



**FUNDAÇÃO EDSON QUEIROZ
UNIVERSIDADE DE FORTALEZA**

Elan Oliveira Lima

**Sumarização de vídeos de futebol com baixo
custo computacional utilizando ritmo visual: descritores de cor e
movimento de câmera**

Fortaleza
Dezembro / 2006

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.



Elan Oliveira Lima

**Sumarização de vídeos de futebol com baixo
custo computacional utilizando ritmo visual: descritores de cor e
movimento de câmera**

Dissertação apresentada ao curso de
Mestrado em Informática Aplicada da
Universidade de Fortaleza como parte dos
requisitos necessários para a obtenção do
grau de Mestre em Informática Aplicada.

Orientador: Prof. Francisco Nivando Bezerra, Docteur

Fortaleza

Dezembro / 2006

Elan Oliveira Lima

**Sumarização de vídeos de futebol com baixo
custo computacional utilizando ritmo visual: descritores de cor e
movimento de câmera**

**Área de Concentração: Ciências da Computação
Linha de Pesquisa: Processamento de Imagens**

Data de Aprovação: _____

Banca Examinadora:

**Prof. Francisco Nivando Bezerra, Docteur
(Orientador- UNIFOR)**

**Prof^a. Maria Andréia Formico Rodrigues, Ph.D
(Membro - UNIFOR)**

**Prof. Neucimar Jerônimo Leite, Docteur
(Membro - Universidade de Campinas/UNICAMP)**

LIMA, Elan Oliveira

Sumarização de vídeos de futebol com baixo custo computacional utilizando ritmo visual: descritores de cor e movimento de câmera.

Fortaleza. Universidade de Fortaleza (UNIFOR),

Dissertação de Mestrado, 2006.

83 p.: il. 210 x 297 mm (MIA/UNIFOR, Mestrado em Informática Aplicada)

1. Processamento de imagens
2. Morfologia matemática
3. Análise de vídeo

I. MIA/UNIFOR

II. Título CDU:

Dedico esse trabalho aos meus pais Leal e Elena, minhas irmãs Antoniele e Ellen, e minha Sara, pelo amor que tenho recebido e que me impulsiona a tornar sonhos em realidade.

Agradecimentos

A Deus, porque Dele por Ele e para Ele são todas as Coisas. Pela força que nunca me deixou faltar.

Meus sinceros agradecimentos ao Prof. Francisco Nivando Bezerra, pela orientação e força que tornaram possível a realização desse trabalho.

Aos meus pais Leal e Elena, por terem me ensinado a caminhar desde os meus primeiros passos sempre com muito amor e dedicação.

Agradeço às minhas irmãs Antoniele e Ellen, pelo carinho e incentivo que nunca faltou em momento algum.

Agradeço à minha Sara, pelo carinho, paciência e apoio na conclusão desse trabalho.

Aos demais professores do mestrado, pelas contribuições indiretas.

Ao Luiz Cláudio e Raul, pelos vídeos usados nesse trabalho.

Aos demais colegas aqui não citados, que de alguma forma me ajudaram e incentivaram.

À secretaria do MIA, pela atenção e presteza sempre imediatas.

A todos os meus amigos, que entenderam minha ausência em alguns momentos.

Ao SERPRO, por ter contribuído na minha participação no mestrado e ter me liberado para cumprimento dos compromissos acadêmicos deste projeto.

“Tudo tem seu tempo determinado, e há tempo para todo propósito debaixo do céu.”

Eclesiaste 3:1

Resumo da Dissertação apresentada ao MIA/UNIFOR como parte dos requisitos necessários para a obtenção do grau de Mestre em Informática Aplicada.

Sumarização de vídeo de futebol com baixo custo computacional utilizando ritmo visual: descritores de cor e movimento de câmera

Elan Oliveira Lima

Dezembro / 2006

Orientador: Prof. Francisco Nivando Bezerra, Docteur
Programa: Informática Aplicada

O crescimento da produção de conteúdo multimídia criou a necessidade de novas técnicas para o armazenamento e recuperação eficiente desses conteúdos. A área de recuperação de imagens baseada no conteúdo (CBIR) é um campo de pesquisa que objetiva métodos baseados na visão computacional para indexar e recuperar vídeos e imagens. Características visuais, como cor, forma, textura e movimento, são usadas como descritores computados diretamente de uma imagem ou vídeo em aplicações CBIR. Esses descritores são usados em várias aplicações, tais como detecção de transições, classificação de tomadas, reconhecimento de objetos, dentre outras. O ritmo visual é um exemplo de representação espaço-temporal dos dados de um vídeo provendo informação compacta preservando vários tipos de eventos. Nós exploramos essas propriedades no presente trabalho para propor dois novos descritores de baixo nível para análise de vídeos de futebol computados diretamente do ritmo visual. Os descritores são relacionados à cor dominante e à direção de movimento de câmera. Os novos descritores são aplicados em diferentes tarefas, envolvendo a análise de vídeos de futebol, tais como detecção de transições, classificação de tomadas e estimativas de direção de ataque. Apresentamos também, uma aplicação simples de sumário automático de vídeos de futebol para ilustrar o uso dos novos descritores.

Palavras-Chaves: descritores de baixo nível, cor dominante, estimativa de movimento, análise de vídeos de futebol, morfologia matemática.

Abstract of the Dissertation presented to MIA/UNIFOR as a partial fulfillment of the requirements for the degree of Master in Applied Informatics.

Low cost soccer video summaries based on visual rhythm: descriptors of color and camera motion.

Elan Oliveira Lima

Dezembro / 2006

Adviser: Prof. Francisco Nivando Bezerra, Docteur

Program: Applied Informatics

The growth of multimedia contents production has created the necessity of new techniques for efficient storage and retrieval of these contents. The area of Content Based Image Retrieval is a research field whose goal is to develop Computer Vision based methods for indexing and retrieval of images and videos. Visual features such as color, shape, texture and motion are used as descriptors computed directly from an image or video in CBIR applications. These descriptors are applied in different tasks such as shot transition detection, shot classification and objects recognition. The visual rhythm is a spatio-temporal sampled representation of video data providing compact information while preserving several types of video events. We exploit these properties in the current work to propose two new low level descriptors for the analysis of soccer videos computed directly from the visual rhythm. The descriptors are related to dominant color and camera motion estimation. The new descriptors are applied in different tasks aiming the analysis of soccer videos, such as shot transition detections, shots' classification and attack direction estimation. We also present a simple automated soccer summary application to illustrate the use of the new descriptors.

Keywords: low level video features, dominant color, motion estimation, soccer video analysis, mathematical morphology.

Sumário

1.	Introdução	15
1.1	Motivação	15
1.2	Objetivo.....	17
1.3	Trabalhos Relacionados.....	17
1.4	Estrutura da Dissertação.....	21
2.	Noções Básicas	22
2.1	Morfologia Matemática	22
2.1.1	Elemento Estruturante	23
2.1.2	Operações Morfológicas.....	23
2.1.2.1	Erosão.....	24
2.1.2.2	Dilatação.....	25
2.1.2.3	Abertura.....	26
2.1.2.4	Fechamento.....	27
2.2	Orientação de estruturas direcionais.....	28
2.2.1	Orientação Global.....	28
2.2.2	Orientação Local.....	29
2.3	Ritmo Visual	30
2.4	Modelo HSV de representação de cores.....	32
3.	Descritores de baixo nível para vídeos de futebol.....	34
3.1	Considerações preliminares	34
3.1.1	Cor	34
3.1.2	Movimento de câmera	36
3.2	Detecção da cor dominante	38
3.3	Detecção de Transições e Classificação das Tomadas	42
3.3	Detecção de Movimento de Câmera.....	47

4. Sumarização: experimentos e resultados	54
4.1 Experimentos	54
4.1.1 Sumarização.....	54
4.2 Resultados.....	56
5. Conclusão.....	62
5.1 Resumo e Contribuições.....	62
5.2 Trabalhos Futuros	63
Referências Bibliográficas	67
Apêndice I	71
Apêndice II	79

Lista de figuras

2.1	Exemplo de elementos estruturantes	23
2.2	Exemplo de erosão	25
2.3	Exemplo de dilatação	26
2.4	Comparação entre erosão e dilatação	26
2.5	Exemplo de abertura e fechamento	28
2.6	Segmentos discretos de reta utilizados como elementos estruturantes ..	30
2.7	Geração do ritmo visual	29
2.8	Ritmo visual.....	32
2.9	Modelo HSV	33
3.1	Cenas típicas de um vídeo de futebol.....	35
3.2	Quadros de um dissolve	35
3.3	Direção da câmera X tempo	36
3.4	Ritmo visual de um jogo típico	37
3.5	Histograma da componente H da imagem da figura 3.4	40
3.6	Detecção de regiões com H dentro do intervalo de cor dominante	41
3.7	Ritmo visual do 5000 quadros iniciais de um vídeo	43
3.8	Quantidade de pixels na cor dominante por quadro.	40
3.9	Resultado do <i>K-means</i>	44
3.10	Quadros antes e após um corte	45
3.11	Quadros em um <i>dissolve</i>	45
3.12	Tomadas de visão de campo	46
3.13	Tomadas de <i>close</i>	46
3.14	Outros closes comuns em um vídeo de futebol	46
3.15	Elementos estruturantes usados na solução.....	47
3.16	Ritmo visual.....	49
3.17	Resultado da erosão da figura 3.16 usando os elementos estruturantes da figura 3.15	50
3.18	Volume da coluna central da imagem da figura 3.16 para cada elemento estruturante da figura 3.15	51
3.19	Orientação de movimento da câmera para cada quadro	52

3.20 Filtragem fechamento+abertura nos valores do gráfico da figura 3.19.....	53
4.1 Resultado da solução como aplicação Web.....	56
4.2 Quadros de diferentes jogos	57
5.1 Classificando <i>closes</i>	63
5.2 Ritmo visual vertical de um trecho de vídeo contendo um <i>Zoom</i>.....	64
5.3 Adaptação da visualização de quadros	65

Tabelas

4.1	Tempo de cada tipo de tomada para cada jogo	58
4.2	Classificação das tomadas	55
4.3	Estatística de posse de bola para os 7 jogos usados como amostra.....	59
4.4	Quantidade de <i>c/oses</i> selecionados por estratégia de corte.....	60
4.4	Comparação de tempos dos vídeos sumários com o vídeo original para cada estratégia de corte.....	61

1. Introdução

Apresentaremos nas seções seguintes as questões que motivaram o desenvolvimento desse trabalho, seu objetivo, trabalhos relacionados e sua estrutura.

1.1. Motivação

O uso de serviços multimídia na internet e em redes sem fio tem aumentado bastante devido ao avanço da tecnologia associada à computação e à comunicação. Hoje, através de recursos como *Streaming*, grandes vídeos podem ser disponibilizados, mas, mesmo assim, a largura de banda continua sendo um limitador. Daí a necessidade de se extrair e disponibilizar apenas partes significativas do vídeo em questão. Para tal, é necessário o uso de um sistema de recuperação de vídeo por conteúdo.

Uma etapa inicial de análise de conteúdo identifica objetos e/ou eventos de interesse no vídeo. Em seguida, a informação da fase de análise é usada para anotação e indexação do vídeo. Diversas aplicações podem ser implementadas a partir de bibliotecas digitais de vídeos devidamente anotados e indexados. Alguns exemplos são:

- ⊕ Transcodificação e adaptação para pequenos dispositivos. A mudança de resolução, quantização e enquadramento pode ser orientada para se obter melhores resultados em dispositivos com baixa resolução como PDA's, celulares, iPods (LIU Feng & GLEICHER Michael, 2006).
- ⊕ Interfaces inteligentes para *browsing* e consulta de grandes acervos (GOOGLE, 2006; QBIC, 2006).

- ⊕ Sumarização: produção de vídeos de menor duração, no qual os eventos mais relevantes de um vídeo inicial são preservados (EKIN *et al.*, 2003; LI & SEZAN, 2003).

Durante o processo de análise são usados *descritores de baixo nível* computados diretamente do vídeo. No contexto de imagens estáticas, os descritores estão relacionados a medidas de cor, textura e formas (MANJUNATH *et al.*, 2001; WALKER *et al.*, 2002). O processamento de vídeo considera, além destes, descritores relacionados a movimento de câmera e de objetos presentes na cena, bem como descritores de áudio.

Os descritores de baixo nível devem, em seguida, ser associados a descritores semânticos de mais alto nível. Essa associação constitui um grande desafio para a área. Nem sempre é possível relacionar a ocorrência de determinada cor a um personagem ou determinado movimento de câmera ao acontecimento de um gol, por exemplo. Esta dificuldade é conhecida na literatura como *gap semântico* (BROECKER, 2001).

Dentre os vídeos que despertam grande interesse, podemos citar os de esportes. Normalmente esses vídeos têm longa duração e necessitam de grande largura de banda para que possam ser disponibilizados e acessados. Além disso, uma grande quantidade de vídeos é produzida diariamente, inviabilizando o tratamento manual de acervos.

O processamento e análise de vídeos de esporte baseada no conteúdo, com o objetivo de detectar eventos importantes e geração de sumários, facilita a sua distribuição, uma vez que o resultado desse processamento geralmente corresponde a um vídeo com duração significativamente menor que o vídeo inicial.

Normalmente, isso é realizado de forma manual onde o vídeo é verificado e cenas e informações são extraídas pela observação humana. Este procedimento é custoso e, portanto, inviável para uso em larga escala. Como os eventos de um vídeo de esporte são bem definidos e se repetem em vídeos do mesmo tipo, podemos propor mecanismos de automatização dessa atividade. O foco de nosso trabalho será vídeos de futebol.

1.2. Objetivo

O objetivo dessa dissertação é propor descritores de baixo nível para vídeos de futebol baseados no ritmo visual (CHUNG,1999;NGO *et al.*, 1999; ZHANG *et al.*, 1994) além disso, visamos um menor custo computacional para a extração de informações, possibilitando o processamento mais próximo do tempo real. Ilustraremos também o uso desses descritores em aplicações simples de sumarização.

1.3. Trabalhos relacionados

Tomamos por base alguns trabalhos para proposição de nossa solução. Várias abordagens foram avaliadas, porém buscamos dentre aquelas relacionadas ao nosso objetivo, as que poderiam contribuir para uma nova proposta com ganhos relacionados ao custo computacional e a simplicidade de implementação. Os diferentes trabalhos são baseados em características específicas que motivam a extração de informações como utilização de cor dominante (EKIN & TEKALP, 2003; WANG *et al.*, 2004), Trajetória da bola (YOW *et al.*, 1995;GONG *et al.*, 1995;YU *et al.*, 2003; D’ORAZIO *et al.*, 2002), proximidade do gol (WAN *et al.*, 2003), identificação de jogadores (BERTINI *et al.*, 2005;FIGUEROA *et al.*, 2003) dentre

outras (LI e SEZAN, 2003). Alguns propõem a sumarização completa de vídeos de futebol com a detecção de eventos específicos como gols (EKIN et al., 2003) e ataques (WANG et al., 2004).

Ekin et al. propõem uma segmentação temporal baseada no conceito de “ataque”. É apresentada a noção semântica de ataque como sendo a unidade temporal para análise de vídeos de esporte. O vídeo de futebol é primeiramente dividido em tomadas, que são constituídas por um conjunto de quadros que representam uma ação contínua no tempo. Essas, por sua vez, são classificadas baseando-se na extração automática da cor do campo e dos uniformes dos jogadores. O movimento da câmera associado a cada quadro é estimado como um indicador da direção do ataque em tomadas de visão global do campo. Quadros adjacentes com a mesma direção de ataque são agrupados no mesmo segmento de ataque, juntamente com as tomadas de visão não global que os seguem.

A divisão de um vídeo em segmentos de ataque facilita a detecção de eventos, dentro de um período de tempo. O ataque não é apenas um critério de segmentação temporal para detecção de eventos, mas também fornece informações úteis para a sumarização de vídeos e extração de detalhes. No vídeo de futebol, a frequência e o tempo do ataque são componentes importantes do percentual de posse de bola. A proposta é detectar ataques pela análise do movimento dos jogadores, que normalmente coincide com o movimento de câmera, desde que esta está sempre tentando mostrar a bola. Tomadas não são a melhor escolha para unidade de análise do vídeo. Elas têm pouco significado semântico em vídeos de esporte. Vários ataques podem ocorrer dentro de uma mesma tomada e ela também pode ser apenas um *replay* de uma jogada anterior. A segmentação do vídeo em tomadas é feita pela diferença de quadros. A cor do campo é extraída através de picos de

distribuição do histograma e é estimada usando o algoritmo *Máxima Estimação* (EKIN *et al.*, 2003). A cor dos uniformes é extraída automaticamente pela localização da região dos corpos dos jogadores usando detecção de face. As tomadas são classificadas utilizando-se a razão dos *pixels* da cor do campo R_g e a razão dos *pixels* dos uniformes R_p (que são a razão entre o número de *pixels* dos correspondentes e o número de *pixels* de um quadro). As tomadas são então classificadas em: visão global quando $R_g > T_g$ e $R_p \leq T_p$, fora do campo quando $R_g \leq T_g$ e $R_p \leq T_p$ e *close* quando $R_p > T_p$, onde T_g e T_p são valores encontrados através de uma calibração utilizando-se algoritmo de reconhecimento de padrões em 10 minutos de vídeo.

A detecção do ataque é feita segmentando-se quadros adjacentes com mesmo movimento de câmera. O movimento de câmera em visões globais normalmente é consistente com a direção do ataque. O movimento é estimado utilizando-se estimativa de movimento baseada em blocos como utilizado por codificadores MPEG (MANJUNATH *et al.*, 2001). Os ataques são então classificados em: esquerda para direita e direita para esquerda. Alguns ataques são agrupados segundo alguns critérios, como tamanho mínimo definido.

A partir dos ataques, é calculada a posse de bola somando-se o tempo de ataque de cada time e calculando-se a razão. Um algoritmo de análise de textura é utilizado neste trabalho para detectar a rede e conseqüentemente lances de gol.

Xu *et al.* propõem uma solução cuja unidade semântica da segmentação do vídeo de futebol é representada pelos conceitos de “jogo” e “pausa” que são utilizados como fundamento para classificação de eventos e sumarização. O jogo acontece quando a bola está dentro do campo e não parada pelo juiz; a pausa é o complemento quando a bola está fora do campo ou quando o juiz pára o jogo.

