

CENTRO UNIVERSITÁRIO DA FEI

JOSÉ EDUARDO DE MELLO FREIRE

**CLASSIFICAÇÃO AUTOMÁTICA DE VÍDEOS APLICANDO TÉCNICAS
DE INTELIGÊNCIA ARTIFICIAL**

São Bernardo do Campo

2014

JOSÉ EDUARDO DE MELLO FREIRE

**CLASSIFICAÇÃO AUTOMÁTICA DE VÍDEOS APLICANDO TÉCNICAS
DE INTELIGÊNCIA ARTIFICIAL**

Dissertação apresentada ao Centro
Universitário da FEI como parte dos
requisitos necessários para a obtenção do
título de Mestre em Engenharia Elétrica,
orientado pelo Prof. Dr. Flavio Tonidandel.

São Bernardo do Campo

2014

Freire, José Eduardo de Mello.

Classificação automática de vídeos aplicando técnicas de inteligência artificial / José Eduardo de Mello Freire. São Bernardo do Campo, 2014.

82 f. : il.

Dissertação - Centro Universitário da FEI.

Orientador: Prof. Dr. Flávio Tonidandel.

1. Classificação de vídeos. 2. Análise automática de conteúdo de vídeos. I. Tonidandel, Flávio, orient. II. Título.

CDU 007.5

 CENTRO UNIVERSITÁRIO DA FEI	APRESENTAÇÃO DE DISSERTAÇÃO ATA DA BANCA EXAMINADORA	Mestrado
	Programa de Pós-Graduação Stricto Sensu em Engenharia Elétrica	PGE-10

Aluno: José Eduardo de Mello Freire

Matrícula: 111106-1

Título do Trabalho: Classificação automática de vídeos aplicando técnicas de inteligência artificial.

Área de Concentração: Inteligência Artificial Aplicada à Automação

Orientador: Prof. Dr. Flavio Tonidandel

Data da realização da defesa: 30/05/2014

ORIGINAL ASSINADA

A Banca Examinadora abaixo-assinada atribuiu ao aluno o seguinte:

APROVADO ☒

REPROVADO ☐

São Bernardo do Campo, 30 de Maio de 2014.

MEMBROS DA BANCA EXAMINADORA

Prof. Dr. Flavio Tonidandel

Ass.: _____

Prof. Dr. Paulo Sérgio Silva Rodrigues

Ass.: _____

Prof. Dr. Olímpio Murilo Capeli

Ass.: _____

VERSÃO FINAL DA DISSERTAÇÃO

**ENDOSSO DO ORIENTADOR APÓS A INCLUSÃO DAS
RECOMENDAÇÕES DA BANCA EXAMINADORA**

Aprovação do Coordenador do Programa de Pós-graduação

Prof. Dr. Carlos Eduardo Thomaz

Ao criador, princípio indiscutível e indispensável, aos meus pais pelo apoio e suporte, e a minha filha Luiza eterna e infinita inspiração.

AGRADECIMENTOS

Agradeço a Deus e, acima de tudo, aos meus pais José Maria e Solange que me conduziram sempre de maneira carinhosa e compreensiva.

A todos meus professores que, por vocação, não medem sacrifícios na diária doação de conhecimentos.

Meu sincero obrigado ao meu orientador Flavio Tonidandel, pela paciência na minha orientação. Ao professor Paulo Sérgio Rodrigues, pelos conhecimentos transmitidos que tanto me motivaram.

Aos colegas de mestrado, em especial Andrey Masiero, sempre me encorajando com palavras de incentivo.

Ao colega Celso Alves da Silva, pela fundamental colaboração na correção dos desafiantes códigos desenvolvidos.

*Não existe caminho para a paz,
a paz é o caminho.
Mohandas Karamchand
(Mahatma Gandhi)*

RESUMO

A digitalização de imagens e vídeos ocasionou um aumento exponencial na captação, armazenamento e distribuição de filmes e animações. Como consequência, permitiu a proliferação de canais de televisão, seja via rádio frequência ou redes de dados, tornando a tarefa de identificar e selecionar vídeos demorada e ineficiente. Entretanto, técnicas de Inteligência Artificial podem ser aplicadas no desenvolvimento de sistemas que venham classificar e até bloquear vídeos, conforme seu conteúdo. Este trabalho analisa as técnicas comumente utilizadas para esta tarefa e propõe alguns métodos visando a aprimorar esta classificação.

Palavras-chave: Classificação de vídeos. Análise automática de conteúdo de vídeos.

ABSTRACT

The advent of digital image and video has led to an exponential increase in capture, storage and distribution of films and animations. As a result they allowed the proliferation of television channels by radio frequencies or data networks making the identification and selection of videos a slow and inefficient task. However, Artificial Intelligence techniques can be applied in the development of systems which could classify or block videos according their content. This work analyzes commonly used techniques, and proposes methods to improve this classification.

Key words: Video classification. Automatic video content analyze.

LISTA DE FIGURAS

Figura 1- Exemplo de imagem e seu respectivo histograma.....	26
Figura 2- Padrões de representação das cores nos espaços RGB e HSV.....	28
Figura 3- Histograma de cores HSV com 40 raias (bins).....	29
Figura 4- Conjunto de treinamento formado elementos pertencentes a duas classes.....	31
Figura 5- Elementos não classificados.....	32
Figura 6- Exemplo de três retas que separam os elementos do conjunto de treinamento.....	32
Figura 7- Três funções lineares usadas na classificação dos elementos.....	33
Figura 8- Três funções lineares e suas respectivas margens de separação.....	33
Figura 9- Representação do conjunto de treinamento, vetores de suporte.....	34
Figura 10- Elementos não linearmente separáveis.....	36
Figura 11- Detecção de transições com o uso do ritmo visual.....	39
Figura 12- Linha do tempo de um vídeo com dez cenas.....	40
Figura 13- Principais etapas desenvolvidas durante os testes.....	42
Figura 14- Os cinco gêneros escolhidos para classificação dos vídeos.....	43
Figura 15- Histograma médio de um vídeo.....	45
Figura 16- Seleção dos quadros-chave com o programa mplayer.....	48
Figura 17- Arquivo contendo os histogramas dos quadros-chave de um filme.....	48
Figura 18- Histograma médio de um vídeo com dados estatísticos do ritmo visual.....	49
Figura 19- Comandos de treinamento e classificação da biblioteca LibSVM.....	50
Figura 20- Resultados da classificação com diferentes valores de C e Gama.....	51
Figura 21- Resultados da classificação com funções de núcleo linear.....	52
Figura 22- Porcentagem de acertos dos tipos de classificação em função da constante C.....	54
Figura 23- Histogramas médios dos cinco gêneros analisados.....	56

SUMÁRIO

1 INTRODUÇÃO.....	11
1.1 Objetivos.....	11
1.2 Justificativa.....	12
1.3 Estrutura do trabalho.....	12
2 REVISÃO BIBLIOGRÁFICA.....	14
2.1 Breve histórico da animação, cinema e vídeo.....	14
2.2 A Classificação de vídeos.....	15
2.3 Métodos de classificação de vídeos.....	16
2.3.1 Classificação baseada no áudio.....	19
2.3.2 Classificação usando aspectos visuais.....	20
2.3.3 Classificação por reconhecimento de objetos e textos.....	22
2.3.4 Identificação de vídeos.....	23
2.3.5 Comparação dos tipos de classificação de vídeos.....	24
3 AS TÉCNICAS PARA A CLASSIFICAÇÃO DE VÍDEOS.....	26
3.1 Histogramas de imagens.....	26
3.1.1 Intersecção de histogramas.....	28
3.1.2 Histogramas RGB e HSV.....	28
3.2 Entropia de Shannon.....	29
3.3 Distância de Kullback-Liebler.....	30
3.4 Máquinas de suporte vetorial (SVMs).....	31
3.4.1 Margem flexível.....	35
3.4.2 Classificação não linear.....	36
3.4.3 Classificação multi-classe.....	37
3.5 Transições de cenas.....	38
3.5.1 Quadros-chave.....	38
3.6 Ritmo visual.....	38
3.7 Linha do tempo.....	39
3.8 Bases de vídeos.....	41

4 METODOLOGIA EMPREGADA.....	42
5 TESTES REALIZADOS.....	47
5.1 Resultados obtidos.....	51
6 CONCLUSÕES.....	55
7 TRABALHOS FUTUROS.....	57
REFERÊNCIAS.....	59
APÊNDICES.....	63
Apêndice A – Tabelas de resultados.....	63
Apêndice B – Códigos.....	72
A biblioteca de visão computacional OpenCV.....	73
Código para o cálculo dos histogramas HSV.....	73
A biblioteca LibSVM.....	76
Código para implementação da distância de Kullback-Liebler.....	76
A linguagem AWK.....	77
Código para o cálculo das estatísticas de duração das cenas.....	77
Apêndice C – Relação de vídeos.....	80

1 INTRODUÇÃO

Filmes e vídeos são, hoje, onipresentes e, cada vez mais, dominam espaços antes ocupados por outros meios de armazenamento de informação (ADOBE, 2013). Se séculos atrás era necessário manter um escrivão como Pero Vaz Caminha para documentar uma expedição, nas últimas décadas este trabalho tem sido feito por cinegrafistas ou mesmo câmeras automatizadas. Nas escolas, livros didáticos estão sendo, em parte, substituídos por material audiovisual. (SOARES ; BARBOSA, 2009). Hoje é mais comum ter uma informação registrada com a câmera do celular, do que usar papel e lápis para anotar, por exemplo, o telefone escrito em uma placa.

Organizar a informação sempre foi importante, porém, só recentemente grande parte da informação passou a ser gerada e distribuída sob a forma de vídeos digitais. Formas de se analisar, classificar, catalogar e armazenar vídeos terão um papel similar ao das bibliotecas como guardiães da informação. Por isso, dominar mecanismos de busca de informações nos vídeos será uma necessidade cotidiana que se encontra em seus estágios iniciais. Este trabalho propõe a análise de tal cenário com a intenção de descobrir os caminhos que levarão à manipulação eficiente da informação contida nos milhões de horas de material audiovisual disponível. Estudando as técnicas atuais de busca e classificação de imagens e vídeos pode-se concluir quais são as mais promissoras, e, portanto, devem ser as escolhidas no desenvolvimento de tais sistemas.

1.1 Objetivos

Este trabalho tem por objetivo analisar os métodos usuais empregados na classificação de imagens quando aplicados na classificação de vídeos de entretenimento. Buscando também investigar como alguns dos aspectos visuais relevantes dos vídeos podem inseri-los em determinados gêneros televisivos e, ainda, como tornar estes métodos mais eficientes.

1.2 Justificativa

Após aproximadamente noventa anos, os padrões adotados na transmissão da televisão analógica, provavelmente alguns dos mais bem sucedidos e longevos da história, estão sendo substituídos por uma família de padrões digitais indiscutivelmente mais eficientes quanto aos custos, recursos e qualidade (BASTOS ; FERNANDES, 2005). Dentre as inúmeras consequências deste processo está a possibilidade da aplicação direta das atuais técnicas de processamento digital de sinais para análise e processamento do conteúdo veiculado seja pelas emissoras, seja pela rede mundial de computadores.

A indústria do entretenimento, em especial emissoras de televisão e produtoras de conteúdo para a indústria de radio difusão, tem grande interesse em sistemas de arquivamento e catalogação, já que possui décadas de material audio visual, em grande parte ainda armazenado de maneira analógica, que, para não ser perdido, terá, necessariamente, que passar por um processo de digitalização nos próximos anos.

A mineração de dados (*data mining*) (VIJAYAKUMAR; NEDUNCHEZHIAN, 2012), por vezes, também esbarra na dificuldade de conseguir extrair automaticamente informações que não estejam explicitamente descritas, como é o caso de fotografias e vídeos, o que a torna menos abrangente. Para ser completo é preciso que um sistema seja capaz de procurar informações independentes, sendo indiferente a maneira que estes dados estejam representados.

A classificação de vídeos deve atender principalmente a três grandes demandas: a dos geradores/detentores de conteúdo, que possuem grandes arquivos de vídeos não catalogados, a dos usuários que precisam pesquisar em grandes acervos de vídeos de maneira rápida e eficiente, incluindo-se nestes acervos redes como a internet, e a de usuários que desejam buscar e/ou filtrar vídeos transmitidos pelas emissoras abertas, por assinatura ou rede de dados.

1.3 Estrutura do Trabalho

Este trabalho está estruturado da seguinte forma:

Na **seção 2** é apresentado um breve histórico das tecnologias de imagens e são apontados trabalhos anteriores sobre classificação de áudio e vídeo comparando técnicas utilizadas e resultados obtidos.

Na **seção 3** são explicadas as técnicas de classificação usando o recurso conhecido como SVMs (*Support Vector Machines*), e cada método empregado na extração dos parâmetros que compõem os vetores que serão submetidos à classificação, como os histogramas de imagens, e a análise do ritmo visual.

Logo em seguida, na **seção 4**, são apresentadas as etapas da pesquisa, bem como as ferramentas utilizadas e desenvolvidas na coleta e análise dos dados (vídeos). As características e limitações dessas ferramentas também serão detalhadas nesta seção.

Os procedimentos e testes realizados são descritos na **seção 5**, juntamente com os resultados obtidos.

As conclusões na **seção 6**, são compostas da análise dos resultados, suas causas, expectativas e implicações.

Na **seção 7**, sob o título de trabalhos futuros, são sugeridas linhas de pesquisa que possam dar continuidade a esta pesquisa.

2 REVISÃO BIBLIOGRÁFICA

O interesse por filmes e vídeos motivou importantes trabalhos, dentre eles muitos serviram de ponto de partida para esta pesquisa e são descritos a seguir.

2.1 Breve histórico da Animação, Cinema e Vídeo

No amplo contexto das tecnologias atuais, o registro e projeção de objetos animados são relativamente recentes, já que seus aproximadamente cem anos podem ser considerados pouco tempo quando comparados à evolução tecnológica em geral (HEAD ; STERLING, 1987).

Como a característica humana da persistência da retina só foi entendida em 1830, os primeiros dispositivos que reproduziram a sensação de movimento foram criados nos anos de 1820 (ANDERSON ; ANDRESON, 1993). Destes destaca-se o Fenaquistiscópio, do grego espectador ilusório, inventado por Joseph Plateau, em 1829, (VAN DER MENSBRUGGHE, 1885) justamente para demonstrar a sua teoria da persistência na retina. Estudos subsequentes comprovaram que a partir de aproximadamente uma dezena de imagens projetadas em um segundo consegue-se gerar a ilusão de movimento contínuo e natural. Sua primeira aplicação popular pode ser considerada os livretos de animação também conhecidos pelo termo original em inglês *Flip-Books*. Posteriormente, uma célebre experiência, realizada em 1876 por Eadweard Muybridge (HAAS, 1976), alinhou 24 câmeras fotográficas ao longo de um hipódromo e tirou várias fotos da passagem de um cavalo projetando-as em um Zoopraxiscópio, equipamento circular usado para exibir sequências de imagens, o que comprovou a utilidade destes estudos.

Aproximadamente uma década depois, os irmãos Lumière (DE ANGELIS, 2003) aproveitando-se da fita de celulose desenvolvida pelos laboratórios de Thomas Edson, inventaram o Cinematógrafo, um aparelho portátil com tripla função: máquina de filmar, de revelar e projetar. Em 1895, eles organizaram, na cidade de Paris, aquela que é tida como a primeira exibição pública de cinema.

Em pouco mais de um século, muitas tecnologias foram agregadas a fim de facilitar a criação de filmes e animações. Algumas substituíram tecnologias obsoletas, outras acrescentaram novos recursos, como exemplos significativos vale mencionar, a sonorização, a cor, a substituição da celulose por poliéster, a projeção tridimensional, a válvula, a televisão mecânica e posteriormente a eletrônica, o transistor, o vídeo-tape, a digitalização e, bem recentemente, a realidade aumentada. Duas tecnologias não diretamente relacionadas tiveram grande impacto na atual onipresença dos presentemente denominados vídeos (do latim vejo, uma vez que não são mais usados filmes), os aparelhos celulares que acumularam a função de captar imagens a qualquer hora e lugar, e a rede mundial de computadores que permite sua distribuição imediata.

2.2 A Classificação de Vídeos

Antes de ser considerado um dos problemas clássicos da Inteligência Artificial, a tarefa de classificação já desafiava a humanidade a desenvolver critérios e métodos de forma a otimizar recursos. A separação de sementes ou a garimpagem de pedras preciosas são exemplos milenares de classificação. É interessante notar que a quase totalidade dos problemas de classificação são atividades meio, ou seja, normalmente uma etapa de uma tarefa mais complexa. Este detalhe é importante pois as atividades meio são atividades normalmente elegíveis para a automação por serem repetitivas, volumosas e, conseqüentemente, dispendiosas e menos críticas.

A classificação automática de vídeos vem sendo pesquisada há mais de uma década (ROACH et al., 2002), motivada principalmente por três fatores: a difusão do vídeo digital iniciada com o desenvolvimento do padrão MPEG (*Moving Picture Experts Group*)¹, a partir de 1988, a aplicação doméstica dos *Digital Versatile Disk* (DVDs, padronizados em 1994) e o crescimento exponencial de vídeos na internet.

Sendo o vídeo (analógico ou digital) por princípio uma sequência de imagens, a classificação de vídeos tem como embrião a classificação de imagens, baseando-se, principalmente, em áreas como visão computacional e reconhecimento de padrões, agregando,

1 Página oficial (<http://mpeg.chiariglione.org/home>)

posteriormente, técnicas auxiliares, como reconhecimento textual (*Optical Character Recognition*) e reconhecimento de áudio (PFEIFER ; FISHER ; EFFELSBURG, 1996).

Um histórico da classificação de vídeos foi realizado por Matt Roach (ROACH et al., 2002), no qual se demonstrou que, mesmo com a supervisão humana, problemas de imprecisão são característicos deste tipo de classificação, devidos às sutilezas de interpretação ou de definição de gêneros. Apesar de parecer contraditório, a imprecisão da classificação quando feita por pessoas manualmente, torna a classificação automática mais interessante, uma vez que, mesmo apresentando imprecisões de origens diversas, no atual estágio de desenvolvimento em que as técnicas se encontram, o processo automático alcança níveis de precisão próximos do processo manual (BARLA; ODOONE; VERRI, 2003). Porém, apesar do avanço contínuo, perceptível e com resultados que comprovam sua viabilidade, os índices de acerto em torno de 80% (BARLA; ODOONE; VERRI, 2003) (~10% falsos positivos e ~10% falsos negativos) ainda tornam seu uso restrito ao campo da pesquisa.

2.3 Métodos Utilizados na Classificação de Vídeos

Vídeos são o meio para armazenar e transmitir informação através de imagens e sons (áudio e vídeo), a informação contida no vídeo pode ser interpretada por seu significado (informação cognitiva) ou por sua forma (informação formativa ou morfológica) (SMEULDERS et al., 2000). Este aspecto duplo de uma mesma informação explica, em grande parte, as diferenças da classificação feita por humanos, que se baseia no significado, e da classificação automática, baseada na forma. Esta diferença é denominada por especialistas em visão computacional como GAP SEMANTICO (BEEBE, 2006). O conceito de *Gap*, que em uma tradução literal do inglês significa fenda ou separação, coloca em diferentes patamares a visão natural, semântica, e a visão artificial, morfológica; considera-se que a primeira é de alto nível, enquanto a segunda possui baixo nível, sendo, então, estabelecido o *gap*. Entre as duas formas de classificação. O Gap Semântico é apontado como o principal motivo que torna difícil a criação de sistemas análogos à visão humana para interpretar imagens. Além disso, as características utilizadas pelos homens e mulheres para reconhecer os objetos, como a memória visual ou o contexto em que se encontram, são baseadas em conhecimento *a priori*, de difícil implementação em câmeras, que, na maioria dos casos,

convertem a imagem em sinais eletrônicos digitais para analisá-los matematicamente em tempo real.

Muitos pesquisadores afirmam que a solução para muitos problemas da interpretação de imagens e vídeos está justamente na construção de uma ponte eficiente sobre o Gap Semântico (JUN ; SHI-JIAO, 2012). Uma abordagem considerada bastante eficiente é a classificação indireta, que usa primeiramente elementos menos genéricos (eventos ou objetos da imagem) que possuam características mais formativas ou morfológicas, e posteriormente, relaciona-os às classes semânticas que podem representar os gêneros dos vídeos (esporte, jornalismo, filmes, adulto, infantil, entre outros).

A classificação de vídeos baseia-se, portanto, na extração destas características formativas ou morfológicas que contenham informação suficientemente relevante para definir se o vídeo analisado pertence a uma das classes propostas, ou excluí-lo de classes cujas características são incompatíveis.

Por definição, os vídeos são sequências de imagens estáticas, que podem ou não estar vinculados a uma trilha de áudio. Mais recentemente, vídeos armazenados em formato digital podem conter, além de imagens e sons, informações adicionais acrescentadas através de textos conhecidos como metadados. Os metadados, quando existem, não são mostrados no processo normal de exibição, mas, costumam conter informações autorais. Estas informações são normalmente inseridas pelo responsável pela digitalização do vídeo em questão e, na maioria das vezes, contém informações como nomes de atores e diretores, ano de criação, o gênero, a sinopse ou qualquer outra informação textual que se deseje acrescentar.

É fato que uma pessoa, ao observar uma cena, usa todos os sentidos simultaneamente para coletar informações de sons, cores, formas e demais características, e a partir delas identifica o gênero ao qual a cena pertence. Os sistemas artificiais são mais adaptados a uma classificação sequencial, isto é, o encadeamento de aspectos em uma espécie de árvore de busca. Assim, enquanto uma pessoa classifica uma cena como uma partida de futebol reconhecendo o ruído da torcida, ou escutando pessoas chutando uma bola em um estádio, um sistema identificaria, por exemplo, a preponderância da cor verde, o que restringiria a cena a um filme com cenas campestres ou a um esporte de campo. Em seguida, o sistema procuraria, então, um objeto parecido com uma bola e confirmaria a opção esporte com a análise do ruído do estádio. Isto demonstra que na classificação automática é necessário um maior número de etapas para o êxito da categorização, já que ela não se utiliza do paralelismo multis-sensorial humano. Se uma pessoa, ao observar a imagem de um personagem empunhando uma arma,

imediatamente identifica a cena como parte de um filme policial, um sistema, por sua vez, teria que procurar por estampidos de tiros, brilho de explosões ou outros objetos para depois associá-los ao gênero policial.

