

2 Segmentação de vídeo

A segmentação do vídeo em unidades de tomada de câmera fornece uma maneira eficiente de acesso ao seu conteúdo. Uma tomada de câmera é composta por uma seqüência de gravação contínua no tempo e no espaço. Desta forma, o conteúdo das diversas imagens interiores a uma tomada de câmera é similar. O processo de segmentação do vídeo em tomadas de câmera é caracterizado pela busca das fronteiras entre duas tomadas de câmera consecutivas, analisando a similaridade entre os quadros. Os diversos tipos de algoritmos para detecção dessas fronteiras possuem diferentes abordagens para caracterizar tal similaridade.

Este capítulo faz um levantamento de diversas abordagens para o problema de segmentação de vídeo em tomadas de câmera. A Seção 2.1 apresenta as diversas abordagens em alto nível, enquanto que a Seção 2.2 apresenta, com maior aprofundamento, alguns trabalhos selecionados para ilustrar os conceitos adotados na solução apontada por esta pesquisa.

2.1. Tipos de transição

A transição mais simples existente entre duas tomadas de câmera é chamada de corte seco. O corte seco representa a transição abrupta entre as duas seqüências, sem nenhum efeito de edição. Nessa transição, o último quadro da primeira tomada de câmera é seguido pelo primeiro quadro da segunda tomada.

Transições mais sofisticadas são chamadas de transições graduais e são resultantes de combinações cromáticas ou espaciais de quadros das duas tomadas. As principais transições graduais são *dissolve*, *wipe*, *fade-in* e *fade-out* (Figura 2).

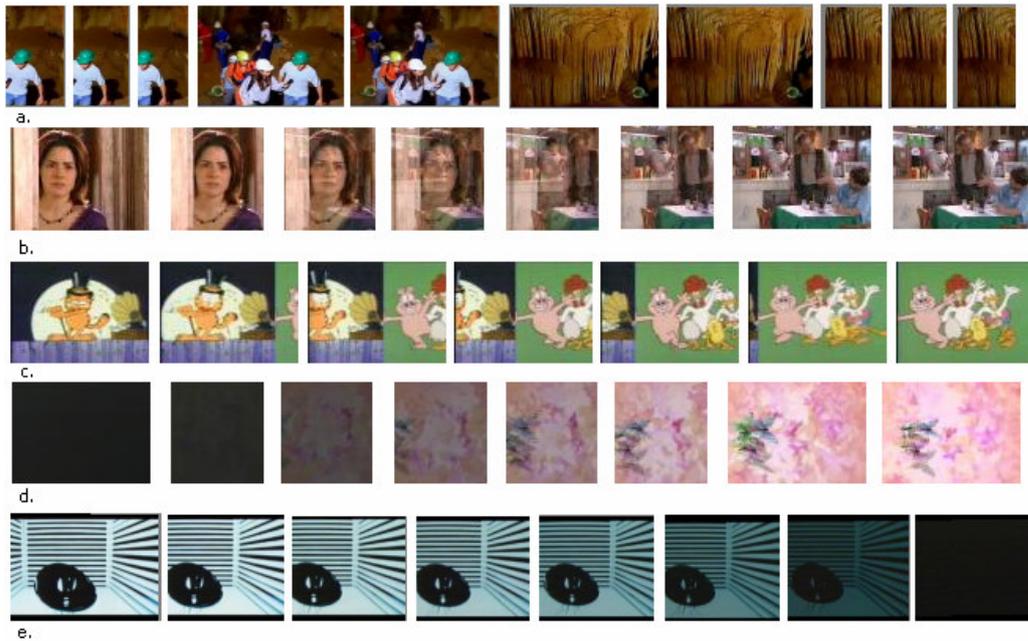


Figura 2– Tipos de Transição: a.Corte Seco; b.*Dissolve*; c.*Wipe*; d.*Fade-in*; e.*Fade-out*.