Primeiramente, os quadros são classificados em 3 tipos de visões: global, *zoom-in* e *closes* utilizando-se a razão da área de campo. Uma grande área de campo existe em visão global, uma média quantidade no *zoom-in* e quase nenhuma nos *closes*.

O cálculo da área de campo é feito utilizando-se a detecção de cor dominante, nesse caso são usados 50 quadros escolhidos aleatoriamente dos 5 minutos iniciais do jogo e é então montado um histograma cumulativo. O pico corresponde à cor dominante (cor do campo). Esse processo é repetido 8 vezes resultando em um intervalo de variação formado pelos valores obtidos. Assim, é calculada a razão do número de *pixels* da cor do campo para cada quadro (número de *pixels* da cor do campo pelo total de *pixels* do quadro).

A classificação do tipo de quadro é feita utilizando-se valores pré-definidos (*thresholds*). Dependendo do valor da razão, os quadros são classificados. A direção da jogada é obtida através da orientação do gramado que está ligada ao movimento de câmera. Para isso, são usados os quadros que representam visões globais. São calculados vetores gradientes (g_x , g_y) da grama, usando um operador Sobel no canal de luminância. Depois, 50% dos *pixels* da cor do campo com o maior gradiente $G = |g_x| + |g_y|$ são declarados como arestas no quadro. A orientação do campo é calculada como a direção média dos vetores de gradiente de todos os *pixels* de aresta na área de grama. Finalmente, o ângulo médio α é comparado com 2 valores pré-definidos para classificar cada quadro (para direita ou esquerda). Segmentos de quadros com a mesma classificação são então agrupados, de acordo com regras pré-definidas relacionadas ao tempo e a mudança de classe.

1.4. Estrutura da dissertação

O restante dessa dissertação é apresentado da seguinte forma: o Capítulo 2 faz referência aos conceitos básicos utilizados na solução proposta envolvendo morfologia matemática, informações direcionais em texturas, ritmo visual e modelo de cores HSV. O Capítulo 3 propõe descritores de baixo nível para extração de características em vídeos de futebol utilizando ritmo visual. Propomos descritores: cor dominante e movimento de câmera. Mostramos ainda o uso destes descritores para detecção de transições de tomadas (cortes e transições graduais) e para a classificação de tomadas. O Capítulo 4 apresenta os experimentos realizados com o uso da solução proposta. Testes são apresentados, bem como exemplos de heurísticas que podem ser aplicadas no resultado, a fim de obtermos informações diversas do vídeo. Finalmente, o Capítulo 5 apresenta as conclusões, bem como sugestões de trabalhos futuros relacionados à solução proposta.

2. Noções básicas

Propomos neste trabalho descritores de baixo nível que podem ser calculados sem muito esforço computacional. Grande parte da informação do vídeo é desconsiderada por meio do método de amostragem conhecido por *ritmo visual*. Com isso, a quantidade de dados a ser analisada é bastante reduzida. Além disso, os descritores propostos se baseiam em operações da morfologia matemática com baixa complexidade e de fácil implementação.

Nas seções a seguir, revisamos os principais conceitos relacionados ao ritmo visual e às operações morfológicas empregadas nos descritores propostos no Capítulo 3.

2.1. Morfologia matemática

Operações morfológicas extraem componentes de interesse contidos em uma imagem original pela comparação desta com outra, de tamanho menor, a qual chamamos de elemento estruturante (SERRA, 1988; SOILLE, 2003). As características geométricas desse elemento estruturante estão relacionadas às informações que desejamos extrair da imagem original. As operações básicas da morfologia digital são a *erosão*, em que *pixels* que não atendem a um dado padrão são apagados da imagem, e a *dilatação*, em que uma pequena área relacionada a um *pixel* é alterada para um dado padrão. A morfologia pode ser aplicada em diversas áreas de processamento de imagens, como realce, filtragem, segmentação, esqueletização e outras.

2.1.1. Elemento estruturante

É um conjunto definido e conhecido (forma e tamanho), que é usado em uma operação sobre a imagem original, para salientar determinado aspecto. Ele pode assumir várias formas, dependendo do efeito a ser obtido. Sua origem pode ser definida em qualquer ponto e utilizada nas operações morfológicas. Um elemento estruturante no *pixel* x significa que sua origem coincide com x .

Exemplos de elementos estruturantes 3x3, com a origem no centro são mostrados na Figura 2.1.

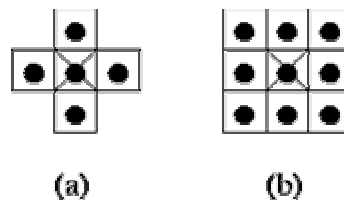


Figura 2.1 : Exemplos de elementos estruturantes, com origem no centro.

2.1.2. Operações morfológicas

A morfologia matemática foi originalmente criada para ser aplicada sobre imagens binárias, e posteriormente, estendida para imagens em tons de cinza. Quando a morfologia é aplicada a imagens binárias, na vizinhança de cada *pixel* da imagem original, é procurada uma configuração de pontos pretos e brancos. Quando a configuração é encontrada, ao *pixel* correspondente da imagem resultante é dado o rótulo “verdadeiro”; se não, o *pixel* resultante recebe o rótulo “falso”. Uma operação morfológica binária é, portanto, completamente determinada a partir da vizinhança examinada ao redor do ponto central da configuração de pontos pretos e brancos nessa vizinhança e pelo algoritmo. Quando a morfologia é aplicada às imagens em tons de cinza, na vizinhança de cada *pixel*, ou numa parte da vizinhança da imagem

original, é necessário conhecer o valor do *pixel* mais escuro MIN e o valor do *pixel* mais claro MAX. O valor do *pixel* resultante corresponde a uma combinação particular de MIN e MAX. O tamanho e forma da vizinhança, os valores de MIN e MAX e o algoritmo determinam completamente uma operação de morfologia em níveis de cinza. A seguir, recapitulamos as operações em níveis de cinza usadas no restante do texto.

2.1.2.1. Erosão

A erosão da imagem $f:Z^2 \rightarrow Z$ pelo elemento estruturante B é definida na equação 2.1.

$$\varepsilon_B(f) = \min\{f_h : h \in B\}, \quad (2.1)$$

em que $f_h(x) = f(x + h)$ é a translação de f por h .

Para calcular a erosão de uma imagem f pelo elemento estruturante B faz-se o seguinte:

1. Posiciona-se a origem do elemento estruturante sobre o primeiro *pixel* da imagem f que está sofrendo erosão;
2. Acha-se o valor mínimo de todos os níveis de cinza em f na vizinhança determinada pelo elemento estruturante B ;
3. Repete-se este processo para cada *pixel* da imagem que sofre erosão.

A erosão elimina picos positivos menores que o elemento estruturante e expande os picos negativos como pode ser observado na Figura 2.2.



Figura 2.2 : Exemplo de erosão. O elemento estruturante B é um círculo de raio 3 *pixels* com origem no centro

2.1.2.2. Dilatação

A dilatação da imagem $f:Z^2 \rightarrow Z$ pelo elemento estruturante B é definida pela Equação 2.2.

$$\delta_B(f) = \max\{f_h : h \in B\}, \quad (2.2)$$

em que $f_h(x) = f(x + h)$ é a translação de f por h .

Para calcular a dilatação de uma imagem f por um elemento estruturante B faz-se o seguinte:

1. Posiciona-se a origem do elemento estruturante sobre o primeiro *pixel* da imagem f a ser dilatada;
2. Acha-se o valor máximo de todos os níveis de cinza em f na vizinhança determinada pelo elemento estruturante B ;
3. Repete-se este processo para cada *pixel* da imagem a ser dilatada.

A dilatação elimina picos negativos, menores do que o elemento estruturante, e expande os picos positivos como pode ser observado na Figura 2.3.

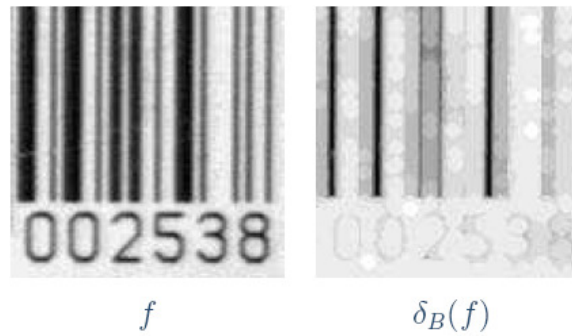


Figura 2.3 : Exemplo de dilatação. O elemento estruturante B é um círculo de raio 3 *pixels* com origem no centro

A Figura 2.4 mostra a comparação de resultados entre operações de dilatação e erosão sobre uma imagem unidimensional. Podemos verificar a variação nos níveis de cinza das imagens resultantes.

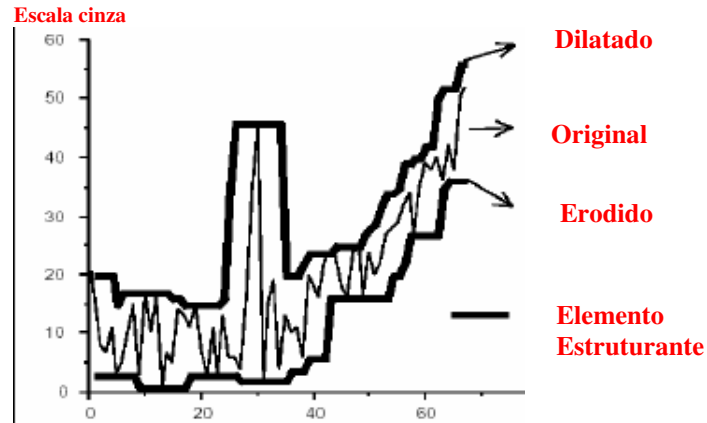


Figura 2.4 : Comparação entre erosão e dilatação.

2.1.2.3. Abertura

A abertura em geral suaviza o contorno de uma imagem, quebra estreitos e elimina proeminências delgadas. A operação de abertura é usada também para

remover ruídos da imagem. A abertura de uma imagem f por um elemento estruturante B é definida na Equação 2.3:

$$\gamma_B(f) = \delta_B(\varepsilon_B(f)) \quad (2.3)$$

A aplicação de uma erosão, seguida de uma dilatação, usando o mesmo elemento estruturante, é uma operação de abertura. Ela tende a abrir pequenos vazios ou espaços entre objetos próximos numa imagem.

2.1.2.4. Fechamento

O fechamento funde pequenas quebras, alarga golfos estreitos e elimina pequenos orifícios. Se uma abertura cria pequenos vazios na imagem, um fechamento irá preencher ou fechar os vazios. Essas operações podem remover muitos dos *pixels* brancos com ruídos, ou seja, basicamente são duais à abertura, i.e., primeiramente, é feita a dilatação e, após, é feita a erosão. Assim, essa operação é definida na Equação 2.4:

$$\phi_B(f) = \varepsilon_B(\delta_B(f)). \quad (2.4)$$

A Figura 2.5 mostra o resultado das duas operações sobre uma imagem em níveis de cinza.

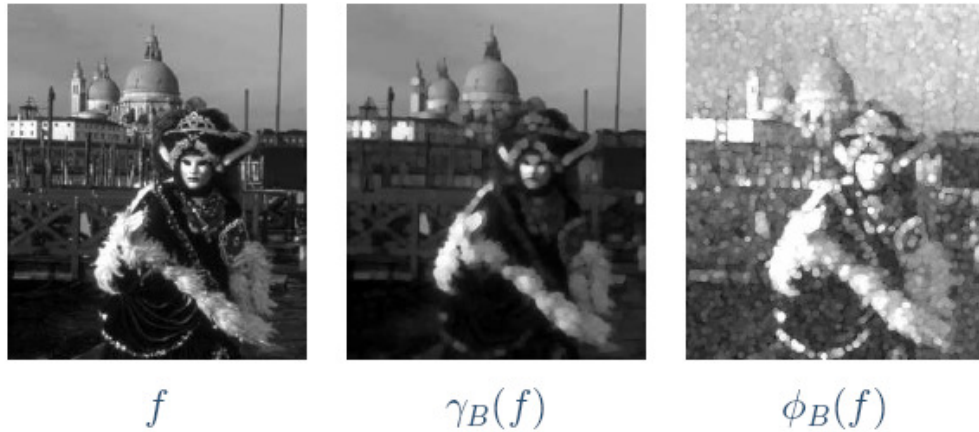


Figura 2.5 : Exemplos de abertura e fechamento.

2.2. Orientação de estruturas direcionais

A morfologia matemática é útil na extração de informações direcionais relacionadas à textura de uma imagem (SOILLE & BREEN, 1996; SOILLE & TALBOT, 2001; YAMADA *et al.*, 1993; DAUGMAN, 1985; DAVIES, 1997). A caracterização dessas informações direcionais pode ser obtida aplicando-se uma série de filtros direcionais (BAMBERGER *et al.*, 1992; BEAUCHEMAN & BARRON, 1995) na imagem original e, para cada direção associada ao filtro, calculando-se o volume da imagem resultante (soma dos valores de nível cinza dos *pixels* da imagem). Essa abordagem é conhecida como *orientação global*, pois fornece a orientação predominante, sem identificar a posição dessa estrutura direcional dentro da imagem. Esse é o objetivo da orientação local, onde para cada ponto da imagem é dado um valor que identifica sua orientação dentro da estrutura direcional a que pertence.

2.2.1. Orientação global

O resultado da informação direcional, em uma dada direção, pode ser estimado pelo cálculo do volume da imagem transformada por um elemento estruturante

direcional nesta direção. Orientações predominantes aparecem picos do volume ao longo das direções consideradas.

2.2.2. Orientação local

O resultado da informação direcional é dado para cada *pixel* da imagem original. Nesse caso, ao contrário do volume da imagem, que é utilizado na orientação global, é realizado uma análise da evolução dos valores da escala cinza da imagem transformada, para cada *pixel* individualmente. Essa variação é obtida a partir de operações de abertura ou fechamento na imagem original. Na prática, o uso de aberturas determina a direção de estruturas com brilho intenso na imagem original e, o fechamento, em estruturas escuras.

A orientação por abertura de um determinado *pixel* x de uma imagem f , em níveis de cinza, para uma determinada escala λ é dada pela orientação α da abertura morfológica direcional de comprimento λ , que menos modifica o valor da imagem original na posição x . A orientação por abertura é denotada por dir_λ^γ , e a orientação por fechamento dir_λ^ϕ , sendo definidas por dualidade conforme as Equações 2.5 e 2.6:

$$[dir_\lambda^\gamma(f)](x) = \{\alpha_i \mid [\gamma_{L_\lambda, \alpha_i}(f)](x) \geq [\gamma_{L_\lambda, \alpha_j}(f)](x), \forall \alpha_i \neq \alpha_j\}, \quad (2.5)$$

$$[dir_\lambda^\phi(f)](x) = \{\alpha_i \mid [\phi_{L_\lambda, \alpha_i}(f)](x) \geq [\phi_{L_\lambda, \alpha_j}(f)](x), \forall \alpha_i \neq \alpha_j\}; \quad (2.6)$$

em que a possível orientação discreta α depende do elemento estruturante envolvido na operação. A escolha do elemento estruturante é importante. Utilizamos

segmentos de retas com variações angulares diversas. A Figura 2.6 mostra exemplos de elementos estruturantes utilizados.



Figura 2.6 : Segmentos discretos de reta, com diferentes inclinações, usados como elementos estruturantes.

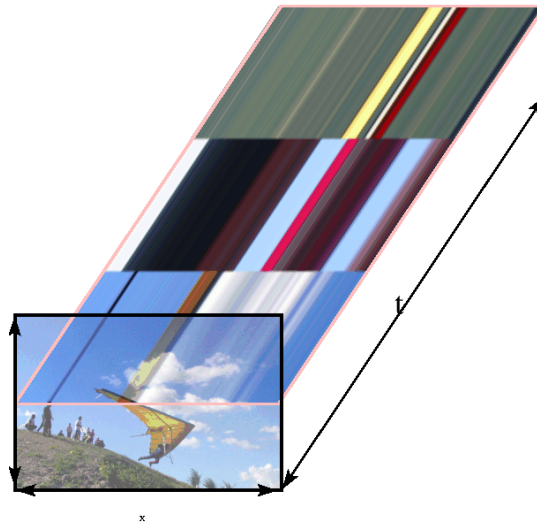
2.3. Ritmo Visual

O ritmo visual consiste na representação de um vídeo por uma imagem 2D obtida por amostragem. Permite, através da aplicação de métodos de processamento de imagem, a extração de informações relativas ao vídeo, como detecção de cortes e outros eventos (BEZERRA, 2004; GUIMARAES *et al.*, 2004; GUIMARAES *et al.*, 2001; FREEMAN & ADELSON, 1991).

Essa representação é obtida através de um esquema de amostragem de pontos do vídeo, como por exemplo, a extração de cada linha central dos quadros que o compõem. Um exemplo é mostrado na Figura 2.7. As linhas centrais horizontais de cada quadro se transformam em colunas do ritmo visual.

Um vídeo é representado por um volume de imagem $V(x, y, t)$ com um quadro (x, y) espacial bi-dimensional mais uma dimensão temporal t . Um ritmo visual é uma imagem bi-dimensional no volume, com uma dimensão sendo t e as outras sendo x ou y . Cada coluna de *pixel* na imagem do ritmo visual horizontal $VR_H(t, x) = V(x, h/2, t)$ corresponde a um momento no tempo e é composto pela linha central horizontal de um quadro do vídeo, em que h é a altura do quadro e w a sua largura. Analogamente, o ritmo visual vertical $VR_V(t, y) = V(w/2, y, t)$ é composto pela coluna central vertical de um quadro do vídeo original V . Vários eventos de vídeo como

movimento de câmera ou objetos e transições podem estar associadas a padrões no ritmo visual. A Figura 2.8 mostra um exemplo de ritmo visual horizontal.



(a) Ritmo visual



(b) Quadros do vídeo

Figura 2.7 : Geração do ritmo visual.

Esta simplificação proporciona uma análise do vídeo de maneira mais eficiente, pois transformamos o problema de segmentação do vídeo em um problema de segmentação de imagens bi-dimensionais, com custo computacional menor, uma vez que não precisamos usar os quadros em sua totalidade e, sim, apenas uma linha ou coluna de cada um.

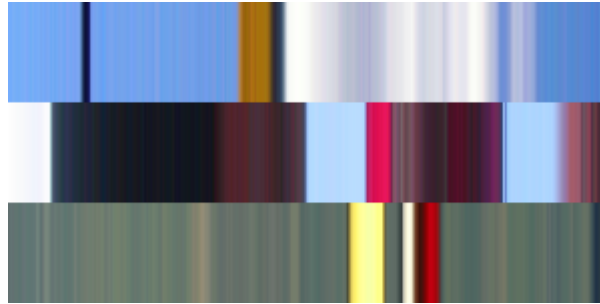


Figura 2.8 : Ritmo visual.