Grande parte do desenvolvimento na classificação de vídeos tem como objetivo distinguir conteúdo adulto de conteúdo apropriado a todas as pessoas, ou seja, busca a detecção de nudez ou pornografia (FENG et al., 2001). Mesmo estando intimamente relacionado à classificação mais genérica, este caso é uma aplicação particular, em que, por exemplo, parâmetros relacionados à cor (tons de pele) são mais significativos que parâmetros relacionados à forma. Esse modelo de classificação emprega certas técnicas com mais ênfase do que outros, o que não necessariamente funcionará em classificações com outros objetivos como, por exemplo, esportes. Este exemplo demonstra que não apenas a escolha das técnicas aplicadas, mas também a forma como essas diversas técnicas serão inter-relacionadas terá uma grande influência no resultado final. Fablet e Bouthemy (FABLET ; BOUTHEMY, 1999) desenvolveram estudos de detecção por movimento bastante eficientes em vídeos com características cinéticas específicas, como imagens de um jogo de tênis ou pingue-pongue. No entanto, esses estudos não tem grande valia na maioria dos casos, indicando assim, que a classificação eficiente deve ser feita também por eliminação e não apenas por identificação direta.

Um dos aspectos mais importantes de um sistema de classificação está na escolha do modelo de classificação adotado. Uma vez definidos os elementos, objetos, eventos, gêneros e tipos de dados a serem identificados, é fundamental determinar de forma criteriosa a técnica que norteará o sistema. Um trabalho bastante completo foi realizado por Olivier Chapelle (CHAPELLE ; HAFFNER ; VAPNIK, 1999) em que as principais técnicas desenvolvidas até então são comparadas partindo da classificação que essas técnicas realizaram de uma popular biblioteca de fotografias. O estudo compara, tanto seus aspectos teóricos quanto seus resultados práticos com o intuito de apontar quais são as mais indicadas em aplicações que dependem da análise de imagens. Entre as características normalmente avaliadas na seleção dessas técnicas, pode-se enumerar a robustez, entendida como a sensibilidade a ruídos e a outras fontes de erro, o custo computacional e a adequação ao tipo de dados coletados.

Darin Brezeale (BREZEALE ; COOK, 2007) e mais recentemente Hu et al. (2011) fazem estudos bastante completos destas técnicas, porém, em muitos casos, não é possível compará-las uma vez que usam bases diferentes com características bem distintas.

É necessário considerar que dentre as diversas aplicações de análise eficiente de vídeos está a monitoração do tráfego de informações pelas redes de computadores, sendo que este tipo de monitoração tem o intuito de garantir aos governos a segurança de seus países. Para citar apenas um exemplo, existe uma agência do governo dos Estados Unidos da América, a National Institute of Standards and Technology (NIST), patrocina desde 2001 pesquisas nesta área, assim como uma conferência anual conhecida como TRECVID. Durante a organização deste evento são definidos quais vídeos serão usados e quais tarefas serão propostas, diversos grupos do mundo todo se inscrevem e participam para apresentar e conhecer as soluções desenvolvidas por todos. Os resultados são, então, divulgados e servem de base para inúmeros trabalhos mundo afora.

Este trabalho, sempre que possível, fez uso de parte dos dados e vídeos utilizados por este grupo, como uma tentativa de padronização, facilitando assim a comparação com trabalhos correlatos. Porém, como parte destes dados e vídeos possuem algumas restrições de uso, este material não pode ser usado em sua integralidade.

Os métodos adotados na classificação de vídeos podem ser divididos segundo o tipo de informação que analisam, ou seja, basicamente áudio, vídeo ou texto. Cabe ressaltar que muitos estudos usam mais de um tipo de informação para aprimorar a classificação.

2.3.1 Classificação Baseada no Áudio

Os sons contidos nos vídeos podem ser analisados utilizando técnicas empregadas em outras aplicações já bastante estudadas como o reconhecimento de voz. Uma das vantagens de se usar o áudio é seu custo computacional mais baixo, já que ele se constitui de dados significativamente menores em termos de memória e consequentemente processamento (a taxa de transmissão de áudio é da ordem de quilobits por segundo enquanto a de vídeo é de megabits por segundo). As características dos sons podem ser analisadas tanto no domínio do tempo como no domínio da frequência. Como características importantes do domínio da frequência, temos a largura do espectro do sinal (*bandwidth*), sua distribuição de energia, seu centroide, a frequência base ou fundamental, entre outras. No domínio do tempo temos como principais: a potência RMS, Root Mean Square; (LATHI,1988), que é reconhecida pelos humanos como o volume, sua variação (dinâmica), o nível de ruído, a quantidade de períodos

de silêncio e número de cruzamentos com zero. Liu (LIU ; WANG ; CHEN, 1998) analisa a maioria destas características de segmentos de 1,5 segundos do áudio de 20 diferentes trechos do vídeo coletando dados para formar um vetor para treinar uma Hidden Markov Model (RABINER, 1989, SUTTON ; BARTTO, 1998) para cada classe de vídeo que se deseje classificar, no seu caso, cinco são as categorias: comercial, basquete, futebol, notícias e previsão do tempo. Outros autores como Dinh et al. (DINH ; DORAI ; VENKATESH, 2002) seguem a mesma linha, porém, usando como classificador o K-Nearest Neighbor.(HASTIE; TIBSHIRANI; FRIEDMAN, 2009).

2.3.2 Classificação Usando Aspectos Visuais

Uma vez que a maior parte da informação contida em um vídeo está em suas características visíveis, grande parte dos trabalhos desenvolvidos (BREZEALE ; COOK, 2007) leva em conta características das imagens para coletar os dados utilizados nos mecanismos de classificação. Os vídeos são compostos por uma ou mais cenas e cada cena por uma sequência de quadros. Como a quantidade de dados em um vídeo é muito grande torna-se conveniente segmentá-lo em pequenos trechos (conhecidos como clips) de cada cena ou até mesmo extrair apenas uma imagem (*key frame*) representativa de cada cena. Como um filme pode conter centenas de cenas, o mais comum é concentrar-se em um determinado número das cenas iniciais, partindo-se do princípio que um filme não muda de gênero em seu decorrer.

Dentre as características visuais mais relevantes temos: a duração das cenas e os aspectos relacionadas à cor, à luminosidade, ao movimento, à forma e à textura. Retomando no exemplo do jogo de futebol, é possível perceber que o vídeo possui cenas curtas, preponderância de tons de verde e muitos movimentos horizontais de câmera. Já um filme com muitos diálogos, classificado como drama, contém cenas longas, preponderância de tons de pele, com formas de rostos, e poucos movimentos de câmera (BREZEALE ; COOK, 2007).

A análise e a comparação das imagens frequentemente tira proveito de histogramas (CHAPELLE ; HAFFNER ; VAPNIK, 1999, SWAIN ; BALLARD, 1991). Histogramas têm a capacidade de agrupar a informação relevante em relação às cores quantificando sua

frequência. Outra vantagem do uso de histogramas é a eliminação de algumas características não tão relevantes como translações ou rotações de objetos.

Para se gerar os histogramas utilizam-se normalmente os padrões RGB ou HSV de cores (GONZALES, 2006), sendo o segundo mais eficiente uma vez que não é tão sensível às variações de luminosidade como aquelas causadas por cenas filmadas à luz do dia ou em recintos fechados. Para evitar este problema de luminosidade, Drew e Au (DREW ; AU, 2000) recorrem à normalização das cores das imagens para, posteriormente, gerar seu histograma. Como toda a informação espacial se perde na geração dos histogramas, alguns autores dividem as imagens em regiões para obter histogramas regionais. Outro aspecto espacial relacionado ao espalhamento da cor pode ser mensurado utilizando a entropia da imagem (ESQUEF, 2002).

Um ponto que se provou importante é a segmentação das cenas (LIENHART, 1999), uma das maneiras mais eficientes de fazê-lo é usar o fluxo do vídeo calculado a partir da variação dos pontos da diagonal das imagens para resolver este problema. Após a etapa da segmentação das cenas, sua duração passa a ser um parâmetro passível de uso na classificação.

No âmbito das características cinemáticas, os movimentos dos pontos das imagens podem ser de dois tipos: um deles baseados nos movimentos de câmera, ou seja, todos os pontos da imagem movem-se no mesmo sentido, e o outro, no movimento de objetos, neste caso apenas parte dos pixels se move. Determinados esportes costumam conter muito movimento de câmera (AYYILDIZ ; CONRAD, 2011), comédias por sua vez, possuem movimentos de objetos e dramas quase não possuem movimento (BREZEALE ; COOK, 2007). Muitos trabalhos partem do fluxo óptico (HORN ; RHUNCK, 1980) para mensurar e classificar os movimentos, entretanto, em trabalhos recentes, o mais comum é tirar proveito dos vetores de movimento que a maioria dos formatos de arquivos de vídeo já armazena como forma de compressão temporal. Estes vetores representam, com precisão, translações e rotações dos objetos e, por terem sido gerados durante a fase de captura do vídeo, têm baixo custo computacional para serem lidos e analisados. Outros parâmetros, como a média das cores de uma determinada região, podem ser conseguidas desta forma.

2.3.3 Classificação por Reconhecimento de Objetos e Textos

A detecção de objetos, em especial rostos humanos, foi abordada por Dimitrova, Agnihotri e Wei (2000) e Wei, Agnihotri e Dimitrova (2000), primeiramente, identificando tons de pele e, posteriormente, buscando aspectos morfológicos característicos às faces. Outros estudos detectam objetos com formas bem específicas como carros, móveis, ou casas. Nota-se aí uma falta de generalização nestas técnicas o que torna seu uso menos abrangente, mas não menos interessante.

Os textos normalmente são tratados como um tipo particular de objeto. A presença de textos na imagem por si só pode servir para indicar uma maior probabilidade de se tratar de notícias ou esportes. Logotipos e vinhetas também podem ser usados para este fim, mas além de especificar a identificação a um universo pequeno de vídeos estes métodos exigem algum tipo de conhecimento prévio, como uma base de dados dos objetos a serem reconhecidos.

Para se usar o significado do texto embutido no vídeo é necessário aplicar técnicas custosas como reconhecimento de caracteres (OPTICAL CHARACTER RECOGNITION ou OCR) e, na sequência, outro sistema com listas de palavras associadas aos gêneros do tipo BAG OF WORDS (DESELAERS; PIMENIDIS; NEY, 2008) que para classificação de vídeos não se mostrou muito eficiente, principalmente pelo fato do texto não estar presente na grande parte dos vídeos. Na tentativa de se contornar esta limitação, pode-se recorrer a determinados tipos de textos mais frequentemente presentes em muitos vídeos. Legendas, muito comuns em DVDs, textos para deficientes auditivos (*close caption*), obrigatórios em grande partes dos países inclusive no Brasil, e metadados em campos pre-definidos, previstos em formatos como MPEG, podem ser facilmente coletados, mas não estão sempre presentes e, no caso da internet, são raros os casos em que podem ser identificados.

Por sua comprovada eficiência, quase a totalidade dos sistemas de reconhecimento de padrões baseia-se em alguma variante de Redes Neurais ou Máquinas de Vetores de Suporte (SVMs) (CHAPELLE; HAFFNER; VAPNIK, 1999). No caso de classificação de imagens, inúmeras pesquisas demonstram a eficiência das redes com funções de base radial ou RBFs. As diversas configurações das RBFs permitem ajustes que tornaram possível um índice de acerto sem precedentes e explicam seu uso frequente.

Mais recentemente, técnicas baseadas em SVMs passaram a ser preferidas pelos pesquisadores nessa área, principalmente por terem a capacidade de classificar com maior

facilidade informações de alta dimensionalidade, como imagens constituídas de milhares de pontos (pixels), predominantes nos vídeos digitalizados. As SVMs têm, como característica, não serem multi-classe, ou seja, não é possível fazer uma classificação em cinco gêneros de uma só vez, mas sim uma classificação do tipo “um contra todos”. Este tipo de classificação tem como desvantagem a necessidade de uma classificação para cada gênero, porém permite que as características mais importantes de cada classe sejam priorizadas individualmente por classe.

A diversidade das características apresentadas por vídeos de diferentes fontes acarretam em dados absolutamente heterogêneos, fato que evidencia que nenhuma técnica isolada irá contemplar uma solução satisfatória. Tal constatação é corroborada pelo fato de que nenhuma técnica sozinha atingiu níveis razoáveis de precisão (BREZEALE ; COOK, 2007) portanto, a associação cuidadosa de mais de uma técnica sempre acarreta em melhora dos resultados (BARLA; ODONE; VERRI, 2003), o que supera sensivelmente qualquer técnica que seja empregada de forma isolada.

Vídeos de um mesmo gênero quando capturados por diferentes câmeras, em diferentes formatos de armazenamento, com resoluções diferentes, entre outras características podem diferir muito entre si, mas o conjunto de algumas de suas características, de alguma forma, contém a informação necessária para inseri-lo em uma mesma classe. Definir quais são estas características e quais são as técnicas mais precisas para mensurá-las é a chave para o desenvolvimento de um sistema de classificação realmente eficiente.

2.3.4 Identificação de Vídeos

Pode-se indagar se a solução definitiva seria a identificação dos vídeos e não sua classificação. Isto pode ser feito com o uso de bancos de dados de grande capacidade que catalogariam uma assinatura, ou seja, uma característica única, como por exemplo 4 quadros idênticos, na mesma posição, em trechos distintos, de praticamente todos os vídeos produzidos. Este modelo já funciona com músicas, é muito popular e entre outras aplicações em dispositivos móveis, para usá-lo, basta cantarolar a música em um aparelho celular para que esta seja rapidamente identificada e reproduzida com o auxílio de um servidor on-line (ex.

aplicativos para celular Shazam² e Midomi³). Mesmo eficiente este método exigiria submissão prévia do vídeo e a alimentação das informações sobre ele. Além disso, a quantidade de vídeos produzidos hoje supera a de músicas, uma vez que não é necessário conhecimento específico como os de um músico para se criar um vídeo. No entanto, esta técnica pode ser útil e com certeza será usada na catalogação de filmes e vídeos de maior importância como os produzidos para cinema e televisão, mas dificilmente conseguirá atingir bons resultados em ambientes dinâmicos como a Internet.

2.3.5 Comparação dos Tipos de Classificação de vídeos

Cada tipo de classificação apresenta suas vantagens, desvantagens e áreas de aplicação. A tabela 1 faz uma comparação entre os tipos mencionados. Deve-se ter em mente que a eficiência não é a única característica importante na escolha do tipo de classificação. Este trabalho, como a maioria dos trabalhos correlatos, pesquisou apenas métodos de classificação baseados em aspectos visuais. Isto se deve às limitações acarretadas pelos demais métodos. As classificações baseadas em objetos são muito específicas, sendo, então, indicadas para o reconhecimento facial, de veículos ou objetos determinados. Já os métodos baseados em áudio ou em textos, possuem a limitação do idioma, o que os torna pouco abrangentes. A identificação de vídeos depende da catalogação prévia, o que restringe seu uso a uma fração dos vídeos existentes. Assim sendo, mesmo por vezes não atingindo a eficiência dos demais métodos, a classificação baseada em aspectos visuais é a mais abrangente, e foi, portanto, a escolhida para esta pesquisa.

2 Página oficial (<http://www.midomi.com>)

3 Página oficial (<http://www.shazam.com>)

Tabela 1- Comparação entre os tipos de classificadores de vídeos

Tipo de Classificação	Vantagens	Desvantagens
Baseada nos Aspectos Visuais	Independente da língua Baixo custo computacional Fácil implementação	Média eficiência Muito generalista
Baseada no Áudio	Alta eficiência Interpretação Semântica	Dependente da língua Alto custo computacional Difícil implementação
Baseada no Texto	Alta eficiência Interpretação Semântica	Nem sempre presente Dependente da língua
Baseada em Objetos	Independente da língua Alta eficiência	Nem sempre presente Alto custo computacional Difícil implementação
Identificação de Vídeos	Alta eficiência	Difícil implementação Exige catalogação manual, conhecimento <i>a priori</i>

Fonte: Elaborado pelo autor.

3 AS TÉCNICAS PARA A CLASSIFICAÇÃO DE VÍDEOS

Por sua importância a classificação de vídeos, como visto no capítulo 2, já foi tema de inúmeras pesquisas. Como resultado, é possível determinar de antemão, com relativa certeza, quais técnicas são mais promissoras. Norteado por este critério e pelos motivos mencionados no item 2.3.5, foram escolhidas as técnicas descritas a seguir, que serão as adotadas neste trabalho.

3.1 Histogramas de Imagens

Histogramas de imagens (CASTLEMAN, 1996) são gráficos gerados a partir da função de densidade e probabilidade (PDF) de uma imagem. Como pode ser visto na figura 1, no eixo horizontal estão representadas as cores ou tons de cinza, e no eixo vertical a quantidade de pontos da imagem que possuem esta cor. No caso de histogramas normalizados, o eixo vertical indica a probabilidade correspondente a um ponto da imagem possuir esta cor ou tom.

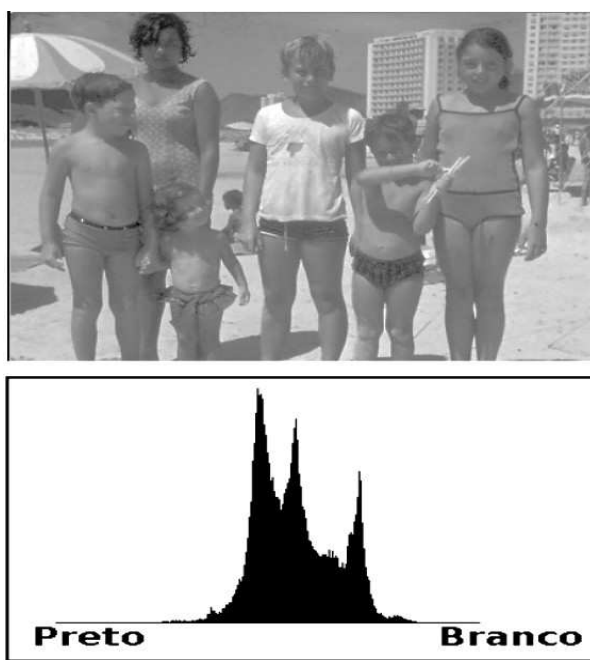


Figura 1 - Exemplo de imagem e seu respectivo histograma

Fonte: Foto pertencente ao acervo pessoal do autor e histograma elaborado pelo autor

Por se tratar de uma probabilidade e o número de pontos (*pixels*) das imagens ser variável, uma normalização é aplicada com a divisão pelo número total de pontos da imagem (equação 1). Histogramas podem ser aplicados em imagens monocromáticas ou coloridas, sendo que no último caso, pode-se criar um gráfico para cada componente de cor ou agrupá-los em um só.

$$P_r(r_k) = \frac{n_k}{n} \quad (1)$$

Onde,

$$0 < r_k < 1$$

$k = [0, 1, 2, 3, \dots, L-1]$, onde L é o número de cores da imagem.

n = número total de pontos da imagem.

n_k = número de pontos cuja cor ou tom de cinza corresponde à k .

$P_r(r_k)$ = Probabilidade do K -ésima cor.

Intuitivamente, os histogramas representam quais cores estão presentes na imagem e com qual frequência elas aparecem. Como as imagens digitais são capazes de armazenar milhões de cores, a representação de cada cor individualmente no histograma é inviável. Por isto, as cores são separadas em faixas conhecidas como *raias*, ou pelo equivalente em língua inglesa, *BINS*, o que além de tornar a interpretação do histograma mais fácil permite o agrupamento de cores com tonalidades próximas tornando a comparação dos tons de cores de imagens distintas muito mais eficiente. É importante ressaltar que histogramas independem de forma, tamanho, texturas, posição ou outras características morfológicas, uma vez que apenas a informação de cor (ou tom de cinza no caso de imagens monocromáticas) é utilizada. Mesmo com esta limitação os histogramas até o presente, são a base da maioria dos sistemas de classificação (SWAIN ; BALLARD, 1991) de vídeos, visto que a cor contém a maior parte da informação, e a forma dos objetos no mundo real, além de variar, depende fundamentalmente da posição do observador (câmera) tornando-se um desafio de difícil transposição no reconhecimento dos objetos.

3.1.1 Intersecção de Histogramas

Uma das técnicas mais adotadas na classificação de imagens é a intersecção de histogramas. Nela duas imagens são comparadas analisando-se as diferenças euclidianas entre seus histogramas, este método se mostrou razoavelmente eficaz na determinação dos gêneros de imagens estudadas (CHAPELLE ; HAFFNER ; VAPNIK, 1999), mas apresenta um erro considerável tanto de falsos positivos quanto falsos negativos, por isso inspira aprimoramentos. Dentre estes aprimoramentos, um bastante utilizado é a distância de Kullback-Liebler em substituição à distância euclidiana (SOBRIECKI et al., 2010). Com isto consegue-se dar uma importância maior as cores mais presentes, também conhecidas como cores preponderantes, e praticamente descarta-se a comparação daquelas cores ou tons que são pouco presentes nas imagens e que, portanto não devem ser relevantes.

3.1.2 Histogramas RGB e HSV

Quando as imagens analisadas são coloridas, os histogramas são gerados a partir de suas componentes de cor. No caso de imagens RGB, um histograma é gerado para a componente vermelha (*Red*), um para a componente verde (*Green*), e um para a azul (*Blue*). O mesmo acontece quando a imagem está armazenada no padrão de cores HSV (HUE, SATURATION, VALUE).

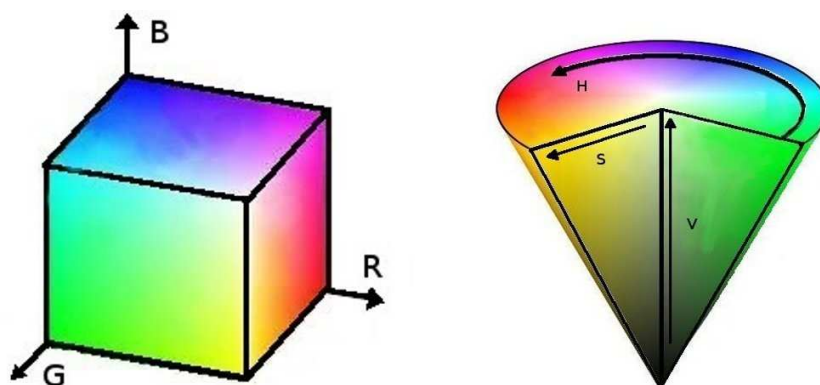


Figura 2 - Padrões de representação das cores nos espaços RGB e HSV
Fonte: Elaborada pelo autor

Isto torna possível e até interessante a manipulação dos pesos destas componentes. Como os vídeos podem ser capturados com diferentes tipos de iluminação (dia, noite, luz artificial, luz natural etc.), o tom da cor (HUE) possui uma informação mais relevante que a luminosidade (VALUE). Portanto, para maior precisão, é recomendado o padrão HSV (SWAIN ; BALLARD, 1991) com ênfase na componente H (HUE).

