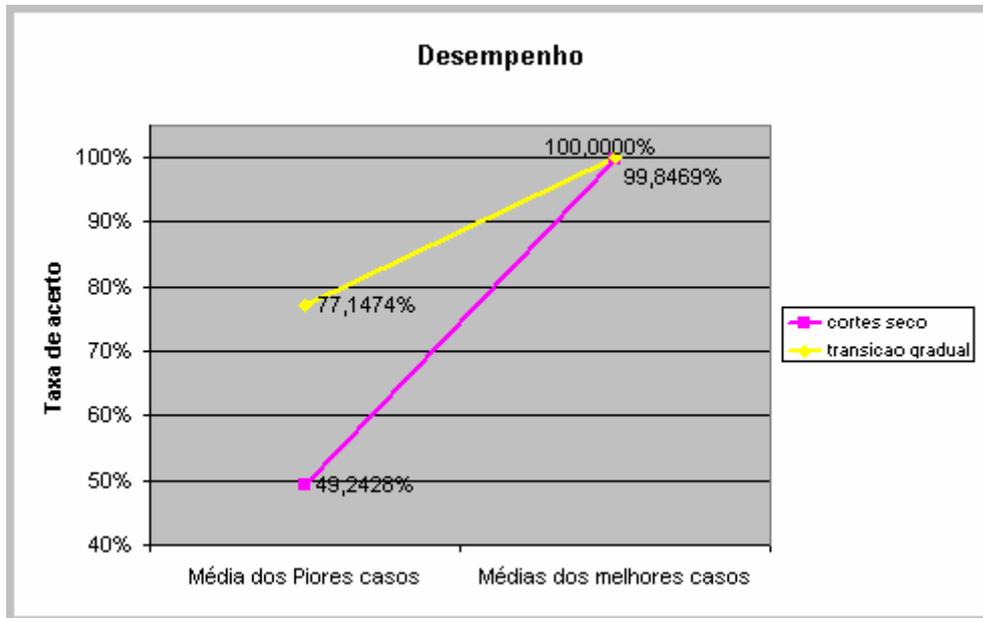


Figura 27 - Desempenho

#### 4.3. Etapas de Refinamento da Detecção

As características dos vídeos alteram os resultados obtidos pelas etapas de refinamento, influenciando diretamente o desempenho da heurística desenvolvida.

A etapa de refinamento por histograma analisa a variação entre as intensidades das imagens nas tomadas de câmera vizinhas. Dessa forma, essa etapa influencia os resultados de acordo com o comportamento geral das imagens do vídeo. Vídeos com pouca variação de intensidade (Figura 22-c) não apresentam diferenças significativas durante a alteração entre as tomadas de câmera, sendo necessário descartar essa análise. A ordenação dos ângulos encontrados pelo histograma permite distinguir os filmes onde não há diferença significativa, daqueles filmes que apresentam diferenças sutis (Figura 22-e).

Em geral, os algoritmos apresentados na literatura indicam como causa para a detecção de falsos cortes os segmentos de vídeos com alterações rápidas de conteúdo no interior de uma mesma tomada de câmera. Essas alterações rápidas

podem ser causadas pelo movimento rápido de objetos, pela própria câmera, ou pela fonte de iluminação, alterações na fonte luminosa, efeito de flash de câmera, filmagens de explosões, entre outras (Pickering, 2002). Esses eventos provocam a diminuição da semelhança entre os quadros vizinhos no interior dessas seqüências provocando os cortes falsos.

A estrutura hierárquica da heurística proposta por esta pesquisa permite eliminar a detecção de um grupo específico desses cortes falsos ocasionados por alterações rápidas de conteúdo. Algumas dessas mudanças rápidas de conteúdo acontecem apenas em alguns quadros que são imediatamente seguidos de quadros semelhantes aos anteriores à mudança, dentro da mesma tomada de câmera. Nesses casos, quando o evento ocorre no interior do intervalo entre dois cortes I, a etapa de histograma analisa a tomada de câmera como semelhante, descartando o grupo de quadros no intervalo onde ocorre a mudança rápida.

As etapas de análise de vizinhança são extremamente dependentes do conteúdo semântico do vídeo, sendo influenciadas pela velocidade da movimentação dos objetos e da câmera e pela duração das tomadas de câmera. O aumento da vizinhança influencia na diminuição das taxas de cortes falsos, mas não devem ser usado em vídeos como os de esportes e comerciais, pois nesses casos, com o crescimento da vizinhança, tendem a perder cortes.

A marcação dos quadros posicionados em tomadas de câmera nos dados codificados é feita durante o processo de codificação do vídeo ou de sua edição. Dessa forma, o desempenho e a relevância da etapa responsável pela identificação desse tipo de quadro dependem da existência desse pré-processamento durante a composição do vídeo.

O desempenho das etapas de refinamento baseados em análises da Função de Probabilidade Adaptada está vinculado a dois fatores principais: a resolução dos quadros e a IPB-ratio dos GOPs.