2.4. Modelo HSV de representação de cores

A escolha do modelo HSV deve-se à propriedade que ele tem de representar separadamente as informações de luminância e cromaticidade, se aproximando do sistema visual humano (FOLEY & VAN, 1996).

O modelo HSV é a concepção intuitiva da técnica utilizada por uma artista ao misturar cores básicas para obtenção de outros tons compostos. A seleção e obtenção de cores no modelo HSV é mais intuitiva que nos modelos RGB e CMY. Seu princípio baseia-se no controle dos valores de Matiz (Hue), Saturação (Saturation) e Valor (Value) (FOLEY & VAN, 1996).

Matiz é a componente que seleciona a "tinta" em uso, sendo controlada pela posição angular de um ponteiro num círculo de cores definida de 0 a 359. Por exemplo, vermelho e azul diferem no matiz. *Saturação* é a componente que determina a pureza da cor selecionada em Matiz. Todos os tons de cinza possuem Saturação = 0 e todos os Matizes puros possuem Saturação = 1. *Valor* regula o brilho da cor determinada por Matiz e Saturação. A cor preta possui brilho zero e qualquer valor de Matiz ou Saturação. O valor 1 de Valor determina uma intensidade pura de Matiz+Saturação. A Figura 2.9 ilustra a representação geométrica deste modelo por um cone. Cada cor corresponde a um ponto no cone, especificado pelo ângulo na base (Matiz), distância do eixo central (Saturação) e altura (Valor).

O descritor de cor dominante proposto no Capítulo 3 leva em consideração a componente Matiz. Para o descritor de direção de câmera levamos em consideração a componente valor.

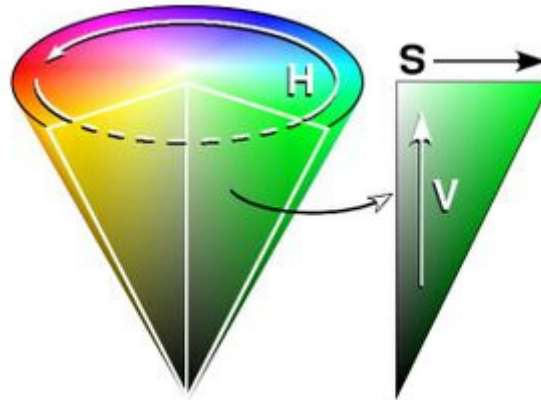


Figura 2.9 : representação geométrica do modelo HSV.

3. Descritores de baixo nível para vídeos de futebol

Contextualizaremos o problema da automatização da sumarização de vídeos, especificamente de futebol, que é o domínio de nosso interesse neste trabalho. A idéia é poder extrair informações referentes a uma partida de futebol contida em um vídeo (BEZERRA & LIMA, 2006; EKIN *et al.*, 2003; ASSFALG *et al.*, 2003) para levantamento de estatísticas do jogo e seleção de jogadas para compor o sumário. A solução proposta usa a informação de cor dominante e movimento de câmera como principais descritores de baixo nível. Assim, classificamos os tipos de tomadas. As que representam jogadas são analisadas para determinação da direção de movimentação da câmera, que indica o time que está atacando em um determinado momento. Com essas características detectadas, podemos extrair várias informações referentes ao jogo. Detalharemos os problemas encontrados, bem como a solução proposta nas seções seguintes.

3.1. Considerações preliminares

3.1.1. Cor

Um vídeo de futebol é composto por vários tipos de cena: visão do campo, *closes* e visão da torcida dentre outros. Exemplos dessas cenas são mostrados na Figura 3.1. Algumas características associadas a esse tipo de vídeo dificultam a utilização de métodos de segmentação temporal usados para decompor o vídeo em tomadas. Uma tomada é um conjunto de quadros temporalmente adjacentes e representa uma ação contínua no tempo (BIMBO, 1999). Tomadas são separadas por transições que podem variar do corte simples a transições graduais que incluem efeitos de edição. O movimento de câmera freqüente bem como a utilização de

transições graduais dificultam a detecção das mesmas. Um exemplo de transição gradual (*dissolve*) é mostrado na Figura 3.2.



Figura 3.1 : Cenas típicas em um vídeo de futebol.

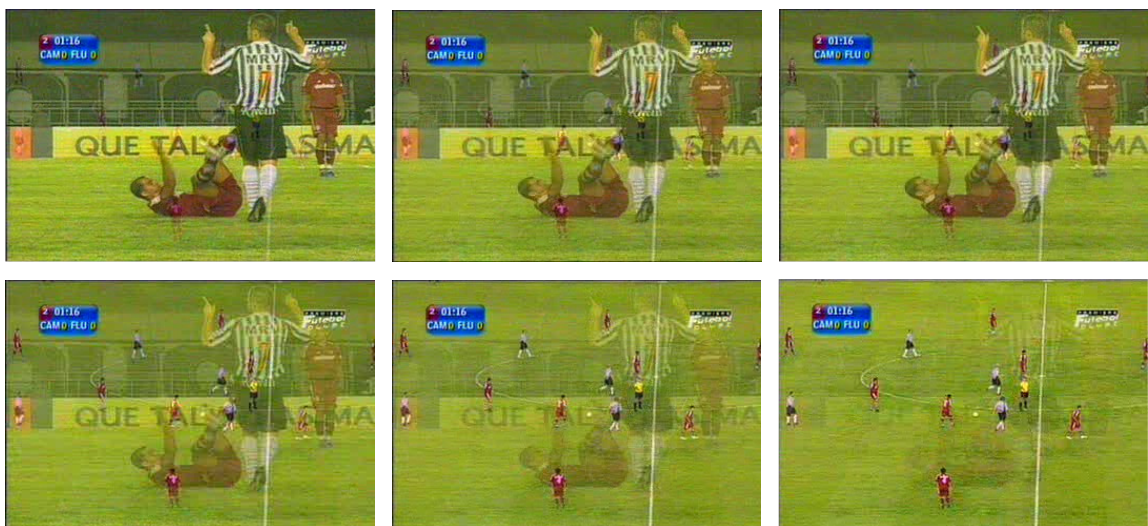


Figura 3.2 : Quadros de um *dissolve*.

Outras características são úteis para detecção de transições, como por exemplo, a cor do campo, predominantemente verde. Quando há uma transição de uma tomada de visão do campo para tomada de *close* ou outro tipo, há uma redução sensível do percentual dessa cor no quadro. Dessa forma, usamos essa característica como base para a solução de algumas tarefas: estimativa da cor de campo, detecção de transições de tomadas e classificação de tomadas.

3.1.2. Movimento de câmera

A direção constante do movimento da câmera permite a detecção do time que está executando uma jogada de ataque num determinado momento, visto que a direção da câmera corresponde geralmente à direção da bola. Isto é ilustrado na Figura 3.3. Notamos que a direção do movimento da câmera define o time que está atacando. Isso pode ser observado na maior parte do tempo de uma partida. Dessa forma, o time B está atacando, uma vez que o sentido do seu ataque é da direita para esquerda e o movimento da câmera, no exemplo, é para a esquerda.



Figura 3.3 : Direção da câmera X tempo.

Muitas técnicas de detecção de transições são baseadas na definição de métricas de similaridade entre os quadros. As métricas mais usuais são baseadas na diferença de intensidade dos *pixels* (ZHANG *et al.*, 1993; OSIAN & GOOL, 2003), histograma das cores (NAGASAKA & TANAKA, 1992) e características da imagem como localização de bordas (ZABIH *et al.*, 1995; WHITEHEAD *et al.*, 2004).

Nossa solução, ao contrário de outras propostas (ASSFALG *et al.*, 2003; TOVINKERE & QIAN, 2001; XIE *et al.*, 2002; ZABIH *et al.*, 1995) que utilizam os quadros em sua totalidade para detecção de transições, será baseada no ritmo

visual gerado a partir dos quadros do vídeo. Um exemplo pode ser visto na Figura 3.4, na qual temos o ritmo visual dos 1500 quadros iniciais de um vídeo. A partir do ritmo visual, podemos extrair informações, como cores dominantes, transições de tomadas (NGO *et al.*, 1999; GUIMARAES *et al.*, 2003; BEZERRA, 2004; KIM *et al.*, 2001; BEZERRA & LEITE, 2003; GUIMARAES *et al.*, 2001) e movimentação da câmera. Estamos interessados apenas no movimento horizontal da câmera. Portanto, usaremos o ritmo visual composto pelas linhas horizontais centrais do vídeo, que preserva de forma compacta a informação de deslocamento horizontal e desconsidera o deslocamento vertical. Exploraremos o fato de que a orientação da textura corresponde à direção do movimento. Padrões horizontais indicam ausência de movimento, padrões entre 0 e 90° indicam movimento para direita e de 90° a 180°, para a esquerda. Outra vantagem dessa representação é que evitamos áreas do quadro normalmente ocupadas por propagandas e outros detalhes incluídos na edição do vídeo como o placar ou logotipo da emissora e concentramos a atenção na área central do campo onde acontece o jogo.

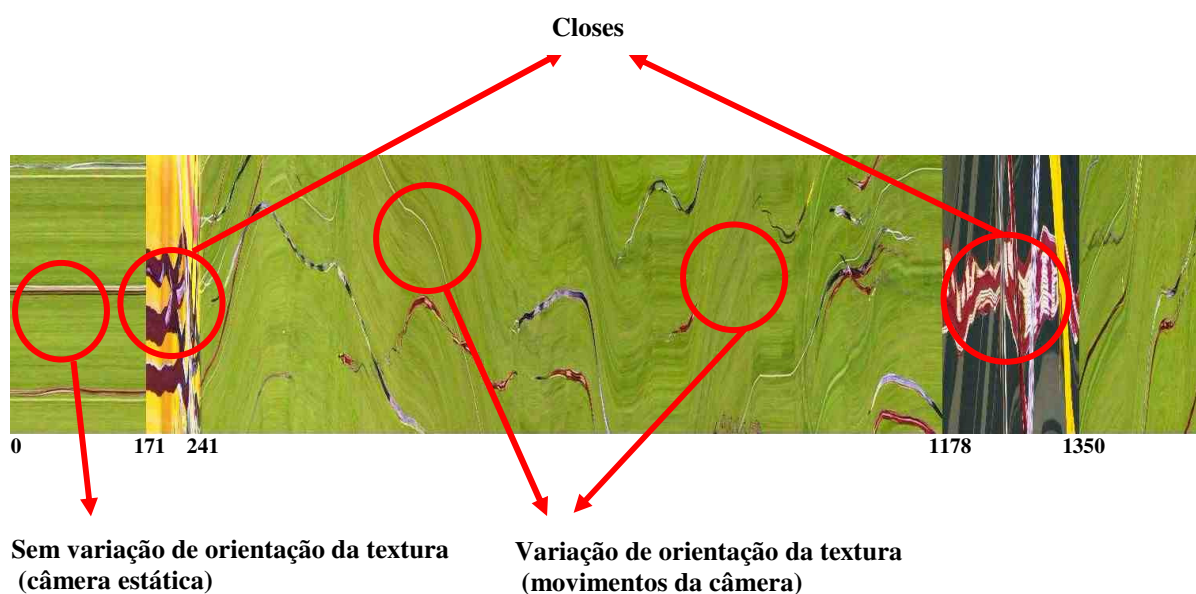


Figura 3.4 : Ritmo visual de um jogo típico.

Detalhamos as etapas de detecção de cor dominante, detecção de transição, classificação de tomadas e estimativa de movimento de câmera nas seções seguintes.

3.2. Detecção da cor dominante

Um campo de futebol tem uma cor dominante específica (um tom de verde), que pode variar de estádio para estádio. Devido a condições climáticas e de iluminação pode haver, também, variação para um mesmo estádio. Para nossa solução, é necessária uma calibragem inicial (fase de aprendizado) para cada vídeo, na qual definimos um intervalo de valores possíveis para a cor dominante (tonalidades de verde) no modelo HSV. Utilizamos o componente H (matiz) conforme abordagem descrita em Ekin *et al.* (2003), que detalhamos a seguir. Primeiramente, calculamos o pico do histograma (i_{pico}) do ritmo visual referente a 1800 quadros, aproximadamente 1 minuto de vídeo. O intervalo $[i_{\text{min}}, i_{\text{max}}]$ representa a cor dominante, em que i_{min} e i_{max} referem-se aos limites mínimo e máximo do intervalo, respectivamente, que satisfazem as condições nas Equações 3.1-3.6, nas quais HISTO refere-se ao histograma da componente H. Essas condições definem o limite mínimo como o menor índice à esquerda de i_{pico} , incluindo o valor do pico que tenha um número predefinido de *pixels*. Da mesma forma, calculamos o limite máximo sendo, nesse caso, o maior índice à direita. Utilizamos o número mínimo de *pixels* equivalente a 20 % do valor do pico, ou seja, $K = 0,2$.

$$\text{HISTO}[i_{\text{min}}] \geq K * \text{HISTO}[i_{\text{pico}}] \quad (3.1)$$

$$\text{HISTO}[i_{\text{min}} - 1] < K * \text{HISTO}[i_{\text{pico}}] \quad (3.2)$$

$$\text{HISTO}[i_{\max}] \geq K * \text{HISTO}[i_{\text{pico}}] \quad (3.3)$$

$$\text{HISTO}[i_{\max} - 1] < K * \text{HISTO}[i_{\text{pico}}] \quad (3.4)$$

$$i_{\min} \leq i_{\text{pico}} \quad (3.5)$$

$$i_{\max} \geq i_{\text{pico}} \quad (3.6)$$

Dessa forma, consideramos cor dominante todas as tonalidades com matiz H no intervalo entre i_{\min} e i_{\max} para um determinado vídeo.

No caso específico de jogo de futebol e, considerando os primeiros minutos do jogo, este intervalo corresponde aos tons de verde do campo. Esta propriedade pode ser verificada no exemplo da Figura 3.4.

Ekin *et al.* (2003) usa o histograma acumulado dos quadros no intervalo de tempo para calibração. Propomos o emprego do histograma do ritmo visual do mesmo intervalo de tempo, visto que preserva suficientemente a informação de distribuição de cor em cada quadro.

Para o ritmo visual da Figura 3.4, encontramos os valores [65,75] conforme mostrado no gráfico da Figura 3.5, que definem o intervalo de tonalidades verdes (valores de H), representando as cores assumidas pelo campo.

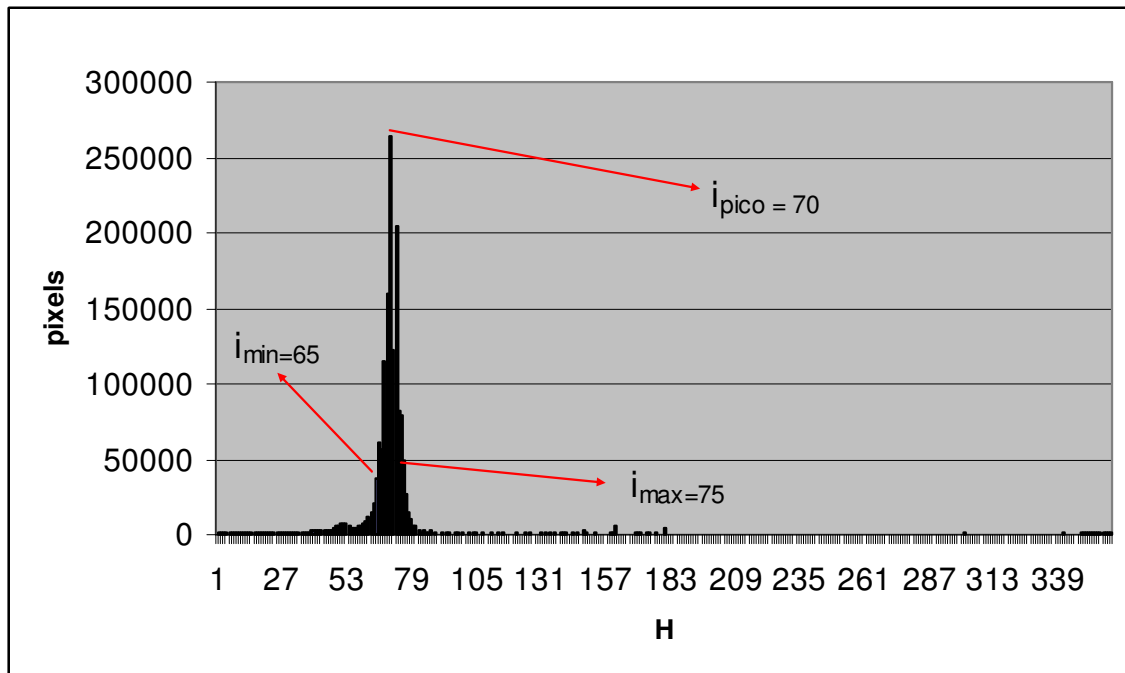


Figura 3.5 : Histograma da componente H da imagem da Figura 3.4.

A figura 3.6 mostra uma seqüência de quadros com as respectivas estimativas de área de campo. As imagens à direita mostram em branco as regiões onde o valor de H se encontra dentro do intervalo de cor dominante. Percebemos que o resultado obtido é uma estimativa razoável da área do campo. Falsos negativos ocorrem principalmente devido a sombras como na Figura 3.6(e-f). A ocorrência de falsos positivos, observado na Figura 3.6(c-d), revela casos em que o componente *matiz* somente é suficiente para a correta seleção da cor do campo. Os componentes saturação e valor podem ser usados em aplicações em que uma maior precisão é necessária. Para a situação proposta o uso apenas do *matiz* produz resultados satisfatórios.



Figura 3.6 : Detecção de regiões com matiz H no intervalo de cor dominante.

3.3. Detecção de transições e classificação das tomadas

Propomos a utilização do intervalo de cor dominante descrito na Seção 3.2 para detectar a ocorrência de transições para o particionamento do vídeo em tomadas e também para classificar as tomadas. No nosso caso específico, estamos interessados na ocorrência de *close*, que gera um corte na seqüência do jogo. A idéia é detectar quando uma coluna tem a quantidade de *pixels* dentro do intervalo de cor dominante, menor que um limiar pré-determinado. Esse limiar deverá ser o número de *pixels* da coluna multiplicado por um fator, que utilizamos o valor de 0,3, obtido experimentalmente, produzindo resultados satisfatórios. Os intervalos de quadros, que têm número de *pixels* menor que o valor pré-determinado, compõem um *close*. Assim, podemos detectar transições.