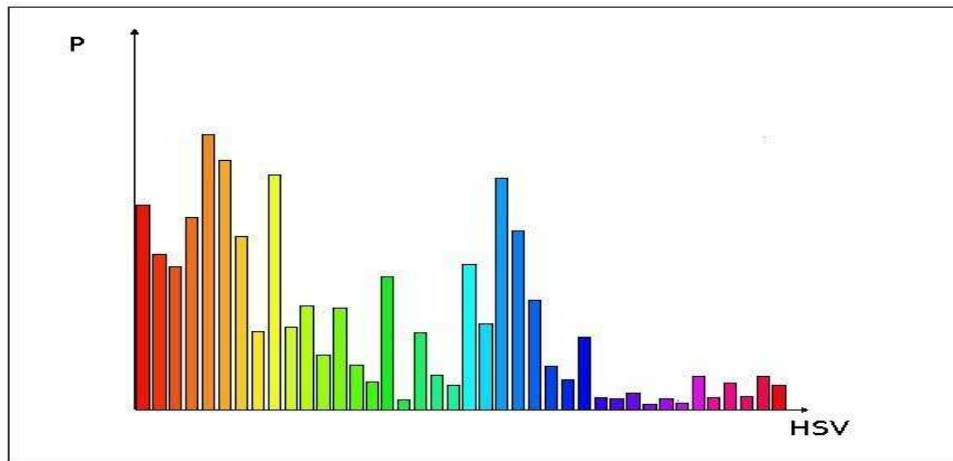


Figura 3 - Histograma de cores HSV com 40 raiais (*bins*)

Fonte: Elaborada pelo autor

3.2 Entropia de Shannon

O conceito de entropia foi inicialmente aplicado em termodinâmica, entretanto, em 1948, Claude Shannon (SHANNON, 1948), um dos precursores da teoria da informação, propôs que conceitos similares poderiam ser aplicados em estudos sobre transmissão de dados deduzindo a equação 2, onde S é a entropia e $p(i)$ a probabilidade de ocorrer a informação i . Esta equação é conhecida como entropia de Boltzmann-Gibbs-Shannon (GALÃO et al., 2010).

$$S = - \sum_i p(i) \ln p(i) \quad (2)$$

No campo do processamento de imagens, a entropia pode ser usada para quantificar o grau de dispersão ou concentração de cores e intensidades. Estudos aplicados em análises de imagens médicas relacionaram a entropia a características específicas de cada tipo de imagem.

A entropia se mostrou bastante eficiente na segmentação de imagens, quando deseja-se separar objetos em imagens. Este trabalho,, inicialmente propunha que a entropia de Shannon fosse concatenada ao histograma no intuito de apontar similaridades ou diferenças entre imagens que possuísem cores semelhantes, porém, na etapa de testes, ficou comprovado que o uso da entropia na classificação de gêneros não é muito eficiente, uma vez que o histograma já contém a informação a respeito da concentração das cores, o que torna, neste caso, a entropia uma informação redundante.

3.3 Distância de Kullback-Leibler

A distância de Kullback-Leibler (SOBRIECKI et al., 2010) também conhecida por entropia relativa (equação 3) é constantemente utilizada em assuntos relacionados à teoria da informação para quantificar semelhanças entre funções de distribuições e probabilidades. Como tal pode ser aplicada na comparação de histogramas já que estes, como mencionado anteriormente, são a representação da distribuição das probabilidades das cores de imagens.

$$D_{KL}(P, Q) = \sum_i \ln \frac{P(i)}{Q(i)} P(i) \quad (3)$$

Onde, $P(i)$ e $Q(i)$, representam a probabilidade da cor i incidir nas imagens P e Q , respectivamente. A principal característica da distância de Kullback-Liebler é ser não simétrica, isso a torna particularmente interessante na análise de cores já que as imagens coloridas, em sua grande maioria, permitem até 16 milhões de cores (24 bits ou 2^{24}). Porém, a quase totalidade das imagens contém apenas um pequeno subconjunto de poucos milhares de cores em uma mesma imagem. Desta observação pode-se deduzir que ao comparar duas imagens as cores semelhantes presentes em ambas têm um significado muito maior que as cores ausentes nas duas imagens.

Ao aplicar a distância de Kullback-Liebler em substituição à distância euclidiana esperava-se conseguir uma maior seletividade na classificação de vídeos, em especial com a diminuição de falsos positivos.

3.4 Máquinas de Vetores de Suporte (Support Vector Machines, SVMs)

Técnicas de classificação são uma importante área do aprendizado de máquina. Nas últimas duas décadas, intensas pesquisas levaram ao aperfeiçoamento tanto de métodos supervisionados como não supervisionados. Nos métodos supervisionados, exemplos pré-classificados são fornecidos de tal forma que o método possa ser calibrado em uma etapa conhecida como treinamento do método. Nas técnicas de aprendizado não supervisionado, o método deve procurar por características comuns aos elementos de forma a separá-los em classes sem exemplos prévios pré-classificados.

Dentre os métodos de aprendizado de máquina supervisionado, as máquinas de vetores de suporte (HASTIE, 2009) têm se destacado por conseguir uma eficiência elevada em casos onde outros métodos como redes neurais, apresentam limitações.

As máquinas de vetores de suporte são métodos, propostos por Vladimir Vapnik (CORTES ; VAPNIK, 1995), bastante eficientes na resolução de problemas tanto de classificação e regressão.

Em especial, as SVMs são indicadas nas aplicações que apresentam alta dimensionalidade e pequeno número de amostras. Estas características normalmente estão presentes na classificação de vídeos e, portanto, como já ficou demonstrado por inúmeras pesquisas, as SVMs são indicadas para estas aplicações.

Vapnik adota como princípio fundamental das SVMs o critério da máxima margem de separação, conceito que será detalhado nas próximas páginas.

Dado um número de elementos pertencentes a duas classes conhecidas, conforme mostrado na figura 4, representados por elementos na cor verde e na cor azul denominado conjunto de treinamento.

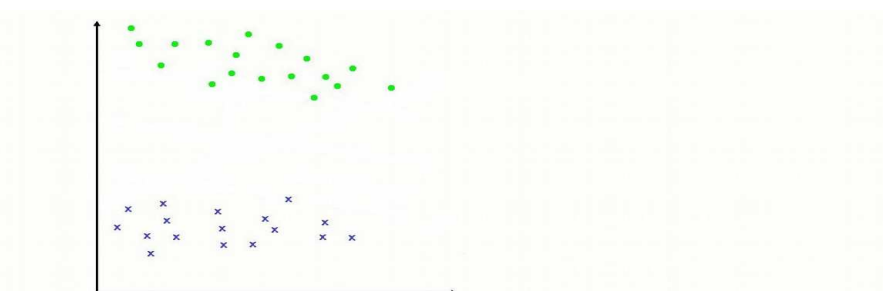


Figura 4 - Conjunto de treinamento formado elementos pertencentes a duas classes. No conjunto de treinamento as classes de todos os elementos devem ser previamente conhecidas.

Fonte: Elaborada pelo autor

Um classificador supervisionado eficiente deve, a partir do conjunto de treinamento, determinar um critério ou uma função capaz de classificar elementos ainda não classificados como os representados na cor amarela da figura 5.

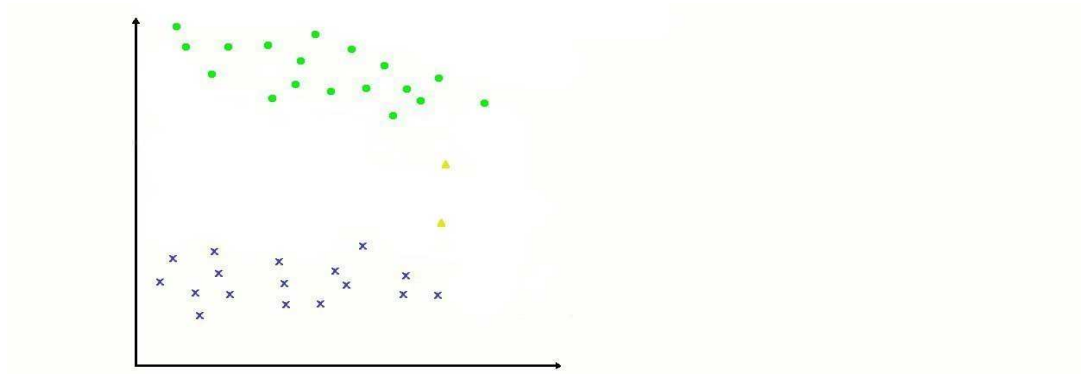


Figura 5 - Elementos do conjunto de treinamento em verde e azul, e dois novos elementos não classificados em amarelo.

Fonte: Elaborada pelo autor

Funções lineares, representadas na figura 6 por cada uma das retas em vermelho, podem ser usadas para separar os elementos ainda não classificados, porém existe um número teoricamente infinito de retas que, corretamente, classificam o conjunto de treinamento como exemplificado pelo conjunto de imagens da figura 6.

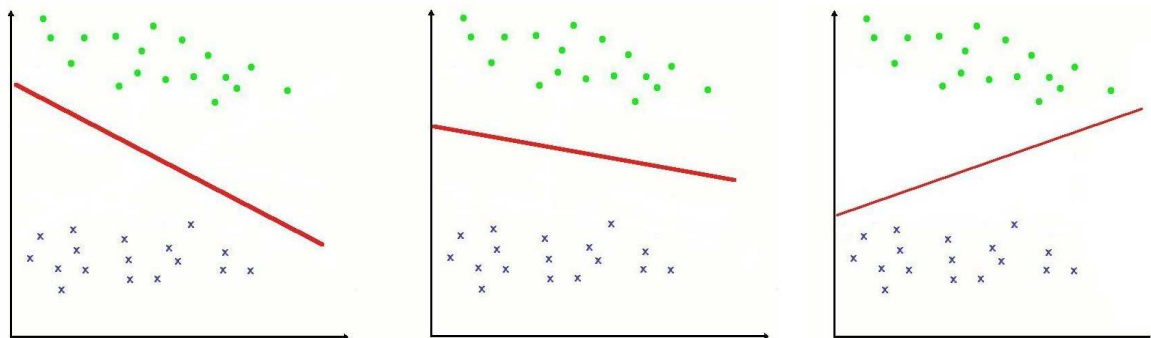


Figura 6 - Exemplo de 3 retas distintas que corretamente separam os elementos do conjunto de treinamento.

Fonte: Elaborada pelo autor

No entanto, no caso da classificação dos elementos ainda não classificados da figura 5, cada uma das retas apresentaria um resultado diferente como ilustrado na figura 7.

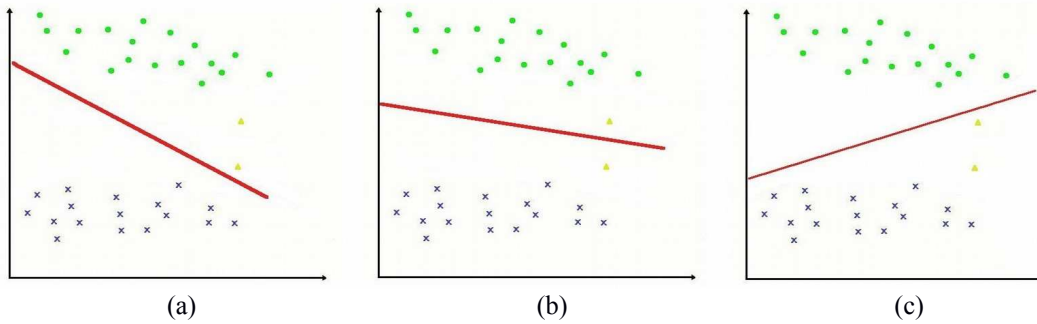


Figura 7 - Três funções lineares usadas na classificação dos elementos em amarelo. Em (a), ambos classificados como da classe verde, em (b), um em verde em azul, em (c), os dois elementos classificados como azuis.
Fonte: Elaborada pelo autor

Segundo o princípio da máxima separação, deve-se eleger a função que apresentar a maior distância (margem) entre as classes, como a apresentada no item b da figura 8.

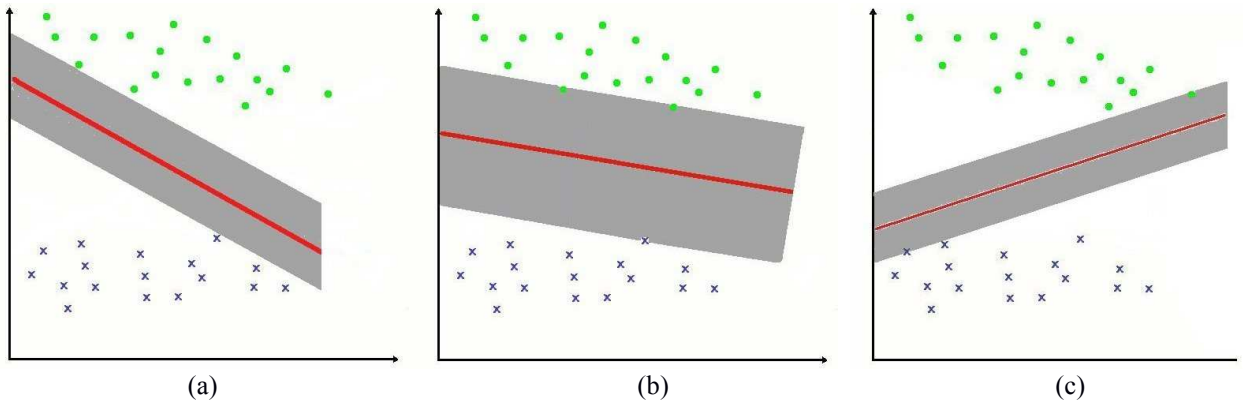


Figura 8 - Três funções lineares e suas respectivas margens de separação dos elementos do conjunto de treinamento. Nota-se que o gráfico de (b) apresenta maior margem e portanto segundo o critério de maior separação, é o mais eficiente para classificar novos elementos.
Fonte: Elaborada pelo autor

Para se determinar a função de decisão que atenda a abordagem da maior separação entre as classes (*widest street approach*), cada elemento de n é definido como um vetor x_i com $x_i \in \mathbb{R}^d$ e a classe a que pertence representada por y_i onde $y_i = +1$ ou $y_i = -1$, como definido na equação 4.

$$D = \{(x_i, y_i) \mid x_i \in \mathbb{R}^d, y_i \in \{-1, 1\}\}_{i=1}^n \quad (4)$$

Portanto, a resposta da função de decisão pode ser expressa usando apenas o sinal de sua saída como representado em (5) por $g(x)$.

$$g(x) = \text{sign}(f(x)) \quad (5) \quad \text{onde,} \quad f(x) = \langle w, x \rangle + b \quad (6).$$

Podemos notar em (6) que a função de decisão é parametrizada pelo vetor w , denominado vetor de pesos e pela constante b , sendo $\langle w, x \rangle$ o produto interno de w e x , cuja definição (equação 7) é dada por :

$$\langle w, x \rangle = \sum w_i \cdot x_i \quad (7)$$

Portanto, o problema de se determinar a função de decisão linear, também conhecida como hiperplano de classificação, no caso de uma função tridimensional, poderia ser enunciado como:

Tendo-se um conjunto de treinamento já classificada, constituído de vetores representados por $x_1, x_2, x_3, \dots, x_n$ pertencentes a uma de duas classes conhecidas representadas por $y_1, y_2, y_3, \dots, y_n$ cujos valores são $+1$ ou -1 , tem-se que determinar w e b em (6) que garanta a maior margem entre os elementos das duas classes.

Nota-se claramente que se trata de um problema de otimização, em que se deve descobrir os parâmetros que maximizem a distância entre as classes respeitando algumas restrições.

Na figura 9, pode-se observar todos os elementos necessários para esta otimização sendo w um vetor unitário perpendicular à função decisória que obedeça a restrição $y_i(\langle w, x_i \rangle + b) > 0$ e escalado de forma cujos elementos mais próximos (x_a e x_b) satisfaçam $|\langle w, x_i \rangle + b| = 1$. Estes vetores serão os vetores de suporte. Os elementos do conjunto de treinamento x_1 a x_n nas cores azul para $y_i = -1$ e verde para $y_i = +1$.

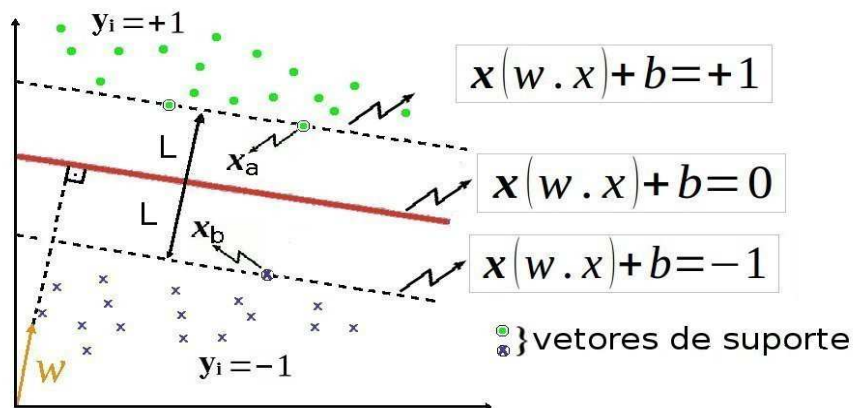


Figura 9 - Representação do conjunto de treinamento, vetores de suporte, vetor de pesos w , hiperplano de decisão e suas margens nas quais se encontram os vetores de suporte.

Fonte: Elaborada pelo autor

Conclui-se que a distância L pode ser determinada por $(X_a - X_b) \cdot W / |W|$ com X_a sendo um elemento exatamente na margem de $y=+1$ e X_b exatamente sobre $y=-1$, o que nos permite constatar que $L = 2 / \|w\|$. Portanto maximizar L é minimizar $|w|$ obedecendo-se a restrição $y_i((w \cdot x_i) + b) \geq 1$.

Como o número de elementos do conjunto de treinamento é finito e determinado, pode-se testar todos os elementos (x_i, y_i) e verificar quais resultam no menor $|w|$, sendo estes os vetores de suporte. A solução então assume a forma da equação 8.

$$f(x) = \sum_{k=1}^n w_k \cdot x_k \quad (8)$$

A equação acima tem solução por Lagrange (9).

$$L(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j k(\mathbf{x}_i, \mathbf{x}_j) \quad (9)$$

Onde, α_i, α_j são os coeficientes de Lagrange e $\mathbf{x}_i^T \mathbf{x}_j$ o produto interno dos vetores x_i, x_j . Pode-se notar na última parte da expressão a substituição do produto de $x_i \cdot x_j$ por uma função de núcleo também conhecida como kernel, $K(x_i, x_j)$ ou seja, uma função, que, neste caso,, é o produto interno entre os elementos de x_i e x_j , mas como será mostrado posteriormente, pode ser substituída por funções que aprimorem a classificação.

3.4.1 Margem Flexível

Como dentre os elementos do conjunto de treinamento existe a probabilidade de alguns deles estarem com sua classe erroneamente definida, um artifício foi adicionado por Cortez e Vapnik (CORTEZ; VAPNIK, 1995), possibilitando a admissão de que um determinado número de elementos do conjunto seja desconsiderado mediante uma penalização. Desta forma, pode-se ponderar a rigidez da margem através do uso de um

parâmetro conhecido como variável de folga (*slack variable*, ξ), reescrevendo a desigualdade relativa às margens para considerar esta variável como mostrado na equação (10).

$$y_i(\mathbf{w} \cdot \mathbf{x}_i - b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \quad (10)$$

3.4.2 Classificação Não Linear

Muitas vezes, no entanto, os elementos de um determinado problema, por sua natureza, não são linearmente separáveis. Como solução, o próprio Vapnik juntamente com outros pesquisadores (BOSER; GUYON; VAPNIK, 1992), baseados em estudos de Aizerman (AIZERMAN et al., 1964), introduziram métodos capazes de transformar o espaço das amostras do conjunto de treinamento, permitindo assim esta separação. Para tanto, utiliza-se um artifício matemático, aumenta-se a dimensionalidade do espaço das amostras como ilustrado na Figura 10. Nota-se claramente que, em um espaço de duas dimensões, as amostras não são separáveis linearmente, mas quando alocadas em um espaço tridimensional é possível separá-las com um hiperplano.

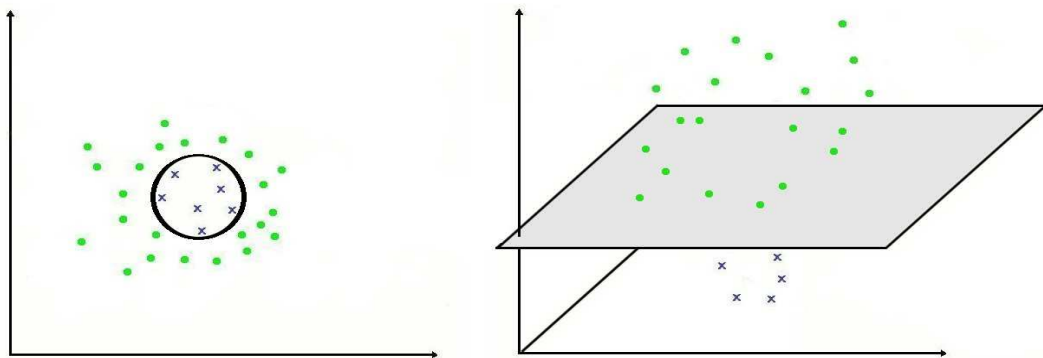


Figura 10 - Elementos não linearmente (a) separáveis remapeados para um espaço de maior dimensionalidade de forma a permitir a separação (b)

Fonte: Elaborada pelo autor

Este artifício, também conhecido como *kernel trick*, é conseguido substituindo-se a função de núcleo ou *kernel* por alguma função não linear. Dentre as funções elegíveis, as funções gaussianas de base radial como a da equação 11 se destacam por apresentarem bons resultados.

$$K(x_i, x_j) = e^{-\frac{(x_i - x_j)^2}{2\sigma^2}} \quad \text{Substituindo, } \gamma \text{ por } \frac{1}{2\sigma^2} \quad \text{Simplifica-se para } K(x_i, x_j) = e^{-\gamma(x_i - x_j)^2} \quad (11)$$

Estudando-se o comportamento desta função é possível entender sua eficiência. Sendo o parâmetro (γ) uma constante empírica, quanto mais similares os vetores X_i e X_j maior será o valor resultante, quanto mais distintos os valores de X_i e X_j menor seu resultado, com isto é possível quantificar a similaridade entre os vetores que, neste caso é a base da classificação.

