

CENTRO UNIVERSITÁRIO DA FEI

MAURÍCIO KENITI HIROTA  
FABRÍCIO LOPES DE SOUZA  
IVAN SHIGUENORI MACHIDA  
THIAGO MIZUTANI

Orientador: Prof. D.Sc. Paulo Sérgio Silva Rodrigues

**M-FIT Motion Finder Integration :  
Editoração Eletrônica de Vídeos**

São Bernardo do Campo  
2008

# CENTRO UNIVERSITÁRIO DA FEI

MAURÍCIO KENITI HIROTA  
FABRÍCIO LOPES DE SOUZA  
IVAN SHIGUENORI MACHIDA  
THIAGO MIZUTANI

Orientador: Prof. D.Sc. Paulo Sérgio Silva Rodrigues

## **M-FIT Motion Finder Integration : Editoração Eletrônica de Vídeos**

Trabalho de Conclusão de curso, apresentado no Centro Universitário da FEI, como parte dos requisitos necessários para obtenção do título de Bacharel em Ciências da Computação, orientado pelo professor Prof. D.Sc. Paulo Sérgio Silva Rodrigues.

São Bernardo do Campo

2007

# Resumo

Na área de Editoração de Vídeos, uma das maiores demandas é a detecção de transição de tomadas, que é basicamente o corte de uma tomada e o início de outra, o que é útil para diversas aplicações, como indexação, recuperação, análise e edição de vídeos. Estas transições podem ocorrer de diversas formas, tais como: zooms, cortes, fades, dissolves. Este trabalho apresenta o M-FIT Motion Finder Integration, um framework que possibilitará o reconhecimento automático das transições de tomadas de cenas a partir de uma determinada amostragem de imagens (frames de um filme) através do uso de técnicas de Ritmo Visual e Morfologia Matemática, permitindo que o ponto exato da transição seja obtido de uma maneira mais prática e rápida. O M-FIT é uma ferramenta desenvolvida em linguagem C, com utilização de OpenCV (biblioteca para computação visual). Sua interface é feita com a utilização da ferramenta QT, tornando o M-FIT um sistema com uma interface simples e de fácil entendimento, a qual terá diversos atalhos para outras funcionalidades do sistema.

# Conteúdo

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Objetivo . . . . .	2
<b>2</b>	<b>Trabalhos Relacionados</b>	<b>3</b>
2.1	Detecção de Transições . . . . .	3
2.2	Técnicas de PDI . . . . .	5
<b>3</b>	<b>Revisão Bibliográfica</b>	<b>7</b>
3.1	Edição de Vídeo . . . . .	8
3.1.1	Fade-in . . . . .	8
3.1.2	Fade-out . . . . .	8
3.1.3	Corte . . . . .	9
3.1.4	Dissolve . . . . .	9
3.2	Sistemas de Vídeo . . . . .	10
3.3	Sistemas de Cor . . . . .	12
3.3.1	RGB . . . . .	12
3.3.2	CMYK . . . . .	13
3.3.3	HSV . . . . .	15
3.3.4	Highcolor . . . . .	16
3.4	Detecção de Transições . . . . .	17
3.5	Ritmo Visual . . . . .	18
3.6	Processamento de Baixo Nível . . . . .	21
3.6.1	Filtragem Espacial . . . . .	21
3.6.2	Detecção de Bordas . . . . .	23
3.7	Processamento de Médio Nível . . . . .	25
3.7.1	Morfologia Matemática . . . . .	25
3.7.2	Segmentação de Imagens . . . . .	28
3.8	Processamento de Alto Nível . . . . .	31
3.8.1	Método de Casamento de Padrões . . . . .	31
3.8.2	Método Baseado em Entropia . . . . .	32
<b>4</b>	<b>Metodologia</b>	<b>33</b>

# Lista de Figuras

3.1	Exemplo de fade-in em uma sequencia de frames. Como pode ser observado não se sabe exatamente onde ocorre a transição. . . . .	8
3.2	Exemplo de Fade-out em uma sequência de frames . . . . .	8
3.3	Exemplo de corte em uma sequência de frames. . . . .	9
3.4	Exemplo de Dissolve em sequência de frames. . . . .	9
3.5	Sistema de Codificação e Decodificação de um vídeo no formato MPEG. . .	10
3.6	Sistema de armazenamento AVI. . . . .	10
3.7	Sistema RGB. . . . .	12
3.8	Misturas de cores no padrão CMYK. . . . .	14
3.9	A mistura de cores representada por um círculo. . . . .	14
3.10	Modelo de representação do padrão HSV. . . . .	15
3.11	Transformação da diagonal principal dos frames de um vídeo em colunas do Ritmo Visual. . . . .	18
3.12	Representações de Ritmo Visual : cortes . . . . .	19
3.13	Representações de Ritmo Visual : fade-in . . . . .	19
3.14	Representações de Ritmo Visual : dissolves . . . . .	19
3.15	Máscara 3x3 sendo utilizada para a filtragem de uma imagem digital. No caso, o ND a ser recalculado é do pixel X. . . . .	21
3.16	Imagem Original . . . . .	22
3.17	Imagem Após Filtragem Passa-Baixa . . . . .	22
3.18	Imagem Após Filtragem Passa-Alta . . . . .	22
3.19	Imagem Original. . . . .	24
3.20	Imagem após sofrer filtragem com operador se sobel. . . . .	24
3.21	Cruz. . . . .	25
3.22	Quadrado. . . . .	25
3.23	Hexagonal. . . . .	25
3.24	Imagem Original. . . . .	26
3.25	Resultado da aplicação da operação de erosão na imagem. . . . .	26
3.26	Imagem Original. . . . .	26
3.27	Resultado da aplicação da operação de dilatação na imagem. . . . .	27
3.28	Exemplo de limiar com histograma. . . . .	28
3.29	Imagem Original. . . . .	29
3.30	Limiar = 10. . . . .	29
3.31	Limiar = 30. . . . .	29
3.32	Limiar = 70. . . . .	30
4.1	Sequência de processos utilizados na metodologia. . . . .	33
4.2	Obtenção do Ritmo Visual. . . . .	34

---

4.3	Filtragem para diminuição de ruídos. . . . .	34
4.4	Detecção das bordas. . . . .	35
4.5	Mapa de bordas. . . . .	35
4.6	Sobreposição do mapa de bordas na imagem do ritmo visual, dando destaque às transições . . . . .	35

# Capítulo 1

## Introdução

O setor de entretenimento está sempre em constantes mudanças e inovações. Isto se deve ao fato do dele estar diretamente ligado a muitos outros setores, como o setor de marketing e cinema. Devido a isto, o número de novos filmes, propagandas, vêm crescendo rapidamente. Entre uma das tarefas mais demandadas da indústria da produção de vídeo, está a edição, que consiste na transformação de um vídeo através da aplicação de filtros e aplicação de efeitos como, por exemplo, ajustes de cor, brilho, contraste e remoção de objetos. Com isso a edição de vídeo torna-se uma atividade com um mercado amplo e que gera uma boa renda anualmente. Um vídeo normalmente é formado por um conjunto de sub-vídeos, que são conhecidos como tomadas. Em uma pré-edição, esses sub-vídeos são agrupados formando o vídeo completo. Todas estas tomadas são interligadas através de efeitos chamados transições, as quais podem ser de vários tipos, desde um simples corte, até uma transição mais complexa como fade ou dissolve. Estas transições são inseridas no vídeo com o objetivo de suavizar as trocas de tomadas. O problema de uma edição de vídeo é que existem várias tomadas em um único vídeo, então se o editor desejar ajustar o brilho de uma única tomada, por exemplo, terá que delimitar o seu início e fim manualmente. Sendo assim, a detecção e localização de transição é uma das tarefas mais importantes no processo de indexação e edição de vídeo digital. Este processo é trabalhoso e demorado, tornando necessária cada vez mais a produção de sistemas automáticos para detecção de transições. Neste trabalho será apresentado um sistema de detecção automática de transições de vídeo baseado em ritmo visual e morfologia matemática.