A análise de vídeos, tomando por base a variação da resolução dos quadros, indica que quanto menor o tamanho dos quadros menos precisa é a classificação obtida pela Função de Probabilidade Adaptada de corte. Esse fenômeno se explica porque a diminuição da resolução implica a diminuição do número de Macroblocks e, conseqüentemente, diminuição da quantidade de referências analisadas pela Função de Probabilidade Adaptada, tornando a detecção do corte menos precisa.

A variação da *IPB-ratio* é observada como um segundo fator de influência aos valores obtidos pela Função de Probabilidade Adaptada. Os casos extremos são definidos pelo aumento ou diminuição do número de quadros B entre dois quadros-âncoras.

As sub-sequências com poucos quadros B apresentam problemas para a análise pela Função de Probabilidade Adaptada pela diminuição do número de referências consideradas. A análise dos quadros pela Função de Probabilidade Adaptada em GOPs de *IPB-ratio* 1:2:3 (ver Figura 28), por exemplo, considera apenas as taxas de referência de um quadro B e um quadro-âncora para aferir a probabilidade dos quadros.

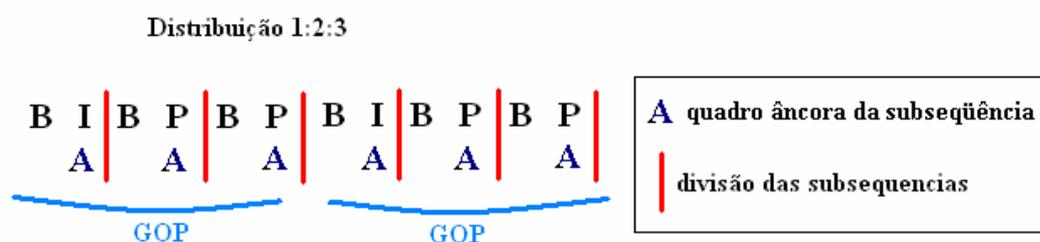


Figura 28 – Distribuição 1:2:3

As equações 12, 13 e 14 representam a equação estendida da Função de Probabilidade Adaptada usada nos possíveis casos de subsequências para *IPB-ratio* 1:2:3 (por simplificação da Equação 4). As versões para a análise dos quadros P e B não são muito confiáveis por levarem em consideração apenas duas taxas de referência que podem ser influenciadas por eventos não diretamente relacionados aos cortes. A versão para a análise dos quadros I é ainda menos precisa, considerando apenas uma taxa de referência. Esses problemas fazem com que a análise direta do valores da probabilidade de corte obtidos no interior de subsequências em distribuições de GOP com poucos quadros B seja altamente suscetível a erros. Este também é um problema contornado pela ponderação por canal sugerida neste trabalho.

$$P(i) = \left(1 - \frac{RR_i^2 + RR_{i+1}^2}{RR_i + RR_{i+1}}\right) \quad \text{se } i \text{ é um quadro B} \quad (12)$$

$$P(i) = \left(1 - \frac{RR_{i-1}^2 + RR_i^2}{RR_{i-1} + RR_i}\right) \quad \text{se } i \text{ é um quadro P} \quad (13)$$

$$P(i) = \left(1 - \frac{RR_{i-1}^2}{RR_{i-1}}\right) \quad \text{se } i \text{ é um quadro } I \quad (14)$$

Onde:

- $i$ : representa a posição do quadro analisado em relação ao interior da máscara.

Por outro lado, o aumento do número de quadros B entre os quadros âncoras proporciona um aumento da continuidade da distribuição da Função de Probabilidade Adaptada. Essa continuidade fornece uma maior precisão para a comparação entre as probabilidades vizinhas. A Figura 29 ilustra essa observação usando a distribuição de probabilidades de um filme com subsequências de tamanho 8, ou seja, sete quadros B a cada quadro âncora.

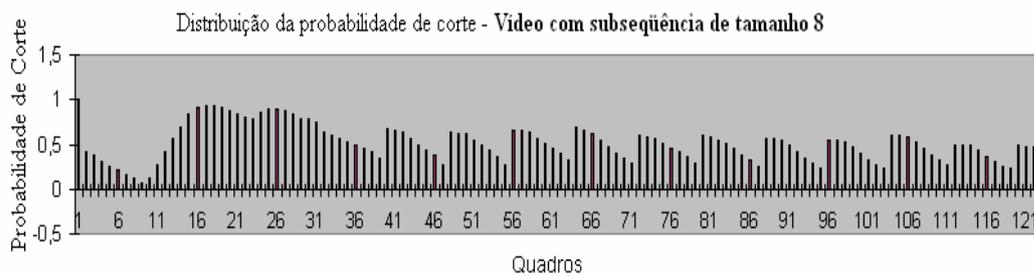


Figura 29 – Distribuição da Função de Probabilidade Adaptada – Vídeo com subsequência de tamanho 8