Este trabalho propõe o aprimoramento desta classificação com a substituição do termo $(x_i - x_j)$, que representa a distância euclidiana entre os vetores, pela distância de Kulback-Liebler, já explicada em 3.3. Reescrevendo a função do núcleo para a equação 12.

$$K(x_i, x_j) = e^{-\gamma \left(\ln \left(\frac{x_i}{x_j} \right) x_i \right)^2} \quad (12)$$

3.4.3 Classificação Multi-classe

SVMs foram criadas inicialmente para classificação binária (duas classes), porém métodos multi-classe podem ser adaptados (CHIH-WEI; CHIH-JEN, 2002). Dois métodos que fazem uso de classificações binárias sucessivas são normalmente implementados: um contra um, e um contra todos, sendo que ambos podem ser aplicados na classificação de vídeos. No método um contra todos, um problema de K classes é subdividido em K problemas do tipo: pertence à classe k ou não. Já no método um contra um, $K*(K-1)/2$ classificações binárias são realizadas. Em geral, os dois métodos apresentam resultados similares, sendo a escolha normalmente uma questão de implementação. As ferramentas utilizadas neste trabalho dão preferência ao método um contra um, que então foi o adotado.

Neste trabalho é proposta a utilização da biblioteca LIBSVM⁴, um código acessório criado com o objetivo de auxiliar no desenvolvimento de aplicativos que necessitem implementar SVMs. Entre suas capacidades destacam-se compatibilidade com java e C++, classificação multi-classe, e por ser um programa livre de qualidade pode ser usado em pesquisas com excelentes resultados.

4 Página do projeto em (<http://www.csie.ntu.edu.tw/~cjlin/libsvm>)

3.5 Transições de Cenas

Vídeos quase sempre são compostos de várias cenas. Estudos anteriores (RASHEED et al., 2005) comprovaram uma forte relação entre características das transições dos vídeos com seu gênero. Muitos trabalhos utilizam o tempo médio entre as transições, que na prática equivale ao tempo médio das cenas do vídeo, como um parâmetro de classificação. Este trabalho propõe o uso não só do tempo médio, mas também de outros dados estatísticos das transições como o desvio padrão, a mediana e a moda para melhor distinguir as diferenças entre os gêneros analisados.

3.5.1 Quadros-chave

Denominam-se quadros-chave imagens que, por alguma de suas características, possam representar uma cena. Não existe um critério único para determinar qual imagem deva ser considerada como quadro-chave, costuma-se adotar o primeiro quadro da cena, um quadro aproximadamente no meio da cena, ou algum quadro que possua objetos de interesse da cena.

3.6 Ritmo Visual

Ritmo Visual é a denominação atribuída à sequência de mudanças bruscas nos aspectos visuais de um vídeo, normalmente está diretamente associada a transições de cenas. Alguns estudos relacionam o ritmo visual a aspectos semânticos do vídeo (CUTTING, 2011).

Dentre as inúmeras técnicas que podem ser adotadas para se determinar o ritmo visual (GUIMARAES et al., 2002), o método da amostragem é comumente usado. Neste método apenas uma amostra de cada quadro é analisada como, por exemplo, os pontos de uma das linhas diagonais de cada quadro, como na figura 11. Com isso a quantidade de pontos analisados é bem inferior ao total, porém bastante representativa.

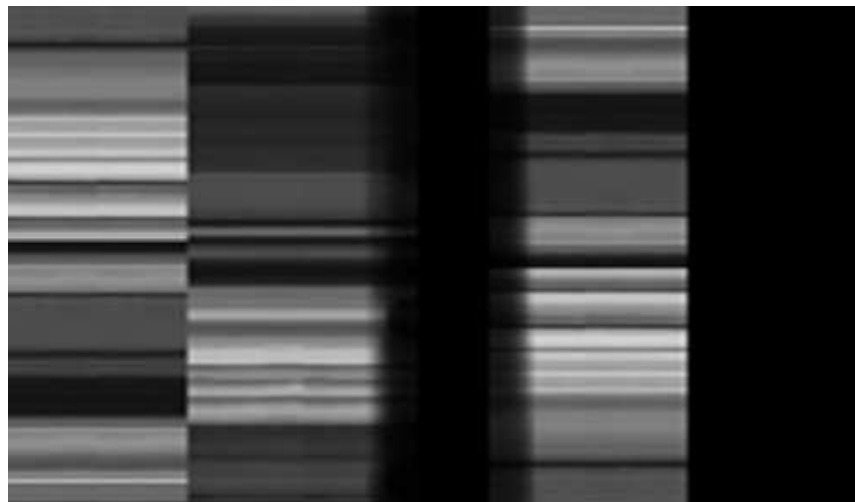


Figura 11 - Detecção de transições com o uso do ritmo visual por amostragem
Fonte: Elaborada pelo autor

Após a extração dos pontos da diagonal, normalmente um filtro passa-baixa é aplicado para eliminar ruídos, e o gradiente entre quadros consecutivos é monitorado. Caso um limiar determinado empiricamente seja ultrapassado, uma mudança de cena ocorreu. Um método análogo pode ser aplicado com os histogramas dos quadros em substituição ao diagonal, neste caso transformações geométricas como translações e rotações não são detectadas o que em alguns casos pode ser útil.

3.7 Linha do Tempo

As Linhas de Tempo, como a do exemplo da figura 12, são representações bastante utilizadas na análise de vídeos. Nelas as cenas são representadas por quadros-chave com a largura ajustada proporcionalmente à duração da cena. Deste modo, é possível identificar o número de cenas, que, neste caso são dez, além de outras informações relevantes como o tempo decorrido do início do vídeo (posicionado na parte superior da figura), e a duração de cada cena (neste caso na parte inferior).

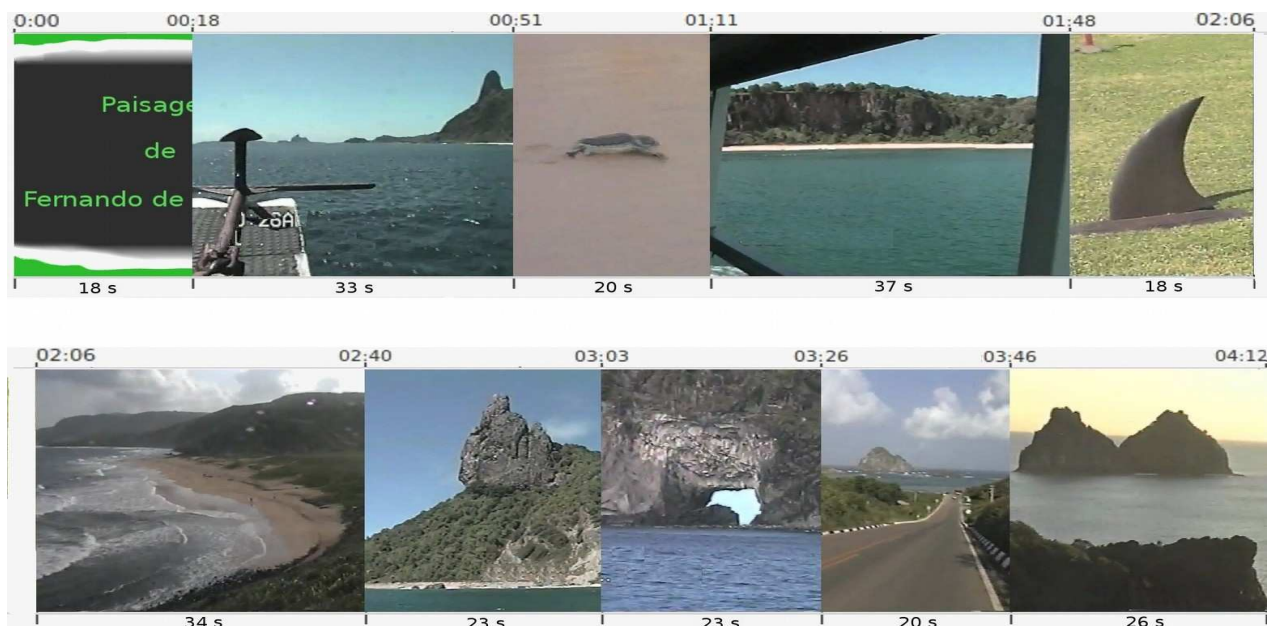


Figura 12 - Linha do tempo de um vídeo com dez cenas, os números na parte inferior indicam a duração de cada cena.

Fonte: Elaborada pelo autor a partir de vídeo de acervo pessoal

Uma vez determinados o número de cenas e suas respectivas durações em segundos, pode-se facilmente calcular os dados estatísticos com o uso das fórmulas estatísticas tradicionais. A tabela 2 mostra o resultado destas operações para o vídeo representado na linha do tempo da figura 12.

Tabela 2- Duração de cenas de um vídeo e seus os dados estatísticos básicos

Número da Cena	Último Quadro	Momento da Transição		Duração da Cena em Segundos	
Cena 1	306	0 mim 18 s	18 seg	18	Segundos
Cena 2	857	0 mim 51 s	51 seg	33	Segundos
Cena 3	1195	01 min 11 s	71 seg	20	Segundos
Cena 4	1813	01 min 48 s	108 seg	37	Segundos
Cena 5	2129	02 min 06 s	126 seg	18	Segundos
Cena 6	2693	02 min 40 s	160 seg	34	Segundos
Cena 7	3099	03 min 03 s	183 seg	23	Segundos
Cena 8	3477	03 min 26 s	206 seg	23	Segundos
Cena 9	3801	03 min 46 s	152 seg	20	Segundos
Cena 10	4246	04 min 12 s	252 seg	26	Segundos
Estatísticas dos Temporais das Cenas do Vídeo					
		Média	25,2	Segundos	
		Moda	18	Segundos	
		Mediana	23	Segundos	
		Desvio Padrão	6,6753277073	Segundos	

Fonte: Elaborada pelo autor

3.8 Bases de Vídeos

A escolha dos vídeos submetidos aos testes exigiu uma atenção especial, pois, além da necessidade de representarem com exatidão o gênero a que pertencem, devem ser bases reconhecidas, e estarem abertamente disponíveis. Um grande esforço de pesquisa foi realizado para que os testes fossem realizados com bases de vídeos que atendessem aos aspectos técnicos e de acessibilidade. A relação completa dos vídeos bem como a localização de suas fontes, onde estão livremente disponíveis, está no apêndice C. Como não foi possível conseguir vídeos de todos os gêneros de uma mesma base, foram usadas fontes distintas em função do gênero. Os vídeos utilizados do gênero FILMES pertencem à base utilizada na conferência TRECVID de 2002 e fazem parte do open vídeo project⁵, porém, nos demais gêneros, houve a necessidade de se recorrer a outras fontes. Nos gêneros MUSICAIS e JORNALISMO foram utilizados vídeos produzidos pela Fundação Padre Anchieta, mantenedora da Rádio e Televisão Cultura de São Paulo, que, por ser uma empresa pública, assumiu o compromisso de disponibilizar sua programação abertamente. Os gêneros ESPORTE e INFANTIL, foram selecionados entre os vídeos disponibilizados sob licença livre (*creative commons*) na biblioteca virtual vimeo⁶.

5 Página oficial (<http://www.open-video.org>)

6 Página oficial (<http://www.vimeo.com/creativecommons>)

4 METODOLOGIA EMPREGADA

Para concluir se as técnicas apresentadas no capítulo anterior, quando aplicadas na classificação de vídeos, alcançaram os resultados esperados, uma série de etapas, com determinados procedimentos foi realizada e cada uma das etapas será descrita nas próximas páginas e estão listadas na figura 13.

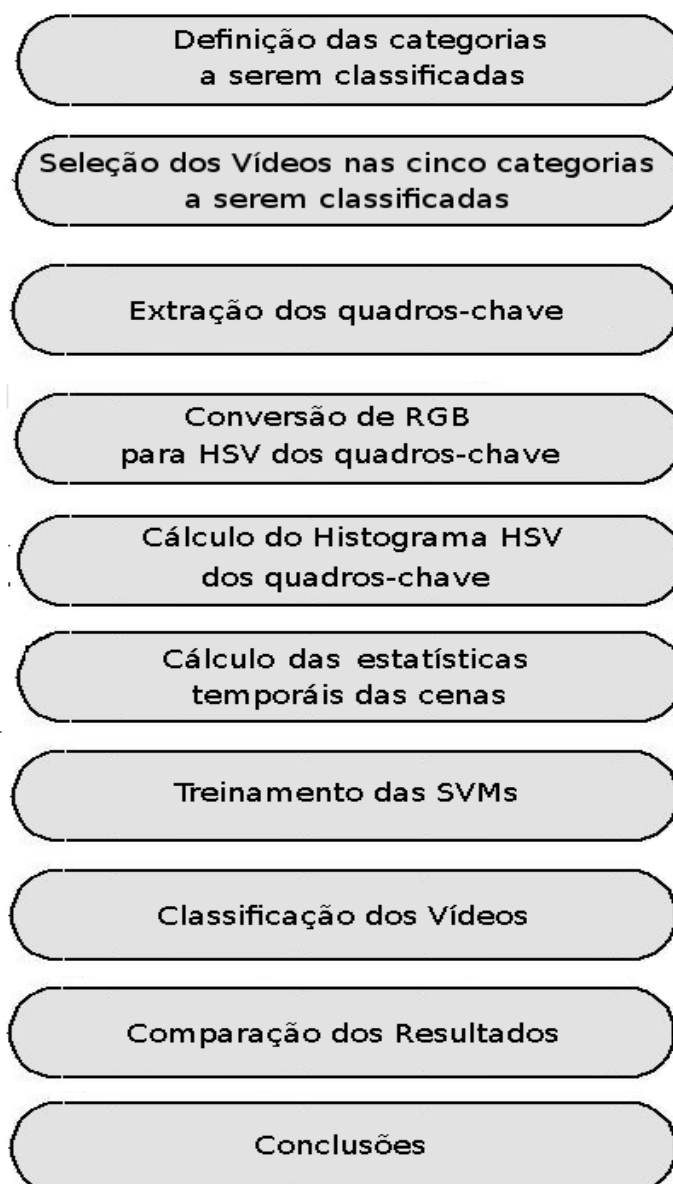


Figura 13- Principais etapas desenvolvidas durante os testes
Fonte: Elaborada pelo autor

Esta pesquisa tem como foco o aprimoramento das técnicas de classificação de vídeos, por isto, com base em estudos anteriores sobre taxonomia de vídeos (ROACH et al., 2002), foram determinadas as cinco categorias a serem classificadas sendo elas ilustradas na figura 14: esporte, notícias, musicais, infantil e filmes. Isto se deve ao fato de que mesmo não existindo um padrão ou consenso estas categorias estão presentes em diversos estudos correlatos. A partir delas foram selecionados trinta e oito vídeos distribuídos entre todas as categorias.



Figura 14 – Os cinco gêneros escolhidos para classificação dos vídeos.

Fonte: Elaborada pelo autor a partir de imagens dos vídeos analisados

De posse de todos os vídeos das cinco categorias, já devidamente classificados durante sua seleção, a extração dos quadros-chave foi realizada criteriosamente em um processo manual. A decisão pelo processo manual foi tomada, pois, mesmo existindo há algum tempo sistemas automáticos com boa eficiência, estes sistemas apresentavam, em maior ou menor grau, alguma possibilidade de erro, e, se usados, estes potenciais erros seriam somados ao resultado dos testes. Portanto, optou-se pela seleção manual dos quadros-chaves, como uma forma de, se não excluir totalmente, pelo menos minimizar ao máximo a possibilidade de algum erro causado na determinação dos quadros-chave. Qualquer imprecisão, por menor

que fosse, iria acarretar na alteração dos resultados dos testes referentes à classificação, que são o alvo principal deste estudo. Outro aspecto importante foi o critério de escolha dos quadros-chave. Dentre os diversos critérios possíveis, foi estipulado que seria considerado quadro-chave a primeira imagem completa da cena, ou seja, o primeiro quadro após o final da transição entre duas cenas consecutivas. Isto porque transições de cenas, como fusões ou efeitos, não são padronizados e prejudicariam a classificação.

Terminada a escolha dos quadros-chave, que, pelo critério de escolha adotado, é sempre o quadro inicial da cena, foi possível determinar o tempo de duração de cada uma das cenas do vídeo que está sendo analisado. Para tanto, basta considerar o tempo decorrido entre a exibição de dois quadros-chave consecutivos. Outro critério adotado foi a limitação da análise aos trezentos primeiros segundos de cada vídeo, ou às primeiras trinta cenas, o que ocorrer primeiro. Isto se deve ao fato de que a comparação de vídeos de duração distintas causariam distorções na análise, uma vez que um maior número de cenas e consequentemente um maior número de quadros-chave acabariam por gerar histogramas médios (que são os histogramas gerados pela média de todos os histogramas dos quadros-chave de um mesmo vídeo) mais espalhados, pois a duração e a quantidade de cenas têm impacto direto no número de cores acumuladas durante a exibição.

Calculada a duração de cada uma das cenas do vídeo em análise, a próxima tarefa seria a geração dos histogramas de cada um dos quadros-chave. Porém como mencionado no item 3.1.2, o uso do espaço de cores HSV apresenta vantagens sobre o espaço RGB no que se refere ao agrupamento das cores por tonalidade, portanto, os quadros-chave depois de extraídos são convertidos para tal espaço para que estes histogramas possam ser gerados. Assim que todos os quadros-chave de um mesmo vídeos, são convertidos, são gerados histogramas com cento e sessenta e duas raiais ou *bins*, resultando em um vetor de cento e sessenta e dois valores para cada quadro. Outro ponto importante em relação aos histogramas é a normalização. É crucial que os histogramas sejam normalizados, ou seja, seus valores representam probabilidades e, portanto, devem estar sempre entre zero e um. Isto garante que mesmo vídeos de características geométricas diferentes possam ser comparados. Desta forma características como: tamanho em pontos, relação de aspecto, resolução ou orientação não interferem no resultado do histograma. Caso não fossem normalizados haveria a necessidade de processamento adicional como redimensionamento ou conversões de formato.

Como o número de quadros-chave varia em função da duração de cada cena dos vídeos, cada vídeo pode possuir um número diferente de quadros-chave. Os diversos

histogramas de um mesmo vídeo são então somados e divididos pelo número de quadros-chave, gerando assim um único histograma por vídeo, que contém a média dos histogramas relativos a ele. Valores muito pequenos desta média são convertidos a zero para se eliminar cores espúrias ou decorrentes de ruído.

Paralelamente, os valores da duração das cenas são utilizados nos cálculos estatísticos da média, mediana, moda e desvio padrão referentes à duração das cenas cada vídeo. Estes dados são, então, concatenados ao vetor da média dos histogramas gerando um vetor de cento e sessenta e dois valores de histogramas de cores, mais quatro valores referentes à duração das cenas, totalizando um vetor de cento e sessenta e seis elementos. Um elemento adicional, contendo um valor de um a cinco é acrescentado identificando o gênero do vídeo em questão (um exemplo de histograma é mostrado na figura 15). Este valor será usado na fase de classificação explicada mais adiante.

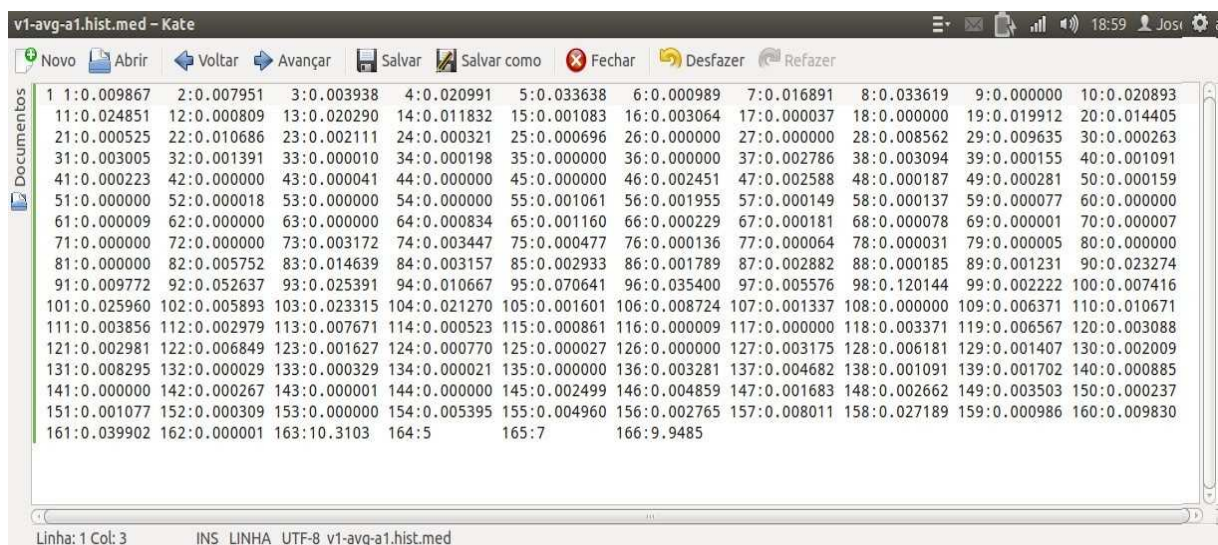


Figura 15 – Histograma médio de um vídeo

Fonte: Elaborado pelo autor no decorrer da pesquisa

O processo é repetido para cada um dos vídeos, gerando assim um vetor contendo as informações de cor e tempo extraídas de cada um deles.

As tarefas executadas até a geração dos vetores constitui a etapa de pré-classificação, que tem como único intuito coletar e formatar os dados para a etapa de classificação.

Uma vez que os dados de todos os vídeos foram coletados e formatados na forma vetorial apropriada, tem início a etapa de treinamento da máquina de vetores de suporte ou SVM. Como visto na seção 3 esta etapa consiste em determinar, através de operações algébricas, quais elementos do conjunto de treinamento serão aqueles que ajudarão na

determinação do hiperplano que garanta a melhor separação entre as classes, que, neste caso, por ser uma SVM é o hiperplano com maior margem de separação.

O treinamento começa com a separação aleatória de uma parte (aproximadamente metade) dos vetores de cada uma das classes dos vídeos. Estes vetores são agrupados em uma matriz e formarão o conjunto de treinamento. Os demais vetores serão usados posteriormente na etapa de teste, gerando outra matriz de igual tamanho. Cabe lembrar que nesta pesquisa todos os vídeos foram previamente classificados, isto por que o objetivo não é desenvolver um classificador, mas sim analisar quais características tornam os classificadores mais eficientes. Os vetores são submetidos então às ferramentas desenvolvidas com o uso da biblioteca LibSVM. Estas ferramentas foram desenvolvidas de modo a permitir o treinamento sequencialmente, variando os valores das variáveis arbitrarias de GAMA e C e alterando as funções do núcleo da SVM, explicados no item 3.4.2.

Após a tarefa de treinamento, o restante dos vetores, ou seja, a outra metade, é testada na fase de classificação propriamente dita, realizada diversas vezes com os classificadores gerados pelos diferentes parâmetros e por diferentes funções de núcleo ou Kernéis. Como todos os vetores, e não só os do conjunto de treinamento, já estavam previamente classificados, estas mesmas ferramentas são capazes de medir precisamente o grau de acerto dos classificadores em função de seus diferentes parâmetros. Portanto a etapa de classificação resumidamente é responsável pelo treinamento, classificação e avaliação dos diferentes parâmetros testados para as SVMs.