## 1.1 Objetivo

Desenvolver um sistema automático de detecção de transição de tomadas e edição de vídeo.



# Capítulo 2

## Trabalhos Relacionados

### 2.1 Detecção de Transições

Uma das técnicas mais comuns e importantes na área de editoração de vídeos é a detecção de transições.

Silvio J. F. Guimarães et al., 2003, propuseram um método para fazer a detecção de transições em um vídeo, aplicando o operador gradiente multi-escalar morfológico em cima de uma determinada amostragem do vídeo. Através deste método, puderam obter bons resultados através do gradiente multi-escalar em relação à detecção de transições graduais, porém o método não foca na dimensão da transição.

Silvio J. F. Guimarães et al., 2004, propuseram um método para detectar transições graduais (fade,dissolve) utilizando ritmo visual e morfologia matemática . O foco principal é usar operadores de morfologia matemática para aumentar os vestígios de transições na imagem do ritmo visual, não alterando os já explícitos. Com a aplicação deste método obtiveram uma taxa de acerto de 93.5

Silvio. J. F. Guimarães et al., 2002, com o objetivo de realizar a identificação e classificação de transições graduais entre duas zonas planas consecutivas (conjunto de pontos consecutivos com os mesmos tons de cinza), aplicaram um método baseado na decomposição hierárquica dos componentes da imagem em resíduos morfológicos. Com o uso deste método era possível retirar as informações necessárias sobre a transição e ainda fazer a reconstrução exata da imagem a partir dos seus resíduos. Desta forma, conseguiram fazer a detecção e identificação das transições facilmente. Apesar de o método funcionar bem em relação às transições graduais, o trabalho não faz referência às transições bruscas

(cortes).

Silvio Jamil Ferzoli Guimarães et al., 2001, desenvolveram um método para a detecção de cortes de vídeos digitais, a partir da aplicação de ritmo visual. Com este método, conseguiram obter com sucesso as transições de corte de um vídeo. A partir deste estudo, é possível desenvolver métodos para detecções de outros tipos de transições.

Nielsen Cassiano Simões, 2004, apresentou um método para detecção de transições abruptas (cortes e flashes) em vídeos, baseado em algoritmos de identificação de padrão em sinais 1D. O método proposto possibilitou a detecção de transições abruptas, ficando aberto a novas propostas para a detecção de outros tipos de transição.

C. W. Ngo et al., 1999, com a intenção de fazer a detecção de transições do tipo corte, wipe e dissolve, propuseram um método baseado em Ritmo Visual. Com a aplicação deste método, foi possível fazer a detecção e classificação das transições de uma maneira eficiente. Ressalta-se ainda que futuramente, pretende-se fazer uma análise de takes de vídeos para fazer indexação de vídeos e recuperação de dados.

Silvio J. F. Guimarães et al., 2002, propuseram um método que faz uso de Ritmo Visual (transformação dos eventos do vídeo em padrões de imagens 2D), juntamente com a aplicação de operadores de Morfologia Matemática e Topologia Digital para fazer a detecção e identificação dos eventos de transição de um vídeo. Aplicando este método, puderam fazer a detecção de fades curtos com uma baixa taxa de erros, sendo possível ainda desenvolver uma extensão do método para detecção de outros tipos de transições, assim como fazer a aplicação do método diretamente nos dados do vídeo.

Sébastien Stevie et al, 2003, apresentaram uma gama de algoritmos referentes a identificação de transições em vídeos sem compressão utilizando métodos baseados em: pixel, histograma, bloco, características, movimento, entre outros. O objetivo era analisar o custo computacional de cada algoritmo e compará-los. Concluiu-se que os algoritmos que utilizam lógicas simples geralmente possuem um custo computacional baixo, porém dispõem de uma qualidade menor nas detecções. Em contrapartida, algoritmos que utilizam lógicas mais complexas, como a análise de movimento e característica, ou um método que use vários algoritmos, conseguem ter como resultado final com boa qualidade em termos de detecção de transição. Porém, possuem um custo computacional mais alto.

Guimarães, S. J. F. et al, 2001, propuseram um método que transforma o problema de segmentação de vídeo em um problema de detecção de padrões , onde cada efeito de

vídeo é transformado em diferentes padrões de ritmo visual. Obtiveram bons resultados em comparação à outros métodos em relação a robustez, falhas e medidas de gama, porém foram piores que outros métodos em relação a quantidade de detecções quando não há nenhuma falsa-deteção (falseless error).

Silvio Jamil Ferzoli Guimarães et al, 2002, propuseram um método para detecção de transições parametrizadas e direcionais, baseado em resíduos morfológicos considerando elementos linearmente estruturados. Com o método proposto, puderam identificar facilmente as transições, fato que não é possível, por exemplo, com o método que utiliza o gradiente Soille [7].

Com o objetivo de identificar o limite entre tomadas consecutivas de um determinado segmento de vídeo, S. J. F. Guimarães et al., 2002, utilizaram um método baseado no ritmo visual e Morfologia matemática, onde cada evento de vídeo é representado por um padrão em uma imagem 2D. Através da aplicação destes métodos, puderam identificar facilmente as transições de um segmento de vídeo, porém foram somente aplicadas aos efeitos de transições fades, cortes e dissolves.

Com o objetivo de identificar cortes de uma cena utilizando um único algoritmo, Jianmin Jiang et al., 2005, propuseram a aproximação da lógica Fuzzy para integrar funcionalidades híbridas para a detecção de cortes de uma cena dentro de um vídeo aleatório. A aproximação da lógica Fuzzy contém dois modos de processamento, onde um é dedicado à detecção de cortes bruscos de cenas e também transições dissolves, e o outro para transições graduais. Estes dois modos são unificados por um seletor de modos que decide qual modo o esquema deve trabalhar para atingir o melhor desempenho de detecção, sendo que não se aplicam a transições flash.

## 2.2 Técnicas de PDI

Para realizar a detecção de transições ou qualquer tipo de editoração de vídeos, são necessárias técnicas de PDI como Morfologia Matemática, Entropia, Filtragem Espacial, Detecção de Bordas, entre outras.

Com o objetivo de segmentar imagens, Gilles Bertrand et al., 1999, utilizaram transformações de níveis e homotópicas (estudos dos pontos de cruzamento da topologia e reconstrução da transformação). Utilizando estas técnicas conseguiram segmentar ima-

gens em tons de cinza, sendo que estas não se aplicam a imagens RGB.

Com o objetivo de analisar a morfologia de imagens em 3D, K. Michielsen et al., 2000, utilizaram o corte dos lados de um cubo 3D e aplicaram a morfologia matemática e a tripla periódica mínima que verifica a continuidade de uma superfície. Com estas técnicas, conseguiram caracterizar a morfologia de uma imagem 3D em tons de cinza, sendo que não se aplicam a imagens em RGB.