A Figura 3.7 refere-se ao ritmo visual dos 5000 quadros iniciais de um vídeo.

No gráfico da Figura 3.8, as transições são definidas pelas fronteiras dos vales.

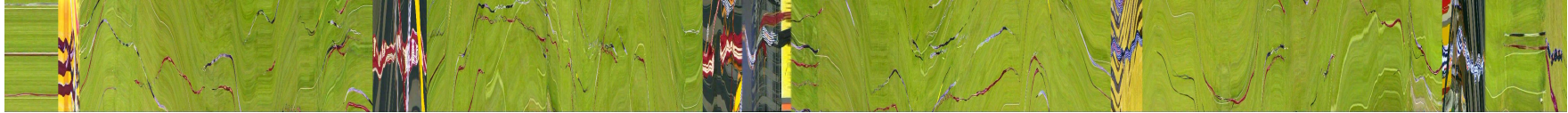


Figura 3.7 : Ritmo visual dos 5000 quadros iniciais de um vídeo

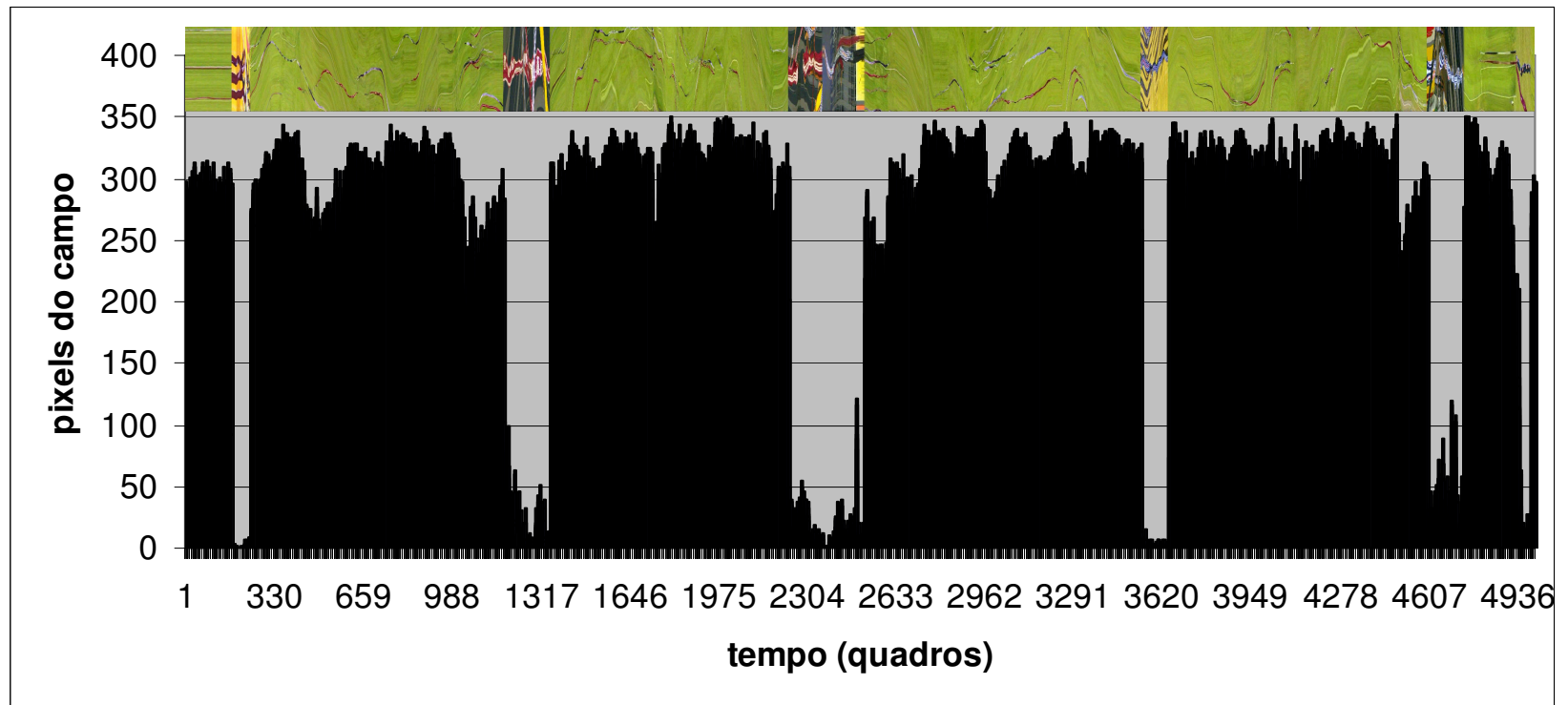


Figura 3.8 : Gráfico representando a quantidade de *pixels* na cor dominante por quadro.

Para evitar o uso de limiar fixo, utilizamos um algoritmo de clusterização (K-means) para separar dois conjuntos possíveis ($K=2$), através de medidas de similaridade. Um conjunto com grande quantidade de *pixels* de tons verdes e outro com pequena quantidade. O resultado pode ser observado no gráfico da Figura 3.9. Para cada quadro, temos a classe a que o mesmo pertence, representada pelos valores 0 (*close*) ou 1 (visão do campo). Podemos detectar então quando ocorre uma nova transição. As cenas do tipo *close* incluem também diversas situações que não dizem respeito ao jogo propriamente. Alguns exemplos são cenas de jogadores e outros membros da equipe fora do campo e cenas da platéia.

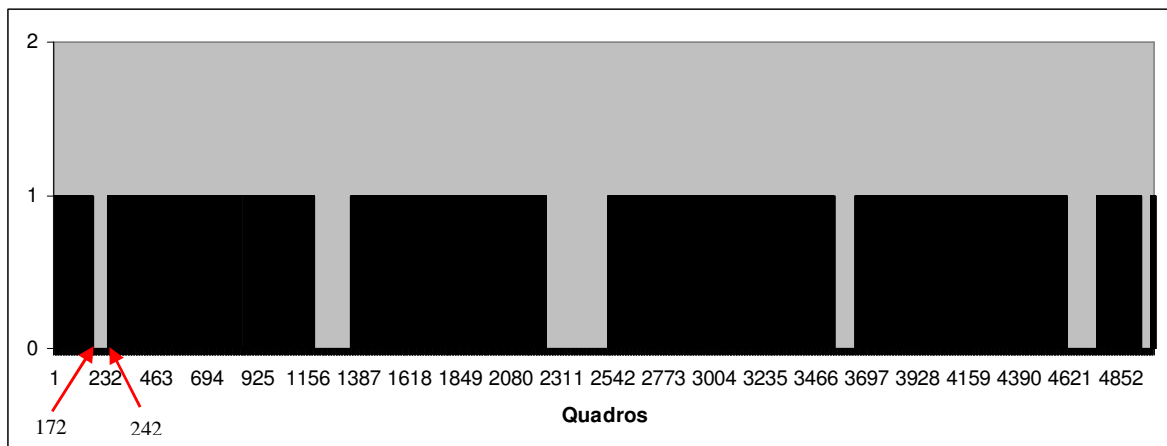
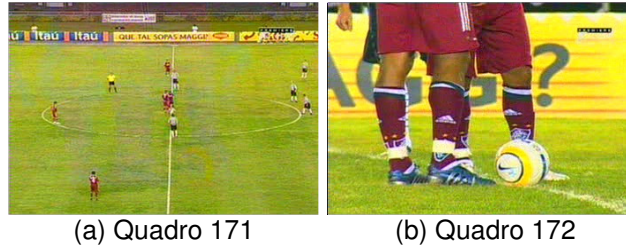


Figura 3.9 : Resultado do K-means.

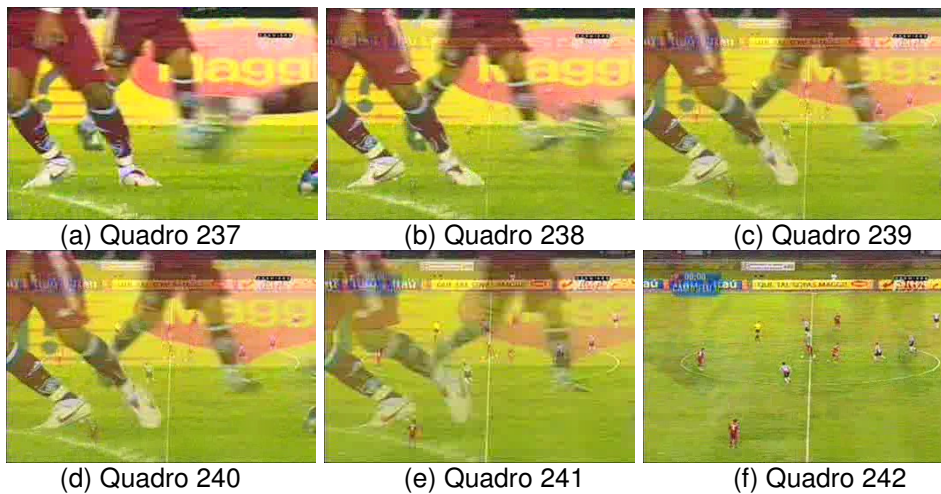
Por exemplo, o gráfico da figura 3.9 mostra que nos quadros 172 e 242 do vídeo aconteceram transições. Nas Figuras 3.10 e 3.11, mostramos quadros do vídeo: 171, 172 e 237 a 242. Podemos ver que, de fato, ocorrem as transições. O método é efetivo para detecção de cortes e *dissolves*.



(a) Quadro 171

(b) Quadro 172

Figura 3.10 : Quadros antes e após um corte.



(a) Quadro 237

(b) Quadro 238

(c) Quadro 239

(d) Quadro 240

(e) Quadro 241

(f) Quadro 242

Figura 3.11 : Quadros em um *dissolve*.

A partir da detecção das transições, podemos então classificar os quadros em duas classes: 1) visão global do campo; 2) *closes* (*closes* propriamente ditos e outras cenas que não sejam visões do campo). Na Figura 3.9, essa classificação fica clara. Analisando um jogo de futebol, podemos verificar que a visão global do campo e os *closes* são as duas cenas mais comuns. O jogo pode estar transcorrendo normalmente em ambas as situações, sendo que no caso dos *closes*, a cena normalmente é a imagem aproximada de um ou mais jogadores. A Figura 3.12 apresenta amostra de quadros da classe 1, conforme gráfico da Figura 3.9, ou seja, visões do campo.



Figura 3.12 : Tomadas de visão do campo.

A Figura 3.13 apresenta amostra de quadros da classe 0, conforme gráfico da Figura 3.9, ou seja, closes.



Figura 3.13 : Tomadas de close.

Em trabalhos futuros pretendemos ampliar a classificação dos quadros desmembrando os *closes* em visão da torcida, dos jogadores, da arbitragem, dentre outras, conforme exemplo da Figura 3.14, através do estudo dos *closes* previamente selecionados por nossa solução.

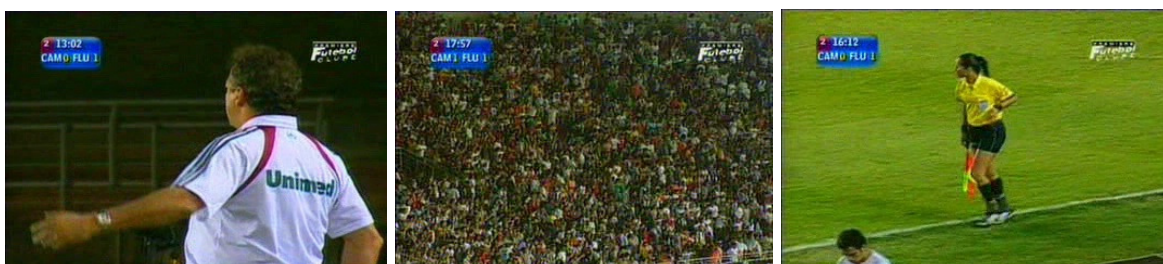


Figura 3.14 : Outros closes comuns em um vídeo de futebol.

3.4. Detecção de movimento de câmara

Após a classificação das tomadas, detectamos o movimento de câmara nas cenas de visão de campo. Em nosso método de sumarização utilizamos uma composição das técnicas de orientação de textura discutidas no Capítulo 2. A ideia é realizar erosões na imagem do ritmo visual, em lugar de aberturas ou fechamentos. Utilizamos 16 elementos estruturantes com tamanho 29×29 pixels, sendo todos segmentos de reta com diferentes ângulos de inclinação. O tamanho dos mesmos foi escolhido em função do número de quadros por segundo de um vídeo que é aproximadamente 29 quadros. Dessa forma, na erosão realizada em cada *pixel* da imagem estará sendo levado em consideração um segundo de vídeo. Os elementos estruturantes escolhidos são mostrados na Figura 3.15. Utilizamos variações na inclinação em torno de 12° . Dessa forma, podemos detectar mudanças sensíveis na orientação da textura, definindo com uma maior precisão a direção do movimento.

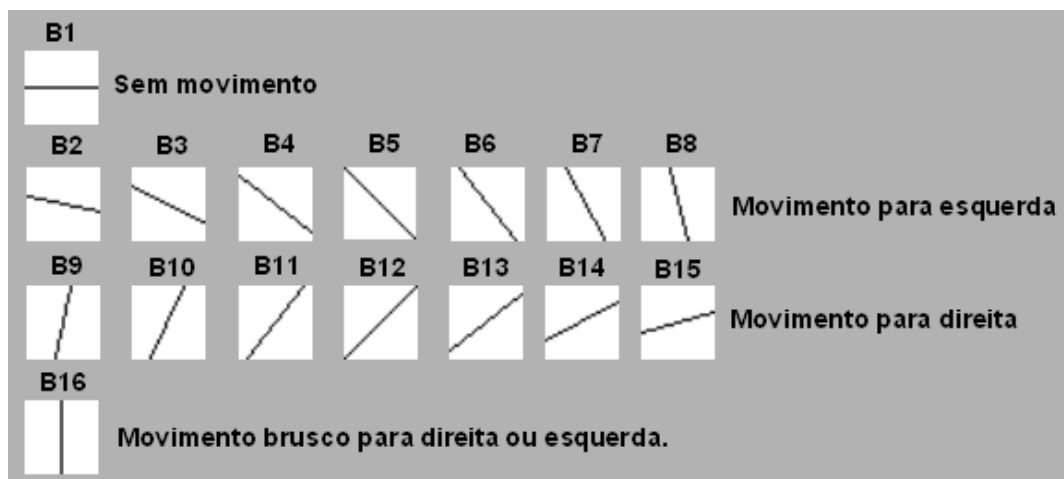


Figura 3.15 : Elementos estruturantes direcionais usados para estimação de orientação na textura do ritmo visual.

Utilizamos os valores em níveis de cinza da componente valor (V) do ritmo visual para a realização das erosões propostas. Realizamos a erosão utilizando os 16 elementos estruturantes. Para cada coluna, verificamos qual o elemento estruturante que produz o maior volume para a mesma.

Seja VR o ritmo visual do vídeo. O volume de uma coluna t de VR é definido como a soma dos níveis de cinza nesta coluna, conforme a Equação 3.7:

$$V(VR, t) = \sum_y VR(y, t) \quad (3.7)$$

A direção dominante $Dir(t)$ para a coluna t é dada pelo elemento estruturante $B \in \{B_1, \dots, B_{16}\}$ que produz a erosão do VR de maior volume $V(\mathcal{E}_B(VR), t)$ definida conforme Equação 3.8.

$$Dir(t) = \{\alpha_i \mid V(\mathcal{E}_{B_i}(VR), t) > V(\mathcal{E}_{B_j}(VR), t), B_i \neq B_j\} \quad (3.8)$$

sendo α_i o ângulo determinado pelo elemento estruturante B_i .

Os elementos de B_2 a B_8 definem movimento para esquerda, B_9 a B_{15} para direita, B_1 ausência de movimento e B_{16} representa um movimento rápido para direita ou esquerda. Neste caso, assumimos que a direção para este quadro é a direção do quadro anterior.

Assim, como cada coluna do ritmo visual representa um quadro do vídeo, podemos detectar qual a direção de câmera associada ao quadro em questão. Os movimentos considerados são: sem movimento, movimento para direita ou movimento para esquerda. Essa classificação pode ser observada na Figura 3.15 na qual, para cada elemento estruturante, temos a direção associada.

A técnica proposta, que chamaremos de regional, difere das técnicas de detecção de orientação de textura local e global abordadas no Capítulo 2. Nesta Seção tratamos cada *pixel* individualmente como na orientação local verificando as variações de níveis de cinza obtidos após as operações de erosão, porém utilizamos o conceito de volume usado na orientação global, com a diferença de que calculamos o mesmo para cada coluna da imagem e não o volume global da imagem. Como cada coluna do ritmo visual representa um quadro do vídeo, temos assim a orientação da textura dos mesmos. Mesclamos as duas técnicas adequando-as à nossa situação.

A Figura 3.17 mostra o resultado da erosão de uma imagem inicial (Figura 3.16) pelos elementos estruturantes B_1 a B_{16} .

Soille propõe algoritmos que otimizam a erosão realizada com elementos estruturantes do tipo segmento de reta (SOILLE, 1998; VICENT, 1992), permitindo um baixo tempo de execução para o cálculo deste descritor.