Ao final dos testes o índice de acerto alcançado para cada tipo de núcleo e para os diversos valores dos parâmetros GAMA e C são relatados e registrados. Além deles são testados também vetores com e sem os dados estatísticos, para se estudar a influência destes dados no resultado da classificação.

Com o resultado dos testes tem início a etapa de análise. Nesta etapa, todos os resultados dos testes são organizados em tabelas numéricas, tornando possível uma análise bastante objetiva, que fornece o embasamento necessário para tecer conclusões consistentes que corroborem as teorias propostas.

5 TESTES REALIZADOS

Para obter todos os dados necessários foram utilizadas as seguintes ferramentas computacionais:

- a) A biblioteca de visão computacional OpenCV⁷;
- b) A biblioteca para implementação de Máquinas de Vetores de Suporte LibSVM;
- c) O sistema operacional livre Linux⁸;
- d) O exibidor de vídeos Mplayer⁹.

Algumas rotinas foram desenvolvidas especificamente para este trabalho. Uma ferramenta em especial, a LibSVM foi adaptada para implementar a distância de Kullback-Liebler, explicada no item 3.3. Os códigos das ferramentas desenvolvidas ou adaptadas nesta pesquisa estão impressos no Apêndice A. As demais ferramentas bem como os vídeos utilizados estão contidos no DVD anexo ou podem ser facilmente conseguidos pelos mecanismos de busca da internet. O Apêndice C contém a lista completa dos vídeos classificados por gênero com respectivo endereço eletrônico e sua fonte.

Seguindo as etapas descritas na seção 4 os testes começam com a extração manual dos quadros-chave. Como mencionado na metodologia, são extraídos os quadros-chave das primeiras trinta cenas do vídeo ou dos primeiros trezentos segundos de cada vídeo, dependendo do que ocorrer primeiro. Este critério foi adotado para evitar um número muito grande de quadros-chave de um mesmo vídeo, que como já foi dito causa uma distorção no histograma médio. Para tanto é executada a rotina cortavid.sh que seleciona os trezentos segundos iniciais do vídeo, e o exibidor mplayer (figura 16) por meio do comando grabframe, extrai o quadro corrente, indica o tempo decorrido desde o início do vídeo e envia a imagem para o programa geraHist, que calcula o histograma deste quadro. A medida que os quadros-chave são extraídos pelo programa mplayer o tempo de exibição é armazenado no arquivo de saída, que, posteriormente, usa a rotina CalcEstatisticas.sh para gerar as estatísticas da duração das cenas deste vídeo resultando em quatro números referentes à média, mediana, moda e ao desvio padrão.

7 Página oficial (<http://opencv.org>)

8 Página oficial (<http://www.linux.org>)

9 Página oficial (<http://www.mplayerhq.hu>)

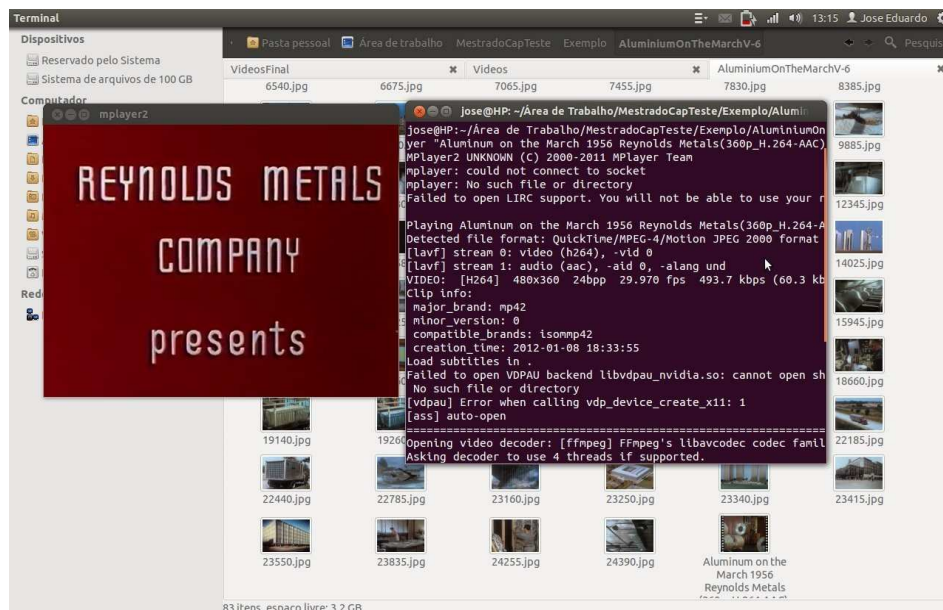


Figura 16 - Seleção dos quadros-chave com o programa mplayer.
Fonte: Imagem gerada pelo autor no desenvolvimento dos testes

Ao final da exibição dos trezentos segundos iniciais do vídeo ou caso o limite de trinta quadros-chave tenha sido atingido, tem-se um arquivo com os histogramas de todos os quadros-chave de um mesmo vídeo, sendo cada quadro representado por uma linha contendo o número da classe, e os cento e sessenta e dois elementos referentes às raiais do histograma. Um exemplo deste arquivo pode ser visto na figura 17. Nota-se pelo valor inicial de cada linha que o vídeo em questão é da classe 1, correspondendo ao gênero Filme, e possui vinte e nove quadros-chave nos seus primeiros trezentos segundos.

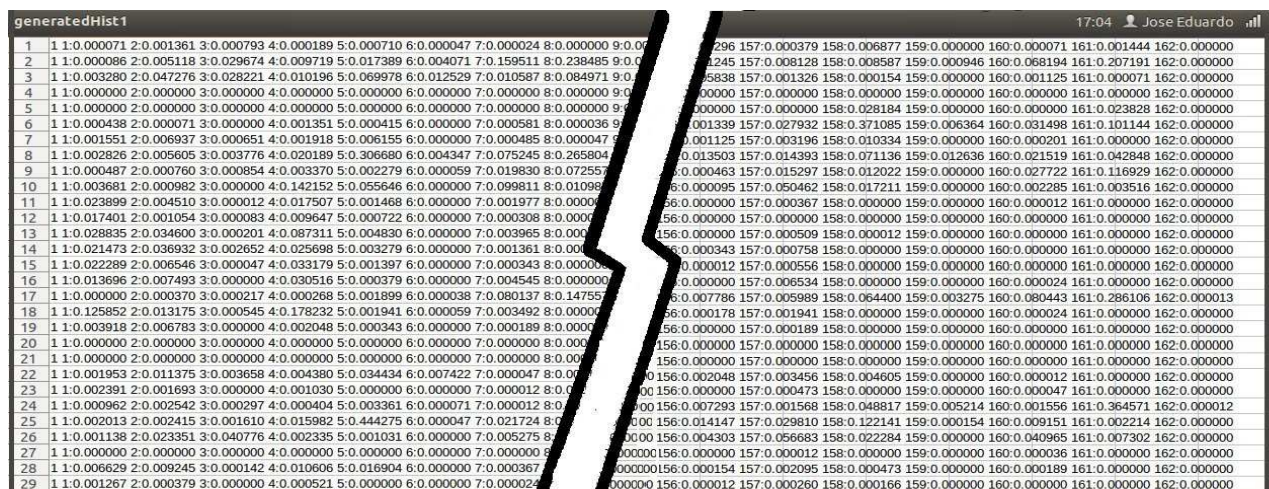
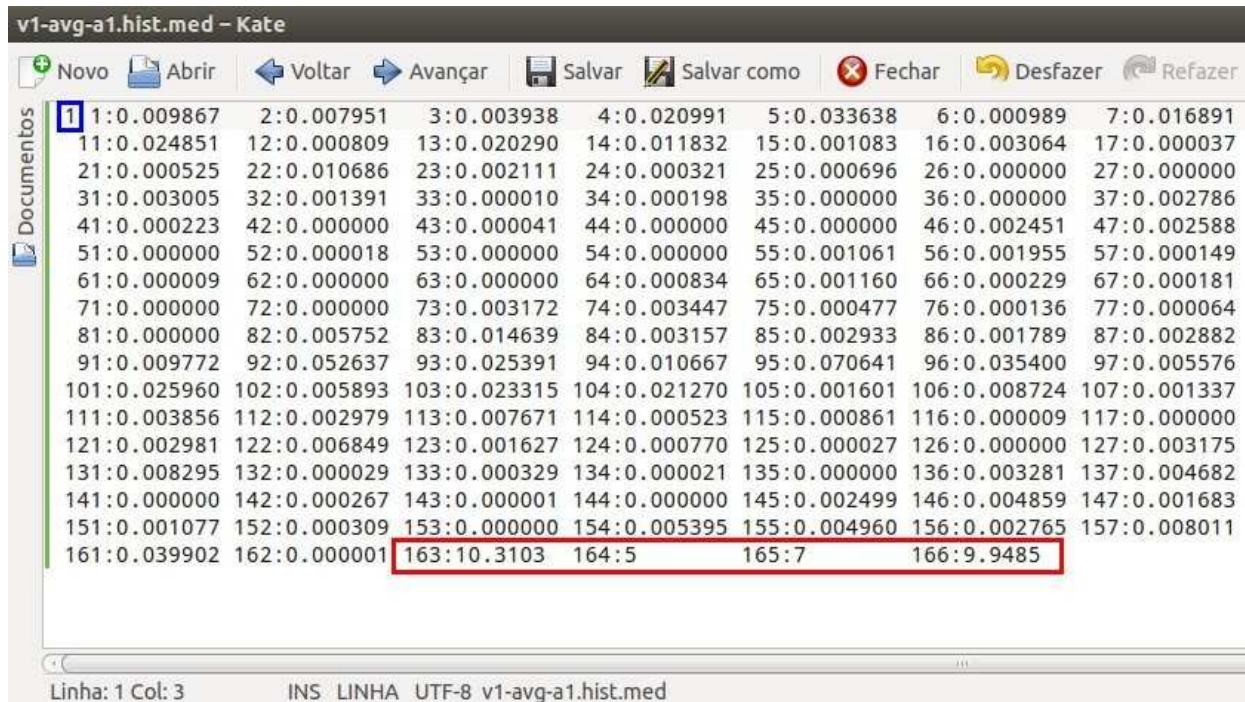


Figura 17 – Arquivo contendo os 29 histogramas de um mesmo vídeo.
Fonte: Exemplo de arquivo de histograma elaborado pelo autor no desenvolvimento da pesquisa.

Como cada vídeo deve ser representado por um único vetor, a média dos elementos de todos quadros-chave é calculada pelo programa `geraMediaEst.sh` que o converte em um arquivo de um único vetor em uma única linha, já com os quatro elementos estatísticos inseridos nos últimos quatro elementos como destacado na figura 18.



1:0.009867	2:0.007951	3:0.003938	4:0.020991	5:0.033638	6:0.000989	7:0.016891
11:0.024851	12:0.000809	13:0.020290	14:0.011832	15:0.001083	16:0.003064	17:0.000037
21:0.000525	22:0.010686	23:0.002111	24:0.000321	25:0.000696	26:0.000000	27:0.000000
31:0.003005	32:0.001391	33:0.000010	34:0.000198	35:0.000000	36:0.000000	37:0.002786
41:0.000223	42:0.000000	43:0.000041	44:0.000000	45:0.000000	46:0.002451	47:0.002588
51:0.000000	52:0.000018	53:0.000000	54:0.000000	55:0.001061	56:0.001955	57:0.000149
61:0.000009	62:0.000000	63:0.000000	64:0.000834	65:0.001160	66:0.000229	67:0.000181
71:0.000000	72:0.000000	73:0.003172	74:0.003447	75:0.000477	76:0.000136	77:0.000064
81:0.000000	82:0.005752	83:0.014639	84:0.003157	85:0.002933	86:0.001789	87:0.002882
91:0.009772	92:0.052637	93:0.025391	94:0.010667	95:0.070641	96:0.035400	97:0.005576
101:0.025960	102:0.005893	103:0.023315	104:0.021270	105:0.001601	106:0.008724	107:0.001337
111:0.003856	112:0.002979	113:0.007671	114:0.000523	115:0.000861	116:0.000009	117:0.000000
121:0.002981	122:0.006849	123:0.001627	124:0.000770	125:0.000027	126:0.000000	127:0.003175
131:0.008295	132:0.000029	133:0.000329	134:0.000021	135:0.000000	136:0.003281	137:0.004682
141:0.000000	142:0.000267	143:0.000001	144:0.000000	145:0.002499	146:0.004859	147:0.001683
151:0.001077	152:0.000309	153:0.000000	154:0.005395	155:0.004960	156:0.002765	157:0.008011
161:0.039902	162:0.000001	163:10.3103	164:5	165:7	166:9.9485	

Figura 18 – Arquivo com vetor referente à média dos histogramas dos quadros-chave: elemento em azul indica classe a que pertence o vídeo e elementos em vermelho dados estatísticos (média, media, moda e desvio padrão do ritmo visual).

Fonte: Elaborado pelo autor

Neste ponto tendo todos os vetores de todos os vídeos já classificados, faz-se a separação aleatória de dois grupos de igual tamanho, divide-se aproximadamente metade dos vetores de cada classe para o grupo de treinamento e a outra metade para o grupo de testes.

Com os grupos definidos, os vetores de cada grupo são concatenados, gerando dois arquivos. Um com todos os vetores do grupo de treinamento e outro com todos os vetores do grupo de testes.

Estes dois arquivos formam a base de dados usada pela biblioteca LibSVM no treinamento e nos testes de classificação descritos a seguir.

Os testes de classificação com a biblioteca LibSVM são compostos por duas etapas, o treinamento executado pela função `svm-train`, e a classificação executada pela função `svm-predict`.

Na primeira etapa, são definidos os parâmetros de tipo de função de núcleo, constante de custo C , valor do coeficiente γ , juntamente com os arquivo contendo os vetores de

teste. A função então processa estes dados gerando um arquivo de modelo. Este arquivo contém as informações dos hiperplanos separadores que delimitam o espaço de cada classe.

A etapa seguinte faz uso da função `svm-predict` do arquivo de modelo gerado durante a etapa anterior e do arquivo contendo os vetores a serem testados. Nesta etapa os vetores do grupo de testes são testados, comparados aos modelos e sua provável classe definida.

Tão logo todos os vetores do conjunto de testes tenham sido classificados esta mesma função compara o resultado com o valor da classe previamente definida, medindo assim o nível de eficiência deste modelo e apresentando sua porcentagem de acerto. A figura 19 apresenta em exemplo de execução do treinamento cujo comando é executado na linha: `./libsvm/svm-train -s 0 -t 2 -c 16 -g 0.125 DTreino.txt`

```

Jose@HP: ~/Area de Trabalho/MestradoCapTeste/LibSVM
$ ./libsvm/svm-train -s0 -t2 -c16 -g0.125 DTreino.txt
*
optimization finished, #iter = 6
nu = 0.857143
obj = -5.929863, rho = -0.976353
nSV = 7, nBSV = 4
*
optimization finished, #iter = 4
nu = 1.000000
obj = -7.824221, rho = -0.026376
nSV = 8, nBSV = 8
*.*
optimization finished, #iter = 9
nu = 0.857143
obj = -5.959054, rho = -0.978654
nSV = 7, nBSV = 4
*
$
$ ./libsvm/svm-predict DPredic.txt DTreino.txt.model output
Accuracy = 33.3333% (6/18) (classification)
$

```

Figura 19 – Comandos de treinamento e classificação com o uso da biblioteca LibSVM.

Fonte: Elaborada pelo autor no decorrer dos testes

Onde, o parâmetro `-s 0` refere-se ao tipo de classificação, no caso classificação multi-classe, o parâmetro `-t 2` determina o tipo de função de núcleo que será usada, no caso para função de base radial, o parâmetro `-c` trata a constante de custo, e `-g` a constante gama, por fim, `DTreino.txt` é o arquivo texto que contém os vetores de treinamento. Ao final da execução da função `svm-train` o arquivo de modelo é gerado com o mesmo nome do arquivo dos vetores de treinamento acrescido da terminação `.model`.

Na sequência, o comando para executar a classificação aparece na penúltima linha, onde se vê: `./libsvm/svm-predict DTreino.txt.model DPredic.txt`, responsável pela classificação dos vetores contidos no arquivo `DPredic.txt`, segundo o modelo

DTreino.txt.model gerado na etapa anterior. A resultado da classificação é apresentado pelos valores da linha seguinte, em percentual de acertos, e pela razão de elementos corretamente classificados sobre total de elementos processados. No exemplo da figura 19 a eficiência atingida foi de 33,3333% com 6 acertos de 18 classificações.

Como o número de combinações de C e Gama é de algumas centenas, a execução manual destes comandos seria demasiadamente trabalhosa, conseqüentemente este processo foi automatizado com o desenvolvimento de uma rotina para esta execução. O resultado desta rotina pode ser visto na figura 20.

```
jose@HP: ~/Área de Trabalho/Mestrado/LibSVM
Accuracy = 50% (9/18) (classification)
Classificado com Valores de C=16 e Gama=0.00390625 K-RBF
--
Accuracy = 27.7778% (5/18) (classification)
Classificado com Valores de C=8 e Gama=4 K-RBF
--
Accuracy = 33.3333% (6/18) (classification)
Classificado com Valores de C=8 e Gama=2 K-RBF
--
Accuracy = 44.4444% (8/18) (classification)
Classificado com Valores de C=8 e Gama=1 K-RBF
--
Accuracy = 55.5556% (10/18) (classification)
Classificado com Valores de C=8 e Gama=0.5 K-RBF
--
Accuracy = 50% (9/18) (classification)
Classificado com Valores de C=8 e Gama=0.25 K-RBF
```

Figura 20 – Trecho do resultado da classificação com diferentes valores de C e Gama.
Fonte: Elaborado pelo autorais

Para facilitar a análise dos resultados, seus valores foram inseridos em planilhas, o que permitiu, entre outros recursos, a geração de gráficos.

5.1 Resultados Obtidos

Os resultados obtidos indicaram um razoável índice de acerto, que, no melhor dos casos atingiu 55%. Deve-se levar em conta que por se tratar de uma classificação com cinco classes, e o número de vídeos de cada classe ser praticamente o mesmo, o valor esperado aleatoriamente está próximo a 0,2, ou seja 20%. Portanto, podemos considerar que o

mecanismo de classificação atua no sentido de aprimorar a classificação bem acima do aleatório. No apêndice A, os resultados completos são detalhadamente listados.

As expectativas de melhora de eficiência com o uso dos dados estatísticos e com a implementação da distância de Kullback-Liebler não se confirmaram e são comentadas nas conclusões.

Os números indicaram que as funções de base radial, na média, obtiveram maior número de acertos, em seguida as funções lineares e, por último, mas não muito inferior, a implementação de Kullbak-Liebler. Porém, é importante mencionar que as funções lineares apresentaram custo computacional pequeno em comparação com as demais. A implementação de Kullback-Liebler, por conter operações com logaritmos, exigiu um tempo de processamento significativamente maior.

Os dados estatísticos não contribuíram para um maior índice de acertos e, por muitas vezes, prejudicaram a classificação, interferindo negativamente no resultado. Considerações a este respeito também estão contidas nas conclusões.

Valor de C	Linear s/ Dados Est.	Linear c/ Dados Est.
0,00048828125	33,3333	27,7778
0,0009765625	33,3333	27,7778
0,001953125	33,3333	27,7778
0,00390625	33,3333	27,7778
0,0078125	33,3333	27,7778
0,015625	33,3333	27,7778
0,03125	33,3333	44,4444
0,0625	33,3333	38,8889
0,125	33,3333	33,3333
0,25	33,3333	38,8889
0,5	33,3333	33,3333
1	33,3333	38,8889
2	33,3333	33,3333
4	33,3333	38,8889
8	33,3333	38,8889
16	33,3333	38,8889
32	44,4444	38,8889
64	55,5556	38,8889
128	55,5556	38,8889
256	55,5556	38,8889
512	55,5556	38,8889
1024	55,5556	38,8889
2048	55,5556	38,8889
4096	55,5556	38,8889

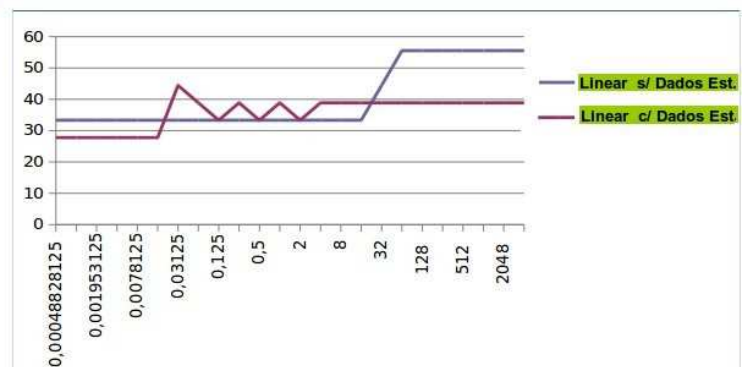


Figura 21 – Resultados da classificação com funções de núcleo linear com e sem o uso dos dados referentes as estatísticas relacionadas ao ritmo visual.

Fonte: Elaborada pelo autor

A figura 21 apresenta os resultados da classificação para a função de núcleo linear com e sem os dados estatísticos referentes ao ritmo visual, juntamente com o gráfico correspondente.

A tabela 3 apresenta apenas os dados referentes às estatísticas do ritmo visual por classe relacionando-os aos gêneros. Esta tabela tem por intuito demonstrar a proximidade dos valores entre os diversos gêneros. Nota-se que os valores da moda, mediana e desvio padrão são semelhantes, não contribuindo assim na distinção das classes. O único valor que apresenta variações significativas é o da média da duração das cenas.

Tabela 3- Estatísticas referentes ao ritmo visual por classe ou gênero.

Estatísticas por Classe				
Classe	Média	Moda	Mediana	Desv. Pad.
1	7,69231	4	6	5,06954
1	9,73333	5	6	10,7453
1	10,5926	10	10	7,97338
1	8,18919	3	7	5,09816
1	7,87179	4	5	6,1485
1	15,0952	5	9	12,5619
V.Médio	9,86	5,17	7,17	7,93
3	14,8421	9	9	17,1932
3	5,875	4	5	3,89778
3	3,27778	1	3	3,01488
3	5,12	5	5	3,12819
3	5,42857	6	6	2,27453
3	9,84	3	5	15,9541
3	3,68182	3	3	2,30388
3	1,67176	1	1	1,50542
3	7,86842	7	7	4,93735
V.Médio	6,40	4,33	4,89	6,02
5	5	4	5	2,44949
5	3,44828	3	3	1,32837
5	17,9412	15	15	8,36288
5	8,88571	12	9	3,69379
5	4,73333	4	4	2,63228
5	21,4	19	19	13,5243
V.Médio	10,23	9,50	9,17	5,33

Classe	Média	Moda	Mediana	Desv. Pad.
2	6,4375	2	4	7,14699
2	5,77778	3	4,5	5,38574
2	6,1	4	5	5,47083
2	4,21127	3	3	3,21929
2	9,36364	5	5	10,5394
2	8,38889	3	7	6,75885
2	3,86076	2	3	3,32895
2	6,7907	3	6	5,96719
V.Médio	6,37	3,13	4,69	5,98
4	4,5	2	3	4,83662
4	4,4878	1	2	5,23181
4	8,75	18	5,5	7,42322
4	8,625	2	3	13,6605
4	4,21127	2	2	4,66004
4	3,54839	2	3	2,74554
V.Médio	5,69	4,50	3,08	6,43

Classe	Gênero
1	Filme
2	Musica
3	Infantil
4	Noticias
5	Esportes

Fonte: Elaborada pelo autor.