Com o objetivo de detectar cortes de cenas e otimizar o custo de tempo humano, Jun Yu et al., 2000, utilizaram a diferença de métrica baseado no histograma de cores de sucessivos frames de uma sequência de vídeo utilizando a entropia de um threshold, onde o threshold é a base da diferença dos histogramas de tons de cinza de um determinado frame. Por outro lado, para confirmar os resultados, são necessárias interações humanas e um procedimento de eliminação de falsa detecção futuramente será desenvolvido para melhorar a precisão.

## Capítulo 3

### Revisão Bibliográfica

Nesta seção, serão definidos os principais conceitos para o entendimento de todo o projeto e a metodologia do sistema proposto.

## 3.1 Edição de Vídeo

A edição de vídeos consiste em qualquer operação que modifique o conteúdo original de um vídeo. Tais operações podem consistir desde a segmentação do vídeo para futuras modificações de indexação ou armazenamento, até a aplicação de efeitos visuais, como por exemplo, uma mudança de cores, o nivelamento da luminosidade entre frames de uma tomada, eliminação de ruídos, equalização de som com imagem, entre outros, os quais são intensamente utilizados no setor cinematográfico, de pesquisa e entretenimento. É comum organizar a edição de vídeos através de tomadas, as quais são divididas por transições. Portanto, é importante que seja feita a detecção destas transições antes da edição do vídeo para que um determinado efeito aplicado ou qualquer outra operação de edição possa ser delimitado às tomadas específicas e não para o vídeo todo.

### 3.1.1 Fade-in

O fade-in é uma transição relativamente lenta e suave entre o preto e uma imagem qualquer. Este tipo de transição geralmente é utilizado no início do vídeo ou no início de um bloco de tomadas.

Podemos visualizar o fade-in na figura 3.1:

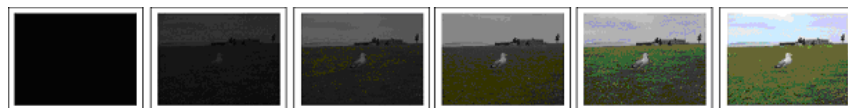


Figura 3.1: Exemplo de fade-in em uma sequência de frames. Como pode ser observado não se sabe exatamente onde ocorre a transição.

### 3.1.2 Fade-out

Assim como o fade-in, o fade-out também é uma transição relativamente lenta e suave, porém com um efeito oposto. Um exemplo desta transição pode ser observado na Figura 3.2:

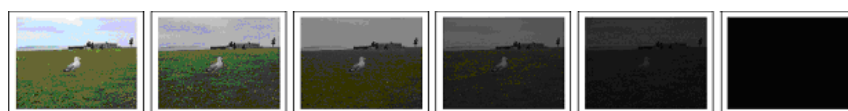


Figura 3.2: Exemplo de Fade-out em uma sequência de frames

### 3.1.3 Corte

Cortes são transições bruscas e secas entre tomadas, onde o último frame de uma determinada tomada é concatenado ao primeiro frame da tomada seguinte diretamente. A Figura 3.3 mostra a ocorrência de um corte:



Figura 3.3: Exemplo de corte em uma sequência de frames.

Note que no exemplo da Figura 3.3 é possível se ver claramente que a transição ocorre entre o terceiro e quarto frame.

### 3.1.4 Dissolve

Dissolves O dissolve é uma transição que mistura fade-out e fade-in, isto é, a primeira tomada dá lugar a uma segunda tomada gradualmente, havendo uma sobreposição dos primeiros frames da próxima tomada sobre os últimos da tomada anterior. Geralmente, este tipo de transição é usado na mudança de paisagens. Podemos visualizar este efeito na Figura 3.4:



Figura 3.4: Exemplo de Dissolve em sequência de frames.

Pode-se ver na Figura 3.4 a primeira tomada dando lugar gradativamente para a segunda tomada até que não haja mais sobreposição de imagens. Uma boa abordagem sobre a definição do dissolve pode ser encontrada em [28].

## 3.2 Sistemas de Vídeo

Os sistemas de vídeo consistem na organização em que uma determinada informação de vídeo é armazenada em forma digital, tais como AVI, MPEG, entre outros. O MPEG (Movie Pictures Experts Group) é um padrão internacional definido pela ISO que tem como característica a compressão de um vídeo. Onde ao comprimir uma determinada amostragem de um vídeo, o mesmo é passado pelo canal compressor, e ao exibir esta mesma amostragem o vídeo passa pelo canal expensor.

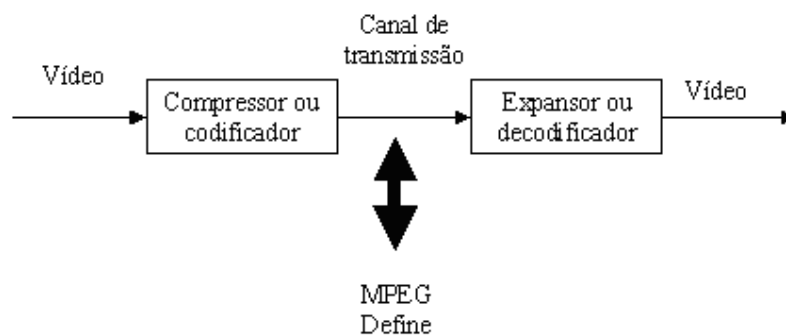


Figura 3.5: Sistema de Codificação e Decodificação de um vídeo no formato MPEG.

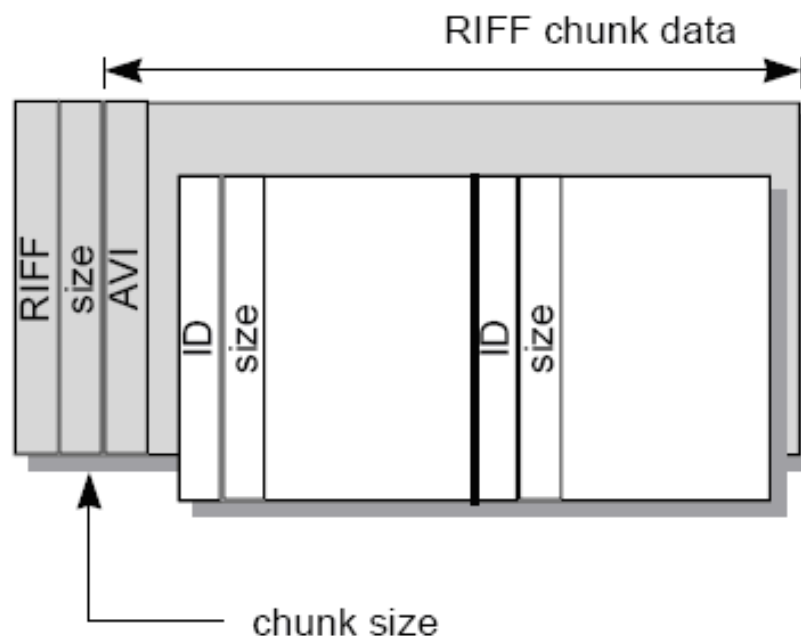


Figura 3.6: Sistema de armazenamento AVI.

O AVI (Audio Video Interleave) é um padrão criado pela Microsoft em 1992, o qual é derivado do padrão RIFF que divide os dados em blocos. Ao contrário dos outros padrões



o AVI não possui compressão.