Figura 3.16 : trecho de ritmo visual

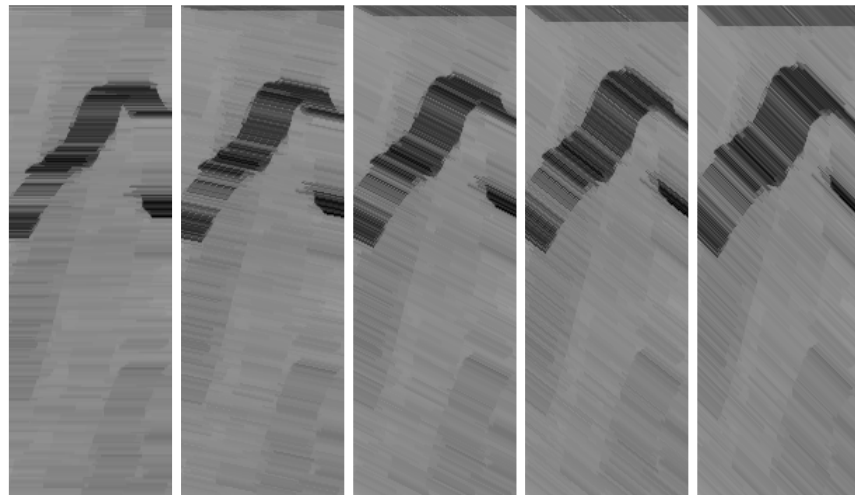
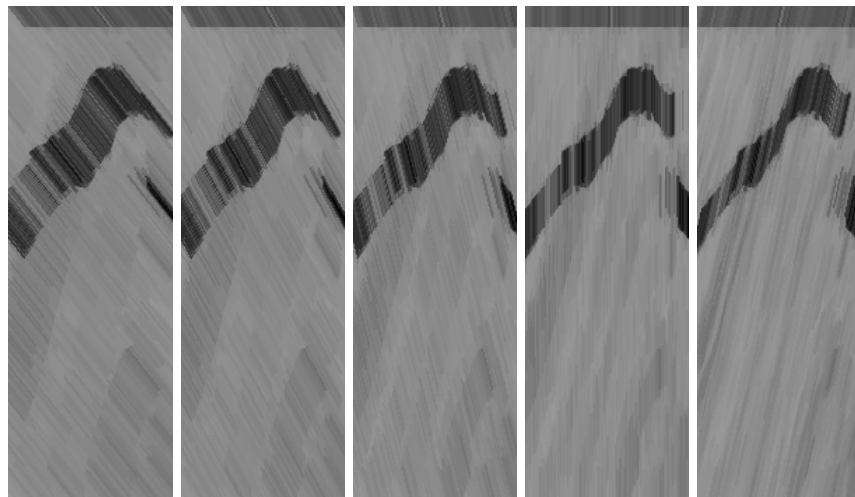
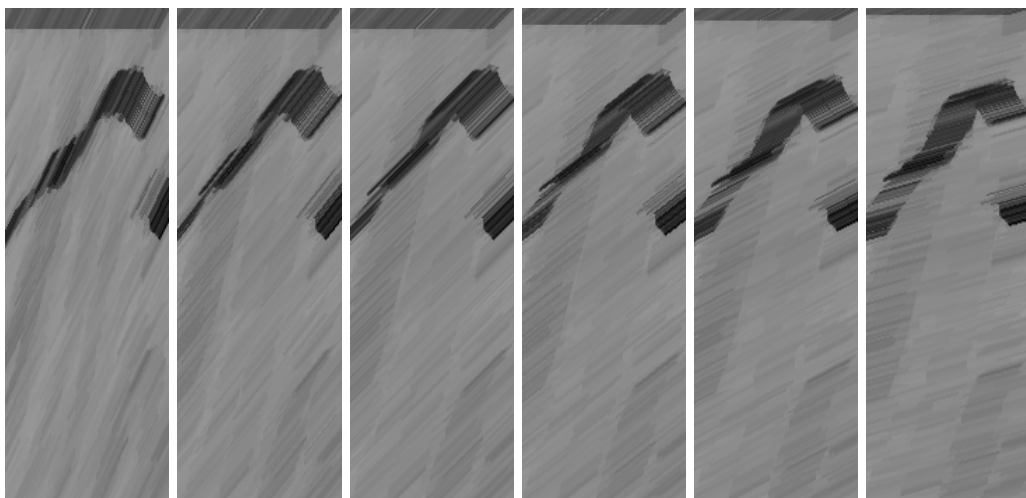
(a) B_1 (b) B_2 (c) B_3 (d) B_4 (e) B_5 (f) B_6 (g) B_7 (h) B_8 (i) B_9 (j) B_{10} (l) B_{11} (m) B_{12} (n) B_{13} (o) B_{14} (p) B_{15} (q) B_{16}

Figura 3.17 : Erosão na imagem da Figura 3.16 usando os elementos estruturantes B_1 a B_{16} .

Podemos verificar na imagem inicial uma tendência vertical para direita de orientação da textura (Figura 3.16). Quando realizamos as erosões, percebemos que os elementos estruturantes (com orientação vertical e para direita) resultam em imagens mais próximas da original (Figura 3.17(j), Figura 3.16(l) e Figura 3.16(m)). Nesses casos, o valor do volume da imagem é maior do que quando utilizamos outros elementos estruturantes. Isso é mostrado na Figura 3.18 que apresenta o volume da coluna central da imagem da Figura 3.16 para cada elemento estruturante (B_1 a B_{16}). Dessa forma, podemos definir a direção associada ao quadro representado por essa coluna como a definida pelo elemento estruturante B_{11} . Ou seja, a direção da câmera para o quadro correspondente à coluna central do ritmo visual é da esquerda para a direita.

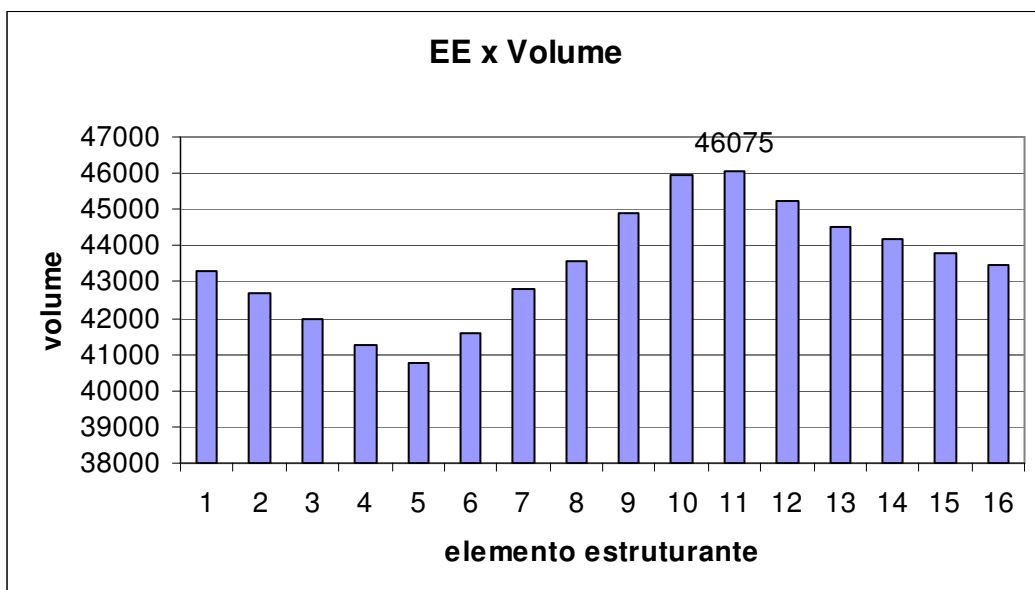


Figura 3.18 : Volume da coluna central da imagem da Figura 3.16 para cada EE.

Observamos a variação da direção de ataque no decorrer do tempo no gráfico da Figura 3.19. As direções são obtidas da detecção de orientação para os 5000 quadros representados pelo ritmo visual da Figura 3.7. A informação de direção é

ausente para quadros em tomadas de *close*. O valor 0 representa ausência de movimento, 1 representa movimento para direita e -1 movimento para esquerda.

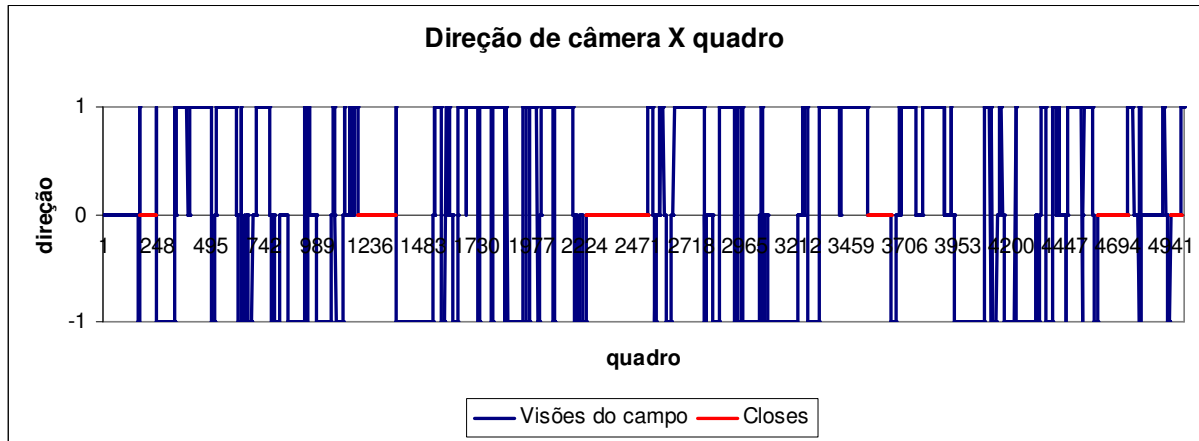


Figura 3.19 : Direção de movimento da câmera para cada quadro.

Normalmente, em uma partida de futebol, um time atacando faz rápidos movimentos contrários à direção do seu ataque (recuos de bola). Isso é ilustrado no gráfico da Figura 3.19 onde existem pequenas inversões de movimento em determinadas seqüências que não correspondem à mudança na direção de ataque. Para resolver este problema, realizamos uma filtragem composta de uma operação de fechamento seguida de uma abertura para remover picos e vales de curta duração, que correspondem a essas inversões. O resultado é ilustrado na Figura 3.20.

Utilizamos nessas operações morfológicas um elemento estruturante horizontal de tamanho 15, correspondente a 0,5s em vídeos de 29 quadros por segundo, que experimentalmente apresentou bom resultado. Assim, conseguimos separar as tomadas que representam as jogadas de ataque de cada time.

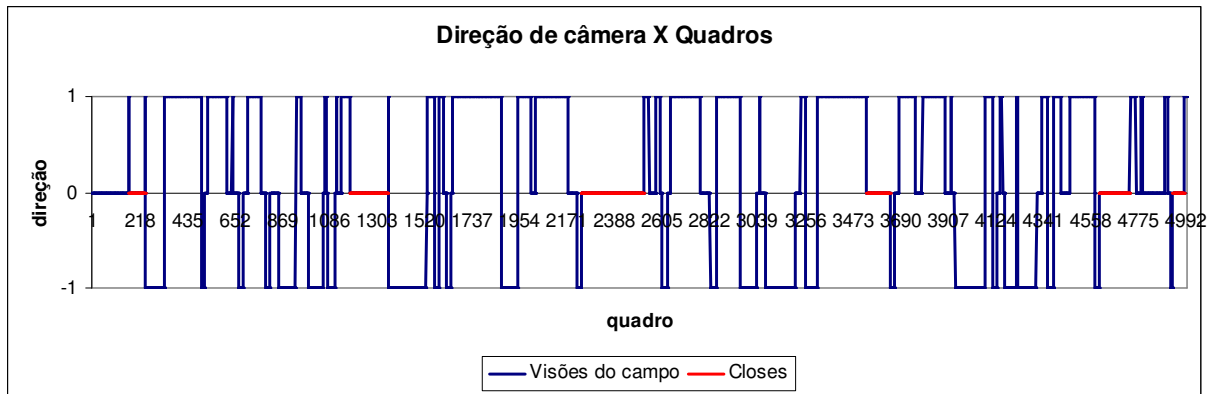


Figura 3.20 : Filtragem da direção de câmera (figura 3.19).

A estimativa de direção do ataque constitui um descritor de baixo custo computacional, se comparado com técnicas mais robustas de detecção de movimento, como o cálculo de campos direcionais (BIMBO, 1999) por exemplo. A técnica proposta mostrou-se efetiva para cenas em que o movimento dominante é o movimento da câmera (e não de objetos isolados), como é o caso das cenas de visão global do campo nos jogos de futebol.

A aplicação da técnica ao ritmo visual vertical pode ser usada para a detecção de movimento tipo *tilt*. A combinação de informação dos ritmos visuais horizontal e vertical permite a detecção de *zooms*.

4. Sumarização: experimentos e resultados

Apresentaremos a seguir o uso dos descritores propostos no Capítulo 3 numa aplicação de sumarização de vídeos de jogos de futebol, assim como os experimentos efetuados e os resultados alcançados.

4.1. Experimentos

Como produto resultante dos descritores de baixo nível temos os quadros classificados em : 1) ausência de movimento, 2) closes, 3) movimento para direita e 4) movimento para esquerda. Podemos então aplicar algumas heurísticas para extrair informações importantes do jogo, o que detalharemos nas seções seguintes.

4.1.1. Sumarização

Utilizamos algumas heurísticas para compor os sumários a partir da classificação dos quadros. Podemos ainda calcular o tempo de posse de bola como produto secundário derivado da classificação. O time que ataca no sentido da esquerda para direita terá seu tempo de posse de bola computado pelo número de quadros classificados como movimento para direita dividido pelo número de quadros por segundo, no nosso caso 29. Temos assim, a posse de bola desse time em segundos. O mesmo fazemos para o outro time. Os tempos dos quadros de *close* e sem movimento de câmera são desconsiderados.

Para composição do sumário, verificamos que em um jogo de futebol, após um lance importante, normalmente ocorre um *close*. Como esses foram detectados em sua totalidade, podemos montar um vídeo formado pelos quadros que antecedem os *closes*. Definimos um intervalo de tempo, como 10s por exemplo, e assim teremos

um vídeo formado com os 10s que antecedem cada *close* acrescidos dos mesmos, ou seja, cerca de 10s de cada lance importante do jogo.

Dependendo do jogo, temos uma grande quantidade de *closes* que nem sempre correspondem a jogadas relevantes. Para obter resultados mais precisos, podemos apenas selecionar os lances com *closes* com duração superior a um valor pré-definido que pode ser utilizado no caso da aplicação da solução para obtenção de resultados mais próximos do tempo real. Se já tivermos o jogo completo para aplicação da sumarização poderemos aplicar *k-means* sobre todos os tempos de *close* e utilizar os *closes* agrupados da classe que possui os maiores valores. Quando dois ou mais *closes* ocorrem em um espaço de menos de 10s, agrupamos os mesmos em uma única jogada importante.

Podemos também separar todas as jogadas de um determinado time, uma vez que temos os quadros classificados pela direção do ataque.

A detecção de contra-ataques acontece quando ocorrer uma inversão brusca no sentido da jogada, ou seja, quando tivermos duas tomadas de sentidos opostos, seguida uma da outra, sem *closes* ou quadros classificados como sem movimento entre elas.

Um sumário em forma de página web com *thumbnails* representando os melhores lances de uma partida, bem como, estatísticas como posse de bola é exibido na Figura 4.1. Cada imagem *thumbnails* pode ser usada para acessar o trecho de vídeo de uma jogada.

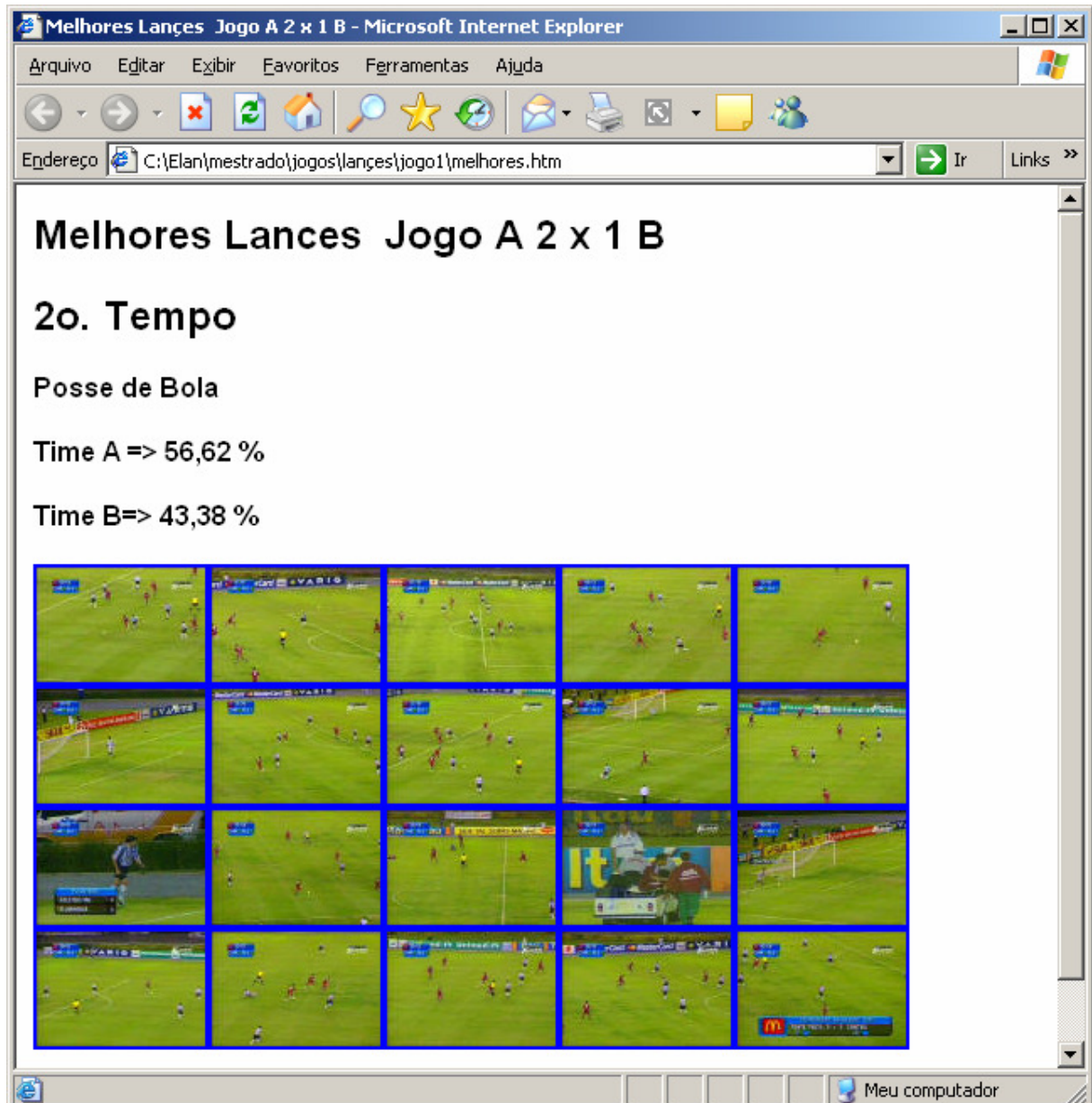


FIGURA 4.1 : Resultado da solução como aplicação Web.