Os tipos de transição entre duas tomadas de câmera podem ser classificados de acordo com as operações 2D realizadas sobre as imagens que compõe a sua produção (Lienhart, 2001). As transições que não alteram o conteúdo das duas tomadas de câmera, como as produzidas por cortes secos, pertencem à Classe Identidade. Um segundo grupo de transições, denominado Classe Espacial, é caracterizado pela aplicação de operações espaciais 2D nas duas tomadas de câmera envolvidas. As transformações mais comuns representantes dessa classe são o *wipe*, *fold*, *slide* e o efeito íris. As transições que operam sobre o esquema de cores das duas tomadas de câmera envolvidas pertencem à Classe Cromática. Nesta classificação se enquadram os *dissolves* e *fades*. A existência simultânea de operações espaciais e cromáticas nas imagens das tomadas de câmera caracteriza a classe mais complexa de ser detectada, denominada Classe Espaço-cromática. Todos os efeitos de *morphing* se enquadram nesta categoria. As seqüências com transformações da Classe Espaço-cromáticas não apresentam unicidade de solução para o problema da segmentação em tomadas de câmera, sendo suas fronteiras altamente ambíguas e relacionadas a critérios semânticos. Por esse motivo, não são encontrados na literatura, algoritmos para o tratamento da detecção de transições pertencente a essa classe.

A pesquisa realizada nesta dissertação está focalizada na detecção de transições pertencentes à Classe Identidade. A detecção de cortes graduais

pertencentes à Classe Espacial ou à Classe Cromática está vinculada à ampliação do algoritmo desenvolvido para incorporar heurísticas de classificação dos quadros-cortes entre os tipos de transição gradual.

2.2.

Abordagens ao problema de detecção de tomada de câmera

As pesquisas sobre algoritmos para detecção dos cortes de câmera podem ser agrupadas de acordo com as características usadas para medir a similaridade entre os quadros do vídeo.

As abordagens mais intuitivas para resolução desse problema, adotadas pelas primeiras iniciativas de pesquisa, utilizam métodos comparativos da intensidade dos *pixels* entre dois quadros consecutivos (Zhang et al., 1993). A diferença total encontrada entre cada par de imagens é comparada a um limiar para classificação da existência do corte. Essa abordagem recebe o nome de *Pair-Wise Comparison*.

Além do alto custo da descompressão e da operação sobre grande volume de dados, esse grupo de algoritmos apresenta como desvantagem uma extrema sensibilidade a movimentações de câmera, a ruídos e a qualquer tipo de movimentações bruscas por grupos de *pixels* na cena. O caso típico de alarme falso acontece pela movimentação de grandes objetos.

Buscando superar o problema da localidade na comparação feita por *pixels*, um segundo grupo de algoritmos, denominados *Likelihood Ratio Comparison*, propõe a definição de uma medida de similaridade entre duas imagens pela comparação entre regiões da imagem, ao invés da comparação direta entre *pixels* (Kasturi & Jain, 1991).

Os algoritmos de *Pair-Wise Comparison* e *Likelihood Ratio Comparison* não detectam transições graduais. A comparação apenas entre dois quadros consecutivos não apresenta características suficientes para que essas transições sejam detectadas.

Normalmente, dois quadros consecutivos relativos à mesma tomada de câmera apresentam pequenas variações sobre um mesmo cenário e mesmos objetos, de forma a manter a distribuição cromática das imagens. A partir dessa observação, surge uma terceira abordagem para a definição da similaridade entre

os quadros, chamada *Histogram Comparision*. Nesses algoritmos a classificação do corte é obtida comparando a um limiar, a diferença entre os histogramas de intensidade ou de cromaticidade de duas imagens consecutivas. A melhoria obtida por essa classe de algoritmos é associada à insensibilidade das medidas fornecidas pelos histogramas a movimentos no interior das imagens (Ceccarelli et al., 1997).

A etapa de maior complexidade na implementação dos algoritmos baseados em histogramas é a escolha do limiar de comparação (Ceccarelli et al., 1997). Sua dificuldade está relacionada ao problema de configurar um limiar ótimo ou encontrar métodos que permitam a adaptação desse limiar, de acordo com o contexto da seqüência analisada.

Essa classe de algoritmos é amplamente usada por diversos sistemas (O'Toole, 1998; Pickering, 2002; Pye et al., 1998; Meng et al., 1995). Uma das adaptações, propostas à sua versão original, procura superar alterações cromáticas locais, utilizando o cálculo de histogramas, a partir das médias de blocos das imagens (O'Toole, 1998). Outra adaptação, bastante aceita, procura adaptar esses algoritmos para a detecção de transições graduais usando dois limiares distintos (Zhang, 1999). O segundo limiar é usado para uma comparação acumulativa entre diversos quadros consecutivos.

Os falsos cortes detectados, resistentes às três abordagens apresentadas acima, têm como causa principal a ocorrência de mudanças bruscas nos conteúdos dos quadros, causadas por outros motivos, além da mudança da tomada de câmera (Pickering, 2002). Essas grandes alterações no conteúdo dos quadros podem ser causadas por movimentações rápidas de grandes objetos, assim como o seu aparecimento ou desaparecimento repentino, por mudanças de iluminação ou por movimentações bruscas de câmera (Frag & A-Wahab, 2001). Os cortes perdidos por algoritmos com comparação de histogramas podem ser causados por tomadas de câmera com distribuições cromáticas parecidas.