### 3.3 Sistemas de Cor

Com a grande conclusão de Newton, no século XVII, de que a cor branca é composta por todas as outras cores, onde as três características básicas de um espectro de luz são, a matriz (definida pelo comprimento da onda), brilho (intensidade da luz fonte) e saturação (predominância da matriz). A formação do sistema de cores no processo aditivo é obtida a partir de duas fontes luminosas onde cores diferentes são projetadas em duas regiões, sendo assim, a interseção entre as cores há a formação de uma nova cor. Já o processo subtrativo é geralmente feito com a utilização de slides. Um exemplo é quando se emite uma luz branca (que possui todos os comprimentos de onda) sobre um filtro verde, o qual filtra todos os comprimentos de onda deixando passar somente o comprimento de onda relativa à cor verde, produzindo assim o verde. Dentro do sistema de cores podemos citar alguns padrões: RGB, CMYK, HSV, RYB, Highcolor, Imagem binária, SRGB, Munsell, entre outros.

#### 3.3.1 RGB

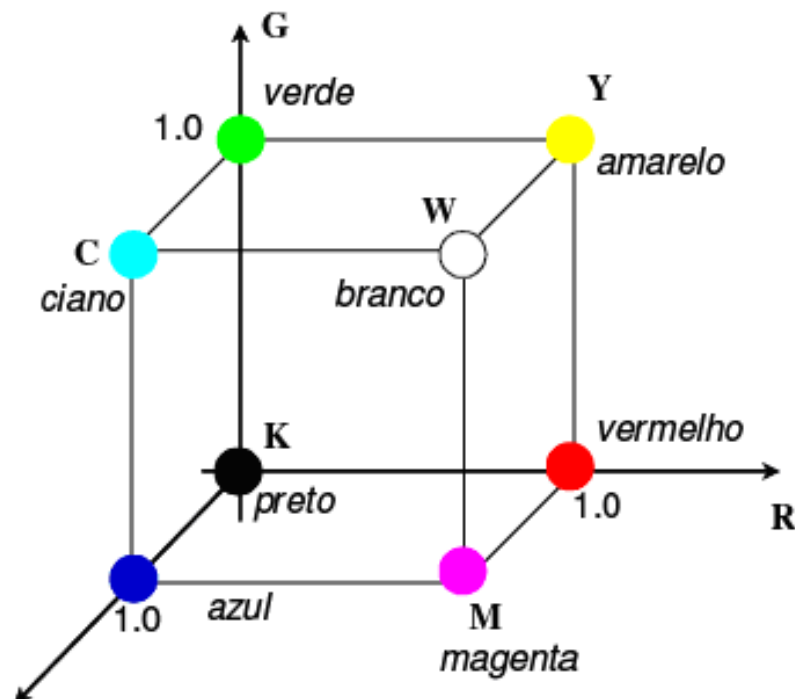


Figura 3.7: Sistema RGB.

Formado pelas cores vermelha, verde e azul. O RGB é um padrão de cores utilizado em monitores/televisores, em tubos de raios catódicos, cristal liquido ou de plasma. Utiliza

8 bits de informação para representar cada uma de suas cores (vermelho, verde, azul), podendo variar de 0 a 255. É possível alcançar mais de 16 milhões de cores com suas combinações de tons. Algumas destas cores estão representadas na Tabela 3.1:

Tabela 3.1: *Tabela com algumas das muitas combinações de cores possíveis com o sistema RGB.*

Cor	R	G	B
Preto	0	0	0
Preto Escuro	0	0	128
Azul	0	0	255
Verde Escuro	0	128	0
Turquesa	0	128	128
Azul Claro	0	128	255
Verde	0	255	0
Verde Água	0	255	128
Ciano	0	255	255
Marrom	128	0	0
Violeta	128	0	128
Azul Marinho	128	0	255
Verde Musgo	128	128	0
Cinza	128	128	128

Na Figura 3.7 - b é possível observar o modelo mapeado em um cubo. Onde o eixo X representado por valores de vermelho incrementa para a esquerda, eixo Y representado pelo azul incrementa para direita inferior e o eixo vertical Z pelo verde que incrementa para cima e a origem representa o preto.

### 3.3.2 CMYK

Também conhecido como modelo de cores baseado na subtração (absorção de luz), é formado pelas cores Ciano, Magenta, Amarelo e Preto, utilizado com maior frequência para impressões gráficas. A Figura ?? representa as misturas de cores de um sistema CMYK.

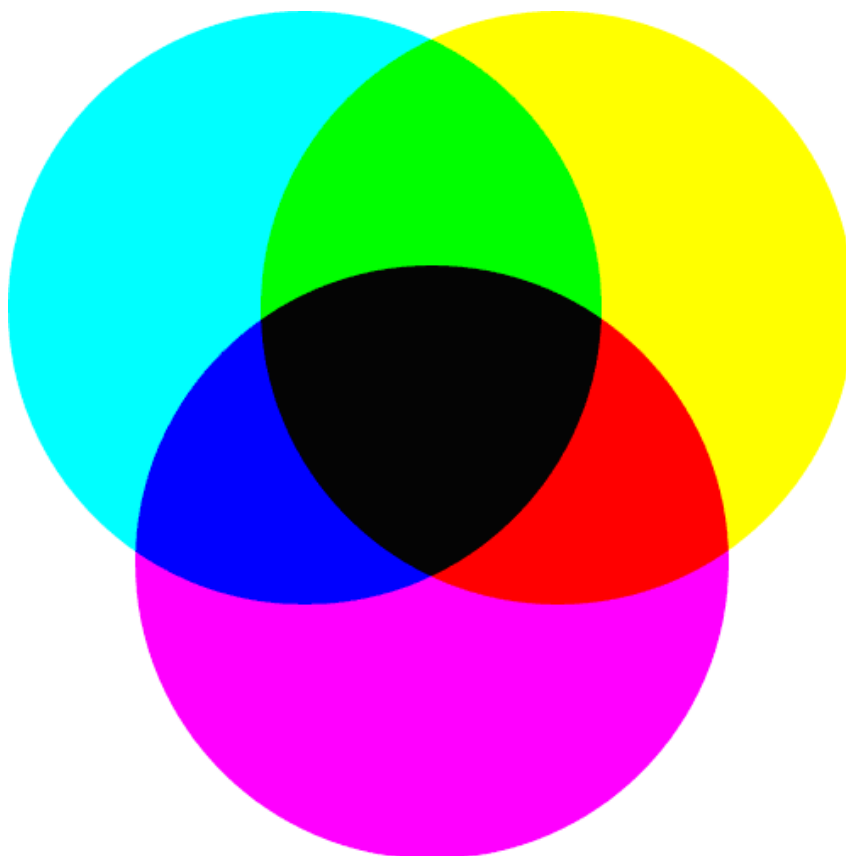


Figura 3.8: Misturas de cores no padrão CMYK.