4.2. Resultados

Foram extraídas informações de 7 jogos de futebol. Analisamos o 1º. tempo de cada jogo, totalizando cerca de 5h e 30 min de vídeo. Todos os vídeos utilizados nos testes possuem quadros com resolução de 352 x 240 e frequência de 29 quadros por segundo. Foram utilizados jogos com situações diferentes de iluminação do campo como pode ser observado na Figura 4.2.

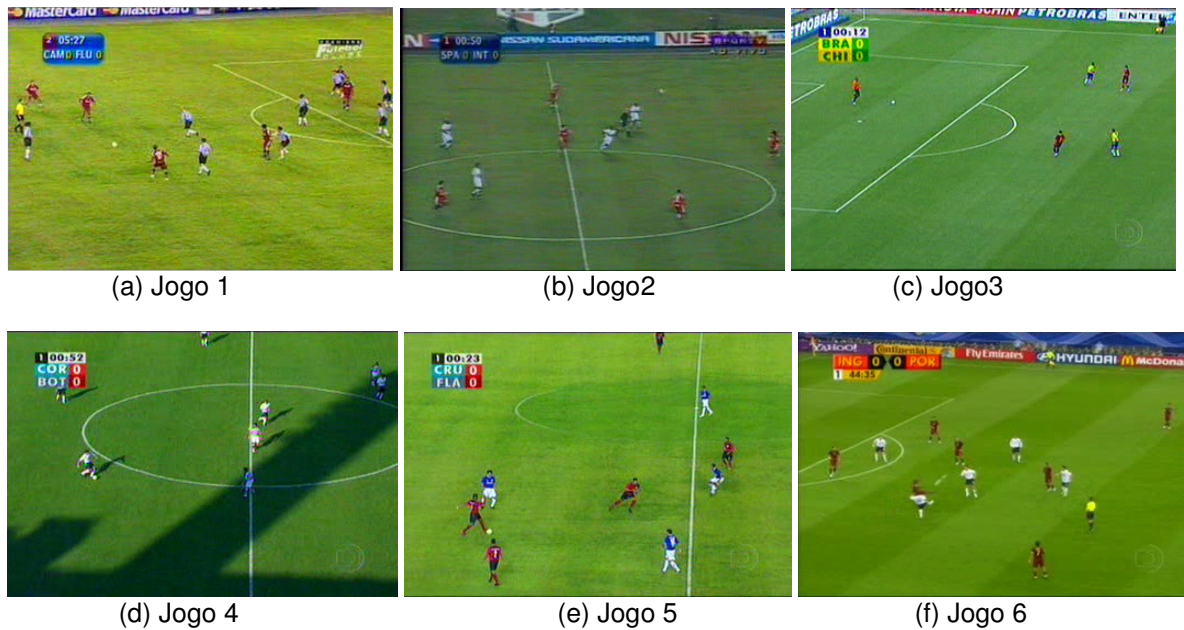


FIGURA 4.2 : Quadros de diferentes jogos.

Um fator determinante dos resultados é o tamanho do elemento estruturante, utilizado para eliminar as pequenas inversões de movimento através de operações morfológicas de fechamento e abertura. Conseguimos bons resultados utilizando um elemento estruturante com tamanho 15. Os resultados obtidos ainda podem ser refinados, agregando pequenas seqüências de quadros ao tipo de movimento anterior. Para tal, precisamos definir um valor mínimo de quadros que uma seqüência de mesma direção pode ter. Assim, quando uma seqüência for menor que esse valor mínimo, agregamos esses quadros à tomada anterior.

A sumarização foi aplicada aos 7 vídeos de jogos utilizados como amostras de testes produzindo sumários para o 1º. tempo de cada jogo. Os resultados obtidos são discutidos a seguir.

A Tabela 4.1 mostra para cada jogo a duração de cada tipo de tomada. A Tabela 4.5 mostra a estimativa de posse de bola para cada equipe.

Jogos	Duração	Tempo em Minutos			
		Close	Mov. Direita	Mov. Esquerda	Sem Mov.
Jogo 1	46 min	8,5 min	18,5 min	14,1 min	4,9 min
Jogo 2	45 min	15,4 min	10,2 min	7,5 min	11,5 min
Jogo 3	46 min	14,5 min	15,5 min	14,5 min	1,5 min
Jogo 4	45 Min	24,1 min	12,3 min	9,7 min	1,4 min
Jogo 5	48 min	15,0 min	17,0 min	14,3 min	2,0 min
Jogo 6	47 min	12,0 min	15,5 min	14,0 min	5,5 min
Jogo 7	49 min	16,5 min	15,0 min	12,5 min	5,0 min

Tabela 4.1 : Tempo de cada tipo de tomada para cada um dos 7 jogos.

Depois de realizadas medições manuais nos jogos em questão, pudemos realizar uma comparação com os resultados obtidos pela solução proposta. A classificação das tomadas para cada jogo pode ser observada na tabela 4.2. Não houve erros de detecção de transição, nem de classificação das tomadas.

Jogo	Quantidade de Tomadas	
	Closes	Visão de campo
Jogo 1	85	86
Jogo 2	119	119
Jogo 3	109	109
Jogo 4	105	105
Jogo 5	114	113
Jogo 6	108	109
Jogo 7	139	138

Tabela 4.2 : Classificação das tomadas.

Podemos observar também que lances como gols, ou jogadas como chutes a gol, são em sua totalidade detectadas pela solução. Como a mesma é dependente dos *closes* para detecção de jogadas importantes, pode acontecer, dependendo da filmagem, que cenas de menor importância como substituição de jogadores ou confusões inerentes a um jogo, possam ser selecionadas. Daí, a necessidade da filtragem dos *closes* utilizando como fator de corte o tempo dos mesmos, uma vez que os lances mais importantes normalmente têm *closes* com duração maior.

A Tabela 4.3 mostra o tempo de posse de bola calculado por nossa solução e pela medição manual(h), divulgada pelas emissoras de tv responsáveis pelas exibições dos jogos. A última coluna apresenta o erro como a diferença absoluta entre os percentuais. Podemos observar que, na maioria dos casos, o erro foi inferior ou bem próximo a 1 %. Uma exceção foi o jogo 4, que apresentou um erro de 6,9 %, fato explicado pela variação de iluminação do campo que apresentava sombras interferindo nos valores encontrados. Um dos quadros desse jogo foi apresentado na Figura 4.1(d).

Estaremos melhorando a solução em trabalhos futuros tratando a variação de iluminação com a utilização de calibragens intermediárias da solução em relação a cor dominante, bem como, o uso de outros modelos de representação de cor que permitam analisar além da intensidade, outras características da cor como o brilho, por exemplo.

Jogos	Time A	Time B	Time A(h)	Time B(h)	Erro
Jogo 1	56,62 %	43,38 %	56,3 %	43,7 %	0,32 %
Jogo 2	58,54 %	41,46 %	58,7 %	43,3 %	0,16 %
Jogo 3	52,21 %	47,79 %	53,3 %	44,7 %	1,09 %
Jogo 4	56,00 %	44,00 %	49,1 %	50,9 %	6,90 %
Jogo 5	54,41 %	45,59 %	53,3 %	46,7 %	1,11 %
Jogo 6	52,50 %	47,50 %	53,0 %	47,0 %	0,50 %
Jogo 7	54,00 %	46,00 %	53,0 %	47,0 %	1,00 %

Tabela 4.3 : Estimativas de posse de bola. Tempos estimados (colunas 2-3), tempos calculados manualmente (colunas 4-5) e erro (coluna 6).

Os sumários obtidos são constituídos por tomadas formadas por 10 segundos de vídeo que antecedem cada *close* selecionado. A Tabela 4.4 mostra resultado da seleção de *closes* relevantes utilizando 4 estratégias. A coluna 1 mostra o número de *closes* selecionados com a aplicação do algoritmo *K-means*, a coluna 2 mostra o

número de *closes* selecionados com tempo superior a média de todos os *closes* do jogo, a coluna 3 e 4 mostram o número de *closes* selecionados com duração superior a um limiar pré-estabelecido, no caso 5 e 10 segundos respectivamente.

Jogo	Número de <i>closes</i> selecionados para definição dos sumários			
	K-means	Média	5 segundos	10 segundos
Jogo 1	11	26	32	12
Jogo 2	10	36	54	29
Jogo 3	13	32	56	24
Jogo 4	18	28	71	41
Jogo 5	25	44	60	31
Jogo 6	23	35	45	23
Jogo 7	14	39	58	28

Tabela 4.4 : Quantidade de *closes* selecionados por estratégia de corte.

O número de *closes* selecionados representa o número de tomadas selecionadas que assumimos ter maior relevância no vídeo.

A Tabela 4.5 mostra os tempos dos sumários e quanto representa, em termos percentuais, em relação ao tamanho do vídeo completo.

A estratégia usando K-means se mostra bem eficiente em um vídeo de acervo. Já o uso de limiares como 10 segundos, pode ser indicado para aplicações de tempo real nas quais não temos todo o conteúdo do vídeo em um dado momento.

Esses resultados são satisfatórios, pois a duração dos sumários é significativamente inferior a dos vídeos iniciais e todos os eventos relevantes são preservados(gols, faltas, dentre outros).

Jogo / Duração em minutos	Tempo aproximado do sumário em segundos				Tamanho do sumário em relação ao jogo (%)			
	K-means	Média	5 seg.	10 seg.	K-means	Média	5 seg.	10 seg.
Jogo 1 (46 min.)	110	260	320	120	4%	9%	12%	4%
Jogo 2 (45 min.)	100	360	540	290	4%	13%	20%	11%
Jogo 3 (46 min.)	130	320	560	240	5%	12%	20%	9%
Jogo 4 (45 min.)	180	280	710	410	7%	10%	26%	15%
Jogo 5 (48 min.)	250	440	600	310	9%	15%	21%	11%
Jogo 6 (47 min.)	230	350	450	230	8%	12%	16%	8%
Jogo 7 (49 min.)	140	390	580	280	5%	13%	20%	10%

Tabela 4.5 : Comparação de tempos dos vídeos sumários com o vídeo original, para cada estratégia de corte.

5. Conclusão

Apresentaremos nas seções seguintes as contribuições do presente trabalho bem como extensões futuras que poderão ser desenvolvidas a partir do mesmo.

5.1. Resumo e contribuições

Tomando por base algumas soluções mencionadas no Capítulo 1 relacionadas à detecção de eventos em vídeos de futebol, propomos uma solução tendo como uma de suas premissas a simplificação dos processos já existentes, buscando um menor custo computacional, bem como a possibilidade de adaptação para obtenção de resultados próximos do tempo real. Dessa forma podemos analisar um vídeo de futebol ao vivo com um mínimo de retardo de tempo. Para isso, fizemos uso do conceito de ritmo visual, que já nos garante um menor esforço computacional, uma vez que não precisamos tratar os quadros do vídeo em sua totalidade. Foram utilizadas operações de morfologia matemática. Muito difundidas e de fácil aplicação, permitiram a detecção do movimento de câmera através da aplicação na textura do ritmo visual. Utilizamos também abordagens de cores dominantes que foram utilizadas com sucesso na detecção e classificação de tomadas.

O trabalho apresentado nesta dissertação suscita uma nova proposta de abordagem para o problema da sumarização de vídeo de futebol. Algumas heurísticas para extração de informações do jogo, a partir da análise dos resultados obtidos por nossa solução, foram propostas e demonstram apenas algumas das utilidades dos resultados obtidos.

5.2. Trabalhos futuros

A eficácia do sumariador foi evidenciada pelos resultados obtidos nos experimentos. Porém a possibilidade de sua aplicação em aplicações de tempo real para difusão de jogos de futebol não foi avaliada. Várias características, já discutidas, sugerem que os tempos de execução são suficientemente baixos. Além disso, as operações de morfologia matemática são propícias ao processamento paralelo, o que diminuiria ainda mais o tempo de execução.

O poder de síntese do sumariador pode ser aumentado diminuindo o número de falsos positivos na seleção de cenas relevantes. Para tanto, será necessária a classificação refinada das tomadas de close. A adição de novas classes tais como Platéia, Fora de Campo e Arbitragem, por exemplo, permitirá a remoção de tomadas tipo close não relevantes. Nesta nova classificação podemos usar descritores baseados em cores dominantes, considerando agora um número maior de cores com forte conteúdo semântico na partida (uniformes). A Figura 5.1 mostra uma possível classificação.



(a) Jogadores

(b) Platéia

(c) Arbitragem

Figura 5.1 : Classificando closes.

O descritor de movimento de câmera pode ser aprimorado pela adição da análise do ritmo visual vertical. Com isso, será possível não apenas a detecção de movimentos verticais de câmera (*tilt*), mas também de movimentos de zoom que têm conteúdo semântico (*zoom-in* evidencia um evento relevante). A Figura 5.2 mostra exemplo de ritmo visual vertical de um trecho de vídeo onde ocorre um *zoom*. Isso pode ser observado pelo direcionamento da textura que define a ocorrência dos movimentos.



Figura 5.2 : Ritmo visual vertical de um trecho de vídeo contendo um Zoom.

Um desafio na distribuição de vídeos digitais é a adaptação de vídeos para pequenos dispositivos. Telefones celulares e PDA's possuem telas de tamanho e resolução reduzidas. Os vídeos produzidos normalmente para dispositivos maiores, como televisores, precisam ser adaptados para visualização nesses dispositivos. Abordagens que utilizam a simples redução da resolução (*zoom out*) não produzem bons resultados, pois inviabilizam a visualização de pequenos detalhes nas imagens (LIU & GLEICHER, 2006). Um sumário mais efetivo deve selecionar não somente as cenas relevantes, mas também os objetos de maior interesse na cena.

O vídeo pode ser re-enquadrado (*crop*) para mostrar apenas a parte mais relevante da imagem. Consegue-se assim, reduzir ou eliminar o *zoom* e exibir os detalhes pequenos com maior qualidade.

A Figura 5.3(a) mostra um quadro em tamanho normal. As Figuras 5.3(b-d) mostram suas possíveis adequações para que o mesmo seja apresentado em um dispositivo com menor resolução. A Figura 5.3(b) é resultado de um *zoom-in* na figura original. O resultado não é bom, pois os objetos presentes no quadro têm suas dimensões reduzidas, dificultando suas visualizações. A Figura 5.3(c) mostra um re-enquadramento centralizado da figura original. Observamos que alguns detalhes importantes do quadro original são perdidos. A Figura 5.3(d) mostra um re-enquadro da figura original com o cuidado de se preservar os objetos relevantes do quadro original.



(a) Quadro em dimensão normal



(b) Zoom-in



(c) Crop normal



(d) Crop com relevância

Figura 5.3 : Adaptação da visualização de quadros.

A solução proposta ainda pode ser melhorada e estendida. Outros eventos como câmera lenta (PAN & SEZAN,2001) e *zoom* podem ser detectados, os *closes* podem ser subclassificados fornecendo variáveis para o desenvolvimento de novas heurísticas para extração de informações estatísticas e de tomadas importantes. Podemos também, a partir da análise do ritmo visual, detectar os quadros onde temos a rede em evidência, podendo dessa forma detectar lances específicos como um gol.

Há a possibilidade de adequação da solução proposta a outros tipos de vídeo de esportes como basquete, futebol de salão dentre outros.

Bibliografia

- ASSFALG, J., BERTINI, M., BIMBO, A. Del, NUNZIAT, W. & PALA, P. *Soccer highlights detection and recognition using hmm*. IEEE (ICME) Processing International Conference on Multimedia and Expo, 2002.
- ASSFALG, J., BERTINI, M., COLOMBO, C., BIMBO, A. Del & NUNZIAT, W. *Automatic interpretation of soccer video for highlights extraction and annotation*. ACM Symposium on Applied Computing. pp. 769-773 , 2003.
- BAMBERGER, R. & SMITH, M. *A filter bank for the directional decomposition of images: theory and design*. IEEE Transactional Signal Processing, vol. 40, pp. 882-893, 1992.
- BEAUCHEMAN, S. S., & BARRON, J. L. *The computation of optical flow*. ACM Computing Surveys, pp. 433–466, 1995.
- BERTINI, M., BIMBO, A. D. & NUNZIATI , W. *Player identification in soccer videos*. 7th ACM SIGMM international workshop on Multimedia information retrieval, pp. 25–32, 2005.
- BEZERRA, F. N. *A longest common subsequence approach to detect cut and wipe video transitions*. In Proceedings of Brazilian Symposium on Computer Graphics and Image Processing, pp. 154–160. Brazilian Computer Society, IEEE CS Press., 2004.
- BEZERRA, Francisco Nivando & LIMA, Elan. *Low cost soccer vídeo summaries based on visual rhythm*. ACM international workshop on Multimedia information retrieval - MIR, 2006
- BEZERRA, Francisco Nivando & LEITE, Neucimar Jerônimo. *Video transition detection using string matching: preliminary results*. XVI Simpósio Brasileiro de Computação Gráfica e Processamento de Imagens - São Carlos, SP, Brasil ,2003.
- BIMBO, A. D. *Visual information retrieval*. Morgan Kaufmann, 1999.
- BROECKER, L.,BOGEN, M. & CREMERS, A. B. *Bridging the semantic gap in content-based image retrieval systems*. Processing SPIE Internet Multimedia Management Systems II pp. 54-62, 2001.
- CHUNG, M.G. *Automatic video segmentation based on spatio-temporal features*. Korea Telecommunication Journal, pp. 4-14, 1999.
- D'ORAZIO, T., ANCONA, N., CICIRELLI, G. & NITTI, M. *A Ball detection algorithm for real soccer image sequences*. IEEE(ICPR) International Conference on Pattern Recognition, pp. 10210, 2002.
- DAUGMAN, J. *Uncertainty Relation for Resolution in Space, Spatial Frequency, and Orientation Optimized by Two-Dimensional Visual Cortical Filters*. Optical Society, vol. 2, no. 7, pp. 1160-1169, 1985.

DAVIES, FR. *Vectorial Strategy for Designing Line Segment Detectors with High Orientation Strategy*. Electronic Letters, vol. 33, no. 21, pp. 1774-1777, 1997.

EKIN, A. & TEKALP, A. M. *A Framework for analysis and tracking of soccer video*. SPIE (VCIP) Visual Computing and Image Processing, 2002.

EKIN, Ahmet & TEKALP, A. Murat. *Generic event detection in sports video using cinematic feature*. IEEE Workshop on Event Mining, pp. 17-24, 2003.