A linha de pesquisa pertencente à Classe Detecção de Arestas (*Edge Detection*) analisa o comportamento das arestas dos objetos através das seqüências de imagens do vídeo, na busca por características geométricas do problema de transição de tomada de câmera. Para isso, analisa o surgimento e o desaparecimento das arestas à procura das transições entre tomadas de câmera (Totterdell, 1998; Kim et al., 2000).

A primeira etapa do algoritmo é constituída pela descompressão do vídeo e geração de imagens contendo apenas o canal de luminância. A aplicação do filtro Sobel (ver Totterdell, 1998; Kim et al., 2000), nessas imagens de luminância, permite a identificação das arestas das imagens. Formado o conjunto de arestas, o algoritmo parte para o processo de acompanhamento da evolução desse conjunto. Uma aresta é considerada nova se surgir a uma distância maior do que um limiar de todas as arestas da imagem anterior. Caso contrário, a aresta é considerada como evolução de uma aresta antiga em movimentação, presente, portanto, na imagem anterior. A partir dessa classificação, são contadas as arestas e calculadas as frações de arestas novas e antigas. O aparecimento de picos nos valores dessas frações é considerado indicativo de transição entre tomadas de câmera.

A fração de novas arestas apresenta picos durante transições de corte seco, *fade-in* e nos quadros finais de um dissolve. A fração de arestas desaparecendo apresenta valores altos em cortes secos, *fade-out* e nos quadros iniciais de *dissolves*.

Os algoritmos da Classe Detecção de Arestas (*Edge Detection*) são muito eficazes, obtendo índices altos de acerto dos cortes. Os falsos cortes detectados ocorrem quando o algoritmo não consegue acompanhar a movimentação das arestas provocadas por mudanças rápidas de câmera ou de intensidade global dos quadros. A desvantagem mais relevante desse grupo de algoritmos, entretanto, é o custo do processamento da descompressão e das operações geométricas envolvidas na extração das arestas e no acompanhamento comparativo para verificação da evolução das arestas.

Considerando como prioridade a diminuição do tempo de processamento e do volume de dados manipulados para obtenção das tomadas de câmera, foi desenvolvido um novo paradigma em relação à natureza dos dados processados. Com esse objetivo, as abordagens dos algoritmos classificados como algoritmos de detecção em Domínio Comprimido procuram características de similaridade entre os quadros, a partir de análises dos dados comprimidos do vídeo MPEG (Kobla et al., 1997; Calic & Izquierdo, 2001; Wills, 2002).

As informações mais utilizadas pelos algoritmos do domínio comprimido são obtidas a partir dos seguintes dados existentes na *stream* MPEG: coeficientes da Transformada Discreta do Cosseno (DCT); vetores de movimento; tipos de

Macroblocks¹; e, taxa de compressão (*bitrate*) das imagens (Koprinska & Carrato, 2001).

Os coeficientes DC (coeficientes de índice zero da Transformada de Cosseno Discreta) podem ser usados para a formação de uma imagem equivalente à imagem original em resolução menor, chamada Imagem DC, possibilitando a aplicação de algoritmos similares aos que operam no domínio descomprimido (Figura 3).