A tabela 4 apresenta um resumo dos resultados em seis colunas, representando as três diferentes funções de núcleo, com e sem o uso dos dados estatísticos do ritmo visual. Os melhores resultados foram alcançados com valores do fator Gama em torno de 0,5, portanto, no resumo somente este valor foi selecionado. A tabela completa encontra-se no Apêndice A.

Tabela 4 – Resultados com alcançados em função da constante C.

RESULTADOS

Constantes		Tipo de Núcleo					
Custo	Gama	RBF Class s/E	RBF Class c/E	Kullback Class. c/E	Kullback Class s/E	Linear Class s/E	Linear Class c/E
1	0,5	33,3333	50	50	27,7778	33,3333	38,8889
2	0,5	33,3333	55,5556	44,4444	33,3333	33,3333	33,3333
4	0,5	33,3333	55,5556	44,4444	33,3333	33,3333	38,8889
8	0,5	33,3333	55,5556	44,4444	33,3333	33,3333	38,8889
16	0,5	33,3333	55,5556	44,4444	50	33,3333	38,8889
32	0,5	44,4444	55,5556	44,4444	38,8889	44,4444	38,8889
64	0,5	55,5556	55,5556	44,4444	38,8889	55,5556	38,8889
128	0,5	55,5556	55,5556	44,4444	38,8889	55,5556	38,8889
256	0,5	55,5556	55,5556	44,4444	38,8889	55,5556	38,8889
512	0,5	55,5556	55,5556	44,4444	38,8889	55,5556	38,8889
1024	0,5	55,5556	55,5556	44,4444	38,8889	55,5556	38,8889
2048	0,5	55,5556	55,5556	44,4444	38,8889	55,5556	38,8889
4096	0,5	55,5556	55,5556	44,4444	33,3333	55,5556	38,8889

Fonte: Elaborada pelo autor

A Figura 22 apresenta o gráfico referente à tabela acima.

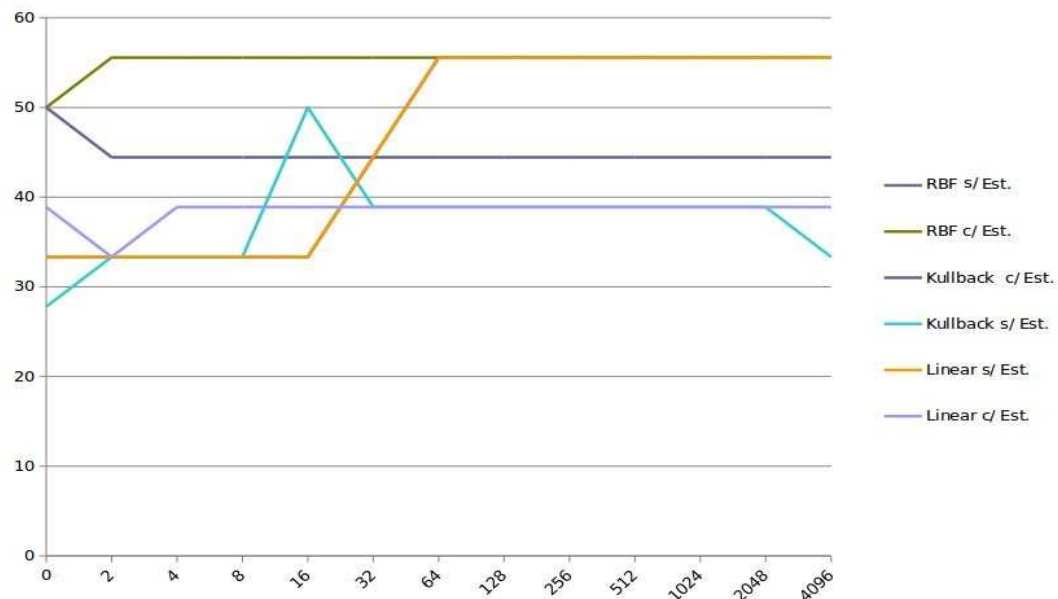


Figura 22 – Porcentagem de acertos dos diversos tipos de de classificação em função da constante C.

Fonte: Elaborada pelo autor.

6 CONCLUSÕES

Este trabalho teve como objetivo o estudo das técnicas de classificação de vídeo mais utilizadas, e propor possíveis aprimoramentos a estas técnicas. Os histogramas de cores há bastante tempo são as ferramentas mais utilizadas em aplicações que necessitam analisar imagens e vídeos, e por isto foram a base deste estudo. As Máquinas de Vetores de Suporte (SVMs) são comprovadamente eficientes nas tarefas de classificação de vídeos. Já a distância de Kullback-Liebler e o ritmo visual aparecem em trabalhos mais recentes, e não se tinha conhecimento amplo de seu uso na classificação de vídeos de entretenimento, por isto sua inclusão neste trabalho.

Com base nos resultados foi possível comprovar a eficiência dos histogramas e a versatilidade das Máquinas de Vetores de Suporte, corroborando trabalhos anteriores mencionados na secção 2.

A distância de Kullback-Liebler, quando diretamente aplicada a função de núcleo, atribuiu um peso maior as cores preponderantes presentes nos vídeos, o que não aprimorou a classificação. Esta fato acabou por indicar um aspecto importante, as cores preponderantes nos vídeos, ou seja aquelas cuja tonalidade está presente em mais de 30% da área da cena, não são as responsáveis pela correta distinção entre as classes. Isto porque estas cores, provavelmente, compõe o fundo da cena e não seus objetos principais. Neste caso as cores que melhor diferenciam as classes, segundo os testes indicaram, são as cores com presença moderada, notadamente entre 5 a 15% da área das cenas analisadas. Cabe ressaltar a diferença entre cores moderadas com cores espúrias, que são as com presença inferior a 0,5% da área da imagem e que representam ruídos ou objetos de pouca importância. Isto faz sentido se considerarmos que as cores normalmente preponderantes como o azul celeste ou tons de pele, estão presentes em todos os gêneros de vídeos. Os gráficos da figura 23 contém histogramas médios dos vídeos das cinco categorias estudadas demonstrando a presença destes tons preponderantes nos diversos gêneros.

A implementação da distância de Kullback-Liebler ao núcleo (equação 12) acaba por acrescentar um peso maior as cores preponderantes, em detrimento das intermediárias, por vezes prejudicando a classificação. Uma opção para contornar este efeito seria a implementação de um filtro que atenuasse o peso das tonalidades preponderantes.

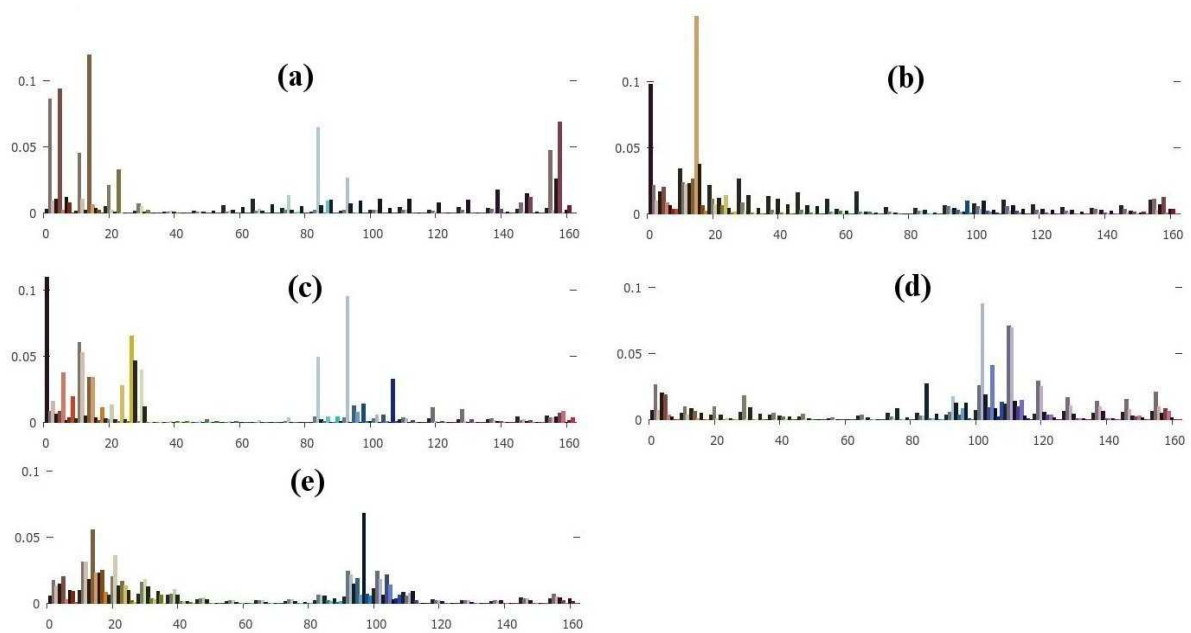


Figura 23- Histogramas médios de vídeos por categoria, (a) esportes (b) filmes (c) infantil (d) jornalismo (e) musicais. Eixo horizontal corresponde a tonalidade (*hue*) eixo vertical porcentagem ocorrência
Fonte: Elaborado pelo autor

Os resultados também demonstraram que nem todos os dados estatísticos referentes ao ritmo visual, contribuíram para uma melhor eficiência da classificação. A análise dos resultados, demonstrados na tabela 3, indicou a semelhança nos valores da moda, mediana e desvio padrão entre todas as classes, o que enfraqueceu sua importância. Somente o valor do tempo médio das cenas apresentou uma variação significativa. Portanto pode-se concluir que o ritmo visual pode contribuir para a classificação, principalmente relação a media das cenas, porém outros parâmetros como a moda, mediana e desvio padrão não possuem relação significativa com o gênero dos vídeos analisados.

De maneira geral este trabalho demonstrou a eficiência dos histogramas e das máquinas de vetores de suporte, e que tanto a distância de Kullback-Liebler como o ritmo visual podem potencialmente contribuir, desde que uma forma mais apropriada seja implementada. Algumas indicações a este respeito serão abordadas na seção a seguir.

7. TRABALHOS FUTUROS

A Classificação Automática de Vídeos é um desafio em contínuo aperfeiçoamento, com muito a ser desenvolvido para atingir a excelência. Como abordado na seção 2, nos aproximadamente vinte anos de pesquisa nesta área, as bases da solução deste problema já foram definidas, mas os resultados comprovam que um aprofundamento maior se faz necessário. Já foi mencionado neste e em outros trabalhos que nenhuma técnica, ou pesquisa, irá sozinha garantir um avanço revolucionário na solução deste problema. A sonhada ponte sobre o fosso semântico exigirá diversos pilares apoiados em diferentes técnicas da inteligência artificial. Reconhecimento de padrões, processamento de sinais e outras áreas da automação, talvez um dia, se bem aplicadas, possam conseguir substituir as ações objetivas e subjetivas intuitivamente realizadas por pessoas ao classificar filmes e programas de televisão.

Dentre as inúmeras possibilidades que a partir deste trabalho podem ser encadeadas, estão a aplicação de outras técnicas que venham caracterizar melhor as diferenças comumente encontradas nos gêneros dos vídeos. A ideia mais imediata é a utilização de transformadas de domínio da frequência como Fourier, ou a transformada discreta do cosseno em relação ao ritmo visual. Desta forma talvez seja possível relacionar melhor a duração das cenas de um vídeo com seu gênero, que como examinado neste trabalho, podem ressaltar as características semânticas destes vídeos. Outro caminho pode estar na divisão da imagem em regiões menores para se aplicar a entropia de Shannon localmente, técnica que sem tal divisão não se mostrou eficiente. A comparação das entropias de quadros distintos de um mesmo vídeo também é uma possibilidade, já que a entropia aplicada de forma temporal aparentemente ainda não foi estudada. A aplicação da distância de Kullback-Liebler especificamente às cores não preponderantes da imagem também pode ser promissora.

São implementações não muito complexas mas que podem ter um resultado significativo, e que portanto merecem ser analisadas.

Outras possibilidades já mencionadas estão na associação de diferentes aspectos do vídeo em especial o seu áudio, ou na identificação de objetos presentes nas cenas, caminhos que vem há anos merecendo atenção, e anualmente são o tema na importante conferência TRECVID.

Nota-se assim que não são poucos os esforços acadêmicos para se estreitar o aparentemente intransponível GAP SEMÂNTICO. E que as dificuldades parecem motivar ainda mais novas pesquisas.

REFERÊNCIAS

ADOBE DIGITAL INDEX, U.S. **Digital video benchmark**, Q4 2013.

AIZERMAN, M. A.; BRAVERMAN, E. M.; ROZONOER, L. I. Theoretical foundations of the potential function method in pattern recognition learning. **Automation and Remote Control**, v. 25, p. 821–837, 1964.

ANDERSON J.; ANDRESON B. The myth of persistence of vision revisited. **Journal of Film and Video**, Champaign: University of Illinois, v. 45, n. 1, 1993.

AYYILDIZ, K.; CONRAD, S. **Video classification by main frequencies of repeating movements**. Dusseldorf : Heinrich Heine University, 2011.

BARLA, A.; ODOE, F.; VERRI, A. Histogram Intersection Kernel For Image Classification. , Proceedings of International Conference of Image Processing, ICIP 2003, Barcelona, v.2 p.513-516, Sep. 2003.

BASTOS A.; FERNADES S., **Televisão Digital**. Rio de Janeiro: Antena Edições Técnicas, 2005.

BEEBE, C. **Bridging The Semantic Gap: Exploring Descriptive Vocabulary For Image Structure**. 2006. Tese (Doutorado em Filosofia) School of Library and Information Science, Indiana University, EUA, 2006. Disponível em:
< http://ils.unc.edu/~wildem/ASIST2007/Beebe_dissertation.pdf > Acesso em 02 maio 2014.

BOSER, B. E.; GUYON, I. M.; VAPNIK, V. N. A training algorithm for optimal margin classifiers. **Proceedings of the fifth annual workshop on Computational learning theory - COLT '92**, Nova York, p. 144, 1992.

BREZEALE, D.; COOK, D. J. Automatic video classification: a survey of the literature. **IEEE, Transactions on systems, man and cybernetics**, part C, p. 416-430, 2008.

CASTLEMAN, K. R. **Digital Image Processing**. Nova York: Prentice Hall, 1996.

CHAPELLE, O.; HAFFNER, P.; VAPNIK, V. N. Support vector machines for histogram-based image classification. **IEEE Transactions on Neural Networks**, v.10, n. 5, 1999.

CHIH-WEI, Hsu; CHIH-JEN, Lin. A Comparison of methods for multi-class support vector machines. **IEEE Transactions on Neural Networks** v. 13, p. 415-425, Mar. 2002.

CORTES, C.; VAPNIK, V. **Support-Vector Networks**. Machine Learning. Kluwer Academic Publishers, v.20, 1995.

CUTTING, J. E.; **Changes in Hollywood film over 75 years**. I-Perception. Londres: Pion , 2011. v. 2, p. 569 – 576.

DE ANGELIS G. **Motion Pictures: making cinema magic**. Oliver, 2003.

DESELAERS, T.; PIMENIDIS, L.; NEY, H. Bag-of-visual-words models for adult image classification and filtering. **International Conference on Pattern Recognition 19th**, p.1-4 ,ICPR 2008,Tampa Dec. 2008. .

DIMITROVA, N.; AGNIHOTRI, L.; WEI, G. Video classification based on HMM using text and faces. **European Signal Processing Conference**, 2000. Disponível em < <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.31.9054> > Acesso em 02 maio. 2014.

DREW, M. S.; AU, J.; Video keyframe production by efficient clustering of compressed chromaticity signatures. In: **Poster session of the ACM international conference on Multimedia 8th**, 2000, p. 365–367. Disponível em < <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.5.1715&rep=rep1&type=pdf> >. Acesso em: 02 maio 2014.

DINH, P. Q.; DORAI, C.; VENKATESH, S. Video genre categorization using audio wavelet coefficients. **Asian Conference on Computer Vision Fifth**, 2002. Disponível em: < <http://www.computing.edu.au/~svetha/cma-current/dinh/accv2002.pdf> >. Acesso em: 02 maio 2014.

ESQUEF, I. A. **Técnicas de entropia em processamento de imagens**. Rio de Janeiro: Centro Brasileiro de Pesquisas Físicas, 2002.

FABLET, R.; BOUTHEMY, P. Spatio-temporal segmentation and general motion characterization for videos indexing and retrieval. **DELOS Workshop on Audio-Visual Libraries, Santorini 10th** , 1999. Disponível em: < <http://public.enst-bretagne.fr/~rfablet/PubliPDF/delos99.pdf> >. Acesso em: 02 maio 2014.

FENG J. et al. **Detecting adult image using multiple features**. Beijing: The Institute of Computing Technology, Chinese Academy of Sciences, 2001.

GALÃO, C. D. et al. **Tsallis entropy and shannon entropy for segmentation of carotida medical image**. São Bernardo do Campo: Centro Universitário da FEI, 2010

GONZALEZ, R. C.; WOODS, R. E. **Processamento de imagens digitais**. São Paulo: Edgard Blücher, 2006.

GUIMARÃES, S. J. F. et al. **Identificação de transições em vídeo digital a partir da análise de ritmo visual**. Belo Horizonte: NPDI/DCC, Universidade Federal de Minas Gerais, 2002.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The elements of statistical learning: data mining, inference, and prediction**. Springer-Verlag, 2009. Disponível em: < <http://statweb.stanford.edu/~tibs/ElemStatLearn/> >. Acesso em: 02 maio 2014.

HAAS, R. B., **Muybridge: man in motion**. Berkeley: University of California, 1976.

HEAD S. W. ; STERLING C. H., **Broadcast in America a survey of eletronic media fifth edition**. Boston: Houghton Mifflin Company, 1987.

HORN, B. K.P.; RHUNCK, B. G. **Determining optical flow**. Cambridge: Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1980.

HU, W. et al. A survey on visual content-based video indexing and retrieval. **IEEE Transaction on systems, man and cybernetics - part C: Applications and reviews**, v. 41, n. 6, nov. 2011.

JUN, Y.; RONG, Y.; HAUPTMANN, A. G.; **Cross-domain video concept detection using adaptive SVMs**. Augsburg Germany: ACM Multimedia, 2007.

JUN, Y.; SHI-JIAO, Z. Narrowing Semantic Gap in Content-based Image Retrieval. **International Conference on Computer Distributed Control and Intelligent Environmental Monitoring, IEEE, CDCIEM 2012**, March 2012.

LATHI, B.P. **Signal processing and linear systems**. EUA: Berkeley-Cambridge, 1988.

LIENHART, R. Comparison of automatic shot boundary detection algorithms. In: **SPIE Conference on Storage and Retrieval for Image and Video Databases VII**, v. 3656, p. 290–301, 1999.

LIU, Z.; WANG, Y.; CHEN, T.; Audio feature extraction and analysis for scene segmentation and classification. **Journal of VLSI signal processing systems**, v. 20, n. 1-2, p. 61–79, 1998.

MIZUTANI, T. et al. **M-Fit**: Edição eletrônica de vídeos com detecção de transições. São Bernardo do Campo: Centro Universitário da FEI, 2008.

PFEIFFER, S.; FISCHER, S.; EFFELSBERG, W. Automatic audio content analysis. In: **Proc. ACM Multimedia**, p. 21–30, 1996.

RABINER, L. A tutorial on hidden markov models and selected applications in speech recognition. **IEEE**, v.77, p.257-286, 1989. Disponível em: <<http://ieeexplore.ieee.org/xpl/abstractAuthors.jsp?tp=&arnumber=18626&url=http%3A%2F%2Fieeexplore.ieee.org%2Fiel5%2F5%2F698%2F00018626>>. Acesso em: 02 maio 2014.

RASHEED, Z.; SHEIHK, Y.; SHAH, M. On the use of computable features for film classification. **IEEE Trans. Circuits Syst. Video Technol.**, v.15, n.1, p. 52-64, Jan 2005.

ROACH, M. et al. Recent trends in video analysis: a taxonomy of video classification problems. Swansea: University of Wales Swansea, 2002.

SHANNON C.E. A mathematical theory of communication. **Bell System Technical Journal**, v. 27, p. 379–423 and 623–656, Jul. and Oct., 1948.

SOBRIECKI, A. et al. **The power of Kullback-Leibler divergence extendable for Tsallis entropy in comparing cone-beam and multi-slice tomography images**. São Bernardo do Campo: Centro Universitário da FEI, 2010.

SMEULDERS, A. et al. Content-based image retrieval at the end of the early years. **IEEE Trans. on Pattern Analysis and Machine Intelligence**, v. 22, n. 12, 2000.

SOARES L.F.G; BARBOSA S. D. J., **Programando em NCL 3.0 desenvolvendo aplicações para TV digital**. Rio de Janeiro: Campus, 2009.

SUTTON, R. S.; BARTO, A. G. **Reinforcement learning: an introduction**. Cambridge: MIT , 1998.

SWAIN, M.; BALLARD, D. **Indexing via color histograms**. Rochester: University of Rochester, 1990.

SWAIN, M.; BALLARD, D. Indexing via color histograms. **Int. J. Comput. Vision**, v. 7, p. 11–32, 1991.

THE MOVING PICTURES EXPERT GROUP WEBSITE. Disponível em: <
http://mpeg.chiariglione.org/home_>. Acesso em: 02 maio 2014.

VAN DER MENSBRUGGHE, G. **Notice sur Joseph Antoine Ferdinand Plateau**, Bruxelles: Académie Royale de Belgique LI, p. 389–486, 1885.

VIJAYAKUMAR, V. ; NEDUNCHEZHIAN, R. A study on video data mining, **International Journal of Multimedia Information Retrieval** , v. 1, p. 153–172 Oct., 2012

WEI, G. ; AGNIHOTRI, L. ; DIMITROVA, N. TV program classification based on face and text processing. **IEEE International Conference on Multimedia and Expo**, v. 3, p. 1345–1348, 2000.

APÊNDICE A - RESULTADOS

RESULTADOS

Na tabela a seguir as duas primeiras colunas indicam os valores adotados da constante de custo C e do fator Gama na classificação para cada tipo de função de núcleo, com e sem o uso dos dados estatísticos do ritmo visual. No caso da função de núcleo linear, que não utiliza o fator Gama, o valor do resultado é repetido para facilitar a comparação.