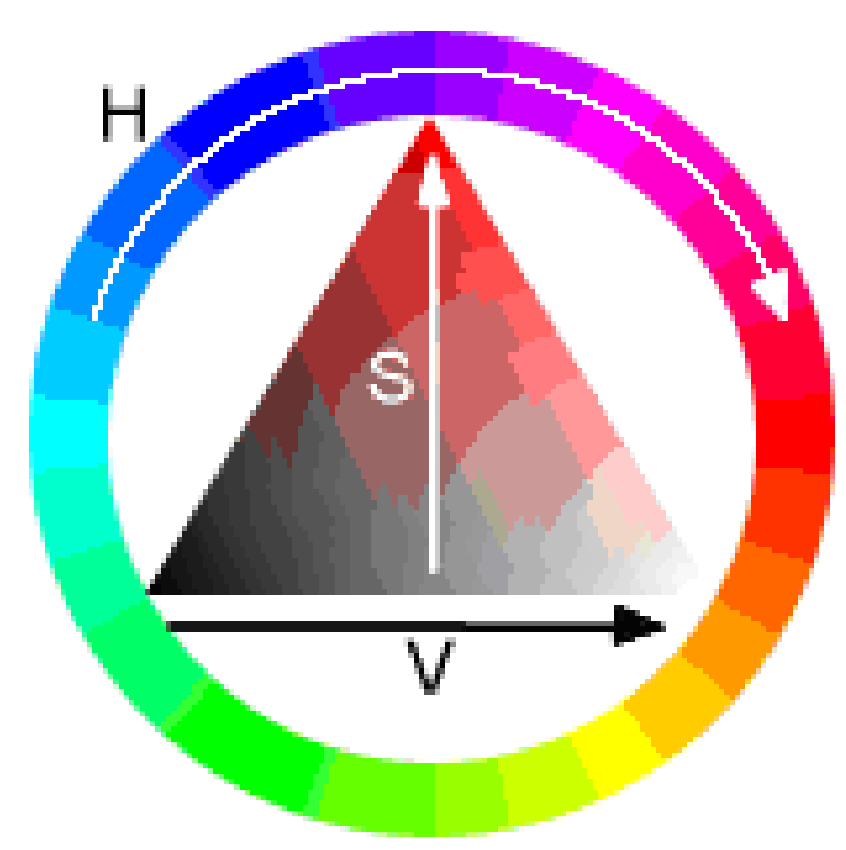


Figura 3.9: A mistura de cores representada por um círculo.

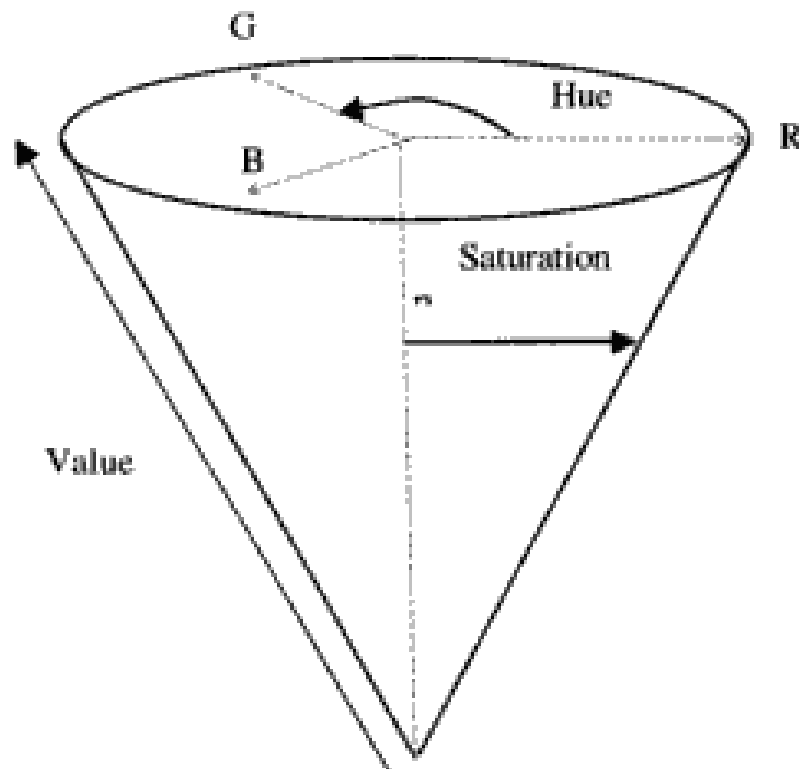


Figura 3.10: Modelo de representação do padrão HSV.

### 3.3.3 HSV

Também conhecido como HSB, este sistema de cores é o mais próximo à percepção da visão humana, o que leva vantagens em relação ao sistema RGB e CMYK. Este sistema é composto pela Tonalidade, Saturação e Valor, onde a tonalidade é a cor definida pela faixa de valores de 0 a 360 graus, representada por um círculo. A saturação varia pelo percentual da presença de tons de cinza de 0 a 100.

Ângulo Cor 0-60 Vermelho 60-120 Amarelo 120-180 Verde 180-240 Ciano 240-300 Azul 300-360 Magenta

Tabela 3.2: Seleção de nível de cores.

Angulo	Cor
0-60	Vermelho
60-120	Amarelo
120-180	Verde
180-240	Ciano
240-300	Azul
300-360	Magenta

3.3.4 Highcolor

No Highcolor cada pixel de cor é apresentado por 2 bytes, uma cor geralmete é representada por 16 bits, mas existem alguns processadores gráficos que armazenam somente 15 bits de informação, neste caso 5 bits são selecionados para definição de uma cor (ilustrada na Tabela 1), onde  $2^5 = 32$ , possibilitando aproximadamente 32 mil cores por pixel.

Tabela 3.3: Seleção de nível de cores.

Angulo	Cor
0-60	Vermelho
60-120	Amarelo
120-180	Verde
180-240	Ciano
240-300	Azul
300-360	Magenta

Bit 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1 Valor Vermelho Verde Azul

Tabela2 - Sistema de cores 15bits no Highcolor Quando os 16 bits são utilizados é possível alcançar aproximadamente 64 mil cores.

### 3.4 Detecção de Transições

A detecção de transições de um vídeo como já visto no item 5.1, é o passo principal para o início da manipulação de qualquer amostra de vídeos, pois representam marcos de início e fim de cada tomada de vídeo. Para cada tipo de transição, existem métodos de maior ou menor eficiência. Dentre os métodos de detecção de transições de vídeo, a aplicação do Ritmo Visual, Morfologia Matemática são as mais utilizadas. Com a aplicação de morfologia matemática nas amostras de vídeo, é possível por exemplo, decompor as imagens em resíduos morfológicos e assim, retirar informações sobre a transição existente. Este método é muito eficiente para detecções de transições graduais porém, ineficiente para transições bruscas. Aplicando Ritmo Visual e Morfologia Matemática, é possível representar um determinado evento de vídeo por um padrão de imagem 2D. Este método permite detectar transições como fades, cortes e dissolves.

### 3.5 Ritmo Visual

Ritmo Visual é definido como a representação de um segmento de imagens 2D (um vídeo) em uma única imagem na dimensão 1D, sendo portanto uma tentativa de tornar o trabalho de análise de vídeo simplificado. Existem dois tipos de representações do ritmo visual. Ritmo Visual por histograma e por sub-amostragem. Neste trabalho iremos focar no ritmo visual por sub-amostragem, o qual se caracteriza pela transformação de determinadas faixas de pixels (amostras) de cada um dos frames de uma tomada do vídeo em uma coluna da representação 1D, como representado na Figura 13. Podemos dizer então que o Ritmo Visual é uma simplificação do vídeo. O padrão desta representação pode ser alterado conforme a escolha dos pixels, portanto, o padrão do ritmo visual é totalmente dependente do tipo da amostragem retirada do vídeo. Bons resultados podem ser obtidos se a sub-amostragem do frame for dada pela sua diagonal principal, uma vez que ela trará características horizontais e verticais do frame, podendo assim fazer uma boa representação do mesmo. Uma melhor abordagem sobre este processo pode ser encontrada em [6,17].