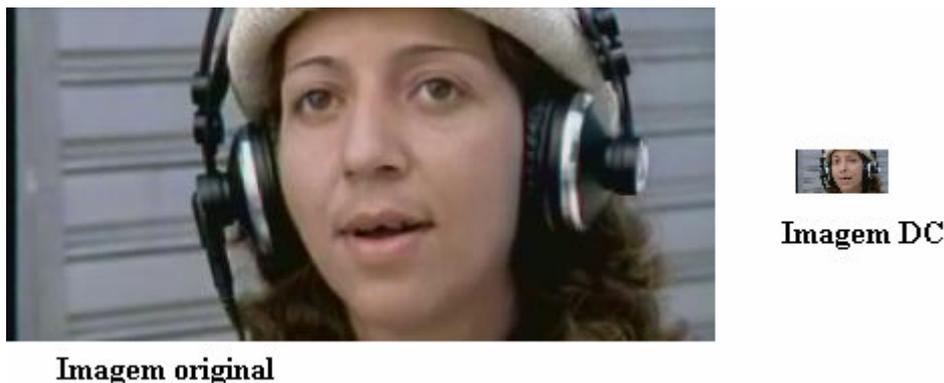


Figura 3 – Imagem DC

Há uma distinção clara sobre a natureza dos dados tratados pelos algoritmos que trabalham no domínio descompactado e dos dados tratados pelos algoritmos que trabalham no domínio compactado (Pickering, 2002; Tobin, 1999; Boccignone et al., 2000b; Calic et al., 2002). No primeiro caso, são feitas análises a partir da informação contida nos *pixels* das imagens seqüenciais formadoras do vídeo. Por outro lado, no domínio compactado os dados representam as freqüências da imagem obtidas pela DCT, segundo o modelo de compactação MPEG-1 e MPEG-2 (Sun, 2002).

Os algoritmos para identificação das tomadas de câmera em domínio descomprimido possui as vantagens de não estar diretamente vinculada ao padrão MPEG-1/2, ou a outro padrão de vídeo qualquer (Pickering, 2002) e permitir a compreensão mais intuitiva da natureza e das operações sobre os dados processados das imagens (Wee, 1999). Entretanto, esses algoritmos exigem que, inicialmente, seja feita a decodificação completa dos vídeos, obtida pelas operações de decodificação de Código de Huffman, inversão da quantização

¹ O termo Macroblock e outros que são usados pelo padrão MPEG podem ser consultados no Apêndice II

MPEG, aplicação da Inversa da Transformada de Cosseno Discreta (IDCT) e da compensação de movimento (Ceccarelli et al., 1997), uma vez que os vídeos são armazenados e transmitidos em versões comprimidas.

Os algoritmos com manipulação de dados no domínio compactado, apesar de sua maior complexidade, possuem as vantagens de processar um menor volume de dados e eliminar o tempo gasto durante o processamento da descompactação (Wee, 2002), apresentando alto grau de eficiência.

Existe uma linha de pesquisa que considera inexistente a distinção entre algoritmos no domínio comprimido e descomprimido. Esses trabalhos consideram que os fundamentos dos algoritmos podem ser reformulados para a aplicação em ambos os domínios (Lienhart, 2001). Consciente da possibilidade de reformulação dos algoritmos, com uma ressalva em relação aos algoritmos baseados em dados específicos da compressão MPEG, esta dissertação faz uso da classificação entre o domínio comprimido e descomprimido, apenas como forma de ilustrar a natureza dos dados manipulados pelos algoritmos propostos em sua versão original. As diversas abordagens são consideradas enriquecedoras para uma melhor compreensão do problema.

2.3. Domínio Comprimido

Algoritmos eficientes de processamento em domínio comprimido são projetados para levar em consideração características próprias da compressão MPEG. Nesta seção são descritos cinco algoritmos da Classe Domínio Comprimido, selecionados para introduzir os conceitos utilizados por esta pesquisa. Entre esses conceitos destacam-se:

- a obtenção de imagens DC a partir dos dados no domínio comprimido;
- a influência das tomadas de câmera na distribuição dos tipos de Macroblocks;
- as taxas de referência entre as imagens;
- o uso de passos hierárquicos; e,
- a necessidade de adaptação dos algoritmos aos comportamentos distintos apresentados pelos diversos gêneros de filmes.

A proposta desenvolvida por O'Toole (1998) estabelece a construção de histogramas cromáticos a partir de imagens extraídas da descompressão parcial do vídeo MPEG. Essa proposta aborda os principais conceitos utilizados nas diversas variações dos algoritmos da classe *Histogram Comparison* e apresenta uma adaptação para a extração das informações do histograma, sem processar as funções de descompressão.

Inicialmente, a etapa de extração de imagens JPEG a partir do vídeo MPEG requer a desvinculação das interdependências entre os quadros do vídeo, estabelecidas durante o processo de compressão. Para otimizar a extração dessas imagens, são utilizadas apenas as informações dos coeficientes DC, diretamente dos dados codificados. Esses coeficientes podem ser descomprimidos sem a utilização da Transformada Inversa de Cosseno Discreta (IDCT), por serem proporcionais à média da intensidade da região ocupada pelo bloco. Esse tipo de otimização gera uma imagem 64 vezes menor que a resolução original do vídeo. Para cada grupo de 64 *pixels* extrai-se apenas um, equivalente à média dos 64, armazenada no coeficiente DC (Figura 3). Diversos algoritmos utilizam a imagem extraída por esse processo como uma versão ícone da imagem original, sendo chamada de Imagem DC (Meng et al., 1995; Ceccarelli et al., 1997; Kim et al. 2000; Farag & A-Wahab, 2001).