Const. C	Fator Gama	RBF	RBF	Kullback	Kullback	Linear	Linear
		s/ Est.	c/ Est	s/ Est.	c/ Est	s/ Est.	c/ Est
0,000488281	0,001953125	33,3333	22,2222	33,3333	27,7778	33,3333	27,7778
0,000488281	0,00390625	33,3333	22,2222	44,4444	27,7778	33,3333	27,7778
0,000488281	0,0078125	33,3333	22,2222	44,4444	27,7778	33,3333	27,7778
0,000488281	0,015625	33,3333	33,3333	44,4444	27,7778	33,3333	27,7778
0,000488281	0,03125	33,3333	50	38,8889	27,7778	33,3333	27,7778
0,000488281	0,0625	33,3333	38,8889	38,8889	27,7778	33,3333	27,7778
0,000488281	0,125	33,3333	38,8889	33,3333	27,7778	33,3333	27,7778
0,000488281	0,25	33,3333	27,7778	33,3333	27,7778	33,3333	27,7778
0,000488281	0,5	33,3333	33,3333	33,3333	27,7778	33,3333	27,7778
0,000488281	1	33,3333	33,3333	33,3333	27,7778	33,3333	27,7778
0,000488281	2	33,3333	27,7778	33,3333	33,3333	33,3333	27,7778
0,000488281	4	27,7778	27,7778	33,3333	33,3333	33,3333	27,7778
0,000488281	8	27,7778	27,7778	33,3333	33,3333	33,3333	27,7778
0,000488281	16	27,7778	27,7778	33,3333	33,3333	33,3333	27,7778
0,000976563	0,001953125	33,3333	22,2222	33,3333	27,7778	33,3333	27,7778
0,000976563	0,00390625	33,3333	22,2222	44,4444	27,7778	33,3333	27,7778
0,000976563	0,0078125	33,3333	22,2222	44,4444	27,7778	33,3333	27,7778
0,000976563	0,015625	33,3333	33,3333	44,4444	27,7778	33,3333	27,7778
0,000976563	0,03125	33,3333	50	38,8889	27,7778	33,3333	27,7778
0,000976563	0,0625	33,3333	38,8889	38,8889	27,7778	33,3333	27,7778
0,000976563	0,125	33,3333	38,8889	33,3333	27,7778	33,3333	27,7778
0,000976563	0,25	33,3333	27,7778	33,3333	27,7778	33,3333	27,7778
0,000976563	0,5	33,3333	33,3333	33,3333	27,7778	33,3333	27,7778
0,000976563	1	33,3333	33,3333	33,3333	27,7778	33,3333	27,7778
0,000976563	2	33,3333	27,7778	33,3333	33,3333	33,3333	27,7778
0,000976563	4	27,7778	27,7778	33,3333	33,3333	33,3333	27,7778
0,000976563	8	27,7778	27,7778	33,3333	33,3333	33,3333	27,7778
0,000976563	16	27,7778	27,7778	33,3333	33,3333	33,3333	27,7778
0,001953125	0,001953125	33,3333	22,2222	33,3333	27,7778	33,3333	27,7778
0,001953125	0,00390625	33,3333	22,2222	44,4444	27,7778	33,3333	27,7778
0,001953125	0,0078125	33,3333	22,2222	44,4444	27,7778	33,3333	27,7778
0,001953125	0,015625	33,3333	33,3333	44,4444	27,7778	33,3333	27,7778
0,001953125	0,03125	33,3333	50	38,8889	27,7778	33,3333	27,7778
0,001953125	0,0625	33,3333	38,8889	38,8889	27,7778	33,3333	27,7778
0,001953125	0,125	33,3333	38,8889	33,3333	27,7778	33,3333	27,7778
0,001953125	0,25	33,3333	27,7778	33,3333	27,7778	33,3333	27,7778
0,001953125	0,5	33,3333	33,3333	33,3333	27,7778	33,3333	27,7778
0,001953125	1	33,3333	33,3333	33,3333	27,7778	33,3333	27,7778
0,001953125	2	33,3333	27,7778	33,3333	33,3333	33,3333	27,7778

Const. C	Fator Gama	RBF	RBF	Kullback	Kullback	Linear	Linear
		s/ Est.	c/ Est	s/ Est.	c/ Est	s/ Est.	c/ Est
0,001953125	4	27,7778	27,7778	33,3333	33,3333	33,3333	27,7778
0,001953125	8	27,7778	27,7778	33,3333	33,3333	33,3333	27,7778
0,001953125	16	27,7778	27,7778	33,3333	33,3333	33,3333	27,7778
0,00390625	0,001953125	33,3333	22,2222	33,3333	27,7778	33,3333	27,7778
0,00390625	0,00390625	33,3333	22,2222	44,4444	27,7778	33,3333	27,7778
0,00390625	0,0078125	33,3333	22,2222	44,4444	27,7778	33,3333	27,7778
0,00390625	0,015625	33,3333	33,3333	44,4444	27,7778	33,3333	27,7778
0,00390625	0,03125	33,3333	50	38,8889	27,7778	33,3333	27,7778
0,00390625	0,0625	33,3333	38,8889	38,8889	27,7778	33,3333	27,7778
0,00390625	0,125	33,3333	38,8889	33,3333	27,7778	33,3333	27,7778
0,00390625	0,25	33,3333	27,7778	33,3333	27,7778	33,3333	27,7778
0,00390625	0,5	33,3333	33,3333	33,3333	27,7778	33,3333	27,7778
0,00390625	1	33,3333	33,3333	33,3333	27,7778	33,3333	27,7778
0,00390625	2	33,3333	27,7778	33,3333	33,3333	33,3333	27,7778
0,00390625	4	27,7778	27,7778	33,3333	33,3333	33,3333	27,7778
0,00390625	8	27,7778	27,7778	33,3333	33,3333	33,3333	27,7778
0,00390625	16	27,7778	27,7778	33,3333	33,3333	33,3333	27,7778
0,0078125	0,001953125	33,3333	22,2222	33,3333	27,7778	33,3333	27,7778
0,0078125	0,00390625	33,3333	22,2222	44,4444	27,7778	33,3333	27,7778
0,0078125	0,0078125	33,3333	22,2222	44,4444	27,7778	33,3333	27,7778
0,0078125	0,015625	33,3333	33,3333	44,4444	27,7778	33,3333	27,7778
0,0078125	0,03125	33,3333	50	38,8889	27,7778	33,3333	27,7778
0,0078125	0,0625	33,3333	38,8889	38,8889	27,7778	33,3333	27,7778
0,0078125	0,125	33,3333	38,8889	33,3333	27,7778	33,3333	27,7778
0,0078125	0,25	33,3333	27,7778	33,3333	27,7778	33,3333	27,7778
0,0078125	0,5	33,3333	33,3333	33,3333	27,7778	33,3333	27,7778
0,0078125	1	33,3333	33,3333	33,3333	27,7778	33,3333	27,7778
0,0078125	2	33,3333	27,7778	33,3333	33,3333	33,3333	27,7778
0,0078125	4	27,7778	27,7778	33,3333	33,3333	33,3333	27,7778
0,0078125	8	27,7778	27,7778	33,3333	33,3333	33,3333	27,7778
0,0078125	16	27,7778	27,7778	33,3333	33,3333	33,3333	27,7778
0,015625	0,001953125	33,3333	22,2222	33,3333	27,7778	33,3333	27,7778
0,015625	0,00390625	33,3333	22,2222	44,4444	27,7778	33,3333	27,7778
0,015625	0,0078125	33,3333	22,2222	44,4444	27,7778	33,3333	27,7778
0,015625	0,015625	33,3333	33,3333	44,4444	27,7778	33,3333	27,7778
0,015625	0,03125	33,3333	50	38,8889	27,7778	33,3333	27,7778
0,015625	0,0625	33,3333	38,8889	38,8889	27,7778	33,3333	27,7778
0,015625	0,125	33,3333	38,8889	33,3333	27,7778	33,3333	27,7778
0,015625	0,25	33,3333	27,7778	33,3333	27,7778	33,3333	27,7778
0,015625	0,5	33,3333	33,3333	33,3333	27,7778	33,3333	27,7778
0,015625	1	33,3333	33,3333	33,3333	27,7778	33,3333	27,7778
0,015625	2	33,3333	27,7778	33,3333	33,3333	33,3333	27,7778
0,015625	4	27,7778	27,7778	33,3333	33,3333	33,3333	27,7778
0,015625	8	27,7778	27,7778	33,3333	33,3333	33,3333	27,7778
0,015625	16	27,7778	27,7778	33,3333	33,3333	33,3333	27,7778
0,03125	0,001953125	33,3333	22,2222	33,3333	27,7778	33,3333	44,4444
0,03125	0,00390625	33,3333	22,2222	44,4444	27,7778	33,3333	44,4444
0,03125	0,0078125	33,3333	22,2222	44,4444	27,7778	33,3333	44,4444

Const. C	Fator Gama	RBF	RBF	Kullback	Kullback	Linear	Linear
		s/ Est.	c/ Est	s/ Est.	c/ Est	s/ Est.	c/ Est
0,03125	0,015625	33,3333	33,3333	44,4444	27,7778	33,3333	44,4444
0,03125	0,03125	33,3333	50	38,8889	27,7778	33,3333	44,4444
0,03125	0,0625	33,3333	38,8889	38,8889	27,7778	33,3333	44,4444
0,03125	0,125	33,3333	38,8889	33,3333	27,7778	33,3333	44,4444
0,03125	0,25	33,3333	27,7778	33,3333	27,7778	33,3333	44,4444
0,03125	0,5	33,3333	33,3333	33,3333	27,7778	33,3333	44,4444
0,03125	1	33,3333	33,3333	33,3333	27,7778	33,3333	44,4444
0,03125	2	33,3333	27,7778	33,3333	33,3333	33,3333	44,4444
0,03125	4	27,7778	27,7778	33,3333	33,3333	33,3333	44,4444
0,03125	8	27,7778	27,7778	33,3333	33,3333	33,3333	44,4444
0,03125	16	27,7778	27,7778	33,3333	33,3333	33,3333	44,4444
0,0625	0,001953125	33,3333	22,2222	33,3333	27,7778	33,3333	38,8889
0,0625	0,00390625	33,3333	22,2222	44,4444	27,7778	33,3333	38,8889
0,0625	0,0078125	33,3333	22,2222	44,4444	27,7778	33,3333	38,8889
0,0625	0,015625	33,3333	33,3333	44,4444	27,7778	33,3333	38,8889
0,0625	0,03125	33,3333	50	38,8889	27,7778	33,3333	38,8889
0,0625	0,0625	33,3333	38,8889	38,8889	27,7778	33,3333	38,8889
0,0625	0,125	33,3333	38,8889	33,3333	27,7778	33,3333	38,8889
0,0625	0,25	33,3333	27,7778	33,3333	27,7778	33,3333	38,8889
0,0625	0,5	33,3333	33,3333	33,3333	27,7778	33,3333	38,8889
0,0625	1	33,3333	33,3333	33,3333	27,7778	33,3333	38,8889
0,0625	2	33,3333	27,7778	33,3333	33,3333	33,3333	38,8889
0,0625	4	27,7778	27,7778	33,3333	33,3333	33,3333	38,8889
0,0625	8	27,7778	27,7778	33,3333	33,3333	33,3333	38,8889
0,0625	16	27,7778	27,7778	33,3333	33,3333	33,3333	38,8889
0,125	0,001953125	33,3333	22,2222	33,3333	27,7778	33,3333	33,3333
0,125	0,00390625	33,3333	22,2222	44,4444	27,7778	33,3333	33,3333
0,125	0,0078125	33,3333	22,2222	44,4444	27,7778	33,3333	33,3333
0,125	0,015625	33,3333	33,3333	44,4444	27,7778	33,3333	33,3333
0,125	0,03125	33,3333	50	38,8889	27,7778	33,3333	33,3333
0,125	0,0625	33,3333	38,8889	38,8889	27,7778	33,3333	33,3333
0,125	0,125	33,3333	38,8889	33,3333	27,7778	33,3333	33,3333
0,125	0,25	33,3333	27,7778	33,3333	27,7778	33,3333	33,3333
0,125	0,5	33,3333	33,3333	33,3333	27,7778	33,3333	33,3333
0,125	1	33,3333	33,3333	33,3333	27,7778	33,3333	33,3333
0,125	2	33,3333	27,7778	33,3333	33,3333	33,3333	33,3333
0,125	4	27,7778	27,7778	33,3333	33,3333	33,3333	33,3333
0,125	8	27,7778	27,7778	33,3333	33,3333	33,3333	33,3333
0,125	16	27,7778	27,7778	33,3333	33,3333	33,3333	33,3333
0,25	0,001953125	33,3333	22,2222	33,3333	27,7778	33,3333	38,8889
0,25	0,00390625	33,3333	22,2222	44,4444	27,7778	33,3333	38,8889
0,25	0,0078125	33,3333	22,2222	44,4444	27,7778	33,3333	38,8889
0,25	0,015625	33,3333	33,3333	44,4444	27,7778	33,3333	38,8889
0,25	0,03125	33,3333	50	38,8889	27,7778	33,3333	38,8889
0,25	0,0625	33,3333	38,8889	38,8889	27,7778	33,3333	38,8889
0,25	0,125	33,3333	38,8889	33,3333	27,7778	33,3333	38,8889
0,25	0,25	33,3333	27,7778	33,3333	27,7778	33,3333	38,8889
0,25	0,5	33,3333	33,3333	33,3333	27,7778	33,3333	38,8889

Const. C	Fator Gama	RBF	RBF	Kullback	Kullback	Linear	Linear
		s/ Est.	c/ Est	s/ Est.	c/ Est	s/ Est.	c/ Est
0,25	1	33,3333	33,3333	33,3333	27,7778	33,3333	38,8889
0,25	2	33,3333	27,7778	33,3333	33,3333	33,3333	38,8889
0,25	4	27,7778	27,7778	33,3333	33,3333	33,3333	38,8889
0,25	8	27,7778	27,7778	33,3333	33,3333	33,3333	38,8889
0,25	16	27,7778	27,7778	33,3333	33,3333	33,3333	38,8889
0,5	0,001953125	33,3333	22,2222	33,3333	27,7778	33,3333	33,3333
0,5	0,00390625	33,3333	22,2222	44,4444	27,7778	33,3333	33,3333
0,5	0,0078125	33,3333	22,2222	44,4444	27,7778	33,3333	33,3333
0,5	0,015625	33,3333	33,3333	33,3333	27,7778	33,3333	33,3333
0,5	0,03125	33,3333	44,4444	33,3333	27,7778	33,3333	33,3333
0,5	0,0625	33,3333	38,8889	38,8889	27,7778	33,3333	33,3333
0,5	0,125	33,3333	38,8889	33,3333	27,7778	33,3333	33,3333
0,5	0,25	33,3333	33,3333	33,3333	27,7778	33,3333	33,3333
0,5	0,5	33,3333	33,3333	33,3333	27,7778	33,3333	33,3333
0,5	1	33,3333	33,3333	33,3333	27,7778	33,3333	33,3333
0,5	2	33,3333	27,7778	33,3333	33,3333	33,3333	33,3333
0,5	4	27,7778	27,7778	33,3333	33,3333	33,3333	33,3333
0,5	8	27,7778	27,7778	33,3333	33,3333	33,3333	33,3333
0,5	16	27,7778	27,7778	33,3333	33,3333	33,3333	33,3333
1	0,001953125	33,3333	22,2222	33,3333	27,7778	33,3333	38,8889
1	0,00390625	33,3333	22,2222	38,8889	27,7778	33,3333	38,8889
1	0,0078125	33,3333	16,6667	33,3333	27,7778	33,3333	38,8889
1	0,015625	33,3333	27,7778	27,7778	27,7778	33,3333	38,8889
1	0,03125	33,3333	22,2222	33,3333	27,7778	33,3333	38,8889
1	0,0625	33,3333	33,3333	27,7778	27,7778	33,3333	38,8889
1	0,125	33,3333	38,8889	44,4444	27,7778	33,3333	38,8889
1	0,25	33,3333	44,4444	50	27,7778	33,3333	38,8889
1	0,5	33,3333	50	50	27,7778	33,3333	38,8889
1	1	33,3333	38,8889	38,8889	27,7778	33,3333	38,8889
1	2	33,3333	27,7778	27,7778	33,3333	33,3333	38,8889
1	4	27,7778	27,7778	33,3333	38,8889	33,3333	38,8889
1	8	27,7778	27,7778	27,7778	38,8889	33,3333	38,8889
1	16	27,7778	27,7778	33,3333	50	33,3333	38,8889
2	0,001953125	33,3333	27,7778	33,3333	27,7778	33,3333	33,3333
2	0,00390625	33,3333	27,7778	44,4444	27,7778	33,3333	33,3333
2	0,0078125	33,3333	33,3333	38,8889	27,7778	33,3333	33,3333
2	0,015625	33,3333	33,3333	33,3333	27,7778	33,3333	33,3333
2	0,03125	33,3333	33,3333	27,7778	27,7778	33,3333	33,3333
2	0,0625	33,3333	33,3333	33,3333	27,7778	33,3333	33,3333
2	0,125	33,3333	44,4444	33,3333	33,3333	33,3333	33,3333
2	0,25	33,3333	50	27,7778	33,3333	33,3333	33,3333
2	0,5	33,3333	55,5556	44,4444	33,3333	33,3333	33,3333
2	1	33,3333	44,4444	38,8889	33,3333	33,3333	33,3333
2	2	33,3333	33,3333	38,8889	33,3333	33,3333	33,3333
2	4	27,7778	27,7778	33,3333	38,8889	33,3333	33,3333
2	8	33,3333	27,7778	33,3333	44,4444	33,3333	33,3333
2	16	44,4444	27,7778	33,3333	44,4444	33,3333	33,3333
4	0,001953125	33,3333	27,7778	27,7778	27,7778	33,3333	38,8889

Const. C	Fator Gama	RBF	RBF	Kullback	Kullback	Linear	Linear
		s/ Est.	c/ Est	s/ Est.	c/ Est	s/ Est.	c/ Est
4	0,00390625	33,3333	27,7778	44,4444	27,7778	33,3333	38,8889
4	0,0078125	33,3333	33,3333	33,3333	27,7778	33,3333	38,8889
4	0,015625	33,3333	44,4444	27,7778	27,7778	33,3333	38,8889
4	0,03125	33,3333	38,8889	27,7778	27,7778	33,3333	38,8889
4	0,0625	33,3333	44,4444	38,8889	33,3333	33,3333	38,8889
4	0,125	33,3333	50	27,7778	33,3333	33,3333	38,8889
4	0,25	33,3333	50	27,7778	33,3333	33,3333	38,8889
4	0,5	33,3333	55,5556	44,4444	33,3333	33,3333	38,8889
4	1	33,3333	44,4444	38,8889	33,3333	33,3333	38,8889
4	2	33,3333	33,3333	38,8889	33,3333	33,3333	38,8889
4	4	38,8889	27,7778	33,3333	38,8889	33,3333	38,8889
4	8	50	27,7778	33,3333	44,4444	33,3333	38,8889
4	16	44,4444	27,7778	33,3333	44,4444	33,3333	38,8889
8	0,001953125	33,3333	27,7778	27,7778	27,7778	33,3333	38,8889
8	0,00390625	33,3333	38,8889	38,8889	27,7778	33,3333	38,8889
8	0,0078125	33,3333	44,4444	38,8889	27,7778	33,3333	38,8889
8	0,015625	33,3333	44,4444	38,8889	27,7778	33,3333	38,8889
8	0,03125	33,3333	38,8889	27,7778	33,3333	33,3333	38,8889
8	0,0625	33,3333	50	38,8889	33,3333	33,3333	38,8889
8	0,125	33,3333	50	27,7778	33,3333	33,3333	38,8889
8	0,25	33,3333	50	27,7778	33,3333	33,3333	38,8889
8	0,5	33,3333	55,5556	44,4444	33,3333	33,3333	38,8889
8	1	33,3333	44,4444	38,8889	27,7778	33,3333	38,8889
8	2	44,4444	33,3333	38,8889	33,3333	33,3333	38,8889
8	4	55,5556	27,7778	33,3333	44,4444	33,3333	38,8889
8	8	50	27,7778	33,3333	44,4444	33,3333	38,8889
8	16	44,4444	27,7778	33,3333	44,4444	33,3333	38,8889
16	0,001953125	33,3333	33,3333	38,8889	27,7778	33,3333	38,8889
16	0,00390625	33,3333	50	38,8889	27,7778	33,3333	38,8889
16	0,0078125	33,3333	44,4444	38,8889	27,7778	33,3333	38,8889
16	0,015625	33,3333	38,8889	44,4444	33,3333	33,3333	38,8889
16	0,03125	33,3333	44,4444	27,7778	33,3333	33,3333	38,8889
16	0,0625	33,3333	50	38,8889	33,3333	33,3333	38,8889
16	0,125	33,3333	50	27,7778	27,7778	33,3333	38,8889
16	0,25	33,3333	50	27,7778	33,3333	33,3333	38,8889
16	0,5	33,3333	55,5556	44,4444	50	33,3333	38,8889
16	1	44,4444	44,4444	38,8889	27,7778	33,3333	38,8889
16	2	55,5556	33,3333	38,8889	33,3333	33,3333	38,8889
16	4	55,5556	27,7778	33,3333	38,8889	33,3333	38,8889
16	8	50	27,7778	33,3333	44,4444	33,3333	38,8889
16	16	44,4444	27,7778	33,3333	44,4444	33,3333	38,8889
32	0,001953125	33,3333	44,4444	44,4444	27,7778	44,4444	38,8889
32	0,00390625	33,3333	44,4444	38,8889	27,7778	44,4444	38,8889
32	0,0078125	33,3333	33,3333	33,3333	33,3333	44,4444	38,8889
32	0,015625	33,3333	33,3333	38,8889	38,8889	44,4444	38,8889
32	0,03125	33,3333	44,4444	33,3333	38,8889	44,4444	38,8889
32	0,0625	33,3333	50	38,8889	22,2222	44,4444	38,8889
32	0,125	33,3333	50	27,7778	27,7778	44,4444	38,8889