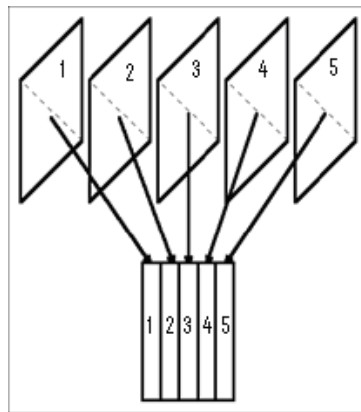


Figura 3.11: transformação da diagonal principal dos frames de um vídeo em colunas do Ritmo Visual.

Na Figura ?? pode ser observado o processo de transformação das diagonais principais dos frames 1, 2, 3, 4 e 5 nas colunas 1, 2, 3, 4, e 5 do Ritmo Visual. Matematicamente, podemos definir ritmo visual como:

Seja  $\mathcal{V}$  um vídeo composto por  $F_i \in \mathcal{V}, i = \{1, 2, 3, \dots, t\}$

Então  $RV = (F_i)$  é o seu Ritmo Visual.

Cada tipo de transição gera uma coluna com características, padrões diferentes no ritmo visual, como podem ser observados nas Figuras: 3.12, 3.13, 3.14:



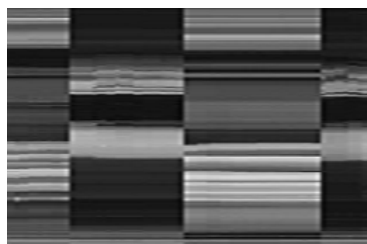


Figura 3.12: Representações de Ritmo Visual : cortes

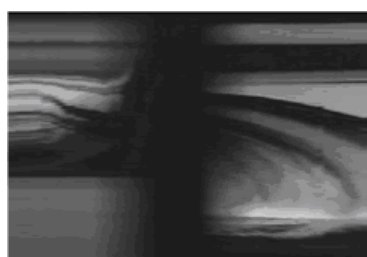


Figura 3.13: Representações de Ritmo Visual : fade-in



Figura 3.14: Representações de Ritmo Visual : dissolves

Em nosso trabalho, as técnicas de Ritmo Visual serão utilizadas para transformar o nosso problema de segmentação de vídeos em um problema de segmentação de imagens. A partir daí, faremos uso de outras ferramentas de PDI como filtragem espacial e detecção de bordas, juntamente com a aplicação de operadores de morfologia matemática, sobre o RV com o objetivo de obter os pontos de transição das tomadas no vídeo  $V$ .

## 3.6 Processamento de Baixo Nível

### 3.6.1 Filtragem Espacial

Filtragem Espacial Filtragem Espacial é uma técnica utilizada na área de processamento de imagens para realizar a suavização ou realce de detalhes da imagem, ou até mesmo a eliminação de elementos indesejáveis dessas imagens, como por exemplo, os ruídos. Para a aplicação de um filtro, utiliza-se o conceito de operadores de vizinhança, que consiste em alterar o Nível Digital (nível de tons de cinza, ou cor) de um pixel através do ND de seus pixels vizinhos. Dependendo do filtro a ser utilizado este cálculo pode ser alterado. Geralmente, cria-se uma máscara de convolução, que nada mais é do que uma matriz (normalmente de tamanho 3x3 : 3.4) que possui um determinado peso em cada uma de suas posições.

Tabela 3.4: Matriz 3x3 para um filtro passa-alta.

-1	-1	-1
-1	8	-1
-1	-1	-1

O pixel a ter seu ND recalculado é sempre aquele que estiver sob a posição central da matriz, como mostrado na Figura 3.15:

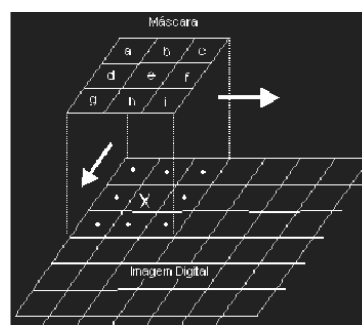


Figura 3.15: Máscara 3x3 sendo utilizada para a filtragem de uma imagem digital. No caso, o ND a ser recalculado é do pixel X.

Segundo [19, 22], o cálculo do novo ND de um pixel é dado pelo processo de convolução, que consiste na soma dos produtos entre os valores da matriz e pixels correspondentes da imagem, como mostrado abaixo: Seja  $M$  uma matriz de convolução 3x3 e  $O$  uma imagem a ser filtrada, então:

$$ND = \sum_{i=n}^n MiOi$$

Onde  $n$  é o número de elementos de  $M$ . Caso  $n$ , então todos os valores de  $M$  devem ser divididos por  $n$ . Durante a filtragem, esta matriz percorre toda a imagem pixel a pixel, partindo da esquerda para a direita, tal deslocamento é obtido através da soma de 1 pixel na posição  $x$  atual da matriz, como pode ser observado na Figura 12. Entre os principais tipos de filtragem está o filtro passa-alta, que pode ser uma boa alternativa para fazer a detecção de bordas e o filtro passa-baixa, que serve para fazer a suavização da imagem, resultando na diminuição de possíveis ruídos presentes na mesma.



Figura 3.16: Imagem Original



Figura 3.17: Imagem Após Filtragem Passa-Baixa

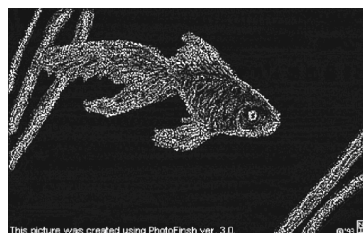


Figura 3.18: Imagem Após Filtragem Passa-Alta

Como pode ser observado na Figura 3.17, ao aplicar um filtro passa-baixa na imagem, ela torna-se mais homogênea e também se perde nitidez, porém os ruídos são suavizados. Já na Figura 3.18, foi aplicado um filtro passa-alta na imagem, havendo um aumento do contraste entre os objetos e o fundo da imagem, ganha-se nitidez e caso existam ruídos na imagem original, estes poderão ser realçados assim como as bordas dos objetos. Uma abordagem mais completa sobre este assunto pode ser encontrada em [19,23].

### 3.6.2 Detecção de Bordas

Detecção de Bordas é uma técnica para localizar os pontos e linhas que dividem regiões distintas de uma determinada imagem. A detecção das bordas pode ser feita através do uso de diversos operadores. Computacionalmente, define-se uma borda como sendo uma região que possui um gradiente de alta magnitude, por este fato, o operador de Sobel é uma interessante forma de se realizar a detecção de bordas, visto que este operador é baseado no cálculo do gradiente e sua magnitude. Segundo [20, 21], o cálculo do gradiente é dado por:

$$\nabla f = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix}$$

E sua magnitude é dada por:

$$mag(\nabla f) = [G_x^2 + G_y^2]^{\frac{1}{2}}$$

Onde:

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \text{ e } \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

são operadores que aplicam o processo de convolução na imagem como visto no tópico anterior. Através destes cálculos é possível obter, além da magnitude, a direção do gradiente.

Observa-se na Figura 3.20, que ao realizar uma filtragem utilizando o operador de Sobel, a imagem tem suas bordas em maior destaque do que na imagem original (Figura 3.19 ) Apesar deste operador realizar uma boa detecção de bordas, ele também aumenta o nível dos ruídos da imagem. Isto se deve ao fato de que ao passar pelos processos de convolução e cálculo de magnitude do gradiente, os ruídos, assim como as bordas, retornam um valor de magnitude elevada, sendo necessário aplicar outros processos de filtragem para a redução dos mesmos. Uma melhor abordagem sobre estes métodos pode ser encontrada em [17, 20, 21, 22, 23]. Em nosso trabalho, utilizaremos algumas técnicas de Filtragem Espacial para fazer o realce das bordas do ritmo visual, possibilitando dar um destaque maior para as bordas que possam possivelmente caracterizar uma transição.