O passo seguinte à extração das imagens ícones é a transformação do espaço cromático YUV da informação contida no vídeo MPEG, para o espaço HVC, usado pelas funções de histograma, tendo o espaço RGB como intermediário para cálculo dessa operação. Traduzidos os dados para o espaço HVC, a detecção dos cortes entre tomadas de câmera é obtida pela comparação dos histogramas de imagens consecutivas, através de uma métrica comparativa dos histogramas.

A métrica utilizada para graduar a semelhança entre imagens consecutivas é o cálculo do ângulo entre os histogramas das duas imagens. A métrica denominada Medida de Similaridade de Cosseno (*Cosine Similarity Metric* - CSM) fornece o cosseno desse ângulo. A definição adotada por O'Toole (1998) associa o valor 1 como indicativo da inexistência de similaridade e o valor 0 como máximo de similaridade (Equação 1).

$$\cos(H1, H2) = (1 - \frac{H1 \bullet H2}{\|H1\| \|H2\|}) \quad (1)$$

Onde:

- □ **H1** e **H2** representam os vetores dos valores obtidos pelos histogramas de dois quadros consecutivos.

Existem alguns trabalhos que definem CSM de forma a associar o valor 0 como máximo de similaridade e o valor 1 como indicativo da inexistência de similaridade (usando a Equação 1 apenas com o segundo termo e com sinal invertido). Nesse caso, a aplicação da função de arco cosseno no valor obtido pela CSM fornece o ângulo entre os dois vetores. A classificação dos quadros como cortes de tomadas de câmera é feita pela comparação dos ângulos encontrados a um limiar.

O segundo trabalho selecionado para ser apresentado nesta dissertação parte de um acervo de 8 horas de vídeos, capturados no padrão MPEG-1, usando uma aplicação de decodificação parcial e a informação dos cortes manualmente identificados (ou seja, faz uma análise a partir da informação de cortes reais previamente identificados). A pesquisa de Brian Tobin (Tobin, 1999) observa, pela análise da listagem de contadores, o aparecimento de padrões diferenciados na distribuição dos tipos de Macroblocks em regiões na vizinhança dos cortes de câmera.

Os padrões identificados variam de acordo com três fatores:

- tipo dos quadros onde ocorre o corte entre tomadas de câmera;
- tipos dos quadros em sua vizinhança; e
- a posição relativa entre os quadros.

Para a comparação e a identificação desses padrões são considerados os GOPs de tamanho fixo (10 quadros) e uma ordenação dos tipos de quadros permanentes durante todo o vídeo (dois quadros B a cada quadro âncora).

A seguir, estão relacionados os casos possíveis para análise do candidato ao corte, na forma de tabelas (Tabela 1, Tabela 2 e Tabela 3). Para cada um dos casos, está descrito o padrão característico esperado para a classificação do quadro como corte. A Figura 4 auxilia o entendimento dos processos envolvidos.

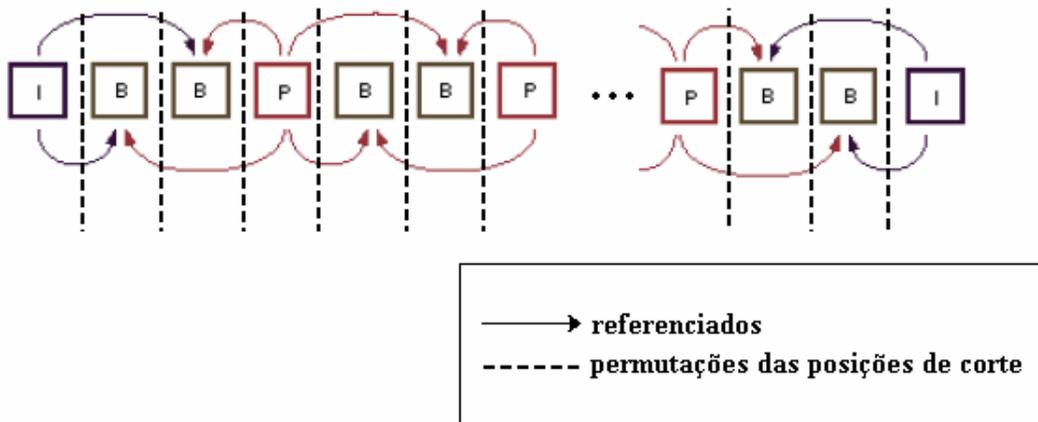


Figura 4 – Possibilidade de posicionamento dos cortes.