EKIN, Ahmet, TEKALP, A. Murat & MEHROTRA, Rajiv. *Automatic soccer video analysis and summarization*. IEEE Transaction on Image Processing, pp. 796-807, 2003.

FIGUEROA, Pascual, LEITE, Neucimar, BARROS, Ricardo M. L., COHEN, Isaac & MEDIONI, Gerard. *Tracking soccer players using the graph representation*. IEEE, 2004.

FOLEY J., DAM Van A., FEINER S. & HUGHES J. *Computer Graphics: Principles and Practice*. Addison Pub, 1996.

FREEMAN, William T. & ADELSON, Edward H. *The Design and Use of Steerable Filters*. IEEE Transactions on Pattern Analysis and Machine Intelligence, v.13 n.9, pp. 891-906, 1991

GONG, Yihong, SIN, Lim Teck, CHUAN, Chua Hock, ZHANG Hongjiang & SAKAUCHI, Masao. *Automatic parsing of TV soccer programs*. IEEE International Conference on Multimedia Computing and Systems, 1995.

GOOGLE, *Pesquisa de imagens*. Disponível em: <<http://images.google.com/>>. Acesso em : novembro/2006.

GUIMARÃES, S. J. F., COUPRIE, M., A. A. & LEITE, N. J. *Video segmentation based on 2D image analysis*. Pattern Recognition Letters, pp. 947–957, 2003.

GUIMARAES, Sílvio Jamil Fazoli, COUPRIE, Michel, LEITE, Neucimar Jerônimo e ARAUJO, Arnaldo de Albuquerque. *Análise de vídeo a partir da sua simplificação*. Anais do II workshop em tratamento de imagens(NPDI/DCC/UFMG), 2001.

GUIMARAES, Sílvio Jamil Fazoli, COUPRIE, Michel, LEITE, Neucimar Jerônimo & ARAUJO, Arnaldo de Albuquerque. *A Method for cut detection base on visual rhythm*. IEEE CS Press., 2001.

JEULIN, D. & KURDY, M. *Directional Mathematical Morphology for Oriented Image Restoration and Segmentation*, Acta Stereologica, vol. 11, pp. 545-550,1992.

KIM, H., LEE, J., YANG, J., KIM,W., SULL, S. & SONG, S. M. *Visual rhythm and shot verification*. Multimedia Tools and Applications, pp. 227–245, 2001.

LI, Baoxin & SEZAN, M. Ibrahim. *Event detection and summarization in sports video*. IEEE(ANSS), 2003.

- LIU, Feng & GLEICHER, Michael. *Video retargeting: Automating pan and scan*. ACM international workshop on Multimedia information retrieval – MIR. ACM Press., 2006
- MANJUNATH, B.S., OHM, J.R., VASUDEVAN, V.V. & YAMADA, A. *Color and texture descriptors*. In *special issue on MPEG-7*, IEEE Transaction on circuits and systems for video technology, pp. 703-715, 2001.
- NAGASAKA , A. & TANAKA Y. *Automatic video indexing and full video search for object appearances*. IFIP Transactions, Visual Database Systems II, pp. 113 -127, 1992.
- NGO, Chong-Wah, PONG, Ting-Chuen & ZHANG, Hong-Jiang. *Motion analysis and segmentation through spatio-temporal slices processing*. IEEE CS Press., 2003.
- NGO,C. W., PONG,T. C. & CHIN, R. T. *Detection of gradual transitions through temporal slide analysis*. In IEEE CVPR, pp. 36–41, 1999.
- OSIAN, M. & GOOL, L. V. *Video shot characterization*. In TREC Video Retrieval Evaluation. NIST, 2003.
- PAN, H., LI, B. & SEZAN, M. I. *Detection of slow-motion replay segments in sports video for highlights generation*. IEEE (ICASSP), 2001.
- QBIC, Query by image content. Disponível em: <<http://www.qbic.almaden.ibm.com>>. Acesso em: novembro/2006.
- SERRA, J. *Image Analysis and Mathematical Morphology, Volume 1*. Academic Press, 1988.
- SOILLE, P. *Morphological image analysis: principles and applications*. Springer, second edition, 2003.
- SOILLE, Pierre & TALBOT, Hugues. *Image structure orientation using mathematical morphology*. IEEE CS Press., 1998.
- SOILLE, Pierre, BREEN, Edmond J. & JONES, Ronald. *Recursive Implementation of Erosions and Dilations Along Discrete Lines at Arbitrary Angles*. IEEE Transactions on Pattern Analysis and Machine Intelligence , pp. 562-567,1996.
- SOILLE,P. & TALBOT, H. *Directional morphological filtering*. IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1313–1329, 2001.
- TOVINKERE, V. & QIAN, R. J. *Detecting semantic events in soccer games: toward a complete solution*. IEEE(ICME), 2001.
- VICENT L., *Morphological Grayscale Reconstruction: Definition, Efficient Algorithms and Applications in Image Analysis*. *Computer Vision and Pattern Recognition*, pp. 633-635, Champaign IL, 1992.

- WALKER, T., HEUER, J. & MARTINEZ, J. M. Introduction to MPEG-7: Multimedia content description interface, cap. 7, 2002.
- WAN, K.W., YAN, X., YU, X. & XU, C. *Real-time goal-mouth detection in MPEG soccer video*. ACM Multimedia, pp. 311-314, 2003.
- WANG, Lei, LEW, Michael & XU, Guargyou. *Offense based temporal segmentation for event detection in soccer video*. ACM Press., 2004.
- WHITEHEAD, A., BOSE, J. & LAGANIERE, R. *Feature based cut detection with automatic threshold selection*. In Proceeding of the International Symposium on Signal Processing and Applications, pp. 410–418, Paris, 2004.
- XIE, L., CHANG S_F., DIVAKARAN, A. & SUN, H. *Structure analysis of soccer video with hidden Markov Models*. IEEE (ICASSP), 2002.
- XU, Peng, XIE, Lexing, CHANG, Shih-Fu, DIVAKARAN, Ajay, VETRO, Avithony & SUN, Huifang. *Algorithms and system for segmentation and structure analysis in soccer video*. IEEE CS Press., 2001.
- YAMADA, H. , YAMAMOTO, K. & HOSOKAWA, K. *Directional Mathematical Morphology and Reformalized Hough Transformation for the Analysis of Topographic Maps*. IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 380-387, 1993
- YOW, D., YEO, B.L., YEUNG, M. & LIU, G. *Analysis and presentation of soccer highlights from digital video*. ACCV, 1995.
- YU, X., XU, C., LEONG, H.W., TIAN, Q., TANG, Q. & WAN, K.W. *Trajectory-based ball detection and tracking with applications to semantic analysis of broadcast soccer video*. ACM Multimedia, pp. 11-20, 2003.
- ZABIH, R., MILLER, J. & MAI, K. *A feature-based algorithm for detecting and classifying scene breaks*. In Proceeding of the 3rd International ACM International Conference on Multimedia, pp. 189–200, 1995.
- ZHANG, H. J. & SMOLIAR, S. W. *Developing power tools for video indexing and retrieval*. Storage and Retrieval for Image and Video Databases II, volume 2185 of SPIE Proceedings Series, pp. 140–149, 1994.
- ZHANG, H. J., KANKANHALLI, A., & SMOLIAR, S. *Automatic partitioning of full motion video*. Multimedia Systems, pp. 10–28, 1993.
- ZHANG, H. J., LOW, C. Y., GONG, Y. & SMOLIAR, S. *Video parsing using compressed data*. SPIE Image and Video Processing II, volume 2172, pp. 142–149, 1994.

Apêndice I

Gráficos de tomadas dos 10 primeiros minutos dos jogos utilizados nos experimentos.