Figura 3.19: Imagem Original.



Figura 3.20: Imagem após sofrer filtragem com operador se sobel.

## 3.7 Processamento de Médio Nível

### 3.7.1 Morfologia Matemática

Morfologia Matemática é basicamente o estudo das formas de objetos através da aplicação de determinadas operações matemáticas. Na área de processamento de imagens a morfologia pode ser aplicada em uma imagem com o objetivo de realizar o seu reconhecimento, a eliminação de ruídos, detecção de bordas, através do uso de algumas operações morfológicas. Segundo [1], a base da morfologia matemática é composta por dois operadores básicos, a erosão e a dilatação. A erosão consiste na retirada de pontos indesejáveis da imagem, enquanto a dilatação é a propagação de um determinado ponto para áreas ao seu redor.

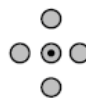


Figura 3.21: Cruz.



Figura 3.22: Quadrado.



Figura 3.23: Hexagonal.

Ambas as operações são sempre baseadas em um elemento estruturante que é um conjunto pequeno definido e conhecido (forma, tamanho) que é sempre comparado a um conjunto desconhecido da imagem. O elemento estruturante pode ser de vários tipos conforme observado nas Figuras 3.21 , 3.22 e ?? . A boa escolha de um elemento estruturante está bastante ligada à qualidade dos resultados das operações. Na Figura 3.24 e Figura 3.25, podem ser observadas a aplicação de uma erosão e dilatação respectivamente:

Segundo [1], a erosão de um element é dada por:

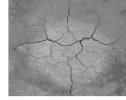


Figura 3.24: Imagem Original.

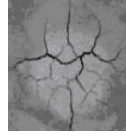


Figura 3.25: Resultado da aplicação da operação de erosão na imagem.

$$[\varepsilon_B(f)](x) = \min_{b \in B} f(x + b) - B(b)$$

Onde  $f$  é a região da imagem a ser erodida e  $B$  é o elemento estruturante. Uma imagem erodida resulta em uma imagem com regiões escuras alargadas e as claras diminuídas, em relação à imagem original, como é possível observar na Figura 3.25.



Figura 3.26: Imagem Original.

A dilatação, segundo [1], é dada por:

$$[\varepsilon_B(f)](x) = \max_{b \in B} f(x + b)$$

Onde  $f$  é a região da imagem a ser dilatada e  $B$  é o elemento estruturante. Como é possível observar na Figura 3.27, uma imagem dilatada tem seus pontos claros expandidos, enquanto os escuros são diminuídos em relação à imagem original.

Em nosso trabalho utilizaremos a Morfologia Matemática para analisar a imagem gerada a partir de técnicas de Ritmo Visual, e após a aplicação de técnicas de filtragem para fazer a extração de suas propriedades geométricas e topológicas definindo os padrões das transições que ocorrem em determinados pontos do vídeo.



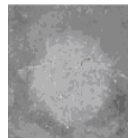


Figura 3.27: Resultado da aplicação da operação de dilatação na imagem.

### 3.7.2 Segmentação de Imagens

Em visão computacional, segmentação é o processo de dividir uma imagem em múltiplas regiões, com a intenção de subdividir partes importantes da imagem, ou simplesmente modificá-las, ambos a fim de facilitar uma análise posterior. Em imagens monocromáticas algoritmos são baseados em uma das propriedades básicas de valores de níveis de cinza: descontinuidade (detecção de bordas) e similaridade (agrupamento de regiões homogêneas). Na descontinuidade, o objetivo é particionar a imagem tendo como regra mudanças bruscas nas escalas de cinza. A técnica de similaridade funciona da seguinte maneira: uma imagem  $f(x,y)$  produz como saída uma imagem  $g(x,y)$ , onde cada pixel da imagem  $f$  é comparado com um limiar  $T$ , se o nível de cinza do pixel for maior que o limiar  $T$ , este pixel é mantido em  $g$ , caso contrário ele é descartado (considerado como fundo). Esta técnica é amplamente utilizada, pois tem um baixo custo computacional. Outra técnica também bastante conhecida é a limiarização ou binarização da imagem. É dado um exemplo de histograma com limiar na Figura 3.28:

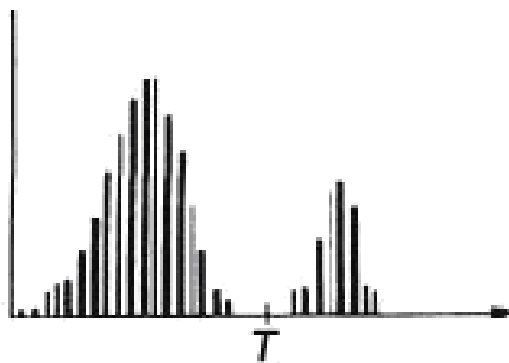


Figura 3.28: Exemplo de limiar com histograma.

Como pode ser observado nos exemplos das Figuras 3.29 , 3.30 , 3.31 e 3.32 , dependendo da limiar escolhida o resultado se altera. (explicar limiarização das imagens [29]).



Figura 3.29: Imagem Original.

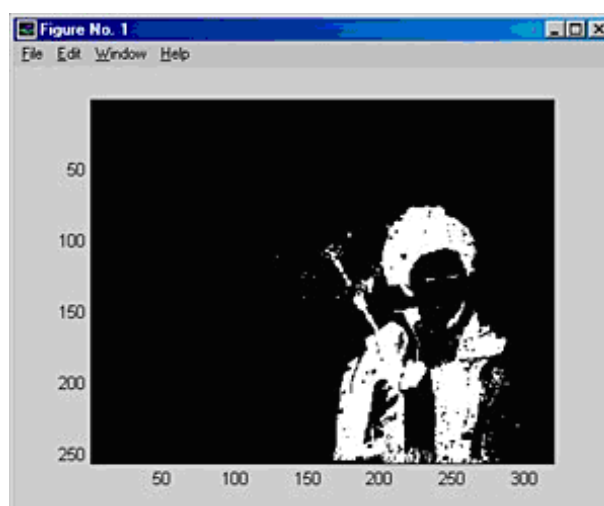


Figura 3.30: Limiar = 10.

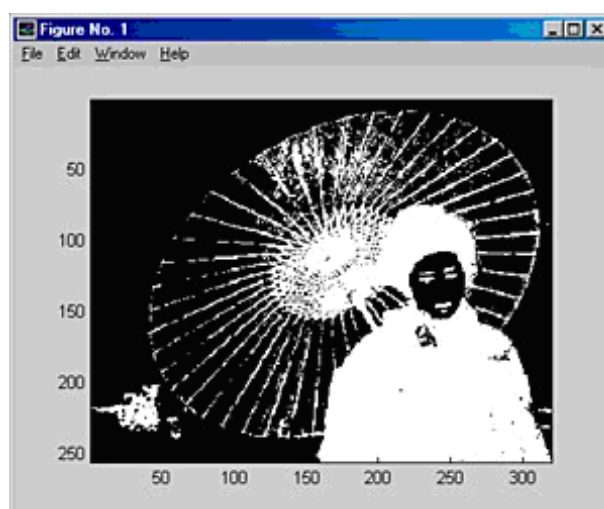


Figura 3.31: Limiar = 30.