Tabela 1 – Análise das características esperadas para cortes em quadros B

Subseqüência	Características esperadas dos Macroblocks nos quadros vizinhos ao corte
Posição Corte	
I BBP Corte no 1º B	<ul style="list-style-type: none"> - O quadro B imediato ao corte possui poucos ou nenhum Macroblock P; - Segundo quadro B: mesmo comportamento; - O quadro P no fim da subseqüência apresenta mais Macroblocks intracodificados do que a média esperada.
IB BP Corte no 2º B	<ul style="list-style-type: none"> - O quadro B anterior ao corte possui poucos ou nenhum Macroblock B e muitos Macroblocks P; - O quadro B imediato ao corte possui poucos ou nenhum Macroblock P e muitos B; - O quadro P no fim da subseqüência apresenta mais Macroblocks intracodificados do que a média esperada.
P BBP Corte no 1º B	<ul style="list-style-type: none"> - O quadro B imediato ao corte possui poucos ou nenhum Macroblock P; - Segundo quadro B - mesmo comportamento; - O quadro P no fim da subseqüência apresenta mais Macroblocks intracodificados do que a média esperada.

PBIBP Corte no 2º B	<ul style="list-style-type: none"> - O quadro B anterior ao corte possui poucos ou nenhum Macroblock B e muitos Macroblocks P; - O quadro B imediato ao corte possui poucos ou nenhum Macroblock P e muitos B; - O quadro P no fim da subsequência apresenta mais Macroblocks intracodificados do que a média esperada.
PBBI Corte no 1º B	<ul style="list-style-type: none"> - O quadro B imediato ao corte possui poucos ou nenhum Macroblock P e muitos B; - Segundo quadro B - mesmo comportamento.
PBIBI Corte no 2º B	<ul style="list-style-type: none"> - O quadro B anterior ao corte possui poucos ou nenhum Macroblock B e muitos Macroblocks P; - O quadro B imediato ao corte possui poucos ou nenhum Macroblock P e muitos B.

Tabela 2 – Análise das características esperadas para cortes em quadros P

Subseqüência	Características esperadas dos Macroblocks nos quadros vizinhos ao corte
Posição Corte	
IBBPBBPBB Corte no 1º P	<ul style="list-style-type: none"> - Os quadros B anteriores ao corte possuem poucos ou nenhum Macroblock com referência B; - O quadro P onde ocorreu o corte possui um aumento de Macroblocks intracodificados e poucos Macroblocks P; - Os quadros B posteriores ao corte não têm seu comportamento alterado.
IBBPBBPBB Corte no 2º P	<ul style="list-style-type: none"> - São esperados os mesmos comportamentos do caso anterior para os Macroblocks dos quadros na vizinhança de quadros P com esse posicionamento e também, nos demais casos possíveis.

Tabela 3 – Análise das características esperadas para cortes em quadros I

Subseqüência	Características esperadas dos Macroblocks nos quadros vizinhos ao corte
Posição Corte	
BB IBBP Corte em I	<ul style="list-style-type: none"> - Os quadros B anteriores ao corte possuem poucos ou nenhum Macroblock com referência B; - Os quadros B posteriores ao corte não têm seu comportamento alterado.

Considerando a existência de apenas um quadro I no interior do GOP, espera-se que um corte em um quadro I implique diminuição das referências entre os quadros B anteriores ao quadro I, por pertencerem a duas tomadas de câmera distintas (Tabela 3).

Os primeiros passos da abordagem adotada pela presente dissertação seguem a direção apontada por Chen & Kuo (2002) no algoritmo desenvolvido no trabalho *A Mask Matching Approach for Video Segmentation*. Esse algoritmo é extremamente rápido, por realizar a análise dos dados no domínio comprimido, sem nenhum tipo de decodificação.

Essa abordagem parte da observação de uma forte relação entre a similaridade dos quadros no interior de uma tomada de câmera e a escolha da codificação dos tipos de Macroblocks, durante o processo de compressão MPEG, em concordância com o trabalho de Tobin (1999), apresentado anteriormente.

A caracterização da existência de corte entre tomadas de câmera se inicia pelo cálculo das taxas de referência (*Reference Ratio* - RR) entre os quadros. De maneira geral, quanto maior a referência entre dois quadros, menor a probabilidade de existir um corte de tomada de câmera entre eles.