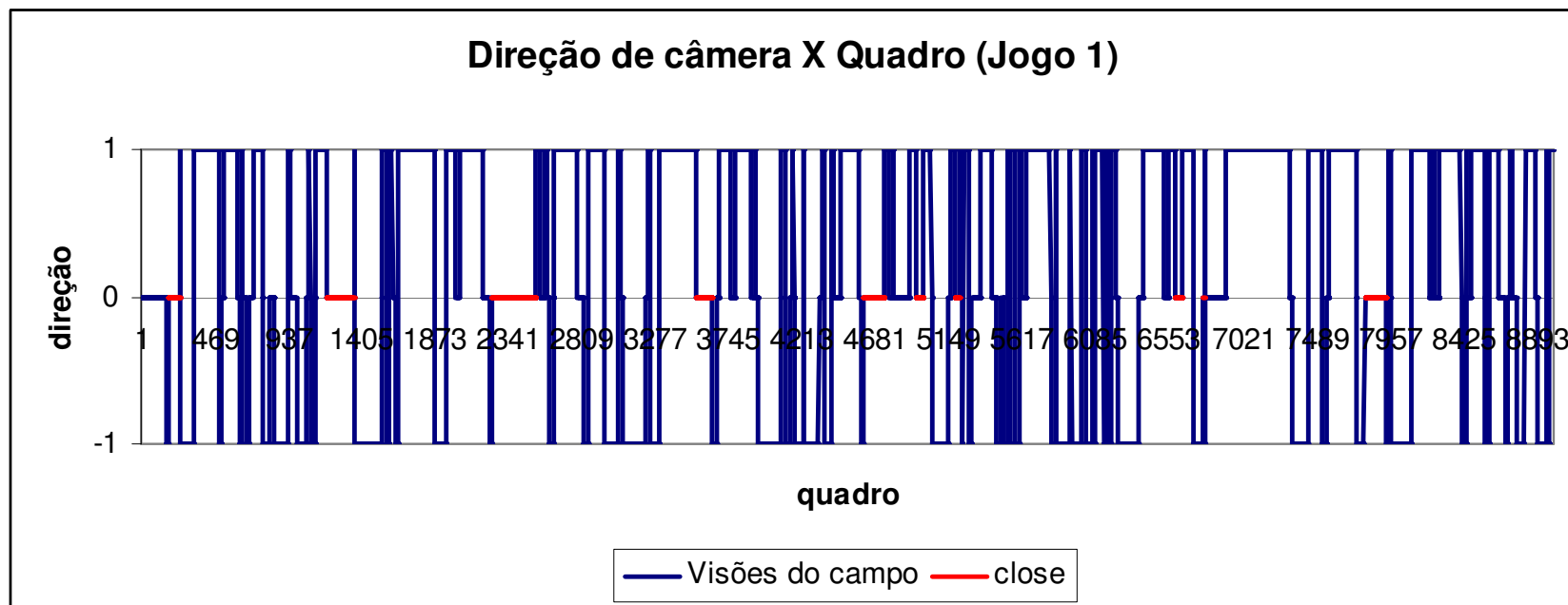


Figura I.1 : Gráfico de tomadas dos 10 minutos iniciais do jogo 1

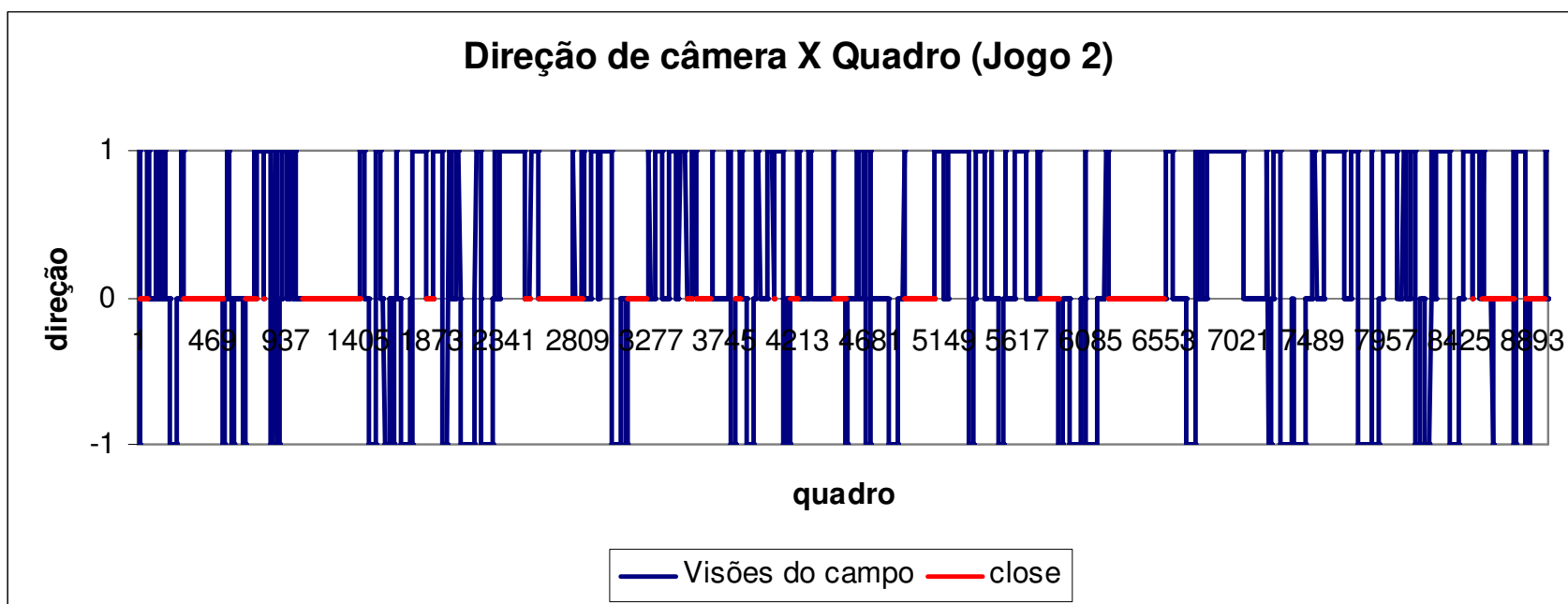


Figura I.2 : Gráfico de tomadas dos 10 minutos iniciais do jogo 1

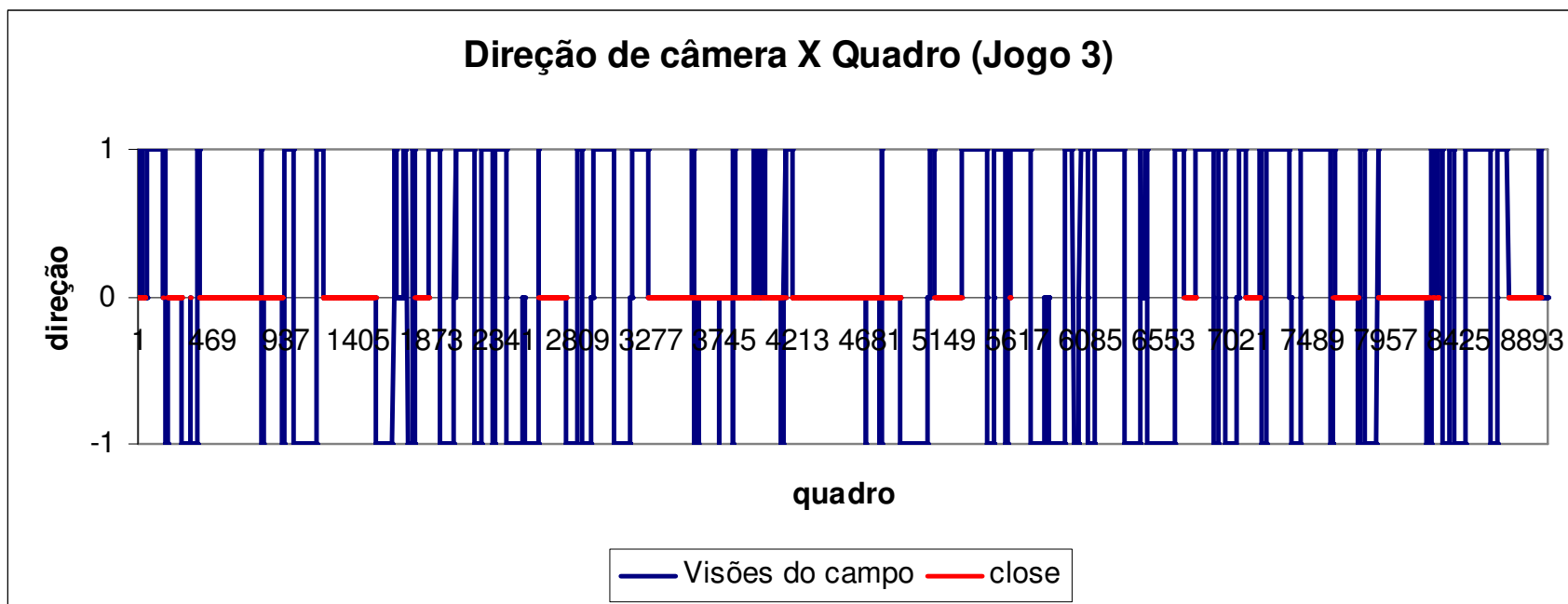


Figura I.3 : Gráfico de tomadas dos 10 minutos iniciais do jogo 1

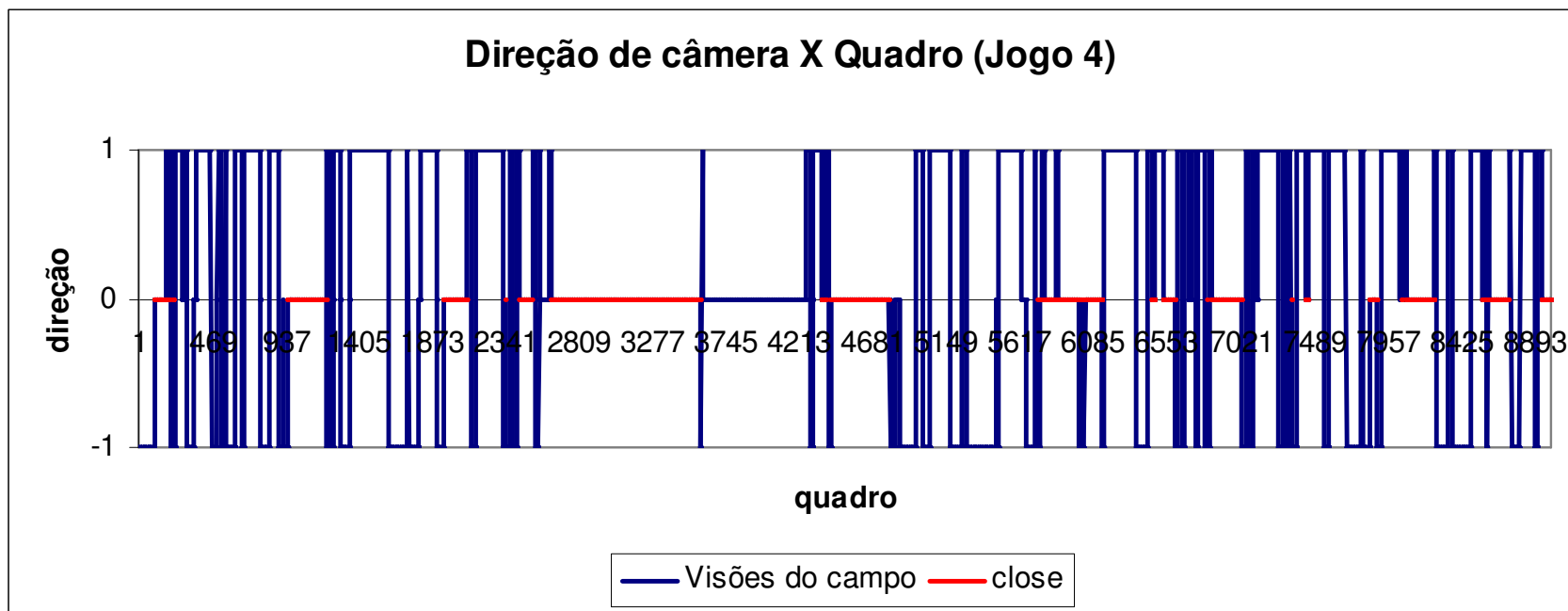


Figura I.4 : Gráfico de tomadas dos 10 minutos iniciais do jogo 1

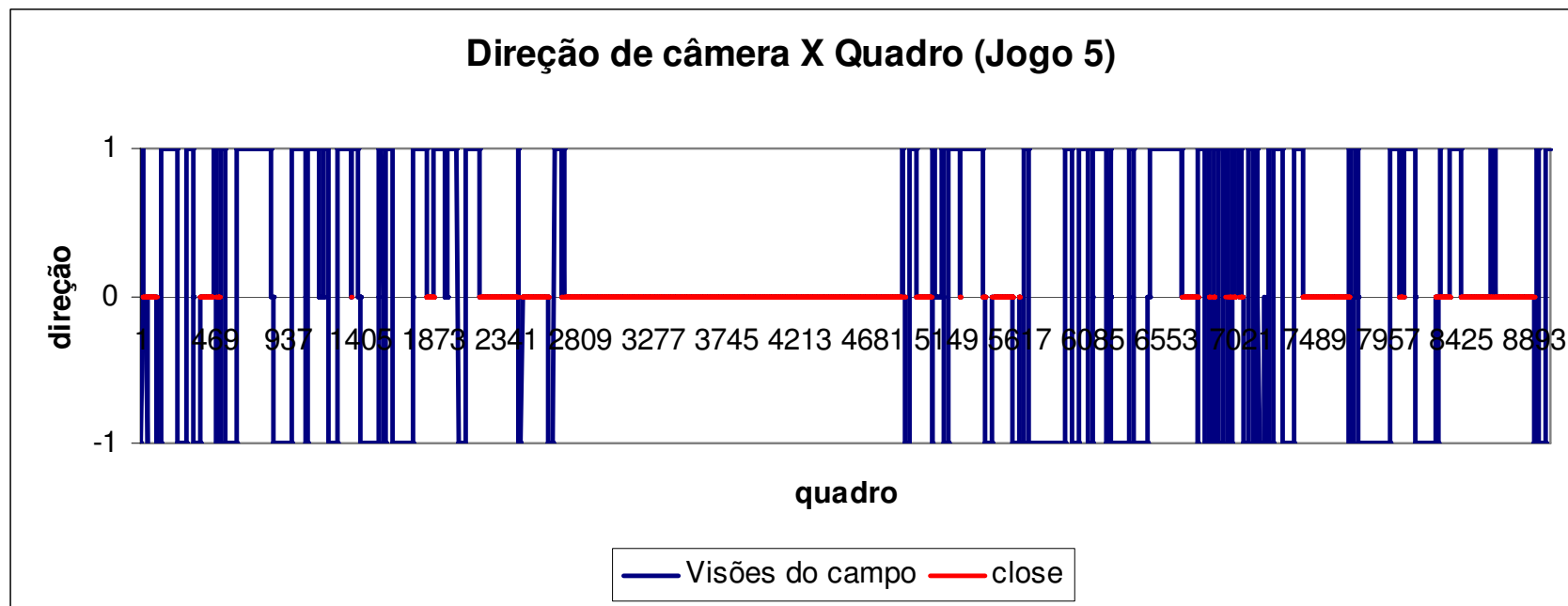


Figura I.5 : Gráfico de tomadas dos 10 minutos iniciais do jogo 1

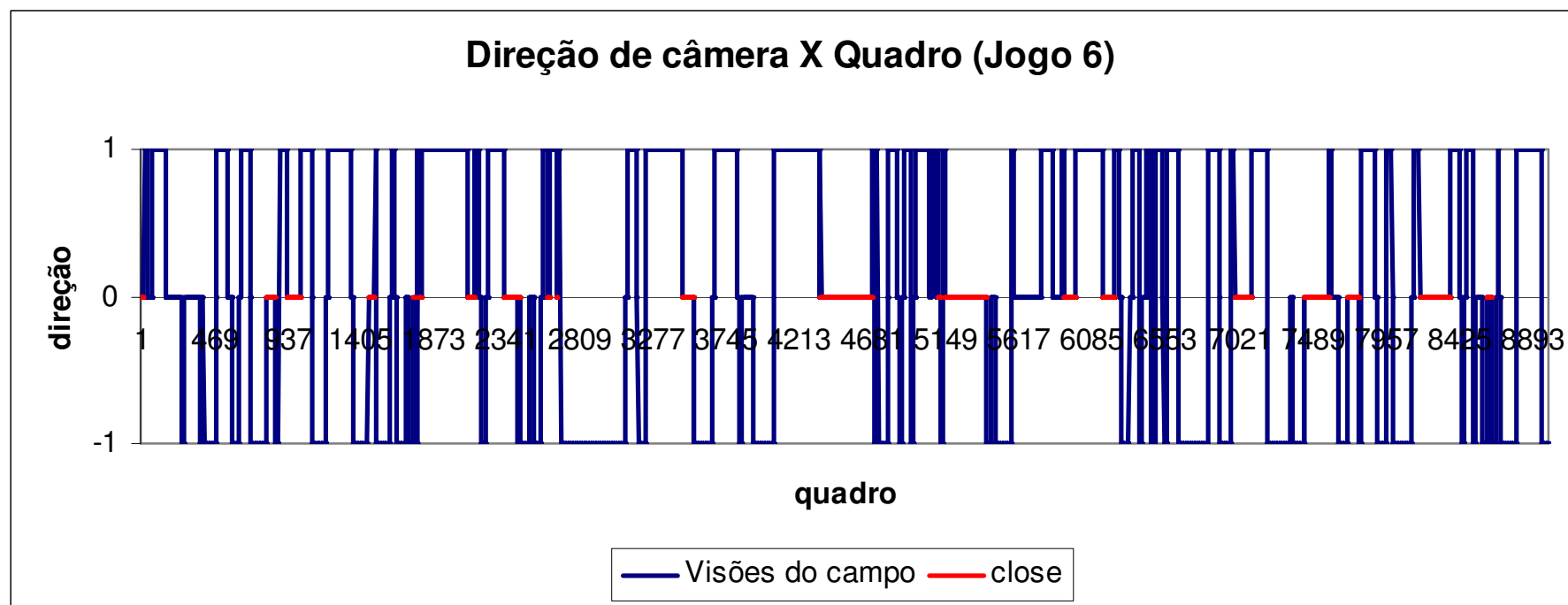


Figura I.6 : Gráfico de tomadas dos 10 minutos iniciais do jogo 1

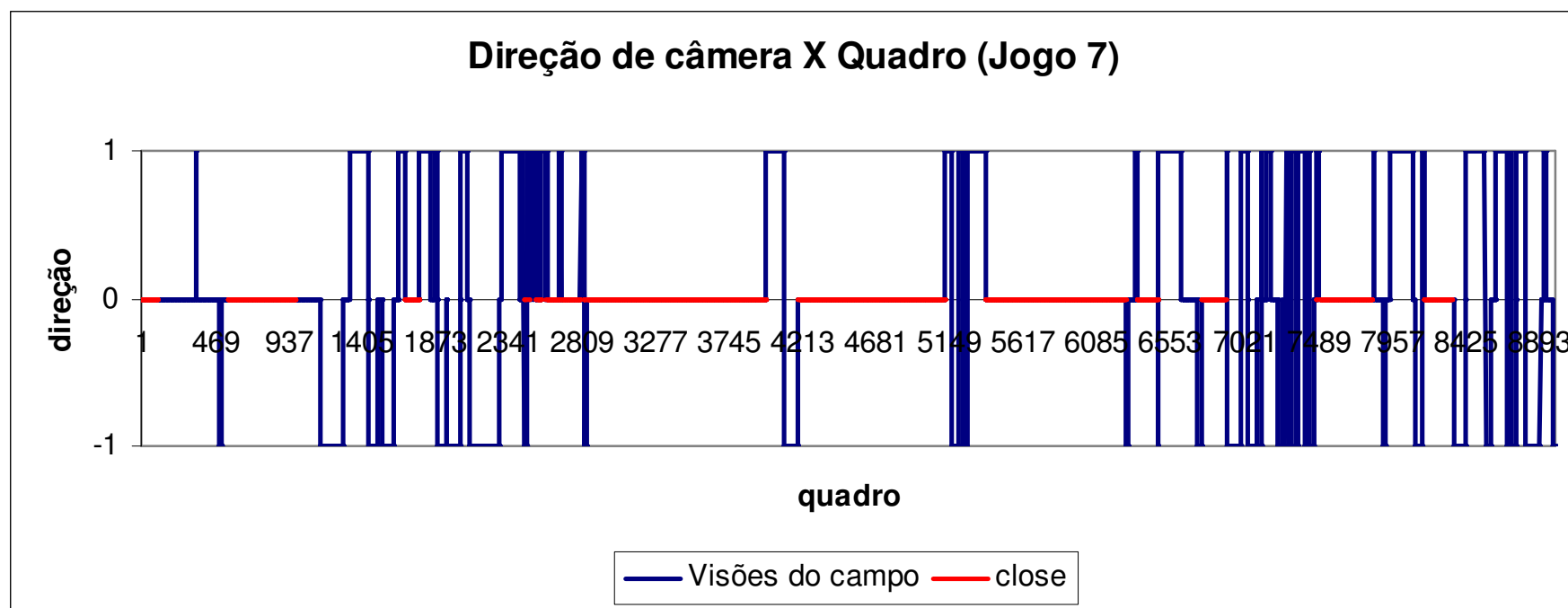


Figura I.7 : Gráfico de tomadas dos 10 minutos iniciais do jogo 1

Apêndice 2

Quadros de um vídeo gerado pela aplicação contendo informações de direcionamento de câmera.

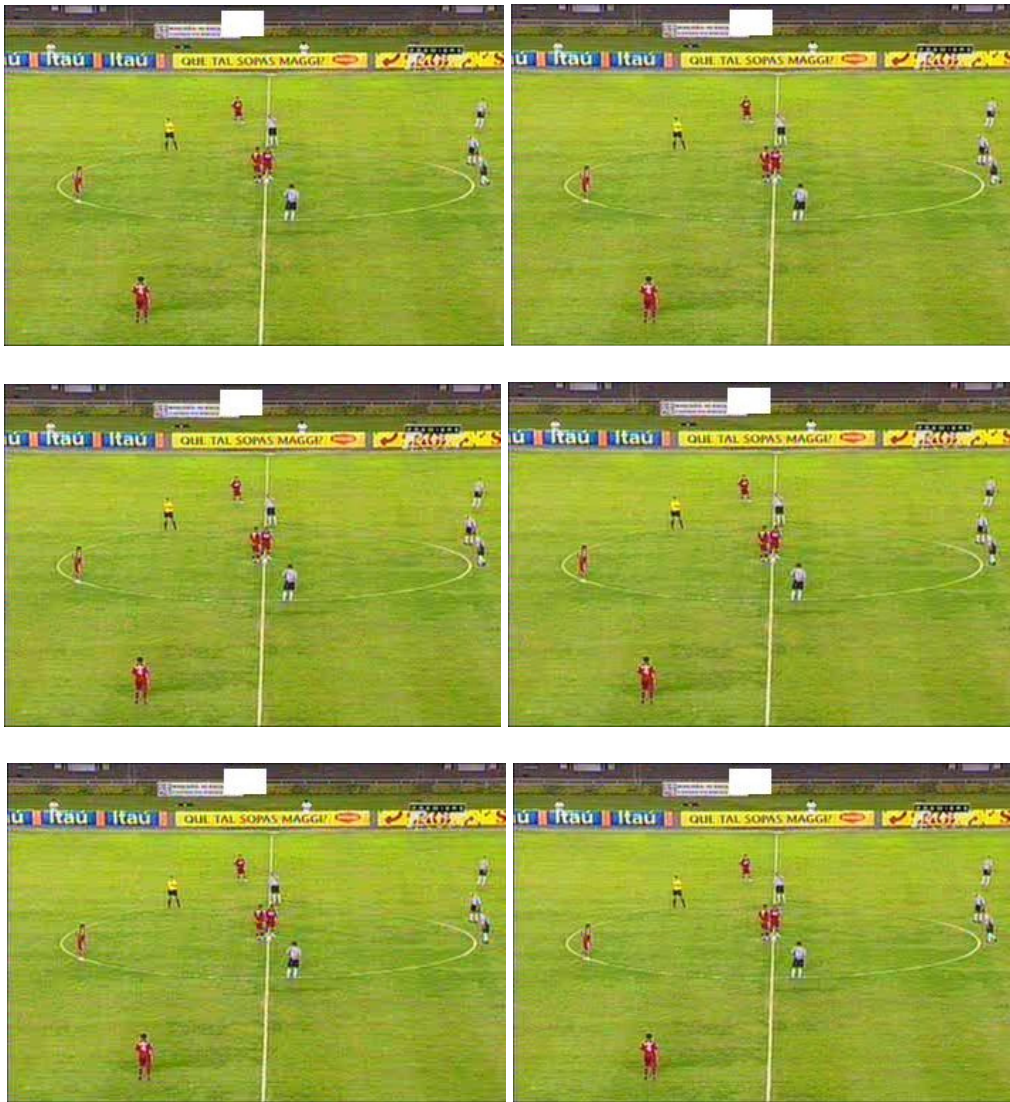


Figura II.1 : Sequência de quadros sem movimento.



Figura II.2 : Seqüência de quadros com closes.



Figura II.3 : Seqüência de quadros com movimento para direita.

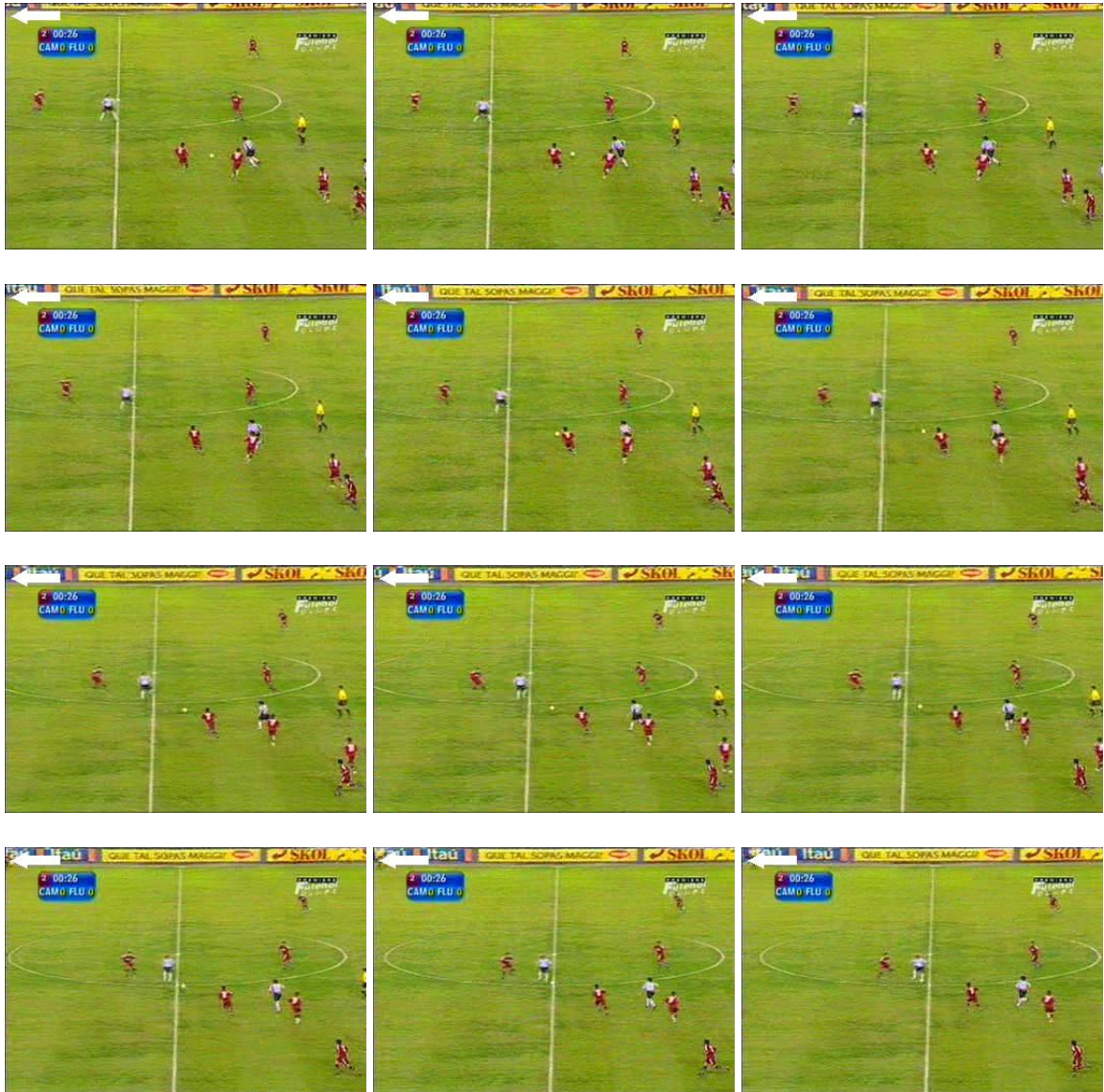


Figura II.4 : Sequência de quadros com movimento para esquerda.

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)