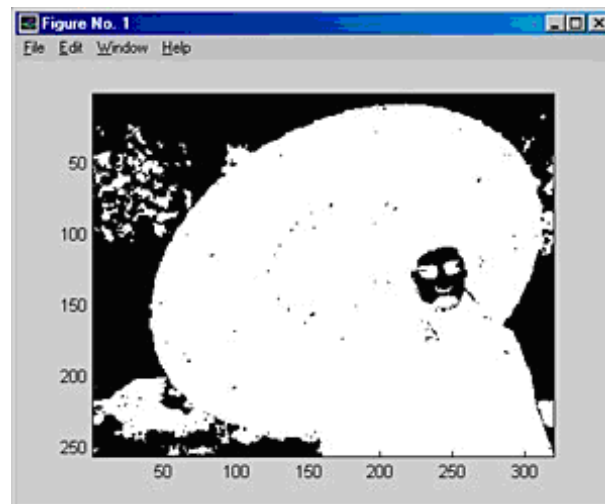


Figura 3.32: Limiar = 70.

## **3.8 Processamento de Alto Nivel**

### **3.8.1 Método de Casamento de Padrões**

### 3.8.2 Método Baseado em Entropia

# Capítulo 4

## Metodologia

Nesta seção será definida a metodologia a ser utilizada para a realização da detecção de transições.

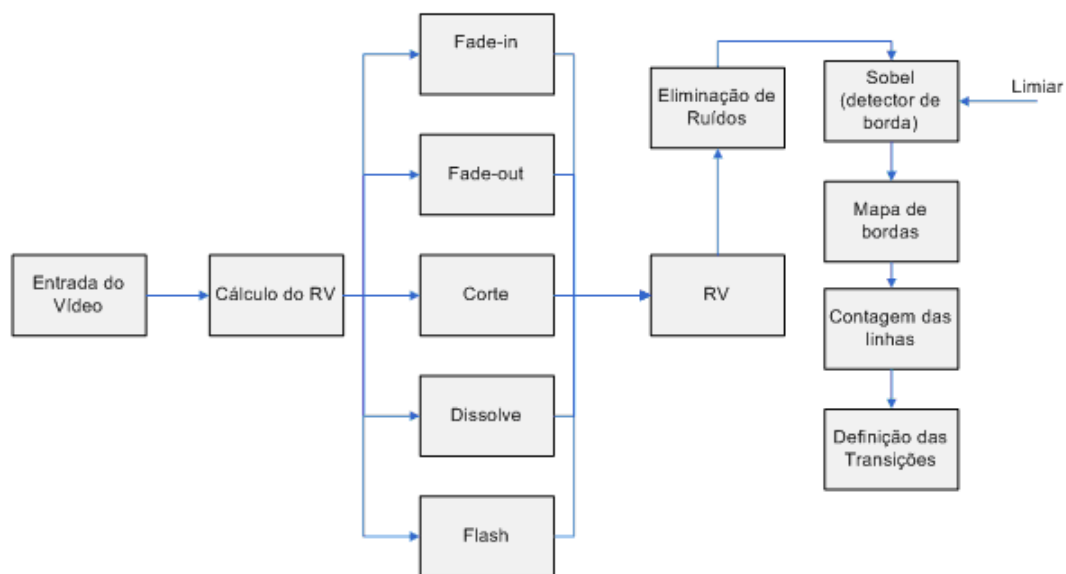


Figura 4.1: Seqüência de processos utilizados na metodologia.

Como pode ser observado na Figura 4.1, será primeiramente realizado o cálculo do Ritmo Visual. Dentro dele podem estar representados vários tipos de transições, somente cortes, somente dissolves, ou vários tipos misturados. Para realizar a detecção de cortes será utilizado o seguinte processo: Após obter o Ritmo Visual do vídeo será realizada a filtragem desta imagem, primeiramente para a diminuição de possíveis ruídos encontrados no vídeo. A eliminação destes ruídos pode ser dada através da aplicação de uma filtragem do tipo passa-baixa como visto mais anteriormente em 3.6.1. Após a suavização desta imagem será feita a binarização da imagem, que consiste na transformação de uma

imagem colorida ou em tons de cinza para uma imagem binária. Este processo é feito baseando-se em cima de seu histograma (vide tópico 3.7.2). Após a binarização será feita a detecção das bordas do ritmo visual, através do uso de operadores morfológicos, no caso o operador de Sobel. Como já dito anteriormente, ele é uma boa opção para realizar a detecção de bordas, visto que estas possuem gradientes de alta magnitude e ao realizar a binarização da imagem, regiões em que não houver borda terão um gradiente de magnitude aproximadamente igual em todos os seus pontos, enquanto as regiões de transição terão um gradiente de magnitude elevada (picos), resultando no mapa de bordas. A partir do mapa de bordas, é possível obter as transições que ocorrem no ritmo visual e com isso determinar onde ocorrem as transições no vídeo.



Figura 4.2: Obtenção do Ritmo Visual.



Figura 4.3: Filtragem para diminuição de ruídos.

Nas Figuras 4.2 , 4.3 , 4.4 , 4.5 e 4.6 , pode-se observar que os cortes de um vídeo são



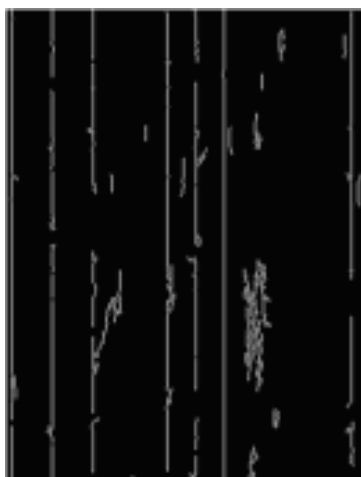


Figura 4.4: Detecção das bordas.

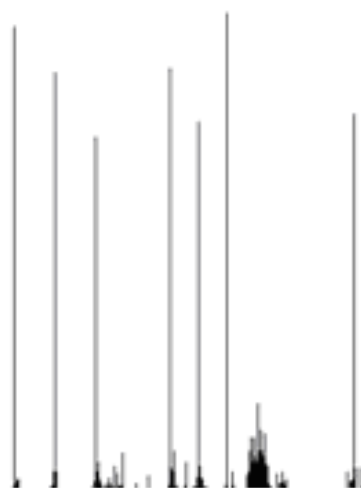


Figura 4.5: Mapa de bordas.

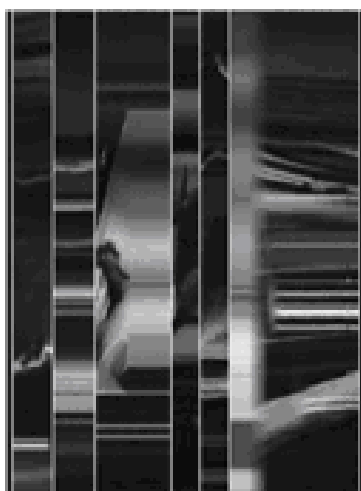


Figura 4.6: Sobreposição do mapa de bordas na imagem do ritmo visual, dando destaque às transições

transições verticais na representação do Ritmo Visual. Através de técnicas de filtragem e uso de operadores para a detecção das bordas, podemos obter as linhas verticais que caracterizam uma transição de corte, porém é importante lembrar que nem toda reta vertical encontrada no ritmo visual pode ser considerada uma transição.