São definidas duas taxas de referência *Forward Reference Ratio* e *Backward Reference Ratio*, representando as taxas de Macroblocks com predição para frente (*Forward*) e para trás (*Backward*), respectivamente. Essas taxas representam a fração do número total de Macroblocks em cada tipo de referência (Equação 2 e Equação 3).

$$FRR = \frac{R_f}{N} \text{ onde } 0 \leq FRR \leq 1 \quad (2)$$

Onde:

- **Rf** é o número total de Macroblocks com predição para frente (*forward*) no quadro.
- **N** é o número total de Macroblocks no quadro.

$$BRR = \frac{Rb}{N} \quad \text{onde } 0 \leq BRR \leq 1 \quad (3)$$

Onde:

- **Rb** é o número total de Macroblocks com predição para trás (*backward*) no quadro.
- **N** é o número total de Macroblocks no quadro.

Como os quadros I não fazem referência a nenhum outro quadro, possuindo apenas Macroblocks intracodificados, esse tipo de quadro não apresenta as taxas de FRR e de BRR. Assim, por serem independentes, não fornecem nenhum indicativo direto do comportamento das imagens em sua vizinhança. Da mesma forma, não faz sentido computar a BRR para quadros P, por não fazerem predição *backward*. Os quadros P não fornecem indicativos do comportamento das imagens posteriores a eles.

Valores altos de FRR indicam similaridade do quadro com o quadro âncora anterior, enquanto que valores altos de BRR apontam similaridade ao quadro âncora seguinte. Por essa razão, é esperado que quadros P ou B, similares ao quadro âncora anterior, possuam FRR alta. Os indicativos de similaridade ao quadro âncora posterior são obtidos pelas taxas altas de BRR nos quadros B. Desta forma, os comportamentos padrões esperados pelos cortes em quadros dos diferentes tipos seguem a análise de Tobin.

Em uma segunda etapa, cada quadro é analisado por uma máscara formada por um operador lógico, com o objetivo de definir quais vizinhos participam da análise do quadro candidato e, também, as suas contribuições para a análise, ou seja, quais tipos de taxa de referência serão considerados pela máscara. Máscaras distintas são usadas para cada tipo de quadro e para cada posição do candidato analisado em relação a uma subsequência. As máscaras de Chen & Kuo (2002) automatizam a verificação das características discutidas na análise de Tobin (1999).

Para a criação das máscaras, Chen & Kuo (2002) partem de algumas premissas que restringem a aplicação desse algoritmo a vídeos genéricos. Essas premissas assumem o seguinte:

- o tamanho fixo dos GOPs, em número de quadros;
- o uso de apenas um quadro I a cada GOP; e
- uma distribuição padrão entre os tipos de quadros nos GOPs.

A distribuição dos tipos de quadros no interior de um GOP é chamada de *IPB-ratio* e define os intervalos preenchidos por quadros B, a cada quadro âncora.

Essas restrições possibilitam a definição de um grupo limitado de máscaras para análise de todos os tipos de quadros e seus possíveis posicionamentos em relação a subsequências de tamanho fixo. O resultado da máscara seleciona as taxas RR de interesse para o cálculo da função de probabilidade de mudança de cena.

Considerando um quadro I como candidato ao corte, é preciso analisar as referências entre os quadros B anteriores ao candidato e posteriores ao último quadro âncora. Para que o candidato seja um corte, os quadros B no intervalo considerado devem possuir baixos valores de BRR. O comportamento da predição para frente (*forward*) dos quadros B no intervalo considerado é irrelevante para a análise deste candidato. Também são irrelevantes as taxas dos quadros posteriores ao candidato. As máscaras de quadros I selecionam para a análise as taxas BRR dos quadros B interiores a essa vizinhança de interesse.

Considerando como candidato ao corte um quadro P, os quadros B entre esse candidato e o quadro âncora anterior são analisados. A diferença da análise entre um candidato I e um candidato P é a possibilidade de existirem referências para frente (*forward*) no candidato P. Para existir um corte, o candidato não deve fazer referência a quadros anteriores ao corte, sendo esperadas taxas baixas de referência tipo *forward* a quadros na tomada de câmera anterior e, portanto, baixa taxa FRR no quadro candidato. As máscaras de quadros P selecionam para análise as taxas BRR dos quadros B interiores a essa vizinhança de interesse e a taxa FRR do candidato.

A análise de um quadro B como candidato ao corte é feita no interior da subsequência entre os dois quadros âncoras mais próximos ao quadro B. Os quadros B anteriores ao candidato na subsequência devem apresentar baixas BRR,

enquanto o candidato e os quadros até o quadro âncora seguinte devem apresentar baixas FRR.