Const. C	Fator Gama	RBF	RBF	Kullback	Kullback	Linear	Linear
		s/ Est.	c/ Est	s/ Est.	c/ Est	s/ Est.	c/ Est
32	0,25	33,3333	50	27,7778	38,8889	44,4444	38,8889
32	0,5	44,4444	55,5556	44,4444	38,8889	44,4444	38,8889
32	1	55,5556	44,4444	38,8889	27,7778	44,4444	38,8889
32	2	50	33,3333	38,8889	27,7778	44,4444	38,8889
32	4	55,5556	27,7778	33,3333	44,4444	44,4444	38,8889
32	8	50	27,7778	33,3333	44,4444	44,4444	38,8889
32	16	44,4444	27,7778	33,3333	44,4444	44,4444	38,8889
64	0,001953125	33,3333	44,4444	38,8889	27,7778	55,5556	38,8889
64	0,00390625	33,3333	38,8889	33,3333	33,3333	55,5556	38,8889
64	0,0078125	33,3333	44,4444	33,3333	33,3333	55,5556	38,8889
64	0,015625	33,3333	44,4444	33,3333	38,8889	55,5556	38,8889
64	0,03125	33,3333	44,4444	33,3333	27,7778	55,5556	38,8889
64	0,0625	33,3333	50	38,8889	38,8889	55,5556	38,8889
64	0,125	33,3333	50	27,7778	27,7778	55,5556	38,8889
64	0,25	44,4444	50	27,7778	27,7778	55,5556	38,8889
64	0,5	55,5556	55,5556	44,4444	38,8889	55,5556	38,8889
64	1	50	44,4444	38,8889	27,7778	55,5556	38,8889
64	2	55,5556	33,3333	38,8889	27,7778	55,5556	38,8889
64	4	55,5556	27,7778	33,3333	44,4444	55,5556	38,8889
64	8	50	27,7778	33,3333	44,4444	55,5556	38,8889
64	16	44,4444	27,7778	33,3333	44,4444	55,5556	38,8889
128	0,001953125	33,3333	44,4444	33,3333	27,7778	55,5556	38,8889
128	0,00390625	33,3333	38,8889	38,8889	33,3333	55,5556	38,8889
128	0,0078125	33,3333	33,3333	38,8889	38,8889	55,5556	38,8889
128	0,015625	33,3333	44,4444	38,8889	33,3333	55,5556	38,8889
128	0,03125	33,3333	44,4444	33,3333	38,8889	55,5556	38,8889
128	0,0625	33,3333	50	38,8889	16,6667	55,5556	38,8889
128	0,125	44,4444	50	27,7778	27,7778	55,5556	38,8889
128	0,25	55,5556	50	27,7778	33,3333	55,5556	38,8889
128	0,5	55,5556	55,5556	44,4444	38,8889	55,5556	38,8889
128	1	55,5556	44,4444	38,8889	22,2222	55,5556	38,8889
128	2	55,5556	33,3333	38,8889	27,7778	55,5556	38,8889
128	4	55,5556	27,7778	33,3333	44,4444	55,5556	38,8889
128	8	50	27,7778	33,3333	44,4444	55,5556	38,8889
128	16	44,4444	27,7778	33,3333	44,4444	55,5556	38,8889
256	0,001953125	33,3333	44,4444	33,3333	33,3333	55,5556	38,8889
256	0,00390625	33,3333	38,8889	38,8889	38,8889	55,5556	38,8889
256	0,0078125	33,3333	33,3333	33,3333	33,3333	55,5556	38,8889
256	0,015625	33,3333	44,4444	38,8889	27,7778	55,5556	38,8889
256	0,03125	33,3333	44,4444	33,3333	22,2222	55,5556	38,8889
256	0,0625	44,4444	50	38,8889	22,2222	55,5556	38,8889
256	0,125	55,5556	50	27,7778	22,2222	55,5556	38,8889
256	0,25	55,5556	50	27,7778	27,7778	55,5556	38,8889
256	0,5	55,5556	55,5556	44,4444	38,8889	55,5556	38,8889
256	1	55,5556	44,4444	38,8889	27,7778	55,5556	38,8889
256	2	55,5556	33,3333	38,8889	27,7778	55,5556	38,8889
256	4	55,5556	27,7778	33,3333	44,4444	55,5556	38,8889
256	8	50	27,7778	33,3333	44,4444	55,5556	38,8889

Const. C	Fator Gama	RBF	RBF	Kullback	Kullback	Linear	Linear
		s/ Est.	c/ Est	s/ Est.	c/ Est	s/ Est.	c/ Est
256	16	44,4444	27,7778	33,3333	44,4444	55,5556	38,8889
512	0,001953125	33,3333	44,4444	38,8889	38,8889	55,5556	38,8889
512	0,00390625	33,3333	38,8889	38,8889	33,3333	55,5556	38,8889
512	0,0078125	33,3333	33,3333	38,8889	33,3333	55,5556	38,8889
512	0,015625	33,3333	44,4444	38,8889	33,3333	55,5556	38,8889
512	0,03125	44,4444	44,4444	33,3333	38,8889	55,5556	38,8889
512	0,0625	55,5556	50	38,8889	27,7778	55,5556	38,8889
512	0,125	55,5556	50	27,7778	27,7778	55,5556	38,8889
512	0,25	55,5556	50	27,7778	33,3333	55,5556	38,8889
512	0,5	55,5556	55,5556	44,4444	38,8889	55,5556	38,8889
512	1	55,5556	44,4444	38,8889	27,7778	55,5556	38,8889
512	2	55,5556	33,3333	38,8889	27,7778	55,5556	38,8889
512	4	55,5556	27,7778	33,3333	44,4444	55,5556	38,8889
512	8	50	27,7778	33,3333	44,4444	55,5556	38,8889
512	16	44,4444	27,7778	33,3333	44,4444	55,5556	38,8889
1024	0,001953125	33,3333	44,4444	44,4444	27,7778	55,5556	38,8889
1024	0,00390625	33,3333	38,8889	38,8889	22,2222	55,5556	38,8889
1024	0,0078125	33,3333	33,3333	38,8889	33,3333	55,5556	38,8889
1024	0,015625	44,4444	44,4444	38,8889	27,7778	55,5556	38,8889
1024	0,03125	55,5556	44,4444	33,3333	27,7778	55,5556	38,8889
1024	0,0625	55,5556	50	38,8889	22,2222	55,5556	38,8889
1024	0,125	55,5556	50	27,7778	22,2222	55,5556	38,8889
1024	0,25	55,5556	50	27,7778	33,3333	55,5556	38,8889
1024	0,5	55,5556	55,5556	44,4444	38,8889	55,5556	38,8889
1024	1	55,5556	44,4444	38,8889	27,7778	55,5556	38,8889
1024	2	55,5556	33,3333	38,8889	27,7778	55,5556	38,8889
1024	4	55,5556	27,7778	33,3333	44,4444	55,5556	38,8889
1024	8	50	27,7778	33,3333	44,4444	55,5556	38,8889
1024	16	44,4444	27,7778	33,3333	44,4444	55,5556	38,8889
2048	0,001953125	33,3333	44,4444	38,8889	27,7778	55,5556	38,8889
2048	0,00390625	33,3333	38,8889	38,8889	27,7778	55,5556	38,8889
2048	0,0078125	44,4444	33,3333	38,8889	27,7778	55,5556	38,8889
2048	0,015625	55,5556	44,4444	38,8889	33,3333	55,5556	38,8889
2048	0,03125	55,5556	44,4444	33,3333	27,7778	55,5556	38,8889
2048	0,0625	55,5556	50	38,8889	27,7778	55,5556	38,8889
2048	0,125	55,5556	50	27,7778	16,6667	55,5556	38,8889
2048	0,25	55,5556	50	27,7778	33,3333	55,5556	38,8889
2048	0,5	55,5556	55,5556	44,4444	38,8889	55,5556	38,8889
2048	1	55,5556	44,4444	38,8889	27,7778	55,5556	38,8889
2048	2	55,5556	33,3333	38,8889	27,7778	55,5556	38,8889
2048	4	55,5556	27,7778	33,3333	44,4444	55,5556	38,8889
2048	8	50	27,7778	33,3333	44,4444	55,5556	38,8889
2048	16	44,4444	27,7778	33,3333	44,4444	55,5556	38,8889
4096	0,001953125	33,3333	44,4444	38,8889	27,7778	55,5556	38,8889
4096	0,00390625	44,4444	38,8889	38,8889	33,3333	55,5556	38,8889
4096	0,0078125	55,5556	33,3333	38,8889	33,3333	55,5556	38,8889
4096	0,015625	55,5556	44,4444	38,8889	38,8889	55,5556	38,8889
4096	0,03125	55,5556	44,4444	33,3333	27,7778	55,5556	38,8889

Const. C	Fator Gama	RBF	RBF	Kullback	Kullback	Linear	Linear
		s/ Est.	c/ Est	s/ Est.	c/ Est	s/ Est.	c/ Est
4096	0,0625	55,5556	50	38,8889	33,3333	55,5556	38,8889
4096	0,125	55,5556	50	27,7778	16,6667	55,5556	38,8889
4096	0,25	55,5556	50	27,7778	33,3333	55,5556	38,8889
4096	0,5	55,5556	55,5556	44,4444	33,3333	55,5556	38,8889
4096	1	55,5556	44,4444	38,8889	27,7778	55,5556	38,8889
4096	2	55,5556	33,3333	38,8889	27,7778	55,5556	38,8889
4096	4	55,5556	27,7778	33,3333	44,4444	55,5556	38,8889
4096	8	50	27,7778	33,3333	44,4444	55,5556	38,8889
4096	16	44,4444	27,7778	33,3333	44,4444	55,5556	38,8889

APÊNDICE B - CÓDIGOS

A Biblioteca de Visão Computacional OpenCV

OpenCV (www.opencv.org) é um projeto comunitário que desenvolve ferramentas livres para visão computacional e aprendizado de máquina. Com mais de quarenta mil usuários ao redor do mundo e atualmente com mais de 2500 algoritmos implementados com foco em eficiência, o OpenCV é uma base muito atraente para o desenvolvimento de projetos tanto acadêmicos como comerciais. Pela sua qualidade e número elevado de usuários possui uma vasta documentação (<http://docs.opencv.org>) e uma infinidade de exemplos de aplicações disponíveis. Recentemente foram desenvolvidas versões para os mais diversos sistemas operacionais em especial para dispositivos móveis, inclusive telefones celulares, aumentando enormemente sua aplicabilidade e popularidade.

Neste projeto, Opencv é utilizado em rotinas de manipulação de imagens necessárias na preparação dos dados que serão submetidos a classificação.

Código para o Cálculo dos Histogramas HSV

O código em linguagem C++ apresentado a seguir foi desenvolvido com o uso da biblioteca OpenCV para gerar um vetor contendo os valores relativos ao histograma de cores HSV dos quadros-chaves dos vídeos a serem classificados. Seu desenvolvimento levou em consideração a necessidade da geração de dados compatíveis a biblioteca LibSVM, responsável pela subsequente etapa de classificação.

```
/* Programa para gerar histogramas HSV
 * Le , converte imagem BGR para HSV
 * Calcula Histograma e gera saída para libSVM
 * v0.93
 */
#include <opencv2/core/core.hpp>
#include <opencv2/highgui/highgui.hpp>
#include <opencv2/imgproc/imgproc.hpp>
#include <iostream>
#include <stdio.h>

//definicoes
#define H_BINS 18
#define S_BINS 3
#define V_BINS 3
#define H_MAX 180;
```

```

#define V_MAX 255;
#define S_MAX 255;

using namespace std;
using namespace cv;

// declaracao funcoes
void calculaBins(Mat plane, int *hist, int nBins, int hNBins, int sNBins, int vNBins, int
hMax, int vMax, int sMax);
void convertRGB2HSV(Mat* srcRGB, Mat* dst);
int divideIntRoundUp(int vMax, int vNBins);
void ImpValoresFinais(int* hist, int histSize);
void InitArray(int *array, int arrSize);

int main( int argc, char** argv ){
    //variaveis quase globais
    Mat src_image, dst_hsv, dst;
    int hMax = H_MAX;
    int vMax = V_MAX;
    int sMax = S_MAX;
    int hNBins = H_BINS;
    int sNBins = S_BINS;
    int vNBins = V_BINS;
    int histBins = hNBins * sNBins * vNBins;
    int histArray[histBins];

    // Carrega imagem
    src_image = imread(argv[1], 1);
    if (!src_image.data) {
        cout << "Nao foi possivel abrir arquivo " << argv[1] << endl;
        return -1;
    }
    convertRGB2HSV(&src_image, &dst_hsv);

    InitArray(histArray, histBins);
    calculaBins(dst_hsv, histArray, histBins, hNBins, sNBins, vNBins, hMax, vMax,
sMax);
    ImpValoresFinais(histArray, histBins);
    return 0;
}

void calculaBins(Mat plane, int *hist, int nBins, int hNBins,
                int sNBins, int vNBins, int hMax, int vMax, int sMax){
    // matrizes tipo char
    CV_Assert(plane.depth() != sizeof(uchar));
    int constV = divideIntRoundUp(vMax,vNBins);
    int constS = divideIntRoundUp(sMax,sNBins);
    int constH = divideIntRoundUp(hMax,hNBins);
    int constMulVS = vNBins * sNBins;
    int posHist; //posicao no cubo do histograma
    //varre todos os pixels da figura

```

```

MatIterator_<Vec3b> it, end;
for( it = plane.begin<Vec3b>(), end = plane.end<Vec3b>(); it != end; ++it)
{
    posHist = ( (*it)[2]/constV) + (((*it)[1]/constS)* vNBins) + ( ((*it)[0]/constH) *
constMulVS );
    hist[posHist]++;
}
}
void convertRGB2HSV(Mat* srcRGB, Mat* dst){
    cvtColor (*srcRGB, *dst, COLOR_BGR2HSV);
}
//divide e arredonda imediatamente superior
int divideIntRoundUp(int total, int divisor){
    if ( total % divisor ){
        return ((total/divisor) + 1);
    }else{
        return (total/divisor);
    }
}
void ImpValoresFinais(int* hist, int histSize){
    int total=0;
    float minValue;
    for(int i=0; i < histSize; i++){
        //printf(" %d:%d",i+1, hist[i]);
        total += hist[i];
    }
    //printf("\n");
    float totalf = (float)total;
    //printf("\nTotal: %d\n", total);
    for(int i=0; i < histSize; i++){
        minValue = hist[i]/totalf;
// If para eliminar valores despreziveis
//         if (minValue > 0.000001){
//             printf(" %d:%f",i+1, minValue);
//         }
    }
    printf("\n");
}
void separaHSV(Mat hsv, Mat* hPlane, Mat* sPlane, Mat* vPlane){
    vector<Mat> planes;
    split( hsv, planes );
    *hPlane = planes[0];
    *sPlane = planes[1];
    *vPlane = planes[2];
}
void InitArray(int *array, int arrSize){
    for ( int i=0; i < arrSize; i++){
        array[i] = 0;
    }
}

```

A Biblioteca LibSVM

A biblioteca de programas LibSVM foi desenvolvida para proporcionar uma rápida implementação de métodos de classificação e regressão baseados em máquinas de vetores de suporte (SVMs). Dentre suas principais características vale mencionar o suporte a diversas linguagens de programação (C, JAVA, PHYTON, MATLAB entre outras), suporte a classificação multi-classe. Os testes de classificação apresentados neste trabalho foram em sua totalidade realizados com o uso desta biblioteca. Foi alterada a função PRECOMPUTED na implementação da distância de Kullback-Liebler. O código acrescentado e listado a seguir.

```
// Kulback e ^ ( -gamma ( square(x[i] * ln(x[i]/x[j]) * ln(x[i]/x[j]) ) ) )
case PRECOMPUTED:
{
    double sum = 0;
    while(x->index != -1 && y->index !=-1)
    {
        if(x->index == y->index)
        {
            double d = x->value * log( x->value / y->value);
            sum += d*d;
            ++x;
            ++y;
        }
        else
        {
            if(x->index > y->index)
            {
                double d = VALORNAOZERO * log( VALORNAOZERO / y->value);
                sum += d*d;
                ++y;
            }
            else
            {
                double d = x->value * log( x->value / VALORNAOZERO );
                sum += d*d;
                ++x;
            }
        }
    }
    while(x->index != -1)
    {
        double d = x->value * log( x->value / VALORNAOZERO );
        sum += d*d;
        ++x;
    }
    while(y->index != -1)
```

```

    {
        double d = VALORNAOZERO * log( VALORNAOZERO / y->value);
        sum += d*d;
        ++y;
    }
    return exp(-param.gamma*sum);
}

```

A Linguagem AWK

Em virtude da grande quantidade de arquivos analisados durante a fase de coletas de dados, em especial os quadros-chave, a partir dos quais além dos histogramas eram calculados a duração de cada cena, algumas rotinas como a coleta dos tempos de transição, foram gerados com o auxílio da linguagem interpretativa AWK. Estas rotinas determinam valores como a média, mediana, moda, desvio padrão, que juntamente com outros arquivos em lote, todos listados nos apêndices, acrescentavam informações ao vetor que posteriormente eram submetidos a classificação.

Código para o Cálculo das Estatísticas de Duração das Cenas

Na etapa de preparação dos vídeos os quadros chaves são selecionados e armazenados juntamente com sua etiqueta de tempo. Esta etiqueta indica o começo de uma nova cena e a quantos segundos do início do vídeo este quadro é exibido. A partir da etiqueta de tempo de quadros chaves subsequentes é possível com uma simples subtração calcular a duração de cada cena. Neste trabalho a duração das cenas dos primeiros cinco minutos de cada vídeo (300 segundos) são armazenadas e posteriormente utilizados na geração dos dados estatísticos da duração das cenas do vídeo que esta sob análise. Estes dados são acrescentados ao vetor de características que será submetido a classificação.

O código a seguir foi desenvolvido exatamente para realizar esta tarefa. Seu funcionamento exige que os quadros-chave já tenham sido extraídos, e a etiqueta de tempo esteja inserida em seu nome. Existem várias ferramentas de extração de quadros-chave de vídeos, porém todas apresentam em maior ou menor escala algum erro o que torna seu uso

pouco interessante neste trabalho, uma vez que este erro poderia comprometer o resultado da classificação. Sendo assim os quadros-chave foram manualmente extraídos para garantir a exatidão das estatísticas geradas.

```
# O codigo a seguir foi desenvolvido em bash
# e necessita da linguagem de interpretação awk
# para calcular corretamente dados estatisticos
# da duracao das cenas.
# A saida segue o padrao adotado
# pela biblioteca SVM
# com a sintaxe indice:valor
# Por parametro da libSVM
# valores inferiores a 0,001 não são enviados a saida

#Extrair Tempo de exibicao dos quadros chaves
for n in `ls *.jpg | grep s.jpg | cut -f1 -ds`
do
    #echo extraindo $n
    echo $n >> Tframes.txt
done

# Coloca em ordem numerica crescente
cat Tframes.txt | sort -n >> TframesSrt.txt

# Executa a subtracao dos tempos acumulados
b=0
for a in `cat TframesSrt.txt`
do
    #echo $a
    c=`expr $a - $b`
    b=$a
    echo $c >> Tdur.txt
done

#Calcula media, mediana , moda e desvio padrão da duracao das cenas
echo Dados Estatisticos
#media mean = average
awk '{a+=$1} END{print "Media = "a/NR}' Tdur.txt
# echo Desvio Padrao
awk '{sum+=$1; sumsq+=$1*$1}END{print "DesPad = "sqrt(sumsq/NR - (sum/NR)**2)}'
Tdur.txt
#moda
awk '{ a[$1]++;} END { for(i in a) { if (a[i] > freq) {max=i; freq=a[i];}} printf("moda=
%s\n", max) }' Tdur.txt
#echo Mediana
gawk -v max=128 '
function mediana(c,v,j) {
    asort(v,j);
```



```
    if (c % 2) return j[(c+1)/2];
    else return (j[c/2+1]+j[c/2])/2.0;
  }
  { count++;
    values[count]=$1;
    if (count >= max) {
      print mediana(count,values); count=0;
    }
  }
END {
  print "mediana = " mediana(count,values);
}' Tdur.txt
# Apaga arquivos temporarios
rm Tframes.txt
rm TframesSrt.txt
rm Tdur.txt
```

APÊNDICE C – RELAÇÃO DOS VÍDEOS

Relação de Vídeos

No.	Titulo	Fonte
1	Filmes	
10	A Revolutionary New Triumph in Tape	https://archive.org
11	Aluminium On The March	https://archive.org
12	History Of Modern Sidewalls	https://archive.org
13	Los Angelis Fire Dep.	https://archive.org
14	Pan An 6.6	https://archive.org
15	The History Of Newspaper	https://archive.org
16	Chevrolet	https://archive.org
17	1962 American Thrift Part 1	https://archive.org
18	America Energy Crisis 1975	https://archive.org
19	American Engineer Part 1 1956	https://archive.org
	Musicais	
20	Almir Guineto	www.youtube.com
21	Altamiro Carrilho	www.youtube.com
22	Florilegio Musical	www.youtube.com
23	Liu e Leu	www.youtube.com
24	Nalva Aguiar	www.youtube.com
25	Paralamas do Sucesso	www.youtube.com
26	Pout Pourri Viola Minha Viola	www.youtube.com
27	Teo Azevedo	www.youtube.com
28	Gonzaguinha	www.youtube.com
29	Djavan	www.youtube.com
	Infantil	
30	Quatar	www.vimeo.com/creativcommons
31	A Clear View of Public House	www.vimeo.com/creativcommons
32	Atelier Animez Moi Gaspard Qui Pue	www.vimeo.com/creativcommons
33	Atelier Animez Moi Le Bouton Mystérieux	www.vimeo.com/creativcommons

34	Atelier Animez Moi Si Jetais Un Animal Je Serais	www.vimeo.com/creativcommons
35	KOLOBOK-RUSSIAN-TALE	www.vimeo.com/creativcommons
36	The Duck Song Parody Super Funny Duck Jokes	www.vimeo.com/creativcommons
37	Meeting With the Galactic Family Disclosed Ina	www.vimeo.com/creativcommons
	Notícias	
41	Jornal da Cultura Primeira Edição - 17/01/2014	www.youtube.com
42	Jornal da Cultura Primeira Edição - 20/01/2014	www.youtube.com
43	Jornal da Cultura Primeira Edição - 21/01/2014	www.youtube.com
44	Jornal da Cultura Primeira Edição - 22/01/2014	www.youtube.com
45	Jornal da Cultura - Ivo Pugnaroni	www.youtube.com
46	Jornal da Cultura - 05/04/2014	www.youtube.com
47	Jornal da Cultura - 17/02/2014	www.youtube.com
48	Jornal da Cultura - 24/04/2014	www.youtube.com
	Esporte	
50	Kobe Bryant 2010-11	www.vimeo.com/creativcommons
51	Gem Sport 26 Mart 2014	www.vimeo.com/creativcommons
52	Fantastic King Berbatov (Football)	www.vimeo.com/creativcommons
53	Georgia Tech Offensevs X Tech (BasketBall)	www.vimeo.com/creativcommons
54	Lebron James 2010-2011 CourtCut Mix Tape (Basketball)	www.vimeo.com/creativcommons
55	People are Awesone Greatest Sport Moments 2000-2012	www.vimeo.com/creativcommons
56	Staines Lammas X Spelthorne Sports	www.vimeo.com/creativcommons
57	Volei Metisson 21/05	www.vimeo.com/creativcommons
58	FootBall Playoff 05/05	www.vimeo.com/creativcommons