As máscaras de quadros B, dados os quadros interiores à subsequência, selecionam para análise as taxas BRR dos quadros B anteriores ao candidato e as taxas FRR do candidato e dos quadros posteriores a ele. Caso o quadro âncora posterior ao corte seja um quadro I, sua participação é indiferente, enquanto, no caso de ser um quadro P, a máscara considera sua FRR.

As características esperadas são transferidas a uma função de probabilidade de corte que associa probabilidades baixas a quadros candidatos que não representam corte.

A Função de Probabilidade de Corte é definida por:

$$P(i) = \left(1 - \frac{RR_{M_1}^2 + RR_{M_2}^2 + \dots + RR_{M_n}^2}{RR_{M_1} + RR_{M_2} + \dots + RR_{M_n}}\right) \quad \text{onde } 0 \leq i \leq n \quad (4)$$

Onde:

- RR_{M_i} : representa a taxa de referência na direção de interesse.
- M : representa a máscara.
- n : é o tamanho da máscara
- i : representa a posição do quadro analisado em relação ao interior da máscara.

O segundo termo da Equação 4 se comporta como uma média ponderada das taxas de referência, associando pesos maiores às taxas mais altas, de forma a influenciar o resultado para destacar sua existência. Ainda que apenas um quadro da subsequência possua taxa alta da referência considerada, o valor total é altamente influenciado por essa taxa. Assim, a probabilidade resultante sofrerá maior influência das taxas maiores de referência, tendendo a sobressair qualquer indicativo de similaridade no interior da seqüência. A probabilidade total será baixa, na existência de pelo menos uma taxa de referência de interesse alta. Para encontrar os cortes, o resultado da função de probabilidade é comparado a um limiar.

O trabalho seguinte, desenvolvido por Boccignone et al. (2000 a), realiza a tarefa de detectar o corte de cena diretamente no fluxo de dados (*stream*) codificado do MPEG, sem nenhum processo de decodificação. Destaca-se nessa

abordagem a implementação baseada em refinamentos combinados com diferentes critérios para a detecção de corte nos tipos distintos de quadros.

O processo de busca por cortes de forma hierárquica se inicia pela análise dos quadros I. A observação dos critérios de similaridade entre esses quadros define os próximos a serem analisados. Caso as mudanças entre dois quadros I consecutivos sejam consideradas desprezíveis, o próximo GOP passa a ser analisado.

Ao identificar mudanças significativas entre dois quadros I, o processo considera a existência de uma suspeita de corte. Nesse caso, o algoritmo passa a analisar os quadros P no interior do GOP considerado.

Nos casos em que a análise de quadros P consecutivos não aponta características de corte, o algoritmo pode descartar a suspeita de corte, passando para a análise do próximo quadro P. Nos casos contrários, ou seja, detectadas alterações significativas, são feitas análises dos quadros B.

Os critérios utilizados para a suspeita de corte nos variados níveis são distintos e independentes, permitindo a adaptação dessa estrutura de comparação a diversas abordagens de características de similaridades.

O último estudo tomado como referência foi desenvolvido por Farag & A-Wahab (2001). Nessa abordagem é utilizada uma representação abstrata do conteúdo do vídeo como entrada para um módulo de rede neural. A representação abstrata do vídeo é formada pela seqüência de coeficientes DC do canal de luminância.

Os coeficientes DC podem ser obtidos diretamente nas imagens I. Nas imagens P e B, entretanto, os coeficientes DC são reconstituídos pela soma do erro residual codificado nos blocos com a média dos blocos usados como referência. A média dos blocos usados como referência é ponderada pela área da intercessão do bloco codificado com cada bloco referência.

A vantagem apontada por essa abordagem é a maleabilidade natural das redes neurais a situações genéricas. A detecção dos cortes é influenciada pelas múltiplas características dos diversos gêneros de vídeo e pelas possibilidades de movimentação de câmera, luzes e objetos. A desvantagem dessa abordagem está na dificuldade da etapa de treinamento da rede.

Apesar de a abordagem de redes neurais não ser adotada na pesquisa relatada por esta dissertação, o trabalho de Farag & A-Wahab (2001) contribui

por destacar a relevância dos aspectos semânticos do vídeo no comportamento das funções características de similaridades. Desta forma, destacando a importância de adaptações aos múltiplos tipos de vídeos – ação, drama, romance, esportes, documentários, noticiários, entre outros, com características próprias na distribuição das tomadas de câmera.